

Targeting Subsets of Mammalian Neurons

Boris V. Zemelman 

Center for Learning and Memory, Department of Neuroscience, The University of Texas at Austin, Austin, TX, USA.

Neuroscience Insights
Volume 15: 1–4
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2633105520908537



ABSTRACT: Functional dissection of mammalian neuronal circuits depends on accurate targeting of constituent cell classes. Transgenic mice offer precise and predictable access to genetically defined cell populations, but there is the pressing need to target neuronal assemblies in species less amenable to genomic manipulations, such as the primate, which is an important animal model for human perception, cognition, and action. We have developed several virus-based methods for accessing all forebrain inhibitory interneurons as well as the major excitatory and inhibitory neuron subclasses. These methods rely on the wealth of emerging single-cell transcriptome data and harness gene expression variations to refine neuron targeting. Our approach enables nuanced functional studies, including in vivo imaging and manipulation, of the diverse cell populations of the mammalian neocortex, and it represents a timely blueprint for transgenics-independent interrogation of functionally significant cell classes.

KEYWORDS: Genetic targeting, adeno-associated virus, rodents, nonhuman primates, cell type specificity

RECEIVED: January 9, 2020. **ACCEPTED:** January 23, 2020.

TYPE: Commentary

FUNDING: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research described was supported by NINDS BRAIN Initiative grants U01NS094330, U01NS099720 and NIDDC grant R21DC016169.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Boris V. Zemelman, The University of Texas at Austin, 1 University Station C7000, Austin, TX 78712-1139, USA. Email: zemelmanb@mail.clm.utexas.edu

COMMENT ON: Mehta P, Kreeger L, Wylie DC, et al. Functional access to neuron subclasses in rodent and primate forebrain. *Cell Rep.* 2019;26(10):2818.e8–2832.e8. doi:10.1016/j.celrep.2019.02.011. PubMed PMID: 30840900.

Introduction

A cell targeting strategy is shaped by how a neuronal population is defined. For us, the principal goal is to access functionally homologous neuronal circuit elements. This choice tends to restrict other neuron attributes: intrinsic properties, connectivity, and the complement of expressed genes. In other words, we do not expect or aim to capture all cells within traditional, but functionally diverse cell classes often represented a single neurochemical marker, such as parvalbumin or somatostatin. Our methods for accessing such cell populations are based on single-cell transcriptome data and rely on interdependent adeno-associated viruses (AAVs) with different expression specificities to refine targeting.¹ An important benefit of using AAVs is that they are not pathogenic and can infect cells of many species, including nonhuman primates (NHPs).

Our AAVs are engineered with short promoters that support different patterns of gene expression. Identifying such promoters has been a major undertaking because the AAV payload is quite limited, few neuron class-specific promoters and enhancers have been described, and the relationship between a specific promoter sequence and the resulting gene expression pattern is poorly understood. As a result, promoter design is an iterative undertaking requiring extensive empirical testing.

To reduce the trial-and-error aspect of vector engineering, we have recently developed powerful methods for identifying and testing candidate regulatory elements. Along the way, we have made several important observations: (a) co-expressed genes offer multiple equally effective solutions to achieve expression specificity, (b) more genetic content is not necessarily

better—a short regulatory domain can target subsets of neurons that share other key characteristics, (c) sequences conserved between species often function similarly, (d) promoter specificity can vary across brain regions as does the function of ostensibly similar neurons, (e) protein expression variability can be harnessed using intersectional approaches to refine neuron targeting, and (f) promoter strength must be suited to neuron type and intended application.

As we often use multiple viruses in the same preparation, we have also worked hard to achieve uniform and reproducible infections. Variables, such as titer and serotype, are discussed in the following sections.

Intersectional Techniques

Functional studies have revealed that excitatory or inhibitory neurons rarely fall within neat neurochemical boundaries.^{2,3} Likewise, neuronal gene expression is both promiscuous and variable^{4,5} making it difficult to match single molecular markers to functionally distinct neuron classes. We overcome this obstacle by using interdependent viruses whose promoters support different gene expression patterns. In an example of a *set intersection* strategy, one promoter may be active in classes A and B and another in classes B and C. Neither promoter alone is sufficient to access class B, but an intersectional strategy that relies on both promoters will successfully isolate class B. In an alternative *set difference* strategy, the first promoter is active in classes A and B and the second promoter is active only in class B. We can then *subtract* the expression pattern of the second promoter from that of the first to access only class A neurons. In these examples, overlapping endogenous gene expression, normally a hindrance to cellular marker-based genetic targeting, is



harnessed to single out a distinct neuronal class. We have now used these intersectional techniques to investigate inhibitory somatostatin, neuropeptide-Y, and excitatory cholecystokinin (CCK) neurons in multiple species.^{1,6}

Regulatory Motifs and Domains

Short promoter DNA motifs (~10 base pairs) are known to bind transcription factors and have been implicated in the regulation of eukaryotic gene expression,^{7,8} but which motifs are needed for specific expression patterns is largely unknown. We therefore set out to develop an algorithm that can mine single-cell transcriptome data to identify candidate cell type-specific DNA regulatory sequences.

Gene expression variability is usually quantified as a continuous score—fold-change, test-statistic, *P* value—comparing biological classes. Unlike existing approaches, our *de novo* strategy termed Suffix Array Kernel Smoothing (SArKS)⁹ and applies nonparametric kernel smoothing¹⁰ to uncover promoter motifs that correlate with elevated differential expression scores. SArKS detects motifs by smoothing sequence scores over sequence similarity. A second round of spatial proximity smoothing extends and merges motifs to reveal multi-motif domains (MMDs) hundreds of base pairs long. The juxtaposition of such MMDs has allowed us to explore combinatorial aspects of promoter organization.

Importantly, we do not screen for the top motifs nor for the most abundant transcripts; all sequences are scored based on expression differences across chosen cell classes. In addition, SArKS neither relies on nor generates consensus sequences, so that biologically relevant sequence variations and motif context are preserved, enabling nuanced comparisons. When a particular MMD is demonstrated experimentally to improve or hinder cell type-specific targeting, its sequence is incorporated iteratively into the SArKS search algorithm to refine subsequent rounds of motif selection. The ability to assign valence to MMDs—the bias in favor of inclusion or exclusion in a particular expression pattern—is also our starting point for rational promoter design, including to achieve layer-specific expression.

Conservation of Noncoding DNA

SArKS examines differences in gene expression across cell classes based on cell-specific transcriptome data. Such data have now been collected from genetically defined cell classes in rodents,^{11,12} but not from primates. Indeed, this chicken-and-egg problem—needing cell-specific transcriptome data to be able to define and access cell classes—represents a significant hurdle in engineering vectors for NHP research. Fortunately, comparisons of distantly related vertebrate genomes have demonstrated that conserved noncoding DNA, especially in the vicinity of developmentally important genes, can support shared regulatory regimes.¹³⁻¹⁵

To circumvent the lack of primate cell-specific data, we have used SArKS to identify candidate mouse regulatory domains

and have then examined these domains for elevated rodent-primate sequence conservation. Our strategy is supported by the promiscuity of transcription factors, which are known to tolerate subtle sequence variations^{16,17} and has helped us uncover human regulatory regions for accessing GABAergic and parvalbumin-expressing forebrain neurons in both rodent and primate.¹ While we and others are striving to collect transcriptome data from primate cells to aid the search for species-specific regulatory domains, we anticipate that the presence of cross-species sequence conservation within putative promoters will continue to be an important parameter when engineering viral vectors that are active in multiple species. One practical benefit of such conservation is that we can pre-screen many candidate promoters in mice.

Chromatin Accessibility

One important parameter that we consider when selecting differentially expressed genes for SArKS analysis is whether or not the chromatin is accessible in the vicinity of differentially expressed genes, where cell-specific transcription factors must bind. From an experimental perspective, genomic DNA may appear inaccessible because it is epigenetically modified, blocking transcription factor binding; alternatively, a bound transcription factor can render chromatin inaccessible while enabling transcription. We filter promoter regions that are not accessible in every cell population being compared because we wish to harness differential gene expression mechanisms supported entirely by cell-specific transcription factors.¹⁸ Variable gene expression where the binding of a ubiquitous transcription factor is epigenetically regulated is at odds with our sequence-based strategy and cannot be reproduced when using viral vectors whose genomes are not similarly modified. However, a screen for inaccessible chromatin in the cells of interest may be a useful strategy when examining the effects of distal sequences, such as enhancers, on gene expression.¹⁹ There, differential accessibility may indeed result from cell-specific transcription factor binding,²⁰ which can foster cell-specific expression.^{21,22}

Features of AAVs

In addition to gene and promoter, a common distinction among viral vectors is the serotype, the capsid that is selected during virus assembly. Because the sequence of events leading to AAV infection is not fully understood, the effect of a specific serotype on infectivity is hard to define. However, the lack of mechanistic insight has not dampened investigators' convictions about the importance of serotype. As much of this sentiment is based on personal experience, the best we can do is to offer our own.

The method for purifying a virus seems to be as important as a specific serotype. This may be because the coat proteins engaged in nuclear entry are especially sensitive to their environment as they undergo functional transformations. As a result, we generally distrust anecdotal claims about serotype potency: a serotype 9 AAV made in one laboratory may be

neither comparable to a serotype 9 nor better than another serotype made elsewhere. The confusion is compounded by the multitude of promoters and protein variants the viruses encode.

When we compare serotypes, the contents are identical as is the purification scheme. Under these conditions, serotypes 1, 5, 8, and 9 injected into mouse forebrain through a craniotomy lead to similar levels of fluorophore expression within 10–14 days. Serotypes 2 and 7 are weaker. These relationships hold in NHPs, although the onset of expression is delayed. In primary cultures of rat hippocampal neurons, serotype 1 is best: serotype 5 first strongly labels glia and then neurons; serotypes 8 and 9 label neurons well, but nonuniformly—some neurons remain unlabeled, suggesting a bias that we prefer to avoid. Consequently, nearly all of our vectors are serotype 1. Subcortically in mice, serotype 1 has occasionally failed to label all neurons; in these rare instances, we have successfully used a mix of serotypes 1 + 2 or serotype 9 AAVs.

Our advice to colleagues is that dogma is much less important than experimental observations: if a particular reagent works, do not switch. Conversely, do not give up if a reagent does not work as expected—change the source or capsid or promoter. Viral vectors are not standardized and there is much we do not know about how they work, so testing several variants is often unavoidable.

Infectivity Limitations

In many instances, we achieve cell selectivity using mixes of 2 or more viruses. One concern is that AAVs may perform differently when used singly versus in a cocktail. To date, we have seen no evidence of altered selectivity or potency when using mixes irrespective of constituent serotypes or promoters. We do, however, address vector dilution by maximizing the vector that encodes the activity reporter or cell actuator at the expense of recombinase-bearing vectors that provide targeting specificity. Surprisingly, we have also seen little evidence of reduced selectivity at injection site edges, even with the *set difference* strategy.¹ We hypothesize that neurons infected by single-virus particles appear unlabeled; the neurons we can score must therefore be infected by multiple viruses, preserving the regime of expression specificity.

We have also observed that a single region of NHP cortex can be re-infected repeatedly with the same or different AAV with no diminution of vector efficacy (unpublished observations and Seidemann et al²³). This is contrary to reports that implicate the immune response in re-infection failures. While we have not tested injected animals for the appearance of neutralizing antibodies, our findings are consistent with the abundance and low immunogenicity of naturally occurring AAVs in many mammals, including humans. Environmental contaminants and impurities associated with some AAV purification techniques, such as cell debris, leached column matrices and salts, as well as prostheses for neuron imaging and manipulation can irritate and injure injected tissues, causing experiments to fail.

The use of engineered retrograde AAVs that can infect axon fibers and terminals represents an anatomical restriction that can complement promoter-based cell targeting. In the one published example, the capsid amino acid modifications were identified through selection in mouse brain, but the mechanism of virus entry was not elucidated.²⁴ As is often the case, the resulting reagent does not generalize well outside the scope of the initial selection: whichever mouse receptor the engineered virus binds is clearly absent from many forebrain neurons, such as the hippocampal CA3 neurons, or is present only in neuron subsets, such as in the entorhinal cortex. At this early stage, it is difficult to predict how retrograde AAVs will perform in any particular system, and labeled neurons will have to be examined for evidence of bias imposed by this labeling strategy. Nonetheless, even if additional screens will be needed to improve its retrograde capability and to port it to primates, this class of AAVs adds a key functional component to cell-specific targeting.


Conclusions

Our efforts demonstrate that single rAAVs can access forebrain GABAergic neurons broadly and that interdependent viruses can be used to restrict access to specific excitatory and inhibitory subpopulations. The multi-virus techniques provide ample protein expression for nuanced functional studies of the diverse forebrain cell classes, including for *in vivo* imaging and manipulation studies in NHPs. The general strategies of identifying DNA sequences that are conserved between rodents and primates and of relying on combinatorial methods to refine genetic targeting offer a timely blueprint applicable to many neuron classes and species for the transgenics-independent brain-wide interrogations of functionally significant cell populations.

Author Contributions

BVZ conceptualized the referenced study and wrote the commentary.

ORCID iD

Boris V. Zemelman  <https://orcid.org/0000-0003-2463-4887>

REFERENCES

1. Mehta P, Kreeger L, Wylie DC, et al. Functional access to neuron subclasses in rodent and primate forebrain. *Cell Rep.* 2019;26:2818–2832. doi:10.1016/j.celrep.2019.02.011.
2. Soltesz I, Losonczy A. CA1 pyramidal cell diversity enabling parallel information processing in the hippocampus. *Nat Neurosci.* 2018;21:484–493. doi:10.1038/s41593-018-0118-0.
3. Tasic B, Menon V, Nguyen T, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci.* 2016;19:335–346. doi:10.1038/nn.4216.
4. Cembrowski MS, Menon V. Continuous variation within cell types of the nervous system. *Trends Neurosci.* 2018;41:337–348. doi:10.1016/j.tins.2018.02.010.
5. Lein ES, Hawrylycz MJ, Ao N, et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature.* 2007;445:168–176. doi:10.1038/nature05453.
6. Kreeger LJ, Connelly CJ, Mehta P, Zemelman BV, Golding NL. Excitatory cholecystokinin neurons of the midbrain integrate diverse temporal responses and drive auditory thalamic subdomains (submitted).

7. Walhout A. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res.* 2006;16:1445-1454. doi:10.1101/gr.5321506.
8. Wasserman WW, Sandelin A. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet.* 2004;5:276-287. doi:10.1038/nrg1315.
9. Wylie D, Hofmann HA, Zemelman BV. SArKS: de novo discovery of gene expression regulatory motif sites and domains by suffix array kernel smoothing. *Bioinformatics.* 2019;35:3944-3952. doi:10.1093/bioinformatics/btz198.
10. Altman N. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Statistician.* 1992;46:175-185. doi:10.2307/2685209.
11. Hodge RD, Bakken TE, Miller JA, et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature.* 2019;573:61-68. doi:10.1038/s41586-019-1506-7.
12. Mo A, Mukamel EA, Davis FP, et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron.* 2016;86:1369-1384. doi:10.1016/j.neuron.2015.05.018.
13. Woolfe A, Goodson M, Goode DK, et al. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 2005;3:e7. doi:10.1371/journal.pbio.0030007.
14. Hardison RC, Oeltjen J, Miller W. Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res.* 1997;7:959-966. doi:10.1101/gr.7.10.959.
15. Elgar G. Quality not quantity: the pufferfish genome. *Hum Mol Genet.* 1996;5:1437-1442. doi:10.1093/hmg/5.supplement_1.1437.
16. Gumucio DL, Shelton DA, Zhu W, et al. Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the b-like globin genes. *Mol Phylogenet Evol.* 1996;5:18-32. doi:10.1006/mpev.1996.0004.
17. Letovsky J, Dynan WS. Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. *Nucleic Acids Res.* 1989;17:2639-2653. doi:10.1093/nar/17.7.2639.
18. Davidson EH. Emerging properties of animal gene regulatory networks. *Nature.* 2010;468:911-920. doi:10.1038/nature09645.
19. Bell O, Tiwari VK, Thoma NH, Schubeler D. Determinants and dynamics of genome accessibility. *Nat Rev Genet.* 2011;12:554-564. doi:10.1038/nrg3017.
20. Li Q, Harju S, Peterson KR. Locus control regions: coming of age at a decade plus. *Trends Genet.* 1999;15:403-408. doi:10.1016/s0168-9525(99)01780-1.
21. Hrvatin S, Tzeng CP, Nagy AM, et al. PESCO: a scalable platform for the development of cell-type-specific viral drivers. *bioRxiv.* 2019;2019:570895. doi:10.1101/570895.
22. Graybuck LT, Sedeño-Cortés A, Nguyen T, et al. Prospective, brain-wide labeling of neuronal subclasses with enhancer-driven AAVs. *BioRxiv.* 2019;2019:525014. doi:10.1101/525014.
23. Seidemann E, Chen Y, Bai Y, et al. Calcium imaging with genetically encoded indicators in behaving primates. *Elife.* 2016;5:e16178. doi:10.7554/eLife.16178.
24. Tervo GRD, Hwang B-Y, Viswanathan S, et al. A designer AAV variant permits efficient retrograde access to projection neurons. *Neuron.* 2016;92:372-382. doi:10.1016/j.neuron.2016.09.021.