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).

Abstract: Nervous systems are composed of various cell types, but the extent of cell type diversity is poorly understood. We construct a cellular taxonomy of one cortical region, primary visual cortex, in adult mice on the basis of single-cell RNA sequencing. We identified 49 transcriptomic cell types, including 23 GABAergic, 19 glutamatergic and 7 non-neuronal types. We also analyzed cell type–specific mRNA processing and characterized genetic access to these transcriptomic types by many transgenic Cre lines. Finally, we found that some of our transcriptomic cell types displayed specific and differential electrophysiological and axon projection properties, thereby confirming that the single-cell transcriptomic signatures can be associated with specific cellular properties.

INTRODUCTION

The mammalian brain is likely the most complex animal organ, given the variety and scope of functions it controls, the diversity of cells it comprises, and the number of genes it expresses[1,2]. In the mammalian brain, the neocortex is essential for sensory, motor and cognitive behaviors. Although different cortical areas have dedicated roles in information processing, they exhibit a similar layered structure, with each layer harboring distinct neuronal populations[3]. In the adult cortex, many types of neurons have been identified through characterization of their molecular, morphological, connectional, physiological and functional properties[4–8]. Despite much effort, objective classification on the basis of quantitative features has been challenging, and our understanding of the extent of cell-type diversity remains incomplete[4,9,10].

Cell types can be preferentially associated with molecular markers that underlie their unique structural, physiological and functional properties, and these markers have been used for cell classification. Transcriptomic profiling of small cell populations from fine dissections[2,11] on the basis of cell surface[12,13] or transgenic markers[5] has been informative; however, any population-level profiling obscures potential heterogeneity in collected cells. Recently, robust and scalable transcriptomic single cell profiling has emerged as a powerful approach to characterization and classification of single cells, including neurons[14–17]. We used single-cell RNA-seq to characterize and classify more than 1,600 cells from the primary visual cortex in adult male mice. The

annotated data set and a single-cell gene expression visualization tool are freely accessible via the Allen Brain Atlas data portal (http://casestudies.brain-map.org/celltax).

RESULTS

Cell type identification

To minimize the potential variability in cell types due to differences in cortical region, age and sex, we focused on a single cortical area in adult (8-week-old) male mice. We selected the primary visual cortex (VISp or V1), which processes and transforms visual sensory information, and is one of the main models for understanding cortical computation and function[18]. To access both abundant and rare cell types in VISp, we selected a set of transgenic mouse lines in which Cre recombinase is expressed in specific subsets of cortical cells[19] (**Supplementary Table 1**). Each Cre line was crossed to the *Ai14* Cre reporter line, which expresses the fluorescent protein tdTomato (tdT) after Cre-mediated recombination (**Supplementary Fig. 1a**, **Supplementary Table 2** and Online Methods). To label more specific cell populations, we combined Cre lines with Dre or Flp recombinase lines and intersectional reporter lines (*Ai65* or *Ai66*; **Supplementary Fig. 1a**, **Supplementary Table 2** and Online Methods). To isolate individual cells for transcriptional profiling, we sectioned fresh brains from adult transgenic male mice, microdissected the full cortical depth, combinations of sequential layers or individual layers (L1, 2/3, 4, 5 and 6) of VISp, and generated single-cell suspensions using a previously published procedure[5] with some modifications (**Fig. 1a**, **Supplementary Fig. 1b** and Online Methods). We developed a robust procedure for isolating individual adult live cells from the suspension by fluorescence-activated cell sorting (FACS), reverse transcribed and amplified full-length poly(A)-RNA with the SMARTer protocol, converted the cDNA into sequencing libraries by tagmentation (Nextera XT), and sequenced them by next generation sequencing (**Fig. 1a**, **Supplementary Fig. 1b** and Online Methods). We established quality control (QC) criteria to monitor the experimental process (**Supplementary Fig. 2**) and data quality (**Supplementary Figs. 3b** and **4–7**, and Online Methods). Our final QC-qualified data set contains 1,679 cells, with more than 98% of cells sequenced to a depth of at least 5,000,000 total reads (median ~8,700,000, range ~3,800,000–84,300,000; **Supplementary Table 3**).

To identify cell types, we developed a classification approach that takes into account all expressed genes and is agnostic as to the origin of cells (**Fig. 1b**, **Supplementary Fig. 3** and Online Methods). Briefly, we applied two parallel and iterative approaches for dimensionality reduction and clustering, iterative principal component analysis (PCA) and iterative weighted gene coexpression network analysis (WGCNA), and validated the cluster membership from each approach using a non-deterministic machine learning method (random forest). The results from these two parallel cluster identification approaches were intersected (**Supplementary Fig. 8**) and subjected to another round of cluster membership validation. This step assessed the consistency of individual cell classification: we refer to the 1,424 cells that were consistently classified into the same cluster as ‘core’ cells and refer to the 255 cells that were classified into more than one cluster by the random forest approach as ‘intermediate’ cells (**Fig. 1b**, **Supplementary Fig. 3** and Online Methods).

This analysis segregated cells into 49 distinct core clusters (**Fig. 1c**). On the basis of known markers for major cell classes, we identified 23 GABAergic neuronal clusters (*Snap25+*,

Slc17a7−, *Gad1+*), 19 glutamatergic neuronal clusters (*Snap25+*, *Slc17a7+*, *Gad1−)* and 7 non-neuronal clusters (*Snap25−*, *Slc17a7−*, *Gad1−*) (**Fig. 1c**).We assigned location and identity to cell types within VISp on the basis of three complementary lines of evidence: layer-enriching dissections from specific Cre lines (**Fig. 2**), expression of previously reported and/or newly discovered marker genes in our RNA-seq data (**Fig. 3**), and localized expression patterns of marker genes determined by RNA in situ hybridization (ISH; **Supplementary Figs. 9** and **10**).

As expected, most layer-specific Cre lines labeled specific types of glutamatergic neurons (**Fig. 2** and **Supplementary Table 4**). Some GABAergic types also displayed laminar enrichment that was uncovered by dissections containing one or several layers (usually upper (L1–4) or lower (L5–6) layers combined; **Fig. 2** and **Supplementary Table 5**). Cells in the seven non-neuronal types were mostly isolated as tdT^{-} cells from layer-specific Cre lines (**Fig. 2b**).

Our single-cell analysis detects most previously known marker genes and identifies many new differentially expressed genes. For each type, if available, we defined 'unique markers', which are genes expressed only in that type among all of the cells sampled. We also identified 'combinatorial markers', which are differentially expressed genes not restricted to a single cell type. Together, these genes produce a unique pattern of expression among all cells sampled (**Fig. 3** and Online Methods). For a select set of markers, we employed single- and double-label RNA ISH (**Supplementary Figs. 9** and **10**) and quantitative RT-PCR (**Supplementary Fig. 11**) to confirm predicted specificity of marker expression or confirm cell location obtained from layer-enriching dissections.

Our Cre-line based approach also enabled the characterization of specificity of these lines, thereby informing their proper use for labeling and perturbing specific cellular populations[19–22]. In general, we found that the examined Cre lines mostly label the expected cell types based on promoters and other genetic elements that control Cre recombinase expression in each line (**Fig. 2** and Online Methods)[19]. However, all but one Cre line (*Chat-IRES-Cre*) labeled more than one transcriptomic cell type.

Cortical cell types: markers and relationships

To provide an overall view of the transcriptomic cell types that we identified, we integrated our data into constellation diagrams that summarize the identity, select marker genes and putative location of these types along the pia-to-white-matter axis (**Fig. 4a–c** and **Supplementary Table 6**). In these diagrams, each transcriptomic cell type is represented by a disk, whose surface area corresponds to the number of core cells in our data set belonging to that type. Intermediate cells are represented by lines connecting the disks; the line thickness is proportional to the number of intermediate cells. We separately present GABAergic, glutamatergic and non-neuronal constellations, as we detected only a single intermediate cell between these major classes. This mode of presentation paints the overall phenotypic landscape of cortical cell types as a combination of continuity and discreteness: the presence of a large number of intermediate cells between a particular pair of core types suggests a phenotypic continuum, whereas a lack of intermediate cells connecting one type to others suggests its more discrete character (**Fig. 4a–c**). We represent the overall similarity of gene expression between the transcriptomic cell types by hierarchical clustering of groups of their core cells based on all genes expressed above a variance threshold (**Fig. 4d**). These two views of transcriptomic cell types are complementary; one shows

the extent of intermediate phenotypes and the other shows the overall similarity in gene expression between cluster cores (**Supplementary Fig. 12**).

We identified 18 transcriptomic cell types belonging to three previously described major classes of GABAergic cells named after the corresponding markers Vip (vasoactive intestinal peptide), Pvalb (parvalbumin) and Sst (somatostatin)[6,23,24]. In a substantial portion of these cells, we detected more than one of these markers, but our method, which takes into account genome-wide gene expression, usually classified these double-expressing cells into the major type corresponding to the most highly expressed major marker in that cell (Online Methods).

We identified five additional GABAergic types. In accord with a previous report[25], we detected *Tnfaip8l3* and *Sema3c* in these types. We named two of them on the basis of a gene for a putative neuropeptide, neuron-derived neurotrophic factor (*Ndnf*), and we found that they corresponded to neurogliaform cells (see below). We refer to the three other types according to markers they express: synuclein gamma (*Sncg*), interferon gamma-induced GTPase (*Igtp*) and SMAD family member 3 (*Smad3*).

Beyond the major types, correspondence of our transcriptomic types to those previously described in the literature was not straightforward and relied on the existence of a Rosetta stone: a shared reagent, feature or molecular marker with unambiguous translational power. Potential inferences on correspondence to previously proposed types were further complicated by previous studies' employment of a variety of animal models, at varying ages, and with focus on different cortical areas. Moreover, most studies have relied on a small set of molecular markers (for example, *Calb1* (calbindin), *Calb2* (calretinin), *Cck*, *Crh*, *Htr3a*, *Nos1*, *Npy* and *Reln*)[4,6] (**Supplementary Table 7**).

We found only one Sst type (Sst-Cbln4) that was prevalent in upper cortical layers, whereas all of the other Sst types appeared to be enriched in lower layers (**Figs. 2b** and **4a**). On the basis of upper-layer enrichment and *Calb2* expression of the Sst-Cbln4 type (**Fig. 3a**), we propose that it likely corresponds to previously characterized Calb2-positive Martinotti cells that are enriched in the upper cortical layers[26] and are fluorescently labeled in transgenic GIN mice[27]. Our analysis revealed only one additional *Calb2*-positive Sst type, which we refer to as Sst-Chodl (**Figs. 2b and 3a**). On the basis of the expression of tachykinin-receptor 1 (*Tacr1*), neuropeptide Y (*Npy*), high levels of nitric oxide synthase (*Nos1*) and the absence of *Calb1* (**Fig. 3a** and **Supplementary Fig. 9**), we conclude that this type most likely corresponds to Nos1 type I neurons[28], which are enriched in L5 and 6 (ref. 29), and are likely long-range projecting[30], sleep-active neurons[31].

The Pvalb types are highly interconnected in the constellation diagrams (**Fig. 4a**). Using layer-enriching dissections (**Fig. 2b**), we found that some types were preferentially present in upper (Pvalb-Tpbg, Pvalb-Tacr3, Pvalb-Cpne5) or lower layers (Pvalb-Gpx3 and Pvalb-Rspo2).To relate our transcriptomic types to previously described Pvalb types, we isolated cells from the upper layers of the *Nkx2.1-CreERT2* line, which, when induced with tamoxifen perinatally, labels a subset of neocortical interneurons, including chandelier cells[32]. Our analysis classified cells from this line in all three upper layer-enriched Pvalb types (**Fig. 4a**). We suggest that Pvalb-Cpne5 corresponds to chandelier cells because it was most transcriptionally distinct among Pvalb

types, it was enriched in upper layers and it did not express *Etv1* (also known as *Er81*), as previously shown for chandelier cells[33] (**Supplementary Fig. 12**).

The Vip major class can be divided into several transcriptomic cell types, all of which appeared to be enriched in upper cortical layers, except the Vip-Gpc3 type (**Fig. 4a**). In accord with previous reports[23,34], our Vip-Chat transcriptomic type was located in upper cortical layers (**Fig. 2a**) and it displayed unique expression of choline acetyltransferase (*Chat*) in Vip-positive cells. These cells have been reported to either express[34] or not express Calb2 at the protein level[23]; we found that they robustly expressed *Calb2* mRNA.

For glutamatergic cells, we identified six major classes of transcriptomic types—L2/3, L4, L5a, L5b, L6a and L6b—on the basis of the layer-specific expression of marker genes and layer-enriching dissections; this is consistent with many previous studies[1,7,8,35]. We discovered subdivisions among all of these layer-specific major types. In L2/3, we identified two major types, one of which (L2-Ngb) appeared to be located more superficially based on marker gene expression (for example, *Ngb*, *Fst*, *Syt17* and *Cdh13*; **Fig. 3b** and **Supplementary Fig. 9**). In L4, we identified three types (L4-Ctxn3, L4-Scnn1a and L4-Arf5) with high gene expression similarity (**Fig. 4d**) and a large number of intermediate cells (**Fig. 4b**). We identified eight different transcriptomic types in L5. Four of these types expressed the L5a marker *Deptor* (L5a-Hsd11b1, L5a-Tcerg1l, L5a-Batf3 and L5a-Pde1c), whereas three expressed the L5b marker *Bcl6* (L5b-Cdh13, L5b-Tph2 and L5b-Chrna6; **Fig. 3b**). One of these L5b types (L5b-Chrna6), together with the L5-Ucma type, appeared most distinct among L5 types, both on the basis of gene expression and the small number of intermediate cells between them and other L5 types (**Fig. 4b**). We identified six transcriptomic cell types in L6: four L6a types and two L6b types. Among L6a types, two highly related types (L6a-Sla, and L6a-Mgp) expressed the marker *Foxp2* (refs. 7,35,36) and were primarily derived from the *Ntsr1-Cre* line (**Fig. 2b**), whereas the other two (L6a-Syt17 and L6a-Car12) did not express *Foxp2* and were isolated as tdT$^-$ cells from L6 of the same Cre line. For the latter two types, we discovered several new markers that can be used to identify them (*Car12*, *Prss22*, *Syt17* and *Penk*; **Fig. 3b** and **Supplementary Figs. 9b** and **10j–k**). The two L6b types (**L6b-Serpinb11** and **L6b-Rgs12**) expressed the known L6b marker Ctgf[7,35,36] and several other previously identified L6b markers (for example, *Trh*, *Tnmd* and *Mup5*; **Fig. 3b** and **Supplementary Fig. 9b**)[7].

Despite the neuronal focus of this study, our sampling strategy captured enough cells to also identify the major non-neuronal classes. We found seven non-neuronal types: astrocytes, microglia, oligodendrocyte precursor cells (OPCs), two types of oligodendrocytes, endothelial cells and smooth muscle cells. In accord with previous population-level studies[12,13], these types could be distinguished by many combinatorial and unique markers (**Figs 3c and 4c**, **Supplementary Fig. 12** and Online Methods).

Comparative analysis of cell types

After defining cell types, we examined additional cellular properties that could be extracted from our data set. We found that neurons contained more total RNA than non-neuronal cells (median 11.5 versus 2.5 pg) and expressed more genes when sequenced to the same depth (mean 7,278 versus 4,274) (**Supplementary Fig. 13a,c**). We estimate that some neuronal types had >20-fold higher RNA content than some glial types (for example, L5b-Tph2 ~37.0 pg per cell versus

microglia ~1.6 pg per cell; **Supplementary Fig. 13b**). We also found differences in the distribution of gene abundances among cell types; overall, neurons expressed more genes at low or intermediate levels than non-neuronal cells, whereas non-neuronal cells expressed more genes at high levels (**Supplementary Fig. 13e,f**). Together, the number of genes and the gene distributions suggest larger variety or complexity of neuronal compared to non-neuronal functions.

Our approach for RNA-seq, which is based on full-length cDNAs, enabled examination of alternative promoter use, polyadenylation and splicing between cell types. We found a total of 567 exons in 320 genes that displayed differential pre-mRNA processing in a cell type–specific manner at various levels of cellular taxonomy (**Fig. 5** and **Supplementary Table 8**). In particular, *Gria1* and *Gria2* displayed highly cell type–specific alternative splicing for two consecutive exons (previously named flip and flop)[37], of which only a single one is included in each mature mRNA (**Fig. 5d,e**). Each exon encodes a small segment of the predicted fourth transmembrane region, which imparts different electrophysiological properties to the receptors[37]. In agreement with relatively low-resolution RNA ISH data[37], we found that the L2-Ngb and L2/3-Ptgs2 types preferentially used the flip exons, L4 types used the flop exons and L6a types utilized both (**Fig. 5d,e**). Moreover, our single-cell analysis and data-driven aggregation of cells into types enabled examination of differential exon use in less abundant cell types and at a higher resolution, revealing additional differential mRNA processing between GABAergic, L5 and L6 types. Many of these differences in mRNA processing would not be apparent if populations containing a mixture of transcriptomic cell types were profiled. Our approach therefore allowed cells belonging to the same cell type to be analyzed together to discover robust cell type–specific signatures of mRNA processing.

In this genome-wide data set, we also explored the expression of genes particularly relevant for neuronal development and function. Examination of transcription factors revealed a number of genes that have been shown to be involved in the specification of neuronal types (**Supplementary Fig. 12**). As expected, many more ion channel genes were expressed in neurons than glia, and many were differentially expressed, but rarely unique, for specific cell types (**Supplementary Fig. 14**). We observed widespread neuronal expression of many glutamate and GABA receptors, including both ionotropic and metabotropic types, whereas the receptors for other, mostly modulatory, neurotransmitters were generally expressed at lower levels and more selectively in certain cell types (**Supplementary Fig. 15**). Neuropeptide genes and their receptors were frequently expressed in specific yet different cell types, suggesting specific cell-cell interactions (**Supplementary Fig. 16**).

Transcriptomic cell types and neuronal properties

To determine whether the transcriptomic cell types that we defined display specific anatomical and physiological properties, we analyzed axonal projections and electrophysiology for a subset of transcriptomic types. To assess the correspondence between the transcriptomic cell types and axonal projection patterns, we combined single cell RNA-sequencing with viral retrograde tracing using canine adenovirus expressing Cre recombinase (CAV-Cre) in the Cre-reporter *Ai14* mice (**Fig. 6a**). We then classified the individual retrogradely labeled cells using a genome-wide gene expression classifier (**Supplementary Fig. 3c** and Online Methods). Cells labeled

retrogradely from the ipsilateral visual thalamus were classified into L5b-Tph2, L5b-Cdh13, L5-Chrna6, L6a-Mgp and L6a-Sla types. In contrast, cells labeled retrogradely from the contralateral VISp were classified into L5a-Batf3, L6a-Car12 and L6a-Syt17 cell types (**Fig. 6**).

These results are in excellent agreement with previous reports that have correlated specific molecular markers or Cre-dependent labeling with neuronal projection patterns. L5a neurons, which express *Deptor*, have been shown to have intra-telencephalic projections and have been designated as cortico-cortical and cortico-striatal projection neurons[7,35]. In contrast, L5b neurons, which express *Bcl6*, have been shown to project subcortically and have been designated as cortico-fugal projection neurons[7,35]. Accordingly, our cells labeled from contralateral VISp and ipsilateral thalamus were classified into transcriptomic L5a and L5b types, respectively (**Fig. 6**). The retrograde labeling of L6a types was also consistent with previous findings. Among the L6a projection neurons, corticothalamic (CT) projecting cells have been shown to express *Foxp2* (ref. 7), and are labeled by *Ntsr1-Cre* in VISp[38,39], which, in our data set, correspond to L6a-Mgp and L6a-Sla types (**Fig. 2b**). In comparison, the *Ntsr1-Cre*$^{-}$ cells (which correspond to L6a-Car12 and L6a-Syt17 types; **Fig. 2b**) have been shown to be corticocortical (CC) projecting cells that do not project to the thalamus[38,39].

To examine the correspondence of electrophysiological features with genome-wide expression signatures and our cell type classification, we focused on the Ndnf types, which, based on their superficial location and expression of *Reln* (**Fig. 3a**), may correspond to neurogliaform cells[6]. We used the *Ndnf* gene to generate a Cre line that should enable specific access to these cells (Online Methods). Indeed, in accord with our *Ndnf* mRNA ISH data (**Supplementary Figs. 9a** and **10b,d**), we found that this Cre line labeled neurons that were highly enriched in L1 (**Fig. 7a–d**) and that the neurons profiled transcriptomically from L1 of this Cre line were classified into the two Ndnf types (**Fig. 2b**).

Previously reported physiological characteristics of neurogliaform cells include a depolarizing ramp voltage near threshold, late spiking[40,41], accelerating spike frequency[42], gap junctional coupling[43,44] and slow GABA-mediated synaptic transmission[43,45]. Some neurogliaform cells have been shown to exhibit one or two action potentials at the onset of the long current pulse near threshold[40,41]. Neurogliaform cells can also form GABA-mediated autaptic synapses[45].

On the basis of whole-cell current-clamp recordings of tdT^{+} cells from *Ndnf-IRES2-dgCre;Ai14* mice in L1 of VISp, we grouped cells into two categories: late spiking (LS), and non–late spiking (NLS). LS neurons showed depolarizing ramp voltage near threshold, late spiking and accelerating spike frequency (**Fig. 7e,f**). NLS neurons displayed an initial depolarizing response that was sufficient to induce an action potential at the onset of the current step in some trials (**Fig. 7e,f**). The NLS neurons, to differing degrees, exhibited an initial depolarizing response that sagged (**Fig. 7e,f**). At slightly higher current intensities, all NLS neurons initiated a bout of late spiking after a period of quiescence (**Fig. 7e**). In multi-patch recordings, we observed frequent electrical coupling (**Fig. 7g**) and autaptic and synaptic transmission between tdT^{+} neurons that was blocked by the $GABA_A$ receptor antagonist SR95531 (**Fig. 7h**). Reconstruction of two biocytin-filled, tdT^{+} neurons revealed that one of them had a tight, dense axonal arbor with small, bouton-like structures and a relatively small dendritic tree that is typical of neurogliaform cells[46]. The other neuron displayed axonal and dendritic arbors like those of the recently described neurogliaform sparse-axon cells (**Fig. 7i**)[47]. Together, our molecular, physiological and

morphological analyses of L1 neurons labeled by the *Ndnf-IRES2-dgCre* line revealed that they correspond to neurogliaform cells.

DISCUSSION

The adult mouse visual cortex contains about 1,000,000 cells, of which about half are neurons[48] that can be divided into glutamatergic (80%) and GABAergic cells (20%)[49]. We defined cell types in the primary visual cortex on the basis of thousands of genes with single-cell resolution. Our description of the 49 transcriptomic cortical cell types includes all the major types reported in the literature, some additional new types, as well as subdivisions among the major types (**Supplementary Table 7**). Our approach also provides an experimental and computational workflow to systematically catalog cell types in any region of the mouse brain and relate them to the tools used to examine those cell types (Cre lines and viruses). The discovery of new marker genes (**Fig. 3**) enables generation of new specific Cre lines (**Fig. 7**) and provides guidance for intersectional transgenic strategies (such as the one in **Supplementary Fig. 1a**) to enable specific access to cortical cell types that do not express unique marker genes.

Our method relies on dissociation and FACS-isolation of single cells, thereby exposing them to stress that might lead to changes in gene expression. However, in our data set, the majority of marker genes showed excellent correspondence to RNA ISH data from the Allen Brain Atlas[1] (~72% of $N = 228$ examined genes; **Supplementary Table 9**), suggesting that our procedure did not markedly alter the transcriptional signatures of cell types. Most of the other examined transcripts in this set (**Supplementary Table 9**), which appeared to be very specific markers based on RNA-seq and qRT-PCR (for example, *Chodl*), were not detected by the Allen Brain Atlas in VISp. This discrepancy is probably a consequence of low sensitivity for a subset of ISH probes.

To classify cells based on their transcriptomes, we employed two iterative clustering methods and one machine learning–based validation method. The latter assessed the robustness of cluster membership for each cell and suggested the existence of cells with intermediate transcriptomic phenotypes. Previous studies either excluded intermediate cells explicitly[17] or allowed cells to have only a single identity[14–16]. We chose to develop a data analysis approach that accommodates these intermediate cells, as they may be a reflection of actual phenotypic continua. However, as in any approach, both biological and technical aspects contributed to our data sets. For example, similarly to a previous single-cell transcriptomic study[16], we estimate that we detected only ~23% of mRNA molecules present in a cell (**Supplementary Fig. 4c**). Employment of a highly efficient transcriptomic method that samples the cells in their native environment and in proportion to their abundance would provide a more complete and accurate description of the transcriptomic cell type landscapes. Inclusion of additional cells, even with the current method, is likely to segregate some of the types we defined here into additional subtypes. This is already apparent in our data set, as we observed more subtypes if we decreased the threshold for the minimal number of core cells required to define a type (Online Methods). In contrast, additional cell sampling may also reveal previously undetected intermediate cells that would define new continua between discrete types. Finally, although we attempted to cover all

major types by choosing a variety of Cre lines, including pan-glutamatergic and pan-GABAergic lines, it is still possible we did not sample some rare types.

We employed substantially deeper sequencing per cell than several other studies[14,17,50]. One of the main advantages of low-depth sequencing is reduction of experimental cost. However, we note that if we downsampled our data from full depth to 1,000,000 or 100,000 mapped reads per cell, we lost the power to detect many types (**Supplementary Table 10**). Thus, when subsampling to 100,000 reads, we only found 35 instead of 49 types. This decrease in resolution could be compensated for by sampling many more cells, but the appropriate balance between the sequencing depth and cell number depends on a variety of factors, including the selected RNA-seq method, informative transcript abundance, tissue and cell type abundance/accessibility, and desired resolution between cell types.

Our study, with its focus on profiling neurons in adult mice from a single cortical region using Cre lines, complements a recent transcriptomic study of single cells from somatosensory cortex and hippocampus in P21–31 mice[16]. Based on the expression of key marker genes, we found both commonalities and differences among the cell types identified in that study and ours (**Supplementary Fig. 17**). For neuronal cells, we identified more transcriptomic glutamatergic (19 versus 7) and GABAergic Sst (six versus three), Pvalb (seven versus one), and Vip (five versus three) types, but fewer other GABAergic types (five versus nine) (**Supplementary Fig. 17**). For non-neuronal cortical cells, the previous study[16] defined many more types that mostly correspond to subdivisions of our non-neuronal types, with the exception of oligodendrocyte precursor cells (OPCs), which are only present in our study. It is important to note that the two studies differed in a number of experimental and data analysis parameters. For example, given the different sampling strategies (Cre line–based versus mostly unbiased), we analyzed more neocortical neurons (1,525 versus 563), and given the differences in RNA-seq procedures (SMARTer versus 5′-end focused STRT) and sequencing depths, we detected more genes in these neurons (~7,200 versus ~4,500) (**Supplementary Fig. 17**). In addition, our study differed from the previous one[16] in the genetic background (mostly C57BL/6J versus CD-1) and age of analyzed mice, as well as the cell isolation procedures (FACS versus mostly Fluidigm C1 microfluidics). Overall, the two studies overlap in their identification of some transcriptomic types, but differ in their focus: the previous study[16] offers deeper insight into non-neuronal transcriptomic types, hippocampal excitatory cells and cells from brain ventricles, whereas our results provide a more comprehensive classification of adult neocortical neurons.

Our results suggest many new directions for further investigation. At the forefront is the question of the correspondence and potential causal relationships between transcriptomic signatures and specific morphological, physiological and functional properties. For example, do the two transcriptomic Ndnf subtypes and the two detected electrophysiological phenotypes (LS and NLS) correspond to each other, and which genes are responsible for these physiological differences? Do the two corticothalamic L6a subtypes (L6a-Sla and L6a-Mgp) correspond to two previously described morphological classes, which terminate their apical dendrites in L1 or L4 (ref. 21)? Are certain transcriptomic differences representative of cell state or activity, rather than cell type? In fact, is there a clear distinction between the state and the type? For example, recent evidence suggests that Pvalb basket cells acquire specific firing properties in an activity-dependent manner that may result in a continuum of basket cell phenotypes[33], perhaps mirroring the large numbers of intermediate cells that we found for upper layer *Etv1*(*Er81*)-positive Pvalb

cells (**Fig. 4a**). Although these questions await further studies, our approach provides an overview of adult cell types in a well-defined cortical area based on a highly multidimensional data set and is an essential step toward understanding the most complex animal organ, the mammalian brain.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Next generation sequencing data have been deposited to the Gene Expression Omnibus under accession number **GSE71585**. Accession numbers for individual cells characterized in this study can be found in **Supplementary Table 3**.

Acknowledgments. We would like to thank M. Chillon Rodrigues (Universitat Autònoma de Barcelona) for providing CAV-Cre, S. Mihalas for advice on data analysis, H. Gu, M. Mills, H. Gill and K. Hadley for technical assistance, C. Ye and A. Kaykas for help with the next generation sequencing, and the Department of In vivo Sciences, especially R. Larsen, L. Pearson and J. Harrington for mouse husbandry. We thank J. Waters and E. Lein for comments on the manuscript. The authors thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement and support. This work was funded by the Allen Institute for Brain Science, and by US National Institutes of Health grants R01EY023173 and U01MH105982 to H.Z.

Author contributions. B.T. and H.Z. designed and supervised the study. T.N.N., T.K.K. and B.T. performed single-cell RNA-seq. V.M. and Z.Y. performed transcriptome data analysis with contributions from L.T.G., T.N.N., B.T., T.K.K., C.L. and M.H. T.N.N. performed stereotaxic injections. T.J. performed electrophysiology and associated data analysis. T.N.N., T.K.K. and B.T. performed single-cell isolation with contributions from B.L., N.S. and S.P. S.A.S. performed imaging of biocytin-filled cells and morphological reconstructions. D.B., J.G., K.S. and A.B. performed qRT-PCR and RNA DFISH in collaboration with T.N.N. and B.T. L.M. generated transgenic mice. T.D. designed the online scientific vignette in collaboration with B.T. and V.M. S.M.S. provided program management support. L.G., T.N., V.M. and B.T. prepared the figures. B.T., V.M., H.Z. and C.K. wrote the manuscript in consultation with all authors.

Competing Financial Interests. The authors declare no competing financial interests.

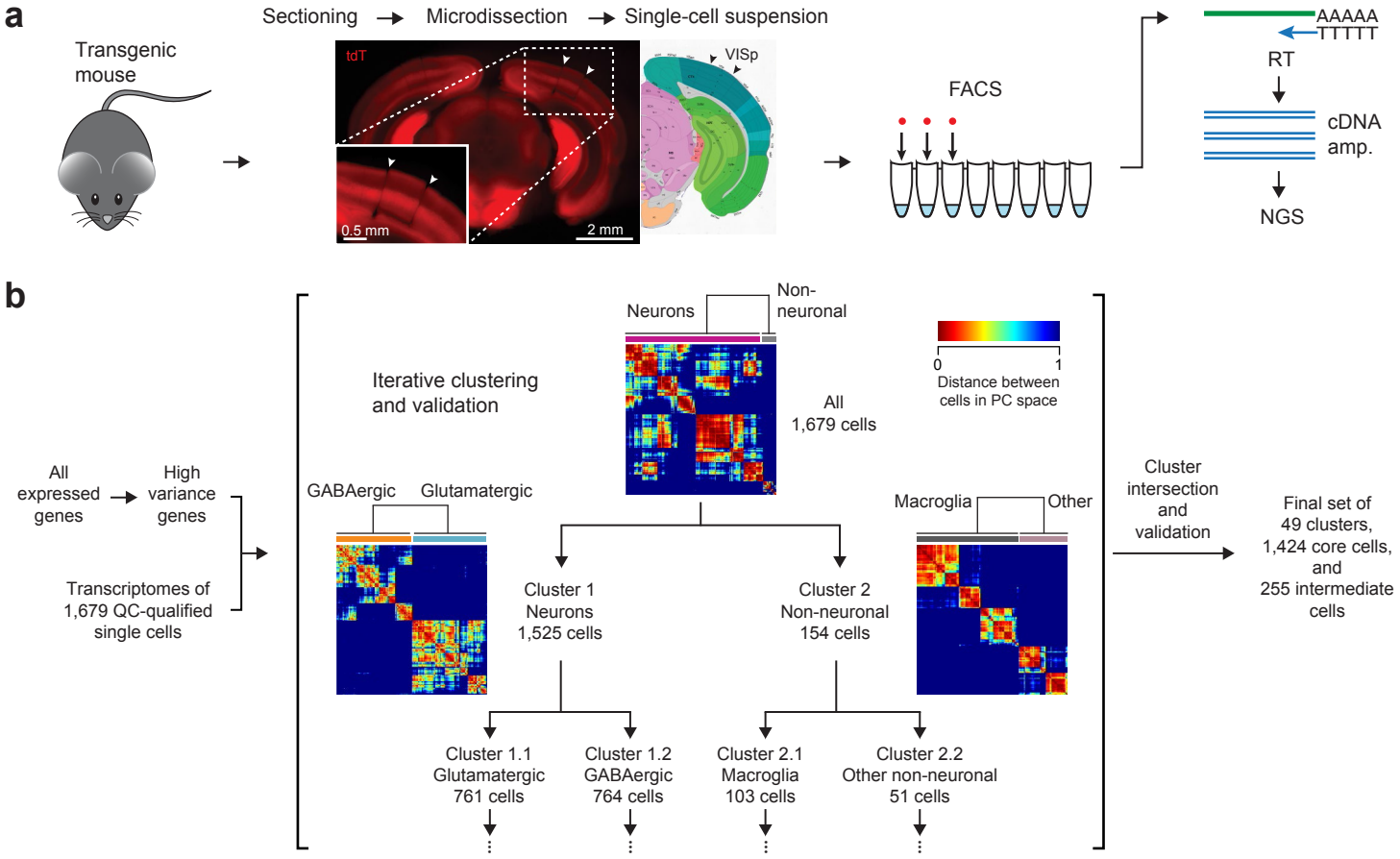

a
Sectioning
Microdissection
Single-cell suspension
Transgenic mouse
tdT
0.5 mm
2 mm
VISp
FACS
AAAAA
TTTTT
RT
cDNA amp.
NGS
b
All expressed genes
High variance genes
Transcriptomes of 1,679 QC-qualified single cells
Iterative clustering and validation
Neurons
Non-neuronal
All 1,679 cells
Distance between cells in PC space
GABAergic
Glutamatergic
Macroglia
Other
Cluster 1 Neurons 1,525 cells
Cluster 2 Non-neuronal 154 cells
Cluster 1.1 Glutamatergic 761 cells
Cluster 1.2 GABAergic 764 cells
Cluster 2.1 Macroglia 103 cells
Cluster 2.2 Other non-neuronal 51 cells
Cluster intersection and validation
Final set of 49 clusters, 1,424 core cells, and 255 intermediate cells
c
1,424 core cells
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 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
Snap25 1.2x10⁴
Gad1 3.8x10³
Vip 2.7x10⁴
Sst 6.5x10⁴
Pvalb 1.9x10³
Slc17a7 1.6x10²
Rorb 5.1x10²
Foxp2 8.6x10¹
Aqp4 9.2x10²
Pdgfra 6.5x10²
Mog 2.7x10³
Itgam 1.1x10³
Flt1 4.6x10²
Bgn 2.2x10³

Figure 1. Workflow overview. (**a**) Experimental workflow started with the isolation, sectioning and microdissection of the primary visual cortex from a transgenic mouse. The tissue samples were converted into a single-cell suspension, single cells were isolated by FACS, poly(A)-RNA from each cell was reverse transcribed (RT), cDNA was amplified and fragmented, and then sequenced on a next-generation sequencing (NGS) platform. (**b**) Analysis workflow started with the definition of high-variance genes and iterative clustering based on two different methods, PCA (shown here) and WGCNA, and cluster membership validation using a random forest classifier. Cells that are classified consistently into one cluster are referred to as core cells (N = 1,424), whereas cells that are mapped to more than one cluster are labeled as intermediate cells (N = 255). After the termination criteria are met, clusters from the two methods are intersected, and iteratively validated until all core clusters contain at least four cells (**Supplementary Fig. 3** and Online Methods). (**c**) The final 49 clusters were assigned an identity based on cell location (**Fig. 2**) and marker gene expression (**Fig. 3**). Each type is represented by a color bar with the name and number of core cells representing that type. The violin plots represent distribution of mRNA expression on a linear scale, adjusted for each gene (maximum RPKM on the right), for major known marker genes: *Snap25* (pan-neuronal); *Gad1* (pan-GABAergic); *Vip*, *Sst* and *Pvalb* (GABAergic); *Slc17a7* (pan-glutamatergic); *Rorb* (mostly L4 and L5a); *Foxp2* (L6); *Aqp4* (astrocytes); *Pdgfra* (oligodendrocyte precursor cells, OPCs); *Mog* (oligodendrocytes); *Itgam* (microglia); *Flt1* (endothelial cells); and *Bgn* (smooth muscle cells, SMC).

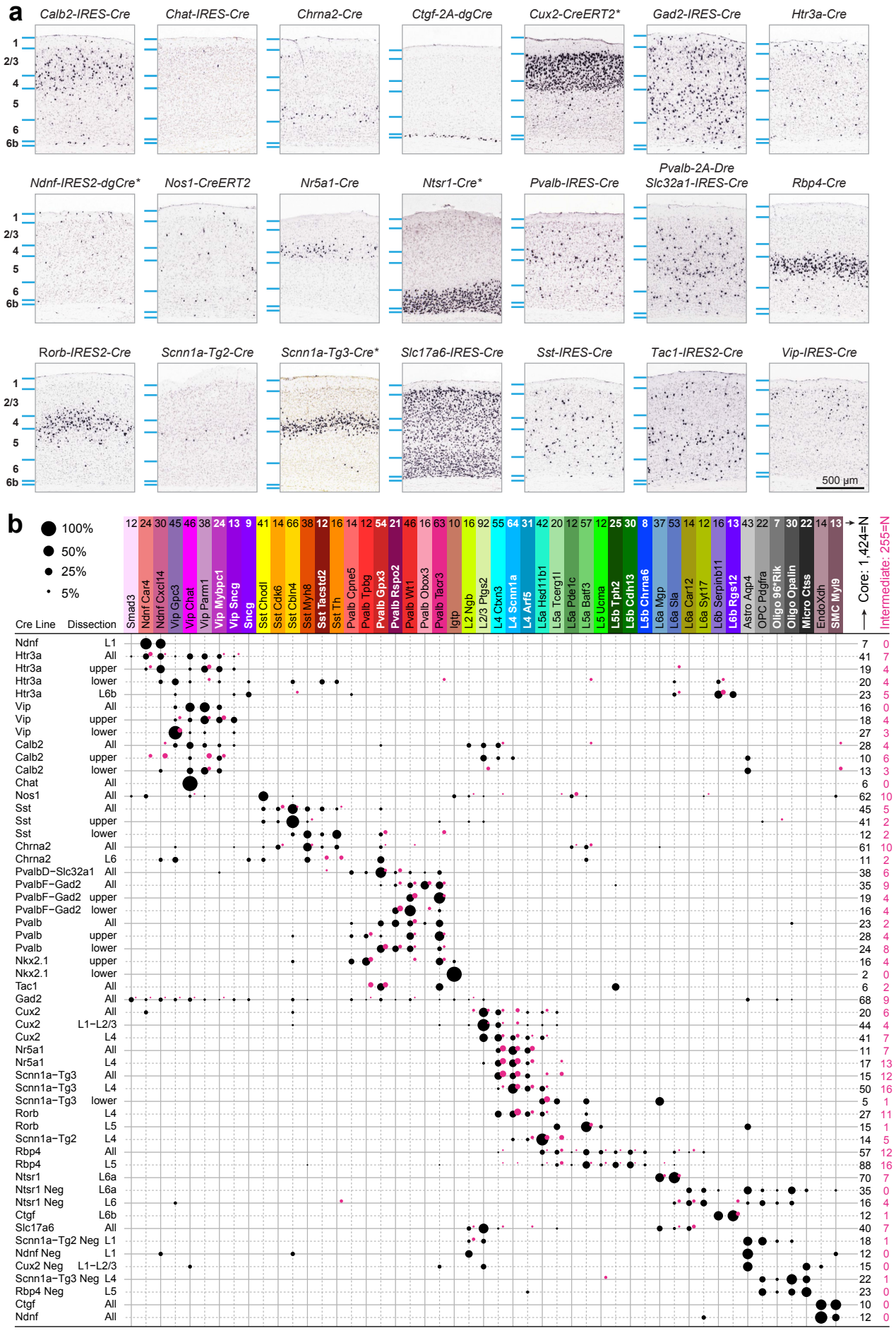

a
Calb2-IRES-Cre
Chat-IRES-Cre
Chrna2-Cre
Ctgf-2A-dgCre
Cux2-CreERT2*
Gad2-IRES-Cre
Htr3a-Cre
Ndnf-IRES2-dgCre*
Nos1-CreERT2
Nr5a1-Cre
Ntsr1-Cre*
Pvalb-IRES-Cre
Pvalb-2A-Dre
Slc32a1-IRES-Cre
Rbp4-Cre
Rorb-IRES2-Cre
Scnn1a-Tg2-Cre
Scnn1a-Tg3-Cre*
Slc17a6-IRES-Cre
Sst-IRES-Cre
Tac1-IRES2-Cre
Vip-IRES-Cre
1
2/3
4
5
6
6b
500 µm
b
100%
50%
25%
5%
Cre Line
Dissection
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 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
EndoXdh
SMC Myl9
Core: 1,424=N
Intermediate: 255=N
Ndnf L1
Htr3a All
Htr3a upper
Htr3a lower
Htr3a L6b
Vip All
Vip upper
Vip lower
Calb2 All
Calb2 upper
Calb2 lower
Chat All
Nos1 All
Sst All
Sst upper
Sst lower
Chrna2 All
Chrna2 L6
PvalbD−Slc32a1 All
PvalbF−Gad2 All
PvalbF−Gad2 upper
PvalbF−Gad2 lower
Pvalb All
Pvalb upper
Pvalb lower
Nkx2.1 upper
Nkx2.1 lower
Tac1 All
Gad2 All
Cux2 All
Cux2 L1−L2/3
Cux2 L4
Nr5a1 All
Nr5a1 L4
Scnn1a−Tg3 All
Scnn1a−Tg3 L4
Scnn1a−Tg3 lower
Rorb L4
Rorb L5
Scnn1a−Tg2 L4
Rbp4 All
Rbp4 L5
Ntsr1 L6a
Ntsr1 Neg L6a
Ntsr1 Neg L6
Ctgf L6b
Slc17a6 All
Scnn1a−Tg2 Neg L1
Ndnf Neg L1
Cux2 Neg L1−L2/3
Scnn1a−Tg3 Neg L4
Rbp4 Neg L5
Ctgf All
Ndnf All

Figure 2. Cell types: genetic access and laminar distribution. (**a**) Characterization of Cre lines by RNA ISH detection of *tdT* mRNA from the *Ai14* transgene. Representative images of VISp were obtained from the Allen Connectivity Atlas, Transgenic Characterization[19]. Sections are coronal except when indicated by asterisks (sagittal); images are representative of at least two brain-wide experiments, except for *Scnn1a-Tg3-Cre*, which is represented by one experiment (average of ~2.9 experiments per Cre line). *Pvalb-2A-Dre;Slc32a1-IRES-Cre* characterization used *Ai66*. Transgenic characterization data for *Pvalb-2A-FlpO;Gad2-IRES-Cre;Ai65* and *Nkx2.1-CreERT2;Ai14* (corresponding to our induction criteria, Online Methods) are not available. Scale bar applies to all panels in **a**. (**b**) Cre line specificity characterized by transcriptomic cell types (N = 1,424 core cells, 255 intermediate cells). The size of each black disk represents the proportion of cells classified as core in each transcriptomic type isolated from a particular Cre line and microdissection combination (rows). Pink disks correspond to the proportion of cells that were classified as intermediate. Upper dissection corresponds to layers 1–4, and lower to layers 5–6 of VISp. The number of cells from each Cre line and microdissection combination for core cells (black) and intermediate cells (pink) is indicated on the right; the number of core cells for each type is indicated on top. Note that the relative proportions of cell types obtained in these experiments are not representative of the ones in the intact brain because of the targeted sampling approach using Cre lines and possible cell type–specific differences in survival during the isolation procedure. Cell numbers and percentages represented in **b** are available in **Supplementary Table 4**.

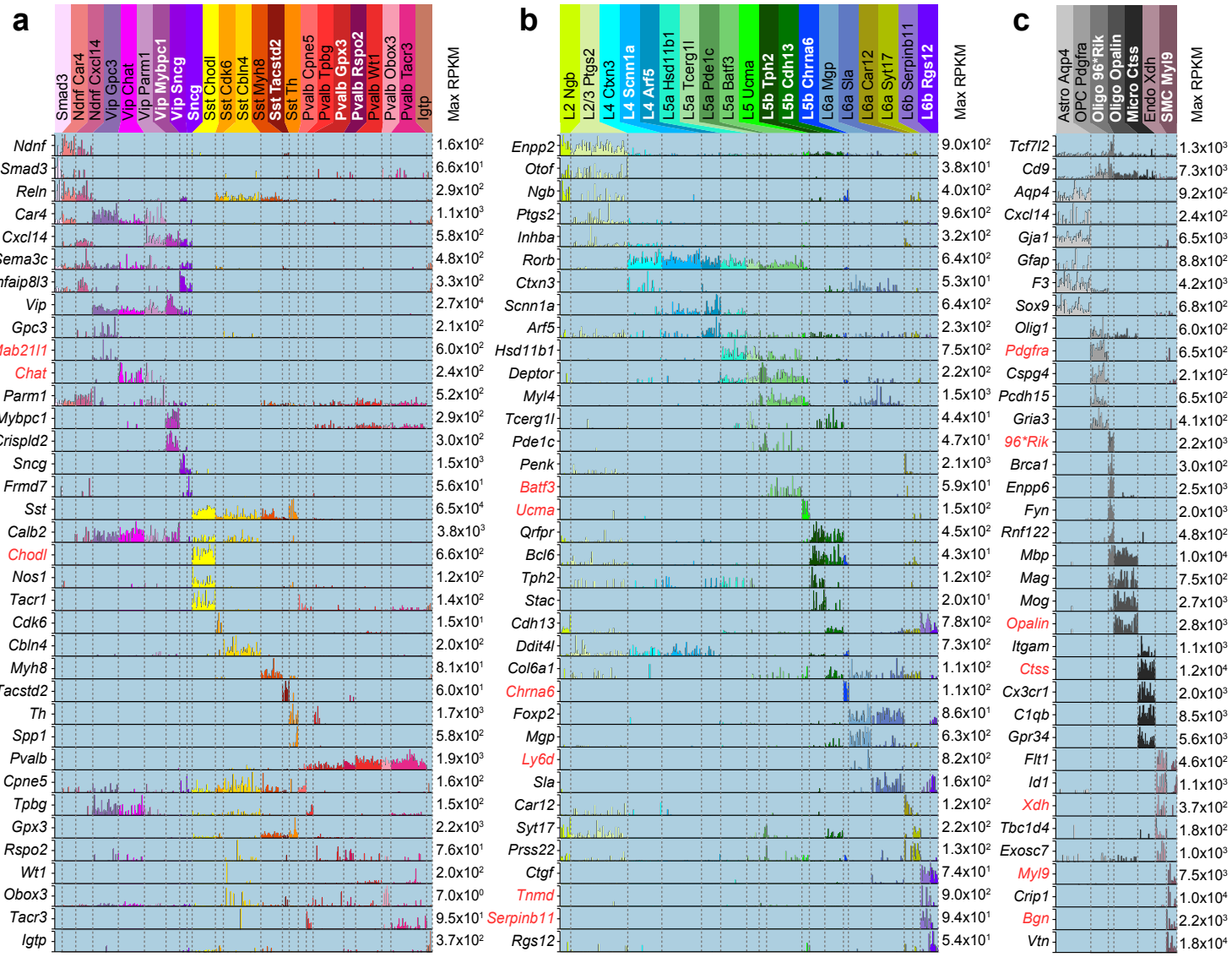

a
Smad3
Ndnf Car4
Ndnf Cxcl14
Vip Gpc3
Vip Chat
Vip Parm1
Vip Mybpc1
Vip Sncg
Sncg
Sst Chodl
Sst Cdk6
Sst Cbln4
Sst Myh8
Sst Tacstd2
Sst Th
Pvalb Cpne5
Pvalb Tpbg
Pvalb Gpx3
Pvalb Rspo2
Pvalb Wt1
Pvalb Obox3
Pvalb Tacr3
Igtp
Max RPKM
Ndnf 1.6x10²
Smad3 6.6x10¹
Reln 2.9x10²
Car4 1.1x10³
Cxcl14 5.8x10²
Sema3c 4.8x10²
Tnfaip8l3 3.3x10²
Vip 2.7x10⁴
Gpc3 2.1x10²
Mab21l1 6.0x10²
Chat 2.4x10²
Parm1 5.2x10²
Mybpc1 2.9x10²
Crispld2 3.0x10²
Sncg 1.5x10³
Frmd7 5.6x10¹
Sst 6.5x10⁴
Calb2 3.8x10³
Chodl 6.6x10²
Nos1 1.2x10²
Tacr1 1.4x10²
Cdk6 1.5x10¹
Cbln4 2.0x10²
Myh8 8.1x10¹
Tacstd2 6.0x10¹
Th 1.7x10³
Spp1 5.8x10²
Pvalb 1.9x10³
Cpne5 1.6x10²
Tpbg 1.5x10²
Gpx3 2.2x10³
Rspo2 7.6x10¹
Wt1 2.0x10²
Obox3 7.0x10⁰
Tacr3 9.5x10¹
Igtp 3.7x10²
b
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
Max RPKM
Enpp2 9.0x10²
Otof 3.8x10¹
Ngb 4.0x10²
Ptgs2 9.6x10²
Inhba 3.2x10²
Rorb 6.4x10²
Ctxn3 5.3x10¹
Scnn1a 6.4x10²
Arf5 2.3x10²
Hsd11b1 7.5x10²
Deptor 2.2x10²
Myl4 1.5x10³
Tcerg1l 4.4x10¹
Pde1c 4.7x10¹
Penk 2.1x10³
Batf3 5.9x10¹
Ucma 1.5x10²
Qrfpr 4.5x10²
Bcl6 4.3x10¹
Tph2 1.2x10²
Stac 2.0x10¹
Cdh13 7.8x10²
Ddit4l 7.3x10²
Col6a1 1.1x10²
Chrna6 1.1x10²
Foxp2 8.6x10¹
Mgp 6.3x10²
Ly6d 8.2x10²
Sla 1.6x10²
Car12 1.2x10²
Syt17 2.2x10²
Prss22 1.3x10²
Ctgf 7.4x10¹
Tnmd 9.0x10²
Serpinb11 9.4x10¹
Rgs12 5.4x10¹
c
Astro Aqp4
OPC Pdgfra
Oligo 96*Rik
Oligo Opalin
Micro Ctss
Endo Xdh
SMC Myl9
Max RPKM
Tcf7l2 1.3x10³
Cd9 7.3x10³
Aqp4 9.2x10²
Cxcl14 2.4x10²
Gja1 6.5x10³
Gfap 8.8x10²
F3 4.2x10³
Sox9 6.8x10²
Olig1 6.0x10²
Pdgfra 6.5x10²
Cspg4 2.1x10²
Pcdh15 6.5x10²
Gria3 4.1x10²
96*Rik 2.2x10³
Brca1 3.0x10²
Enpp6 2.5x10³
Fyn 2.0x10³
Rnf122 4.8x10²
Mbp 1.0x10⁴
Mag 7.5x10²
Mog 2.7x10³
Opalin 2.8x10³
Itgam 1.1x10³
Ctss 1.2x10⁴
Cx3cr1 2.0x10³
C1qb 8.5x10³
Gpr34 5.6x10³
Flt1 4.6x10²
Id1 1.1x10³
Xdh 3.7x10²
Tbc1d4 1.8x10²
Exosc7 1.0x10³
Myl9 7.5x10³
Crip1 1.0x10⁴
Bgn 2.2x10³
Vtn 1.8x10⁴

Figure 3. Cortical cell types and corresponding marker genes. (**a-c**) Gene expression (rows) in individual cells (columns) arranged according to the cell type (top bar) and grouped according to major classes: GABAergic neurons (**a**), glutamatergic neurons (**b**), and non-neuronal cells (**c**). The scale is linear and adjusted to the maximum for each gene within each panel (maximum RPKM on the right). Only core cells are represented (N = 1424); for numbers of core cells per type see **Figure 2b**. *Tacr1* encodes neurokinin-1 receptor or substance P receptor; *96*Rik* is *9630013A20Rik*. Unique marker genes are in red.

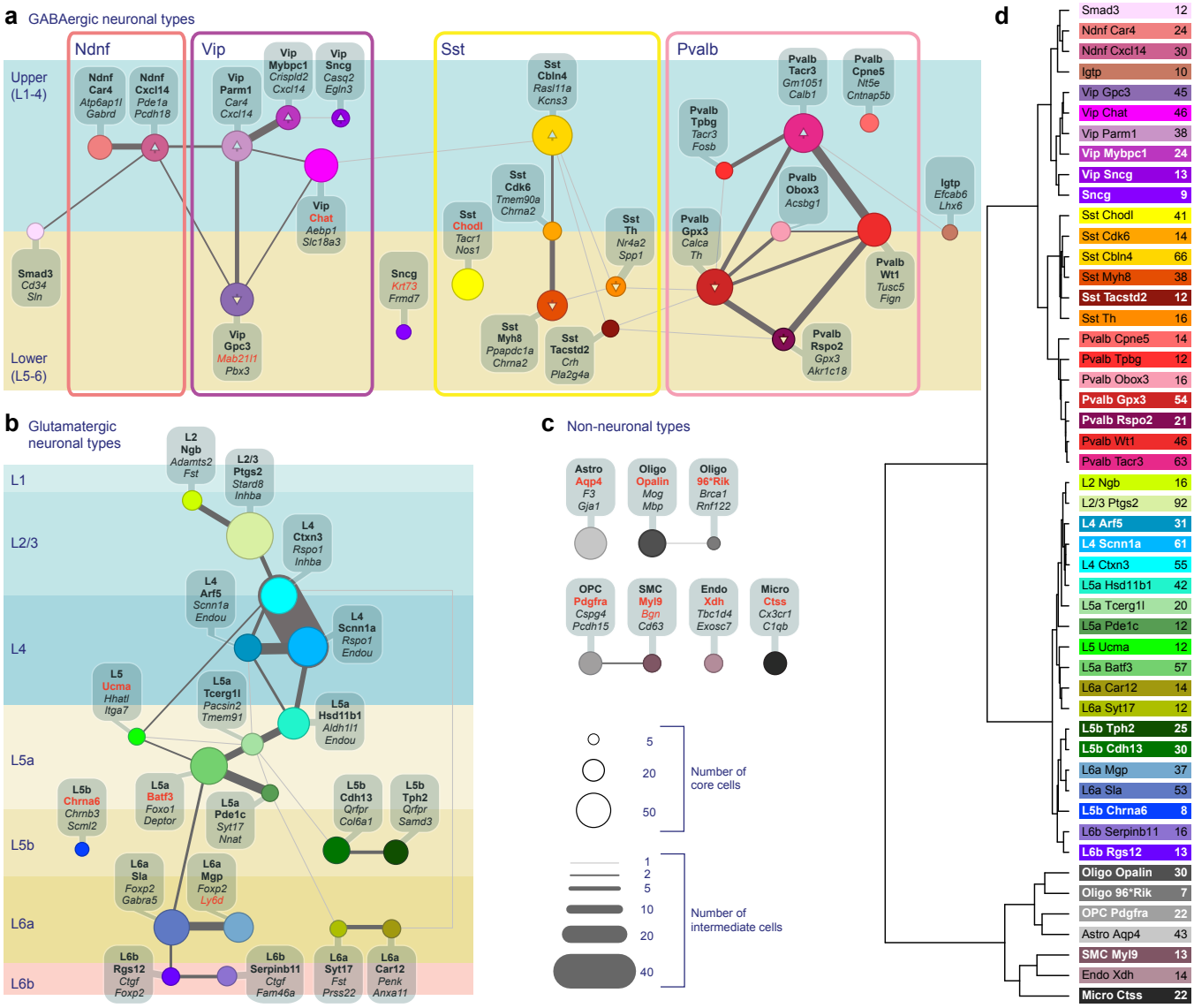

a GABAergic neuronal types
Ndnf
Vip
Sst
Pvalb
Upper (L1-4)
Lower (L5-6)
Ndnf Car4 Atp6ap1l Gabrd
Ndnf Cxcl14 Pde1a Pcdh18
Vip Parm1 Car4 Cxcl14
Vip Mybpc1 Crispld2 Cxcl14
Vip Sncg Casq2 Egln3
Vip Chat Aebp1 Slc18a3
Vip Gpc3 Mab21l1 Pbx3
Smad3 Cd34 Sln
Sncg Krt73 Frmd7
Sst Cbln4 Rasl11a Kcns3
Sst Cdk6 Tmem90a Chrna2
Sst Chodl Tacr1 Nos1
Sst Th Nr4a2 Spp1
Sst Myh8 Ppapdc1a Chrna2
Sst Tacstd2 Crh Pla2g4a
Pvalb Tacr3 Gm1051 Calb1
Pvalb Cpne5 Nt5e Cntnap5b
Pvalb Tpbg Tacr3 Fosb
Pvalb Obox3 Acsbg1
Pvalb Gpx3 Calca Th
Pvalb Wt1 Tusc5 Fign
Pvalb Rspo2 Gpx3 Akr1c18
Igtp Efcab6 Lhx6
b Glutamatergic neuronal types
L1
L2/3
L4
L5a
L5b
L6a
L6b
L2 Ngb Adamts2 Fst
L2/3 Ptgs2 Stard8 Inhba
L4 Ctxn3 Rspo1 Inhba
L4 Arf5 Scnn1a Endou
L4 Scnn1a Rspo1 Endou
L5 Ucma Hhatl Itga7
L5a Tcerg1l Pacsin2 Tmem91
L5a Hsd11b1 Aldh1l1 Endou
L5b Chrna6 Chrnb3 Scml2
L5a Batf3 Foxo1 Deptor
L5a Pde1c Syt17 Nnat
L5b Cdh13 Qrfpr Col6a1
L5b Tph2 Qrfpr Samd3
L6a Sla Foxp2 Gabra5
L6a Mgp Foxp2 Ly6d
L6b Rgs12 Ctgf Foxp2
L6b Serpinb11 Ctgf Fam46a
L6a Syt17 Fst Prss22
L6a Car12 Penk Anxa11
c Non-neuronal types
Astro Aqp4 F3 Gja1
Oligo Opalin Mog Mbp
Oligo 96*Rik Brca1 Rnf122
OPC Pdgfra Cspg4 Pcdh15
SMC Myl9 Bgn Cd63
Endo Xdh Tbc1d4 Exosc7
Micro Ctss Cx3cr1 C1qb
5
20
50
Number of core cells
1
2
5
10
20
40
Number of intermediate cells
d
Smad3 12
Ndnf Car4 24
Ndnf Cxcl14 30
Igtp 10
Vip Gpc3 45
Vip Chat 46
Vip Parm1 38
Vip Mybpc1 24
Vip Sncg 13
Sncg 9
Sst Chodl 41
Sst Cdk6 14
Sst Cbln4 66
Sst Myh8 38
Sst Tacstd2 12
Sst Th 16
Pvalb Cpne5 14
Pvalb Tpbg 12
Pvalb Obox3 16
Pvalb Gpx3 54
Pvalb Rspo2 21
Pvalb Wt1 46
Pvalb Tacr3 63
L2 Ngb 16
L2/3 Ptgs2 92
L4 Arf5 31
L4 Scnn1a 61
L4 Ctxn3 55
L5a Hsd11b1 42
L5a Tcerg1l 20
L5a Pde1c 12
L5 Ucma 12
L5a Batf3 57
L6a Car12 14
L6a Syt17 12
L5b Tph2 25
L5b Cdh13 30
L6a Mgp 37
L6a Sla 53
L5b Chrna6 8
L6b Serpinb11 16
L6b Rgs12 13
Oligo Opalin 30
Oligo 96*Rik 7
OPC Pdgfra 22
Astro Aqp4 43
SMC Myl9 13
Endo Xdh 14
Micro Ctss 22

Figure 4. Cell types summary and relationships. (**a–c**) Constellation diagrams showing core and intermediate cells for all cell types. Core cells (*N* = 1,424 total, 664 GABAergic, 609 glutamatergic, 151 non-neuronal) are represented by colored disks with areas corresponding to the number of core cells for each cluster. Linked tags include cell type names based on marker genes and layers; unique markers are in red. Intermediate cells (*N* = 255 total, 97 GABAergic, 155 glutamatergic, 3 non-neuronal) are represented by lines connecting disks; line thickness corresponds to the number of such cells. (**a**) GABAergic types are grouped according to major classes and arranged by their preferential location (enrichment) in upper versus lower cortical layers. Up and down arrows in disks represent statistically significant enrichment determined by layer-enriching dissections (Supplementary Table 5). Locations for other clusters are estimates that combine marker gene expression or Cre-line expression based on RNA ISH. The position at the border of upper and lower layers represents lack of evidence for location preference. (**b**) Glutamatergic types are arranged according to cortical layer. (**c**) Non-neuronal types share few intermediate cells among one another. *96*Rik*, *9630013A20Rik*. (**d**) Dendrogram depicting relatedness of the mean gene expression pattern for all cell types based on core cells (*N* = 1,424) and genes (*N* = 13,878) with s.d. for expression >1 across all types. The distance metric is Pearson's correlation coefficient over the genes in the $\log_{10}$(RPKM+1) space. The tree was generated by standard hierarchical clustering with average linkage.

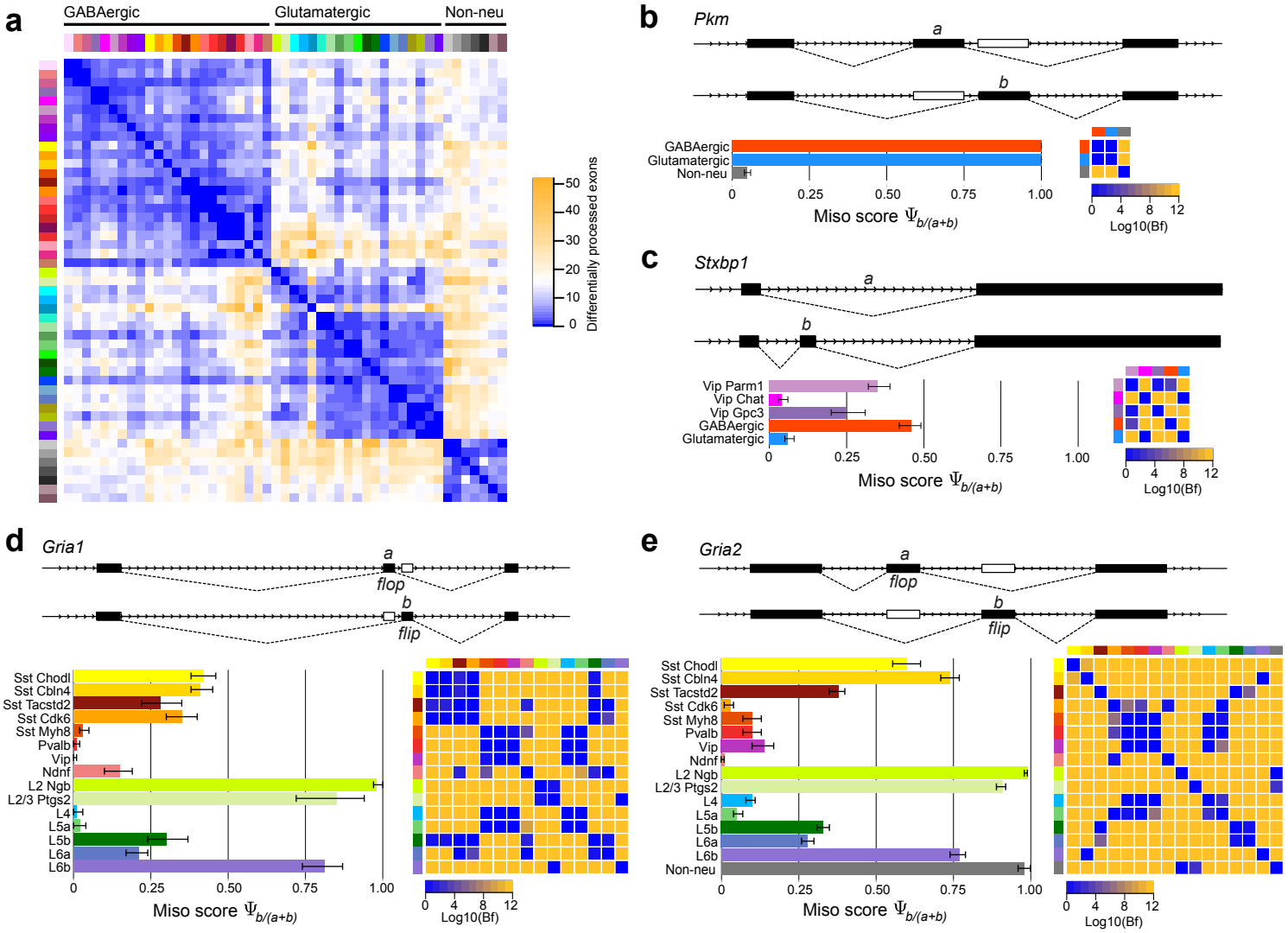

a
GABAergic
Glutamatergic
Non-neu
Differentially processed exons
50
40
30
20
10
0
b
Pkm
a
b
GABAergic
Glutamatergic
Non-neu
0
0.25
0.50
0.75
1.00
Miso score $\Psi_{b/(a+b)}$
0
4
8
12
Log10(Bf)
c
Stxbp1
a
b
Vip Parm1
Vip Chat
Vip Gpc3
GABAergic
Glutamatergic
0
0.25
0.50
0.75
1.00
Miso score $\Psi_{b/(a+b)}$
0
4
8
12
Log10(Bf)
d
Gria1
a
flop
b
flip
Sst Chodl
Sst Cbln4
Sst Tacstd2
Sst Cdk6
Sst Myh8
Pvalb
Vip
Ndnf
L2 Ngb
L2/3 Ptgs2
L4
L5a
L5b
L6a
L6b
0
0.25
0.50
0.75
1.00
Miso score $\Psi_{b/(a+b)}$
0
4
8
12
Log10(Bf)
e
Gria2
a
flop
b
flip
Sst Chodl
Sst Cbln4
Sst Tacstd2
Sst Cdk6
Sst Myh8
Pvalb
Vip
Ndnf
L2 Ngb
L2/3 Ptgs2
L4
L5a
L5b
L6a
L6b
Non-neu
0
0.25
0.50
0.75
1.00
Miso score $\Psi_{b/(a+b)}$
0
4
8
12
Log10(Bf)

Figure 5. Cell type–specific mRNA processing. (**a**) Heat map showing the number of differentially processed exons (*N* = 567 of 256,430 examined) for each pairwise comparison of transcriptomic cell types (Online Methods). (**b–e**) Confirmation of differential exon processing for four gene examples from a using MISO (Online Methods). Schematic of each gene (top) and corresponding quantitation (bottom).The MISO score (Ψ), or 'percent spliced-in', represents the relative exon usage of transcript variant *b* versus *a*, for each gene in each cell type. The significance in pairwise comparisons for all cell types for each alternatively processed exon was measured by the Bayes factor (Bf); Bf > 100 is considered significant. Bf for each alternatively processed mRNA is presented as the heat map to the right; yellow represents strongest statistical significance of Bf = 1×10^{12}. (**b**) In agreement with a population-level transcriptome profiling study[13], pyruvate kinase (*Pkm*) mRNAs displayed differential exon usage among neurons and non-neuronal cells. (**c**) Syntaxin binding protein 1 (*Stxbp1*) mRNAs showed differential processing among broad neuronal types, but also specific Vip types. (**d,e**) mRNAs for AMPA receptor genes, *Gria1* and *Gria2*, both displayed mutually exclusively spliced 'flip' and 'flop' exons. The two *Gria* genes showed similar alternative exon usage in same cell types, suggesting a shared mechanism for alternative splicing. For simplicity, all genes are shown in the same orientation. Error bars represent 95% confidence intervals, as calculated by the MISO software package.

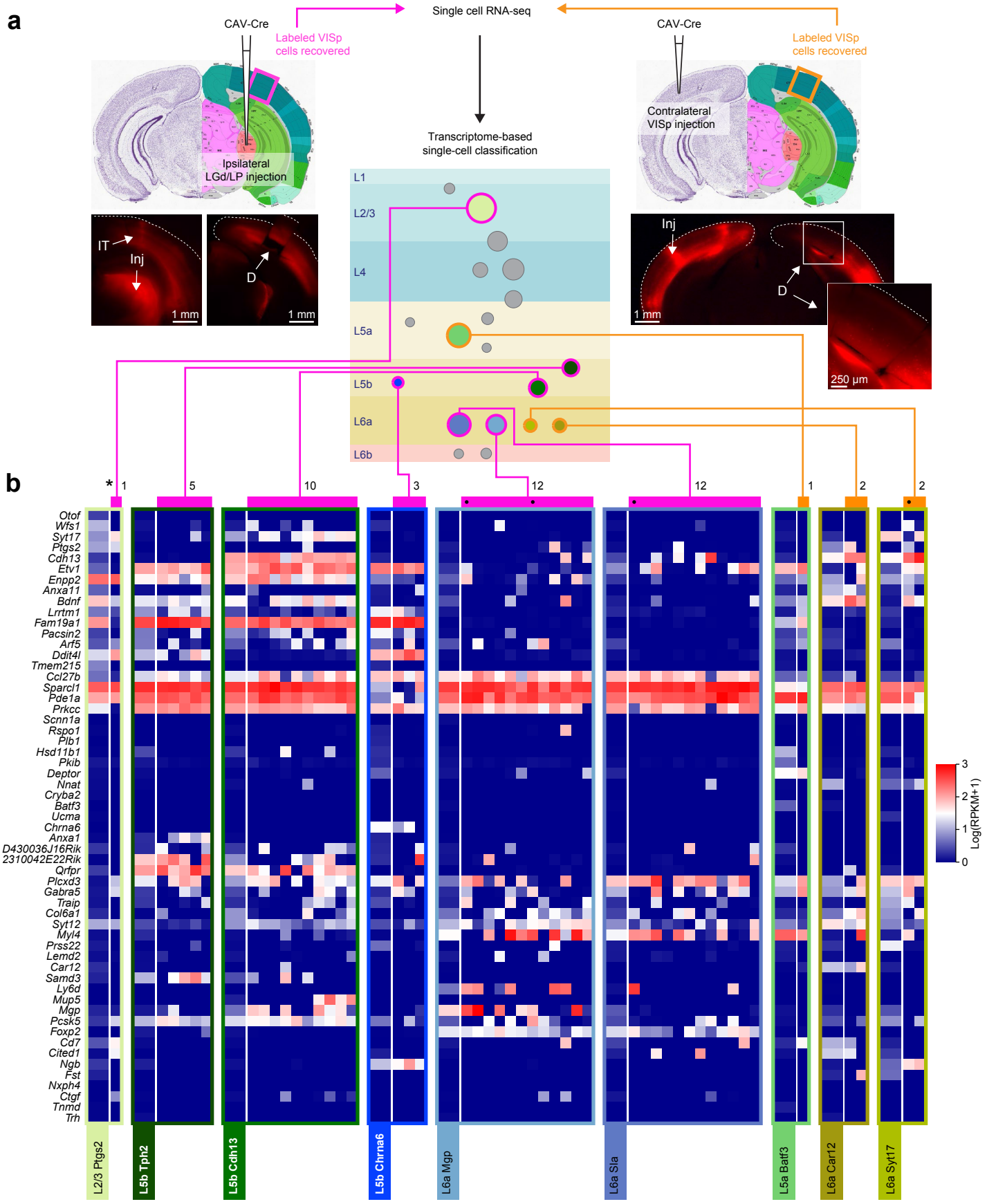

a
CAV-Cre
Labeled VISp cells recovered
Single cell RNA-seq
CAV-Cre
Labeled VISp cells recovered
Contralateral VISp injection
Ipsilateral LGd/LP injection
Transcriptome-based single-cell classification
IT
Inj
1 mm
D
1 mm
L1
L2/3
L4
L5a
L5b
L6a
L6b
Inj
D
1 mm
250 µm
b
*
1
5
10
3
12
12
1
2
2
Otof
Wfs1
Syt17
Ptgs2
Cdh13
Etv1
Enpp2
Anxa11
Bdnf
Lrrtm1
Fam19a1
Pacsin2
Arf5
Ddit4l
Tmem215
Ccl27b
Sparcl1
Pde1a
Prkcc
Scnn1a
Rspo1
Plb1
Hsd11b1
Pkib
Deptor
Nnat
Cryba2
Batf3
Ucma
Chrna6
Anxa1
D430036J16Rik
2310042E22Rik
Qrfpr
Plcxd3
Gabra5
Traip
Col6a1
Syt12
Myl4
Prss22
Lemd2
Car12
Samd3
Ly6d
Mup5
Mgp
Pcsk5
Foxp2
Cd7
Cited1
Ngb
Fst
Nxph4
Ctgf
Tnmd
Trh
L2/3 Ptgs2
L5b Tph2
L5b Cdh13
L5b Chrna6
L6a Mgp
L6a Sla
L5a Batf3
L6a Car12
L6a Syt17
3
2
1
0
Log(RPKM+1)

Figure 6. Transcriptomic signatures and axonal projections. (**a**) Schematic of experimental approach. CAV-Cre was injected into two different VISp projection areas in *Ai14* mice: ipsilateral visual thalamus (LGd/LP) or contralateral visual cortex (VISp). TdT^+ single cells were isolated from VISp by microdissection and FACS. Examples of fresh brain slices from injected animals are presented below. Inj, injection site; IT, injection tract; D, microdissected tissue used for preparation of single-cell suspension and FACS. Single-cell transcriptomes were obtained and used to classify the corresponding cells by the random forest approach (Online Methods) to our previously determined transcriptomic cell types. (**b**) Heat map panels show gene expression in individual projection-labeled cells classified into one of nine (of 49) previously determined transcriptomic types. Median gene expression in each type is shown to the left of each heat map panel. Black dots indicate cells that were classified as intermediate, but were primarily associated with the indicated cell types. The asterisk indicates an unexpected L2/3-Ptgs2 cell, which may have been labeled through the virus injection tract (IT). Note that these projection-labeled cells were not used in the original classification scheme to identify transcriptomic cell types. The number of cells obtained for each type is noted above. Total cells: $n = 43$ for two thalamus injection experiments; $n = 5$ for one contralateral VISp injection experiment.

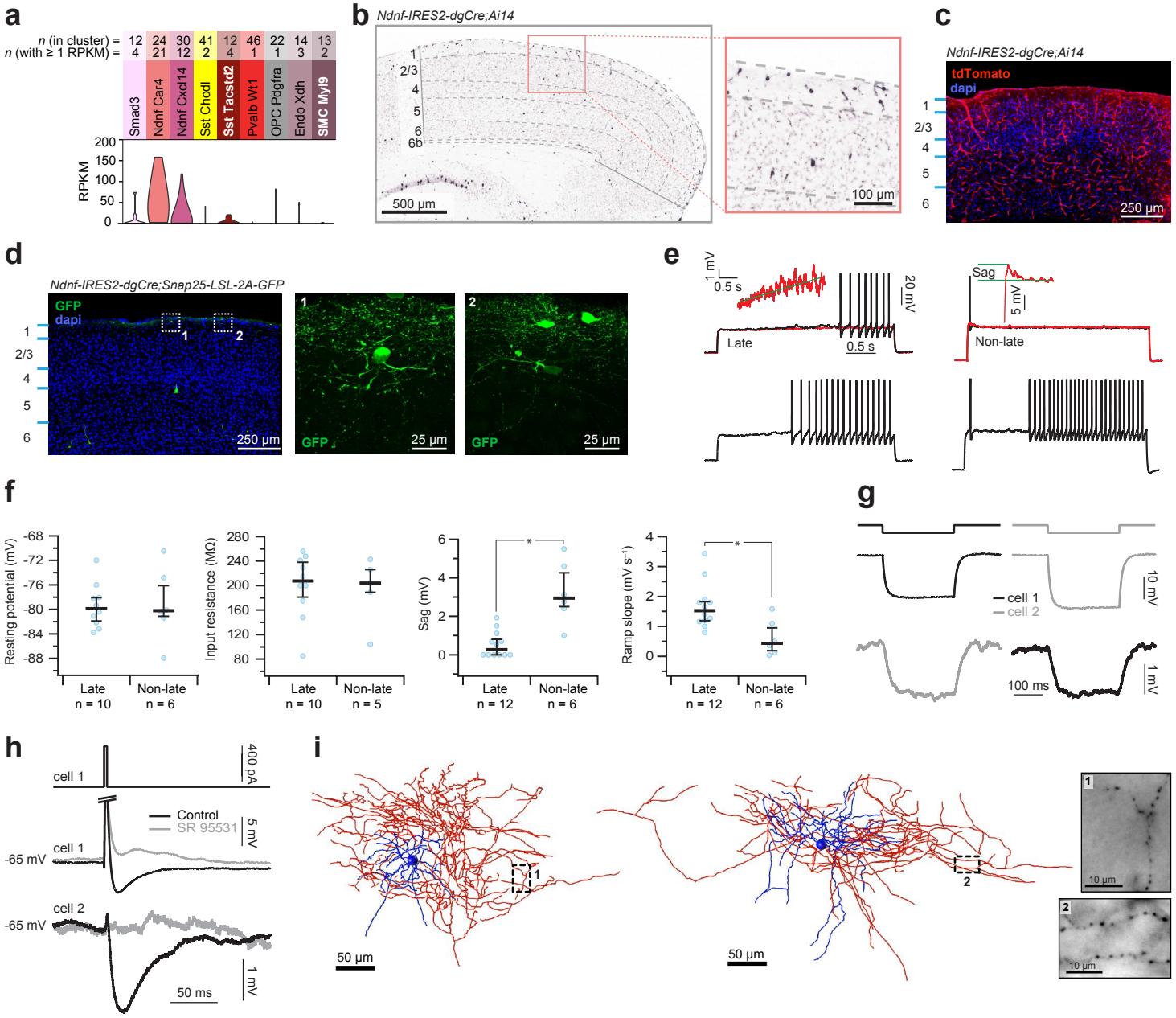

a
n (in cluster) = 12 24 30 41 12 46 22 14 13
n (with ≥ 1 RPKM) = 4 21 12 2 4 1 1 3 2
Smad3
Ndnf Car4
Ndnf Cxcl14
Sst Chodl
Sst Tacstd2
Pvalb Wt1
OPC Pdgfra
Endo Xdh
SMC Myl9
RPKM
200
150
100
50
0
b
Ndnf-IRES2-dgCre;Ai14
1
2/3
4
5
6
6b
500 µm
100 µm
c
Ndnf-IRES2-dgCre;Ai14
tdTomato
dapi
1
2/3
4
5
6
250 µm
d
Ndnf-IRES2-dgCre;Snap25-LSL-2A-GFP
GFP
dapi
1
2
250 µm
25 µm
25 µm
e
1 mV
0.5 s
20 mV
Late
0.5 s
Sag
5 mV
Non-late
f
Resting potential (mV)
-68
-72
-76
-80
-84
-88
Late
n = 10
Non-late
n = 6
Input resistance (MΩ)
280
240
200
160
120
80
Late
n = 10
Non-late
n = 5
Sag (mV)
6
4
2
0
*
Late
n = 12
Non-late
n = 6
Ramp slope (mV s⁻¹)
4
3
2
1
0
*
Late
n = 12
Non-late
n = 6
g
cell 1
cell 2
10 mV
100 ms
1 mV
h
cell 1
400 pA
Control
SR 95531
5 mV
-65 mV
cell 1
cell 2
-65 mV
50 ms
1 mV
i
50 µm
50 µm
1
2
10 µm
10 µm

Figure 7. Ndnf interneurons: genetic access and physiological properties. (**a**) Violin plot for *Ndnf* mRNA expression in cell types containing one or more cells with *Ndnf* RPKM ≥ 1. (**b**) Characterization of the *Ndnf-IRES2-dgCre; Ai14* transgenic mouse by RNA ISH detection of *tdT* mRNA in VISp. Inset focuses on upper layers. The image is from a representative section from one brain-wide experiment. (**c**) Data are presented as in **b** for tdT protein fluorescence. (**d**) The endothelial cells can be avoided if pan-neuronal Cre reporter (*Snap25-LSL-2A-GFP*) is used instead of *Ai14* (Online Methods). Insets, putative neurogliaform cells. The images in **c** and **d** are each representative of two independent experiments. (**e**) Intrinsic properties of tdT^+ neurons in L1 of VISp in *Ndnf-IRES2-dgCre; Ai14*. Sub- (red) and supra- (black) threshold responses to a 3-s square-pulse current injection of representative LS neuron (left) and NLS neuron (right). Top left inset, magnified view of the subthreshold depolarizing ramp. Bottom left, with larger current injection, the same LS neuron spiked earlier. Top right inset, magnified view of subthreshold initial response sag. The same neuron (bottom right) displayed additional late spiking in response to a larger current pulse. (**f**) Resting membrane potential, input resistance, sag and ramp slope for cells in **e** represented as averages ± s.e.m. LS neurons displayed significantly less sag ($P = 4.31 \times 10^{-4}$, Mann-Whitney test with 16 degrees of freedom) and significantly steeper depolarizing ramp than NLS neurons ($P = 6.52 \times 10^{-3}$, Mann-Whitney test with 16 degrees of freedom). (**g**) Recording of electrically coupled tdT^+ cells. Hyperpolarizing current injection (top) into either tdT^+ cell was transmitted to the other cell (bottom). 77% of cells were electrically coupled at an average intersomatic distance of 110 ± 11 μm and mean junctional conductance of 181 ± 41 pS ($n = 14$). Errors represent s.e.m. (**h**) Recording of a synaptically connected pair of tdT^+ cells. Action potential (middle, truncated) induced by a brief (3 ms) current injection (top) caused inhibitory postsynaptic potentials (IPSPs) in both neurons that were blocked by SR 95531 (5 μM). IPSP mean 10−90% rise time = 7.4 ± 0.5 ms; IPSP mean tau decay = 35.3 ± 7.6 ms, $n = 12$. Errors represent s.e.m. (**i**) Three-dimensional reconstructions of two biocytin-filled tdT^+ neurons (LS on the left, and NLS on the right) with cell bodies at the border between L1 and L2/3. Axons are red, dendrites and soma are blue. Insets, magnified views of bouton-like structures from original images. Morphological reconstruction of additional cells would be needed to assess whether the ones presented here on the left and right are generally representative of the LS or NLS spiking cells, respectively.

ONLINE METHODS

Data, reagent and code availability. To explore the annotated data set, an online interactive scientific vignette application has been developed and can be viewed through the Allen Brain Atlas data portal (http://www.brain-map.org) or directly at http://casestudies.brain-map.org/celltax. Note the change in cell type nomenclature in the paper compared to the original version of the online vignette (**Supplementary Table 6**). The newly generated mouse lines have been deposited to the Jackson Laboratory: *Ctgf-2A-dgCre-D* (JAX stock number 028525) and *Ndnf-IRES2-dgCre-D* (JAX stock number 028536). Supplementary Software contains the code for an iteration of the PCA and WGCNA-based clustering methods, the cluster membership validation algorithm, as well as the differential gene expression algorithm.

Mouse breeding and husbandry. All procedures were carried out in accordance with Institutional Animal Care and Use Committee protocols 0703 and 1208 at the Allen Institute for Brain Science. Animals were provided food and water *ad libitum* and were maintained on a regular 12-h day/night cycle at no more than five adult animals per cage. Animals were maintained on the C57BL/6J background. Newly received or generated transgenic lines were also backcrossed to C57BL/6J as much as possible, so that all animals used in this study had ≥75% of C57BL/6J background and on average 96% of C57BL/6J background (**Supplementary Table 11**). For the full list of recombinase and reporter lines, see **Supplementary Tables 1** (refs. 19,20,32,51–59) and **2** (refs. 55,57), respectively. All experimental animals were heterozygous for the recombinase transgenes and the reporter transgenes. Tamoxifen treatment for *CreER* lines was performed with a single dose of tamoxifen (40 μl of 50 mg ml^{-1}) dissolved in corn oil and administered via oral gavage at postnatal day (P)10–14. Tamoxifen treatment for *Nkx2.1-CreERT2* was performed at embryonic day (E)17 (oral gavage of the dam at 1 mg per 10 g of body weight), pups were delivered by cesarean section at E19 and then fostered. Trimethoprim was administered to animals containing *Ctgf-2A-dgCre* by oral gavage at postnatal day 35 ± 5 for three consecutive days (0.015 ml per g of body weight using 20 mg ml^{-1} trimethoprim solution). We excluded any animals with anophthalmia or microphthalmia for downstream experiments.

Generation of transgenic mice (*Ndnf-IRES2-dgCre* and *Ctgf-2A-dgCre*). Targeting constructs were generated using a combination of molecular cloning, gene synthesis (GenScript) and Red/ET recombineering (Gene Bridges). The 129S6/B6 F1 ES cell line, G4, was used to generate all transgenic mice by homologous recombination. Modified ES cell clones were injected into blastocysts to obtain germline transmission. Resulting mice were crossed to the *Rosa26-PhiC31o* mice (JAX Stock # 007743)[60] to delete the selection marker cassette, then backcrossed to C57BL/6J mice and maintained in the C57BL/6J background. The suffix "-*D*" in the name of the strain deposited to the Jackson Laboratory refers to the deletion of the selection marker cassette. The Ndnf-IRES2-dgCre contains an IRES2 sequence and a destabilized EGFP-Cre fusion protein (dgCre) inserted downstream of the Ndnf translational stop codon. The ecDHFR (R12Y/Y100I) domain of dgCre directs the proteosomal degradation of the entire EGFP/Cre fusion protein while administration of the DHFR inhibitor trimethoprim (TMP) via either intraperitoneal injection or oral gavage prevents degradation of the Cre fusion protein[61]. The *Ctgf-2A-dgCre* targeted transgene contains a viral 2A peptide (modified T2A, 5′-GAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCT-3′) and dgCre inserted in-frame and downstream of the coding sequence of the *Ctgf* gene. For the

Ndnf-IRES2-dgCre, the baseline dgCre activity (without TMP induction) was sufficient to label the cells with the *Ai14* and *Snap25-LSL-2A-GFP* reporters.

Retrograde labeling. We injected canine adenovirus expressing Cre recombinase (CAV-Cre, gift of Miguel Chillon Rodrigues, Universitat Autònoma de Barcelona)[62] into brains of heterozygous Ai14 mice using a previously described procedure with modifications[63]. Briefly, mice were anesthetized with 5% isoflurane and then placed into a stereotaxic alignment instrument (Kopf, model 1900). Anesthesia was maintained for the duration of the surgery by administering isoflurane at 1–2% through a nose cone. The skin along the midline of the skull was opened using a scalpel, and a surgical drill was used to create a small hole in the skull. A pulled glass pipette prefilled with CAV-Cre solution was lowered into the brain, and 165–500 nl of the virus solution was delivered to the targeted brain area using a pressure injection system (NanoJect II, Drummond Scientific Company, Catalog# 3-000-204). Stereotaxic coordinates were obtained from Paxinos adult mouse brain atlas[64] for visual thalamus (area LP, AP −2.30, ML 2.00, DV 2.60) and visual cortex (VISp/V1, −4.16, ML −3.00, DV 0.50). After the delivery of virus solution into the brain, the glass pipette was retracted and the incision in the scalp was closed using sutures. The animal was removed from the stereotaxic frame and allowed to recover from anesthesia. Mice were sacrificed 7−14 d after surgery for single cell isolation. TdT+ single cells were isolated from the ipsilateral VISp for thalamic injections (42 cells) or contralateral VISp for VISp injections (5 cells).

Single cell isolation. We adapted a previously described procedure to isolate fluorescently labeled neurons from the mouse brain[5,65]. Individual adult male mice (P56 ± 3) were anesthetized in an isoflurane chamber, decapitated, and the brain was immediately removed and submerged in fresh ice-cold artificial cerebrospinal fluid (ACSF) containing NaCl (126 mM), NaHCO3 (20 mM), dextrose (20 mM), KCl (3 mM), NaH2PO4 (1.25 mM), CaCl2 (2 mM), MgCl2 (2 mM), dl-AP5 sodium salt (50 μM), DNQX (20 μM), and tetrodotoxin (0.1 μM), bubbled with a carbogen gas (95% O2 and 5% CO2). The brain was sectioned on a vibratome (Leica VT1000S) on ice, and each slice (300−400 μm) was immediately transferred to an ACSF bath at room temperature (~25 °C). After the brain slicing was complete (not more than 15 min), individual slices of interest were transferred to a small Petri dish containing bubbled ACSF at room temperature. The regions of interest (all layers of VISp or specific layers of VISp) were microdissected under a fluorescence dissecting microscope, and the slices before and after dissection were imaged to later examine the location of the microdissected tissue and confirm its location within VISp. The dissected tissue pieces were transferred to a microcentrifuge tube and treated with 1 mg/ml pronase (Sigma, Cat#P6911-1G) in carbogen-bubbled ACSF for 70 min at room temperature without mixing in a closed tube. After incubation, with the tissue pieces sitting at the bottom of the tube, the pronase solution was pipetted out of the tube and exchanged with cold ACSF containing 1% fetal bovine serum (FBS). The tissue pieces were dissociated into single cells by gentle trituration through Pasteur pipettes with polished tip openings of 600-μm, 300-μm and 150-μm diameter[37].

Single cells were isolated by FACS into individual wells of 96-well plates or 8-well PCR strips containing 2.275 μl of Dilution Buffer (SMARTer Ultra Low RNA Kit for Illumina Sequencing, Clontech Cat#634936), 0.125 μl RNase inhibitor (SMARTer kit), and 0.1 μl of 1:1,000,000 diluted spike-in RNAs (ERCC RNA Spike-In Mix 1, Life Technologies Cat#4456740). Sorting was performed on a BD FACSAriaII SORP using a 130 μm nozzle, a

sheath pressure of 10 psi, and in the single-cell sorting mode. To exclude dead cells, DAPI (DAPI*2HCl, Life Technologies Cat#D1306) was added to the single cell suspension to the final concentration of 2 ng ml^{-1}. FACS populations were chosen to select cells with low DAPI and high tdT fluorescence. Accuracy of single cell sorting was evaluated as described in **Supplementary Figure 2a**, and confirmed post hoc by observing markedly higher expression of tdT mRNA in tdT^+ than in tdT^- cells (**Supplementary Fig. 2c**). In some cases, we also selected cells that have low DAPI and low tdT fluorescence, to capture tdT^- cells from a sample. To collect all cells in an unbiased manner, we selected all cells with low DAPI fluorescence, regardless of their tdT fluorescence level. Sorted cells were frozen immediately on dry ice and stored at −80 °C.

In total we used 72 animals, with at least two animals per Cre line in most cases. One animal each was used for the *Chat-IRES-Cre*, *Tac1-IRES2-Cre*, *Gad2-IRES-Cre*, and *Slc17a6-IRES-Cre* lines. The 72 animals were used for 55 specific dissection conditions (unique combination of Cre, layer dissection, and tdT labeling; **Supplementary Table 3**), with 34 conditions corresponding to one animal each, 13 conditions corresponding to two animals, and five conditions corresponding to three animals, two conditions corresponding to four animals, and one condition corresponding to five animals.

cDNA amplification and library construction. We used the SMARTer kit (SMARTer Ultra Low RNA Kit for Illumina Sequencing, Clontech Cat#634936) to reverse transcribe poly(A) RNA and amplify cDNA[14,66–68]. To stabilize the RNA after quickly thawing the plates or tubes containing cells on ice, we immediately added to each sample an additional 0.125 µl of RNase inhibitor mixed with SMART CDS Primer II A. All steps downstream were carried out according to the manufacturer's instructions. We performed reverse transcription and cDNA amplification for 19 PCR cycles in 96-well plates or 0.2-ml strip-tubes. Each amplification experiment included a set of controls: 10 pg cortex RNA (isolated from *Rbp4-Cre;Ai14*, P57 male) as positive control for amplification, ERCC-only control to demonstrate the absence of RNases throughout the sorting process, and water-only control, to control for specificity of amplification/absence of contamination. cDNA concentration was quantified using Agilent Bioanalyzer High Sensitivity DNA chips. For most samples, 1 ng of amplified cDNA was used as input to make sequencing libraries with the Nextera XT DNA kit (Illumina Cat#FC-131-1096). For smaller cells (for example, glia), which did not consistently produce more than 1 ng cDNA, we used 0.5–1 ng cDNA as input. We stopped the procedure after PCR clean-up and did not perform library normalization or library pooling. Individual libraries were quantified using Agilent Bioanalyzer DNA 7500 chips. In order to assess sample quality and adjust the concentrations of libraries for multiplexing on HiSeq, all libraries were sequenced first on Illumina MiSeq to obtain approximately 100,000 reads per library, and then on Illumina HiSeq 2000 or 2500 to generate 100-bp reads.

Sequencing data processing and QC. 100 base-pair single-end reads were aligned to GRCm38 (mm10) using the RefSeq annotation gff file downloaded on 6/1/2013. Transcriptome alignment was performed using RSEM69, and unmapped reads were then aligned to the ERCC and tdT sequences using Bowtie[70]. The remaining unmapped reads were aligned to the mm10 genome using Bowtie. Genome-mapped reads were not used further in the analysis. Iterative PCA clustering was performed using RPKM (reads per kilobase per million mapped reads) values, while iterative WGCNA clustering used TPM (transcripts per million) values. Differential

expression analyses with DESeq2 (ref. 71) and DESeq72 both use raw read counts. After the alignment, we performed QC (**Supplementary Fig. 3b**) to exclude 60 out of 1739 cells.

Clustering. We used two independent clustering methods to identify a set of clusters, which were the input into the subsequent validation stage to assess robustness of cluster membership. The first method, iterative principal component analysis, iteratively identifies groups of cells in principal component space, subdividing cells into two groups until a set of termination criteria are met (see below), indicating lack of further structured subdivision. At each iteration, the following steps are carried out, using only data from those cells under consideration at the specific iteration.

1. Identify genes with more variance than technical noise, as determined by ERCCs[71]. Four sets of genes were selected, corresponding to % CVs greater than 0%, 25%, 50% and 100% above the technical noise fit based on ERCCs. At each iteration, the percentage threshold that generated the best separation (as determined by the sigClust P value, described below) was selected. In general, when multiple thresholds yielded significant *P* values for segregation, they resulted in identical clustering.

2. Perform PCA on the log-transformed *z*-scored data matrix and identify the number of relevant PCs by looking for the shoulder in the eigenvalue spectrum. Initially, the number of relevant PCs was selected by shuffling the data matrix 100 times and calculating the mean and s.d. of the first eigenvalue, and selecting those PCs whose eigenvalue was greater than the mean + 2 s.d. However, it was quickly apparent that this method yielded the same results as simply visually inspecting the eigenvalue scree plot for the existence of a shoulder in the spectrum, a standard procedure for this type of application.

3. After selecting the number of relevant PCs, generate a cell-cell distance matrix by calculating the Euclidean distance between cells in PC space, weighting each PC dimension by the corresponding eigenvalue.

4. Cluster cells using Ward's method using the distance matrix generated in step 3 and split cells into two groups based on the top branch of this tree.

5. Assess the significance of the binary split using the sigClust package in R, which generates a *P* value for the null hypothesis that the data points are drawn from a single multivariate Gaussian, as opposed to two Gaussians.

6. Since steps 1–5 are carried out for four different technical noise thresholds, select the one that provides the best separation of cells into two groups, based on the PC spectrum and sigClust *P* value.

The first iteration of this procedure begins with all cells, and then proceeds subsequently for the groups of cells generated at each binary split. A given branch in this iterative tree ends when any of the following termination criteria are met.

1. There are no cellular genes with variance greater than technical noise.
2. There is no significant shoulder in the PC spectrum.
3. sigClust does not return a *P* value < 0.01.
4. If the group of cells at that iteration is smaller than 4.

This procedure results in a final set of PCA-defined clusters.

We also developed an alternative clustering approach which iteratively applies WGCNA[73] to the data set, similar to the iterative PCA approach. At each iteration, the following steps are carried out.

1. Identify genes with more variance than technical noise, as determined by ERCCs[71], with an adjusted *P* value threshold varying from 0.001 at the top level to 0.5 at bottom level to select genes above the technical noise fit curve.

2. Run standard WGCNA with the soft thresholding power set to 4, and minimal gene cluster size at 10.

3. For each WGCNA gene module, cluster the cells based on the member genes into two clusters. If one cluster contains fewer than 4 cells, remove the gene module, which likely marks potential outliers. Then identify the differentially expressed (DE) genes between the two clusters, and compute the DE score as the sum of −log10(adjusted *P* value) of all DE genes. Select only the modules with DE score of at least 60.

4. Take the genes from all remaining gene modules and perform hierarchical clustering with using Ward's method. Select the optimal number of clusters by maximizing the sum of DE scores for all pairwise comparisons between clusters.

5. From this initial clustering, sharpen the boundaries of the groups by identifying DE genes among all pairs of clusters (using the limma package in R)[74], and reclustering using this set of DE genes.

For iterative WGCNA, the clustering terminates if the group of cells at that iteration is smaller than 4 or if there are no significant gene modules at the given DE score threshold. The threshold is chosen based on performing the same analysis on the shuffled data matrix.

Validation of cluster membership. Once cluster identities have been determined, we ran a standard machine learning-based cross-validation approach that consisted of the following steps:

1. Remove 20% of the cells and extract differentially expressed genes among all pairs of clusters (using the remaining 80% of the cells) using the limma package in R[74].

2. Train a random forest classification scheme (with 1000 trees) on every pair of groups within the 80% of cells using the differentially expressed genes from step 1 for each pair of groups.

3. Run the classifier on the 20% of cells that were removed. For every pair of cell clusters, run the appropriate classifier form step 2, and determine which of the two groups the cell belongs to.

4. Repeat steps 1-3 five times with mutually exclusive groups of cells forming the 20%, such that each cell is classified once among every pair of clusters.

5. Repeat steps 1–4 ten times, such that every cell is classified ten times among every pair of clusters.

6. For each cell, tabulate the number of times that cell was classified into each cluster. For each pair of clusters, identify whether one cluster dominates the other for that given cell (the cell is classified 10 out of 10 times into one of the clusters), and retain only the set of non-dominated clusters. These non-dominated clusters are identified as those where the cell is always classified at least 1 time, in all pairwise comparisons with other clusters. Cells that were classified into a single non-dominated cluster 10/10 times are labeled core cells, and the remainder—for which more than one non-dominated cluster remains—are labeled intermediate cells. For every cell, its membership score to each cluster is calculated as the proportion of times it was classified into each non-dominated cluster.

This cross-validation was run on the terminal PCA clusters and the terminal WGCNA clusters separately, and all clusters with fewer than 4 core cells were removed. The remaining

clusters were then intersected to define a consensus set of clusters (see below). The cross-validation was then run on this consensus set of clusters, and any clusters with fewer than 4 cells were removed.

Once an ultimate set of consensus clusters was obtained, the results of this cross-validation technique were used to label all of the original cells as either ‘Core’ or ‘Intermediate’, using the same criteria specified in cross-validation step 6, above.

There are two tunable parameters in this cross-validation algorithm: 1) the number of differentially expressed genes used to distinguish pairs oftranscriptomic types from each other (20 genes, for the cross-validation in the paper), and 2) the *P* value threshold for selecting differentially expressed genes ($P < 0.05$, for the cross-validation in the paper). To assess the impact of the two parameters described above, we ran the cross-validation algorithm multiple times to assess how cell assignments change based on these parameters. For the default values presented in the paper (20 genes per pair of transcriptomic types, genes selected at $P < 0.05$), we obtained 1,424 core and 255 intermediate cells. For 20 genes and $P < 0.01$, we obtained 1,413 core and 266 intermediate cells. Restricting the number of genes to 10 does not have a major effect: we obtained 1,423 core/256 intermediate cells using a differential expression $P < 0.05$ and 1,418 core/261 intermediate cells at $P < 0.01$. Increasing the number of genes to 50 per pair of transcriptomic types, however, results in more cells being classified as intermediate: 1,369 core/310 intermediate cells using $P < 0.05$ and 1,383 core/296 intermediate cells. The full assignments of each cell for each of these conditions are provided in **Supplementary Table 12**. In summary, although the changes in the two parameters affect classification for some of the cells, the number and identity of core clusters is maintained despite the variation in the parameters.

Note on minimal cluster core size. When minimal size of cluster core was set to 3, additional clusters were detected by the iterative PCA and WGCNA approaches. Examination of some of these small clusters suggests that they probably represent genuine cell types that will become more apparent with additional cell sampling:

1. A subset of 4 cells within the SMC-Myl9 type (Ct1988_V, Ct1994_V, Ct1986_V and Nd1968n_V1), which do not express *Myl9* and *Flt1*, but express *Lum*, *Dcn*, *Col1a1* and *Aox3*. Although iterative PCA segregated this set of 4 cells initially as a separate cluster, one cell from this cluster showed similarity to the rest of the SMC-Myl9 cells, and was thus classified as an intermediate cell. The remaining cluster then contained only 3 core cells, and so did not pass the 4-cell minimum requirement; this cluster was re-merged with the SMC-Myl9 cluster for subsequent analyses.

2. Subsets of cells within the Sst-Th type with mutually exclusive expression of *Th* and *Spp1*.

3. A subset of 3 cells (D1217_V, D1222_V, H1418_V6b) that express *Krt73* and *Cyb5r2* but not *Vip* within the Sncg type.

It is important to note that the minimum cluster size (4 cells) is the lowest possible number for the cross-validation algorithm above, because variance estimates for gene expression require at least 3 cells within a group (and one cell will be removed from the group during the membership assessment approach). These gene expression variance estimates are necessary to identify differentially expressed genes between groups, a crucial step in cluster membership assessment. As a result, minimum cluster size is not a parameter that can be decreased when

employing our cross-validation algorithm.

Cluster intersection. Both clustering methods (iterative PCA and iterative WGCNA) yield a set of terminal clusters. For each of the two methods, we identified clusters containing ≥ 4 core cells (as explained above). We then assessed the overlap of these clusters obtained from the two clustering methods. Whereas the majority of clusters obtained by both methods overlap, there were eight cases where one method subdivided a set of cells differently from the other (**Supplementary Fig. 8**). In these cases, we generated a set of clusters based on the finest set of subdivisions, taken as the intersection of partitions from both methods.

Mapping CAV-Cre-labeled cells to RNA-seq clusters. To map the CAV projection-labeled cells to the final set of clusters, a technique very similar to the cross-validation step was performed, except that none of the original (non–projection labeled) cells were removed when training the random forest classifier, and the classifier was used only on the projection-labeled cells.

Identifying discriminatory genes and marker gene sets. To identify key discriminatory genes, we first assembled lists of all differentially expressed genes among all pairs of types in the glutamatergic, GABAergic and non-neuronal major categories. We also identified differentially expressed genes between all neurons and all non-neuronal cells, as well as between all glutamatergic and all GABAergic neurons. In all cases, differential expression was calculated using the DESeq package for R[72]. After assembling lists of significant (adjusted P value < 0.01) differentially expressed genes, we then selected a subset of them using the following criteria:

1. For a given pair of cell types, select only those genes whose 20th percentile expression in type 1 is greater than the 80th percentile expression in type 2. This ensures a good separation of distributions.

2. For a given pair of cell types, the 80th percentile expression for a given gene must be <1 RPKM for one of the types. This ensures close to zero expression for the lower group, helping to generate an approximate on-off separation among the two groups.

Additional marker genes were identified based on the percentage of cells in each cell type in which each differentially expressed gene was detected (>0 RPKM). This was done in using a pairwise comparison method to identify genes expressed specifically in individual or few cell types:

1. For each cell type, each gene was analyzed to determine if its expression was biased significantly toward the selected cell type compared to each other cell type (>95% of cells in the selected type, and <5% in the other).

2. Each gene was scored based on the number of clusters for which the gene was associated with the selected cell type. Genes were ranked according to this score, and the top genes were selected.

3. If no genes were identified for a given cluster in steps 1–2, the 95% threshold for expression was reduced to a minimum of 80%.

4. To detect highly specific but sparsely expressed markers, the upper and lower thresholds were adjusted to 30% and 0%, respectively.

After selecting the genes this way, we also selected genes that distinguished among the maximum number of cell type pairs in the following categories: all glutamatergic types, all

GABAergic types, all non-neuronal types, all Sst types, all Pvalb types, and all Vip or Ndnf types. Genes selected by both methods were visually inspected for type or category-specific expression characteristics by plotting heatmaps of gene expression for all cells in all types. These lists were augmented with known markers from the literature, and the results are presented in **Supplementary Figure 12** and **Supplementary Table 6**.

Evaluation of differential exon usage. We used the limma Bioconductor package[74] to detect differentially expressed exons between every pair of transcriptomic cell types. Input data into limma were $\log_2$-transformed read counts for each exon that were previously scaled by the total number of reads in each sample. We considered significant at least a twofold change and adjusted $P < 0.05$. In addition, we used a custom code to detect exons associated with alternative processing events, defined as those that utilize the same splicing acceptor or donor as another exon within the data set. From these candidates, we selected only the exons that are differentially expressed compared to their corresponding gene. We used MISO[75] to confirm differential exon processing for select examples. The MISO score (Ψ), or 'percent spliced-in', represents the relative exon usage of transcript variant *b* versus *a*, for each gene in each cell type[75]. Because MISO does not accommodate replicates, to calculate MISO Ψ, we pooled ten randomly selected single cell samples for each cell type (20 for broad glutamatergic, GABAergic and non-neuronal types) for each pairwise comparison. The significance in pairwise comparisons for all cell types for each alternatively processed RNA was measured by the Bayes factor (Bf). Bf corresponds to the odds of differential expression (change in Ψ score that is nonzero) over no differential expression (change in Ψ score = zero). Bf > 100 is considered significant.

Estimation of cellular total RNA content. As stated in the single-cell cDNA amplification and library preparation section, we added the same amount of synthetic ERCC transcripts to each sample containing a single cell before reverse transcription and cDNA amplification. After obtaining and mapping the next generation sequencing reads from the samples, we calculated the percentage of ERCC reads in each sample. This ratio of ERCC versus cellular reads was used to estimate the mass of mRNA in each cell. To do this, we converted the known numbers of added ERCC molecules and their weights to femtograms of RNA, and by simple proportion estimated the mass of cellular mRNA in that cell. To estimate the total RNA mass, we assumed that the mRNA to total RNA ratio in all cells is the same as in total cortex RNA, and used the samples containing 10 pg cortex total RNA to estimate the appropriate amounts of single cell total RNA.

RNA double-fluorescence *in situ* hybridization (DFISH). We performed RNA DFISH experiments using a previously described protocol[76], which was based on the Allen Institute's colorimetric RNA ISH protocol[1]. Tissue sections (25 μm) were collected from fresh frozen brains of P53 male C57BL/6J mice. Riboprobes were labeled with digoxigenin (DIG) or dinitrophenyl-11-UTP (DNP, Perkin Elmer) (**Supplementary Table 13**). Probe pairs were simultaneously hybridized onto the tissue sections, and the signal from each probe was sequentially amplified with tyramide (anti-DIG-HRP and tyramide-biotin, or anti-DNP-HRP and tyramide-DNP). The amplified signal was detected by labeling with streptavidin-Alexa-Fluor 488 (Life Technologies) or anti-DNP-Alexa-Fluor 555 (Life Technologies). The DFISH protocol was carried out on Leica autostainers, and images were taken using a 10x objective on a fluorescence microscope (VS110 Virtual Slide Microscope, Olympus).

High-throughput qRT-PCR. *Assay Selection.* PrimeTime qPCR assays (containing forward primer, reverse primer and probe), provided by Integrated DNA Technologies (IDT) were selected using the IDT Assay Selection Tool on the IDT web site. Primer and probe sequences were compared against the mouse UCSC transcripts to identify assays that: 1. maximize detection of all isoforms of a gene; 2. span large introns to minimize detection of corresponding genomic DNA. When a PrimeTime qPCR assay that met our requirements was not available, we used the PrimerQuest Custom Design Tool provided by IDT. If a single assay for detection of all isoforms could not be designed, multiple assays were ordered and subjected to validation.

Assay validation. All assays were validated using a dilution series of total RNA (in pg: 1, 5, 10, 32, 100, 320, 1,000, 3,200, 10,000 and 20,000 per reaction) from whole mouse (RNA pool from 11 mouse cell lines, Agilent Tech quantitative Mouse Ref RNA, Cat#750600), mouse brain (Zyagen MR-201) and mouse cortex (isolated from *Rbp4-Cre;Ai14* P57 male mouse). All dilutions were run in triplicate. To pass validation, each assay had to show linear RNA detection ($R^2 > 0.85$) across a minimum of 5 dilution points in at least one RNA background. Each assay also had to show no detection below the limit of detection (LOD) in water and 50 pg mouse genomic DNA control wells (LOD was set at 2 s.d. above the mean for all single copy ERCC transcripts detected). To assess the specificity of assays, we also tested them against a dilution series of several single cell cDNAs (libraries ranged from 10–1,000 pg) that were previously subjected to RNA-seq and that displayed differential expression of genes of interest. Only assays that showed linear RNA detection, low background and good specificity were used. The sequences for the final set of validated assays are available in **Supplementary Table 14**.

Single cell qRT-PCR. Experiments were performed using Fluidigm BioMark according to manufacturer's instructions. Single cells were isolated as described above, and deposited by FACS into individual wells of 96-well plates containing 5.1 µl of buffer (5 µl of Cells Direct 2× Reaction Mix (Thermo Fisher Scientific) and 0.1 µl of SUPERase In RNase Inhibitor (20 U μl^{-1}; Thermo Fisher Scientific)) and frozen at −80 °C. Synthetic transcripts (ERCC RNA Spike-in Mix1, Life Technologies Cat#4456740, 1 µl of 550,000 x-dilution added per sample) were included in all reverse transcription-specific target amplification (RT-STA) reactions except the two negative, water-only controls. The RT-STA included 20 cycles of PCR. Each RT-STA sample was diluted fivefold and analyzed by 96.96 chip that included control assays and RT-STAs. Control assays corresponded to 9 different ERCC RNAs (**Supplementary Table 14**) and 3 housekeeping genes (Ppia, Gapdh, Tfrc). Control templates included eight different whole mouse RNA dilutions (1, 5, 10, 32, 100, 320, 1,000 and 3,200 pg per RT-STA reaction), gDNA (100 pg per RT-STA reaction), water with ERCCs and water without ERCCs. The ERCC controls allowed monitoring of the PCR efficiency in each sample well (the nine assayed ERCC transcripts cover a range from 1–4,100 RNA copies). Any sample well that did not display linear ERCC transcript amplification was flagged or failed. The bulk RNA dilution series and reference assays allowed us to monitor chip to chip variation.

Electrophysiology. *Slice preparation.* Coronal cortical slices were obtained from P51 ± 10–day-old *Ndnf-IRES2-dgCre;Ai14* mice. Mice were anesthetized with 5% isofluorane, perfused transcardially with ACSF and decapitated. The brain was then removed from the skull and coronal visual cortex slices (300 µm) were prepared using a vibratome. Slices were transferred to an incubation chamber (34 °C) for 10 min and then to a holding chamber at 22 °C. For perfusion and slice incubation, ACSF contained (in mM): 98 NMDG, 98 HCl, 25 D-glucose, 25 $NaHCO_3$,

17.5 HEPES, 12 N-acetyl-L-cysteine, 10 $MgSO_4$, 5 Na-(L)-ascorbate, 3 myoinositol, 3 Na-pyruvate, 2.5 KCl, 2 mM thiourea, 1.25 NaH_2PO_4, and 0.01 taurine, or 73 Tris-HCl, 30 $NaHCO_3$, 28 Tris Base, 25 D-Glucose, 20 HEPES, 10 $MgSO_4$, 5 Na-(L)-ascorbate, 3 Na-pyruvate, 2.5 KCl, 2 thiourea, 1.2 NaH_2PO_4, and 0.5 $CaCl_2$. The holding chamber solution contained (in mM): 97 NaCl, 25 $NaHCO_3$, 25 D-glucose, 14 HEPES, 12.3 N-acetyl-L-cysteine, 5 Na-(L)-ascorbate, 3 myoinositol, 3 Na-pyruvate, 2.5 KCl, 2 $CaCl_2$, 2 $MgSO_4$, 2 thiourea, 1.25 NaH_2PO_4, and 0.01 taurine.

Patch-clamp recording. Recordings were performed in ACSF containing (in mM): 126 NaCl, 26 $NaHCO_3$, 12.5 D-glucose, 2.5 KCl, 2 $CaCl_2$, 1.25 NaH_2PO_4, and 1 $MgSO_4$. Individual slices were held in a small chamber perfused with ACSF at 2.5 mL min^{-1} (32–34 °C) and visualized with an upright, fixed-stage microscope (Scientifica SliceScope) using dodt-gradient contrast, infrared video microscopy. Fluorescent tdT+ neurons were identified using simultaneous epifluorescent imaging. Single to quadruple whole-cell current-clamp recordings were made with MultiClamp 700B (Molecular Devices) amplifier(s) and patch electrodes with an open tip resistance of 5–7 MΩ. The intracellular solution contained (in mM) 126 K-gluconate, 10.0 HEPES, 4 KCl, 4 Mg-ATP, 0.3 EGTA, 0.3 Na-GTP, 10 Na-phosphocreatine, and 0.5% biocytin. Cells were maintained with bias current over the course of the experiment at the resting potential observed 2 min after the whole-cell configuration was achieved, except in synaptic experiments, where cells were held at −65 mV. Synaptic transmission was blocked in experiments investigating the intrinsic properties (**Fig. 7e–g**) of tdT+ neurons with 1 mM kynurenic acid and 0.1 mM picrotoxin. AMPA receptors were blocked in using 2 μM NBQX during experiments investigating electrical coupling and GABAergic synaptic transmission between tdT+ neurons (**Fig. 7h,i**).

Data acquisition and analysis. Data were transferred to a computer during experiments by an ITC-1600 digital-analog converter (Heka). Igor Pro software (Wavemetrics) was used for acquisition and analysis. Electrophysiological records were filtered at 10 kHz and digitally sampled at 50, 67, 100, or 200 kHz. Gj = (1/R2) × CC/(1−CC) was used to calculate the junctional conductance. CC is the coupling coefficient and R2 is the input resistance of the non-injected cell[43].

Morphological reconstruction of biocytin-stained neurons. Brain slices containing biocytin-filled neurons were fixed with 4% paraformaldehyde and stained using ABC-DAB detection kit (Vector Laboratories). Stained slices were post-fixed with 0.05% OsO4 and mounted with MOWIOL 4-88. Biocytin-filled neurons were imaged in 3D with a Zeiss AxioImager at 63× with 1.4-NA objective. Images were then inverted and imported into Vaa3D[77] for semi-automated reconstruction using the virtual finger tool.

Methodological note on Cre transgene expression. Cre transgenes are usually made to mimic the expression of an endogenous gene. However, the Cre transgene expression does not necessarily mimic the corresponding endogenous gene expression as the transgene may have some of the regulatory elements missing or altered, or new regulatory elements present due to position effects. In addition, the expression of a Cre transgene is usually monitored by the activation of a Cre-reporter transgene, which is expressed from a strong and ubiquitous promoter (as is the case for *Ai14*, which is used throughout this study). This approach has two additional consequences. First, the Cre reporter gene expression reflects Cre expression throughout

developmental history of the cell and any of its progenitors, and second, the Cre transgene expression, which is variable and may be low, is converted into very strong and binary Cre-reporter gene expression.

We have extensively characterized expression of Cre transgenes and/or Cre-reporter genes by RNA ISH or fluorescence as part of the transgenic characterization pipeline at the Allen Institute[19]. In many instances, this type of examination has already revealed that the mRNA expression of the endogenous gene does not fully correlate with the corresponding Cre transgene or the Cre transgene-dependent reporter expression. It is also important to note that the identity of the Cre-dependent reporter matters: some reporters are more susceptible to Cre-mediated recombination than others. And finally, additional discrepancies may be encountered if endogenous gene expression is examined at the protein level, while Cre protein expression may not be under the same regulation. Several examples below illustrate the apparent or true discrepancies between transgenic Cre and corresponding endogenous gene expression.

Example 1. Ntsr1 and *Nr5a1* mRNAs are not detectable in the cortex by RNA ISH in the Allen Mouse Brain Atlas, although the Cre-mediated reporter gene expression is detected in L6 and L4, respectively. Cre expression in the cortex could therefore be interpreted as an artifact of transgenesis. However, by single-cell RNA-seq, we clearly detected *Ntsr1* and *Nr5a1* mRNAs in some L6 and L4 cells. The corresponding mRNAs are present at a low level, and not consistently among all tdT^+ cells isolated from the *Ntsr1-Cre;Ai14* and *Nr5a1-Cre;Ai14* lines, respectively. This shows that the Cre expression in this case does reflect the endogenous gene expression, and the fact that tdT expression is broader than endogenous gene expression likely reflects conversion of low and variable endogenous gene expression into strong Cre-dependent reporter expression.

Example 2. Although *Calb2-IRES-Cre* is a knock-in line, we observe discrepancies in its expression compared to endogenous *Calb2* expression in adult. First, we observe Cre-dependent tdT expression (from the *Ai14* reporter) in glutamatergic cells. As we do not detect expression of *Calb2* mRNA by RNA-Seq in glutamatergic cells, this may be a result of developmental *Calb2* expression or a transgenic artifact. Second, based on RNA-Seq, this line should label some Sst cells, but Sst-positive cells were not among tdT^+ cells collected from *Calb2-IRES-Cre;Ai14* mice (**Fig. 2b**). This could be a result of disruption of a regulatory element in transgenesis that is responsible for *Calb2* expression is Sst cells.

Example 3. As previously reported[25,78], we find that *Sst-IRES-Cre* does express in a small number of cells that we classify into a Pvalb type (**Fig. 2b**), although at the protein level, Sst/Pvalb double-positive cells are virtually absent in VISp[24]. The labeling of cells based on *Sst-IRES-Cre* most likely reflects the presence of *Sst* mRNA, and although most of those cells are indeed Sst cell types, some *Sst* mRNA is transcribed, but not translated, in Pvalb types.

Methodological note on single cell classification. *Interneurons.* We detect mRNA (RPKM>0) for at least one of the major GABAergic markers, *Vip*, *Sst* or *Pvalb*, in 99.7% (661/663) of cells that were classified into one of these major types. Although most of them express mRNA for only one of these three genes (411/663, and 410 are classified in accordance with the expression of that marker), a substantial number of cells (250) express more than one, as previously observed[25,79]. Our classification procedure, which takes into account genome-wide gene expression, usually classifies these double-expressing cells into the major type that corresponds

to the highest expressed major marker in that cell (64/65 for Vip, 92/104 for Sst, 81/81 for Pvalb).

Non-neuronal cells. We identify astrocytes based on expression of previously reported markers *Aqp4*, *F3*, and *Gfap12*. Our Oligo-96*Rik type corresponds to previously described newly generated oligodendrocytes based on the unique expression of *Enpp6* and *9630013A20Rik* (abbreviated as *96*Rik*), while our Oligo-Opalin type corresponds to myelinating oligodendrocytes13. Oligo precursor cells (OPC) express *Pdgfra* and *Cspg4* as previously reported[12,13]. Accordingly, microglial cells express *Itgam*, *Cx3cr1* and *C1qb13*. We identify endothelial cells based on expression of *Flt1* (ref. 13), and smooth muscle cells (SMC) based on the expression of *Bgn*[80].

Cells with unexpected combinations of markers. We note three cells that, although they passed our QC criteria (Online Methods and **Supplementary Fig. 3**), have unexpected expression of marker genes. Cell H1122_VU is classified as an intermediate with primary type L6a-Sla and secondary type Pvalb-Gpx3 (**Supplementary Table 3**), and it is the only cell that is an intermediate between a GABAergic and a glutamatergic type. This cell does not express the pan-excitatory marker *Slc17a7*, any of the L6a markers (*Foxp2*, *Crym*), pan-inhibitory markers (*Gad1*, *Gad2*), nor the marker for its classified secondary type, *Pvalb*. Another cell to note is A1612_V, which is classified as an intermediate with primary type L5a-Batf3 and secondary type L6a-Sla. This cell also does not express the pan-excitatory marker Slc17a7, L5a markers (*Deptor*, *Rorb*), nor L6a markers (*Crym*, *Foxp2*). Finally, we also note cell G1766_V, which was isolated from the *Pvalb-2A-Flpo;Gad2-IRES-Cre;Ai65D* line. It is classified into L5b-Tph2 type (**Supplementary Table 3**), and it expresses a combination of markers from many types: the pan-excitatory marker *Slc17a7*, L5a markers (*Deptor*, *Rorb*), L5b markers (*Bcl6*, *Qrfpr*), astrocyte markers (*Gja1*, *F3*), as well as pan-inhibitory markers (*Gad1*, *Gad2*), and *Pvalb*.

Statistical analyses and methodology. *Blinding.* Data collection and analysis were not performed blind to the conditions of the experiments. The authors were not blind to the Cre lines used for cell collection, and no randomization was used to assign experimental groups.

Sample sizes. The sample sizes are similar to or higher than those generally employed in the field.

Parametric tests. To estimate significant differences in the numbers of genes detected among broad cell classes (**Supplementary Fig. 13c**), we used t-tests because the distributions are approximately normal. We did not compare variance estimates between groups, although variances are represented graphically in the figures. As a result, we used the heteroscedastic assumption in the calculation of the *P* value when performing parametric tests. The tests were two-sided.

Non-parametric tests. For all remaining comparisons, we used the appropriate nonparametric test in order to avoid making assumptions of distribution normality. We did not explicitly test whether the distributions (and hence variances) were identical, and thus the *P* values indicate stochastic dominance. The tests were two-sided.

Hypergeometric tests for layer enrichment. For evaluating statistically significant enrichment in upper or lower cortical layers for GABAergic cell types (**Supplementary Table 5**), we calculated the cumulative hypergeometric probability of sampling M or fewer cells of a given type from the upper layer of a Cre line, given N total upper layer and P total lower layer cells from that Cre line, and T cells total from that Cre line belonging to the cell type of interest.

In other words, this is the probability of getting M or fewer red balls in T draws from an urn containing N red balls and P non-red balls. For cases where the given cell type contained both upper and lower layer-derived cells, the selection criterion was cumulative hypergeometric probability (hypergeometric *P* value in Supplementary Table 5) > 0.975 for enrichment. For corner cases where the given cell type contained only upper or only lower layer-derived cells, the selection criterion was cumulative hypergeometric probability < 0.025 for the non-enriched case. This criterion is required because the cumulative hypergeometric probability for having T or fewer successes in T draws is, by definition, equal to 1, so the criterion described above is not informative for significance. Finally, we also considered the corner case where the sampling is too sparse to ever obtain a *P* value less than $P < 0.025$ for either of the extreme cases (all upper layer or all lower layer cells). These cases are marked in **Supplementary Table 5** as having "too few cells for significance". Note there are no degrees of freedom associated with the hypergeometric test because it is an exact test. The tests were two-sided.

Other tests. For the differential gene expression tests, we used the DESeq and DESeq2 packages, both of which derive estimates for the underlying distributions (in the form of negative binomial distribution) for the read counts.

Corrections for multiple comparisons. We used Benjamini-Hochberg correction for FDRs and Bonferroni correction for *P* value-based tests.

Plotting. The ggplot2 package[81] for R was used to generate violin plots, dot plots, bar plots, heat maps, and jittered point plots throughout the figures.

A **Supplementary Methods Checklist** is available.

References:

1. Lein, E.S., *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168-176 (2007).
2. Hawrylycz, M.J., *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391-399 (2012).
3. Harris, K.D. & Shepherd, G.M. The neocortical circuit: themes and variations. *Nature neuroscience* 18, 170-181 (2015).
4. DeFelipe, J., *et al.* New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nature reviews. Neuroscience* 14, 202-216 (2013).
5. Sugino, K., *et al.* Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature neuroscience* 9, 99-107 (2006).
6. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Developmental Neurobiology* 71, 45-61 (2011).
7. Sorensen, S.A., *et al.* Correlated Gene Expression and Target Specificity Demonstrate Excitatory Projection Neuron Diversity. *Cereb Cortex* (2013).
8. Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H. & Macklis, J.D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nature reviews. Neuroscience* 14, 755-769 (2013).
9. Toledo-Rodriguez, M., *et al.* Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. *Cereb Cortex* 14, 1310-1327 (2004).
10. Ascoli, G.A., *et al.* Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature reviews. Neuroscience* 9, 557-568 (2008).
11. Belgard, T.G., *et al.* A transcriptomic atlas of mouse neocortical layers. *Neuron* 71, 605-616 (2011).
12. Cahoy, J.D., *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28, 264-278 (2008).
13. Zhang, Y., *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34, 11929-11947 (2014).
14. Pollen, A.A., *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature biotechnology* 32, 1053-1058 (2014).
15. Usoskin, D., *et al.* Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nature neuroscience* (2014).
16. Zeisel, A., *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138-1142 (2015).
17. Macosko, E.Z., *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202-1214 (2015).
18. Glickfeld, L.L., Reid, R.C. & Andermann, M.L. A mouse model of higher visual cortical function. *Current opinion in neurobiology* 24, 28-33 (2014).
19. Harris, J.A., *et al.* Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Frontiers in neural circuits* 8, 76 (2014).
20. Taniguchi, H., *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* 71, 995-1013 (2011).
21. Olsen, S.R., Bortone, D.S., Adesnik, H. & Scanziani, M. Gain control by layer six in cortical circuits of vision. *Nature* 483, 47-52 (2012).
22. Huang, Z.J. Toward a genetic dissection of cortical circuits in the mouse. *Neuron* 83, 1284-1302 (2014).
23. Gonchar, Y., Wang, Q. & Burkhalter, A.H. Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining. *Frontiers in neuroanatomy* 2 (2008).
24. Xu, X., Roby, K.D. & Callaway, E.M. Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *The Journal of comparative neurology* 518, 389-404 (2010).
25. Pfeffer, C.K., Xue, M., He, M., Huang, Z.J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience* 16, 1068-1076 (2013).
26. Xu, X., Roby, K.D. & Callaway, E.M. Mouse cortical inhibitory neuron type that coexpresses somatostatin and calretinin. *The Journal of comparative neurology* 499, 144-160 (2006).

27. Oliva, A.A., Jr., Jiang, M., Lam, T., Smith, K.L. & Swann, J.W. Novel hippocampal interneuronal subtypes identified using transgenic mice that express green fluorescent protein in GABAergic interneurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20, 3354-3368 (2000).
28. Seress, L., Abraham, H., Hajnal, A., Lin, H. & Totterdell, S. NOS-positive local circuit neurons are exclusively axo-dendritic cells both in the neo- and archi-cortex of the rat brain. *Brain research* 1056, 183-190 (2005).
29. Lee, J.E. & Jeon, C.J. Immunocytochemical localization of nitric oxide synthase-containing neurons in mouse and rabbit visual cortex and co-localization with calcium-binding proteins. *Molecules and cells* 19, 408-417 (2005).
30. Tomioka, R., *et al.* Demonstration of long-range GABAergic connections distributed throughout the mouse neocortex. *The European journal of neuroscience* 21, 1587-1600 (2005).
31. Gerashchenko, D., *et al.* Identification of a population of sleep-active cerebral cortex neurons. *Proceedings of the National Academy of Sciences of the United States of America* 105, 10227-10232 (2008).
32. Taniguchi, H., Lu, J. & Huang, Z.J. The spatial and temporal origin of chandelier cells in mouse neocortex. *Science* 339, 70-74 (2013).
33. Dehorter, N., *et al.* Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch. *Science* 349, 1216-1220 (2015).
34. von Engelhardt, J., Eliava, M., Meyer, A.H., Rozov, A. & Monyer, H. Functional characterization of intrinsic cholinergic interneurons in the cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27, 5633-5642 (2007).
35. Molyneaux, B.J., Arlotta, P., Menezes, J.R. & Macklis, J.D. Neuronal subtype specification in the cerebral cortex. *Nature reviews. Neuroscience* 8, 427-437 (2007).
36. Zeng, H., *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* 149, 483-496 (2012).
37. Sommer, B., *et al.* Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science* 249, 1580-1585 (1990).
38. Velez-Fort, M., *et al.* The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing. *Neuron* 83, 1431-1443 (2014).
39. Bortone, D.S., Olsen, S.R. & Scanziani, M. Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron* 82, 474-485 (2014).
40. Kawaguchi, Y. Physiological subgroups of nonpyramidal cells with specific morphological characteristics in layer II/III of rat frontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 15, 2638-2655 (1995).
41. Hestrin, S. & Armstrong, W.E. Morphology and physiology of cortical neurons in layer I. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 16, 5290-5300 (1996).
42. Povysheva, N.V., *et al.* Electrophysiological differences between neurogliaform cells from monkey and rat prefrontal cortex. *Journal of neurophysiology* 97, 1030-1039 (2007).
43. Chu, Z., Galarreta, M. & Hestrin, S. Synaptic interactions of late-spiking neocortical neurons in layer 1. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23, 96-102 (2003).
44. Simon, A., Olah, S., Molnar, G., Szabadics, J. & Tamas, G. Gap-junctional coupling between neurogliaform cells and various interneuron types in the neocortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25, 6278-6285 (2005).
45. Karayannis, T., *et al.* Slow GABA transient and receptor desensitization shape synaptic responses evoked by hippocampal neurogliaform cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30, 9898-9909 (2010).
46. Kawaguchi, Y. & Kubota, Y. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb Cortex* 7, 476-486 (1997).
47. Muralidhar, S., Wang, Y. & Markram, H. Synaptic and cellular organization of layer 1 of the developing rat somatosensory cortex. *Frontiers in neuroanatomy* 7, 52 (2013).
48. Herculano-Houzel, S., Watson, C. & Paxinos, G. Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones. *Frontiers in neuroanatomy* 7, 35 (2013).
49. DeFelipe, J. Cortical interneurons: from Cajal to 2001. *Progress in brain research* 136, 215-238 (2002).
50. Jaitin, D.A., *et al.* Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science* 343, 776-779 (2014).

51. Raymond, C.S. & Soriano, P. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PloS one* 2, e162 (2007).
52. Rossi, J., *et al.* Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell metabolism* 13, 195-204 (2011).
53. Gerfen, C.R., Paletzki, R. & Heintz, N. GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* 80, 1368-1383 (2013).
54. Franco, S.J., *et al.* Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science* 337, 746-749 (2012).
55. Dhillon, H., *et al.* Leptin directly activates SF1 neurons in the VMH, and this action by leptin is required for normal body-weight homeostasis. *Neuron* 49, 191-203 (2006).
56. Madisen, L., *et al.* Transgenic Mice for Intersectional Targeting of Neural Sensors and Effectors with High Specificity and Performance. *Neuron* 85, 942-958 (2015).
57. Hippenmeyer, S., *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS biology* 3, e159 (2005).
58. Madisen, L., *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature neuroscience* 13, 133-140 (2010).
59. Vong, L., *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* 71, 142-154 (2011).
60. Tong, Q., Ye, C.P., Jones, J.E., Elmquist, J.K. & Lowell, B.B. Synaptic release of GABA by AgRP neurons is required for normal regulation of energy balance. *Nature neuroscience* 11, 998-1000 (2008).
61. Sando, R., 3rd, *et al.* Inducible control of gene expression with destabilized Cre. *Nature methods* 10, 1085-1088 (2013).
62. Hnasko, T.S., *et al.* Cre recombinase-mediated restoration of nigrostriatal dopamine in dopamine-deficient mice reverses hypophagia and bradykinesia. *Proceedings of the National Academy of Sciences of the United States of America* 103, 8858-8863 (2006).
63. Harris, J.A., Oh, S.W. & Zeng, H. Adeno-associated viral vectors for anterograde axonal tracing with fluorescent proteins in nontransgenic and cre driver mice. *Current protocols in neuroscience / editorial board, Jacqueline N. Crawley ... [et al.]* Chapter 1, Unit 1 20 21-18 (2012).
64. Franklin, K.B.J.a.P., G. *Mouse brain in stereotaxic coordinates* (Academic Press, 2008).
65. Hempel, C.M., Sugino, K. & Nelson, S.B. A manual method for the purification of fluorescently labeled neurons from the mammalian brain. *Nature protocols* 2, 2924-2929 (2007).
66. Ramskold, D., *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature biotechnology* 30, 777-782 (2012).
67. Shalek, A.K., *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510, 363-369 (2014).
68. Treutlein, B., *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371-375 (2014).
69. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 12, 323 (2011).
70. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 10, R25 (2009).
71. Brennecke, P., *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nature methods* 10, 1093-1095 (2013).
72. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* 11, R106 (2010).
73. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 9, 559 (2008).
74. Ritchie, M.E., *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43, e47 (2015).
75. Katz, Y., Wang, E.T., Airoldi, E.M. & Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature methods* 7, 1009-1015 (2010).
76. Thompson, C.L., *et al.* Genomic anatomy of the hippocampus. *Neuron* 60, 1010-1021 (2008).
77. Peng, H., Ruan, Z., Long, F., Simpson, J.H. & Myers, E.W. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nature biotechnology* 28, 348-353 (2010).

78. Hu, H., Cavendish, J.Z. & Agmon, A. Not all that glitters is gold: off-target recombination in the somatostatin-IRES-Cre mouse line labels a subset of fast-spiking interneurons. *Frontiers in neural circuits* 7, 195 (2013).
79. Rossier, J., *et al.* Cortical fast-spiking parvalbumin interneurons enwrapped in the perineuronal net express the metallopeptidases Adamts8, Adamts15 and Neprilysin. *Molecular psychiatry* 20, 154-161 (2015).
80. Nikkari, S.T., Jarvelainen, H.T., Wight, T.N., Ferguson, M. & Clowes, A.W. Smooth muscle cell expression of extracellular matrix genes after arterial injury. *The American journal of pathology* 144, 1348-1356 (1994).
81. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer, 2009).

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.

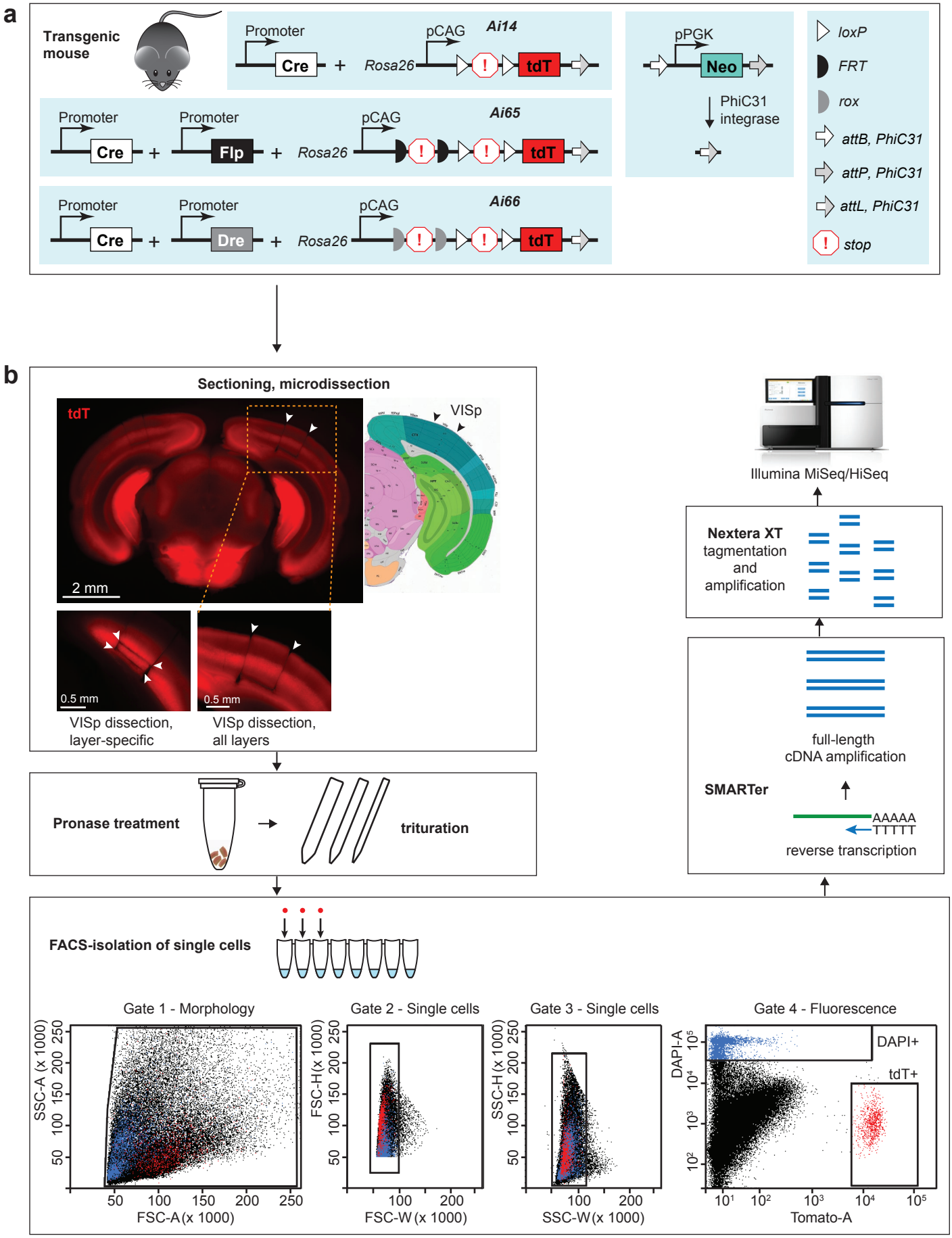
Supplementary Table 11. Genetic background estimate for all animals used in the study. In our experimental animals, the percent of C57BL/6J genetic background ranges from 75% to

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.

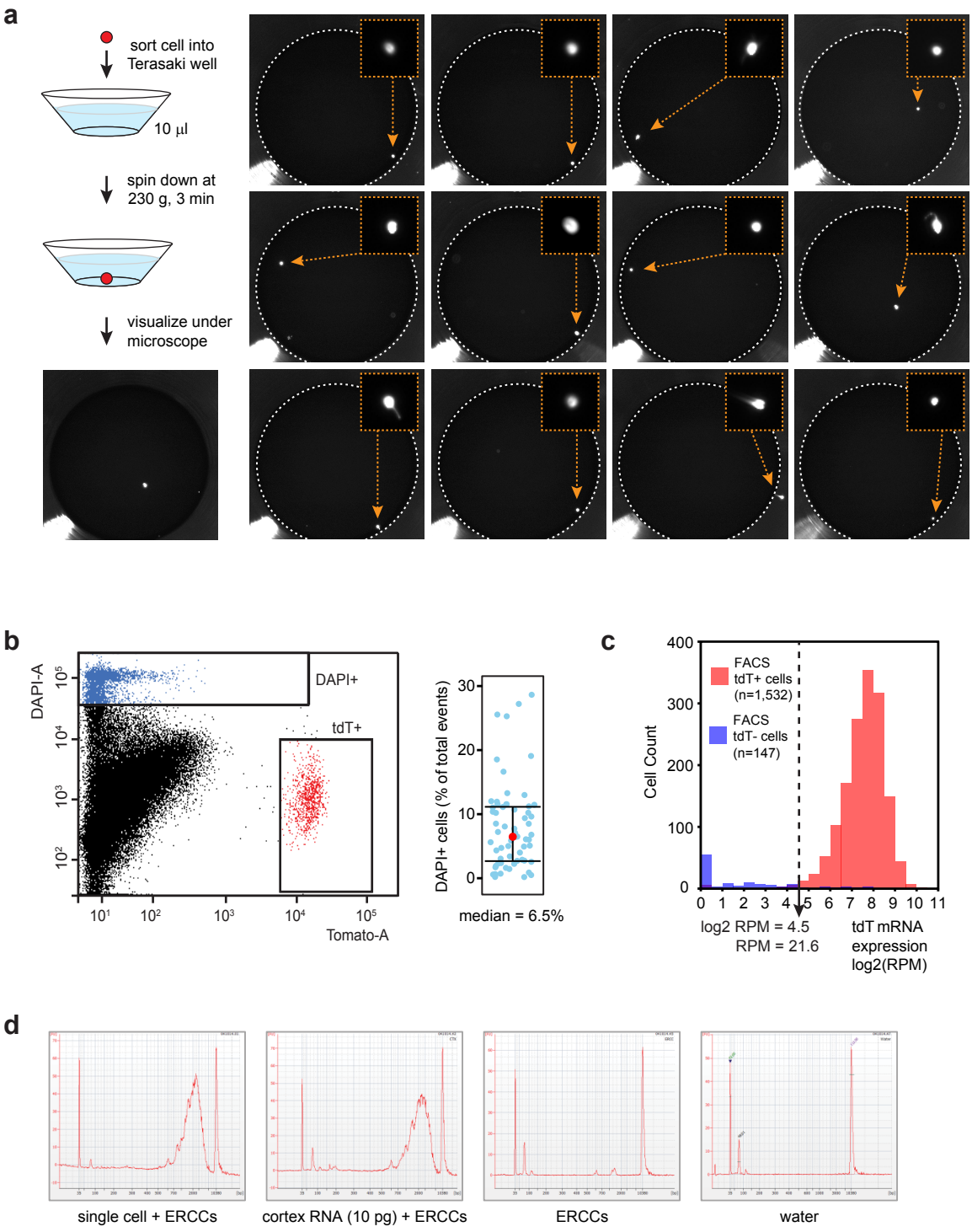
Supplementary Table 12. Consequences of cluster validation parameter change on single cell classification. Cluster identity assignment for each cell is listed for our default parameters (20 genes, $p < 0.01$), and after changes in these parameters: decrease or increase in the number of genes to 10, and 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.

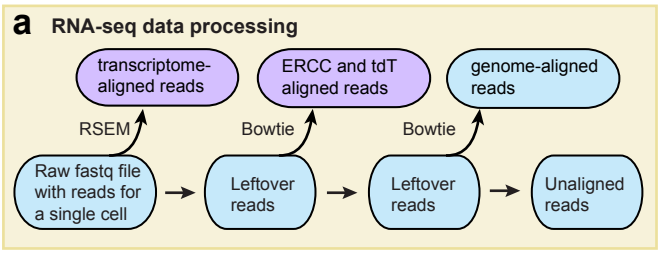
Supplementary Table 14. Quantitative RT-PCR assays.

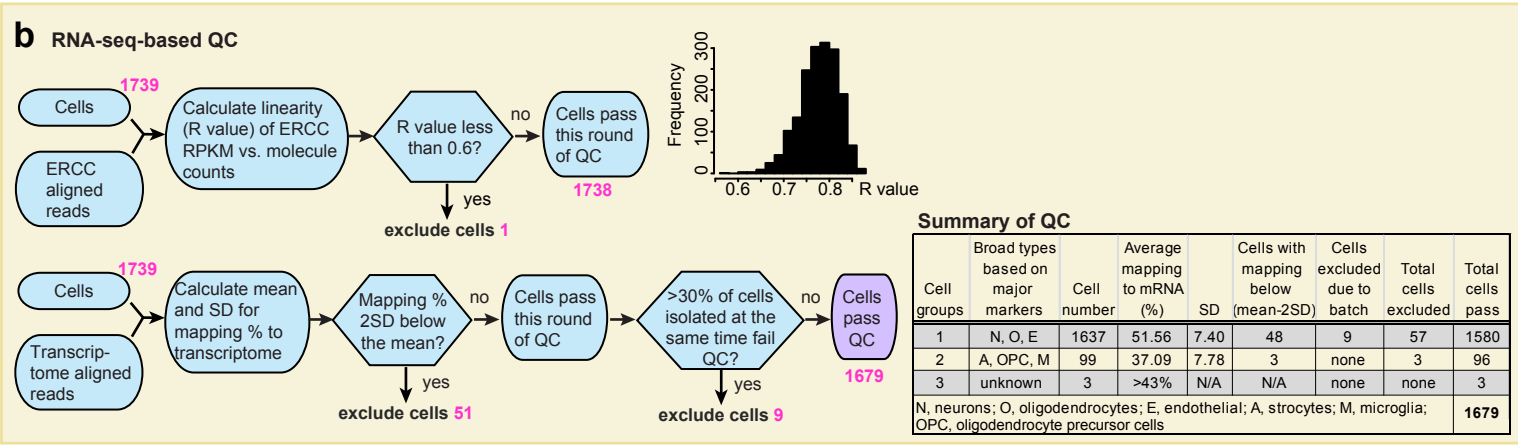

a
Transgenic mouse
Promoter
Cre
+
Rosa26
pCAG
Ai14
!
tdT
Promoter
Cre
+
Promoter
Flp
+
Rosa26
pCAG
Ai65
Promoter
Cre
+
Promoter
Dre
+
Rosa26
pCAG
Ai66
pPGK
Neo
PhiC31 integrase
loxP
FRT
rox
attB, PhiC31
attP, PhiC31
attL, PhiC31
stop
b
Sectioning, microdissection
tdT
VISp
2 mm
0.5 mm
0.5 mm
VISp dissection, layer-specific
VISp dissection, all layers
Pronase treatment
trituration
Illumina MiSeq/HiSeq
Nextera XT tagmentation and amplification
full-length cDNA amplification
SMARTer
AAAAA
TTTTT
reverse transcription
FACS-isolation of single cells
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

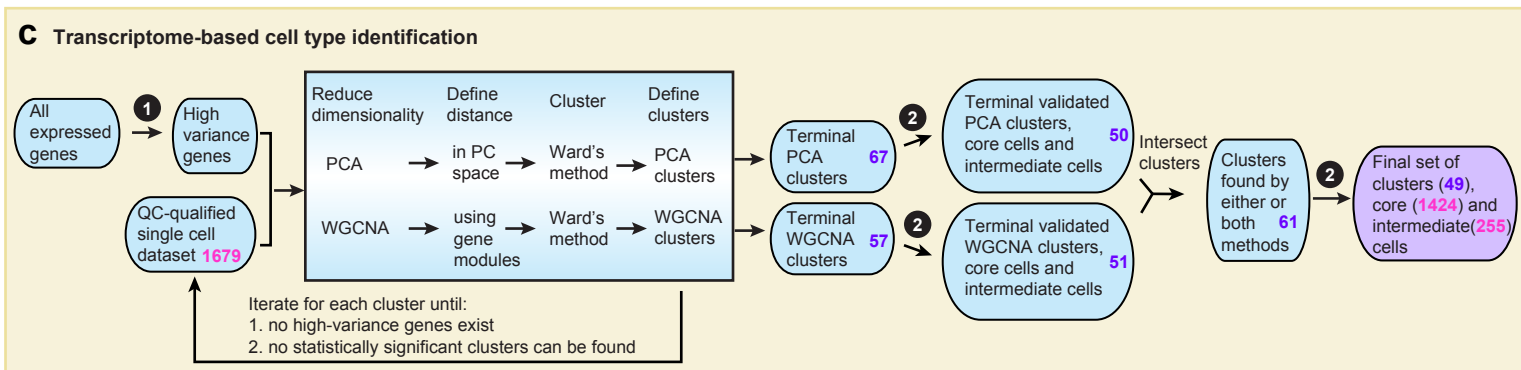
Supplementary Figure 1. Detailed experimental workflow. (**a**) Schematic representation of transgenes used for the lines mentioned in this paper; polyadenylation sites in transgenes are omitted for clarity. Each Cre recombinase line was crossed to *Ai14* or to a second recombinase line (*Flp* or *Dre*) 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.

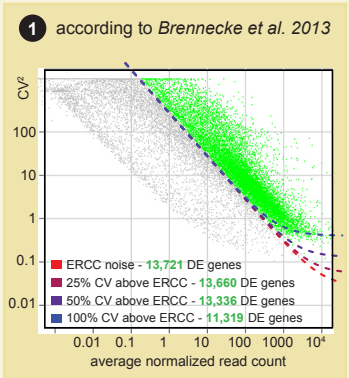

a
sort cell into Terasaki well
10 µl
spin down at 230 g, 3 min
visualize under microscope
b
DAPI-A
DAPI+
tdT+
Tomato-A
DAPI+ cells (% of total events)
median = 6.5%
c
FACS tdT+ cells (n=1,532)
FACS tdT- cells (n=147)
Cell Count
log2 RPM = 4.5
RPM = 21.6
tdT mRNA expression log2(RPM)
d
single cell + ERCCs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

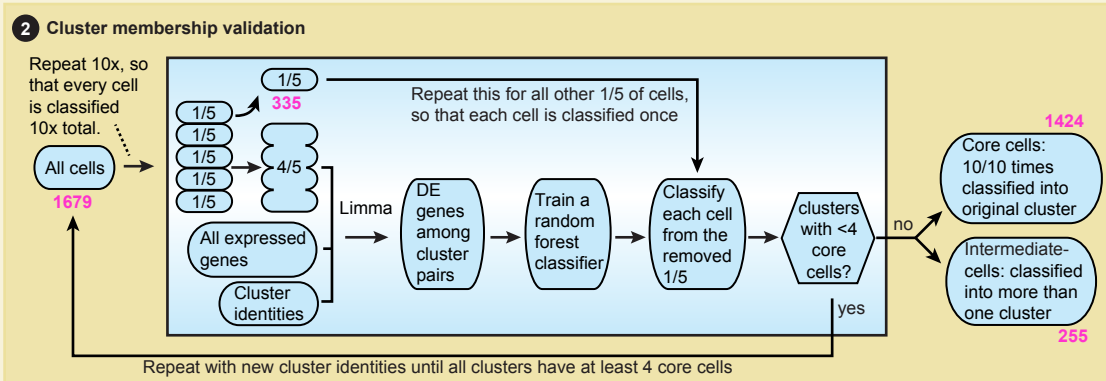
Supplementary Figure 2. Experimental QC. (**a**) A representative control experiment for assessing FACS specificity and efficiency. Before sorting cells into 8-well strips or 96-well plates for transcriptional profiling, the cells were sorted using the same setup into Terasaki plates containing 5 or 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

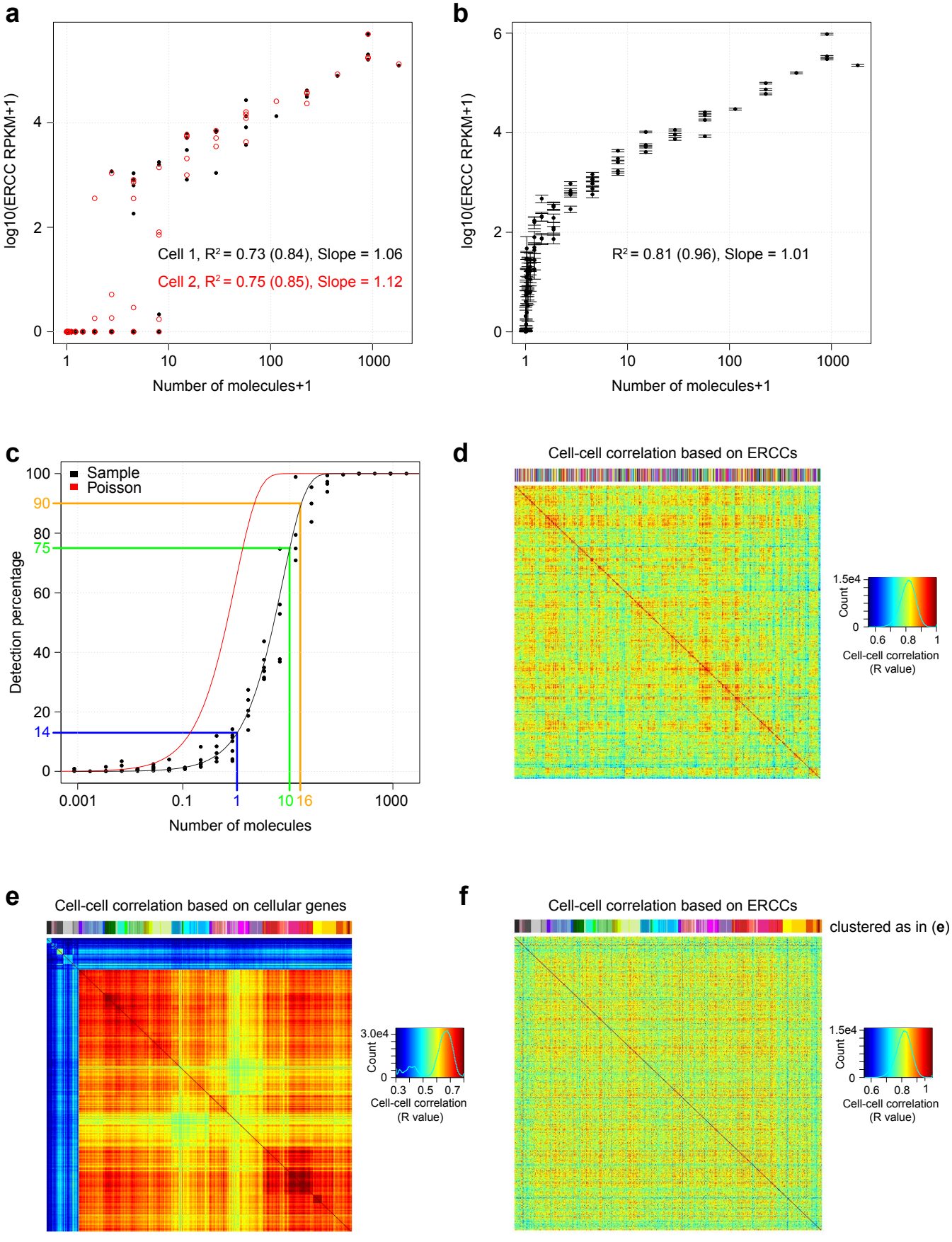

b
RNA-seq-based QC
1739
Cells
ERCC aligned reads
Calculate linearity (R value) of ERCC RPKM vs. molecule counts
R value less than 0.6?
no
Cells pass this round of QC
1738
yes
exclude cells 1
Frequency
0
100
200
300
0.6
0.7
0.8
R value
1739
Cells
Transcrip-tome aligned reads
Calculate mean and SD for 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

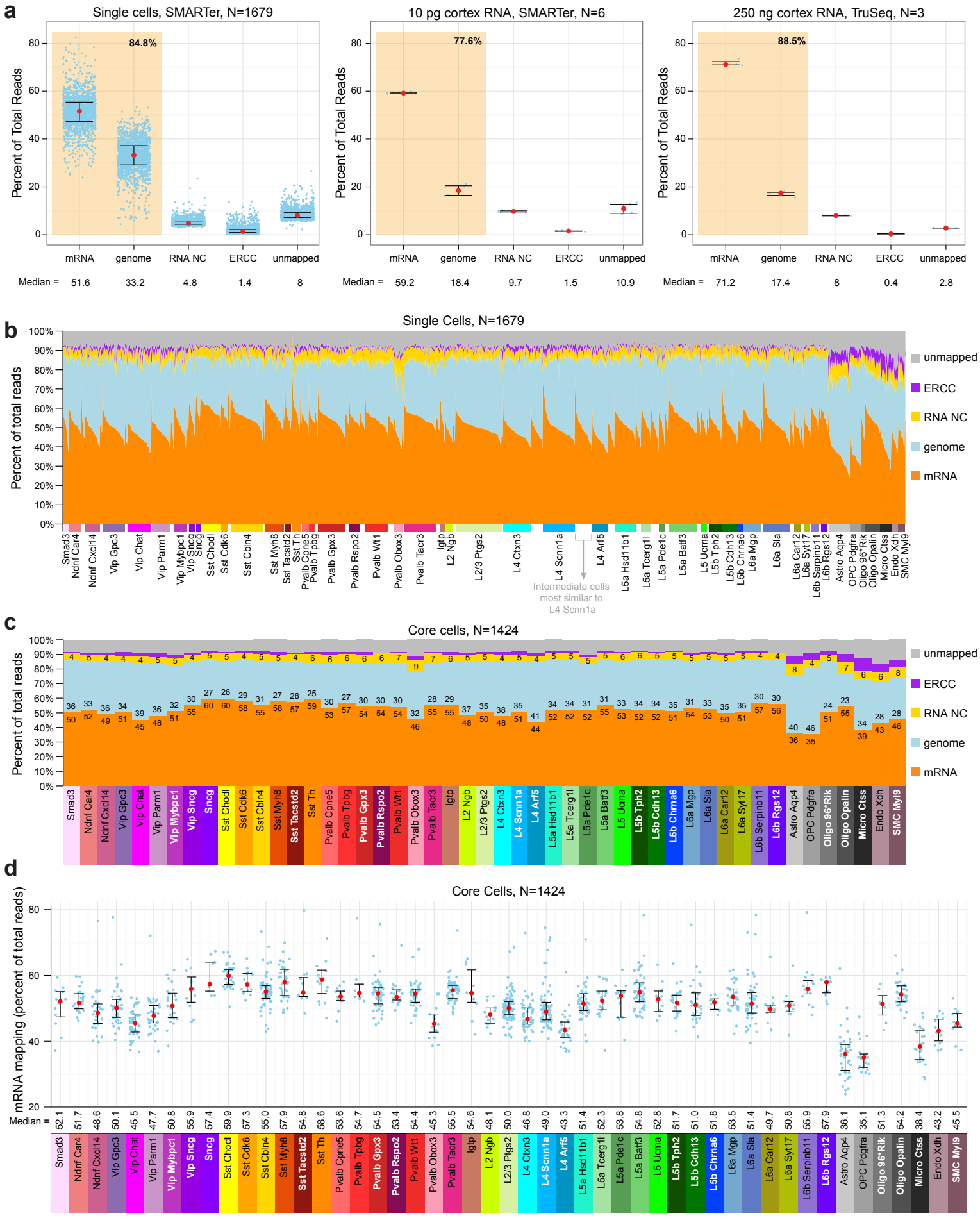

1 according to Brennecke et al. 2013
CV²
100
10
1
0.1
0.01
ERCC noise - 13,721 DE genes
25% CV above ERCC - 13,660 DE genes
50% CV above ERCC - 13,336 DE genes
100% CV above ERCC - 11,319 DE genes
0.01
0.1
1
10
100
1000
10⁴
average normalized read count


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

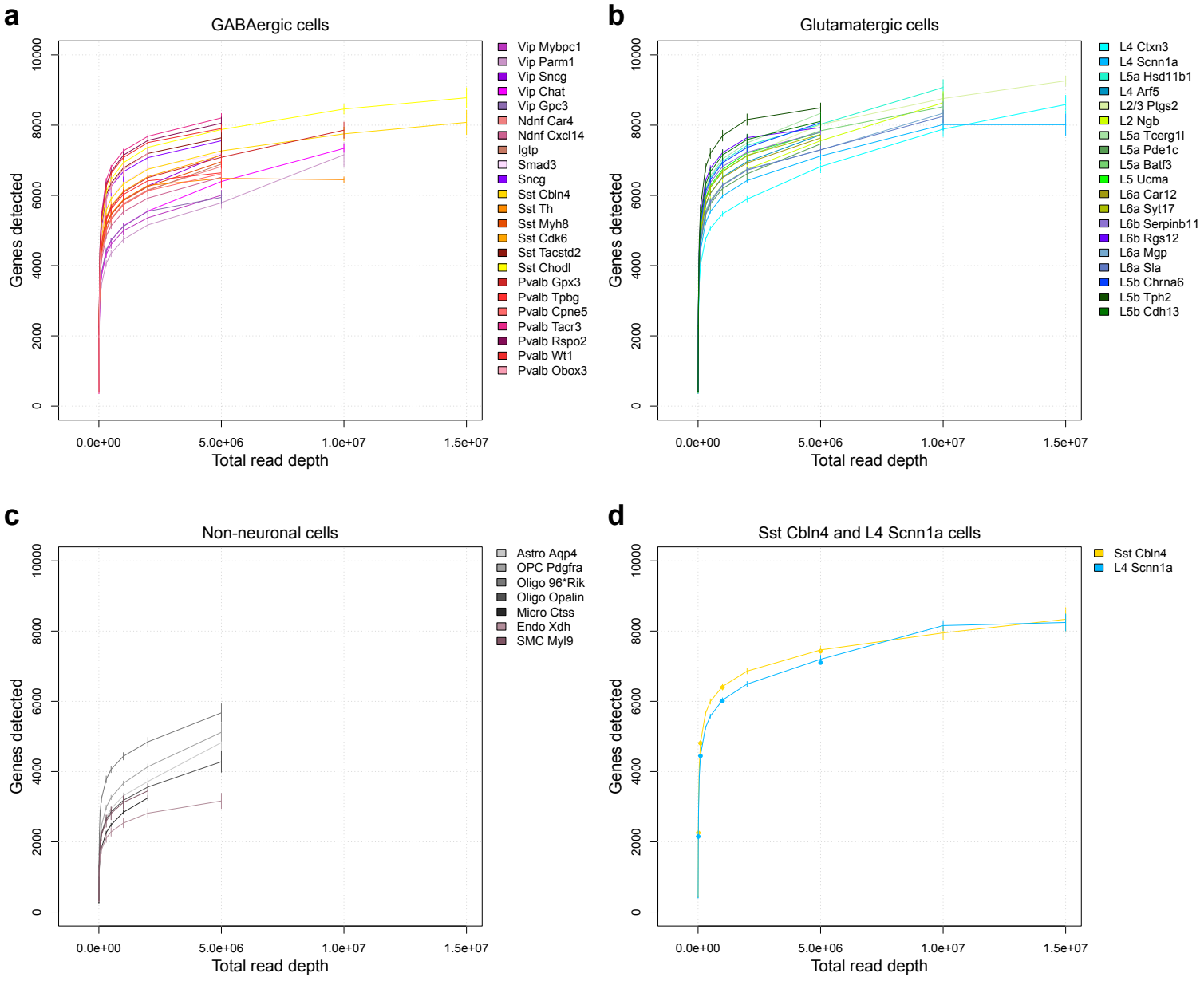
Supplementary Figure 3. Detailed data analysis workflow. (**a**) Workflow diagram for processing of raw sequencing reads to generate counts and RPKM values for genes, as well as total read counts aligning to *ERCC* RNAs, *tdT* mRNA, and genomic regions. (**b**) Quality control steps based on *ERCC* detection 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)

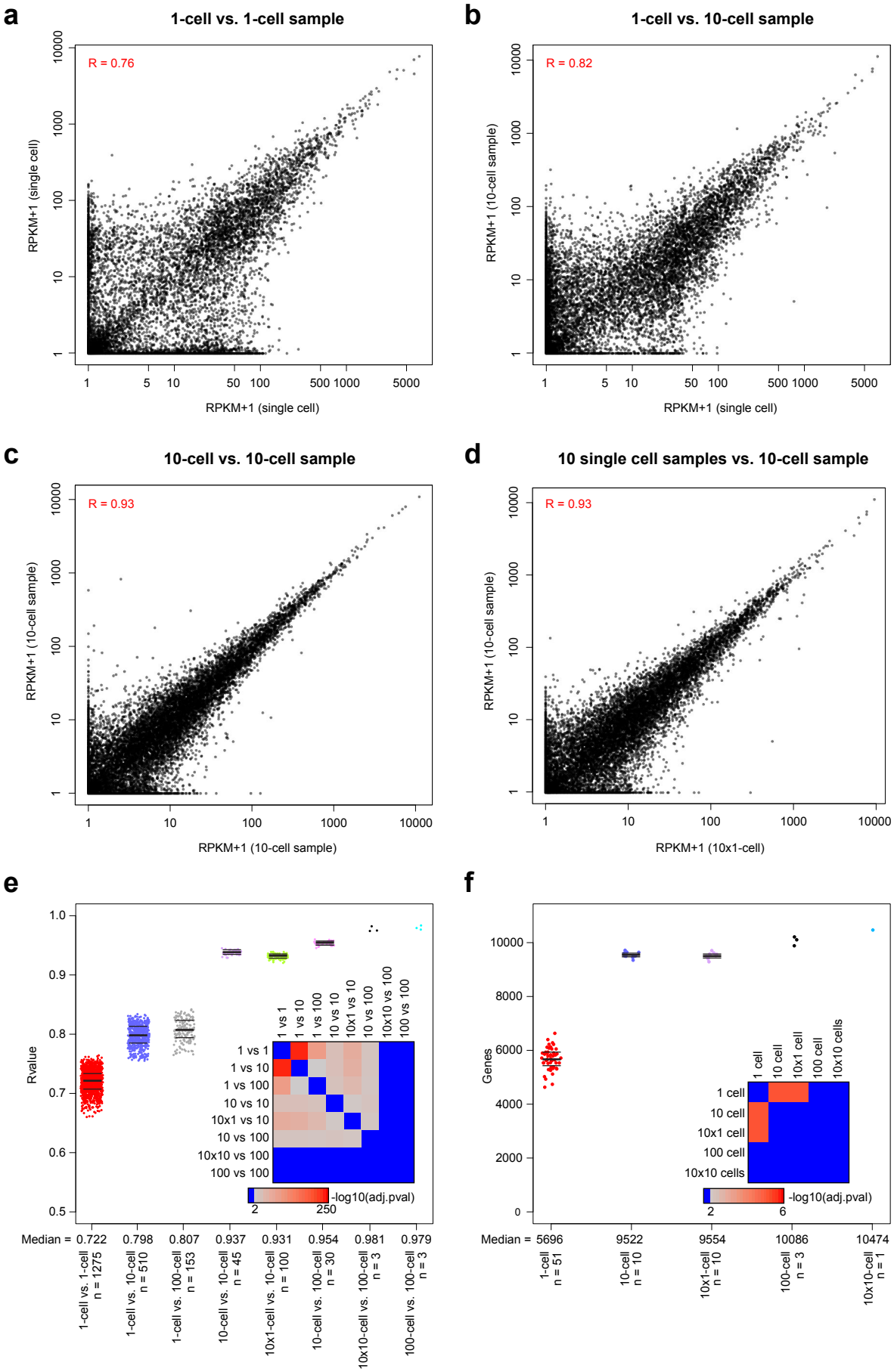
Supplementary Figure 4. QC based on spike-in *ERCC* RNAs. (**a**) Plots of *ERCC* RPKM values (where the RPKM values were calculated with respect to total reads mapped to *ERCC* RNAs only, $N = 92$ species) versus putative number of molecules for two different cells shows good linearity as determined 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 =

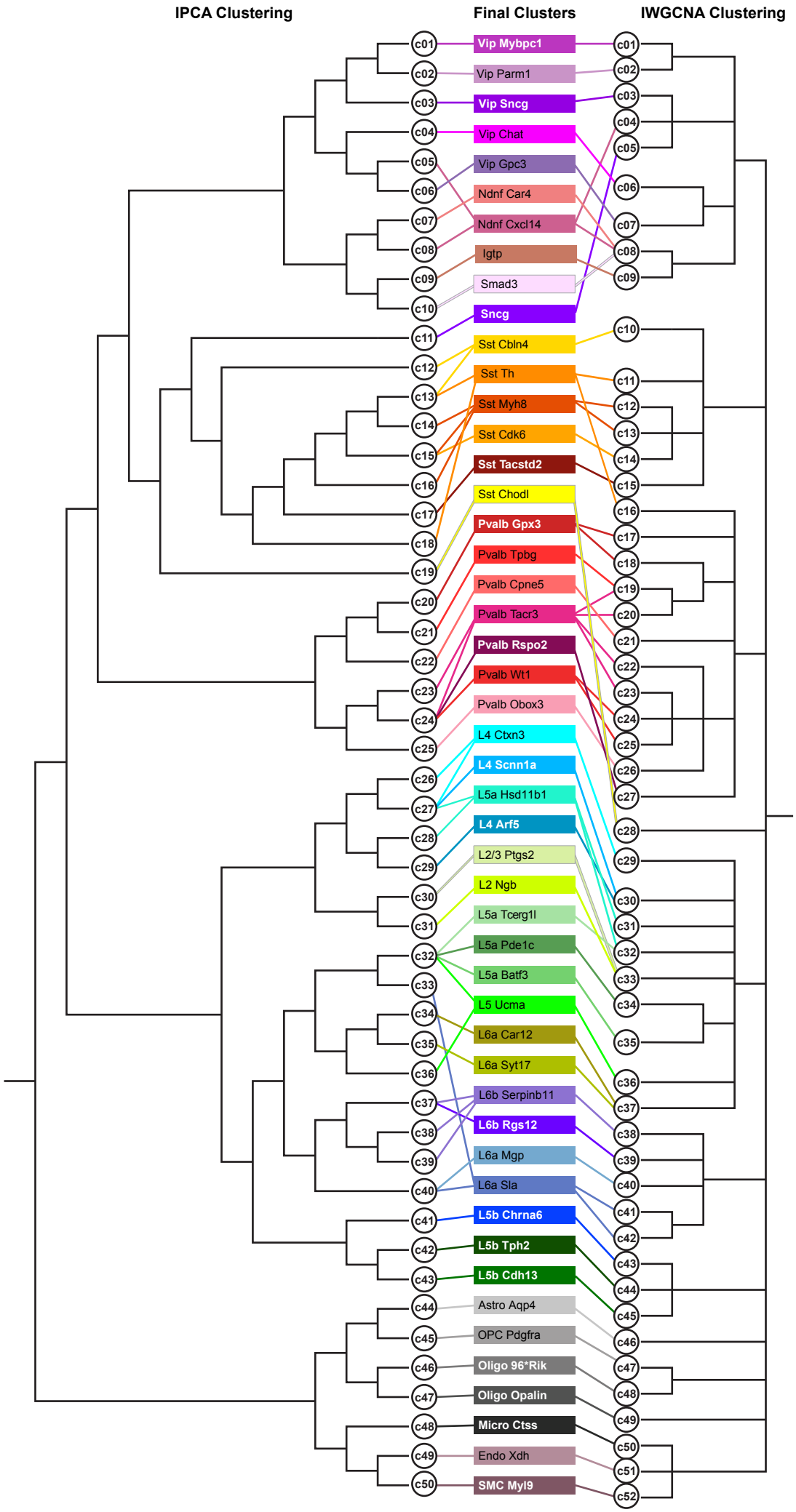
Supplementary Figure 5. Data QC. (**a**) Mapping of transcriptomic data to mRNA (RefSeq mm10 assembly), genome, non-coding RNA (RNA NC) and *ERCC* RNAs for all 1679 single cells (left), 6 replicates of 10 pg total cortex RNA processed like the single cells (middle), and 3 replicates of 250 ng of unamplified cortex RNA prepared by TruSeq (right). 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

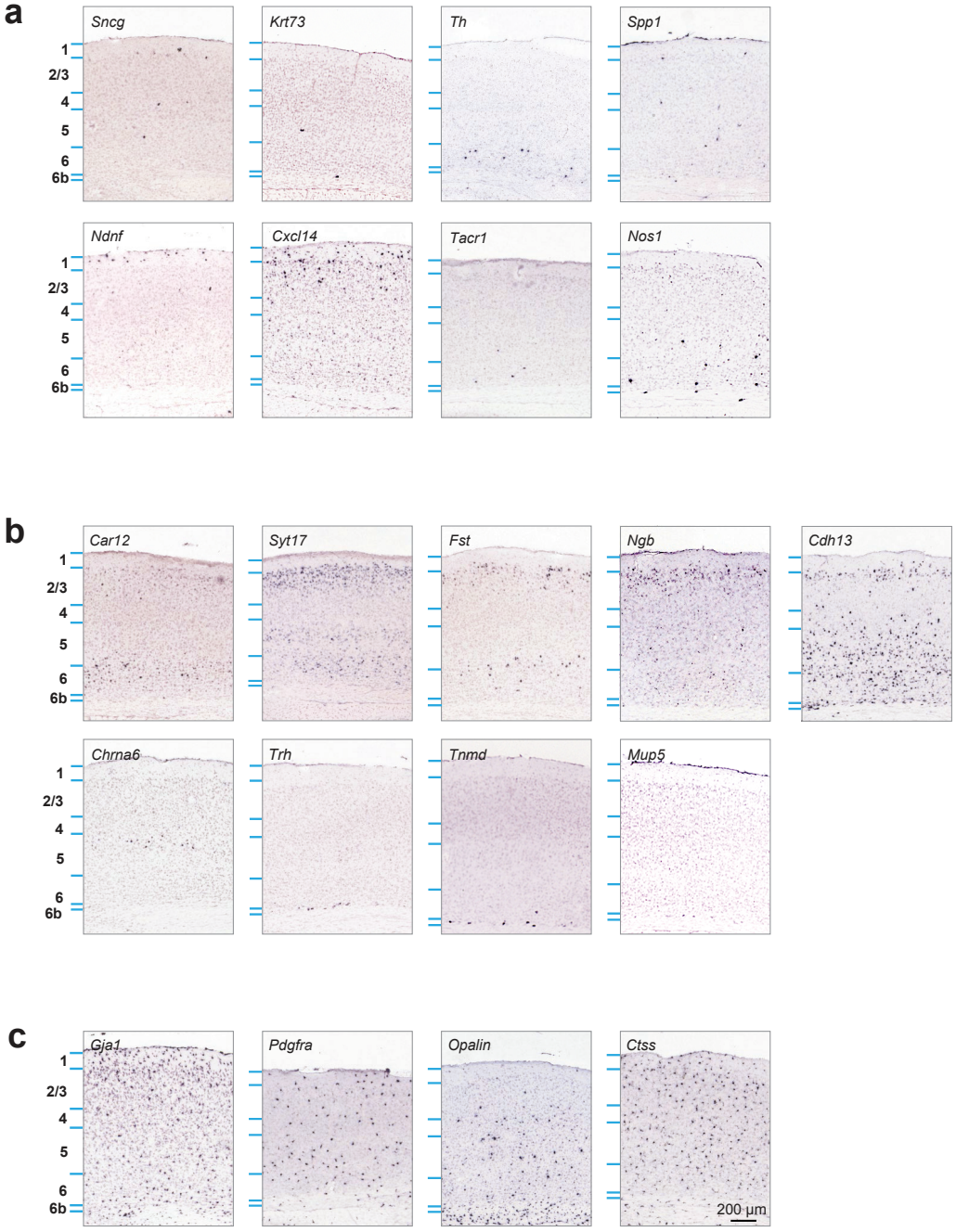
Supplementary Figure 6. Gene detection and sequencing depth. Plots showing the number of genes detected (≥ 1 read) for each of the (**a**) GABAergic (*N* = 23), (**b**) glutamatergic (*N* = 19), and (**c**) non-neuronal (*N* = 7) transcriptomic cell types as a function of post-alignment subsampling to 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

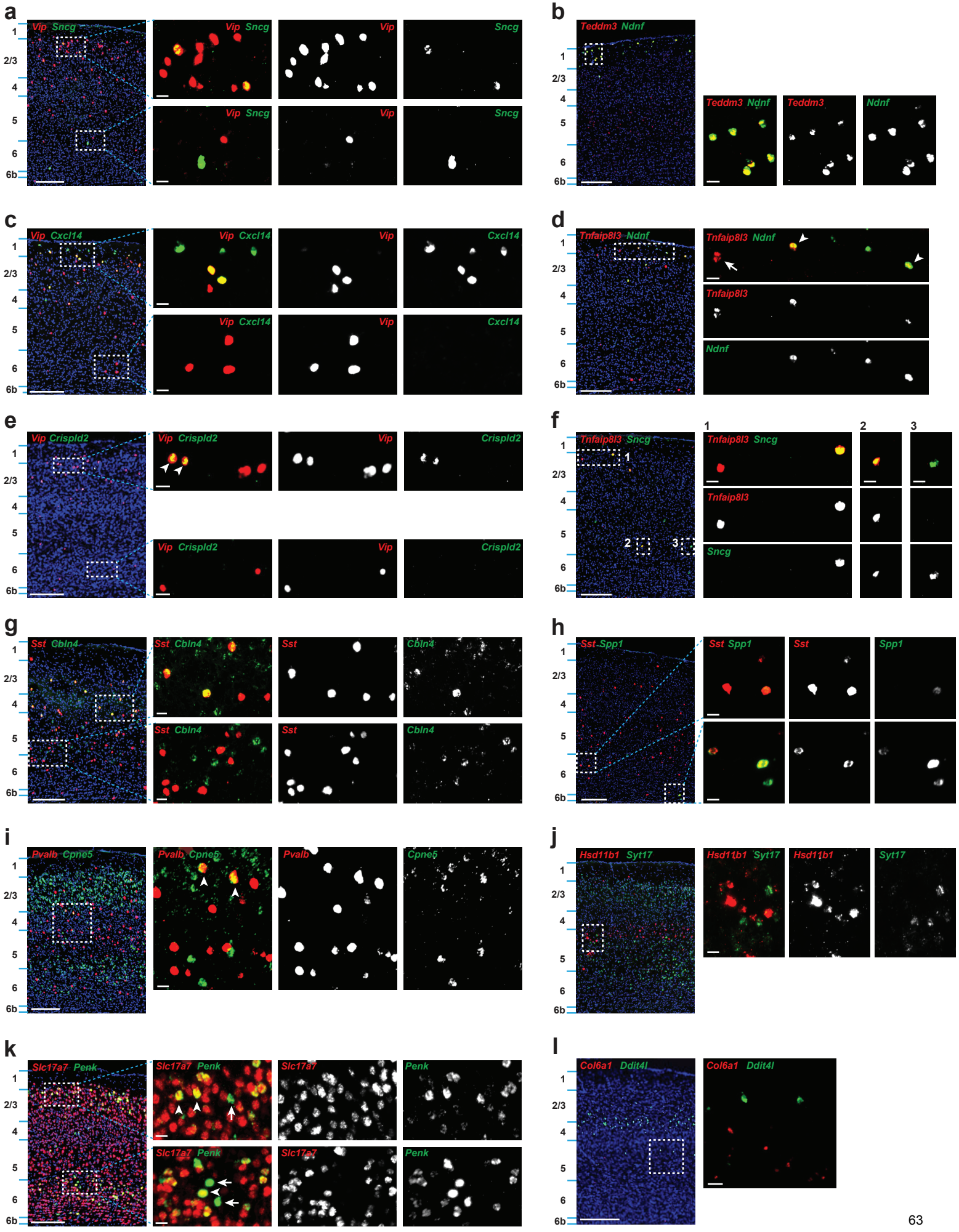
Supplementary Figure 7. Gene detection in single cells and cell populations. Comparison of RNA-seq data generated by the SMARTer/Nextera approach from individual *tdT*$^+$ cells and small populations of *tdT*$^+$ cells isolated from layer 6 of VISp in the *Ntsr1-Cre;Ai14* line. Examples of gene expression 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

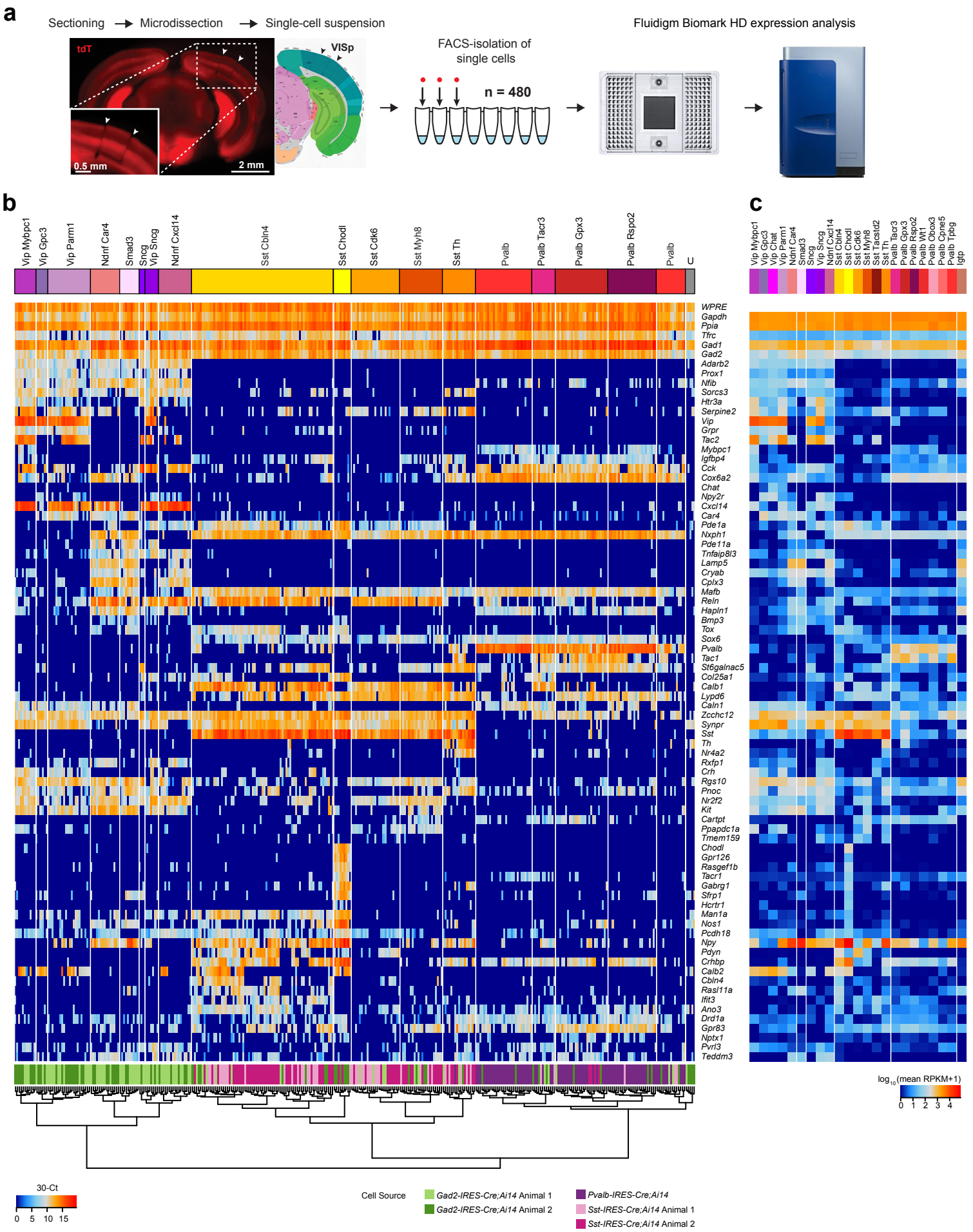
Supplementary Figure 8. Cluster intersection for iterative PCA and iterative WGCNA. Schematic showing the iterative splits leading to the final set of PCA-based clusters (left) and WGCNA-based clusters (right), and their subsequent intersection to generate the final set of clusters described in the paper. 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

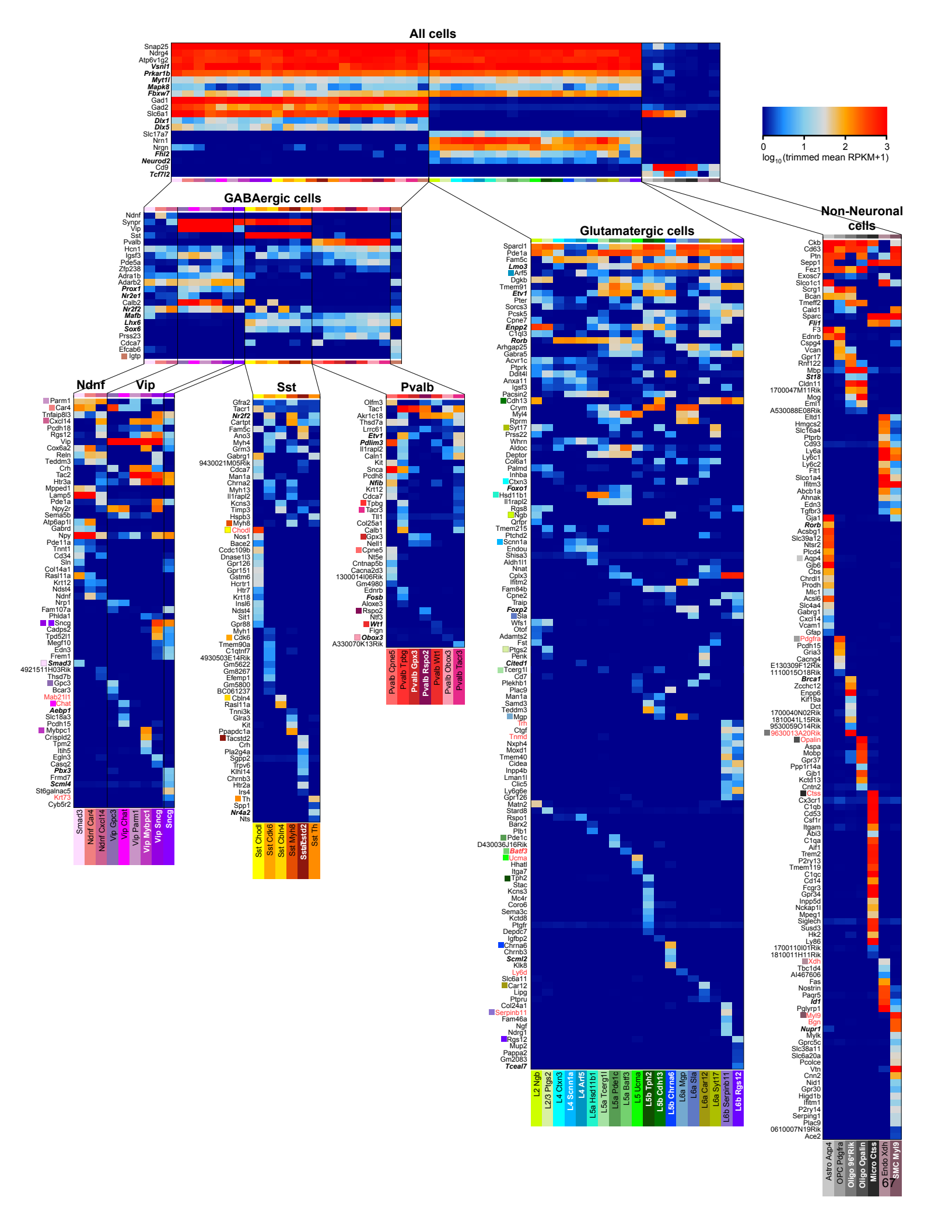
Supplementary Figure 9. Chromogenic RNA *in situ* hybridization confirms select gene expression and confirms/refines spatial positioning of cell types. Images were obtained from the Allen Brain Atlas[1]. Each focuses on VISp, and is part of at least two brain-wide experiments, except for a single experiment 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).


a
Vip Sncg
Vip
Sncg
b
Teddm3 Ndnf
Teddm3
Ndnf
c
Vip Cxcl14
Cxcl14
d
Tnfaip8l3 Ndnf
Tnfaip8l3
e
Vip Crispld2
Crispld2
f
Tnfaip8l3 Sncg
g
Sst Cbln4
Cbln4
h
Sst Spp1
Spp1
i
Pvalb Cpne5
Pvalb
Cpne5
j
Hsd11b1 Syt17
Hsd11b1
Syt17
k
Slc17a7 Penk
Slc17a7
Penk
l
Col6a1 Ddit4l
1
2/3
4
5
6
6b

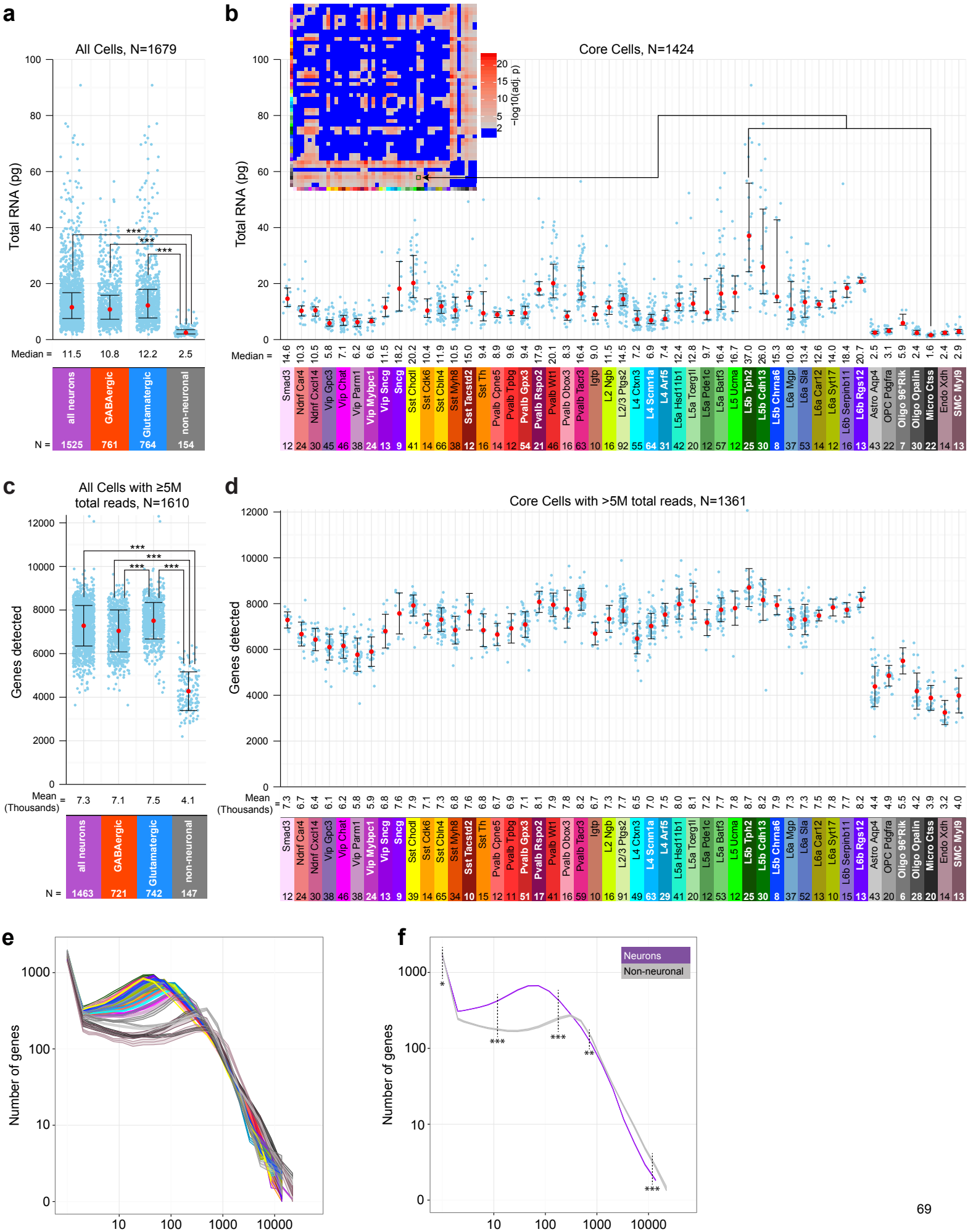
Supplementary Figure 10. Double-label fluorescence RNA *in situ* hybridization (DFISH) confirms coexpression, mutually exclusive expression, and spatially restricted expression of select genes. (**a**) *Sncg* mRNA is expressed sparsely in VISp: in a subset of *Vip*$^+$ cells in upper layers, and independently 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
tdT
0.5 mm
2 mm
VISp
FACS-isolation of single cells
n = 480
Fluidigm Biomark HD expression analysis
b
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
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
c
Vip Mybpc1
Vip Gpc3
Vip Chat
Vip Parm1
Ndnf Car4
Smad3
Sncg
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
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.

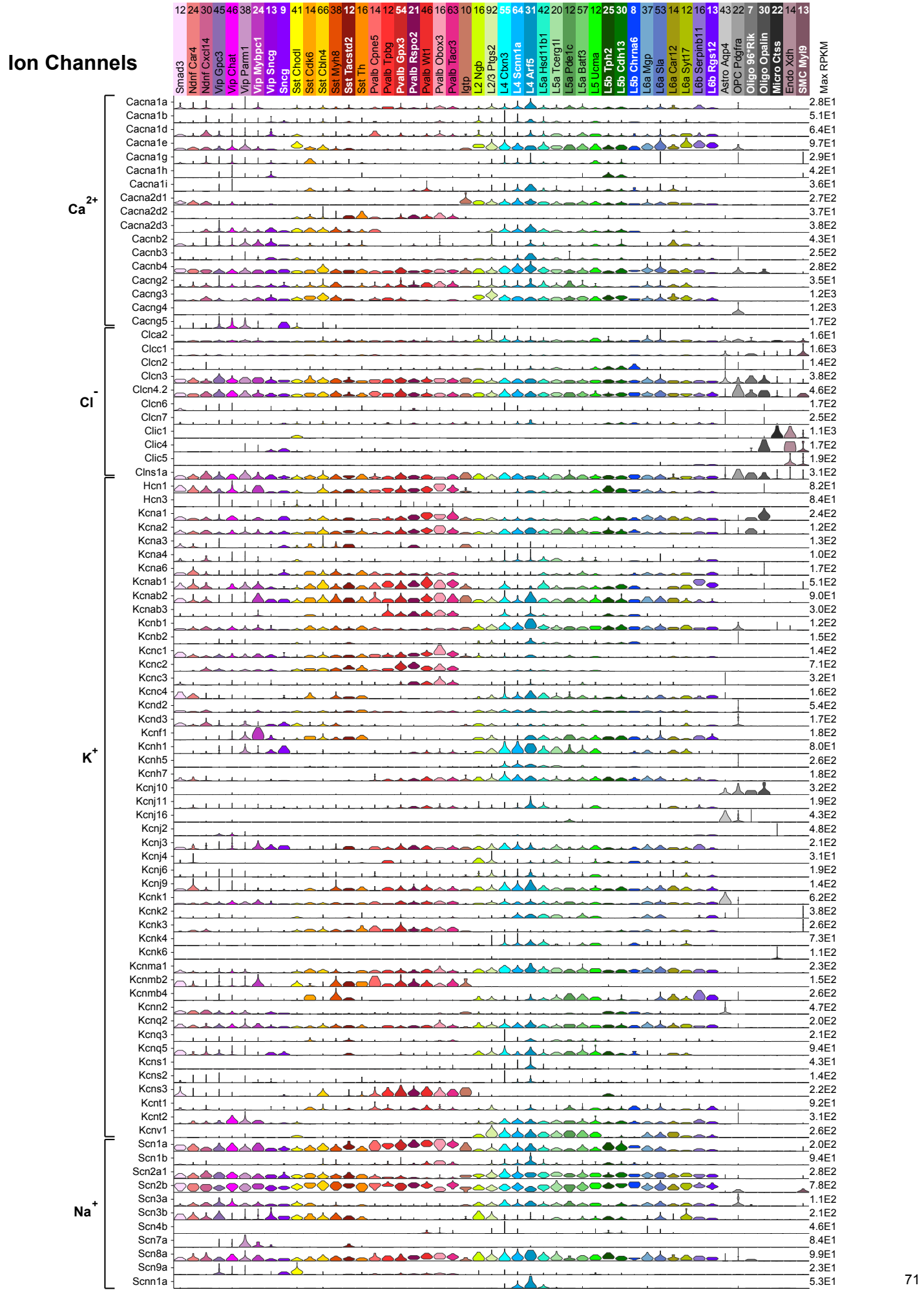

All cells
GABAergic cells
Glutamatergic cells
Non-Neuronal cells
Ndnf
Vip
Sst
Pvalb
0 1 2 3
log10(trimmed mean RPKM+1)
Snap25 Ndrg4 Atp6v1g2 Vsnl1 Prkar1b Myt1l Mapk8 Fbxw7 Gad1 Gad2 Slc6a1 Dlx1 Dlx5 Slc17a7 Nrn1 Nrgn Fli2 Neurod2 Cd9 Tcf7l2
Ndnf Synpr Vip Sst Pvalb Hcn1 Igsf3 Pde5a Zfp238 Adra1b Adarb2 Prox1 Nr2e1 Calb2 Nr2f2 Mafb Lhx6 Sox6 Prss23 Cdca7 Efcab6 Igtp
Parm1 Car4 Tnfaip8l3 Cxcl14 Pcdh18 Rgs12 Vip Cox6a2 Reln Teddm3 Crh Tac2 Htr3a Mpped1 Lamp5 Pde1a Npy2r Sema5b Atp6ap1l Gabrd Npy Pde11a Tnnt1 Cd34 Sln Col14a1 Rasl11a Krt12 Ndst4 Ndnf Nrp1 Fam107a Phlda1 Sncg Cadps2 Tpd52l Megf10 Edn3 Frem1 Smad3 4921511H03Rik Thsd7b Gpc3 Bcar3 Mab21l1 Chat Aebp1 Slc18a3 Pcdh15 Mybpc1 Crispld2 Tpm2 Itih5 Egln3 Casq2 Pbx3 Frmd7 Scml4 St6galnac5 Krt73 Cyb5r2
Smad3 Ndnf Car4 Ndnf Cxcl14 Vip Gpc3 Vip Chat Vip Parm1 Vip Mybpc1 Vip Sncg Sncg
Gfra2 Tacr1 Nr2f2 Cartpt Fam5c Ano3 Myh4 Grm3 Gabrg1 9430021M05Rik Cdca7 Man1a Chrna2 Myh13 Il1rapl2 Kcns3 Timp3 Hspb3 Myh8 Chodl Nos1 Bace2 Ccdc109b Dnase1l3 Gpr126 Gpr151 Gstm6 Hcrtr1 Htr7 Krt18 Insl6 Ndst4 Sit1 Gpr88 Myh1 Cdk6 Tmem90a C1qtnf7 4930503E14Rik Gm5622 Gm8267 Efemp1 Gm5800 BC061237 Cbln4 Rasl11a Tnni3k Glra3 Kit Ppapdc1a Tacstd2 Crh Pla2g4a Sgpp2 Trpv6 Klhl14 Chrnb3 Htr2a Irs4 Th Spp1 Nr4a2 Nts
Sst Chodl Sst Cdk6 Sst Cbln4 Sst Myh8 Sst Tacstd2 Sst Th
Olfm3 Tac1 Akr1c18 Thsd7a Lrrc61 Etv1 Pdlim3 Il1rapl2 Caln1 Kit Snca Pcdh8 Nfib Krt12 Cdca7 Tpbg Tacr3 Tll1 Col25a1 Calb1 Gpx3 Nell1 Cpne5 Nt5e Cntnap5b Cacna2d3 1300014I06Rik Gm4980 Ednrb Fosb Aloxe3 Rspo2 Ntf3 Wt1 Fign Obox3 A330070K13Rik
Pvalb Cpne5 Pvalb Tpbg Pvalb Gpx3 Pvalb Rspo2 Pvalb Wt1 Pvalb Obox3 Pvalb Tacr3
Sparcl1 Pde1a Fam5c Lmo3 Arf5 Dgkb Tmem91 Etv1 Pter Sorcs3 Pcsk5 Cpne7 Enpp2 C1ql3 Rorb Arhgap25 Gabra5 Acvr1c Ptprk Ddit4l Anxa11 Igsf3 Pacsin2 Cdh13 Crym Myl4 Rprm Syt17 Prss22 Whrn Aldoc Deptor Col6a1 Palmd Inhba Ctxn3 Foxo1 Hsd11b1 Il1rapl2 Rgs8 Ngb Qrfpr Tmem215 Ptchd2 Scnn1a Endou Shisa3 Aldh1l1 Nnat Cplx3 Ifitm2 Fam84b Cpne2 Traip Foxp2 Sla Wfs1 Otof Adamts2 Fst Ptgs2 Penk Cited1 Tceerg1l Cd7 Plekhb1 Plac9 Man1a Samd3 Teddm3 Mgp Trh Ctgf Tnmd Nxph4 Moxd1 Tmem40 Cidea Inpp4b Lman1l Clic5 Ly6g6e Gpr126 Matn2 Stard8 Rspo1 Barx2 Plb1 Pde1c D430036J16Rik Batf3 Ucma Hhatl Itga7 Tph2 Stac Kcns3 Mc4r Coro6 Sema3c Kctd8 Ptgfr Depdc7 Igfbp2 Chrna6 Chrnb3 Scml2 Klk8 Ly6d Slc6a11 Car12 Lipg Ptpru Col24a1 Serpinb11 Fam46a Ngf Ndrg1 Rgs12 Mup2 Pappa2 Gm2083 Tceal7
L2 Ngb L2/3 Ptgs2 L4 Ctxn3 L4 Scnn1a L4 Arf5 L5a Hsd11b1 L5a Tceerg1l L5a Pde1c L5a Batf3 L5 Ucma L5b Tph2 L5b Cdh13 L5b Chrna6 L6a Mgp L6a Sla L6a Car12 L6a Syt17 L6b Serpinb11 L6b Rgs12
Ckb Cd63 Ptn Sepp1 Fez1 Exosc7 Slco1c1 Scrg1 Bcan Tmeff2 Cald1 Sparc Fli1 F3 Ednrb Cspg4 Vcan Gpr17 Rnf122 Mbp St18 Cldn11 1700047M11Rik Mog Eml1 A530088E08Rik Eltd1 Hmgcs2 Slc16a4 Ptprb Cd93 Ly6a Ly6c1 Ly6c2 Flt1 Slco1a4 Ifitm3 Abcb1a Ahnak Edn3 Tgfbr3 Gja1 Rorb Acsbg1 Slc39a12 Ntsr2 Plcd4 Aqp4 Gjb6 Cbs Chrdl1 Prodh Mlc1 Acsl6 Slc4a4 Gabrg1 Cxcl14 Vcam1 Gfap Pdgfra Pcdh15 Gria3 Cacng4 E130309F12Rik 1110015O18Rik Brca1 Zcchc12 Enpp6 Kif19a Dct 1700040N02Rik 1810041L15Rik 9530059O14Rik 9630013A20Rik Opalin Aspa Mobp Gpr37 Ppp1r14a Gjb1 Kctd13 Cntn2 Ctss Cx3cr1 C1qb Cd53 Csf1r Itgam Abi3 C1qa Aif1 Trem2 P2ry13 Tmem119 C1qc Cd14 Fcgr3 Gpr34 Inpp5d Nckap1l Mpeg1 Siglech Susd3 Hk2 Ly86 1700110I01Rik 1810011H11Rik Xdh Tbc1d4 AI467606 Fas Nostrin Pacs1 Id1 Pglyrp1 Myl9 Bgn Nupr1 Mylk Gprc5c Slc38a11 Slc6a20a Pcolce Vtn Cnn2 Nid1 Gpr30 Higd1b Ifitm1 P2ry14 Serping1 Plac9 0610007N19Rik Ace2
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*.


a
All Cells, N=1679
Total RNA (pg)
Median = 11.5 10.8 12.2 2.5
N = 1525 761 764 154
all neurons
GABAergic
Glutamatergic
non-neuronal
b
Core Cells, N=1424
−log10(adj. p)
Total RNA (pg)
c
All Cells with ≥5M total reads, N=1610
Genes detected
Mean (Thousands) = 7.3 7.1 7.5 4.1
N = 1463 721 742 147
d
Core Cells with >5M total reads, N=1361
Genes detected
e
Number of genes
TPM
f
Neurons
Non-neuronal
Number of genes
TPM

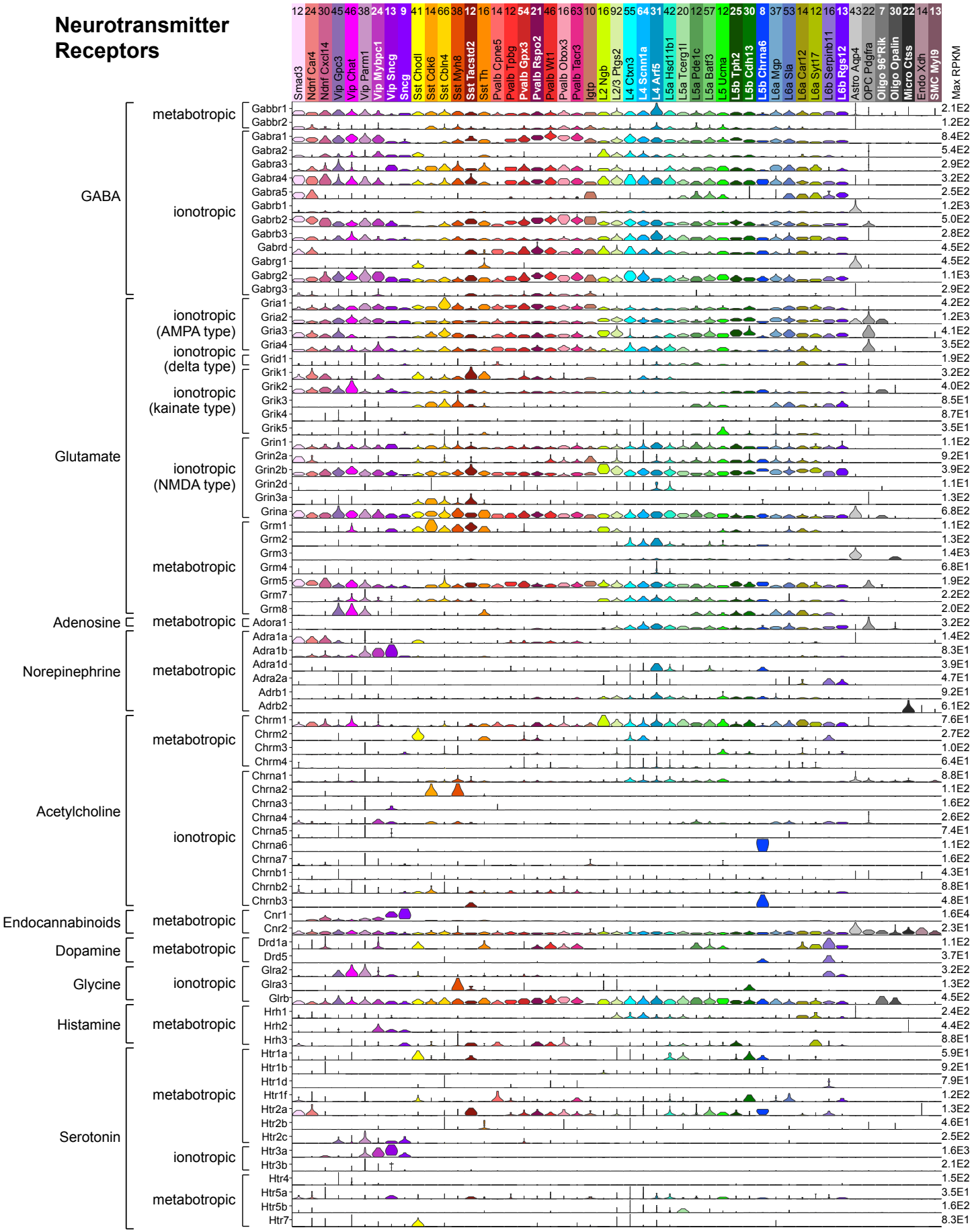
Supplementary Figure 13. RNA content, gene count and distribution of gene abundances for single cells belonging to different cell types. (**a**) Estimation of single cell RNA content based on the ratio of synthetic spike-in ERCC reads and cellular reads (**Methods**) for all cells from major cell classes (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



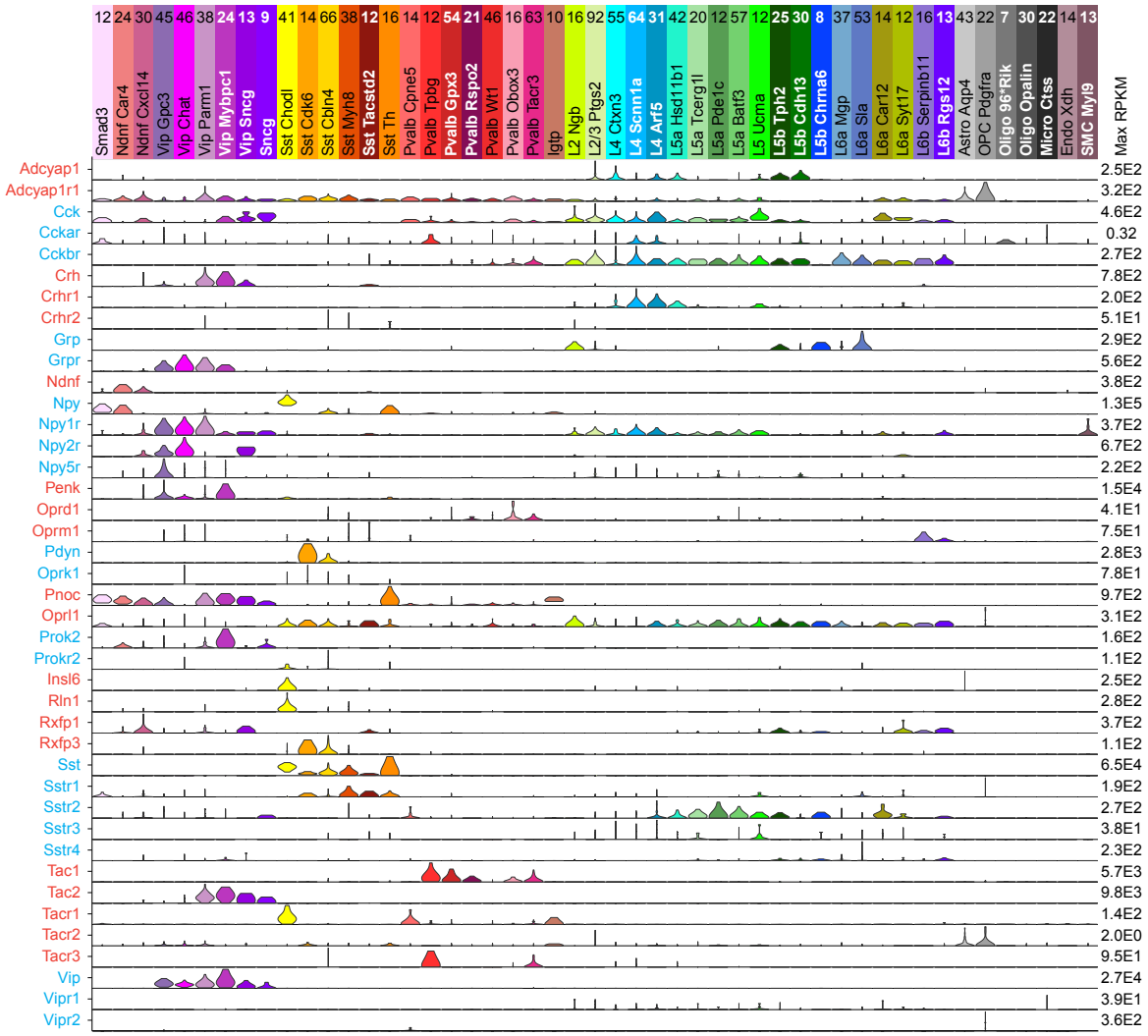
Supplementary Figure 14. Expression of ion channels in cell types. Violin plots represent the gene expression distributions (rows) among single cells within each of the 49 transcriptomic cell types (columns). Only core cells are used (N = 1424). Expression is on a linear scale and is normalized to the 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



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



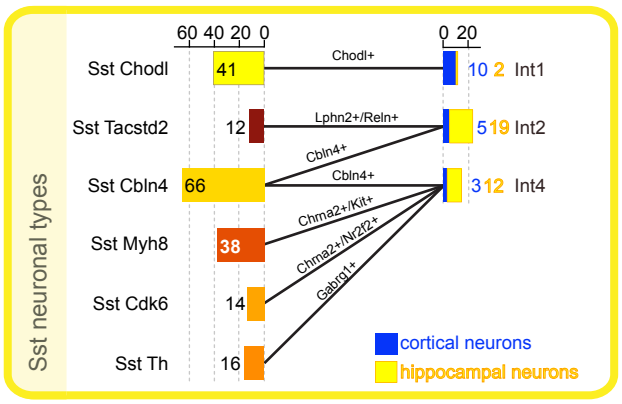
Supplementary Figure 16. Expression of neuropeptides and their receptors in cell types. Violin plots represent the gene expression distributions (rows) among single cells within each of the 49 transcriptomic cell types (columns). Only core cells are used (N = 1424). Expression is on a linear scale and 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*.

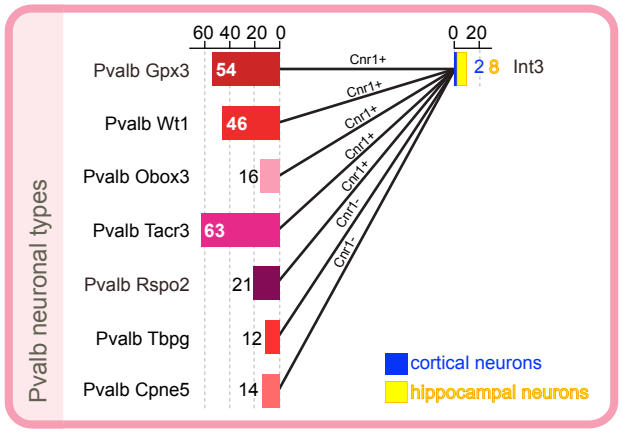
a

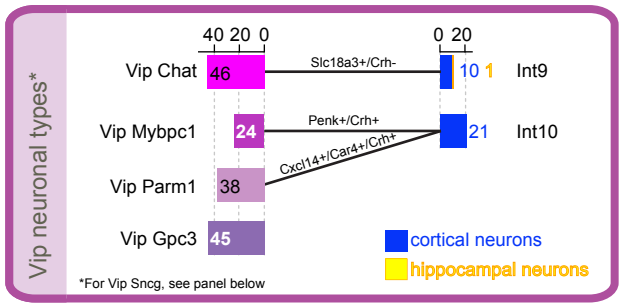
mean number of genes detected
cell number

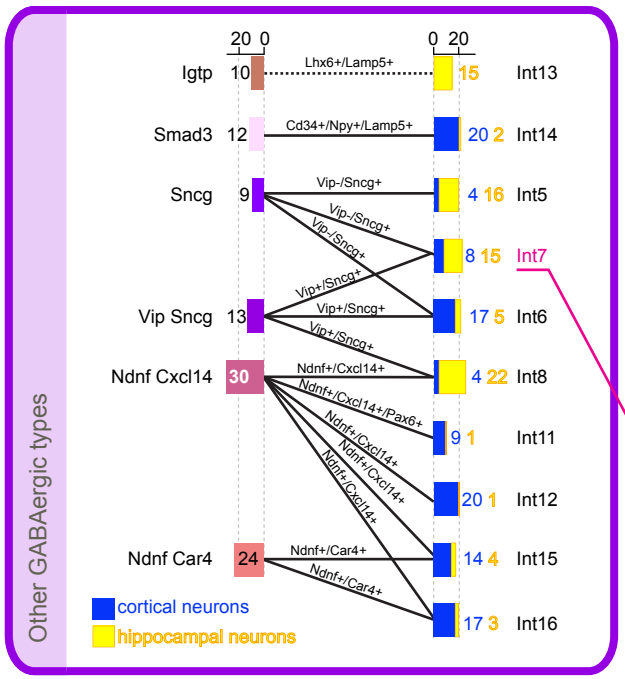
	This study		Zeisel et al. Neocortex		Zeisel et al. Hippocampus	
GABAergic neurons	**761**	7,042	**164**	4,650	**175**	4,911
Glutamatergic neurons	**764**	7,507	**399**	4,543	**939***	4,909
Non-neuronal cells	**154**	4,274	**1128**	2,494	**249**	2,645

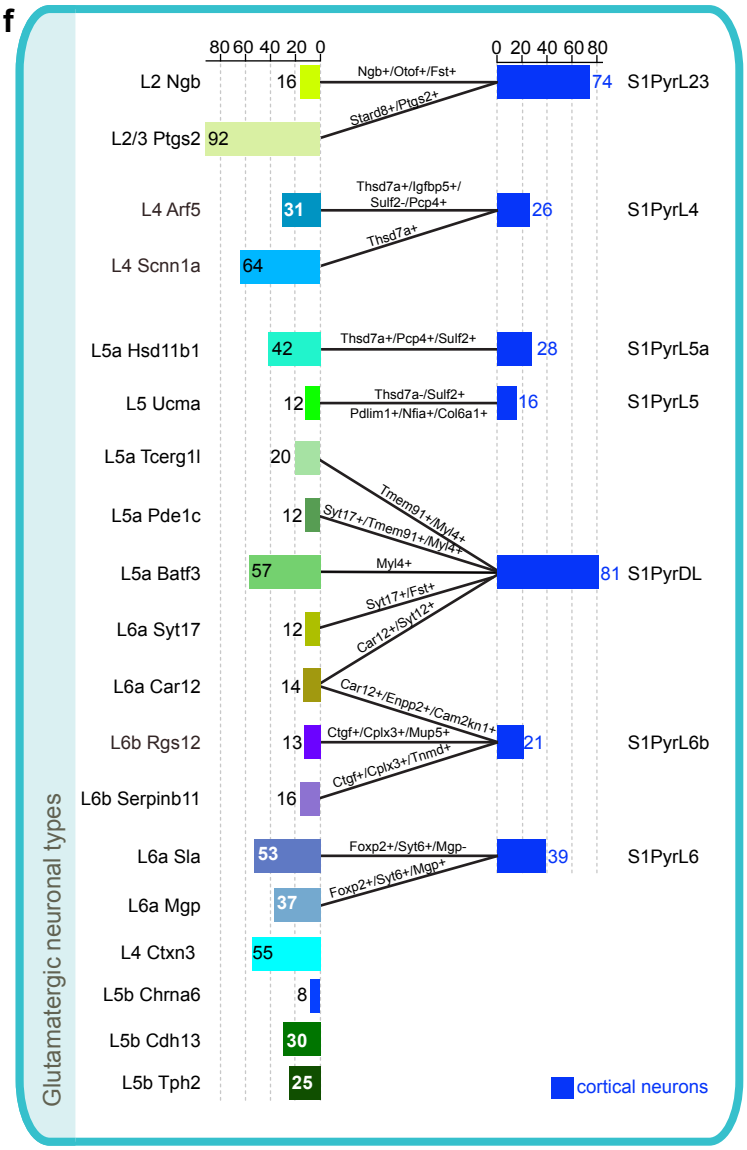
*Not used for the comparison

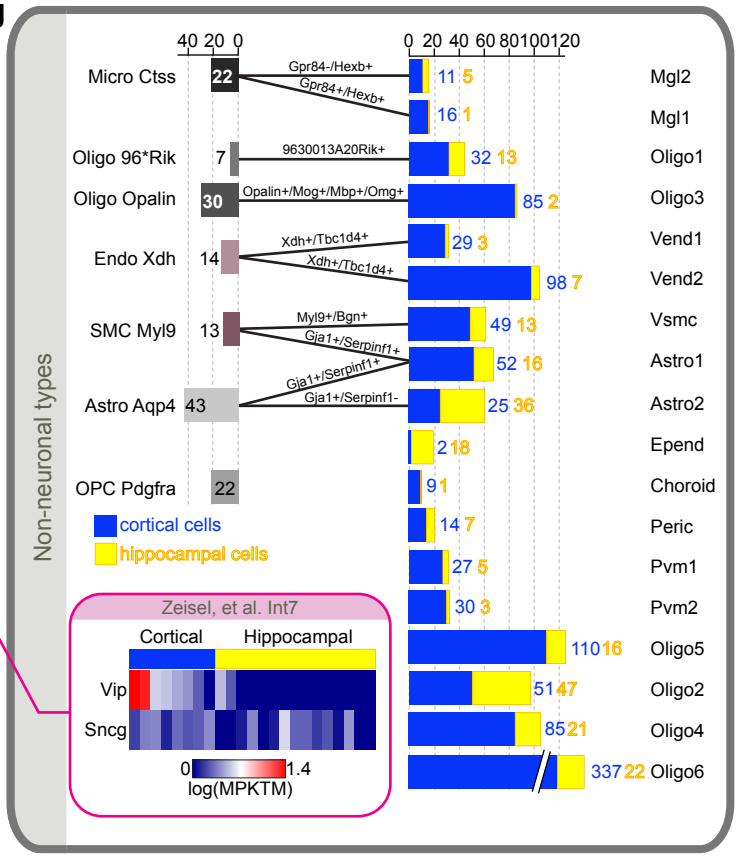












Supplementary Figure 17. Comparison of cell types defined in our study and Zeisel *et al.*[16] (**a**) Summary statistics for cell sampling and gene detection. Compared to the Zeisel *et al.* study, we sampled more cortical neurons and sequenced the individual cells more deeply to detect more genes. 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.