



Published in final edited form as:

*Nat Neurosci.* 2016 February ; 19(2): 335–346. doi:10.1038/nn.4216.

## Adult Mouse Cortical Cell Taxonomy by Single Cell Transcriptomics

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

<sup>1</sup>Allen Institute for Brain Science, Seattle, WA, USA.

### Abstract

Nervous systems are composed of various cell types, but the extent of cell type diversity is poorly understood. Here, we construct a cellular taxonomy of one cortical region, primary visual cortex, in adult mice based on single cell RNA-sequencing. We identify 49 transcriptomic cell types including 23 GABAergic, 19 glutamatergic and seven non-neuronal types. We also analyze cell-type specific mRNA processing and characterize genetic access to these transcriptomic types by many transgenic Cre lines. Finally, we show that some of our transcriptomic cell types display 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 due to the variety and scope of functions it controls, the diversity of cells it comprises, and the number of genes it expresses<sup>1,2</sup>. Within the mammalian brain, the neocortex plays essential roles in 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<sup>3</sup>. In the adult cortex, many types of neurons have been identified through characterization of their molecular, morphological, connective, physiological and functional properties<sup>4–8</sup>. Despite much effort, objective classification

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>3</sup>Correspondence to: Bosiljka Tasic (bosiljkat@alleninstitute.org).

<sup>2</sup>These authors contributed equally to this work.

**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.

based on quantitative features has been challenging, and our understanding of the extent of cell type diversity remains incomplete<sup>4, 9, 10</sup>.

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<sup>2, 11</sup>, based on cell surface<sup>12, 13</sup> or transgenic markers<sup>5</sup> has been informative; however, any population-level profiling obscures potential heterogeneity within 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<sup>14–17</sup>. Here, we use single cell RNA-seq to characterize and classify more than 1,600 cells from the primary visual cortex in adult male mice. The annotated dataset 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<sup>18</sup>. 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<sup>19</sup> (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, Methods). To label more specific cell populations, Cre lines were combined with Dre or Flp recombinase lines and intersectional reporter lines (*Ai65* or *Ai66*, Supplementary Fig. 1a, Supplementary Table 2, 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<sup>5</sup> with some modifications (Fig. 1a, Supplementary Fig. 1b, 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, Methods). We established quality control (QC) criteria to monitor the experimental process (Supplementary Fig. 2) and data quality (Supplementary Fig. 3b,4,5,6,7, Methods). Our final QC-qualified dataset contains 1679 cells, with more than 98% of cells sequenced to a depth of at least 5 million total reads (median ~8.7 million, range ~3.8–84.3 million, 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, 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 name the 1424 cells that are consistently classified into the same cluster as “core” cells, in contrast to 255 “intermediate” cells, which we define as cells that are classified into more than one cluster by the random forest approach (Fig. 1b, Supplementary Fig. 3, Methods).

This analysis segregated cells into 49 distinct core clusters (Fig. 1c). Based on known markers for major cell classes, we identified 23 GABAergic neuronal clusters (*Snap25*<sup>+</sup>, *Slc17a7*<sup>-</sup>, *Gad1*<sup>+</sup>), 19 glutamatergic neuronal clusters (*Snap25*<sup>+</sup>, *Slc17a7*<sup>+</sup>, *Gad1*<sup>-</sup>), and seven non-neuronal clusters (*Snap25*<sup>-</sup>, *Slc17a7*<sup>-</sup>, *Gad1*<sup>-</sup>) (Fig. 1c). We assigned location and identity to cell types within VISp based on 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. 3a–c); and localized expression patterns of marker genes determined by RNA *in situ* hybridization (ISH) (Supplementary Fig. 9, 10).

As expected, most layer-specific Cre lines label specific types of glutamatergic neurons (Fig. 2a,b; Supplementary Table 4). Some GABAergic types also display laminar enrichment that was uncovered by dissections containing one or several layers (usually upper (L1–4) or lower (L5–6) layers combined, Fig. 2a,b, Supplementary Table 5). Cells within the seven non-neuronal types were mostly isolated as tdT<sup>-</sup> 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 define “unique markers”, which are genes expressed only in that type among all cells sampled. We also identify “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, Methods). For a select set of markers, we employed single and double label RNA ISH (Supplementary Fig. 9, 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<sup>19–22</sup>. In general, we find that the examined Cre lines mostly label expected cell types based on promoters and other genetic elements that control Cre recombinase expression in each line (Fig. 2a,b, Methods: Supplementary Note 1)<sup>19</sup>. However, all but one Cre line (*Chat-IRES-Cre*) label more than one transcriptomic cell type.

### Cortical cell types: markers and relationships

To provide an overall view of the transcriptomic cell types 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). Within these diagrams, each transcriptomic cell type is represented by a disc, whose surface area corresponds to the number of core cells in our dataset belonging to that type. Intermediate cells are represented by lines connecting the discs; the line thickness is proportional to the number of intermediate cells. We separately present GABAergic, glutamatergic and non-neuronal constellations as we detect 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: presence of a large number of intermediate cells between a particular pair of core types suggests a phenotypic continuum, while 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, while the other, the overall similarity in gene expression between cluster cores. We summarize expression of select marker genes in Supplementary Table 6 and Supplementary Fig. 12.

Our analysis identifies 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)<sup>6, 23, 24</sup>. In a substantial portion of these cells, we detect more than one of these markers, but our method, which takes into account genome-wide gene expression, usually classifies these double-expressing cells into the major type corresponding to the most highly expressed major marker in that cell (Methods: Supplementary Note 2).

We identify five additional GABAergic types. In agreement with a previous report<sup>25</sup>, we detect *Tnfrsf813* and *Sema3c* in these types. We name two of them based on a gene for a putative neuropeptide, neuron-derived neurotrophic factor (*Ndnf*), and we provide evidence that they correspond to neurogliaform cells (see below). We name 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 is not straightforward and relies 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 are 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 (e.g., *Calb1* (calbindin), *Calb2* (calretinin), *Cck*, *Crh*, *Htr3a*, *Nos1*, *Npy*, *Reln*)<sup>4, 6</sup>. We describe the comparison with the existing literature below, and summarize it in Supplementary Table 7.

We find only one Sst type (Sst-Cbln4) that is prevalent in upper cortical layers, while all other Sst types appear enriched in lower layers (Fig. 2b, 4a). Based on the upper layer-enrichment and *Calb2* expression of the Sst-Cbln4 type, we propose that it likely corresponds to previously characterized Calb2-positive Martinotti cells that are enriched in

the upper cortical layers<sup>26</sup>, and are fluorescently labeled in transgenic “GIN” mice<sup>27</sup>. Our analysis reveals only one additional Calb2-positive Sst type, which we name Sst-Chodl (Fig. 2b). Based on the expression of tachykinin-receptor 1 (*Tacr1*), neuropeptide Y (*Npy*), high levels of nitric oxide synthase (*Nos1*) and absence of *Calb1* (Fig. 3a, Supplementary Fig. 9), this type most likely corresponds to Nos1 Type I neurons<sup>28</sup>, which are enriched in L5 and 6 (ref. 29), and are likely long-range projecting<sup>30</sup>, sleep-active neurons<sup>31</sup>.

The Pvalb types are highly interconnected in the constellation diagrams (Fig. 4a). Using layer-enriching dissections (Fig. 2b), we find that some types are 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<sup>32</sup>. Our analysis classifies cells from this line within all three upper layer-enriched Pvalb types (Fig. 4a). We suggest that Pvalb-Cpne5 corresponds to chandelier cells because it is most transcriptionally distinct among Pvalb types, it is enriched in upper layers, and it does not express *Etv1* (also known as *Er81*) as previously shown for chandelier cells<sup>33</sup> (Supplementary Fig. 12).

The Vip major type can be divided into several transcriptomic cell types, all of which appear enriched in upper cortical layers, except the Vip-Gpc3 type (Fig. 4a). In agreement with previous reports<sup>23, 34</sup>, our Vip-Chat transcriptomic type is located in upper cortical layers (Fig 2a), and it displays unique expression of choline acetyltransferase (*Chat*) in Vip-positive cells. These cells were reported to either express<sup>34</sup> or not express Calb2 at the protein level<sup>23</sup>; we find that they robustly express *Calb2* mRNA.

For glutamatergic cells, we identify six major classes of transcriptomic types – L2/3, L4, L5a, L5b, L6a, and L6b – based on the layer-specific expression of marker genes and layer-enriching dissections; this is in agreement with many previous studies<sup>1, 7, 8, 35</sup>. In this study, we discover subdivisions among all of these layer-specific major types. Within L2/3, we identify two major types, one of which (L2-Ngb) appears to be located more superficially based on marker gene expression (e.g., *Ngb*, *Fst*, *Syt17*, and *Cdh13*, Fig. 3, Supplementary Fig. 9). Within L4, we identify 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 identify eight different transcriptomic types within L5. Four of these types express the L5a marker *Deptor* (L5a-Hsd11b1, L5a-Tcerg11, L5a-Batf3, and L5a-Pde1c), while three express the L5b marker *Bcl6* (L5b-Cdh13, L5b-Tph2, and L5b-Chrna6, Fig 3b). One of those L5b types (L5b-Chrna6), together with the L5-Ucma type, appear most distinct among L5 types, both based on gene expression and the small number of intermediate cells between them and other L5 types (Fig. 4b). We identify six transcriptomic cell types within L6: four L6a types, and two L6b types. Among L6a types, two highly related types (L6a-Sla, and L6a-Mgp) express the marker *Foxp2* (refs. 7, 35, 36), and were primarily derived from the *Ntsr1-Cre* line (Fig 2b), while the other two (L6a-Syt17 and L6a-Car12) do not express *Foxp2*, and were isolated as tdT<sup>-</sup> cells from L6 of the same Cre line. For the latter two types, we discover several new markers that can be used to identify them (*Car12*, *Prss22*, *Syt17* and *Penk*, Fig 3b, Supplementary Fig. 9b, 10j–k). The two L6b types (L6b-Serp1b11 and L6b-

Rgs12) express the known L6b marker *Ctgf*<sup>7, 35, 36</sup>, and several other previously identified L6b markers (e.g., *Trh*, *Tnmd*, *Mup5*, Fig. 3b; Supplementary Fig. 9b)<sup>7</sup>.

Despite the neuronal focus of this study, our sampling strategy captured enough cells to identify the major non-neuronal classes as well. We find seven non-neuronal types: astrocytes, microglia, oligodendrocyte precursor cells (OPCs), two types of oligodendrocytes, endothelial cells and smooth muscle cells. In agreement with previous population-level studies<sup>12, 13</sup>, these types can be distinguished by many combinatorial and unique markers (Fig. 4c, Supplementary Fig. 12, Methods: Supplementary Note 2).

### Comparative analysis of cell types

After defining cell types, we examined additional cellular properties that can be extracted from our dataset. We show that neurons contain more total RNA than non-neuronal cells (median 11.5 vs. 2.5 pg) and express more genes when sequenced to the same depth (mean 7278 vs. 4274) (Supplementary Fig. 13a,c). We estimate that some neuronal types can have >20-fold higher RNA content than some glial types (e.g., L5b-Tph2 ~37.0 pg/cell vs. microglia ~1.6 pg/cell, Supplementary Fig. 13b). We also find differences in the distribution of gene abundances among cell types: overall, neurons express more genes at low/intermediate levels than non-neuronal cells, while non-neuronal cells express 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 find a total of 567 exons within 320 genes that display differential pre-mRNA processing in a cell type-specific manner at various levels of cellular taxonomy (Fig. 5a, Supplementary Table 8). Several examples are shown in Fig. 5b–e: mRNAs for pyruvate kinase (*Pkm*), syntaxin binding protein 1 (*Stxbp1*), and subunits of the AMPA receptors, *Gria1* and *Gria2*. The last two display highly cell-type specific alternative splicing for two consecutive exons (previously named “flip” and “flop”)<sup>37</sup>, of which only a single one is included in each mature mRNA. Each exon encodes a small segment of the predicted fourth transmembrane region, which imparts different electrophysiological properties to the receptors<sup>37</sup>. In agreement with relatively low-resolution RNA ISH data<sup>37</sup>, we find that L2-Ngb and L2/3-Ptgs2 types preferentially use the flip exons, L4 types use the flop exons, while L6a types utilize 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 splicing 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 thus allowed cells belonging to the same cell type to be analyzed together to discover robust cell type-specific signatures of RNA processing.

Within this genome-wide dataset, we also explored the expression of genes particularly relevant for neuronal development and function. Examination of transcription factors reveals a number of genes that have been previously shown to be involved in specification of

neuronal types (Supplementary Fig. 12). As expected, many more ion channel genes are expressed in neurons than glia, and many are differentially expressed but rarely unique for specific cell types (Supplementary Fig. 14). We observe widespread neuronal expression of many glutamate and GABA receptors, including both ionotropic and metabotropic types, while the receptors for other, mostly modulatory neurotransmitters, are generally expressed at lower levels, and more selectively in certain cell types (Supplementary Fig. 15). Neuropeptide genes are usually selectively expressed in one or a few GABAergic cell types, while the receptors for these neuropeptides can be specific for other cell types, suggesting specific cell-cell interactions (Supplementary Fig. 16).

### Transcriptomic cell types and neuronal properties

To inquire if the transcriptomic cell types defined here 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 (CAVCre) 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, 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<sup>7, 35</sup>. In contrast, L5b neurons, which express *Bcl6*, have been shown to project subcortically and have been designated as cortico-fugal projection neurons<sup>7, 35</sup>. Accordingly, our cells labeled from contralateral VISp and ipsilateral thalamus are classified respectively into transcriptomic L5a and L5b types (Fig. 6). The retrograde labeling of L6a types is also in agreement with the previous literature. 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<sup>38, 39</sup>, which in our dataset correspond to L6a-Mgp and L6a-Sla types (Fig. 2b). In comparison, the *Ntsr1-Cre*-negative 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<sup>38, 39</sup>.

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<sup>6</sup>. We used the *Ndnf* gene to generate a Cre line that should enable specific access to these cells (Methods). Indeed, in agreement with the *Ndnf* mRNA ISH data (Supplementary Fig. 9a, 10h, 10l), we find that this Cre line labels neurons that are

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<sup>40, 41</sup>, accelerating spike frequency<sup>42</sup>, gap junctional coupling<sup>43, 44</sup>, and slow GABA-mediated synaptic transmission<sup>43, 45</sup>. Some neurogliaform cells have been shown to exhibit one or two action potentials at the onset of the long current pulse near threshold<sup>40, 41</sup>. Neurogliaform cells can also form GABA-mediated autaptic synapses<sup>45</sup>.

Based on whole cell current clamp recordings of tdT<sup>+</sup> 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,g). NLS neurons displayed the initial depolarizing response that was sufficient to induce an action potential at the onset of the current step in some trials (Fig. 7f,g). The NLS neurons, to differing degrees, exhibited an initial depolarizing response that sagged (Fig. 7f, top, inset, 7g). At slightly higher current intensities, all NLS neurons initiated a bout of late spiking after a period of quiescence (Fig. 7f, bottom). In multi-patch recordings, we observed frequent electrical coupling (Fig. 7h) and autaptic and synaptic transmission between tdT<sup>+</sup> neurons that was blocked by the GABA<sub>A</sub> receptor antagonist SR95531 (Fig. 7i). Reconstruction of two biocytin-filled, tdT<sup>+</sup> neurons revealed one of them to have the tight, dense axonal arbor with small, bouton-like structures, and a relatively small dendritic tree that is typical of neurogliaform cells<sup>46</sup>. The other neuron displayed axonal and dendritic arbors more similar to the recently described neurogliaform sparse-axon cells (Fig. 7j)<sup>47</sup>. Together, molecular, physiological and morphological analyses of L1 neurons labeled by the *Ndnf-IRES2-dgCre* line show that they correspond to neurogliaform cells.

## DISCUSSION

The adult mouse visual cortex contains about one million cells, of which about half are neurons<sup>48</sup> that can be divided into glutamatergic (80%) and GABAergic cells (20%)<sup>49</sup>. We define cell types within the primary visual cortex based on 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 catalogue 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 (like 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 dataset, the majority of marker genes show excellent correspondence to RNA ISH data from the Allen Brain



Atlas<sup>1</sup> (~72% out of N = 228 examined genes, Supplementary Table 9), suggesting that our procedure does not dramatically alter the transcriptional signatures of cell types. Most of the other examined transcripts within this set (Supplementary Table 9), which appear to be very specific markers based on RNA-seq and qRT-PCR (e.g., Chodl), are 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<sup>17</sup> or allowed cells to have only a single identity<sup>14–16</sup>. 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 contribute to our datasets. For example, similarly to a previous single-cell transcriptomic study<sup>16</sup>, we estimate that we detect only ~23% of mRNA molecules present in a cell (Supplementary Fig. 4). 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 define here into additional subtypes. This is already apparent in our dataset, as we observe more subtypes if we decrease the threshold for the minimal number of core cells required to define a type (Methods: Supplementary Note 2). 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<sup>14, 17, 50</sup>. One of the main advantages of low-depth sequencing is reduction of experimental cost. However, we note that if we downsample our data from full depth to 1 million or 100,000 mapped reads per cell, we lose the power to detect many types (Supplementary Table 10). Thus, when subsampling to 100,000 reads, we only find 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<sup>16</sup>. Based on the expression of key marker genes, we find both commonalities and differences in the cell types identified in these two studies (Supplementary Fig. 17). For neuronal cells, we identify more transcriptomic glutamatergic (19 vs. seven) and GABAergic Sst (six vs. three), Pvalb (seven vs. one), and Vip (five vs. three) types, but fewer other GABAergic types (five vs. nine) (Supplementary Fig. 17). For non-neuronal cortical cells, Zeisel *et al.*<sup>16</sup> 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 differ in a number of experimental and data analysis parameters. For example, due to different sampling strategies (Cre line-based versus mostly unbiased), we analyzed more neocortical neurons (1525 vs. 563); due to the differences in RNA-seq procedures (SMARTer vs. 5'-end focused STRT) and sequencing depths, we detect more genes in these neurons (~7200 vs. ~4500) (Supplementary Fig. 17). Our studies also differ in the genetic background (mostly C57BL/6J versus CD-1) and age of analyzed mice, as well as the cell isolation procedures (FACS vs. mostly Fluidigm C1 microfluidics). Overall, the two studies overlap in their identification of some transcriptomic types, but differ in their focus: Zeisel *et al.*<sup>16</sup> offer deeper insight into non-neuronal transcriptomic types, hippocampal excitatory cells and cells from brain ventricles, while our study provides a more comprehensive classification of adult neocortical neurons.

Our study suggests 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 (late and non-late spiking) correspond to each other, and which genes are responsible for these physiological differences? Do the two corticothalamic L6a subtypes (L6a-S1a 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<sup>33</sup>, perhaps mirroring the large numbers of intermediate cells we find for upper layer *Etv1(Er81)*-positive Pvalb cells (Fig. 3a). While these questions await further studies, the approach detailed here provides an overview of adult cell types within a well-defined cortical area based on a highly multidimensional dataset, and is an essential step towards understanding the most complex animal organ, the mammalian brain.

## METHODS

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

### Data, reagent, and code availability

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. 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 (Supplemental Table 6). The newly generated mouse lines are in the process of being deposited to the Jackson Laboratory.

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.

## METHODS

### Mouse breeding and husbandry

All procedures were carried out in accordance with IACUC 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 5 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, such 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 Table 1<sup>19, 20, 32, 52–60</sup> and Supplementary Table 2<sup>56, 58</sup>, 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) 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/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/g of body weight using 20 mg/ml 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, Piscataway, US) and Red/ET recombineering (Gene Bridges, Heidelberg, DE). 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)<sup>51</sup> to delete the selection marker cassette, then backcrossed to C57BL/6J mice and maintained in the C57BL/6J background. 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<sup>61</sup>. The *Ctgf-2A-dgCre* targeted transgene contains a viral 2A peptide (modified T2A, 5'-gagggcagaggaagtcttctaactgcggtgacgtggaggagaatccccgcct-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 (CAVCre, gift of Miguel Chillon Rodrigues, Universitat Autònoma de Barcelona, Spain)<sup>62</sup> into brains of heterozygous *Ai14* mice using a previously described procedure with modifications<sup>63</sup>. 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 CAVCre 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<sup>64</sup> 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 days after surgery for single cell isolation. TdT<sup>+</sup> 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<sup>5, 65</sup>. 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), NaHCO<sub>3</sub> (20 mM), dextrose (20 mM), KCl (3 mM), NaH<sub>2</sub>PO<sub>4</sub> (1.25 mM), CaCl<sub>2</sub> (2 mM), MgCl<sub>2</sub> (2 mM), DL-AP5 sodium salt (50 μM), DNQX (20 μM), and tetrodotoxin (0.1 μM), bubbled with a carbogen gas (95% O<sub>2</sub> and 5% CO<sub>2</sub>). 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. After the brain slicing was complete (not more than 15 minutes), individual slices of interest were transferred to a small Petri dish containing bubbled room temperature ACSF. 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 minutes 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. 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<sup>37</sup>.

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  $\mu$ l RNase inhibitor (SMARTer kit), and 0.1  $\mu$ l of 1:1,000,000 diluted RNA spike-in RNAs (ERCC RNA Spike-In Mix 1, Life Technologies Cat#4456740). Sorting was performed on a BD FACSAriaII SORP using a 130  $\mu$ m 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. 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 Fig. 2a, and confirmed post-hoc by observing dramatically higher expression of *tdT* mRNA in tdT<sup>+</sup> than in tdT<sup>-</sup> cells (Supplementary Fig. 2c). In some cases, we also selected cells that have low DAPI and low tdT fluorescence, in order to capture tdT<sup>-</sup> 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 polyA-RNA and amplify cDNA<sup>14, 66–68</sup>. 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  $\mu$ l 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 (e.g., 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 RSEM<sup>69</sup>, and unmapped reads were then aligned to the ERCC and tdT sequences using Bowtie<sup>70</sup>. The remaining unmapped reads were aligned to the mm10 genome. 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<sup>71</sup> and DESeq<sup>72</sup> 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<sup>71</sup>. 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 SD of each eigenvalue, and selecting those PCs whose eigenvalue was greater than the mean + 2 SDs. 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  $< 0.01$ .

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

We also developed an alternative clustering approach which iteratively applies Weighted Gene Coexpression Network Analysis (WGCNA)<sup>73</sup> 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<sup>71</sup>, 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  $-\log_{10}(\text{adjusted p-value})$  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)<sup>74</sup>, and reclustering using this set of DE genes.

For iterative WGCNA, the clustering terminates 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<sup>74</sup>.
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 from 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 of transcriptomic 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 1424 core and 255 intermediate cells. For 20 genes and  $p < 0.01$ , we obtained 1413 core and 266 intermediate cells. Restricting the number of genes to 10 does not have a major effect: we obtained 1423 core/256 intermediate cells using a differential expression  $p < 0.05$  and 1418 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: 1369 core/310 intermediate cells using  $p < 0.05$  and 1383 core/296 intermediate cells. The full assignments of each cell for each of these conditions is 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-My19 type (Ct1988\_V, Ct1994\_V, Ct1986\_V and Nd1968n\_V1), which do not express *My19* and *Flt1*, but express *Lum*, *Dcn*, *Colla1*, 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-My19 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-My19 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 CAVCre-labeled cells to RNA-seq clusters

In order 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 within 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<sup>72</sup>. 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 20<sup>th</sup> percentile expression in type 1 is greater than the 80<sup>th</sup> percentile expression in type 2. This ensures a good separation of distributions.
2. For a given pair of cell types, the 80<sup>th</sup> 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 towards 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 within 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 Fig. 12 and Supplementary Table 6.

### Evaluation of differential exon usage

We used the limma Bioconductor package<sup>74</sup> 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 two-fold 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 dataset. From these candidates, we selected only the exons that are differentially expressed compared to their corresponding gene. We used MISO<sup>75</sup> to confirm differential exon processing for select examples. The MISO score ( $\Psi$ ), or “percent spliced-in”, represents the relative exon usage of transcript variant  $b$  vs.  $a$ , for each gene in each cell type<sup>75</sup>. Because MISO does not accommodate replicates, to calculate MISO  $\Psi$ , we pooled 10 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  $\Psi$  score that is non-zero) over no differential expression (change in  $\Psi$  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 vs. 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<sup>76</sup>, which was based on the Allen Institute’s colorimetric RNA ISH protocol<sup>1</sup>. Tissue sections (25  $\mu\text{m}$ ) 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, 1000, 3200, 10000 and 20000 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 standard deviations 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 to 1000 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  $\mu$ l of buffer (5  $\mu$ l of Cells Direct 2X Reaction Mix (Thermo Fisher Scientific) and 0.1  $\mu$ l of SUPERase In RNase Inhibitor (20 U/ $\mu$ l; Thermo Fisher Scientific)) and frozen at  $-80^\circ\text{C}$ . Synthetic transcripts (ERCC RNA Spike-in Mix1, Life Technologies Cat#4456740, 1  $\mu$ l 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 5-fold 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 8 different whole mouse RNA dilutions (1, 5, 10, 32, 100, 320, 1000 and

3200 pg / RT-STA reaction), gDNA (100 pg / RT-STA reaction), water with ERCCs and water without ERCCs. The ERCC controls allowed monitoring of the PCR efficiency in each sample well (the 9 assayed ERCC transcripts cover a range from 1 to 4100 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-IRESS2-dgCre;Ai14* mice. Mice were anesthetized with 5% isoflurane, perfused transcardially with artificial cerebrospinal fluid (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 minutes and then to a holding chamber at room temperature (22°C). For perfusion and slice incubation, aCSF contained (in mM): 98 NMDG, 98 HCl, 25 D-glucose, 25 NaHCO<sub>3</sub>, 17.5 HEPES, 12 N-acetyl-L-cysteine, 10 MgSO<sub>4</sub>, 5 Na-(L)-ascorbate, 3 myoinositol, 3 Na-pyruvate, 2.5 KCl, 2 mM thiourea, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, and 0.01 taurine, or 73 Tris-HCl, 30 NaHCO<sub>3</sub>, 28 Tris Base, 25 D-Glucose, 20 HEPES, 10 MgSO<sub>4</sub>, 5 Na-(L)-ascorbate, 3 Na-pyruvate, 2.5 KCl, 2 thiourea, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, and 0.5 CaCl<sub>2</sub>. The holding chamber solution contained (in mM): 97 NaCl, 25 NaHCO<sub>3</sub>, 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<sub>2</sub>, 2 MgSO<sub>4</sub>, 2 thiourea, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, and 0.01 taurine.

**Patch-clamp recording**—Recordings were performed in aCSF containing (in mM): 126 NaCl, 26 NaHCO<sub>3</sub>, 12.5 D-glucose, 2.5 KCl, 2 CaCl<sub>2</sub>, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, and 1 MgSO<sub>4</sub>. Individual slices were held in a small chamber perfused with aCSF at 2.5 mL/min (32–34°C) and visualized with an upright, fixed-stage microscope (Scientifica SliceScope) using dot-gradient contrast, infrared video microscopy. Fluorescent tdT<sup>+</sup> 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 minutes 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, f and g) of tdT<sup>+</sup> 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<sup>+</sup> neurons (Fig. 7h and 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.  $G_j = (1/R_2) \times CC/$

(1-CC) was used to calculate the junctional conductance. CC is the coupling coefficient and R2 is the input resistance of the noninjected cell<sup>43</sup>.

**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, Burlingame, CA). Stained slices were post-fixed with 0.05% OsO<sub>4</sub> and mounted with MOWIOL 4–88. Biocytin-filled neurons were imaged in 3D with a Zeiss AxioImager (Thornwood, NY) at 63x with 1.4-NA objective. Images were then inverted and imported into Vaa3D<sup>77</sup> for semi-automated reconstruction using the virtual finger tool.

### Supplementary Note 1 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<sup>19</sup>. 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<sup>+</sup> 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 points No.1 and/or No.2 above.

**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 due to developmental *Calb2* expression or a transgenic artefact. Second, based on RNA-Seq, this line should label some Sst cells, but *Sst*-positive cells were not among tdT<sup>+</sup> cells collected from *Calb2-IRES-Cre;Ai14* mice (See Fig. 2b). This could be due to disruption of a regulatory element in transgenesis that is responsible for *Calb2* expression in Sst cells.

**Example 3**—As previously reported<sup>25, 78</sup>, 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<sup>24</sup>. 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.

### Supplementary Note 2 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 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<sup>25, 79</sup>. 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 *Gfap*<sup>12</sup>. 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 oligodendrocytes<sup>13</sup>. Oligo precursor cells (OPC) express *Pdgfra* and *Cspg4* as previously reported<sup>12, 13</sup>. Accordingly, microglial cells express *Itgam*, *Cx3cr1*, and *C1qb*<sup>13</sup>. We identify endothelial cells based on expression of *Flt1*<sup>13</sup>, and smooth muscle cells (SMC) based on the expression of *Bgn*<sup>80</sup>.

**Cells with unexpected combinations of markers**—We note three cells that, although they passed our QC criteria (Methods, 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.

A Supplementary Methods Checklist is available.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank Miguel Chillon Rodrigues for providing CAVCre, Stefan Mihalas for advice on data analysis, Hong Gu, Maya Mills, Harminder Gill and Kristen Hadley for technical assistance, Chaoyang Ye and Ajamete Kaykas for help with the next generation sequencing, and the Department of In Vivo Sciences, especially Rachael Larsen, Laura Pearson and James Harrington for mouse husbandry. We thank Jack Waters and Ed Lein for comments on the manuscript. This work was funded by the Allen Institute for Brain Science, and by NIH grants R01EY023173 and U01MH105982 to H.Z. The authors thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement and support.

## References

1. Lein ES, et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*. 2007; 445:168–176. [PubMed: 17151600]
2. Hawrylycz MJ, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012; 489:391–399. [PubMed: 22996553]
3. Harris KD, Shepherd GM. The neocortical circuit: themes and variations. *Nature neuroscience*. 2015; 18:170–181. [PubMed: 25622573]
4. DeFelipe J, et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nature reviews. Neuroscience*. 2013; 14:202–216.
5. Sugino K, et al. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature neuroscience*. 2006; 9:99–107. [PubMed: 16369481]
6. Rudy B, Fishell G, Lee S, Hjerling-Leffler J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Developmental Neurobiology*. 2011; 71:45–61. [PubMed: 21154909]
7. Sorensen SA, et al. Correlated Gene Expression and Target Specificity Demonstrate Excitatory Projection Neuron Diversity. *Cereb Cortex*. 2013
8. Greig LC, Woodworth MB, Galazo MJ, Padmanabhan H, Macklis JD. Molecular logic of neocortical projection neuron specification, development and diversity. *Nature reviews. Neuroscience*. 2013; 14:755–769. [PubMed: 24105342]
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*. 2004; 14:1310–1327. [PubMed: 15192011]
10. Ascoli GA, et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature reviews. Neuroscience*. 2008; 9:557–568. [PubMed: 18568015]
11. Belgard TG, et al. A transcriptomic atlas of mouse neocortical layers. *Neuron*. 2011; 71:605–616. [PubMed: 21867878]
12. Cahoy JD, 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*. 2008; 28:264–278. [PubMed: 18171944]

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*. 2014; 34:11929–11947. [PubMed: 25186741]
14. Pollen AA, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature biotechnology*. 2014; 32:1053–1058.
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*. 2015; 347:1138–1142. [PubMed: 25700174]
17. Macosko EZ, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015; 161:1202–1214. [PubMed: 26000488]
18. Glickfeld LL, Reid RC, Andermann ML. A mouse model of higher visual cortical function. *Current opinion in neurobiology*. 2014; 24:28–33. [PubMed: 24492075]
19. Harris JA, et al. Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Frontiers in neural circuits*. 2014; 8:76. [PubMed: 25071457]
20. Taniguchi H, et al. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron*. 2011; 71:995–1013. [PubMed: 21943598]
21. Olsen SR, Bortone DS, Adesnik H, Scanziani M. Gain control by layer six in cortical circuits of vision. *Nature*. 2012; 483:47–52. [PubMed: 22367547]
22. Huang ZJ. Toward a genetic dissection of cortical circuits in the mouse. *Neuron*. 2014; 83:1284–1302. [PubMed: 25233312]
23. Gonchar Y, Wang Q, Burkhalter AH. Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining. *Frontiers in neuroanatomy*. 2008; 2
24. Xu X, Roby KD, Callaway EM. Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *The Journal of comparative neurology*. 2010; 518:389–404. [PubMed: 19950390]
25. Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience*. 2013; 16:1068–1076. [PubMed: 23817549]
26. Xu X, Roby KD, Callaway EM. Mouse cortical inhibitory neuron type that coexpresses somatostatin and calretinin. *The Journal of comparative neurology*. 2006; 499:144–160. [PubMed: 16958092]
27. Oliva AA Jr, Jiang M, Lam T, Smith KL, Swann JW. 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*. 2000; 20:3354–3368. [PubMed: 10777798]
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*. 2005; 1056:183–190. [PubMed: 16102735]
29. Lee JE, Jeon CJ. Immunocytochemical localization of nitric oxide synthase-containing neurons in mouse and rabbit visual cortex and co-localization with calcium-binding proteins. *Molecules and cells*. 2005; 19:408–417. [PubMed: 15995359]
30. Tomioka R, et al. Demonstration of long-range GABAergic connections distributed throughout the mouse neocortex. *The European journal of neuroscience*. 2005; 21:1587–1600. [PubMed: 15845086]
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*. 2008; 105:10227–10232. [PubMed: 18645184]
32. Taniguchi H, Lu J, Huang ZJ. The spatial and temporal origin of chandelier cells in mouse neocortex. *Science*. 2013; 339:70–74. [PubMed: 23180771]
33. Dehorter N, et al. Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch. *Science*. 2015; 349:1216–1220. [PubMed: 26359400]

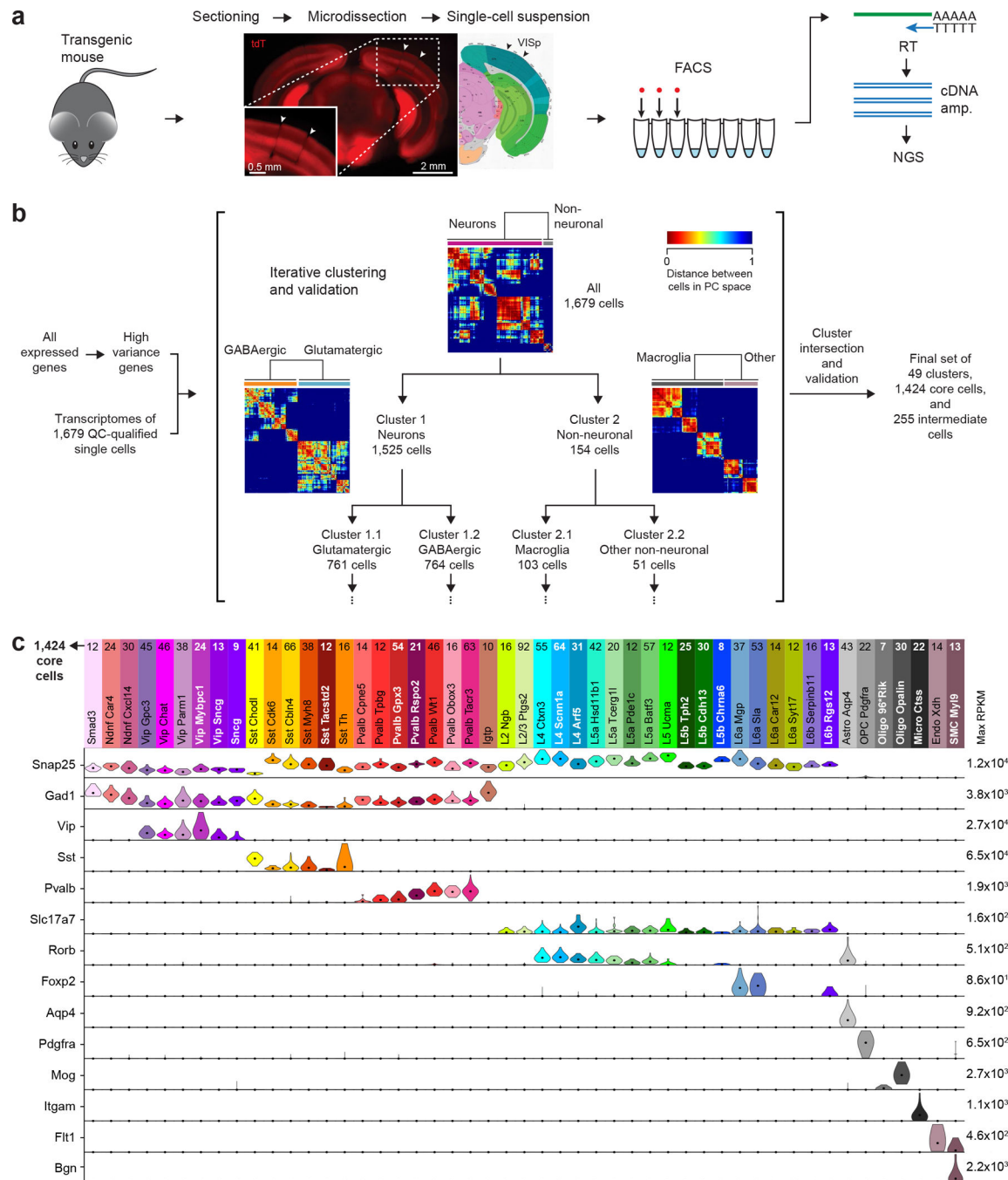
34. von Engelhardt J, Eliava M, Meyer AH, Rozov A, Monyer H. Functional characterization of intrinsic cholinergic interneurons in the cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2007; 27:5633–5642. [PubMed: 17522308]
35. Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD. Neuronal subtype specification in the cerebral cortex. *Nature reviews. Neuroscience*. 2007; 8:427–437. [PubMed: 17514196]
36. Zeng H, et al. Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell*. 2012; 149:483–496. [PubMed: 22500809]
37. Sommer B, et al. Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science*. 1990; 249:1580–1585. [PubMed: 1699275]
38. Velez-Fort M, et al. The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing. *Neuron*. 2014; 83:1431–1443. [PubMed: 25175879]
39. Bortone DS, Olsen SR, Scanziani M. Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron*. 2014; 82:474–485. [PubMed: 24656931]
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*. 1995; 15:2638–2655. [PubMed: 7722619]
41. Hestrin S, Armstrong WE. Morphology and physiology of cortical neurons in layer I. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 1996; 16:5290–5300. [PubMed: 8757242]
42. Povysheva NV, et al. Electrophysiological differences between neurogliaform cells from monkey and rat prefrontal cortex. *Journal of neurophysiology*. 2007; 97:1030–1039. [PubMed: 17122314]
43. Chu Z, Galarreta M, Hestrin S. Synaptic interactions of late-spiking neocortical neurons in layer I. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2003; 23:96–102. [PubMed: 12514205]
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*. 2005; 25:6278–6285. [PubMed: 16000617]
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*. 2010; 30:9898–9909. [PubMed: 20660272]
46. Kawaguchi Y, Kubota Y. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb Cortex*. 1997; 7:476–486. [PubMed: 9276173]
47. Muralidhar S, Wang Y, Markram H. Synaptic and cellular organization of layer 1 of the developing rat somatosensory cortex. *Frontiers in neuroanatomy*. 2013; 7:52. [PubMed: 24474905]
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*. 2013; 7:35. [PubMed: 24155697]
49. DeFelipe J. Cortical interneurons: from Cajal to 2001. *Progress in brain research*. 2002; 136:215–238. [PubMed: 12143384]
50. Jaitin DA, et al. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science*. 2014; 343:776–779. [PubMed: 24531970]

## Additional references

51. Raymond CS, Soriano P. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS one*. 2007; 2:e162. [PubMed: 17225864]
52. Rossi J, et al. Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell metabolism*. 2011; 13:195–204. [PubMed: 21284986]
53. Gerfen CR, Paletzki R, Heintz N. GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron*. 2013; 80:1368–1383. [PubMed: 24360541]
54. Franco SJ, et al. Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science*. 2012; 337:746–749. [PubMed: 22879516]

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*. 2006; 49:191–203. [PubMed: 16423694]
56. Madisen L, et al. Transgenic Mice for Intersectional Targeting of Neural Sensors and Effectors with High Specificity and Performance. *Neuron*. 2015; 85:942–958. [PubMed: 25741722]
57. Hippenmeyer S, et al. A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS biology*. 2005; 3:e159. [PubMed: 15836427]
58. Madisen L, et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature neuroscience*. 2010; 13:133–140. [PubMed: 20023653]
59. Vong L, et al. Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron*. 2011; 71:142–154. [PubMed: 21745644]
60. Tong Q, Ye CP, Jones JE, Elmquist JK, Lowell BB. Synaptic release of GABA by AgRP neurons is required for normal regulation of energy balance. *Nature neuroscience*. 2008; 11:998–1000. [PubMed: 19160495]
61. Sando R 3rd, et al. Inducible control of gene expression with destabilized Cre. *Nature methods*. 2013; 10:1085–1088. [PubMed: 24056874]
62. Hnasko TS, 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*. 2006; 103:8858–8863. [PubMed: 16723393]
63. Harris JA, Oh SW, Zeng H, Crawley, Jacqueline N., et al. Adeno-associated viral vectors for anterograde axonal tracing with fluorescent proteins in nontransgenic and cre driver mice. *Current protocols in neuroscience*. 2012; 20:21–18. Chapter 1, Unit 1.
64. Franklin, KBJaPG. *Mouse brain in stereotaxic coordinates*. Academic Press; 2008.
65. Hempel CM, Sugino K, Nelson SB. A manual method for the purification of fluorescently labeled neurons from the mammalian brain. *Nature protocols*. 2007; 2:2924–2929. [PubMed: 18007629]
66. Ramskold D, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature biotechnology*. 2012; 30:777–782.
67. Shalek AK, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*. 2014; 510:363–369. [PubMed: 24919153]
68. Treutlein B, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*. 2014; 509:371–375. [PubMed: 24739965]
69. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*. 2011; 12:323. [PubMed: 21816040]
70. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 2009; 10:R25. [PubMed: 19261174]
71. Brennecke P, et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nature methods*. 2013; 10:1093–1095. [PubMed: 24056876]
72. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome biology*. 2010; 11:R106. [PubMed: 20979621]
73. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008; 9:559. [PubMed: 19114008]
74. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015; 43:e47. [PubMed: 25605792]
75. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature methods*. 2010; 7:1009–1015. [PubMed: 21057496]
76. Thompson CL, et al. Genomic anatomy of the hippocampus. *Neuron*. 2008; 60:1010–1021. [PubMed: 19109908]
77. Peng H, Ruan Z, Long F, Simpson JH, Myers EW. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nature biotechnology*. 2010; 28:348–353.
78. Hu H, Cavendish JZ, 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*. 2013; 7:195. [PubMed: 24339803]

79. Rossier J, et al. Cortical fast-spiking parvalbumin interneurons enwrapped in the perineuronal net express the metallopeptidases Adamts8, Adamts15 and Neprilysin. *Molecular psychiatry*. 2015; 20:154–161. [PubMed: 25510509]
80. Nikkari ST, Jarvelainen HT, Wight TN, Ferguson M, Clowes AW. Smooth muscle cell expression of extracellular matrix genes after arterial injury. *The American journal of pathology*. 1994; 144:1348–1356. [PubMed: 8203472]

**Figure 1. Workflow overview**

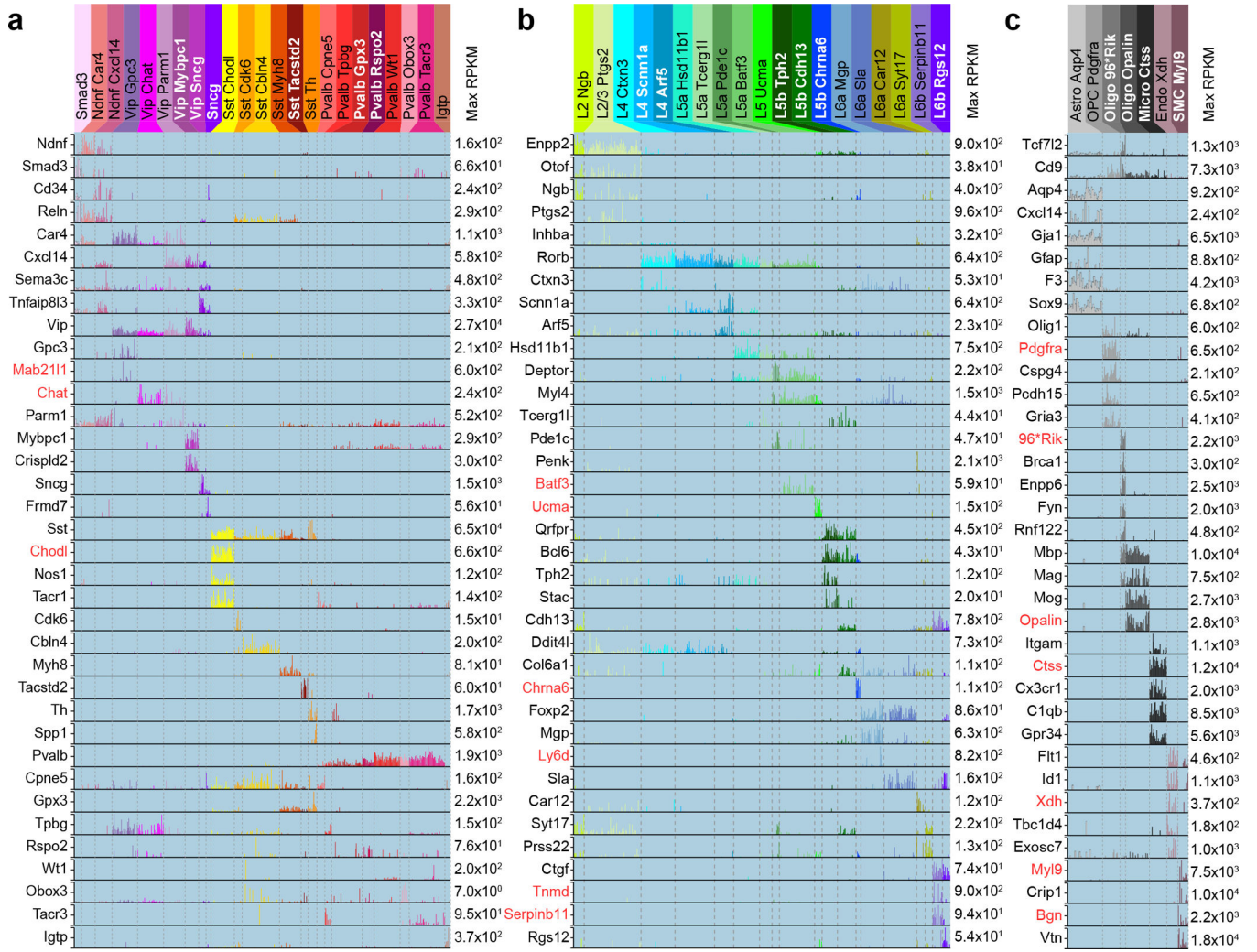
(a) Experimental workflow starts with the isolation, sectioning and micro-dissection of the primary visual cortex from a transgenic mouse. The tissue samples are converted into a single-cell suspension, single cells are isolated by FACS, poly(A)-RNA from each cell is reverse transcribed (RT), cDNA is amplified and fragmented, and sequenced on a next-generation sequencing (NGS) platform. (b) Analysis workflow starts 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 called ‘core’ cells (N = 1424), while cells that are mapped to more than one cluster are labeled ‘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 4 cells. (Supplementary Fig. 3, Methods). (c) The final 49 clusters were assigned an identity based on cell location (Fig. 2) and marker genes (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 (max. 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).



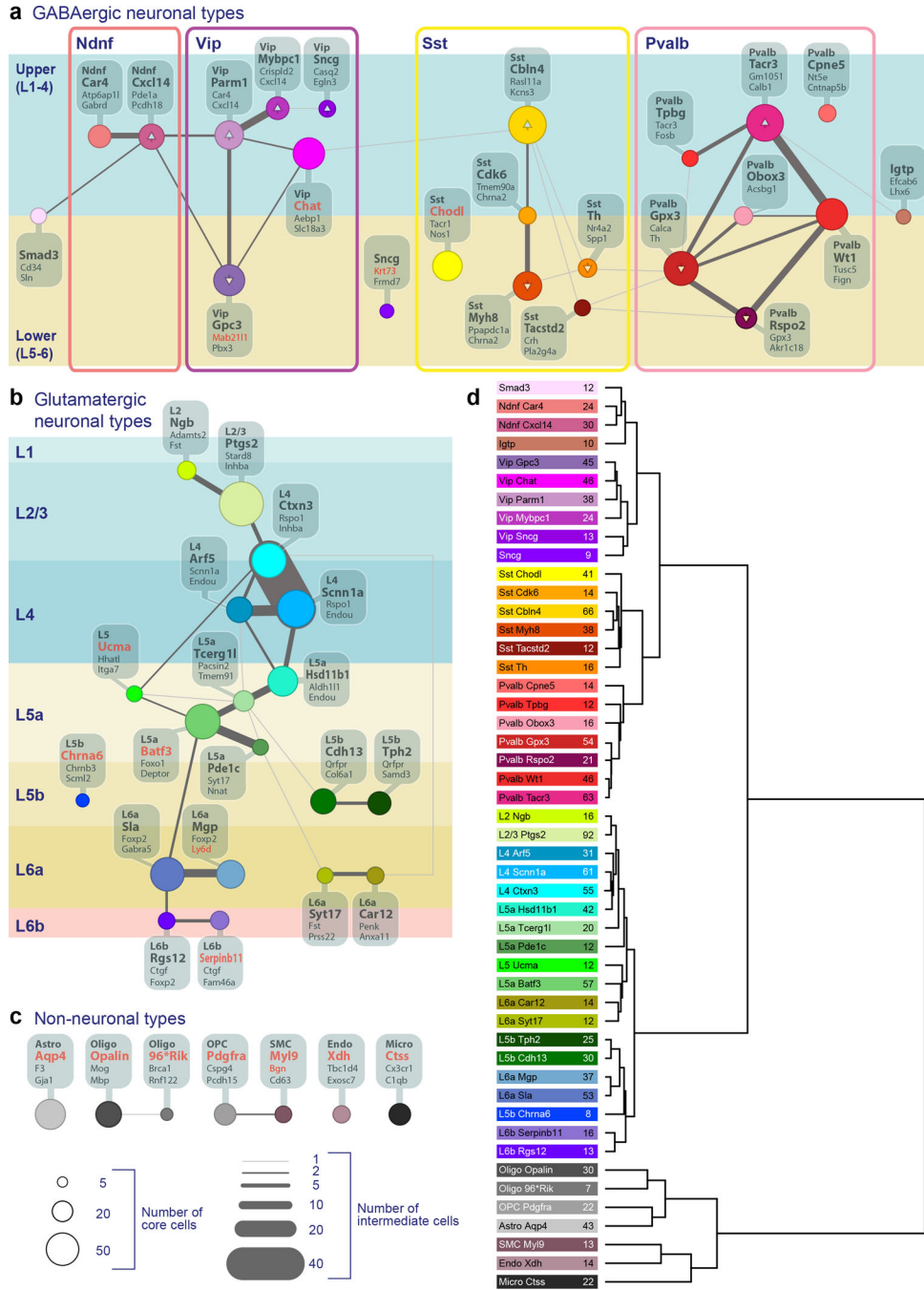


Scale bar in the last panel applies to all. **(b)** Cre line specificity characterized by transcriptomic cell types (N = 1424 core cells, 255 intermediate cells). The size of each black disk represents the proportion of cells classified as core in each transcriptomic type isolated from a particular Cre 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 core 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.



**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 (max. RPKM on the right). Only core cells are represented (N = 1424); for numbers of core cells per type see Fig. 2b, top. *Tacr1* encodes neurokinin-1 receptor or substance P receptor; *96\*Rik* is *9630013A20Rik*. Unique marker genes are in red.



**Figure 4. Cell types summary and relationships**  
**(a–c)** Constellation diagrams showing core and intermediate cells for all cell types. Core cells (N = 1424 total, 664 GABAergic, 609 glutamatergic, 151 non-neuronal) are represented by colored discs with area corresponding to the number of core cells for each cluster. Linked tags include cell type names based on marker genes/layers; unique markers are in pink. Intermediate cells (N = 255 total, 97 GABAergic, 155 glutamatergic, 3 non-neuronal) are represented by lines connecting discs; line thickness corresponds to the number of such cells. **(a)** GABAergic neuron types are grouped according to major classes

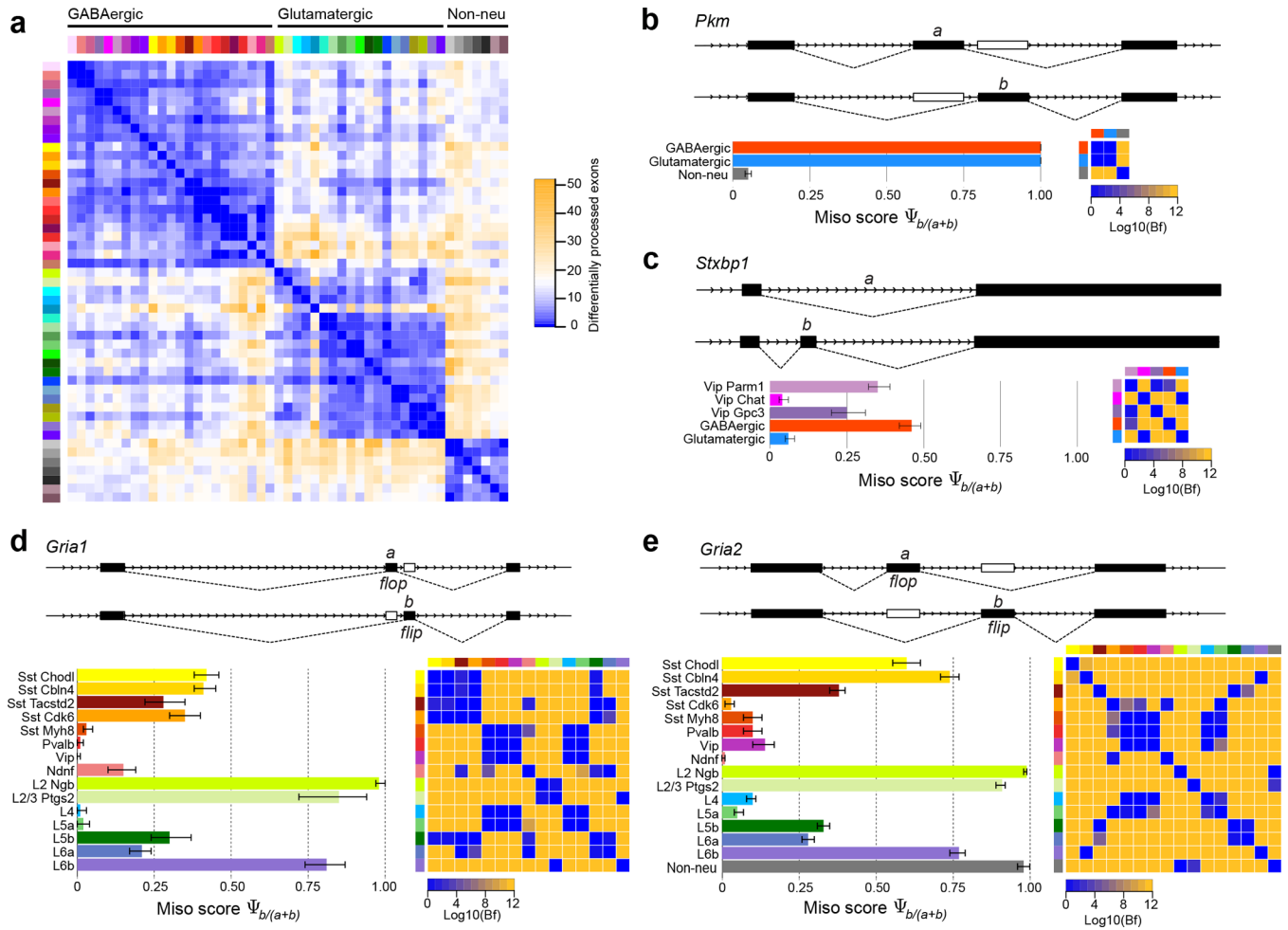
Author Manuscript

Author Manuscript

Author Manuscript

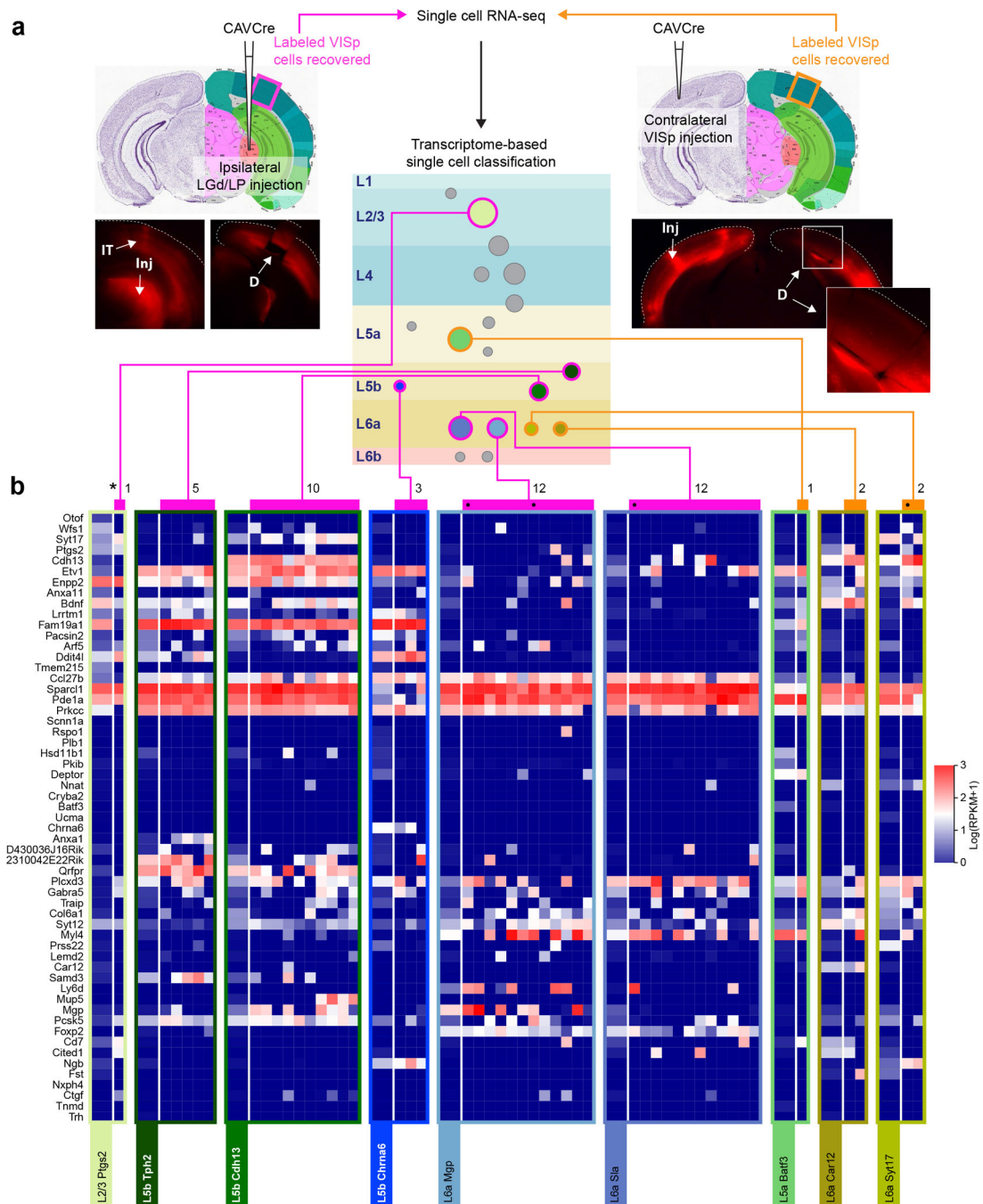
Author Manuscript

and arranged by their preferential location (enrichment) in upper vs. lower cortical layers. Up/down arrows within discs 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 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 = 1424) and genes (N = 13,878) with standard deviation for expression > 1 across all types. The distance metric is Pearson's correlation coefficient over the genes in the  $\log_{10}(\text{RPKM}+1)$  space. The tree was generated by standard hierarchical clustering with average linkage.



**Figure 5. Cell-type specific mRNA processing**

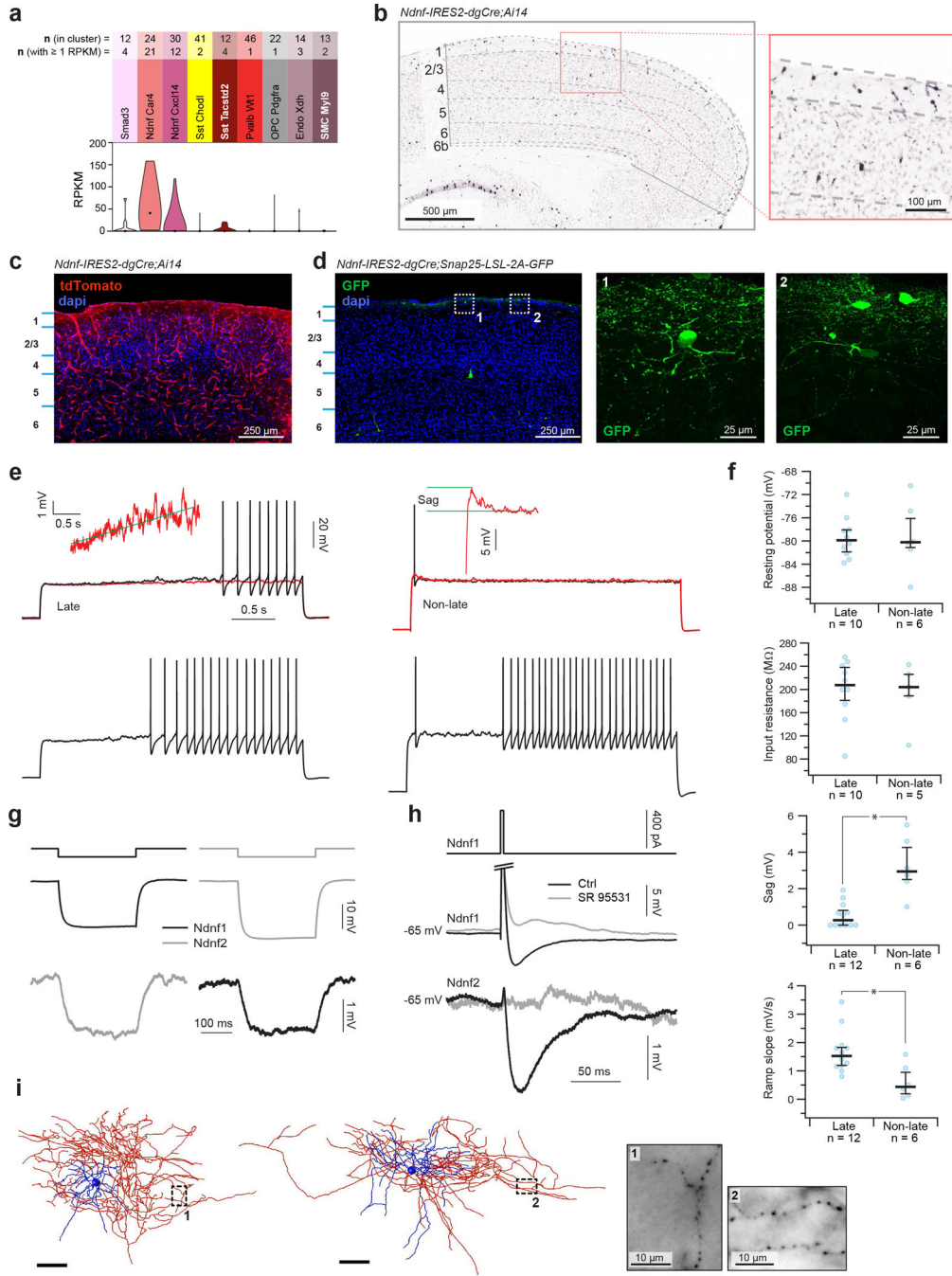
(a) Heat map showing the number of differentially processed exons (N = 567 out of 256,430 examined) for each pairwise comparison of transcriptomic cell types (Methods). (b–e) Confirmation of differential exon processing for four gene examples from (a) using MISO (Methods). Schematic of each gene (top) and corresponding quantitation (bottom). The MISO score ( $\Psi$ ), or “percent spliced-in”, represents the relative exon usage of transcript variant *b* vs. *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 = 10<sup>12</sup>. (b) In agreement with a population-level transcriptome profiling study<sup>13</sup>, pyruvate kinase (*Pkm*) mRNAs display differential exon usage among neurons and non-neuronal cells. (c) Syntaxin binding protein 1 (*Stxbp1*) mRNAs show differential processing among broad neuronal types, but also specific Vip types. (d,e) mRNAs for AMPA receptor genes, *Gria1* and *Gria2*, both display mutually exclusively spliced “flip” and “flop” exons. The two *Gria* genes show similar alternative exon usage within same cell types, suggesting a shared mechanism for alternative splicing. For simplicity, all genes are shown in the same orientation.



**Figure 6. Transcriptomic signatures and axonal projections**

(a) Schematic of experimental approach shows that canine adenovirus expressing Cre recombinase (CAVCre) was injected into two different VISp projection areas in *Ai14* mice: ipsilateral visual thalamus (LGd/LP) or contralateral visual cortex (VISp). TdT<sup>+</sup> 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 (Methods) to our previously determined transcriptomic cell types. **(b)** Heatmap panels show gene expression in individual projection-labeled cells classified into one of 9 (out of 49) previously determined transcriptomic types. Median gene expression in each type is shown to the left of each heatmap panel. Black dots indicate cells that are classified as intermediate, but are 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 labeled on top. Total cells: n = 43 for two thalamus injection experiments; n = 5 for one contralateral VISp injection experiment.



**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  $\geq 1$ . (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) Same as (b), but showing *tdT* protein fluorescence. (d) The endothelial cells can be avoided if pan-neuronal Cre reporter (*Snap25-LSL-2A-GFP*) is used instead of *Ai14* (Methods). Insets: putative neurogliaform cells. The images in (c) and (d) are each representative of two independent experiments. (e)



Intrinsic properties of tdT<sup>+</sup> 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 spikes earlier. Top right inset: magnified view of subthreshold initial response sag. The same neuron (bottom right) displays 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  $\pm$  SEM. LS neurons display 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<sup>+</sup> cells. Hyperpolarizing current injection (top) into either tdT<sup>+</sup> cell was transmitted to the other cell (bottom). 77% of cells were electrically coupled at an average intersomatic distance of  $110 \pm 11 \mu\text{m}$  and mean junctional conductance of  $181 \pm 41 \text{ pS}$ ,  $n = 14$ . Errors represent SEM. **(h)** Recording of a synaptically connected pair of tdT<sup>+</sup> cells. Action potential (middle, truncated) induced by a brief (3 ms) current injection (top) caused IPSPs in both neurons that were blocked by SR 95531 ( $5 \mu\text{M}$ ). IPSP mean 10–90% rise time =  $7.4 \pm 0.5 \text{ ms}$ ; IPSP mean tau decay =  $35.3 \pm 7.6 \text{ ms}$ ,  $n = 12$ . Errors represent SEM. **(i)** 3-D reconstructions of two biocytin-filled tdT<sup>+</sup> 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. Scale bar,  $50 \mu\text{m}$ . Insets: magnified views of bouton-like structures from original images. Morphological reconstruction of additional cells would be needed to assess if the ones presented here on the left and right are generally representative of the LS or NLS spiking cells, respectively.