Supplemental Information

Massively Parallel Single Nucleus Transcriptional

Profiling Defines Spinal Cord Neurons

and Their Activity during Behavior

Anupama Sathyamurthy, Kory R. Johnson, Kaya J.E. Matson, Courtney I. Dobrott, Li Li, Anna R. Ryba, Tzipporah B. Bergman, Michael C. Kelly, Matthew W. Kelley, and Ariel J. Levine
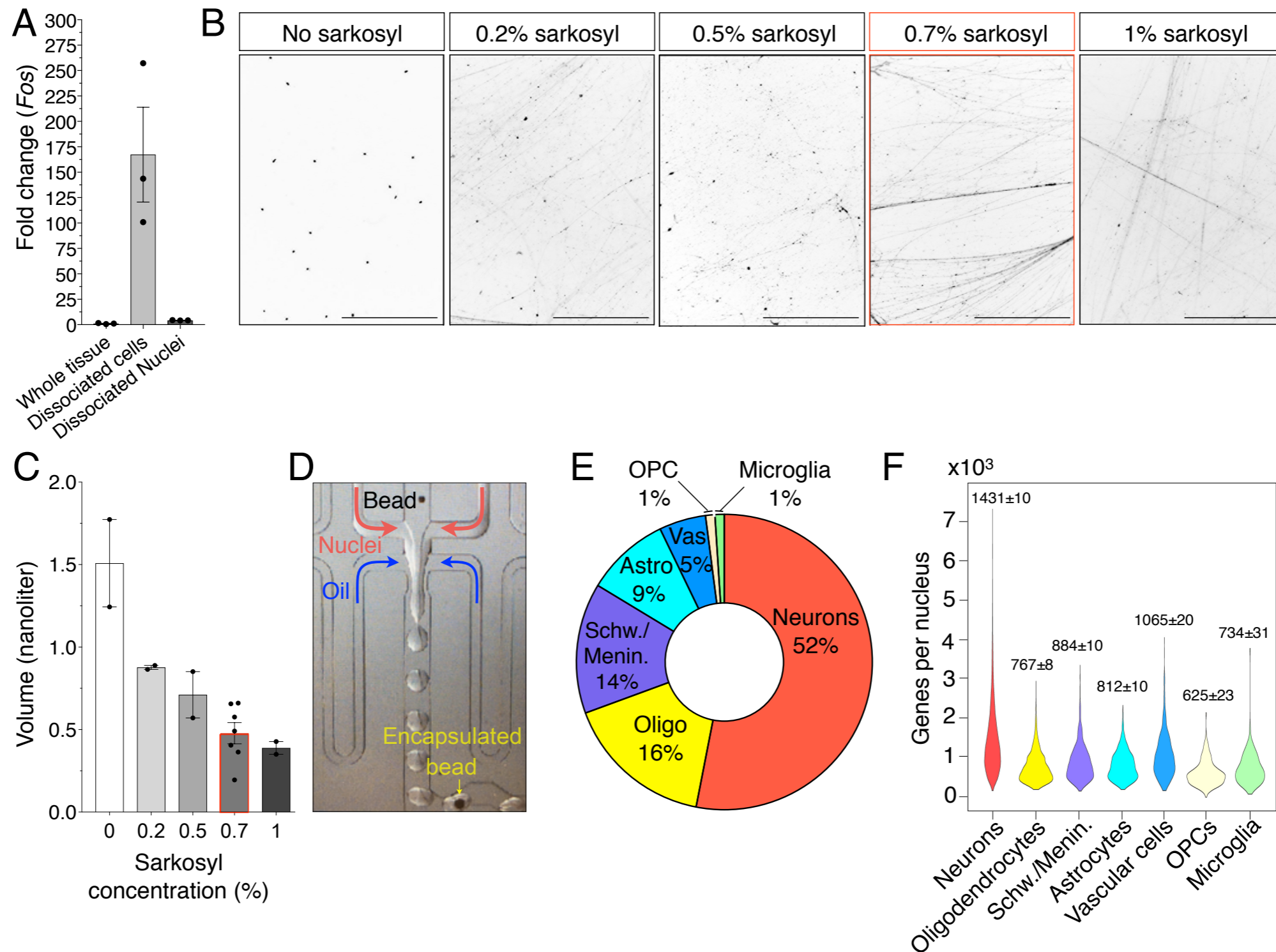
Figure S1. Characterization of the massively parallel single nucleus RNA-seq approach to identify spinal cord cell types. Related to Figure 1. Characterization of the massively parallel single nucleus RNA-seq approach to identify spinal cord cell types. (A) RT-qPCR detection of *Fos* RNA from whole spinal cord tissue, dissociated cells, and dissociated nuclei (normalized to *Gapdh*, performed from three independent biological samples for each). Mean ± s.e.m. are shown. (B) DNA visualized with DAPI stain from isolated nuclei in lysis buffer containing varying concentrations of the detergent sarkosyl. 0.7% sarkosyl was selected for further experiments. Scale bars are 500 μm. (C) Droplet volume with varying concentrations of sarkosyl. Mean ± s.e.m. are shown. (D) Drop-Seq apparatus image showing bead inflow, nuclei inflow, and oil inflow, as well as the formation of droplets and the encapsulation of beads. (E) Pie chart showing the overall contribution of each major cell type to the total population of 17,354 nuclei: Neurons, Oligodendrocytes (Oligo), Meningeal/Schwann cells (Menin./Schw.), Astrocytes (Astro), Vascular cells (Vas), Oligodendrocyte precursor cells (OPC), and Microglia. (F) Genes detected per nucleus, by major cell type. A violin plot distribution is shown, as well as the mean ± s.e.m.
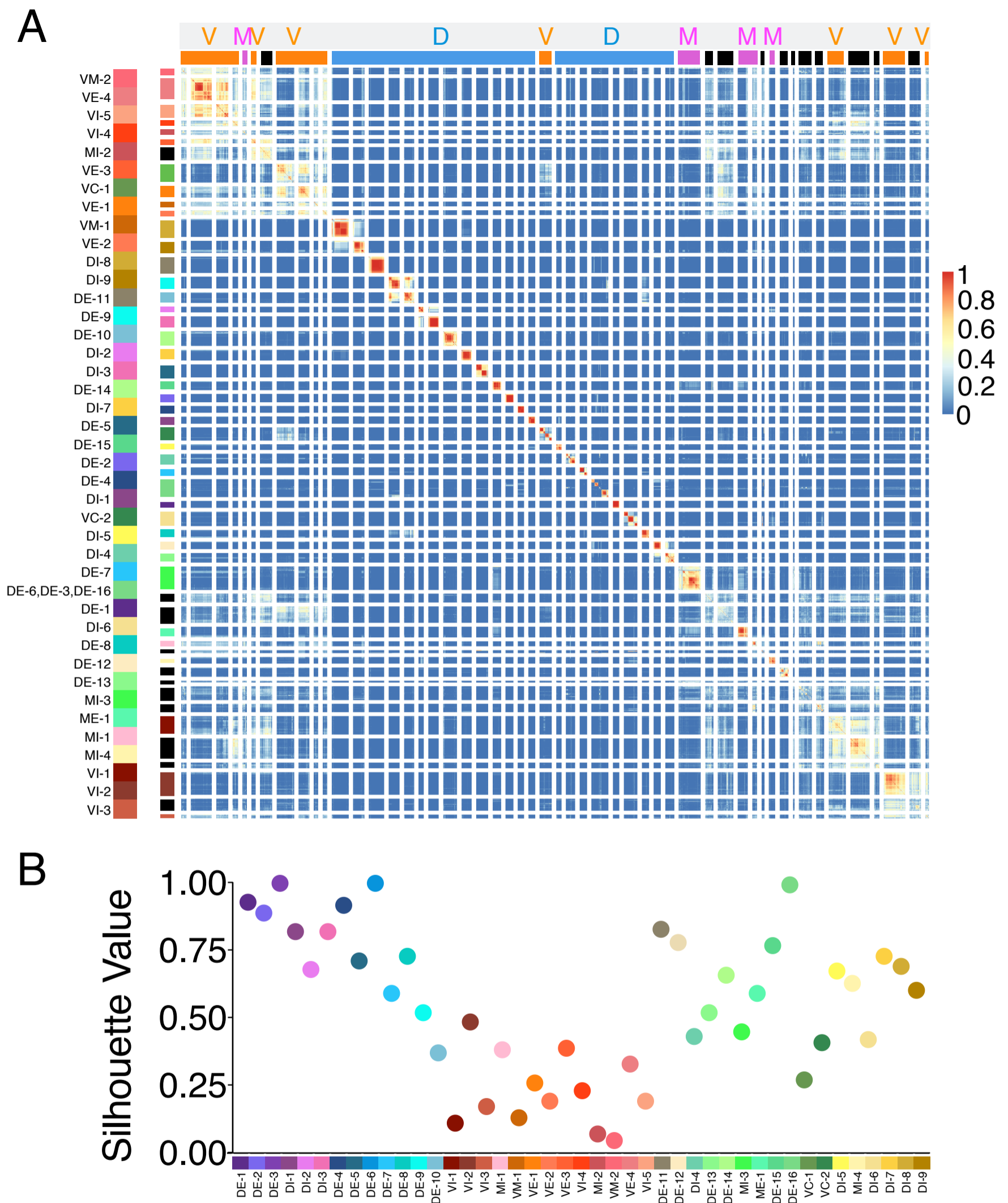
Figure S2. Cluster consensus metrics. Related to Figure 2. (A) Nucleus-by-nucleus consensus matrix based on co-clustering within SC3. Consensus is measured on a scale from 0 (blue) to 1 (red). Fifty-two clusters are shown, including those that were discarded (see Experimental Procedures; black). Clusters DE-3, DE-6, and DE-16 were obtained by sub-clustering the indicated cluster. The spatial location of each cluster is shown on top (D – dorsal, V – ventral, M – intermediate zone/deep dorsal/"mid"). Note that the cluster order is distinct from other figures and is shown on the left. (B) Silhouette width consensus values, plotted by cluster for the final set of forty-three clusters. This measure represents the "diagonality" of the consensus matrix.
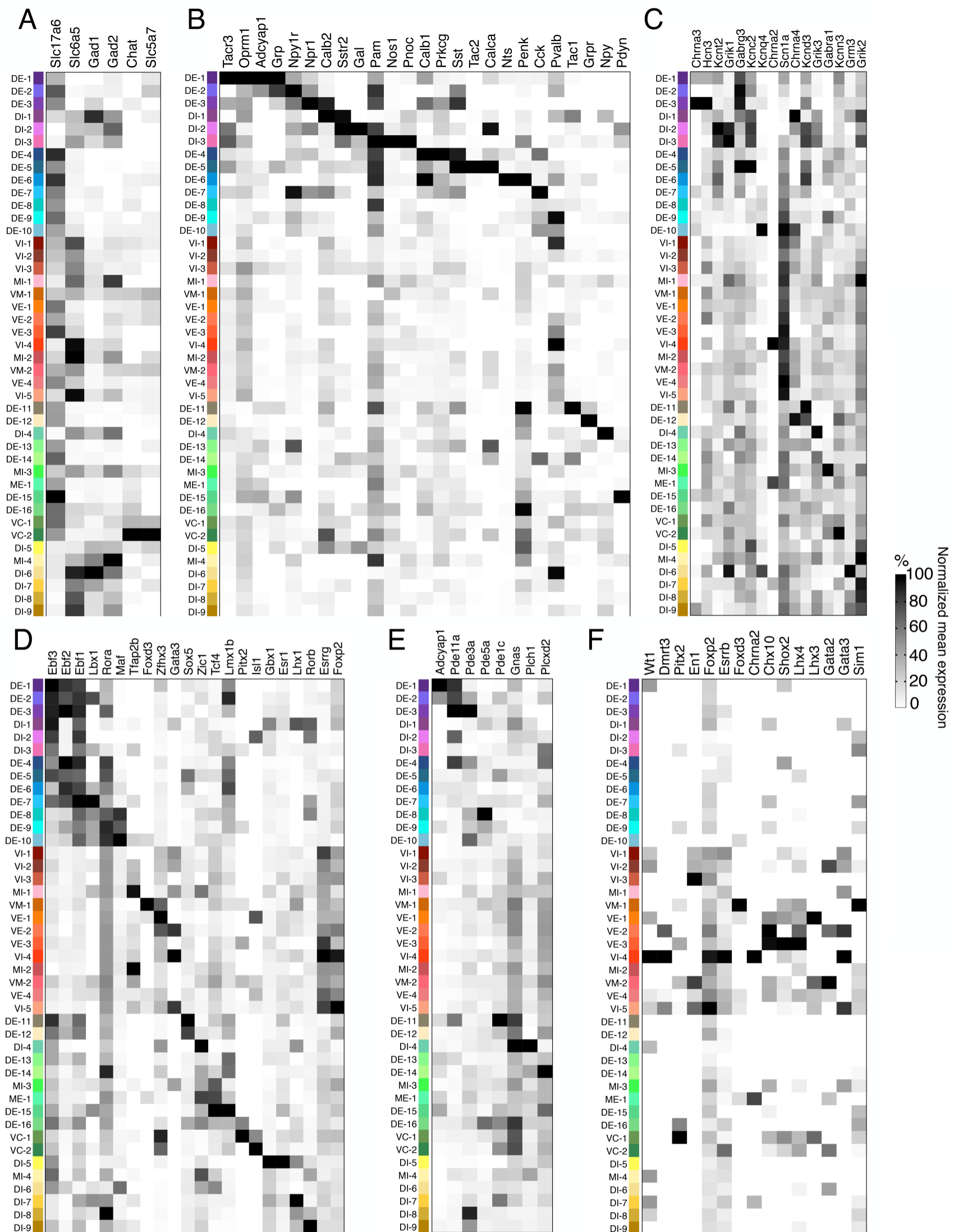
Figure S3. Gene expression across clusters. Related to Figures 2 and 3. (A) Normalized mean gene expression for neurotransmitter markers using excitatory marker *Slc17a6*/vGlut2, inhibitory markers *Slc6a5*/GlyT2 (glycinergic), *Gad1* (gaba-ergic), and *Gad2* (gaba-ergic), and cholinergic markers *Chat* and *Slc5a7*/CHT. (B-F) Normalized mean gene expression across clusters for genes that are classic spinal cord markers (B), neurotransmitter receptors and channel proteins (C), transcription factors (D), cAMP pathway components (defined by GO analysis) (E), and embryonic-lineage domain markers (F).
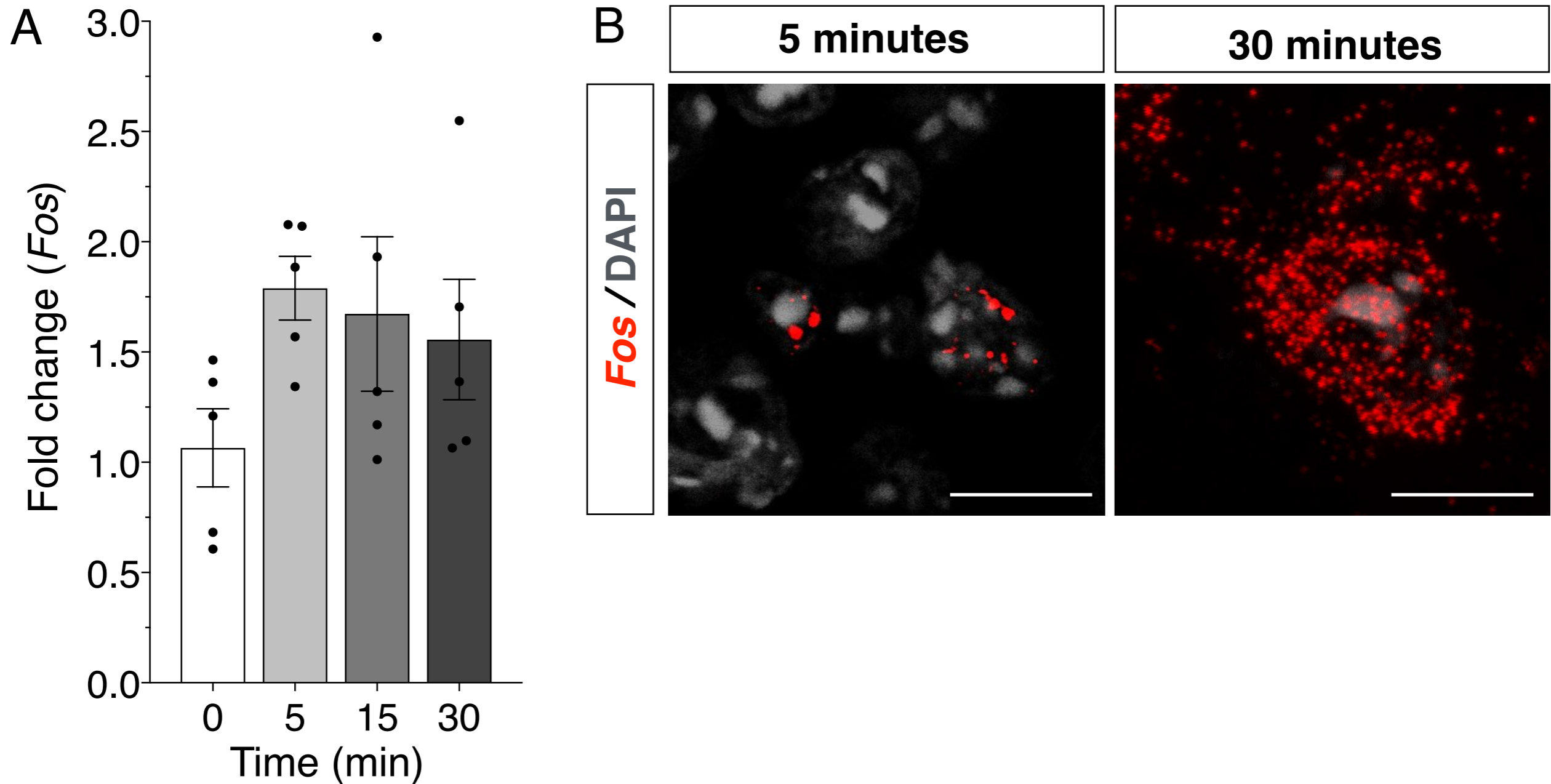
Figure S4. *Fos* mRNA detected in nuclei following formalin administration. Related to Figure 5. *Fos* mRNA detected in nuclei following formalin administration. (A) Fos mRNA as detected by RT-qPCR from bulk nuclei isolated following formalin administration at 0, 5, 15, or 30 minutes. mRNA levels are presented as fold-change from baseline, after normalization to *Gapdh* levels. Five independent biological samples were analyzed for each time point. Mean ± s.e.m are shown. (B) Fos mRNA as detected by fluorescent in situ hybridization at 5 minutes (left) and 30 minutes (right) following formalin administration, showing sub-cellular distribution of mRNA. Scale bars are 10 μm.

SUPPLEMENTAL TABLES

Table S1. Related to Figure 1. All genes significantly associated with non-neuronal clusters.

Table S2. Related to Figure 2. All genes significantly associated with neuronal clusters.

Table S3. Related to Main Text descriptions of clusters and Figures 2-4. Gene expression within each cluster for all detected genes, presented as the mean gene expression of each gene in each cluster and the percent of nuclei in each cluster that express the indicated gene.

Reagent	Source	Catalog	Antibody dilution/RNA	Usage Notes
Cacna1e	ACD	449211	RNA	
cFOS-Gt	Santa Cruz Biotechnology	sc-52-G	1:500	
cFOS-Rb	Santa Cruz Biotechnology	sc-52	1:500	
Chx10	Santa Cruz Biotechnology	sc-8047	1:500	
FOS	ACD	316921	RNA	
Ebf1	EMDMillipore	AB10523	1:500	antigen retrieval
Ebf2	R&D Systems	AF7006	1:500	
Grik3	ACD	493981 (HS)	RNA	
Lbx1	gift of C. Birchmeier	N/A	1:10,000	
Necab1	ACD	428541-C2	RNA	
Nrgn	R&D Systems	MAB7947-SP	01:50	
Npy	ACD	313321-C2	RNA	
PKCg	Santa Cruz Biotechnology	sc-211	1:500	
Rorb	ACD	444271	RNA	
Satb1	Santa Cruz Biotechnology	sc-376096	1:500	
Snca	ACD	313281	RNA	
Syt1	ACD	491831	RNA	

Table S4. Related to Experimental Procedures. Antibodies and RNA probes used in validation.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Single Cell Dissociation. Cells were dissociated using the Miltenyi Neural Dissociation Kit (P) (130-092-628), with manual dissociation according to the manufacturer's protocol.

Bulk Nuclei RT-qPCR. Three samples (one animal each) were used for Figure S1A. Five samples (one animal each) were analyzed for each time point in Figure S4A. Nuclei or cells were prepared as above, then spun at 3,200xg for 10 minutes. The supernatant was removed and 350 μl of buffer RLT (Qiagen RNeasy Mini Kit) was used for lysis. Total RNA was then extracted using the Qiagen RNeasy Mini Kit, including on-column DNase digestion. cDNA (and a no RT control) was prepared using SuperScript IV First-Strand Synthesis System and qPCR (with technical triplicates for each sample) was performed using the following primers. cFOS_F: CTGAGAAGACTGGATAGAGC; cFOS_R: CGTTGAAACCCGAGAACATC; Gapdh_F: AGGTCGGTGTGAACGGATTTG; Gapdh_R: GGGGTCGTTGATGGCAACA.

Drop-Seq: For sample preparation, nuclei in 1mL of PBS with 0.02% BSA were counted with a hemocytometer and adjusted to a final concentration of 225 nuclei/μl. Beads (Chemgenes, Macosko-2011-10, lot 011717C) were prepared at a concentration of 250 beads/μl. Lysis buffer detergent concentration was adjusted to 0.7% sarkosyl. For the apparatus, a PDMS device from FlowJem was used (according to the Drop-Seq CAD design), CorSolutions pumps were used to replace syringe pumps for the cell and oil lines, and the flow was visually monitored using a Photronics High Speed camera (5000 frames per second). The following flow rates were used: beads (35μl/min), nuclei (35 μl/min), oil (200 μl/min). We analyzed individual droplets to characterize the bead occupancy and droplet size that result from these new parameters (n = 581 droplets from 5 samples). We found that 5.7 ± 1% (s.e.m.) of droplets contained a bead, all droplets with beads had a single bead, and the calculated droplet size was 0.46 ± 0.08 nl (s.e.m.). Droplet volume was calculated as the % droplet occupancy divided by 125 beads/μl (which was one half of the initial bead concentration).

For the cDNA amplification and library preparation, multiple PCR reactions were performed for each sample, pooled, purified with AMPure beads (0.6X), quantified with Qubit 3.0, and analyzed on a High Sensitivity DNA Bioanalyzer chip. Multiple tagmentation and library preparation reactions were performed for each sample, each with 600 pg input. Parallel reactions were then pooled, double purified with AMPure beads (0.8X), quantified with Qubit 3.0, and analyzed on a High Sensitivity DNA Bioanalyzer chip. Libraries were sequenced at a minimum depth of 20,000 reads per nucleus using HiSeq 2500 Rapid Mode with the following parameters. Read 1: 25bp, index, custom primer; Read 2: 50bp, Illumina primer.

snRNA-Seq Data Processing. Raw sequence reads (.fastq) per sample were pre-processed, mapped against MM10, and the number of mapped reads per gene enumerated (exon+intron) by cell barcode using commands described in the "Drop-seq Alignment Cookbook" (http://mccarrolllab.com/Drop-Seq/). Quality inspection and filtering of the count data returned was performed using functions supported in the "scater" package (https://bioconductor.org/packages/release/bioc/html/scater.html). First, metrics for

mitochondrial (MT) genes per cell barcode were generated using the calculateQCMetrics() function then inspected using the hist() and scater::plotPhenoData() functions. Among the metrics generated per cell barcode was the percent total reads mapped to MT genes. This specific metric was next used to filter-discard nuclei with a percent total reads mapped to MT genes ≥ 20%. For surviving cell barcodes/nuclei, MT read counts and immediate early genes were filter-removed and metrics generated for remaining genes by nuclei using the same calculateQCMetrics() function. Among these metrics generated was the total number of genes with mapped reads per nuclei and the total number of mapped reads across genes per nuclei. These two metrics were inspected using the same hist() and scater::plotPhenoData() functions used prior and threshold filters defined to discard outlier nuclei. Specifically, nuclei having ≤ 200 genes per nucleus were discarded. During sub-clustering of the neuronal populations, nuclei having ≤ 500 genes per nucleus were discarded.

snRNA-Seq Clustering and Analysis. For clustering, the "SC3" package (https://bioconductor.org/packages/release/bioc/html/SC3.html) was used on normalized expression for nuclei not discarded in these two successive steps. The normalization method used was "Counts Per Million"; also referred to as "CPM". All genes in ≥ 3 nuclei were analyzed and the gene filter was set to "false". The first clustering in SC3 step involved ten clusters, while seven clusters were considered further. One cluster was excluded because it was considered to contain low quality nuclei and was not defined by any cell-type specific genes, but rather by genes from a list of the twenty genes with the most mapped reads in the entire set. Two small clusters were discarded because they contained < 0.5% of the total nuclei. We verified that all clusters contained nuclei from at least three biological samples. To visualize SC3-derived clusters of nuclei by t-SNE scatterplot, CPM expression of detected genes expressed across nuclei were passed to the Rtsne() function under default settings and the results returned passed to the plot() function. The second-clustering step was performed on the neuronal nuclei (cluster 1) obtained by SC3 analysis of the total set of nuclei. Preliminary SC3 analysis identified a cluster of low quality nuclei and a cluster that represented doublets between nuclei from neurons and oligodendrocytes and both of these clusters were discarded from further analysis. SC3 analysis of the remaining nuclei then identified fifty-two clusters. The complete fifty-two cluster set can be seen in Figure S2A. This set of clusters was determined through an iterative process of clustering, followed by visual inspection of gene lists and comparison to established spinal cord markers. At this step, clusters that represented doublets, that did not have at least three significantly associated annotated genes, or that contained less than 0.7% of the total neuronal nuclei were discarded from further analysis. These discarded clusters are indicated by black bars in Figure S2A. After these exclusions, there were forty-one remaining clusters. One cluster (#30 in the original analysis) was further sub-clustered, following inspection of the tSNE plot and cell consensus matrix, which suggested three sub-groups that were not related to each other. Thus, a final list of forty-three clusters was analyzed. All of these clusters had nuclei from at least three independent biological samples. To assess relatedness of clustered nuclei, the mean expression per gene (with CPM expression > 0) per cluster was calculated in R and then passed to the dist() function to produce a Euclidean-based distance metric for each pair-wise combination of clusters. These distance metrics were then used in the performance of hierarchical clustering using "complete" agglomeration via the hclust() function. The

dendrogram describing the results of how clusters of nuclei are related was visualized via the as.phylo() function. GO term analysis was performed using GO David Functional Annotation Tool with the top ten most significant genes for each cluster, using functional annotation clustering on "molecular function" GO terms with low stringency. Functional annotation groupings with an enrichment score > 1.3 (-log of mean p-value) are presented.

Immunofluorescence. Animals were euthanized with avertin either at baseline or one hour after behavior and perfused with PBS and 4% paraformaldehyde. Lumbar spinal cords were extracted and washed in PBS, incubated overnight in 30% sucrose, and embedded in OCT. Frozen sections of 50 μm were cut and stained, using a blocking buffer that contained 1% IgG-free BSA, 10% donkey serum, and 0.1% Triton-X 100 in PBS. Antibodies used are presented in Table S4. At least eight sections from at least two male and two female animals, ages 8-12 weeks old, were analyzed for expression patterns and a representative example was imaged by confocal microscopy (Zeiss 800 LSM).

In Situ Hybridization. Animals were euthanized by cervical dislocation at baseline or thirty minutes after behavior and the lumbar spinal cords were extracted and fresh frozen in OCT. Frozen sections of 16-18 μm were cut and RNAscope (ACD) was used for in situ hydridization using the manufacturer's protocol. Probes used are presented in Table S4. At least five sections from at least two male and two female animals, ages 8-12 weeks old, were analyzed for expression patterns and a representative example was imaged by confocal microscopy (Zeiss 800 LSM).