



An arginine-rich nuclear localization signal (ArgiNLS) strategy for streamlined image segmentation of single cells

Eric R. Szelenyi^{a,b,1} , Jovana S. Navarrete^{a,b,c} , Alexandria D. Murry^{a,b}, Yizhe Zhang^{a,b}, Kasey S. Girven^d, Lauren Kuo^{a,e,2} , Marcella M. Cline^{a,f,3} , Mollie X. Bernstein^{a,f}, Mariia Burdnyiuk^g, Bryce Bowler^b , Nastacia L. Goodwin^{a,b,c}, Barbara Juarez^{a,f,h,4}, Larry S. Zweifel^{a,f,h}, and Sam A. Golden^{a,b,1}

Affiliations are included on p. 11.

Edited by Jan Funke, HHMI, Janelia Farm Research Campus, Ashburn, VA; received November 29, 2023; accepted June 25, 2024 by Editorial Board Member Jeremy Nathans

High-throughput volumetric fluorescent microscopy pipelines can spatially integrate whole-brain structure and function at the foundational level of single cells. However, conventional fluorescent protein (FP) modifications used to discriminate single cells possess limited efficacy or are detrimental to cellular health. Here, we introduce a synthetic and nondeleterious nuclear localization signal (NLS) tag strategy, called “Arginine-rich NLS” (ArgiNLS), that optimizes genetic labeling and downstream image segmentation of single cells by restricting FP localization near-exclusively in the nucleus through a poly-arginine mechanism. A single N-terminal ArgiNLS tag provides modular nuclear restriction consistently across spectrally separate FP variants. ArgiNLS performance *in vivo* displays functional conservation across major cortical cell classes and in response to both local and systemic brain-wide AAV administration. Crucially, the high signal-to-noise ratio afforded by ArgiNLS enhances machine learning-automated segmentation of single cells due to rapid classifier training and enrichment of labeled cell detection within 2D brain sections or 3D volumetric whole-brain image datasets, derived from both staining-amplified and native signal. This genetic strategy provides a simple and flexible basis for precise image segmentation of genetically labeled single cells at scale and paired with behavioral procedures.

NLS | single cell | whole-brain | brain-wide | volumetric

The structure of the mammalian brain is complex and spans multiple spatial scales, requiring the development and application of spatially integrated methods to reveal key relationships with its function. High-throughput volumetric fluorescent microscopy pipelines, facilitated by light-sheet fluorescence microscopy (LSFM) (1) and serial section block-face microscopy (2, 3), have accelerated these efforts. Notably, they have democratized experimental access to spatially integrated intact whole-brain anatomy, at the resolution of single cells and within computationally feasible datasets that are compatible with brain atlas mapping. For example, in combination with genetic fluorescent labeling, these techniques have facilitated the mapping of cellular activity (4–16), cell-type distributions (17–20), and endogenous genetic (21) and connectivity patterns (22).

However, while significant hardware-based advancements have been made in microscopy systems (2, 23, 24), and numerous software-based image analysis tools are readily available (5, 13, 14, 25–29), there has been a relative lack of genetic tool development that maximizes single-cell detection and quantification through the optimization of fluorescent signal. As opposed to hardware- and software-based approaches, genetic approaches may offer improved cellular detection without the prohibitively expensive acquisition of cutting-edge microscopy and computational equipment. Past genetic-based approaches have relied on fluorescent protein (FP) expression in its unmodified or modified forms (30), but the efficacy of these prior efforts has not been systematically examined at the level of single-cell discrimination across image datasets of various sizes. As a result, the quantitative accuracy of single-cell brain mapping measurements remains ambiguous both within and between studies.

Typically, single-cell discrimination is improved by targeted genetic modifications to FPs by i) N- or C-terminal tags of a minimal nuclear localization signal (NLS) sequence from a naturally occurring protein like the simian virus 40 large T antigen (SV40nls) (31–33), or ii) fusion with a full-length nuclear protein, like histone 2B (17, 22, 34). Both approaches target FP localization to the cell nucleus to reduce extranuclear fluorescence. However, these approaches have limitations. Minimal NLS tag modifications only achieve partial nuclear localization and therefore are suboptimal for image segmentation. Conversely, full-length nuclear protein fusions guarantee nuclear restriction but depend on overexpression of the

Significance

Quantifying labeled cells in fluorescent microscopy is a fundamental aspect of modern biology. Critically, the use of short nuclear localization signal (NLS) is a key genetic modification for discriminating single cells labeled with fluorescent proteins (FPs). However, mainstay NLS approaches typically localize proteins to the nucleus with limited efficacy, while alternative non-NLS tag strategies can enhance efficacy at the cost of cellular health. Thus, quantitative cell counting using FP labels remains suboptimal or not compatible with health and behavior. Here, we present a genetic tagging strategy—named Arginine-rich NLS—that flexibly and safely achieves FP nuclear restriction across the brain to facilitate machine learning-based segmentation of single cells at scale, delivering a timely update to the behavioral neuroscientist’s toolkit.

Competing interest statement: E.R.S. and S.A.G. are inventors of a provisional patent related to ArgiNLS genetic technology (W149-0052USP1/49409.01US1). All other authors have no competing interests.

This article is a PNAS Direct Submission. J.F. is a guest editor invited by the Editorial Board.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: szelenyi@uw.edu or sagolden@uw.edu.

²Present address: Allen Institute for Cell Science, Seattle, WA 98109.

³Present address: Cajal Neuroscience, Seattle, WA 98109.

⁴Present address: Department of Neurobiology, University of Maryland School of Medicine, Baltimore, MD 21201.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2320250121/-DCSupplemental>.

Published July 29, 2024.

fused nuclear gene, which can have significant deleterious effects on neuronal function and animal health (35).

Two major bottlenecks have slowed the development of new ectopic NLSs—including those identified *in silico* (36) or synthetic versions developed for genome editing (37, 38)—for use as FP fusion tags. First, NLS performance when ectopically tagged as an FP fusion is unpredictable. This is because the nuclear import efficiency of naturally occurring NLSs is influenced by many factors, including intrinsic NLS sequence affinity to nuclear transport receptors KPNA and KPNB (39), and their interplay with the amino acids surrounding the NLS in its native state (40, 41). Second, due to the barrier posed by challenging biological properties, there is not a high-throughput pipeline to systematically screen novel NLSs for whole-brain mapping compatibility. Those properties include i) minimal extranuclear FP localization for high SNR, ii) a simple and flexible sequence with minimal length and a single location of fused tag for unaltered FP function, iii) modular compatibility with FP variants for spectral flexibility in imaging design, including multiplexing, iv) equal functionality across cell types for unbiased labeling of single cells, and v) nondeleterious effects of cellular and animal health. Thus, a genetic tag strategy that simultaneously synergizes and stereotypes FP-tagged NLS function while maintaining cellular health could satisfy these requirements.

Here, we present a NLS genetic tag strategy named “Arginine-rich NLS” or “ArgiNLS,” that improves nuclear localization of genetically encoded FPs and enhances image segmentation of single cells through a poly-arginine-driven nucleolar enrichment mechanism. This method does not alter neural physiology or mouse behavior after standard AAV expression windows and is therefore compatible with AAV-paired behavioral procedures. It is functionally modular when using spectrally separate FPs and possesses equal efficacy when expressed within genetically distinct cortical cell types. Moreover, it significantly improves two- and three-dimensional image segmentation that decreases computational expense associated with machine learning (ML)-based segmentation.

Results

Design of a Nuclear-Restricting ArgiNLS. Arginine (R) is the most hydrophilic, basic natural amino acid residue with a pKa over 12 and possession of a positive charge within all biological environments. Because of these properties, synthetic polyarginine (poly-R) tags (i.e., a peptide sequence consisting of multiple, R repeats) serve a variety of experimental functions, including their use as a protein purification tag (42–45), cell-penetrating peptide signal, multi-DNA-binding domain linker, and nucleolar trafficking signal (42, 46–48). We hypothesized that a poly-R stretch coded in between a NLS sequence and a downstream FP could act as an improved nuclear-localizing agent within neurons (Fig. 1*A*).

To explore this idea *in vitro*, we first tested the effects of a range of poly-R lengths on the nuclear localization of two known classical NLS (cNLS) types: i) monopartite (MP), composed of a single cluster of 4 to 8 basic amino acids and consensus recognition motif of K (K/R) X (K/R) (X is any amino acid), and ii) bipartite (BP), composed of two clusters of 2 to 3 positively charged amino acids, separated by a 9 to 12 amino acid linker region, and with a consensus sequence R/K(X)10–12KRXX (39, 49) (Fig. 1*B* and *SI Appendix, Table S1*). We selected cNLS sequences derived from the widely applied and characterized MP SV40 (31, 50) or BP nucleoplasmin2 (NPM2) (51) genes, whose NLS protein sequences differ in length and net positive charge but share the

same isoelectric point (pI) across poly-R length additions (Fig. 1*C* and *D*). We performed cotransfections of AAV-(SV40nls or NPM2nls)-poly-R-EGFP overexpression constructs with untagged mKate2 in mouse neuroblastoma 2A (N2A) cells. Nuclear or cellular fluorescence localization and brightness (see *Materials and Methods* for a detailed description quantification approach) in individual EGFP+/mKate2+ cells were measured using DAPI staining as a nuclear marker and untagged mKate2 expression to label entire cells.

Cotransfected EGFP+/mKate+ cells revealed a poly-R length-dependent increase in nucleolar localization for both cNLSs. NPM2nls, and not SV40nls, exhibited decreased extranuclear fluorescence regardless of poly-R length (Fig. 1*E*). Measured composite values across poly-R lengths was 77% (± 1.48) for SV40nls and 91% (± 0.85) for NPM2nls, where NPM2nls significantly obtained stable, higher nuclear localization across all poly-R lengths (Fig. 1*F*). Additionally, unlike the linear decrease observed with SV40nls, nuclear brightness with NPM2nls peaked with poly-R7 and decayed with longer lengths (Fig. 1*G*).

We validated the optimal NPM2nls-R7 configuration in a separate heterologous mouse brain cell line by comparing untagged performance for each cNLS type (Fig. 1*H* and *I*). The control SV40nls tag exhibited 63% of nuclear EGFP fluorescence that was significantly decreased to 55% with the poly-R7 addition, despite a nucleolar-like localization (Fig. 1*H* and *I*). NPM2nls-R7 enhanced nuclear localization above untagged levels from 70 to 95% (Fig. 1*I*). We next probed the specificity of poly-R7 on BP cNLS activity using a BP version of SV40nls (biSV40nls, *SI Appendix, Fig. S1*). We observed a significant enhancement of nuclear-localized EGFP fluorescence with biSV40nls-R7 (93% ± 1.5) versus untagged biSV40nls (70% ± 7.3) (*SI Appendix, Fig. S1 B and C*).

Prior work has shown that a synthetic poly-R8 tag facilitates DNA-binding protein function when inserted as a linker sequence (47), suggesting that poly-Rs can be used as a linker in other molecular contexts. Next, we tested the ability of the NPM2nls-R7 genetic tag—now termed ArgiNLS—to provide modularity with other FPs beyond EGFP (Fig. 1*J–L*). We attached ArgiNLS to a range of spectrally separate FPs that have diverging N-terminal sequences located within the first 10 amino acids, which could influence fused NLS tag activity (Fig. 1*J*). Monomeric FPs included EGFP (Ex/Em: 488/507), oScarlet (52) (Ex/Em: 569/594), mKate2 (53) (Ex/Em: 588/633), and miRFP670 (54) (Ex/Em: 647/670) (Fig. 1*J, Bottom*). In N2A cells, we measured nuclear restriction of native fluorescence for EGFP (93% ± 0.3), oScarlet (93% ± 1.0), mKate2 (93% ± 1.5), and miRFP670 (91% ± 0.3) (Fig. 1*K* and *L*).

Nuclear Restriction of FP Fluorescence In Vivo. The *in vitro* results show enhanced ArgiNLS restriction of nuclear fluorescence. We next determined whether the performance replicated *in vivo* using AAV overexpression in the mouse brain. Nuclear import of NLS-containing proteins—including SV40 and NPM2—are controlled by nuclear transport receptors called importins (or karyopherins) (49). Importins are dynamically expressed in multiple isoforms across cell types and brain regions (55), providing a potential cause of variable NLS activity of NLSs in a cell-type and brain region-dependent manner. We used a conditional AAV coexpression strategy to quantify nuclear-localized EGFP fluorescence in two genetically distinct cortical cell classes, GABAergic or glutamatergic neurons. We reasoned that each class comprises a divergent set of multiple neuronal cell subtypes that have functionally and genetically distinct characteristics (20, 56), and therefore are well suited for benchmarking the nuclear localizing efficacy of ArgiNLS- or conventional SV40nls-tagged EGFP.

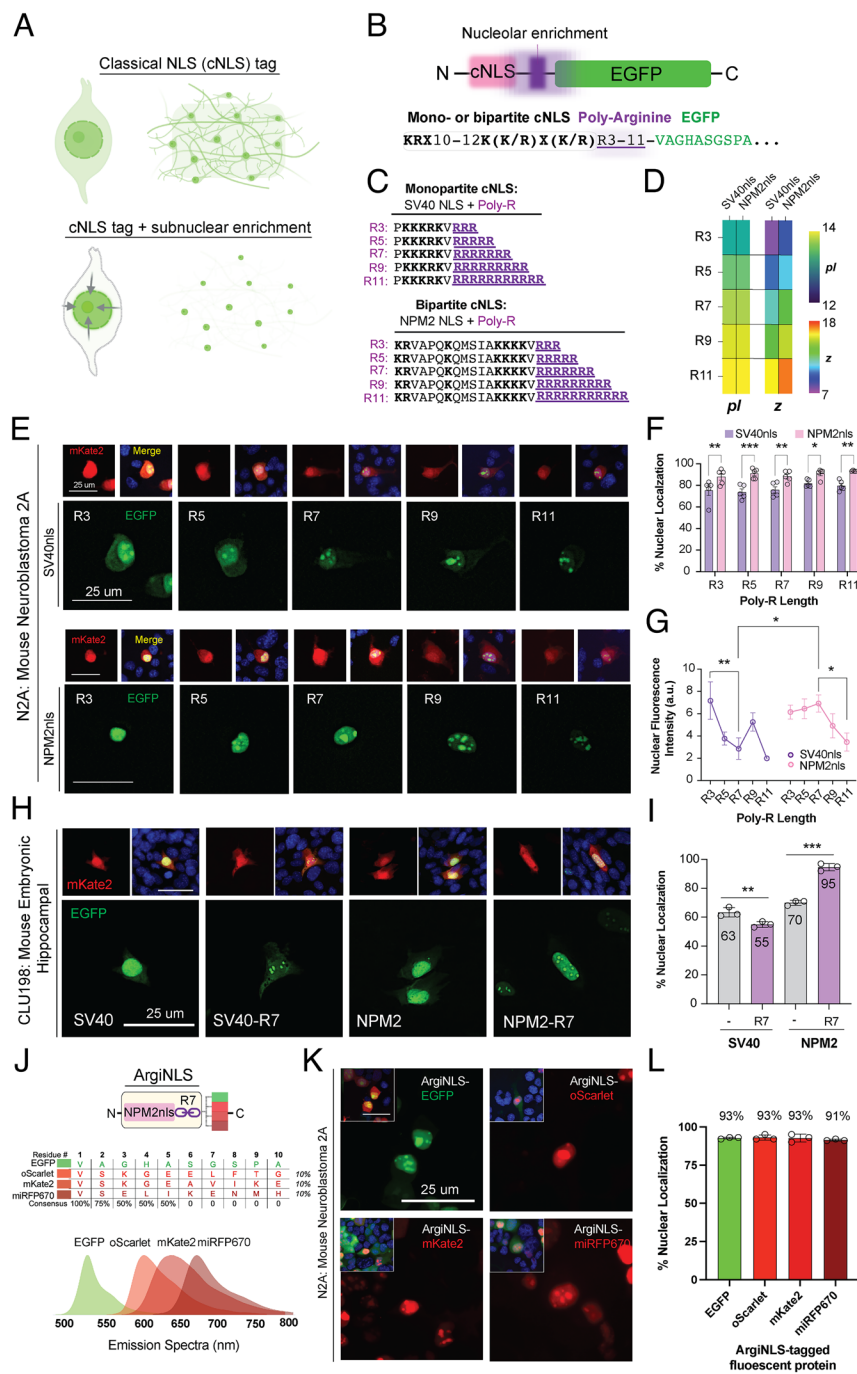


Fig. 1. Strategy and in vitro characterization of an arginine-rich NLS (ArgiNLS). (A) Schematic depicting extracellular (EC) signal caused by classical NLS (cNLS) tag use, and depicting the theoretical effects of an optimized NLS tag that eliminates extranuclear signal artifact through a subnuclear enrichment strategy. (B) Top: Configuration of cNLS, poly-arginine nucleolar enrichment tag, and EGFP in AAV overexpression vectors created for in vitro testing. Bottom: Corresponding amino acid composition of mono- or bipartite cNLS consensus sequences followed by poly-R stretch separating the coding sequence (CDS) of EGFP (first 10 amino acids shown). (C) Amino acid sequences of SV40nls- and NPM2nls-poly-R tags. Basic amino acids are indicated in bold. (D) Heat maps of isoelectric point (pI) and net positive charge (z) of each tag configuration. (E) Representative images of transiently cotransfected single N2A cells displaying untagged mKate2 and SV40nls-poly-R (Top) or NPM2nls-poly-R (Bottom) tagged EGFP expression, spanning 3 to 11 R. (F) Percent nuclear localization of each NLS across poly-R lengths (n = 5/condition). (G) Nuclear fluorescence intensity (a.u. = arbitrary units) across poly-R lengths for each NLS from same cells measured in F. (H) Representative images of transient cotransfections with untagged mKate2 and each experimental EGFP-expressing construct within immortalized CLU198 embryonic hippocampal mouse cell line. (I) Percent nuclear localization with/without R7 addition for each cNLS. (J) Molecular schematic of ArgiNLS tag (NPM2nls-R7) and amino acid sequence comparison and consensus of the first 10 N-terminal residues to EGFP, oScarlet, mKate2, and mRFP670. (K) Representative images of transient cotransfections with (from L to R and top to bottom) ArgiNLS-EGFP/mKate2, ArgiNLS-oScarlet/EGFP, ArgiNLS-mKate2/EGFP, and ArgiNLS-mRFP670/EGFP combinations within N2A mouse cell line. (L) Percent nuclear localization of ArgiNLS-tagged FPs. I and L cell culture summary data are from five blinded and randomly selected cells quantified per transfection condition across three replicate cultures. The data in F were statistically analyzed using a two-way ANOVA followed by Holm-Sidak's multiple comparison post hoc test. The data in G were statistically analyzed using a two-way ANOVA followed by Holm-Sidak's multiple comparison post hoc test for effects of poly-R length both within and across each cNLS. The data in L were statistically analyzed using a two-way (NLS (3) × R7 (2)) ANOVA followed by the Sidak post hoc test. *P < 0.05; **P < 0.01; ***P < 0.005.

We selected the barrel cortex—where GABAergic and glutamatergic cell types are abundant (20, 56)—for AAV coinjections using SV40nls or ArgiNLS-tagged EGFP (AAV-EF1 α -SV40nls- or ArgiNLS-EGFP), combined with a conditional Cre-dependent virus to express somatic tdTomato (AAV-CAG-FLEX-tdTomato) in GABAergic or glutamatergic cells within *vGat-Cre* or *vGlut1-Cre* animals, respectively (Fig. 2A). Confocal images were acquired of the injection site, and nuclear (as defined by DAPI+ costain) and somatic (as defined by tdTomato+ somatic fluorescence) EGFP fluorescence was quantified at the level of single cells across NLS tags and cell types at both high (Fig. 2B and C) and low (SI Appendix, Fig. S2) magnification.

Barrel cortex neurons infected with SV40nls-EGFP exhibited somatic EGFP concomitant with high EC background, putatively from SV40nls-EGFP localized in cellular processes (Fig. 2B and C, Left). Conversely, ArgiNLS-EGFP did not exhibit EC EGFP

background, which corresponded with nuclear-restricted EGFP and localization in presumptive nucleolar puncta (Fig. 2B and C, Right). Sampled EGFP+/tdTomato+ single cells used for nuclear/somatic EGFP fluorescence quantifications contained equal distributions of nuclear localization (corrected total nuclear fluorescence, CTNF; see *Materials and Methods*) levels across tags and cell-type (Fig. 2D). However, when CTNF was expressed as a percent cellular fluorescence (corrected total cellular fluorescence, CTCF; see *Materials and Methods*), nuclear localization variance was significantly different across tags; SV40nls showed high variance while ArgiNLS showed low variance (Fig. 2E). Despite differences in variance, tag type had a significant effect on nuclear localization, where ArgiNLS-EGFP contained significantly higher nuclear localization (GABAergic cells: 88% $\pm$ 0.9; glutamatergic cells: 87% $\pm$ 0.7) than SV40nls-EGFP (GABAergic cells: 65% $\pm$ 1.8; glutamatergic cells: 67% $\pm$ 2.0) (Fig. 2F–H).

We next determined how ArgiNLS impacts EC background fluorescence at the site of injection and within anatomically distal areas (Fig. 2 *I–L*). First, we compared EC pixel intensity measurements taken from field of views (FOVs) positioned in between EGFP-labeled cells of the injected barrel cortex (i.e., ipsilateral), and in between DAPI+ cells of the uninjected (i.e., contralateral) barrel cortex, as a control measurement, from the same mice (Fig. 2*J*). Ipsilateral SV40nls FOVs contained a significant four-fold increase in EC fluorescence compared to uninjected contralateral FOVs, that was also significantly higher than between-subject

ipsilateral and contralateral ArgiNLS FOVs. EC fluorescence within ipsilateral ArgiNLS FOVs was unchanged from contralateral FOV levels from both SV40nls and ArgiNLS injected samples.

Neurons within the barrel cortex send diverse axonal projections throughout the brain (57). ArgiNLS consistently eliminated non-nuclear labeling of axon fiber of passage in the CC, as well as axonal terminations in the dorsolateral thalamus (dlTH), that was observed with SV40nls-EGFP injections (Fig. 2*J*). We then confirmed decreased CC axon labeling by ArgiNLS in separate

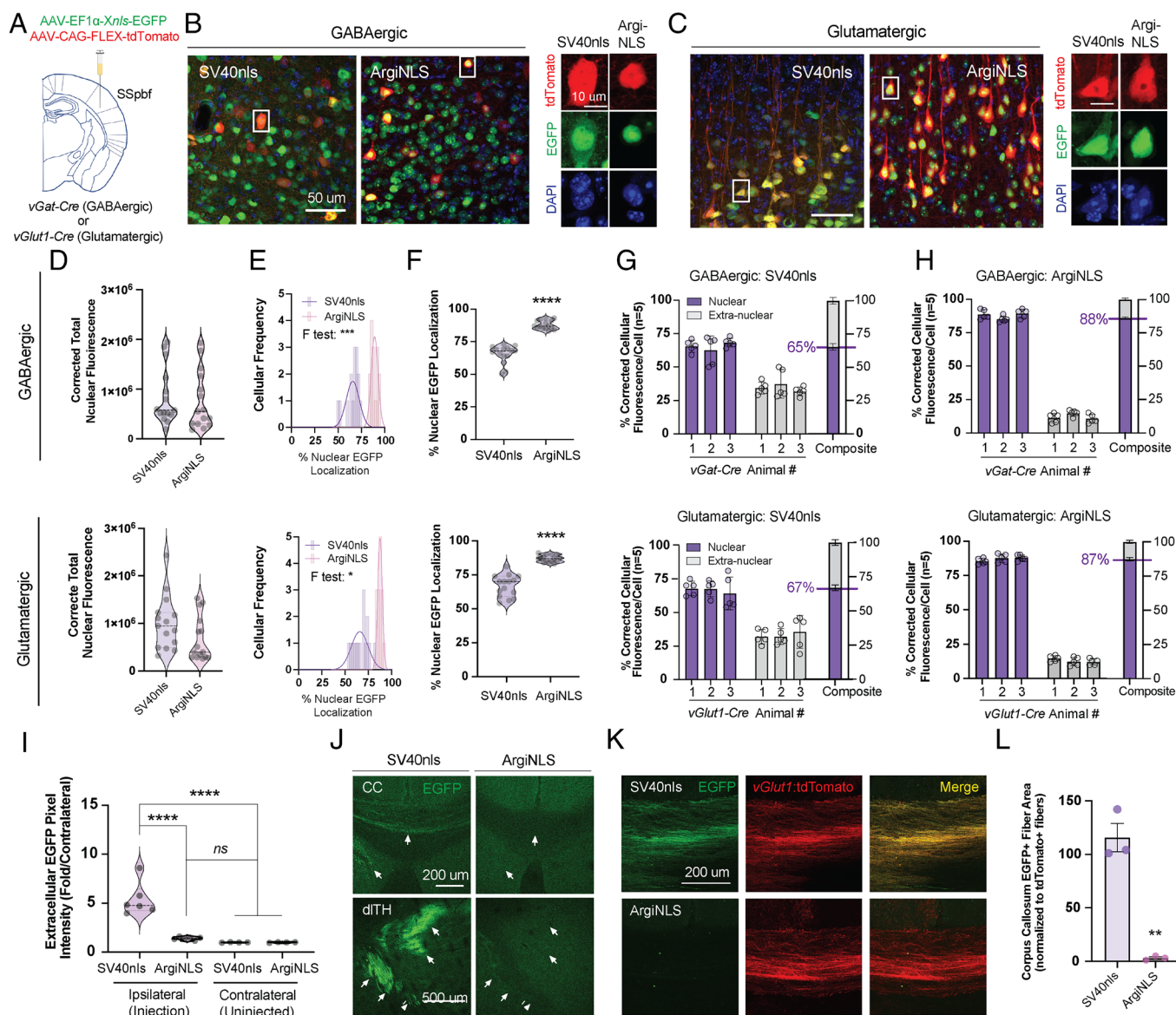


Fig. 2. Optimized nuclear localization of fluorescence across major cortical cell-type classes in vivo. (A) Schematic of stereotaxic AAV coinjection in the barrel field of the primary somatosensory cortex (SSpBF) of *vGat-Cre* or *vGlut1-Cre* transgenic mice. (B and C) *Left*: Representative composite 40x confocal images of SV40nls-EGFP/FLEX-tdTomato and ArgiNLS-EGFP/FLEX-tdTomato viral expression in (B) *vGat-Cre* or (C) *vGlut1-Cre* mice. *Right*: Cropped representative image of an example EGFP+/tdTomato+ cell displayed across separate red (tdTomato), green (EGFP), and blue (DAPI) channels. (D–H) Quantification of NLS tag performance in GABAergic (*Top*) or glutamatergic (*Bottom*) EGFP+/tdTomato+ cells (*n* = random 5 cells/animal; *n* = 3 animals per Cre-driver/tag). (D) EGFP corrected total nuclear fluorescence (CTNF). (E) Cellular frequency histogram of percent nuclear EGFP localization. (F) Percent nuclear EGFP fluorescence comparison across tags. (G and H) Nuclear and extranuclear EGFP fluorescence for all quantified cells of each animal across (G) SV40nls- or (H) ArgiNLS-labeled cells. The composite percent nuclear EGFP across all animals is indicated to the right. (I) Total pixel intensity of EC EGFP signal ipsilateral to the injection site, expressed as a fold change from noninjected contralateral side (*n* = 3 EC quantifications/3 *vGlut1-Cre* mice). (J) Representative images of extranuclear EGFP signal in axon bundles of passage in the corpus callosum (CC; *Top*) or terminations in the dorsolateral thalamus (dlTH; *Bottom*). Arrows indicate the location of labeling observed with SV40nls and are superimposed on ArgiNLS-EGFP. (K) Representative 40x confocal images of SV40nls- or ArgiNLS-EGFP with *vGlut1:tdTomato* viral expression in fibers of the CC (*n* = 3 EC quantifications/3 *vGlut1-Cre* mice). (L) Quantification of EGFP+ fiber area in CC normalized to *vGlut1:tdTomato*+ fiber area. All bar graphs represent mean $\pm$ SEM. The data in D, F, and L was statistically analyzed using an unpaired *t* test. Group differences in population variance of data in E were statistically compared using an F-test. The data in I were statistically analyzed using a 1-way ANOVA followed by Tukey's multiple comparison post hoc test. Violin plots display median and interquartile range as dashed lines. **P* < 0.01; ***P* < 0.05; ****P* < 0.01; *****P* < 0.001.

quantifications made of EGFP sampled in high-magnification FOVs. To control for error in sample-to-sample variability of viral delivery, we restricted our analysis within the *vGlut1-Cre* injected sections and normalized EGFP+ fiber area to *vGlut1:tdTomato*+ fiber area (Fig. 2 *K* and *L*). Accordingly, SV40nls-EGFP labeled 115% (± 13) of *vGlut1:tdTomato*+ fibers, indicating that extranuclear EGFP localization produced from constitutively expressed SV40nls-EGFP labels CC fibers within both *vGlut1*(+) and *vGlut1*(-) neurons of the barrel cortex (Fig. 2*L*). Conversely, ArgiNLS-EGFP labeled 3% (± 1.2) of *vGlut1:tdTomato*+ fibers.

Nondeleterious to Neuronal and Behavioral Health. Full-length nuclear protein-FP fusions, such as H2B-EGFP (58), demonstrate high levels of nuclear-restricted fluorescence, but cause detrimental effects to cellular and animal health when overexpressed in the mouse brain (35). Since ArgiNLS-tagged FPs display similar nuclear-restricted fluorescence, we next determined whether ArgiNLS-EGFP viral overexpression in the brain causes detrimental effects to baseline neuronal physiology or animal behavior (despite it lacking the fusion of an endogenous nuclear protein) (Fig. 3). We first examined baseline physiological properties of glutamatergic neurons within the barrel cortex of *vGlut1-Cre* animals. We used a local unilateral injection of Cre-dependent tdTomato virus to create control *vGlut1-Cre:tdTomato*(+) labeled neurons, and a contralateral coinjection of Cre-dependent tdTomato with ArgiNLS-EGFP viruses to create within-sample, experimental ArgiNLS-EGFP(+)/*vGlut1-Cre:tdTomato*(+) labeled neurons (Fig. 3*A*). Analysis of baseline physiological parameters indicated no effect of ArgiNLS-EGFP overexpression on resting membrane potential (Fig. 3*B*) and action potential generation across a range of injected currents (Fig. 3 *C* and *D*).

Next, we used the PHP.eB viral serotype to examine the effect of systemic whole-brain ArgiNLS-EGFP or control virus expression—including a nonexpressing negative Cre-dependent tdTomato control virus, and positive SV40nls-EGFP and H2B-EGFP expressing controls—on behaviors including the OFT and EPM to study effects on acute exploratory drive, and a 60-h locomotion monitoring assay to study chronic locomotion and sensorimotor performance across dark/light cycles (Fig. 3 *G–M*). We found no differences in acute OFT performance across viral groups as measured by mean total distance (cm) traveled (Fig. 3*G*), or percent center time duration (Fig. 3*H*). Additionally, acute EPM performance as measured by open arm entry frequency (Fig. 3*I*), or open arm total duration (s) (Fig. 3*J*) was the same across viral groups. Consecutive monitoring of homecage distance traveled (cm) across three nights and two days indicated a lack of effect from ArgiNLS-EGFP or SV40nls-EGFP systemic overexpression (Fig. 3 *K–M*). Conversely, H2B-EGFP-injected animals displayed dysregulated locomotor activity during the light phase. We confirmed that animals used for behavioral phenotyping contained robust systemic expression in the brain by performing 2D histology and imaging in a subset of animals (*SI Appendix*, Fig. S3).

Enhanced Compatibility with ML Classification. Image classification of objects can be reliably automated and scaled through the creation of ML algorithms or “classifiers” that rapidly extract features based on pixel content to create image segmentations for quantification (59, 60). For fluorescently labeled images, the manual effort taken to compile training datasets is directly tied to the fluorescent signal’s fidelity to label a target class. Thus, poorly defined labeling may require extensive training and user annotation. Since ArgiNLS enables consistent FP restriction in the nucleus of neurons, we next asked whether this influences

the creation—or performance—of a single-cell ML classifier, as compared to a classifier trained with cSV40nls tag labeling (Fig. 4).

We systemically labeled single cells with SV40nls-EGFP or ArgiNLS-EGFP-expressing AAV.eB viruses through RO injections (Fig. 4*A*) and used DAPI counterstained images of coronal sections (2.11 and -0.52 from bregma) to benchmark single-cell ML classifier training and testing at varying anatomical locations throughout the brain (Fig. 4*B*). We observed that SV40nls displayed variable extranuclear fluorescence and background, and ArgiNLS conversely showed nuclear-restriction and an overall higher density of labeled single cells (Fig. 4 *C* and *D*). The difference in FP localization within brain cells generalized to the liver within a separate cohort of animals systemically infected with a higher titer of virus (*SI Appendix*, Fig. S4).

We next created single-cell classifiers for SV40nls and ArgiNLS labels and benchmarked their performance using F1 scores: a ML metric that determines model accuracy based on a combination of precision (i.e., positive predictions) and recall (i.e., correct identification of positive class) measurements. We performed multiple training iterations consisting of two sample FOVs per session until maximum classifier performance, as measured by F1 scoring in FOVs with similar ground truth (GT) cell detections, was reached (Fig. 4*E* and *SI Appendix*, Fig. S5). ArgiNLS (5 rounds, 10 input FOV) resulted in a substantial reduction in the total number of training iterations required to reach max classifier performance at both anatomical planes compared to SV40nls (14 rounds, 28 input FOV) (Fig. 4 *F–H*). Specifically, for AP 2.11, SV40nls classification started with a 93% greater total amount of cell detections than GT, that was composed of 11% correct (24 ± 3.1) and 89% (222 ± 3.8) incorrect detections. Both correct and incorrect detections displayed an exponential-like scaling toward GT levels as function of training round number (Fig. 4 *F*, *Left*). Initial ArgiNLS classification consisted of 63% more total cell detections (184 ± 7.5 ; Correct: 87 ± 7.0 ; Incorrect: 96 ± 13.9) than GT, with correct cells matching 90% of GT (Fig. 4 *G*, *Left*), and incorrect cells progressively reduced over each training session.

Total cell detections for both classifiers at AP -0.52 were stable across all training iterations and corresponded closely to GT detection levels with training (Fig. 4 *F* and *G*, *Right*). Like the raw data, ArgiNLS classification outperformed SV40nls classification in both precision (Fig. 4 *H*, *Top*) and recall (Fig. 4 *H*, *Middle*). ArgiNLS significantly outperformed SV40nls in F-score across nearly all training tile input FOV iterations (Fig. 4 *H*, *Bottom*).

Enriched Brain-Wide Segmentation of Single Cells. We next sought to determine ArgiNLS tag’s impact on single-cell segmentation within a whole-brain mapping dataset using large-scale ML classification. We systemically delivered PHP.eB SV40nls-EGFP and ArgiNLS-EGFP expressing viruses and performed iDISCO+ whole-mount clearing and EGFP staining to fluorescently label all the infected cells throughout the brain (Fig. 5 *A*, *Top* and *B*). Whole-brain samples were imaged with LSFM, and image datasets were run through a pipeline consisting of reference atlas registration, ML cellular segmentation, atlas segmentation of brain regions, and voxelized cell density analyses (Fig. 5 *A*, *Bottom*). Like our observations in 2D histology, SV40nls-EGFP staining produced higher image artifact and less cell counts than ArgiNLS-EGFP (Fig. 5 *B*, *Left*). We therefore trained separate ML classifiers using ilastik software (61) to achieve optimal single-cell classification performance, separately for each tag (*SI Appendix*, Fig. S6*A*). ArgiNLS reduced the model training requirements in 3D, where SV40nls classifier training required a total dataset of 44 training images and ArgiNLS required 20 training images (*SI Appendix*, Fig. S6*B*).

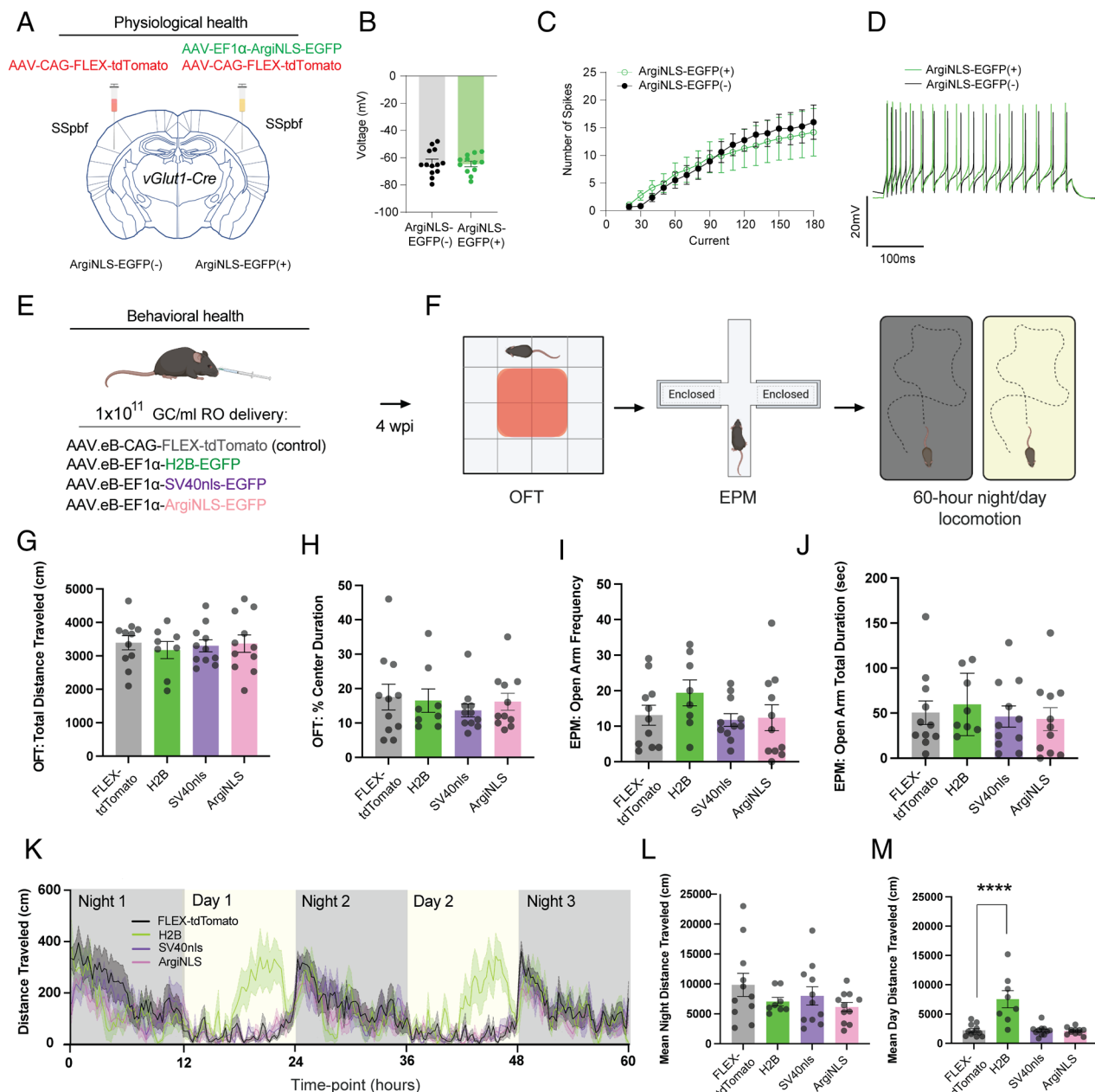


Fig. 3. ArgiNLS-EGFP viral expression is not associated with physiological or behavioral deficits. (A) Schematic illustration of bilateral stereotaxic AAV injection/coinjection in barrel cortices (SSpBF) of *vGlut1-Cre* transgenic mice, and electrophysiological recordings conducted on resulting *vGlut1-Cre*:tdTomato(+) neurons that are either ArgiNLS-EGFP(−) or ArgiNLS-EGFP(+). (B) Resting membrane potential of ArgiNLS-EGFP(−) ($n = 13$) or ArgiNLS-EGFP(+) ($n = 12$) neurons. (C) Current injection curve from ArgiNLS-EGFP(−) or ArgiNLS-EGFP(+) neurons. (D) Example trace current injection at 100 pA. (E) AAV.eB delivery and vector information schematic. Groups of mice were injected with 1×10^{11} GC of a negative control AAV.eB-CAG-FLEX-tdTomato ($n = 11$), AAV.eB-EF1 α -H2B-EGFP ($n = 8$), AAV.eB-EF1 α -SV40nls-EGFP ($n = 11$), or AAV.eB-EF1 α -ArgiNLS-EGFP ($n = 11$). (F) Baseline behavioral battery 4 wk post viral injection (wpi). Mice were sequentially tested in the open field test (OFT), elevated plus maze (EPM), and 60-h night/dark locomotion recording. (G) Total distance traveled in the OFT. (H) Percent time duration spent in the center of the OFT arena. (I) Frequency of entries in the open arm of the EPM. (J) Total duration of time spent in the open arm of the EPM. (K) Locomotion XY plot showing total distance traveled across time of the 60-h recording. Day (yellow)/night (gray) phase of the light cycle is indicated according to time of 60-h recording. (L) Mean night distance traveled across the 60-h recording period. (M) Mean day distance traveled across the 60-h recording period. All bar graphs represent mean $\pm$ SEM. The data for each panel from G–J, L, and M were statistically analyzed using a one-way ANOVA with Dunnett's multiple comparison post hoc test of each experimental group versus control FLEX-tdTomato group. **** $p < 0.0001$.

Following extensive classifier validation (SI Appendix, Text), we applied our whole-brain image processing pipeline to create maps of systemic AAV.eB infection (62). We found that ArgiNLS-EGFP significantly increases AAV.eB labeled whole-brain cell counts by 2.6-fold (SV40nls: $514,422 \pm 90,388$; ArgiNLS: $1,362,466 \pm 268,889$) (Fig. 5C). Additionally, we asked whether the whole-brain count difference was anatomically distributed preferentially or evenly across gross areas of the brain. Statistically significant increases of ArgiNLS-labeled cell densities were found specifically in the isocortex

(ISO) (SV40nls: $1,798 \pm 277$, ArgiNLS: $5,526 \pm 1,156$) and olfactory area (OLF) (SV40nls: $1,162 \pm 211$, ArgiNLS: $3,753 \pm 609$), suggesting these two brain areas are major drivers of the differences in whole-brain cell counts.

To identify subregional differences in labeled cell densities, we used 1,216 anatomical annotations provided by the Unified atlas (63). First, we asked whether the general distribution of labeled cell patterning across subregions differs among genetic tags (Fig. 5E). Linear regression analysis revealed a statistically significant positive

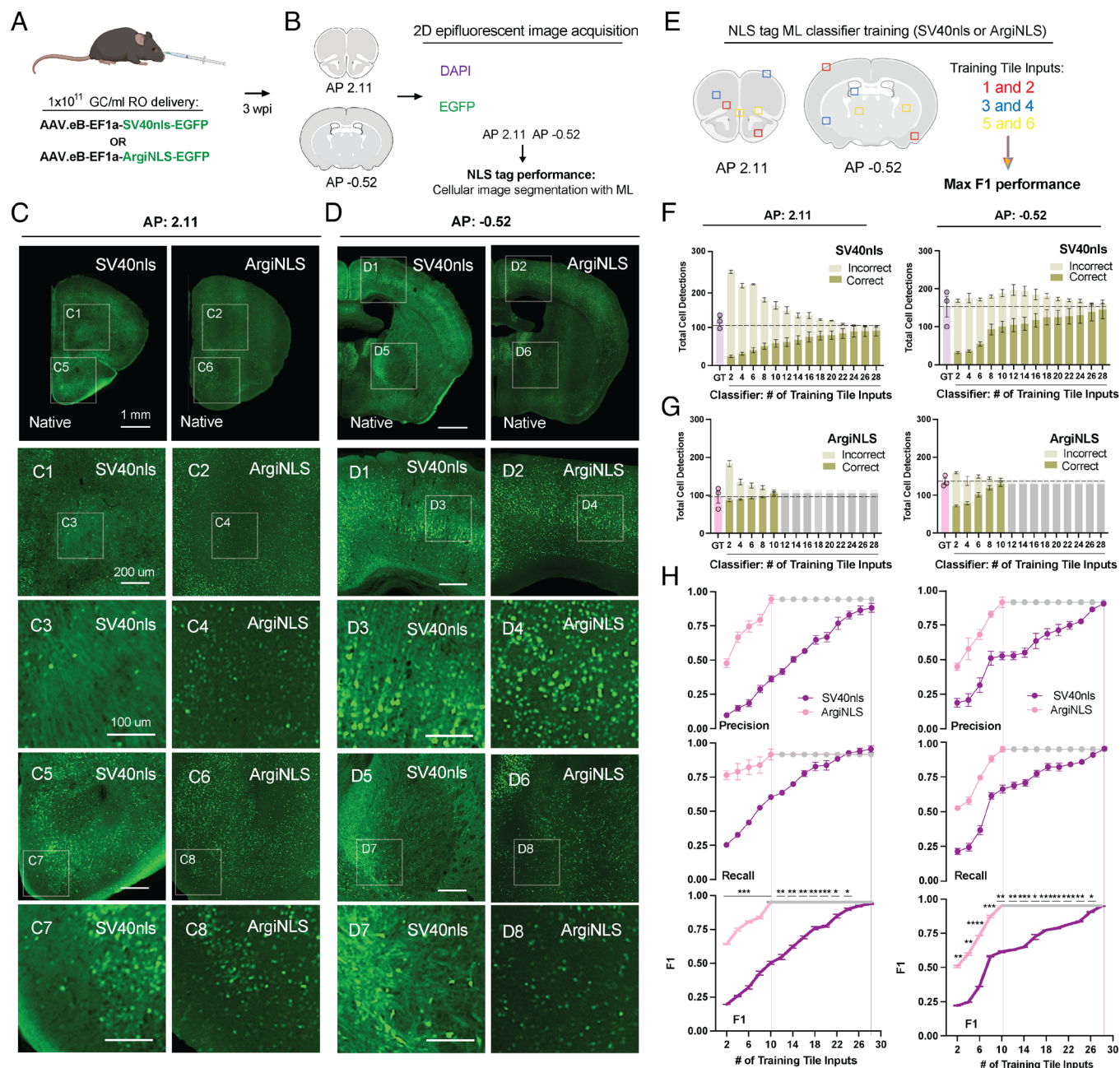


Fig. 4. Enhanced performance of single-cell ML classification. (A) Experimental schematic of systemic AAV.eB delivery of NLS-tagged EGFP. (B) Microscopy acquisition schematic. (C and D) Representative images of EGFP viral expression at matching (C) AP 2.11 or (D) -0.52 hemispheric coronal planes, across two FOVs at three magnification levels. (E) Schematic of SV40nls and ArgiNLS classifier training workflow for single-cell segmentation of 2D images. (F-H) Cell segmentation classifier benchmarking and performance metrics for (Left) AP: 2.11 and (Right) AP: -0.52 brain sections. Total cell detections for (F) SV40nls-EGFP or (G) ArgiNLS-EGFP for expert rater ground truth (GT; left bar) ($n = 3$) and ML classifier across iterations of two-increment additive training input rounds. Gray bars represent iterations after peak performance criteria were met. The dashed line indicates mean GT cell detections. (H) Precision, recall, and F1 harmonic scoring of single-cell detection for SV40nls-EGFP and ArgiNLS-EGFP. Gray data points represent iterations after max F1 performance was met. All bar graphs represent mean $\pm$ SEM. The F1 data in H were statistically analyzed using a two-way repeated measures ANOVA followed by Holm-Sidak's post hoc test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

correlation of ArgiNLS versus SV40nls labeled cell densities of subregions, suggesting an equivalent anatomical distribution of labeled cells across subregions. However, with the exception of 30 subregions, 98% contained greater cell densities when labeled with ArgiNLS (Fig. 5 E and F).

Second, we screened for the identities of subregions significantly enriched in ArgiNLS-labeled cell density (Fig. 5 F, Bottom and G, SI Appendix, Figs. S7 and S8, and Dataset S1). The most enriched gross regions—the ISO and OLF—contained a compilation of 60 [208/345; e.g., primary somatosensory area (S1)] and 23% [16/69; e.g., anterior olfactory area (AO), piriform cortex (Pir)] significant

subregions, respectively (Fig. 5 G and SI Appendix, Fig. S6). We found that ISO cortical enrichment was predominant in layers 2/3 to 5, with heavy concentration in areas related to sensory processing (SI Appendix, Fig. S8). We additionally found significant enrichment of ArgiNLS-labeled cell density in subregions broadly located in every major division of the brain [HPF: 13/82 = 16%, e.g., dorsolateral entorhinal cortex (DLEnt), subiculum (SUB); CTXsp: 1/23 = 4%, e.g., intermediate endopiriform claustrum (IEn); CNU: 4/120 = 3%, olfactory tubercle (Tu); CB: 1/65 = 1.5%, e.g., lateral cerebellar nucleus parvicellular part (LatPC); TH: 6/94 = 6%, e.g., medial geniculate nucleus (MG); HY: 1/104 = 1%, medial mammillary

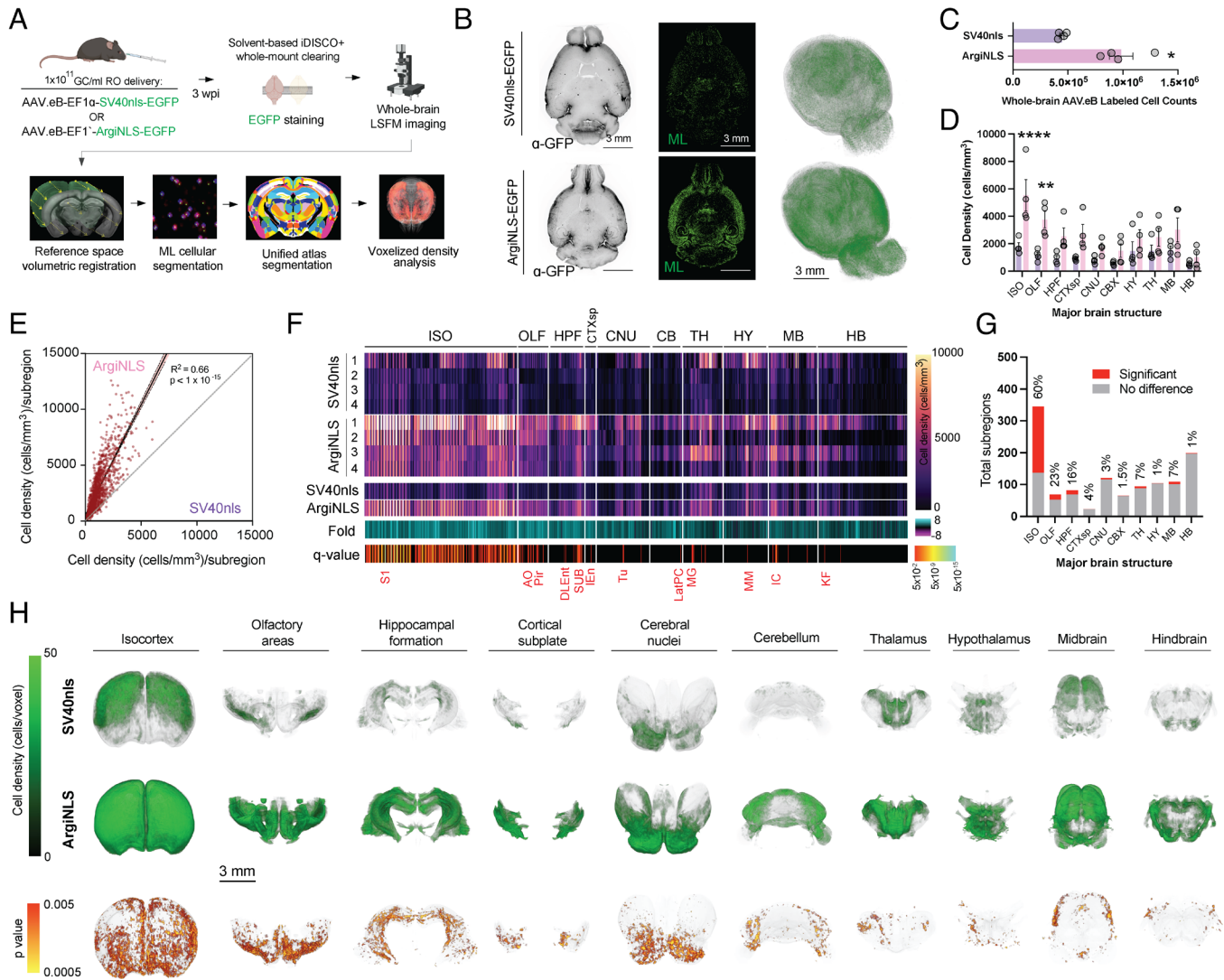


Fig. 5. Enriched brain-wide segmentation of single cells. (A) Schematic of experimental and image processing workflow used to quantify systemic SV40nls- or ArgiNLS-EGFP-labeled cells across the whole mouse brain. (B) *Left:* Raw LSFM image in horizontal orientation from a representative SV40nls-EGFP (*Top*) or ArgiNLS-EGFP (*Bottom*) expressing brain sample. *Middle:* Corresponding ML-segmented cells within the same optical section. *Right:* Whole-brain 3D renderings of voxelized cell density for each sample. (C) Comparison of whole-brain AAV.eB-labeled cell counts. (D) Comparison of cell density across major ontological brain structures. (E) Scatterplot and linear regression of ArgiNLS (Y axis) versus SV40nls (X axis)-labeled mean cell density across all subdivisions of the brain. (F) 2D heatmap plot of (from top to bottom): cell density across each individual sample, mean cell density, fold change over ArgiNLS, and statistical significance of mean cell density differences across all anatomical subdivisions by structure order. Major ontological brain structure positioning is indicated with white lines and written above. Example statistically significant subdivisions are listed in red. (G) Bar graph of total statistically significant subdivisions versus no difference plotted for each major brain structure. Percent significant subdivisions are listed above each bar. (H) Volumetric renderings of mean (*Top*) SV40nls-EGFP or (*Middle*) ArgiNLS-EGFP voxelized cell densities and the (*Bottom*) statistically significant voxels ($P < 0.005$) of the mean differences across all major brain structures. All bar graphs represent mean $\pm$ SEM. The data for panel C were statistically analyzed using an unpaired Student's t test. The data for panel D were statistically analyzed using a two-way ANOVA with the Bonferroni post hoc test for mean significant differences at each major region. The data for panel E were analyzed with linear regression. The data for panel F were analyzed with FDR-corrected multiple t tests using the two-stage step-up method of Benjamini, Krieger, and Yekutieli. $*P < 0.05$; $**P < 0.01$. Acronyms: ISO = isocortex; OLF = olfactory areas; HPF = hippocampal formation; CTXsp = cortical subplate; CNU = cerebral nuclei; CBX = cerebellum; HY = hypothalamus; TH = thalamus; MB = midbrain; HB = hindbrain; AO = accessory olfactory area; Pir = piriform cortex; DLent = dorsolateral entorhinal cortex; SUB = subiculum; Tu = olfactory tubercle; MG = medial geniculate; MM = medial mammillary area; IC = inferior colliculus; and KF = Kolliker-Fuse subnucleus.

nucleus (MM); MB: 8/109 = 7%, e.g., inferior colliculus (IC); and HB: 2/198 = 1%, e.g., Kolliker-fuse nucleus (KF)] (Fig. 5G and SI Appendix, Fig. S8). We additionally performed an orthogonal, voxel-by-voxel analysis of differences in brain-wide cell density similar to recent studies (10, 64–66). To account for the uncorrected nature of voxel comparisons, we limited the visualization of significant voxels to a more conservative cutoff at $P < 0.005$, maintaining a threshold in 3D that more closely reflects the FDR-corrected approach used in more common ROI analyses. Like our region-by-region analysis, significantly greater voxels were visualized across all major brain areas confirming wide-spread enrichment of labeled cell density by ArgiNLS (Fig. 5H and Movie S1).

ArgiNLS-Optimized Applications in Volumetric Brain Cell Counting. Spatially integrated features of brain structure and function can be studied through volumetric imaging paired with genetically encoded fluorescent labels. For example, many transgenic lines are available for, among other purposes, guiding FP labeling among a constellation of cell types of choice (67). Further, AAV technology is advancing rapidly as seen by recent serotype development directing FP labeling in retrograde, transsynaptic, or systemic fashions (68). Since ArgiNLS optimizes single-cell FP labeling and segmentation, we established its utility for cell counting applications when incorporated into a variety of molecular genetic tools.

We first compiled an AAV vector toolkit consisting of multiple ArgiNLS-tagged FPs, including bright green- [AausFP1 (69) and mGreenLantern (70), yellow- [mVenus ME (71) and red-emitting [oScarlet (52) variants within constitutive and conditional formats to offer experimental range. We included several DREADD receptor (72)-expressing constructs that coexpress ArgiNLS-tagged FPs as a reporter gene, which enables high-fidelity labeling of DREADD-modulated single cells with behavioral compatibility (SI Appendix, Table S2). We applied aqueous-based SHIELD clearing (73) and axial swept LSMF imaging (24) to demonstrate ease of ArgiNLS-assisted segmentation of single cells and their total count quantification using 3D ML from native fluorescence in 3D volumes. We demonstrated ArgiNLS-assisted quantification by generating and injecting a retrograde AAV-EF1 α -ArgiNLS-mKate2 virus into the basomedial amygdala (BMA)—an amygdalar structure involved in anxiety and fear modulation (74)—to create an anatomical input map (Fig. 6A–C). Coinjected EGFP-expressing virus confirmed BMA injection placement, and native retro-ArgiNLS-mKate2+ single cells were readily identifiable at local and long-range distances from the injection site (Fig. 6A). 3D ML classification of brain-wide ArgiNLS-mKate2+ cells quantified at total of 2.4×10^5 BMA inputs that were mainly distributed continuously throughout the ipsilateral cortical subplate including among multiple amygdalar subregions (Amy) and claustrum (Cl), as well as within the subiculum of the hippocampus (Fig. 6B). Major long-range BMA input cells resided in five nodes including (from anterior to posterior) the accessory olfactory bulb (AOB), infralimbic cortex (IL), anterior paraventricular nucleus (aPVT) of the thalamus, contralateral amygdala, and ventral premammillary nucleus of the hypothalamus (PMv). The aPVT represented one of the largest long-range and singular thalamic node containing 5.6×10^3 input cells, highlighting the existence of a putative aPVT-BMA circuit motif of unknown behavioral function.

Second, we generated and locally injected AAV.eB-CAG-FLEX-ArgiNLS-EGFP and AAV.eB-CAG-FDIO-ArgiNLS-mKate2 virus into the barrel cortex of *vGat-Cre* or *vGat-Flp* driver mice, respectively, to conditionally label and 3D segment GABAergic single cells (Fig. 6D and E). High SNR-labeled single cells were identically observed in Cre- or Flp-dependent conditions in 2D (Fig. 6D and E, Left). In 3D, a clear distribution of single-cell-labeled viral spread was observed and devoid of extranuclear signal artifact in both Cre- and Flp-dependent conditions (Fig. 6D and E, Middle), with 3D ML segmenting 1.8×10^4 and 2.5×10^4 GABAergic cells, respectively (Fig. 6D and E, Right).

Finally, we generated and systemically delivered an AAV.eB-EF1 α -ArgiNLS-oScarlet virus to demonstrate robust native fluorescence labeling of single cells and their ease of segmentation of large cell amounts ubiquitously throughout the brain (Fig. 6F and Movie S2). To gauge sensitivity of detection levels enabled by ArgiNLS, we systemically delivered three amounts— 1×10^{10} (1x), 1×10^{11} (10x), and 2×10^{11} (20x)—of viral GC payloads. 3D renderings of native signal displayed sharply contrasted single cells distributed widely throughout the brain, and scaled with viral GC amount (Fig. 6F, Left). 3D ML quantification of single hemispheres segmented 1.3×10^5 , 1.8×10^6 (14-fold), and 3.4×10^6 (26-fold) single cells for 1×10^{10} , 1×10^{11} , and 2×10^{11} GC payloads, respectively, or 1 cell labeled for every 58,173 viral GCs delivered (Fig. 6F, Left and SI Appendix, Fig. S9A, Left). Expressing fold change values over the lowest 1×10^{10} GC values indicated the utility of ArgiNLS-oScarlet to precisely label and segment 1.3-fold cell count increases for every 1-fold increase in viral GC delivered, within the 1- to 20-fold (i.e., 1×10^{10} to 2×10^{11}) GC range tested (SI Appendix, Fig. S9A, Right). This total single-cell detection by viral GC relationship was maintained evenly

throughout the brain as observed in cell counts across separate horizontal planes located in dorsal, central, and ventral anatomical positions (Fig. 6F, Right and SI Appendix, Fig. S9B).

Discussion

Volumetric imaging methods can be used to unveil spatially integrated representations—or maps—of brain structure and function at the direct level of single cells. Pivotal to this approach is the ability of genetically encoded fluorescent labels to discriminate single cells within dense and complex brain tissue. Here, we developed and characterized a nuclear-localizing genetic tag strategy named ArgiNLS that optimizes FP labeling of single cells, without off-target effects on cellular physiology or animal behavior. By extension, ArgiNLS provides seamless compatibility and enrichment of ML-automated brain cell segmentation across small-to-large scale image datasets of various applications, demonstrating its use as an improved single-cell resolution brain mapping genetic tool.

Unlike other recently developed synthetic NLS approaches created for enhanced genome editing (37, 38, 75), the ArgiNLS approach for image segmentation is based on a well-understood molecular foundation. This approach uses a poly-R-based rerouting mechanism to reduce extranuclear NLS-tagged FP localization by attracting it into the nucleolus. Optimal tag efficacy of ArgiNLS occurs when the BP cNLS sequence is combined with a specific poly-R length (R7) that titrates FP localization into a balanced nucleoplasmic/nucleolar composition with optimal native brightness. MP cNLS combinations fail to reduce extranuclear artifact in combinations up to 11 Rs, suggesting that the function of R7 synergism on ectopically fused NLS activity requires sequence and/or length characteristics of a BP NLS. The function of this configuration within cells outside of the CNS and liver, and among varied species was not determined in our current study, and therefore context-specific modification may be required. However, as demonstrated in the mouse brain here, the ArgiNLS strategy lays a robust framework for tailoring function through the parameterization of BP NLS sequence selection and/or poly-R lengths.

The nature of the poly-R7 component of ArgiNLS additionally provides functional fidelity across a wide variety of FPs, that can provide spectral flexibility for single-cell labeling, and potential nuclear localization of other non-FP proteins, through its dual role as a protein linker. This capability was demonstrated on an initial group of spectrally separate FPs. However, further validation among a larger FP selection and with non-FP proteins for applications outside of image segmentation is necessary to strengthen the generalization of this linker-like modularity. Additionally, while our current ArgiNLS strategy applies N terminus tags to FPs, we did not test linker-like capability or function when attached to the C terminus, thereby limiting its current principles of use to a single fusion location.

NLS performance of ArgiNLS in vivo demonstrates several advantages over current SV40nls tag or H2B fusion single-cell labeling approaches. First, ArgiNLS causes FP nuclear restriction that is equally robust whether an AAV is delivered locally or systemically to the brain. ArgiNLS eliminates background fluorescence-producing, extranuclear signal that is consistently observed with SV40nls. ArgiNLS elicits consistent nuclear-restricted FP labeling across major cortical cell classes. This property enables improved ML image classification of intact whole brain tissue, which is characterized by a reduction in model training requirements and coupled to increased precision and recall of segmentations. Unlike H2B fusion (35), these benefits were not associated with deleterious effects to neural physiology or animal behavior. Therefore,

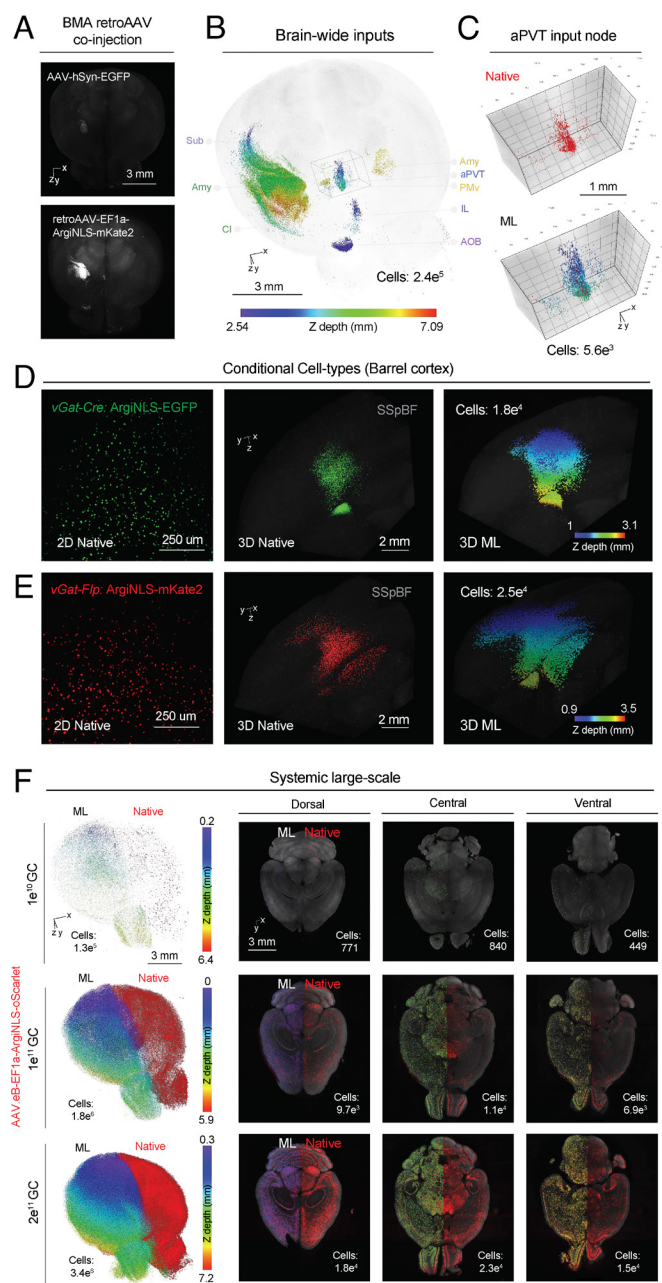


Fig. 6. Optimized volumetric brain cell counting applications with ArgiNLS viral vectors. (A–C) Example circuit mapping with ArgiNLS-mKate2-expressing retroAAV into the basomedial amygdala (BMA). (A) *Top*: Whole-brain volumetric rendering of AAV-hSyn-EGFP native expression labeling the injection site in the green channel. *Bottom*: Corresponding red channel image displaying native fluorescence of ArgiNLS-mKate2+ single-cell inputs. (B) Volumetric rendering of single-cell brain-wide input segmentations and their total counts using 3D ML, gradient color coded by Z depth. Total BMA input cells are listed in the bottom right. Main anatomical areas of local and long-range inputs are listed to the left and right, respectively. Sub = subiculum; Amy = amygdala; Cl = claustrum; aPVT = anterior paraventricular nucleus of the thalamus; PMv = ventral premammillary nucleus; IL = infralimbic cortex; AOB = accessory olfactory bulb. (C) Single cells (native fluorescence; *Top*; ML segmentation; *Bottom*) and their total count from the anterior PVT (aPVT) input node cropped and expanded from the outlined inlay in B. (D and E) Conditional labeling of GABAergic single cells in the barrel cortex using local (D) Cre- or (E) Flp-dependent AAV injections in vGAT-Cre or vGAT-Flp driver mice, respectively. *Left*: 2D confocal image stacks displaying native conditional single-cell labeling from separate example brain samples. *Middle*: Example volumetric renderings of native conditional single-cell labeling from separate brain samples. *Right*: Volumetric renderings of 3D ML-segmented single cells and their total counts for each sample. (F) Large-scale labeling of single cells using systemic delivery of AAV.eB-EF1a-ArgiNLS-oScarlet virus at 1×10^{10} , 1×10^{11} , or 2×10^{11} GC payloads. *Left*: Whole-brain volumetric renderings of combined hemispheric native ArgiNLS-oScarlet fluorescence (left hemisphere) and 3D ML segmentation (right hemisphere) of single cells from separate example brain samples per viral condition. Total hemispheric cell counts are listed in the bottom left. *Right*: 2D horizontal planes of each sample across (top to bottom) dorsal, central, and ventral positions. Images display native ArgiNLS-oScarlet fluorescence with overlaid 3D ML classification on the right hemisphere and total cell counts per plane listed in the bottom right.

ArgiNLS is appropriate for genetic labeling experiments of single cells that require concomitant behavioral procedures paired with AAV infection. Future time-course studies of in vivo expression durations longer than studied here (i.e., 3 to 4 wk), will help understand the total range of temporal compatibility with cellular and animal health.

The in vivo use of ArgiNLS highlighted how the selection of genetic labeling strategy can significantly influence whole-brain single-cell segmentations, with an effective doubling of cell counts across subregions located in all major brain structures. The exact nature of this improvement is unclear; however, it is possibly related to depletion of extranuclear FP localization that enhances single-cell SNR. Additionally, since a fraction of wild-type FP molecules misfold or do not form a proper chromophore (76, 77), poly-R driven interaction with nucleolar ribosomal RNA (46) and/or nucleolus-dependent changes of pH could provide a stabilizing influence in aggregation or folding of ArgiNLS-tagged FPs, leading to more single cells detected.

Finally, ArgiNLS can be applied in various volumetric cell labeling applications to resolve cell counts seamlessly and rapidly with 3D ML segmentation. Our proof-of-concept experiments begin to demonstrate ArgiNLS compatibility with, and enhancement of, anatomical input mapping, conditional segmentation of cell types, and mapping a wide range of titrated cell counts within 3D brain volumes using various AAV vectors. However, ArgiNLS may also be incorporated into other molecular genetic tools that will provide improvements to single-cell counting applications including intersectional mapping (24) with or without multiplexing of spectrally separate ArgiNLS-tagged FPs, transsynaptic anatomical tracing (68), activity-dependent reporting (78), and transgenic reporter mouse lines for genomically encoded expression (67). Indeed, while we demonstrate improved single-cell labeling capacity of ArgiNLS over other tags, our in vivo characterization was solely based on AAV expression. The creation of future ArgiNLS-tagged expression vectors for gene delivery from other commonly used viral vectors (79), as well as gene-targeting vectors for the creation of reporter mouse

lines, will help expand optimized single-cell labeling and segmentation among broad genetic delivery options.

In conclusion, the simple yet powerful ArgiNLS genetic tag strategy described here advances streamlined quantification of single cells from small-to-large scale brain image datasets with enriched labeling capacity and enhanced ML capabilities.

Materials and Methods

Mice. Adult male and female mice (8 to 12 wk old) were used for all experiments unless otherwise noted. Animals were group housed on a reverse 12-h light-dark cycle (0700 OFF/1900 ON) and had access to water and food ad libitum. All experimental procedures were performed in accordance with University of Washington IACUC guidelines. Wild-type and transgenic mice lines were purchased from the following vendors: C57Bl6/J (JAX, 000664); CD-1 [Charles River, Crl:CD1(ICR)]; vGlut1-Cre (Slc17a7-IRES2-Cre-D; JAX, 023527); vGat-Cre [Vgat-ires-cre knock-in (C57BL/6J); JAX, 028862]; vGat-Flp (Slc32a1-IRES2-FlpO-D knock-in; JAX, 031331). Descriptions of stereotaxic and retroorbital viral injections, electrophysiology, and behavioral testing are provided in [SI Appendix, Supporting Information Methods](#).

Fluorescent Protein-Expressing AAV Vector Construction. Untagged and NLS-tagged fluorescent proteins were cloned through PCR amplification using standard primers with or without extension overhangs (Integrated DNA Technologies). The monopartite SV40nls coding tag sequence used was from amino acids 126 to 132 of large T antigen (Betapolyomavirus macacae; NCBI Reference sequence: NP_043127.1). Amino acids 156 to 172 of the bipartite mouse nucleoplasmin2 NLS (NCBI accession number: NP_001398932.1) was used for NPM2nls. Isoelectric point (pI) and net charge (z) calculations of each poly-R7 modified or unmodified cNLS protein tag were determined through the open-source Prot pi online tool (<http://protpi.ch>). Starting Addgene plasmids and cloning strategies for construction are described in [SI Appendix, Supporting Information Methods](#).

Characterization of Nuclear Localization In Vitro. Immortalized embryonic mouse hippocampal cells [Cedarlane; mHippoE-1; cat. CLU198] or mouse Neuroblastoma 2A cells (ATCC; CCL-131) were cultured in DMEM (high glucose and L-glutamine; Gibco 11965-092), supplemented with 10% fetal bovine serum (Gibco; 1600) and 1% penicillin-streptomycin (Invitrogen; 15140122) at 37 °C in 5% CO₂. Cells were subcultured following the manufacturer's instructions. For transient DNA transfections, cells were seeded onto Poly-L-lysine-coated glass coverslips (Electron Microscopy Science, 72292-08) in a 24-well plate format. Once cells reached ~80% confluence, transient transfections were performed in duplicate with 250 ng of plasmid DNA per construct using Lipofectamine 3000 transfection reagent (ThermoFisher Scientific, L3000015). After 48 h, cells were fixed with formalin, counterstained with DAPI, and mounted onto glass slides. All NLS-tagged variants were cotransfected with a spectrally separate untagged fluorescent protein to fluorescently label total cellular area in the following combinations: all NLS tagged EGFP constructs with AAV-EF1 α -mKate2; all NLS tagged RFP constructs with p-EGFP-N1 (Addgene; 6085-1). To qualitatively verify Cre-dependent ArgiNLS variant expression, each plasmid was cotransfected with pCAG-Cre (Addgene; 13775), respectively. To qualitatively verify Flp-dependent ArgiNLS-mKate2 expression, we PCR-amplified and subcloned FlpO from PGK-NLS-FlpO-bpA [a gift from Dr. Phillip Soriano (80)] into pTCAV-12a (81) at AgeI/EcoRV restriction sites to create the FlpO-expressing pTCAV-FlpO-12a which was used for cotransfections. Images of all transfection conditions were acquired with a Keyence BZ-X800E inverted fluorescence microscope using a 20x/0.75 NA objective (Nikon). Details of nuclear localization quantification are provided in [SI Appendix, Supporting Information Methods](#).

Adeno-Associated Virus Production.

Php.eB and retroAAV2 packaging plasmid construction. A single packaging plasmid to produce PHP.eB viruses, pDG-PHP.eB, was constructed by subcloning the PHP.eB capsid from pUCmini-iCAP-PHP.eB (Addgene, 103005) into pDG1 plasmid (82) to replace the AAV1 capsid at ClaI/SwaI restriction enzyme sites. The final vector was produced at 1 mg/mL by Nature Technologies (Lincoln, NE).

A single packaging plasmid for retroAAV2 was created by subcloning the retro-AAV2 capsid (83) into the pDG plasmid in a manner identical to pDG-PHP.eB to create pDG_retroAAV2. The pDG-PHP.eB plasmid is publicly available through Addgene. AAV generation and genome copy (GC) titration methods are provided in [SI Appendix, Supporting Information Methods](#).

Characterization of ArgiNLS-EGFP Nuclear Localization In Vivo. The effects of SV40nls and ArgiNLS tags on cellular localization of EGFP fluorescence were quantified in EGFP+/tdTomato+ single cells (n = 5 randomly selected cells/3 vGlut1-Cre or vGat-cre animals) from 40x confocal images in a blinded fashion using ImageJ in a manner similarly described for in vitro quantification. Details regarding histology, imaging, and quantification are provided in [SI Appendix, Supporting Information Methods](#).

ML Classification Performance in 2D Images. Mice RO-injected with AAV.eB-EF1 α -ArgiNLS-EGFP or AAV.eB-EF1 α -SV40nls-EGFP were terminally anesthetized with isoflurane and transcardially perfused and histologically processed as described above. Single-plane epifluorescent images of entire coronal planes at AP 2.11 and -0.52 were acquired in montages with a Keyence BZ-X800E inverted fluorescence microscope using a 10x/0.45 NA objective (Nikon). Consistent DAPI and EGFP channel exposure settings were applied across conditions. Montages were stitched using BZ-X800 Analyzer software (Keyence, v 1.1.1.8) and composite images were created in ImageJ and exported as TIFF images. ML classifiers for ArgiNLS- and SV40nls-EGFP labels were created using random forest classification with QuPath (84) software applied to composite TIF images ([SI Appendix, Supporting Information Methods](#)).

3D Volumetric Characterization of ArgiNLS-EGFP Performance. Two independent approaches were applied to systematically quantify and characterize single-cell labeling across the intact mouse brain with ArgiNLS-tagged FPs: 1) iDISCO+ solvent-based brain clearing and EGFP staining based on methods previously described (9, 13) with modification, and 2) SHIELD aqueous-based brain clearing ([SI Appendix, Supporting Information Methods](#)).

Whole-Brain Image Processing.

Brain-wide SV40nls versus ArgiNLS-tagged EGFP cell count comparisons. Volumetric LSFM-acquired images were processed for whole-brain template registration, automated cell segmentation, and regional segmentation as previously described (9) with modifications ([SI Appendix, Supporting Information Methods](#)).

3D Visualization and ML. Zeiss arivis Pro (formerly, Vision4D) software was used to generate 3D renderings of brain samples throughout the manuscript ([SI Appendix, Supporting Information Methods](#)).

Data, Materials, and Software Availability. Original data for each figure, including whole-brain image stack data has been deposited to figshare (62). Raw cell count from Fig. 5 is included as [Dataset S1](#).

ACKNOWLEDGMENTS. We would like to acknowledge support from the University of Washington Center of Excellence in Addiction Pain and Emotion Molecular Genetics Resource Core (P30DA048736). We would like to thank Valerie Tsai, Aida Chan, Axel Salazar, and Kevin Bai for their technical support. We thank the entire Golden and Heshmati Laboratories for their discussions and critiques of the manuscript.

Author affiliations: ^aCenter of Excellence in Neurobiology of Addiction, Pain, and Emotion, University of Washington, Seattle, WA 98195; ^bDepartment of Biological Structure, University of Washington, Seattle, WA 98195; ^cGraduate Program in Neuroscience, University of Washington, Seattle, WA 98195; ^dDepartment of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA 98195; ^eUndergraduate Program in Biochemistry, University of Washington, Seattle, WA 98195; ^fDepartment of Pharmacology, University of Washington, Seattle, WA 98195; ^gCarl Zeiss Microscopy LLC, White Plains, NY 10601; and ^hDepartment of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195

Author contributions: E.R.S., L.S.Z., and S.A.G. designed research; E.R.S., J.S.N., A.D.M., Y.Z., K.S.G., L.K., M.M.C., M.X.B., B.B., N.L.G., and B.J. performed research; E.R.S., M.B., L.S.Z., and S.A.G. contributed new reagents/analytic tools; E.R.S., J.S.N., K.S.G., and B.J. analyzed data; and E.R.S. and S.A.G. wrote the paper.

1. S. Corsetti, F. Gunn-Moore, K. Dholakia, Light sheet fluorescence microscopy for neuroscience. *J. Neurosci. Methods* **319**, 16–27 (2019).
2. T. Ragan *et al.*, Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat. Methods* **9**, 255–258 (2012).
3. A. Li *et al.*, Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain. *Science* **330**, 1404–1408 (2010).
4. Y. Kim *et al.*, Whole-brain mapping of neuronal activity in the learned helplessness model of depression. *Front. Neural. Circuits* **10**, 3 (2016).
5. Y. Kim *et al.*, Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Rep.* **10**, 292–305 (2015).
6. H. Azevedo, M. Ferreira, A. Mascarello, P. Osten, C. R. W. Guimarães, Brain-wide mapping of c-fos expression in the single prolonged stress model and the effects of pretreatment with ACH-000029 or prazosin. *Neurobiol. Stress* **13**, 100226 (2020).
7. G. Cai *et al.*, Brain-wide mapping of c-Fos expression with fluorescence micro-optical sectioning tomography in a chronic sleep deprivation mouse model. *Neurobiol. Stress* **20**, 100478 (2022).
8. A. V. Roland *et al.*, Alcohol dependence modifies brain networks activated during withdrawal and reaccess: A c-Fos-based analysis in mice. *Biol. Psychiatry* **94**, 393–404 (2023), 10.1016/j.biopsych.2023.01.018.
9. R. Madangopal *et al.*, Incubation of palatable food craving is associated with brain-wide neuronal activation in mice. *bioRxiv* [Preprint] (2022). <https://www.biorxiv.org/content/10.1101/2022.05.31.494210v1> (Accessed 1 September 2022).
10. L. A. DeNardo *et al.*, Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nat. Neurosci.* **22**, 460–469 (2019).
11. D. R. Rijksket *et al.*, UNRAVELing the synergistic effects of psilocybin and environment on brain-wide immediate early gene expression in mice. *Neuropsychopharmacology* **48**, 1798–1807 (2023), 10.1038/s41386-023-01613-4.
12. H. H. Hansen *et al.*, Whole-brain activation signatures of weight-lowering drugs. *Mol. Metab.* **47**, 101171 (2021).
13. N. Renier, Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* **133**, 1789–1802 (2016).
14. E. A. Susaki *et al.*, Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. *Cell* **157**, 726–739 (2014).
15. H. Yamaguchi *et al.*, Brain-wide mapping of neuronal architecture controlling torpor. *bioRxiv* [Preprint] (2023). <https://www.biorxiv.org/content/10.1101/2023.03.03.531064v1> (Accessed 15 June 2023).
16. D. S. Roy *et al.*, Brain-wide mapping reveals that engrams for a single memory are distributed across multiple brain regions. *Nat. Commun.* **13**, 1799 (2022).
17. Y. Kim *et al.*, Brain-wide maps reveal stereotyped cell type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171**, 456–469.e22 (2017).
18. S. Son *et al.*, Whole-brain wiring diagram of oxytocin system in adult mice. *J. Neurosci.* **42**, 5021–5033 (2022).
19. M. Habart, G. Lio, A. Soumier, C. Demily, A. Sirigu, An optimized iDISCO+ protocol for tissue clearing and 3D analysis of oxytocin and vasopressin cell network in the developing mouse brain. *STAR Protoc.* **4**, 101968 (2023).
20. K. S. Matho *et al.*, Genetic dissection of the glutamatergic neuron system in cerebral cortex. *Nature* **598**, 182–187 (2021).
21. E. R. Szelenyi *et al.*, Distributed X chromosome inactivation in brain circuitry is associated with X-linked disease penetrance of behavior. *Cell Rep.* **43**, 114068 (2024).
22. Y. W. Wang, Y. Hu, J. T. Qi, Z. Zhang, M. Q. Luo, Comparative whole-brain mapping of isoflurane and ketamine-activated nuclei and functional networks. *bioRxiv* [Preprint] (2023). <https://www.biorxiv.org/content/10.1101/2023.06.03.543576v1> (Accessed 15 June 2023).
23. E. H. K. Stelzer *et al.*, Light sheet fluorescence microscopy. *Nat. Rev. Methods Primers* **1**, 1–25 (2021).
24. K. M. Dean *et al.*, Isotropic imaging across spatial scales with axially swept light-sheet microscopy. *Nat. Protoc.* **17**, 2025–2053 (2022).
25. F. Claudi *et al.*, Visualizing anatomically registered data with brainrender. *eLife* **10**, e65751 (2021).
26. C. J. Nieuwrook *et al.*, aMAP is a validated pipeline for registration and segmentation of high-resolution mouse brain data. *Nat. Commun.* **7**, 11879 (2016).
27. A. L. Tyson *et al.*, A deep learning algorithm for 3D cell detection in whole mouse brain image datasets. *PLoS Comput. Biol.* **17**, e1009074 (2021).
28. A. L. Tyson *et al.*, Accurate determination of marker location within whole-brain microscopy images. *Sci. Rep.* **12**, 867 (2022).
29. M. Negwer *et al.*, FriendlyClearMap: An optimized toolkit for mouse brain mapping and analysis. *Gigascience* **12**, giad035 (2022).
30. D. Choquet, M. Sainlos, J.-B. Sibarita, Advanced imaging and labelling methods to decipher brain cell organization and function. *Nat. Rev. Neurosci.* **22**, 237–255 (2021).
31. R. E. Lanford, J. S. Butel, Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell* **37**, 801–813 (1984).
32. D. Goertsen *et al.*, AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nat. Neurosci.* **25**, 106–115 (2022).
33. M. J. Jang *et al.*, Spatial transcriptomics for profiling the tropism of viral vectors in tissues. *Nat. Biotechnol.* **41**, 1272–1286 (2023), 10.1038/s41587-022-01648-w.
34. K. Matsumoto *et al.*, Advanced CUBIC tissue clearing for whole-organ cell profiling. *Nat. Protoc.* **14**, 3506–3537 (2019).
35. S. Ito *et al.*, Loss of neuronal 3D chromatin organization causes transcriptional and behavioural deficits related to serotonergic dysfunction. *Nat. Commun.* **5**, 4450 (2014).
36. M. Bernhofer *et al.*, NLSdb—Major update for database of nuclear localization signals and nuclear export signals. *Nucleic Acids Res.* **46**, D503–D508 (2018).
37. K. Luk *et al.*, Optimization of nuclear localization signal composition improves CRISPR-Cas12a editing rates in human primary cells. *GEN Biotechnol.* **1**, 271–284 (2022).
38. T. Thumberger *et al.*, Boosting targeted genome editing using the hei-tag. *eLife* **11**, e70558 (2022).
39. B. Cautain, R. Hill, N. de Pedro, W. Link, Components and regulation of nuclear transport processes. *FEBS J.* **282**, 445–462 (2015).
40. B. Friedrich, C. Quensel, T. Sommer, E. Hartmann, M. Köhler, Nuclear localization signal and protein context both mediate importin alpha specificity of nuclear import substrates. *Mol. Cell Biol.* **26**, 8697–8709 (2006).
41. C. Y. Xiao, P. Jans, D. A. Jans, Negative charge at the protein kinase CK2 site enhances recognition of the SV40 large T-antigen NLS by importin: Effect of conformation. *FEBS Lett.* **440**, 297–301 (1998).
42. F. Bartnicki, P. Bonarek, E. Kowalska, W. Strzalka, The Argi system: One-step purification of proteins tagged with arginine-rich cell-penetrating peptides. *Sci. Rep.* **7**, 2619 (2017).
43. S. M. Fuchs, R. T. Raines, Polyarginine as a multifunctional fusion tag. *Protein Sci.* **14**, 1538–1544 (2005).
44. S. J. Brewer, H. M. Sassenfeld, The purification of recombinant proteins using C-terminal polyarginine fusions. *Trends Biotechnol.* **3**, 119–122 (1985).
45. K. Hristova, W. C. Wimley, A look at arginine in membranes. *J. Membr. Biol.* **239**, 49–56 (2011).
46. R. M. Martin *et al.*, Principles of protein targeting to the nucleolus. *Nucleus* **6**, 314–325 (2015).
47. N. Schmidt, A. Mishra, G. H. Lai, G. C. L. Wong, Arginine-rich cell-penetrating peptides. *FEBS Lett.* **584**, 1806–1813 (2010).
48. M. Imanishi, W. Yan, T. Morisaki, Y. Sugiura, An artificial six-zinc finger peptide with polyarginine linker: Selective binding to the discontinuous DNA sequences. *Biochem. Biophys. Res. Commun.* **333**, 167–173 (2005).
49. N. Freitas, C. Cunha, Mechanisms and signals for the nuclear import of proteins. *Curr. Genomics* **10**, 550–557 (2009).
50. N. Ishii *et al.*, Analysis of a nuclear localization signal of simian virus 40 major capsid protein Vp1. *J. Virol.* **70**, 1317 (1996).
51. C. Dingwall, S. V. Sharnick, R. A. Laskey, A polypeptide domain that specifies migration of nucleoplasm into the nucleus. *Cell* **30**, 449–458 (1982).
52. L. E. Fenno *et al.*, Comprehensive dual- and triple-feature intersectional single-vector delivery of diverse functional payloads to cells of behaving mammals. *Neuron* **107**, 836–853.e11 (2020).
53. D. Shcherbo *et al.*, Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* **418**, 567–574 (2009).
54. D. M. Shcherbakova *et al.*, Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat. Commun.* **7**, 12405 (2016).
55. K. Hosokawa, M. Nishi, H. Sakamoto, Y. Tanaka, M. Kawata, Regional distribution of importin subtype mRNA expression in the nervous system: Study of early postnatal and adult mouse. *Neuroscience* **157**, 864–877 (2008).
56. Z. J. Huang, A. Paul, The diversity of GABAergic neurons and neural communication elements. *Nat. Rev. Neurosci.* **20**, 563–572 (2019).
57. I. M. Zakiewicz, J. G. Bjaalie, T. B. Leergaard, Brain-wide map of efferent projections from rat barrel cortex. *Front. Neuroinf.* **8**, 5 (2014).
58. T. Kanda, K. F. Sullivan, G. M. Wahl, Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **8**, 377–385 (1998).
59. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: A generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
60. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation. *arXiv* [Preprint] (2015). <https://arxiv.org/abs/1505.04597v1> (Accessed 10 October 2023).
61. S. Berg *et al.*, ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
62. E. R. Szelenyi, S. A. Golden, Data from "An arginine-rich nuclear localization signal (ArgiNLS) strategy for streamlined image segmentation of single cells." Figshare. <https://figshare.com/s/9ffdc47e42a8aac6e>. Deposited 8 July 2024.
63. U. Chon, D. J. Vanselow, K. C. Cheng, Y. Kim, Enhanced and unified anatomical labeling for a common mouse brain atlas. *Nat. Commun.* **10**, 5067 (2019).
64. M. J. Ortuño *et al.*, Melanocortin 4 receptor stimulation prevents antidepressant-associated weight gain in mice caused by long-term fluoxetine exposure. *J. Clin. Invest.* **131**, e151976 (2021).
65. M. Schneberger *et al.*, Regulation of energy expenditure by brainstem GABA neurons. *Cell* **178**, 672–685.e12 (2019).
66. A. Ilanges *et al.*, Brainstem ADCYAP1+ neurons control multiple aspects of sickness behaviour. *Nature* **609**, 761–771 (2022).
67. S. Navabpour, J. L. Kwapis, T. J. Jarome, A neuroscientist's guide to transgenic mice and other genetic tools. *Neurosci. Biobehav. Rev.* **108**, 732–748 (2020).
68. D. L. Haggerty, G. G. Grecco, K. C. Reeves, B. Atwood, Adeno-associated viral vectors in neuroscience research. *Mol. Therapy: Methods Clin. Dev.* **17**, 69–82 (2020).
69. G. G. Lambert *et al.*, Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing. *PLoS Biol.* **18**, e3000936 (2020).
70. B. C. Campbell *et al.*, mGreenLantern: A bright monomeric fluorescent protein with rapid expression and cell filling properties for neuronal imaging. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30710–30721 (2020).
71. R. S. Cox, M. J. Dunlop, M. B. Elowitz, A synthetic three-color scaffold for monitoring genetic regulation and noise. *J. Biol. Eng.* **4**, 10 (2010).
72. B. L. Roth, DREADDs for neuroscientists. *Neuron* **89**, 683–694 (2016).
73. Y.-G. Park *et al.*, Protection of tissue physicochemical properties using polyfunctional crosslinkers. *Nat. Biotechnol.* **37**, 73–83 (2019).
74. A. Adhikari *et al.*, Basomedial amygdala mediates top-down control of anxiety and fear. *Nature* **527**, 179–185 (2015).
75. P. Hu, X. Zhao, Q. Zhang, W. Li, Y. Zu, Comparison of various nuclear localization signal-fused Cas9 proteins and Cas9 mRNA for genome editing in zebrafish. *G3 (Bethesda)* **8**, 823–831 (2018).
76. O. V. Stepanenko, O. V. Stepanenko, I. M. Kuznetsova, V. V. Verkhusha, K. K. Turoverov, "Chapter Four: Beta-barrel scaffold of fluorescent proteins: Folding, stability and role in chromophore formation" in *International Review of Cell and Molecular Biology*, K. W. Jeon, Ed. (Academic Press, 2013), pp. 221–278.
77. B. P. Cormack, R. H. Valdivia, S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
78. L. DeNardo, L. Luo, Genetic strategies to access activated neurons. *Curr. Opin. Neurobiol.* **45**, 121–129 (2017).
79. S. Ghosh, A. M. Brown, C. Jenkins, K. Campbell, Viral vector systems for gene therapy: A comprehensive literature review of progress and biosafety challenges. *Appl. Biosaf.* **25**, 7–18 (2020).
80. C. S. Raymond, P. Soriano, High-efficiency FLP and ΦC31 site-specific recombination in mammalian cells. *PLoS ONE* **2**, e162 (2007).
81. E. J. Kremer, S. Boutin, M. Chillon, O. Danos, Canine adenovirus vectors: An alternative for adenovirus-mediated gene transfer. *J. Virol.* **74**, 505–512 (2000).
82. D. Grimm, A. Kern, K. Rittner, J. A. Kleinschmidt, Novel tools for production and purification of recombinant adenoassociated virus vectors. *Hum. Gene Ther.* **9**, 2745–2760 (1998).
83. D. G. R. Tervo *et al.*, A designer AAV variant permits efficient retrograde access to projection neurons. *Neuron* **92**, 372–382 (2016).
84. P. Bankhead *et al.*, QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).