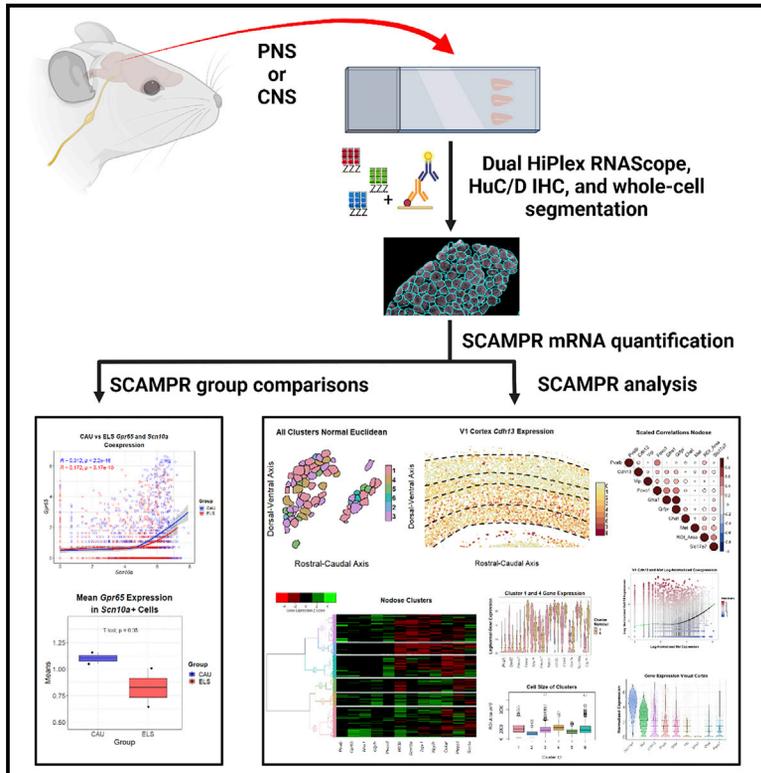Article

SCAMPR, a single-cell automated multiplex pipeline for RNA quantification and spatial mapping

Graphical abstract

Authors

Ramin Ali Marandi Ghoddousi,
Valerie M. Magalong,
Anna K. Kamitakahara, Pat Levitt

Correspondence

plevitt@med.usc.edu

In brief

The widespread use of smFISH methods will benefit greatly from accessible quantification and analysis tools. Ali Marandi Ghoddousi et al. present SCAMPR, an easy-to-use, open-source pipeline for accurately quantifying smFISH signal at a whole-cell level. SCAMPR facilitates topological gene expression analyses and identification of gene expression differences between experimental groups.

Highlights

- SCAMPR is a pipeline for quantification and spatial analysis of gene expression

- Dual HiPlex-IHC allows for identification of neuronal boundaries in the PNS and CNS

- Gene expression differences between groups are identified by ImageJ/R code

- SCAMPR is rapid, easy to use, and cost effective

CelPress

Cell Reports Methods

Article

SCAMPR, a single-cell automated multiplex pipeline for RNA quantification and spatial mapping

Ramin Ali Marandi Ghoddousi,[1,2] Valerie M. Magalong,[1] Anna K. Kamitakahara,[1,2,3] and Pat Levitt[1,2,3,4,*]
[1]Children's Hospital Los Angeles, Los Angeles, CA 90027, USA
[2]University of Southern California, Los Angeles, CA 90007, USA
[3]Keck School of Medicine, Los Angeles, CA 90033, USA
[4]Lead contact
*Correspondence: plevitt@med.usc.edu
https://doi.org/10.1016/j.crmeth.2022.100316

MOTIVATION   Quantitative analysis of spatial mRNA expression in neurons presents challenges in terms of accuracy and being computationally and time intensive. Existing methods that rely on nuclear labeling (DAPI) to distinguish adjoining cells lack the precision to detect mRNA expression in the cytoplasm. In addition, quantification methods that rely on puncta counts can generate large, variable datasets that potentially undercount highly expressed mRNAs. To overcome these methodological barriers, we developed the SCAMPR pipeline that allows for fast, accurate segmentation of neuronal cell body boundaries, topographic gene expression mapping, and high-dimensional quantification and analysis of mRNA expression in tissue sections.

SUMMARY

Spatial gene expression, achieved classically through *in situ* hybridization, is a fundamental tool for topographic phenotyping of cell types in the nervous system. Newly developed techniques allow for visualization of multiple mRNAs at single-cell resolution and greatly expand the ability to link gene expression to tissue topography, yet there are challenges in efficient quantification and analysis of these high-dimensional datasets. We have therefore developed the single-cell automated multiplex pipeline for RNA (SCAMPR), facilitating rapid and accurate segmentation of neuronal cell bodies using a dual immunohistochemistry-RNAscope protocol and quantification of low- and high-abundance mRNA signals using open-source image processing and automated segmentation tools. Proof of principle using SCAMPR focused on spatial mapping of gene expression by peripheral (vagal nodose) and central (visual cortex) neurons. The analytical effectiveness of SCAMPR is demonstrated by identifying the impact of early life stress on gene expression in vagal neuron subtypes.

INTRODUCTION

The murine brain contains approximately 70 million individual neurons (Herculano-Houzel et al., 2006). Understanding a tissue of such complexity requires knowledge about its component parts, both at the level of single neurons and at the level of cell groups that constitute structural and functional nodes in circuits. Analysis of the individual characteristics of single neurons, such as their topographical organization, morphology, and molecular signature (among other properties), can be used to categorize them into distinct functional cell groups (Zeng and Sanes, 2017; Erö et al., 2018). The commercialization and ongoing development of single-molecule fluorescent *in situ* hybridization (smFISH) techniques (e.g., RNAscope, HCR, RNA-FISH, MERFISH, etc.), supplemented by a variety of image analysis tools and software that quantitate molecular signal with cellular

specificity, allow for multimodal comparison of transcriptomics and histology at a single-cell level (Wang et al., 2012; Chen et al., 2015; Ståhl et al., 2016; Choi et al., 2018; Eng et al., 2019; Goh et al., 2020). Existing methods for quantitative analysis of smFISH face several challenges, including incomplete single-cell segmentation and a loss of cell size information due to the use of the nuclear marker DAPI to delineate cell boundaries, limited compatibility with multiround versions of smFISH that are required for analysis of more than 4 genes, and resource-/time-intensive pipelines (Carine Stapel et al., 2016; Codeluppi et al., 2018; Dries et al., 2019; Maynard et al., 2020; Wang et al., 2021). Here, we introduce SCAMPR, a quantitative analysis pipeline that combines commercialized smFISH (HiPlex RNAscope) with immunohistochemistry (IHC), and employs existing and newly developed semi-automated image processing tools for accurate segmentation and quantification of mRNA
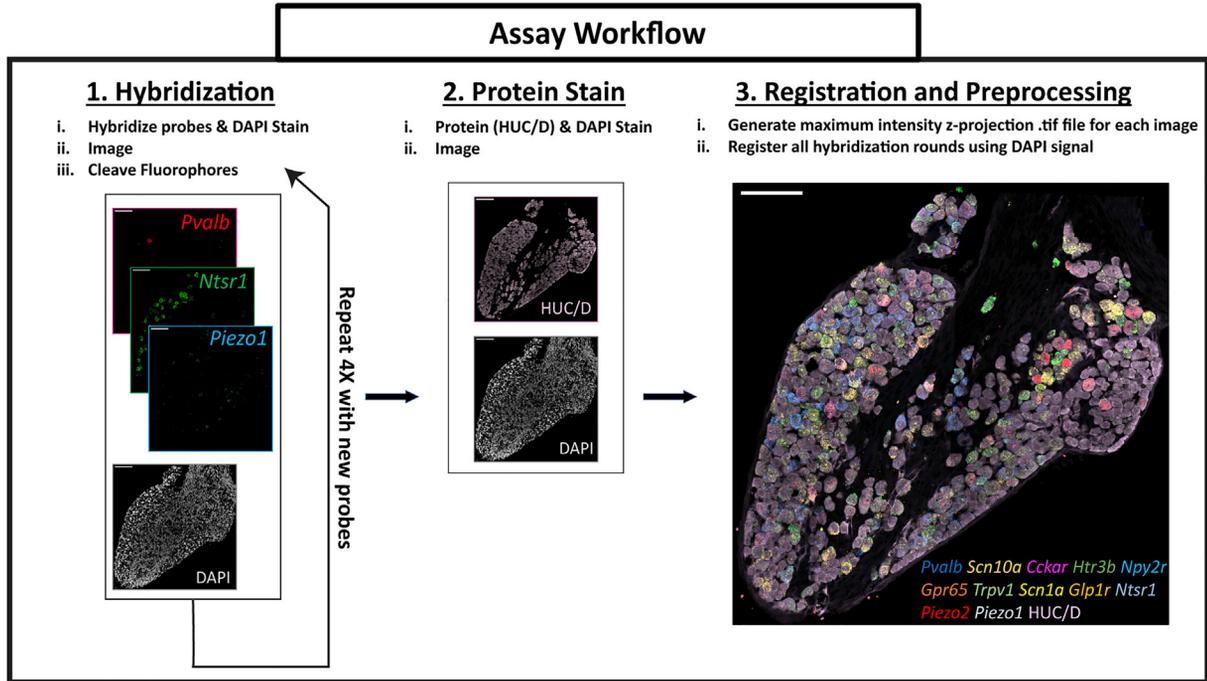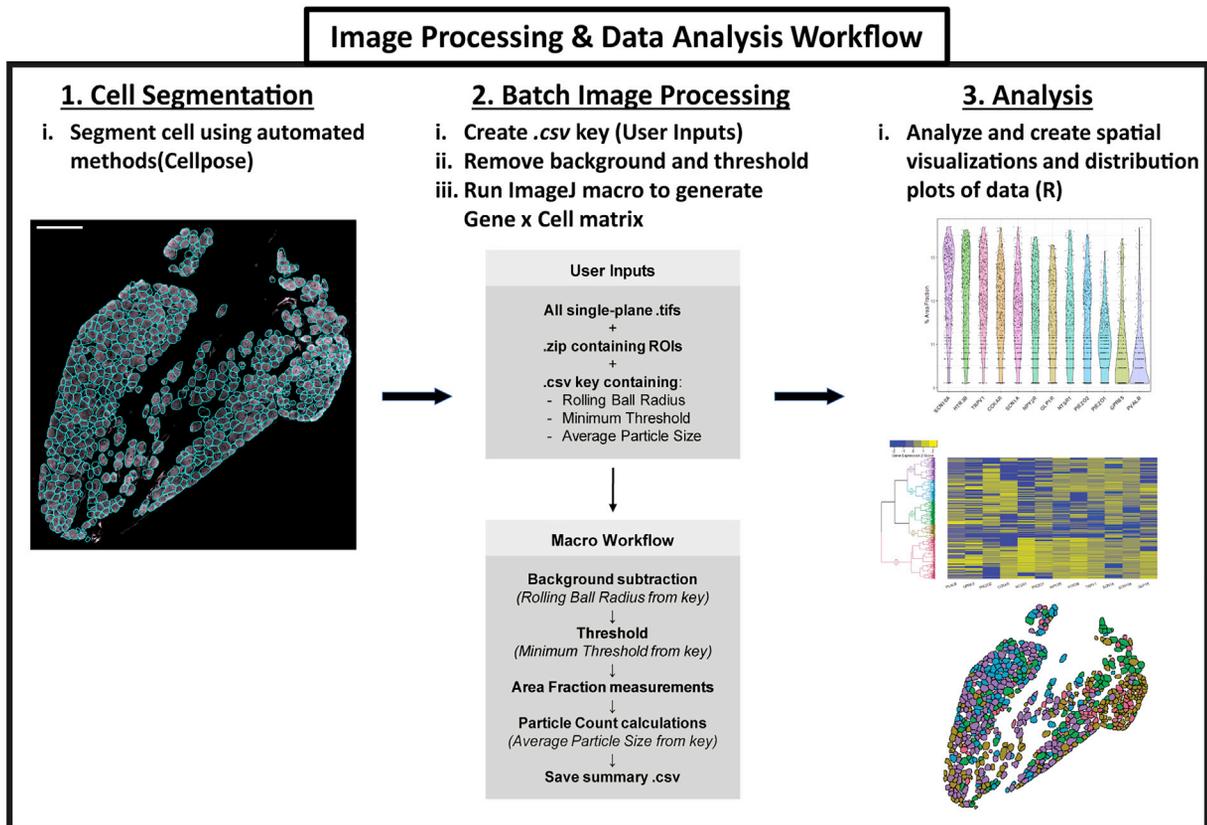
expression in neurons (Wang et al., 2012). SCAMPR includes open-source code for quantitative analysis and spatial mapping of mRNA expression in neurons, as well as precise comparison of gene expression between distinct experimental groups. In addition, neuron size is accurately preserved due to whole-soma segmentation, and SCAMPR generates datasets that require relatively small amounts of digital storage and processing power for downstream analysis. Lastly, the entirety of the SCAMPR pipeline, from the benchwork to the generation of quantitative plots, takes 5 days to complete when multiplexing 12 genes, making it possible to perform comparative, multi-modal gene expression experiments in a short amount of time.

We validate the compatibility of SCAMPR on two mouse nervous system structures that have distinct neuronal packing densities and topographical organizations: the jugular-nodose complex of the peripheral nervous system, hereto referred to as the nodose ganglion (NG) for simplicity, and the primary visual cortex (V1) of the central nervous system. In V1, we demonstrate that SCAMPR can be used to correlate single and multidimensional gene expression to cortical layer topography and to distinguish cell types based on soma size, gene expression, and location. We use SCAMPR in the NG to demonstrate cell-type-specific gene expression changes resulting from early life stress (ELS). Our findings demonstrate the accuracy and utility of SCAMPR for descriptive and comparative analysis of neuronal gene expression and topography.

RESULTS

Overview of SCAMPR

A comprehensive snapshot of the SCAMPR pipeline is provided, outlining the steps involved in processing and analyzing images from a single tissue section (Figure 1). SCAMPR has two work-flows: (1) an assay workflow (Figure 1A), which comprises benchtop *in situ* hybridization, immunostaining, microscopy, and image pre-processing, and (2) an image processing and data analysis workflow (Figure 1B), which includes cell segmentation, additional batch image processing in FIJI/ImageJ, mRNA quantification, and data analysis (Schindelin et al., 2012; Schneider et al., 2012). The SCAMPR assay workflow (Figure 1A) starts with the published protocol for the Advanced Cell Diagnostics (ACD) HiPlex RNAscope assay (Wang et al., 2012). Briefly, probes targeting three distinct mRNAs are hybridized to the tissue, each with a different fluorescent signal, a DAPI counterstain is performed, the tissue is imaged, the fluorophores are cleaved, and this process is repeated with the hybridization of three new mRNA probes. This is repeated in the same tissue until all genes of interest have been hybridized and imaged. If infrared confocal imaging is available, four probes can be hybridized simultaneously. After all probe signals have been imaged, DAPI counterstaining and HuC/D IHC are performed to mark

the boundaries of cellular nuclei and the neuronal cytoplasm, respectively, and the tissue is imaged one final time. Flattened, maximum intensity projection images from the different hybridization/staining rounds are spatially registered (overlayed) using the DAPI signal with the ACD RNAscope HiPlex Image Registration Software or in ImageJ. The resulting registered images then proceed through the SCAMPR image processing and data analysis workflow (Figure 1B).

In this second workflow (Figure 1B), cell segmentation and batch image processing are performed. The HuC/D image is used to segment the cells into individual regions of interest (ROIs). These ROIs may be achieved by manual drawing in FIJI/ImageJ or, to save time, by using an open-source automated cell segmentation program such as Cellpose (Stringer et al., 2021). In parallel, the images containing the mRNA signal are visually inspected by the user in ImageJ to determine the optimal parameters for applying a threshold to convert the mRNA signal into a binary signal/no-signal image. To add flexibility to this process, an alternative, semiautomated method was designed to determine the thresholding parameters (Figure S1). The ROIs are then used in combination with these parameters to quantify signal for each mRNA probe on a single-cell level in each tissue section. New macros were written to automate this process in ImageJ. These macros produce a "gene by count" matrix in the form of a CSV file, which is used for subsequent data analyses. The data analysis component of the SCAMPR pipeline contains numerous methods to quantify gene expression, co-expression, and global expression patterns, as well as methods for spatial mapping and comparative gene expression analysis between groups. Each analysis method has its own R script that can be accessed on GitHub.

Dual IHC-RNAscope accurately demarcates cell boundaries in peripheral and central nervous system tissues

HuC and HuD are mRNA-binding proteins that are involved in transcript regulation and are selectively expressed in neurons starting in embryonic development, initially appearing in progenitor cells, and continuing to be expressed in postmitotic neurons into adulthood (Okano and Darnell, 1997). HuC and HuD are present at the protein level in both the nucleus and cytosol (Diaz-Garcia et al., 2021). The whole-cell expression, neuron specificity, and broad expression time course of HuC/D make them excellent cell boundary markers for quantitative studies that require neuronal segmentation. Two experiments employing SCAMPR demonstrate that HuC/D IHC is compatible in multiple nervous system tissues at different developmental time points. First, multiround HiPlex RNAscope is successfully applied in postnatal day 9 (P9) NG of the peripheral nervous system as well as in adult V1 in the central nervous system (Figures 2A, 2B, 2E, and 2F). Subsequent HuC/D immunostaining is

---

Figure 1. Schematic of SCAMPR pipeline

(A) Assay workflow. (1) All rounds of HiPlex RNAscope are performed and sections are imaged. (2) HuC/D IHC is performed. (3) Images from all rounds are pre-processed and registered using the DAPI signal as the anchor.

(B) Image processing and data analysis workflow. (1) Batch cell boundary segmentation using Cellpose. (2) Images are inspected by the user, then processed using macros to automate gene expression quantification for each cell in ImageJ. (3) Gene expression (top) and co-expression, clustering (middle), and spatial mapping (bottom) are performed in R. Scale bars: 100 μm.
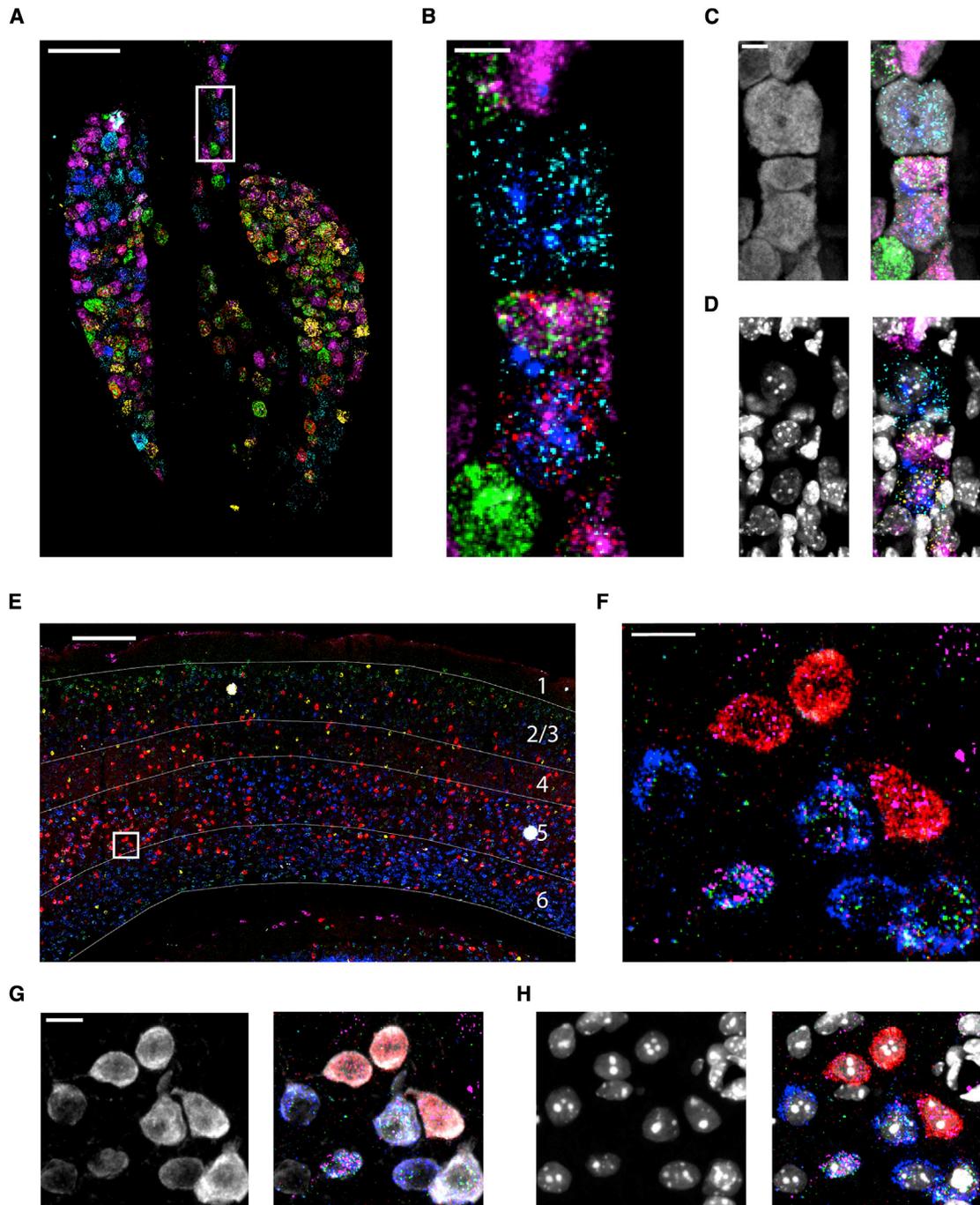
Figure 2. Combined IHC and HiPlex RNAscope allows for high-resolution, single-cell mRNA expression in the central and peripheral nervous system

(A) HiPlex RNAscope visualized in postnatal day 9 nodose ganglion. Scale bar: 100 μm.

(B) High-magnification inset from (A). Scale bar: 15 μm.

(C and D) High-magnification insets from (A) co-visualized with HuC/D (C) and DAPI (D). Scale bar: 10 μm.

(E) HiPlex RNAscope visualized in adult mouse visual cortex. Scale bar: 100 μm.

(F) High-magnification inset from (E). Scale bar: 5 μm.

(G and H) High-magnification inset from (E) co-visualized with HuC/D (G) and DAPI (H). Scale bar: 5 μm.

compatible with HiPlex RNAscope in these same tissues and accurately labels neuronal cell bodies following multiround Hi-Plex RNAscope (Figures 2C and G). Comparison of HuC/D labeling with nuclear DAPI labeling demonstrates that HuC/D more precisely captures the complete individual cellular profile than the nuclear DAPI signal, resulting in accurate localization of single-cell RNAscope signals in both tissues (Figures 2C, 2D, 2G, and 2H).

Additional validation of RNAscope-HuC/D co-staining was performed in the embryonic NG and early postnatal brainstem (data not shown). In all instances, an intense HuC/D signal allowed for distinct segmentation between individual neurons. There is minimal immunolabeling of neuronal processes, further enhancing the ability to distinguish between closely apposed neuronal somata.

Automated methods for accurate segmentation of single neurons

Quantitative analysis of spatial transcriptomics data requires segmentation of cells into single, discrete units. The most accurate segmentation of cells in a tissue section will be obtained with more precise cell boundary markers that demarcate the cytoplasm rather than using a single organelle label such as the nucleus. DAPI marks cell nuclei, and because mRNA is localized to both nuclear and cytosolic cell compartments, segmentation based on the nucleus can result in mRNA fluorescent signal that is unassigned to a cell profile and therefore not measured. Using expanded nuclear labeling (cytoDAPI) allows for more accurate segmentation of whole neurons but makes it difficult to separate neurons from non-neurons in densely packed tissues with large amounts of non-neuronal cell types (Wang et al., 2021). Using the localization of two genes, manual segmentation was performed to demonstrate potential differences in mRNA signal detection between nuclear segmentation using DAPI signal and whole-cell segmentation using HuC/D staining after HiPlex RNAscope in a sample section of the V1 cortex (Figure 3A). The HuC/D signal allows for segmentation of the cell body and more accurately encircles the entirety of the mRNA signals (here *Pvalb* and *Chd13*) in the segmented cells compared with the DAPI signal (Figure 3B). Quantification of gene expression after DAPI segmentation and HuC/D segmentation on a subset of images further demonstrates that the HuC/D segmentation captures more of the RNA signal and therefore higher average gene expression across most of the genes assayed (Figure S2A). The size of this increase is not consistent across all genes. Thus, some genes such as *Cckar* are affected more than others by the incomplete capture of RNA signal after DAPI segmentation (Figure S2A). These gene-specific differences are likely due to the predominant cytosolic localization of certain transcripts such as *Cckar* (Figures S2B and S2C).

Furthermore, using HuC/D staining for segmentation facilitates selective capture of signal in neurons, as both the V1 and NG contain large numbers of non-neuronal cells. Segmentation of a portion of the V1 cortex using the nuclear DAPI signal results in a 28% increase in segmented cells (1,213 HuC/D versus 1,547 DAPI). This difference is much larger in the NG, where there was a 209% increase in segmented cells (398 HuC/D versus 1,230 DAPI), likely due to the large number of non-neuronal satellite cells.

The time requirement for manual segmentation increases dramatically as the surface area of the tissue of interest and the cell packing density increases. Popular automated segmentation methods such as watershed segmentation greatly reduce this time requirement but perform poorly on images of densely packed cell bodies, a particular challenge in most developing and certain adult tissues. More recent methods such as StarDist are trained on specific image sets (DAPI signal) and require further user-guided training to accurately define whole-cell boundaries that are more closely apposed (Weigert et al., 2020). Cellpose, a generalized, deep-learning, automated segmentation algorithm that has been trained on multiple different image sets, obviates these challenges (Stringer et al., 2021). Application of Cellpose on HuC/D immunolabeling signal produces highly comparable results to manual segmentation in the same tissue sections. Small deviations (missed cells, incorrect boundaries, etc.) can be corrected manually after Cellpose segmentation (Corrected Cellpose) (Figure 3C). On a per-cell basis, data from two different operators demonstrate that there is an approximately 20-fold difference in segmentation time between using Cellpose and manual segmentation, with Cellpose segmentation requiring 0.25 s per cell and manual segmentation requiring 4.77 s per cell. Corrected Cellpose takes ~6 times longer than Cellpose alone but is also ~3 times faster than manual segmentation (Figure 3D). The time-savings estimates from limited ROIs are likely an undervaluation, as the rate of Cellpose segmentation scales up in a non-linear fashion with an increase in number of analyzed cells. Thus, Cellpose segmentation on ~6,500 cells using the entire V1 cortex tissue section results in a time reduction from 0.25 to 0.085 s per cell. This translates into Cellpose segmentation being approximately 56 times faster than manual segmentation. The calculation of an F1 score, representing the accuracy of Cellpose and Corrected Cellpose segmentation compared with manual segmentation, demonstrates the precision of this segmentation method (see STAR Methods for details). To generate the F1 score, ROIs that showed a ≥70% overlap with the manually generated ROIs were considered true positives (Caicedo et al., 2019). This comparison resulted in an F1 score of 0.76 for Cellpose and 0.79 for Corrected Cellpose (Figures 3D and 3E). To better understand whether this accuracy affected the overall dynamics of segmentation in each image, the number of segmented cells and the distribution of ROI areas were compared between each method. Cellpose and Corrected Cellpose resulted in only 1.6%–3.0% fewer ROIs than manual segmentation, and the distribution of the ROI areas of the neurons between the three methods were nearly identical, providing further evidence for the overall accuracy of the fully automated segmentation method compared with the manual method (Figure 3F). These data demonstrate that Cellpose segmentation alone can be employed for both high segmentation accuracy and substantial time savings.

Quantification and spatial projection of pairwise mRNA expression patterns

One of the main advantages of SCAMPR is the ability to accurately quantify a multiplexed fluorescent *in situ* signal on a single-cell level while preserving the topographic location of each cell in the tissue. The pipeline provides an opportunity to
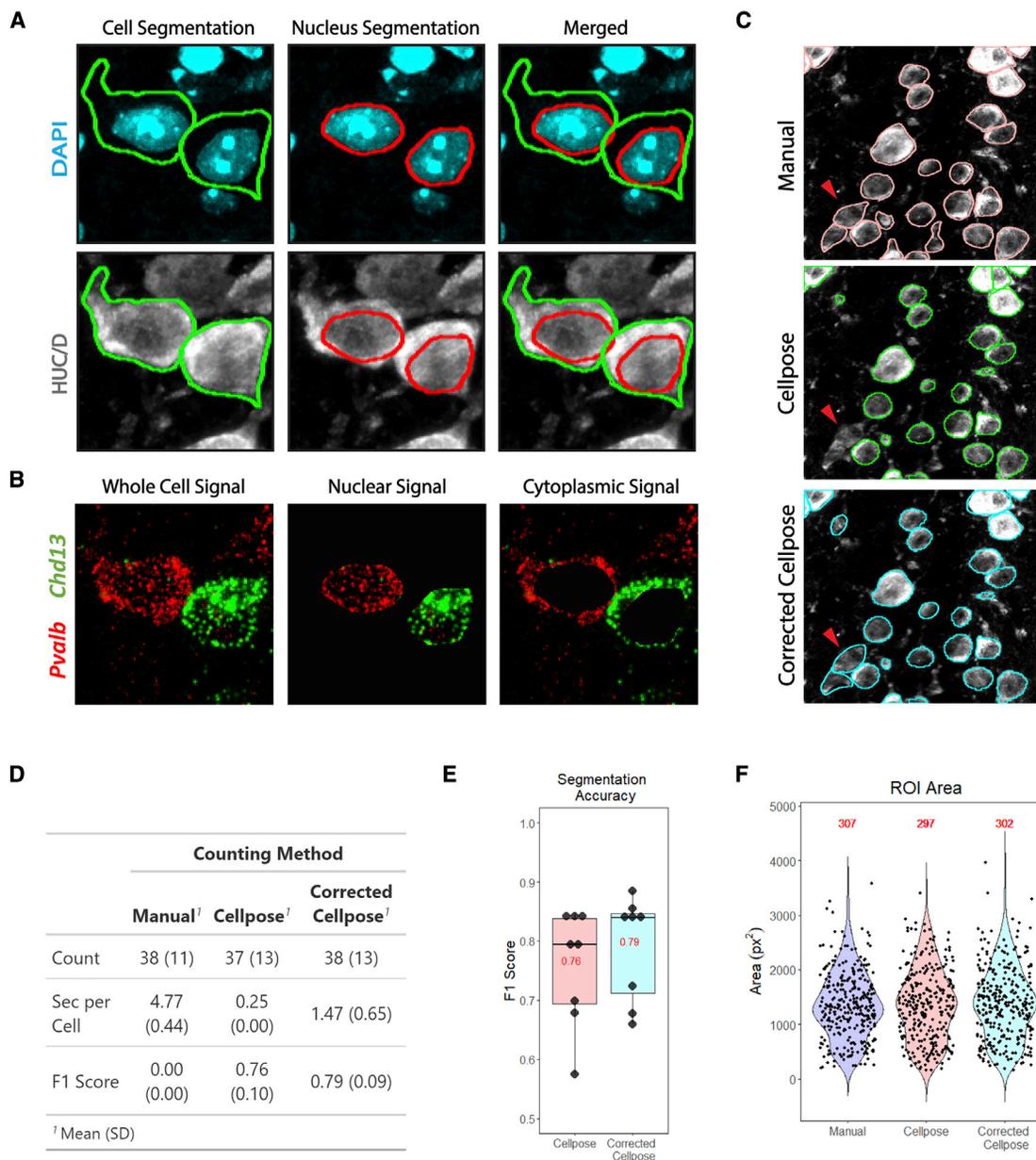
Figure 3. Automated segmentation of cytoplasm of single cells using Cellpose

(A) Differences in nuclear segmentation boundaries (red outline) and whole-soma segmentation boundaries (green outline) in V1 cortex.

(B) Nuclear signal that is present in the cytoplasm for two genes is not included when using nuclear segmentation.

(C) Visualization of whole-cell segmentation. Automated Cellpose segmentation is compared with manual segmentation and a mix of Cellpose and manual segmentation (Corrected Cellpose). Red arrowhead denotes a cell identified during manual segmentation but not during automated Cellpose segmentation.

(D) Mean cell counts, time burden, and F1 scores for manual, Cellpose, and Corrected Cellpose segmentation methods. n = 8 images per group.

(E) Mean (red text) and individual F1 scores for ROIs generated by Cellpose and Corrected Cellpose segmentation compared with the ground-truth ROIs generated by manual segmentation. n = 8 images per group.

(F) The total number of ROIs generated after Cellpose, Corrected Cellpose, or manual segmentation (see total number of cells in red). Note that cell area distributions are nearly indistinguishable between the three methods. n = 8 images per group.

analyze both average gene expression and expression topography with single-cell spatial resolution across all neurons in a single tissue section. Single-neuron expression of nine different genes was quantified in V1 (Figure 4A). The analysis of V1 not only reveals differences in mean gene expression between

transcripts (e.g., *Slc17a7* versus *Chat*) but also demonstrates that for some genes with similar mean expression (*Chd13* versus *Piezo1*), the neurons have distinct expression distribution ranges. Specifically, *Piezo1* expression is normally distributed and exhibits a smaller distribution range, while *Chd13*
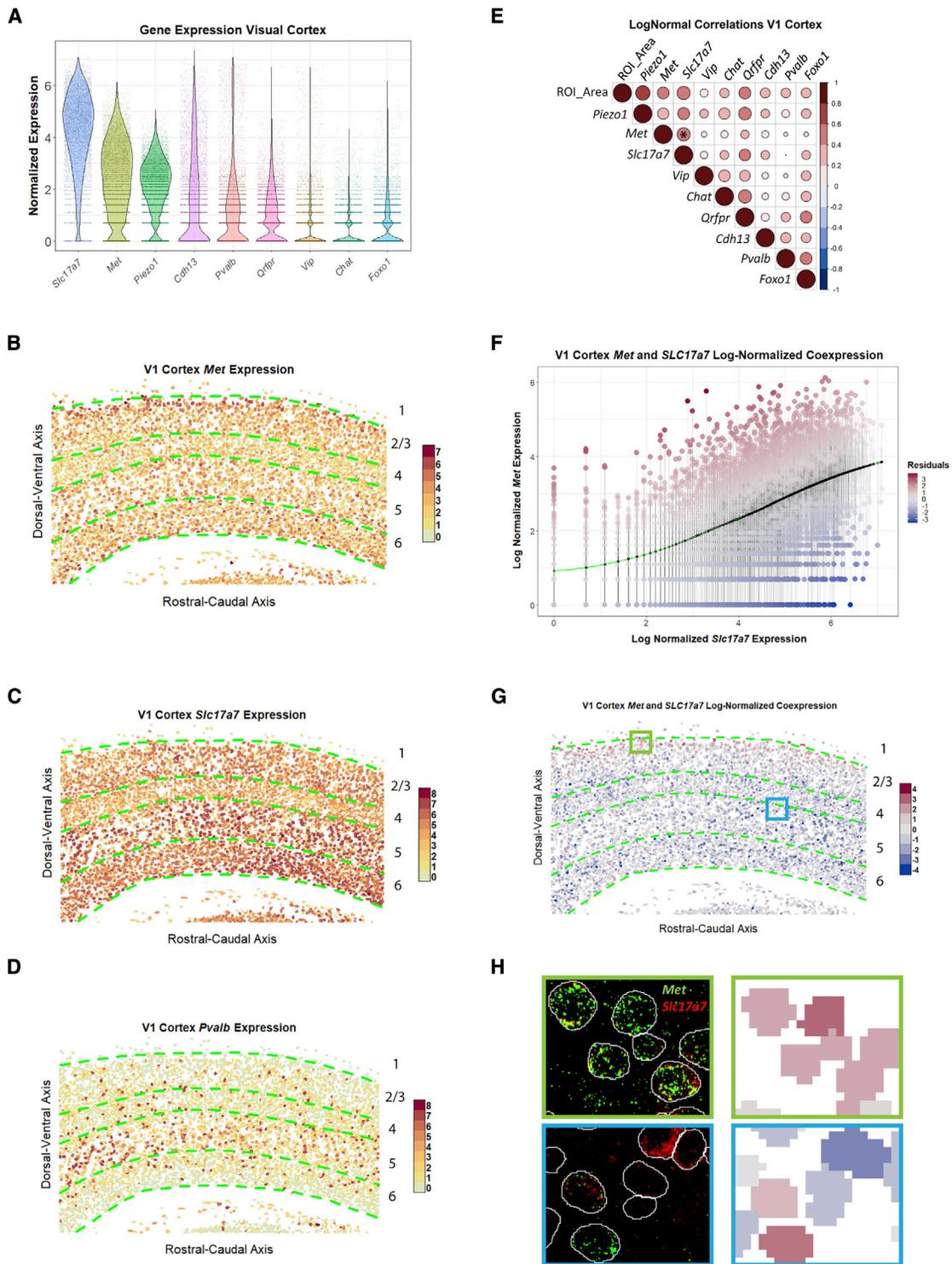
Figure 4. Analysis and spatial visualization of gene expression and co-expression of HiPlex RNAscope

(A) Violin plot of log-normalized gene expression for nine genes expressed in V1 cortex.

(B) Spatial map of log-normalized *Met* expression in the V1 cortex. Note the greater expression in layers 2–3 and low expression in layer 4.

(C) Spatial map of log-normalized *Slc17a7* expression in V1 cortex. Note the greater expression in layers 5 and 6 and medium-low expression in layer 4.

(D) Spatial map of log-normalized *Pvalb* expression in V1 cortex. Note high-expressing red profiles scattered across all layers, with an enrichment in layer 5.

(E) Pairwise correlation plot for log-normalized gene expression in V1 cortex. Asterisk denotes correlation between *Met* and *Slc17a7*, which is analyzed further in (F) and (G).

(legend continued on next page)

expression is skewed leftward toward zero and exhibits a larger overall range of expression (Figure 4A).

SCAMPR also facilitates the mapping of gene expression and gene co-expression patterns at single-cell spatial resolution onto the tissue of origin. This is demonstrated by using SCAMPR to map 3 genes, *Met*, *Slc17a7*, and *Pvalb*, each having distinct levels of expression and neuronal distributions in V1 cortical layers. Consistent with prior results from our laboratory, neurons expressing the highest levels of *Met* are distributed predominantly in layer 2, and those expressing the lowest levels of *Met* localize to layer 4 (Figure 4B) (Judson et al., 2009; Eagleson et al., 2016). *Slc17a7* is expressed in neurons across all layers in V1, with slightly lower expression in layer 4 (Figure 4C). Neurons expressing *Pvalb* at high levels also are scattered across all layers of V1, with noticeable enrichment in layer 5 (Figure 4D). In addition to single-gene spatial mapping, SCAMPR also can be used to quantify and spatially map gene co-expression patterns. Gene co-expression was quantitated by generating pairwise linear correlation coefficients between all nine genes in V1. A variety of relations are revealed, with some genes highly co-expressed (*Slc17a7* and *Met*), and others discordant (*Slc17a7* and *Pvalb*) (Figure 4E). In V1, *Slc17a7* and *Pvalb* are known to be expressed in two different neuronal populations (excitatory versus inhibitory neurons), whereas *Met* is highly enriched in excitatory neurons. Comparison of the expression levels of these genes across the layers of V1 at single-cell resolution as well as co-expression analysis confirm these prior findings. As an added dimension, the accurate segmentation of neuron cell bodies, rather than nuclei, provides cell size (ROI_Area) as a variable in this co-expression analysis. Matching prior reports in AMPA and NMDA gene expression in motor neurons, the quantitative expression levels of most genes in V1 exhibited positive correlations with cell area (Rana et al., 2020). One obvious exception was noted, with *Vip* (expressed in inhibitory neurons) exhibiting expression levels that were independent of cell size (Figure 4E).

A unique feature of data analysis using SCAMPR is the single-cell, topographic characterization of gene co-expression patterns between two genes. This is accomplished by fitting a trend line to the co-expression scatterplot of the two genes and subsequently measuring residual distances between each cell and the trend line. Here, the residual distances were mapped back onto the cells in the tissue. Fitting a locally estimated scatterplot smoothing (LOESS) line to the *Met* and *Slc17a7* co-expression scatterplot shows a positive relationship between *Met* expression and *Slc17a7* expression for the large majority of *Met*/ *Slc17a7* co-expressing cells. Cells that fall on or near the trend line represent the majority or "expected" co-expression pattern for *Met* and *Slc17a7*, whereas those that fall far from the trend line express relatively more or relatively less *Met* and *Slc17a7*

than expected. On the tissue section rendition, these subtypes were marked in increasingly darker shades of red if they expressed more *Met* and/or less *Slc17a7* than expected and increasingly darker shades of blue if they expressed less *Met* and/or less *Slc17a7* than expected (Figure 4F). The positive (red), negative (blue), or neutral (grey) distances of these cells from the trend line were then mapped back onto their location in the V1 tissue, showing that the *Met*/*Slc17a7* ratio was larger than the trend in layer 2 and smaller than the trend in multiple layers, particularly in layer 6a (Figure 4G). To validate the accuracy of this method for characterizing both trends, and thus the heterogeneity in the co-expression pattern of two genes, high-magnification insets from layers 2 and 5 of the spatial map were compared with the original HiPlex RNAscope images from the same regions (Figure 4H).

Clustering and spatial mapping of high-dimensional mRNA expression patterns

In addition to single and dual gene expression patterns, SCAMPR can utilize hierarchical clustering methods to categorize neurons into groups based on global gene expression patterns. Using nine genes assayed in the V1 cortex HiPlex RNAscope experiments, SCAMPR was used to cluster the cells into eight distinct groups. The clusters are represented as a joint dendrogram/heatmap (Figure 5A). Because ROI identity and spatial coordinates are preserved during the clustering, each cluster can be spatially mapped back onto the tissue section. Notably, some of the clusters segregate spatially. Cells that comprise cluster 1 are predominantly located in layer 4, whereas cells in cluster 8 are predominantly located in layers 5 and 6a (Figure 5B). Additionally, the cells that comprise cluster 8 are larger than those in cluster 1 (Figure 5C). Cells in cluster 8 also exhibit either higher or equivalent log-normalized gene expression when compared with cells in cluster 1, with prominent differences in genes such as *Cdh13* and *Foxo1* (Figure 5D). Examination of the original images for mRNA fluorescence validates significantly higher *Cdh13* and *Foxo1* expression in deeper, compared with superficial, layers, with the highest co-expression in layer 5 and a near absence of these two genes in layers 2–4 (Figure 5E).

Using SCAMPR to identify gene expression pattern differences between two experimental groups

Differences in expression levels of specific genes due to a manipulation may either vary across all neurons that express a specific transcript or in specific subtypes and can be resolved using quantitative analyses at the single-cell level. Thus, high-dimensional *in situ* hybridization can be used to compare the expression of a carefully selected set of genes, such as cell

---

(F) LOESS regression plot between log-normalized *Met* and *Slc17a7* expression in V1 cortex. Points denote single cells and are colored based on their distance from the regression line (residual size).

(G) Spatial map of residuals from (F). Red cells have large positive residuals and express more *Met* and/or less *Slc17a7* than is predicted by the model. Blue cells have large negative residuals and express less *Met* and/or more *Slc17a7* than is predicted by the model. Note layer 2 enrichment of red cells.

(H) Left: high-magnification micrographs, denoted by green (layer 2) and blue (layer 4–5 border) boxes in (G). Top image illustrates the relatively high expression of *Met* (green) compared with *Slc17a7* (red) in layer 2. The resulting large positive residuals (pink/red) are shown in the top right graphic. Bottom image illustrates cells with a high *Slc17a7* (red) in layer 4, lower co-expression in one profile, and higher expression of *Met* (green) in a layer 5 neuron. These expression patterns match with their residual-denoted colors in the spatial map (bottom right).
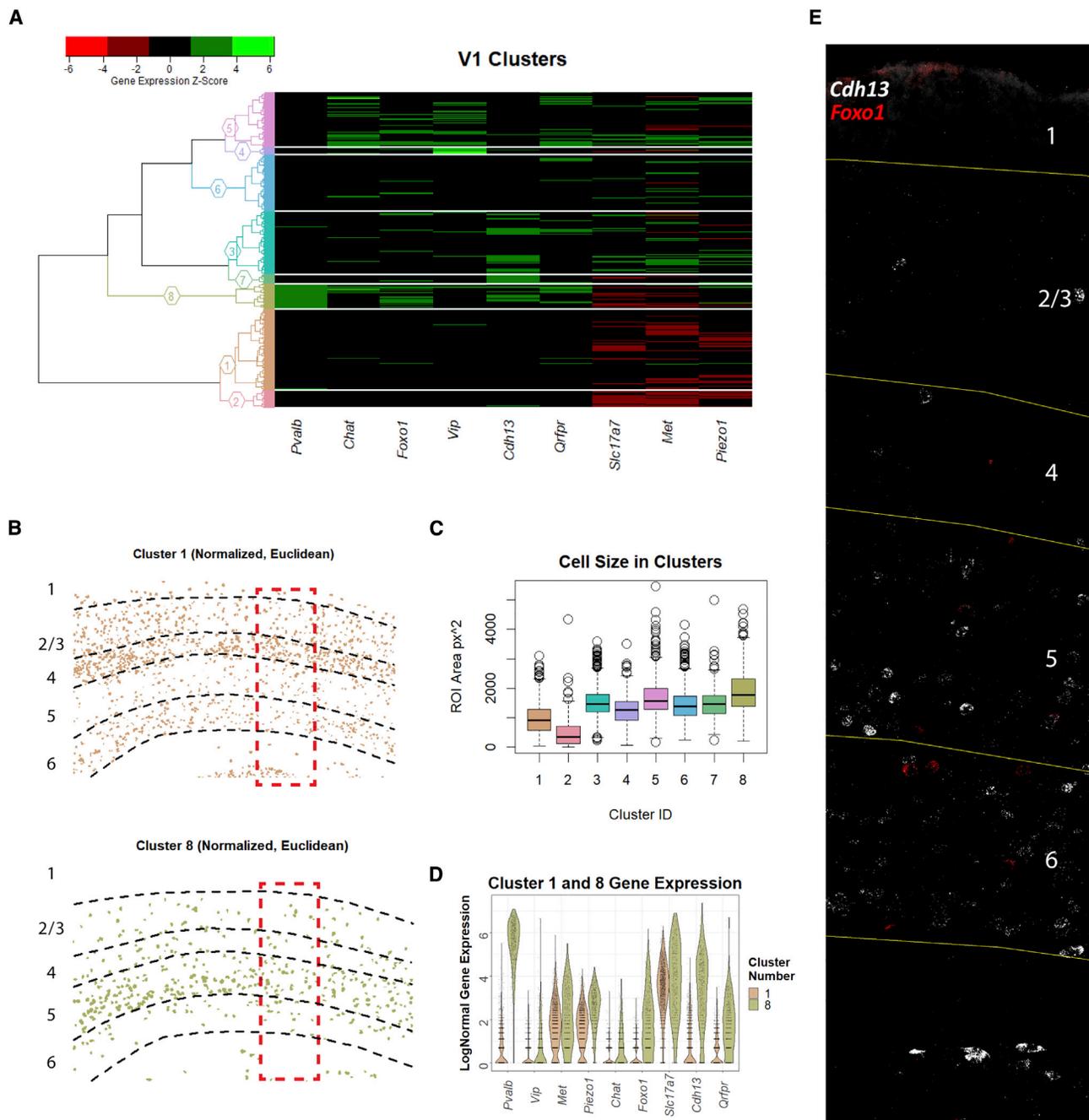
Figure 5. High-dimensional clustering analysis of HiPlex RNAscope data

(A) Heatmap of scaled gene expression in V1 cortex after hierarchical clustering.

(B) Spatial maps of cells in cluster 1, which are mostly localized in layers 2/4/6, and cluster 8, which are predominantly localized in layers 5/6a.

(C) Boxplots demonstrating differences in cell size (area) between the clusters. Largest difference in cell size is between cluster 1 (brown) and cluster 8 (green). Data are represented as medians.

(D) Violin plots demonstrating higher expression of several genes in cluster 1 compared with cluster 8.

(E) Hiplex RNAscope image of *Chd13* (white) and *Foxo1* (red), which are enriched in cluster 8 and absent in cluster 1. Note that both genes are more highly expressed in layers 5 and 6a, the primary location of cluster 8 cells.

subtype markers, to determine potential phenotypic differences that occur due to an experimental manipulation between groups. Here, we used a well-validated model of limited bedding and nesting (LBN), applied from P2 to P9 in mice. LBN disrupts maternal care and produces an ELS response in pups (Rice et al., 2008; Heun-Johnson and Levitt, 2016; Eagleson et al.,

2020). Care-as-usual (CAU) pups were raised under normal conditions. Development of vagal circuitry is sensitive to ELS, though information is limited at a molecular level (Card et al., 2005; Banihashemi and Rinaman, 2010). To investigate prospective ELS-induced vagal gene expression changes in early postnatal mice, we used SCAMPR to analyze the expression of 12 genes that demarcate adult subtypes of vagal sensory neurons located in the NG of ELS and CAU mice (Kupari et al., 2019). Tissue was harvested for processing on P9 at the end of the period of ELS exposure. Pairwise correlation coefficients and least squares regression analyses were performed for both groups separately to identify possible gene co-expression differences between groups (Figure S3). Visual inspection of the plot matrix was sufficient to narrow possible differences in gene co-expression between CAU and ELS. Analyses began with the identification of well-correlated gene pairs for all cells from all groups ($R > 0.40$). From these, gene pairs that had both prominent between-group differences in correlation coefficients and linear regression line slopes were selected. Two gene pairs fit these criteria: *Scn10a/Ntsr1* and *Scn1a/Pvalb* (Figure S3). To identify differences between the co-expression of these genes in CAU and ELS groups, scatterplots of *Scn10a/Ntsr1* co-expression and *Scn1a/Pvalb* co-expression were generated, and LOESS models were fitted for each group. These plots demonstrated lower *Pvalb* and *Ntsr1* expression with rising values of *Scn10a* and *Scn1a* in the ELS group when compared with the CAU group (Figures 6A and 6B). Since *Scn10a* and *Scn1a* have been shown to be broad cell-type markers in the NG, and because *Pvalb* and *Ntsr1* have been shown to be differentially expressed in NG populations with varying putative functions (control of pulmonary volume versus gastrointestinal chemo/mechano sensation), these analyses at single-cell resolution suggest that there are subtype-specific decreases in *Pvalb* and *Ntsr1* expression in NG cells after ELS exposure (Kupari et al., 2019).

The subtype-specific reductions observed could be driven either by an average decrease in expression of *Pvalb or Ntsr1* across many cells or by a decrease in the proportion of cells that express each gene. To investigate these possibilities, *Scn1a*-expressing (*Scn1a+*), *Scn10a*-expressing (*Scn10a+*), non-*Scn1a*-expressing (*Scn1a−*), and non-*Scn10a*-expressing (*Scn10a−*) data subsets were generated, and mean *Pvalb* and *Ntsr1* expression was calculated for NG neurons from each ELS and CAU pup. The averages of these mean expression values were compared between groups for each subset. There was no difference in mean *Pvalb* and *Ntsr1* expression in the *Scn1a−* and *Scn10a−* subsets of neurons, and while mean *Pvalb* and *Ntsr1* expression in ELS mice trended toward a decrease in in the *Scn1a+* and *Scn10a+* subsets of neurons, these differences did not reach statistical significance (Figures 6C and 6D).

It has been shown that the proportion of cells expressing a particular gene can be altered in an experience-dependent manner during development (Cheng et al., 2022). To determine whether the reduction of *Pvalb* and *Ntsr1* expression in ELS animals reflects a reduction in the proportion of cells expressing the genes rather than overall mean expression levels in the specific subpopulations, the percentage of cells expressing *Pvalb* and *Ntsr1* in the different cell-type-specific subsets was quantitated (Figures 6E and 6F). In *Scn1a+* cells, there was a significant

24.1% decrease in the percentage of cells expressing *Pvalb* in ELS animals when compared with CAU animals (51.8% CAU versus 39.3% ELS, $p \leq 0.05$), whereas in *Scn1a−* cells, there was no significant difference between groups (23.5% CAU versus 19.7% ELS, p = 0.56) (Figure 6E). In both *Scn10a+* and *Scn10a−* cells, there was a trending decrease in *Ntsr1*-expressing cells in ELS animals when compared with CAU animals (64.7% CAU versus 52.2% ELS, p = 0.12 and 25.8% CAU versus 21.1% ELS, p = 0.38) (Figure 6F). These analyses suggest that in ELS animals, *Pvalb* expression is reduced in a subset of NG cells that express *Scn1a* and that this reduction is partially driven by a reduction in the fraction of cells expressing *Pvalb* transcript. In addition, broad gene expression comparisons of the remaining ten genes demonstrated a decrease in mean *Gpr65* expression in ELS animals, yet no difference in the percentage of cells expressing this gene (Figure S4). Together, these experiments demonstrate that SCAMPR can be used as a tool to quantitatively compare global and cell-type-specific differences at single-cell spatial resolution, as well as gene co-expression patterns across experimental groups.

DISCUSSION

Here, we present SCAMPR, an open-source, user-friendly, time-saving, bench-to-desk pipeline for accurate quantification and analysis of spatial neuronal gene expression with single-cell resolution. The quantification and analytical tools that comprise SCAMPR allow for comparison of the expression patterns of multiple genes, spatial mapping of relative gene expression and co-expression patterns across a tissue, and cell-type-specific comparisons of gene expression between experimental groups, all at single-neuron resolution. The comparative analyses of cell labeling methods demonstrated superior savings of time, fully completing bench and analytical work in approximately 5 days, without sacrificing accuracy, while using fully automated or semi-automated segmentation of the entire neuronal soma.

Achieving accurate spatial analyses at single-neuron resolution

Utilization of HuC/D staining in the SCAMPR pipeline allows for demarcation of neurons across multiple developmental time points and multiple subregions of the nervous system that exhibit different cell packing densities and spatial organization patterns, allowing broad application of SCAMPR for analyzing topographic gene expression in neurons. The stability of the anti-HuC/D antibody under the harsh protease digestions and antigen retrieval processing steps of HiPlex RNAscope allows for HuC/D staining at the very end of hybridization protocol, circumventing the early occupation of an imaging channel or the need for protein stain degradation/quenching or fluorophore cleaving. Other sufficiently stable primary antibodies marking glial cells in the nervous system, or other antibodies marking non-nervous system cell classes present in other tissue types, can be validated for tissue processing stability and used with SCAMPR to quantify gene expression differences in myriad cell types.

SCAMPR utilizes a combination of Cellpose and ImageJ scripts to quantify gene expression for each neuron. In contrast to methods that perform puncta counting, which rely on intense particles standing out from their surroundings,
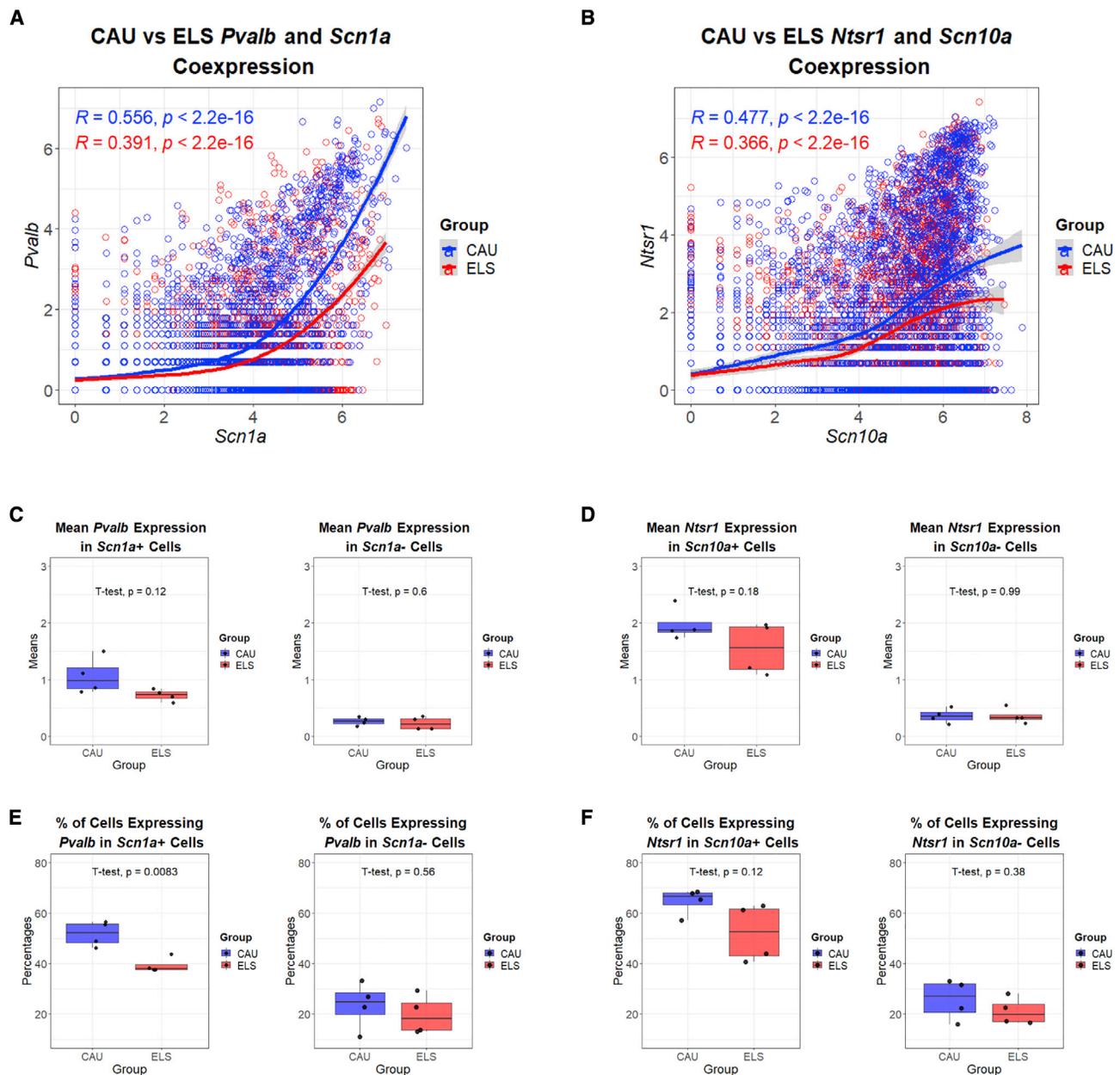
Figure 6. Utilization of HiPlex RNAscope for comparison of gene expression at single-cell resolution in nodose ganglion neurons from CAU and ELS experimental groups

(A and B) *Pvalb*/*Scn1a* and *Ntsr1*/*Scn10a* co-expression scatterplots for all cells in the dataset. Points denote single cells in CAU (blue) and ELS (red) mice. Solid blue and red lines denote the LOESS lines of CAU and ELS mice, respectively. Pearson's r and corresponding p values for the correlations were determined for each group separately.

(C) Mean, per-animal *Pvalb* expression in ELS and CAU *Scn1a*-expressing (left) and non-*Scn1a*-expressing (right) cells (n = 4 per group). No statistically significant difference noted between groups. Data are represented as medians.

(D) Mean, per-animal Ntsr1 expression in ELS and CAU *Scn10a*-expressing (left) and non-*Scn10a*-expressing (right) cells (n = 4 per group). No statistically significant difference noted between groups. Data are represented as medians.

(E) Percentage of cells from CAU and ELS animals that express *Pvalb* in *Scn1a*+ (left) and *Scn1a*− cells (right) (n = 4 per group). Note the statistically significant decrease in the percentage of neurons expressing *Pvalb* in *Scn1a*+ neurons in the ELS mice. Data are represented as medians.

(F) Percentage of cells from CAU and ELS animals that express *Ntsr1* in *Scn10a*+ (left) and*Scn10a*− cells (right) (n = 4 per group). No statistically significant difference noted between groups. Data are represented as medians.

SCAMPR employs a detection-by-thresholding method to quantify the number of pixels in a binary image that are positive for a fluorescent signal representing a particular gene. Because adjoining fluorescent molecules do not need to be separated from one another based on intensity in detection by thresholding, this method allows for accurate quantification of genes expressed at low or high levels within the same tissue section or even in adjacent neurons. Furthermore, SCAMPR quantifies gene expression using flattened, z-projected images, and outputs gene expression as the percentage of the cell area or the number of pixels within a cell area that are positive for a gene. This contrasts with some puncta count methods, which match puncta with cellular ROIs by assigning x, y, and z coordinates to each individual puncta, thus generating large data matrices that increase the computational burden in the requisite downstream analyses. This is an important feature when considering data management and computational time savings. Lastly, segmentation based on the use of both DAPI and HuC/D, in conjunction with modified versions of the ImageJ quantification macros, can be used to obtain cell compartment information by separately quantifying nuclear and cytosolic mRNA signals.

Spatial mapping tools enable visualization of complex expression, co-expression, and cell-type clustering patterns

The capabilities of SCAMPR for fast and accurate segmentation of neurons and quantification of in situ hybridization signal were combined with its tools for spatial analysis of gene expression, co-expression, and clustering-driven cell-type localization in the V1, a highly topographic structure with a laminar organization. We posited that challenging the ability of SCAMPR to accurately segment, measure, and map gene expression in a highly organized brain structure that is characterized by region-specific differences in packing density would reveal strengths and weaknesses of the pipeline. The data demonstrate the capabilities of SCAMPR for detecting layer-enriched differences in the expression of single and multiple genes in this brain region. Analyses of the co-expression of two genes (*Met* and *Slc17a7*), in which each cell was binned based on residual proximity to a fitted line, facilitated spatial mapping of the general co-expression patterns of these genes in the majority of neurons. Using this same analysis method, SCAMPR was able to extract minority populations of neurons with more unique co-expression patterns based on their deviation from the general co-expression patterns in the majority population of neurons. This holistic analysis method will compliment more direct strategies of separating neuron populations based on gene co-expression patterns, such as the semi-quantitative categorization of neurons based on the expression levels of both genes (low-low, low-high, high-low, and high-high) or by considering their expression distributions and summary statistics (min, max, mean, quartiles, etc.). Lastly, it is possible to integrate HiPlex RNAscope and other smFISH data with published single-cell RNA sequencing datasets (Hashikawa et al., 2020; Wang et al., 2021). These integrated datasets could theoretically be used with SCAMPR to spatially map and validate the expression levels of genes from single-cell RNA sequencing experiments.

SCAMPR enables analysis of differentially expressed genes in subsets of cell types

SCAMPR's utility for comparing gene expression between neurons developing under different environmental conditions was demonstrated. In the NG, a tissue without a known cell-type-specific topography, employing SCAMPR successfully identified unique cell-type-specific changes in gene expression due to ELS. Furthermore, we demonstrated varied responses to ELS across cell subtypes, where some genes had lower mean expression across a cell type (*Gpr65*), while others were expressed in a lower number of neurons in specific cell types (*Pvalb*) (Figure S2, 6D, and 6F). It should be noted that many of the genes assayed in these experiments are expressed in both the jugular and NG of the jugular-nodose complex. Future experiments can employ ganglion specific markers such as *Phox2b* to interrogate both cell-type- and tissue-specific changes in gene expression after ELS. In summary, SCAMPR also provides a cost-effective approach to high-resolution detection of potential, single-cell differences in gene expression in experimental models, with spatial details retained.

Limitations of study

SCAMPR, which utilizes HuC/D signal for cellular segmentation and ROI generation, is currently limited to the study of neurons in the nervous system. While the stability of reagents under multiplex in situ hybridization conditions would need to be assessed, it is likely that antibodies detecting proteins to mark other non-neural cells can be utilized with SCAMPR to assay gene expression in unique cell populations. This would facilitate analysis in pathogenic tissues, organoids, and other sources of tissue to assay gene expression spatially.

The SCAMPR pipeline utilizes three cellular attributes—gene expression, soma size, and cellular topography—to distinguish neuron populations. Yet, neuronal innervation patterns, often in combination with the three attributes utilized by SCAMPR, are also important for categorizing neurons into distinct subtypes. It has been demonstrated that one round of RNAscope is compatible with retrograde tracers, making it feasible to incorporate a fourth modality, circuit topology, into the SCAMPR pipeline (Rana et al., 2020). For this to be successfully incorporated with multiround HiPlex RNAscope and SCAMPR, the signal from the traced cells would require bleaching or degradation after an initial round of imaging, making the microscopy channel available for use in the proceeding rounds to assay mRNA expression.

To accurately quantify mRNA signal in image sets with large variations in gene expression and signal intensity, SCAMPR utilizes either manual or semi-automated threshold selection to distinguish a positive signal from noise in an image-specific manner. To mitigate the effects of between-user variation during threshold selection, it is imperative that the user utilize proper experimental blinding procedures to the identity of the image and the experimental group during imaging and throughout the SCAMPR pipeline.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

Cell Reports Methods
Article

SUPPLEMENTAL INFORMATION

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

R.A.M.G. and V.M.M. conceived of and developed the data analysis methodology; R.A.M.G., V.M.M., A.K.K., and P.L. provided recommendations for modification of the studies; A.K.K. generated the experimental material using the ELS model and validated the code; R.A.M.G. wrote the paper; and V.M.M., A.K.K., and P.L. edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Banihashemi, L., and Rinaman, L. (2010). Repeated brief postnatal maternal separation enhances hypothalamic gastric autonomic circuits in juvenile rats. Neuroscience 165, 265–277. https://doi.org/10.1016/J.NEUROSCIENCE.2009.09.081.

Caicedo, J.C., Roth, J., Goodman, A., Becker, T., Karhohs, K.W., Broisin, M., Molnar, C., McQuin, C., Singh, S., Theis, F.J., and Carpenter, A.E. (2019). Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. Cytometry A. 95, 952–965. https://doi.org/10.1002/CYTO.A.23863.

Card, J.P., Levitt, P., Gluhovsky, M., and Rinaman, L. (2005). Early experience modifies the postnatal assembly of autonomic emotional motor circuits in rats. J. Neurosci. 25, 9102–9111. https://doi.org/10.1523/JNEUROSCI.2345-05.2005.

Stapel, L.C., Lombardot, B., Broaddus, C., Kainmueller, D., Jug, F., Myers, E.W., and Vastenhouw, N.L. (2016). Automated detection and quantification of single RNAs at cellular resolution in zebrafish embryos. Development (Camb.) 143, 540–546. https://doi.org/10.1242/dev.128918.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348. https://doi.org/10.1126/SCIENCE.AAA6090.

Cheng, S., Butrus, S., Tan, L., Xu, R., Sagireddy, S., Trachtenberg, J.T., Shekhar, K., and Zipursky, S.L. (2022). Vision-dependent specification of cell types and function in the developing cortex. Cell 185, 311–327.e24. https://doi.org/10.1016/J.CELL.2021.12.022.

Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development 145. https://doi.org/10.1242/DEV.165753/48466.

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

Diaz-Garcia, S., Ko, V.I., Vazquez-Sanchez, S., Chia, R., Arogundade, O.A., Rodriguez, M.J., Traynor, B.J., Cleveland, D., and Ravits, J. (2021). Nuclear depletion of RNA binding protein ELAVL3 (HuC) in sporadic and familial amyotrophic lateral sclerosis. Preprint at bioRxiv. https://doi.org/10.1101/2021.06.03.446017.

Dries, R., Zhu, Q., Dong, R., Eng, C.H.L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., and George, R.E. (2019). Giotto, a toolbox for integrative analysis and visualization of spatial expression data. https://doi.org/10.1101/701680.

Eagleson, K.L., Lane, C.J., McFadyen-Ketchum, L., Solak, S., Wu, H.H., and Levitt, P. (2016). Distinct intracellular signaling mediates C-MET regulation of dendritic growth and synaptogenesis. Dev. Neurobiol. 76, 1160–1181. https://doi.org/10.1002/DNEU.22382.

Eagleson, K.L., Villaneuva, M., Southern, R.M., and Levitt, P. (2020). Proteomic and mitochondrial adaptations to early-life stress are distinct in juveniles and adults. Neurobiol Stress 13, 100251. https://doi.org/10.1016/J.YNSTR.2020.100251.

Eng, C.H.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.C., and Cai, L. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature 568, 235–239. https://doi.org/10.1038/S41586-019-1049-Y.

Erö, C., Gewaltig, M.O., Keller, D., and Markram, H. (2018). A cell atlas for the mouse brain. Front. Neuroinf. 12, 84. https://doi.org/10.3389/FNINF.2018.00084/BIBTEX.

Goh, J.J.L., Chou, N., Seow, W.Y., Ha, N., Cheng, C.P.P., Chang, Y.C., Zhao, Z.W., and Chen, K.H. (2020). Highly specific multiplexed RNA imaging in tissues with split-FISH. Nat. Methods 17, 689–693. https://doi.org/10.1038/S41592-020-0858-0.

Hashikawa, Y., Hashikawa, K., Rossi, M.A., Basiri, M.L., Liu, Y., Johnston, N.L., Ahmad, O.R., and Stuber, G.D. (2020). Transcriptional and spatial resolution of cell types in the mammalian habenula. Neuron 106, 743–758.e5. https://doi.org/10.1016/j.neuron.2020.03.011.

Herculano-Houzel, S., Mota, B., and Lent, R. (2006). Cellular scaling rules for rodent brains. Proc. Natl. Acad. Sci. 103, 12138–12143. https://doi.org/10.1073/PNAS.0604911103.

Heun-Johnson, H., and Levitt, P. (2016). Early-life stress paradigm transiently alters maternal behavior, dam-pup interactions, and offspring vocalizations in mice. Front. Behav. Neurosci. 10, 142. https://doi.org/10.3389/fnbeh.2016.00142.

Judson, M.C., Bergman, M.Y., Campbell, D.B., Eagleson, K.L., and Levitt, P. (2009). Dynamic gene and protein expression patterns of the autism-associated met receptor tyrosine kinase in the developing mouse forebrain. J. Comp. Neurol. 513, 511–531. https://doi.org/10.1002/cne.21969.

Kupari, J., Häring, M., Agirre, E., Castelo-Branco, G., and Ernfors, P. (2019). An atlas of vagal sensory neurons and their molecular specialization. Cell Rep. 27, 2508–2523.e4. https://doi.org/10.1016/j.celrep.2019.04.096.

Maynard, K.R., Tippani, M., Takahashi, Y., Phan, B.N., Hyde, T.M., Jaffe, A.E., and Martinowich, K. (2020). dotdotdot: an automated approach to quantify multiplex single molecule fluorescent in situ hybridization (smFISH) images

in complex tissues. Nucleic Acids Res. *48*, e66. https://doi.org/10.1093/nar/gkaa312.

Okano, H.J., and Darnell, R.B. (1997). A hierarchy of hu RNA binding proteins in developing and adult neurons. J. Neurosci. *17*, 3024–3037. https://doi.org/10.1523/JNEUROSCI.17-09-03024.1997.

Peña, C.J., Kronman, H.G., Walker, D.M., Cates, H.M., Bagot, R.C., Purushothaman, I., Issler, O., Loh, Y.H.E., Leong, T., Kiraly, D.D., and Goodman, E. (2017). Early Life Stress Confers Lifelong Stress Susceptibility in Mice via Ventral Tegmental Area OTX2. Science *356*, 1185–1188.

Rana, S., Sieck, G.C., and Mantilla, C.B. (2020). Heterogeneous glutamatergic receptor mRNA expression across phrenic motor neurons in rats. J. Neurochem. *153*, 586–598. https://doi.org/10.1111/jnc.14881.

Rice, C.J., Sandman, C.A., Lenjavi, M.R., and Baram, T.Z. (2008). A novel mouse model for acute and long-lasting consequences of early life stress. Endocrinology *149*, 4892–4900. https://doi.org/10.1210/EN.2008-0633.

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

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., and Mollbrink, A. (2016).

Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Am. Assoc. Adv. Sci. *353*, 78–82. https://doi.org/10.1126/SCIENCE.AAF2403.

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

Wang, F., Flanagan, J., Su, N., Wang, L.C., Bui, S., Nielson, A., Wu, X., Vo, H.T., Ma, X.J., and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J. Mol. Diagn. *14*, 22–29. https://doi.org/10.1016/j.jmoldx.2011.08.002.

Wang, Y., Eddison, M., Fleishman, G., Weigert, M., Xu, S., Wang, T., Rokicki, K., Goina, C., Henry, F.E., Lemire, A.L., and Schmidt, U. (2021). EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization. Cell *184*, 6361–6377.e24. https://doi.org/10.1016/J.CELL.2021.11.024/ATTACHMENT/D0CE8F78-B9C5-4C98-8D79-3FCF32F06C0F/MMC4.

Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (2020). Starconvex polyhedra for 3D object detection and segmentation in microscopy. In In Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 3655–3662. https://doi.org/10.1109/WACV45572.2020.9093435.

Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. Nat. Rev. Neurosci., 530–546. https://doi.org/10.1038/nrn.2017.85.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
HuC/HuD	ThermoFisher Scientific	RRID:AB_221448
Alexa Fluor 594 AffiniPure F(ab')₂ Fragment Donkey Anti-Mouse IgG (H + L)	Jackson ImmunoResearch	RRID:AB_2340858
Chemicals, peptides, and recombinant proteins		
Tissue-Tek Optimal Cutting Temperature Compound	Sakura Finetek	4583
ProLong Gold Antifade Mountant	Jackson ImmunoResearch	Cat. #017-000-121
Tween-20	Sigma-Aldrich	P2287
SSC	G. Biosciences	786-023
Bovine Serum Albumin	Gemini Bio	700-106P
Normal Donkey Serum	Jackson ImmunoResearch	017-000-121
Critical commercial assays		
RNAScope Intro Pack for HiPlex12 Reagent Kit – Fresh/Fixed Frozen – Mm	ACD Biosciences	324102
RNAscope HiPlex Probe – Mm – Pvalb – T1	ACD Biosciences	421931-T1
RNAscope HiPlex Probe – Mm – Met – T2	ACD Biosciences	405301-T2
RNAscope HiPlex Probe – Mm – Piezo1 – T3	ACD Biosciences	400181-T3
RNAscope HiPlex Probe – Mm – Chat– T4	ACD Biosciences	408731-T4
RNAscope HiPlex Probe – Mm – Foxo1 – T5	ACD Biosciences	485761-T5
RNAscope HiPlex Probe – Mm – Slc17a7 – T6	ACD Biosciences	416631-T6
RNAscope HiPlex Probe – Mm – Cdh13 – T7	ACD Biosciences	443251-T7
RNAscope HiPlex Probe – Mm – Qrfpr – T8	ACD Biosciences	566341-T8
RNAscope HiPlex Probe – Mm – Mab21l1 – T9	ACD Biosciences	485051-T9
RNAscope HiPlex Probe – Mm – Vip – T10	ACD Biosciences	415961-T10
RNAscope HiPlex Probe – Mm – Gfra1 – T11	ACD Biosciences	431781-T11
RNAscope HiPlex Probe – Mm – Cckar – T12	ACD Biosciences	313751-T12
RNAscope HiPlex Probe – Mm – Ntsr1 – T2	ACD Biosciences	422411-T2
RNAscope HiPlex Probe – Mm – Piezo1 – T3	ACD Biosciences	400181-T3
RNAscope HiPlex Probe – Mm – Npy2r – T4	ACD Biosciences	315951-T4
RNAscope HiPlex Probe – Mm – Httr3b – T5	ACD Biosciences	497541-T5
RNAscope HiPlex Probe – Mm – Trpv1 – T6	ACD Biosciences	313331-T6
RNAscope HiPlex Probe – Mm – Scn1a – T7	ACD Biosciences	434181-T7
RNAscope HiPlex Probe – Mm – Scn10a – T8	ACD Biosciences	426011-T8
RNAscope HiPlex Probe – Mm – Glp1r – T9	ACD Biosciences	418851-T9
RNAscope HiPlex Probe – Mm – Gpr65 – T10	ACD Biosciences	431431-T10
RNAscope HiPlex Probe – Mm – Piezo2 – T11	ACD Biosciences	400191-T11
Experimental models: Organisms/strains		
C57BL/6J Mice	The Jackson Laboratory	000664
Software and algorithms		
SCAMPR_FilenamesCSV	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_AreaFraction ImageJ Macro	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_AreaFraction_SemiAuto ImageJ Macro	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_ViolinPlots	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_ExpressionTopomaps	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_LoessCoexpressionTopomaps	GitHub	https://zenodo.org/badge/latestdoi/443220284

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SCAMPR_CorrelationCirclePlots	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_hClustHeatmapTopomaps	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_hClustViolinPlotsCellAreas	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_LoessScatterplotComparisons	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_MeanComparisonsCellTypes	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_PercentCellsComparisonsCellTypes	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_CorrelationRegressionComparisonsMatrix	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_MeanComparisonsMatrix	GitHub	https://zenodo.org/badge/latestdoi/443220284
SCAMPR_PercentCellsComparisonsMatrix	GitHub	https://zenodo.org/badge/latestdoi/443220284

RESOURCE AVAILABILITY

Lead contact
Further information and help with the pipeline should be directed to the lead contact, Pat Levitt (plevitt@med.usc.edu).

Materials availability
This study did not generate new unique reagents or mouse lines.

Data and code availability
- Microscopy data, HiPlex images, and any additional data reported in this paper will be shared by the lead contact upon request.
- HiPlex Gene-Count matrices and all code used to analyze each Gene-Count matrix are publicly available at GitHub as of the date of publication (https://github.com/ramin-ali-marandi-ghoddousi/SCAMPR). DOI numbers are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Animals
One female C57BL/6J mouse, sacrificed at postnatal day 100 (P100), was used for the V1 cortex HiPlex RNAscope experiments. Four ELS and four CAU litters were used for the ELS experiments. From each litter, 1 male and 1 female postnatal day 9 (P9) C57BL/6J mouse was chosen at random and sacrificed for the NG HiPlex RNAscope experiments.

All experimental procedures were performed in accordance with the Institutional Animal Care and Use Committee of The Saban Research Institute, Children's Hospital Los Angeles.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Sequential steps for SCAMPR pipeline
Tissue collection and preparation
Prior to collection of adult visual cortex and early postnatal nodose ganglion, mice were anesthetized with a Ketamine:Xylazine mixture (100 mg/kg: 10 mg/kg) and transcardially perfused with 4% paraformaldehyde (pH 7.4). Due to the small size of the tissue, a dissecting microscope was used to collect the nodose ganglion from each mouse. All collected tissue was postfixed in 4% para-formaldehyde for 2–3 hours at 4°C, then cryoprotected overnight in 20% sucrose. Tissue samples were embedded in Tissue-Tek Optimal Cutting Temperature Compound (OCT) (Sakura Finetek, Cat. #4583), frozen over dry ice, and stored at −80°C.

20 μM cryosections were collected in six series through the entirety of the nodose ganglion and mounted directly onto slides. For the visual cortex, 20 μM cryosections were collected in the sagittal orientation in 4 series. All slides were stored at −20°C until processing.

Dual Immunohistochemistry-HiPlex RNAscope and imaging
The RNAscope HiPlex Assay was performed according to the manufacturer's standard protocol using the RNAscope HiPlex12 Kit (Advanced Cell Diagnostics Cat. #324194). Tissue sections were baked for 1 hour at 60°C and dehydrated in an ethanol series, followed by antigen retrieval (5 min at 100°C) and protease treatment (Protease IV for 30 min at room temperature for NG, Protease III for 30 min at 40°C for V1). Probes for twelve target genes were hybridized for 2 hours at 40°C, washed, and hybridized

with target-binding amplifiers allowing for signal amplification of single RNA transcripts. Hybridization with negative control probes targeting bacterial genes was performed in parallel. The final step of the first round of hybridization attached three fluorophores to the first three of twelve of the target genes (T1-T3). Once the fluorophores were hybridized to the three genes, the sections were counterstained with DAPI for 30 seconds, then mounted/coverslipped with ProLong Gold Antifade Mountant (Jackson ImmunoResearch, Cat. #017-000-121). For the NG samples, signal detection was performed in four rounds, where three target genes were labeled with three cleavable fluorophores and imaged with a 40X water objective each round on a Zeiss LSM 710 confocal microscope. Experimenters were blind to the genotype of all animals during imaging and signal quantification of the NG sections. For the V1 sample, signal detection was performed in three rounds (4 probes each round) due to the availability of an infrared fluorophore (ACD Biosciences Catalog No. 322830) and a Leica Stellaris 5 confocal microscope with infrared detector. The V1 sections were imaged using a 40x water objective on a Leica Stellaris 5 and automated, inter-tile registration and blending was performed for the visual cortex images using the Stellaris imaging software (LAS X 4.1) to resolve any coordinate discrepancies between adjoining tiles.

For each section, the gain and laser power were qualitatively optimized by the experimenter for each channel (DAPI, Alexa Fluor 488, Atto 550, Atto 647N, AF750 – V1 only) and a 20 μM z-stack image was obtained (1.33 μM step-size). No signal was observed in the negative control tissue sections, or after cleaving all fluorophores, demonstrating the specificity of the assay. For the NG, only the middle 6–7 μM of each image was used for downstream analysis to minimize packing density-related segmentation issues.

After the sections were imaged, the coverslips were removed in 4X SSC buffer (G. Biosciences, Cat. #786-023) and the fluorophores were cleaved using the cleaving solution provided in the kit. A new set of fluorophores targeting the next three genes (T4-T6 for NG, T5-T9 for V1) were hybridized onto the tissue sections, another round of DAPI counterstaining was performed, and the sections were reimaged as described above. This was repeated until all 12 target genes were imaged. In order to identify the neurons in the NG and V1, immunofluorescence labeling was performed in the same tissue sections following the cleaving of the fluorophores from the last round of the HiPlex Assay. The coverslips were washed off once again, tissue sections were briefly washed in 0.005% Tween-20 (Sigma-Aldrich Cat. #P2287) in PBS before incubation for 30 min in blocking solution containing 10% normal donkey serum (Jackson ImmunoResearch Cat. #017-000-121) and 1% bovine serum albumin (Gemini Bio Cat. #700-106P) in PBS. Slides were incubated overnight at room temperature with antibodies against the neuronal marker HuC/HuD (1:500, ThermoFisher Scientific Cat. #A-21271, RRID:AB_221448) with 1% BSA in PBS. Slides were washed in 0.005% PBST, then incubated for 1 hour at room temperature in Alexa Fluor 594 AffiniPure F(ab')₂ Fragment Donkey Anti-Mouse IgG (H + L) (1:500, Jackson ImmunoResearch Cat. #715-586-151, RRID:AB_2340858) with 1% BSA in PBS. Following three washes in 0.005% PBST and one wash in PBS, slides were counterstained with DAPI (Advanced Cell Diagnostics Cat. #324108) and coverslipped using ProLong Gold Antifade Mountant (ThermoFisher Scientific Cat. #36930). One final round of imaging was performed as described above to capture the HuC/D and DAPI signals.

Image preprocessing

Each confocal LSM (NG) or LIF (V1) image file was loaded into FIJI ImageJ. A maximum intensity z-projection was created, each channel (corresponding to a specific gene / HuC/D / DAPI) was split into a separate image file, and each new image was saved as a TIFF file. Generating flattened z-projections at this step provides **major advantages to efficiency** in the downstream processing; image registration and cell segmentation only needs to be performed once for each z-projected image, in contrast to separate registration and segmentation of each of the multiple z-planes in each image. To increase the speed of this step, a macro was generated in FIJI ImageJ to create and save maximum intensity z-projections for all images in a folder or set of subfolders.

Next, images from multiple rounds of imaging the same tissue were registered together using the ACD Biosciences Image Registration Software, following the manufacturers protocol (ACD Biosciences Document No. 30065UM). In brief, a DAPI image from one of the rounds of imaging was used as the reference image. DAPI images from all other rounds of imaging were registered to this reference image, generating a transformation matrix of coordinate conversions that was applied to the remaining gene and HuC/D images from each imaging round to create one, unified coordinate system for all images from all rounds. In the same software, non-overlapping regions around the edges of the images are cropped, and each registered image is saved separately as a TIFF file for downstream processing. An alternative registration method using ImageJ is detailed on the SCAMPR GitHub.

Cell segmentation and signal quantification

Manual cell segmentation was performed by opening all registered HuC/D maximum intensity projection images in FIJI ImageJ, then using the freehand selection tool to circle individual cells, which were each outlined as separate regions of interest (ROIs). ROIs for each image were then saved into a ZIP file for use in later processing. Automated cell segmentation was performed using Cellpose (Stringer et al., 2021). A Google Colab notebook that was provided by the authors of Cellpose was modified and used to generate the ROIs for each image based on the HuC/D images. The ROI outlines that are generated by Cellpose were converted into FIJI ImageJ compatible ROIs using the Cellpose_Outline_to_ROI_Converter.py script and saved as a ZIP file for use in later processing. We demonstrate that even in densely packed tissues such as the NG, the maximum intensity projections are suitable for cellular ROI generation as changes in cell morphology or the appearance/disappearance of cells across the z-stack have minimal impact on segmentation accuracy (Figures S5A and S5B).

Semi-automated signal quantification using SCAMPR requires limited user inputs. In ImageJ, background subtraction was used to remove imaging artifacts such as autofluorescence and other noise. Next, global image thresholding was performed to designate each pixel in the image as either signal or background. Both steps require the selection of specific parameters: a rolling-ball radius

pixel size for the background subtraction, and a minimum threshold value for the global image thresholding. Through trial and error, a rolling-ball radius of radius 1 works well across all images and is therefore set as the default. This can be changed by the user when analyses are performed. Due to differential tissue quality, imaging, and mRNA expression levels between multiple genes, using only one set of threshold parameters or one automated thresholding method does a poor job of separating signal from background for all mRNAs across multiple tissues (Figure S1A). For this reason, experimenter-blinded manual selection of the minimum threshold value is required for each image prior to signal quantification. The rolling-ball radii and threshold values for image files of interest are saved in one CSV file, which is utilized by the FIJI ImageJ macros that perform automated signal quantification in downstream steps.

For users that require a more time friendly alternative for determining the minimum threshold value, we have included a semi-automated method to SCAMPR that calculates image-specific threshold values for the entire image set by using image intensities and manual threshold specification of one of the images. To do this: 1) calculate the mean pixel intensity of all background corrected images in an experiment, 2) calculate the maximum of these mean intensity values (Max Mean Intensity), 3) identify the image for which the mean is the closest to the Max Mean (Representative Image), 4) perform manual thresholding of the Representative Image and save the lower threshold value (Representative Threshold), 5) divide the mean intensity of each individual image (Individual Mean Intensity) by the Max Mean Intensity to obtain an image-specific thresholding weight, and 3) multiply the Representative Threshold by the image-specific thresholding weight to obtain the intensity-adjusted lower threshold (Figures S1B–S1E).

$$\textit{Individual Threshold} = \textit{Representative Threshold} \times \frac{\textit{Individual Mean Intensity}}{\textit{Max Mean Intensity}}$$

A FIJI ImageJ macro was developed as part of SCAMPR to quantify whole cell mRNA signal from RNAscope experiments. The inputs for this macro, which were generated and processed in previous steps, are as follows: 1) the maximum intensity z-projected images each gene, 2) the cellular ROIs generated manually or through utilization of Cellpose, and 3) the CSV file containing the optimal, image-specific rolling-ball radius and minimum threshold as determined by the experimenter. Using this information, these macros determine the expression of all genes in all cells for the input images, thereby generating a Gene by Cell matrix file with the results. Each row of this matrix corresponds to a single cell and contains the animal ID, the image ID, the ROI (cell) area, and the expression level of each gene of interest. The gene expression level is displayed as an area-fraction calculation (the number of pixels expressing the gene within the cell divided by the area of the cell in pixels). This area-fraction expression value can be used as-is for the downstream analysis or can be converted to a value representing the number of pixels expressing the gene by multiplying the area-fraction by the cell area (i.e. pixel coverage in a cell).

Quantitative analysis and spatial mapping
The first step prior to analysis is to import the Gene by Cell matrix created in the previous step into RStudio. Cells not expressing any genes can be removed from the dataset if desired. Depending on the genes chosen for the experiment, this can help remove noise due to misclassification of tissue imperfections or other structures as cells. Next, the columns containing the gene-expression data can be left as area-fraction calculations or converted into pixel coverage. These values are then natural log normalized to reduce high-expressing gene bias during analysis and data visualization and used to generate violin plots, co-expression charts, spatial maps, and cluster heatmaps/dendrograms.

Violin plots. Violin plots can be generated in two ways: 1) by plotting the gene expression levels of each gene in each cell, or 2) by only plotting the expression levels of genes after excluding cells that do not express the gene of interest. The latter provides a comparison of gene expression levels, while the former will give a comparison of both expression levels as well as a relative comparison of the number of cells that express a particular gene.

Expression, co-expression, and spatial mapping. Cellular ROIs are uploaded into R and matched to their appropriate cells in the Gene by Cell matrix. If desired, manual annotations that mark structural boundaries (e.g the layers in the V1) or other image features are also generated in ImageJ and uploaded into R. The gene expression levels for a gene of interest are binned into different expression ranges and added as a column to the matrix. Each bin is assigned a color, and these assigned colors are also appended to the matrix. Each cell is then plotted based on its ROI coordinates and colored according to its gene expression bin.

Graphical representation of gene co-expression is generated in R by using the log normalized gene expression values. Correlation coefficient matrices can be plotted alone or combined with ordinary least squares regression lines for each gene pair. For these matrices, the data can be stratified based on group, making it useful for exploratory analysis of differences in gene co-expression between groups.

For co-expression spatial mapping, a local polynomial regression model (LOESS) is fitted to the log normalized gene expression values of two genes of interest. Residual distances are calculated for each cell then binned based on size, and a color code from a red-blue scale is assigned to each bin. This color code is added as a column to the Gene by Cell matrix. Cells are plotted according to their ROI coordinates, and the new matrix is then used to color each cell according to its residual size bin.

Clustering and heatmaps. Log normalized data are hierarchically clustered. Depending on the dataset, different distance measures (Euclidean, Pearson, and Spearman) can be utilized as inputs into the clustering algorithm, and a choice can be made between multiple agglomeration methods (Ward.d2, complete, etc.). Here, we employed Ward.d2 clustering on Spearman correlation coefficient distance measures. Cluster identities are assigned to each cell, and heatmaps are generated using normalized gene expression data. The rows (cells) in the heatmap are organized by hierarchical clusters.

Cellular cluster identities are appended to the Gene by Cell matrix and used in conjunction with ROIs to spatially map cells based on cluster. Cluster identities also can be utilized in conjunction with gene expression levels and cell area to generate violin and boxplots, allowing one to look for gene expression differences between clusters.

Early life stress paradigm

A limited bedding and nesting (LBN) model was used to expose pups in our experimental group to chronic early life stress (ELS) from P2 to P9 (Peña et al., 2017). In brief, each ELS cage is lined with a mesh platform and just enough bedding is placed under the platform to sparsely cover the cage floor. The dam is provided with half of standard nesting material. Combined limited bedding and nesting materials have been shown to act as a stressor on the dam, and consequently leads to fragmented and erratic maternal care of the pups. Care-as-usual (CAU), control litters are raised in cages with normal bedding and nesting resources. At P9, one male and one female mouse were chosen at random from each of 4 different litters and sacrificed for HiPlex experiments.

12 genes that are involved in classifying NG neurons into 18 putative functionally projecting cell-types were compared between groups in these experiments (Kupari et al., 2019)

Manual and cellpose segmentation comparisons

Eight 384 x 344 pixel ROIs, containing a combined total of 307 cells with a mean of 38 cells per ROI, were randomly cropped from different locations in the visual cortex image and were used to assess the time demand and accuracy associated with manual segmentation, automated segmentation using Cellpose, and semi-automated segmentation by combining Cellpose and manual correction.

Time requirement experiments. To determine the time required for manual segmentation, each image was opened in ImageJ, and the amount of time needed to segment each cell (ROI) by hand was determined. For these experiments, each cell was segmented using a stylus on a touchscreen laptop or tablet in combination with the free hand selection tool on ImageJ. For Cellpose segmentation, the amount of time required to load all eight files into the Cellpose Google Colab notebook, to run the program, and to output a ZIP file of all ROI masks was calculated. For the Corrected Cellpose time calculation, Cellpose ROIs were overlayed onto each image in ImageJ, and the amount of time needed to delete inaccurate Cellpose ROIs and redraw boundaries by hand using the free hand selection tool was calculated. To determine the total time required for Corrected Cellpose segmentation, the time to perform ROI corrections was added to the average time for segmentation using Cellpose alone.

Accuracy experiments. Segmentation accuracy was calculated by comparing the ROIs from Cellpose and Corrected Cellpose to the ROIs from the manual segmentation. Using the centroid of each ROI, the nearest Cellpose or Cellpose Corrected neighbor for each Manual ROI was determined in ImageJ. The percentage of overlap was calculated for each neighboring pair. In R, the number of true positive (TPs), false positives (FPs), and false negatives (FNs) was determined at a 70% overlap threshold. The F1 score was calculated by using the following equation:

$$F_1 = \frac{2TP}{2TP + FP + FN}$$

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistics were performed in R and are described in the results section as well as the figure legends. In the segmentation method comparisons, n represents the number of images that were used for the time and accuracy calculations in each group. In the ELS NG comparison experiments, n represents the number of animals per group. For all comparisons, means were compared using a Student's T test and significance was reached at $p < 0.05$.