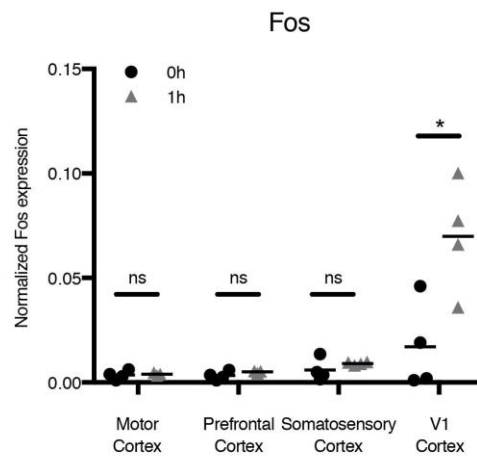


In the format provided by the authors and unedited.

Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex

Sinisa Hrvatin¹, Daniel R. Hochbaum^{2,3}, M. Aurel Nagy¹, Marcelo Cicconet⁴, Keiramarie Robertson³, Lucas Cheadle¹, Rapolas Zilionis^{5,6}, Alex Ratner⁷, Rebeca Borges-Monroy⁸, Allon M. Klein⁵, Bernardo L. Sabatini^{3*} and Michael E. Greenberg^{1*}

¹Department of Neurobiology, Harvard Medical School, Boston, MA, USA. ²Society of Fellows, Harvard University, Cambridge, MA, USA. ³Department of Neurobiology, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. ⁴Image and Data Analysis Core, Harvard Medical School, Boston, MA, USA. ⁵Department of Systems Biology, Harvard Medical School, Boston, MA, USA. ⁶Vilnius University Institute of Biotechnology, Vilnius, Lithuania. ⁷ICCB-L Single Cell Core, Harvard Medical School, Boston, MA, USA. ⁸Program for Bioinformatics and Integrative Genomics, Graduate School of Arts and Science, Division of Medical Sciences, Harvard University, Cambridge, MA, USA. Sinisa Hrvatin, Daniel R. Hochbaum and M. Aurel Nagy contributed equally to this work. *e-mail: bernardo_sabatini@hms.harvard.edu; michael_greenberg@hms.harvard.edu

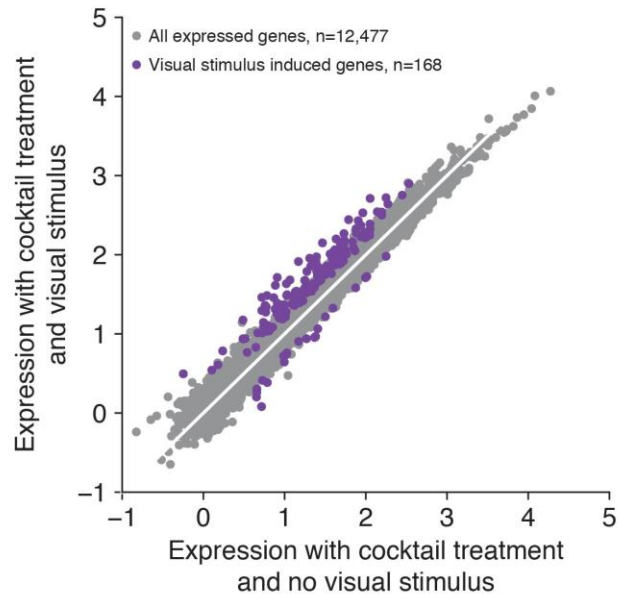


Supplementary Figure 1

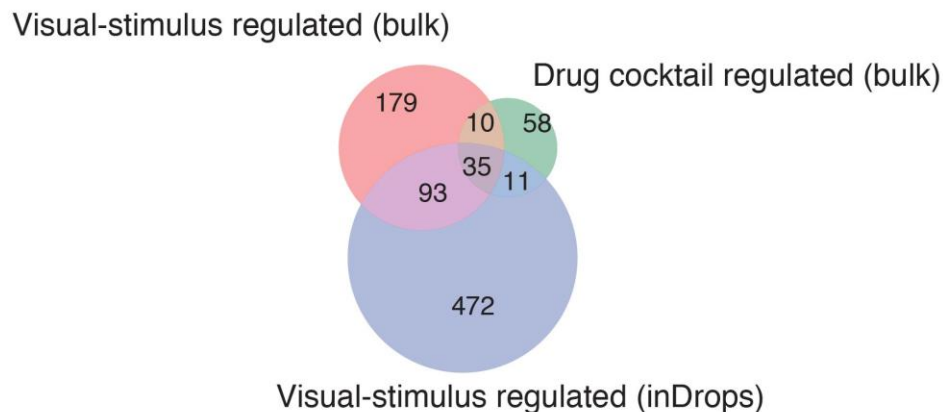
Fos induction by visual stimulus is restricted to V1

Fos expression measured by qRT-PCR at 0 and 1 h conditions following inDrops sample preparation in 4 brain tissues. * $p=0.021$, unpaired t-test, two-sided. $n=4$ animals for all 0 h samples, $n=3$ animals for 1 h motor and prefrontal cortex, $n=4$ animals for 1 h somatosensory and visual cortex. Bar indicates mean.

a



b

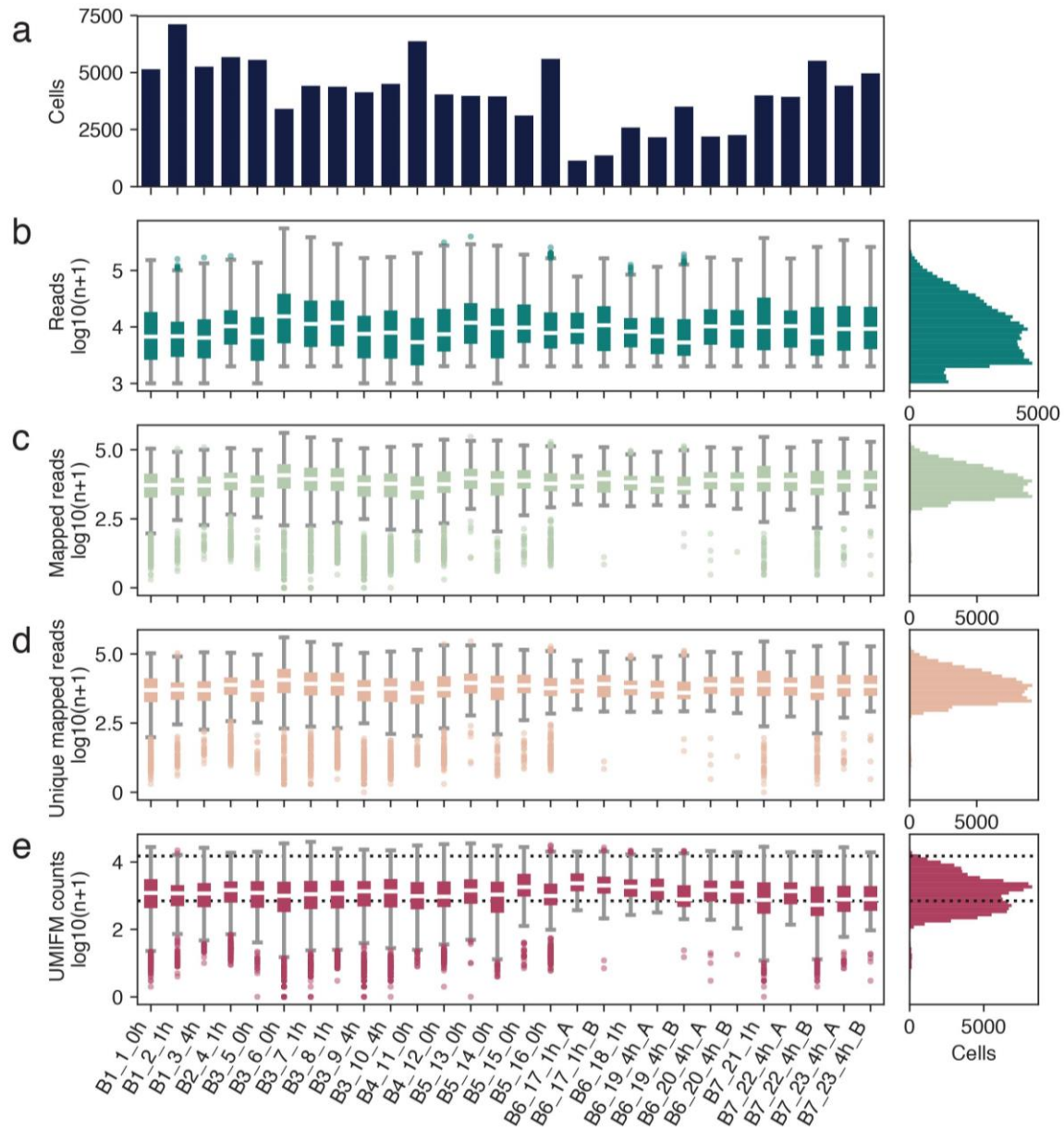


Supplementary Figure 2

Drug cocktail does not grossly alter transcription in visual cortex as measured by bulk RNA-seq

(a) Scatter plot of gene expression (log10 [TMM-normalized CPM + .01]) for 12,477 genes in visual cortex from drug-cocktail-perfused animals that were dark- (0 h) or light- (1 h) exposed (n=4 animals for each condition). 168 significantly differentially expressed genes (DEG) corresponding to the bulk visual experience-dependent program are denoted in purple (FDR corrected q-value<.01, fold change >=2, limma, Online Methods).

(b) Venn diagram demonstrating the overlap of the following three gene sets (in clockwise order): DEG from panel A, DEG from **Fig. 1d**, and the 611 cell-type-specific DEG from **Fig. 2b**. Approximately one half of genes (56/114) that are repressed by the cocktail were identified as visual stimulus induced in either the bulk or single-cell RNA-seq experiments. GO analysis results for all 114 genes are listed in **Supplementary Table 1**.



Supplementary Figure 3

inDrops sample sequencing metrics

The samples described in the manuscript comprise 28 libraries collected across 23 animals denoted by the sample names along the x-axis. All panels in this figure represent sequencing metrics across all cells with $\geq 1,000$ total reads (cells=114,601). For each metric, whisker plots demarcate quartiles. Data outside 1.5-times the interquartile range ($Q3-Q1$) were labeled as dots. The median of each distribution is marked in white (left). Each distribution is also aggregated across all samples and this histogram is plotted.

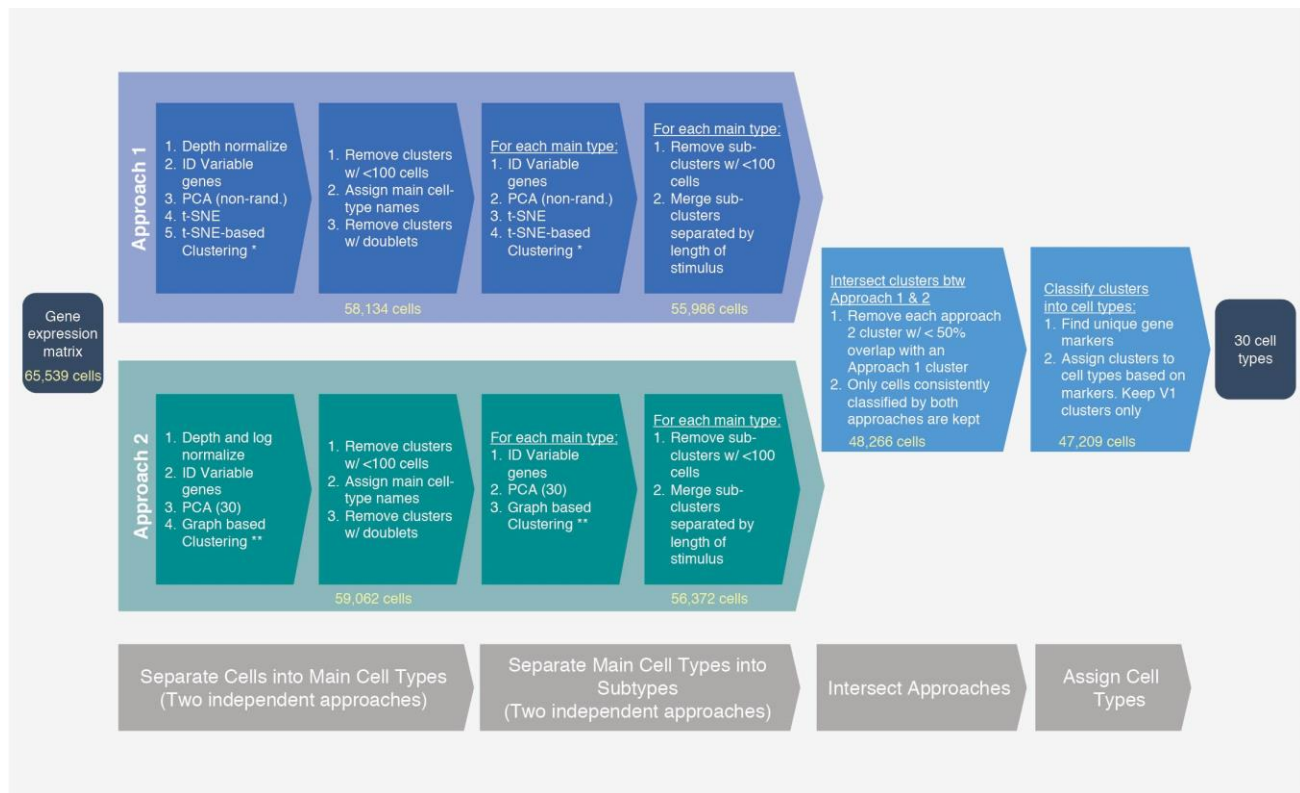
(a) Number of cells collected by sample.

(b) Total reads per cell by sample.

(c) Total mapped reads per cell by sample.

(d) Reads mapping to a unique genomic location per cell by sample.

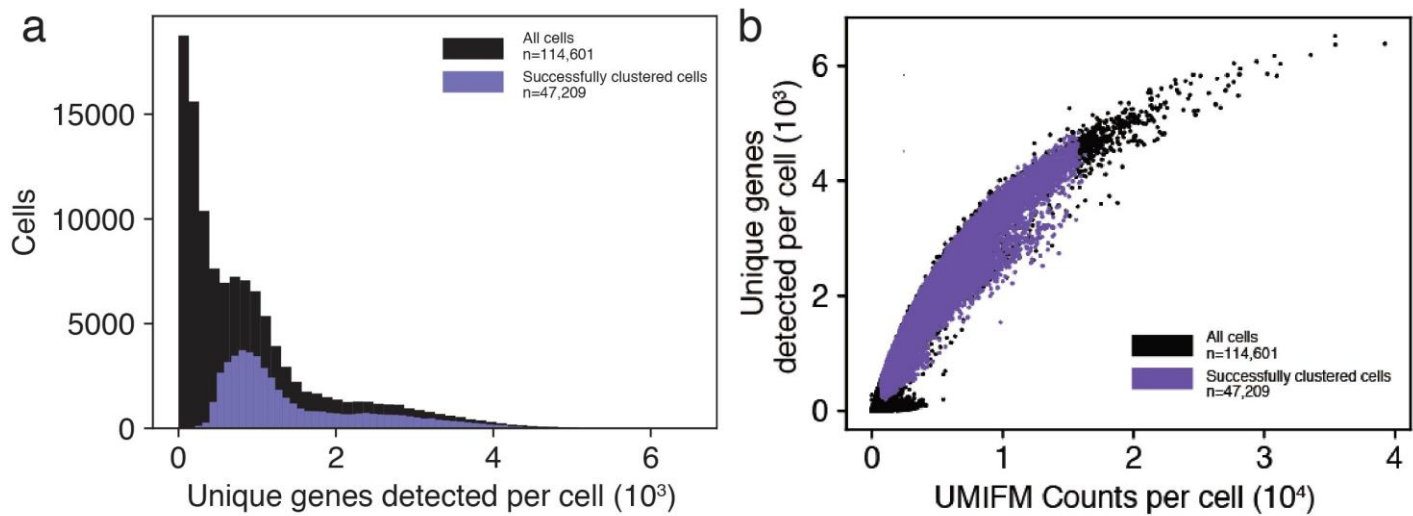
(e) UMI filtered, uniquely mapping counts per cell by sample.



Supplementary Figure 4

Flowchart of bioinformatic clustering and cell type assignment

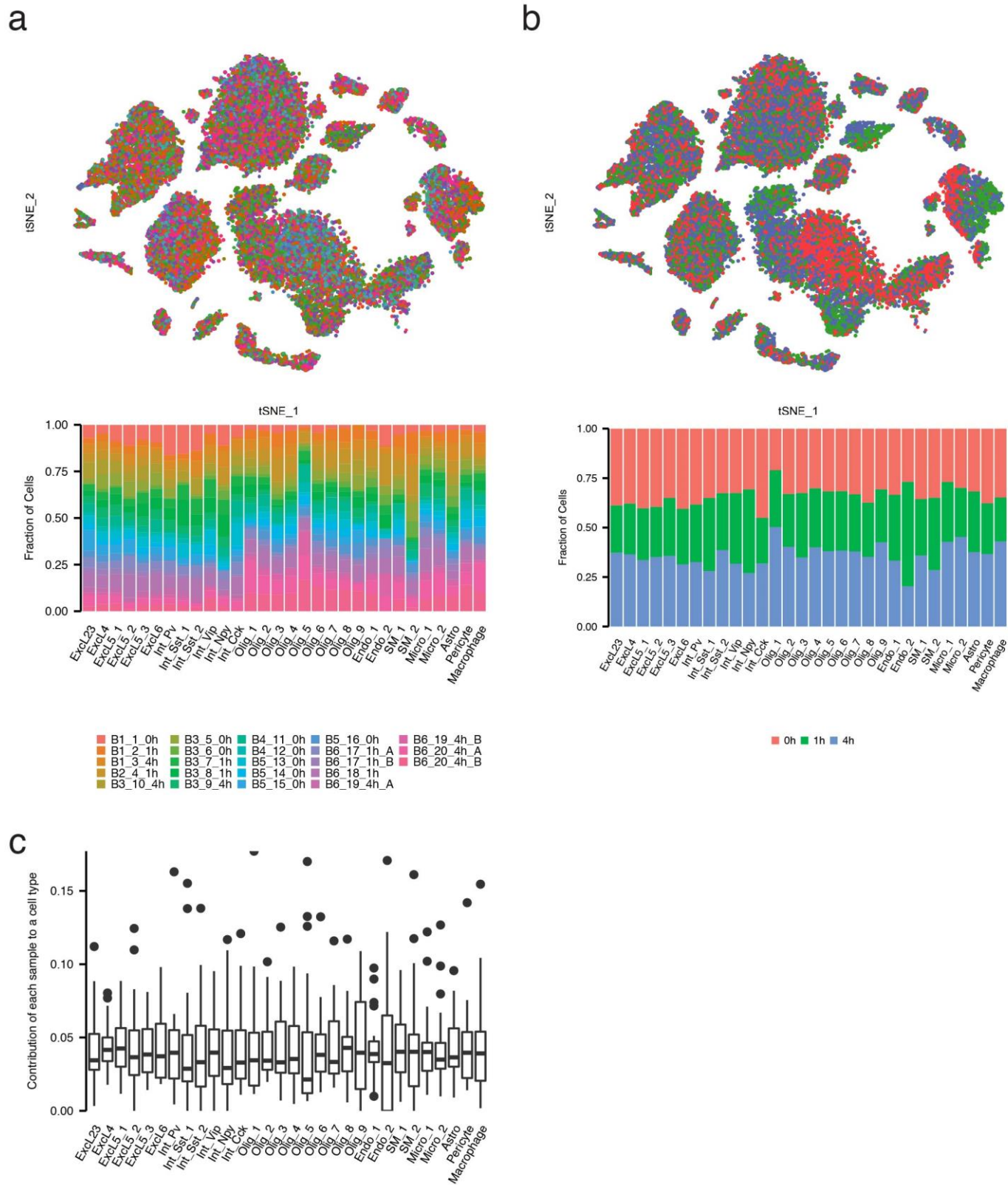
Workflow diagram showing bioinformatic clustering and cell type assignment pipeline (Online Methods). The number of cells passing each stage of the pipeline are indicated in light yellow.



Supplementary Figure 5

inDrops sample gene detection metrics

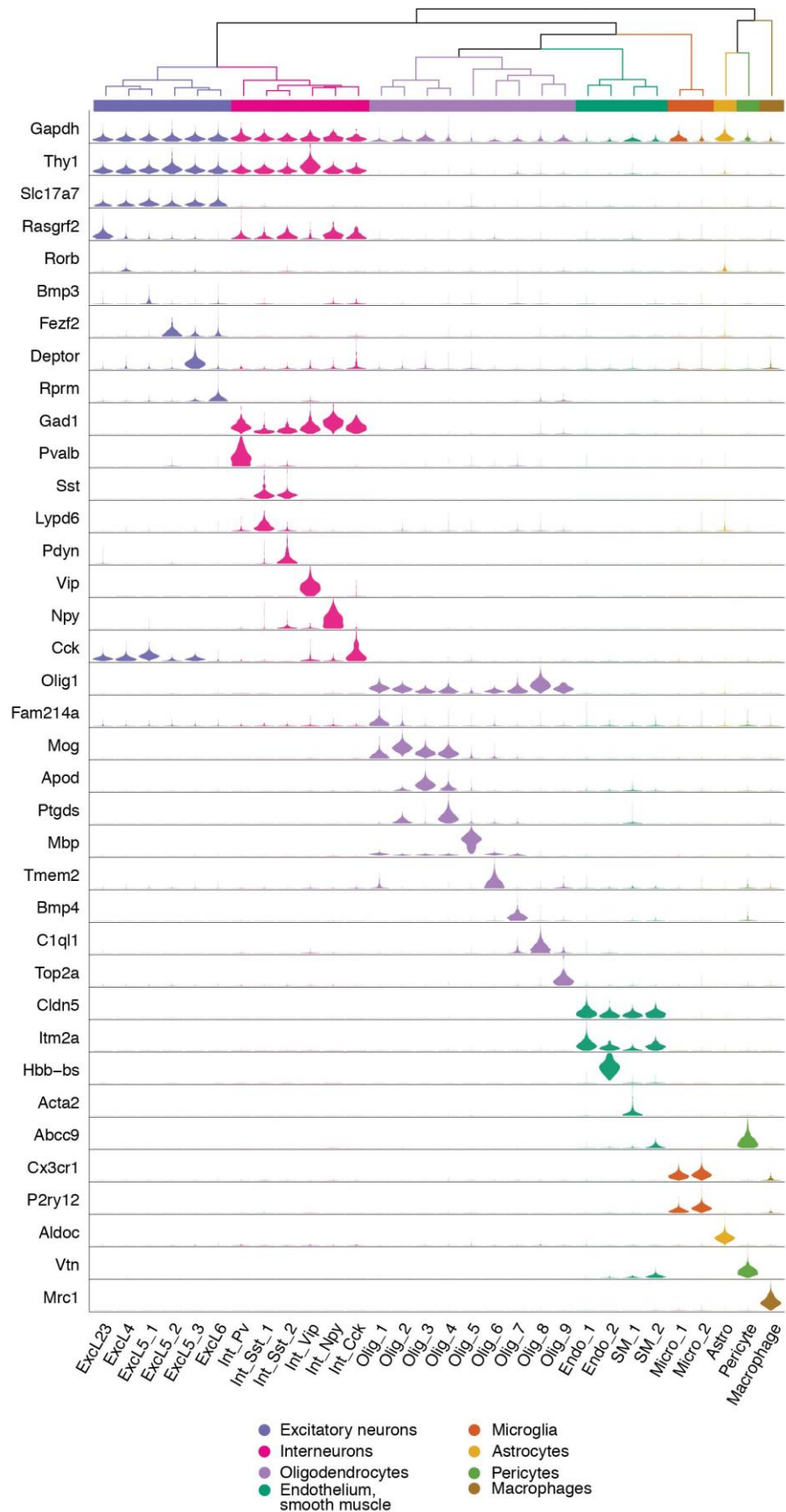
- (a) The number of unique genes detected per cell for all cells acquired with $\geq 1,000$ reads per cell (cells=114,601, black), and for the subset of cells that were successfully assigned a cell type designation (cells=47,209, purple) as a histogram.
- (b) Same data plotted on a scatter plot showing unique genes per cell as a function of UMIFM counts per cell.



Supplementary Figure 6

Samples and stimulus conditions are evenly distributed across cell types

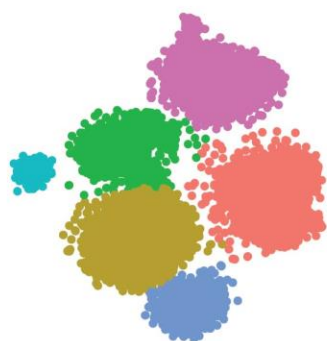
- (a) Distribution of 23 samples across 30 cell types shown on top of total t-SNE (top) and denoted as fraction of cells for each cell type (bottom).
- (b) Same plots showing distribution of stimulus paradigm across 23 samples and 30 cell types.
- (c) Quantification of the distribution shown in (a) Box plot with median (Tukey). Data outside 1.5-times the interquartile range ($Q3-Q1$) were labeled as dots.



Supplementary Figure 7

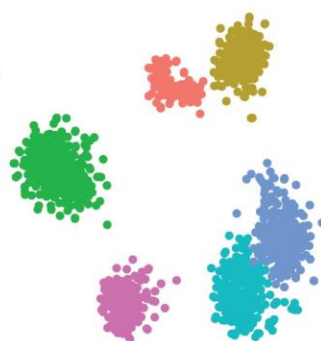
Expression of marker genes across all cell types

Dendrogram of cell types and violin plots showing the distribution of expression of selected marker genes across all 30 analyzed cell types. Cell types are hierarchically clustered based on gene expression across all variable genes.



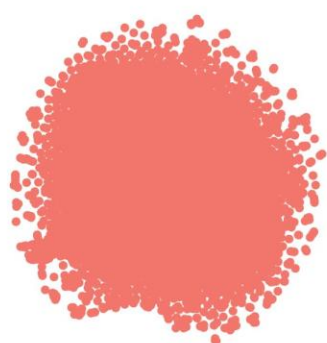
Excitatory Neurons

- Excl23
- Excl4
- Excl5_1
- Excl5_2
- Excl5_3
- Excl6



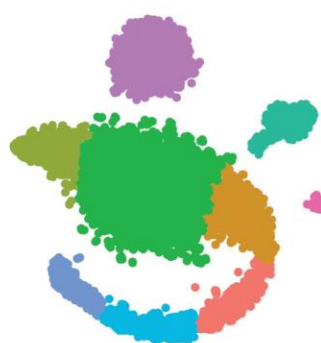
Interneurons

- Int_Cck
- Int_Npy
- Int_Pv
- Int_Sst_1
- Int_Sst_2
- Int_Vip



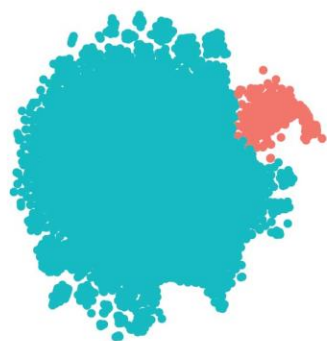
Astrocytes

- Astro



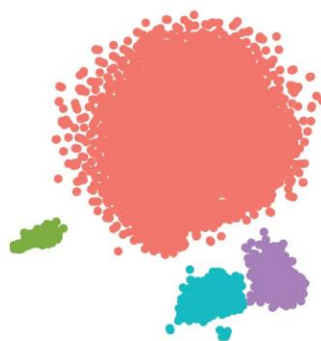
Oligodendrocytes

- Olig_1
- Olig_2
- Olig_3
- Olig_4
- Olig_5
- Olig_6
- Olig_7
- OPC_1
- OPC_2



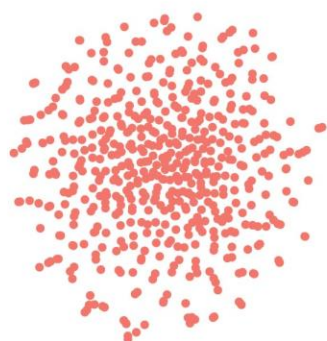
Microglia

- Micro_1
- Micro_2



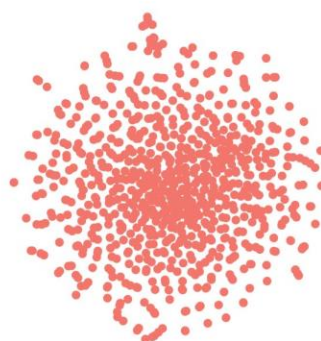
Endothelium, smooth muscle

- Endo_1
- Endo_2
- SM_1
- SM_2



Macrophages

- Macrophage



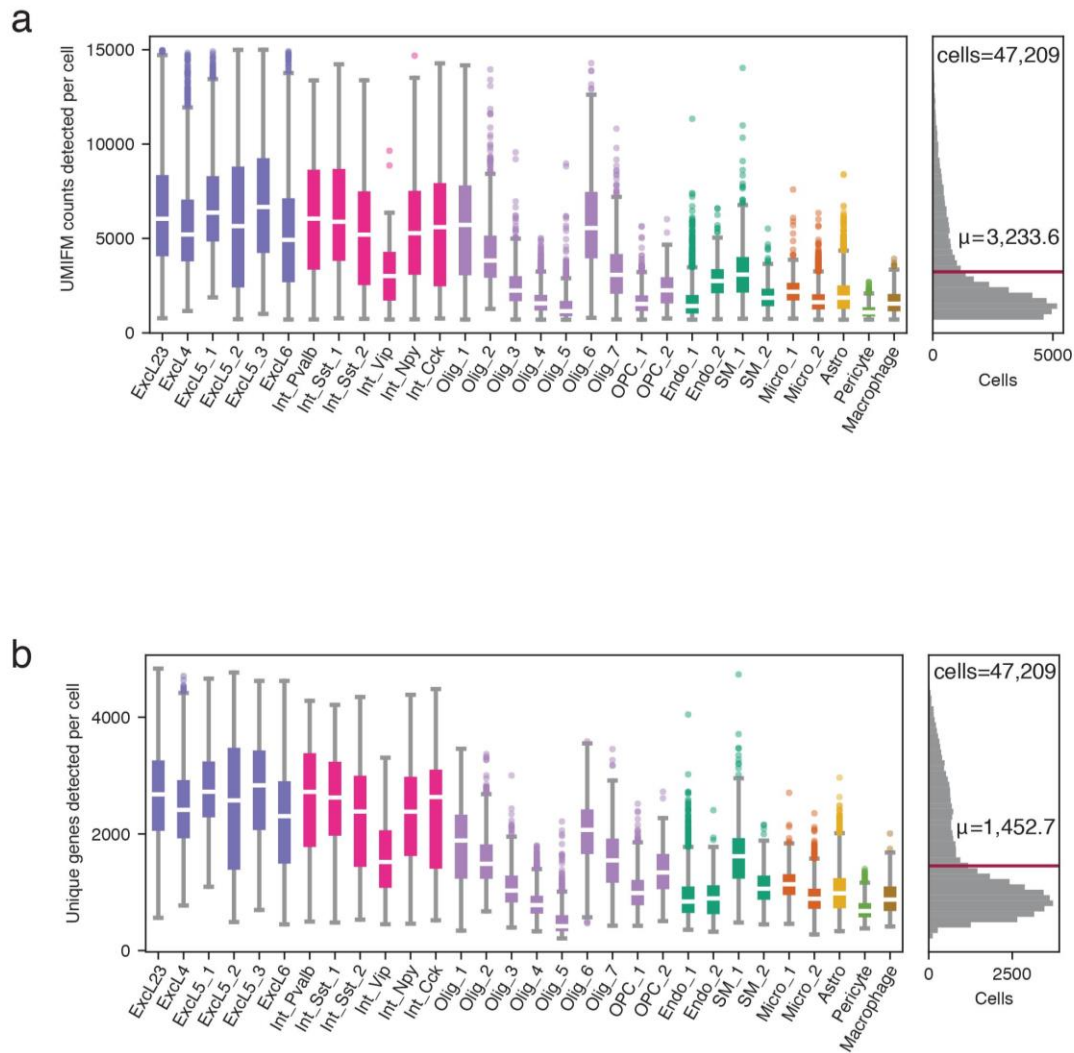
Pericytes

- Pericyte

Supplementary Figure 8

t-SNE plots separating cell types within each main cell class

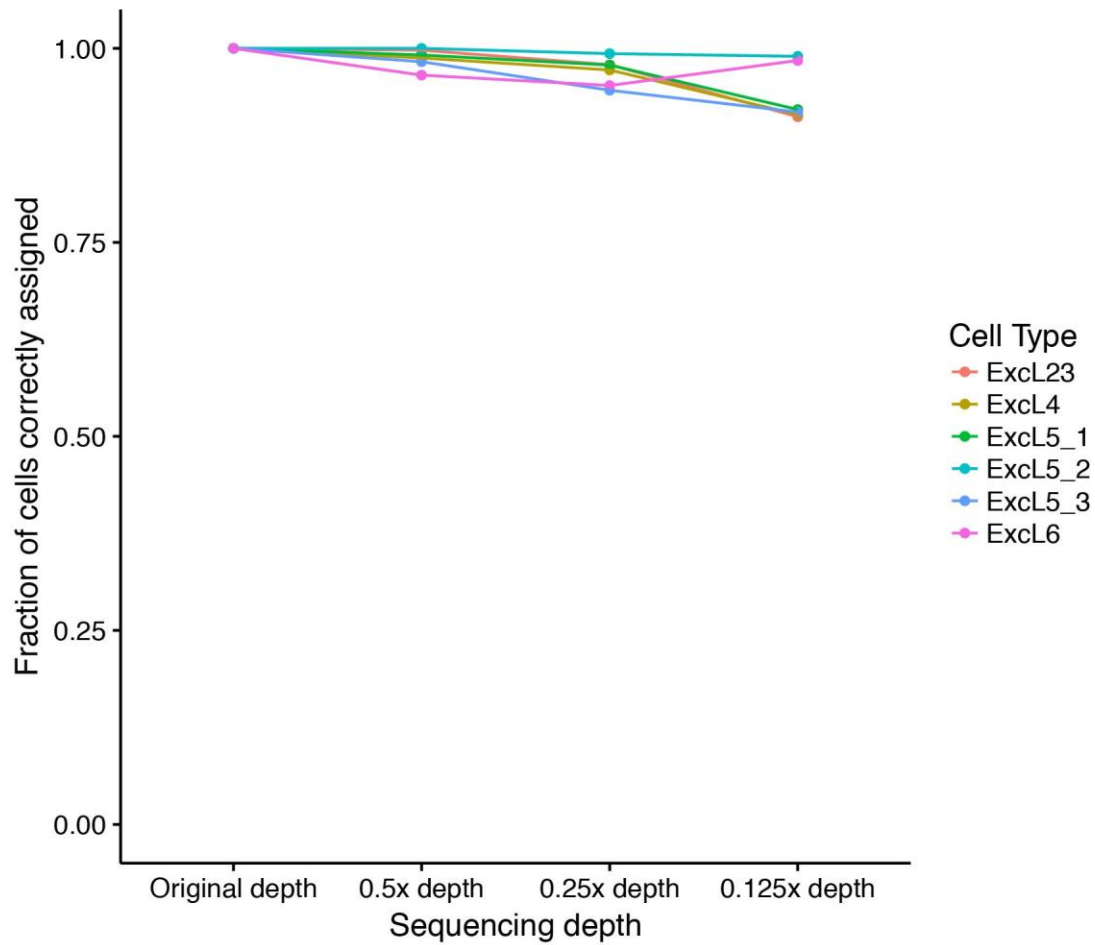
t-SNE plots are shown for each of the eight main cell types: excitatory neurons (14287 cells), inhibitory neurons (936 cells), astrocytes (7039 cells), oligodendrocytes and oligodendrocyte precursor cells (OPCs; 10,456 cells), microglia (10158 cells), endothelial and smooth muscle cells (4071 cells), macrophages (537 cells), and pericytes (782 cells). Colors within each plot indicate different subtypes.



Supplementary Figure 9

UMIFM counts and genes detected per cell by cell type

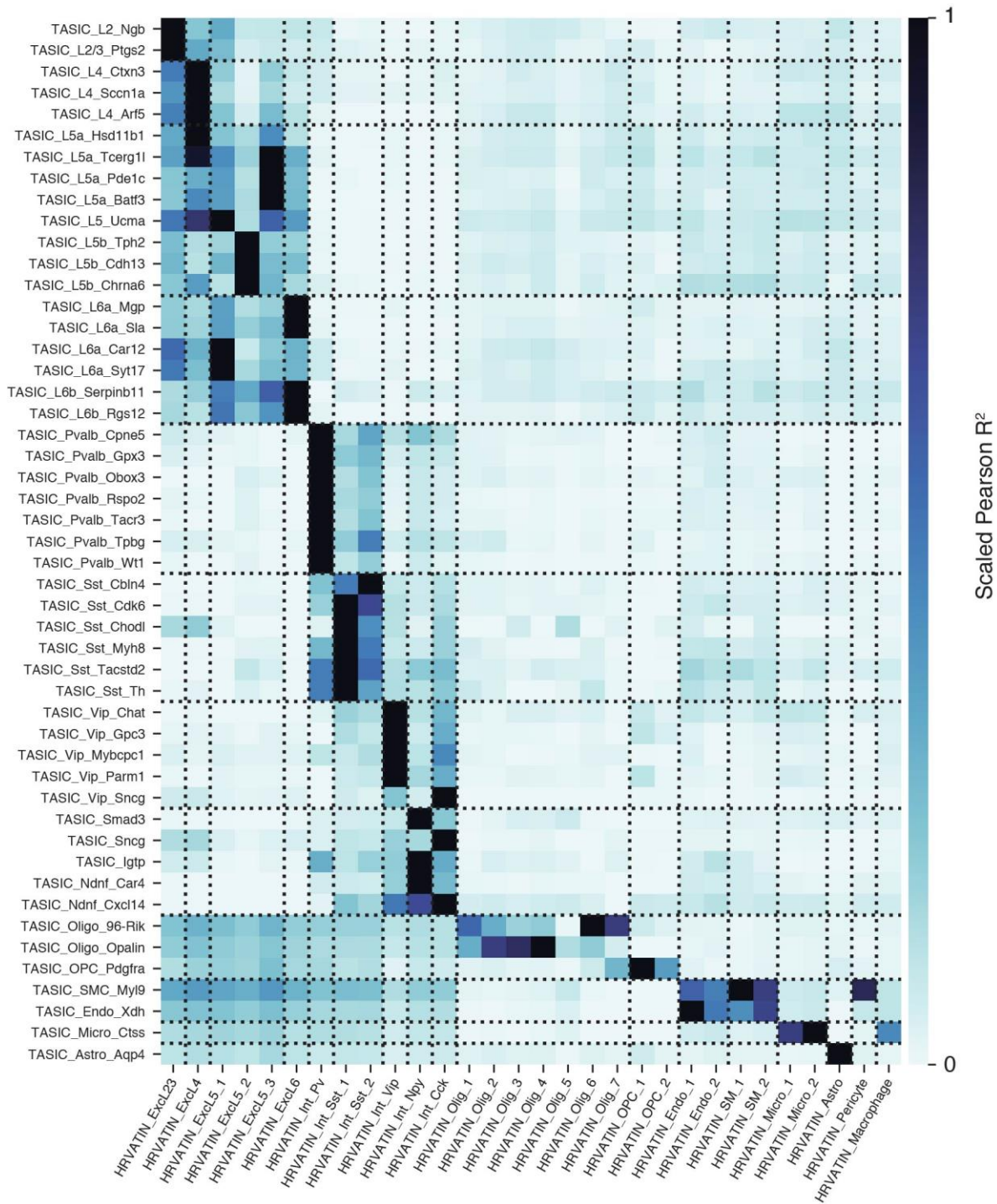
Distributions of **(a)** UMIFM counts per cell and **(b)** unique genes detected per cell plotted by cell type as whisker plots showing quartiles (median in white) for all cells that were successfully assigned to a cell type (cells=47209, left). Summary histograms (right) show the relevant metric aggregated across all clustered cells with means denoted by the red line. Across the 47209 clustered cells, an average of 3233.6 UMIFM counts and 1452.7 genes were obtained per cell. Data outside 1.5-times the interquartile range (Q3-Q1) were labeled as dots.



Supplementary Figure 10

Dependence of clustering on sequencing depth

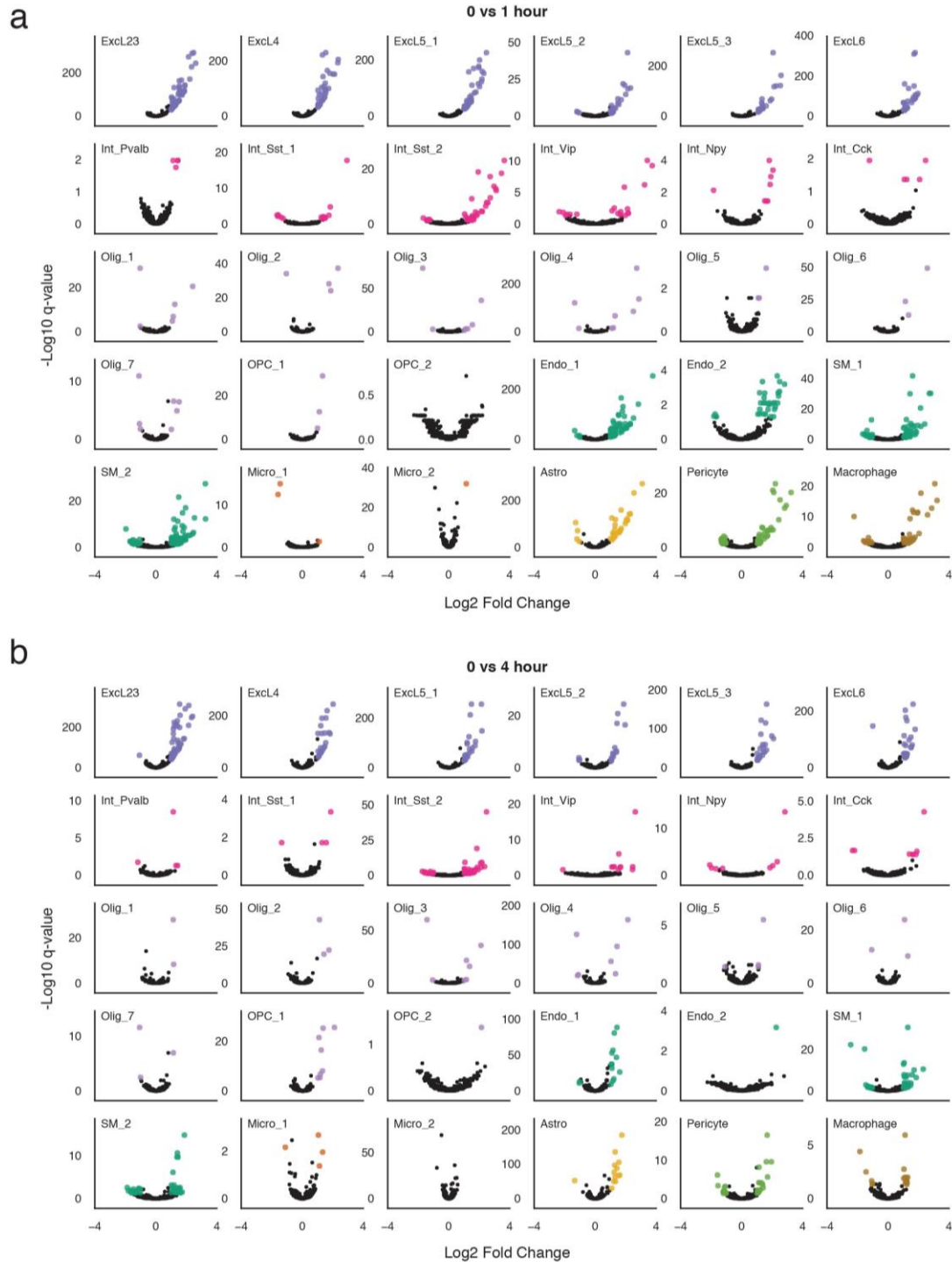
All excitatory neurons (L23, L4, L5_1, L5_2, L5_3 and L6, total cell number=13,230, average UMI per cell=6,020) were pooled and three additional datasets created by randomly subsampling each cell to contain $\frac{1}{2}$, $\frac{1}{4}$ or $\frac{1}{8}$ its starting number of transcripts. All four datasets were analyzed and clustered independently using Seurat. Marker genes were used to assign each new cluster back to one of the original excitatory cell types (L23, L4, L5_1, L5_2, L5_3 and L6). Graph shows the fraction of cells in each cluster derived from the subsetting data that were correctly assigned to the original cluster in the non-subsetting dataset.



Supplementary Figure 11

Correlations between inDrops- and Tasic et al.-identified cell types

Heatmap of pairwise Pearson R^2 correlations between all cell types identified in Ref. ²³ and this study. Magnitude of correlation denoted by color intensity as shown on scale bar. For clarified representation, values are scaled such that the highest obtained R^2 across each TASIC-labeled cell type is 1. Broadly defined cell types within each of the two studies are demarcated by dotted lines.

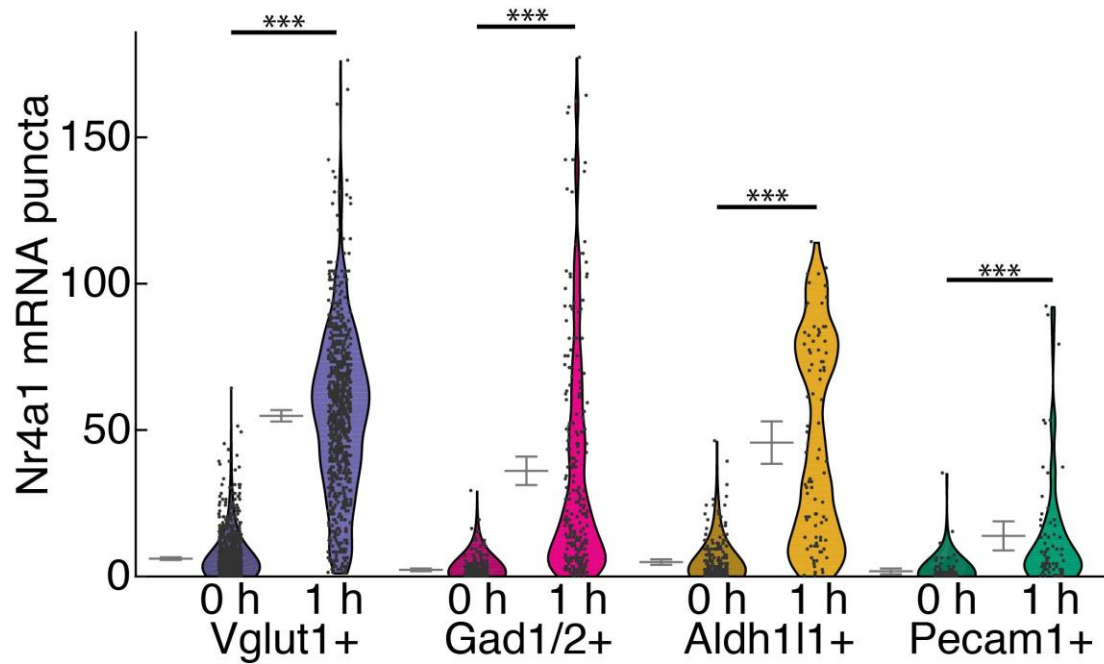


Supplementary Figure 12

Differential expression analysis results by stimulus for all cell types

Volcano plots of $-\log_{10}$ Monocle2 derived q-values against $\log_2$ fold change between 0 and 1 h (**a**) and 0 and 4 h (**b**) stimulus comparisons. Cells that passed $FDR < .05$ and $|\log_2 FC| > 1$ thresholds are shown in their representative colors. Number of cells for the 0, 1, 4h condition respectively: ExcL23 = 1150, 706, 1107; ExcL4 = 1211, 819, 1168; ExcL5_1 = 795, 516, 665; ExcL5_2 = 162, 104, 144; ExcL5_3 = 493, 412, 502; ExcL6 = 1328, 916, 1032; Int_Pv = 87, 66, 74; Int_Sst_1 = 61, 64, 49; Int_Sst_2 = 59, 52, 70; Int_Vip = 41,

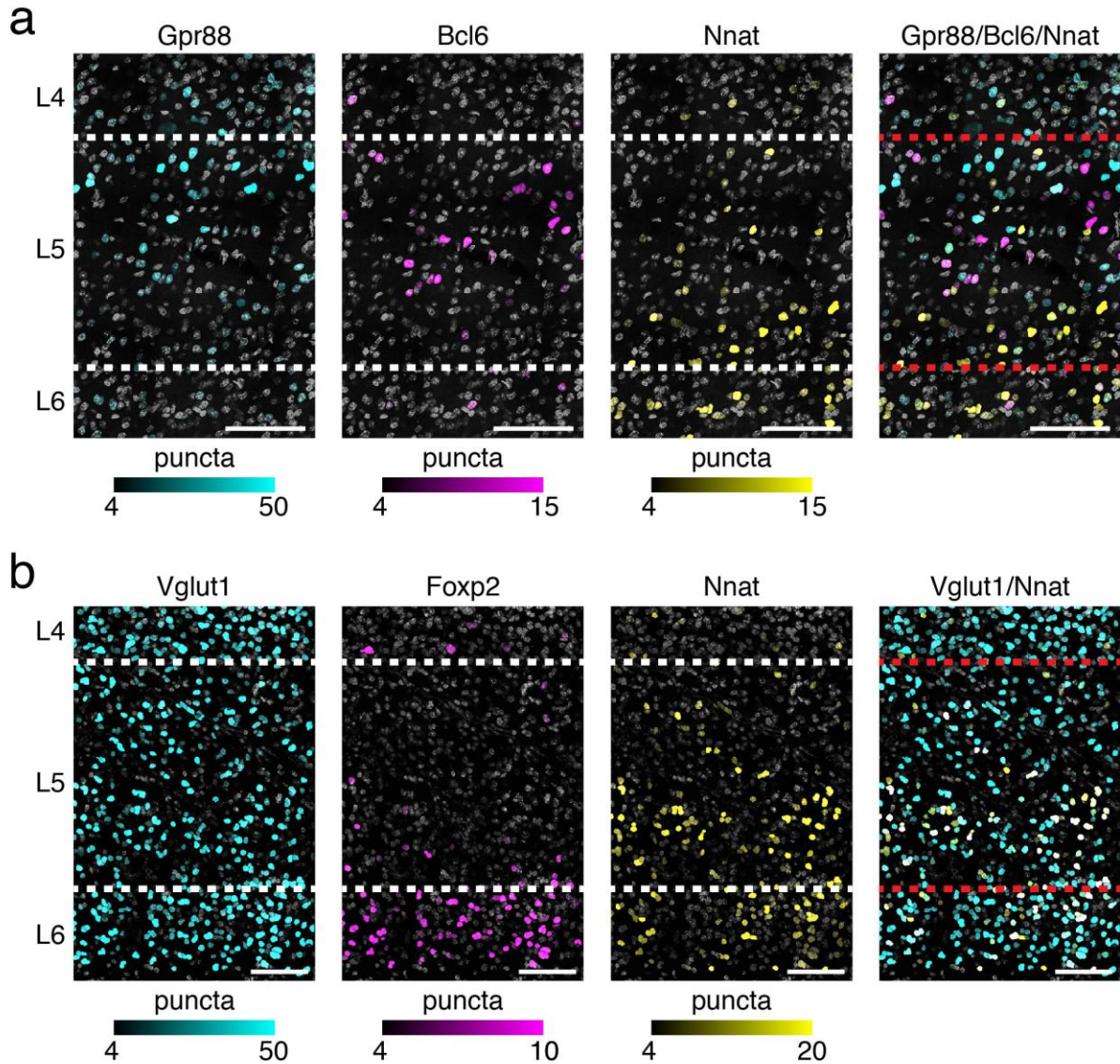
45, 40; Int_Npy = 42, 58, 37; Int_Cck = 41, 21, 29; Olig_1 = 128, 175, 307; Olig_2 = 318, 257, 389; Olig_3 = 275, 275, 296; Olig_4 = 1220, 1204, 1623; Olig_5 = 237, 225, 285; Olig_6 = 248, 236, 302; Olig_7 = 209, 182, 239; OPC_1 = 645, 472, 608; OPC_2 = 31, 27, 43; Endo_1 = 1111, 1107, 1109; Endo_2 = 33, 65, 25; SM_1 = 115, 92, 116; SM_2 = 104, 109, 85; Micro_1 = 148, 166, 235; Micro_2 = 2873, 2381, 4355; Astro = 2234, 2156, 2649; Pericyte = 295, 200, 287; Macrophage = 187, 119, 231.



Supplementary Figure 13

***Nr4a1* is induced across multiple cell types as measured by FISH**

Nr4a1 expression as measured by FISH in 0 and 1 h stimulus conditions in excitatory (*Vglut1*+, cells_{0h}=987, cells_{1h}=778) and inhibitory (*Gad1/2*+, cells_{0h}=238, cells_{1h}=243) neurons, astrocytes (*Aldh1l1*+, cells_{0h}=220, cells_{1h}=89), and endothelial and smooth muscle cells (*Pecam1*+, cells_{0h}=82, cells_{1h}=67). *** $p < 10^{-12}$, Mann-Whitney U-test, two-sided. Mean and 95% confidence intervals denoted by gray lines.

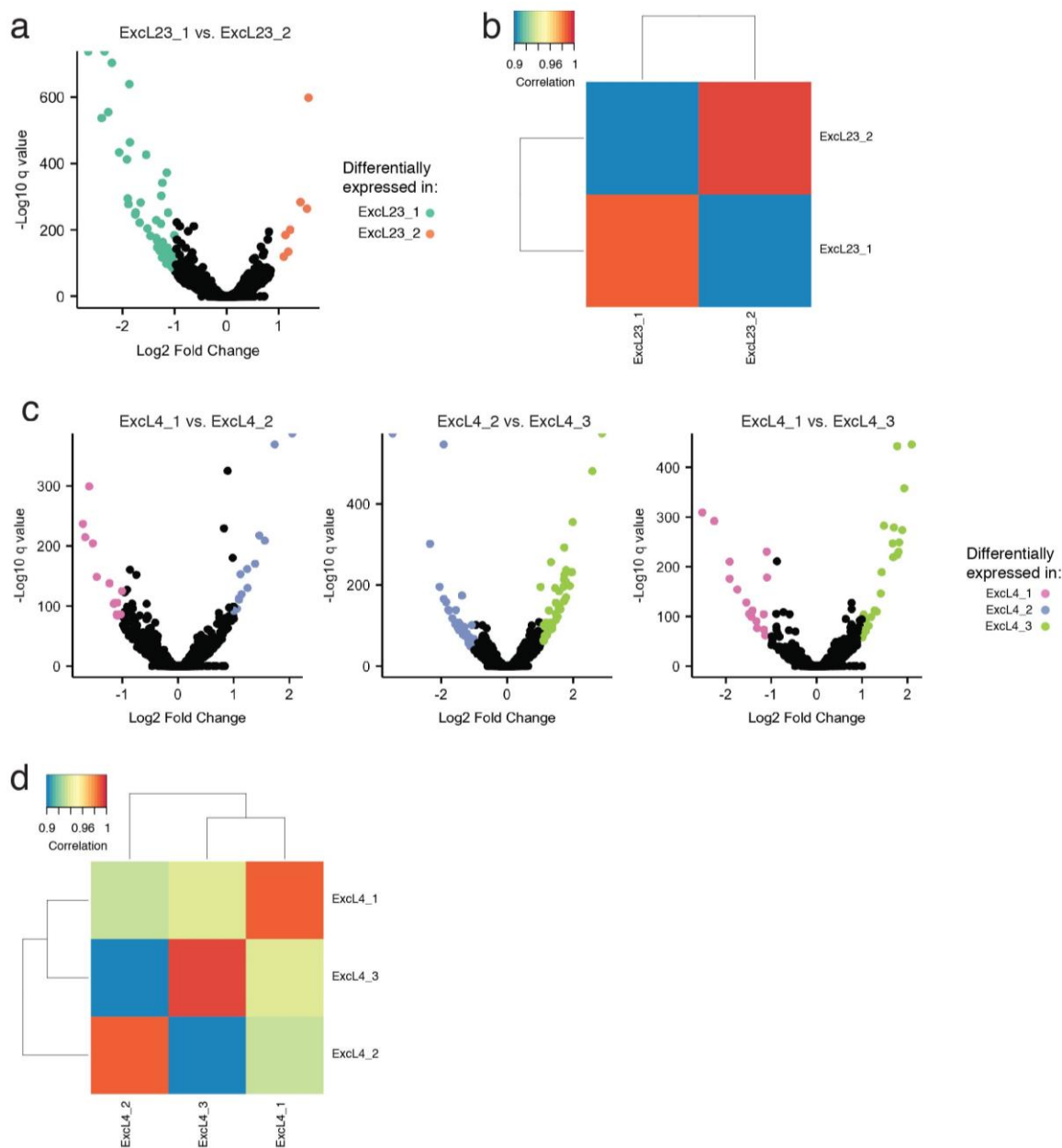


Supplementary Figure 14

Anatomical location of putative layer 5 cell types

(a) 3-color FISH for cell type enriched markers of Excl5 subtypes (from left to right): *Gpr88* (Excl5_3), *Bcl6* (Excl5_2), and *Nnat* (Excl5_1). The combined image (right) shows largely non-overlapping markers defining the Excl5 subtypes. The *Nnat* marker labeled cells in the ventral section of layer 5 and layer 6.

(b) To confirm expression of *Nnat* across both layer 5 and 6, we carried out 3-color FISH (from left to right): *Vglut1*, *Foxp2* (a canonical layer 6 marker), and *Nnat*. The presence of *Vglut1/Nnat* positive cells dorsal of and within *Foxp2* defined layer 6 supports the layer 5/6 anatomical location of the Excl5_1 population. Scale bars=100 μ m.

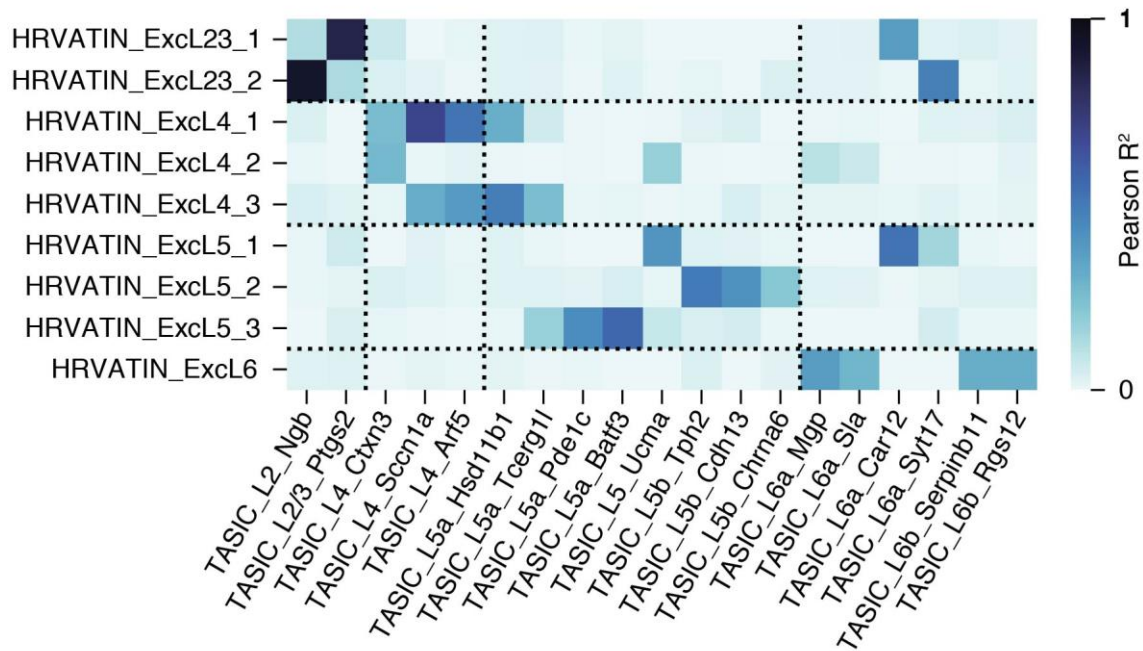


Supplementary Figure 15

Newly identified Excl23 and Excl4 subpopulations show differential gene expression

(a, c) Volcano plots of $-\log_{10}$ Monocle2 derived q-values against log2 fold change for all genes between the two Excl23 subpopulations (Excl23_1/2) described in Fig. 4c and for all three pairwise comparisons of Excl4 subtypes (Excl4_1,2,3) described in Fig. 4d respectively. Genes that passed $q < 0.05$ and $|\log_2 FC| > 1$ thresholds are shown in their representative colors.

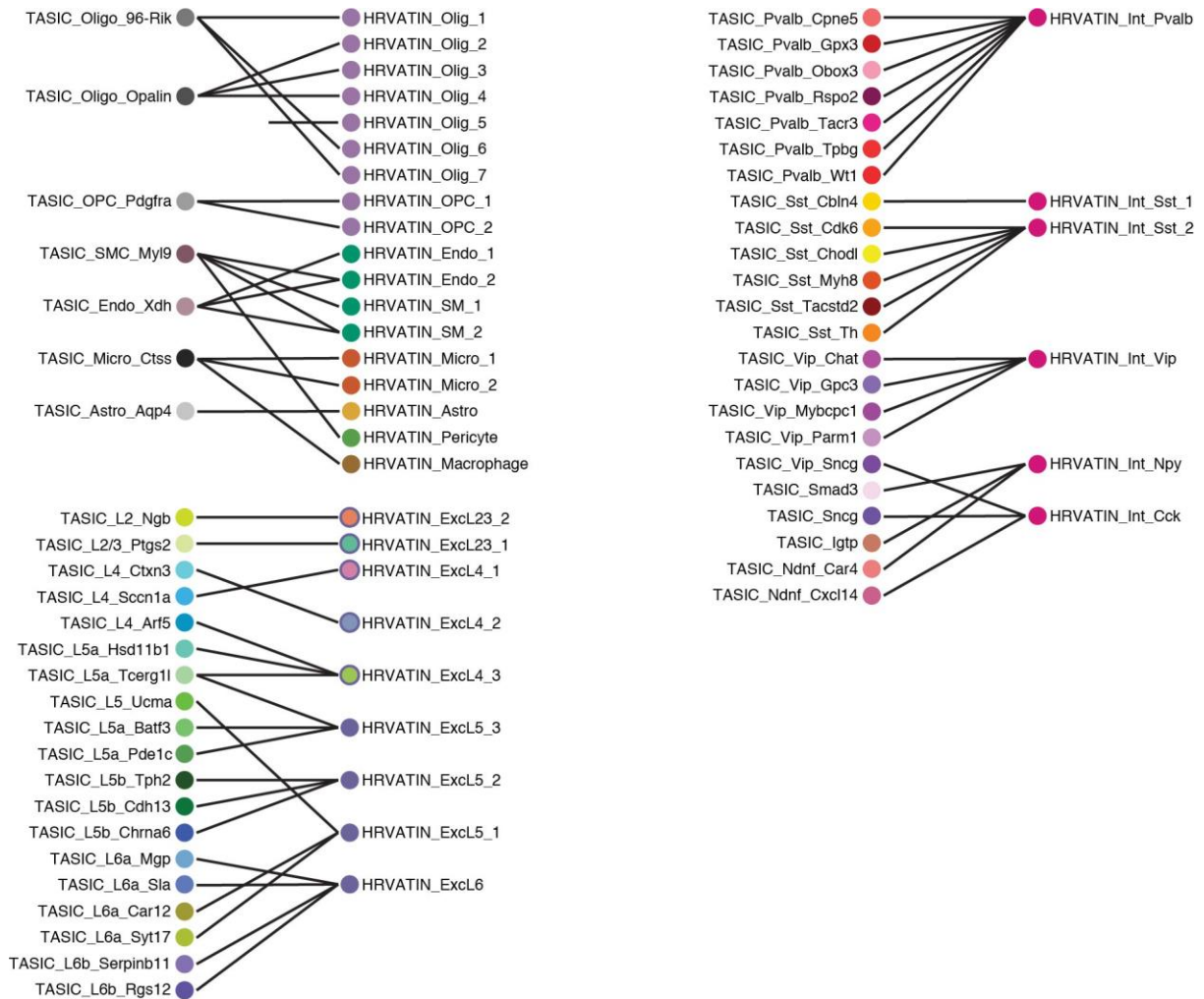
(b, d) Sets comprising 100 random cells were generated from each of Excl23_1/2 or Excl4_1/2/3 populations. The pairwise Pearson correlations between Excl23 subpopulations (b) and Excl4 subpopulations (d) was calculated using mean gene expression within each set. The analysis was repeated 100 times and the mean correlations between different subpopulations of the relevant layer are plotted.



Supplementary Figure 16

Excitatory subpopulation comparison between inDrops dataset and Tasic et al.

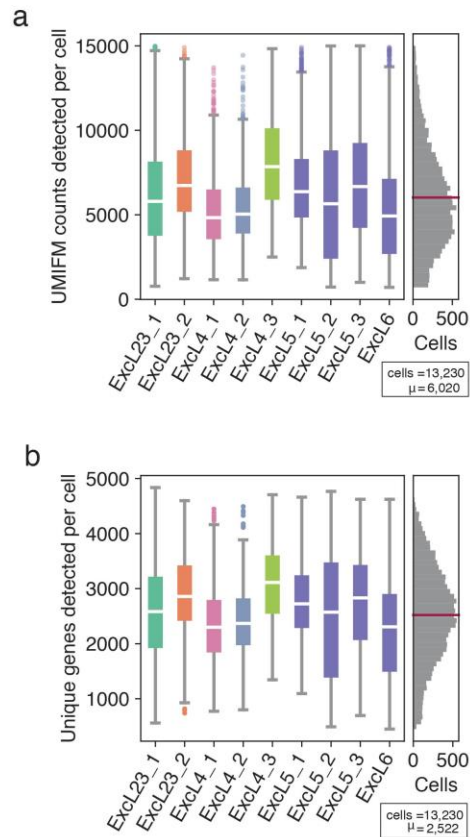
(a, c) Volcano plots of $-\log_{10}$ Monocle2-derived FDR against $\log_2$ fold change for all genes between the two ExcL23 subpopulations (ExcL23_1/2) described in **Fig. 4c** and for all three pairwise comparisons of ExcL4 subtypes (ExcL4_1,2,3) described in **Fig. 4d** respectively. Genes that passed $FDR < .05$ and $|\log_2 FC| > 1$ thresholds are shown in their representative colors. $n_{ExcL23_1} = 774, 539, 907$ cells; $n_{ExcL23_2} = 362, 166, 193$ cells; $n_{ExcL4_3} = 136, 137, 175$ cells; $n_{ExcL4_1} = 732, 468, 783$ cells; $n_{ExcL4_2} = 343, 214, 210$ cells for 0, 1, and 4 h timepoints respectively.



Supplementary Figure 17

Summary of inDrops- and Tasic et al.-derived cell types

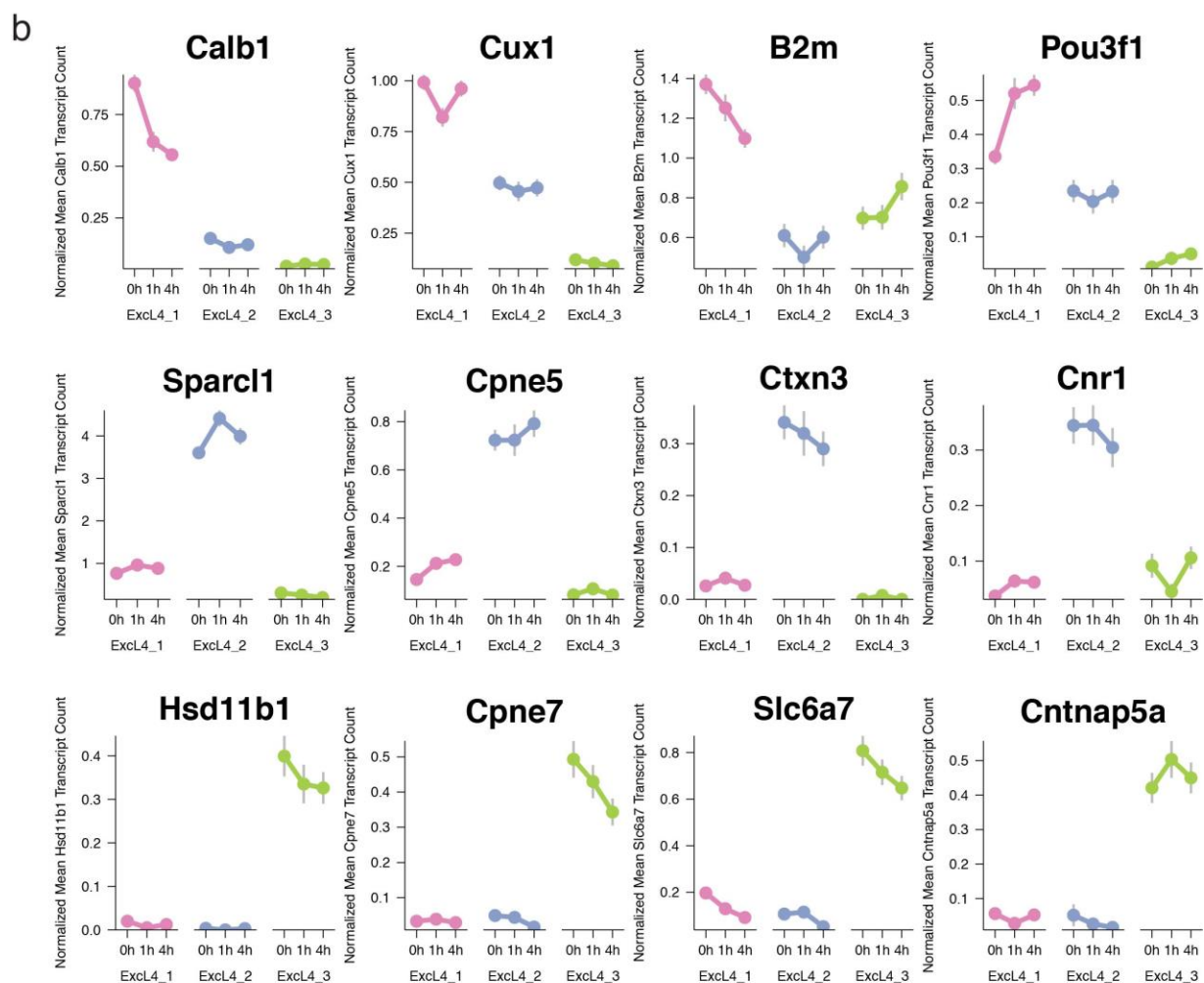
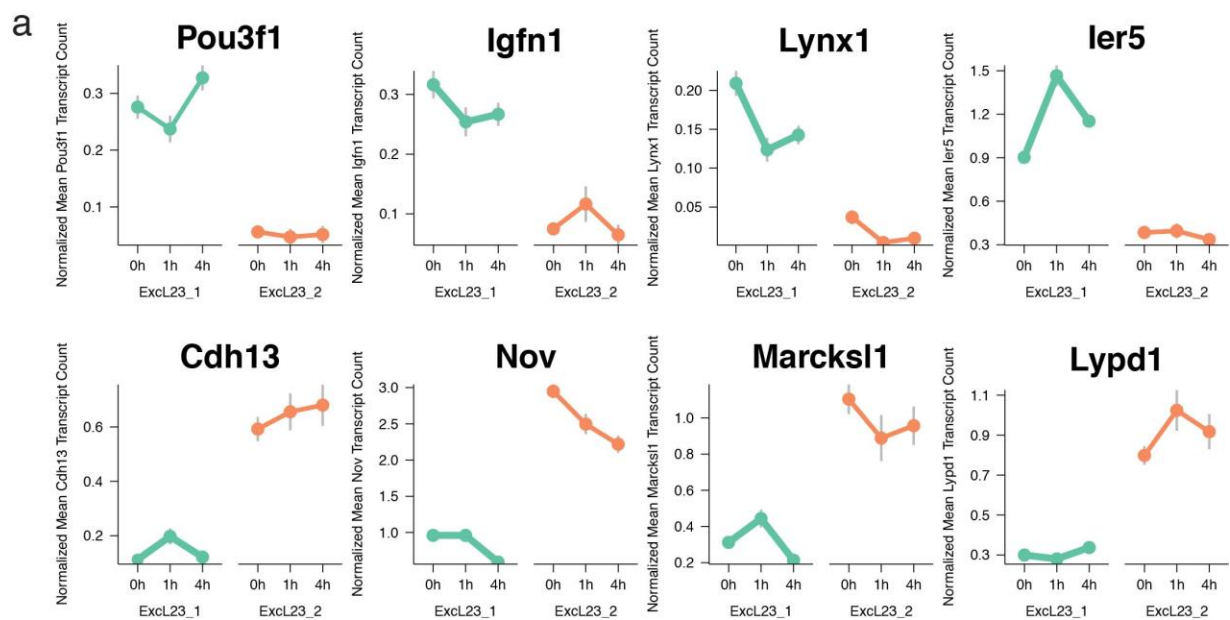
A summary mapping of the Tasic et al. transcriptionally-defined cell types²³ and those described in this study constructed from analyses shown in **Supplemental Fig. 7, 11, 16, 19, 21, 22, 23**.



Supplementary Figure 18

Excitatory subpopulations are not defined by sequencing depth

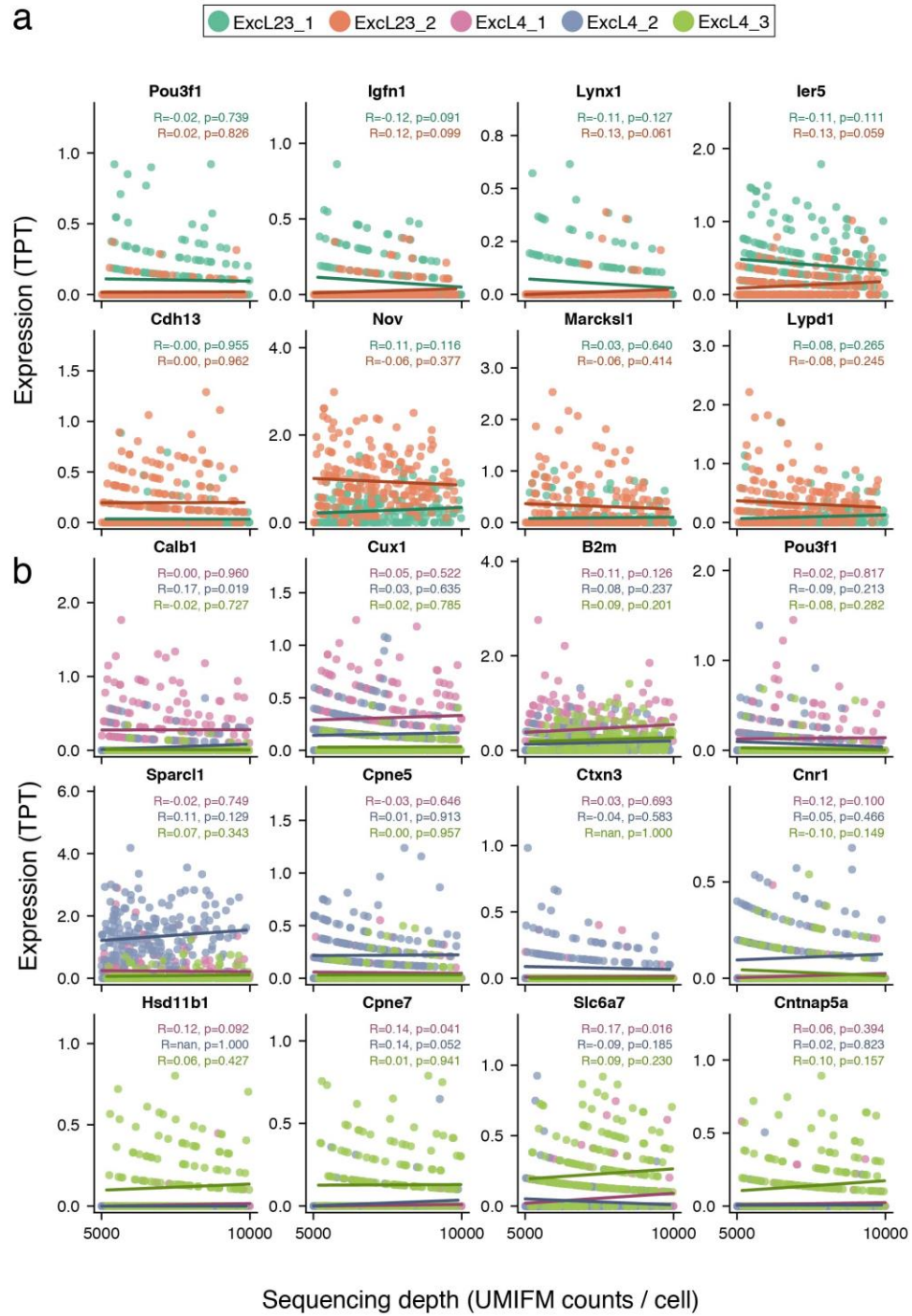
Left: Distributions of UMIFM counts (**a**) and unique genes (**b**) detected per cell represented as whisker plots demarcating quartiles and the median (white) for each excitatory subtype across all successfully clustered excitatory cells (cells=13,230). Right: Aggregated distributions represented as histograms. Means of each metric are denoted by the red lines. Data outside 1.5-times the interquartile range (Q3-Q1) were labeled as dots.



Supplementary Figure 19

L23 and L4 excitatory subtype marker genes are not stimulus-regulated

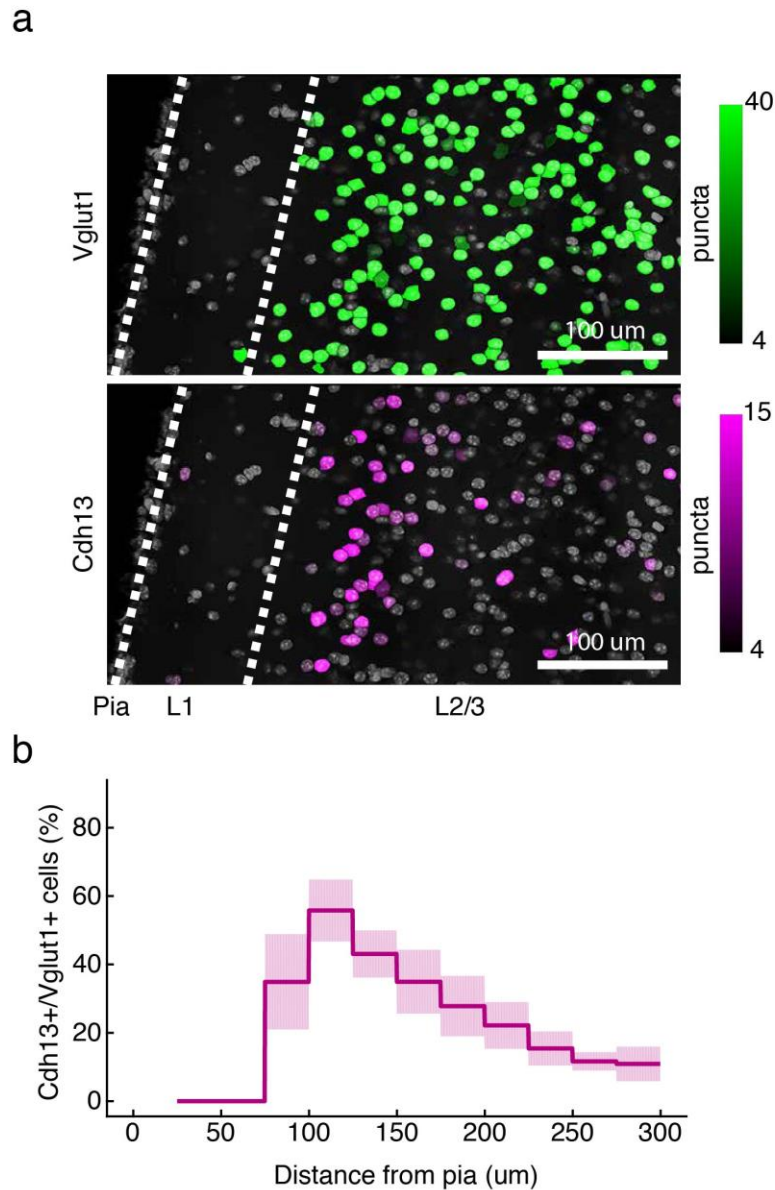
Mean normalized expression for several non-stimulus regulated marker genes between (a) L23 excitatory and (b) layer 4 excitatory subtypes. Error bars indicate s.e.m. $n_{\text{ExcL23}_1}=774, 539, 907$ cells; $n_{\text{ExcL23}_2}=362, 166, 193$ cells; $n_{\text{ExcL4}_3}=136, 137, 175$ cells; $n_{\text{ExcL4}_1}=732, 468, 783$ cells; $n_{\text{ExcL4}_2}=343, 214, 210$ cells for 0, 1, and 4 h time points respectively.



Supplementary Figure 20

L23 and L4 subtype marker gene expression is not correlated with sequencing depth

L23 (a) and L4 (b) subtype marker gene expression (transcripts per thousand [TPT] calculated across each cell type) plotted as functions of sequencing depth (mean UMIFM counts) across 200 randomly sampled cells (with $5000 \leq \text{UMIFM count/cell} \leq 10000$) within L23 and L4 respectively (markers described in **Fig. 4d, 4e, Supplemental Fig. S19**, Online Methods). Within each plot, the lines of best fit are shown together with the calculated Pearson correlation and the associated p-value. The color of each data point or trend-line denotes the cell type represented by that object. We find that marker gene expression is not significantly correlated with sequencing depth. We also see that for each marker gene, the relative expression of the marker is appropriately greater in its marked sublayer despite restricting the sequencing depth ranges for the cells examined. This analysis confirms that the cell-type-specific expression of these marker genes is not due to differences in sequencing depth.

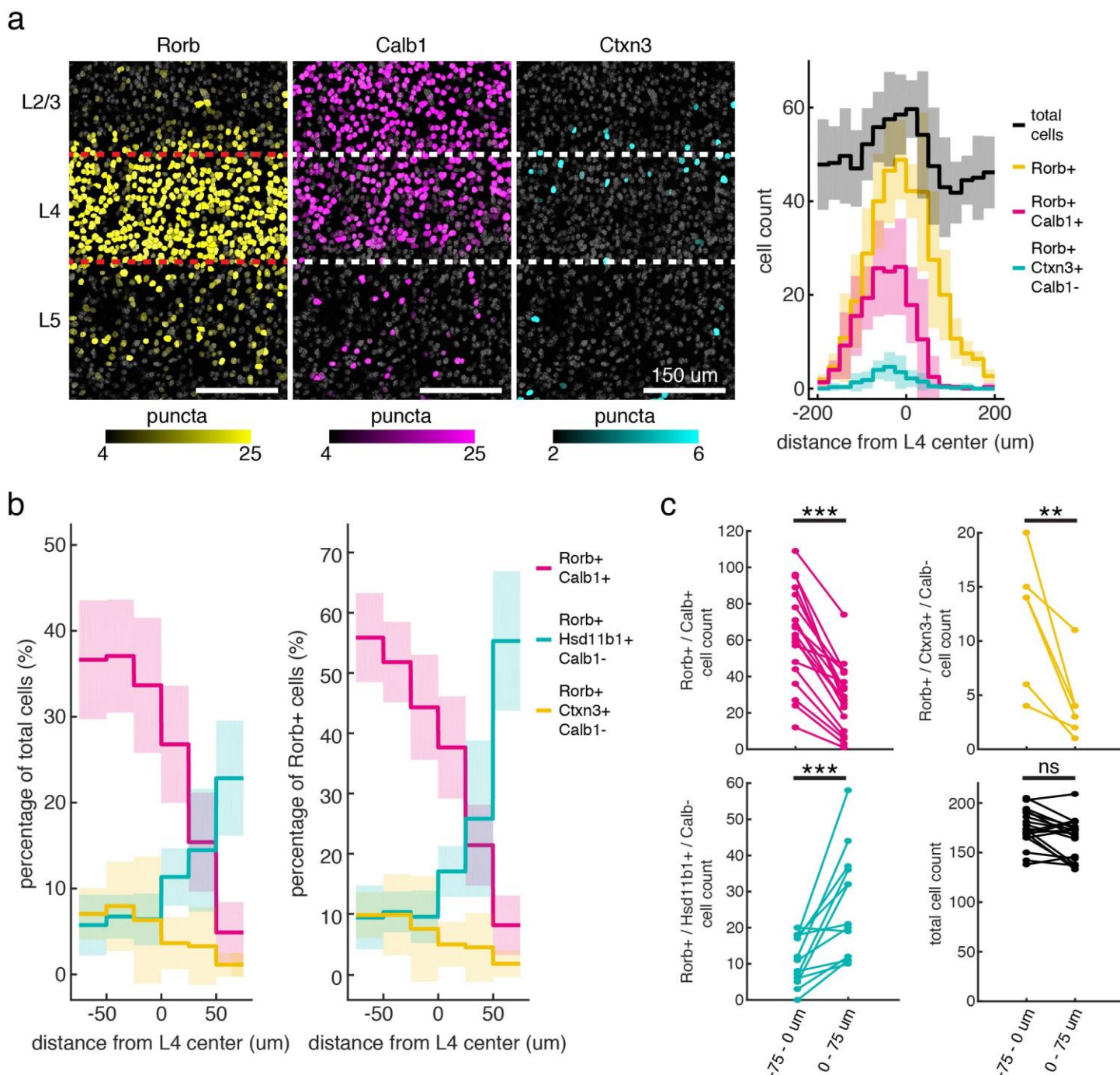


Supplementary Figure 21

***Cdh13*+ excitatory neurons are superficially enriched in layer 2/3**

(a) FISH images showing superficial enrichment of *Cdh13*+ cells in Layer 2/3. Nuclei are pseudo-colored by expression level of *Vglut1* (green) or *Cdh13* (magenta) FISH probes (Online methods). Scale bar=100 um.

(b) Quantification of FISH data showing fraction of *Cdh13*+ excitatory neurons (*Vglut1*+) as a function of distance from the pia (cells=436 *Cdh13*+/*Vglut1*+, of cells=1,624 *Vglut1*+ cells). Data from 8 cortical slices.



Supplementary Figure 22

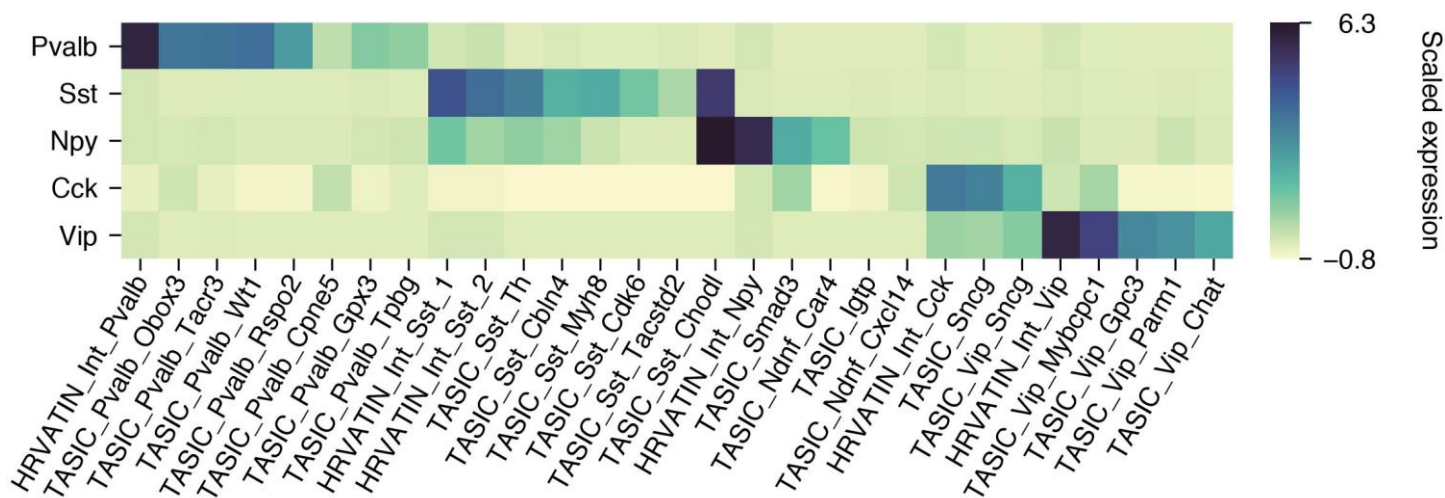
FISH confirmation of L4 subtypes

(a) We carried out 3-color FISH for (from left) *Rorb* (yellow), *Calb1* (magenta), and *Ctxn3* (cyan). We found that *Ctxn3*+/*Rorb*+/*Calb1*- cells were intermingled with *Calb1*+/*Rorb*+ cells superficially in layer 4. Right: summary of the anatomical distribution of the *Ctxn3* and *Calb1* transcriptionally defined cell types. Data collected from 6 slices of mouse visual cortex. Negative/positive values on x-axis correspond to distance from the center of layer 4 towards the slice surface (negative) or towards deeper cortical layers (positive).

(b) Summary data for all layer 4 excitatory subtypes as a percentage of total cells (left), or of *Rorb*+ cells (right). *Calb1*+/*Rorb*+ and *Ctxn3*+/*Rorb*+/*Calb1*- cell types (data collected from 19 and 6 slices respectively that contained visual cortex) were enriched superficially within layer 4 by both measures, and *Hsd11b1*/*Rorb*+/*Calb1*- defined cells (data collected from 13 slices containing visual

cortex) were enriched deep within layer 4 and peaked at the border between layer 4 and 5 (see Fig. 4h). Shaded area around lines indicate 95% confidence intervals.

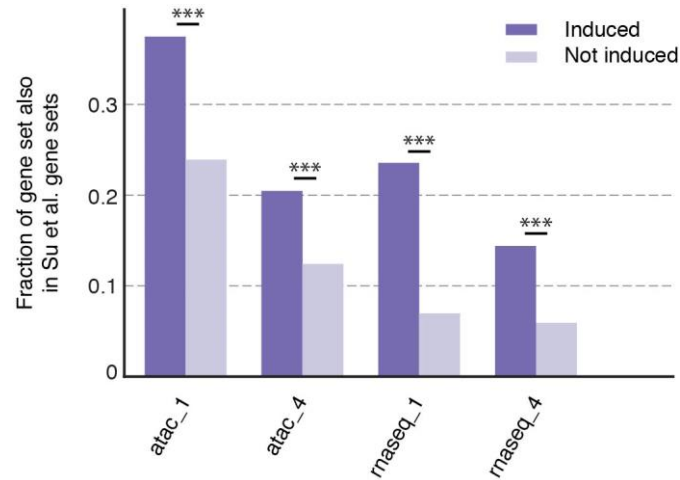
(c) Quantification of cell population enrichment to superficial (-75 to 0 μ m from L4 center) or deep (0 to 75 μ m from L4 center) layer 4. The transcriptionally defined cell types *Calb1*⁺/*Rorb*⁺ and *Ctnn3*/*Rorb*⁺/*Calb1*⁻ were significantly enriched superficially ($p=0.0002$, $p=0.015$, Mann-Whitney U-test, two-sided; data collected from 19 and 6 slices containing visual cortex, respectively), and *Hsd11b1*⁺/*Rorb*⁺/*Calb1*⁻ cells were enriched deep within layer 4 ($p=0.0034$, Mann-Whitney U-test, two-sided; data collected from 13 slices containing visual cortex). On the other hand, the total number of cells in superficial and deep layer 4 was not significantly different ($p=0.19$, Mann-Whitney U-test, two-sided; data collected from 19 slices containing visual cortex). Scale bars=150 μ m. Shaded area around lines indicate 95% confidence intervals.



Supplementary Figure 23

Interneuron comparison between inDrops and Tasic et al. datasets

Scaled expression of interneuronal markers used to assign cell types across all interneuronal populations identified in this work and by Ref. ²³. Expression levels were scaled for each gene from each cell type independently (centered to mean with unit variance, Online Methods).



Supplementary Figure 24

V1-identified induced genes are highly enriched for genes associated with inducibly accessible loci in dentate gyrus

Induced gene sets from dentate gyrus comprising the genes with the most proximal TSS to inducibly accessible loci identified by Su *et al.* at 1 and 4 hours of stimulation (atac_1, atac_4) or the genes directly identified by Su *et al.* via bulk RNA-seq (maseq_1, maseq_4) were tested for enrichment among the induced genes identified in this study (Online Methods)⁴⁸. The fractions of the visual cortex induced gene set identified in this study (611 genes) and its complement that were also induced in Su *et al.*'s paradigm are plotted. All gene sets show significant enrichment. Relative probability=1.84, 1.75, 3.67, 2.50; FDR-corrected q-value=3.2e-13, 3.4e-8, 3.4e-35, 2.7e-13 (in order shown in Figure). ***FDR-corrected $q < 1e-5$.