Spatial biology tools to accelerate and refine adeno-associated virus engineering and application

Thesis by

Gerard Michael Coughlin

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
(Defended February 28th, 2025)

Gerard Michael Coughlin
ORCID: 0000-0003-0544-4721

ACKNOWLEDGEMENTS

I'm extremely grateful to my advisor, Dr. Viviana Gradinaru. Thank you for your support throughout my PhD, for the hard work in funding the lab, and for modeling what a good mentor should be. Thank you also for reminding me that not everything needs to be perfect before it's submitted (though I'm not sure that one has fully sunk in yet). Lastly, as someone with broad scientific interests, I am most appreciative of the environment that Viviana has created for her trainees. As GLab members, we are exposed to and encouraged to engage with topics ranging from structural biology of protein-protein interactions to how the structure and function of neural circuits shape animal behaviour. However, this breadth of approach is in service of a singular vision: creating neurotechnologies for understanding and repairing nervous systems through targeted gene delivery. Mirroring this breadth of focus in the lab, my own academic trajectory has taken me through many diverse topics: neural circuits underlying reproduction, understanding and treatment of neurodevelopmental disorders, profiling gene delivery vectors through high throughput spatial transcriptomics. The fact that something came from each of these endeavors is a testament to another of Viviana's strengths: recruiting and inspiring phenomenal colleagues.

Some of these colleagues have served as mentors. Anat Kahan was my rotation mentor, and together we worked on understanding neural circuits underlying fertility in female mice. I worked closely with Elliott Robinson on profiling, then later attempting to rescue, neurological deficits in a mouse model of Neurofibromatosis Type I. Finally, Min Jee Jang and I worked together on developing methods for characterizing engineered gene delivery vectors through high-throughput spatial transcriptomics. Each of these mentors modeled drive and determination, but also balance; I'm grateful for the perspective that their mentorship has brought.

I'm especially thankful to have Máté Borsos as a friend and close collaborator. It's invaluable having someone who shares my scientific interests to bounce ideas back and forth with. Thank you for showing us around Vienna, and hopefully we'll be able to visit again soon.

I've had the privilege of mentoring (formally or informally) numerous scientists throughout my PhD: George Hudson, Yameng Zhang, Rana Eser, Camille Chossis, Nathan Appling, Bre'Anna Barcelona, Karan Mahe, Cameron Jackson, Tyler Brittain, Alex Chung and Santi Bhattarai-Kline. Thank you all for your patience and allowing me to become a better teacher. Rana, Karan,

Cameron, Tyler, and Alex — I'm always impressed to see how you've built your projects. Nathan — I'm grateful to have another lab member that shares my enthusiasm for the art of molecular cloning. Bre'Anna — your enthusiasm and high standards in tackling challenging experiments is inspiring. You just need to brush up on your Connect 4 skills.

The Gradinaru lab is fortunate to have outstanding administrative support. Elisha Mackey has been as invaluable asset, both scientifically and administratively. Thank you for the many primary neuron cultures and for keeping our many animal protocols up to date. Pat Anguiano — thank you for placing all my orders (by my count, at least 1400 total orders). Catherine Oikonomou — thank you for your efforts and expertise in improving our scientific communication.

Lastly, I've been fortunate to cross paths with many excellent scientists that have made my PhD more enjoyable and more interesting. In particular, I'm thankful for the diverse perspectives and expertise of Tim Shay, Acacia Mayfield, Xinhong Chen, Xiaozhe Ding, David Goertson, Miggy Chuapoco, Michael Altermatt, Priya Ravindra Kumar, Rajib Schubert, Sayan Dutta, and Yujie Fan. Thank you as well to Grigorios Oikonomou, Chanpreet Singh, Alon Greenbaum, Jenny Treweek, Collin Challis, Rose Challis, Mengying Zhang, Jimin Park, Cynthia Arokiaraj, Changfan Lin, Seongmin Jang, Jonathan Hoang, Ken Chan, Claire Bedbrook, Nick Flytzanis, Nick Goeden, David Brown, Tatyana Dobreva, Nikhila Swarna, Keith Beadle, Yaping Lei, Zhe Qu, Izabela Giriat, Damien Wolfe, Erin Sullivan, Carrie Saada, Shinae Park, Chelsie Steele, Sirisha Gudavalli, Josette Medicielo, Kristen De La Torre, Bill Kavvathas, and Fiona Ristic.

I'm also thankful to my thesis and defense committee members: Drs. Paul Sternberg, David Prober, Yuki Oka, and Carlos Lois for support throughout my PhD. I'm especially grateful to Drs. Prober and Oka for allowing me to rotate in their labs, further broadening my scientific experience. I'm grateful to previous mentors — Cam Teskey, Andrew Brown, Deborah Kurrasch, and Max Anikovskiy for providing guidance and allowing me to explore a multitude of scientific questions.

Thank you to members of my family who have enabled and encouraged me to pursue a PhD: Jackie MacLaren, Blair Rasmussen, Judy Leyden, Tomo and George. Blair and Judy — thank you for providing a second home throughout undergraduate, and for hosting Thanksgivings and celebrations.

Figure 0.1. George

Finally, I have to thank my partner, Tasha Cammidge, who has supported me throughout my undergraduate and graduate training, despite my aimless wandering. I don't know how you put up with me.

ABSTRACT

The transfer of exogenous genetic material into living cells is a fundamental technique for basic research and, increasingly, for the treatment of human disease. Adeno-associated viruses (AAVs) are small, unenveloped viruses that can carry a limited DNA cargo of 4.4 kb (plus 0.3 kb inverted terminal repeats). These vectors are workhorses for *in vivo* gene transfer into mammalian systems, both for fundamental research and for therapeutic purposes. Natural serotypes of AAVs generally show broad tropism for easy to access tissues. Engineering of AAVs, through modification to the capsid surface and/or to the DNA genome, can enable access to otherwise privileged organs (e.g., brain) and can refine tropism to specific cell types (e.g., Purkinje cells of the cerebellum). Such engineering efforts can generate hundreds to thousands of interesting variants, but there is a dearth of high-throughput methods to characterize these variants. Furthermore, despite widespread usage, including in human patients, many questions on fundamental AAV biology remain unanswered.

In this thesis, I attempt to address some of these outstanding bottlenecks and open questions. In Chapter 2, I address the lack of high-throughput methods for broadly characterizing engineered AAV vectors *in vivo*, by developing and applying high-throughput spatial transcriptomics for AAV transcripts. In Chapter 3, I focus on understanding the biology of AAV genome processing, illuminated by novel spatial genomics methods. Using these novel methods, I then profile and mechanistically dissect transcriptional crosstalk between codelivered AAV vectors (Chapter 4). Finally, in Chapter 5, I address the limited packaging capacity of AAV vectors by leveraging AAV transcriptional crosstalk to enable minimally invasive, all-AAV cell type-specific gene editing in wildtype animals, with enough efficiency to recapitulate known phenotypes.

The work presented in this thesis will help to accelerate and refine AAV engineering and application. Furthermore, this thesis highlights potential confounds for AAV genome engineering, but also opens new avenues for AAV-powered functional genetics in mammalian systems.

PUBLISHED CONTENT AND CONTRIBUTIONS

Coughlin, G.M.[#], Borsos, M.[#], Barcelona, B.H.[$], Appling, N.[$], Mayfield, A.M.H., Mackey, E.D., Eser, R.A., Jackson, C.R., Chen, X., Ravindra Kumar, S., and Gradinaru, V.* (2025). Spatial genomics of AAV vectors reveals mechanism of transcriptional crosstalk that enables targeted delivery of large genetic cargo. Nat. Biotechnol. https://doi.org/10.1038/s41587-025-02565-4

 # equal contribution. G.M.C. designed, conducted, and analyzed experiments, and prepared the manuscript.

Jang, M.J., **Coughlin, G.M.**, Jackson, C.R., Chen, X., Chuapoco, M.R., Vendemiatti, J.L., Wang, A.Z., and Gradinaru, V.* (2023). Spatial transcriptomics for profiling the tropism of viral vectors in tissues. Nat. Biotechnol. *41*, 1272-1286. https://doi.org/10.1038/s41587-022-01648-w

G.M.C designed and conducted experiments, and contributed to the preparation of the manuscript.

Fan, Y., Hoang, J.D., Steele, C., Benchetrit, Y., Dutta, S., **Coughlin, G.M.,** Appling, N., Ding, X., Lin, C., Qu, Z., and Gradinaru, V.* (2025). Genetically encoded self-assembling synuclein for tunable Parkinsonian pathology *in vitro* and *in vivo*. bioRxiv 2025.03.10.641009. https://doi.org/10.1101/2025.03.06.641009.

G.M.C designed and conducted experiments, and contributed to the preparation of the manuscript.

Dutta, S., Pang, M., **Coughlin, G.M.**, Gudavalli, S., Roukes M.L., Chou, T., and Gradinaru, V.* (2025). Molecularly-guided spatial proteomics captures single-cell identity and heterogeneity of the nervous system. bioRxiv 2025.02.10.637505. https://doi.org/10.1101/2025.02.10.637505.

G.M.C designed and conducted experiments, and contributed to the preparation of the manuscript.

Kahan, A.*, **Coughlin, G.M.**, Borsos, M., Swarna, N., Altermatt, M., Brunton, B., and Gradinaru, V.* (2023). VIP neurons of the circadian pacemaker integrate light information to regulate the

mammalian estrous cycle. bioRxiv 2023.01.145240753. https://doi.org/10.1101/2023.01.14.524075.

G.M.C designed and conducted experiments, and contributed to the preparation of the manuscript.

Challis, R.C., Ravindra Kumar, S., Chen, X., Goertsen, D., **Coughlin, G.M.**, Hori, A.M., Chuapoco, M.R., Otis, T.S., Miles, T.F., and Gradinaru, V.* (2022). Adeno-associated virus toolkit to target diverse brain cells. Annu. Rev. Neurosci. *45*: 447-469. https://doi.org/10.1146/annurev-neuro-111020100834

G.M.C. wrote sections of manuscript on AAV-delivered genome editing, and provided feedback and proofreading for the rest of the manuscript.

J. Elliott Robinson, **Gerard M. Coughlin**, Acacia M. Hori, Juonhong Ryan Cho, Elisha D. Mackey, Zeynep Turan, Tommaso Patriarchi, Lin Tian, and Viviana Gradinaru*. Optical dopamine monitoring with dLight1 reveals mesolimbic phenotypes in a mouse model of neurofibromatosis type 1. eLife: e48983 (2019). https://doi.org/10.7554/eLife.48983

G.M.C designed, created and validated the Th-VAST system, conducted experiments, and contributed to the preparation of the manuscript.

Grigorios Oikonomou#, Michael Altermatt#, Rong-wei Zhang, **Gerard M. Coughlin**, Christin Montz, Viviana Gradinaru*, and David A. Prober*. The serotonergic raphe promote sleep in zebrafish and mice. Neuron *103:* 686-701 (2019). https://doi.org/10.1016/j.neuron.2019.05.038

G.M.C performed mouse histology and quantification, and contributed to the preparation of the manuscript.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

AAV	adeno-associated virus
ATAC-seq	assay for transposase-accessible chromatin using sequencing
Atm	ataxia-telangiectasia mutated
BBB	blood-brain barrier
bp	base pairs
BSA	bovine serum albumin
C57BL/6J	C57 black 6 mouse strain
Cacna1a	*calcium voltage-gated channel subunit alpha1 A*
CAG	cytomegalovirus early enhancer, chicken beta-actin promoter and intron, rabbit beta-globin splice acceptor
Cas	clustered regularly interspersed short palindromic repeats (CRISPR)-associated protein
cDNA	complementary deoxyribonucleic acid
ChIP-seq	chromatin immunoprecipitation sequencing
CI	chimeric intron
CLARITY	clear lipid-exchanged acrylamide-hybridized rigid imaging/immunostaining/in situ hybridization-compatible tissue hydrogel
CMV	cytomegalovirus
CMVe	cytomegalovirus immediate-early enhancer
CNS	central nervous system
CRISPR	clustered regularly interspersed short palindromic repeats
DAPI	4′,6-diamidino-2-phenylindole
Dlx	*distal-less homeobox,* often used to refer to AAV enhancers deriving from *Dlx* gene family members
ddPCR	digital droplet polymerase chain reaction
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNA-PK	deoxyribonucleic acid (DNA)-dependent protein kinase
DPBS	Dulbecco's phosphate buffered saline
DREADD	designer receptor exclusively activated by designer drugs
DRG	dorsal root ganglion

dsDNA	double-stranded deoxyribonucleic acid
E. coli	*Escherichia coli*
EEG	electroencephalography
EGFP	enhanced green fluorescent protein
EtOH	ethanol
FBS	fetal bovine serum
FISH	fluorescence in-situ hybridization
GC	guanine and cytosine
GFP	green fluorescent protein
HCR	hybridization chain reaction
HCR-FISH	hybridization chain reaction fluorescence in-situ hybridization
HEK 293T	human embryonic kidney cell line
HGPRT	hypoxanthine-guanine phosphoribosyl transferase
HPRT1	gene that codes for hypoxanthine-guanine phosphoribosyl transferase (HGPRT)
IHC	immunohistochemistry
i.p.	intraperitoneal
Itpr1	*inositol 1,4,5-triphosphate receptor type 1*
ITR	inverted terminal repeat
IV	intravenous
kb	kilobase
kD	kilodalton
mRNA	messenger ribonucleic acid
minBG	minimal beta-globin promoter
miRNA	micro ribonucleic acid
miRNA TS	micro ribonucleic acid target site
MOI	multiplicity of infection
MSN	medium spiny neuron
NA	numerical aperture
Nf1	neurofibromatosis type 1
NHP	non-human primate
nt	nucleotide
ORF	open reading frame
pA	polyadenylation signal sequence
pAAV	plasmid adeno-associated virus,

	transfer plasmid for adeno-associated virus packaging
PACT	passive clear lipid-exchanged acrylamide-hybridized rigid imaging/immunostaining/in situ hybridization -compatible tissue hydrogel (CLARITY) technique
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PC	Purkinje cell
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEI	polyethyleneimine
PFA	paraformaldehyde
PNS	peripheral nervous system
Prkdc	protein-kinase, DNA-activated, catalytic subunit
qPCR	quantitative polymerase chain reaction
RCA	rolling circle amplification
RCAHCR	rolling circle amplification followed by hybridization chain reaction
RIMS	refractive index matching solution
RISC	ribonucleic acid-induced silencing complex
RO	retro-orbital
RNA	ribonucleic acid
S. pneumoniae	*Streptococcus pneumoniae*
SaCas9	*Staphylococcus aureus* Cas9
scAAV	self-complementary adeno-associated virus
scATAC-seq	single cell assay for transposase-accessible chromatin using sequencing
SCID	severe combined immunodeficiency
SCP1	super core promoter 1
SDS	sodium dodecyl sulfate
SL	simian vacuolating virus 40 (SV40) late polyadenylation signal
SpECTr	SpECTr enables AAV concatemer tracking
spGFP1-10	split green fluorescent protein fragment consisting of first 10 domains
SSC	saline sodium citrate buffer
ssDNA	single-stranded deoxyribonucleic acid
Sst	somatostatin

ssAAV	single stranded adeno-associated virus
SV40	simian vacuolating virus 40
SWD	spike-and-wave discharge
TALEN	transcription activator-like effector nuclease
Th	tyrosine hydroxylase
TRE	tetracycline response element
TSS	transcriptional start site
tTA	tetracycline transactivator
UGI	universal gene identifier
USeqFISH	ultrasensitive sequential fluorescence in-situ hybridization
UTR	untranslated region
vg	viral genomes
VP1/2/3	viral protein 1/2/3
VR	variable region
WPRE	woodchuck hepatitis virus posttranscriptional regulatory element
W3	truncation of woodchuck hepatitis virus posttranscriptional regulatory element (WPRE)
W3SL	truncation of woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) plus simian vacuolating virus 40 (SV40) late polyadenylation signal
ZFN	zinc-finger nuclease
Zombie	Zombie is optical measurement of barcodes by *in situ* expression

Chapter 1

BACKGROUND AND INTRODUCTION

Contains work from:

Challis, R.C., Ravindra Kumar, S., Chen, X., Goertsen, D., **Coughlin, G.M.**, Hori, A.M., Chuapoco, M.R., Otis, T.S., Miles, T.F., and Gradinaru, V.* (2022). Adeno-associated virus toolkit to target diverse brain cells. Annu. Rev. Neurosci. *45*: 447-469. https://doi.org/10.1146/annurev-neuro-111020100834

&

J. Elliott Robinson, **Gerard M. Coughlin**, Acacia M. Hori, Juonhong Ryan Cho, Elisha D. Mackey, Zeynep Turan, Tommaso Patriarchi, Lin Tian, and Viviana Gradinaru*. Optical dopamine monitoring with dLight1 reveals mesolimbic phenotypes in a mouse model of neurofibromatosis type 1. eLife: e48983 (2019). https://doi.org/10.7554/eLife.48983

1.1 Gene delivery for research, biotechnology, and therapy

Transfer of endogenous genetic material into a cell can enable organisms to adapt to a changing environment and take on new properties. In 1928, Frederick Griffith demonstrated that exposing non-virulent *Streptococcus pneumoniae* to heat-inactivated virulent *S. pneumoniae* could transform the former into a virulent form[1]. Subsequent work[2–4], especially that by Avery and colleagues[5], showed that the factor important for this transformation was DNA and that in addition to transformation, bacterial horizontal gene transfer can occur through bacterial conjugation[6] or through transduction via a bacteriophage intermediate[7,8].

Just as horizontal gene transfer between bacteria can allow for adaptation to changing conditions, gene therapy may facilitate rescue or relief from human disease. This principle of therapeutic gene transfer was demonstrated by Elizabeth and Waclaw Szybalski[9], who rescued hypoxanthine-guanine phosphoribosyl transferase (HGPRT) deficiency in cultured human cell lines by transforming HGPRT- cells with DNA extracted from HGPRT+ cells. However this process was inefficient, necessitating the development of vectors with which to shuttle the nucleic acid into recipient cells[10].

Development of such gene transfer vectors was spurred by findings that viruses could stably transfer their genetic information into host cells. For example, in 1961, Howard Temin[11] demonstrated heritable gene transfer into chicken embryonic cells through Rous sarcoma virus, and Sambrook and colleagues[12] showed that polyomaviruses, such as SV40, integrate viral DNA into the host genome. Further elucidation of viral lifecycles and gene transfer mechanisms, especially those of retroviruses, accelerated gene transfer methods[10].

Engineering of viral nucleic acid, to confer new properties, was described in 1968 by Rogers and Pfuderer[13], who chemically appended adenines to purified tobacco mosaic virus RNA, then transformed tobacco plants with the modified RNA. Harvested leaves showed a higher abundance of tetra- and penta-lysine polypeptides, though the authors were unable to show any effect on the viral progeny or changes to the viral capsid. Development of recombinant DNA technology enabled more refined engineering of viral nucleic acids, and subsequent work[14–16] demonstrated transfer of entire genes into retroviral vectors, with integration of exogenous sequence into the host genome following infection by the recombinant vector. Such alteration of the viral DNA to contain exogenous sequences and subsequent packaging into viral particles yielded the first viral vectors.

The ability to both manipulate and efficiently deliver exogenous nucleic acids through viral vectors allowed scientists to better demonstrate the principle of therapeutic gene delivery. Using a retrovirus, Moloney murine sarcoma virus, Williams and colleagues[17] successfully integrated exogenous DNA into mouse hematopoietic stem cells *ex vivo*, then transplanted these cells into donor mice, where they were able to engraft. Retroviral gene therapy was further advanced with successful rescue of hypoxanthine phosphoribosyltransferase (HGPRT) deficiency following transduction by a retroviral vector carrying the cDNA for *HPRT1*[18,19].

1.2 Adeno-associated viruses (AAVs) are workhorses for *in vivo* gene delivery

Adeno-associated viruses (AAVs) were initially discovered as contaminants in preparations of adenovirus[20] and named so due to their reliance on adenovirus for replication. As with adenovirus[21], and unlike many well studied model viruses, AAVs do not integrate into host genomes at high frequency[22], though double-stranded breaks can drive increased AAV genome integration[23]. The low rate of AAV genome integration in the absence of double-stranded breaks mitigates, but does not eliminate, adverse events that can arise from insertional mutagenesis. This property, along with their high efficiency, broad tropism (a preference for infecting certain cell or tissue types), and lack of known pathogenicity[24,25] have made AAVs attractive as vectors for therapeutic gene delivery. To distinguish natural AAV from AAV vectors, the term recombinant

AAV (rAAV) is often used. However, for simplicity I will use “AAV” only, but restrict discussion to AAV vectors unless otherwise indicated.

Since their discovery in the 1960s and subsequent vectorization, AAVs have become increasingly popular as gene delivery vehicles, both for basic science and in pre-clinical and clinical usage[26,27]. In neuroscience, AAVs are commonly used for long-term expression of cargo to label, modulate, and/or monitor specific cell types. Commonly used AAV cargo include fluorescent proteins, opsins, designer receptors exclusively activated by designer drugs (DREADDs), genetically encoded calcium or voltage indicators, neurotransmitter and neuropeptide sensors, site-specific recombinases, and programmable endonucleases. AAVs can transduce differentiated and non- or slowly dividing cells, producing long-lasting and stable expression. Spatial and temporal targeting can be accomplished through strategic combination of capsid serotype, regulatory elements, and delivery route[26,27]. These factors are discussed below.

AAVs are also popular as gene delivery vectors for pre-clinical and clinical delivery of therapeutic cargo[24,25,28]. Several FDA- or EMA-approved AAV gene therapies are available, with more in clinical trials. A large proportion of these trials are targeting nervous system disorders[24]. In particular, Zolgensma (Onasemnogene abeparvovec), packaging the coding sequence for *SMN1*, has been shown to provide therapeutic benefit for spinal muscular atrophy, especially when administered prophylactically[29–33]. Zolgensma is FDA- and EMA-approved for the treatment of spinal muscular atrophy and, as of March 2024, has been administered to over 3700 patients.

Despite these successes in AAV-based gene therapies, several challenges remain. Inefficient production, purification, and quality control can impede scaling up and result in high cost[25,34]. Pre-existing immunity, due to neutralizing antibodies, can preclude patients from receiving an AAV-based therapy, whereas a humoral immune response to an AAV-based therapy can prevent subsequent re-dosing[24,25,35]. Lastly, numerous adverse reactions have been observed, including immunogenicity, genotoxicity, hepatotoxicity, expression in non-target organs or cells, and insertional mutagenesis[24]. These adverse events raise important questions about the long-term safety of AAV therapeutics and thus necessitate further understanding of AAV’s transduction pathway and development of methods for precision gene therapies.

1.3 Structure and transduction pathway of AAVs

AAVs consist of a roughly 25 nm icosahedral protein capsid, comprised of 60 subunits, encapsidating a 4.7 kb single-stranded DNA genome of either polarity[24,25,36]. The capsid is made up of 3 separate structural proteins, VP1, VP2, and VP3, in a roughly 1:1:10 ratio, respectively.

These viral proteins are encoded by the *cap* gene from one open reading frame (ORF) with alternative start codons and from alternative splicing. Alternative ORFs in *cap* encode assembly activating protein (AAP) and membrane associated-accessory protein (MAAP), which have functions in capsid assembly[37,38] and viral egress[39,40], respectively. The AAV genome also encodes multiple proteins important for genome replication in the *rep* gene. The entire genome is flanked by 140 nt palindromic inverted terminal repeats (ITRs) that form T-shaped hairpin structures. These ITRs are the only components necessary for packaging. Thus, vectorization of the AAV can be accomplished through removal of all AAV genomic sequence except the ITRs, allowing for packaging of up to 4.4 kb.

Multiple natural AAV serotypes have been described, differing in capsid amino acid sequence as well as genome sequence[24,25,41]. The differences in capsid sequence between serotypes are pronounced in nine so-called variable regions (numbered VR-I through VR-IX), which are important in facilitating interaction with host cell factors and as target domains for neutralizing antibodies[36]. Generally, AAV vectors consist of a capsid sequence from one natural serotype, often chosen to enable genetic access to a target organ or cell type, combined with AAV2 ITRs flanking exogenous sequence. Interactions between the capsid surface and host receptors and co-receptors largely determine the capsid's tropism. However, it is important to note that differences between the capsid amino acid sequence could also affect other steps in the AAV transduction pathway and thereby the tropism[24,25]. Conventionally, AAV vectors are named using the capsid serotype (e.g., AAV9), the ITR serotype (most often AAV2), and elements in the exogenous sequence necessary for understanding the behaviour of the expression cassette (e.g., CAG-EGFP-W3SL), yielding a name in the form AAV9/2.CAG-EGFP-W3SL. Due to the near ubiquity of using AAV2 ITRs, the ITR serotype is often omitted from AAV names.

A model schematic of AAV transduction is shown below (Figure 1.1). The AAV particle interacts with the cell surface through an AAV-host receptor interaction[24,25]. Co-receptors may also play an important role here. The receptor-AAV complex is endocytosed. Subsequent intracellular trafficking and acidification of the AAV-containing endosome may lead to conformational changes in the capsid, enabling endosomal escape. The released AAV can then enter the nucleus through the nuclear pore complex and uncoat, releasing its DNA genome. Note that rather than being imported into the nucleus, AAVs may also be degraded by the proteosome or may even transcytose through the cell. For example, AAV9 has been shown to undergo limited transcytosis through the blood-brain barrier, enabling sparse transduction of brain cell types following systemic administration[42]. In this thesis, I will use the term 'transduction' to refer to this entire pathway, from interaction with the host cell to production of RNA and protein

production. Thus, for delivery of a fluorescent protein, translation and folding of the polypeptide is part of transduction.

Once inside the nucleus, the genome must become double-stranded before transgene expression (Figure 1.1). In most cases, the AAV's DNA is a single stranded AAV (ssAAV) genome, which is comprised of one DNA molecule of either plus- or minus-sense. This ssAAV genome must be converted to a double-stranded form either through second-strand synthesis or by annealing to another ssAAV genome of the opposite polarity[24,25,36]. Identification of AAV genomes comprised of dimeric inverted repeats that can self-anneal yielded self-complementary AAV (scAAV) genomes[36,43]. The inverted dimerization of scAAV genomes is due to mutation in one ITR that prevents nicking of that ITR by the Rep proteins during AAV genome replication[36]. Due to their inverted dimeric structure, these genomes do not require second strand synthesis or annealing to another ssAAV genome before expression. Thus, scAAV genomes can express faster than ssAAV genomes. However, because scAAV genomes contain both the plus and minus strand in an inverted dimer, the packaging capacity is halved relative to a ssAAV genome (about 2.2 kb total).

AAV genomes may persist in the nucleus in various states. Intramolecular recombination between ITRs on the same genome yields circular monomers, whereas recombination between ITRs of separate AAV genomes can yield linear or circular concatemeric episomes[44–49] of varying sizes (dimers, trimers, etc.). Importantly, these states are not fixed. AAV episomes may integrate, recombine, and/or increase in size over time[46,49–51]. Numerous DNA-repair factors have been implicated in AAV genome circularization and concatemerization[52–57], including factors involved in DNA damage detection and those involved in non-homologous end joining. For example, deficiency in *Atm*, which is important for detection of double-stranded breaks, results in reduced expression from circularization-dependent AAVs, without any effect on expression from circularization-independent vectors[55]. Likewise, *Prkdc*$^{scid/scid}$ mice, which have a loss of function in the catalytic subunit of DNA-dependent protein kinase (DNA-PK) show reduced concatemer formation in muscle[53] and liver[54], and reduced expression from concatemerization-dependent AAVs[52]. Given that multiple pathways are involved, it is likely that AAV genome concatemerization depends on cell type and disease-state[58], necessitating methods for understanding AAV genome processing at a single cell level and with cell type information.

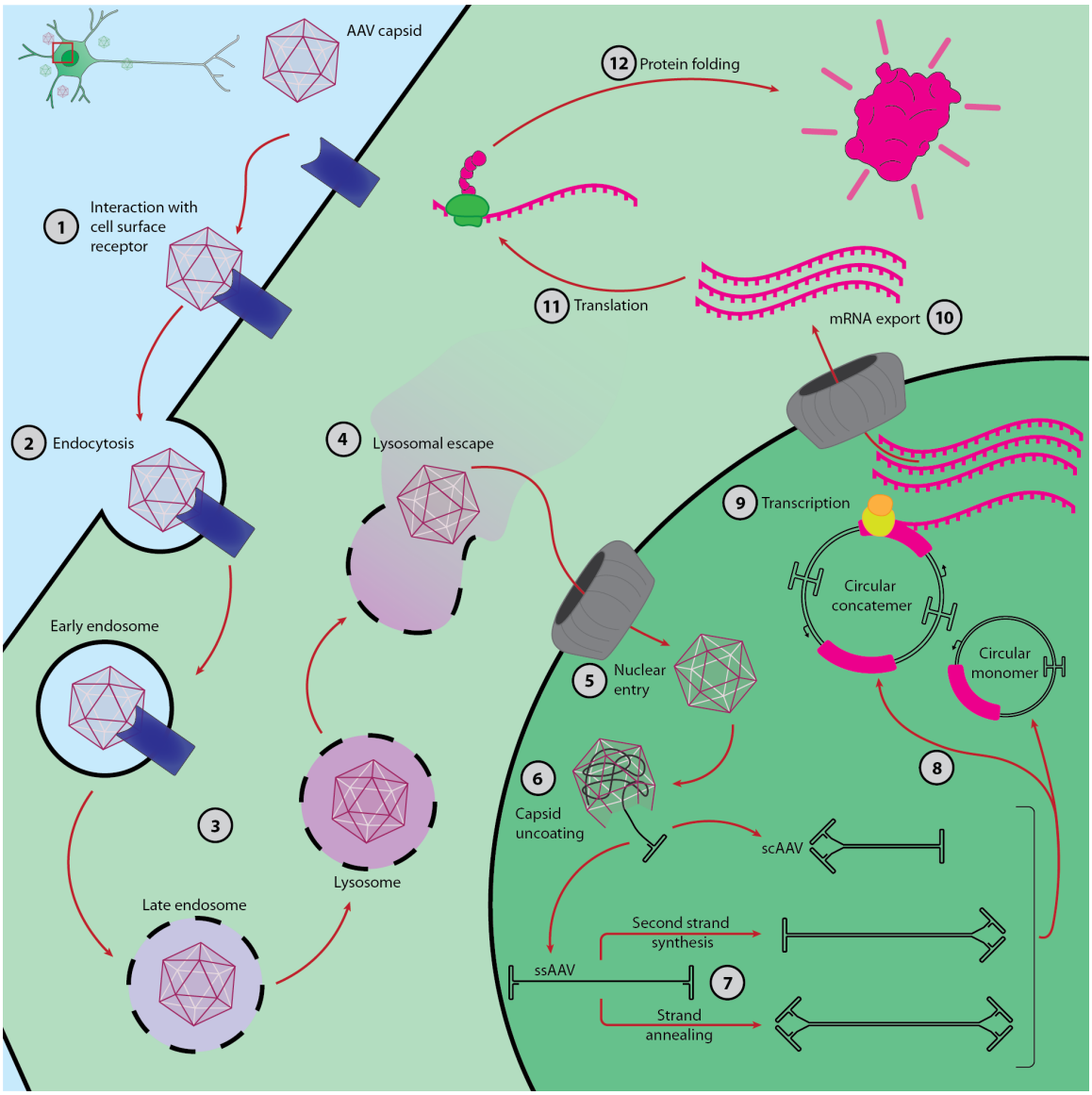


Figure 1.1. Schematic of productive AAV transduction. **(1)** Extracellular AAV capsids interact with cell surface receptors, and **(2)** the receptor-AAV complex is internalized through endocytosis. **(3)** The endosome becomes acidified, and **(4)** at the lysosome stage the AAV can escape from the lysosome. **(5)** AAV capsids enter the nucleus and **(6)** uncoat, releasing their DNA genomes. AAV genomes are most often single-stranded in either + or – orientation (ssAAV), but can be engineered to be double-stranded (scAAV); **(7)** if the genomes are single-stranded, then the genome will need to be made double stranded, either through second strand synthesis or by annealing of + and – strands. **(8)** Double-stranded AAV genomes may then form circular monomeric or concatemeric episomes, through co-opting of host DNA repair machinery. **(9)** The transgene is transcribed, and **(10)** mature mRNA is exported from the nucleus into the cytoplasm, where it can be **(11)** translated and **(12)** the nascent polypeptide can fold into the proper conformation.

There are some important caveats for the above discussion of AAV transduction. Firstly, the precise nature of each of these steps is likely to depend on both the host cell and serotypes of the capsid and ITRs. Furthermore, much of the current knowledge about AAV transduction pathways has been elucidated using AAV serotype 2 (AAV2), in cultured cell lines, and with low resolution readouts, such as Southern blotting or reporter gene expression[59,60]. Finally, many steps in the AAV transduction pathway represent bottlenecks to productive transduction; nuclear import[61], genome uncoating[62], and second-strand synthesis[63,64]. Thus, development of novel methods to monitor and probe AAV transduction may also yield novel insights that improve AAV-based gene therapies.

1.4 Engineering AAV capsids

Engineering of AAV capsids involves changes to the capsid structure, thereby modifying interactions with host cell factors and potentially with the AAV genome. Multiple strategies have been used for AAV capsid engineering. Here I roughly classify these as rational engineering and library-based approaches.

Rational engineering approaches involve making defined changes based on previous information. For example, the capsid AAV2.5 was generated by altering the AAV2 capsid at five separate positions, substituting or adding amino acid residues to better match capsids that transduce muscle more efficiently[65]. The resulting AAV2.5 capsid transduced mouse muscle tissue more efficiently than the parent capsid, AAV2. Furthermore, neutralizing antibodies against AAV2 were less effective at neutralizing AAV2.5[65]. Entire domains may also be grafted onto another serotype capsid to confer new behaviour[66,67]. Identification and engraftment of a galactose-binding domain from AAV9 onto AAV2 generated AAV2G9, which can utilize both the parent capsid's glycan receptor, heparan sulfate, as well as galactose. AAV2G9 provides higher liver transduction efficiency *in vivo*[66]. Rational engineering approaches can be used to modify intracellular processing of AAV vectors. Point mutations of surface-exposed tyrosines on AAV2 can reduce capsid ubiquitination, thereby mitigating vector degradation and improving transduction[68]. This specific strategy seems to translate across capsids, resulting in increased transduction by similar mutants of AAV8 and AAV9[69], and can be applied to other ubiquitinated residues such as lysines or serines to enhance transduction[70].

Library-based approaches are distinguished from rational approaches by being mechanism-agnostic, but requiring a means through which to screen many variants for the property or properties of interest. Ideally, this screening method would impose a strong selective pressure to effectively filter out poor performing variants. Error-prone PCR was used to generate

an AAV2 *cap* library, which were screened for enhanced transduction and the ability to evade antibody neutralization[71]. Diversity can also be generated by shuffling sequences from multiple capsids. This strategy was adapted to create AAV-DJ, which contains sequence from AAV2, AAV8, and AAV9. This engineered capsid outperforms natural serotypes for *in vitro* transduction in multiple cell lines[72]. Insights into AAV structure also inform these DNA shuffling approaches, by suggesting rational break points that minimize disruption to the protein's 3-dimensional structure[73,74]. Ojala and colleagues[73] used this SCHEMA-based[75] approach to create a library of AAV capsids containing sequences from six natural serotypes. Screening of this library in mouse brain, through intra-cerebroventricular injection identified one variant, SCH9, that can efficiently transduce mouse neural stem cells in subventricular zones.

Capsid diversification can also be accomplished with a peptide display library, in which short random amino acid sequences can be inserted or substituted into AAV capsids at defined locations. Such insertions or substitutions are most frequently placed into VR-VIII, though other surface exposed sites have also been explored[76]. Incorporating a Cre recombinase-based selection to impose stringency, Deverman and colleagues[77] identified AAV-PHP.B from an AAV9 VR-VIII peptide insertion library. This engineered AAV can efficiently cross the blood-brain barrier and transduce brain neurons and astrocytes following systemic administration. Re-diversification of the 7 amino acid insertion in AAV-PHP.B, as well as flanking amino acids, yielded AAV-PHP.eB, which further improved upon the brain tropism of AAV-PHP.B and reduced liver transduction as compared to AAV9[78,79]. Goertsen and colleagues[79] further evolved AAV-PHP.eB through diversification of VR-IV by amino acid substitution. The resulting AAV.CAP-B10 shows markedly reduced transduction of liver and astrocytes, with no detected change to neuronal tropism[79]. Additional screening of peptide display libraries has uncovered AAV variants that more effectively transduce the murine peripheral nervous system[78,80], brain endothelial cells[81,82], and muscle[83,84], and lung[85], and those that enhance genetic access to non-human primate nervous systems[79,80,82,86,87].

Rational and library-based approaches to AAV capsid engineering are not exclusive of one another. That is, multiple engineering strategies may be combined. Tervo and colleagues[88] generated AAV capsid libraries through a combination of peptide insertion, random mutagenesis, and domain shuffling. These libraries were screened for efficient retrograde transport by intraparenchymal injection at axon terminals followed by AAV genome recovery at the cell bodies. This pipeline identified the engineered capsid, AAV2-retro, which incorporates both a 10 amino acid VR-VIII insertion and 2 additional point mutations. In addition, peptides identified in a library screen may be grafted onto another AAV capsid to rationally modify that capsid's

behaviour[89]. El Andari and colleagues[90] combined capsid shuffling and rational peptide transfer approaches to generate a muscle-tropic AAV capsid with reduced liver tropism. Interestingly, behaviours conferred by one mutation or motif do not always translate to another capsid. Engraftment of the 7-mer peptide from the VR-IV substitution of AAV.CAP-B10 onto the brain endothelial cell-tropic capsid AAV-X1 did not reduce mouse liver transduction[82], potentially reflecting context-specific behaviour of AAV.CAP-B10's VR-IV substitution, as well as compensation by interaction with liver-expressed receptors[91].

1.5 Engineering AAV genomes

Engineering of the AAV genomes involves consideration and alteration to the DNA sequences packaged in AAV vectors. Sequences that may be altered include promoters, enhancers, coding sequences, introns, untranslated regions (UTRs), transcriptional terminators, flanking sequences, or even the AAV ITRs themselves. Due to relevance for and high activity in AAV engineering and usage, I will restrict discussion to promoters, enhancers, and UTRs. For the latter, I will specifically discuss the use of miRNA target sites (miRNA TS). Furthermore, these components will be discussed from the perspective of targeting specific cell types with AAV vectors. The ability to access specific cell types has implications for basic research, enabling genetic access to targeted cell types without relying on costly transgenics and/or time-consuming classical genetics (Figure 1.2). In gene therapy, specific targeting can help to reduce side-effects by restricting expression of the therapeutic effector to the desired cell type.

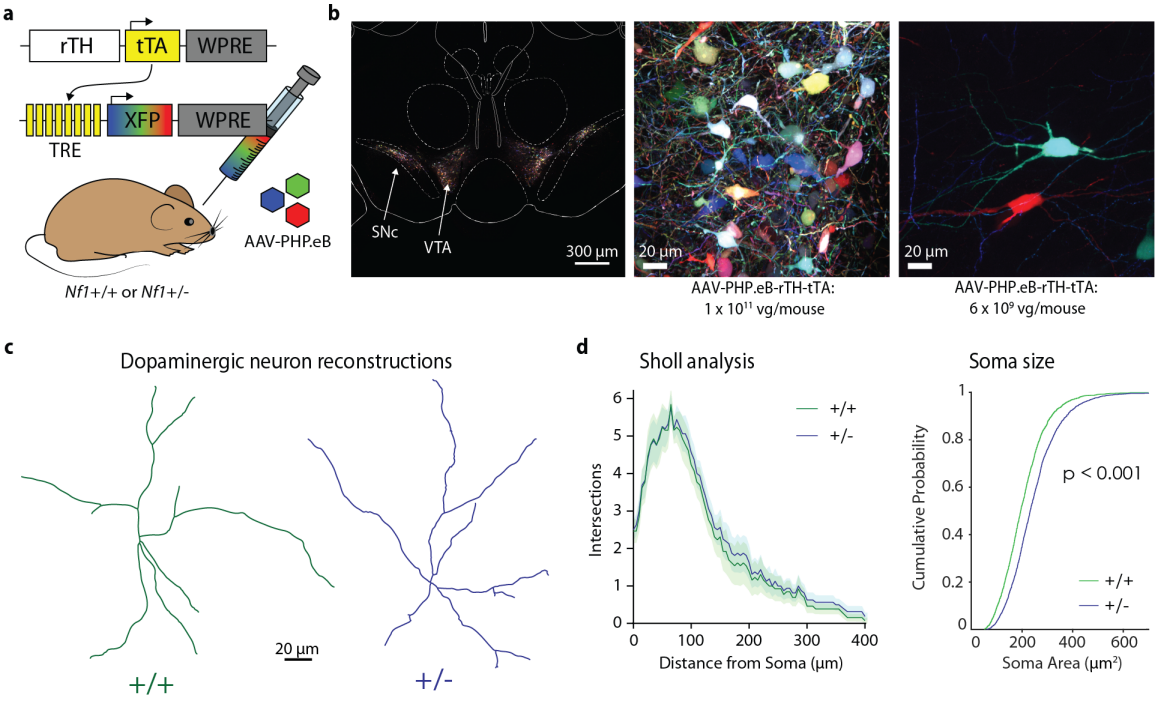


Figure 1.2. Promoter-driven labeling of dopaminergic neurons with systemically administered AAVs, applied to a mouse model of neurofibromatosis type 1. a, Experimental design. A cocktail of 4 viruses is delivered systemically, using the blood-brain barrier-penetrant engineered serotype, AAV-PHP.eB. Three viruses encode spectrally distinct fluorescent proteins, under control of a tetracycline response element (TRE). The other virus expresses a tetracycline transactivator (tTA) under the control of the rat tyrosine hydroxylase (TH) promoter. In cells where this promoter is active, the tTA is expressed, which activates expression of fluorescent proteins. **b**, Due to the stochastic nature of AAV transduction, cells will receive random mixtures of fluorescent proteins, yielding a range of hues that can be used to distinguish neighboring cells. Due to the dependence upon the tetracycline transactivator for fluorescent protein expression, the density of labeling can be controlled by altering the dose of the tTA-encoding AAV, enabling both dense (middle) and sparse labeling (right). **c**, Multispectral labeling allows for reconstruction of the desired cell type's dendritic arbors. **d**, Scholl analysis reveals no significant difference between dendritic complexity of dopaminergic midbrain neurons from *Nf1*+/- and *Nf1*+/+ animals (left). However, we do see a significant difference in soma area.

Promoters are stretches of DNA where RNA polymerase binds and initiates transcription[1]. The strength of promoters varies greatly from gene to gene, despite highly conserved elements within promoter regions, suggesting the importance for context sequence and interaction between elements. Within promoter sequences, the core, or minimal, promoter refers to the minimal sequence required for effective transcription. In the context of AAV engineering and usage, the term promoter is more broadly applied, including to elements that contain multiple functional units. For example, the commonly used CAG 'promoter'[93] contains a promoter from chicken beta-actin, but also includes the cytomegalovirus (CMV) early enhancer[94,95], exonic and

intronic sequence from chicken beta-actin, and a splice acceptor from rabbit beta-globin. In this thesis, I will use 'promoter' in the more strictly defined sense and differentiate elements in hybrid constructs when appropriate. Furthermore, the term "minimal promoter" will be used to refer to promoter sequence that has been truncated from the 5' end to define the minimal sequence necessary for transcriptional activity.

Identification of promoter elements involves mapping the 5' ends of transcripts back to the genome to identify transcriptional start sites (TSSs). Proximal upstream, and sometimes downstream, sequences from TSSs represent putative promoters that can be validated and further dissected. Due to the complexity of this problem, it is unsurprising that much of the fundamental work in promoter characterization was performed on viral promoters. For example, the CMV immediate early promoter, which has activity in eukaryotes, was identified and characterized in the late 1970s and early 1980s[96,97], and a strong *cis*-acting enhancer element was identified soon after[94,95]. Some eukaryotic promoters have also been characterized in depth. For example, the beta-globin promoter has been well characterized and a minimal sequence identified[98,99]. Both the CMV immediate early promoter and the minimal beta-globin promoter have been used extensively in AAV expression vectors. Furthermore, the insights from such fundamental work have facilitated rational design of minimal promoters with ideal combinations and positioning of sequence motifs[100]. This work has yielded engineered promoters that may outperform viral or endogenous promoters, such as the super core promoter 1 (SCP1)[100].

Cell type-specific promoters can also be identified in a similar manner, by identifying the promoter of a cell type-specific marker gene. Using transgenic founder analysis, Liu and colleagues identified a 2.5 kb fragment upstream of the rat tyrosine hydroxylase TSS that could direct expression to mouse midbrain dopaminergic neurons[101]. This size, though restrictive, is compatible with AAV vectors and can similarly direct expression of AAV-delivered transgenes to midbrain dopaminergic neurons[102] (Figure 1.2). Similarly sized, AAV-compatible, mouse brain cell type-targeted promoters have been identified by the Pleiades Promoter Project, enabling targeting of diverse brain cell types, including endothelial cells, astrocytes, oligodendrocytes, dopaminergic neurons, serotonergic neurons, and Purkinje neurons[103–107]. This general strategy of extracting large sequences is not without confounds though. Putative promoter sequences defined in this way may contain other elements, including enhancers.

Enhancers are DNA elements that serve to increase expression from a promoter, often in a cell type-dependent manner[108,109]. Enhancers are *cis*-acting elements, meaning that they interact with promoter elements on the same DNA molecule, and are typically found in intergenic or intronic regions. Furthermore, enhancers act in a position- and orientation-independent manner. In

the genome, enhancers may be separated from the promoters that they act upon by large distances. Chromatin looping can bring enhancers and promoters into proximity. Hundreds of thousands of putative enhancers have been identified in the human genome[108,109]. This diversity, coupled with fact that enhancers can act in cell type, developmental stage, and/or disease state-specific manners, make enhancers attractive as regulatory elements for targeted expression from AAVs.

As with promoters, much of our fundamental understanding of enhancers originates with studies of viral enhancer elements. Banerji and colleagues[110] described a 72 bp DNA element from simian vacuolating virus 40 (SV40) that could activate expression of beta-globin in a position- and orientation-independent manner, but only when placed in *cis* to the beta-globin gene. They termed this sequence the SV40 enhancer. The same research group later identified a strong enhancer in the CMV genome[94]. Subsequent advances in sequencing, and availability of whole genome sequences, facilitated the identification of cell type-specific enhancers in eukaryotic genomes. For example, two enhancers in the intergenic region between divergently transcribed genes *Dlx5* and *Dlx6* were identified based on strong conservation with sequences in the zebrafish *dlx4* and *dlx6* intergenic region[111]. The identified sequences were able to direct expression of a minimal beta-globin promoter-driven reporter gene to specific forebrain populations, recapitulating the expression patterns of the corresponding genes in both species[111,112]. These identified enhancers, as well as related enhancers identified from the intergenic region of *Dlx1* and *Dlx2*, were later transplanted into AAV genomes, enabling genetic access to forebrain interneurons[113–115].

Advances in high-throughput assays for characterizing genome structure have accelerated identification of AAV-compatible enhancers. In cell types where a given enhancer is active, the sequence will be occupied by transcription factors and other regulators, and thus be deficient of nucleosomes[108,109]. This property means that putative enhancers can be identified using open-chromatin analyses, such as DNase I-hypersensitivity site sequencing or assay for transposase-accessible chromatin using sequencing (ATAC-seq)[116], or through targeted methods, such as ChIP-seq[117] to profile histone modifications predictive of active or repressive regulatory elements. In particular, single-cell ATAC-seq (scATAC-seq) can be used to identify regions of open chromatin in dissociated single nuclei[118]; these regions can be mapped back to cluster single cells and classify cell types[119]. This general strategy has been used by multiple groups to identify AAV-compatible enhancers that can boost transgene expression in specific cell types[120–131].

In addition to enhancers, AAV transduction can also be honed by incorporation of miRNA TSs. miRNAs are small RNAs, often deriving from sequence present within introns or

non-coding transcripts[132,133]. In canonical miRNA biogenesis, processing of longer primary miRNAs (pri-miRNAs) by the Microprocessor complex yields a hairpin pre-miRNA that can be exported to the cytoplasm, cleaved by Dicer into a duplex structure, which associates with Argonaute (Ago) protein family members. One strand of the duplex is discarded while the other strand facilitates recognition of RNA sequences by the RNA-induced silencing complex (RISC). Transcripts containing the miRNA TS may be cleaved by the Ago protein, translationally repressed, or degraded through recruitment of additional factors[132,133].

Many miRNAs are expressed in developmental stage-, tissue- and/or cell type-specific manners[134–136]. Thus, inclusion of miRNA TSs into the UTRs of AAV transcripts can dampen expression in select tissues or cells, providing a strategy to hone expression patterns to desired cell type or to reduce toxic overexpression. For example, inclusion of target sites for miR122 into AAV transcript UTRs can effectively reduce liver expression, even when using highly liver-tropic capsids such as AAV9[137,138]. Likewise, miR183 TSs can be used to reduce AAV transgene expression in the dorsal root ganglion (DRG) of mice and non-human primates[139]. Such DRG overexpression is of concern for potential adverse off-target effects of AAV vector administration[140]. As miRNAs function post-transcriptionally, miRNA TSs can be easily combined with cell type specific promoter or enhancers, to further hone expression[141]. Further exploration of using miRNA TSs to refine AAV expression may yield new targeting strategies and means of reducing off-target effects.

There are two important caveats to the above discussion on achieving cell type specificity using the described regulatory elements. First, promoter, enhancer, and miRNA TS sequences pulled from the genome of one species will not necessarily perform the same in other species. Jüttner and colleagues[142] mined mouse transcriptomic and genomic information to design multiple promoters for targeting mouse retinal cell types. Only subset of these promoters (about 1/3) maintained some cell type specificity when tested in non-human primate or human retina samples[142]. Prioritizing sequences that are conserved across species can facilitate translation across species. Secondly, the specificity and efficiency achieved by an AAV is dependent upon many factors, including regulatory sequences in the genome, the chosen capsid, the delivery method, and the dose. Furthermore, the capsid itself may play a role in expression from the AAV genome, potentially by modifying the genome's epigenetic state[143–146].

1.6 Characterizing engineered AAV capsid and genome variants

For both AAV genome and capsid engineering approaches, the number of possible variants may be larger than what can be reasonably screened in depth. Hrvatin and colleagues

identified[120] more than 36000 putative enhancer elements for increased expression in mouse somatostatin+ (Sst+) cortical neurons, filtered these *in silico* and selected the top 287 putative enhancers for coarse screening with single-cell sequencing, and then performed in-depth characterization of only 3 enhancers. Similarly, the miRNATissueAtlas[147,148] currently lists more than 2600 miRNAs in the human genome and more than 1900 in the mouse genome. The space of potential miRNA TSs is much larger though, as miRNAs do not need to be perfectly complementary to their target site to exert a regulatory function[149,150].

Low throughput in-depth screening methods also represent a bottleneck for library-based AAV capsid engineering. As previously mentioned, screening AAV capsid libraries requires strong selective pressure to whittle down a large library to a number of variants that can be reasonably characterized in depth. For a 7 amino acid peptide insertion library the theoretical library size is 1.28e9 total variants. Even stringent selection pipelines applied to such a library may still yield hundreds to thousands of potentially interesting variants, with researchers characterizing a small number (e.g., three to five) in depth. Methods to facilitate higher throughput characterization would take some guesswork out of both AAV capsid and genome engineering pipelines. Increased screening capacity may also allow for identification of capsid or genome variants that allow genetic access to rare cell types.

1.7 Integrating engineered AAVs with gene editing and modulation technologies

Coupled with advances in AAV delivery vectors, tools for editing and manipulating the genome offer exciting new possibilities for understanding gene function and for intervening in gene dysfunction. These techniques are based on natural or engineered DNA or RNA endonucleases, including zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short-palindromic repeat (CRISPR)-associated (Cas) proteins[1]. Whereas targeting with ZFNs and TALENs is based on protein-DNA interaction, Cas proteins are targeted to specific sequence with an RNA guide via Watson-Crick base-pairing. The targeting capabilities of CRISPR-based tools are broad, primarily constrained by the necessary protospacer adjacent motif (PAM) sequence adjacent to the target sequence.

Over the past decade, the number of CRISPR-derived tools has grown substantially, enabling researchers to manipulate the genome, epigenome, and transcriptome of a cell. Such techniques can allow researchers to produce loss-of-function mutations, to alter, add or remove sequence, or to activate or repress transcription[153–159]. If coupled with efficient and targeted AAV-based delivery, the broad functionality of the CRISPR toolbox may provide previously unrealized capabilities in preclinical research.

The large size of many Cas proteins and fusion proteins has limited the integration of CRISPR-based tools with AAV delivery. The well-characterized and widely used *Streptococcus pyogenes* Cas9 is 4.1 kb in length, which precludes easy packaging into an AAV vector with guide RNA cassette, promoter and terminator elements. Smaller Cas proteins have been characterized, including *Staphylococcus aureus* Cas9 (SaCas9, 3.16 kb)[160], *Campylobacter jejuni* Cas9 (2.95 kb)[161], and recently Cas12f (1.26 kb)[162,163] which enable packaging into a single AAV vector, alongside the guide RNA cassette. Further optimization and characterization of these tools (e.g., to reduce PAM constraints) will be instrumental for integration into the AAV toolbox. AAV-mediated delivery of CRISPR-based tools can also be achieved by separating the components onto separate vectors. This can be accomplished by delivering the guide RNA in a separate AAV from the Cas effector, through DNA recombination between AAV genomes, or through mRNA- or protein *trans*-splicing[27]. Finally, refinement of AAV transgene cassettes, through minimalization and optimization of promoters, enhancers, and terminators may be necessary to fully integrate CRISPR-based tools with AAV delivery[164,165].

1.8 Thesis overview

In this thesis, I attempt to address some of the outlined problems and questions in AAV engineering and application (Figure 1.3). I have taken a tool development approach in tackling these problems, first by developing and validating tools and then by applying these tools. This work extends across spatial dimensions, from single molecule characterization of AAV transduction (Chapter 2) and intracellular processing (Chapter 3), to mechanistic understanding of interaction between co-delivered AAV genomes (Chapter 4), and finally using this understanding to manipulate animal behaviour (Chapter 5).

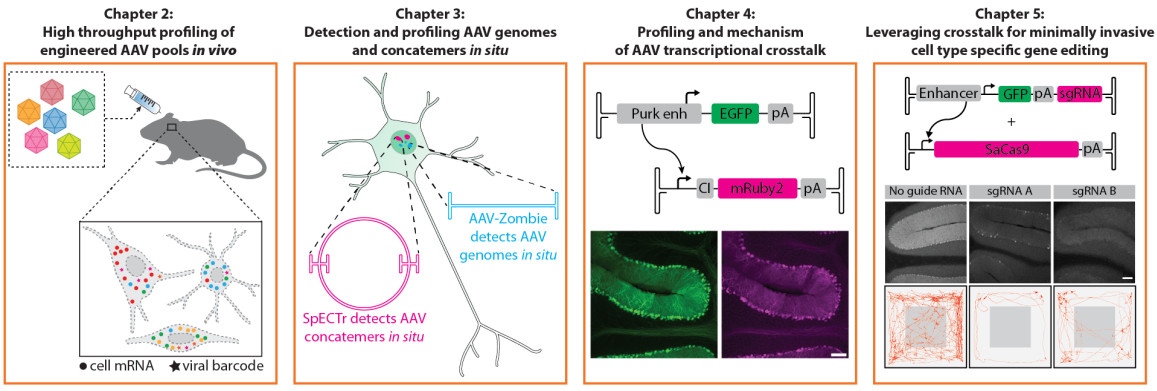


Figure 1.3. Overview of thesis work.

In Chapter 2, I address the lack of high-throughput methods for broadly characterizing engineered AAV vectors *in vivo*. Working with a colleague, I developed and applied a high-throughput, high-resolution ultrasensitive sequential fluorescence *in situ* hybridization (USeqFISH) technique for profiling the transduction of engineered AAV vectors. Our methods are scalable, allowing for tropism profiling of multiple capsid or genome variants in the same animal, and with high transcriptomic depth for nuanced cell typing. We applied this method to pools of AAV capsid or genome variants, profiling transduction in as many as 26 defined cell types and across multiple brain regions. Further development of these methods may enable screening of larger pools or libraries *in vivo*.

In addition to increased capacity to characterize engineered AAVs, we need to better understand AAV biology to successfully apply these vectors to therapeutic purposes. In Chapter 3, I describe validated methods for detecting and tracking the AAV's DNA genome inside cells, with protocol modifications that can be used to specifically detect and quantify concatemeric episomes, enigmatic structures consisting of multiple AAV genomes in one molecule. These methods work both *in vitro* and *in vivo* and are compatible with readout of AAV capsid localization and expression, enabling multiparameter characterization of AAV transduction, from capsid to concatemer formation.

The formation of AAV concatemers can have unexpected consequences for expression from co-injected AAV vectors. In Chapter 4, I demonstrate how concatemerization of AAV genomes can lead to interaction between an enhancer on one AAV genome with a promoter on another, leading to unexpected expression from the latter. I identified and profiled this transcriptional crosstalk occurring across multiple cell type-specific enhancers and in multiple central and peripheral tissues. Furthermore, I identified the necessary components for this behaviour and, using our novel spatial genomics methods, mechanistically linked concatemer formation to transcriptional crosstalk. These findings highlight important confounds for pooled characterization of novel enhancers in AAV genomes, which is becoming a common workflow thanks to the increasing abundance of transcriptomic and epigenomic datasets.

Though transcriptional crosstalk presents a confound for pooled enhancer screening, it also represents an opportunity to address the limited cargo capacity of AAV vectors. I explore this in Chapter 5. By splitting enhancers and short minimal promoters onto separate genomes, we can achieve broad and cell type-specific expression of even large cargo following minimally invasive delivery. In wildtype animals and using Cas9 as a large cargo, I demonstrated that we could restrict efficient editing to the desired cell type and with sufficient coverage of the targeted population to recapitulate known loss-of-function behavioural phenotypes. Notably, this approach

to cell type-specific gene function interrogation is faster than conventional mouse genetics; we can generate a cohort of animals with cell type-targeted disruption of a specific gene within 2 weeks (vs. months to years with conventional approaches). We envision that such techniques can be used to perform cell type-specific reverse genetics screens in mammalian systems. Furthermore, this crosstalk-enabled approach may allow for cell type-targeted genome modulation in therapeutic contexts, which may mitigate side effects due to editing in off-target cells.

Chapter 2

SPATIAL TRANSCRIPTOMICS FOR PROFILING THE TROPISM OF VIRAL VECTORS IN TISSUE

Adapted from:

Jang, M.J., **Coughlin, G.M.**, Jackson, C.R., Chen, X., Chuapoco, M.R., Vendemiatti, J.L., Wang, A.Z., and Gradinaru, V.* (2023). Spatial transcriptomics for profiling the tropism of viral vectors in tissues. Nat. Biotechnol. *41*, 1272-1286. https://doi.org/10.1038/s41587-022-01648-w

2.1 Summary

Engineering of AAV capsids and genomes can redirect and refine transduction, enabling genetic access to tissues and cell types of interest. Engineering strategies, such as directed evolution of AAV capsids or mining regulatory elements from the host genome, can yield tens to hundreds of potentially interesting variants. Typically, these variants are characterized one at a time, in one animal per variant and using immunohistochemistry to label major cell classes. Dissociation-based single-cell RNA sequencing can be used to profile multiple variants in a single animal with broader and deeper cell typing[166]. However, these methods lack spatial resolution and can be challenging to adapt for different tissues, due to the need for tissue-specific cell dissociation protocols. We sought to address this bottleneck using spatial transcriptomics, which allow for detection of hundreds to thousands of unique RNA transcripts with spatial resolution and can be readily adapted for different tissue types. We developed ultrasensitive sequential fluorescence *in situ* hybridization (USeqFISH), enabling detection of short barcode sequences placed into transcribed regions of the AAV genome. Combined with tissue clearing, USeqFISH can detect endogenous and viral transcripts in intact tissue volumes. Using USeqFISH, we profiled the tropism of six engineered capsids, in multiple regions of the mouse brain and with high transcriptomic depth for nuanced cell typing. Our capsid pool contained one previously uncharacterized variant, PHP.AX, which shows relatively unbiased tropism of cortical cell types. We further demonstrated the capacity of USeqFISH for high-throughput AAV characterization by profiling a pool of thirteen AAV genome variants, containing different miRNA target sites for refinement of expression. We envision that further development of USeqFISH could enable screening of even larger pools or libraries *in vivo*.

2.2 USeqFISH for in situ profiling of endogenous and viral gene expression

To enable high-throughput profiling of AAV transduction, we developed ultrasensitive sequential fluorescence in situ hybridization (USeqFISH; Fig. 2.1a, "Signal amplification with RCAHCR"). This technique combines signal amplification by rolling circle amplification[1] (RCA) and hybridization chain reaction[169–172] (HCR), to produce strong signal, even from low abundance transcripts and using a small number of probes. First the tissue is embedded in an RNA-retaining hydrogel, then lipids are removed. Primer and padlock probes are hybridized to the target transcript; the necessity of for having both primer and padlock probes hybridized to the target reduces noise from non-specific binding. Ligation of the padlock probe enables rolling circle amplification, creating multiple copies of a 19-nt unique gene identifier (UGI) sequence. We then hybridize initiator probes to the RCA amplicon, and trigger HCR through addition of fluorophore-conjugated hairpins.

As RCAHCR does not require any signal deposition through formation of covalent bonds, the signal can be removed by disassembling HCR amplicons (Figure 2.1a, "Two-step stripping for sequential labeling"). This is accomplished by incorporating a toehold sequence onto the HCR hairpins; addition of a displacement oligo causes disassembly of the HCR amplicon. Initiator probes can then by removed chemically, allowing for further rounds of initiator hybridization and HCR. Using USeqFISH, we were able to detect multiple endogenous transcripts in mouse cortex (Figure 2.1b). To enable detection of AAV transcripts, we incorporated a barcoding strategy, adding 4 probe binding sites into the transcribed regions of the AAV genomes (Figure 2.1c). We first assessed whether placing barcodes before or after the WPRE[173,174] would affect detection of barcoded transcripts by USeqFISH, and found that both positions enabled efficient detection (Supplementary Figure 2.1a). We then tested whether our barcode design enabled specific detection. Indeed, barcodes were detected only when using complementary probes (Supplementary Figure 2.1b). We then packaged barcoded AAV genomes into AAV-PHP.eB[175] and systemically administered these capsids into wildtype mice. Using USeqFISH, we were able to detect barcoded transcripts in multiple brain regions, noting overlap between barcode spots and expression of the fluorescent reporter (Figure 2.1d).

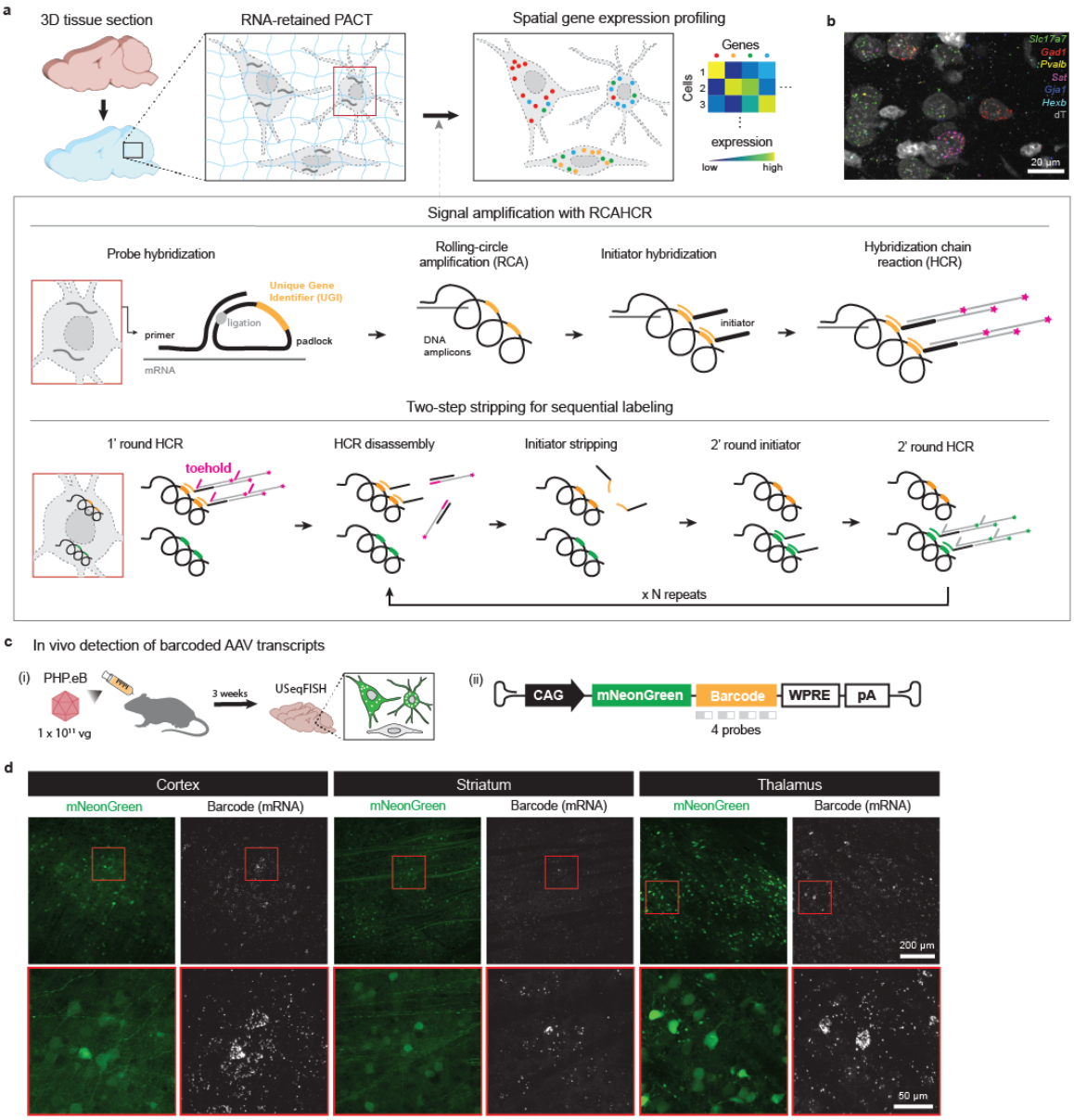


Figure 2.1. USeqFISH for multiplex and sensitive gene expression profiling in 3D tissue. a, USeqFISH procedure. Tissue sections are embedded and cleared using hydrogel chemistry optimized for RNA retention. Primer and padlock probes are hybridized to target transcripts in tissue, and the padlock probe is ligated into a circular structure. Rolling circle amplification (RCA) is then carried out, using the 3' OH on the primer probe. Initiator probes are hybridized to this nascent DNA amplicon and are used to initiate hybridization chain reaction (HCR) with fluorophore-conjugated hairpin probes. For sequential labeling, a two-step stripping process can be used. Toehold sequences appended to HCR hairpins enable HCR amplicon disassembly by introduction of a displacement oligo. The initiator probe can then be removed with formamide. This enables detection of a new RCA amplicon using a different initiator probe. **b**, USeqFISH detection of 6 endogenous transcripts in mouse cortex. The cytoplasm is labeled with an Alexa Fluor 647-conjugated polyT probe (dT). **c**, Detection of barcoded AAV transcripts in mouse brain. AAV-PHP.eB was used to deliver AAV genome, containing barcode between mNeonGreen coding

sequence and WPRE. **d**, Representative images showing detection of barcoded transcript in cells expressing mNeonGreen in cortex, striatum and thalamus.

2.3 Assessment of AAV transcript detection dosage sensitivity

To facilitate *in situ* profiling of AAV pools, we further optimized the viral cargo by incorporating a non-fluorescent, but antigenically detectable coding sequence (spGFP1-10), and a shortened 3' untranslated region and terminator sequence (W3SL[165]). As the total dose for systemically administered AAVs is limited, each variant in the pool must be delivered at a lower dose. Thus, we assessed the minimum dose of a well-performing variant (AAV-PHP.eB) that could be readily profiled with USeqFISH. We constructed a barcoded pool with different doses (10^{11}, 10^{10}, 10^{9}, 10^{8}, 10^{7} vg per animal) represented by different barcodes, and administered this pool to wildtype mice (Figure 2.2a,b). USeqFISH profiling of the tissue revealed a strong dose-dependence on transduction rate and spot number per cell (Figure 2.2c-e). Though we were able to detect transduced cells at the 10^{7} vg dose, we conclude that a minimal dose of 10^{9} to 10^{10} vg is optimal to provide reliable profiling of cells and enough dynamic range to profile even poorer performing variants. At this dose, and assuming a conservative upper limit of 1e12 total vg per animal, USeqFISH could enable profiling of 100 to 1000 variants co-injected into the same animal.

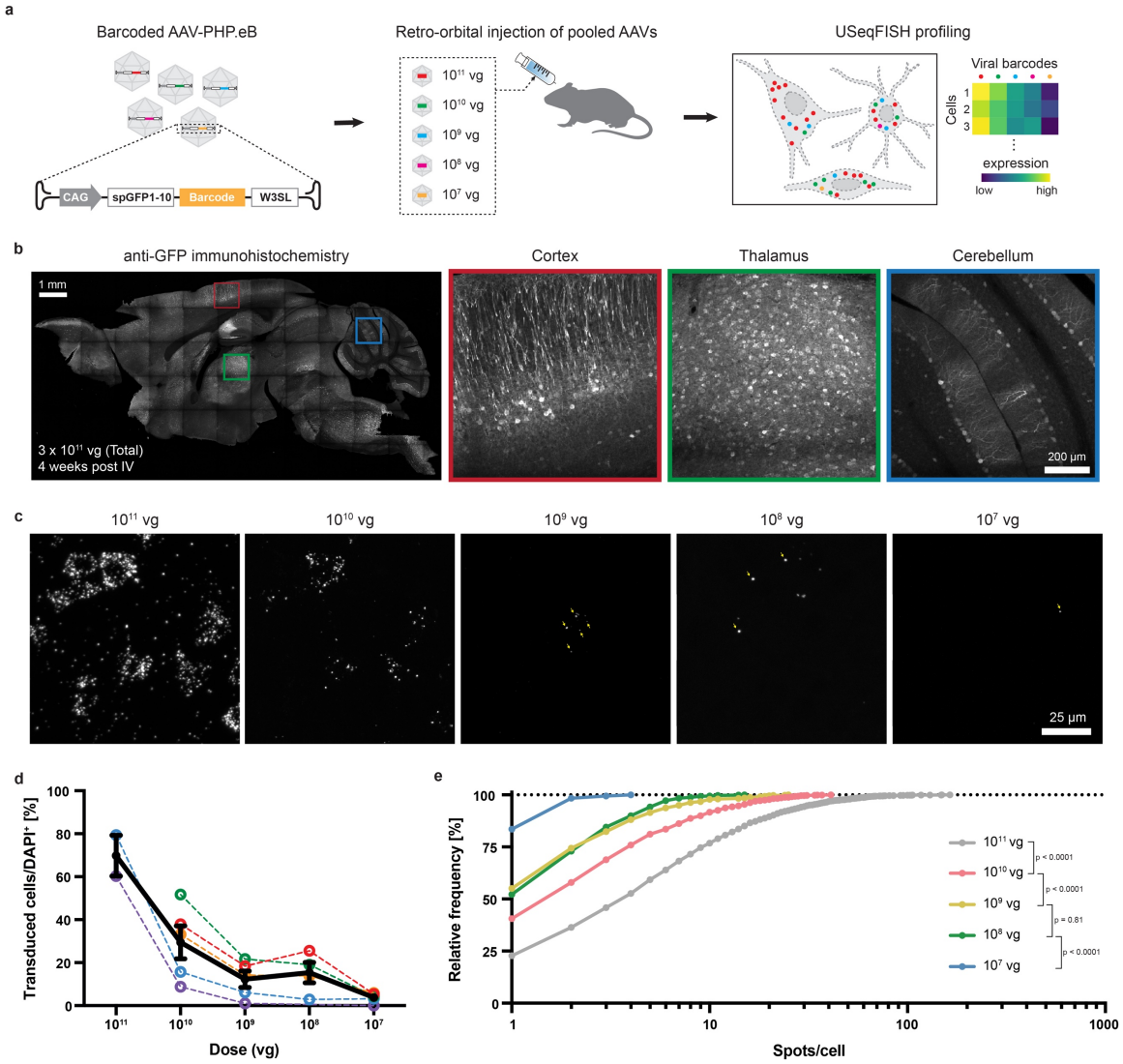


Figure 2.2. Assessment of AAV transcript detection dosage sensitivity. Barcoded AAV genomes were cloned and separately packaged into AAV-PHP.eB. To maximize fluorescence channels for readout of barcoded transcripts, we used a non-fluorescent, but antigenically detectable coding sequence, spGFP1-10. Packaged AAVs were then diluted into one pool with different doses (covering 10^7 to 10^{11} vg per mouse) represented by different barcodes, and administered as a single RO injection. Four weeks later, the tissue was collected and processed. **b**, Immunohistochemistry against spGFP1-10 shows widespread viral transduction in brain. **c**, Detection of barcoded transcripts delivered at different doses. Representative images are shown; at lower doses, yellow arrows indicate transcript spots **d**, Quantification of transduction efficiency at different doses, from 5 separate mice (indicated by coloured lines). Black line represents mean ± s.e.m. Cells with 1 or more detected viral transcripts were considered as transduced. **e**, Cumulative distribution of viral spot number in transduced cells at different doses. Significance was determined by Kolmogorov-Smirnov test.

2.4 High-throughput, high-resolution profiling of AAV capsid pools in mouse brain with USeqFISH

We next tested the ability of USeqFISH to profile a pool of capsid variants *in situ*. We constructed a pool containing previously characterized capsids (AAV-PHP.eB[175], AAV.CAP-B10[79], AAV-PHP.N[81], AAV-PHP.V1[81], AAV-PHP.B8[81]), as well as one uncharacterized variant, AAV-PHP.AX, which contains a 7 amino acid substitution[176] into AA452-458 of AAV-PHP.eB. Each variant was separately packaged with a unique barcode, titered, pooled together at equal concentrations, then administered to wildtype animals at a dose of 5×10^{10} vg per capsid (Figure 2.3a,b). The tissue was collected 4 weeks post-injection and the cortex was profiled with USeqFISH, using probes to detect both AAV transcript barcodes as well as endogenous transcripts for cell typing[177–182] (Figure 2.3b). We first determined the overall transduction by each variant (Figure 2.3b), noting that trends in transduction efficiency matched our expectations and that the previously uncharacterized AAV-PHP.AX showed slightly lower total transduction than the strongly performing capsid variants in the pool, AAV-PHP.eB and AAV.CAP-B10. We then generated cell type clusters based on analysis of endogenous gene expression, and analyzed the expression of each barcode within cell type clusters (Figure 2.3d). We defined two relevant metrics of transduction: enrichment and relative tropism bias. Enrichment is the mean of log-transformed spot numbers. Relative tropism bias is the z-scored spot counts for a single variant normalized to the sum of all barcode spot counts for a given cell. Thus, relative tropism bias accounts for differential transduction of cell types by all variants, highlighting differences between variants in targeting rarely transduced cell types.

To validate tropism profiling by USeqFISH, we compared our tropism profiles to those obtained through immunohistochemistry[79,81,175] or single-cell RNA sequencing[166]. We noted that variants with strong cell type or regional tropisms (AAV-PHP.N for vascular cells and AAV-PHP.B8 for thalamus and cerebellum), show low overall transduction in cortex (Figure 2.3c). Likewise, AAV-PHP.eB showed higher transduction of astrocytes than AAV.CAP-B10, AAV-PHP.N showed a neuronal bias, and AAV-PHP.V1 had a bias towards transduction of vascular cells, all consistent with previous characterization[79,81]. Importantly, we observed consistent cell typing and transduction profiles between the two mice (Supplementary Figure 2.2a-c), demonstrating the reproducibility of USeqFISH for AAV tropism profiling.

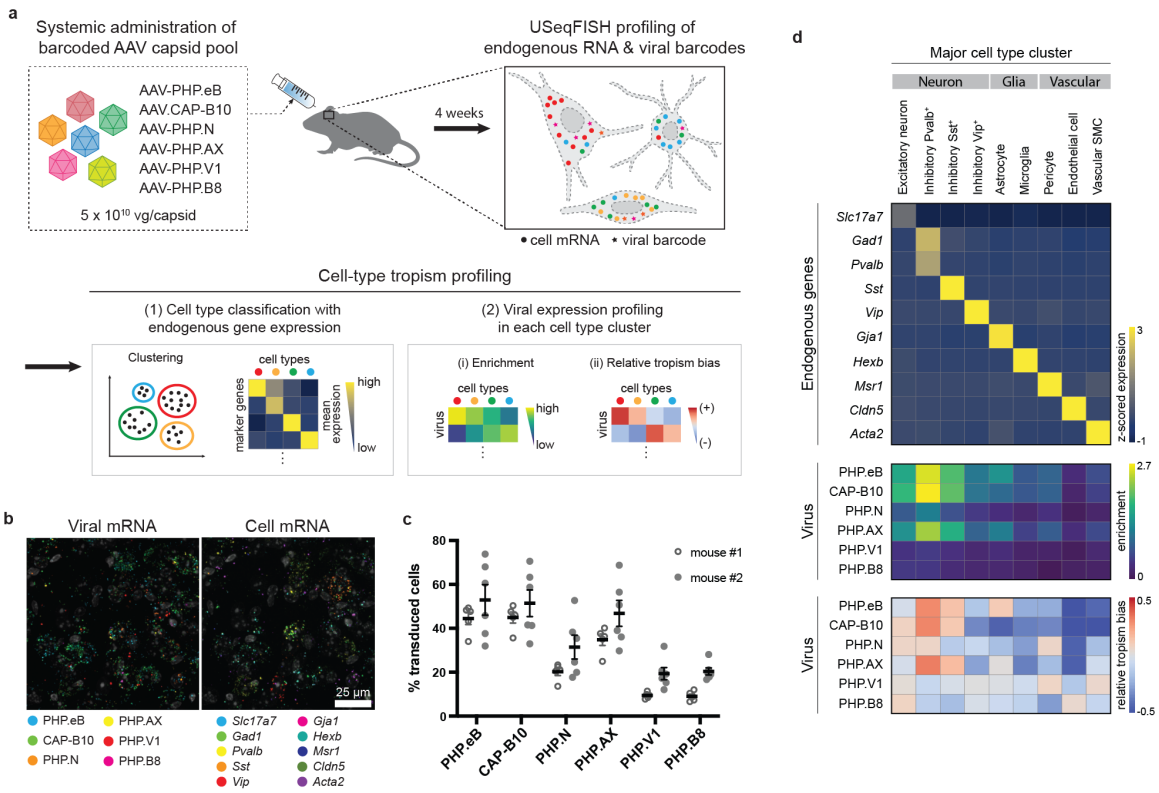


Figure 2.3. Profiling of barcoded capsid variants in mouse cortex. **a**, Experimental protocol. Molecular barcodes were separately packaged into six different capsid variants, pooled at equal titers (5e10 vg per capsid), then administered into adult wildtype mice through retro-orbital injection. 4 weeks after administration, brain tissue was collected and assayed using USeqFISH to profile viral transduction in multiple cell types. Cell types were defined through automated clustering, and the viral transduction of each cell type was profiled. **b**, Representative images of detected viral barcodes (left) and marker gene transcripts (right). Both images show the same field of view. **c**, Quantification of transduction efficiency by different capsids, for each of the 2 mice. Each gray spot represents data from 1 field of view. Black line represents mean ± s.e.m of different fields of view; $n = 5$ for mouse 1 and $n = 6$ for mouse 2). Cells with 1 or more detected viral transcripts were considered as transduced. **d**, Endogenous (top) and viral expression profiles (middle and bottom), for mapped cell types. Enrichment: mean of log-transformed spot counts. Relative tropism bias: z-scored spot counts log-normalized to total barcode count.

The ability to profile cells with greater transcriptomic depth can enable more nuanced cell typing. Thus, we explored the tropism profiles of the 6 capsids across 26 molecularly defined cell types in mouse cortex (Figure 2.4a,b). We observed that AAV-PHP.eB is biased towards L5 and inhibitory neurons, whereas AAV.CAP-B10 shows bias towards L2/3 and L4. Comparing inhibitory and excitatory subclasses also revealed that whereas AAV-PHP.eB and AAV.CAP-B10 are biased towards inhibitory neurons, AAV-PHP.N is biased towards excitatory subtypes (Figure

2.4d). Notably, we observed a relatively low bias for AAV-PHP.AX (Figure 2.4d), suggesting that this capsid could be used for broad targeting.

In addition to profiling 6 engineered AAV capsids in the cortex, we also assessed capsid tropism in the striatum, thalamus, and cerebellum, incorporating relevant cell type markers for those regions (Figure 2.4e,f). As opposed to dissociation-based single cell or single nucleus profiling, *in situ* profiling methods should be easier to translate across tissues or brain regions, as the latter do not require optimization of dissociation protocols. Our transduction profiles from multiple brain regions, obtained without any adaptation to the protocol, supports this advantage for *in situ* profiling.

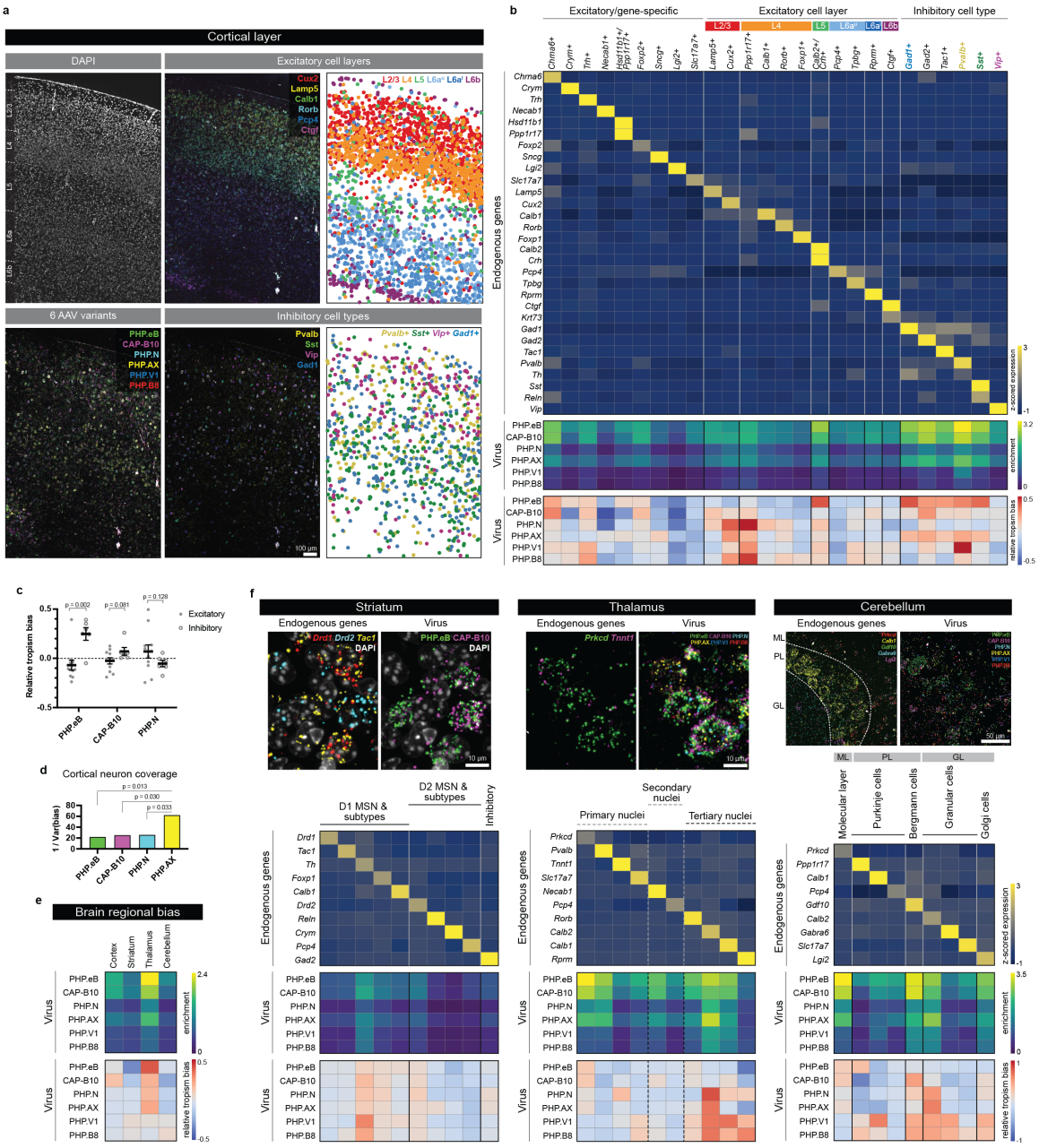


Figure 2.4. In-depth profiling of transduction by capsid variants, in neuronal subtypes, in cortical layers, and in other brain regions. a, Representative images showing mapping of excitatory and inhibitory cell types to different cortical layers, as well as transduction by 6 capsid variants. Right panels for excitatory and inhibitory marker gene show mapped cell types and localization. **b**, Endogenous (top) and viral expression profiles (middle and bottom), for mapped cell types. Enrichment: mean of log-transformed spot counts. Relative tropism bias: z-scored spot counts log-normalized to total barcode count. **c**, Relative tropism bias for 3 capsid variants from pool (AAV-PHP.eB, AAV.CAP-B10, and AAV-PHP.N), for excitatory (n = 11) and inhibitory clusters (n = 6). Black line represents mean ± s.e.m. Significance was determined by two-sided unpaired t-tests. **d**, Cortical neuron coverage for 4 neuron enriched capsid variants from the pool (AAV-PHP.eB, AAV.CAP-B10, AAV-PHP.N, and AAV-PHP.AX), quantified as the inverse

variance of relative tropism bias across cell type clusters (*F*-test on variance). As compared to the other variants, AAV-PHP.AX shows relatively broad coverage of neuronal subtypes. **e**, Profiling of regional transduction bias for 6 capsid variants across 4 brain regions (cortex, striatum, thalamus, and cerebellum). **f**, Profiling of viral transduction in striatum, thalamus, and cerebellum, in identified cell type clusters. The thalamus was also divided into putative cell-groups based on spatial localization. Representative images are shown. Heatmaps show endogenous (top) and viral expression profiles (middle and bottom), for identified cell types. Enrichment: mean of log-transformed spot counts. Relative tropism bias: z-scored spot counts log-normalized to total barcode count.

2.5 Profiling of AAV genome pools in mouse brain with USeqFISH

The altered tropism of engineered AAV vectors can be further refined through incorporation of regulatory elements into the AAV genome. Enhancers can boost transcription from of AAV genomes in a cell type-specific manner, whereas miRNA target sites (TSs) can dampen expression in specific cell types by promoting the degradation of AAV transcripts. As USeqFISH detects AAV transcripts, it is compatible with profiling of such regulatory elements. To demonstrate this, we generated a pool of AAV genomes containing 13 genome variants: 12 unique miRNA target sites (miRNA TSs), chosen based on previous miRNA sequencing results[183–185], plus one control with no TS, all uniquely identified by a USeqFISH barcode (Supplementary Figure 2.5a). As the genome variants were co-packaged, we then titered each genome variant individually using digital droplet PCR (ddPCR), and used the resulting concentrations to normalize the transduction profiles to the composition of the pool (Supplementary Figure 2.5b).

These genomic variants were co-packaged into AAV-PHP.eB and administered through RO injection into wildtype mice. The tissue was harvested 4 weeks later, and the cortex and hippocampus were profiled with USeqFISH across 16 different molecularly defined cell types (Figure 2.5b,c). Consistent with previous profiling of AAV-PHP.eB, we observed strong transduction of *Pvalb*+ cells and a bias towards transduction of L5/6. Comparing miRNA TS-containing genomes to the no TS control, we found that some miRNA TS (miRa1-1 TS and miR433-3p TS) strongly repressed expression across cell types. Furthermore, miR204-5p reduced expression in excitatory neurons, while sparing expression in inhibitory neurons. Interestingly, we also observed increased expression with miR126a-3p across profiled cell types.

Taken together, these results demonstrate that USeqFISH can be used for high-throughput, high-resolution characterization of AAV capsid and genome variants in intact tissues.

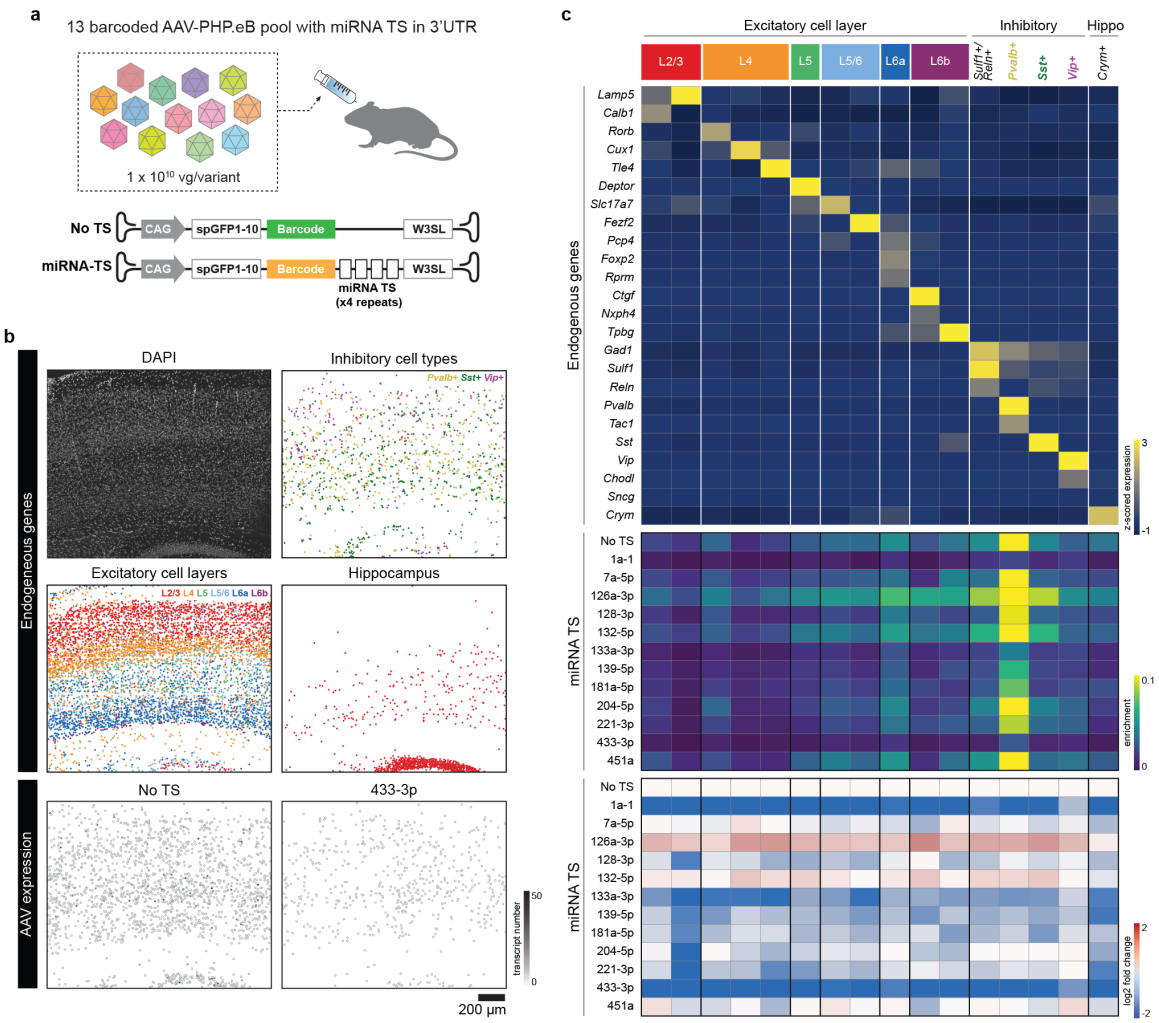


Figure 2.5. In-depth profiling of transduction by AAV genome variants, carrying different miRNA target sites (miRNA TSs), in neuronal subtypes in cortical layers. a, Experimental design. 13 barcoded genomes, with different miRNA TSs or a no target site (no TS) control were co-packaged into AAV-PHP.eB, and delivered at 1.3e11 vg total dose via RO injection. USeqFISH was used to assay viral transduction profiles. **b**, Representative images showing mapping of identified cell types, as well as transduction by no TS control (left) and miR433-3p TS-containing AAV (right). **c**, Profiling of endogenous (top) and viral expression profiles (middle and bottom), for mapped cell types. Enrichment: mean of log-transformed spot counts. $\log_2$-fold change: $\log_2$-fold change of enrichment vs. no target site control.

2.6 Discussion

High-throughput engineering of AAV capsids and genomes yields hundreds to thousands of potentially interesting variants that require thorough characterization. One-at-a-time characterization in small rodent models is time- and resource-intensive and will not scale well for characterization in more translationally relevant models, such as non-human primates. USeqFISH is a novel spatial transcriptomics method that can address this unmet need, by enabling parallel

characterization of multiple AAV capsid or genome variants in the same animal. USeqFISH maintains spatial information, and thus can be used to profile transduction of rare cell types, can be easily adapted to different tissue types, and is sensitive. Here we used USeqFISH to profile as many as 13 AAV variants, in multiple brain regions and across as many as 26 molecularly defined cell types in the same animal.

The high sensitivity of USeqFISH is advantageous, as it can enable detection of even rare transduction events. Indeed, USeqFISH was able to detect transcripts following a low dose transduction of 1 x 10^8 vg. Assuming a conservative maximum dose of 1 x 10^{12} vg and a desired per variant dose of 1 x 10^9 vg to 1 x 10^{10} vg, to allow for detection of worse-transducing variants, we estimate that USeqFISH could enable profiling of 100s to 1000s of variants in a single animal. Furthermore, the single-probe sensitivity of USeqFISH means that more space can be devoted to regulatory sequence, rather than barcode sequence. Thus, longer promoters or enhancers, or even combinations of regulatory elements, may be profiled using USeqFISH.

In the current work, transcript barcodes were read out sequentially over as many as 13 rounds of labeling. With 4 spectrally distinct fluorophores per round, this translates to at least 52 AAV variants plus cell markers that can be profiled in a single experiment. Advancing USeqFISH through integration of combinatorial or temporal barcoding schemes[186] can increase the number of variants and cell markers that can be profiled, pushing this technology closer to the goal of parallel characterization of large AAV variant pools or libraries.

2.7 Methods

Chemicals

Polyethylenimine (PEI-MAX, 24765, Polysciences), PBS (AM9625, Invitrogen), ethanol (EtOH), paraformaldehyde (PFA, RT 15714-S, Electron Microscopy Sciences), Tween 20 (P7949, MilliporeSigma), saline sodium citrate (SSC, AM9763, Invitrogen), formamide (AM9342, Invitrogen), Ribonucleoside Vanadyl Complex (RVC, S1402S, New England Biolabs (NEB)), salmon sperm DNA (15632011, Invitrogen), T4 ligase (EL0011, Thermo Fisher Scientific), BSA (B9000S, NEB), SUPERase inhibitor (AM2696, Invitrogen), Phi29 polymerase (EP0094, Thermo Fisher Scientific), dNTP (18427088, Invitrogen), 5-(3-aminoallyl)-dUTP (AM8439, Invitrogen), acrylic acid N-hydroxysuccinimide ester (AA-NHS, A8060, MilliporeSigma), acrylamide (1610140, Bio-Rad), bisacrylamide (1610142, Bio-Rad), tetramethylethylenediamine (TEMED, T7024, MilliporeSigma), ammonium persulfate (APS, A3678, MilliporeSigma), Gel Slick (50640, Lonza), HCR hairpins (Molecular Technologies), VA-044 (27776-21-2, FUJIFILM Wako Pure Chemical Corporation), sodium dodecyl sulfate (SDS, 7990-OP, Calbiochem), proteinase K

(P8107S, NEB), 4′,6-diamidino-2-phenylindole (DAPI, 62248, Thermo Fisher Scientific), dimethyl sulfoxide (DMSO, D8418, Milli- poreSigma), poly-l-lysine (PLL, P8920, Sigma-Aldrich), poly-d-lysine (PDL, P6407, MilliporeSigma), laminin (230017105, Thermo Fisher Scientific), ethylene carbonate (EtCB, E26258, MilliporeSigma) and dextran sulfate sodium salt (D6001, MilliporeSigma).

Barcode and UGI sequence generation

We computationally generated unique barcodes and UGIs with the following criteria. We designed a random sequence (20 nt for barcodes and 19 nt for UGIs) that consisted of only three letters, A, C or T, to enhance hybridization efficiency[187]. For both barcodes and UGIs, we excluded those with more than four repeats of each letter and that had a hit against the mouse transcriptome via a BLAST search. The GC range and melting temperature (Tm) selected for the barcode and the UGI were different (barcode: $40\% \leq GC \leq 60\%$, $Tm < 70$ °C; UGI: $10\% \leq GC \leq 20\%$, $Tm < 40$ °C), as were their hybridization conditions. We also performed pairwise comparisons of the new sequence with previously designed barcodes or UGIs to prevent cross-hybridization.

Probe design for endogenous genes

To optimize probe design for USeqFISH (and HCR v3[172]), we improved our first version of the probe design script for HCR v3 based on MATLAB and BLAST[188] by importing it to Python and using Bowtie2[189]. This improvement made the code run much faster (<1 minute per gene) than the previous version (tens of minutes per gene). In brief, from the entire coding sequence, we selected 20-nt regions with $40\% \leq GC \leq 60\%$, no more than three (for C and G) or four (for A and T) repeats, Gibb's free energy (dG) of ≤ -9 kcal per mol and unique under a Bowtie2 search. Once the target sequence candidates were identified, we aligned the whole sequence of each primer and padlock, including linkers and UGIs, with Bowtie2 again to prevent their unexpected binding to any other endogenous genes. To make the script applicable across species, we built Bowtie2 databases from GenBank genome databases: mm10 (mouse, Mus musculus). All designed probes were ordered through IDT and diluted in Ultrapure water before use. Probe sequences are provided in Supplementary Table 2 of Jang et al., 2023[190].

Plasmid DNA

Standard molecular cloning techniques were used to generate DNA constructs in this study. Double-stranded DNA was synthesized by Integrated DNA Technologies and inserted into

pAAV backbones with NEBuilder HIFI (New England Biolabs, E2621). Barcoded AAV genome plasmids were based on pAAV-CAG- mNeonGreen-WPRE-hGHpA (Addgene #99134, RRID:Addgene_99134). Double-stranded DNA (dsDNA) fragments containing the spGFP(1–10) coding sequence and the W3SL sequence, with appropriate overhangs, were synthesized as dsDNA fragments (IDT) and inserted into pAAV-CAG-mNeonGreen-WPRE-hGHpA with NEBuilder HiFi (NEB) to generate pAAV-CAG-spGFP(1–10)-W3SL. For the six-pool experiment, barcodes with 40-nt flanking sequences complementary to the acceptor vector were synthesized as dsDNA fragments (IDT). For the miRNA TS-pool experiment, AAV genomes with barcodes and miRNA TSs were generated by a commercial vendor (Alta Biotech).

pUCmini-iCAP-AAV-PHP.eB[1] (Addgene #103005; RRID:Addgene_103005), pUCmini-iCAP-AAV.CAP-B10[79] (Addgene #175004; RRID:Addgene_175004), pUCmini-iCAP-AAV-PHP.N[81] (Addgene #127851; RRID:Addgene_127851), pUCmini-iCAP-AAV-PHP.AX (Addgene #195218; RRID:Addgene_195218) and pHelper (Agilent, #240071) plasmids were used for production of AAVs. Prior to use, all plasmids were sequence verified via Sanger sequencing and SmaI-digest or via whole-plasmid sequencing through Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation.

AAV production

Detailed protocols for AAV production and titration are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjnew5gk5/v1 and dx.doi.org/10.17504/protocols.io.e6nvw1n47lmk/v1). AAVs were produced and purified according to published methods[191], with some minor alterations. Briefly, HEK293T cells (ATCC, CRL-3216; RRID:CVCL_0063) were triple transfected with PEI-MAX (Polysciences, #24765) to deliver the rep-cap or iCAP, pHelper, and genome packaging plasmids. Viruses were harvested from cells and media, then purified over 15%, 25%, 40%, and 60% iodixanol (OptiPrep, Serumwerk, #1893) step gradients. A Type 70 Ti fixed-angle titanium rotor (Beckman Coulter, #337922) at 58.4k rpm for 1.5 hr, or a Type 70.1 Ti fixed-angle titanium rotor (Beckman Coulter, #342184) at 61.7k rpm for 1.25 hr was used, depending on the scale and number of AAVs to be purified simultaneously. Viruses were concentrated using Amicon Ultra-15 or Amicon Ultra-4 filters with a 100 kD size cutoff (MilliporeSigma, UFC9100 and UFC8100) and formulated in sterile DPBS (ThermoFisher, #14190144) with 0.001% Pluronic F-68 (ThermoFisher, #24040032). AAVs were titered with qPCR by measuring the number of DNase I-resistant viral genomes, relative to a linearized genome plasmid standard. Prior to injection, AAVs were diluted in sterile saline.

Viruses containing a WPRE were titered with the following primers:

Forward: 5'-GGCTGTTGGGCACTGACAAT-3'

Reverse: 5'-CCGAAGGGACGTAGCAGAAG-3'

Viruses containing spGFP(1-10) were titered with the following primers:

Forward: 5'-GGTTACGTGCAAGAAAGAACAA-3'

Reverse: 5'-GGTTAACCAAAGTATCTCCTTCAAA-3'

For dose and capsid pools, viruses were packaged, purified and titered separately, then combined to ensure equal dosing. For miRNA TS pool, the pAAV plasmids were pooled at equimolar amounts before transfection, and variants were thus packaged, purified, and titered simultaneously. Digital droplet PCR was used to individually titer variants in the pool.

Digital droplet PCR

For titration of individual variants in co-packaged pools, we designed sets of primers and double-quenched FAM-labeled and HEX-labeled probes (Table 2.1; IDT, resuspended in pH 8 TE buffer) targeting each miRNA TS, barcode and spGFP sequence. We extracted viral genomes[191] and performed six ten-fold serial dilutions of the extracted DNA. The final two dilutions were used for ddPCR. We loaded 3 µl of DNA into 25-µl PCR reactions (Bio-Rad, 1863024) and generated droplets from 22 µl of that PCR reaction by using droplet generation oil (Bio-Rad, 1863005) and a QX200 Droplet Generator (Bio-Rad). After transferring 40 µl of droplets to a 96-well PCR plate and sealing the plate with a pierceable heat seal (Bio-Rad, 1814040 and 1814000), we ran the PCR according to the manufacturer's protocol. After PCR, we measured droplets with a QX200 Droplet Reader and analyzed the data with the QX Manager software (Bio-Rad, 12010213). Within each well, the concentrations of one specific genome variant (miRNA TS and barcode) and all AAV genomes (spGFP) were measured to calculate the ratio of [genome variant] to [total genome], and the mean was used for normalization.

Common primer / probe set			
Target	**Forward primer**	**Reverse primer**	**Probe**
spGFP	TCATCTGTACTACTGGTAAACTTCC	TTTCATATGGTCTGGGTATCTCG	TGGCCGACTCTCGTAACAACGCTT

miRNA target site-specific primer / probe sets (common reverse primer and probe)			
Target	**Forward primer**	**Reverse primer**	**Probe**
miR-126a-3p TS	ACTCACGGTACGATATCGATAATC	GGGAAGCAATAGCATGATACAAAG	ACGCTATGTGGATACGCTGCTTT
miR-128-3p_TS	CGGTTCACTGTGATATCGATAATC		
miR-132-5p_TS	AAAGCCACGGTTTATCGATAATC		
miR-133a-3p TS	TGAAGGGGACCAAATATCGATAATC		
miR-139-5p_TS	CGTGCACTGTAGATATCGATAATC		
miR-181-5p_TS	ACAGCGTTGAATGTTTATCGATAATC		
miR-1a-1 TS	ACATACTTCTTTACATTCCATATCGATAATC		
miR-204-5p_TS	GGATGACAAAGGGAATATCGATAATC		
miR-221-3p_TS	GCAGACAATGTAGCTTATCGATAATC		
miR-433-3p_TS	GAGCCCATCATGATTATCGATAATC		
miR-451a TS	AGTAATGGTAACGGTTTTATCGATAATC		
miR-7a-5p_TS	CAAAATCACTAGTCTTCCATATCGATAATC		
no_TS	GATGGGTATAGGATAGGTATCGATAATC		

Barcode-specific primer / probe sets (common forward primer and probe)			
Target	**Forward primer**	**Reverse primer**	**Probe**
BC4	GGTCCGGTACTTCTTCCTG	ACATACCCTCAACCTGATATCG	ACCTCTCAACACAAACAGTCCTGAGC
BC7		ACATACCTAAACACCCTGATATCG	
BC12		AACCACTACTCATACACTGATATCG	
BC13		CACTACTAATCCTTCCCTGATATCG	
BC20		ACCCATATCATACCCATGATATCG	
BC28		AACTAACTACTCTCCACTTGATATCG	
BC1		ACCTTACCACCTATCTTGATATCG	
BC32		CCTCTACTATCCAACTAACTGATATCG	
BC36		CCTAACCTATCCTCCTATGATATCG	
BC61		TTACACCCAATCCCTTGATATCG	
BC62		CCCTAACCACCCTGATATCG	
BC70		CACCTATTCACCTCATTGATATCG	
BC76		CTTACCTACACTACCCTATGATATCG	

Table 2.1. Digital droplet PCR primers and probes for individual titering of miRNA TS variants from pool packaged AAVs.

Tissue culture

For AAV production, and for some *in vitro* experiments, HEK293T cells were used (ATCC, CRL-3216; RRID:CVCL_0063). Cells were grown in DMEM (ThermoFisher, #10569010) supplemented with 10% defined FBS (Cytiva, SH30070.03).

For comparison of barcode placement and to test probe specificity (Supplementary Figure 2.1), HEK293T cells were used, and were transfected with indicated pAAVs (final concentration of 100 ng/mL), using PEI-MAX (1:4). Three days after transfection, cells were fixed with 4% PFA for 10 min at room temperature, then stored in 70% EtOH at -20 °C until use.

Animals

Animal husbandry and all procedures involving animals were performed in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and approved by the Institutional Animal Care and Use Committee (IACUC) and by the Office of Laboratory Animal Resources at the California Institute of Technology.

8-week old, male C57BL/6J (strain #: 000664; RRID:IMSR_JAX:000664) mice were obtained from the Jackson Laboratory. Mice were housed 3-4 per cage, on a 12 hr light/dark cycle, and had *ad libitum* access to food and water. For animal experiments, mice were 8.5 to 9.5 weeks old at the time of injection.

Retro-orbital injection

A detailed protocol for systemic AAV administration through retro-orbital injection is available on protocols.io (dx.doi.org/10.17504/protocols.io.36wgqnw73gk5/v1). AAVs were administered via retro-orbital injection during isoflurane anesthesia (1-3% in 95% O_2/5% CO_2, provided by nose cone at 1 L/min), followed by administration of 1-2 drops of 0.5% proparacaine to the corneal surface[191].

Tissue harvest and slice preparation

After 3–4 weeks of expression, the animals were sacrificed by transcardiac perfusion with 1× PBS, followed by 4% PFA. The brain was harvested and post-fixed in 4% PFA at 4 °C for overnight. Once harvested, the brains were sliced with a vibratome to a thickness of 50 μm. The slices were post-fixed with 4% PFA for 10 minutes at room temperature, followed by EtOH for >15 minutes at −20 °C. The slices stored in EtOH were gradually rehydrated in 75% and 50% EtOH and then washed in 1× PBS for ~30 minutes before use.

Immunohistochemistry

Free-floating mouse brain slices were incubated in blocking buffer (1× PBS with 10% donkey serum and 1% BSA) with primary antibodies (Aves GFP-1020, 1:1,000) at room temperature overnight. After being washed twice with 1× PBS for 30 minutes, the samples were incubated in blocking buffer with secondary antibodies (goat anti-chicken IgY, Alexa Fluor 633, A21103, Invitrogen, 1:1,000) for 1 hour at room temperature. Samples were then washed twice with 1× PBS for 30 minutes, then mounted onto glass slides with Prolong Diamond Antifade Mountant (P36970, Molecular Probes).

USeqFISH protocol

For cell culture, we washed the samples with 1× PBST (0.1% Tween 20 in 1× PBS) for 1 hour and incubated with 10 nM probes in hybridization mixture (2× SSC, 10% formamide, 1% Tween 20, 20 mM RVC and 0.1 mg/mL of salmon sperm DNA) at 37 °C, overnight. Then, we washed the samples with wash buffer (2× SSC with 10% formamide) at 37 °C for 20 minutes twice and 2× SSC at 37 °C for 20 minutes twice. Next, we added the ligation mixture (T4 ligase (100 U/mL) in 1× T4 ligase buffer with 1% BSA and 0.2 U/mL of SUPERase inhibitor) at room temperature, overnight. After a brief wash with 1× PBST, we added the polymerization mixture (Phi29 polymerase (200 U/ml) in 1× Phi29 polymerase buffer with 1% BSA, 0.2 U/µl of SUPERase inhibitor, 250 µM dNTP and 20 µM 5-(3-aminoallyl)-dUTP) at 30 °C for 2 hours. The samples were washed with 1× PBST and then treated with AA-NHS (400 µM in 1× PBST) at room temperature for 2 hours. Next, we embedded the sample in hydrogel. The samples were immersed in hydrogel monomer solution (4% acrylamide and 0.2% bisacrylamide in 2× SSC) for 30 minutes and flattened on a glass slide. We dropped the same hydrogel solution with 0.2% TEMED and 0.2% APS to the sample and covered it with Gel Slick-coated slides. Once the gel formed in 1 hour, we detached the slide. For HCR, the initiators (10 nM in 2× SSC with 10% formamide) were added to the samples at room temperature for 30 minutes. HCR hairpins were heated at 95 °C for 90 seconds, followed by cool-down at room temperature for >30 minutes. After a brief wash with 2× SSC, hairpins (60 nM in 2× SSC) were added at room temperature for 1 hour.

USeqFISH on tissue slices was performed as described above for cultured cells, save for a few modifications. First, once rehydrated, the samples were kept in the PACT monomer solution (4% acrylamide, 1% PFA and 0.25% VA-044 in 2× SSC) at 4 °C, overnight. Next, we formed the PACT gel by purging the solution with N_2 for 5–10 minutes and immediately incubating it at 37 °C for 2 hours. After aspirating the excess gel, we washed the samples with 2×

SSC 3–4 times and cleared them in 8% SDS (in 2× SSC) at 37 °C, overnight. Once cleared, the samples were washed with 2× SSC 3–4 times at room temperature for 1 day. Then, we proceeded with probe hybridization as described above. Second, before each enzyme reaction, we washed the samples with each enzyme buffer briefly. Third, for polymerization, we immersed the samples in the Phi29 polymerization mixture at 4 °C overnight before starting the reaction at 30 °C. Finally, once embedded in the hydrogel, the samples were treated with proteinase K (0.2 mg/mL in 1× PBST) at 37 °C for 1 hour before HCR amplification.

For sequential labeling, we detached hairpin assemblies and initiators from the amplicon using a two-step stripping method and added another set of initiators and hairpins for the next round. For two-step stripping, we added unique 10-nt toehold sequences to one of the hairpin pairs (Supplementary Table 2; Integrated DNA Technologies (IDT)). Each imaging round was performed as follows. We labeled the samples with DAPI (1:5,000 in 2× SSC) for 10 minutes and imaged in 2× SSC. Next, we added the displacement oligos (Supplementary Table 2; 1 μM for cell culture and 3 μM for tissue in 2× SSC) to the sample (30 minutes for cell culture and 1 hour for tissue) and, subsequently, formamide (60% in 2× SSC, 30 minutes for cell culture; 70% in 2× SSC, 1 hour for tissue) at room temperature. After washing the samples with 2× SSC, we added the initiators for the next round. Once all imaging rounds were completed, we treated the sample with DAPI for 10 minutes and dT(30) conjugated with Alexa Fluor 647 (1 μM in 2× SSC, IDT) for 1 hour for cytosolic labeling. Hairpin and displacement oligo sequences are provided in Supplementary Table 2 of Jang et al., 2023[190].

Imaging

We used a Keyence fluorescence microscope (BZX-710) for cultured cells. For tissue slices, a confocal microscope (LSM 880, Zeiss, Zen for software control) with a ×10 air/×40 water immersion objective and a spinning disk confocal microscope (SDCM; Dragonfly, Andor, Fusion for software control) with a ×40/×100 oil immersion objective (Leica) and an sCMOS camera (Zyla, Andor) were used. For sequential labeling, we established an automated imaging and fluidic solution change system on the SDCM; the sample was attached to a glass coverslip pre-coated with poly-L-lysine (1 mg/mL) and embedded in the hydrogel. After the proteinase K treatment, the sample on the coverslip was assembled with a flow cell (FCS2, Bioptechs) connected with tubing to apply various solutions as needed at each step. Solution selection and flow control were carried out using a peristaltic pump (Minipuls 3, Gilson) and valves (MVP valves and positioners, Hamilton Company). The resolution of volume imaging was 0.151 μm per pixel in the *x* and *y* axes and 0.4–0.5 μm per pixel in the *z* axis (with a ×40 oil immersion

objective). All parts of the system were automatically controlled through RS232 and REST by a custom-built Python script.

Data analysis

For USeqFISH dose-dependency experiment (Fig. 2.2), we used a maximum intensity projection of the ~20-μm-thick volume. Quantification of the RNA signal intensity was performed as follows. We subtracted the background calculated by applying the area_opening function in scikit-image. After removing small objects with double erosion, we identified the foreground pixels and measured the mean intensity of the background to calculate the cumulative histogram of the signal intensity and the SBR (the intensity of the foreground pixels / the mean value of the background intensity). For quantifying RNA spots, we created a mask of the DAPI signal manually using Fiji and processed other channels with RNA spots as follows. We subtracted the background as described above and applied the Laplacian of Gaussian filter to detect RNA spots. Then, we calculated the distance of each spot to all nuclei and assigned it to the closest nucleus only if the distance was <10 μm.

For the pool studies (Figures 2.3–2.5), we developed a computational analysis pipeline that includes registration, spot detection and cell segmentation in a 3D volume (Supplementary Figure 2.5). First, we exploited Cellpose[192] with the dT(30)-labeled image to segment single-cell bodies in the 3D volume. We found that downsampling the dT-labeled image to make each cell have an estimated diameter of ~30 pixels worked quickly and produced the best result in single-cell segmentation. With the labeled mask of each cell, we performed a convex hull operation to smooth cell boundaries. Second, we identified the RNA spots of each channel in each round by applying the Laplacian of a Gaussian filter. Third, we acquired the transformation matrix by using phase cross-correlation of the DAPI image at each round to the last DAPI image. We added this transformation matrix to the one to correct optical aberration that we obtained with fluorescence microbeads (FocalCheck Fluorescence Microscope Test slides #1, F36909, Thermo Fisher Scientific) before the experiment. Finally, we combined all three pieces of information (segmented cells, detected spots, and registration coordinates) to assign the spots to individual cells and finally obtained the expression matrix of endogenous and viral genes in each cell. For tiled datasets (cortex layers; Figures 2.4a and 2.5b), we processed all individual tiles to get the expression matrices of each, and cells in the overlap between tiles (10%) were excluded from one tile for the clustering analysis below. Tiled images were stitched in Fusion (Andor) for visualization. The pipeline was parallelized using Dask to accelerate the processing. The total processing time was 10–20 minutes for Cellpose cell segmentation (with GPU) and 10–20

minutes per round (for registration and spot detection of all four channels) on clusters at the Caltech Resnick High Performance Computing Center.

The quantitative analysis of the expression matrix was conducted mainly with Scanpy[193] (Supplementary Figure 2.5b) on a standalone laptop. In brief, we used only the endogenous gene expression matrix of all cells to identify cell types as follows. Based on the distribution of total spot counts, we filtered cells with no RNA spots or too many (usually <5 cells from the entire dataset). After normalization and z-standardization of the data, we applied principal component analysis and Leiden clustering to the data to identify cell type clusters. We performed subclustering with the large clusters and merged the clusters based on Ward distance until the elbow point. Once the type of each cell was determined, we calculated (1) enrichment by calculating mean of log-transformed (log1p) spot counts per cell and (2) relative tropism bias by calculating mean of log-normalized and z-scored spot counts per cell. For additional information, we provide (1) transduction efficiency measured by dividing the number of cells having one or more of each viral barcode by the total cell number in each cluster and (2) mean spot numbers per cell measured by averaging the spot numbers of each virus in transduced cells in each cluster for all data in Supplementary Figure 2.4. Images were visualized using Napari, Fiji or Imaris 9.5 for 3D views.

Statistics and reproducibility

All *in vitro* experiments were repeated at least three times with similar results. All *in vivo* experiments with mice were repeated at least twice using 2–5 animals with similar results.

Data availability

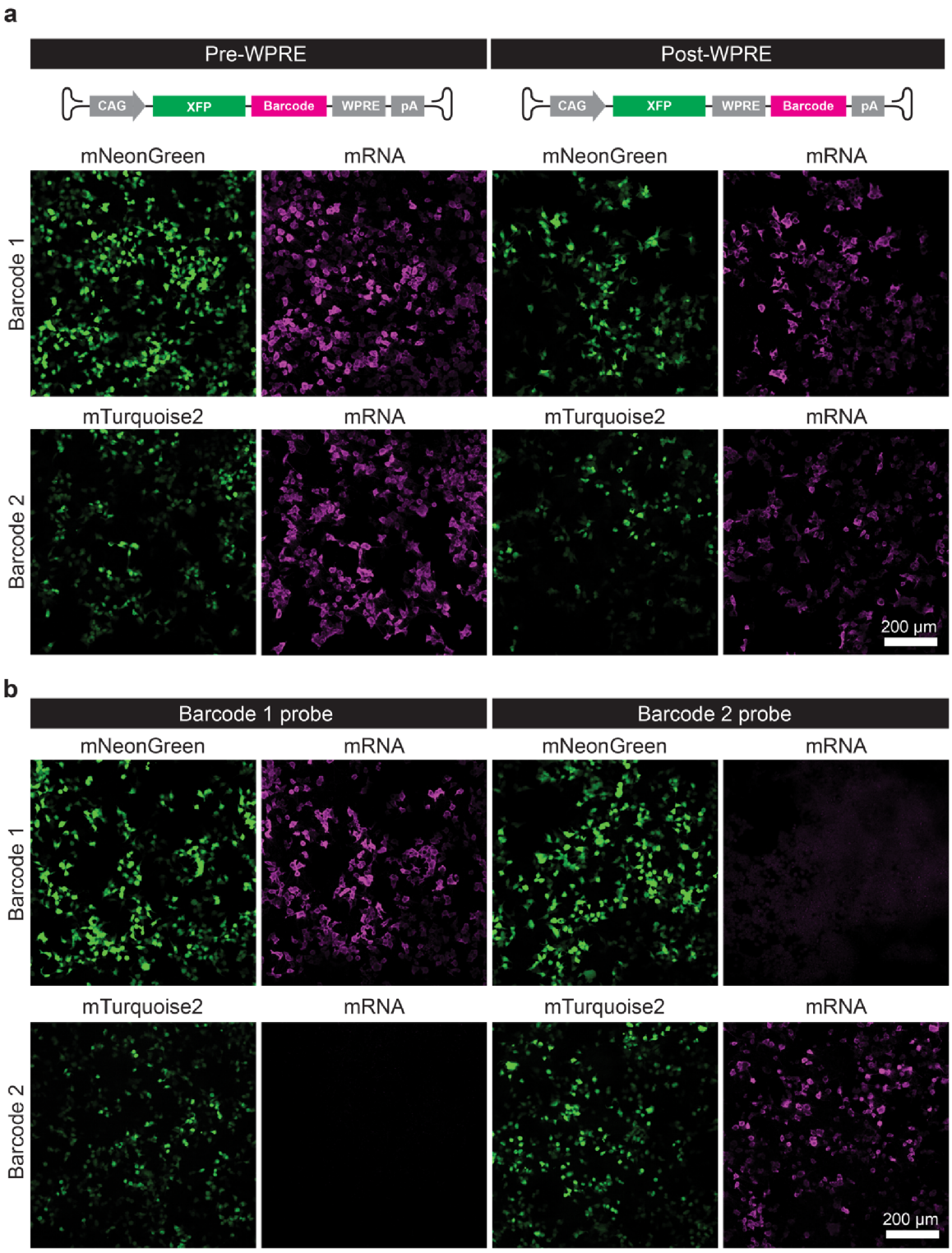
All sequences of probes and primers used in this study are provided in Supplementary Tables 2 and 3 of Jang et al., 2023[190]. We used the mm10 GenBank genome assemblies for *Mus musculus* to build Bowtie2 databases for probe design. The vector plasmid used to produce AAV-PHP.AX is available at Addgene (195218). Raw image datasets for pooled screening experiments are deposited in the Brain Image Library (https://doi. org/10.35077/g.529). Other data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

All custom Python code used in this study and an example dataset to test are available at https://github.com/GradinaruLab/useqfish_probede- sign (ref. 87; probe/barcode design for

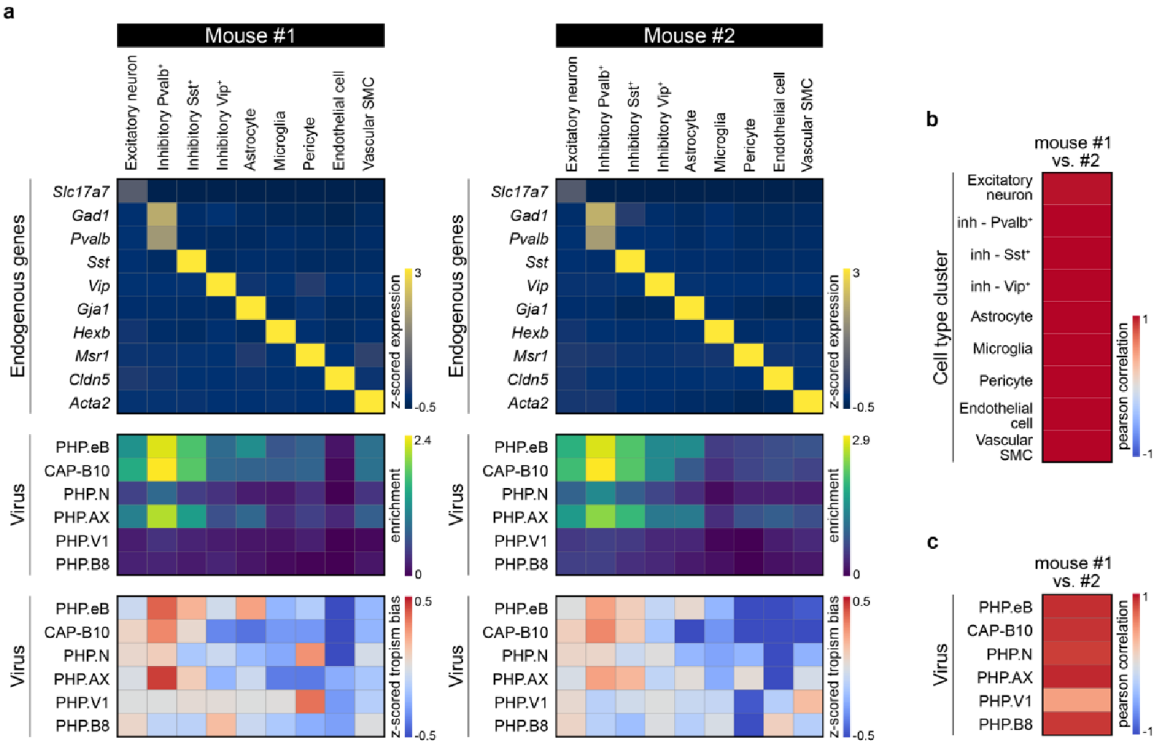
USeqFISH and HCR v3), https:// github.com/GradinaruLab/useqfish_imaging (ref. 88; automated imaging and fluidics system control) and https://github.com/GradinaruLab/ useqfish_analysis (ref. 89; image processing and data analysis).

2.8 Supplementary Figures

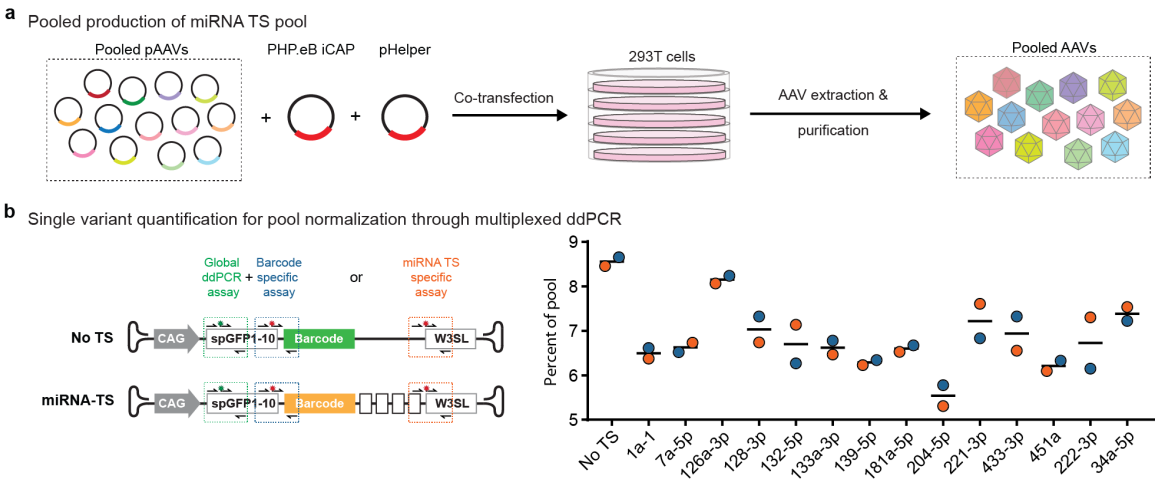


Supplementary Figure 2.1. Effect of barcode position and specificity of USeqFISH detection. a, Two barcodes ("Barcode 1" and "Barcode 2") were both tested in two transcribed locations within AAV genome: before the WPRE or after the WPRE. HEK293T were transfected

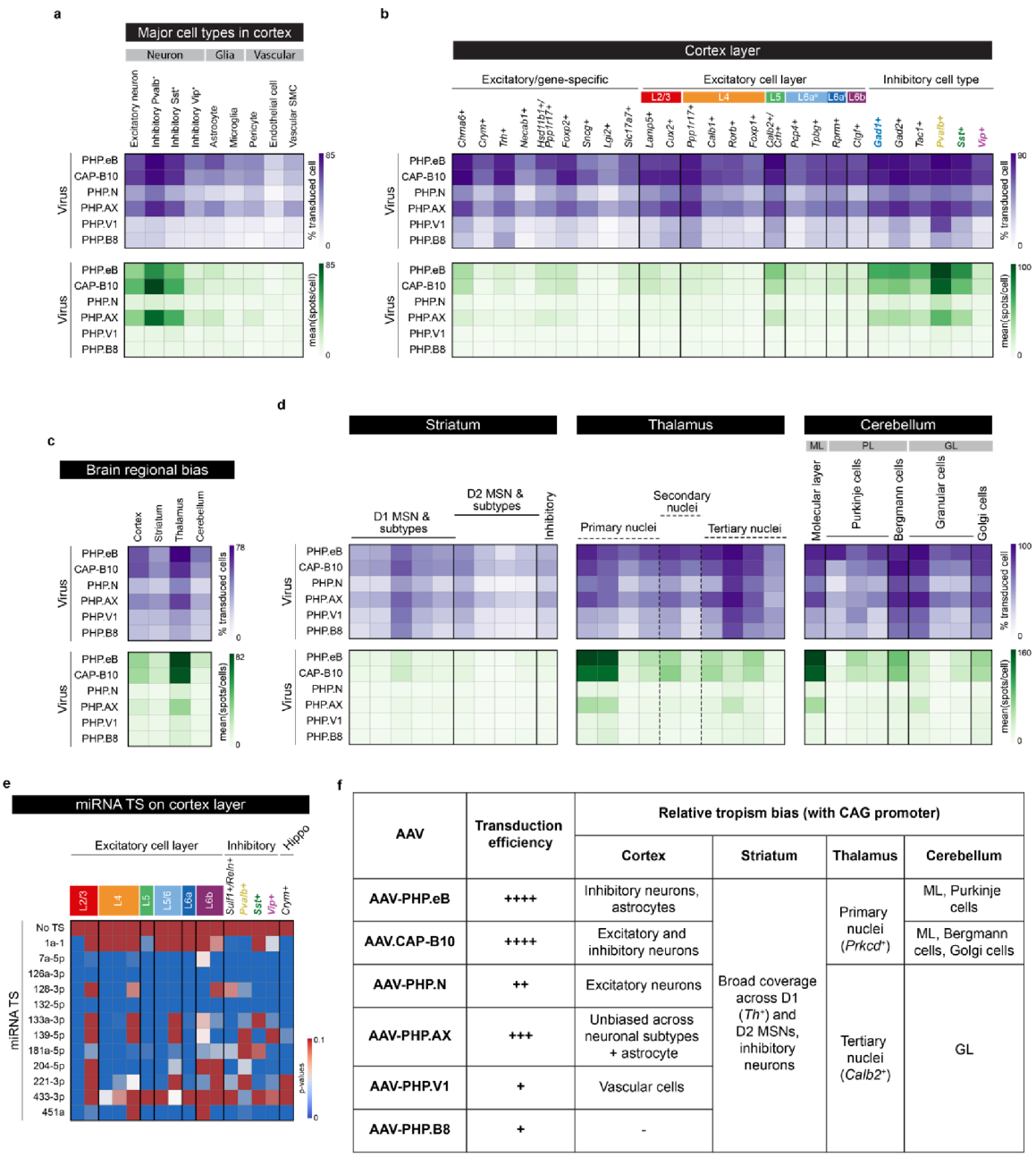
with pAAVs. Barcodes in either position could be readily detected with USeqFISH. Pre-WPRE barcode placement was used for the rest of the experiments. **b**, Specificity of barcode probes for complimentary barcode, in transfected HEK293T cells. Barcodes are only detected by complimentary probes, and no cross-reactivity was observed.



Supplementary Figure 2.2. Comparison of capsid transduction profiles between two mice. Related to Figure 2.3 and Figure 2.4. We examined the reproducibility of USeqFISH for AAV profiling by separately analyzing and comparing data from twe mice. **a**, Endogenous (top) and viral expression profiles (middle and bottom), for mapped cell types, separated into two mice. Enrichment: mean of log-transformed spot counts. Relative tropism bias: z-scored spot counts log-normalized to total barcode count. **b**, Pearson correlation between mean endogenous gene expression for mapped cell type clusters. **c**, Pearson correlation between mean enrichment for each AAV between two mice.



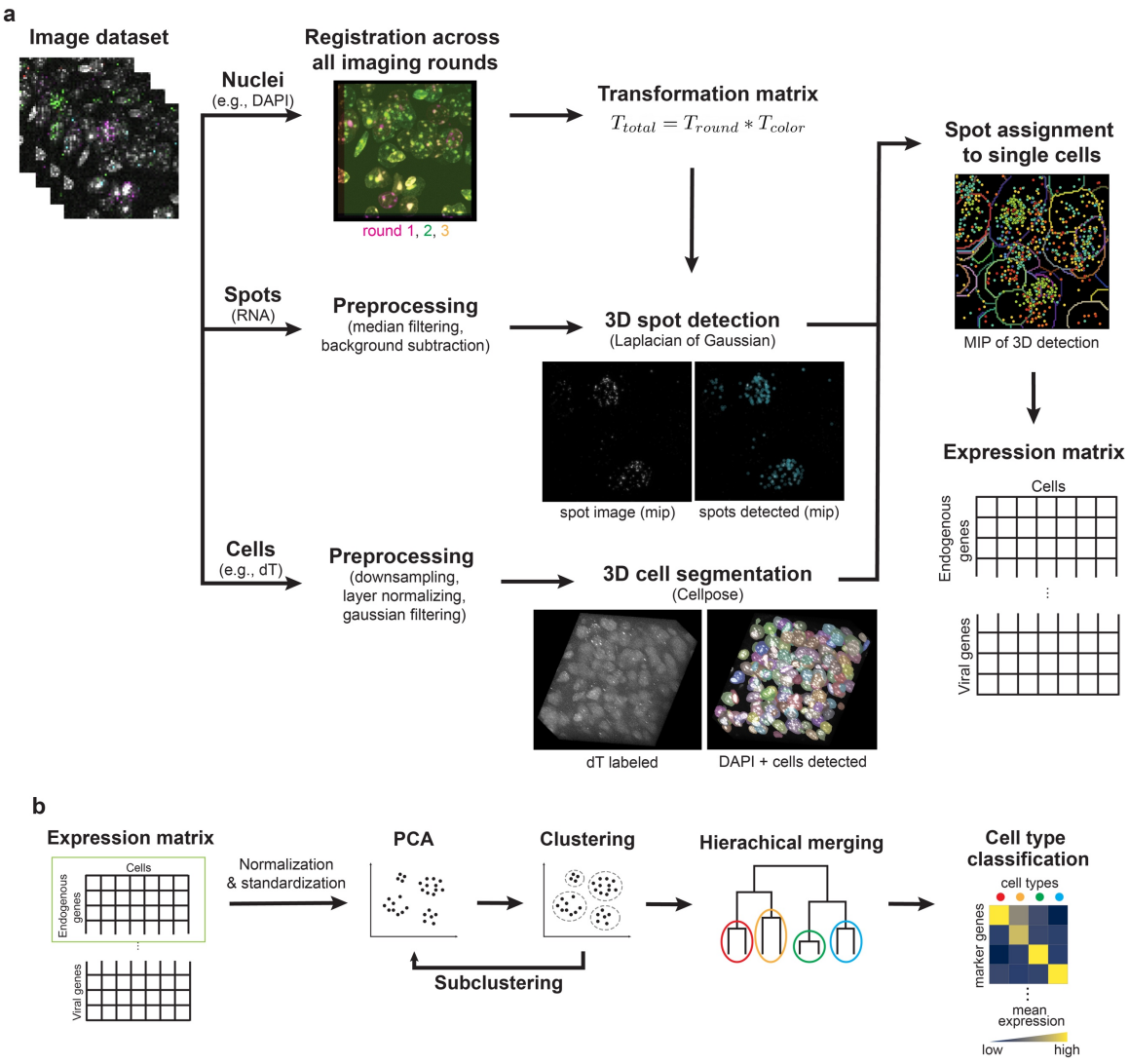
Supplementary Figure 2.3. Packaging and normalization for miRNA TS genome variant pool. Related to Figure 2.5. **a**, AAV packaging strategy. Barcoded genome packaging plasmids, with miRNA TSs or no TS control, were pooled at equimolar amounts and co-transfected into HEK293T cells, along with AAV-PHP.eB iCAP and pHelper plasmids. The pooled AAVs were then extracted, purified, and titered using primers universal to all variants in the pool. **b**, Single variant titering and normalization strategy. Multiplexed digital droplet PCR (ddPCR) was used, with probe-based assays. A global assay designed against spGFP1-10 coding sequence was used to measure total AAV genomes within each well. Barcode- or miRNA TS-specific assays were used to quantify specific genomes variants in the pool. For each variant, we calculated the proportion of genomes corresponding to that variant within the well. Barcode- and miRNA TS specific assays were conducted separately, and the mean was used for normalization. miR222-3p and miR34a-5p were also packaged in the pool, but were not assayed through USeqFISH.



f

AAV	Transduction efficiency	Relative tropism bias (with CAG promoter)			
		Cortex	Striatum	Thalamus	Cerebellum
AAV-PHP.eB	++++	Inhibitory neurons, astrocytes	Broad coverage across D1 (*Th*⁺) and D2 MSNs, inhibitory neurons	Primary nuclei (*Prkcd*⁺)	ML, Purkinje cells
AAV.CAP-B10	++++	Excitatory and inhibitory neurons			ML, Bergmann cells, Golgi cells
AAV-PHP.N	++	Excitatory neurons		Tertiary nuclei (*Calb2*⁺)	GL
AAV-PHP.AX	+++	Unbiased across neuronal subtypes + astrocyte			
AAV-PHP.V1	+	Vascular cells			
AAV-PHP.B8	+	-			

Supplementary Figure 2.4. Other measures of transduction for capsid variant pool, statistical tests for miRNA TS pool, and summary for capsid variant pool. a-d, transduction profiling for capsid pool, quantified by transduction rate (percent of cells with 1 or more AAV transcript spots) and transcript quantity (mean number of AAV transcripts spots per cell), for **a**, major cortical cell types, related to Figure 2.3d; **b**, cortical layers, related to Figure 2.4b; **c**, brain regions, related to Figure 2.4e; **d**, cell type clusters in striatum, thalamus, and cerebellum, related to Figure 2.4f. **e**, Related to Figure 2.5. Statistical comparison of miRNA TS genomes to no TS control. Significance was determined by two-sided unpaired t-tests. **f**, Related to Figure 2.3 and Figure 2.4. Summary of transduction profiles for six capsid variants.



Supplementary Figure 2.5. Automated 3D image processing and quantitative data analysis pipeline for USeqFISH. a, A collection of volume imaging data for each USeqFISH experiment consists of nuclei labeling and RNA spots for each round and cytosolic labeling for the last round. Using the nuclei labeling for each round, we calculated the rigid transformation matrix to the last image for registration across imaging rounds. We combined this transformation matrix with one for correcting optical aberration obtained with fluorescent microbeads prior to the experiment. For RNA spot detection, we proceeded with smoothing by 3-pixel median filtering and background subtraction for each volume and applied a Laplacian of Gaussian filter to obtain the location of each spot. For cell body segmentation, we preprocessed the dT labeled image and used it to identify single cells by applying Cellpose. These three calculations were then combined to register all volumes into the same coordinates, to assign each spot to each cell, and finally to obtain the cells-by-genes expression matrix. **b**, The expression matrix of endogenous genes was then normalized and z-standardized and used to cluster cell types with endogenous genes by applying principal component analysis (PCA), followed by Leiden clustering. The viral gene expression profiles were analyzed along the clusters identified.

Chapter 3

SPATIAL GENOMICS TOOLS FOR TRACKING AAV GENOMES AND CONCATEMERS

Adapted from:

Coughlin, G.M.[#], Borsos, M.[#], Barcelona, B.H.[$], Appling, N.[$], Mayfield, A.M.H., Mackey, E.D., Eser, R.A., Jackson, C.R., Chen, X., Ravindra Kumar, S., and Gradinaru, V.* (2025). Spatial genomics of AAV vectors reveals mechanism of transcriptional crosstalk that enables targeted delivery of large genetic cargo. Nat. Biotechnol. https://doi.org/10.1038/s41587-025-02565-4

3.1 Summary

Refined understanding of AAV transduction can inform both research and therapeutic application of these vectors. Profiling transduction by assaying reporter gene protein or RNA may miss transduction events in which DNA becomes epigenetically silenced or degraded before adequate transgene expression. Furthermore, methods that preserve spatial information can enable monitoring of AAV trafficking in early stages of transduction. Thus, we sought to develop methods for detection and quantification of the AAV's DNA genome with spatial resolution. AAV-Zombie reveals intracellular AAV genome localization in cultured cells and in tissue. When used with self-complementary AAV genomes, AAV-Zombie can detect AAV genomes at any stage of transduction. We use this method, in combination with immunofluorescence staining, to profile the interaction between the AAV genome and capsid in primary neurons. Furthermore, using AAV-Zombie to investigate DNA-level transduction by two generations of engineered capsids, recapitulates results obtained through protein- and RNA-level profiling, suggesting that differences between these engineered capsids are due to cell entry, rather than transcriptional or post-transcriptional differences.

Within the host cell, AAV genomes undergo concatemerization, forming episomal linear or circular molecules with multiple AAV genomes. We adapted our AAV-Zombie protocol for specific detection of concatemerized genomes. Using this method, termed SpECTr, we then constructed multiparametric views of AAV transduction over time in primary neurons, and explored relationships between AAV genome form, localization, and reporter expression. These results support a role for AAV concatemers in promoting strong transgene expression.

3.2 AAV-Zombie reveals intracellular genome localization

To better understand AAV vector biology, we wanted to develop methods to visualize and quantify AAV transduction at the DNA level. Thus, we adapted the Zombie method[194], by incorporating phage RNA polymerase promoters and barcodes into the AAV genome (Figure 3.1a). *In situ* transcription and HCR-FISH against the nascent barcoded transcript allow for subcellular localization of both single-stranded AAV (ssAAV) and self-complementary AAV (scAAV) genomes (Figure 3.1b). Importantly, fixation by methanol and acetic acid is sufficient to release the AAV genome, enabling detection of scAAV genomes irrespective of processing by the host cell (Supplementary Figure 3.2).

Understanding AAV trafficking and processing at early stages of transduction can provide invaluable insights into the vector's biology. To investigate the dynamics of AAV capsid-genome interaction, we paired AAV-Zombie with immunohistochemistry (IHC) and profiled transduction in primary neuron culture over 24 hours (Figure 3.1c and Supplementary Figure 3.2a-c). As expected, capsid puncta were transient, in both the cytoplasm and nucleus, peaking early in transduction and dropping back to baseline by 12 hours. scAAV genomes were more stable over time in both compartments. Importantly, more than 96% of capsid puncta colocalized with a genome (across all time points); the fraction of genome puncta colocalizing with a capsid was lower and decreased over time (Supplementary Figure 3.2b,c).

Given these promising results of AAV-Zombie in cultured cells, we then tested its performance in mouse brain and liver, comparing scAAV genome localization at 1 day post-injection between two generations of engineered capsids and their parent AAV9 (Figure 3.1d and Supplementary Figure 3.2d). Consistent with known protein- and RNA-level transduction patterns[1,2] (Figure 3.1d, bottom), AAV9 accumulated in the liver, but was rarely observed in the brain, while AAV-PHP.eB[175] and AAV.CAP-B10[79] both strongly localized to the brain, with reduced liver signal for AAV-PHP.eB and no detected liver signal for AAV.CAP-B10. At this early time point, both AAV-PHP.eB and AAV.CAP-B10 showed very strong accumulation in brain vasculature. Tracking of AAV.CAP-B10-delivered genomes over time shows a progressive loss in this vascular signal (Supplementary Figure 3.2e). These results demonstrate the power of AAV-Zombie for exploring AAV transduction, both in cultured cells and in tissue.

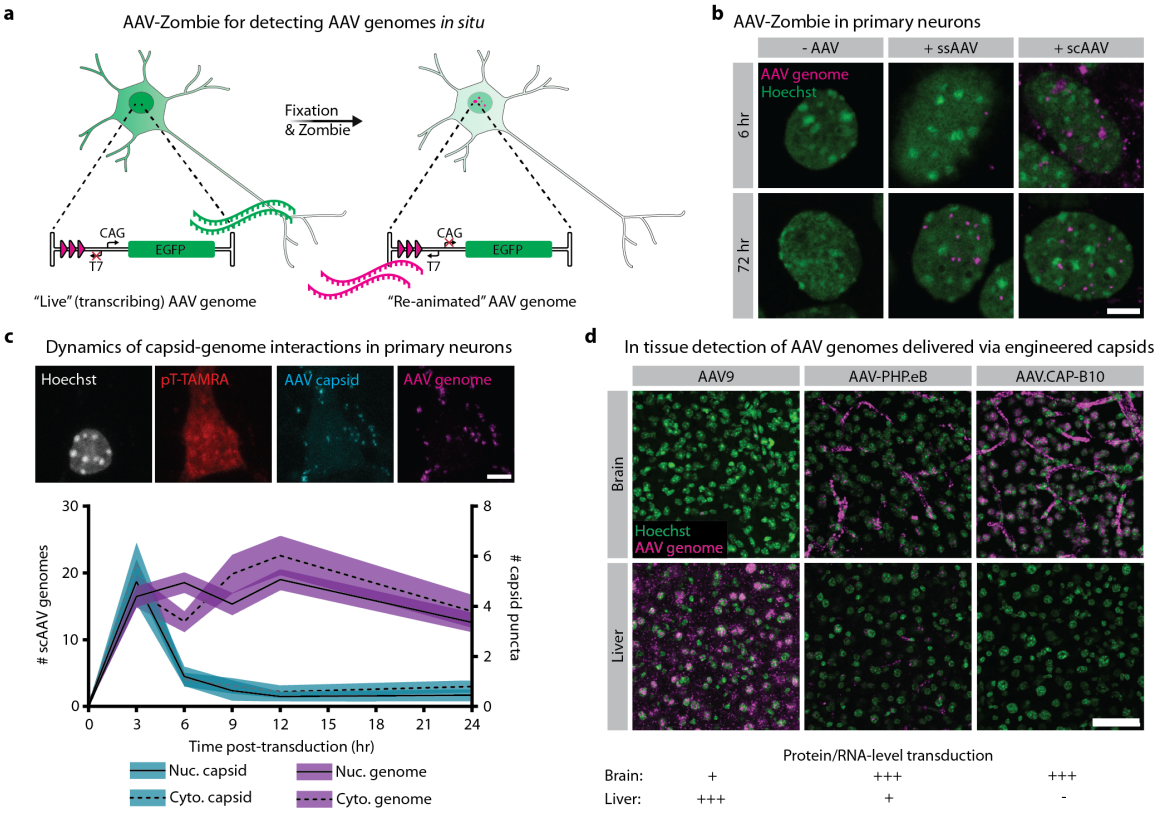


Figure 3.1. AAV-Zombie reveals intracellular AAV genome localization in cultured cells and in tissue. a, Schematic of AAV-Zombie. A barcode and phage RNA polymerase promoter are integrated into the AAV genome. While the cell is alive, the barcode is not transcribed. After fixation, *in situ* transcription of the barcode by phage RNA polymerase yields barcoded transcripts that can be detected by HCR-FISH. These transcripts serve as a proxy for the AAV genome. **b**, Detection of single-stranded and self-complementary AAV genomes (ssAAV and scAAV, respectively) in cultured primary neurons. At 6 hours post-transduction, ssAAV genomes are rarely detected due to the necessity of second strand synthesis, whereas scAAV genomes are readily detected in and outside the nucleus. At 72 hours, genomes of both formats are detected in the nucleus. All genomes delivered at 1e5 MOI in AAV6. Scale bar = 5 μm. **c**, Time course of AAV capsids and scAAV genomes in nucleus and cytoplasm of primary neurons. Capsids were detected through immunofluorescence with an antibody against linear epitopes. Cytoplasm was labeled with a TAMRA-conjugated polyT probe. Genomes were delivered at 1e6 MOI in AAV-DJ. Black line is mean; shaded area is 95% confidence interval. n = 243 (t = 0 hr), 191 (3 hr), 317 (6 hr), 212 (9 hr), 220 (12 hr), 255 (24 hr) neurons per time point. Scale bar = 5 μm. **d**, AAV-Zombie detection of scAAV genomes in C57BL/6J mouse brain and liver 1 day post-injection, following systemic delivery by AAV9, AAV-PHP.eB, or AAV.CAP-B10, at 3e11 vg dose. Distribution of AAV genomes recapitulates known protein- and RNA-level transduction profiles (bottom). Representative images from n = 3 animals per condition. Scale bar = 50 μm.

3.3 SpECTr reveals dynamics of AAV concatemerization

Within host cells, AAV genomes are often processed into linear or circular concatemeric episomes, through the activity of host DNA repair factors[44–49]. These concatemers are thought to be important for persistence of the AAV genome and expression. Thus, we also sought to develop methods for specific detection and quantification of AAV concatemers.

To enable detection of concatemerized AAV genomes, we adapted AAV-Zombie by separating the barcode and T7 RNA polymerase promoter into separate AAV genomes (termed "Genome A" and "Genome B," respectively). (Figure 3.2a). Concatemerization of these two genomes orients the T7 promoter and its barcode (hereafter referred to as ConcBC) such that T7 polymerase can transcribe the barcode. Genome B also contains a barcode (GenBC) driven by an SP6 RNA polymerase promoter, allowing detection of that AAV genome independent of concatemerization. The short length of the phage promoters and barcodes (~20 nt and 100-250 nt, respectively) leaves ample space for strong mammalian promoters and reporter genes. Thus, following cotransduction, fixation, and Zombie, we could detect the concatemer-independent barcode, concatemer-dependent barcode, as well as reporter gene transcripts (Figure 3.2b), providing single-molecule information about AAV transduction, concatemer formation, and expression in single cells. We term this method SpECTr, for "SpECTr Enables AAV Concatemer Tracking."

To confirm that the ConcBC transcript arises from a single molecule containing both the T7 promoter and ConcBC, we performed *in situ* restriction enzyme digests on AAV-DJ-transduced and fixed HEK293T cells before barcode transcription. Digestion with SmaI (which cuts within the AAV ITR) or MscI (which cuts immediately downstream of the T7 promoter) significantly reduced the number of detected ConcBC spots, without affecting the number of GenBC spots. Conversely, digestion with BstEII (which cuts immediately downstream of the SP6 promoter) significantly reduced the number of GenBC spots without affecting the number of ConcBC spots (Figure 3.2c-e and Supplementary Figure 3.3). These results provided confidence that SpECTr specifically detects AAV concatemers *in situ*.

To test the utility of SpECTr for exploring AAV transduction, we conducted a time course of AAV-DJ transduction in primary neurons, collecting samples at 14 time points over 360 hours post-transduction (Figure 3.2f and Supplementary Figure 3.4a). As expected, we observed an immediate and steadily increasing count of AAV genomes in both the nucleus and cytoplasm. Nuclear concatemeric genome counts began to rise between 12 and 24 hours post-transduction, followed shortly after by EGFP transcript intensity. The relative order of these increases (genomes, concatemers, transcript) further supports that SpECTr is detecting AAV concatemers.

Consistent with specific detection of AAV concatemers, cytoplasmic concatemer counts were low at all time points measured (mean < 1 and median = 0, per cell, for each time point).

SpECTr provides subcellular and multiparametric data about AAV transduction, enabling us to explore relationships between genome forms, their localization, and expression at the single-cell level (Figure 3.2g and Supplementary Figure 3.4b). Notably, we observed a weak correlation of reporter transcript intensity with cytoplasmic genome counts (R^2 = .087), a moderate correlation with nuclear genome counts (R^2 = .317), and a strong correlation with nuclear concatemer counts (R^2 = .541) (Figure 3.2g).

Previous work has demonstrated that AAV concatemers can increase in size over time[46,49]. Likewise, we observed variation in the measured concatemer spot area over time (Supplementary Figure 3.4c), with larger spots more frequently observed at later time points. To assess whether the spot area is indeed related to the size of the concatemer, we transfected HEK293T cells with plasmids containing increasing numbers of T7-barcode repeats and performed Zombie (Supplementary Figure 3.5a,b). As expected, plasmids with more T7-BC repeats yielded larger spots.

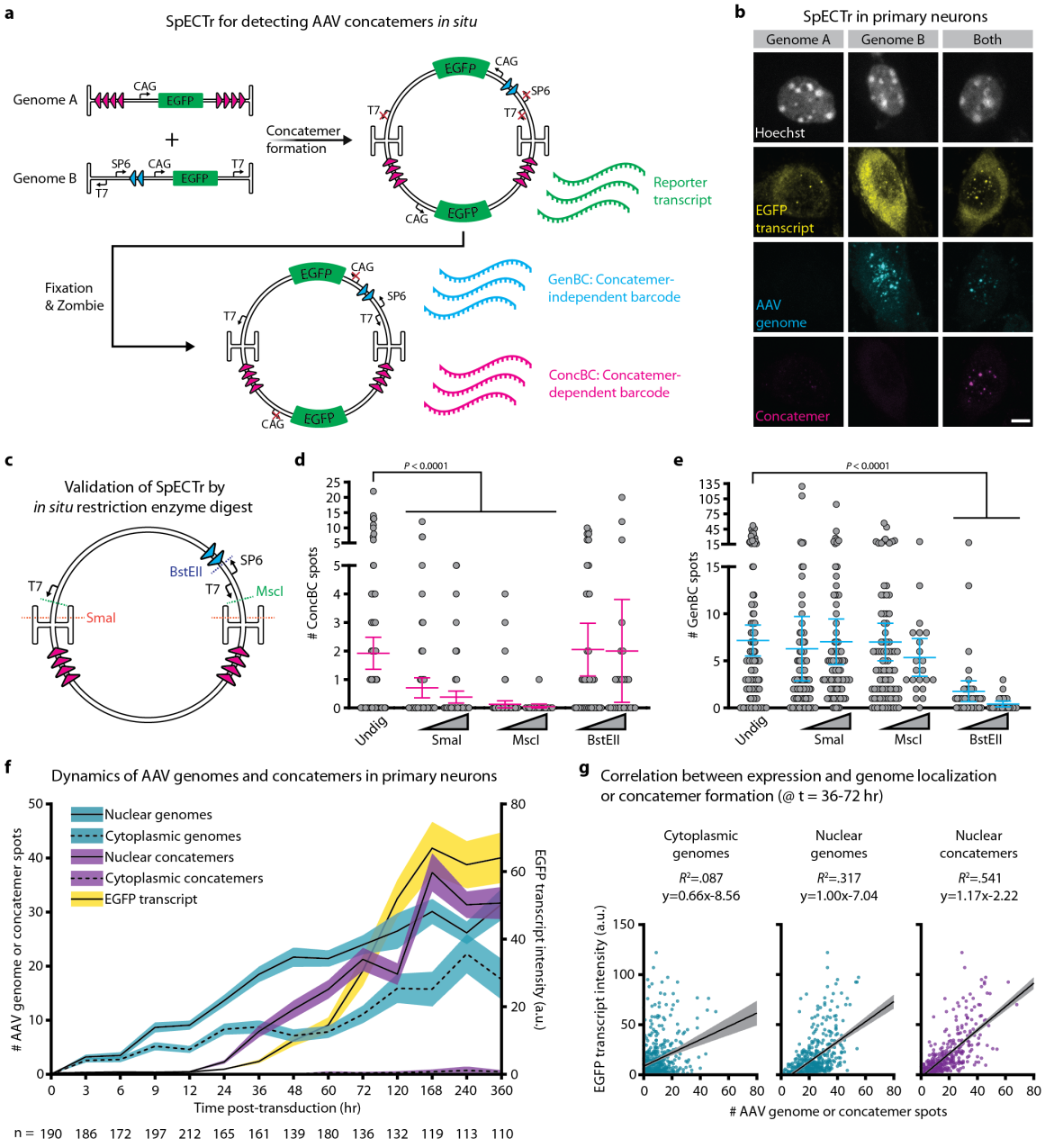


Figure 3.2. SpECTr reveals spatiotemporal dynamics of AAV concatemerization. a, Two AAV genomes are used: Genome A delivers a concatemerization-dependent barcode (ConcBC) and Genome B delivers the T7 RNA polymerase promoter. Concatemerization of these two genomes orients the T7 promoter and the ConcBC such that T7 RNA polymerase can transcribe the ConcBC. Genome B also contains a concatemerization-independent barcode (GenBC), driven by an SP6 RNA polymerase promoter. Both genomes carry a CAG-driven EGFP. **b**, Specificity of SpECTr in primary neurons, 72 hours post-transduction. Scale bar = 5 μm. **c-e**, Validation of SpECTr through *in situ* restriction enzyme digest of HEK293T cells transduced with SpECTr genomes. **c**, Model AAV concatemer containing 1 copy of Genome A and 1 copy of the Genome B, showing SmaI, MscI and BstEII restriction enzyme sites. **d**,**e**, Number of ConcBC spots (**d**) and GenBC spots (**e**) detected following *in situ* restriction enzyme digests, with low (20 U/mL)

and high (200 U/mL) restriction enzyme concentrations. “Undig”: undigested condition in which fixed cells were incubated at 37 °C in restriction enzyme buffer, without any enzyme present. Statistical significance was determined using Kruskal-Wallis ($P < 0.0001$) with Dunn’s test against the undigested condition. Bars are mean ± s.e.m. $n = 138$ (Undigested), 99 (low SmaI), 89 (high SmaI), 87 (low MscI), 22 (high MscI), 40 (low BstEII), 26 (high BstEII) cells per condition. **f**, Time course of AAV transduction, concatemer formation, and EGFP reporter transcription in primary neurons. Cytoplasm was labeled with a TAMRA-conjugated polyT probe, and nucleus with Hoechst. EGFP transcript intensity was quantified in entire soma; AAV genomes and concatemers were quantified in nucleus and cytoplasm separately. Black line is mean; shaded area is 95% confidence interval. Number of neurons per time point is indicated on figure. **g**, Correlation between EGFP reporter expression and indicated genome states. $n = 616$ primary neurons, pooled from t = 36, 48, 60, and 72 hr time points (chosen for detectable EGFP transcript that had not yet plateaued). Shaded area is 95% confidence interval. For all experiments, genomes were delivered at 1e6 MOI in AAV-DJ.

3.4 Discussion

Greater understanding of natural and engineered AAV transduction pathways may yield insights that can improve gene delivery. Characterization at a single molecule level, with subcellular resolution and in concert with other readouts of transduction, will further this understanding. AAV-Zombie and SpECTr allow for single molecule detection of AAV DNA genomes and concatemers. Paired with HCR-FISH and immunohistochemistry, these methods allow for multiparametric characterization of AAV transduction, both *in vitro* and *in vivo*. Though we restricted our profiling to one AAV genome per sample, incorporating a barcoding strategy to allow for multiplexed profiling of DNA-level transduction should be trivial, and may allow for profiling of AAV pools or libraries.

Profiling AAV transduction at the DNA level can reveal transduction events that are undetectable at the RNA or protein level, due to epigenetic silencing or degradation of transgene products. For example, microglia have been observed to be resistant to AAV transduction. Using DNA FISH, Wang and colleagues[195] demonstrated that microglia do take up AAVs, but seem to degrade the AAV genome prior to transgene expression. Similarly, our profiling of *in vivo* brain transduction by AAV-PHP.eB and AAV.CAP-B10 revealed a potential bottleneck in transduction. We observe strong accumulation of AAVs in brain endothelial cells at 1 day post-injection, but sparse presence of AAVs in parenchyma 14 days post-injection, suggesting that a majority of these AAVs entered endothelial cells, but failed to transcytose and transduce cells in the brain parenchyma.

Our profiling of AAV concatemer formation suggests a functional role for AAV concatemerization in productive transduction. Specifically, we observed that reporter transcript

levels began to rise after we detected nuclear concatemers. Furthermore, reporter transcript expression was most strongly correlated with counts of nuclear concatemers, rather than nuclear or cytoplasmic genomes. Use and further development of AAV-Zombie and SpECTr may provide additional insights into AAV biology, which can in turn be used to improve AAV vectors.

3.5 Methods

Key resources

Key resources for this work are listed in Supplementary Table 2 of Coughlin et al., 2025.

Plasmid DNA

Standard molecular cloning techniques were used to generate DNA constructs in this study. Double-stranded DNA was synthesized by Integrated DNA Technologies and inserted into pAAV backbones with NEBuilder HIFI (New England Biolabs, E2621). Plasmids used in Supplementary Figure 3.5 were constructed from PCR-amplified DNA fragments (Integrated DNA Technologies) assembled via Golden Gate Assembly (New England Biolabs, E1602S).

pUCmini-iCAP-AAV-PHP.eB[175] (Addgene #103005; RRID:Addgene_103005), pUCmini-iCAP-AAV.CAP-B10[79] (Addgene #175004; RRID:Addgene_175004), AAV-DJ rep-cap (Cell Biolabs, VPK-420-DJ), AAV6 rep-cap (Cell Biolabs, VPK-426), and pHelper (Agilent, #240071) plasmids were used for production of AAVs. Prior to use, all plasmids were sequence verified via whole-plasmid sequencing through Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation.

AAV production

Detailed protocols for AAV production and titration are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjnew5gk5/v1 and dx.doi.org/10.17504/protocols.io.e6nvw1n47lmk/v1). AAVs were produced and purified according to published methods[191], with some minor alterations. Briefly, HEK293T cells (ATCC, CRL-3216; RRID:CVCL_0063) were triple transfected with PEI-MAX (Polysciences, #24765) to deliver the rep-cap or iCAP, pHelper, and genome packaging plasmids. Viruses were harvested from cells and media, then purified over 15%, 25%, 40%, and 60% iodixanol (OptiPrep, Serumwerk, #1893) step gradients. A Type 70 Ti fixed-angle titanium rotor (Beckman Coulter, #337922) at 58.4k rpm for 1.5 hr, or a Type 70.1 Ti fixed-angle titanium rotor (Beckman Coulter, #342184) at 61.7k rpm for 1.25 hr was used, depending on the scale and number of AAVs to be purified simultaneously. Viruses were concentrated using Amicon Ultra-15 or Amicon Ultra-4

filters with a 100 kD size cutoff (MilliporeSigma, UFC9100 and UFC8100) and formulated in sterile DPBS (ThermoFisher, #14190144) with 0.001% Pluronic F-68 (ThermoFisher, #24040032). AAVs were titered with qPCR by measuring the number of DNase I-resistant viral genomes, relative to a linearized genome plasmid standard. Prior to injection, AAVs were diluted in sterile saline. The following qPCR primers were used for titering AAV viruses.

ssAAV viruses were titered with the following primers against the W3 sequence:

Forward: 5'-TGGTATTCTTAACTATGTTGCTCCT-3'

Reverse: 5'-AAGCCATACGGGAAGCAATAG-3'

scAAV viruses were titered with the following primers against the EGFP sequence:

Forward: 5'- TCTTCAAGTCCGCCATGC-3'

Reverse: 5'- CGCCCTCGAACTTCACC-3'

Tissue culture

For AAV production, and for some *in vitro* experiments, HEK293T cells were used (ATCC, CRL-3216; RRID:CVCL_0063). Cells were grown in DMEM (ThermoFisher, #10569010) supplemented with 10% defined FBS (Cytiva, SH30070.03).

For small-scale HEK293T experiments, cells were seeded at optimal confluence (50% for transduction, 90% for transfection) in the morning, and transfected or transduced in the afternoon. For transfection, Lipofectamine LTX (ThermoFisher, #15338100) was used, with 500 ng total of DNA and 3 µL of transfection reagent. To avoid saturating SpECTr signal, 1000 dsDNA copies per cell (for Supplementary Figure 3.5) was used, with pUC19 (New England Biolabs, N3041S; RRID:Addgene_50005) used as filler to ensure efficient transfection. For *in situ* restriction enzyme digest of AAV concatemers (Figure 3.2c-e & Supplementary Figure 3.3), an MOI of 1e6 AAV-DJ was used and cells were collected 3 days later. On the morning of collection, we passaged cells 1:10 onto poly-D-lysine coated coverslips (Neuvitro, GG-12-1.5h-PDL). Once HEK293T cells had attached, the coverslips were washed three times in DPBS and then fixed. For analysis of fluorescent protein expression, cells were fixed with ice-cold 4% paraformaldehyde (PFA, Electron Microscopy Sciences, #15714-S) in 1x PBS for 15 min at 4 °C and stored in 1x PBS at 4 °C until use. For AAV-Zombie or SpECTr, cells were fixed with ice-cold 3:1 methanol:acetic acid (MAA, Sigma-Aldrich, #322415 and A6283) for 15 min at -20°C, then stored at -20 °C in 70% ethanol until use.

For primary neuron cultures, coverslips (Neuvitro, GG-12-1.5h-pre) were prepared by coating with poly-D-lysine (0.1 mg/mL overnight, Sigma-Aldrich, P6407), poly-L-ornithine

(0.01% overnight, Sigma-Aldrich, P4957), and laminin (0.02 mg/mL overnight, ThermoFisher, #23017015). Primary neurons were prepared by pooling cortices and hippocampi from several E16.5 embryos and digesting the tissue in 15 U/mL papain (Sigma-Aldrich, P3125). The cell suspension was then treated with DNase I and cells triturated in Hanks balanced salt solution (ThermoFisher, #14025092), with 5% horse serum (ThermoFisher, #16050130), then centrifuged through 4% bovine serum albumin. The cell pellet was resuspended in NeuroCult Neuronal Plating Medium (STEMCELL Technologies, #05713), supplemented with 1:50 NeuroCult SM1 (STEMCELL Technologies, #05711), 0.5 mM GlutaMAX (ThermoFisher, #35050061), and 3.7 µg/mL L-Glutamic acid (Sigma-Aldrich, #49449), and plated at a density of 60,000 cells per coverslip. At 5 days *in vitro* (DIV), half the media was exchanged for BrainPhys Neuronal Media (STEMCELL Technologies, #05790), also supplemented with 1:50 NeuroCult SM1. For transduction, AAV was diluted in the added growth media. The removed plating media was saved and combined 1:1 with complete BrainPhys media. To minimize prolonged transduction due to AAVs in culture media, we used the 1:1 mix of conditioned plating media and BrainPhys media to perform a complete media change at 3 hr post-transduction, with 3 washes in pre-warmed BrainPhys between the aspiration of the virus-containing media and addition of fresh conditioned media. Subsequently, the media was half-changed with supplemented BrainPhys media every 3 days. Primary neurons were harvested and fixed as described for HEK293T cells above.

Animals

Animal husbandry and all procedures involving animals were performed in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and approved by the Institutional Animal Care and Use Committee (IACUC) and by the Office of Laboratory Animal Resources at the California Institute of Technology.

8-week old, male C57BL/6J (strain #: 000664; RRID:IMSR_JAX:000664) mice were obtained from the Jackson Laboratory. Mice were housed 3-4 per cage, on a 12 hr light/dark cycle, and had *ad libitum* access to food and water. For animal experiments, mice were 8.5 to 9.5 weeks old at the time of injection.

For primary neuron cultures, timed pregnant C57BL/6N (RRID:MGI:2159965) dams were obtained from Charles River Laboratories.

Retro-orbital injection

A detailed protocol for systemic AAV administration through retro-orbital injection is available on protocols.io (dx.doi.org/10.17504/protocols.io.36wgqnw73gk5/v1). AAVs were

administered via retro-orbital injection during isoflurane anesthesia (1-3% in 95% O_2/5% CO_2, provided by nose cone at 1 L/min), followed by administration of 1-2 drops of 0.5% proparacaine to the corneal surface[191].

Tissue harvest and processing

Tissue was collected 4 weeks post-AAV administration. Animals were euthanized via i.p. injection of 100 mg/kg euthasol.

For AAV-Zombie in tissue, animals were transcardially perfused with 30 mL of ice-cold heparinized 1x PBS, and liver and brain were dissected out. For analysis of fluorescent protein expression, one hemisphere of brain and one lobe of liver were submerged in ice-cold 4% PFA formulated in 1x PBS and fixed overnight at 4 °C. The other hemisphere and another lobe of liver were manually dissected into 1 mm^3 pieces with regions of interest and flash frozen in O.C.T. Compound (Scigen, #4586) using a dry ice-ethanol bath. O.C.T. blocks were kept at -70 °C until sectioning.

Sections were obtained using a cryostat (Leica Biosystems). Tissue for AAV-Zombie or SpECTr was sectioned at 20 µm, collected on a clean glass slide (Brain Research Laboratories, #2575-plus), allowed to dry, then stored at -70 °C until use.

AAV-Zombie and SpECTr of cultured cells

A detailed protocol for AAV-Zombie and SpECTr on cultured cells is available on protocols.io (dx.doi.org/10.17504/protocols.io.36wgqnz53gk5/v3). AAV-Zombie and SpECTr protocols, and sequences of Zombie barcodes and their split initiator probes were adapted from Askary *et al.* (2020)[194]. Split initiator probes against endogenous genes and reporter transcripts were designed according to Jang *et al*. (2023)[190]. Sequences of HCR-FISH probes against reporter and endogenous transcripts and against Zombie/SpECTr barcodes are provided in Supplementary Table 3 of Coughlin et al., 2025.

For detection of ssAAV and scAAV genomes in cell-free conditions (Supplementary Figure 3.1), we embedded packaged AAVs (AAV-DJ serotype) in high-concentration Matrigel (Corning, #354262). AAVs were first diluted in ice-cold 1x PBS, and 30 µL of that dilution was added to a pre-chilled tube with 270 µL of high-concentration Matrigel. After mixing by pipetting and brief vortexing, 100 µL of this suspension was spread onto a PDL-coated coverslip, in a 24-well plate on ice. Following gelation for 30 min at 37 °C, the samples were incubated for 15 min in ice-cold 1x PBS at 4 °C or in MAA at -20 °C.

For AAV-Zombie and SpECTr of Matrigel-embedded AAV samples and of cultured cells on coverslips, a humidified reaction chamber consisting of a 1 mL pipette tip box filled with pre-warmed RNase-free water was used. Parafilm placed on the wafer of the box served as a surface for the *in situ* transcription reaction. Coverslips, previously fixed in MAA and stored in 70% ethanol, were first washed twice in 1x PBS. 20 µL of transcription mixture per coverslip was prepared according to the manufacturer's protocol (ThermoFisher, AM1334 and AM1330). For simultaneous T7 and SP6 reactions, the T7 buffer was used with 1 µL of each RNA polymerase. For single polymerase reactions, 2 µL of the polymerase was used. 20 µL droplets were pipetted onto the surface of the parafilm. The coverslips were dipped in UltraPure water (ThermoFisher, #10977015), quickly dried by touching their edges to a Kimwipe, then placed cell-side down over the droplets. This reaction was incubated at 37 °C for 3 hr.

Once the transcription reaction was finished, the coverslips were placed cell-side up into a clean 24-well plate and fixed for 20 min at 4 °C with ice-cold PFA in 1x PBS. This was followed by two 5 min washes in 1x PBS, followed by two 5 min washes in 5x SSC (ThermoFisher, AM9770). Samples were then incubated for 15-30 min in pre-warmed probe hybridization buffer, consisting of 2x SSC, 10% ethylene carbonate (Sigma-Aldrich, E26258), and 10% dextran sulfate (Sigma-Aldrich, #3730), at 37 °C. Following this incubation, the coverslips were incubated for 12-16 hr at 37 °C in hybridization buffer plus 2 nM of each probe. Probes for Zombie barcodes, reporter transcripts, and endogenous transcripts were pooled.

After probe hybridization, samples were washed twice for 30 min in stringent wash buffer (2x SSC, 30% ethylene carbonate) at 37 °C, then three times for 15 min in 5x SSC with 0.1 % Tween-20 (Sigma-Aldrich, P1379), and then incubated in HCR amplification buffer (2x SSC, 10% ethylene carbonate) for 20-30 min. HCR hairpins (Molecular Technologies) were heated to 95 °C for 90 s, then cooled to room temperature for 30 min in the dark. For HCR on cultured cells, 30 nM hairpin in amplification buffer was used in a 1-hr amplification reaction. The samples were then washed four times in 5x SSC with 0.1% Tween-20 (10 min per wash, at room temperature).

In some cases, the cytoplasm was labeled with a fluorophore-conjugated poly(dT_{30}) probe (Integrated DNA Technologies). Coverslips were incubated with 100 nM of poly(dT_{30}) probe in 5x SSC with 0.1% Tween-20 for 1 hr, followed by four 10 min, room temperature washes in 5x SSC with 0.1% Tween-20. Finally, Hoechst 33342 was used to label cell nuclei. Samples were mounted with Prolong Diamond Antifade Mountant.

For co-detection of AAV genomes and capsids, a mouse anti-AAV VP1/VP2/VP3 monoclonal antibody conjugated to Alexa Fluor 488 was used (Clone B1, Progen, #61058-488,

RRID:AB_3107170). Following poly(dT) labeling, the samples were immunolabeled as described above, with an overnight 4 °C incubation with a 1:100 dilution of the primary antibody in blocking buffer.

For *in situ* restriction enzyme digest, coverslips were treated with restriction enzymes after MAA fixation and before *in situ* transcription. Restriction enzyme digests were carried out overnight, at 25 °C for SmaI (New England Biolabs, R0141), and at 37 °C for MscI (New England Biolabs, R0534) and BstEII-HF (New England Biolabs, R3162).

AAV-Zombie of tissue sections

A detailed protocol for AAV-Zombie on tissue sections is available on protocols.io (dx.doi.org/10.17504/protocols.io.14egn6k7yl5d/v1). AAV-Zombie was performed on tissue sections as described above for cultured cells, save for a few differences. Incubations in tissue were performed in a staining tray (Simport, M918), and fixation and washes were done in Coplin jars.

Sliced fresh tissue was first removed from -70 °C storage and allowed to warm to room temperature. Slides were then briefly washed with 1x PBS to remove O.C.T. compound, then fixed for 3 hr in MAA at -20 °C. Residual fixative was washed off with 1x PBS while the transcription mix was prepared. A total of 200 µL of transcription mix was used per slide, which was pipetted onto the slide and spread out with a clean glass coverslip. T7 RNA polymerase was used at a 1:10 dilution. As with cultured cells, *in situ* transcription was carried out at 37 °C for 3 hr.

For the HCR-FISH steps on tissue sections, we used 4 nM of each probe in an overnight 37 °C hybridization. The HCR hairpin concentration was also doubled to 60 nM.

Controls for AAV-Zombie and SpECTr

Guidelines for designing, imaging, and analyzing AAV-Zombie and SpECTr experiments are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjn72pgk5/v1). Both AAV-Zombie and SpECTr can produce signals due to hybridization of probes directly to single-stranded AAV genomes and/or transcriptional activity of the AAV ITRs producing barcoded transcripts (e.g., faint ‘concatemer’ signal in Genome A condition, Figure 3.2b). Thus, controls are necessary for setting thresholds for determining real vs. artifactual signal. A non-transduced/non-transfected control sample was used for all AAV-Zombie and SpECTr experiments. For SpECTr experiments, a barcode-only control was used to define signal from probe hybridizing to the AAV genome and/or barcoded transcripts produced due to transcriptional

activity of the ITR. As the transcriptional activity of the AAV ITR may differ between cell types, these control experiments were performed in each cell and tissue of interest, and processed side-by-side with experimental samples to mitigate assay-to-assay variability. Depending on the needs of the experiment, other controls may have been included and are outlined in the description of those experiments.

Imaging

For imaging, a Zeiss LSM 880 was used. Imaging of fluorescent protein expression was accomplished with a 10x, 0.45 NA air objective. Imaging of AAV-Zombie and SpECTr signal in Matrigel, cultured cells, and in tissue was performed with a 63x, 1.4 NA oil immersion objective. Imaging settings were chosen to capture full dynamic range of the signal without saturating pixels. When possible, laser power was adjusted before adjusting detector gain. Imaging settings were first optimized on control samples, before imaging of experimental samples. Fields of view were chosen while imaging non-experimental channels (e.g., Hoechst or Nissl).

Image analysis for AAV-Zombie and SpECTr

Guidelines for designing, imaging, and analyzing AAV-Zombie and SpECTr experiments are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjn72pgk5/v1).

For all cell and nuclear segmentation, Cellpose[192] (v3.0.7; https://www.cellpose.org/; RRID:SCR_021716) was used. Images were batch processed using napari[196] (v0.4.19.post1; https://napari.org/stable/; RRID:SCR_022765) and the serialcellpose plugin (v0.2.2; https://www.napari-hub.org/plugins/napari-serialcellpose). An Anaconda (v2.5.4; https://www.anaconda.com/) distribution of Python (v3.10.14; https://www.python.org/; RRID:SCR_008394) was used. For primary neurons and HEK293T cells, cell body masks were generated from poly(dT)-TAMRA signal and nuclear masks from Hoechst signal.

Quantification and measurement of AAV genomes and concatemers in PCs was accomplished using CellProfiler. For both HEK293T cells and primary neurons, masks were size filtered, using empirically determined thresholds. Primary neuron masks were further filtered for presence of an overlapping nuclear mask, and a cytoplasmic mask was generated by subtracting the nuclear mask from the cell body mask. Genome, concatemer, and capsid spots were identified within segmented nuclear and cytoplasmic masks, using empirically determined spot size thresholds and robust background intensity thresholding, chosen due to the sparse foreground signal. EGFP transcript intensity was measured in the entire cell body mask. For HEK293T cells, only nuclear AAV genomes and concatemers were measured.

Statistics and reproducibility

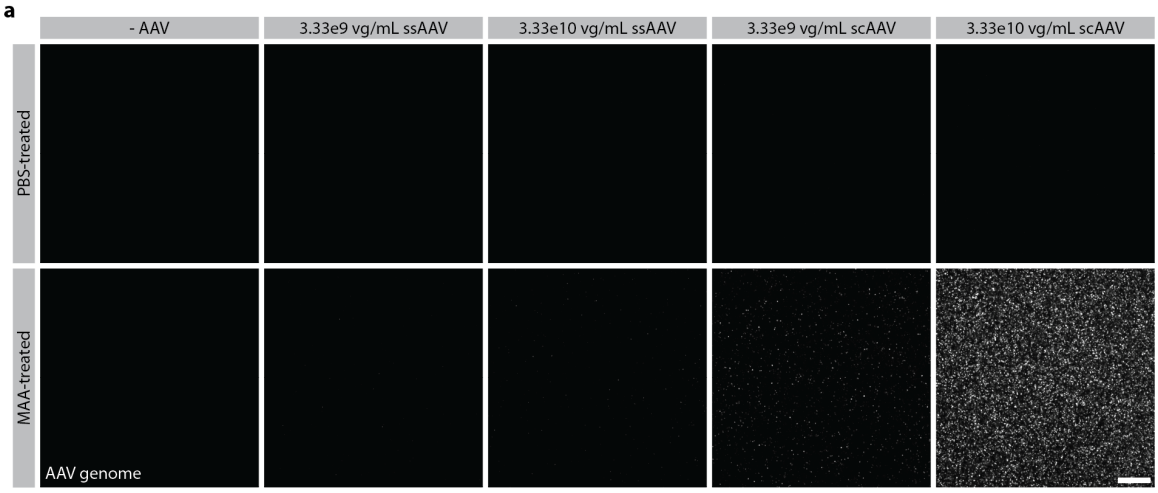
The number of biological replicates for each experiment are included in the corresponding figure legends. No data were excluded from analyses. For all violin plots, the middle dashed line is the median and the upper and lower dashed lines are quartiles. Statistical analysis was performed with GraphPad Prism (v10.0.3, GraphPad Software; RRID:SCR_002798) as described in figure legends. Where relevant, all tests were two-tailed and corrected for multiple comparisons to maintain an experiment-wide alpha of 0.05.

The following *in vivo* experiments were repeated once ($n > 2$ animals per experimental condition) with similar results: Figure 3.1d and Supplementary Figure 3.2d,e. All *in vitro* experiments were repeated at least twice with similar results.

Data availability

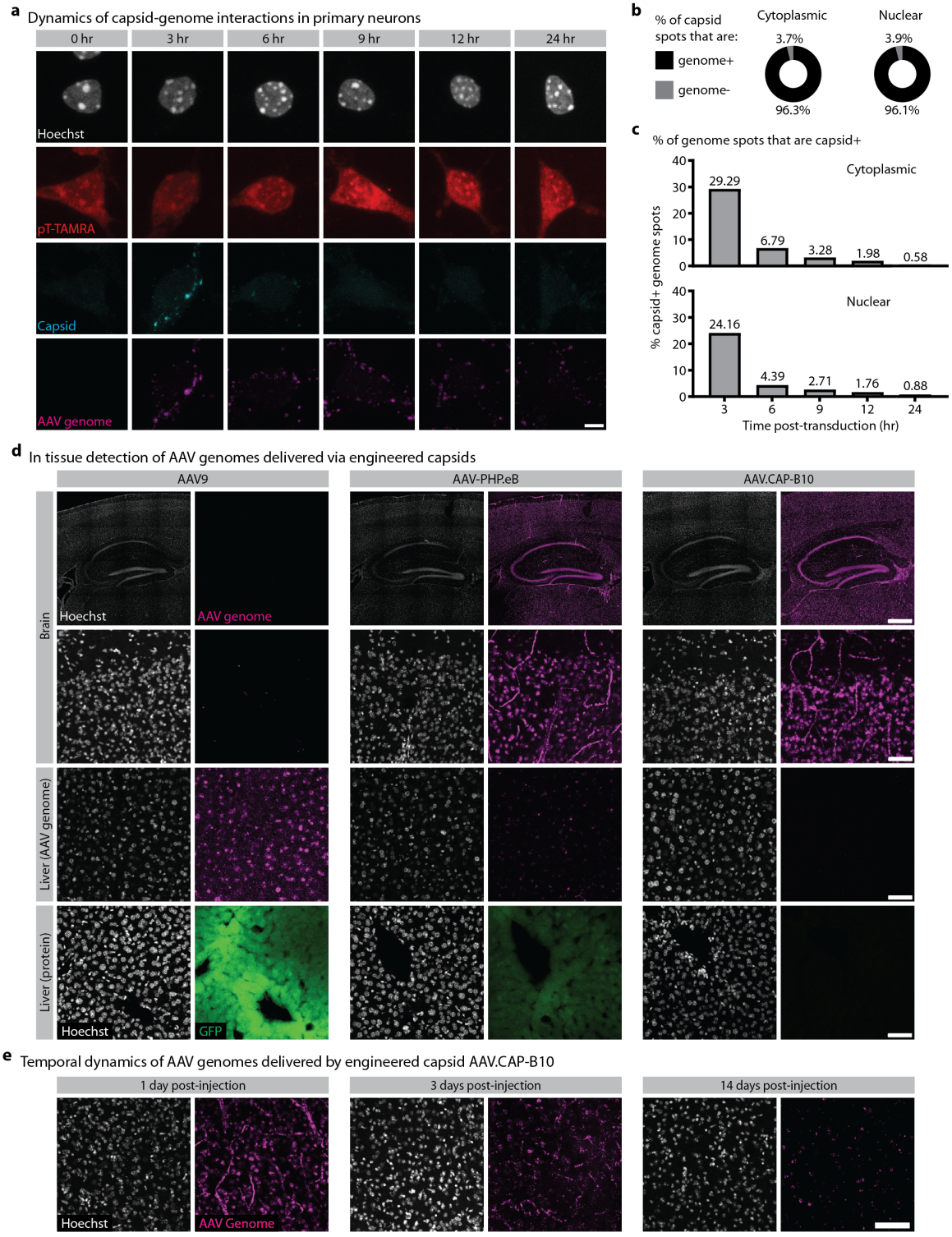
All sequences of primers, probes, sgRNAs, and other sequence elements are provided in Supplementary Table 3 of Coughlin et al., 2025. Images of brain tissue that are quantified in this work are deposited in the Brain Image Library (https://doi.org/10.35077/g.1163). Tabular datasets supporting conclusions of this work are available on Zenodo (10.5281/zenodo.13952929). All other data that support the findings of this study are available from the corresponding authors upon reasonable request.

3.6 Supplementary figures



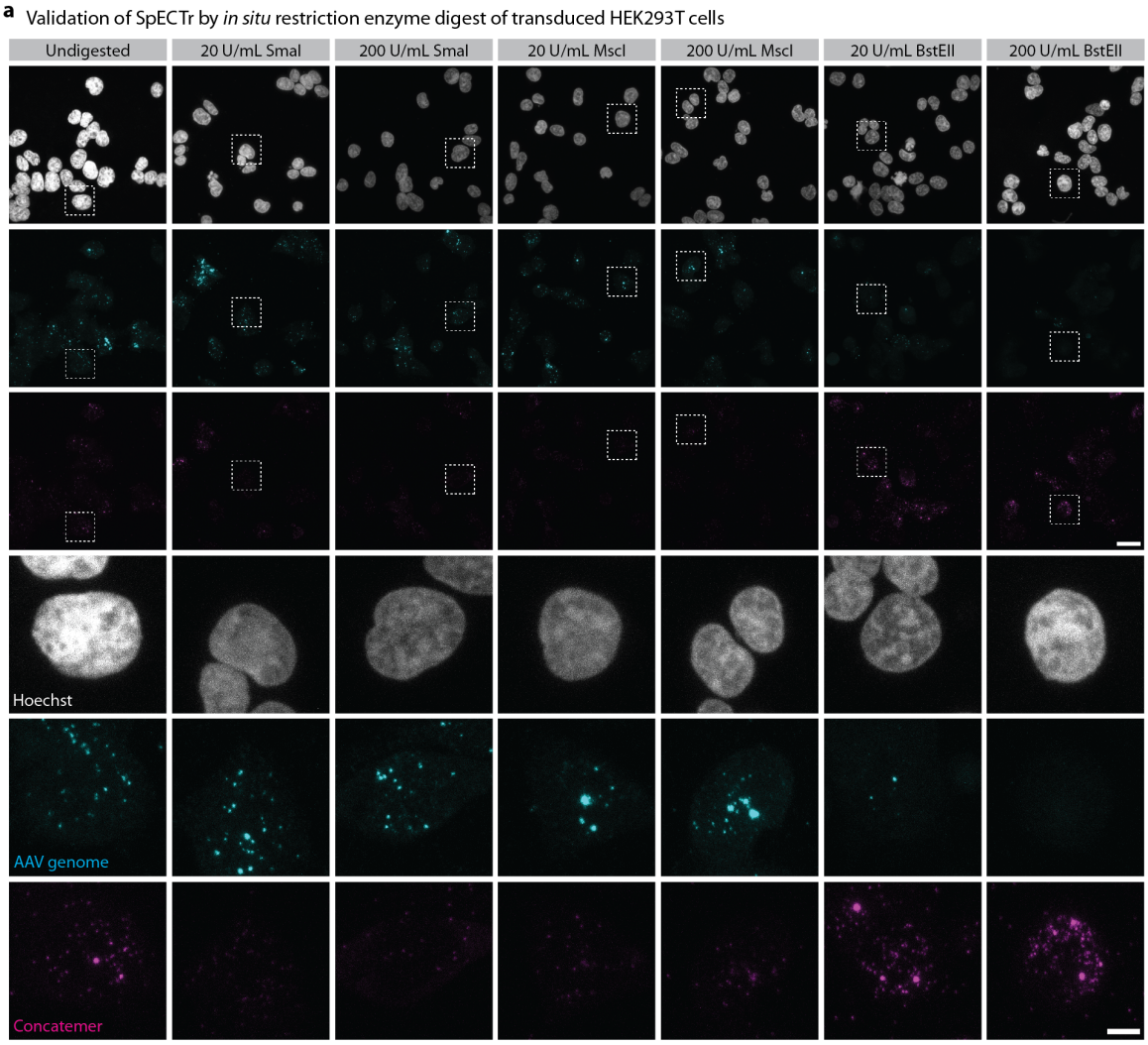
Supplementary Figure 3.1. Methanol and acetic acid fixation is sufficient to denature the AAV capsid, enabling genome detection by AAV-Zombie. To investigate whether processing of the AAV genome by the host cell was necessary for detection by AAV-Zombie, we devised a cell-free system. Single-stranded and self-complementary genomes were packaged into AAV-DJ, and

then embedded in high-concentration Matrigel. Following gelation, the samples were treated with ice-cold 1x PBS (top row) or MAA (bottom row). MAA treatment resulted in an increase in Zombie signal from scAAV samples, as compared to PBS controls, suggesting that MAA treatment can denature the capsid, releasing the AAV genome. Scale bar = 20 μm.

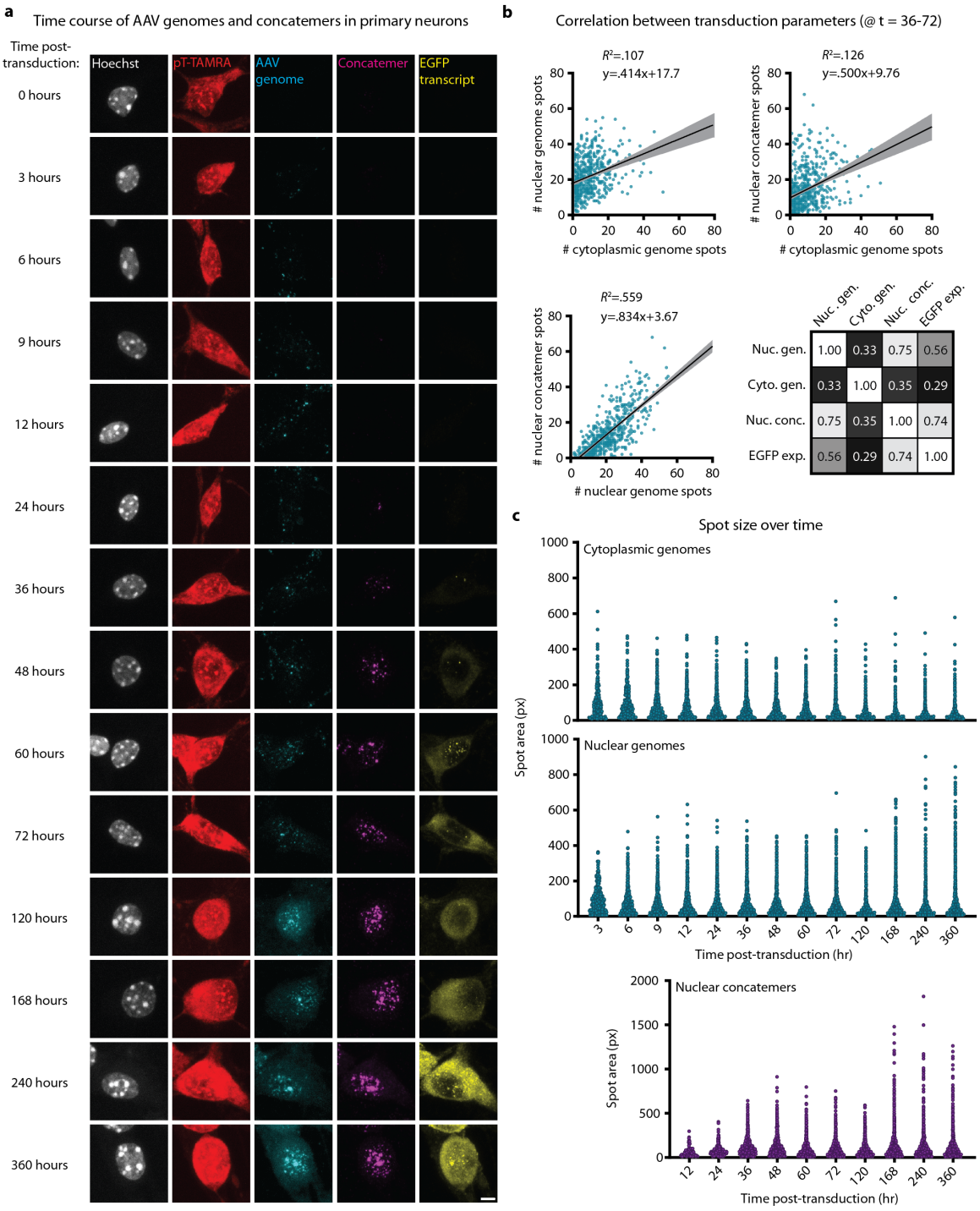


Supplementary Figure 3.2. Application of AAV-Zombie to understand AAV transduction *in vitro* and *in vivo*. a-c, Related to Figure 3.1c. **a**, Representative images of AAV capsids and scAAV genomes, over 24 hrs post-transduction. *n* = 243 (t = 0 hr), 191 (3 hr), 317 (6 hr), 212 (9 hr), 220 (12 hr), 255 (24 hr) neurons per time point. Scale bar = 5 μm. **b**, Percent of capsid puncta that overlap with a scAAV genome, quantified from capsid puncta identified at all time points.

The high overlap presumably reflects high encapsidation of the genome. *n* = 763 (nuclear), 1133 (cytoplasmic) capsid puncta. **c**, Percent of genome puncta that overlap with a capsid punctum, for cytoplasmic and nuclear fractions. The decrease in overlap over time reflects uncoating of the AAV genome and degradation of the capsid. For the cytoplasmic fraction, *n* = 3305 (t = 3 hr), 4022 (6 hr), 4201 (9 hr), 5092 (12 hr), 3593 (24 hr) AAV genomes. For the nuclear fraction, *n* = 2421 (t = 3 hr), 3987 (6 hr), 2512 (9 hr), 3469 (12 hr), 2279 (24 hr) AAV genomes. **d**, Related to Figure 3.1d. Top 3 rows: AAV-Zombie detection of scAAV genomes in C57BL/6J brain and liver, following delivery by engineered capsids AAV-PHP.eB and AAV.CAP-B10 and compared to parent capsid AAV9. Tissue was collected 1 day post-injection. Bottom row: EGFP protein in liver, following 2 weeks of expression. Data shows that reduced liver protein expression with AAV-PHP.eB and AAV.CAP-B10 is due to reduced DNA-level transduction, rather than a transcriptional or post-transcriptional difference. Genomes were delivered at 3e11 vg dose. Representative images from *n* = 3 animals per condition. Scale bars = 500 μm for top row, 50 μm for rest. **e,** Time course of AAV.CAP-B10-delivered scAAV genomes in mouse cortex, showing decreasing vascular signal over time. Genomes were delivered at 3e11 vg dose. Representative images from *n* = 3 animals per condition. Scale bar = 500 μm.

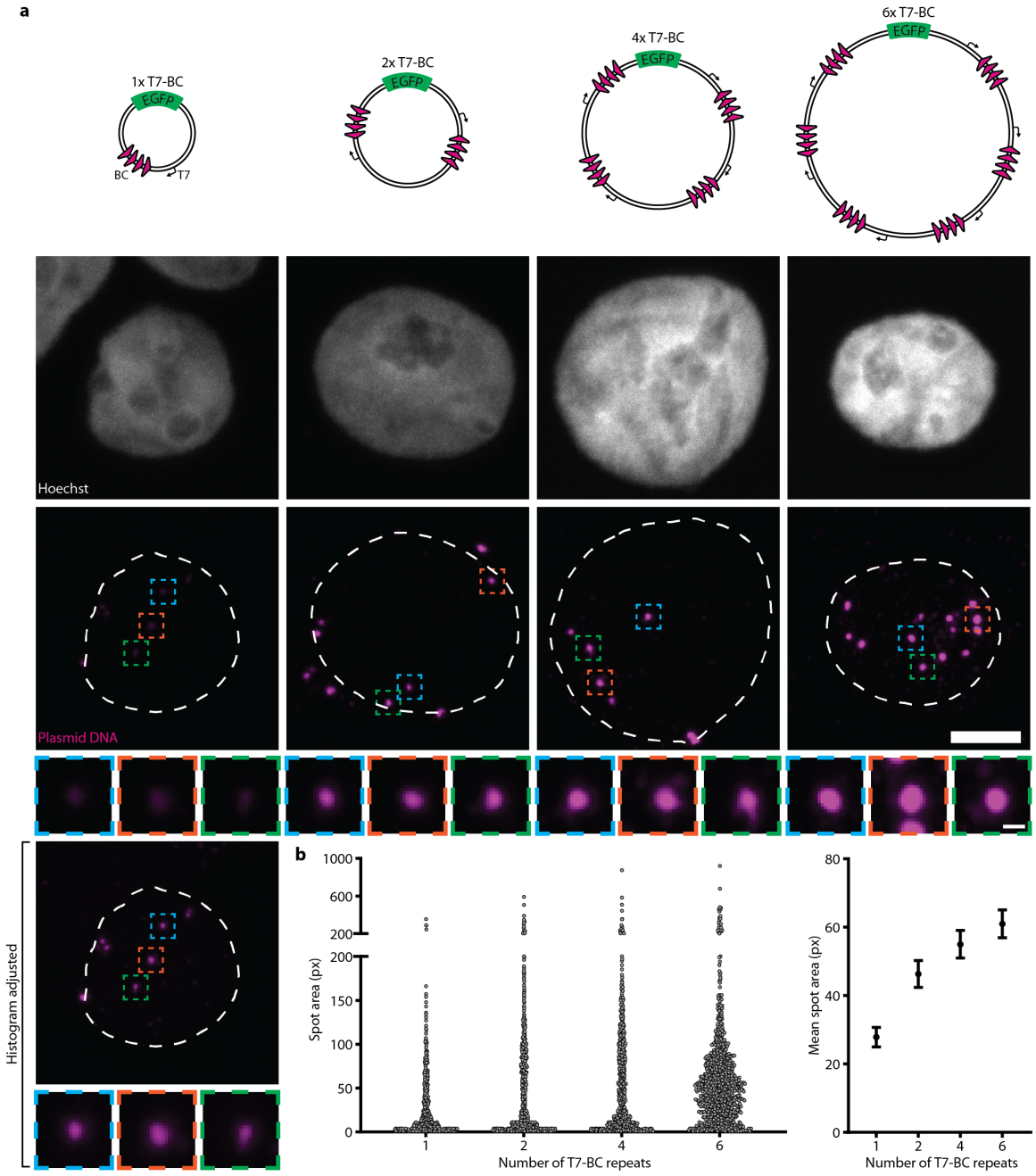


Supplementary Figure 3.3. Validation of SpECTr by *in situ* restriction enzyme digest. Related to Figure 3.2c-e. Representative images from *in situ* digests of HEK293T cells transduced with SpECTr genomes. "Undig": undigested condition in which fixed cells were incubated at 37 °C in restriction enzyme buffer without any restriction enzyme. Genomes delivered at 1e6 MOI in AAV-DJ. Scale bars = 20 μm for top 3 rows, 5 μm for rest.



Supplementary Figure 3.4. Time course of AAV transduction, concatemerization, and expression in primary neurons. a-c, Related to Figure 3.2f,g. **a**, Representative images from time course of AAV transduction, concatemer formation, and EGFP reporter transcription in primary neurons. TAMRA-conjugated polyT probe (pT-TAMRA) was used to label cell bodies. Scale bar = 5 µm. **b**, Linear correlations between cytoplasmic AAV genomes, nuclear AAV genomes, and nuclear concatemers, and summary of correlation coefficients for all correlations measured. *n* = 616 primary neurons, pooled from t = 36, 48, 60, and 72 hr time points (time

points chosen for detectable EGFP transcript that has not yet reached a plateau). Shaded area is 95% confidence interval. **c**, Distribution of spot sizes for cytoplasmic genomes (top), nuclear genomes (middle), and nuclear concatemers (bottom) over time. *n* = 476 - 2098 (cytoplasmic genomes), 657-4078 (nuclear genomes), 111-4226 (nuclear concatemers) spots per time point.



Supplementary Figure 3.5. Relationship between number of T7-barcode repeats and spot area. a, Plasmids with increasing numbers of T7-barcode (T7-BC) repeats were constructed and transfected at equimolar amounts into HEK293T cells. Zombie was then used to detect the individual plasmids. Scale bar = 20 μm for larger images, 2 μm for insets. **b,** Quantification of spot area as a function of number of T7-BC repeats. Plasmids with more T7-BC repeats yielded larger spots. *n* = 706 (1 T7-BC repeat), 932 (2 T7-BC repeats), 1052 (4 T7-BC repeats), 999 (6 T7-BC repeats) spots per condition. Points represent mean and bars are 95% confidence interval.

Chapter 4

TRANSCRIPTIONAL CROSSTALK BETWEEN AAV GENOMES IS DEPENDENT UPON CONCATEMER FORMATION

Adapted from:

Coughlin, G.M.[#], Borsos, M.[#], Barcelona, B.H.[$], Appling, N.[$], Mayfield, A.M.H., Mackey, E.D., Eser, R.A., Jackson, C.R., Chen, X., Ravindra Kumar, S., and Gradinaru, V.* (2025). Spatial genomics of AAV vectors reveals mechanism of transcriptional crosstalk that enables targeted delivery of large genetic cargo. Nat. Biotechnol. https://doi.org/10.1038/s41587-025-02565-4

4.1 Summary

Integration of cell type-specific enhancers into AAV vectors can refine transduction to specific cell classes. This approach pairs well with engineering of novel AAV capsids that can provide easy access to target tissues from minimally invasive delivery routes. Indeed, the broad access to the mouse brain afforded by the engineered capsid AAV-PHP.eB[78] has enabled multiple groups to better screen for AAV-compatible enhancers with activity in the mouse brain. This general strategy has been used by multiple groups to identify AAV-compatible enhancers that can boost transgene expression in specific cell types[121–124,127–131,197]. Multiplexed screening, in which multiple regulatory elements are characterized in the same animal, can be used to reduce animal numbers and inter-animal variability and to increase throughput. Likewise, targeting of multiple cell classes with different effectors may provide new modalities for gene therapy.

Here we describe a confound to simultaneous delivery of multiple enhancer-driven AAVs. Transcriptional crosstalk arises when an enhancer element on one AAV genome interacts with and alters expression from the promoter on a co-delivered AAV genome. Here, we explore and mechanistically dissect this behaviour. We identify and profile transcriptional crosstalk across multiple enhancer and promoter sequences and occurring in multiple central and peripheral tissues. Using the novel spatial genomics methods developed in Chapter 3, we mechanistically link transcriptional crosstalk to formation of AAV concatemers. These results support a model in which concatemerization of codelivered AAV genomes places elements delivered in *trans* into a *cis* conformation, facilitating interaction of enhancers and promoters from separate AAV genomes. These results both highlight important confounds for pooled characterization of AAV-delivered enhancers and multiclass targeting, and suggest strategies to mitigate such confounds.

4.2 Crosstalk between regulatory elements of separate AAV genomes

Transcriptional crosstalk between AAV genomes can occur when regulatory elements in one genome interact with those of another. The Ple155 element[104] drives strong expression in mouse cerebellar Purkinje cells (PCs) following systemic delivery via a blood-brain barrier-penetrant engineered AAV (AAV-PHP.eB[175]). Conversely, the mDLX enhancer[115] paired with a minimal beta-globin promoter (mDLX-minBG) directs expression to forebrain interneurons, but not PCs. However, following co-transduction of these viruses, we observed strong expression of the mDLX-minBG-driven transgene in PCs (Figure 4.1a,b and Supplementary Figure 4.1a). This result suggests that elements in the Ple155 sequence can interact with elements in the mDLX-minBG genome and increase expression of the latter in a cell type-specific manner.

To identify which elements in the mDLX-minBG sequence are necessary for this crosstalk, we serially truncated the mDLX-minBG genome (Figure 4.1c,d and Supplementary Figure 4.1b,c). Removal of the mDLX enhancer did not produce a detectable effect on crosstalk (truncation Δ_i), whereas removal of the minBG promoter decreased both the percent of mRuby2-positive PCs and the PC mRuby2 intensity (truncations Δ_{ii} and Δ_{iii}). These data point to a model in which elements in the Ple155 interact with the minBG promoter, reminiscent of the classical description of enhancer-promoter interaction[108,109].

Given this model for transcriptional crosstalk, we expect to observe this behaviour with multiple promoter and enhancer sequences. Indeed, we observed robust crosstalk between the Ple155 element and 3 commonly used minimal promoters: Ef1s, the CMV promoter, and the super core promoter 1 (SCP1)[100] (Figure 4.1e and Supplementary Figure 4.2). Furthermore, we screened a panel of 9 characterized cortical enhancer sequences[115,121,122], using a minBG-driven mRuby2 crosstalk reporter virus (Figure 4.1f and Supplementary Figure 4.3a-c). In all 9 cases, presence of the enhancer resulted in an increase in expression from the reporter genome delivered in *trans*, when compared to a 'no enhancer' condition (Figure 4.1f).

To further demonstrate the generalized nature of transcriptional crosstalk, we used the ubiquitous cytomegalovirus immediate-early enhancer[94] (CMVe) and SCP1 in combination with a cocktail of blood-brain barrier-penetrant and peripheral nervous system-tropic engineered AAV capsids (AAV-PHP.eB and AAV-MaCPNS2[80]), to provide broad central and peripheral nervous system coverage (Supplementary Figure 4.3d,e). We observed increased tdTomato crosstalk reporter expression in cerebellum, proximal colon, dorsal root ganglia, and liver with an enhancer delivered in *trans* versus the 'no enhancer' condition.

These results support a generalized model for transcriptional crosstalk, in which enhancer elements on one AAV genome can interact with and drive expression from a promoter on another AAV genome. As this interaction is more likely to occur between elements in *cis,* we and others[198] propose that concatemerization of AAVs could enable transcriptional crosstalk, by placing elements delivered in *trans* into a *cis* conformation (Figure 4.1g).

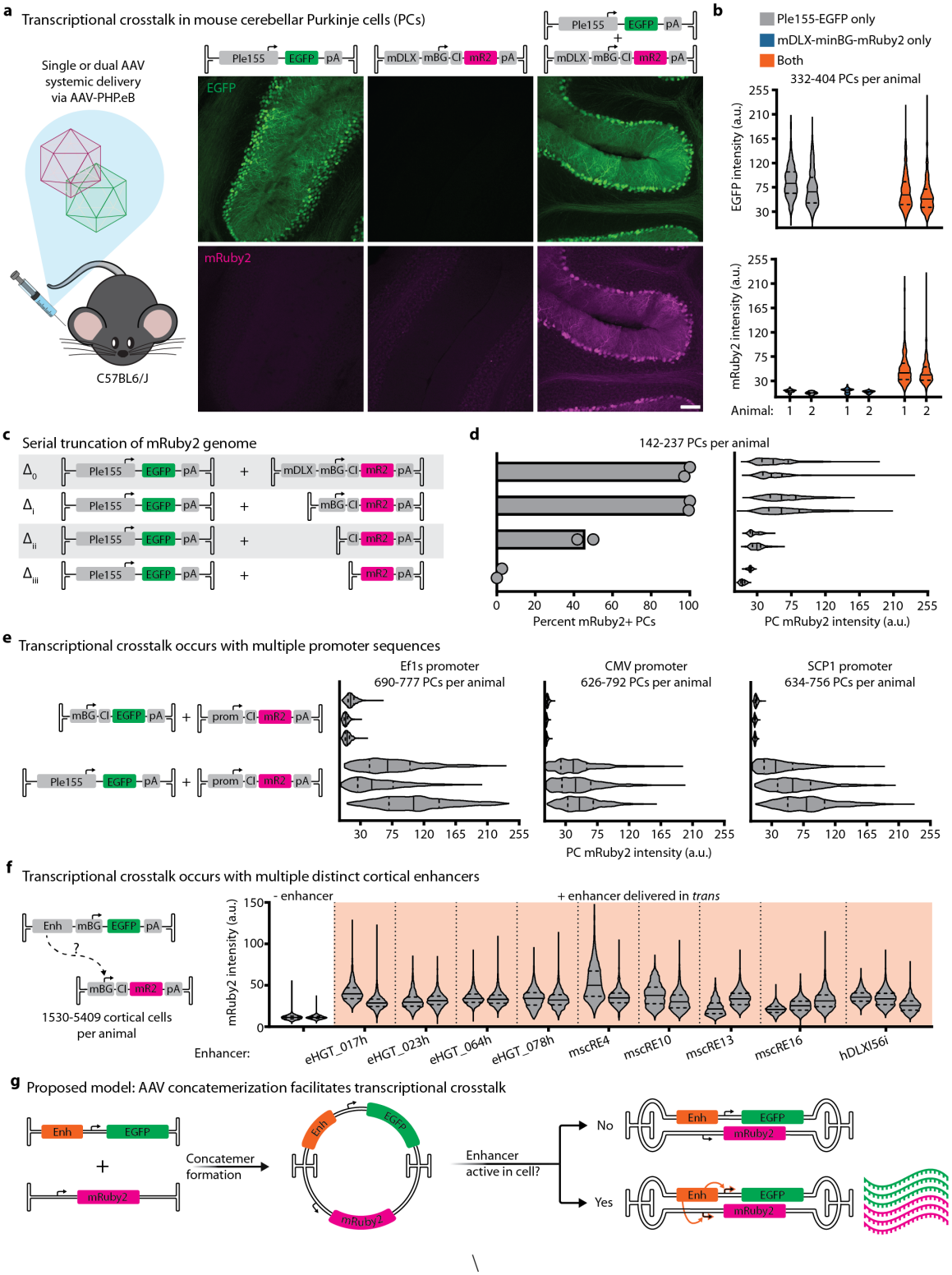


Figure 4.1. Broad transcriptional crosstalk between enhancers and promoters delivered in separate AAV genomes. a, Transcriptional crosstalk. Left column: when injected alone, the AAV-delivered Ple155 element directs strong expression to cerebellar Purkinje cells (PCs). Middle: AAV-delivered mDLX-minBG-driven mRuby2 does not yield any detectable PC transduction.

Right: co-administration of both AAVs results in unexpected mRuby2 expression in PCs. All genomes delivered at 1e12 vg dose in AAV-PHP.eB. Scale bar = 100 μm. **b**, Distribution of PC cell body EGFP (top) and mRuby2 (bottom) intensities from animals shown in (**a)**. *n* = 2 animals per condition. **c**, Schematic of serially-truncated mDLX-minBG-mRuby2 constructs, coinjected with Ple155-EGFP, to assess necessity of elements for transcriptional crosstalk. **d**, Quantification of results for truncation conditions shown in (**c**), quantified as percent of PCs positive for mRuby2 (left) and PC mRuby2 fluorescence intensity (right). Bars represent mean. *n* = 2 animals per condition. All genomes delivered at 5e11 vg dose in AAV-PHP.eB. **e,** Transcriptional crosstalk between Ple155 and three commonly used minimal promoters (Ef1s, CMV promoter, and SCP1). *n* = 3 animals per condition. All genomes delivered at 1e12 vg dose in AAV-PHP.eB. **f**, Screen of 9 cortical enhancers for ability to upregulate expression of minBG promoter-driven mRuby2 delivered in *trans*. *n* = 2 animals per condition, except mscRE16 and hDLXI56i, in which *n* = 3. All genomes delivered at 1e12 vg dose in AAV-PHP.eB. **g**, Proposed model for transcriptional crosstalk. Formation of concatemeric episomes places enhancer and promoter elements that were delivered in *trans* into a *cis* conformation. This concatemerization facilitates interaction of the enhancer with the promoter that was delivered in *trans*, resulting in increased expression in cells where the enhancer is active. Each violin plot represents data from one animal. CI = chimeric intron.

4.3 Reducing AAV concatemer formation decreases crosstalk

Using SpECTr to visualize AAV concatemers, we next explored the mechanistic connection between concatemerization and transcriptional crosstalk. If concatemerization of AAV genomes enables transcriptional crosstalk (Figure 4.1g), then we expect reductions in concatemer formation to reduce transcriptional crosstalk. We first tested this hypothesis in HEK293T cells with the ubiquitous CMVe, comparing AAV-DJ transduction to plasmid transfection (Figure 4.2a-e). As expected, transcriptional crosstalk was apparent following cotransduction by AAVs, but not after cotransfection of the corresponding genome plasmids. Transfection of a “plasmid concatemer,” consisting of the entire tdTomato-containing genome inserted outside the ITRs of the TagBFP-containing genome plasmid, recapitulated the co-transduction result. These data suggest that transcriptional crosstalk occurs when the enhancer and promoter are in a *cis* conformation.

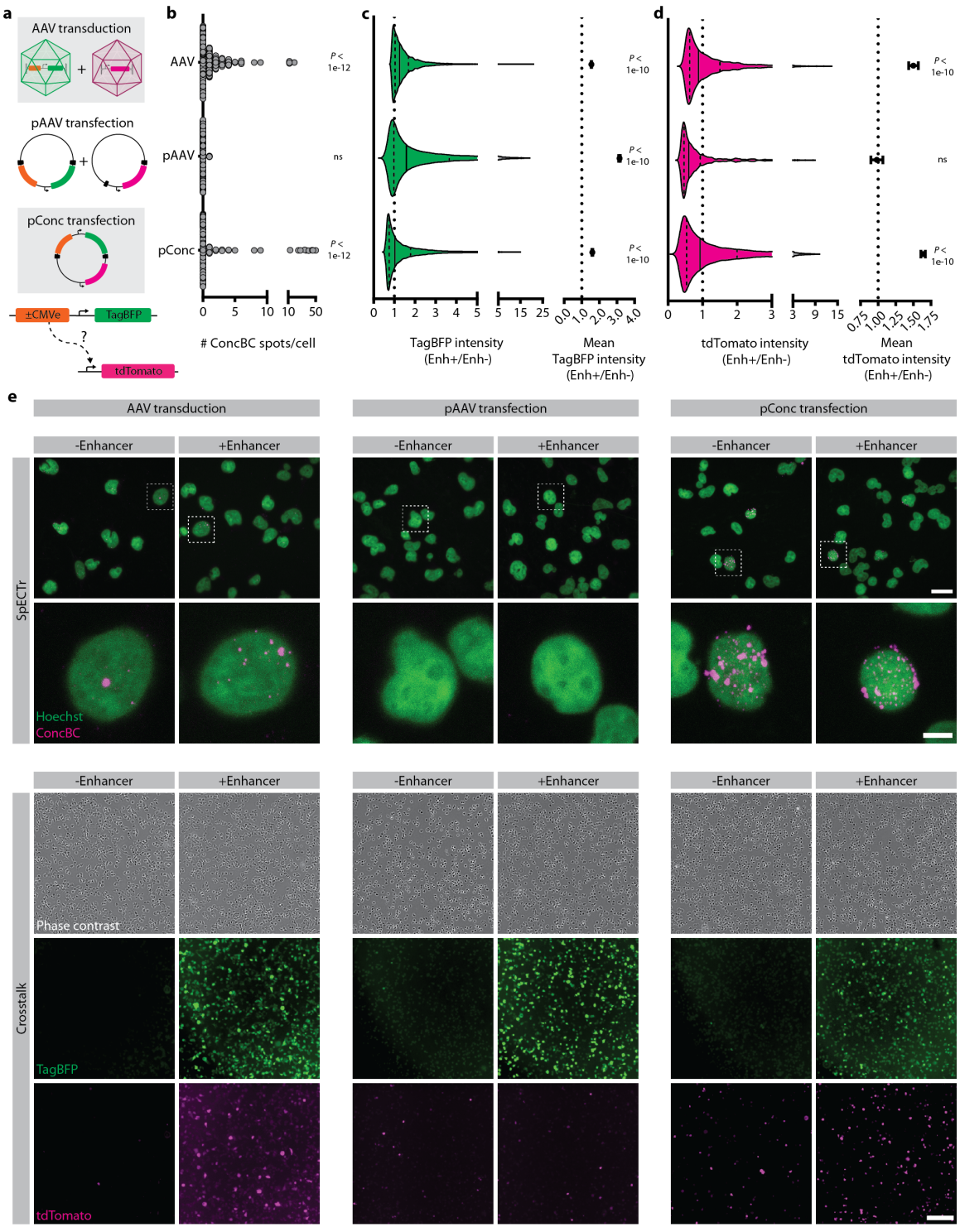


Figure 4.2. *In vitro* exploration of transcriptional crosstalk mechanisms. a, Schematic of experiment. HEK293T cells were either transduced with two cross-talking AAV genomes (top), transfected with two packaging plasmids (middle), or transfected with a single plasmid concatemer (pConc) containing both genomes in *cis* (bottom). For each delivery method, a '+ CMV enhancer' and a 'no enhancer' condition were tested. Genomes also contained SpECTr elements. **b**, Quantification of number of ConcBC spots per cell. Statistical significance for

SpECTr signal was determined using a Wilcoxon signed rank test, against the null hypothesis that spot count = 0. *n* = 190 (AAV transduction), 220 (pAAV transfection), 227 (pConc transfection) HEK293T cells per condition. **c**,**d**, TagBFP intensity (**c**), and tdTomato crosstalk reporter intensity (**d**), represented as distribution of cell intensities (violin plots) as well as mean ± 95% confidence interval. Fluorescent protein intensity is normalized to the mean of the no-enhancer condition. In all conditions, presence of the enhancer increased expression of TagBFP delivered in *cis*. However, presence of the enhancer increased expression of the tdTomato crosstalk reporter only in the AAV transduction and pConc transfection conditions. Statistical significance for fluorescent protein intensity was determined using one sample t-tests, against the null hypothesis that normalized intensity = 1 (shown by dashed line). For TagBFP: *n* = 14334 (AAV transduction), 18773 (pAAV transfection), 11606 (pConc transfection) TagBFP-positive HEK293T cells per condition. For tdTomato: *n* = 2675 (AAV transduction), 696 (pAAV transfection), 11613 (pConc transfection) tdTomato-positive HEK293T cells per condition. ns = not significant. **e**, Representative images for data quantified in (**b-d**), showing SpECTr signal (upper panel) and reporter fluorescence (lower panel). Scale bars = 20 μm for top row, 5 μm for second row, and 100 μm for rest.

We next tested this hypothesis *in vivo.* Previous research has implicated DNA repair pathways in recognizing and processing free ITR ends, resulting in formation of concatemeric AAV episomes[52–57,199]. In particular, *Prkdc*$^{scid/scid}$ mice (hereafter referred to as SCID mice), which have a loss of function in the DNA double-strand break repair enzyme Prkdc, show reduced concatemer formation in bulk muscle[53] and liver[54], and lower expression from concatemerization-dependent AAVs[52]. However, neither concatemer formation at a single-cell level nor transcriptional crosstalk have been explored in SCID mice.

We first validated SpECTr for detection of AAV concatemers in tissues, including cortex, liver, and cerebellum (Supplementary Figure 4.4). Then, to enable paired measurement of concatemer formation and transcriptional crosstalk in the same animals, we integrated SpECTr components into the Ple155 and mDLX-minBG AAV genome pair. Reasoning that the high doses of AAVs we used previously (1e12 vg, Figure 4.1a) may yield many large indistinguishable spots and thus confound accurate measurement of concatemers, we injected 3e11 total vg into C57BL/6J-background SCID mice and wildtype C57BL/6J controls. Even at this reduced dose, transcriptional crosstalk was readily apparent in PCs of wildtype animals; transduction with both genomes resulted in significantly more mRuby2-positive PCs and a significant increase in PC mRuby2 intensity, compared to either single transduction condition (Figure 4.3a,b and Supplementary Figure 4.4b). These effects were not observed in SCID mice.

To determine whether SCID PCs were deficient in AAV concatemerization, we applied AAV-Zombie and SpECTr to cerebellum sections from the same animals. PCs were identified using HCR-FISH against *Itpr1* transcript[200]. With AAV-Zombie, we measured a 2.6-fold higher

AAV genome count in SCID than wildtype PCs (Supplementary Figure 4.4c,d). In a separate cohort of mice, we similarly observed significantly higher DNA-level transduction of SCID brains by AAV-PHP.eB, with no significant differences in protein-level transduction (Supplementary Figure 4.5a-c). Despite higher DNA-level transduction of SCID brains, SpECTr revealed significantly fewer and smaller ConcBC spots in PCs of SCID mice than wildtype controls (Figure 4.3c,d and Supplementary Figure 4.4e), indicating reduced concatemer formation in the absence of functional Prkdc.

Finally, we assessed whether the SCID mutation would affect crosstalk of other enhancers as well. We chose two additional enhancers, targeting GABAergic interneurons (hDLXI56i)[115] and layer 5 pyramidal tract excitatory neurons (mscRE4)[121], and coinjected these with an mRuby2 crosstalk reporter (Figure 4.3e-h and Supplementary Figure 4.4f,g). Consistent with our observations from the Ple155 and mDLX-minBG pair, we observed reduced transcriptional crosstalk with both enhancers in SCID mice, quantified by both number of mRuby2-positive cells per mm^3 and fluorescence intensity of mRuby2-positive cells. We did not detect any difference in transduction between genotypes, as assessed by number and intensity of EGFP-positive cells. Taken together, these *in vitro* and *in vivo* results strongly suggest that AAV concatemer formation enables transcriptional crosstalk.

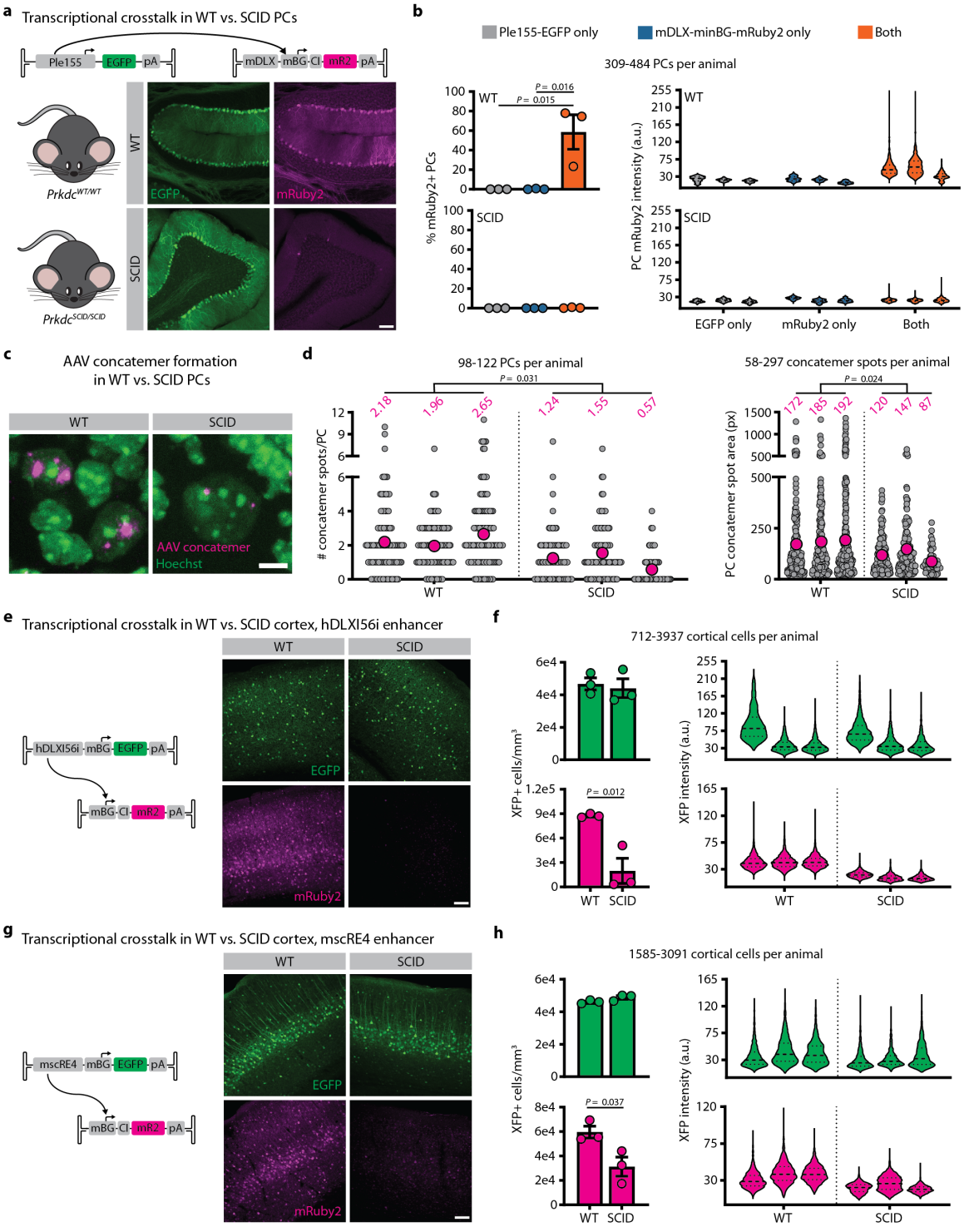


Figure 4.3. Figure 4. Reducing AAV concatemer formation decreases transcriptional crosstalk between AAV genomes. **a**, Representative images of transcriptional crosstalk between Ple155 and minBG promoter, in dual-injected wildtype and SCID mouse Purkinje cells (PCs). Both genomes delivered at 3e11 vg dose in AAV-PHP.eB. Scale bar = 100 µm. **b**, Quantification of transcriptional crosstalk shown in (**a**), comparing single injection conditions to dual injection condition, and measured as percent of PCs positive for mRuby2 (left) and PC mRuby2

fluorescence intensity (right). Statistical significance was determined using one-way ANOVA (P = 0.010) and Tukey's multiple comparison test. n = 3 animals per condition. **c**, Representative images of AAV concatemers detected with SpECTr in PCs of dual-injected wildtype and SCID animals shown in (**a**). Scale bar = 5 μm. **d**, Quantification of PC concatemer spot count (left) and spot size (right), in wildtype and SCID PCs. Each grey dot corresponds to a single PC (left) or a single concatemer spot (right). Magenta dot and number indicate mean of animal. n = 3 animals per condition. **e-h**, Representative images and quantification of reduced transcriptional crosstalk in SCID animals with the GABAergic interneuron enhancer hDLXI56i (**e,f**) and the layer 5 pyramidal tract neuron enhancer mscRE4 (**g**,**h**). Fluorescent protein (XFP) signal was amplified through IHC. Quantification is presented as number of XFP-positive cells per mm^3 and XFP fluorescence intensity. n = 3 animals per condition. Scale bars = 100 μm. Bars in (**b**), (**f**), and (**h**) represent mean ± s.e.m. Statistical significance in (**d**), (**f**), and (**h**) was determined using unpaired t-tests.

4.4 Discussion

The mammalian genome contains a vast diversity of regulatory elements, including hundreds of thousands of putative enhancer sequences[108,109]. Multiplexed profiling using systemically administered AAVs is an attractive strategy for exploring the diversity of enhancer sequences, with the goal of identifying AAV-compatible enhancers that can confer cell type specificity in a range of animal models and in healthy and diseased humans. Enhancer sequences identified in mouse may not function the same in other species[142]. Thus, identifying cell type-specific enhancers in other animal models, including non-human primates, will likely require such multiplexed characterization platforms. However, multiplexed enhancer characterization is complicated by transcriptional crosstalk between enhancer and promoter elements on separate AAV genomes.

In exploring this transcriptional crosstalk, we demonstrated generalizability across promoter and enhancer elements as well as tissue and cell types. Furthermore, using SCID mice and our novel single molecule AAV concatemer detection method SpECTr, we mechanistically linked AAV crosstalk with concatemer formation. This generalizability suggests that AAV transcriptional crosstalk is likely to be a source of noise in multiplexed enhancer screening and confound for simultaneous targeting of multiple cell types. Some groups have used pooled screening of AAV-compatible enhancers[120,125,197,201], and others have observed unexpected noise resulting from AAV crosstalk in pooled enhancer screens[130,197].

Methods to reduce or eliminate AAV crosstalk would be beneficial for multiplexed enhancer screening. Our results with SCID mice suggest one method to reduce crosstalk, and indeed, Hunker and colleagues confirmed that crosstalk could be mitigated using *Prkdc*$^{scid/scid}$ mice[202]. However, the success of this strategy is likely to depend on cell type(s), delivery method,

and dosage, among other factors. Differences in the effect of Prkdc loss-of-function between cell types may reflect differences in dominant DNA repair pathways. Further understanding of transcriptional crosstalk may suggest other strategies to mitigate this confound, including the use of insulator elements[203–205].

By enabling separation of enhancers and coding sequences onto separate AAV genomes, transcriptional crosstalk may also represent an opportunity for cell type-specific expression of large coding sequences, such as genome editing and modulation machinery. This is particularly exciting given the diversity of CRISPR-based tools that can disrupt genes, change, add or remove sequence, or activate or repress transcription[153–159]. Though dual vector strategies based in mRNA or protein *trans*-splicing have been used for AAV delivery of CRISPR-based tools[27], these methods are often inefficient and require some engineering to determine optimal points to split transcripts or proteins. Thus, further exploration of transcriptional crosstalk as a novel strategy for cell type-targeted genome modulation is warranted.

4.5 Methods

Key resources

Key resources for this work are listed in Supplementary Table 2 of Coughlin et al., 2025.

Plasmid DNA

Standard molecular cloning techniques were used to generate DNA constructs in this study. Double-stranded DNA was synthesized by Integrated DNA Technologies and inserted into pAAV backbones with NEBuilder HIFI (New England Biolabs, E2621). Sequences of utilized DNA elements (e.g., promoters and enhancers) are provided in Supplementary Table 3 of Coughlin et al., 2025.

pUCmini-iCAP-AAV-PHP.eB[175] (Addgene #103005; RRID:Addgene_103005), pUCmini-iCAP-AAV.MaCPNS2[80] (Addgene #185137;RRID:Addgene_185137), AAV-DJ rep-cap (Cell Biolabs, VPK-420-DJ), and pHelper (Agilent, #240071) plasmids were used for production of AAVs. Prior to use, all plasmids were sequence verified via whole-plasmid sequencing through Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation.

AAV production

Detailed protocols for AAV production and titration are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjnew5gk5/v1 and dx.doi.org/10.17504/protocols.io.e6nvw1n47lmk/v1). AAVs were produced and purified

according to published methods[191], with some minor alterations. Briefly, HEK293T cells (ATCC, CRL-3216; RRID:CVCL_0063) were triple transfected with PEI-MAX (Polysciences, #24765) to deliver the rep-cap or iCAP, pHelper, and genome packaging plasmids. Viruses were harvested from cells and media, then purified over 15%, 25%, 40%, and 60% iodixanol (OptiPrep, Serumwerk, #1893) step gradients. A Type 70 Ti fixed-angle titanium rotor (Beckman Coulter, #337922) at 58.4k rpm for 1.5 hr, or a Type 70.1 Ti fixed-angle titanium rotor (Beckman Coulter, #342184) at 61.7k rpm for 1.25 hr was used, depending on the scale and number of AAVs to be purified simultaneously. Viruses were concentrated using Amicon Ultra-15 or Amicon Ultra-4 filters with a 100 kD size cutoff (MilliporeSigma, UFC9100 and UFC8100) and formulated in sterile DPBS (ThermoFisher, #14190144) with 0.001% Pluronic F-68 (ThermoFisher, #24040032). AAVs were titered with qPCR by measuring the number of DNase I-resistant viral genomes, relative to a linearized genome plasmid standard. Prior to injection, AAVs were diluted in sterile saline. ssAAV genomes were used for all experiments. The following qPCR primers against the W3 sequence were used for titering AAV viruses:

Forward: 5'-TGGTATTCTTAACTATGTTGCTCCT-3'
Reverse: 5'-AAGCCATACGGGAAGCAATAG-3'

Tissue culture

For AAV production, and for some *in vitro* experiments, HEK293T cells were used (ATCC, CRL-3216; RRID:CVCL_0063). Cells were grown in DMEM (ThermoFisher, #10569010) supplemented with 10% defined FBS (Cytiva, SH30070.03).

For small-scale HEK293T experiments, cells were seeded at optimal confluence (50% for transduction, 90% for transfection) in the morning, and transfected or transduced in the afternoon. For transfection, Lipofectamine LTX (ThermoFisher, #15338100) was used, with 500 ng total of DNA and 3 μL of transfection reagent. To avoid saturating SpECTr or fluorescent protein signal, 50 ng of DNA (for Supplementary Figure 4.2) was used, with pUC19 (New England Biolabs, N3041S; RRID:Addgene_50005) used as filler to ensure efficient transfection. For investigation of transcriptional crosstalk with transfection and transduction *in vitro* (Supplementary Figure 4.2), we transduced cells with a 1e5 multiplicity of infection (MOI) of AAV-DJ and cells were collected 5 days later. On the morning of collection, we passaged cells 1:10 onto poly-D-lysine coated coverslips (Neuvitro, GG-12-1.5h-PDL). Once HEK293T cells had attached, the coverslips were washed three times in DPBS and then fixed. For analysis of fluorescent protein expression, cells were fixed with ice-cold 4% paraformaldehyde (PFA, Electron Microscopy

Sciences, #15714-S) in 1x PBS for 15 min at 4 °C and stored in 1x PBS at 4 °C until use. For AAV-Zombie or SpECTr, cells were fixed with ice-cold 3:1 methanol:acetic acid (MAA, Sigma-Aldrich, #322415 and A6283) for 15 min at -20°C, then stored at -20 °C in 70% ethanol until use.

Animals

Animal husbandry and all procedures involving animals were performed in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and approved by the Institutional Animal Care and Use Committee (IACUC) and by the Office of Laboratory Animal Resources at the California Institute of Technology.

8-week old, male C57BL/6J (strain #: 000664; RRID:IMSR_JAX:000664) and C57BL/6J-background *Prkdc*$^{scid/scid}$ (strain #: 001913; RRID:IMSR_JAX:001913) mice were obtained from the Jackson Laboratory. Mice were housed 3-4 per cage, on a 12 hr light/dark cycle, and had *ad libitum* access to food and water. For animal experiments, mice were 8.5 to 9.5 weeks old at the time of injection.

Retro-orbital injection

A detailed protocol for systemic AAV administration through retro-orbital injection is available on protocols.io (dx.doi.org/10.17504/protocols.io.36wgqnw73gk5/v1). AAVs were administered via retro-orbital injection during isoflurane anesthesia (1-3% in 95% O_2/5% CO_2, provided by nose cone at 1 L/min), followed by administration of 1-2 drops of 0.5% proparacaine to the corneal surface[191].

Tissue harvest and processing

Tissue was collected 4 weeks post-AAV administration. Animals were euthanized via i.p. injection of 100 mg/kg euthasol.

For AAV-Zombie or SpECTr in tissue, animals were transcardially perfused with 30 mL of ice-cold heparinized 1x PBS, and liver and brain were dissected out. For analysis of fluorescent protein expression, one hemisphere of brain and one lobe of liver were submerged in ice-cold 4% PFA formulated in 1x PBS and fixed overnight at 4 °C. The other hemisphere and another lobe of liver were manually dissected into 1 mm^3 pieces with regions of interest and flash frozen in O.C.T. Compound (Scigen, #4586) using a dry ice-ethanol bath. O.C.T. blocks were kept at -70 °C until sectioning.

For measurement of viral genomes from bulk DNA, tissue was processed as above, except that unfixed tissue was used immediately for genomic DNA extraction (DNeasy Blood and Tissue Kit, Qiagen, #69504), rather than frozen.

If animals were not used for AAV-Zombie, SpECTr, or bulk DNA extraction, then following perfusion with PBS, animals were perfused with 30 mL of ice-cold 4% PFA in 1x PBS. Relevant tissues were then extracted and post-fixed overnight in 4% PFA in 1x PBS at 4 °C. For sectioning, brain and liver were cryoprotected through immersion in 30% sucrose in 1x PBS. Once the tissue had sunk, it was flash-frozen in O.C.T. Compound using a dry ice-ethanol bath and kept at -70 °C until sectioning.

Sections were obtained using a cryostat (Leica Biosystems). Fixed tissue was sectioned at 80 µm, collected in 1x PBS, and stored at 4 °C until use. Tissue for AAV-Zombie or SpECTr was sectioned at 20 µm, collected on a clean glass slide (Brain Research Laboratories, #2575-plus), allowed to dry, then stored at -70 °C until use.

Immediately prior to imaging, gut tissue and DRGs were optically cleared by overnight room-temperature incubation in RIMS[206,207], then mounted in RIMS with an iSpacer (SunJin Lab). Gut tissue was cut longitudinally before incubation in RIMS and mounted with the myenteric plexus up.

Digital Droplet PCR

A detailed protocol for quantification of AAV genomes from total DNA with digital droplet PCR is available on protocols.io dx.doi.org/10.17504/protocols.io.8epv5r84dg1b/v1). To measure viral genomes from bulk cortex and liver DNA, digital droplet PCR was used. 1 µg of total DNA was first digested overnight with 20 U of SmaI (New England Biolabs, R0141) at 25 °C, or with 20 U each of KpnI-HF and SpeI-HF (New England Biolabs, R3142 and R3133) at 37 °C. The digests were diluted 1:10, and 5 µL of each dilution was loaded into a 25 µL PCR reaction (Bio-Rad, #1863024). 23 µL of the PCR reaction was used to generate droplets (Bio-Rad, #1863005) on a QX200 Droplet Generator (Bio-Rad). 40 µL of droplets were transferred to a PCR plate, which was sealed with a pierceable heat seal (Bio-Rad, #1814040 and #1814000) and the PCR was run according to the manufacturer's protocol. Post-PCR, droplets were measured with a QX200 Droplet Reader and analyzed using the QX Manager software (Bio-Rad, #12010213). Double-quenched FAM- and HEX-labeled probe assays (Integrated DNA Technologies) were used to detect EGFP sequence and W3SL sequence in the same droplets, and the mean of the two resultant concentrations was used. Sequences of ddPCR primer and probe

sets are provided in Supplementary Table 3 of Coughlin et al., 2025. SmaI and KpnI-HF/SpeI-HF digests yielded similar results; only SmaI digests are shown.

Immunohistochemistry

A detailed protocol for immunohistochemistry (IHC) on mouse brain slices is available on protocols.io (dx.doi.org/10.17504/protocols.io.5qpvokmq7l4o/v1). IHC was performed on free-floating sections. Sections were first blocked in BlockAid Blocking Solution (ThermoFisher, B10710) with 0.1% Triton X-100 (Sigma-Aldrich, #93443). Primary and secondary antibodies were diluted in this blocking buffer. Tissue was incubated with primary antibody overnight at 4 °C and with secondary antibody for 2 hr at room temperature. Following each antibody incubation step, sections were washed 3 times for 10 min each in 1x PBS with 0.1% Triton X-100. For Hoechst labeling, sections were incubated for 10 min with 1/10000 Hoechst 33342 (ThermoFisher, H3570) in 1x PBS, followed by 3 washes in 1x PBS. For segmentation of Purkinje cells, sections were Nissl stained with 1/50 NeuroTrace 435/455 (ThermoFisher, N21479) in 1x PBS, followed by two 1-hr room temperature washes and one overnight wash at 4 °C in 1x PBS with 0.1% Triton X-100. Sections were allowed to dry on slides, and then a coverslip was mounted using Prolong Diamond Antifade Mountant (ThermoFisher, P36965).

The following primary antibodies and dilutions were used: chicken anti-GFP (1:1000, Aves Labs, #1020; RRID:AB_10000240) and rabbit anti-TagRFP (for detection of mRuby2, 1:1000, a generous gift from Dr. Dawen Cai, University of Michigan, Cancer Tools, #155266; RRID:AB_3107169). Fluorophore-conjugated $F(ab')_2$ fragment secondary antibodies (Jackson ImmunoResearch) were used at a 1:1000 working concentration.

AAV-Zombie and SpECTr of cultured cells

A detailed protocol for AAV-Zombie and SpECTr on cultured cells is available on protocols.io (dx.doi.org/10.17504/protocols.io.36wgqnz53gk5/v3). AAV-Zombie and SpECTr protocols, and sequences of Zombie barcodes and their split initiator probes were adapted from Askary *et al.* (2020)[194]. Split initiator probes against endogenous genes and reporter transcripts were designed according to Jang *et al.* (2023)[190]. Sequences of HCR-FISH probes against reporter and endogenous transcripts and against Zombie/SpECTr barcodes are provided in Supplementary Table 3 of Coughlin et al., 2025.

For AAV-Zombie and SpECTr of cultured cells on coverslips, a humidified reaction chamber consisting of a 1 mL pipette tip box filled with pre-warmed RNase-free water was used. Parafilm placed on the wafer of the box served as a surface for the *in situ* transcription reaction.

Coverslips, previously fixed in MAA and stored in 70% ethanol, were first washed twice in 1x PBS. 20 µL of transcription mixture per coverslip was prepared according to the manufacturer's protocol (ThermoFisher, AM1334 and AM1330). For simultaneous T7 and SP6 reactions, the T7 buffer was used with 1 µL of each RNA polymerase. For single polymerase reactions, 2 µL of the polymerase was used. 20 µL droplets were pipetted onto the surface of the parafilm. The coverslips were dipped in UltraPure water (ThermoFisher, #10977015), quickly dried by touching their edges to a Kimwipe, then placed cell-side down over the droplets. This reaction was incubated at 37 °C for 3 hr.

Once the transcription reaction was finished, the coverslips were placed cell-side up into a clean 24-well plate and fixed for 20 min at 4 °C with ice-cold PFA in 1x PBS. This was followed by two 5 min washes in 1x PBS, followed by two 5 min washes in 5x SSC (ThermoFisher, AM9770). Samples were then incubated for 15-30 min in pre-warmed probe hybridization buffer, consisting of 2x SSC, 10% ethylene carbonate (Sigma-Aldrich, E26258), and 10% dextran sulfate (Sigma-Aldrich, #3730), at 37 °C. Following this incubation, the coverslips were incubated for 12-16 hr at 37 °C in hybridization buffer plus 2 nM of each probe. Probes for Zombie barcodes, reporter transcripts, and endogenous transcripts were pooled.

After probe hybridization, samples were washed twice for 30 min in stringent wash buffer (2x SSC, 30% ethylene carbonate) at 37 °C, then three times for 15 min in 5x SSC with 0.1 % Tween-20 (Sigma-Aldrich, P1379), and then incubated in HCR amplification buffer (2x SSC, 10% ethylene carbonate) for 20-30 min. HCR hairpins (Molecular Technologies) were heated to 95 °C for 90 s, then cooled to room temperature for 30 min in the dark. For HCR on cultured cells, 30 nM hairpin in amplification buffer was used in a 1-hr amplification reaction. The samples were then washed four times in 5x SSC with 0.1% Tween-20 (10 min per wash, at room temperature).

In some cases, the cytoplasm was labeled with a fluorophore-conjugated poly(dT_{30}) probe (Integrated DNA Technologies). Coverslips were incubated with 100 nM of poly(dT_{30}) probe in 5x SSC with 0.1% Tween-20 for 1 hr, followed by four 10 min, room temperature washes in 5x SSC with 0.1% Tween-20. Finally, Hoechst 33342 was used to label cell nuclei. Samples were mounted with Prolong Diamond Antifade Mountant.

AAV-Zombie and SpECTr of tissue sections

A detailed protocol for AAV-Zombie and SpECTr on tissue sections is available on protocols.io (dx.doi.org/10.17504/protocols.io.14egn6k7yl5d/v1). AAV-Zombie and SpECTr were performed on tissue sections as described above for cultured cells, save for a few

differences. Incubations in tissue were performed in a staining tray (Simport, M918), and fixation and washes were done in Coplin jars.

Sliced fresh tissue was first removed from -70 °C storage and allowed to warm to room temperature. Slides were then briefly washed with 1x PBS to remove O.C.T. compound, then fixed for 3 hr in MAA at -20 °C. Residual fixative was washed off with 1x PBS while the transcription mix was prepared. A total of 200 µL of transcription mix was used per slide, which was pipetted onto the slide and spread out with a clean glass coverslip. We found that simultaneous T7 and SP6 transcription in tissue yielded relatively few and small spots from the SP6-driven barcode. Thus, we carried out T7 and SP6 transcription reactions on separate slides. Likewise, T7 RNA polymerase was used at a 1:10 dilution, whereas SP6 RNA polymerase was used at 1:5 dilution. As with cultured cells, *in situ* transcription was carried out at 37 °C for 3 hr.

For the HCR-FISH steps on tissue sections, we used 4 nM of each probe in an overnight 37 °C hybridization. The HCR hairpin concentration was also doubled to 60 nM. Short HCR incubations may result in low signal for endogenous transcripts, whereas long incubations can yield large, unresolvable Zombie barcode spots. Thus, we did an overnight incubation with only hairpins for endogenous transcripts, then switched the amplification solution to one containing all hairpins, for 1 hr.

Controls for AAV-Zombie and SpECTr

Guidelines for designing, imaging, and analyzing AAV-Zombie and SpECTr experiments are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjn72pgk5/v1). Both AAV-Zombie and SpECTr can produce signals due to hybridization of probes directly to single-stranded AAV genomes and/or transcriptional activity of the AAV ITRs producing barcoded transcripts (e.g., faint ‘concatemer’ signal in Genome A condition, Figure 3.2b). Thus, controls are necessary for setting thresholds for determining real vs. artifactual signal. A non-transduced/non-transfected control sample was used for all AAV-Zombie and SpECTr experiments. For SpECTr experiments, a barcode-only control was used to define signal from probe hybridizing to the AAV genome and/or barcoded transcripts produced due to transcriptional activity of the ITR. As the transcriptional activity of the AAV ITR may differ between cell types, these control experiments were performed in each cell and tissue of interest, and processed side-by-side with experimental samples to mitigate assay-to-assay variability. Depending on the needs of the experiment, other controls may have been included and are outlined in the description of those experiments.

Imaging

For imaging of fluorescent protein expression in cultured cells and for obtaining whole section images of mouse brain and liver, a Keyence BZ-X710 epifluorescence microscope was used, with a 10x, 0.45 NA air objective.

For all other imaging, a Zeiss LSM 880 was used. Imaging of fluorescent protein expression and IHC-stained tissue was accomplished with a 10x, 0.45 NA air objective. Imaging of AAV-Zombie and SpECTr signal in cultured cells and in tissue was performed with a 63x, 1.4 NA oil immersion objective. Imaging settings were chosen to capture full dynamic range of the signal without saturating pixels. When possible, laser power was adjusted before adjusting detector gain. Imaging settings were first optimized on control samples, before imaging of experimental samples. Fields of view were chosen while imaging non-experimental channels (e.g., Hoechst or Nissl).

Image analysis for fluorescent protein expression

For all cell and nuclear segmentation, except segmentation of PCs, Cellpose[192] (v3.0.7; https://www.cellpose.org/; RRID:SCR_021716) was used. Images were batch processed using napari[196] (v0.4.19.post1; https://napari.org/stable/; RRID:SCR_022765) and the serialcellpose plugin (v0.2.2; https://www.napari-hub.org/plugins/napari-serialcellpose). An Anaconda (v2.5.4; https://www.anaconda.com/) distribution of Python (v3.10.14; https://www.python.org/; RRID:SCR_008394) was used. For HEK293T cells, masks were generated from phase-contrast images. For images of cortex, the fluorescent protein signal was used to generate masks.

PC cell bodies were segmented manually, using the Fiji[208] distribution of ImageJ (v1.54f; https://fiji.sc/; RRID:SCR_002285), from images of Nissl-stained tissue (Supplementary Figure 4.1d). The large size and intense Nissl-staining of the PC cell body, relative to neighboring cells, was used to identify PCs.

For analysis of fluorescent protein intensity in HEK293T cells, cortical cells, and PCs, CellProfiler[209] (v4.2.5; https://cellprofiler.org/; RRID:SCR_007358) was used. Classification of cortical cells and PCs as XFP-positive or XFP-negative was also done using CellProfiler. For PCs, we determined the threshold for using empirically determined thresholds based on negative control tissue. For classification of PCs as mRuby2+ or mRuby2- (Figure 4.1d & Figure 4.3b), a threshold of 25.5 a.u. was used, based on measured intensity of mRuby2 signal in wildtype animals injected with 3e11 vg of only mDLX-minBG-CI-mRuby2 (Figure 4.3a,b). For classification of cortical cells as mRuby2+ or mRuby2- (Supplementary Figure 4.3b,c & Figure

4.3f,h) a threshold of 19.125 a.u. was used, based on measured intensity of segmented cortical cells from the mRuby2 channel for 'no enhancer' control animals (Figure 4.1f). As Cellpose reliably did not detect GFP cells in the 'no enhancer' condition (Figure 4.1f), no threshold was necessary for classification of cortical cells as EGFP+ or EGFP- (Figure 4.3f,h). The same threshold was used for all relevant experiments and was measured in animals injected with the lowest dose of the relevant AAV, in order to provide the most stringent threshold. For these analyses of cortical cells and PCs, 3 planes (850 μm x 850 μm) from at least 4 non-adjacent sagittal sections were quantified (i.e., at least 12 volumes per animal).

Bulk protein quantification of SCID and wildtype mice was performed using Fiji, from 3 non-adjacent 100 μm sections per tissue per animal. Cortex and cerebellum were manually segmented from sagittal sections; liver sections were analyzed whole.

Image analysis for AAV-Zombie and SpECTr

Guidelines for designing, imaging, and analyzing AAV-Zombie and SpECTr experiments are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjn72pgk5/v1). For analysis of AAV-Zombie and SpECTr spots, segmentation was performed as described above. For HEK293T cells nuclear mask were generated from Hoechst signal. PC nuclei were manually segmented in Fiji, using large nucleus size, euchromatic nuclear staining, and the presence of *Itpr1* transcript to positively identify PCs.

Quantification and measurement of AAV genomes and concatemers in PCs was accomplished using CellProfiler. Genome and concatemer spots were identified within segmented nuclear masks, using empirically determined spot size thresholds and robust background intensity thresholding, chosen due to the sparse foreground signal.

AAV concatemers were identified in HEK293T cells as described above, with some exceptions. Masks were size filtered, using empirically determined thresholds and only nuclear AAV genomes and concatemers were measured.

Statistics and reproducibility

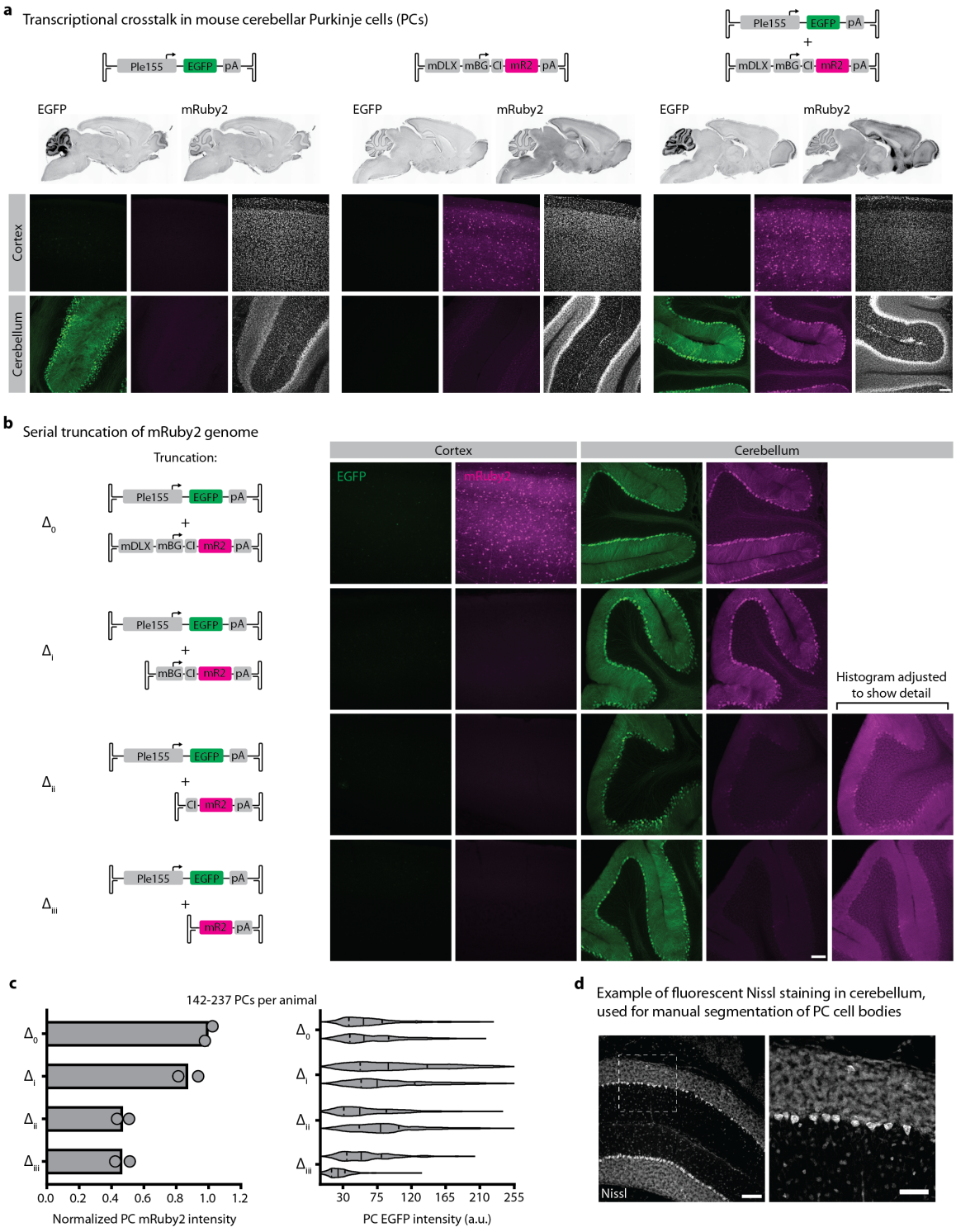
The number of biological replicates for each experiment are included in the corresponding figure legends. No data were excluded from analyses. For all violin plots, the middle dashed line is the median and the upper and lower dashed lines are quartiles. Statistical analysis was performed with GraphPad Prism (v10.0.3, GraphPad Software; RRID:SCR_002798) as described in figure legends. Where relevant, all tests were two-tailed and corrected for multiple comparisons to maintain an experiment-wide alpha of 0.05.

The following *in vivo* experiments were repeated once ($n > 2$ animals per experimental condition) with similar results: Figure 4.1a,b and Supplementary Figure 4.1a, Figure 4.3a-d and Supplementary Figure 4.4a-e, and Supplementary Figure 4.5a-c. The remaining *in vivo* experiments were not independently repeated. All *in vitro* experiments were repeated at least twice with similar results.

Data availability

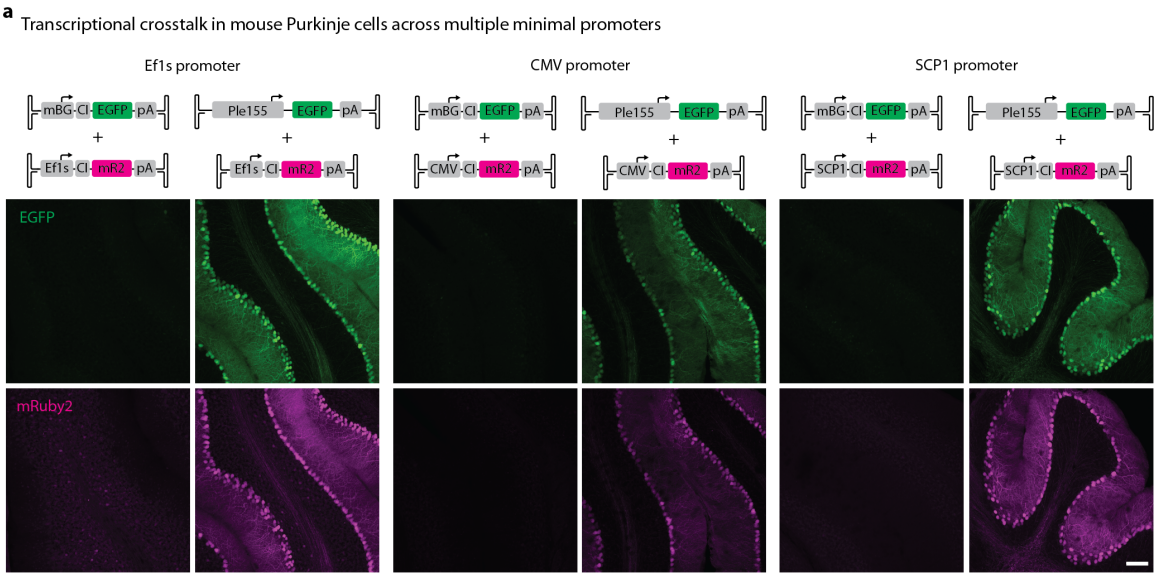
All sequences of primers, probes, sgRNAs, and other sequence elements are provided in Supplementary Table 3 of Coughlin et al., 2025. Images of brain tissue that are quantified in this work are deposited in the Brain Image Library (https://doi.org/10.35077/g.1163). Tabular datasets supporting conclusions of this work are available on Zenodo (10.5281/zenodo.13952929). All other data that support the findings of this study are available from the corresponding authors upon reasonable request.

4.6 Supplementary figures

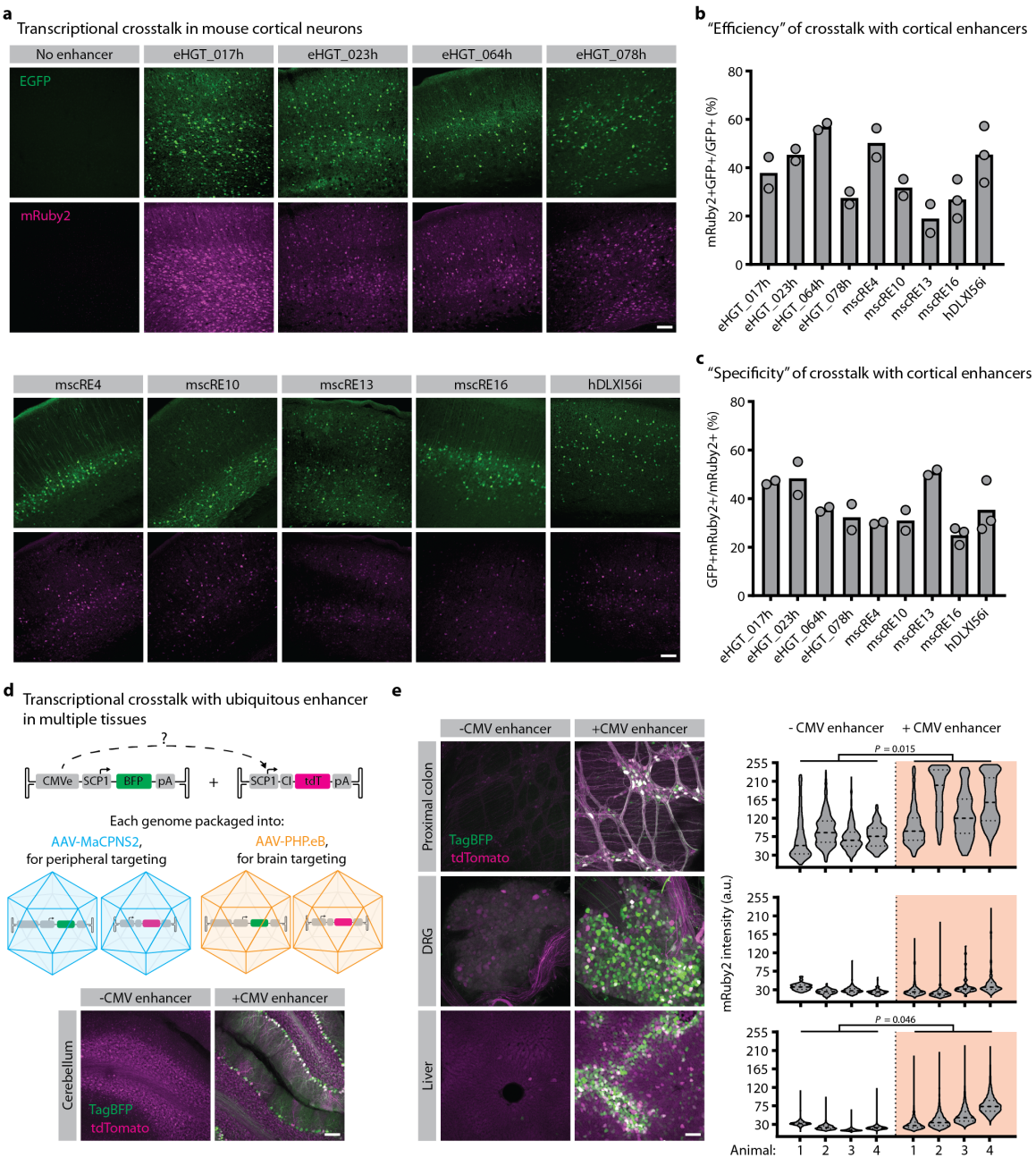


Supplementary Figure 4.1. Transcriptional crosstalk in cerebellar PCs between Ple155 and mDLX-minBG. a, Related to Figure 4.1a,b. Representative images of whole brain, cortical, and cerebellar expression patterns after single or double AAV injections. All genomes delivered at

1e12 vg dose in AAV-PHP.eB. Scale bar = 100 µm. **b**,**c**, Related to Figure 4.1c,d. **b**, Representative images of transcriptional crosstalk between Ple155 element and serially-truncated mDLX-minBG. All genomes delivered at 5e11 vg dose, in AAV-PHP.eB. Scale bar = 100 µm. **c**, Quantification of results for truncation conditions shown in (**b**), quantified as normalized mean PC mRuby2 intensity (left) and PC EGFP fluorescence intensity (right). mRuby2 fluorescence intensity is normalized to the mean of full-length mRuby2 genome (construct Δ_0). Bars represent mean. Each violin plot represents EGFP intensity from PCs in one animal. $n = 2$ animals per condition. All genomes delivered at 5e11 vg dose in AAV-PHP.eB. **d**, Example of fluorescent Nissl staining in cerebellum, showing intense signal in PCs. Nissl staining was used for all manual segmentation of PC cell bodies. Scale bar = 100 µm for left image, 50 µm for right image. CI = chimeric intron.

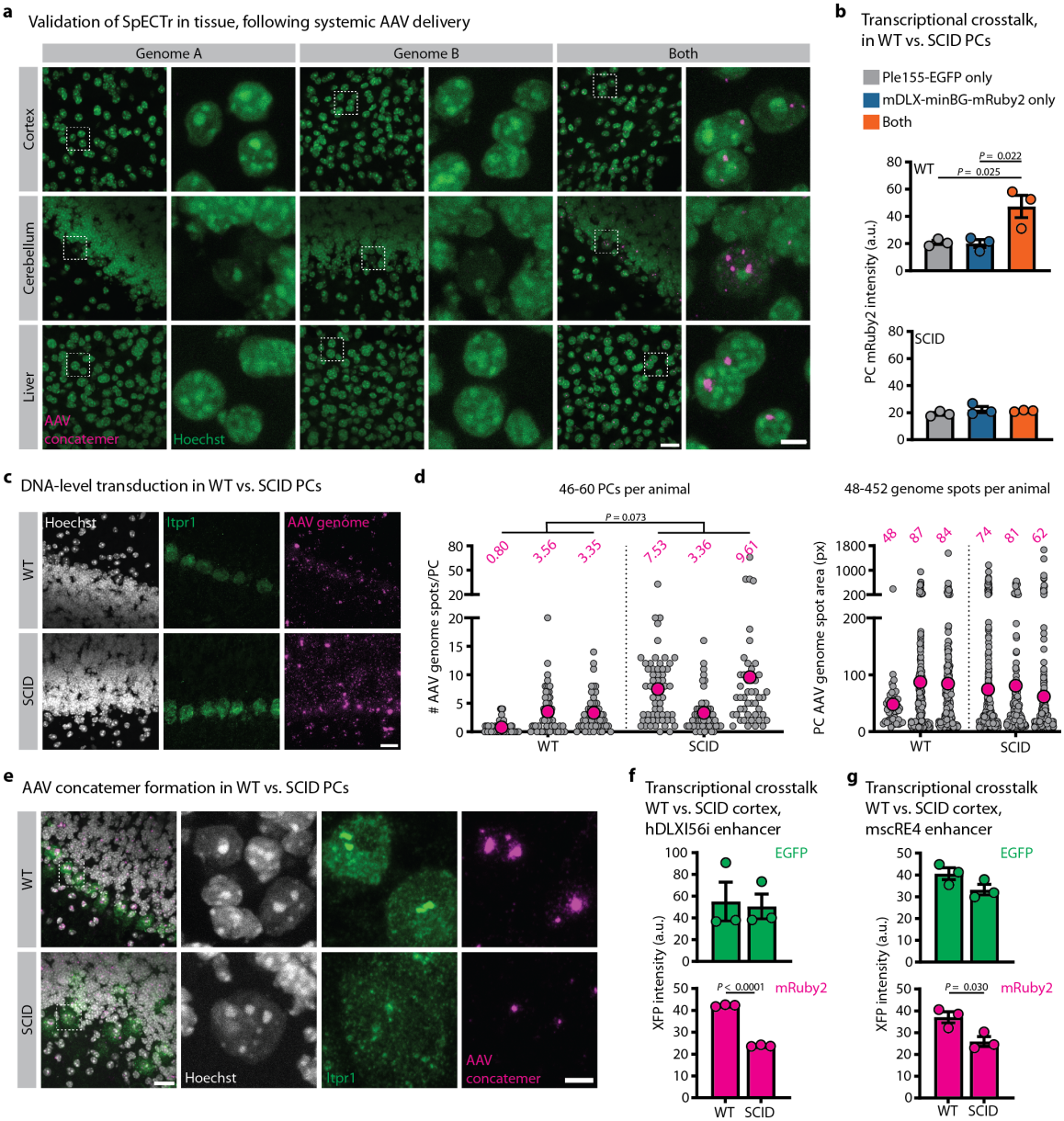


Supplementary Figure 4.2. Transcriptional crosstalk occurs across multiple minimal promoters. Related to Figure 4.1e. Representative images of transcriptional crosstalk between Ple155 and commonly used minimal promoters: Ef1s (left), CMV promoter (middle), and SCP1 (right). A minBG-driven EGFP serves as a 'no enhancer' control. All genomes delivered at 1e12 vg dose in AAV-PHP.eB. Scale bar = 100 µm.



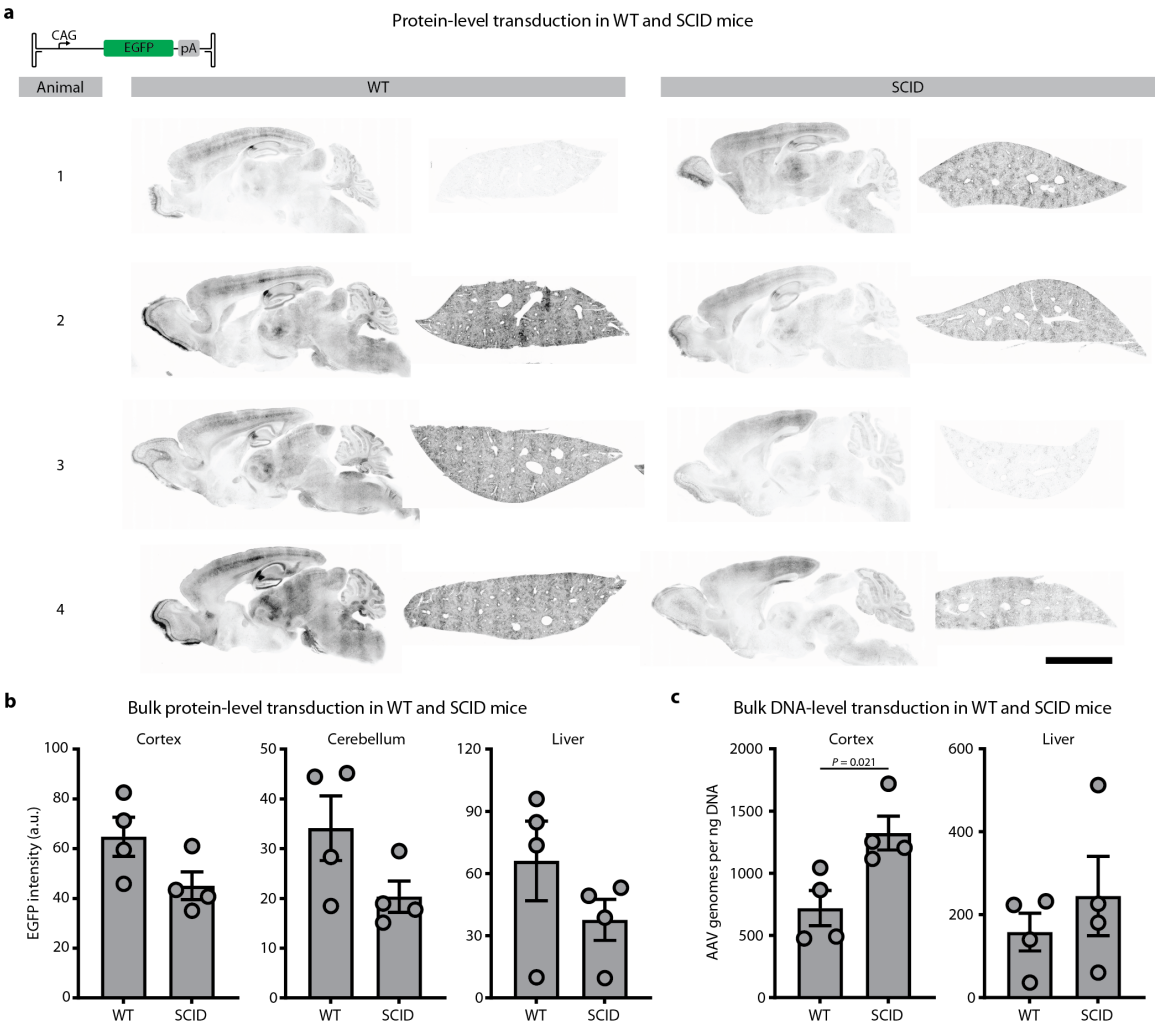
Supplementary Figure 4.3. Transcriptional crosstalk in cortex and periphery. a-c, Related to Figure 4.1f. **a**, Representative images of cortical expression patterns. EGFP reports activity of promoter in *cis* to the enhancer. mRuby2 reports activity of promoter in *trans* to the enhancer. All genomes delivered at 1e12 vg dose in AAV-PHP.eB. Fluorescent protein signal was amplified through IHC. Scale bar = 100 µm. **b**, "Efficiency" of transcriptional crosstalk with various cortical enhancers, quantified as percent of EGFP-positive cells that are also mRuby2-positive. *n* = 2 animals per condition, except mscRE16 and hDLXI56i, for which *n* = 3. **c**, "Specificity" of transcriptional crosstalk with various cortical enhancers, quantified as percent of mRuby2-positive cells that are also EGFP-positive. *n* = 2 animals per condition, except mscRE16 and hDLXI56i, for which *n* = 3. **d**, Strategy for investigating transcriptional crosstalk in multiple

tissues. A cocktail of AAV-PHP.eB and AAV-MaCPNS2 was used to provide broad CNS and PNS coverage (1e12 vg per genome-capsid pair, 4e12 vg per animal) and deliver the ubiquitous CMVe and SCP1 promoter. Crosstalk between CMVe and SCP1 was readily apparent in cerebellar PCs (bottom). **e,** Representative images of proximal colon, dorsal root ganglion (DRG), and liver, and quantification of mRuby2 fluorescence intensity. Each violin plot represents mRuby2 intensity from segmented cells in one animal. $n = 4$ animals per condition. Statistical significance was determined using unpaired t-tests. Scale bar = 100 μm. Bars in **b** and **c** represent mean.



Supplementary Figure 4.4. AAV concatemer formation and transcriptional crosstalk in wildtype and SCID animals. a, Validation of SpECTr for detecting AAV concatemers in tissue, using same genomes as in Figure 3.2a,b. Wildtype C57BL/6J mice were transduced with only

Genome A (left 2 columns), only Genome B (middle), or both genomes (right). AAV concatemers were nuclear and only detected with SpECTr following co-transduction of both genomes. Scale bars = 20 μm for left images of pairs, 5 μm for right. All genomes delivered at 3e11 vg dose in AAV-PHP.eB. **b**, Related to Figure 4.3a,b. Quantification of transcriptional crosstalk in wildtype and SCID PCs, quantified as mean PC mRuby2 fluorescence intensity. Statistical significance was determined using one-way ANOVA ($P = 0.015$) and Tukey's multiple comparison test. $n = 3$ animals per condition. Bars represent mean ± s.e.m. **c-e**, Related to Figure 4.3c,d. **c,** Representative images of AAV genomes detected and quantified with AAV-Zombie in PCs of dual AAV-injected wildtype and SCID animals shown in Figure 4.3a. HCR-FISH against *Itpr1* transcript serves as a marker for PCs. Scale bar = 20 μm. **d**, Quantification of PC AAV genome spot count (left) and spot size (right), in wildtype and SCID PCs. Each grey dot corresponds to a single PC (left) or a single genome spot (right). Magenta dot and number indicate mean of animal. $n = 3$ animals per condition. Statistical significance was determined using unpaired t-tests. **e**, Representative images of AAV concatemers detected with SpECTr in PCs of dual-injected wildtype and SCID animals shown in Figure 4.3a. HCR-FISH against *Itpr1* transcript serves as a marker for PCs. Scale bars = 20 μm for left column, 5 μm for others. **f,g**, Related to Figure 4.3e-h. Quantification of transcriptional crosstalk in wildtype and SCID cortical cells, quantified as mean XFP fluorescence intensity, with hDLXI56i enhancer (**f**) and mscRE4 enhancer (**g**). $n = 3$ animals per condition. Bars in (**f**), and (**g**) represent mean ± s.e.m. Statistical significance was determined using unpaired t-tests.



Supplementary Figure 4.5. AAV transduction of wildtype and SCID mice. a, Wildtype and SCID C57BL/6J animals were transduced with 3e11 vg of AAV-PHP.eB packaging a CAG-EGFP reporter, and tissue was collected 4 weeks later. Representative sagittal brain (left) and liver (right) sections are shown. Scale bar = 5 mm. **b**, Quantification of bulk protein in cortex (left), cerebellum (middle), and liver (right) of wildtype and SCID animals, assessed by mean EGFP intensity. **c**, Quantification of bulk viral DNA in wildtype and SCID cortex and liver, assayed through digital droplet PCR. SmaI digests and KpnI-HF/SpeI-HF digests yielded similar results; results from SmaI-digested samples are shown. For (**b**) and (**c**), statistical significance was determined using unpaired t-tests. $n = 4$ animals per genotype. Bars are mean ± s.e.m.

Chapter 5

AAV TRANSCRIPTIONAL CROSSTALK ENABLES ALL-AAV CELL TYPE-SPECIFIC GENE EDITING

Adapted from:

Coughlin, G.M.[#], Borsos, M.[#], Barcelona, B.H.[$], Appling, N.[$], Mayfield, A.M.H., Mackey, E.D., Eser, R.A., Jackson, C.R., Chen, X., Ravindra Kumar, S., and Gradinaru, V.* (2025). Spatial genomics of AAV vectors reveals mechanism of transcriptional crosstalk that enables targeted delivery of large genetic cargo. Nat. Biotechnol. https://doi.org/10.1038/s41587-025-02565-4

5.1 Summary

Tools for editing and modulating the mammalian genome can facilitate new understanding of gene function and offer new modalities to address disease. Clustered regularly interspaced short-palindromic repeat (CRISPR)-associated (Cas)-based tools[151–154] are particularly attractive for this goal, as they are easily programmed and can target a broad range of sequence space. Engineering of CRISPR-based tools has yielded a toolkit that allows researchers knockout genes, alter, add or remove sequence, and to increase or decrease expression from an endogenous locus[153–159]. Integrating CRISPR-based tools with broad, efficient, and cell type-targeted AAV delivery can provide previously unrealized capabilities in preclinical research. However, many Cas proteins are large, especially well-characterized ones such as *Staphylococcus aureus* Cas9 (SaCas9; 3.2 kb)[160], and additional functionality, through fusion of domains to Cas proteins, increases this size further. This large size precludes simple incorporation of enhancers for AAV-based cell type-specific gene editing.

Our profiling of transcriptional crosstalk in Chapter 4 suggests a novel method for achieving cell type-specific expression of large coding sequences. Here we explore this application using a Purkinje cell-targeting element[104] and SaCas9 as a large coding sequence. These 2 sequences (3.2 kb and 1.65 kb, respectively) are too large to fit into a single AAV genome together. Using a reporter assay and systemically administered AAVs, we demonstrate that our crosstalk-mediated targeting could achieve cell type-specific gene editing in the mouse brain. We then assessed whether this approach could be used to explore cell type-specific gene function in wildtype animals, by disrupting *Cacna1a* in either a ubiquitous or Purkinje cell-targeted manner. Targeted disruption of this gene recapitulated known loss-of-function

phenotypes while avoiding phenotypes due to loss-of-function in non-target tissues. This approach is faster than cell type-targeted gene disruption using conventional approaches with recombinase driver lines, and could be used for cell type-specific reverse genetics experiments in mammalian systems. Furthermore, transcriptional crosstalk can allow for cell type-targeted genome modulation in more translationally relevant species, as well as for therapeutic purposes.

5.2 Crosstalk enables all-AAV cell type-specific gene editing

We reasoned that transcriptional crosstalk might enable cell type-specific delivery of larger cargo, by separating bulky gene regulatory elements from minimal promoters and coding sequences in another AAV (Figure 5.1a). We explored the feasibility of this approach using *Staphylococcus aureus* Cas9 (SaCas9) as a large cargo and targeting Purkinje cells with Ple155. Notably, these 2 sequences (3.2 kb and 1.65 kb, respectively) are too large to fit into a single AAV genome together. We adopted a commonly-used reporter assay based on Ai14 mice (*Rosa26*$^{CAG\text{-}LSL\text{-}tdTomato}$)[210–213].

Minimal expression of SaCas9 with no enhancer resulted in low efficiency of editing in all tissues examined (Figure 5.1b,c “No enhancer”). Conversely, when SaCas9 was strongly expressed with a ubiquitous enhancer (CMVe) delivered in *cis*, we observed a strong increase in editing efficiency in tissues of interest, compared to the no-enhancer condition. We saw a 31-fold increase in editing in the liver, a 17-fold increase in the cortex, a 9-fold increase in non-Purkinje cerebellar cells (non-PCs), and a 107-fold increase in PCs (Figure 5.1b,c “Ubiquitous enhancer”). Using transcriptional crosstalk to direct SaCas9 expression specifically to PCs with the Ple155 element in the companion AAV genome, we restricted efficient editing to PCs, yielding a 177-fold increase in PC editing efficiency compared to the no-enhancer condition, with no significant increases in other tissues and cerebellar cell types (Figure 5.1b,c “Crosstalk”). These results establish the utility of transcriptional crosstalk for AAV-based cell type-specific genome editing and manipulation, bypassing the need for transgenic driver lines to restrict expression to a target population or Cas9 reporter lines to deliver editing machinery.

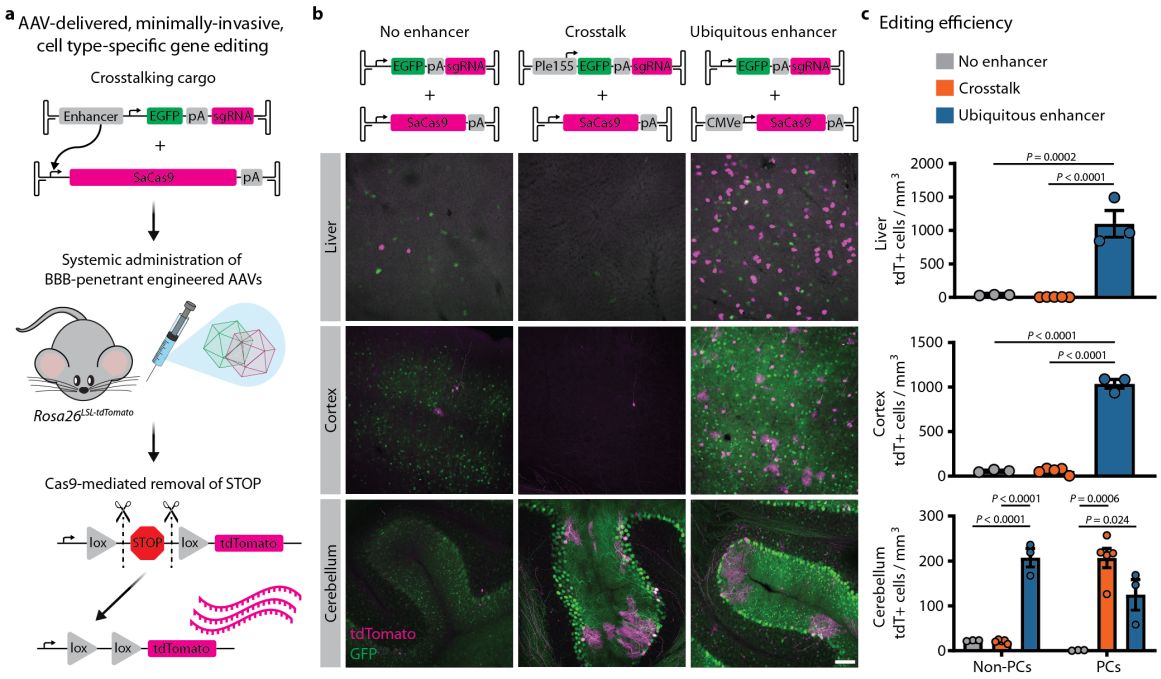


Figure 5.1. Transcriptional crosstalk enables all-AAV cell type-specific genome editing with CRISPR-Cas9. a, Schematic of AAV-delivered, minimally-invasive, cell type-specific gene editing. *Staphylococcus aureus* Cas9 (SaCas9), packaged with minimal elements (total size 4.2 kb), is delivered with a bulky enhancer element in *trans*, resulting in upregulation of SaCas9 expression in a cell type-specific manner. As a proof of principle, we used a common reporter assay with *Rosa26*$^{LSL\text{-}tdTomato}$ mice, in which guide RNAs direct SaCas9 to remove the stop cassette, enabling tdTomato expression. For all conditions, the sgRNAs were expressed by the ubiquitous U6 promoter. All genomes delivered at 1e12 vg dose in AAV-PHP.eB. **b**, Demonstration of crosstalk-enabled gene editing using the Ple155 element to drive SaCas9 expression in Purkinje cells (PCs). As controls, we included a 'no enhancer' condition, as well as a condition in which SaCas9 is strongly expressed by the ubiquitous CMVe delivered in *cis*. Representative images from liver, cortex, and cerebellum. Scale bars = 100 μm. **c**, Quantification of editing efficiency, assessed by number of tdTomato-positive cells per mm^3 of tissue. PCs and non-PCs were quantified separately. Using crosstalk to drive strong SaCas9 expression specifically in PCs restricted high efficiency editing to that cell type. Statistical significance was determined using one-way ANOVAs (for liver, cortex and non-PCs, $P < 0.0001$; for PCs, $P = 0.0008$) and Tukey's multiple comparison test. $n = 3$ (no enhancer or ubiquitous enhancer) or 5 (crosstalk) animals per condition. Bars represent mean ± s.e.m.

5.3 Harnessing crosstalk for cell targeted functional genetics

Efficient and specific gene editing through transcriptional crosstalk from systemically-delivered AAVs offers a means to explore gene function in a cell type-specific manner. Importantly, this strategy does not rely on transgenic lines, enabling rapid and cost-effective generation of large cohorts from easily obtained wildtype animals.

To test this approach, we targeted *Cacna1a* in wildtype C57BL/6J mice, in either a ubiquitous or PC-specific manner using transcriptional crosstalk (Figure 5.2a, top). *Cacna1a* is broadly expressed in the brain, and global knockout leads to dystonia, ataxia, cerebellar degeneration, absence seizures, and early lethality[214,215]. Forebrain-specific deletion of *Cacna1a* causes learning and memory deficits[216], and leads to the emergence of absence seizures[217,218]. Purkinje cell targeted loss-of-function through a Pcp2-cre driver line leads to ataxia[219], and, unexpectedly, absence seizures[220]. This epileptiform activity was attributed to recombinase activity from this driver line in forebrain populations[220,221]. Thus, understanding the function of *Cacna1a* in PCs requires methods to specifically target PCs, thereby avoiding confounds due to loss-of-function in other brain regions.

As a control for the effects of off-target editing, we used two sequence-independent guide RNAs (sgCacna1a A and sgCacna1a B), comparing these to an unguided condition in which no guide RNA sequence was present. To assess whether this approach could recapitulate known ataxia phenotypes resulting from PC-specific loss-of-function of *Cacna1a*, we assessed a battery of behaviours before and for five weeks after AAV administration (Figure 5.2a, bottom).

In both ubiquitous and PC-specific paradigms, we observed a strong reduction in Cacna1a staining in the cerebellum that was consistent between both guide RNAs (Figure 5.2b and Supplementary Figure 5.1a,b). Importantly, we saw similar reductions in Cacna1a staining intensity with both ubiquitous and PC-specific SaCas9 expression.

Both ubiquitous and PC-specific *Cacna1a* disruption also recapitulated several hallmarks of ataxia: reduced locomotion in an open field (Figure 5.2c,d and Supplementary Figure 5.1c), impairments in skilled motor behaviour, as assessed by narrowing beam crossing (Figure 5.2e, Supplementary Figure 5.1d and Supplementary Videos 1,2), reduced limb strength (Figure 5.2f), and gait deficits (Figure 5.2g,h and Supplementary Figure 5.1e-h). Both guide RNAs resulted in similar phenotypes, suggesting that the deficits observed were not due to off-target editing, and no deficits were observed in animals that did not receive a guide RNA. Whereas ubiquitous expression of SaCas9 led to significantly reduced weight by 3 weeks post-injection, we did not observe any significant difference in weight until 5 weeks post-injection with PC-specific expression (Supplementary Figure 5.1i).

We next assessed whether the specificity afforded by transcriptional crosstalk in the PC-specific paradigm could circumvent the epileptic activity observed with forebrain-specific disruption of *Cacna1a*. Thus, in a separate cohort of animals, we conducted longitudinal cortical EEG recordings, starting one week before and continuing six weeks after AAV injection (Figure

5.2i), monitoring for spike-and-wave discharges (SWDs) that are characteristic of absence seizures.

In the ubiquitous paradigm, both sgRNAs against *Cacna1a* resulted in significant increases in detected SWDs by 3 weeks post-injection (Figure 5.2j,k, Supplementary Figure 5.2). This was not observed in the unguided condition. Importantly, transcriptional crosstalk-mediated PC-specific *Cacna1a* disruption did not result in a significant increase in detected SWDs. However, in a small number of animals targeted through transcriptional crosstalk, we did detect slight increases in SWDs (Supplementary Figure 5.2e), potentially reflecting some leakiness due to basal activity of the minBG promoter.

Taken together, the results of our behavioural and EEG experiments demonstrate that transcriptional crosstalk can be leveraged for cell type-specific gene manipulation in wildtype animals.

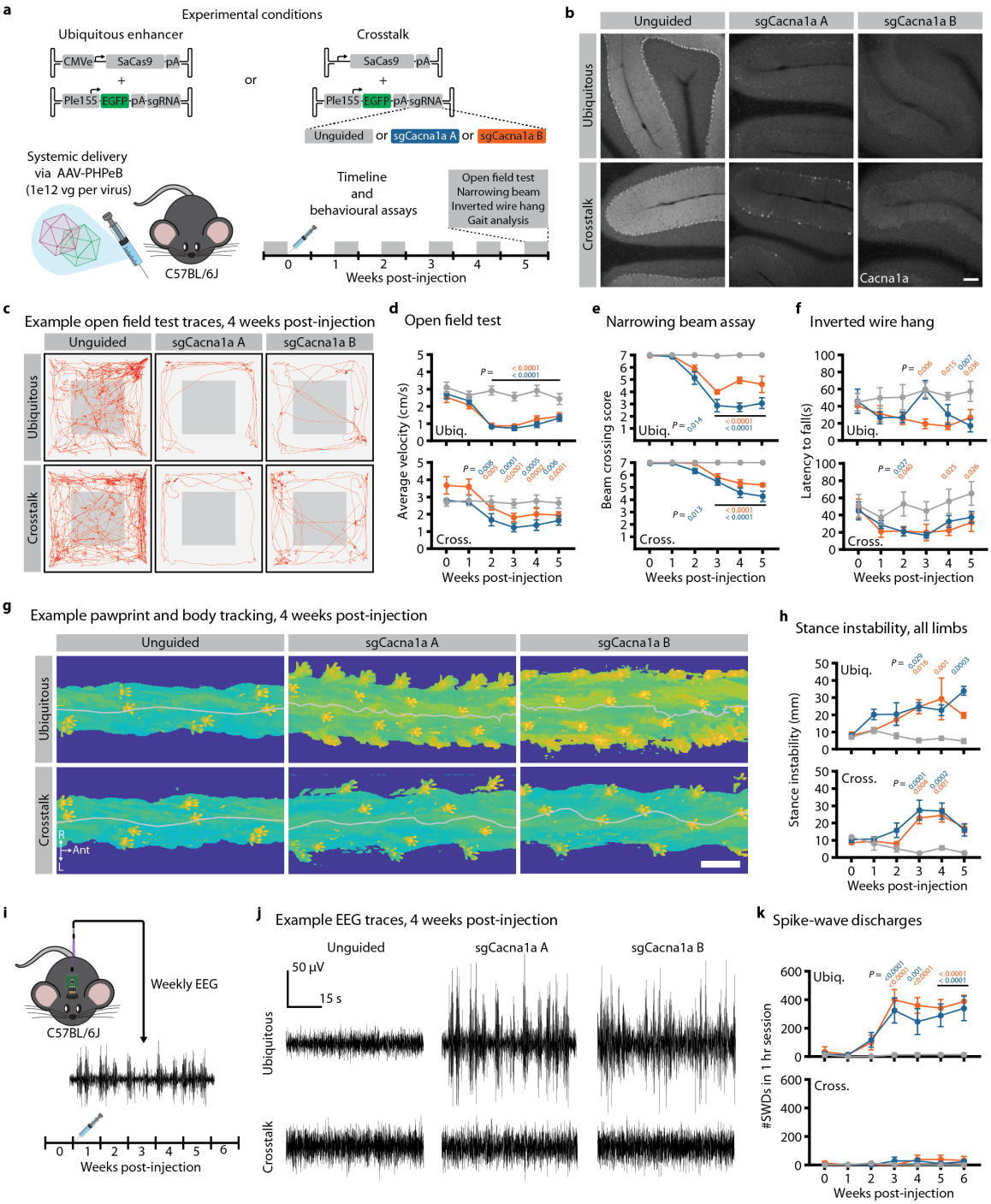


Figure 5.2. Transcriptional crosstalk enables efficient cell type-specific gene disruption. a, Two conditions were tested. In the ubiquitous condition, SaCas9 was strongly expressed with CMVe delivered in *cis*. In the crosstalk condition, SaCas9 expression was restricted to Purkinje cells (PCs) through inclusion of the Ple155 element delivered in *trans*. Two sequence-independent sgRNAs targeting *Cacna1a* were used and compared to an unguided condition. sgRNAs were expressed by the ubiquitous U6 promoter. Behavioural assays were performed weekly, before and five weeks after AAV administration. **b**, Representative IHC against Cacna1a in cerebellum. Scale bar = 100 µm. **c-h**, Characterization of ataxic phenotypes following

ubiquitous ("Ubiq.," top graphs) and PC-specific ("Cross.," bottom graphs) disruption of *Cacna1a*. Motor behaviour was assessed with an open field test (**c**,**d**), skilled locomotion with the narrowing beam assay (**e**), limb strength with inverted wire hang (**f**), and gait using automated pawprint and body tracking (**g**,**h**). Red lines in (**c)** represent animal position over a 10 min trial, 4 weeks post-injection. Heatmaps in (**g)** show pawprint positions and body tracking over a small segment of the elevated platform used for gait analysis, 4 weeks post-injection. Grey line indicates midline of body. Scale bar = 3 cm. **i-k**, Characterization of epileptic activity in cortex for same experimental manipulations as in (**a**). **i**, Weekly 90 min EEG recordings were collected, before and for six weeks after AAV administration. **j**, Sample cortical EEG traces, 4 weeks post-injection. **k**, Quantification of spike-and-wave discharges (SWDs) for animals in ubiquitous (top) and PC-specific conditions (bottom). Statistical significance for open field test, beam crossing, stance instability, and SWDs was determined by two-way repeated-measures ANOVA and Dunnett's multiple comparison tests against 0-week time point. Statistical significance for inverted wire hang was determined by two-way repeated-measures ANOVA and Dunnett's multiple comparison tests against the unguided condition. Points and bars represent mean ± s.e.m. For all groups, $n = 5$ except (ubiquitous + sgCacna1a B behaviour group), (crosstalk + sgCacna1a A behaviour group) and (ubiquitous + sgCacna1a B EEG group), in which $n = 6$. Grey line = unguided controls, blue = sgCacna1a A, and orange = sgCacna1a B.

5.4 Discussion

By separating cell type-specific enhancers and coding sequences onto separate AAV genomes, AAV transcriptional crosstalk can be used for cell type-targeted delivery of large coding sequences. We profiled this using SaCas9 as a large coding sequence and a Purkinje cell-targeting regulatory element, and in two separate contexts. Using a reporter assay, we demonstrated that transcriptional crosstalk-mediated targeting could restrict efficient editing to the target cell type. To demonstrate the utility of this approach, we then targeted *Cacna1a* in a ubiquitous or Purkinje cell-targeted manner via crosstalk. In both the ubiquitous and the crosstalk targeted paradigm, disruption of *Cacna1a* recapitulated ataxic phenotypes observed following conventional ubiquitous[214,215] or Purkinje-cell specific[219] gene knockout techniques. However, with transcriptional crosstalk-mediated Purkinje cell-targeted disruption, we did not observe significant increases in cortical epileptiform activity, despite significant increases in such activity following ubiquitous disruption. Previous reports have observed increased cortical epileptiform activity following Purkinje cell-specific *Cacna1a* knockout[220], this has been attributed to leakiness of the utilized recombinase driver line in the forebrain[220,221].

Though we did not observe significant increases in absence seizures following Purkinje cell-targeted *Cacna1a* disruption, some animals did show relatively small increases in epileptiform activity. This observation suggests some leak expression of SaCas9 in forebrain. Other regulatory strategies could be adopted to reduce this leakiness. As miRNAs regulate

expression post-transcriptionally and their target sites are relatively short, transcriptional crosstalk and miRNA TS-mediated repression could be combined to further hone expression. Furthermore, in our crosstalk-mediated gene editing experiments, the sgRNA was under the control of a ubiquitous U6 Pol III promoter. Expressing the sgRNA from the untranslated region of the Purkinje cell-specific Pol II promoter-driven transcript may confer further specificity.

Our gene disruption experiments were performed using wildtype animals and all components for cell type-specific gene disruption were delivered using systemically administered AAVs. Thus, we were able to generate sufficiently powered behavioural cohorts in ~2 weeks, using efficient production pipelines for systemically administered AAVs and easily obtained wildtype animals. This strategy contrasts with conventional generation of cell type-specific loss-of-function mouse models, involving crossing of cell type-specific driver lines to floxxed reporter lines, which can take months to years, assuming appropriate lines already exist. A systemic AAV crosstalk-based strategy could be used for rapid cell type-specific reverse genetics screens in mammalian systems. However, one potential limitation of this strategy arises from the stochastic and cell type- or organ-dependent transduction afforded by systemically administered AAVs. Thus, development of efficient gene delivery vectors and identification of cell type-specific enhancers can advance the use of systemic AAVs for reverse genetics experiments in mammalian systems.

5.5 Methods

Key resources

Key resources for this work are listed in Supplementary Table 2 of Coughlin et al., 2025.

Plasmid DNA

Standard molecular cloning techniques were used to generate DNA constructs in this study. Double-stranded DNA was synthesized by Integrated DNA Technologies and inserted into pAAV backbones with NEBuilder HIFI (New England Biolabs, E2621). sgRNA sequences were synthesized as overlapping single-stranded DNA oligos (Integrated DNA Technologies) that were then annealed together and ligated into sgRNA expression cassettes using T4 DNA ligase (New England Biolabs, M0202). Sequences of sgRNAs and utilized DNA elements (e.g., promoters and enhancers) are provided in Supplementary Table 3 of Coughlin et al., 2025.

pUCmini-iCAP-AAV-PHP.eB[175] (Addgene #103005; RRID:Addgene_103005) and pHelper (Agilent, #240071) plasmids were used for production of AAVs. Prior to use, all

plasmids were sequence verified via whole-plasmid sequencing through Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation.

AAV production

Detailed protocols for AAV production and titration are available on protocols.io (dx.doi.org/10.17504/protocols.io.n2bvjnew5gk5/v1 and dx.doi.org/10.17504/protocols.io.e6nvw1n47lmk/v1). AAVs were produced and purified according to published methods[191], with some minor alterations. Briefly, HEK293T cells (ATCC, CRL-3216; RRID:CVCL_0063) were triple transfected with PEI-MAX (Polysciences, #24765) to deliver the rep-cap or iCAP, pHelper, and genome packaging plasmids. Viruses were harvested from cells and media, then purified over 15%, 25%, 40%, and 60% iodixanol (OptiPrep, Serumwerk, #1893) step gradients. A Type 70 Ti fixed-angle titanium rotor (Beckman Coulter, #337922) at 58.4k rpm for 1.5 hr, or a Type 70.1 Ti fixed-angle titanium rotor (Beckman Coulter, #342184) at 61.7k rpm for 1.25 hr was used, depending on the scale and number of AAVs to be purified simultaneously. Viruses were concentrated using Amicon Ultra-15 or Amicon Ultra-4 filters with a 100 kD size cutoff (MilliporeSigma, UFC9100 and UFC8100) and formulated in sterile DPBS (ThermoFisher, #14190144) with 0.001% Pluronic F-68 (ThermoFisher, #24040032). AAVs were titered with qPCR by measuring the number of DNase I-resistant viral genomes, relative to a linearized genome plasmid standard. Prior to injection, AAVs were diluted in sterile saline. ssAAV genomes were used for all experiments. The following qPCR primers against the W3 sequence were used for titering AAV viruses:

Forward: 5'-TGGTATTCTTAACTATGTTGCTCCT-3'
Reverse: 5'-AAGCCATACGGGAAGCAATAG-3'

Animals

Animal husbandry and all procedures involving animals were performed in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and approved by the Institutional Animal Care and Use Committee (IACUC) and by the Office of Laboratory Animal Resources at the California Institute of Technology.

8-week old, male C57BL/6J (strain #: 000664; RRID:IMSR_JAX:000664) and C57BL/6J-background *Rosa26*$^{CAG\text{-}LSL\text{-}tdTomato}$ (strain #: 007914; RRID:IMSR_JAX:007914) mice were obtained from the Jackson Laboratory. Mice were housed 3-4 per cage, on a 12 hr light/dark

cycle, and had *ad libitum* access to food and water. For behavioural experiments and EEG recordings, animals were kept in a reverse light cycle; all behavioural assays and recordings were conducted during the dark cycle, between ZT13 and ZT17.

For motor behaviour experiments, mice were 8.5 to 9.5 weeks old at the time of injection. For EEG experiments the animals were 8.5 to 12.5 weeks old at the time of surgery, as they needed to be staggered to accommodate the large cohort size. In this case, care was taken to ensure no systemic assignment to experimental groups based on age at experiment onset.

Retro-orbital injection

A detailed protocol for systemic AAV administration through retro-orbital injection is available on protocols.io (dx.doi.org/10.17504/protocols.io.36wgqnw73gk5/v1). AAVs were administered via retro-orbital injection during isoflurane anesthesia (1-3% in 95% O_2/5% CO_2, provided by nose cone at 1 L/min), followed by administration of 1-2 drops of 0.5% proparacaine to the corneal surface[191].

EEG implantation surgery

A detailed protocol for mouse EEG implantation surgery and EEG data collection is available on protocols.io (dx.doi.org/10.17504/protocols.io.81wgbzj2ygpk/v1). Mice were anesthetized with isoflurane (5% induction, 1% maintenance), then subcutaneously injected with ketoprofen (5 mg/kg) and buprenorphine XR (3.25 mg/kg). The animals' heads were fixed in a stereotaxic frame (David Kopf Instruments), with a heating pad to maintain body temperature. The scalp was then sterilized, subcutaneously injected with 1-2 drops of 0.5% bupivacaine, and a 1.5 cm anterior-posterior incision made to expose the skull. The skull surface was scored with a scalpel and the EEG headmount (Pinnacle Technology, #8201) was glued to the surface of the skull using cyanoacrylate adhesive. The anterior edge of the headmount was targeted to be 3.5 mm anterior to bregma. A sterile 23g needle was used to pierce the skull underneath each hole in the headmount. EEG screws were implanted through the headmount and into the craniotomy hole; 0.10" screws (Pinnacle Technology, #8209) were used for the anterior holes and 0.12" EEG screws (Pinnacle Technology, #8212) were used for posterior holes. A small amount of silver epoxy (Pinnacle Technology, #8226) was applied to each screw before fully tightening to ensure electrical connection between the screw and the headmount. Continuity of the contacts was assessed with a multimeter. Adhesive cement (C&B Metabond, Parkell, S398, S371, S396) was used to secure screws and headmount in place, followed by dental cement to cover the edges of

the headmount. Ibuprofen (30 mg/kg) was provided in drinking water for at least 3 days following surgery. Animals were allowed to recover for at least 1 week before EEG recordings.

Tissue harvest and processing

Tissue was collected 4 weeks post-AAV administration, except for animals used in the *Cacna1a* knockout experiments in which tissue was collected 6 weeks (for motor behaviour cohort) or 8 weeks (for EEG cohort) post-AAV administration. Animals were euthanized via i.p. injection of 100 mg/kg euthasol.

Animals were transcardially perfused with 30 mL of ice-cold heparinized 1x PBS, followed by 30 mL of ice-cold 4% PFA in 1x PBS. Relevant tissues were then extracted and post-fixed overnight in 4% PFA in 1x PBS at 4 °C. For sectioning, brain and liver were cryoprotected through immersion in 30% sucrose in 1x PBS. Once the tissue had sunk, it was flash-frozen in O.C.T. Compound using a dry ice-ethanol bath and kept at -70 °C until sectioning.

Sections were obtained using a cryostat (Leica Biosystems). Fixed tissue was sectioned at 80 µm, collected in 1x PBS, and stored at 4 °C until use.

Immunohistochemistry

A detailed protocol for immunohistochemistry (IHC) on mouse brain slices is available on protocols.io (dx.doi.org/10.17504/protocols.io.5qpvokmq7l4o/v1). For IHC detection of Cacna1a, sections were first mounted onto slides and subjected to heat-induced epitope retrieval by boiling in 1x citrate buffer, pH 6 (Sigma-Aldrich, C9999) for 10 min in a microwave, followed by thorough washing with 1x PBS.

Sections were blocked in BlockAid Blocking Solution (ThermoFisher, B10710) with 0.1% Triton X-100 (Sigma-Aldrich, #93443). Primary and secondary antibodies were diluted in this blocking buffer. Tissue was incubated with primary antibody overnight at 4 °C and with secondary antibody for 2 hr at room temperature. Following each antibody incubation step, sections were washed 3 times for 10 min each in 1x PBS with 0.1% Triton X-100. For Hoechst labeling, sections were incubated for 10 min with 1/10000 Hoechst 33342 (ThermoFisher, H3570) in 1x PBS, followed by 3 washes in 1x PBS. Sections were allowed to dry on slides, and then a coverslip was mounted using Prolong Diamond Antifade Mountant (ThermoFisher, P36965).

Rabbit anti-Cacna1a (1:100, Alomone Labs, ACC-001; RRID:AB_2039764) was used. Fluorophore-conjugated F(ab')$_2$ fragment secondary antibodies (Jackson ImmunoResearch) were used at a 1:1000 working concentration.

Imaging

All other imaging, a Zeiss LSM 880 was used, with a 10x, 0.45 NA air objective. Imaging settings were chosen to capture full dynamic range of the signal without saturating pixels. When possible, laser power was adjusted before adjusting detector gain. Imaging settings were first optimized on control samples, before imaging of experimental samples. Fields of view were chosen while imaging non-experimental channels (e.g., Hoechst or Nissl).

Image analysis for fluorescent protein expression and IHC

To quantify CRISPR-Cas9 editing of the Ai14 locus, tdTomato-positive cells were manually counted using Fiji. Three volumes (850 µm x 850 µm x 64 µm) were captured from each of at least 4 non-adjacent sections per animal. PCs and non-PCs were differentiated based on distinct cell morphology and location.

For analysis of Cacna1a expression in cerebellum, Fiji was also used. Four maximum intensity projections of 850 µm x 850 µm x 30 µm volumes were analyzed per animal. In each image, the molecular layer (ML) and granular layer (GL) were manually segmented, and the total average fluorescence intensity was measured in those regions. For each image, the ML intensity was divided by the GL intensity, and then a per-animal average was determined.

Animal behaviour

Detailed protocols for the following behavioural assays are available on protocols.io (dx.doi.org/10.17504/protocols.io.6qpvr8jbzlmk/v2). On each day of behavioural training and data collection, animals were acclimated to the testing room for at least 30 min before measurements were taken. Animals were trained on beam crossing and gait measurement assays 1-2 weeks before experimental measurements started. Behaviour equipment was disinfected and deodorized between each animal or, in the case of the open-field test, between each cage.

The open-field apparatus consisted of four square arenas (27 cm x 27 cm), with a camera (EverFocus, EQ700) placed 1.83 m above the floor of the arenas. EthoVision XT (Noldus, v17.5; https://www.noldus.com/ethovision-xt; RRID:SCR_000441) was used to capture and subsequently analyze animal locomotion. Each trial consisted of a 2 min habituation period, followed by a 10 min test period. To avoid confounds due to odours from non-cagemates, only

animals from the same cage were recorded simultaneously. The average velocity over the course of the experimental period was determined.

The inverted wire hang test was used to measure limb strength[222]. Animals were placed onto a wire mesh screen (6 mm x 6 mm mesh), which was then inverted over the top of a 45 cm tall cylinder with clean bedding in the bottom. A blinded experimenter recorded the latency to fall within a max trial period of 120 s. Three trials were recorded and the average of those three trials used.

To measure skilled locomotion using the narrowing beam assay, a clear plexiglass beam consisting of three 25 cm segments (widths 3.5 cm, 2.5 cm, and 1.5 cm) was elevated above the table surface using empty clean cages, according to published protocol[223]. At the narrow end, an empty cage was placed on its side and bedding from the animal's home cage placed inside. A white light was also placed over the broad end to motivate animals to move across the beam. For each trial, animals were placed at the end of the widest segment, with all 4 limbs touching the beam surface. Each trial was recorded with a video camera placed to the side and perpendicular to the beam's length, affording a view of both left and right hindlimbs. A trial was considered complete once the animal had traversed the beam, without turning around, and entered the goal cage. Once an animal had completed three trials, the session was completed. For each trial, a blinded experimenter measured the animal's time to cross the beam (ignoring time spent paused), and assigned a neurological score[224]: (7) traverses the beam successfully, with no more than 4 foot slips and does not grip the side of the beam, (6) traverses the beam successfully, using hindlimbs to aid in more than 50% of strides, (5) traverses the beam successfully, using hindlimbs to aid in less than 50% of strides, (4) traverses the beam successfully, using a hindlimb at least once to push forward, but without bearing load on limb, (3) traverses beam successfully, by dragging hindlimbs without using them to push forward, (2) moves at least 1 body length, but fails to traverse beam in 120 s trial period or falls off, (1) fails to traverse beam or falls off, and does not move more than 1 body length. The average score and traversal time of the three trials was used for data presentation and statistics.

For gait analysis, we used MouseWalker, according to published protocols for hardware design and analysis[225,226]. A clear acrylic platform, 80 cm long, with a 5.3 cm corridor flanked by 12.5 cm high walls was used. LED lights positioned around the platform enable tracking of animal contacts with the platform surface, through frustrated total internal reflection (fTIR) that is captured using a camera (iPhone 12 Pro, Apple) positioned under the platform. Mice were placed on one end of the corridor, and fTIR recorded as the animal moved across the platform. Animals

were recorded until they had completed 3 continuously moving traversals of the field of view. Data were analyzed using MouseWalker.

EEG recording and analysis

A detailed protocol for mouse EEG implantation surgery and EEG data collection is available on protocols.io (dx.doi.org/10.17504/protocols.io.81wgbzj2ygpk/v1). EEG recordings were conducted in clear Plexiglas cylinders (25 cm wide, 30 cm high) with *ad libitum* water. Mice were connected to a pre-amplifier (100x gain, 0.5 Hz high-pass EEG filter; Pinnacle Technology, #8208-SL), which was attached to a commutator (Pinnacle Technology, #8204). Data were acquired by Sirenia Acquisition (Pinnacle Technology, v2.2.12; https://www.pinnaclet.com/software.html; RRID:SCR_016183), using a Pinnacle data conditioning and acquisition system (Pinnacle Technology, #8206), at a sampling rate of 400 Hz.

Mice were first habituated to the chamber for 1 session, at least 1 day before recordings began. For each session, a minimum of 90 min was recorded; only the last 60 min were analyzed. To assess ethosuximide blockade of absence seizures, mice were recorded for 90 min, then received a single i.p. injection of ethosuximide (200 mg/kg in sterile saline; Sigma-Aldrich, E7138), and then recorded for another 90 min. Only the last 60 min of the pre- and post-ethosuximide recordings were analyzed. Ethosuximide blockade experiments were performed 8 weeks after AAV injection.

EEG signal was analyzed using Sirenia Seizure Pro (Pinnacle Technology, v2.2.13; https://www.pinnaclet.com/software.html; RRID:SCR_016184). The raw EEG signal was first bandpass filtered (1-25 Hz). A sliding window (0.8 s wide, 0.4 s increments) was used to automatically detect absence seizures using the following criteria: a root mean square (RMS) power exceeding 50 μV in the 5-8 Hz band and a mean amplitude at least 2-fold higher than the baseline defined during the pre-injection recording session.

Statistics and reproducibility

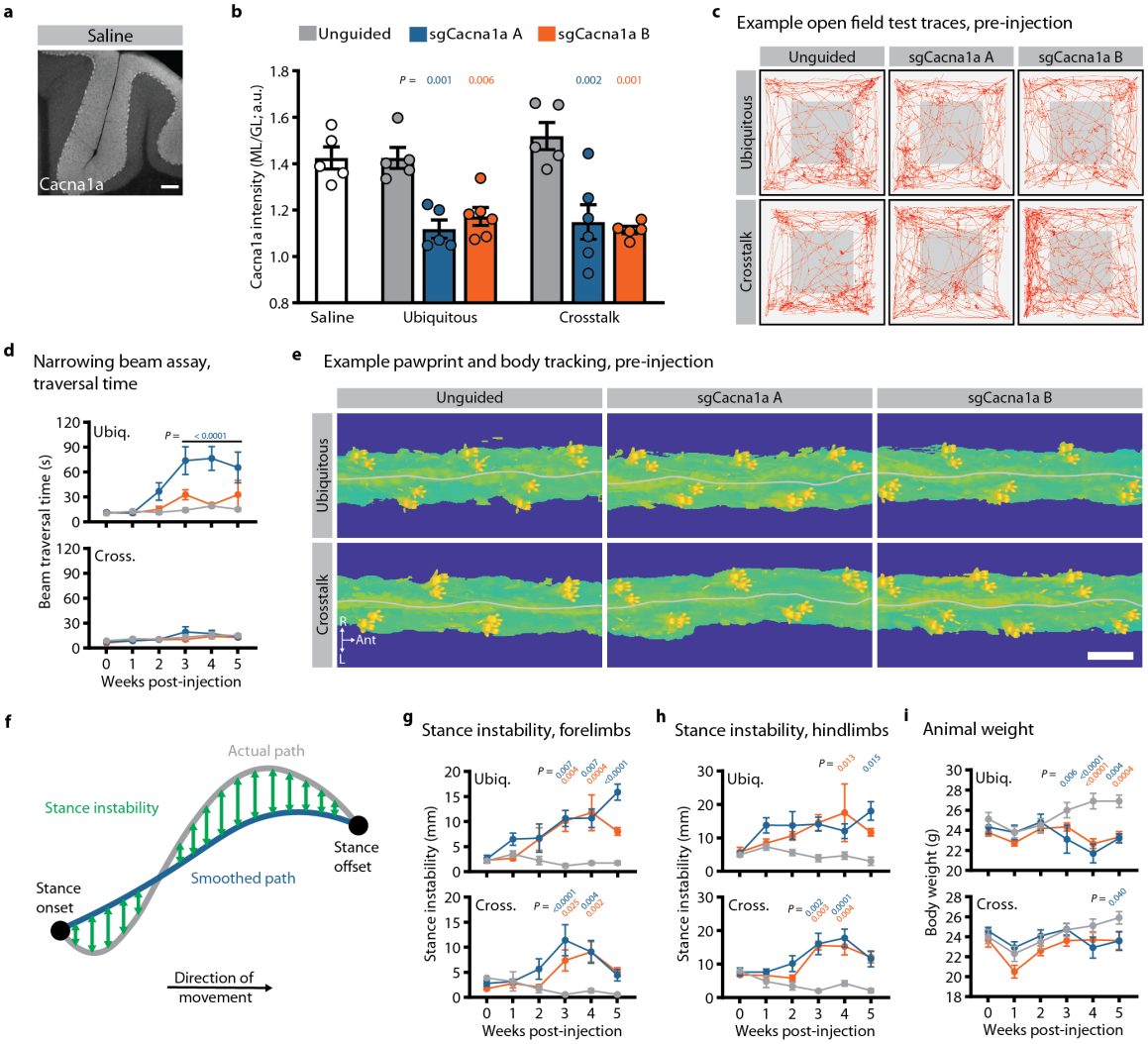
The number of biological replicates for each experiment are included in the corresponding figure legends. No data were excluded from analyses. For all violin plots, the middle dashed line is the median and the upper and lower dashed lines are quartiles. Statistical analysis was performed with GraphPad Prism (v10.0.3, GraphPad Software; RRID:SCR_002798) as described in figure legends. Where relevant, all tests were two-tailed and corrected for multiple comparisons to maintain an experiment-wide alpha of 0.05.

The following *in vivo* experiments were repeated once ($n > 2$ animals per experimental condition) with similar results: Figure 5.2b and Supplementary Figure 5.1a,b. The remaining *in vivo* experiments were not independently repeated.

Data availability

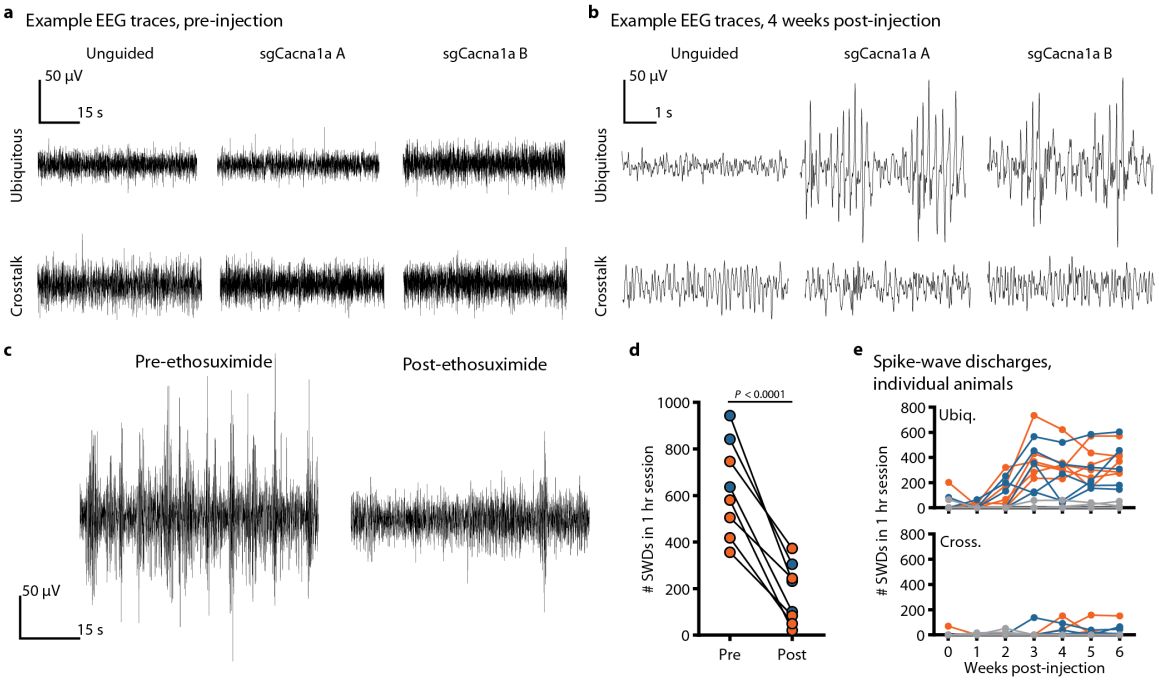
All sequences of primers, probes, sgRNAs, and other sequence elements are provided in Supplementary Table 3 of Coughlin et al., 2025. Images of brain tissue that are quantified in this work are deposited in the Brain Image Library (https://doi.org/10.35077/g.1163). Tabular datasets supporting conclusions of this work are available on Zenodo (10.5281/zenodo.13952929). All other data that support the findings of this study are available from the corresponding authors upon reasonable request.

5.6 Supplementary Figures



Supplementary Figure 5.1. Transcriptional crosstalk-mediated knockout of Cacna1a in Purkinje cells results in ataxic phenotypes. a,b, Related to Figure 5.2b. **a**, Representative image of IHC against Cacna1a in cerebellum of saline-injected control animal. Scale bar = 100 µm. **b**, Quantification of Cacna1a staining in cerebellum, normalized as the intensity in the molecular layer (ML) divided by the intensity in the granular layer (GL). Statistical significance was determined by one-way ANOVA and Dunnett's multiple comparison tests against Cacna1a intensity in age-matched saline-injected mice. **c**, Related to Figure 5.2c,d. Example open field test traces acquired pre-injection, showing same animals as in Figure 5.2c. Red lines represent animal position over a 10 min trial. **d**, Related to Figure 5.2e. Beam traversal time for narrowing beam assay following ubiquitous and PC-specific disruption of *Cacna1a*. **e-h**, Related to Figure 5.2g,h. e, Example pre-injection pawprint and body tracking, over a small segment of the elevated plexiglass platform used for gait analysis, showing same animals as in Figure 5.2g. Grey line indicates midline of body. Scale bar = 3 cm. **f**, Schematic to demonstrate stance instability metric. For each stance cycle, tracking of body center relative to paw location yields a stance trace (grey

line) and a smoothed version of that trace (blue line). The summed difference between the actual path and the smoothed path corresponds to the stance instability. **g,h**, Stance instability for forelimbs (g) and hindlimbs (h) following ubiquitous and PC-specific disruption of *Cacna1a*. **i**, Animal weights following ubiquitous and PC-specific disruption of *Cacna1a*. Statistical significance for beam crossing and stance instability was determined by two-way repeated-measures ANOVA and Dunnett's multiple comparison tests against behavioural performance at 0-week time point. Statistical significance for animal weight was determined by two-way repeated-measures ANOVA and Dunnett's multiple comparison tests against the unguided condition. Points and bars represent mean ± s.e.m. For all groups, $n = 5$ except (ubiquitous + sgCacna1a B) and (crosstalk + sgCacna1a A) groups, in which $n = 6$. For all plots, grey line = unguided controls, blue = sgCacna1a A, and orange = sgCacna1a B.



Supplementary Figure 5.2. Transcriptional crosstalk-mediated knockout of Cacna1a in Purkinje cells does not result in absence seizures. Related to Figure 5.2i-k. **a**, Example cortical EEG traces, acquired pre-injection, showing same animals as in Figure 5.2j. **b**, Higher temporal resolution cortical EEG traces, acquired post-injection, corresponding to those shown in Fig 5.2j. **c**, Example EEG traces acquired before and after ethosuximide administration. **d**, Number of spike-and-wave discharges (SWDs) detected in 1 hr sessions, before and after ethosuximide administration, in animals from the ubiquitous condition. The observed reduction in the number of detected SWDs provides confidence in our SWD detection and quantification pipeline. Statistical significance was determined using a paired t-test. **e**, Quantification of SWDs for individual animals in ubiquitous (top) and PC-specific conditions (bottom). For all plots, grey line = unguided controls, blue = sgCacna1a A, and orange = sgCacna1a B.

Chapter 6

CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, I attempted to address some outstanding bottlenecks in AAV engineering and usage: low throughput in characterizing the performance of engineered AAV vectors (Chapter 2), limited understanding of AAV genome processing (Chapter 3), interactions between AAV genomes that can confound AAV usage and characterization (Chapter 4), and lastly, the limited cargo space of AAVs (Chapter 5).

In Chapter 2, I tackled the issue of low throughput in characterizing engineered AAV vectors by developing and applying a novel high-throughput, high-sensitivity spatial transcriptomics platform (USeqFISH). This work demonstrated profiling of up to 13 AAV variants, in multiple brain regions and across as many as 26 molecularly defined cell types in the same animal. This ability to comprehensively profile tropism in a single animal can mitigate inter-animal variability, such as that arising from poorly reproducible injections. Furthermore, though parallel AAV characterization with cell type resolution has been accomplished through dissociation-based single-cell RNA sequencing[166,227,228], these methods result in a loss of spatial information and can be difficult to adapt to different sample types.

Our profiling with USeqFISH also exposed some novel biology. Though designed as proof-of-principle experiments to demonstrate the advantage of our approach, our work uncovered previously unrealized tropism biases of some engineered capsids, such as an excitatory neuron bias for the capsid AAV-PHP.N[81]. We also profiled a novel AAV capsid, PHP.AX, which shows a relatively low cell type bias. This capsid may have utility for screening of regulatory elements across a broad range of cell types, mitigating confounds due to the intrinsic bias from the AAV capsid. Further development of USeqFISH, through integration of a combinatorial or sequential barcoding scheme could enable profiling of larger pools or even libraries of AAV variants. Adaptation of this screening platform to more translationally relevant eukaryotic species, such as non-human primates, could accelerate identification of AAV variants with desirable properties for therapeutic applications.

Increased capacity to characterize AAV capsid and genome variants can accelerate existing AAV engineering pipelines. Further insight into AAV transduction may open new avenues for improving AAV-based gene delivery. In Chapter 3, I developed novel methods for subcellular localization of AAV genomes (AAV-Zombie) and concatemers (SpECTr), and applied these methods to provide detailed, multiparameter time courses of AAV transduction and genome

concatemerization *in vitro*. Furthermore, these methods work *in vivo*, providing single-cell and DNA-level views of brain transduction by engineered AAVs. These *in vivo* experiments highlight another potential bottleneck in systemic delivery of AAVs to the brain. With engineered AAV capsids, AAV-PHP.eB and AAV.CAP-B10, we observe strong accumulation of AAVs in brain endothelial cells at 1 day post-injection, but sparse presence of AAVs in parenchyma 14 days post-injection. Understanding the fate of the genomes that entered endothelial cells but failed to access the brain parenchyma may yield insights that can improve gene delivery to the brain.

The importance of AAV genome concatemerization is unknown, though circularization has been suggested to be important for persistence of active AAV episomes[24,25]. It is also feasible that concatemerization of AAV genomes increases expression by sequestering transcription factors into areas with a high concentration of AAV genomes. Indeed, our results provide some evidence that concatemerization is important for transgene expression. In our time course of AAV genome localization and concatemerization in primary neurons, we observed that the rise in reporter gene transcript began after we detected nuclear concatemers. Furthermore, in visualizing the correlation between reporter expression and genome localization and state, we observed the strongest correlation between reporter expression and counts of concatemers in the nucleus. Our newly developed spatial genomics techniques may facilitate further exploration into the importance of AAV genome processing.

Concatemerization of AAV genomes can also facilitate interactions between elements on co-delivered genomes. I explored this transcriptional crosstalk in Chapter 4. First, I demonstrated generalizability across promoters, enhancers, and tissue types. Additionally, I identified necessary components for this behaviour. These results support a model in which enhancer elements on one genome interact with promoter elements on the other genome. Using *Prkdc*$^{scid/scid}$ mice and the novel spatial genomics methods developed in Chapter 3, I then mechanistically linked AAV transcriptional crosstalk to genome concatemerization.

This work identified a hurdle for co-delivery of enhancer-driven AAV genomes. In particular, enhancer screening efforts using AAVs can benefit from pooled characterization of vectors by reducing animal numbers and variability and by increasing throughput. Pooled characterization will likely be necessary for enhancer screening efforts in more translationally relevant non-human primate species. Indeed, some groups have already used pooled approaches for AAV enhancer screening[120,125,197,201]. Consistent with our findings, two groups recently noted unexpected interaction between enhancer constructs arising from transcriptional crosstalk in their pooled screens[130,197]. Hunker and colleagues[229] further characterized the noise in AAV enhancer pools arising from transcriptional crosstalk, noting that the degree of transcriptional crosstalk was

dependent on the particular enhancer and targeted cell types. As with our results, transcriptional crosstalk could be mitigated, but not eliminated, using *Prkdc*$^{scid/scid}$ mice[202]. Thus, eliminating the confound of AAV crosstalk in pooled enhancer screening will require development of additional strategies, such as incorporation of insulator elements[203–205]. Methods to better understand AAV genome processing, including AAV-Zombie and SpECTr, may yield additional insights that can direct development of such strategies. Finally, transcriptional crosstalk is not the only source of noise for pooled packaging and characterization of AAV genome variants. Notably, Lalanne and colleagues[229] demonstrated a high incidence (20%-60%) of barcode swapping during AAV packaging. Thus, strategies to prevent barcode swapping during AAV packaging will also be necessary to accelerate identification and characterization of AAV-compatible enhancer sequences.

Finally, transcriptional crosstalk also presents an opportunity for cell type-specific delivery of large cargo, by allowing for separation of long coding sequences and enhancer elements onto separate AAV genomes. Using a long Purkinje cell-targeting element (1.65 kb) and SaCas9 (3.2 kb), along with systemically administered, blood-brain barrier-penetrant AAVs, I explored this application of transcriptional crosstalk in Chapter 5. Capitalizing on a common reporter assay, I first demonstrated cell type-targeted editing that avoided off-target editing, before demonstrating that our targeting was efficient enough to recapitulate known loss-of-function phenotypes. Knockout of *Cacna1a* in Purkinje cells through classical genetics manipulation results in ataxia, whereas loss-of-function in forebrain is expected to lead to emergence of absence epilepsy. Transcriptional crosstalk-mediated cell type-targeted disruption of *Cacna1a* recapitulated these ataxic phenotypes. Reassuringly, our crosstalk-mediated targeted gene disruption was specific enough to prevent emergence of cortical epileptiform activity, which has been observed using Purkinje cell cre-driver lines but attributed to potential leakiness of those lines. Usually, such experiments would take months to years if employing classical genetics approaches for cell type targeting based on recombinase driver lines. Making use of easily obtained wildtype animals and systemically administered AAVs, we were able to generate sufficiently powered cohorts within 2 weeks. Potentially, this strategy could be used for cell type-specific reverse genetics screens in mammalian systems.

In addition to these applications in genetics and disease modeling, crosstalk-mediated cell type targeting may provide a strategy for targeted therapeutic gene delivery. Cell type-specific expression of therapeutic effectors, including functional copies of endogenous genes as well as machinery for precision gene editing, may reduce side effects. Enhancer-driven AAV vectors are an attractive strategy to achieve such high and/or cell type-specific expression, but the limited

cargo capacity of AAV vectors can complicate integration of enhancers with longer coding sequences. Indeed, there are several human disorders in which enhancer-mediated targeted would be challenging due to coding sequence length; a sample of these is presented in Table 6.1.

Disorder (reference(s))	Human Gene (bp)	Target Cell Type	Targeting Rationale
Alpha-1 antitrypsin deficiency[230,231]	SERPINA1 (1254)	Lung, Liver	Mutations in SERPINA1 cause accumulation in the liver and underexpression in the lung. Enhancing lung expression while minimizing liver expression would be beneficial. Though this transcript is relatively short, crosstalk-mediated enhancement may be necessary due to potential need for several regulatory motifs on the genome.
CDKL5 Deficiency Disorder[232]	CDKL5 (3427)	Cerebral Cortex, Cerebellum	CDKL5 is primarily expressed in neurons, both excitatory and inhibitory, and highly expressed in early development. Inclusion of a temporally sensitive enhancer element could promote high expression during the critical period, while maintaining a lower baseline expression level.
Cornelia De Lange Syndrome[233–236]	BRD4 (4089), SMC1A (3969)	Cerebral Cortex Neurons, Epithelial Cells, Smooth Muscle and Enteric Nervous System	Brd4 overexpression is linked to cancer progression and inflammatory dysregulation, emphasizing the incentive to restrict unnecessary expression in off-target cell types. Smc1a overexpression can dysregulate chromosomal segregation in mitosis and meiosis.
Cystic Fibrosis[237]	CFTR (4440)	Secretory Epithelia in Lung, Pancreas	The coding sequence length for CFTR limits promoter and regulatory element choices in a single AAV genome. Additionally, CFTR is not expressed in the muscle, heart, neurons or blood cells. Expressing chloride channels in non-target cell types could perturb homeostatic processes
Duchenne Muscular Dystrophy[238]	mini- or micro-dystrophin (varies)	Skeletal Muscle	Dystrophin, including miniaturized constructs, have long sequences, which limit the inclusion of gene regulatory elements.
Osteogenesis Imperfecta; Ehlers Danlos Syndrome[239,240]	COL1A1 (4392)	FSP+ mesenchymal cells, fibroblasts	Subpopulations of COL1A1 producing cells are more specifically responsible for the OI phenotype. Aberrant overexpression of COL1A1 may lead to extensive tissue fibrosis and complications, particularly in the vascular system.
Pyruvate Carboxylase Deficiency[241]	PC (3534)	Hepatocytes, Astrocytes	Overexpression or off target expression could lead to metabolic imbalances, including overaccumulation of downstream metabolites
Spinocerebellar Ataxia 2; Parkinsons Disease[242,243]	Atxn2 (3948)	Purkinje Cells, Basal Ganglia Neurons	Different cell populations can be targeted for different pathologies using the same transgene; Purkinje cells for SCA2, but to Basal Ganglia for PD
Spinocerebellar Ataxia 44[244,245]	Grm1 (3585)	Purkinje Cells	Some patients display both cerebellar and cortical involvement, but Grm1 is expressed at much higher levels in Purkinje cells.
SYNGAP1-related intellectual disability[246]	SYNGAP1 (3879)	Hippocampal neurons, Cerebellar granule cells	SYNGAP1 is not expressed in non-neuronal cell types in the CNS, and its sequence length limits inclusion of additional regulatory elements in AAV genome.
Wilson's Disease[247]	ATP7B (4410)	Hepatocytes, Basal Ganglia Neurons, Cornea	Hepatocytes express ATP7B at a much higher level than other tissues, thus enhancement may be desirable there while not overexpressing in other tissues

Table 6.1. Indications where transcriptional crosstalk may be therapeutically beneficial to achieve high and/or cell type-specific expression.

Integration of additional regulatory schemes with transcriptional crosstalk may further hone expression. Though our transcriptional crosstalk-mediated Purkinje cell-targeted knockout of *Cacna1a* in Chapter 5 did not result in a significant increase in epileptiform activity in the cortex, there were some individual animals that did show relatively small increases in epileptiform activity. This result may suggest some leak expression of SaCas9 and thus loss-of-function in the cortex. As enhancers regulate gene expression transcriptionally, post-transcriptional regulators may be easily integrated with transcriptional crosstalk. Our profiles of miRNA target site (miRNA TS)-based repression (Chapter 2) provide some starting points to further hone crosstalk-mediated targeting. Furthermore, because miRNA TSs are short, this strategy would have minimal disruption for cell type-targeted expression of large coding sequences. Placing the expression of the sgRNA under control of the enhancer-driven RNA polymerase II promoter, to achieve cell type-specific expression of both sgRNA and the effector protein, may provide an additional strategy that is also orthogonal to miRNA TS-based repression. Layering levels of regulation may be necessary to achieve the desired expression pattern and strength.

Throughout this thesis, I have explored and emphasized specificity in gene delivery. I have also emphasized the importance of understanding gene delivery vehicles at a subcellular level, and at multiple levels of analysis: DNA, RNA, and protein. The significance of this work extends beyond natural and engineered AAVs, to include non-viral and non-AAV viral gene delivery methods. Approaches used for engineering of viral delivery vectors are also being applied to non-viral vectors, such as virus-like particles[248,249]. Likewise, as summarized in Chapter 1, a small number of non-AAV viral vectors have been used to both scientific and therapeutic ends. The space of possible viral vectors is vastly larger than the number of viruses that have been vectorized; there are an estimated total of 10^9 virus species on Earth, with only 0.001% of those currently recognized and even fewer studied beyond being merely identified based on sequence[250]. Furthering our understanding of viruses and gene delivery vectors will necessitate diversifying our toolkit to probe viral and vector biology. In turn, this understanding may allow humans to address the current and rapidly changing disease landscape.

BIBLIOGRAPHY

1. Griffith, F. (1928). The significance of pneumococcal types. J. Hyg. *27*, 113–159. https://doi.org/10.1017/S0022172400031879.

2. Dawson, M.H., and Sia, R.H.P. (1931). In vitro transformation of pneumococcal types. J. Exp. Med. *54*, 681–699. https://doi.org/10.1084/jem.54.5.681.

3. Alloway, J.L. (1932). The transformation in vitro of R pneumococci into S forms of different specific types by the use of filtered pneumococcus extracts. J. Exp. Med. *55*, 91–99. https://doi.org/10.1084/jem.55.1.91.

4. Alloway, J.L. (1933). Further observations on the use of pneumococcus extracts in effecting transformation of type in vitro. J. Exp. Med. *57*, 265–278. https://doi.org/10.1084/jem.57.2.265.

5. Avery, O.T., MacLeod, C.M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J. Exp. Med. *79*, 137–158. https://doi.org/10.1084/jem.79.2.137.

6. Tatum, E.L., and Lederberg, J. (1947). Gene recombination in the bacterium *Escherichia coli*. J. Bacteriol. *53*, 673–684. https://doi.org/10.1128/jb.53.6.673-684.1947.

7. Zinder, N.D., and Lederberg, J. (1952). Genetic exchange in Salmonella. J. Bacteriol. *64*, 679–699. https://doi.org/10.1128/jb.64.5.679-699.1952.

8. Freeman, V.J. (1951). Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. J. Bacteriol. *61*, 675–688. https://doi.org/10.1128/jb.61.6.675-688.1951.

9. Szybalska, E.H., and Szybalski, W. (1962). Genetics of human cell lines, IV. DNA-mediated heritable transformation of a biochemical trait. Proc. Natl. Acad. Sci. U.S.A. *48*, 2026–2034. https://doi.org/10.1073/pnas.48.12.2026.

10. Friedmann, T. (1992). A brief history of gene therapy. Nat. Genet. *2*, 93–98. https://doi.org/10.1038/ng1092-93.

11. Temin, H.M. (1961). Mixed infection with two types of Rous sarcoma virus. Virology *13*, 158–163. https://doi.org/10.1016/0042-6822(61)90049-6.

12. Sambrook, J., Westphal, H., Srinivasan, P.R., and Dulbecco, R. (1968). The integrated state of viral DNA in SV40-transformed cells. Proc. Natl. Acad. Sci. U.S.A. *60*, 1288–1295. https://doi.org/10.1073/pnas.60.4.1288.

13. Rogers, S., and Pfuderer, P. (1968). Use of viruses as carriers of added genetic information. Nature *219*, 749–751. https://doi.org/10.1038/219749a0.

14. Shimotohno, K., and Temin, H.M. (1981). Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase gene into DNA of an avian retrovirus. Cell *26*, 67–77. https://doi.org/10.1016/0092-8674(81)90034-9.

15. Tabin, C.J., Hoffmann, J.W., Goff, S.P., and Weinberg, R.A. (1982). Adaptation of a retrovirus as a eucaryotic vector transmitting the herpes simplex virus thymidine kinase gene. Mol. Cell. Biol. *2*, 426–436. https://doi.org/10.1128/mcb.2.4.426-436.1982.

16. Wei, C.M., Gibson, M., Spear, P.G., and Scolnick, E.M. (1981). Construction and isolation of a transmissible retrovirus containing the src gene of Harvey murine sarcoma virus and the thymidine kinase gene of herpes simplex virus type 1. J. Virol. *39*, 935–944. https://doi.org/10.1128/jvi.39.3.935-944.1981.

17. Williams, D.A., Lemischka, I.R., Nathan, D.G., and Mulligan, R.C. (1984). Introduction of new genetic material into pluripotent haematopoietic stem cells of the mouse. Nature *310*, 476–480. https://doi.org/10.1038/310476a0.

18. Miller, A.D., Jolly, D.J., Friedmann, T., and Verma, I.M. (1983). A transmissible retrovirus expressing human hypoxanthine phosphoribosyltransferase (HPRT): gene transfer into cells obtained from humans deficient in HPRT. Proc. Natl. Acad. Sci. U.S.A. *80*, 4709–4713. https://doi.org/10.1073/pnas.80.15.4709.

19. Willis, R.C., Jolly, D.J., Miller, A.D., Plent, M.M., Esty, A.C., Anderson, P.J., Chang, H.C., Jones, O.W., Seegmiller, J.E., and Friedmann, T. (1984). Partial phenotypic correction of human Lesch-Nyhan (hypoxanthine-guanine phosphoribosyltransferase-deficient) lymphoblasts with a transmissible retroviral vector. J. Biol. Chem. *259*, 7842–7849. https://doi.org/10.1016/S0021-9258(17)42870-5.

20. Hoggan, M.D., Blacklow, N.R., and Rowe, W.P. (1966). 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. https://doi.org/10.1073/pnas.55.6.1467.

21. Harui, A., Suzuki, S., Kochanek, S., and Mitani, K. (1999). Frequency and stability of chromosomal integration of adenovirus vectors. J. Virol. *73*, 6141–6146. https://doi.org/10.1128/JVI.73.7.6141-6146.1999.

22. McCarty, D.M., Young, S.M., and Samulski, R.J. (2004). Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu. Rev. Genet. *38*, 819–845. https://doi.org/10.1146/annurev.genet.37.110801.143717.

23. Hanlon, K.S., Kleinstiver, B.P., Garcia, S.P., Zaborowski, M.P., Volak, A., Spirig, S.E., Muller, A., Sousa, A.A., Tsai, S.Q., Bengtsson, N.E., et al. (2019). High levels of AAV vector integration into CRISPR-induced DNA breaks. Nat. Commun. *10*, 4439. https://doi.org/10.1038/s41467-019-12449-2.

24. Wang, J.-H., Gessler, D.J., Zhan, W., Gallagher, T.L., and Gao, G. (2024). Adeno-associated virus as a delivery vector for gene therapy of human diseases. Sig. Transduct. Target. Ther. *9*, 78. https://doi.org/10.1038/s41392-024-01780-w.

25. Wang, D., Tai, P.W.L., and Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. Nat. Rev. Drug Discov. *18*, 358–378. https://doi.org/10.1038/s41573-019-0012-9.

26. Bedbrook, C.N., Deverman, B.E., and Gradinaru, V. (2018). Viral strategies for targeting the central and peripheral nervous systems. Annu. Rev. Neurosci. *41*, 323–348. https://doi.org/10.1146/annurev-neuro-080317-062048.

27. Challis, R.C., Ravindra Kumar, S., Chen, X., Goertsen, D., Coughlin, G.M., Hori, A.M., Chuapoco, M.R., Otis, T.S., Miles, T.F., and Gradinaru, V. (2022). Adeno-associated virus toolkit to target diverse brain cells. Annu. Rev. Neurosci. *45*, 447–469. https://doi.org/10.1146/annurev-neuro-111020-100834.

28. Pupo, A., Fernández, A., Low, S.H., François, A., Suárez-Amarán, L., and Samulski, R.J. (2022). AAV vectors: the Rubik's cube of human gene therapy. Mol. Ther. *30*, 3515–3541. https://doi.org/10.1016/j.ymthe.2022.09.015.

29. Mendell, J.R., Al-Zaidy, S., Shell, R., Arnold, W.D., Rodino-Klapac, L.R., Prior, T.W., Lowes, L., Alfano, L., Berry, K., Church, K., et al. (2017). Single-dose gene-replacement therapy for spinal muscular atrophy. N. Engl. J. Med. *377*, 1713–1722. https://doi.org/10.1056/NEJMoa1706198.

30. Mercuri, E., Muntoni, F., Baranello, G., Masson, R., Boespflug-Tanguy, O., Bruno, C., Corti, S., Daron, A., Deconinck, N., Servais, L., et al. (2021). Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy type 1 (STR1VE-EU): an open-label, single-arm, multicentre, phase 3 trial. Lancet Neurol. *20*, 832–841. https://doi.org/10.1016/S1474-4422(21)00251-9.

31. Day, J.W., Finkel, R.S., Chiriboga, C.A., Connolly, A.M., Crawford, T.O., Darras, B.T., Iannaccone, S.T., Kuntz, N.L., Peña, L.D.M., Shieh, P.B., et al. (2021). Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy in patients with two copies of SMN2 (STR1VE): an open-label, single-arm, multicentre, phase 3 trial. Lancet Neurol. *20*, 284–293. https://doi.org/10.1016/S1474-4422(21)00001-6.

32. Strauss, K.A., Farrar, M.A., Muntoni, F., Saito, K., Mendell, J.R., Servais, L., McMillan, H.J., Finkel, R.S., Swoboda, K.J., Kwon, J.M., et al. (2022). Onasemnogene abeparvovec for presymptomatic infants with two copies of SMN2 at risk for spinal muscular atrophy type 1: the Phase III SPR1NT trial. Nat. Med. *28*, 1381–1389. https://doi.org/10.1038/s41591-022-01866-4.

33. Strauss, K.A., Farrar, M.A., Muntoni, F., Saito, K., Mendell, J.R., Servais, L., McMillan, H.J., Finkel, R.S., Swoboda, K.J., Kwon, J.M., et al. (2022). Onasemnogene abeparvovec for presymptomatic infants with three copies of SMN2 at risk for spinal muscular atrophy: the Phase III SPR1NT trial. Nat. Med. *28*, 1390–1397. https://doi.org/10.1038/s41591-022-01867-3.

34. Jiang, Z., and Dalby, P.A. (2023). Challenges in scaling up AAV-based gene therapy manufacturing. Trends Biotechnol. *41*, 1268–1281. https://doi.org/10.1016/j.tibtech.2023.04.002.

35. Earley, J., Piletska, E., Ronzitti, G., and Piletsky, S. (2023). Evading and overcoming AAV neutralization in gene therapy. Trends Biotechnol. *41*, 836–845. https://doi.org/10.1016/j.tibtech.2022.11.006.

36. Li, C., and Samulski, R.J. (2020). Engineering adeno-associated virus vectors for gene therapy. Nat. Rev. Genet. *21*, 255–272. https://doi.org/10.1038/s41576-019-0205-4.

37. Sonntag, F., Köther, K., Schmidt, K., Weghofer, M., Raupp, C., Nieto, K., Kuck, A., Gerlach, B., Böttcher, B., Müller, O.J., et al. (2011). The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. J. Virol. *85*, 12686–12697. https://doi.org/10.1128/JVI.05359-11.

38. Sonntag, F., Schmidt, K., and Kleinschmidt, J.A. (2010). A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc. Natl. Acad. Sci. U.S.A. *107*, 10220–10225. https://doi.org/10.1073/pnas.1001673107.

39. Elmore, Z.C., Patrick Havlik, L., Oh, D.K., Anderson, L., Daaboul, G., Devlin, G.W., Vincent, H.A., and Asokan, A. (2021). The membrane associated accessory protein is an adeno-associated viral egress factor. Nat. Commun. *12*, 6239. https://doi.org/10.1038/s41467-021-26485-4.

40. Aksu Kuz, C., Ning, K., Hao, S., Cheng, F., and Qiu, J. (2024). Role of the membrane-associated accessory protein (MAAP) in adeno-associated virus (AAV) infection. J. Virol. *98*, e00633-24. https://doi.org/10.1128/jvi.00633-24.

41. Srivastava, A. (2016). In vivo tissue-tropism of adeno-associated viral vectors. Curr. Opin. Virol. *21*, 75–80. https://doi.org/10.1016/j.coviro.2016.08.003.

42. Foust, K.D., Nurre, E., Montgomery, C.L., Hernandez, A., Chan, C.M., and Kaspar, B.K. (2009). Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes. Nat. Biotechnol. *27*, 59–65. https://doi.org/10.1038/nbt.1515.

43. McCarty, D., Monahan, P., and Samulski, R. (2001). Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther. *8*, 1248–1254. https://doi.org/10.1038/sj.gt.3301514.

44. Berns, K.I., and Muzyczka, N. (2017). AAV: an overview of unanswered questions. Hum. Gene Ther. *28*, 308–313. https://doi.org/10.1089/hum.2017.048.

45. Penaud-Budloo, M., Le Guiner, C., Nowrouzi, A., Toromanoff, A., Chérel, Y., Chenuaud, P., Schmidt, M., Von Kalle, C., Rolling, F., Moullier, P., et al. (2008). Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. J. Virol. *82*, 7875–7885. https://doi.org/10.1128/JVI.00649-08.

46. Duan, D., Sharma, P., Yang, J., Yue, Y., Dudus, L., Zhang, Y., Fisher, K.J., and Engelhardt, J.F. (1998). Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. J. Virol. *72*, 8568–8577. https://doi.org/10.1128/JVI.72.11.8568-8577.1998.

47. Nakai, H., Yant, S.R., Storm, T.A., Fuess, S., Meuse, L., and Kay, M.A. (2001). Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. J. Virol. *75*, 6969–6976. https://doi.org/10.1128/JVI.75.15.6969-6976.2001.

48. Schnepp, B.C., Chulay, J.D., Ye, G.-J., Flotte, T.R., Trapnell, B.C., and Johnson, P.R. (2016). Recombinant adeno-associated virus vector genomes take the form of long-lived, transcriptionally competent episomes in human muscle. Hum. Gene Ther. *27*, 32–42. https://doi.org/10.1089/hum.2015.136.

49. Yang, J., Zhou, W., Zhang, Y., Zidon, T., Ritchie, T., and Engelhardt, J.F. (1999). Concatamerization of adeno-associated virus circular genomes occurs through intermolecular recombination. J. Virol. *73*, 9468–9477. https://doi.org/10.1128/JVI.73.11.9468-9477.1999.

50. Greig, J.A., Martins, K.M., Breton, C., Lamontagne, R.J., Zhu, Y., He, Z., White, J., Zhu, J.-X., Chichester, J.A., Zheng, Q., et al. (2024). Integrated vector genomes may contribute to long-term expression in primate liver after AAV administration. Nat. Biotechnol. *42*, 1232–1242. https://doi.org/10.1038/s41587-023-01974-7.

51. Sabatino, D.E., Bushman, F.D., Chandler, R.J., Crystal, R.G., Davidson, B.L., Dolmetsch, R., Eggan, K.C., Gao, G., Gil-Farina, I., Kay, M.A., et al. (2022). Evaluating the state of the science for adeno-associated virus integration: an integrated perspective. Mol. Ther. *30*, 2646–2663. https://doi.org/10.1016/j.ymthe.2022.06.004.

52. Duan, D., Yue, Y., and Engelhardt, J.F. (2003). Consequences of DNA-dependent protein kinase catalytic subunit deficiency on recombinant adeno-associated virus genome circularization and heterodimerization in muscle tissue. J. Virol. *77*, 4751–4759. https://doi.org/10.1128/jvi.77.8.4751-4759.2003.

53. Song, S., Laipis, P.J., Berns, K.I., and Flotte, T.R. (2001). Effect Of DNA-dependent protein kinase on the molecular fate of the rAAV2 genome in skeletal muscle. Proc. Natl. Acad. Sci. U.S.A. *98*, 4084–4088. https://doi.org/10.1073/pnas.061014598.

54. Nakai, H., Storm, T.A., Fuess, S., and Kay, M.A. (2003). Pathways of removal of free DNA vector ends in normal and DNA-PKcs–deficient SCID mouse hepatocytes transduced with rAAV vectors. Hum. Gene Ther. *14*, 871–881. https://doi.org/10.1089/104303403765701169.

55. Choi, V.W., McCarty, D.M., and Samulski, R.J. (2006). Host cell DNA repair pathways in adeno-associated viral genome processing. J. Virol. *80*, 10346–10356. https://doi.org/10.1128/JVI.00841-06.

56. Choi, Y.-K., Nash, K., Byrne, B.J., Muzyczka, N., and Song, S. (2010). The effect of DNA-dependent protein kinase on adeno-associated virus replication. PLoS ONE *5*, e15073. https://doi.org/10.1371/journal.pone.0015073.

57. Maurer, A.C., and Weitzman, M.D. (2020). Adeno-associated virus genome interactions important for vector production and transduction. Hum. Gene Ther. *31*, 499–511. https://doi.org/10.1089/hum.2020.069.

58. Kim, J.-J., Kurial, S.N.T., Choksi, P.K., Nunez, M., Lunow-Luke, T., Bartel, J., Driscoll, J., Her, C.L., Dhillon, S., Yue, W., et al. (2025). AAV capsid prioritization in normal and steatotic human livers maintained by machine perfusion. Nat. Biotechnol. https://doi.org/10.1038/s41587-024-02523-6.

59. Dhungel, B.P., Bailey, C.G., and Rasko, J.E.J. (2021). Journey to the center of the cell: tracing the path of AAV transduction. Trends Mol. Med. *27*, 172–184. https://doi.org/10.1016/j.molmed.2020.09.010.

60. Riyad, J.M., and Weber, T. (2021). Intracellular trafficking of adeno-associated virus (AAV) vectors: challenges and future directions. Gene Ther. *28*, 683–696. https://doi.org/10.1038/s41434-021-00243-z.

61. Nicolson, S.C., and Samulski, R.J. (2014). Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus. J. Virol. *88*, 4132–4144. https://doi.org/10.1128/JVI.02660-13.

62. Hauck, B., Zhao, W., High, K., and Xiao, W. (2004). Intracellular viral processing, not single-stranded DNA accumulation, is crucial for recombinant adeno-associated virus transduction. J. Virol. *78*, 13678–13686. https://doi.org/10.1128/JVI.78.24.13678-13686.2004.

63. Ferrari, F.K., Samulski, T., Shenk, T., and Samulski, R.J. (1996). Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J. Virol. *70*, 3227–3234. https://doi.org/10.1128/jvi.70.5.3227-3234.1996.

64. Fisher, K.J., Gao, G.P., Weitzman, M.D., DeMatteo, R., Burda, J.F., and Wilson, J.M. (1996). Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. J. Virol. *70*, 520–532. https://doi.org/10.1128/jvi.70.1.520-532.1996.

65. Bowles, D.E., McPhee, S.W., Li, C., Gray, S.J., Samulski, J.J., Camp, A.S., Li, J., Wang, B., Monahan, P.E., Rabinowitz, J.E., et al. (2012). Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector. Mol. Ther. *20*, 443–455. https://doi.org/10.1038/mt.2011.237.

66. Shen, S., Horowitz, E.D., Troupes, A.N., Brown, S.M., Pulicherla, N., Samulski, R.J., Agbandje-McKenna, M., and Asokan, A. (2013). Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency. J. Biol. Chem. *288*, 28814–28823. https://doi.org/10.1074/jbc.M113.482380.

67. Girod, A., Ried, M., Wobus, C., Lahm, H., Leike, K., Kleinschmidt, J., Deléage, G., and Hallek, M. (1999). Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. Nat. Med. *5*, 1052–1056. https://doi.org/10.1038/12491.

68. Zhong, L., Li, B., Mah, C.S., Govindasamy, L., Agbandje-McKenna, M., Cooper, M., Herzog, R.W., Zolotukhin, I., Warrington, K.H., Weigel-Van Aken, K.A., et al. (2008). Next generation of adeno-associated virus 2 vectors: point mutations in tyrosines lead to high-efficiency transduction at lower doses. Proc. Natl. Acad. Sci. U.S.A. *105*, 7827–7832. https://doi.org/10.1073/pnas.0802866105.

69. Petrs-Silva, H., Dinculescu, A., Li, Q., Min, S.-H., Chiodo, V., Pang, J.-J., Zhong, L., Zolotukhin, S., Srivastava, A., Lewin, A.S., et al. (2009). High-efficiency transduction of the mouse retina by tyrosine-mutant AAV serotype vectors. Mol. Ther. *17*, 463–471. https://doi.org/10.1038/mt.2008.269.

70. Ling, C., Li, B., Ma, W., and Srivastava, A. (2016). Development of optimized AAV serotype vectors for high-efficiency transduction at further reduced doses. Hum. Gene Ther. Methods *27*, 143–149. https://doi.org/10.1089/hgtb.2016.054.

71. Maheshri, N., Koerber, J.T., Kaspar, B.K., and Schaffer, D.V. (2006). Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. Nat. Biotechnol. *24*, 198–204. https://doi.org/10.1038/nbt1182.

72. Grimm, D., Lee, J.S., Wang, L., Desai, T., Akache, B., Storm, T.A., and Kay, M.A. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J. Virol. *82*, 5887–5911. https://doi.org/10.1128/JVI.00254-08.

73. Ojala, D.S., Sun, S., Santiago-Ortiz, J.L., Shapiro, M.G., Romero, P.A., and Schaffer, D.V. (2018). 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. *26*, 304–319. https://doi.org/10.1016/j.ymthe.2017.09.006.

74. Ho, M.L., Adler, B.A., Torre, M.L., Silberg, J.J., and Suh, J. (2013). SCHEMA computational design of virus capsid chimeras: calibrating how genome packaging, protection, and transduction correlate with calculated structural disruption. ACS Synth. Biol. *2*, 724–733. https://doi.org/10.1021/sb400076r.

75. Voigt, C.A., Martinez, C., Wang, Z.-G., Mayo, S.L., and Arnold, F.H. (2002). Protein building blocks preserved by recombination. Nat. Struct. Biol. https://doi.org/10.1038/nsb805.

76. Büning, H., and Srivastava, A. (2019). Capsid modifications for targeting and improving the efficacy of AAV vectors. Mol. Ther. Methods Clin. Dev. *12*, 248–265. https://doi.org/10.1016/j.omtm.2019.01.008.

77. Deverman, B.E., Pravdo, P.L., Simpson, B.P., Kumar, S.R., Chan, K.Y., Banerjee, A., Wu, W.-L., Yang, B., Huber, N., Pasca, S.P., et al. (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. *34*, 204–209. https://doi.org/10.1038/nbt.3440.

78. Chan, K.Y., Jang, M.J., Yoo, B.B., Greenbaum, A., Ravi, N., Wu, W.-L., Sánchez-Guardado, L., Lois, C., Mazmanian, S.K., Deverman, B.E., et al. (2017). Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. *20*, 1172–1179. https://doi.org/10.1038/nn.4593.

79. Goertsen, D., Flytzanis, N.C., Goeden, N., Chuapoco, M.R., Cummins, A., Chen, Y., Fan, Y., Zhang, Q., Sharma, J., Duan, Y., et al. (2022). AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. Nat. Neurosci. *25*, 106–115. https://doi.org/10.1038/s41593-021-00969-4.

80. Chen, X., Ravindra Kumar, S., Adams, C.D., Yang, D., Wang, T., Wolfe, D.A., Arokiaraj, C.M., Ngo, V., Campos, L.J., Griffiths, J.A., et al. (2022). Engineered AAVs for non-invasive gene delivery to rodent and non-human primate nervous systems. Neuron *110*, 2242-2257.e6. https://doi.org/10.1016/j.neuron.2022.05.003.

81. Ravindra Kumar, S., Miles, T.F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Greenbaum, A., et al. (2020). Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods *17*, 541–550. https://doi.org/10.1038/s41592-020-0799-7.

82. Chen, X., Wolfe, D.A., Bindu, D.S., Zhang, M., Taskin, N., Goertsen, D., Shay, T.F., Sullivan, E.E., Huang, S.-F., Ravindra Kumar, S., et al. (2023). Functional gene delivery to and across brain vasculature of systemic AAVs with endothelial-specific tropism in rodents and broad tropism in primates. Nat. Commun. *14*, 3345. https://doi.org/10.1038/s41467-023-38582-7.

83. Tabebordbar, M., Lagerborg, K.A., Stanton, A., King, E.M., Ye, S., Tellez, L., Krunnfusz, A., Tavakoli, S., Widrick, J.J., Messemer, K.A., et al. (2021). Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. Cell *184*, 4919-4938.e22. https://doi.org/10.1016/j.cell.2021.08.028.

84. Weinmann, J., Weis, S., Sippel, J., Tulalamba, W., Remes, A., El Andari, J., Herrmann, A.-K., Pham, Q.H., Borowski, C., Hille, S., et al. (2020). Identification of a myotropic AAV by massively parallel in vivo evaluation of barcoded capsid variants. Nat. Commun. *11*, 5432. https://doi.org/10.1038/s41467-020-19230-w.

85. Goertsen, D., Goeden, N., Flytzanis, N.C., and Gradinaru, V. (2022). Targeting the lung epithelium after intravenous delivery by directed evolution of underexplored sites on the AAV capsid. Mol. Ther. Methods Clin. Dev. *26*, 331–342. https://doi.org/10.1016/j.omtm.2022.07.010.

86. Chuapoco, M.R., Flytzanis, N.C., Goeden, N., Christopher Octeau, J., Roxas, K.M., Chan, K.Y., Scherrer, J., Winchester, J., Blackburn, R.J., Campos, L.J., et al. (2023). Adeno-associated viral vectors for functional intravenous gene transfer throughout the non-human primate brain. Nat. Nanotechnol. *18*, 1241–1251. https://doi.org/10.1038/s41565-023-01419-x.

87. Stanton, A.C., Lagerborg, K.A., Tellez, L., Krunnfusz, A., King, E.M., Ye, S., Solomon, I.H., Tabebordbar, M., and Sabeti, P.C. (2023). Systemic administration of novel engineered AAV capsids facilitates enhanced transgene expression in the macaque CNS. Med *4*, 31-50.e8. https://doi.org/10.1016/j.medj.2022.11.002.

88. Tervo, D.G.R., Hwang, B.-Y., Viswanathan, S., Gaj, T., Lavzin, M., Ritola, K.D., Lindo, S., Michael, S., Kuleshova, E., Ojala, D., et al. (2016). A designer AAV variant permits efficient retrograde access to projection neurons. Neuron *92*, 372–382. https://doi.org/10.1016/j.neuron.2016.09.021.

89. Michelfelder, S., Varadi, K., Raupp, C., Hunger, A., Körbelin, J., Pahrmann, C., Schrepfer, S., Müller, O.J., Kleinschmidt, J.A., and Trepel, M. (2011). Peptide ligands incorporated into the threefold spike capsid domain to re-direct gene transduction of AAV8 and AAV9 in vivo. PLoS ONE *6*, e23101. https://doi.org/10.1371/journal.pone.0023101.

90. El Andari, J., Renaud-Gabardos, E., Tulalamba, W., Weinmann, J., Mangin, L., Pham, Q.H., Hille, S., Bennett, A., Attebi, E., Bourges, E., et al. (2022). Semirational bioengineering of AAV vectors with increased potency and specificity for systemic gene therapy of muscle disorders. Sci. Adv. *8*, eabn4704. https://doi.org/10.1126/sciadv.abn4704.

91. Shay, T.F., Jang, S., Brittain, T.J., Chen, X., Walker, B., Tebbutt, C., Fan, Y., Wolfe, D.A., Arokiaraj, C.M., Sullivan, E.E., et al. (2024). Human cell surface-AAV interactomes identify LRP6 as blood-brain barrier transcytosis receptor and immune cytokine IL3 as AAV9 binder. Nat. Commun. *15*, 7853. https://doi.org/10.1038/s41467-024-52149-0.

92. Deal, C., De Wannemaeker, L., and De Mey, M. (2024). Towards a rational approach to promoter engineering: understanding the complexity of transcription initiation in prokaryotes. FEMS Microbiol. Rev. *48*, fuae004. https://doi.org/10.1093/femsre/fuae004.

93. Hitoshi, N., Ken-ichi, Y., and Jun-ichi, M. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene *108*, 193–199. https://doi.org/10.1016/0378-1119(91)90434-D.

94. Boshart, M., Weber, F., Jahn, G., Dorschhler, K., Fleckenstein, B., and Schaffner, W. (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell *41*, 521–530. https://doi.org/10.1016/S0092-8674(85)80025-8.

95. Stinski, M.F., and Roehr, T.J. (1985). Activation of the major immediate early gene of human cytomegalovirus by cis-acting elements in the promoter-regulatory sequence and by virus-specific trans-acting components. J. Virol. *55*, 431–441. https://doi.org/10.1128/jvi.55.2.431-441.1985.

96. Stinski, M.F., Thomsen, D.R., Stenberg, R.M., and Goldstein, L.C. (1983). Organization and expression of the immediate early genes of human cytomegalovirus. J. Virol. *46*, 1–14. https://doi.org/10.1128/jvi.46.1.1-14.1983.

97. Thomsen, D.R., Stenberg, R.M., Goins, W.F., and Stinski, M.F. (1984). Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. Proc. Natl. Acad. Sci. U.S.A. *81*, 659–663. https://doi.org/10.1073/pnas.81.3.659.

98. Dierks, P., Van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell *32*, 695–706. https://doi.org/10.1016/0092-8674(83)90055-7.

99. Grosveld, G.C., Rosenthal, A., and Flavell, R.A. (1982). Sequence requirements for the transcription of the rabbit β-globin gene *in vivo* : -80 region. Nucleic Acids Res. *10*, 4951–4971. https://doi.org/10.1093/nar/10.16.4951.

100. Juven-Gershon, T., Cheng, S., and Kadonaga, J.T. (2006). Rational design of a super core promoter that enhances gene expression. Nat. Methods *3*, 917–922. https://doi.org/10.1038/nmeth937.

101. Liu, J., Merlie, J.P., Todd, R.D., and O'Malley, K.L. (1997). Identification of cell type-specific promoter elements associated with the rat tyrosine hydroxylase gene using transgenic founder analysis. Mol. Brain Res. *50*, 33–42. https://doi.org/10.1016/S0169-328X(97)00163-0.

102. Oh, M.S., Hong, S.J., Huh, Y., and Kim, K.-S. (2009). Expression of transgenes in midbrain dopamine neurons using the tyrosine hydroxylase promoter. Gene Ther. *16*, 437–440. https://doi.org/10.1038/gt.2008.148.

103. Portales-Casamar, E., Swanson, D.J., Liu, L., Leeuw, C.N.D., Banks, K.G., Ho Sui, S.J., Fulton, D.L., Ali, J., Amirabbasi, M., Arenillas, D.J., et al. (2010). A regulatory toolbox of MiniPromoters to drive selective expression in the brain. Proc. Natl. Acad. Sci. U.S.A. *107*, 16589–16594. https://doi.org/10.1073/pnas.1009158107.

104. De Leeuw, C.N., Korecki, A.J., Berry, G.E., Hickmott, J.W., Lam, S.L., Lengyell, T.C., Bonaguro, R.J., Borretta, L.J., Chopra, V., Chou, A.Y., et al. (2016). rAAV-compatible MiniPromoters for restricted expression in the brain and eye. Mol. Brain *9*, 52. https://doi.org/10.1186/s13041-016-0232-4.

105. De Leeuw, C.N., Dyka, F.M., Boye, S.L., Laprise, S., Zhou, M., Chou, A.Y., Borretta, L., McInerny, S.C., Banks, K.G., Portales-Casamar, E., et al. (2014). Targeted CNS delivery using human MiniPromoters and demonstrated compatibility with adeno-associated viral vectors. Mol. Ther. Methods Clin. Dev. *1*, 5. https://doi.org/10.1038/mtm.2013.5.

106. Korecki, A.J., Cueva-Vargas, J.L., Fornes, O., Agostinone, J., Farkas, R.A., Hickmott, J.W., Lam, S.L., Mathelier, A., Zhou, M., Wasserman, W.W., et al. (2021). Human MiniPromoters for ocular-rAAV expression in ON bipolar, cone, corneal, endothelial, Müller glial, and PAX6 cells. Gene Ther. *28*, 351–372. https://doi.org/10.1038/s41434-021-00227-z.

107. Hickmott, J.W., Chen, C., Arenillas, D.J., Korecki, A.J., Lam, S.L., Molday, L.L., Bonaguro, R.J., Zhou, M., Chou, A.Y., Mathelier, A., et al. (2016). PAX6 MiniPromoters drive restricted expression from rAAV in the adult mouse retina. Mol. Ther. Methods Clin. Dev. *3*, 16051. https://doi.org/10.1038/mtm.2016.51.

108. Field, A., and Adelman, K. (2020). Evaluating enhancer function and transcription. Annu. Rev. Biochem. *89*, 213–234. https://doi.org/10.1146/annurev-biochem-011420-095916.

109. Panigrahi, A., and O'Malley, B.W. (2021). Mechanisms of enhancer action: the known and the unknown. Genome Biol. *22*, 108. https://doi.org/10.1186/s13059-021-02322-1.

110. Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308. https://doi.org/10.1016/0092-8674(81)90413-X.

111. Zerucha, T., Stühmer, T., Hatch, G., Park, B.K., Long, Q., Yu, G., Gambarotta, A., Schultz, J.R., Rubenstein, J.L., and Ekker, M. (2000). A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. J. Neurosci. *20*, 709–721. https://doi.org/10.1523/JNEUROSCI.20-02-00709.2000.

112. Stühmer, T., Puelles, L., Ekker, M., and Rubenstein, J.L.R. (2002). Expression from a Dlx gene enhancer marks adult mouse cortical GABAergic neurons. Cereb. Cortex *12*, 75–85. https://doi.org/10.1093/cercor/12.1.75.

113. Lee, A.T., Gee, S.M., Vogt, D., Patel, T., Rubenstein, J.L., and Sohal, V.S. (2014). Pyramidal neurons in prefrontal cortex receive subtype-specific forms of excitation and inhibition. Neuron *81*, 61–68. https://doi.org/10.1016/j.neuron.2013.10.031.

114. Lee, A.T., Vogt, D., Rubenstein, J.L., and Sohal, V.S. (2014). A class of GABAergic neurons in the prefrontal cortex sends long-range projections to the nucleus accumbens and elicits acute avoidance behavior. J. Neurosci. *34*, 11519–11525. https://doi.org/10.1523/JNEUROSCI.1157-14.2014.

115. Dimidschstein, J., Chen, Q., Tremblay, R., Rogers, S.L., Saldi, G.-A., Guo, L., Xu, Q., Liu, R., Lu, C., Chu, J., et al. (2016). A viral strategy for targeting and manipulating interneurons across vertebrate species. Nat. Neurosci. *19*, 1743–1749. https://doi.org/10.1038/nn.4430.

116. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.

117. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502. https://doi.org/10.1126/science.1141319.

118. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490. https://doi.org/10.1038/nature14590.

119. Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol. Cell *71*, 858-871.e8. https://doi.org/10.1016/j.molcel.2018.06.044.

120. Hrvatin, S., Tzeng, C.P., Nagy, M.A., Stroud, H., Koutsioumpa, C., Wilcox, O.F., Assad, E.G., Green, J., Harvey, C.D., Griffith, E.C., et al. (2019). A scalable platform for the development of cell-type-specific viral drivers. eLife *8*, e48089. https://doi.org/10.7554/eLife.48089.

121. Graybuck, L.T., Daigle, T.L., Sedeño-Cortés, A.E., Walker, M., Kalmbach, B., Lenz, G.H., Morin, E., Nguyen, T.N., Garren, E., Bendrick, J.L., et al. (2021). Enhancer viruses for combinatorial cell-subclass-specific labeling. Neuron *109*, 1449-1464.e13. https://doi.org/10.1016/j.neuron.2021.03.011.

122. Mich, J.K., Graybuck, L.T., Hess, E.E., Mahoney, J.T., Kojima, Y., Ding, Y., Somasundaram, S., Miller, J.A., Kalmbach, B.E., Radaelli, C., et al. (2021). Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. Cell Rep. *34*, 108754. https://doi.org/10.1016/j.celrep.2021.108754.

123. Mich, J.K., Sunil, S., Johansen, N., Martinez, R.A., Leytze, M., Gore, B.B., Mahoney, J.T., Ben-Simon, Y., Bishaw, Y., Brouner, K., et al. (2023). Enhancer-AAVs allow genetic access to oligodendrocytes and diverse populations of astrocytes across species. BioRxiv 2023.09.214558718. https://doi.org/10.1101/2023.09.20.558718.

124. Lawler, A.J., Ramamurthy, E., Brown, A.R., Shin, N., Kim, Y., Toong, N., Kaplow, I.M., Wirthlin, M., Zhang, X., Phan, B.N., et al. (2022). Machine learning sequence prioritization for cell type-specific enhancer design. eLife *11*, e69571. https://doi.org/10.7554/eLife.69571.

125. Lambert, J.T., Su-Feher, L., Cichewicz, K., Warren, T.L., Zdilar, I., Wang, Y., Lim, K.J., Haigh, J.L., Morse, S.J., Canales, C.P., et al. (2021). Parallel functional testing identifies enhancers active in early postnatal mouse brain. eLife *10*, e69479. https://doi.org/10.7554/eLife.69479.

126. Rubin, A.N., Malik, R., Cho, K.K.A., Lim, K.J., Lindtner, S., Robinson Schwartz, S.E., Vogt, D., Sohal, V.S., and Rubenstein, J.L.R. (2020). Regulatory elements inserted into AAVs confer preferential activity in cortical interneurons. eNeuro *7*, ENEURO.0211-20.2020. https://doi.org/10.1523/ENEURO.0211-20.2020.

127. Vormstein-Schneider, D., Lin, J.D., Pelkey, K.A., Chittajallu, R., Guo, B., Arias-Garcia, M.A., Allaway, K., Sakopoulos, S., Schneider, G., Stevenson, O., et al. (2020). Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. Nat. Neurosci. *23*, 1629–1636. https://doi.org/10.1038/s41593-020-0692-9.

128. Kussick, E., Johansen, N., Taskin, N., Wynalda, B., Martinez, R., Groce, E.L., Reding, M., Liang, E., Shulga, L., Huang, C., et al. (2024). Enhancer AAVs for targeting spinal motor neurons and descending motor pathways in rodents and macaque. bioRxiv 2024.07.30.605684, https://doi.org/10.1101/2024.07.30.605864.

129. Furlanis, E., Dai, M., Leyva Garcia, B., Vergara, J., Pereira, A., Pelkey, K., Tran, T., Gorissen, B.L., Vlachos, A., Hairston, A., et al. (2024). An enhancer-AAV toolbox to target and manipulate distinct interneuron subtypes. bioRxiv 2024.07.17.603924, https://doi.org/10.1101/2024.07.17.603924.

130. He, J., Phan, B.N., Kerkhoff, W.G., Alikaya, A., Brull, O.R., Fredericks, J.M., Hong, T., Sedorovitz, M., Srinivasan, C., Leone, M.J., et al. (2024). Cell type specific enhancers for dorsolateral prefrontal cortex. bioRxiv 2024.12.01.626253, https://doi.org/10.1101/2024.12.01.626253.

131. Nagy, M.A., Price, S., Wang, K., Gill, S., Ren, E., Jayne, L., Pajak, V., Deighan, S., Liu, B., Lu, X., et al. (2024). Cis-regulatory elements driving motor neuron-selective viral payload expression within the mammalian spinal cord. Proc. Natl. Acad. Sci. U.S.A. *121*, e2418024121. https://doi.org/10.1073/pnas.2418024121.

132. Shang, R., Lee, S., Senavirathne, G., and Lai, E.C. (2023). microRNAs in action: biogenesis, function and regulation. Nat. Rev. Genet. *24*, 816–833. https://doi.org/10.1038/s41576-023-00611-y.

133. Iwakawa, H., and Tomari, Y. (2022). Life of RISC: formation, action, and degradation of RNA-induced silencing complex. Mol. Cell *82*, 30–43. https://doi.org/10.1016/j.molcel.2021.11.026.

134. He, M., Liu, Y., Wang, X., Zhang, M.Q., Hannon, G.J., and Huang, Z.J. (2012). Cell-type-based analysis of microRNA profiles in the mouse brain. Neuron *73*, 35–48. https://doi.org/10.1016/j.neuron.2011.11.010.

135. Pomper, N., Liu, Y., Hoye, M.L., Dougherty, J.D., and Miller, T.M. (2020). CNS microRNA profiles: a database for cell type enriched microRNA expression across the mouse central nervous system. Sci. Rep. *10*, 4921. https://doi.org/10.1038/s41598-020-61307-5.

136. Jovičić, A., Roshan, R., Moisoi, N., Pradervand, S., Moser, R., Pillai, B., and Luthi-Carter, R. (2013). 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. https://doi.org/10.1523/JNEUROSCI.0600-12.2013.

137. Xie, J., Xie, Q., Zhang, H., Ameres, S.L., Hung, J.-H., Su, Q., He, R., Mu, X., Seher Ahmed, S., Park, S., et al. (2011). MicroRNA-regulated, systemically delivered rAAV9: a step closer to CNS-restricted transgene expression. Mol. Ther. *19*, 526–535. https://doi.org/10.1038/mt.2010.279.

138. Qiao, C., Yuan, Z., Li, J., He, B., Zheng, H., Mayer, C., Li, J., and Xiao, X. (2011). Liver-specific microRNA-122 target sequences incorporated in AAV vectors efficiently inhibits transgene expression in the liver. Gene Ther. *18*, 403–410. https://doi.org/10.1038/gt.2010.157.

139. Hordeaux, J., Buza, E.L., Jeffrey, B., Song, C., Jahan, T., Yuan, Y., Zhu, Y., Bell, P., Li, M., Chichester, J.A., et al. (2020). MicroRNA-mediated inhibition of transgene expression reduces dorsal root ganglion toxicity by AAV vectors in primates. Sci. Transl. Med. *12*, eaba9188. https://doi.org/10.1126/scitranslmed.aba9188.

140. Buss, N., Lanigan, L., Zeller, J., Cissell, D., Metea, M., Adams, E., Higgins, M., Kim, K.H., Budzynski, E., Yang, L., et al. (2022). Characterization of AAV-mediated dorsal root ganglionopathy. Mol. Ther. Methods Clin. Dev. *24*, 342–354. https://doi.org/10.1016/j.omtm.2022.01.013.

141. Gleichman, A.J., Kawaguchi, R., Sofroniew, M.V., and Carmichael, S.T. (2023). A toolbox of astrocyte-specific, serotype-independent adeno-associated viral vectors using microRNA targeting sequences. Nat. Commun. *14*, 7426. https://doi.org/10.1038/s41467-023-42746-w.

142. Jüttner, J., Szabo, A., Gross-Scherf, B., Morikawa, R.K., Rompani, S.B., Hantz, P., Szikra, T., Esposti, F., Cowan, C.S., Bharioke, A., et al. (2019). Targeting neuronal and glial cell types with synthetic promoter AAVs in mice, non-human primates and humans. Nat. Neurosci. *22*, 1345–1356. https://doi.org/10.1038/s41593-019-0431-2.

143. Gonzalez-Sandoval, A., Pekrun, K., Tsuji, S., Zhang, F., Hung, K.L., Chang, H.Y., and Kay, M.A. (2023). The AAV capsid can influence the epigenetic marking of rAAV delivered episomal genomes in a species dependent manner. Nat. Commun. *14*, 2448. https://doi.org/10.1038/s41467-023-38106-3.

144. Loeb, E.J., Havlik, P.L., Elmore, Z.C., Rosales, A., Fergione, S.M., Gonzalez, T.J., Smith, T.J., Benkert, A.R., Fiflis, D.N., and Asokan, A. (2024). Capsid-mediated control of adeno-associated viral transcription determines host range. Cell Rep. *43*, 113902. https://doi.org/10.1016/j.celrep.2024.113902.

145. Powell, S.K., Samulski, R.J., and McCown, T.J. (2020). AAV capsid-promoter interactions determine CNS cell-selective gene expression in vivo. Mol. Ther. *28*, 1373–1380. https://doi.org/10.1016/j.ymthe.2020.03.007.

146. Bohlen, M.O., McCown, T.J., Powell, S.K., El-Nahal, H.G., Daw, T., Basso, M.A., Sommer, M.A., and Samulski, R.J. (2020). Adeno-associated virus capsid-promoter interactions in the brain translate from rat to the nonhuman primate. Hum. Gene Ther. *31*, 1155–1168. https://doi.org/10.1089/hum.2020.196.

147. Rishik, S., Hirsch, P., Grandke, F., Fehlmann, T., and Keller, A. (2025). miRNATissueAtlas 2025: an update to the uniformly processed and annotated human and mouse non-coding RNA tissue atlas. Nucleic Acids Res. *53*, D129–D137. https://doi.org/10.1093/nar/gkae1036.

148. Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grässer, F.A., Lenhof, H.-P., et al. (2019). An estimate of the total number of true human miRNAs. Nucleic Acids Res. *47*, 3353–3364. https://doi.org/10.1093/nar/gkz097.

149. Houzet, L., Klase, Z., Yeung, M.L., Wu, A., Le, S.-Y., Quiñones, M., and Jeang, K.-T. (2012). The extent of sequence complementarity correlates with the potency of cellular miRNA-mediated restriction of HIV-1. Nucleic Acids Res. *40*, 11684–11696. https://doi.org/10.1093/nar/gks912.

150. Denzler, R., McGeary, S.E., Title, A.C., Agarwal, V., Bartel, D.P., and Stoffel, M. (2016). Impact of microRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. Mol. Cell *64*, 565–579. https://doi.org/10.1016/j.molcel.2016.09.027.

151. Porteus, M.H. (2019). A new class of medicines through DNA editing. N. Engl. J. Med. *380*, 947–959. https://doi.org/10.1056/NEJMra1800729.

152. Doudna, J.A. (2020). The promise and challenge of therapeutic genome editing. Nature *578*, 229–236. https://doi.org/10.1038/s41586-020-1978-5.

153. Wang, D., Zhang, F., and Gao, G. (2020). CRISPR-based therapeutic genome editing: strategies and in vivo delivery by AAV vectors. Cell *181*, 136–150. https://doi.org/10.1016/j.cell.2020.03.023.

154. Zhang, F. (2019). Development of CRISPR-Cas systems for genome editing and beyond. Quart. Rev. Biophys. *52*, e6. https://doi.org/10.1017/S0033583519000052.

155. Liu, X.S., and Jaenisch, R. (2019). Editing the epigenome to tackle brain disorders. Trends Neurosci. *42*, 861–870.

156. Xu, X, and Lei S.Q. (2019). A CRISPR-dCas toolbox for genetic engineering and synthetic biology. J. Mol. Biol. *431*, 34-47. https://doi.org/10.1016/j.jmb.2018.06.037.

157. Porto, E.M., Komor, A.C., Slaymaker, I.M., and Yeo, G.W. (2020). Base editing: advances and therapeutic opportunities. Nat. Rev. Drug Discov. *19*, 839–859. https://doi.org/10.1038/s41573-020-0084-6.

158. Chen, P.J., and Liu, D.R. (2023). Prime editing for precise and highly versatile genome manipulation. Nat. Rev. Genet. *24*, 161–177. https://doi.org/10.1038/s41576-022-00541-1.

159. Raguram, A., Banskota, S., and Liu, D.R. (2022). Therapeutic in vivo delivery of gene editing agents. Cell *185*, 2806–2827. https://doi.org/10.1016/j.cell.2022.03.045.

160. Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191. https://doi.org/10.1038/nature14299.

161. Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.-Y., Song, D.W., Lee, K.J., Jung, M.H., Kim, S., et al. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat. Commun. *8*, 14500. https://doi.org/10.1038/ncomms14500.

162. Wu, Z., Zhang, Y., Yu, H., Pan, D., Wang, Y., Wang, Y., Li, F., Liu, C., Nan, H., Chen, W., et al. (2021). Programmed genome editing by a miniature CRISPR-Cas12f nuclease. Nat. Chem. Biol. *17*, 1132-1138. https://doi.org/10.1038/s41589-021-00868-6.

163. Xu, X., Chemparathy, A., Zeng, L., Kempton, H.R., Shang, S., Nakamura, M., and Qi, L.S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol. Cell *81*, 4333-4345. https://doi.org/10.1016/j.molcel.2021.08.008.

164. Lau, C.-H., and Suh, Y. (2017). In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease. F1000Res *6*, 2153. https://doi.org/10.12688/f1000research.11243.1.

165. Choi, J.-H., Yu, N.-K., Baek, G.-C., Bakes, J., Seo, D., Nam, H.J., Baek, S.H., Lim, C.-S., Lee, Y.-S., and Kaang, B.-K. (2014). Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. Mol. Brain *7*, 17. https://doi.org/10.1186/1756-6606-7-17.

166. Brown, D., Altermatt, M., Dobreva, T., Chen, S., Wang, A., Thomson, M., and Gradinaru, V. (2021). Deep parallel characterization of AAV tropism and AAV-mediated transcriptional changes via single-cell RNA sequencing. Front. Immunol. *12*, 730825. https://doi.org/10.3389/fimmu.2021.730825.

167. Gyllborg, D., Langseth, C.M., Qian, X., Choi, E., Salas, S.M., Hilscher, M.M., Lein, E.S., and Nilsson, M. (2020). Hybridization-based *in situ* sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. Nucleic Acids Res. *48*, e112–e112. https://doi.org/10.1093/nar/gkaa792.

168. Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science *361*, eaat5691. https://doi.org/10.1126/science.aat5691.

169. Choi, H.M.T., Chang, J.Y., Trinh, L.A., Padilla, J.E., Fraser, S.E., and Pierce, N.A. (2010). Programmable in situ amplification for multiplexed imaging of mRNA expression. Nat. Biotechnol. *28*, 1208–1212. https://doi.org/10.1038/nbt.1692.

170. Choi, H.M.T., Beck, V.A., and Pierce, N.A. (2014). Next-generation *in situ* hybridization chain reaction: higher gain, lower cost, greater durability. ACS Nano *8*, 4284–4294. https://doi.org/10.1021/nn405717p.

171. Choi, H.M.T., Calvert, C.R., Husain, N., Huss, D., Barsi, J.C., Deverman, B.E., Hunter, R.C., Kato, M., Lee, S.M., Abelin, A.C.T., et al. (2016). Mapping a multiplexed zoo of mRNA expression. Development *143*, 3632–3637. https://doi.org/10.1242/dev.140137.

172. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753. https://doi.org/10.1242/dev.165753.

173. Zufferey, R., Donello, J.E., Trono, D., and Hope, T.J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. J. Virol. *73*, 2886–2892. https://doi.org/10.1128/JVI.73.4.2886-2892.1999.

174. Loeb, J.E., Cordier, W.S., Harris, M.E., Weitzman, M.D., and Hope, T.J. (1999). Enhanced expression of transgenes from adeno-associated virus vectors with the woodchuck hepatitis virus posttranscriptional regulatory element: implications for gene therapy. Hum. Gene Ther. *10*, 2295–2305. https://doi.org/10.1089/10430349950016942.

175. Chan, K.Y., Jang, M.J., Yoo, B.B., Greenbaum, A., Ravi, N., Wu, W.-L., Sánchez-Guardado, L., Lois, C., Mazmanian, S.K., Deverman, B.E., et al. (2017). Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. *20*, 1172–1179. https://doi.org/10.1038/nn.4593.

176. Terashima, T., Ogawa, N., Nakae, Y., Sato, T., Katagi, M., Okano, J., Maegawa, H., and Kojima, H. (2018). Gene therapy for neuropathic pain through siRNA-IRF5 gene delivery with homing peptides to microglia. Mol. Ther. Nucleic Acids *11*, 203–215. https://doi.org/10.1016/j.omtn.2018.02.007.

177. Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., Van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. Nat. Methods *15*, 932–935. https://doi.org/10.1038/s41592-018-0175-z.

178. Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature *563*, 72–78. https://doi.org/10.1038/s41586-018-0654-5.

179. Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346. https://doi.org/10.1038/nn.4216.

180. Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Südhof, T.C., and Quake, S.R. (2016). Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-seq. Cell Rep. *16*, 1126–1137. https://doi.org/10.1016/j.celrep.2016.06.059.

181. Kozareva, V., Martin, C., Osorno, T., Rudolph, S., Guo, C., Vanderburg, C., Nadaf, N., Regev, A., Regehr, W.G., and Macosko, E. (2021). A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. Nature *598*, 214–219. https://doi.org/10.1038/s41586-021-03220-z.

182. Phillips, J.W., Schulmann, A., Hara, E., Winnubst, J., Liu, C., Valakh, V., Wang, L., Shields, B.C., Korff, W., Chandrashekar, J., et al. (2019). A repeated molecular architecture across thalamic pathways. Nat. Neurosci. *22*, 1925–1935. https://doi.org/10.1038/s41593-019-0483-3.

183. Pena, J.T.G., Sohn-Lee, C., Rouhanifard, S.H., Ludwig, J., Hafner, M., Mihailovic, A., Lim, C., Holoch, D., Berninger, P., Zavolan, M., et al. (2009). miRNA in situ hybridization in formaldehyde and EDC–fixed tissues. Nat. Methods *6*, 139–141. https://doi.org/10.1038/nmeth.1294.

184. He, M., Liu, Y., Wang, X., Zhang, M.Q., Hannon, G.J., and Huang, Z.J. (2012). Cell-type-based analysis of microRNA profiles in the mouse brain. Neuron *73*, 35–48. https://doi.org/10.1016/j.neuron.2011.11.010.

185. Isakova, A., Fehlmann, T., Keller, A., and Quake, S.R. (2020). A mouse tissue atlas of small noncoding RNA. Proc. Natl. Acad. Sci. U.S.A. *117*, 25634–25645. https://doi.org/10.1073/pnas.2002277117.

186. Eng, C.-H.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature *568*, 235–239. https://doi.org/10.1038/s41586-019-1049-y.

187. Moffitt, J.R., Hao, J., Wang, G., Chen, K.H., Babcock, H.P., and Zhuang, X. (2016). High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl. Acad. Sci. U.S.A. *113*, 11046–11051. https://doi.org/10.1073/pnas.1612826113.

188. Patriarchi, T., Cho, J.R., Merten, K., Howe, M.W., Marley, A., Xiong, W.-H., Folk, R.W., Broussard, G.J., Liang, R., Jang, M.J., et al. (2018). Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. Science *360*, eaat4422. https://doi.org/10.1126/science.aat4422.

189. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

190. Jang, M.J., Coughlin, G.M., Jackson, C.R., Chen, X., Chuapoco, M.R., Vendemiatti, J.L., Wang, A.Z., and Gradinaru, V. (2023). Spatial transcriptomics for profiling the tropism of viral vectors in tissues. Nat. Biotechnol. *41*, 1272–1286. https://doi.org/10.1038/s41587-022-01648-w.

191. 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., et al. (2019). Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc. *14*, 379–414. https://doi.org/10.1038/s41596-018-0097-3.

192. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106. https://doi.org/10.1038/s41592-020-01018-x.

193. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

194. Askary, A., Sanchez-Guardado, L., Linton, J.M., Chadly, D.M., Budde, M.W., Cai, L., Lois, C., and Elowitz, M.B. (2020). In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription. Nat. Biotechnol. *38*, 66–75. https://doi.org/10.1038/s41587-019-0299-4.

195. Wang, S.K., Lapan, S.W., Hong, C.M., Krause, T.B., and Cepko, C.L. (2020). In situ detection of adeno-associated viral vector genomes with SABER-FISH. Mol. Ther. Methods Clin. Dev. *19*, 376–386. https://doi.org/10.1016/j.omtm.2020.10.003.

196. Ahlers, J., Althviz Moré, D., Amsalem, O., Anderson, A., Bokota, G., Boone, P., Bragantini, J., Buckley, G., Burt, A., Bussonnier, M., et al. (2023). napari: a multi-dimensional image viewer for Python. Version v0.4.18 (Zenodo). https://doi.org/10.5281/ZENODO.3555620

197. Hunker, A.C., Wirthlin, M.E., Gill, G., Johansen, N.J., Hooper, M., Omstead, V., Taskin, N., Weed, N., Vargas, S., Bendrick, J.L., et al. (2024). Enhancer AAV toolbox for accessing and perturbing striatal cell types and circuits. bioRxiv 2024.09.27.615553. https://doi.org/10.1101/2024.09.27.615553.

198. Duan, D., Yue, Y., Yan, Z., and Engelhardt, J.F. (2000). A new dual-vector approach to enhance recombinant adeno-associated virus-mediated gene expression through intermolecular cis activation. Nat. Med. *6*, 595–598. https://doi.org/10.1038/75080.

199. Nakai, H., Fuess, S., Storm, T.A., Meuse, L.A., and Kay, M.A. (2003). Free DNA ends are essential for concatemerization of synthetic double-stranded adeno-associated virus vector genomes transfected into mouse hepatocytes in vivo. Mol. Ther. *7*, 112–121. https://doi.org/10.1016/S1525-0016(02)00034-5.

200. Maeda, N., Niinobe, M., and Mikoshiba, K. (1990). A cerebellar Purkinje cell marker P400 protein is an inositol 1,4,5-trisphosphate (InsP3) receptor protein. Purification and characterization of InsP3 receptor complex. EMBO J. *9*, 61–67. https://doi.org/10.1002/j.1460-2075.1990.tb08080.x.

201. He, J., Phan, B.N., Kerkhoff, W.G., Alikaya, A., Brull, O.R., Fredericks, J.M., Hong, T., Sedorovitz, M., Srinivasan, C., Leone, M.J., et al. (2024). Cell type specific enhancers for dorsolateral prefrontal cortex. bioRxiv 2024.12.01.626253. https://doi.org/10.1101/2024.12.01.626253.

202. Hunker, A.C., Mich, J.K., Taskin, N., Torkelson, A., Pham, T., Bertagnolli, D., Chakka, A.B., Chakrabarty, R., Donadio, N.P., Ferrer, R., et al. (2025). Technical and biological sources of noise confound multiplexed enhancer AAV screening. bioRxiv 2025.01.14.633018. https://doi.org/10.1101/2025.01.14.633018.

203. Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. Proc. Natl. Acad. Sci. U.S.A. *99*, 16433–16437. https://doi.org/10.1073/pnas.162342499.

204. Dorman, E.R., Bushey, A.M., and Corces, V.G. (2007). The role of insulator elements in large-scale chromatin structure in interphase. Semin. Cell Dev. Biol. *18*, 682–690. https://doi.org/10.1016/j.semcdb.2007.08.009.

205. Brasset, E., and Vaury, C. (2005). Insulators are fundamental components of the eukaryotic genomes. Hered. *94*, 571–576. https://doi.org/10.1038/sj.hdy.6800669.

206. Yang, B., Treweek, J.B., Kulkarni, R.P., Deverman, B.E., Chen, C.-K., Lubeck, E., Shah, S., Cai, L., and Gradinaru, V. (2014). Single-cell phenotyping within transparent intact tissue through whole-body clearing. Cell *158*, 945–958. https://doi.org/10.1016/j.cell.2014.07.017.

207. Treweek, J.B., Chan, K.Y., Flytzanis, N.C., Yang, B., Deverman, B.E., Greenbaum, A., Lignell, A., Xiao, C., Cai, L., Ladinsky, M.S., et al. (2015). Whole-body tissue stabilization and selective extractions via tissue-hydrogel hybrids for high-resolution intact circuit mapping and phenotyping. Nat. Protoc. *10*, 1860–1896. https://doi.org/10.1038/nprot.2015.122.

208. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

209. Stirling, D.R., Swain-Bowden, M.J., Lucas, A.M., Carpenter, A.E., Cimini, B.A., and Goodman, A. (2021). CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics *22*, 433. https://doi.org/10.1186/s12859-021-04344-9.

210. Tabebordbar, M., Zhu, K., Cheng, J.K.W., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A., et al. (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science *351*, 407–411. https://doi.org/10.1126/science.aad5177.

211. Stahl, E.C., Sabo, J.K., Kang, M.H., Allen, R., Applegate, E., Kim, S.E., Kwon, Y., Seth, A., Lemus, N., Salinas-Rios, V., et al. (2023). Genome editing in the mouse brain with minimally immunogenic Cas9 RNPs. Mol. Ther. *31*, 2422–2438. https://doi.org/10.1016/j.ymthe.2023.06.019.

212. Monteys, A.M., Hundley, A.A., Ranum, P.T., Tecedor, L., Muehlmatt, A., Lim, E., Lukashev, D., Sivasankaran, R., and Davidson, B.L. (2021). Regulated control of gene therapies by drug-induced splicing. Nature *596*, 291–295. https://doi.org/10.1038/s41586-021-03770-2.

213. Lang, J.F., Toulmin, S.A., Brida, K.L., Eisenlohr, L.C., and Davidson, B.L. (2019). Standard screening methods underreport AAV-mediated transduction and gene editing. Nat. Commun. *10*, 3415. https://doi.org/10.1038/s41467-019-11321-7.

214. Jun, K., Piedras-Rentería, E.S., Smith, S.M., Wheeler, D.B., Lee, S.B., Lee, T.G., Chin, H., Adams, M.E., Scheller, R.H., Tsien, R.W., et al. (1999). Ablation of P/Q-type Ca $^{2+}$ channel currents, altered synaptic transmission, and progressive ataxia in mice lacking the α_{1A} -subunit. Proc. Natl. Acad. Sci. U.S.A. *96*, 15245–15250. https://doi.org/10.1073/pnas.96.26.15245.

215. Fletcher, C.F., Lutz, C.M., O'Sullivan, T.N., Shaughnessy, J.D., Hawkes, R., Frankel, W.N., Copeland, N.G., and Jenkins, N.A. (1996). Absence epilepsy in tottering mutant mice is associated with calcium channel defects. Cell *87*, 607–617. https://doi.org/10.1016/S0092-8674(00)81381-1.

216. Mallmann, R.T., Elgueta, C., Sleman, F., Castonguay, J., Wilmes, T., Van Den Maagdenberg, A., and Klugbauer, N. (2013). Ablation of $Ca_V2.1$ voltage-gated Ca2+ channels in mouse forebrain generates multiple cognitive impairments. PLoS ONE *8*, e78598. https://doi.org/10.1371/journal.pone.0078598.

217. Rossignol, E., Kruglikov, I., Van Den Maagdenberg, A.M.J.M., Rudy, B., and Fishell, G. (2013). $Ca_V2.1$ ablation in cortical interneurons selectively impairs fast-spiking basket cells and causes generalized seizures. Ann. Neurol. *74*, 209–222. https://doi.org/10.1002/ana.23913.

218. Bomben, V.C., Aiba, I., Qian, J., Mark, M.D., Herlitze, S., and Noebels, J.L. (2016). Isolated P/Q calcium channel deletion in layer VI corticothalamic neurons generates absence epilepsy. J. Neurosci. *36*, 405–418. https://doi.org/10.1523/JNEUROSCI.2555-15.2016.

219. Todorov, B., Kros, L., Shyti, R., Plak, P., Haasdijk, E.D., Raike, R.S., Frants, R.R., Hess, E.J., Hoebeek, F.E., De Zeeuw, C.I., et al. (2012). Purkinje cell-specific ablation of $Ca_V2.1$ channels is sufficient to cause cerebellar ataxia in mice. Cerebellum *11*, 246–258. https://doi.org/10.1007/s12311-011-0302-1.

220. Mark, M.D., Maejima, T., Kuckelsberg, D., Yoo, J.W., Hyde, R.A., Shah, V., Gutierrez, D., Moreno, R.L., Kruse, W., Noebels, J.L., et al. (2011). Delayed postnatal loss of P/Q-type calcium channels recapitulates the absence epilepsy, dyskinesia, and ataxia phenotypes of genomic *Cacna1A* mutations. J. Neurosci. *31*, 4311–4326. https://doi.org/10.1523/JNEUROSCI.5342-10.2011.

221. Zhang, X., Ng, A.H., Tanner, J.A., Wu, W., Copeland, N.G., Jenkins, N.A., and Huang, J. (2004). Highly restricted expression of Cre recombinase in cerebellar Purkinje cells. Genesis *40*, 45–51. https://doi.org/10.1002/gene.20062.

222. Maejima, T., Wollenweber, P., Teusner, L.U.C., Noebels, J.L., Herlitze, S., and Mark, M.D. (2013). Postnatal loss of P/Q-type channels confined to rhombic-lip-derived neurons alters synaptic transmission at the parallel fiber to Purkinje cell synapse and replicates genomic *Cacna1a* mutation phenotype of ataxia and seizures in mice. J. Neurosci. *33*, 5162–5174. https://doi.org/10.1523/JNEUROSCI.5442-12.2013.

223. Fleming, S.M., Ekhator, O.R., and Ghisays, V. (2013). Assessment of sensorimotor function in mouse models of Parkinson's disease. J. Vis. Exp. *76*, 50303. https://doi.org/10.3791/50303.

224. Carter, R.J., Morton, J., and Dunnett, S.B. (2001). Motor coordination and balance in rodents. Curr. Protoc. Neurosci. *15*, 8.12.1-8.12.14. https://doi.org/10.1002/0471142301.ns0812s15.

225. Mendes, C.S., Bartos, I., Márka, Z., Akay, T., Márka, S., and Mann, R.S. (2015). Quantification of gait parameters in freely walking rodents. BMC Biol. *13*, 50. https://doi.org/10.1186/s12915-015-0154-0.

226. Kiven, S., Wang, Y., Aich, A., Argueta, D.A., Lei, J., Sagi, V., Tennakoon, M., Bedros, S.J., Lambrecht, N., and Gupta, K. (2020). Spatiotemporal alterations in gait in humanized transgenic sickle mice. Front. Immunol. *11*, 561947. https://doi.org/10.3389/fimmu.2020.561947.

227. Xi, Z., Öztürk, B.E., Johnson, M.E., Turunç, S., Stauffer, W.R., and Byrne, L.C. (2022). Quantitative single-cell transcriptome-based ranking of engineered AAVs in human retinal explants. Mol. Ther. Methods Clin. Dev. *25*, 476–489. https://doi.org/10.1016/j.omtm.2022.04.014.

228. Öztürk, B.E., Johnson, M.E., Kleyman, M., Turunç, S., He, J., Jabalameli, S., Xi, Z., Visel, M., Dufour, V.L., Iwabe, S., et al. (2021). scAAVengr, a transcriptome-based pipeline for quantitative ranking of engineered AAVs with single-cell resolution. eLife *10*, e64175. https://doi.org/10.7554/eLife.64175.

229. Lalanne, J.-B., Mich, J.K., Huynh, C., Hunker, A.C., McDiarmid, T.A., Levi, B.P., Ting, J.T., and Shendure, J. (2025). Extensive length and homology dependent chimerism in pool-packaged AAV libraries. bioRxiv 2025.01.14.632594. https://doi.org/10.1101/2025.01.14.632594.

230. Chiuchiolo, M.J., and Crystal, R.G. (2016). Gene therapy for alpha-1 antitrypsin deficiency lung disease. Ann. Am. Thorac. Soc. *13*, S352–S369. https://doi.org/10.1513/AnnalsATS.201506-344KV.

231. Stoller, J.K., Hupertz, V., and Aboussouan, L.S. (1993). Alpha-1 antitrypsin Deficiency. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

232. Benke, T.A., Demarest, S., Angione, K., Downs, J., Leonard, H., Saldaris, J., Marsh, E.D., Olson, H., and Haviland, I. (1993). CDKL5 Deficiency Disorder. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

233. Deardorff, M.A., Noon, S.E., and Krantz, I.D. (1993). Cornelia de Lange Syndrome. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

234. Deciphering Developmental Disorders Study, Olley, G., Ansari, M., Bengani, H., Grimes, G.R., Rhodes, J., Von Kriegsheim, A., Blatnik, A., Stewart, F.J., Wakeling, E., et al. (2018). BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange–like syndrome. Nat. Genet. *50*, 329–332. https://doi.org/10.1038/s41588-018-0042-y.

235. Donati, B., Lorenzini, E., and Ciarrocchi, A. (2018). BRD4 and cancer: going beyond transcriptional regulation. Mol. Cancer *17*, 164. https://doi.org/10.1186/s12943-018-0915-9.

236. Sarogni, P., Palumbo, O., Servadio, A., Astigiano, S., D'Alessio, B., Gatti, V., Cukrov, D., Baldari, S., Pallotta, M.M., Aretini, P., et al. (2019). Overexpression of the cohesin-core subunit SMC1A contributes to colorectal cancer development. J. Exp. Clin. Cancer Res. *38*, 108. https://doi.org/10.1186/s13046-019-1116-0.

237. Savant, A., Lyman, B., Bojanowski, C., and Upadia, J. (1993). Cystic Fibrosis. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

238. Duan, D., Goemans, N., Takeda, S., Mercuri, E., and Aartsma-Rus, A. (2021). Duchenne muscular dystrophy. Nat. Rev. Dis. Primers *7*, 13. https://doi.org/10.1038/s41572-021-00248-3.

239. Steiner, R.D., and Basel, D. (1993). COL1A1/2 Osteogenesis Imperfecta. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

240. Chen, Y., Yang, S., Lovisa, S., Ambrose, C.G., McAndrews, K.M., Sugimoto, H., and Kalluri, R. (2021). Type-I collagen produced by distinct fibroblast lineages reveals specific function during embryogenesis and Osteogenesis Imperfecta. Nat. Commun. *12*, 7199. https://doi.org/10.1038/s41467-021-27563-3.

241. Duque Lasio, M.L., Lehman, A.N., Ahmad, A., and Bedoyan, J.K. (1993). Pyruvate Carboxylase Deficiency. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

242. Pulst, S.M. (1993). Spinocerebellar Ataxia Type 2. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

243. Casse, F., Courtin, T., Tesson, C., Ferrien, M., Noël, S., Fauret-Amsellem, A., Gareau, T., Guegan, J., Anheim, M., Mariani, L., et al. (2023). Detection of *ATXN2* expansions in an exome dataset: an underdiagnosed cause of Parkinsonism. Movement Disord. Clin. Pract. *10*, 664–669. https://doi.org/10.1002/mdc3.13699.

244. Perlman, S. (1993). Hereditary Ataxia Overview. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

245. Watson, L.M., Bamber, E., Schnekenberg, R.P., Williams, J., Bettencourt, C., Lickiss, J., Jayawant, S., Fawcett, K., Clokie, S., Wallis, Y., et al. (2017). Dominant mutations in GRM1 cause spinocerebellar ataxia type 44. Am. J. Hum. Genet. *101*, 451–458. https://doi.org/10.1016/j.ajhg.2017.08.005.

246. Holder, J.L., Hamdan, F.F., and Michaud, J.L. (1993). SYNGAP1-Related Intellectual Disability. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

247. Weiss, K.H., and Schilsky, M. (1993). Wilson Disease. In GeneReviews® (M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds.), University of Washington, Seattle

248. An, M., Raguram, A., Du, S.W., Banskota, S., Davis, J.R., Newby, G.A., Chen, P.Z., Palczewski, K., and Liu, D.R. (2024). Engineered virus-like particles for transient delivery of prime editor ribonucleoprotein complexes in vivo. Nat. Biotechnol. *42*, 1526–1537. https://doi.org/10.1038/s41587-023-02078-y.

249. Raguram, A., An, M., Chen, P.Z., and Liu, D.R. (2024). Directed evolution of engineered virus-like particles with improved production and transduction efficiencies. Nat. Biotechnol. https://doi.org/10.1038/s41587-024-02467-x.

250. Koonin, E.V., Krupovic, M., and Dolja, V.V. (2023). The global virome: How much diversity and how many independent origins? Environ. Microbiol. *25*, 40–44. https://doi.org/10.1111/1462-2920.16207.