Supplementary Information for

Adult Mouse Cortical Cell Taxonomy Revealed by Single Cell Transcriptomics

Bosiljka Tasic[1, 2, 3], Vilas Menon[1, 2], Thuc Nghi Nguyen[1], Tae Kyung Kim[1], Tim Jarsky[1], Zizhen Yao[1], Boaz Levi[1], Lucas T. Gray[1], Staci A. Sorensen[1], Tim Dolbeare[1], Darren Bertagnolli[1], Jeff Goldy[1], Nadiya Shapovalova[1], Sheana Parry[1], Changkyu Lee[1], Kimberly Smith[1], Amy Bernard[1], Linda Madisen[1], Susan M. Sunkin[1], Michael Hawrylycz[1], Christof Koch[1], Hongkui Zeng[1]

[1] Allen Institute for Brain Science, Seattle, WA, USA.

[2] These authors contributed equally to this work.

[3] Correspondence to: Bosiljka Tasic (bosiljkat@alleninstitute.org).

The following Supplementary Tables are included as Excel files:

Supplementary Table 1. Transgenic driver lines.

Supplementary Table 2. Transgenic reporter lines.

Supplementary Table 3. Single cell samples.

Supplementary Table 4. Cre line and cell type relationships. The percentage of cell types detected in each Cre line/dissection combination separated by core and intermediate cells. These data were used to generate the graphical representation in **Fig. 2b.**

Supplementary Table 5. Evaluation of enrichment of interneuron types in upper or lower cortical layers using the hypergeometric test. For details on statistics methodology see **Methods**. Note that lack of statistically significant enrichment does not necessarily indicate that there is no enrichment, as our sampling did not allow comprehensive evaluation of spatial enrichment for all types. We do not claim lower layer-enrichment for the Pvalb-Wt1 type because we obtained statistical significance only in one of the two examined recombinase lines. Additional information on spatial enrichment for some of these types can be obtained by examination of cell-type-specific markers by RNA ISH.

Supplementary Table 6. Marker genes for transcriptomic cell types. The table also contains an earlier version of the cell type nomenclature used in the original release of the online scientific vignette.

Supplementary Table 7. Transcriptomic cell types: correspondence to literature.

Supplementary Table 8. Differentially processed exons among cell types.

Supplementary Table 9. Evaluation of correspondence between RNA-seq and Allen Brain Atlas chromogenic RNA ISH data. Out of 228 genes examined, 72% show agreement between single cell RNA-seq and Allen Brain Atlas data. For most of the other genes, the disagreement is due to the absence of signal in the Allen Brain Atlas ISH (17%). Small numbers of genes display apparently ubiquitous signal in VISp by ISH (2%), specificity of the signal that is difficult to interpret (2%), or the ISH pattern that, in fact, disagrees with RNA-seq (2%). For about 4% of the genes, no data is available in the Allen Brain Atlas.

Supplementary Table 10. Cluster identities after subsampling of single cell RNA-seq data. Cluster identities obtained using the full depth sequencing data (median of ~4.4 million mapped reads or ~8.7 million total reads) are compared to cluster identities obtained when data from each cell were subsampled to 100,000 and 1,000,000 mapped reads per cell. We detect fewer clusters with decreased sequencing depth.

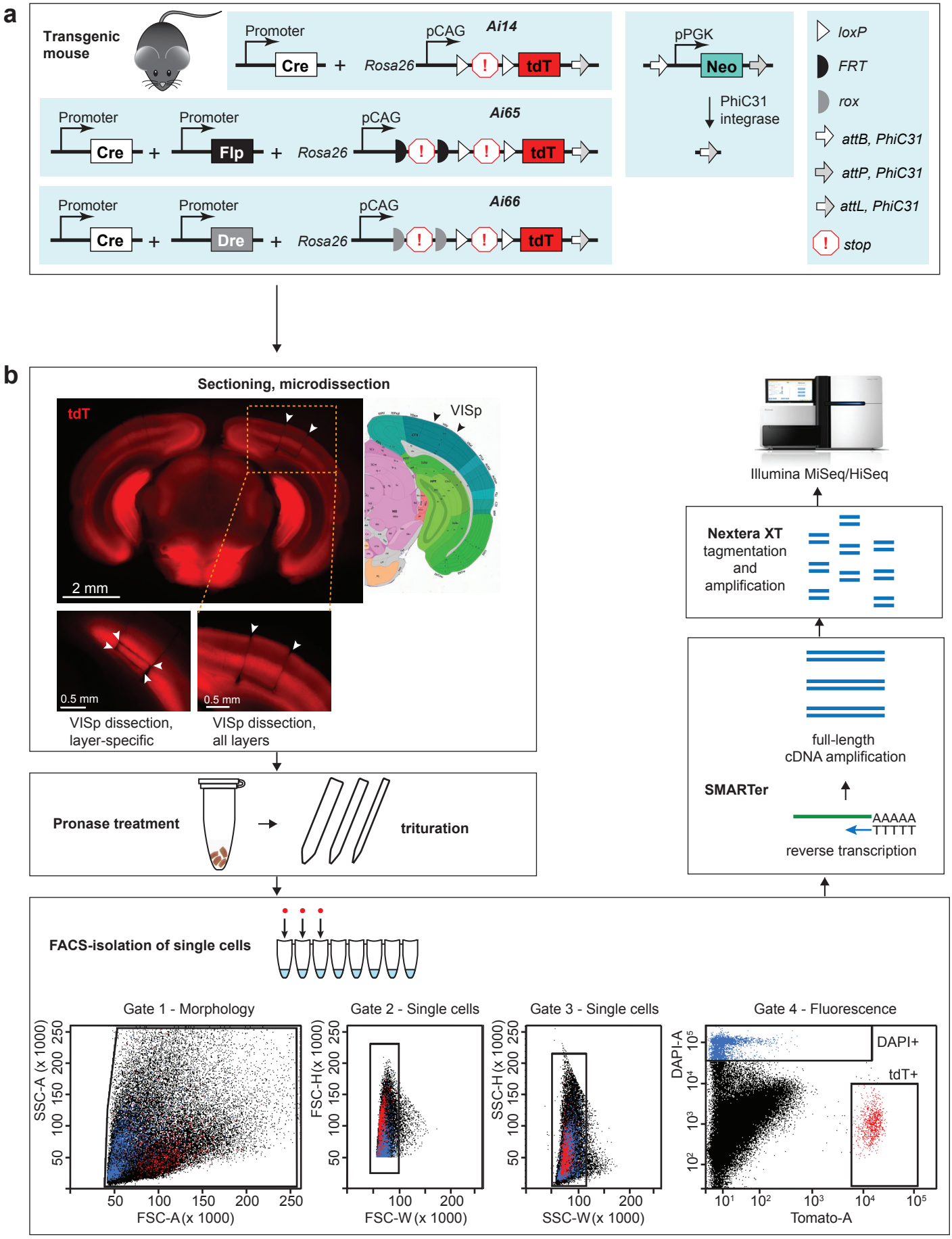
Supplementary Table 11. Genetic background estimate for all animals used in the study. In our experimental animals, the percent of C57BL/6J genetic background ranges from 75% to

100% with the average of ~96%. In cases where the original ancestor we obtained was on mixed background, we adopted a conservative estimate of 0% C57BL/6J in that ancestor.

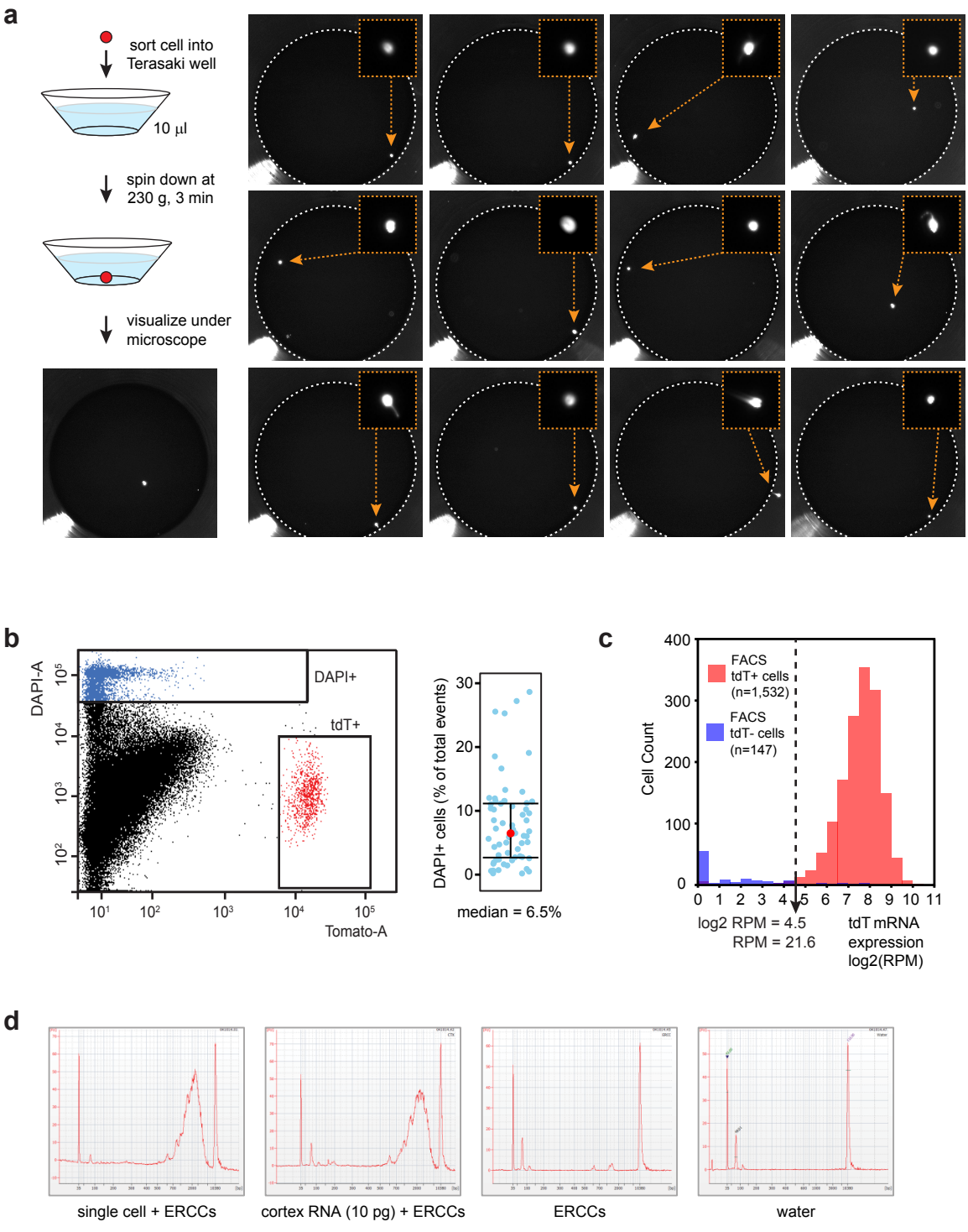
Supplementary Table 12. Consequences of cluster validation parameter change on single cell classification. Cluster identity assignment for each cell is listed for our default parameters (20 genes, $p < 0.01$), and after changes in these parameters: decrease or increase in the number of genes to 10, and 50, respectively, and change in the p value to $p < 0.05$. With parameter change, on average, ~3% of the cells change cluster identity (from one core to another core, from one core to intermediate connecting two different cores, or from intermediate connecting two cores to an intermediate connecting two different cores or becoming a third core), while ~18% change from core to intermediate and vice versa (but stay within same core/intermediate identity combination). However, the total number of core clusters is preserved for all parameter changes.

Supplementary Table 13. Probes for DFISH.

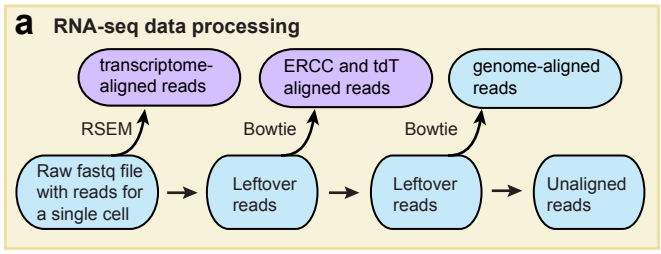
Supplementary Table 14. Quantitative RT-PCR assays.

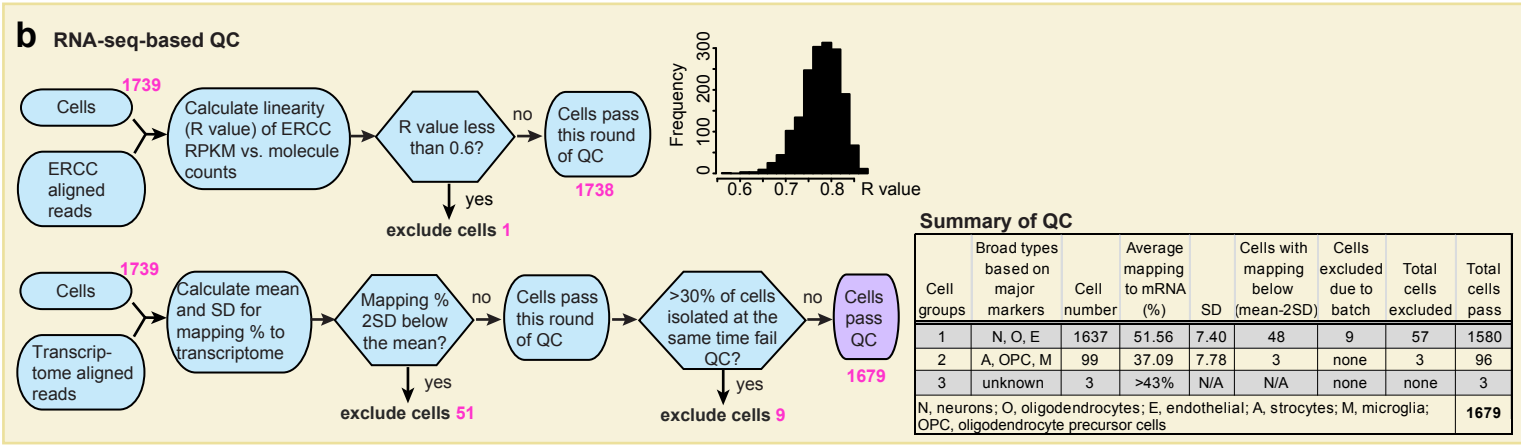

a
Transgenic mouse
Promoter
Cre
+
Rosa26
pCAG
Ai14
!
tdT
Promoter
Cre
+
Promoter
Flp
+
Rosa26
pCAG
Ai65
Promoter
Cre
+
Promoter
Dre
+
Rosa26
pCAG
Ai66
pPGK
Neo
PhiC31 integrase
loxP
FRT
rox
attB, PhiC31
attP, PhiC31
attL, PhiC31
stop
b
Sectioning, microdissection
tdT
VISp
2 mm
0.5 mm
0.5 mm
VISp dissection, layer-specific
VISp dissection, all layers
Pronase treatment
trituration
FACS-isolation of single cells
Illumina MiSeq/HiSeq
Nextera XT tagmentation and amplification
full-length cDNA amplification
SMARTer
AAAAA
TTTTT
reverse transcription
Gate 1 - Morphology
SSC-A (x 1000)
FSC-A (x 1000)
Gate 2 - Single cells
FSC-H (x 1000)
FSC-W (x 1000)
Gate 3 - Single cells
SSC-H (x 1000)
SSC-W (x 1000)
Gate 4 - Fluorescence
DAPI-A
DAPI+
tdT+
Tomato-A

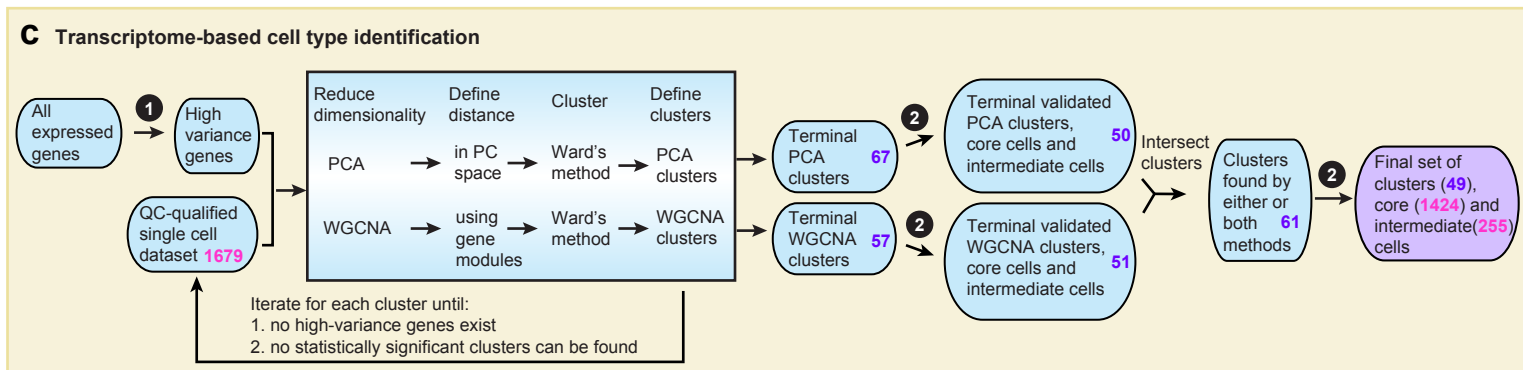
Supplementary Figure 1. Detailed experimental workflow. (**a**) Schematic representation of transgenes used for the lines mentioned in this paper; polyadenylation sites in transgenes are omitted for clarity. Each Cre recombinase line was crossed to *Ai14* or to a second recombinase line (*Flp* or *Dre*) and then to an appropriate reporter (*Ai65* or *Ai66*, see **Supplementary Table 2**). We used *Ai65* with or without the *Neo* gene present (it can be excised by a cross to a *PhiC31o* integrase line)[51]. (**b**) Detailed experimental workflow. Starting with an adult male transgenic mouse, age P56 ± 3, fresh brain was isolated, sectioned and microdissected. The microdissection was performed to isolate tissue within VISp that spans the whole cortical depth or was focused on one or several contiguous layers of VISp. The microdissected tissue was treated with protease and triturated with pipettes with increasingly smaller tip diameter (600 µm, 300 µm, and 150 µm). We isolated single cells from the cell suspension by FACS. We applied the presented set of gates and "single cell sorting mode," which excludes any cell-containing droplets if adjacent droplets also contain any cells or debris. Gate 1 was applied to exclude debris, while gates 2 and 3 exclude cell doublets. Gate 4 was used to select cells with high tdT fluorescence and low DAPI fluorescence. Single cell mRNA was reverse transcribed, amplified into cDNA (SMARTer, Clontech), and tagmented using Nextera XT (Illumina). Single cell libraries were sequenced on Illumina HiSeq and/or MiSeq.

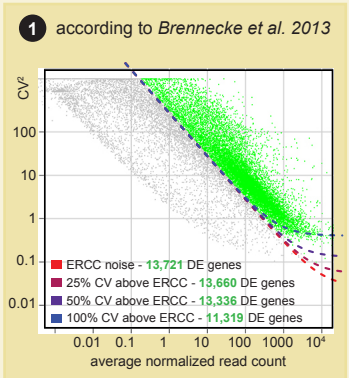

a
sort cell into Terasaki well
10 µl
spin down at 230 g, 3 min
visualize under microscope
b
DAPI-A
DAPI+
tdT+
Tomato-A
DAPI+ cells (% of total events)
median = 6.5%
c
FACS tdT+ cells (n=1,532)
FACS tdT- cells (n=147)
Cell Count
log2 RPM = 4.5
RPM = 21.6
tdT mRNA expression log2(RPM)
d
single cell + ERCCs
cortex RNA (10 pg) + ERCCs
ERCCs
water

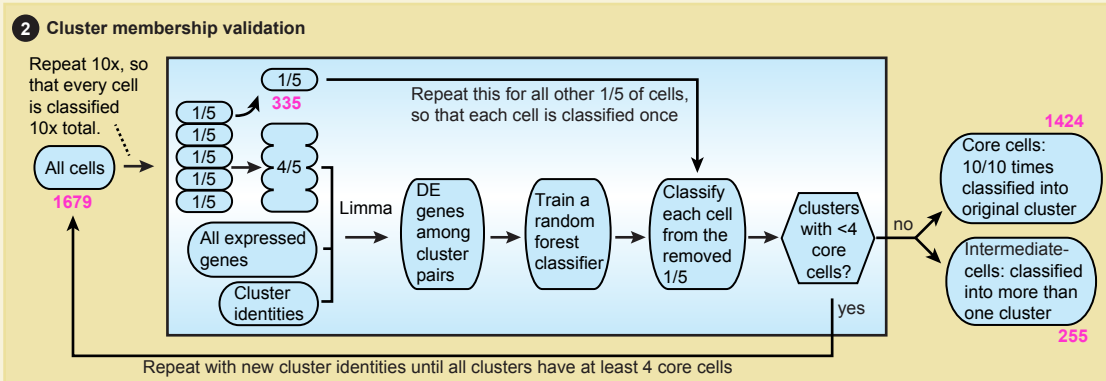
Supplementary Figure 2. Experimental QC. (**a**) A representative control experiment for assessing FACS specificity and efficiency. Before sorting cells into 8-well strips or 96-well plates for transcriptional profiling, the cells were sorted using the same setup into Terasaki plates containing 5 or 10 µl of artificial cerebrospinal fluid (ACSF). Terasaki wells were examined for presence of a single cell, more than one cell, or absence of a cell. In total, we scored 425 wells over 39 experiments, with 6-12 wells per experiment, and found that on average 96 ± 6% (standard deviation) wells contained one cell. No wells were found to contain two or more cells. (**b**) Assessing the percentage of dead cells in a sample of dissociated single cells by FACS. Left: A representative FACS plot for sorting tdT^+ cells, and assessing the percentage of DAPI-positive cells in the sample. Right: Average percentage of cells within the DAPI-positive gate for 64 out of 72 FACS experiments performed in this study. Red dot represents the median, whiskers represent the 25^{th} and 75^{th} percentiles. (**c**) The distribution of *tdT* mRNA expression in single cells as measured by RNA-seq in tdT^+ (red) and tdT^- (blue) cells, for all classified neurons. More than 99% of cells sorted as tdT^+ show higher expression of *tdT* mRNA than all classified neurons sorted as tdT^-. (**d**) Representative electrophoretograms obtained by Bioanalyzer (Agilent) for 53 batches of cDNA amplifications showing amplified cDNA from a single cell and standard positive (cortex RNA) and negative (ERCCs and water) controls.


a
RNA-seq data processing
transcriptome-aligned reads
ERCC and tdT aligned reads
genome-aligned reads
RSEM
Bowtie
Bowtie
Raw fastq file with reads for a single cell
Leftover reads
Leftover reads
Unaligned reads

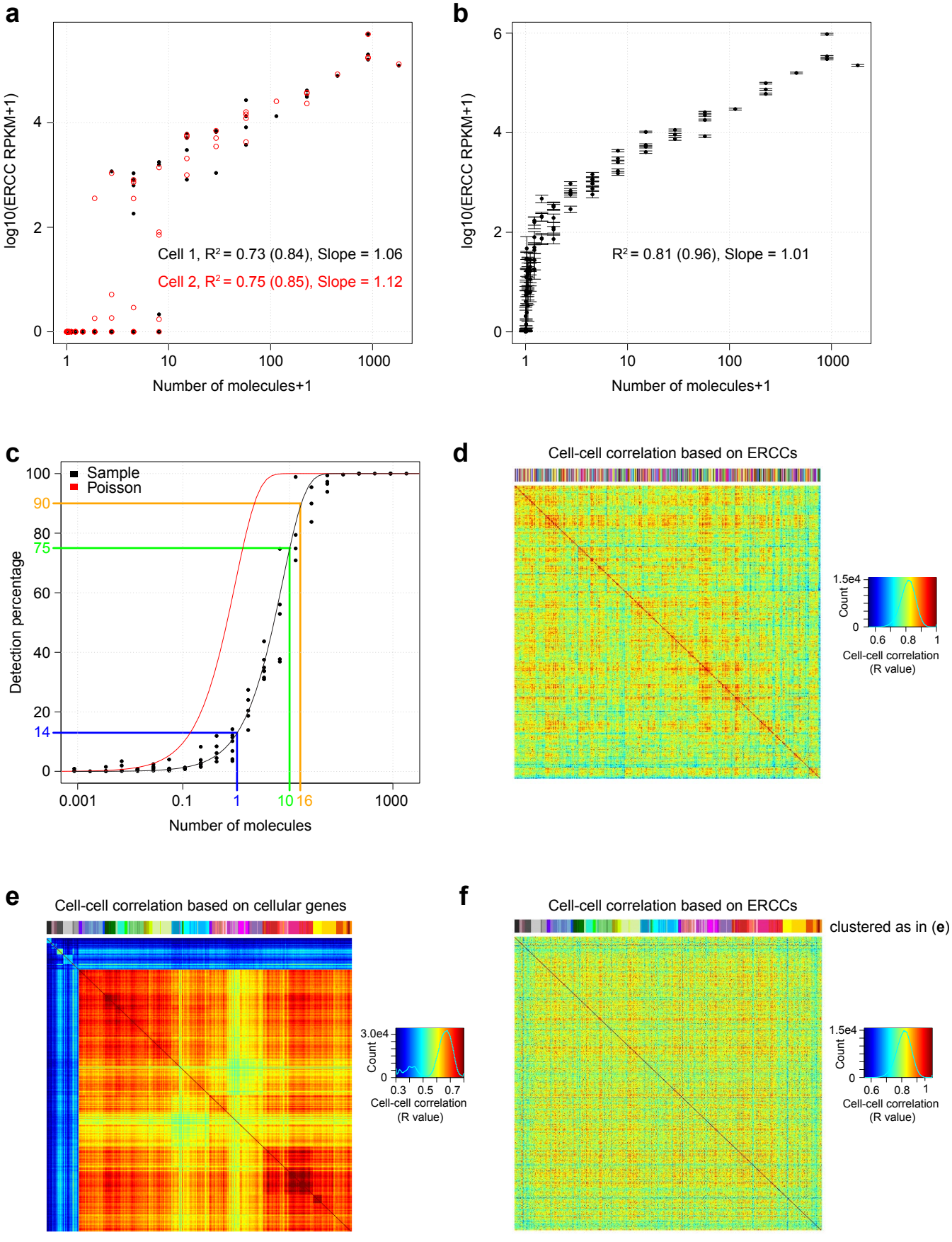

b
RNA-seq-based QC
1739
Cells
ERCC aligned reads
Calculate linearity (R value) of ERCC RPKM vs. molecule counts
R value less than 0.6?
no
Cells pass this round of QC
1738
yes
exclude cells 1
Frequency
0
100
200
300
0.6
0.7
0.8
R value
1739
Cells
Transcrip-tome aligned reads
Calculate mean and SD for mapping % to transcriptome
Mapping % 2SD below the mean?
no
Cells pass this round of QC
>30% of cells isolated at the same time fail QC?
no
Cells pass QC
1679
yes
exclude cells 51
yes
exclude cells 9
Summary of QC
Cell groups
Broad types based on major markers
Cell number
Average mapping to mRNA (%)
SD
Cells with mapping below (mean-2SD)
Cells excluded due to batch
Total cells excluded
Total cells pass
1
N, O, E
1637
51.56
7.40
48
9
57
1580
2
A, OPC, M
99
37.09
7.78
3
none
3
96
3
unknown
3
>43%
N/A
N/A
none
none
3
N, neurons; O, oligodendrocytes; E, endothelial; A, strocytes; M, microglia; OPC, oligodendrocyte precursor cells
1679


c
Transcriptome-based cell type identification
All expressed genes
1
High variance genes
QC-qualified single cell dataset 1679
Reduce dimensionality
Define distance
Cluster
Define clusters
PCA
in PC space
Ward's method
PCA clusters
WGCNA
using gene modules
Ward's method
WGCNA clusters
Terminal PCA clusters 67
Terminal WGCNA clusters 57
2
Terminal validated PCA clusters, core cells and intermediate cells 50
Terminal validated WGCNA clusters, core cells and intermediate cells 51
Intersect clusters
Clusters found by either or both 61 methods
2
Final set of clusters (49), core (1424) and intermediate(255) cells
Iterate for each cluster until:
1. no high-variance genes exist
2. no statistically significant clusters can be found

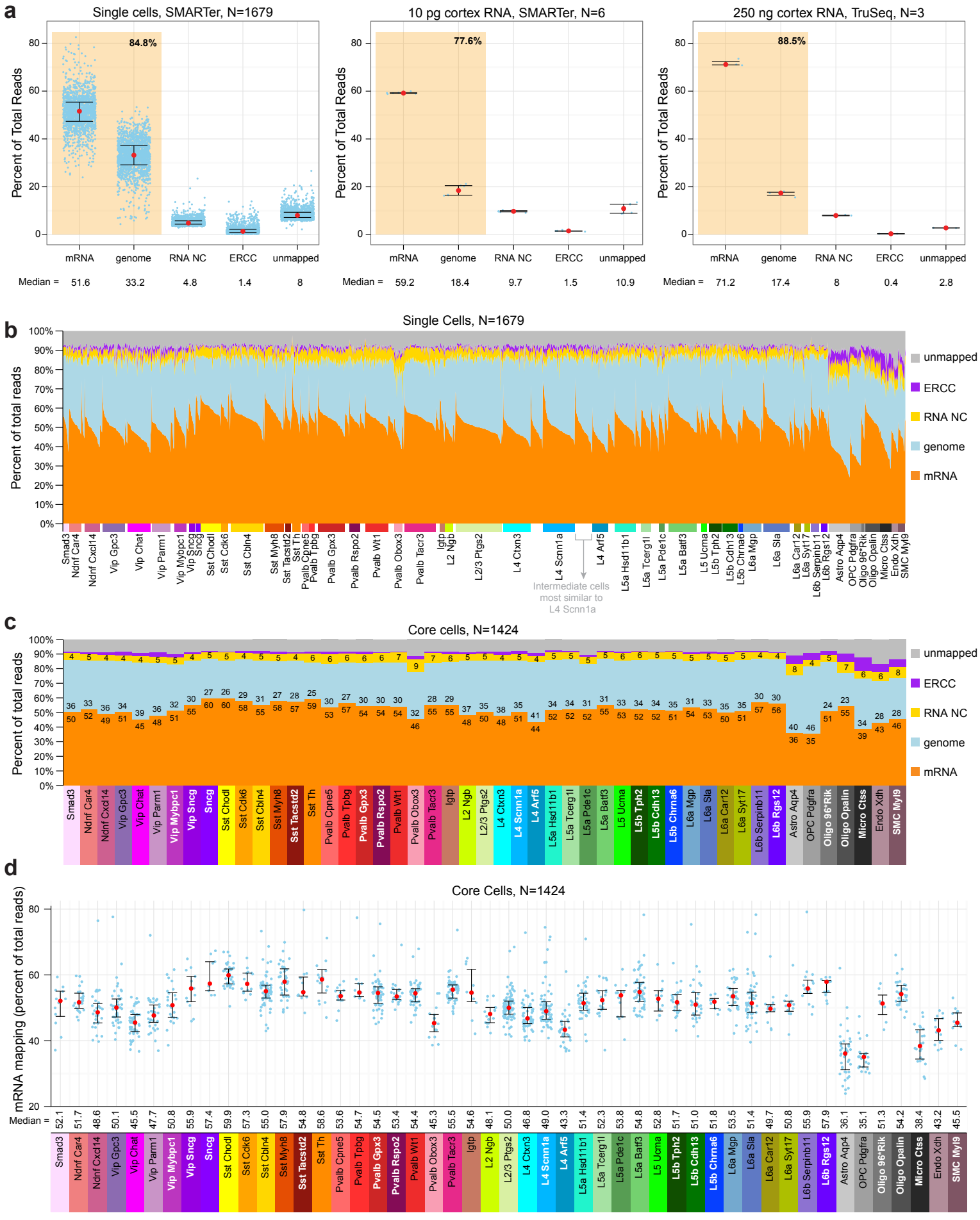

1 according to Brennecke et al. 2013
CV²
100
10
1
0.1
0.01
ERCC noise - 13,721 DE genes
25% CV above ERCC - 13,660 DE genes
50% CV above ERCC - 13,336 DE genes
100% CV above ERCC - 11,319 DE genes
0.01
0.1
1
10
100
1000
10⁴
average normalized read count


2 Cluster membership validation
Repeat 10x, so that every cell is classified 10x total.
All cells
1679
1/5
335
Repeat this for all other 1/5 of cells, so that each cell is classified once
4/5
All expressed genes
Cluster identities
Limma
DE genes among cluster pairs
Train a random forest classifier
Classify each cell from the removed 1/5
clusters with <4 core cells?
no
yes
1424
Core cells: 10/10 times classified into original cluster
Intermediate-cells: classified into more than one cluster
255
Repeat with new cluster identities until all clusters have at least 4 core cells

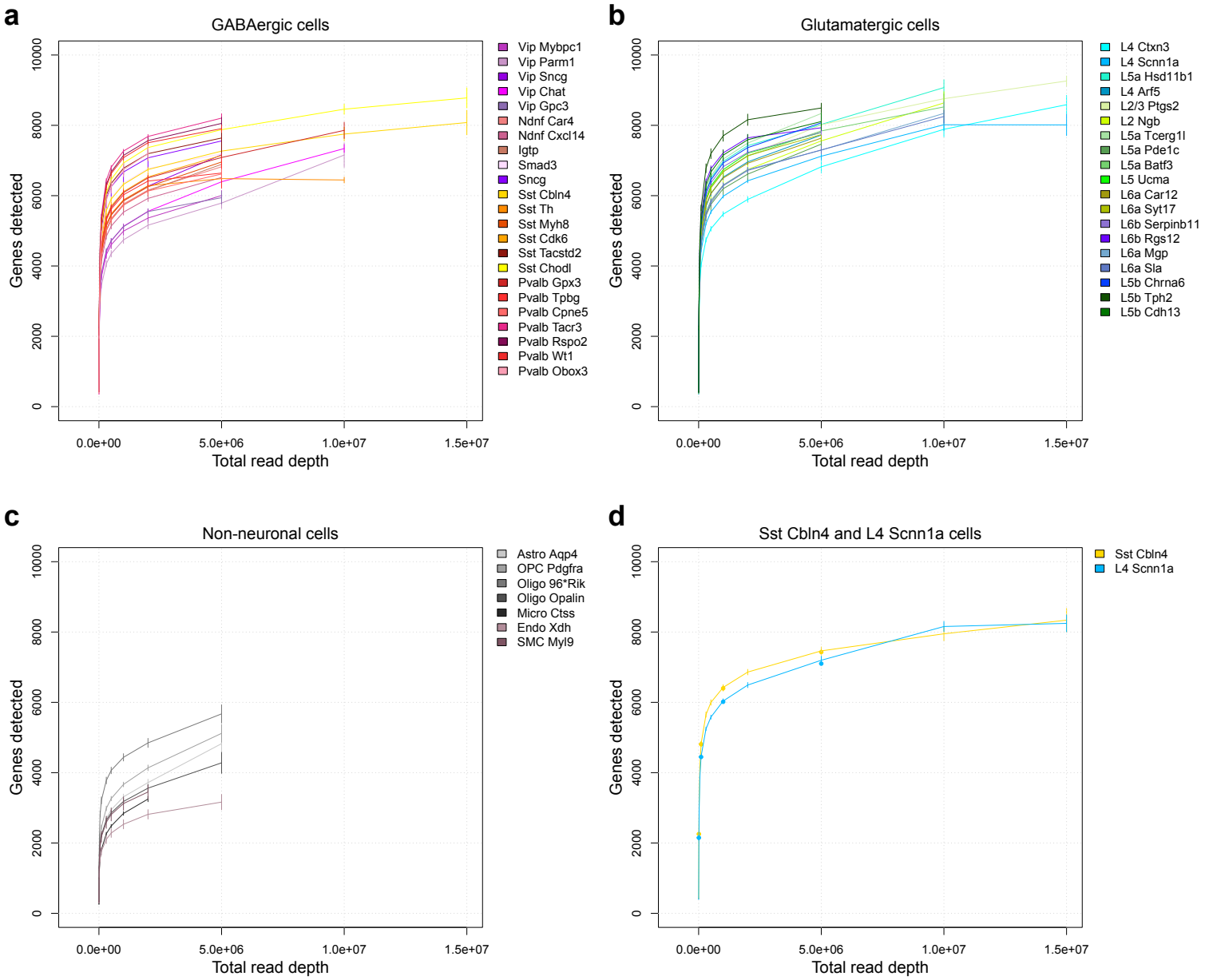
Supplementary Figure 3. Detailed data analysis workflow. (**a**) Workflow diagram for processing of raw sequencing reads to generate counts and RPKM values for genes, as well as total read counts aligning to *ERCC* RNAs, *tdT* mRNA, and genomic regions. (**b**) Quality control steps based on *ERCC* detection linearity and transcriptome mapping percentage, including the number of cells that were excluded at each stage. The single cell that was excluded based on low *ERCC* linearity was also excluded based on low transcriptome mapping percentage. (**c**) Details of the iterative cell type identification workflow, starting with the identification of high variance genes (shown as green dots in inset 1), and proceeding through the repeated use of the validation procedure (explained in detail in inset 2) that tests cluster membership to identify core cells and intermediate cells. The latter procedure also results in reintegration of small clusters that contain less than 4 cells. Numbers in pink indicate the number of cells used at each point in the analysis; numbers in purple represent the numbers of clusters.


a
log10(ERCC RPKM+1)
Number of molecules+1
Cell 1, R² = 0.73 (0.84), Slope = 1.06
Cell 2, R² = 0.75 (0.85), Slope = 1.12
b
log10(ERCC RPKM+1)
Number of molecules+1
R² = 0.81 (0.96), Slope = 1.01
c
Sample
Poisson
Detection percentage
Number of molecules
d
Cell-cell correlation based on ERCCs
Count
Cell-cell correlation (R value)
e
Cell-cell correlation based on cellular genes
Count
Cell-cell correlation (R value)
f
Cell-cell correlation based on ERCCs
clustered as in (e)
Count
Cell-cell correlation (R value)

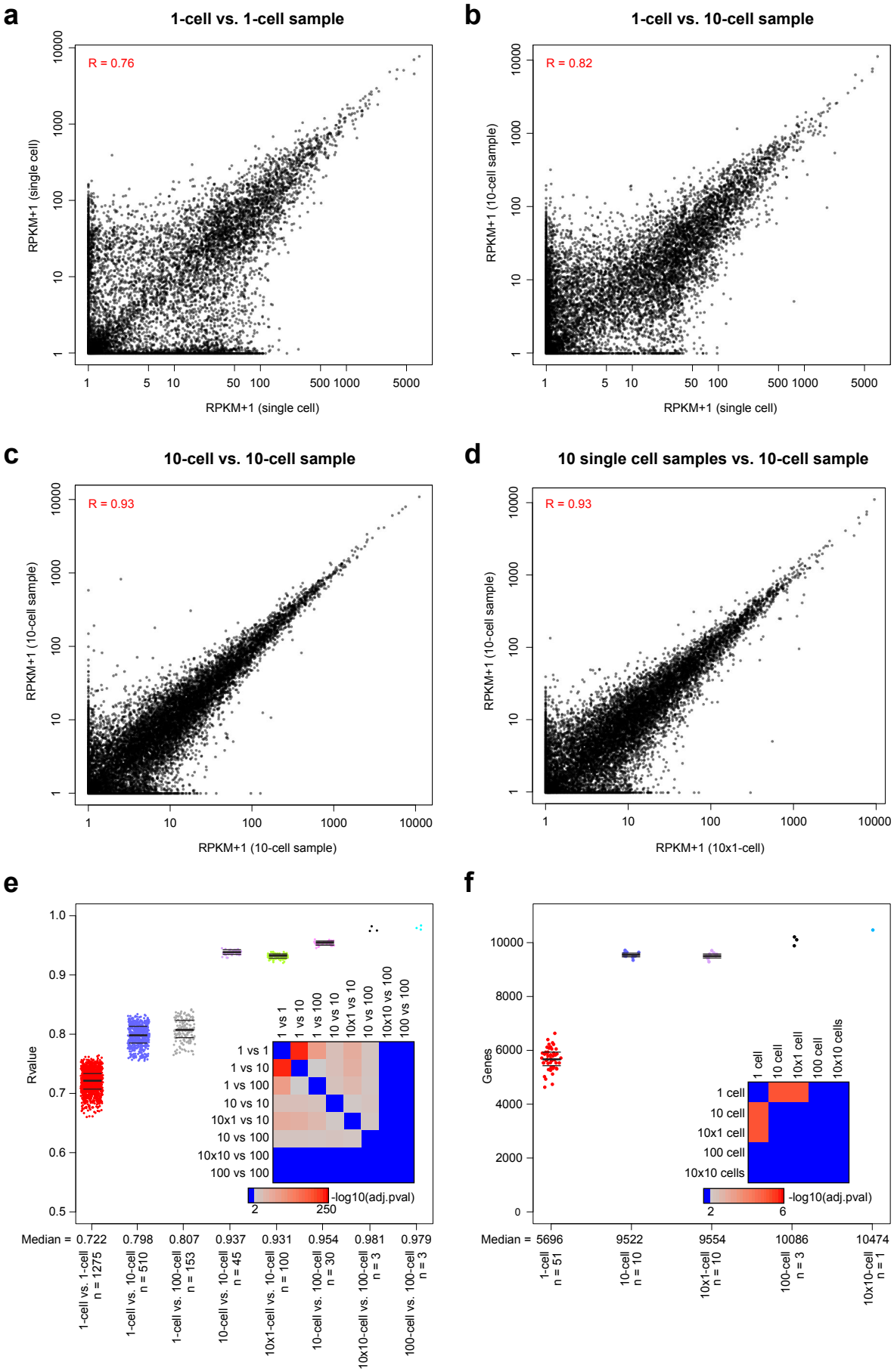
Supplementary Figure 4. QC based on spike-in *ERCC* RNAs. (**a**) Plots of *ERCC* RPKM values (where the RPKM values were calculated with respect to total reads mapped to *ERCC* RNAs only, $N = 92$ species) versus putative number of molecules for two different cells shows good linearity as determined by R^2 value (in parentheses, R^2 value was calculated using only the 38 *ERCC* RNA species present at > 1 molecule per sample) and slope close to 1. (**b**) Same as (a), but aggregated for all 1679 cells. Error bars represent SEM. (**c**) Percentage of times a given *ERCC* RNA species was detected (out of 1679 cells) versus putative molecule count. Red line shows the expected detection based on Poisson statistics of dilution. Blue and green lines indicate 1 and 10 molecules, respectively, while the orange line indicates 90% detection. Assuming that *ERCC* spike-ins follow Poisson statistics in dilution, an *ERCC* RNA species diluted down to one molecule per sample should be present in approximately 63% of the samples. In our samples, a single molecule of *ERCC* RNA, which is about 500-2000 nucleotides long, is detected ~14.7% of the time. This suggests that our method reliably detects ~23% of all molecules, given Poisson statistics. (**d**) Clustered heatmap showing Pearson's correlation R values based on *ERCC* RPKM values for each pairwise comparison between all 1679 cells. Color bar on top indicates final cluster identity. Cells do not group into clusters related to cell types based on their *ERCC* RPKM values. (**e**) Same as (d), but with cellular genes ($N = 24,057$), showing block-like structures related to cell types, in contrast with the *ERCC*-only clustering shown in (d). (**f**) Same as (d), but with cells ordered as in (e), showing that there is no bias in *ERCC* RNA detection and quantification that is related to transcriptomic cell types.


a
Single cells, SMARTer, N=1679
10 pg cortex RNA, SMARTer, N=6
250 ng cortex RNA, TruSeq, N=3
Percent of Total Reads
84.8%
77.6%
88.5%
mRNA
genome
RNA NC
ERCC
unmapped
Median = 51.6 33.2 4.8 1.4 8
Median = 59.2 18.4 9.7 1.5 10.9
Median = 71.2 17.4 8 0.4 2.8
b
Single Cells, N=1679
Percent of total reads
unmapped
ERCC
RNA NC
genome
mRNA
Intermediate cells most similar to L4 Scnn1a
c
Core cells, N=1424
Percent of total reads
d
Core Cells, N=1424
mRNA mapping (percent of total reads)
Median =

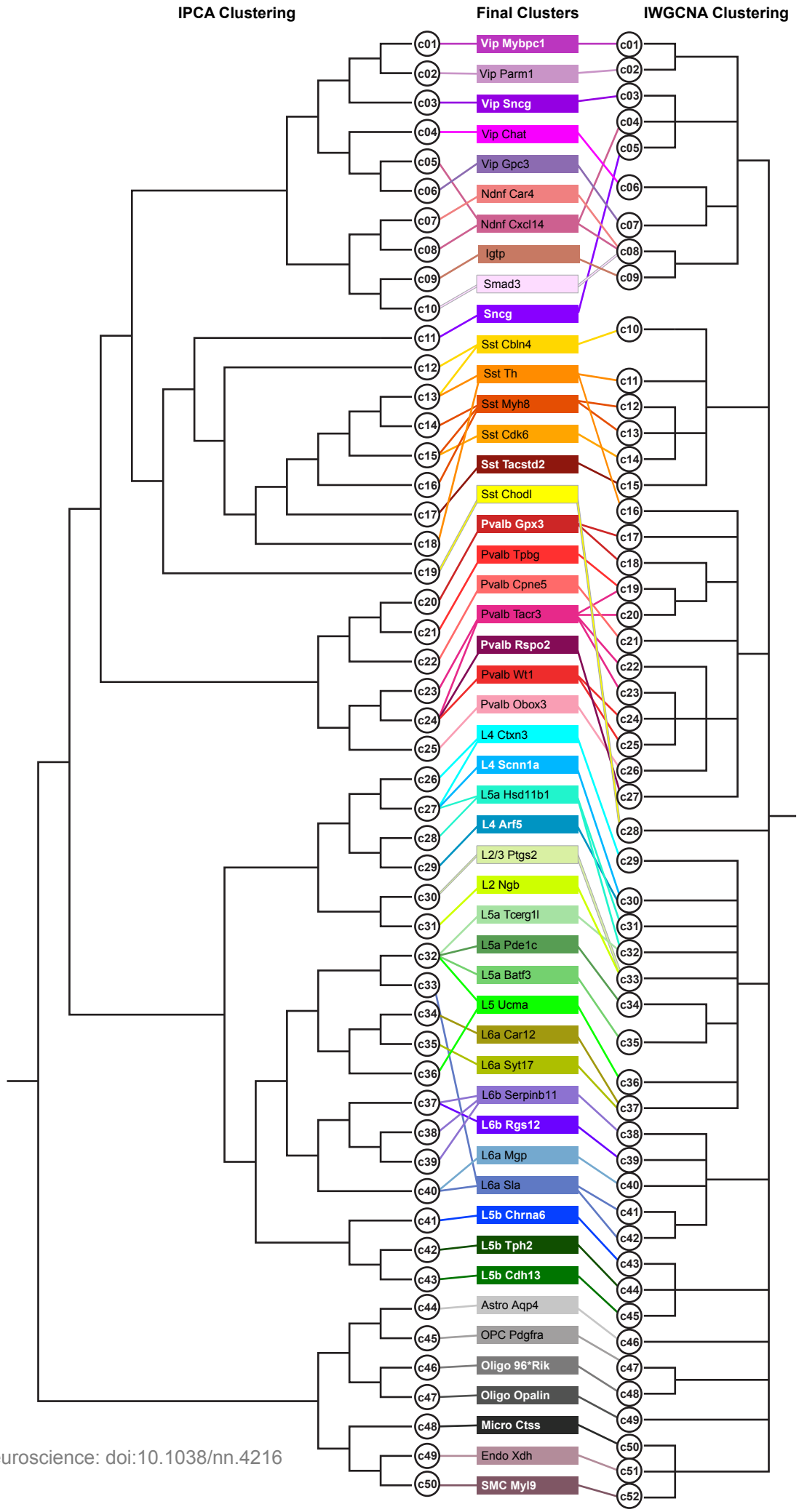
Supplementary Figure 5. Data QC. (**a**) Mapping of transcriptomic data to mRNA (RefSeq mm10 assembly), genome, non-coding RNA (RNA NC) and *ERCC* RNAs for all 1679 single cells (left), 6 replicates of 10 pg total cortex RNA processed like the single cells (middle), and 3 replicates of 250 ng of unamplified cortex RNA prepared by TruSeq (right). The samples for unamplified cortex RNA were prepared from two *Rbp4-Cre;Ai14* mice and one *Trib2-2A-CreERT2;Snap25-LSL-2A-GFP* mouse. Red dots represent medians (values reported at the bottom), whiskers represent 25^{th} and 75^{th} percentiles. (**b**) Mapping statistics for individual cells that passed the QC arranged by the cell type as defined in **Fig. 1b**. Intermediate cells are labeled white and are positioned to the right of the cell type with which they are most strongly associated by random forest classification. (**c**) Mean mapping percentages of each category described in (a) for all 49 cell types based on 1424 core cells. (**d**) Percent of total reads mapping to mRNA for all 1424 core cells for all 49 cell types. Red dots represent medians (values reported at the bottom). Whiskers represent 25^{th} and 75^{th} percentiles.


a
GABAergic cells
Genes detected
Total read depth
Vip Mybpc1
Vip Parm1
Vip Sncg
Vip Chat
Vip Gpc3
Ndnf Car4
Ndnf Cxcl14
Igtp
Smad3
Sncg
Sst Cbln4
Sst Th
Sst Myh8
Sst Cdk6
Sst Tacstd2
Sst Chodl
Pvalb Gpx3
Pvalb Tpbg
Pvalb Cpne5
Pvalb Tacr3
Pvalb Rspo2
Pvalb Wt1
Pvalb Obox3
b
Glutamatergic cells
L4 Ctxn3
L4 Scnn1a
L5a Hsd11b1
L4 Arf5
L2/3 Ptgs2
L2 Ngb
L5a Tcerg1l
L5a Pde1c
L5a Batf3
L5 Ucma
L6a Car12
L6a Syt17
L6b Serpinb11
L6b Rgs12
L6a Mgp
L6a Sla
L5b Chrna6
L5b Tph2
L5b Cdh13
c
Non-neuronal cells
Astro Aqp4
OPC Pdgfra
Oligo 96*Rik
Oligo Opalin
Micro Ctss
Endo Xdh
SMC Myl9
d
Sst Cbln4 and L4 Scnn1a cells
Sst Cbln4
L4 Scnn1a

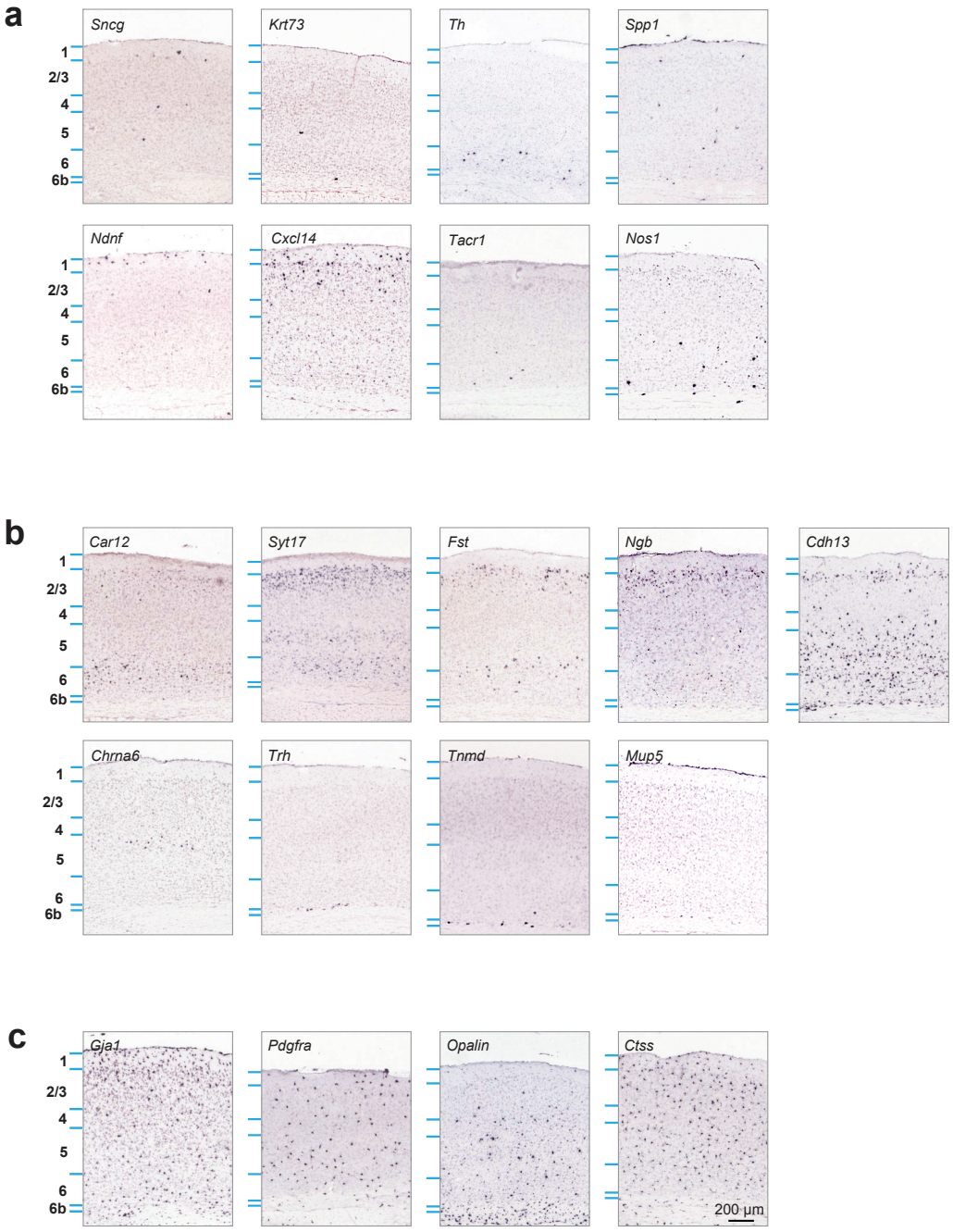
Supplementary Figure 6. Gene detection and sequencing depth. Plots showing the number of genes detected (≥ 1 read) for each of the (**a**) GABAergic (N = 23), (**b**) glutamatergic (N = 19), and (**c**) non-neuronal (N = 7) transcriptomic cell types as a function of post-alignment subsampling to a specified number of total reads. Each curve represents the mean number of genes detected over all the cells in that group, and error bars represent SEM. (**d**) Comparison of the number of genes detected (≥ 1 read) for two representative cell types upon post-alignment subsampling (lines) or upon subsampling raw reads and rerunning the alignment (dots). The minor differences between the two approaches for subsampling on gene detection suggest that the computationally simpler post-alignment subsampling is a valid way to simulate subsampling of raw reads.


a
1-cell vs. 1-cell sample
R = 0.76
RPKM+1 (single cell)
RPKM+1 (single cell)
b
1-cell vs. 10-cell sample
R = 0.82
RPKM+1 (10-cell sample)
RPKM+1 (single cell)
c
10-cell vs. 10-cell sample
R = 0.93
RPKM+1 (10-cell sample)
RPKM+1 (10-cell sample)
d
10 single cell samples vs. 10-cell sample
R = 0.93
RPKM+1 (10-cell sample)
RPKM+1 (10x1-cell)
e
Rvalue
1 vs 1
1 vs 10
1 vs 100
10 vs 10
10x1 vs 10
10 vs 100
10x10 vs 100
100 vs 100
-log10(adj.pval)
Median = 0.722 0.798 0.807 0.937 0.931 0.954 0.981 0.979
1-cell vs. 1-cell n = 1275
1-cell vs. 10-cell n = 510
1-cell vs. 100-cell n = 153
10-cell vs. 10-cell n = 45
10x1-cell vs. 10-cell n = 100
10-cell vs. 100-cell n = 30
10x10-cell vs. 100-cell n = 3
100-cell vs. 100-cell n = 3
f
Genes
1 cell
10 cell
10x1 cell
100 cell
10x10 cells
-log10(adj.pval)
Median = 5696 9522 9554 10086 10474
1-cell n = 51
10-cell n = 10
10x1-cell n = 10
100-cell n = 3
10x10-cell n = 1

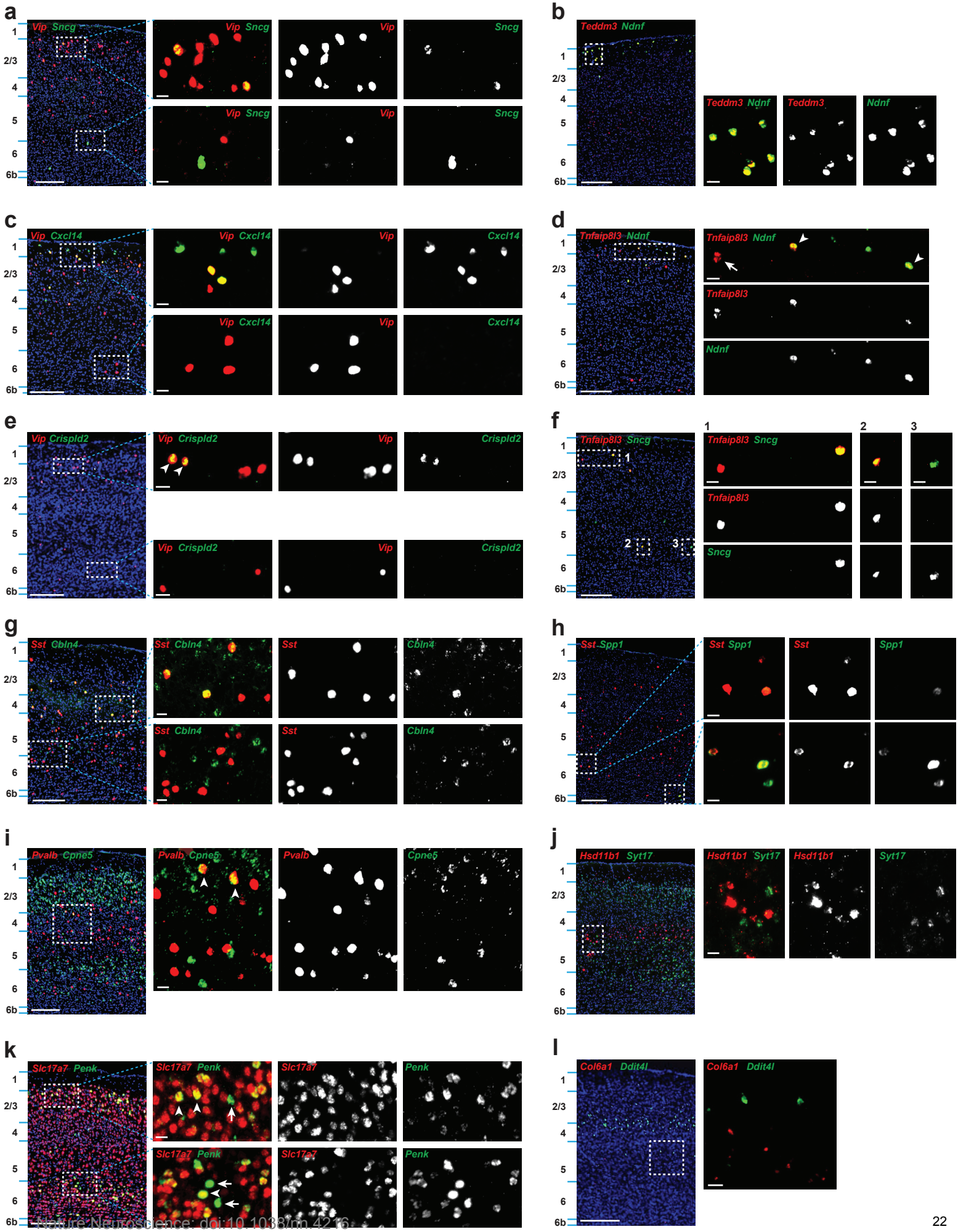
Supplementary Figure 7. Gene detection in single cells and cell populations. Comparison of RNA-seq data generated by the SMARTer/Nextera approach from individual *tdT*$^+$ cells and small populations of *tdT*$^+$ cells isolated from layer 6 of VISp in the *Ntsr1-Cre;Ai14* line. Examples of gene expression correlation between (**a**) two single cell samples, (**b**) one single cell sample and one 10-cell population, (**c**) two 10-cell populations, and (**d**) ten single cell samples pooled computationally and one 10-cell population. All samples were subsampled down to 5 million mapped reads, total number of genes for all comparisons is 24,057. (**e**) Distributions of Pearson's R values (on log-transformed data) for all pairwise comparisons between 77 single cell samples, ten 10-cell samples, three 100-cell samples, and ten computationally pooled single cell samples of 10 cells; n indicates the number of such pairwise comparisons in each group. Statistical significance between distributions of R values was evaluated by Mann-Whitney test with Bonferroni correction and is represented as a heatmap at the bottom-right corner of the panel. The medians for Pearsons's R values for the "10-cell vs. 10-cell" and "10×1 cell vs. 10-cell" comparisons, while statistically significantly different, are less than 0.01 apart (0.937 and 0.931, respectively), indicating that computational pooling of the data from 10 individual cells provides essentially the same information as profiling 10 cells together in an experimental batch. Black bars represent medians, whiskers represent 25th and 75th percentiles. (**f**) Genes detected (RPKM ≥1) in a single cell samples, 10-cell samples, 100-cell samples, computationally pooled ten single cells, and computationally pooled ten 10-cell samples. The difference in gene detection between single cells and 10-cell samples is eliminated when ten single cells are pooled computationally, suggesting the lower gene detection in single cell samples is due to biological variation rather than technical issues due to limited sensitivity of the employed method. Computationally pooled samples are labeled as: 10×1, ten single-cell samples pooled together; 10×10, ten samples derived from 10-cell populations pooled together. Black bars represent medians, whiskers represent 25th and 75th percentiles.


IPCA Clustering
Final Clusters
IWGCNA Clustering
Vip Mybpc1
Vip Parm1
Vip Sncg
Vip Chat
Vip Gpc3
Ndnf Car4
Ndnf Cxcl14
Igtp
Smad3
Sncg
Sst Cbln4
Sst Th
Sst Myh8
Sst Cdk6
Sst Tacstd2
Sst Chodl
Pvalb Gpx3
Pvalb Tpbg
Pvalb Cpne5
Pvalb Tacr3
Pvalb Rspo2
Pvalb Wt1
Pvalb Obox3
L4 Ctxn3
L4 Scnn1a
L5a Hsd11b1
L4 Arf5
L2/3 Ptgs2
L2 Ngb
L5a Tcerg1l
L5a Pde1c
L5a Batf3
L5 Ucma
L6a Car12
L6a Syt17
L6b Serpinb11
L6b Rgs12
L6a Mgp
L6a Sla
L5b Chrna6
L5b Tph2
L5b Cdh13
Astro Aqp4
OPC Pdgfra
Oligo 96*Rik
Oligo Opalin
Micro Ctss
Endo Xdh
SMC Myl9

Supplementary Figure 8. Cluster intersection for iterative PCA and iterative WGCNA. Schematic showing the iterative splits leading to the final set of PCA-based clusters (left) and WGCNA-based clusters (right), and their subsequent intersection to generate the final set of clusters described in the paper. For iterative PCA, every split is binary, whereas for WGCNA, the number of clusters at each iteration was determined as described in the **Methods**.

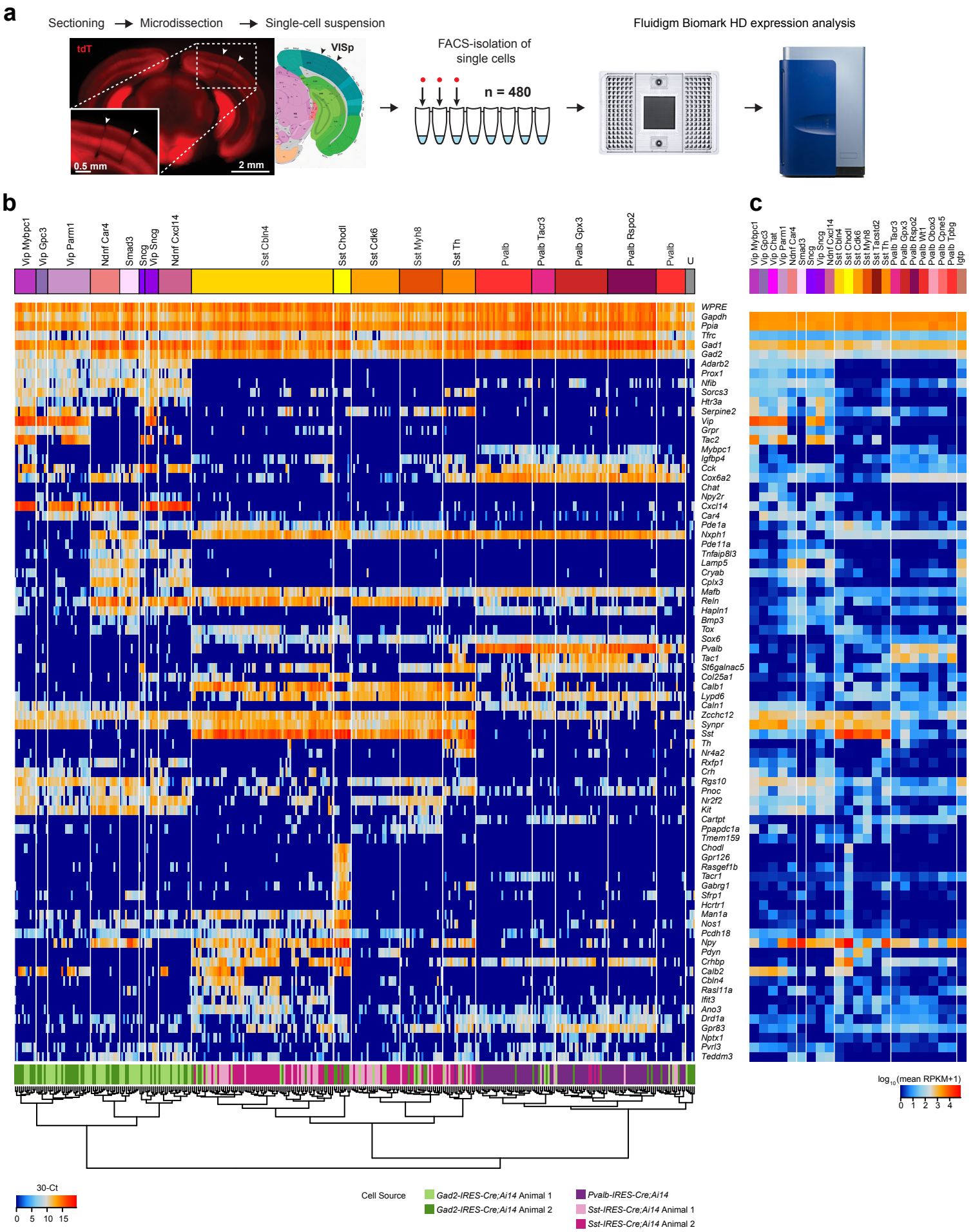

a
Sncg
Krt73
Th
Spp1
Ndnf
Cxcl14
Tacr1
Nos1
b
Car12
Syt17
Fst
Ngb
Cdh13
Chrna6
Trh
Tnmd
Mup5
c
Gja1
Pdgfra
Opalin
Ctss
1
2/3
4
5
6
6b
200 µm

Supplementary Figure 9. Chromogenic RNA *in situ* hybridization confirms select gene expression and confirms/refines spatial positioning of cell types. Images were obtained from the Allen Brain Atlas[1]. Each focuses on VISp, and is part of at least two brain-wide experiments, except for a single experiment for *Opalin*. Scale bar in the *Ctss* panel applies to all. Select genes are shown for (**a**) GABAergic, (**b**) glutamatergic and (**c**) non-neuronal cell types. (**a**) *Sncg* mRNA labels cells very sparsely distributed throughout VISp – this agrees with low-abundant Vip-Sncg and Sncg types defined by RNA-seq. As the Vip-Sncg type is enriched in upper layers (**Supplementary Table 5**), the lower layer *Scng*$^{+}$ cells likely belong to the Sncg type. *Krt73* is expressed in a very rare set of cells mostly in lower layers of VISp. As *Krt73* is shown by RNA-seq to be present in a subset of cells of the Sncg type, the *Krt73* ISH agrees with the enrichment of the Sncg type in lower layers of VISp. *Th* mRNA labels cells enriched in lower layers in agreement with its unique expression in the Sst-Th and Pvalb-Gpx3 types, which are predominantly located in lower cortical layers (**Fig. 2b, Supplementary Table 5**). *Spp1* is expressed in a small set of cells dispersed throughout VISp. This agrees with RNA-seq, as only subsets of cells within the Sst-Th and SMC-Myl9 types express this marker. The cells along the pia may belong to the SMC-Myl9 type. *Ndnf* (*A930038C07Rik*) mRNA is expressed strongly in L1 in agreement with its RNA-seq expression in Ndnf types, which are enriched in upper layers (**Supplementary Table 5**). *Cxcl14* mRNA is expressed mostly in upper-layers in agreement with its RNA-seq-based expression in Ndnf and Vip types that are enriched in upper layers (**Supplementary Table 5**). *Cxcl14* mRNA is also expressed in small cell bodies throughout VISp – those, in agreement with RNA-seq data, likely respresent astrocytes. *Tacr1* mRNA sparsely labels cells mostly confined to L5 and 6, and since the majority of *Tacr1*$^{+}$ cells belong to Sst-Chodl type, this suggests Sst-Chodl cells are enriched in L5/6. In agreement with this, *Nos1* mRNA strongly labels cells enriched in lower layers and based on RNA-seq is strongly expressed in the Sst-Chodl type. Therefore, the Sst-Chodl type is likely enriched in lower layers, based *Tacr1* and *Nos1* ISH. (**b**) In agreement with RNA-seq data, mRNAs for *Car12*, *Syt17*, *Fst*, and *Ngb* are expressed in subsets of L6 cells. These likely correspond to L6a-Car12 type (labeled by *Car12*), and L6a-Syt17 type (labeled by *Syt17*, *Fst*, and *Ngb*). *Syt17* is also expressed in L2/3 corresponding to L2-Ngb and L2/3-Ptgs2 types, and sparsely in L5, corresponding to L5a-Syt17, and L5b-Tph types. *Fst*, *Ngb* and *Cdh13* are expressed in superficial L2/3 cells, corresponding to the L2-Ngb type. *Chrna6* is expressed in a very small subset of L5 cells, corresponding to the L5b-Chrna6 type. *Trh*, *Tnmd* and *Mup5* are expressed in subsets of L6b cells. (**c**) Expression of several non-neuronal markers showing typical non-neuronal labeling: *Gja1*, astrocytes; *Pdgfra*, OPCs; *Opalin*, oligodendrocytes (note white matter-enrichment below L6b); *Ctss*, microglia. Mean RNA-seq expression for each gene in this figure within each transcriptomic cell type is shown in **Supplementary Fig. 12**. To examine gene expression determined by RNA-seq in individual cells within any of the types, refer to the online visualization tool via the Allen Brain Atlas data portal (http://casestudies.brain-map.org/celltax).

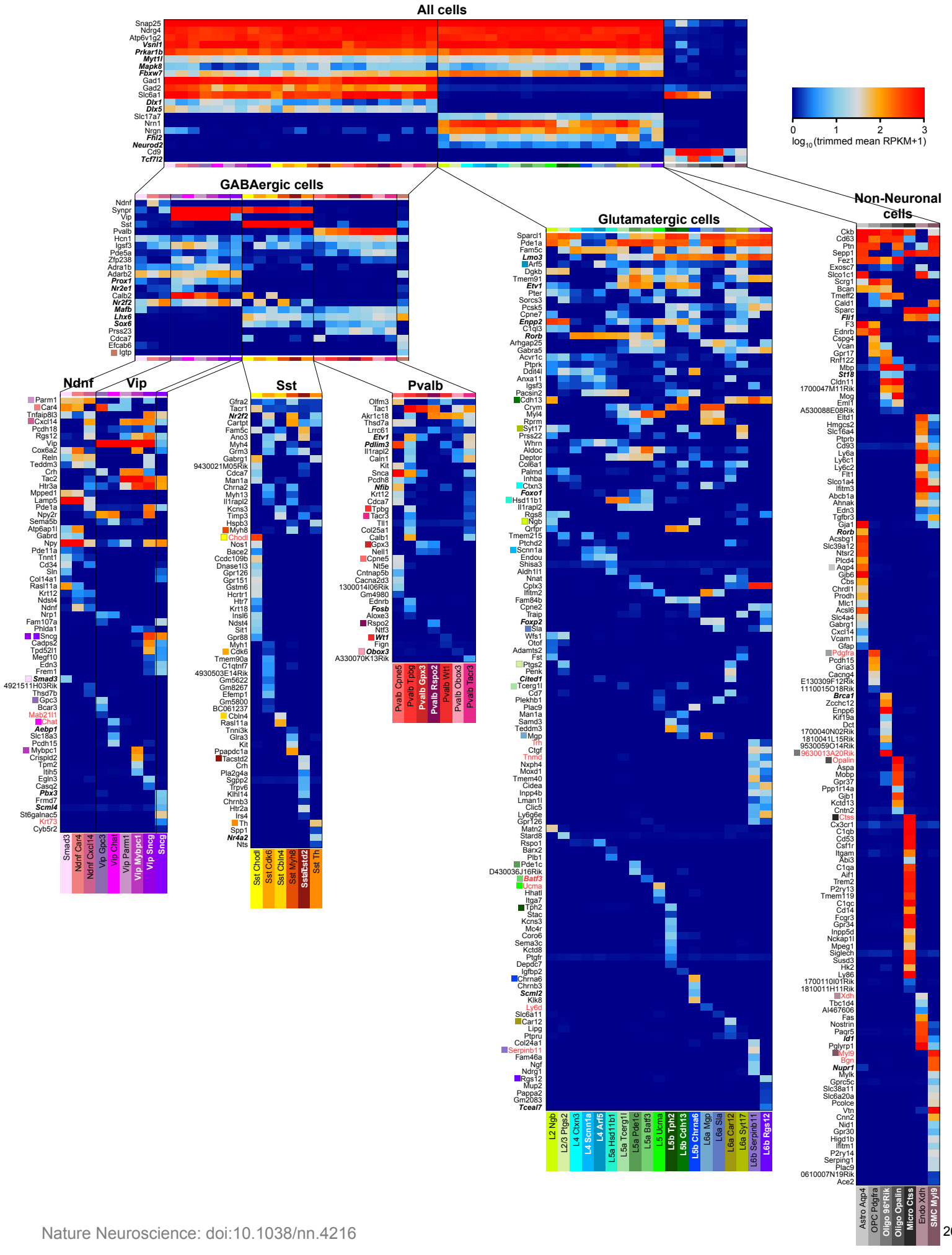


Nature Neuroscience: doi:10.1038/nn.4216

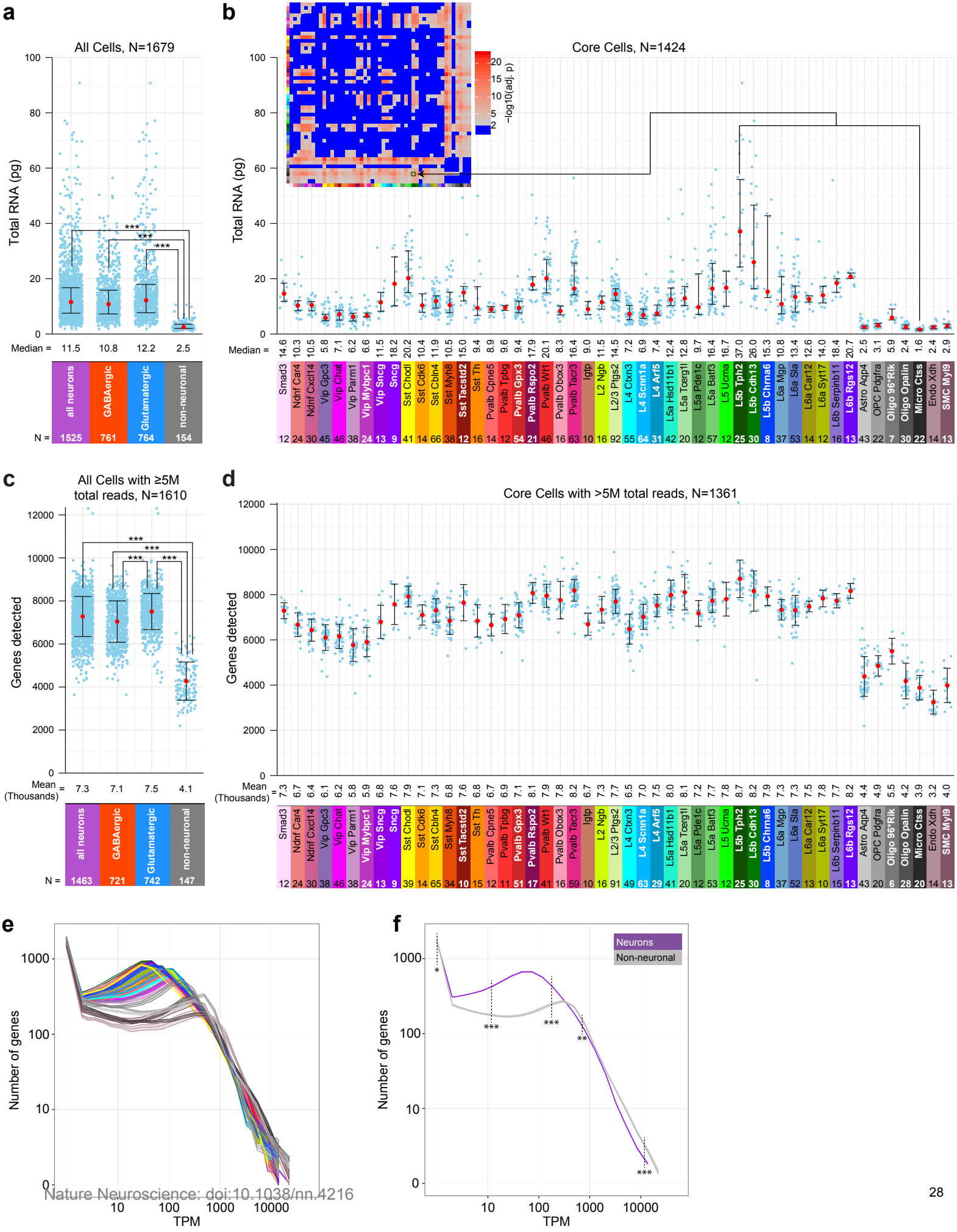
Supplementary Figure 10. Double-label fluorescence RNA *in situ* hybridization (DFISH) confirms coexpression, mutually exclusive expression, and spatially restricted expression of select genes. (**a**) *Sncg* mRNA is expressed sparsely in VISp: in a subset of *Vip*$^+$ cells in upper layers, and independently from *Vip* in lower layers, likely corresponding to Vip-Sncg and Sncg types, respectively. (**b**) *Teddm3* (*2310042E22Rik*) and *Ndnf* (*A930038C07Rik*) are co-localized in L1, corresponding to cells from Ndnf and Smad3 types. *Teddm3* also labels cells in L5, likely corresponding to L5b-Tph2 and L5b-Cdh13 types. (**c**) *Cxcl14* mRNA is expressed in a subset of *Vip*$^+$ cells only in upper cortical layers that most likely correspond to the Vip-Parm1, Vip-Mybpc1, and Vip-Sncg cell types. In lower layers, *Vip*$^+$ cells, do not express *Cxcl14*, likely corresponding to the Vip-Gpc3 type. (**d**) *Tnfaip8l3* and *Ndnf* are coexpressed in upper layers and likely correspond to the Ndnf types (arrowheads). *Tnfaip8l3*$^+$*/Ndnf*$^-$ neurons (arrow) are also present, and most likely represent the Vip-Sncg, Sncg, and Igtp types. (**e**) *Crispld2* mRNA is expressed only in *Vip*$^+$ cells enriched in upper cortical layers (arrowheads) that most likely correspond to the Vip-Mybpc1 type. In lower layers, *Crispld2* is not coexpressed with *Vip*. (**f**) *Tnfaip8l3* and *Sncg* are coexpressed in cells that most likely correspond to the Vip-Sncg and Sncg types. (**g**) *Sst* and *Cbln4* mRNAs are coexpressed in a subset of *Sst*$^+$ cells in upper layers only, likely corresponding to the Sst-Cbln4 type. In lower layers, *Sst* and *Cbln4* are mutually exclusive. *Cbln4* is also expressed in many glutamatergic cell types. (**h**) *Spp1* is expressed in a subset of *Sst*$^+$ cells, likely corresponding to the Sst-Th type. (**i**) Coexpression of *Pvalb* and *Cpne5* mRNAs in rare upper-layer cells (arrowheads) likely corresponds to the Pvalb-Cpne5 type. *Cpne5* is also expressed in other non-Pvalb GABAergic and many glutamatergic cells. (**j**) *Hsd11b1* and *Syt17* are mostly mutually exclusively expressed in L5. (**k**) *Penk* is expressed in a subset of glutamatergic cells (labelled by pan-glutamatergic marker *Slc17a7*, arrowheads) in L2/3 and in L6, likely corresponding to L2/3-Ptgs2 and L6a-Car12 types. *Penk* is also expressed in some GABAergic cells (*Slc17a7*$^-$ cells, arrows) of Vip and Sst types. (**l**) *Col6a1* and *Ddit4l* mRNAs are mutually exclusively expressed in L5b cells. White boxes indicate magnified regions. Scale bars are 200 µm in low-magnification images and 20 µm in high-magnification images. Sequence information for DFISH probes is available in **Supplementary Table 13**. Each image is representative of a single experiment containing at least two independent slides; each slide included at least 2 coronal brain sections containing VISp. Mean RNA-seq expression for each gene in this figure within each transcriptomic cell type is shown in **Supplementary Fig. 12**. To examine gene expression determined by RNA-seq in individual cells within any of the types, refer to the online visualization tool via the Allen Brain Atlas data portal (http://casestudies.brain-map.org/celltax).


a
Sectioning → Microdissection → Single-cell suspension
Fluidigm Biomark HD expression analysis
tdT
VISp
0.5 mm
2 mm
FACS-isolation of single cells
n = 480
b
c
Vip Mybpc1
Vip Gpc3
Vip Parm1
Ndnf Car4
Smad3
Sncg
Vip Sncg
Ndnf Cxcl14
Sst Cbln4
Sst Chodl
Sst Cdk6
Sst Myh8
Sst Th
Pvalb
Pvalb Tacr3
Pvalb Gpx3
Pvalb Rspo2
Pvalb
U
Vip Chat
Ndnf Car4
Smad3
Vip Sncg
Ndnf Cxcl14
Sst Cbln4
Sst Chodl
Sst Cdk6
Sst Myh8
Sst Tacstd2
Sst Th
Pvalb Tacr3
Pvalb Gpx3
Pvalb Rspo2
Pvalb Wt1
Pvalb Obox3
Pvalb Cpne5
Pvalb Tpbg
Igtp
WPRE
Gapdh
Ppia
Tfrc
Gad1
Gad2
Adarb2
Prox1
Nfib
Sorcs3
Htr3a
Serpine2
Vip
Grpr
Tac2
Mybpc1
Igfbp4
Cck
Cox6a2
Chat
Npy2r
Cxcl14
Car4
Pde1a
Nxph1
Pde11a
Tnfaip8l3
Lamp5
Cryab
Cplx3
Mafb
Reln
Hapln1
Bmp3
Tox
Sox6
Pvalb
Tac1
St6galnac5
Col25a1
Calb1
Lypd6
Caln1
Zcchc12
Synpr
Sst
Th
Nr4a2
Rxfp1
Crh
Rgs10
Pnoc
Nr2f2
Kit
Cartpt
Ppapdc1a
Tmem159
Chodl
Gpr126
Rasgef1b
Tacr1
Gabrg1
Sfrp1
Hcrtr1
Man1a
Nos1
Pcdh18
Npy
Pdyn
Crhbp
Calb2
Cbln4
Rasl11a
Ifit3
Ano3
Drd1a
Gpr83
Nptx1
Pvrl3
Teddm3
log10(mean RPKM+1)
0 1 2 3 4
30-Ct
0 5 10 15
Cell Source
Gad2-IRES-Cre;Ai14 Animal 1
Gad2-IRES-Cre;Ai14 Animal 2
Pvalb-IRES-Cre;Ai14
Sst-IRES-Cre;Ai14 Animal 1
Sst-IRES-Cre;Ai14 Animal 2

Supplementary Figure 11. Quantitative RT-PCR confirms coexpression or mutually exclusive expression of marker genes identified by RNA-seq. (**a**) Schematic of the workflow for cell isolation and qRT-PCR profiling using the Fluidigm Biomark system. TdT^+ cells were isolated from the *Gad2-IRES-Cre;Ai14* (*N* = 2 animals), *Sst-IRES-Cre;Ai14* (*N* = 2 animals), and *Pvalb-IRES-Cre;Ai14* (*N* = 1 animal) transgenic lines as described in the **Methods**. (**b**) qRT-PCR expression values (30-Ct; Ct stands for 'cycle threshold') for marker genes that discriminate GABAergic types in the RNA-seq data. Single cells (*N* = 480) are represented by individual columns and are grouped by hierarchical clustering of the expression of displayed genes. The color bar above represents putative interneuron identity based on expression of key marker genes; U, unclassified. The color bar below indicates the Cre line and animal from which each individual cell was isolated. Overall, qRT-PCR recapitulates RNA-seq data for key genes that are found to be mutually exclusively expressed or coexpressed in specific subsets of cells. The major GABAergic types (Vip, Ndnf, Pvalb, and Sst) are identified according to assays for the corresponding genes, with the exception of the Ndnf type, which can be identified by expression of *Lamp5*. Among the Vip types, key discriminatory markers include *Tac2*, *Mybpc1*, and *Car4*. Ndnf types can be distinguished from each other by coexpression of *Cox6a2* and *Car4* or *Npy2r* and *Pde1a*. Similar to Ndnf types, Sncg and Vip-Sncg types are labeled by expression of *Pde1a* and *Tnfaip8l3*, but they do not express *Lamp5*. The Smad3 type is identified by coexpression of *Sfrp1* and *Rasl11a*. Coexpression of *Tac1*, *St6galnac5*, *Col25a1*, and *Calb1* is expected in the Pvalb-Tacr3 type, while Pvalb-Rspo2 and Pvalb-Gpx3 types are marked by expression of *Tac1*, *St6galnac5*, and *Lypd6*, but no expression of *Col25a1* and *Calb1*. Pvalb-Gpx3 can be distinguished from Pvalb-Rspo2 by more consistent expression of *Zcchc12*. Other Pvalb types cannot be clearly distinguished by these assays. Among Sst types, *Kit* is only expressed in the Sst-Myh8 type. The Sst-Cdk6 type is identified by expression of *Nr2f2* and absence of *Kit*. In accordance with the RNA-seq data for the Sst-Chodl transcriptomic type, *Chodl*, *Tacr1*, *Gpr126* and *Gabrg1* are specifically coexpressed. The Sst-Tacstd2 type cannot not be distinguished based on these assays. The Sst-Cbln4 type is identified by coexpression of *Cbln4* and *Rasl11a*. *WPRE* is a control probe to determine the expression of *tdTomato-WPRE* mRNA; WPRE stands for woodchuck hepatitis virus posttranscriptional regulatory element. *Lamp5* is also known as *6330527O06Rik*. *Teddm3* is also known as *2310042E22Rik*. qRT-PCR primer and probe sequences are listed in **Supplementary Table 14**. (**c**) Expression of the same genes as in (b) according to RNA-seq data. Each column corresponds to a GABAergic cell type (*N* = 23), with $\log_{10}$(mean RPKM+1) plotted for each gene within that type.

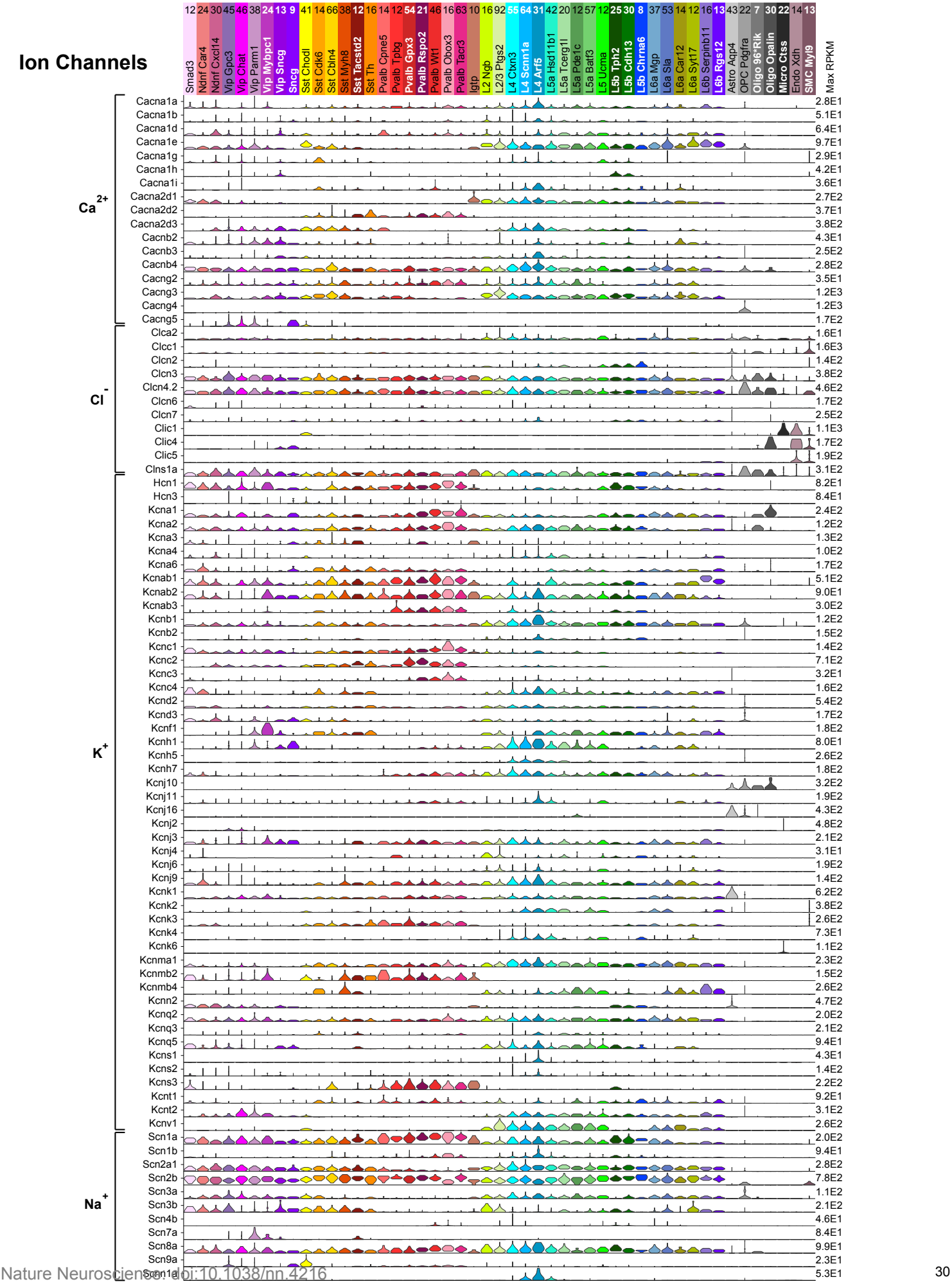

All cells
GABAergic cells
Glutamatergic cells
Non-Neuronal cells
Ndnf
Vip
Sst
Pvalb
0 1 2 3
log10(trimmed mean RPKM+1)
L2 Ngb
L2/3 Ptgs2
L4 Ctxn3
L4 Scnn1a
L4 Arf5
L5a Hsd11b1
L5a Tcerg1l
L5a Pde1c
L5a Batf3
L5 Ucma
L5b Tph2
L5b Cdh13
L5b Chrna6
L6a Mgp
L6a Sla
L6a Car12
L6a Syt17
L6b Serpinb11
L6b Rgs12
Astro Aqp4
OPC Pdgfra
Oligo 96°Rik
Oligo Opalin
Micro Ciss
Endo Xdh
SMC Myl9

Supplementary Figure 12. Hierarchically organized marker genes. Marker gene expression (25% trimmed mean RPKM within each type) represented at different levels of cell type taxonomy. Most discriminating genes were selected as described in the **Methods**, and were arranged hierarchically to illustrate a gene code for all 49 cortical cell types. Additional genes from the literature or discovered by the authors were manually added. The marker genes, which were included into the names of cell types are labeled with a colored flag corresponding to that cell type. Unique markers are labeled red, and transcription factor genes are bold and italicized. Many transcription factors listed here have been previously implicated in development, specification or function of specific cell types. For example, *Lhx6*, which is expressed during the development of medial ganglionic eminence-derived GABAergic neurons, is detected in Sst and Pvalb transcriptomic types, as expected, but also shows robust expression in the Igtp GABAergic type. Similarly, *Prox1* is expressed during the development of caudal ganglionic eminence (CGE)-derived GABAergic neurons, and is detected as expected in the Vip and Ndnf transcriptomic types (see also **Supplementary Fig. 11** for confirmation of *Prox1* expression by qRT-PCR). A second reported CGE-derived neuron marker gene, *Nr2f2*, is detected in Vip and Ndnf transcriptomic types, but also shows high expression in three Sst transcriptomic types. *Ndnf* is also known as *A930038C07Rik*; *Lamp5* as *6330527O06Rik*; and *Teddm3* as *2310042E22Rik*.

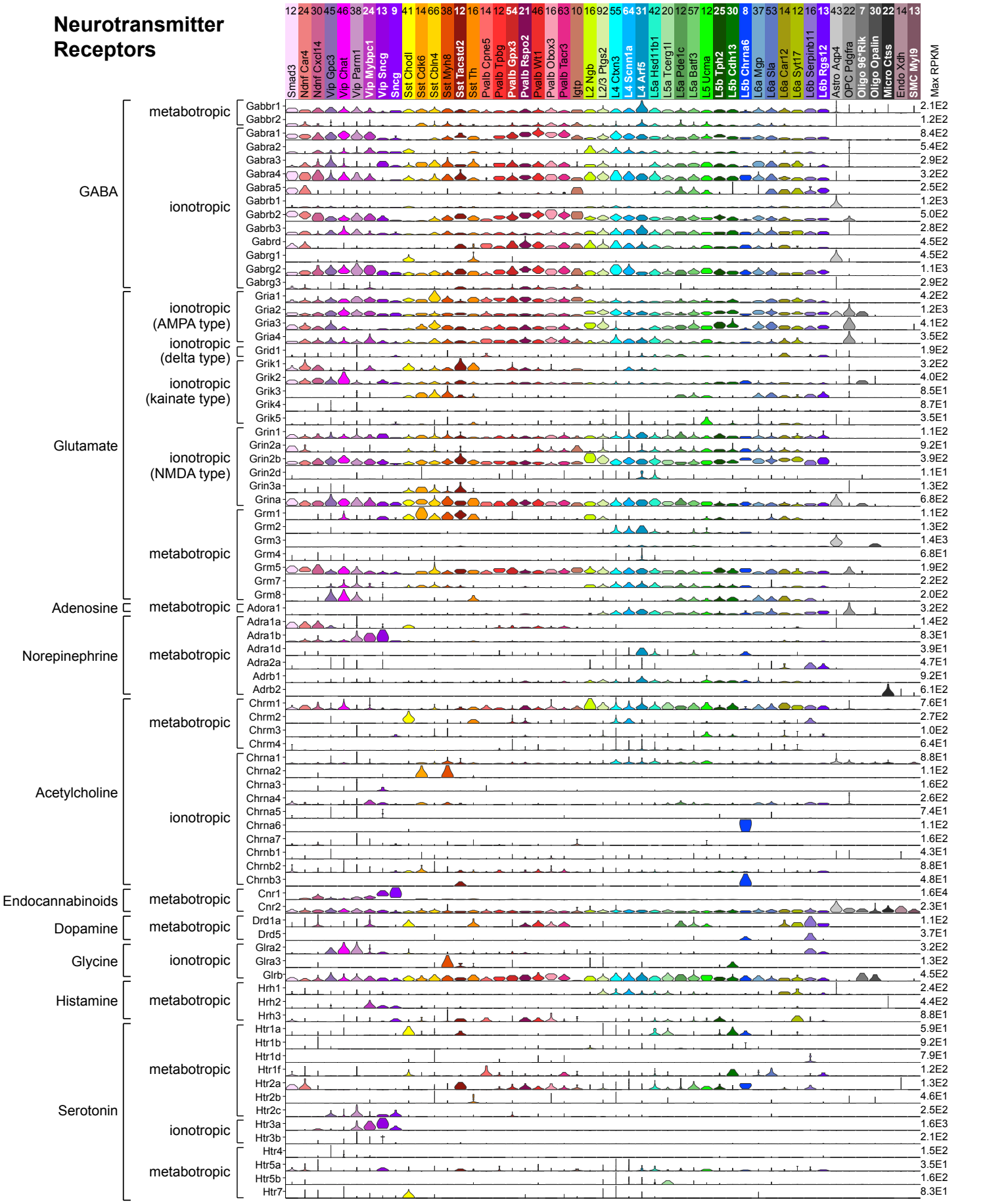


Nature Neuroscience: doi:10.1038/nn.4216

Supplementary Figure 13. RNA content, gene count and distribution of gene abundances for single cells belonging to different cell types. (**a**) Estimation of single cell RNA content based on the ratio of synthetic spike-in ERCC reads and cellular reads (**Methods**) for all cells from major cell classes (N = 1525 for all neurons, 761 for GABAergic neurons, 764 for glutamatergic neurons, and 154 for glia). Red dots represent medians, and whiskers represent 25^{th} and 75^{th} percentiles. Non-neuronal cells contain significantly less RNA than neurons ($p = 1.34 \times 10^{-82}$ for comparison to all neurons, $p = 7.03 \times 10^{-75}$ for comparison to all GABAergic neurons, and $p = 1.49 \times 10^{-76}$ for comparison to all glutamatergic neurons; Mann-Whitney test with Bonferroni correction, the corresponding degrees of freedom are: 1677, 913, and 916). *** $p < 10^{-30}$. (**b**) Same as (a) but for all cell types using only core cells (number listed at the bottom of the corresponding colored label; total N = 1424). The inset heatmap shows p-values for all pairwise Mann-Whitney tests with Bonferroni correction. The highlighted position in the heatmap corresponds to highly significant difference in RNA content between L5b-Tph2 cell type and microglia. (**c**) Average numbers of genes detected (read counts ≥ 1, values at bottom) across major classes. For this analysis, all single cell sequencing results were subsampled to 5 million total reads (69 cells that have total read depth lower than 5 million reads were excluded leaving 1610 total cells). Red dots represent means, and error bars represent standard deviation. We detect significantly fewer genes in non-neuronal cells than in neurons ($p = 8.09 \times 10^{-89}$ for comparison to all neurons, $p = 4.14 \times 10^{-89}$ for comparison to all GABAergic neurons, and $p = 3.02 \times 10^{-98}$ for comparison to all glutamatergic neurons; t-test with unequal variances and Bonferroni correction, the corresponding degrees of freedom are: 1608, 866, and 887). We also detect significantly more genes in glutamatergic than in GABAergic neurons ($p = 1.80 \times 10^{-21}$, t-test with unequal variances and Bonferroni correction, 1461 degrees of freedom). *** $p < 10^{-30}$. The use of the t-test is justified by the approximately normal distribution of the genes detected within samples of a given group. (**d**) Same as (c), but for all cell types using only core cells (number listed at the bottom of the corresponding colored label; total N = 1361). (**e**) The distributions of transcripts per million (TPM) for all genes in each of the 49 transcriptomic cell types; number of core cells for each type is listed in (b). Each central line designates the mean, and each shaded region surrounding it indicates SEM. Line colors correspond to cluster colors used in (b) and (d). (**f**) Same as (e) but for all cells belonging to neuronal types (N = 1525) versus all non-neuronal cells (N = 154). Compared to neurons, non-neuronal cells exhibit significantly fewer transcripts at low and intermediate abundance, and more transcripts at high abundance. (From left to right, starred p-values are 0.018, 2.7×10^{-86}, 1.5×10^{-83}, 3.0×10^{-5}, and 2.1×10^{-14}, * $p < 0.05$, ** $p < 0.001$, *** $p < 10^{-5}$, Mann-Whitney test with Bonferroni correction, 1677 degrees of freedom). Note that it is possible that gene abundance distributions may change in the future as more complete mapping to transcriptome for all types is achieved due to better genome annotation.

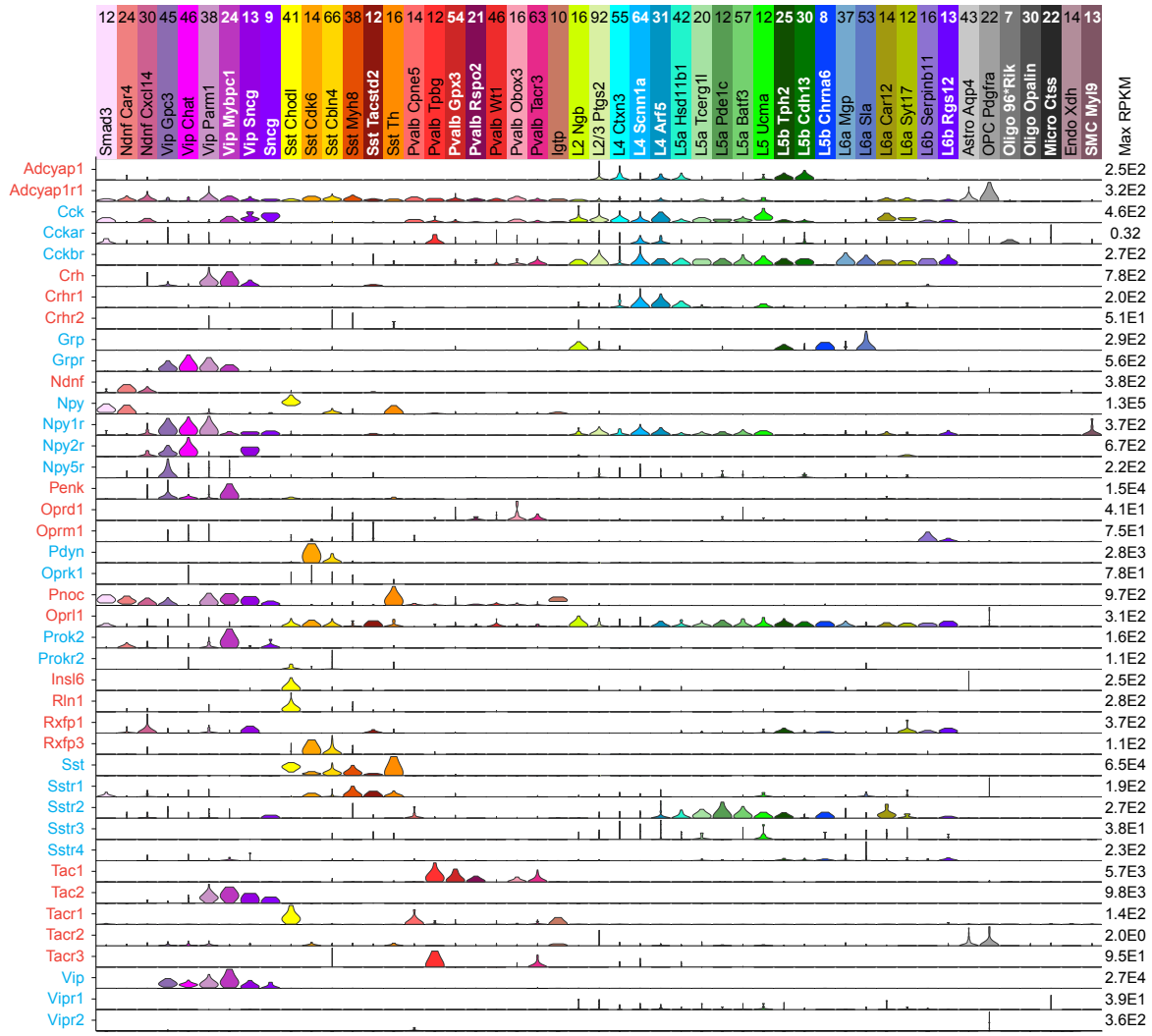

Ion Channels
Smad3
Ndnf Car4
Ndnf Cxcl14
Vip Gpc3
Vip Chat
Vip Parm1
Vip Mybpc1
Vip Sncg
Sncg
Sst Chodl
Sst Cdk6
Sst Cbln4
Sst Myh8
Sst Tacstd2
Sst Th
Pvalb Cpne5
Pvalb Tpbg
Pvalb Gpx3
Pvalb Rspo2
Pvalb Wfl
Pvalb Obox3
Pvalb Tacr3
Igtp
L2 Ngb
L2/3 Ptgs2
L4 Ctxn3
L4 Scnn1a
L4 Arf5
L5a Hsd11b1
L5a Tcerg1l
L5a Pde1c
L5a Batf3
L5 Ucma
L5b Tph2
L5b Cdh13
L5b Chrna6
L6a Mgp
L6a Sla
L6a Car12
L6a Syt17
L6b Serpinb11
L6b Rgs12
Astro Aqp4
OPC Pdgfra
Oligo 96*Rik
Oligo Opalin
Micro Ctss
Endo Xdh
SMC Myl9
Max RPKM
Ca2+
Cacna1a
Cacna1b
Cacna1d
Cacna1e
Cacna1g
Cacna1h
Cacna1i
Cacna2d1
Cacna2d2
Cacna2d3
Cacnb2
Cacnb3
Cacnb4
Cacng2
Cacng3
Cacng4
Cacng5
Cl-
Clca2
Clcc1
Clcn2
Clcn3
Clcn4.2
Clcn6
Clcn7
Clic1
Clic4
Clic5
Clns1a
K+
Hcn1
Hcn3
Kcna1
Kcna2
Kcna3
Kcna4
Kcna6
Kcnab1
Kcnab2
Kcnab3
Kcnb1
Kcnb2
Kcnc1
Kcnc2
Kcnc3
Kcnc4
Kcnd2
Kcnd3
Kcnf1
Kcnh1
Kcnh5
Kcnh7
Kcnj10
Kcnj11
Kcnj16
Kcnj2
Kcnj3
Kcnj4
Kcnj6
Kcnj9
Kcnk1
Kcnk2
Kcnk3
Kcnk4
Kcnk6
Kcnma1
Kcnmb2
Kcnmb4
Kcnn2
Kcnq2
Kcnq3
Kcnq5
Kcns1
Kcns2
Kcns3
Kcnt1
Kcnt2
Kcnv1
Na+
Scn1a
Scn1b
Scn2a1
Scn2b
Scn3a
Scn3b
Scn4b
Scn7a
Scn8a
Scn9a
Scn10a
2.8E1
5.1E1
6.4E1
9.7E1
2.9E1
4.2E1
3.6E1
2.7E2
3.7E1
3.8E2
4.3E1
2.5E2
2.8E2
3.5E1
1.2E3
1.2E3
1.7E2
1.6E1
1.6E3
1.4E2
3.8E2
4.6E2
1.7E2
2.5E2
1.1E3
1.7E2
1.9E2
3.1E2
8.2E1
8.4E1
2.4E2
1.2E2
1.3E2
1.0E2
1.7E2
5.1E2
9.0E1
3.0E2
1.2E2
1.5E2
1.4E2
7.1E2
3.2E1
1.6E2
5.4E2
1.7E2
1.8E2
8.0E1
2.6E2
1.8E2
3.2E2
1.9E2
4.3E2
4.8E2
2.1E2
3.1E1
1.9E2
1.4E2
6.2E2
3.8E2
2.6E2
7.3E1
1.1E2
2.3E2
1.5E2
2.6E2
4.7E2
2.0E2
2.1E2
9.4E1
4.3E1
1.4E2
2.2E2
9.2E1
3.1E2
2.6E2
2.0E2
9.4E1
2.8E2
7.8E2
1.1E2
2.1E2
4.6E1
8.4E1
9.9E1
2.3E1
5.3E1

Supplementary Figure 14. Expression of ion channels in cell types. Violin plots represent the gene expression distributions (rows) among single cells within each of the 49 transcriptomic cell types (columns). Only core cells are used (N = 1424). Expression is on a linear scale and is normalized to the maximum single cell expression value (listed on the right). The following ion channel genes are not shown here due to absent, extremely low or sparse expression: *Cacna1f, Cacna1s, Cacna2d4, Cacnb1, Cacng1, Cacng6, Cacng7, Cacng8, Clca1, Clca3, Clca4, Clca5, Clca6, Clcn1, Clcn5, Clcnka, Clcnkb, Clic3, Clic6, Hcn2, Hcn4, Kcna5, Kcna7, Kcna10, Kcnd1, Kcne1, Kcne2, Kcne3, Kcne4, Kcng1, Kcng2, Kcng3, Kcng4, Kcnh2, Kcnh3, Kcnh4, Kcnh6, Kcnh8, Kcnj1, Kcnj12, Kcnj13, Kcnj14, Kcnj15, Kcnj5, Kcnj8, Kcnk5, Kcnk7, Kcnk9, Kcnk10, Kcnk12, Kcnk13, Kcnk15, Kcnk16, Kcnk18, Kcnmb1, Kcnmb3, Kcnn1, Kcnn3, Kcnn4, Kcnq1, Kcnq4, Kcnv2, Scn10a, Scn11a, Scn4a, Scn5a, Scnn1b, Scnn1g.*


Neurotransmitter Receptors
12 24 30 45 46 38 24 13 9 41 14 66 38 12 16 14 12 54 21 46 16 63 10 16 92 55 64 31 42 20 12 57 12 25 30 8 37 53 14 12 16 13 43 22 7 30 22 14 13
Smad3 Ndnf Car4 Ndnf Cxcl14 Vip Gpc3 Vip Chat Vip Parm1 Vip Mybpc1 Vip Sncg Sncg Sst Chodl Sst Cdk6 Sst Cbln4 Sst Myh8 Sst Tacstd2 Sst Th Pvalb Cpne5 Pvalb Tpbg Pvalb Gpx3 Pvalb Rspo2 Pvalb Wt1 Pvalb Obox3 Pvalb Tacr3 Igtp L2 Ngb L2/3 Ptgs2 L4 Cxcn3 L4 Scnn1a L4 Arf5 L5a Hsd11b1 L5a Tcerg1l L5a Pde1c L5a Batf3 L5 Ucma L5b Tph2 L5b Cdh13 L5b Chrna6 L6a Mgp L6a Sla L6a Car12 L6a Syt17 L6b Serpinb11 L6b Rgs12 Astro Aqp4 OPC Pdgfra Oligo 96*Rik Oligo Opalin Micro Ctss Endo Xdh SMC Myl9
Max RPKM
GABA
metabotropic
Gabbr1 2.1E2
Gabbr2 1.2E2
ionotropic
Gabra1 8.4E2
Gabra2 5.4E2
Gabra3 2.9E2
Gabra4 3.2E2
Gabra5 2.5E2
Gabrb1 1.2E3
Gabrb2 5.0E2
Gabrb3 2.8E2
Gabrd 4.5E2
Gabrg1 4.5E2
Gabrg2 1.1E3
Gabrg3 2.9E2
Glutamate
ionotropic (AMPA type)
Gria1 4.2E2
Gria2 1.2E3
Gria3 4.1E2
Gria4 3.5E2
ionotropic (delta type)
Grid1 1.9E2
ionotropic (kainate type)
Grik1 3.2E2
Grik2 4.0E2
Grik3 8.5E1
Grik4 8.7E1
Grik5 3.5E1
ionotropic (NMDA type)
Grin1 1.1E2
Grin2a 9.2E1
Grin2b 3.9E2
Grin2d 1.1E1
Grin3a 1.3E2
Grina 6.8E2
metabotropic
Grm1 1.1E2
Grm2 1.3E2
Grm3 1.4E3
Grm4 6.8E1
Grm5 1.9E2
Grm7 2.2E2
Grm8 2.0E2
Adenosine
metabotropic
Adora1 3.2E2
Norepinephrine
metabotropic
Adra1a 1.4E2
Adra1b 8.3E1
Adra1d 3.9E1
Adra2a 4.7E1
Adrb1 9.2E1
Adrb2 6.1E2
Acetylcholine
metabotropic
Chrm1 7.6E1
Chrm2 2.7E2
Chrm3 1.0E2
Chrm4 6.4E1
ionotropic
Chrna1 8.8E1
Chrna2 1.1E2
Chrna3 1.6E2
Chrna4 2.6E2
Chrna5 7.4E1
Chrna6 1.1E2
Chrna7 1.6E2
Chrnb1 4.3E1
Chrnb2 8.8E1
Chrnb3 4.8E1
Endocannabinoids
metabotropic
Cnr1 1.6E4
Cnr2 2.3E1
Dopamine
metabotropic
Drd1a 1.1E2
Drd5 3.7E1
Glycine
ionotropic
Glra2 3.2E2
Glra3 1.3E2
Glrb 4.5E2
Histamine
metabotropic
Hrh1 2.4E2
Hrh2 4.4E2
Hrh3 8.8E1
Serotonin
metabotropic
Htr1a 5.9E1
Htr1b 9.2E1
Htr1d 7.9E1
Htr1f 1.2E2
Htr2a 1.3E2
Htr2b 4.6E1
Htr2c 2.5E2
ionotropic
Htr3a 1.6E3
Htr3b 2.1E2
metabotropic
Htr4 1.5E2
Htr5a 3.5E1
Htr5b 1.6E2
Htr7 8.3E1

Supplementary Figure 15. Expression of neurotransmitter receptors in cell types. Violin plots represent the gene expression distributions (rows) among single cells within each of the 49 transcriptomic cell types (columns). Only core cells are used (N = 1424). Expression is on a linear scale and is normalized to the maximum single cell expression value (listed on the right). The following receptor genes are not shown here due to absent, extremely low or sparse expression: *Gabra6, Gabre, Gabrp, Gabrq, Gabrr1, Gabrr2, Gabrr3, Grid2, Grin3b, Grm6, Adora2a, Adora2b, Adora3, Adra2b, Adra2c, Adrb3, Chrm5, Chrna9, Chrna10, Chrnb4, Chrnd, Chrne, Chrng, Drd2, Drd3, Drd4, Glra1, Glra4, Grin2c, Hrh4, Htr6*.

Neuropeptides and Receptors

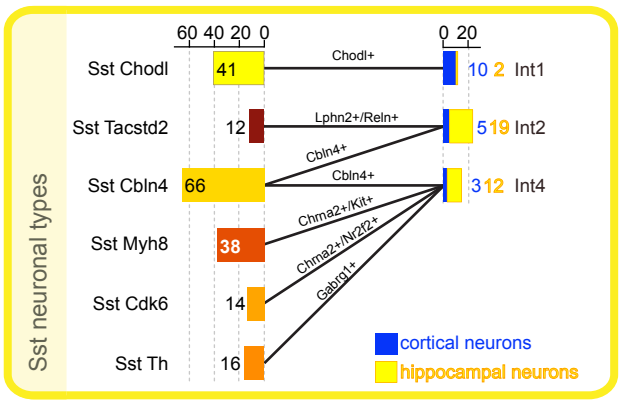


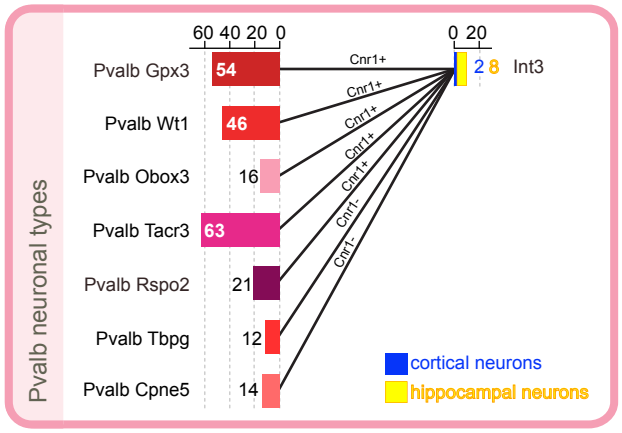
Supplementary Figure 16. Expression of neuropeptides and their receptors in cell types. Violin plots represent the gene expression distributions (rows) among single cells within each of the 49 transcriptomic cell types (columns). Only core cells are used (N = 1424). Expression is on a linear scale and is normalized to the maximum single cell expression value (listed on the right). Each neuropeptide and its receptors are grouped together in like colors, and each set alternates between red and blue. The following receptor genes are not shown here due to absent, extremely low or sparse expression: *Rxfp2*, *Rxfp4*, *Sstr5*.

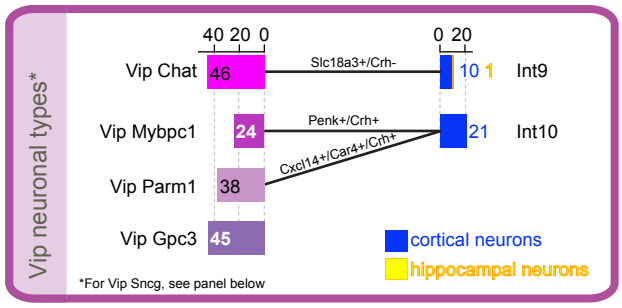
a

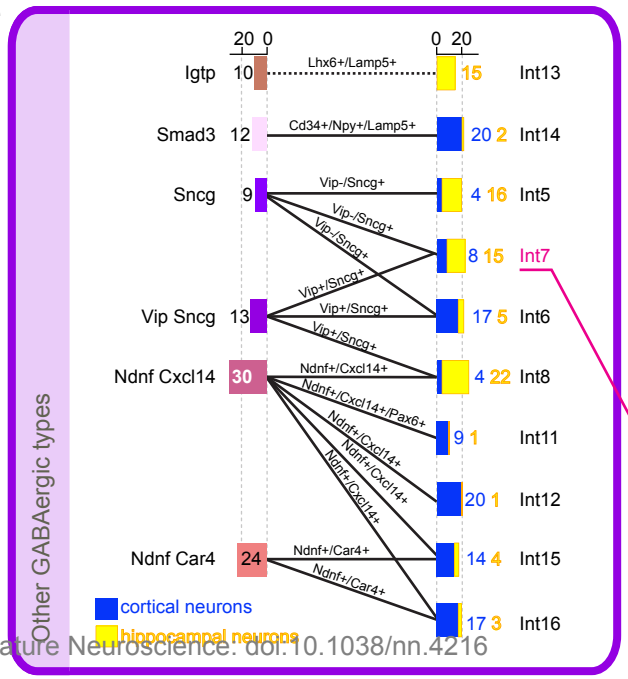
mean number of genes detected / cell number	This study		Zeisel et al. Neocortex		Zeisel et al. Hippocampus	
GABAergic neurons	**761**	7,042	**164**	4,650	**175**	4,911
Glutamatergic neurons	**764**	7,507	**399**	4,543	**939***	4,909
Non-neuronal cells	**154**	4,274	**1128**	2,494	**249**	2,645

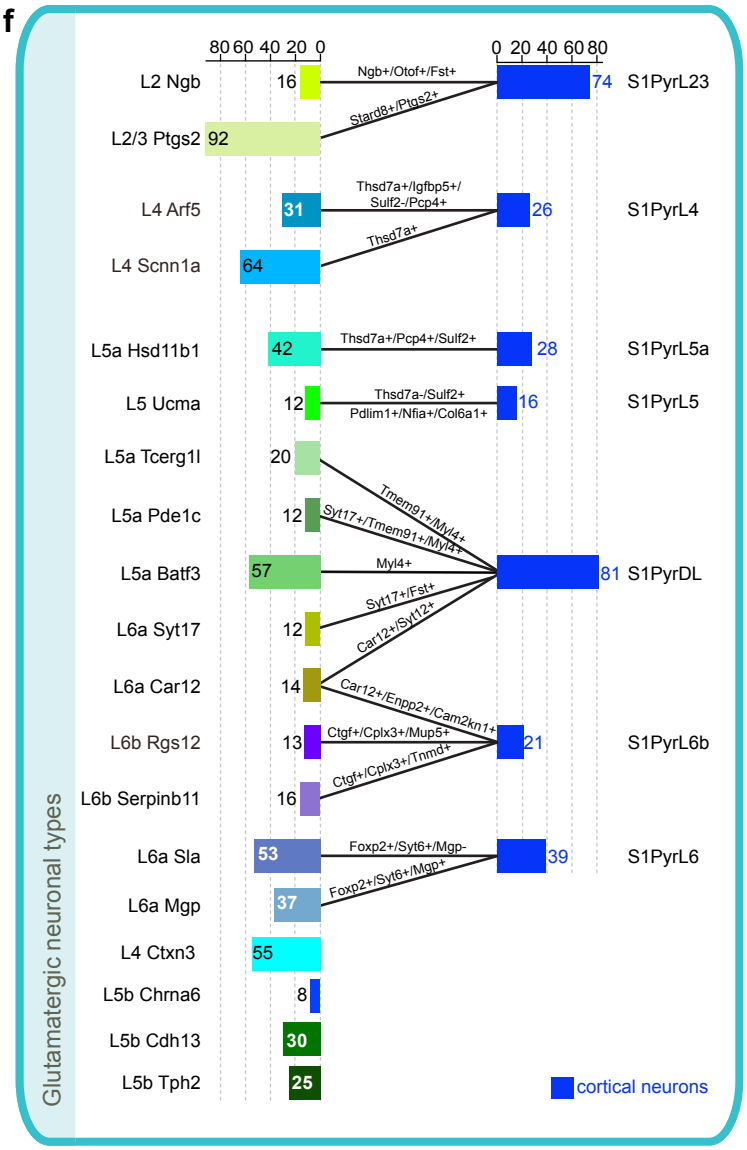
*Not used for the comparison

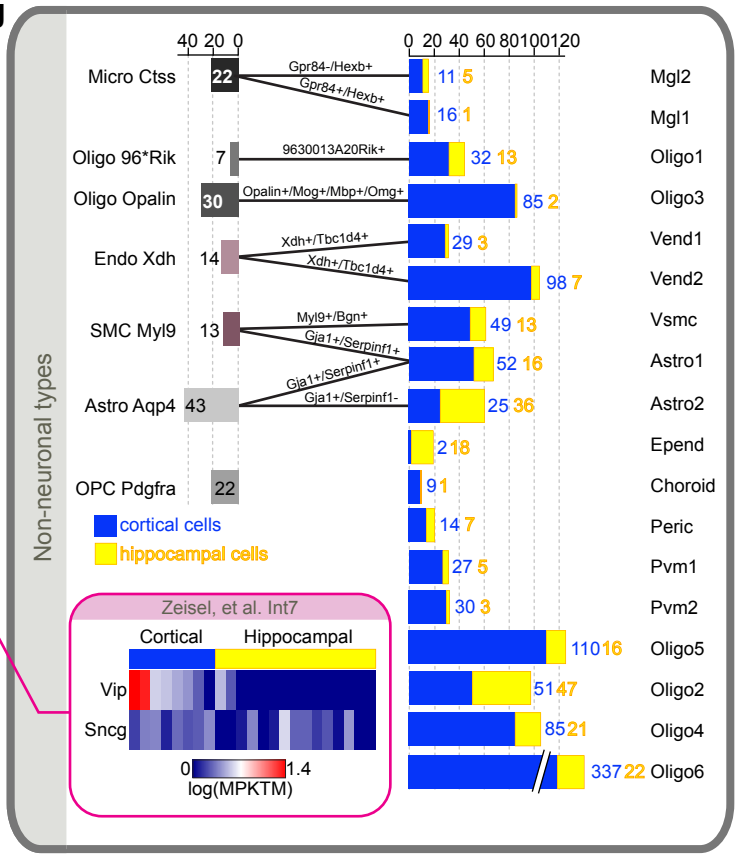












Supplementary Figure 17. Comparison of cell types defined in our study and Zeisel *et al.*[16] (**a**) Summary statistics for cell sampling and gene detection. Compared to the Zeisel *et al.* study, we sampled more cortical neurons and sequenced the individual cells more deeply to detect more genes. Many of those genes are not highly expressed, and are therefore not detected by Zeisel *et al*. Therefore, we used only the genes reported by Zeisel *et al.* to determine cell type correspondences. When performing clustering, Zeisel *et al.* combined the GABAergic neurons and non-neuronal cells from both hippocampus and cortex, but separated the glutamatergic cells based on region of origin. We therefore retained this grouping of GABAergic cells from the Zeisel *et al.* study, but did not analyze the glutamatergic hippocampal cells. (**b-e**) Comparison of GABAergic neurons based on marker genes. Due to the low sampling of Sst and Pvalb types in the Zeisel *et al.* dataset, the only clear correspondence among these types is between our Sst-Chodl type and *Int1* Zeisel *et al.* type (b). For Vip types, the clearest correspondence is between Vip-Chat and *Int9* (d). For other GABAergic types, the correspondences are less clear, and sometimes unexpected. For example, our Igtp type appears to correspond to *Int13* type, which we connect using a dotted line because *Int13* comprises hippocampal cells only (e). Correspondence is sometimes complicated by differences in marker gene expression in cortical vs. hippocampal cells. For example, *Int7* shows marked differences in the prevalence of the marker gene *Vip* between cells from different regions (inset). MPKTM, molecules per thousand total molecules detected. (**f**) We identified 21 glutamatergic types, most of which correspond to subdivisions of the L2/3, L4, L6, and deep-layer types from Zeisel *et al.* However, we also identified distinct types that appear to have no equivalent in the Zeisel *et al.* study: L4-Ctxn3, L5-Chrna6, L5b-Cdh13, and L5b-Tph2. We find that the latter two types contain the largest amount of RNA (**Supplementary Fig. 13**), and are probably the largest cells overall. This characteristic may have prevented their capture on Fluidigm C1 arrays employed by Zeisel *et al.* (**g**) Zeisel *et al.* identified many more non-neuronal types (18 vs. 7 in our case), but no oligodendrocyte precursor cells (OPCs), which are present in our study.