Expanded View Figures

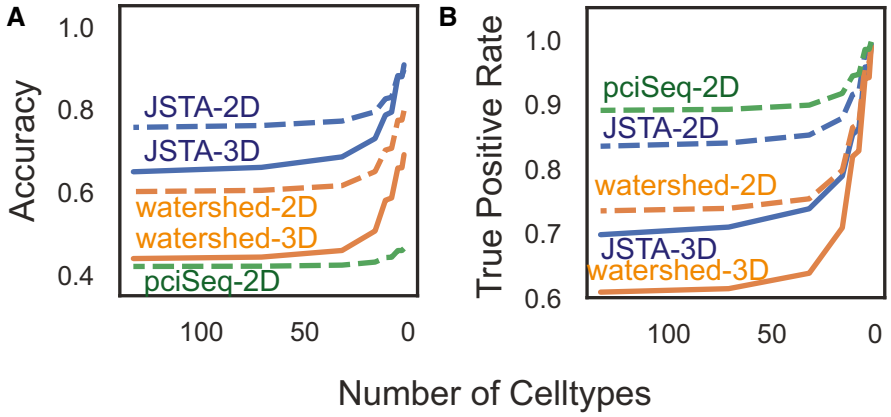

Figure EV1. Performance evaluation of JSTA, pciSeq, and watershed.

A, B pciSeq is unable to run on 3D data (solid line), so we simulated additional 2D data (dotted line). We evaluated these methods on the performance of accuracy of assigning mRNA to the correct cell (A). JSTA is more accurate than pciSeq on the accuracy metric. pciSeq is not very accurate here, because many mRNA are incorrectly assigned to background. We additionally tested these methods on their performance of assigning mRNA to the correct cell while ignoring mRNA assigned to background (B). pciSeq is highlighted here, because it mainly assigns spots close to the nucleus; JSTA is comparable.

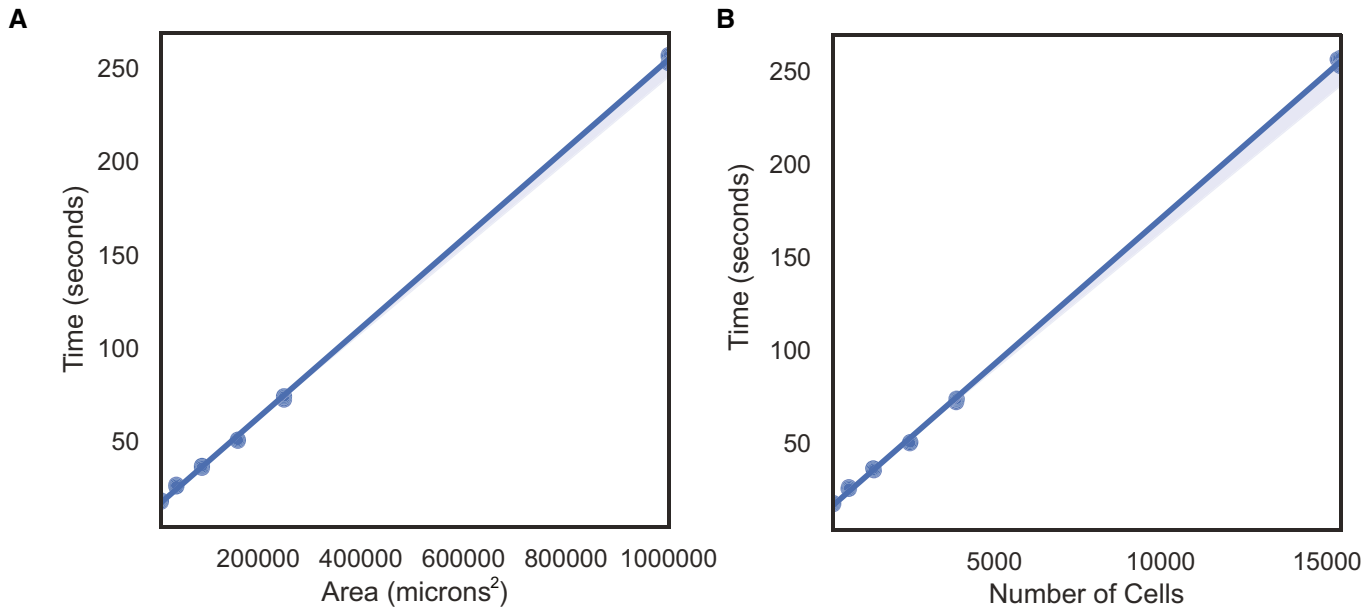

Figure EV2. Application of JSTA to osmFISH data from the mouse somatosensory cortex.

A Glutamatergic neurons are consistent with previously identified spatial patterns of the somatosensory cortex.

B JSTA-mapped high-resolution (sub)types are correlated with their NCTT counterparts in terms of gene expression patterns (Table EV4). Cell types with at least five cells were kept.

Figure EV3. Run time evaluation of JSTA on simulated data.

A, B We ran JSTA on data simulated with a width and height of 100, 200, 300, 400, 500, and 1,000 μm, with three replicates each. We evaluated the time taken to run JSTA by the area of the section (A), and the number of cells in each section (B).

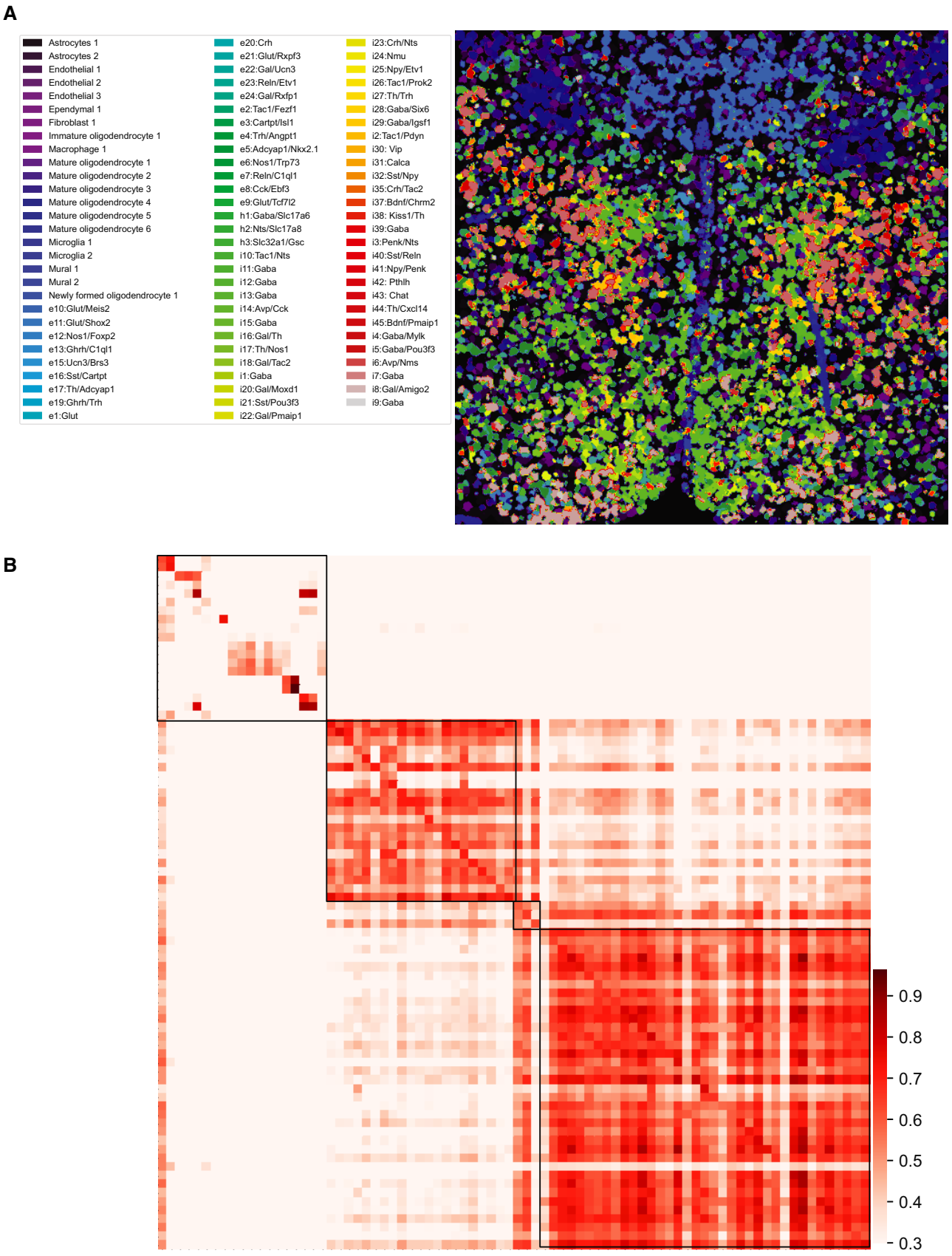

Figure EV3.

Figure EV4. Application of JSTA to MERFISH data from the mouse hypothalamic preoptic region.

A High-resolution cell types identified by JSTA. The spatial mappings of these high-resolution cell types are consistent with the manually annotated data from Moffit *et al* (2018).

B JSTA-mapped high-resolution (sub)types are highly correlated with their scRNAseq reference counterparts in terms of gene expression patterns (Table EV3). Cell types with at least five cells were kept.

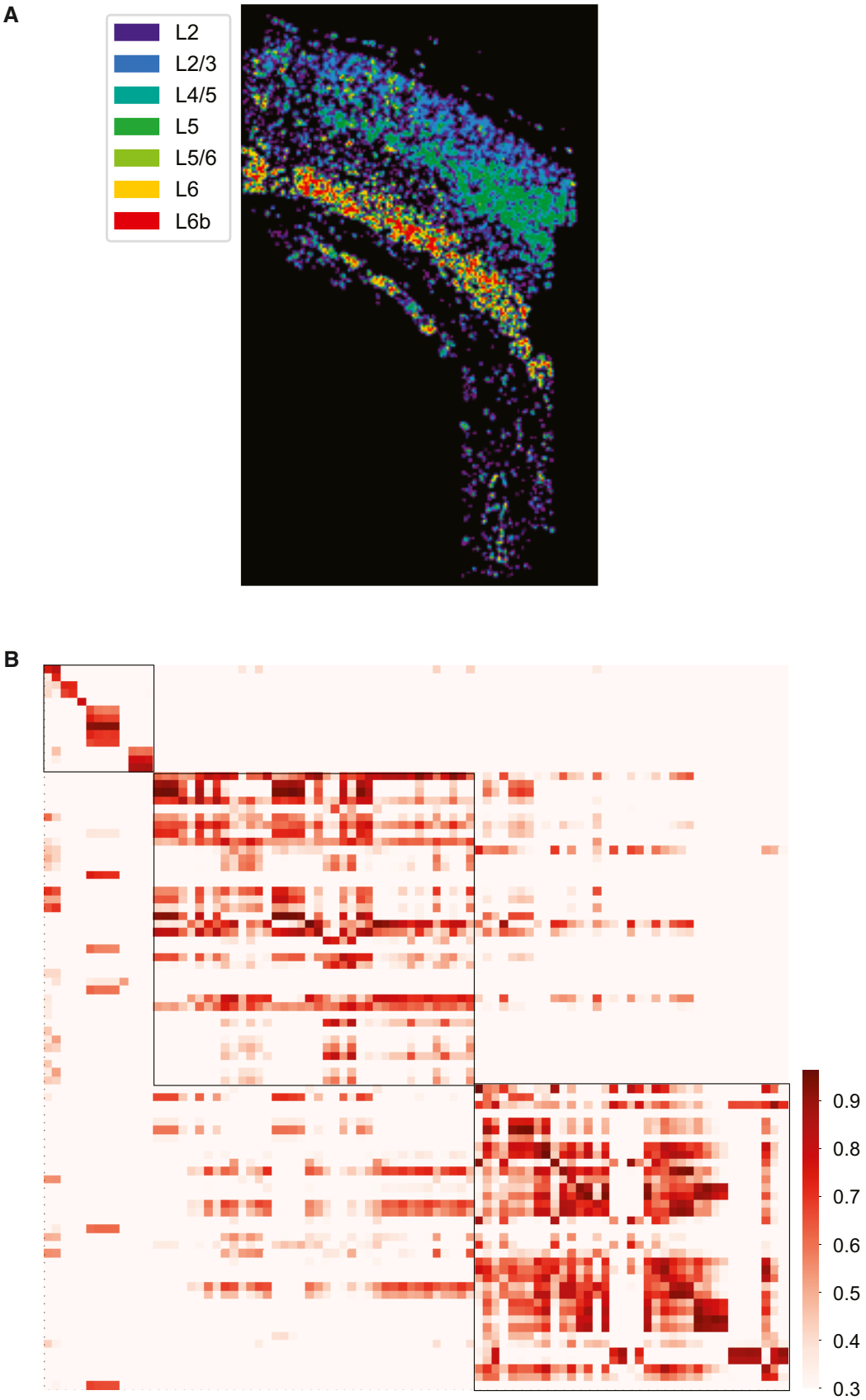

Figure EV4.

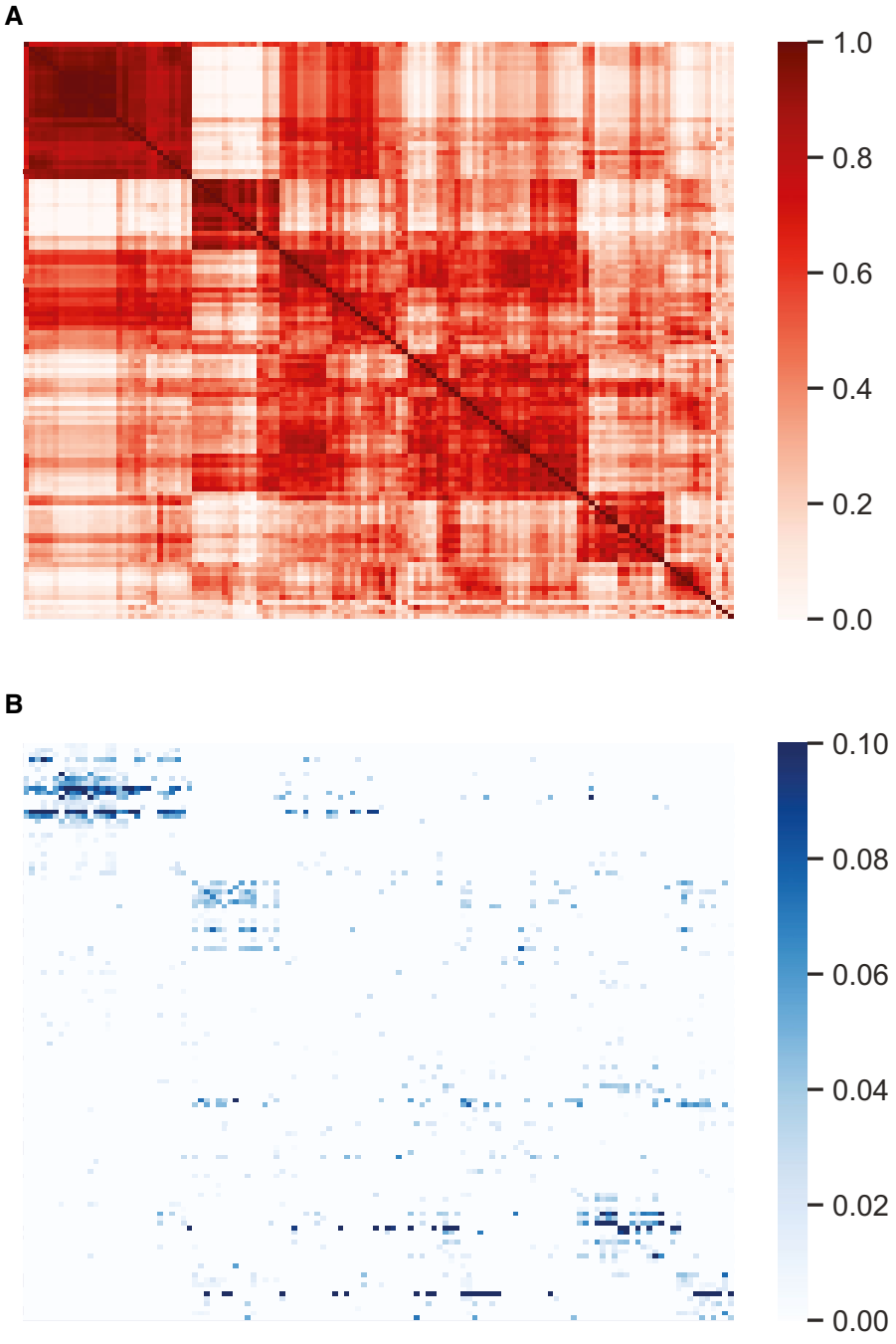

Figure EV5. Correlation structure of cell types compared with their colocalization.

Neuronal subtypes that are highly colocalized are often correlated in their gene expression. Cell types with more than 10 cells were included.

A Pearson correlation of 122 (sub)types across 83 selected genes.

B Frequency of neighbors between each of 122 (sub)types. Only significant (FDR < 0.05) colocalizations are shown. Labels and values are detailed in Tables EV5 and EV6.

Figure EV6. Identification of spatial differentially expressed genes (spDEGs).

spDEGs were computed by comparing the true variance in gene expression between cell subtype neighborhoods to that of randomly permuted cell (sub)type neighborhoods.

A 63 genes across 61 cell types show significant spDEGs. Heatmap values correspond to $-\log_2(P\text{-value})$.

B Number of spDEGs in each of the 61 cell types.

C Number of cell types with each of the 63 spDEGs.

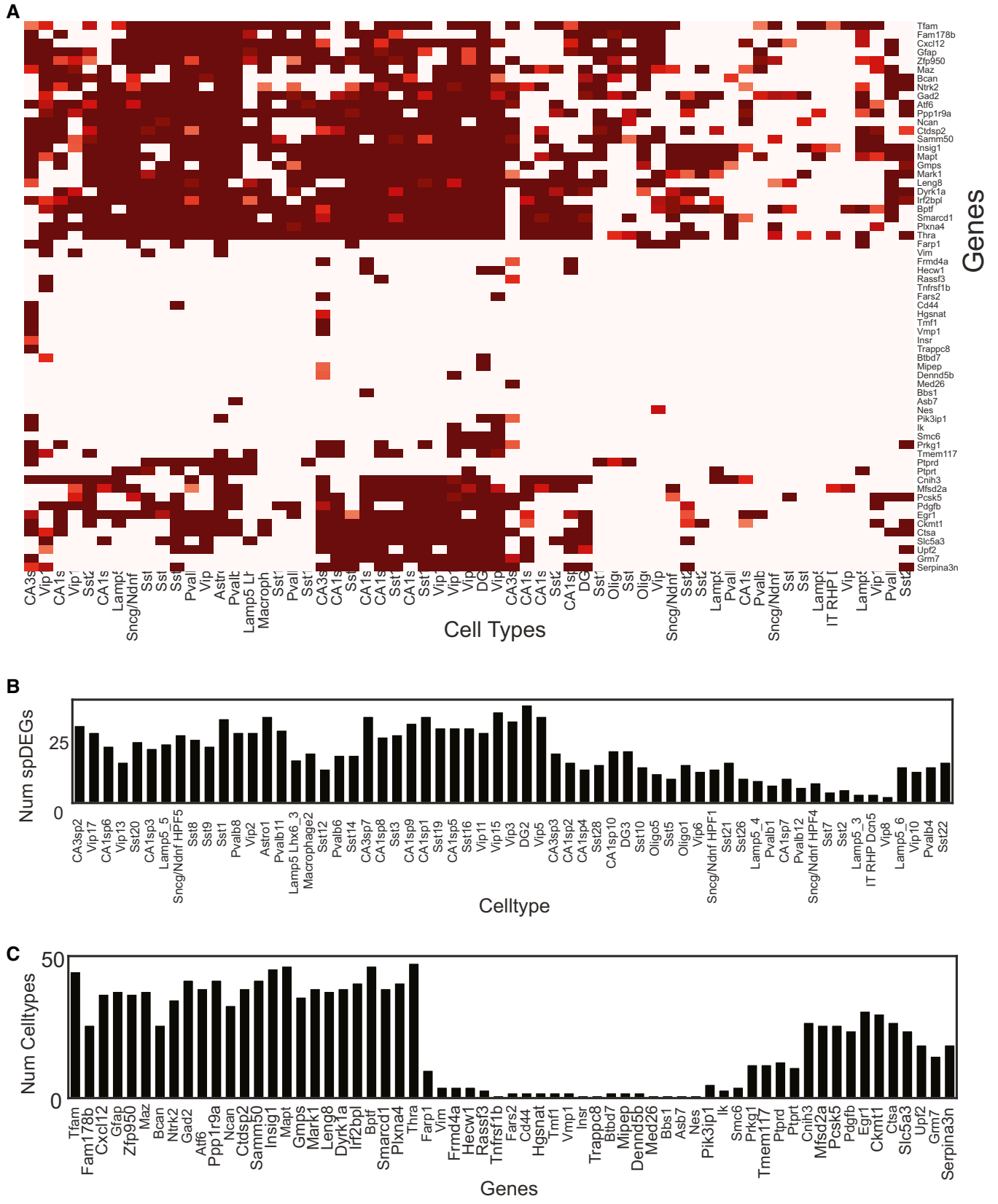
A
Genes
Cell Types
B
Num spDEGs
25
0
Celltype
C
Num Celltypes
50
0
Genes

Figure EV6.

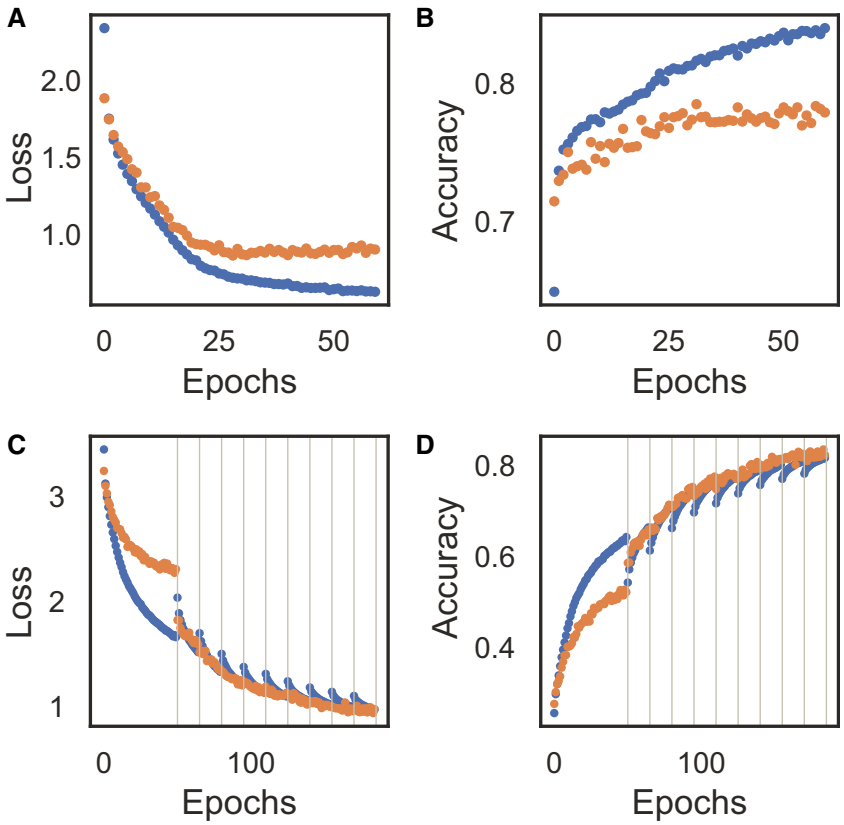

Figure EV7. Cross-entropy loss and accuracy of cell type (A, B) and pixel (C, D) classifier during training for the train (blue) and validation (orange) datasets.

A, B Cross-entropy (A) loss and accuracy (B) during training cell type classifier. The cell type classifier overfits the training data and is mitigated by stopping training after 40 epochs.

C, D Cross-entropy loss (C) and accuracy (D) during training of the pixel classifier. Black lines indicate new training iteration after pixel reassignment.