bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A Single-Cell Atlas of Cell Types, States, and Other Transcriptional Patterns from Nine Regions of the Adult Mouse Brain

Arpiar Saunders[1,2,3], Evan Macosko[1,3,4], Alec Wysoker[2,3], Melissa Goldman[2,3], Fenna Krienen[2,3], Heather de Rivera[2,3], Elizabeth Bien[2,3], Matthew Baum[2,3], Shuyu Wang[3], Aleks Goeva[3], James Nemesh[2,3], Nolan Kamitaki[2,3], Sara Brumbaugh[2,3], David Kulp[2,3] and Steven A. McCarroll[2,3]

1. Equal Contribution 2. Department of Genetics, Harvard Medical School 3. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard 4. Department of Psychiatry, Massachusetts General Hospital

Correspondence to: Arpiar Saunders (asaunders@genetics.med.harvard.edu), Evan Macosko (emacosko@broadinstitute.org), and Steven A. McCarroll (mccarroll@genetics.med.harvard.edu)

The mammalian brain is composed of diverse, specialized cell populations, few of which we fully understand. To more systematically ascertain and learn from cellular specializations in the brain, we used Drop-seq to perform single-cell RNA sequencing of 690,000 cells sampled from nine regions of the adult mouse brain: frontal and posterior cortex (156,000 and 99,000 cells, respectively), hippocampus (113,000), thalamus (89,000), cerebellum (26,000), and all of the basal ganglia – the striatum (77,000), globus pallidus externus/nucleus basalis (66,000), entopeduncular/subthalamic nuclei (19,000), and the substantia nigra/ventral tegmental area (44,000). We developed computational approaches to distinguish biological from technical signals in single-cell data, then identified 565 transcriptionally distinct groups of cells, which we annotate and present through interactive online software we developed for visualizing and re-analyzing these data (DropViz). Comparison of cell classes and types across regions revealed features of brain organization. These included a neuronal gene-expression module for synthesizing axonal and presynaptic components; widely shared patterns in the combinatorial co-deployment of voltage-gated ion channels by diverse neuronal populations; functional distinctions among cells of the brain vasculature; and specialization of glutamatergic neurons across cortical regions to a degree not observed in other neuronal or non-neuronal populations. We describe systematic neuronal classifications for two complex, understudied regions of the basal ganglia, the globus pallidus externus and substantia nigra reticulata. In the striatum, where neuron types have been intensely researched, our data reveal a previously undescribed population of striatal spiny projection neurons (SPNs) comprising 4% of SPNs. The adult mouse brain cell atlas can serve as a reference for analyses of development, disease, and evolution.

Cellular specialization is central to the function of the mammalian brain. At the coarsest level, cells of different classes (for example, neurons, astrocytes, and endothelial cells) interact to maintain homeostasis and enable electrochemical communication. At the finest level, subtle specializations – such as those that distinguish neuron subtypes in the same brain region – can control behaviors such as appetite (Andermann and Lowell, 2017; Sternson, 2013), sex drive (Anderson, 2012), habit formation (O'Hare et al., 2016; Wang et al., 2011), spatial mapping (Moser et al., 2008), and associative learning (Krabbe et al., 2017; Letzkus et al., 2015). Although some of these cell populations have been characterized in detail, many more remain to be characterized or even discovered for the first time.

Systematic efforts to identify cell populations, reveal the RNA repertoires of every cell type and state, and identify markers for these populations, would help to understand the functions and interactions of cells in the brain. Comprehensive datasets would also enable new ways of probing the roles of distinct cell types in circuit function, behavior, and disease. High-throughput single-cell RNA-seq (scRNA-seq) now makes it possible to profile RNA expression in thousands of individual cells in complex tissue (Klein et al., 2015; Macosko et al., 2015; Rosenberg et al., 2018; Zheng et al., 2017). To date, single-cell gene expression studies have yielded cell-type classifications in the mouse cerebral cortex (Tasic et al., 2016a; Zeisel et al., 2015), retina (Shekhar et al., 2016), hypothalamic arcuate nucleus (Campbell et al., 2017), and amygdala (Wu et al., 2017).

Here we sought to analyze cellular diversity across a

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

wide variety of brain regions in a uniform manner that would make it possible to learn from shared and region-specific patterns in cellular composition and gene expression. To do this, we had to overcome several challenges. First, dissociating adult mammalian brain tissue into healthy and representative cell suspensions is difficult; many scRNA-seq studies have therefore used developing mice (younger than 30 days old). However, the strong transcriptional programs associated with development can obfuscate gene-expression differences underlying the functional specializations of cell subtypes. Here we developed approaches, borrowing from techniques used for electrophysiological recordings, that allowed adult tissue to be dissociated into intact cell soma while representing all major cell classes. Second, scRNA-seq data are simultaneously shaped by real underlying cellular categories, continuously varying signals, and technical artifacts; the grouping of cells into discrete clusters often reflects unknown and non-transparent combinations of these effects. Here we developed analytical methods to dissect biological and technical influences on single-cell data sets and enable a more transparent view and revision of the resulting cellular classifications.

Here, we describe an early mouse brain cell atlas, which we created by analyzing (using Drop-seq) 690,000 individual cells from nine regions of the adult mouse brain. By comparing transcriptional patterns from within and across neuron types, we identified and validated a widely-used transcriptional state that supports axon and synapse function and discovered large-scale structure in the expression of ion channels that enable and constrain electrophysiological properties. We determined that glutamatergic neurons tend to be specialized by region in cortex, while non-neuron cell classes, such as those that make up the vasculature, can be variably specialized across cortical and subcortical areas. We also highlight the depth of individual datasets through the classification of neuron types. While we present in-depth analyses of all nine regions through DropViz, here we illustrate our insights into neuronal diversity using examples from the basal ganglia. In the globus pallidus externus (GPe) and substantia nigra reticulata (SNr), two regions where neuron types are not well-understood, we propose complete neuron-type classifications and identify selective markers for each population. In the striatum, where neuronal diversity is well charted, our data rediscovered, then built upon, accepted molecular distinctions, identifying a novel group of principal neurons that have been overlooked despite decades of research.

Our hope is that these data will advance a wide variety of efforts and nominate many unforeseen research questions. To facilitate the exploration and utilization of these data, we developed a flexible, interactive analysis platform ("DropViz", http://dropviz.org/) for comparing cell types to one another, identifying cell populations that express genes of interest, and performing many other kinds of analyses.

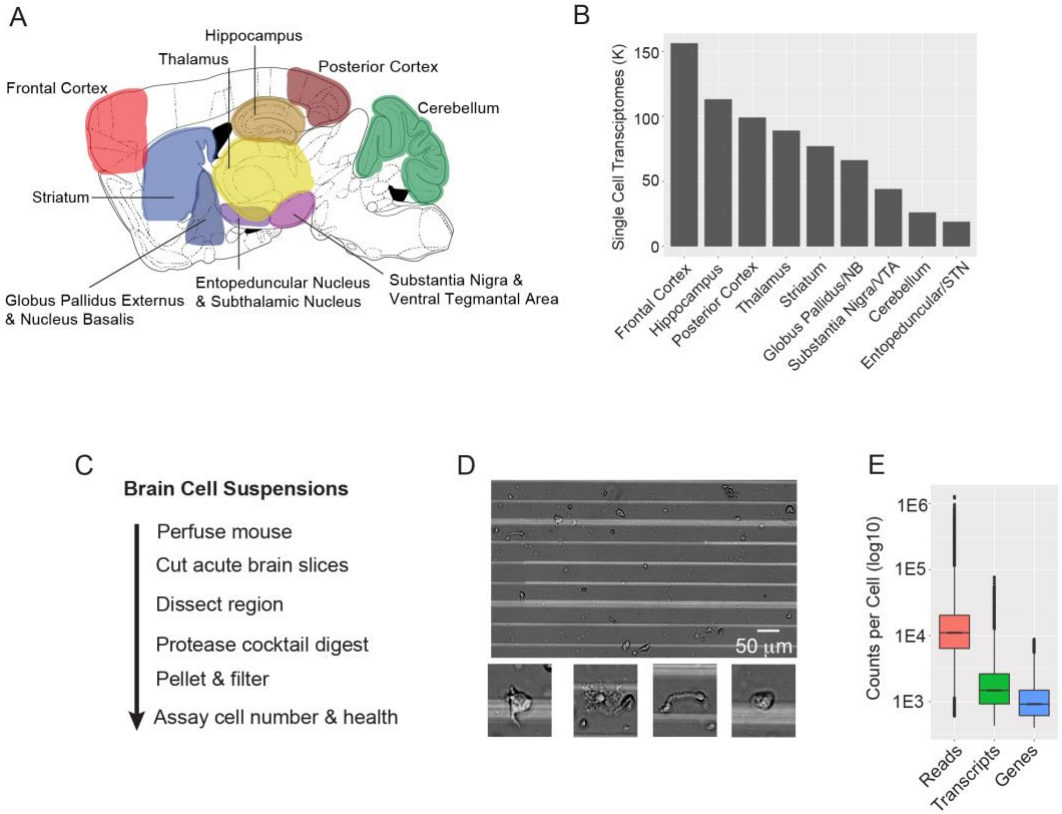


Figure 1. Single-cell transcriptional profiling of the adult mouse brain using Drop-seq
(A) Sagittal brain schematic illustrating the n=9 profiled regions. Detailed schematics are presented in **Figure S1**.
(B) Plot illustrating the number of single-cell transcriptomes per region.
(C) Flowchart describing how acute single-cell suspensions are prepared from adult mouse brain.
(D) Example single-cell suspension generated from frontal cortex that contained little debris and diverse cellular morphologies. Bottom, enlargements of example cells.
(E) Distribution of sequencing reads, transcripts, and genes for all cells (n=690,207)

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

RESULTS

Isolation and Molecular Analysis of Cells for an Adult Brain Cell Atlas

To build an atlas of cell populations and cell-type-specific gene expression across many regions of the adult mouse brain, we prepared single-cell suspensions from nine brain regions (**Table S1**) and used Drop-seq (Macosko et al., 2015) to profile the RNA expression of 690,207 individual cells (**Figure 1A,B**). We profiled cells from the Frontal and Posterior Cerebral Cortex (FC and PC), Hippocampus (HP), Thalamus (TH), Cerebellum (CB), and the Basal Ganglia (consisting of the Striatum [STR], Globus Pallidus Externus / Nucleus Basalis [GP/NB], Entopeduncular Nucleus / Subthalamic Nucleus [EP/STN] (Wallace et al., 2017), and Substantia Nigra/Ventral Tegmental Area [SN/VTA]) (**Figure S1**).

We generated cell suspensions from adult male mice (60–70 days old; C57Blk6/N) by adapting protocols from single-cell patch-clamp recording (Carter and Bean, 2009). We optimized the digestion time for each region to maximize cell recovery rates (see **Methods**, **Figure 1C**). Tissue was prepared in a buffer lacking Ca^{2+} with ionic concentrations designed to avoid activity-induced toxicity by maintaining voltage-gated Na channels in an inactivated state [Estimated V_m = -30.5 mV, **Methods**]). The resulting cell suspensions, which recovered intact 40–50% of cells from most tissues (cortex: 0.46 ± 19 mean ± sem; striatum: 0.39 ± 20, **Figure S4A-C**), had cells with the characteristic morphologies of neurons, astrocytes, and oligodendrocytes, and were largely free of debris (**Figure 1D**). We prepared 3-7 replicate suspensions of each brain region (44 total).

We performed Drop-seq analysis of 690,207 individual cells from these 44 cell suspensions (Macosko et al., 2015). We generated and analyzed 13 billion sequencing reads from the resulting Drop-seq libraries, ascertaining 1.45 billion distinct mRNA transcripts (UMIs), which arose from 31,767 distinct genes. We ascertained an average of 17,480 reads (median = 10,824), 2,218 mRNA transcripts (median= 1,450 UMIs), and 1,169 genes per cell (median = 900) (**Figure 1E**).

Cell-class Composition of Nine Adult Brain Regions

We first separately analyzed data for each of the nine brain regions, analyzing each region in two stages (here termed "global clustering" and "subclustering"). The first round of analysis, which is relatively straightforward (**Methods**), grouped cells into 8-11 broad classes that were then easily identified by known marker genes; these broad classes included neurons, astrocytes, microglia/macrophages, oligodendrocytes, polydendrocytes (oligodendrocyte progenitor cells), and components of the vasculature – the endothelial stalk, endothelial tip, and mural cells (Abbott et al., 2006; Wälchli et al., 2015a). For example, the hippocampus yielded cells from all 11 cell classes (**Figure 2A-C**), including three local cell classes, of which two are native to the ventricle – the choroid plexus and ependymal cells, which had been sampled from ventricular tissue proximal to the hippocampal dissections, and an additional cell class undergoing adult neurogenesis (Habib et al., 2016; Hochgerner et al., 2018; Ming and Song, 2011). Assignment of cells to these broad classes was robust to analysis parameters (**Figure S2**).

Distinct brain regions yielded the major cell classes in different proportions. (**Figure 2D**). The proportion of neurons varied inversely with those of oligodendrocytes, endothelial stalk cells, and mural cells (**Figure 2D**); a natural explanation for this is the regional variation in the abundance of grey matter, with oligodendrocytes (as expected), endothelial stalk cells, and mural cells being more abundant in white matter than in grey matter. Endothelial tip cells comprised a similar proportion of all cells in each region, consistent with similar degrees of vascular sprouting throughout the brain. Polydendrocyte and astrocyte abundance appeared to vary independently of other cell classes, exhibiting enrichment in the GPe as expected based on earlier findings (Cui et al., 2016; Lange et al., 1976). Small fractions of choroid plexus and ependymal cells were sampled from ventricle-adjacent regions, while sparse neurogenic populations were observed in regions adjacent to the subventricular and subgranular zones (frontal cortex, striatum, and hippocampus) (Ming and Song, 2011).

Inference of Cell Types and States Using Independent Components Analysis

The systematic recognition of more-subtle (yet consistent) patterns of variation among cells of the same class presents formidable analytical challenges – particularly when these distinctions are to be made in systematic, inductive (unsupervised), data-driven ways (Mayer et al., 2015; Satija et al., 2015; Shekhar et al., 2016; Tanay and Regev, 2017; Tasic et al., 2016b). The size and diversity of the data sets for these nine brain regions, and the generation of many biological replicates for each region, made apparent many limitations in diverse analysis methods. We initially found clusters that were specific to experimental replicates or that were defined by transcriptional signatures that we recognized as artifacts of tissue digestion (see below). We sought an analysis strategy that would make it possible to (i) dissect biological from technical

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

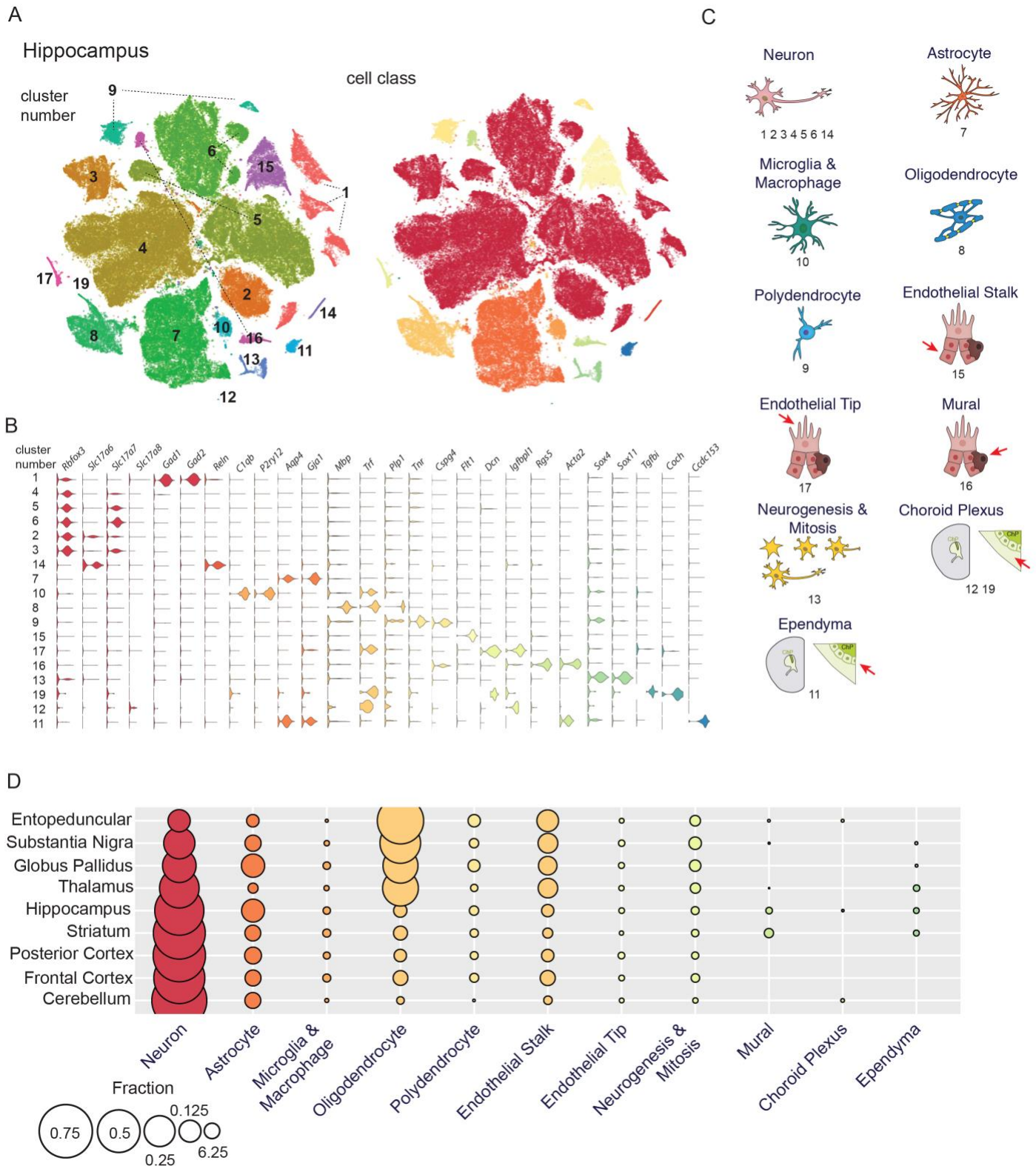


Figure 2. Comprehensive identification of brain cell classes

(A) T-distributed stochastic neighbor embedding (tSNE) plot of gene expression relationships amongst the n=113,171 cell hippocampus dataset based on the first round of ICA-based clustering. The resulting "global" clusters (colored independently and numbered 1-19) are each associated with major cell classes of the brain. Right, global clusters color-coded by cell class.

(B) Violin plots showing example gene markers that distinguish across and within brain cell classes (log10). Genes are color-coded by cell class. Global clusters are ordered by cell class.

(C) Cartoons of the major cell classes of the brain. Numbers below indicate the corresponding global clusters from hippocampus.

(D) Dot plots displaying the proportional representation of individual cell classes across regions.

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

contributions to the expression data, so that technical contributions could be excluded from subclustering analysis; and (ii) generate transparent and meaningful intermediate data sets (downstream of input expression data, and upstream of clustering outcomes) that could be critically evaluated – and, ideally, be meaningful objects of analysis in their own right.

We therefore developed an analysis strategy based on independent components analysis (ICA)(**Figure 3A,B**). ICA reduces large datasets to a smaller number of dimensions in which entities (here cells) have score distributions that are maximally structured – as measured by deviation from a normal distribution (generally due to a spikey or clustered distribution of the cells in that dimension) – and statistically independent (Hyvärinen, 1999). Each of the inferred independent components (ICs) is a weighted combination of many genes (the weight of each gene's contribution to an IC is the gene "loading"). Each cell's gene-expression profile is reduced to a weighted sum of these ICs, with each IC having a score for each cell that reflects the expression of that combination of genes in that cell.

We found that many individual ICs corresponded to easily recognized biological phenomena (**Figure 3C-F**)(Adamson et al., 2016). For example, among cells from frontal cortex cluster 6 (glutamatergic neurons from layer 2/3 and 5a), we identified ICs whose strongly loading genes were readily recognized as known markers for specific cell types, cell states or spatial gradients across anatomical axes (**Figure 3F**). Other ICs exhibited cell-loading distributions that immediately suggested technical effects, such as ICs that distinguished 1) cells (of all classes) from different replicate preparations from the same brain region; 2) single-cell RNA libraries of different sizes; 3) experimentally-identified effects of tissue preparation (**Figure S3C-E**); or 4) cell-cell "doublets" (**Figure 3C-E**). (By contrast, in principal components analysis (PCA), which optimizes components to maximize variance without regard to statistical independence or data distribution, distinct biological signals are typically distributed across many components and mixed with technical sources of noise (**Figure S3A**); PCs still carry maximal information forward to downstream clustering algorithms while reducing data dimensionality, but are not intended to parse statistically distinct signals into separate components). We found that the interpretability of individual ICs allowed us to distinguish ICs corresponding to presumed endogenous signals (which we call here "biological ICs") from ICs related to technical signals described above. Moreover, by removing contributions from artifact ICs, we reduced spurious noise (**Figure 3G**).

We analyzed the data from each class and brain region (109 analyses total, **Table S9**) using a semi-supervised ICA, in which we excluded 1,157 ICs as recognizably technical in origin in that they corresponded doublets of cells from different classes (n=759), individual outlier cells (n=99), or replicate-specific or tissue preparation artifacts (n= 299) (**Table S11**). We classified the remaining 601 ICs as "biological ICs", though it remains possible that additional technical influences on the data were not recognized during curation. We then clustered the cells for each of the 109 brain-region-specific, cell-class-specific analyses based on their combination of loadings for the "biological ICs" in the respective analysis.

To group cells into subclusters, we used network-based clustering (Waltman and van Eck, 2013), which has been utilized in earlier studies (Shekhar et al., 2016). For each subcluster analysis, we used only the putative "biological ICs" as dimensions for graph construction. To create a transparent relationship between biological ICs and subclusters, we utilized clustering-resolution parameters that maximized one-to-one relationships between subcluster assignments and underlying ICs (**Figure 3H** and **Figures S3E,F**). The 109 separate subclustering analyses of the cell classes from each of the nine regions identified 565 total subclusters **(Figure 4A** and **Table S10).** Of these subclusters, 323 were neuronal and were derived from 368 biological ICs (**Figure S4).** Hierarchical clustering of the expression profiles for subclusters (aggregated across their constituent cells) demonstrated that clustering is driven primarily by cell class and not by brain region (**Figure 4B**).

Characteristics of the Cells of the Blood-Brain Barrier

Non-neuronal cells, including glia and endothelial cells, exhibited broadly consistent expression signals across brain regions. To better appreciate diversity among non-neuronal classes, we grouped single-cell libraries from each region by cell class and performed semi-supervised ICA on each of the seven datasets independently, identifying a total of 53 biological ICs (**Figure 5A** and **Figure S5A-F**). We focus here on cell classes that form the blood-brain barrier – mural, endothelial tip and endothelial stalk cells – because they are disease-relevant (Sweeney et al., 2016) and remain incompletely characterized (**Figure 5B**).

Mural cells are morphologically diverse cells embedded within the basement membrane of the endothelium. Mural cells control vascular development, stability, and homeostasis, and exhibit pathology in many disease states (Sweeney et al., 2016; Trost et al., 2016). We identified 7 mural cell subclusters from 7 biological ICs (n=4,713 cells, **Figure 5C and Figure S5E**). Mural cells are known to have two subtypes:

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

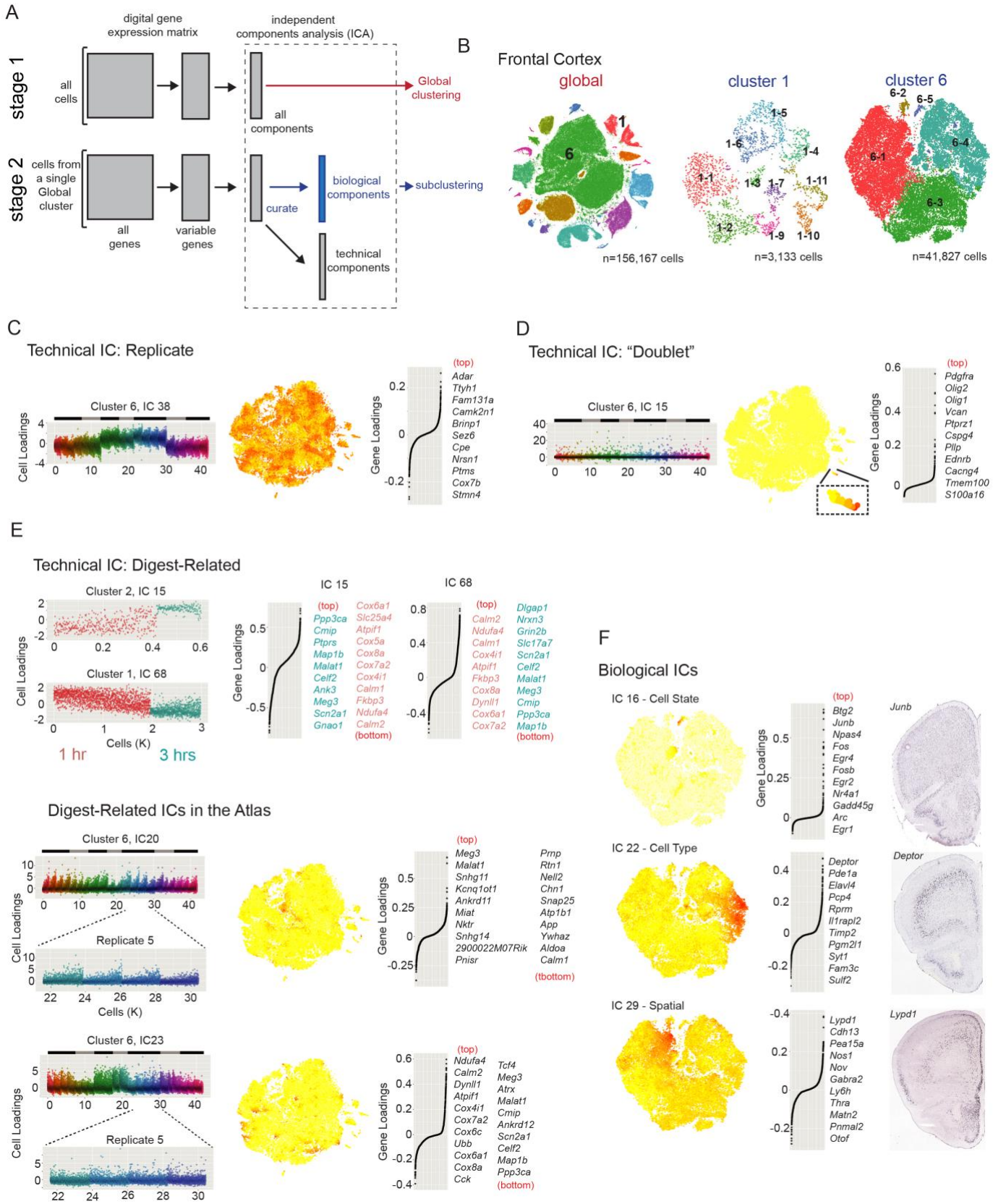


Figure 3 continued…

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

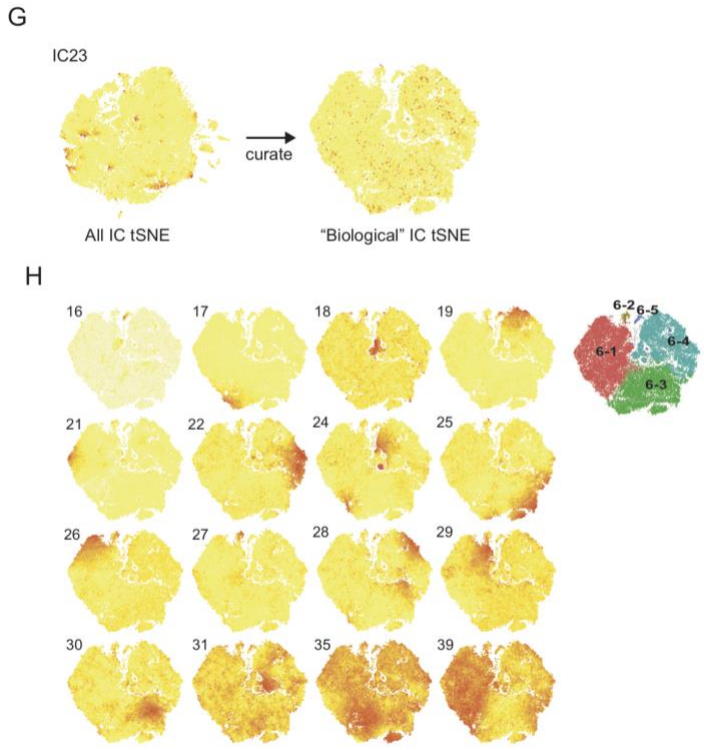


Figure 3. Independent component analysis identifies transcriptional programs allowing transparent description of within-cell class molecular diversity.

(A) Workflow for semi-supervised two-stage ICA-based signal extraction and clustering. In the first stage ("global clustering", red arrows), the full digital gene expression (DGE) is clustered into "global" clusters using a reduced space generated by ICA (Methods). These "global" clusters represent cell classes. The process is then repeated independently for each of the global clusters with >200 cells ("subclustering", blue arrows), except that the resulting Independent Components (ICs) are manually curated into two categories: "technical" components representing signals associated with the 1) replicate (batch effects); 2) libraries containing two cells ("cell-cell doublets") or single-cells with unique transcriptional signatures ("outliers") or 3) the preparation process for acute tissue. "Biological" signals originate from cell types, cell states, or from spatial anatomical relationships. Only "Biological" ICs are used as input to subclustering. Subclustering is performed over a range of neighborhood sizes and resolutions, generating a portfolio of potential solutions. The single solution with the clearest correspondence between Biological IC cell loading and subclusters is selected (Figure S3).

(B) Color-coded tSNE plots for frontal cortex global clustering (left) and two representative subclusterings, GABAergic interneurons (middle, cluster 1) and glutamatergic layer 2/3 and some layer 5 neurons (right, cluster 6).

(C-E) Examples of "technical" ICs from frontal cortex.

(C) Example of an independent component (IC 38, cluster 6) representing a replicate-based signal. Left, plot of cell-loading scores (y-axis). On the x-axis, cells are ordered by library size (largest to smallest) within each sequencing pool (color-coded). Sequencing pools from the same mouse are grouped with black and gray bars above. Middle, cell-loading scores plotted on the subcluster 6 tSNE (tSNE constructed from all ICs). Higher loadings are shown in darker red. Right, gene-loading plot. The n=10 genes with the highest scores are shown at right.

(D) Example of a component (IC 15, cluster 6) representing a "doublet" signal. Plots and layout are similar to C. IC 15 loads heavily on a small number of cells across mice and sequencing pools. The top loading genes are markers of the polydendrocyte cell class, suggesting a signal that represents a "doublet" transcriptome consisting of layer 2/3 neurons and polydendrocytes. Inset shows area of the tSNE plot with high IC 15 loading.

(E) Experimental identification of protease digest-related technical ICs. Top, frontal cortex tissue was under- (1 hr) or over- (3 hr) digested before making single-cell suspensions and Drop-seq libraries. The datasets were co-analyzed. Subcluster IC loadings that correlate with digest time were identified (Figure S3C-E). Example digest-related ICs (IC 15, cluster 2, Syt6+ deep-layer glutamatergic) and (IC 68, cluster 1, Nptxr+ upper-layer glutamatergic) show contributions from similar genes. Genes related to ATP synthesis and Calmodulins (eg., Cox6a1, Ndufa4, Calm2) load on the under-digested cells, while another set of genes including the nuclear-enriched Meg3 and Malat1 load highly onto the over-digested condition. Bottom, similar digest-related ICs were commonly observed in atlas subclustering analyses, suggesting heterogeneity in cellular transcriptional response to identical preparation conditions. Two examples of digest-related ICs from frontal cortex cluster 6 (20 and 23). Left, cell-loading plots and insets highlight the correlation of IC loading to library size. IC 20 tends to load on smallest libraries, while IC 23 loads on the largest. Middle, cell-loadings for IC 20 and 23 demonstrate that digest-related signals contribute to local relationships within the subcluster 6 tSNE. Right, gene-loading plots for the 10 top and bottom loading genes indicate a similar signal to that identified in over and under-digested experiments.

(F) Examples of heterogeneous "Biological" ICs from frontal cortex cluster 6, representing a cell state (top, IC 16), cell type (middle, IC 22), and spatial anatomical signal (bottom, IC 29). For each example, a cell-loading tSNE plot, gene loading plot, and in situ hybridization experiment for a top-loading gene are shown from left to right. IC 16 corresponds to the immediate early gene (IEG) signal. The IC 22 signal originates from layer 5a glutamatergic neurons, exemplified by the spatially restricted expression of top-loading Deptor. IC 29 represents a spatial signal. Top-loading genes such as Lypd1 exhibit an expression gradient from medial (highest) to lateral (lowest). ISH data from the Allen Institute Mouse Atlas.

(G) Removing technical ICs prevents spurious transcriptional similarities. Example tSNE cell-loading plot for cluster 6 digest-related IC 23 before and after IC curation. Before curation, cells with high loading of digest-related IC 23 (E, bottom left) are locally grouped. After curation, technical ICs (including IC 23) are removed, creating a different tSNE plot and preventing the digest-related effects from contributing to clustering.

(H) Correspondence between heterogeneous transcriptional signals (Biological ICs) and subclusters identified by modularity-based clustering (Methods). tSNE cell loading plots for each of the n=16 Biological ICs from frontal cortex cluster 6. Top right, the resulting plot in which the n=5 subclusters are identified. The portfolio of alternative subcluster solutions is shown in Figure S3G.

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

pericytes, which associate with small-diameter capillaries and are enriched for *Rgs5* expression, and smooth muscle alpha actin (SMA) cells, which associate with larger-bore arterioles and express *Acta2* (Bondjers, 2006; He et al., 2016; Hughes and Chan-Ling, 2004; Nehls and Drenckhahn, 1991). One independent component (IC 19) appeared to correspond to this distinction by having *Rgs5* and *Acta2* as the most strongly weighted genes (**Figure 5C**). Notably, the data suggested that the pericyte-SMA distinction may be continuous rather than categorical, as the expression levels of *Rgs5* and *Acta2* were graded and the cell scores for this IC were continuously rather than bimodally distributed across these cells(**Figure 5C**)(Trost et al., 2013). (For those downstream analyses that require comparisons of discrete cell populations, we discretized this signal into three subclusters (2,4,6), but "clusters" cannot be reflexively equated with "discrete cell types".) Interestingly, the expression of *Rgs5* and *Acta2* was not always inversely correlated: a subpopulation of mural cells (cluster 1) expressed both *Rgs5* and *Acta2*, and also uniquely expressed genes such as *Aldh1a1* (**Figure S5E**). A population of *Rgs5*$^{+}$ mural cells (cluster 3) appeared to potentially have a specialized function, as they expressed genes for a potassium channel potently activated by intracellular diphosphate levels (*Kcnj8* and *Abcc9*) and an ADP-ribosyltransferase (*Art3*) – indicating they have signaling machinery that couples dinucleotide metabolites to membrane potential and post-translational modification.

Dynamic remodeling of the vasculature occurs in the postnatal brain, yet the cellular and molecular mechanisms are poorly understood (Wälchli et al., 2015b). One striking morphological example is sprouting angiogenesis, during which endothelial stalk cells transform into endothelial tip cells, which then dynamically grow and fuse to build new vascular networks (Wälchli et al., 2015a). We found 7 subclusters each of endothelial stalk (n=16,248) and tip (n=1,587) cells (**Figure 5D,E**). Among the tip cells, two clusters (4 and 5) were notable for their expression of genes encoding a variety of membrane transporters (e.g., *Slc38a2, Slc4a10*, *Slc26a2*, and *Slc47a1*) and pumps (e.g., *Fxyd5* and *Atp1b1*) (**Figure 5D**). To varying extents, clusters 1, 2, and 3 expressed genes involved with extracellular matrix (ECM) secretion, angiogenesis, and contraction (**Figure S5F**), such as the basement membrane collagen genes (*Col4a1*, *Col4a2*, and *Col15a1*) (**Figure 5D**). Interestingly, cluster 3 expressed a distinct set of collagen genes (*Col1a1* and *Col3a1*) (**Figure S5F**). These examples suggest that subsets of tip cells are transcriptionally specialized to affect membrane transport and ECM production; ECM secretion may involve coordinated expression of distinct sets of collagen genes.

Among endothelial stalk cells (**Figure 5E**), one subpopulation expressed genes related to host immunity; these included genes encoding interferons (*Ifit3*, *Ifit1*, *Ifit2,* and *Ifitm3*), GTPases induced by interferons (*Ligp1*, *Irmg2,* and *Gbp7*), and other proteins involved in the anti-viral response (*Isg15* and *Rsad2*). Cluster 4 had the highest levels of expression of these genes, including *Isg15* and *Ifitm3*. A distinct subpopulation (cluster 6) expressed genes related to growth-factor dependent vascular remodeling (*Mgp*, *Fbln5*, *Cytl1*, *Eln*, *Igfbp4*, and *Clu*) (Boström et al., 2004; Contois et al., 2012; Fitch et al., 2004; Fu et al., 2013; Guadall et al., 2011; Karnik et al.), suggesting that they may be specialized to coordinate a component of angiogenesis. This phenotype appeared to continuous rather than categorical in distribution across cells: While *Cytl1* was expressed exclusively in cluster 6, other genes, such as *Mgp*, were expressed at lower levels in other endothelial stalk cells (**Figure 5E**). These examples define transcriptional specializations in stalk cells related to host defense and remodeling of the adult vasculature. Other groups might reflect cell states specialized for iron handling, calcium signaling, and the stress response (**Figure S5G**).

Functional specializations within endothelial, glial, and other non-neuronal cell classes could be a ubiquitous feature of the adult mouse brain, or could be enriched in particular brain regions. We compared the abundance of endothelial stalk and tip cells from each brain region within each of the clusters (**Figure 5H**) (results for astrocytes, oligodendrocytes, polydendrocytes,

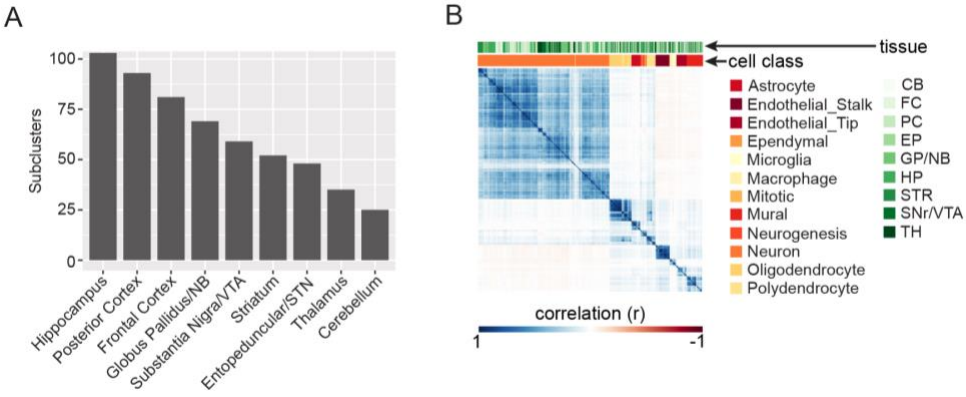


Figure 4. Transcriptional diversity by region and cell class
(A) Number of subclusters by region.
(B) Transcriptional correlations across atlas subclusters are largely explained by cell class and not region of origin. Hierarchical clustering diagram showing pairwise Pearson correlation scores calculated pairwise between average atlas subclusters. The analysis was restricted to genes with significantly variable expression (**Methods**). Color-coded bars at the top of the plot display the ordered region/cell class assignments for the subcluster.

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

microglia/macrophages and mural cells are in **Figure S5H**). While stalk cell subpopulations appeared in similar relative abundance across regions, tip cell subpopulations exhibited different contributions from cortical and subcortical areas: for example, the cortex and hippocampus contributed disproportionately to the stalk cell population that more strongly expressed genes with membrane-transport functions (cluster 6), while collagen-expressing cell populations (clusters 2 and 3) came largely from the basal ganglia and thalamus (**Figure 5H**).

A

Non-Neuronal Cell Class	# Biological ICs
Astrocytes	8
Oligodendrocytes	6
Microglia/Macrophage	3
Polydendrocytes	15
Endothelial Stalk	8
Endothelial Tip	6
Mural	7
Total	**53**

B

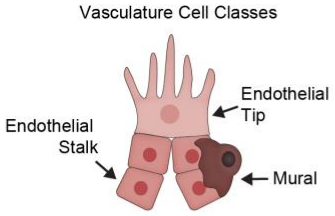
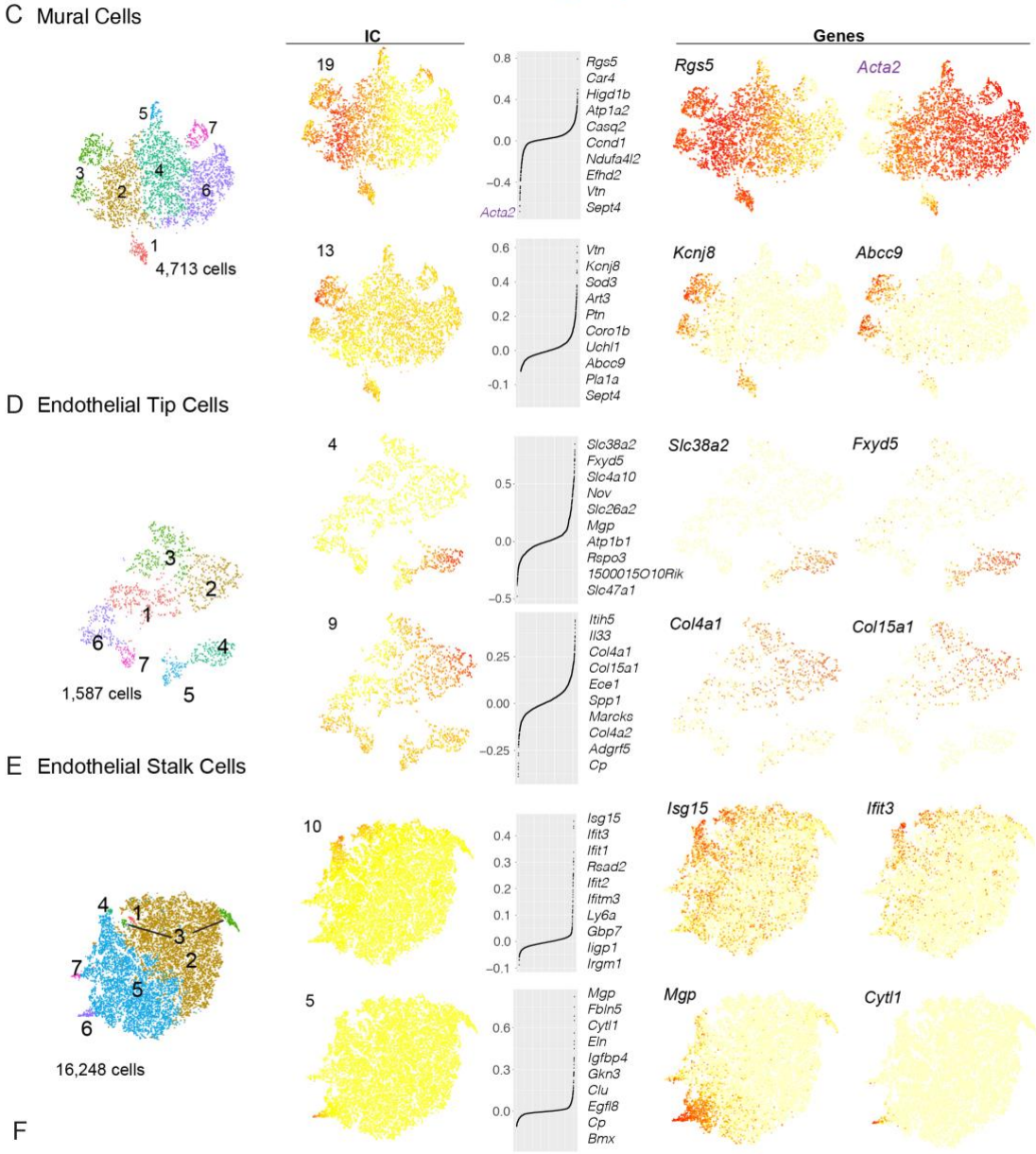
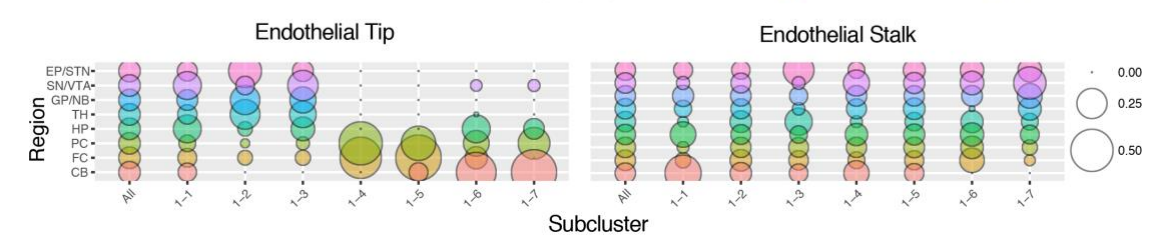


Figure 5. Comprehensive description of transcriptional diversity within non-neurons as illustrated by cell classes of the vasculature.
(A) Table showing the number of biological ICs curated for each non-neuronal cell class. A description of non-neuronal ICs is provided in **Figure S5**.
(B) Cartoon of vasculature cell classes.



(C-E) Subcluster assignments and examples of two biological ICs for each vasculature cell class. Subclusters (color-coded), IC cell-loadings, and gene expression values displayed on tSNE plots. Left, subcluster assignments. Middle, IC cell- and gene-loadings. For each IC, the top ten loading genes are listed. Right, expression plots for individual genes. For Mural Cell IC 19, the bottom loading gene *Acta2* is shown in purple.
(F) Dot plots illustrating fractional representation of cells from each region contributing to each particular endothelial tip and stalk subcluster. Similar plots for other non-neuronal cell classes are shown **Figure S5H**.



bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A Neuronal Transcriptional Program Related to Axon Function

Cell states involve constellations of genes that are transiently co-expressed to enact cellular functions.

In neurons, the most well-studied cell state, involving the expression of the immediate early genes (IEGs), is activated by Ca^{2+} influx after neuronal action potential firing (Bading, 2013; Hrvatin et al., 2017). IEGs (such as *Fos* and *Junb*) encode transcription factors that orchestrate a cellular response to neuronal activity. IEG expression in response to action potential firing is known to be largely uniform across neuronal types(Hrvatin et al., 2017) and brain regions and exhibits little background expression, making it straightforward to detect in the Drop-seq data. Though IEG induction has been reported as an artifact of cell dissociation(Hrvatin et al., 2017; Wu et al., 2017), we had sought to avoid promiscuous expression by digesting tissue in a buffer which lacked Ca^{2+} and inactivated channels necessary for action potential firing (**Methods**). Among the cells in this analysis, elevated expression of IEGs was observed in only a small fraction of the neurons of each type. For example, in FC layer 2/3 and 5a (cluster 6, **Figure 3**) or HP CA3 (cluster 6, **Figure S4C**) – regions prone to epileptiform activity due to recurrent, excitatory connections – IEG$^+$ neurons comprised less than 1.0% of the total population.

Neurons could in principle have other common transcriptional dynamics or cell states. To identify such patterns, we looked for broad transcriptional patterns that, despite being ascertained in many different brain regions or neuronal classes, involve similar combinations of genes. More specifically, we calculated pairwise IC-IC correlations of gene loadings for 368 biological ICs ascertained from neuronal cell classes (**Figure 6A**). One block of 15 highly inter-correlated ICs clearly represented the IEG signal in different regions (a positive control). Two additional correlation blocks were prominent. One of these arose entirely from thalamic neurons (of diverse types), suggesting a tissue-specific transcriptional program (that we did not further analyze). We focused on the other correlation block, which consisted of ICs from many brain regions and neuronal types.

This broad transcriptional pattern – present in almost all of the brain regions and many neuronal types – involved a large constellation of genes that underlie axonal and pre-synaptic function. We call this signal the "Neurofilament IC," because three of most strongly contributing genes encode the neurofilament subunits of the axonal cytoskeleton (*Nefl*, *Nefm*, and *Nefh*) (Yuan et al., 2012). Other strongly co-regulated genes in this module included included *Syt2*, *Vamp1,* and *Cplx1* – which have roles in vesicle exocytosis – and *Pvalb* and *Caln1,* which bind presynaptic Ca^{2+} (**Figure S6A**). Thus, the genes nominated by this transcriptional pattern are functionally related in the maintenance of axonal function and the support or tuning of neurotransmitter release.

Neurofilament ICs were observed in all sampled brain regions and in multiple neuronal classes, including GABAergic, glutamatergic, and neuromodulatory neurons. The expression levels of genes with strongest Neurofilament IC contributions tended to covary strongly both within and across neuronal types. Among interneurons, the Neurofilament IC cell loading was most prominent in *Pvalb*$^+$ interneurons, which tend to have higher firing rates than other interneurons (Hu et al., 2014). For example, in frontal cortex *Sst*$^+$/*Pvalb*$^+$ interneuron cluster 2, this constellation of genes exhibited continuously varying expression across *Pvalb*$^+$ cells, suggesting that it contributes to variation among cells of the same subtype (**Figure 6B-D**). Among *Pvalb*$^+$ interneurons, expression levels of the genes most strongly contributing to the Neurofilament IC – including *Nefh, Kcnc3, Syt2,* and *Nefm* – were continuously distributed and strongly correlated with one another; by contrast, expression levels of these genes were lower and less correlated among *Sst*$^+$ interneurons (**Figure 6D**). In the hippocampus, *Pvalb*$^+$ interneurons exhibited high cell loading for the Neurofilament IC, as did the *Pvalb*$^+$ "prototypical" neurons of the globus pallidus externus (Mallet et al., 2012; Saunders et al., 2016)(**Figure S6A**). We also identified this axis of transcriptional variation among glutamatergic populations, for example the subplate neurons in the frontal cortex (**Figure S6A**).

To determine whether these co-expression relationships are present *in vivo*, we performed three-channel transcript counting using single-molecule fluorescent *in situ* hybridization (smFISH). The Drop-seq data predicted correlated expression of *Pvalb*, *Syt2,* and *Nefm* – but not *Gabra4* (selected as a control gene) – among *Pvalb*$^+$ neurons in the frontal cortex (**Figure 6E**). This prediction was strongly confirmed by smFISH: among *Pvalb*$^+$ interneurons, *Syt2* transcript densities correlated with transcript densities for *Pvalb* and *Nefm* (**Figure 6F**) but not the control gene *Gabra4* (**Figure 6G**). We did not observe any obvious spatial organization of *Syt2/Nefm* expression among *Pvalb*$^+$ cells. To confirm that the Neurofilament signal did not arise from a cell-preparation artifact (such as transcriptional responses to axotomization during cell dissociation), we performed Drop-seq analysis of 28,194 single nuclei isolated from flash-frozen mouse frontal cortex; Neurofilament IC cell loading was still strongly visible among the nuclei of the *Pvalb*$^+$ interneurons (**Figure 6H**).

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

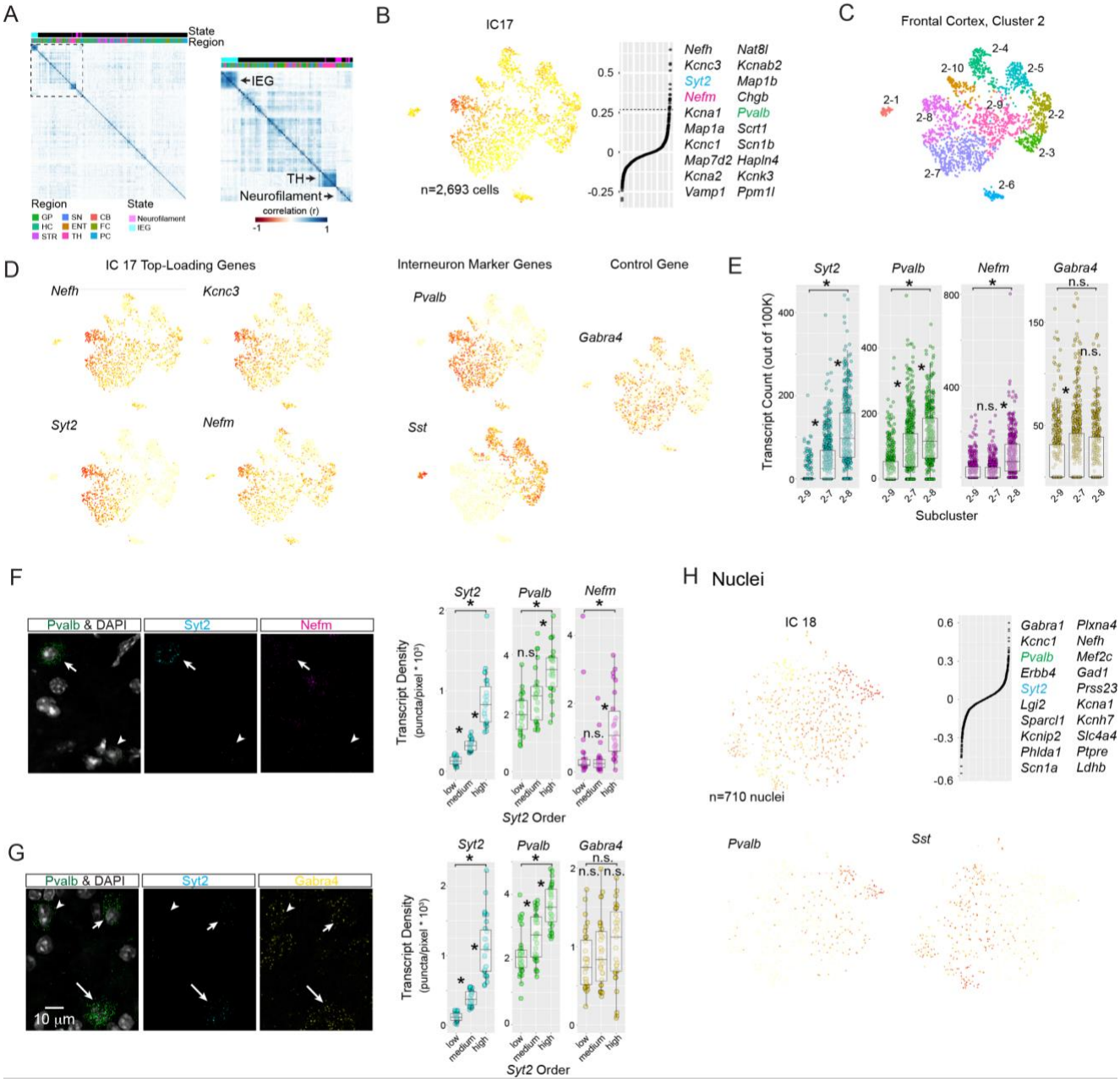


Figure 6. A prevalent expression state in neurons related to axon structure and presynaptic function

(A) Identification of a novel transcriptional state in neurons using biological ICs. ICs encoding transcriptional states utilized by different types of neurons can be identified by high correlations in expressed genes (gene-loadings). Hierarchical clustering diagram displaying pairwise Pearson correlations of gene-loading scores for all biological ICs curated from neuronal subclustering (n=45 subclustering analyses, n=323 total subclusters). Right, enlargement of the boxed region. IC gene-loading correlation blocks identify the immediate early gene ("IEG") transcriptional state, with IC contributions from different regions. Another correlation block ("TH") identifies a thalamus-specific collection of ICs. A third correlation block ("Neurofilament") has IC contributions from different regions and top-loading genes containing neurofilament subunits and other proteins involved in Ca2+ handling, vesicle exocytosis, and membrane excitability. ICs were color-coded for states by assessing the presence of state-marker genes in top loading genes.

(B) The Neurofilament transcriptional signal (IC 17) in frontal cortex Sst+/Pvalb+ interneurons (Cluster 2). Left, IC 17 cell-loadings displayed on subcluster tSNE plot. Right, gene-loading plot, with the top 20 loading genes shown at right. Syt2, Nefm, and Pvalb are color-coded to match ISH experiments below.

(C) Color-coded subcluster identities for frontal cortex cluster 2. N=10 subclusters were based on n=9 biological ICs. The graded loading of IC 17 is discretized into subclusters 2-8, 2-7, and 2-9.

(D) Expression plots for single genes. Left, n=4 top loading genes for IC 17 exhibit gradients of expression in Pvalb+ but not Sst+ interneurons (middle). Right, Gabra4 – a control gene that does not load onto IC 17 – does not exhibit expression gradients across Pvalb+ subclusters.

Figure 6 continued…

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(E) Quantitative comparison of Neurofilament (Syt2, Pvalb, and Nefm) and control gene (Gabra4) single-cell transcript counts across Pvalb+ subclusters from Drop-seq. Transcript means were compared with a two-way Anova followed by a Tukey Honest Significance Difference Test to calculate a p-value for differences across means. Asterisk, $P < 0.05$; n.s $P > 0.05$.
(F-G) Neurofilament gene and control gene in situ transcript count experiments within Pvalb+ frontal cortex cells using smFISH. Left, example single planes from confocal stacks. Right, quantification of transcript densities. Pvalb+ cells were split into n=3 groups based on Syt2 levels (low, medium, and high) mimicking subclusters 2-9, 2-7, and 2-8 from the atlas. Transcript densities were compared with a two-way Anova followed by a Tukey Honest Significance Difference Test to calculate a p-value for differences across means. Asterisk, $P < 0.05$; n.s $P > 0.05$. Longer arrows indicate cells with higher Pvalb expression.
(F) Experiment 1, Pvalb, Syt2, and Nefm.
(G) Experiment 2, Pvalb, Syt2, and Gabra4 (control).
(H) The Neurofilament IC is observed in flash-frozen nuclei from frontal cortex. The Neurofilament (IC 25) cell-loading signal distribution across the Sst+/Pvalb+ interneuron subcluster. Left, cell-loadings displayed on subcluster tSNE plot. Right, gene-loading plots. The top 20 loading genes are shown at right.

We conclude that many neuronal types share a coordinated transcriptional program involving genes whose functions are to facilitate maintenance, expansion, or subcellular transport to the axon and presynaptic terminal. Neuronal types characterized by extensive axonal arbors, long-distance axonal projections, and/or faster firing rates tended to utilize this transcriptional program more than other neurons. At the same time, the magnitude of expression of this program varied among neurons of the same subtype, suggesting that this program contributes to both intra- and inter-type diversity.

Gene-gene Co-expression Relationships Inferred from Hundreds of Cell types and States

Cells have diverse gene expression patterns – yet functional imperatives and regulatory architecture may greatly constrain genes' patterns of co-expression. Gene expression data for the 565 cell populations (identified by the above subclustering analysis) make it possible to analyze genes' co-expression relationships across many cell types, cell states and other functional contexts. We performed such analyses at the level of the 565 transcriptionally distinct cell populations (rather than 690,207 individual cells), a level of organization at which the data sample functional diversity yet are less influenced by single-cell noise and statistical sampling. (The 565 subclusters contained on average 565 cells, with expression patterns defined by an average of 1.9 million UMIs.) (**Figure S7A**).

To assess whether gene-gene correlations across the 565 cell populations could capture known functional relationships, we first focused on genes encoding subunits of the nicotinic acetylcholine receptors (nAChRs, n=16 genes). nAChRs are ligand-gated, pentameric ion channels which we chose as a test system because the channels observed *in vivo* consist of eclectic but well-described subunit combinations that vary by brain region and neuron type (Gotti et al., 2006). We hypothesized that the observed combinations are at least partially achieved through cell-type-specific patterns of RNA expression (in addition to selective protein association).

Across the 565 cell populations, expression levels of nAChR genes exhibited two prominent correlation blocks, each containing genes that encode subunits of known heteromeric α/β channels (Zoli et al., 2015)(**Figure 7A**). For example, the expression of *Chrna3* and *Chrnb4* (known to form a functional heteromeric receptor) were positively correlated across a large range of average transcript abundance (from 0.01 to 100 transcripts per 100K total transcripts) (**Figure 7B**). The expression of other pairs of genes that encode subunits of known heteromeric receptors (*Chrna6/Chrnb3* and *Chrna4/Chrnb2*) were also well-correlated with each other, whereas *Chrnb1* and *Chrnb2* were if anything negatively correlated, consistent with a lack of β1/β2 channels described in the brain (**Figure 7B**). These selective correlations match prominent subunit combinations inferred at the protein level *in vivo*, suggesting that nAChR composition is achieved at least in part by cell-type-specific patterns of RNA expression.

Some nAChR subunits form homomeric functional channels; the absence of obligate partners for such subunits could in principle cause them to lack positive pairwise correlations with other subunit genes. We observed several α-subunit genes with this property, including those known (*Chrna7*, *Chrna9,* and *Chrna10*) and others not yet known (*Chrna1* and *Chrna2*) to form homomeric channels in the brain (Gotti et al., 2006). In summary, gene expression correlations across cell types recapitulate known nAChR receptor subunit combinations (Gotti et al., 2006) and do so more accurately than correlations based on the bulk expression profiles for the nine regions (**Figure S7B**).

We then sought to explore the expression-correlation structure for other vital gene families whose co-function relationships are less well understood. An elegant body of experimental and theoretical work suggests that individual neurons can attain their type-specific electrophysiological properties by expressing unique combinations of voltage-gated ion channels (Marder and Goaillard, 2006). It is not known, however, if neurons can utilize any combination of ion channel genes, or face constraints (functional or regulatory) that limit which

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

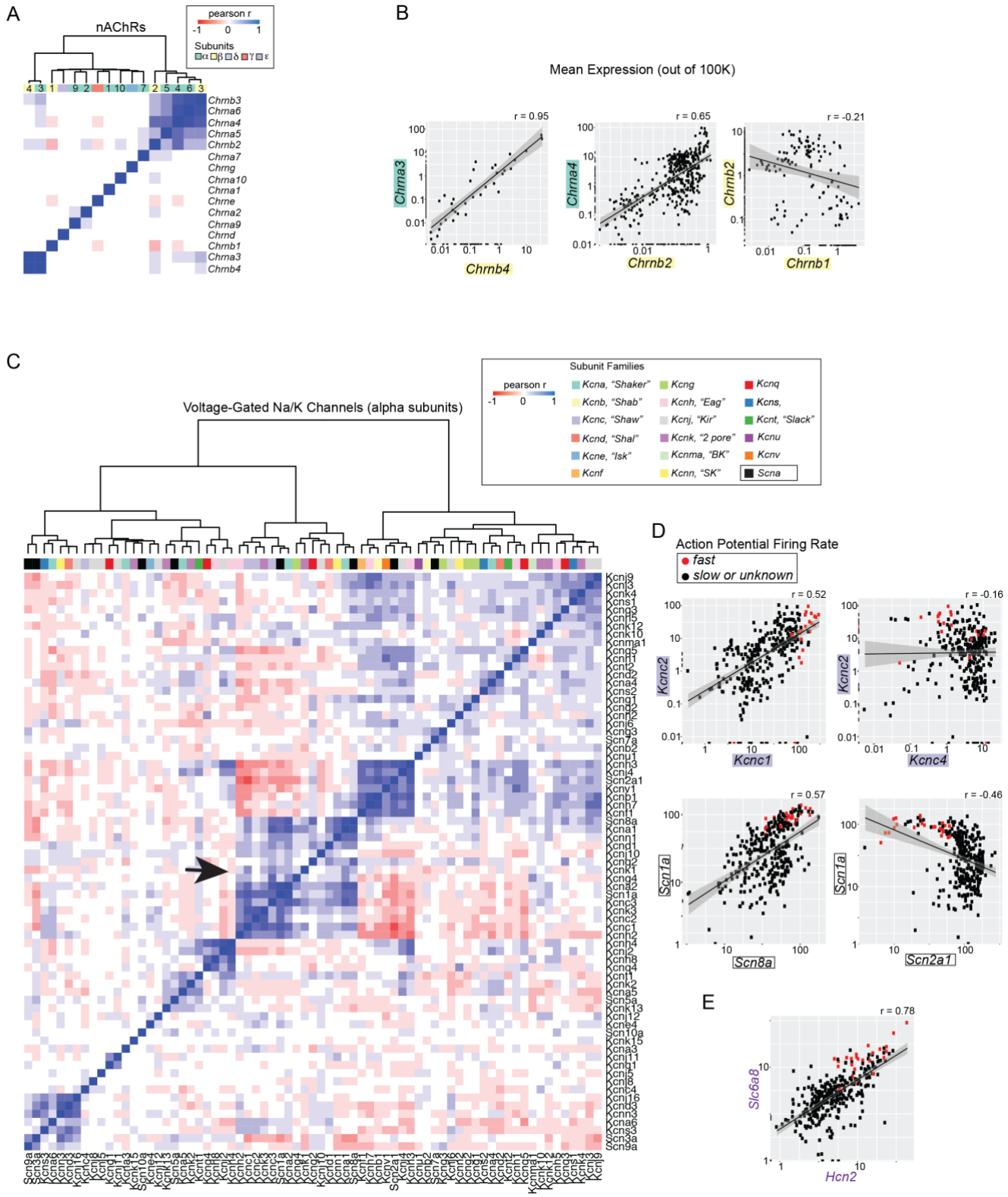


Figure 7. Inferring ion channel gene-gene co-expression relationships by comparing hundreds of brain cell types and states
(A-B) Nicotinic acetylcholine receptor (nAChR) subunit composition can be inferred from gene expression correlations found in 565 transcriptionally distinct atlas cell populations. Pearson correlations calculated from linear expression data.
(A) Hierarchical clustering of pairwise correlations of n=16 nAChR subunit genes. Subunit families are color-coded.
(B) Scatterplot comparisons of subunit expression levels. Expression levels are scaled to 100K UMIs and shown on a log10 scale. α and β subunit genes are color-coded. Note, Chrna3/Chrnb4/Chrna5 are adjacently located on chromosome 9; Chrna4/ Chrnb2 genes are on chromosome 2 and 3, respectively.
(C-E) Correlation structure amongst voltage-gated (VG) Na and K channels measured from n=323 neuronal populations.
Figure 7 continued…

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(C) Hierarchical clustering of pairwise correlations calculated from expression data. The n=17 VGK channel alpha subunit families are color-coded and labeled with a common name if used. The n=1 VGNA channel alpha subunit family is shown in black. The correlation block encompassing channels known to control firing rate is shown with an arrow.
(D-E) Scatterplot comparisons of subunit expression levels. Neuronal populations known to exhibit fast firing rates are shown in red (Figure S7D). Gene families are color-coded. Expression levels are scaled to 100K UMIs and shown on a log10 scale. Slc6a8 and Hcn2 were amongst the genes most frequently correlated with the alpha subunit gene set that putatively encodes firing rate (Figure S7D).

channel subunits are expressed in the same cell types. We therefore evaluated the correlations among the expression patterns of 71 voltage-gated potassium (VGK) and sodium (VGNA) channel genes across 323 neuronal populations, focusing on the genes that encode the alpha (pore-forming) subunits. This analysis identified prominent blocks of positive and negative correlation, with each such block including combinations of VGK and VGNA genes **(Figure 7D**).

To evaluate whether these gene expression–correlation blocks relate to cells' electrophysiological properties, we identified fast-firing cell types among the 323 neuron populations (**Figure S7D**). Among two prominent sets of co-expressed genes (which were negatively correlated to each other), the larger set contained genes that enable fast and persistent action potential firing rates, including the *Kcnc1-3* (Kv3.1), *Kcna1* (Kv1.1) and *Scn8a* ($Na_v1.6$) channels (Chen et al., 2008; Goldberg et al., 2008; Rudy and McBain, 2001) (**Figure 7D**). Fast-firing cell types expressed high levels of the VGK channel genes *Kcnc2* and *Kcnc1* and the VGNA channel genes *Scn1a* and *Scn8a* (the only two VGNA channels in the block). Fast-firing cells expressed variable levels of *Kcnc4* (the fourth Kv3 family member) and consistently low levels of *Scn2a1*, whose expression was inversely correlated with *Kcnc1-3* and *Scn1a* (**Figure 7D**). These relationships may nominate functional hypotheses about channels whose contributions to cells' physiological properties are only partly understood. For example, *Kcnc4* could be a Kv3 family member that helps tailor membrane properties orthogonal to firing frequency, and *Scn2a1* might undermine fast-firing. Interestingly, many of the genes whose expression was correlated with fast-firing properties are transcriptionally activated during postnatal maturation of fast-spiking interneurons (Okaty et al., 2009).

Finally, we asked what other genes were consistently co-expressed with the alpha subunit channel genes that themselves exhibited high expression in fast-firing neurons (**Figure S7D**). This collection contained expected genes (like those found in the Neurofilament signal, **Figure 6**), plausible ion channels (such as *Hcn2*), and unexpected genes, such as the transcription factor *Foxj1* and the creatine transporter *Slc6a8*. Indeed, expression levels of *Slc6a8* and *Hcn2* were strongly correlated and high in fast-firing cell types (**Figure 7F**). Humans with dominant-negative mutations in *Slc6a8* often present with epilepsy, intellectual disability, and motor symptoms (van de Kamp et al., 2013), but the cellular mechanisms remain largely unknown. Our data suggest that *Slc6a8* plays an enhanced role in fast-spiking cells – which would include *Pvalb*$^+$ basket cells, the resident fast-firing inhibitory neurons in cortical areas known to seed seizures. Indeed, male mice with mutations in (x-linked) *Slc6a8* have fewer GABAergic synapses (Baroncelli et al., 2017). These data highlight how analyses of gene co-expression across a large number of cell types can nominate new hypotheses about genes, brain circuitry, and disease.

Cell-type Specialization Between Cortical Poles

The cerebral cortex processes motor, sensory, and associative information and is expanded in primates, especially humans (Buckner and Krienen, 2013). While extrinsic inputs are specific to different cortical areas, little is known about what intrinsic molecular specializations could contribute to region-specific cortical function. We first determined how accurately our cortex datasets represents cellular populations *in vivo* and then systematically identified transcriptional specializations within each non-neuronal cell class and across the two major neuronal types (glutamatergic and GABAergic neurons).

The process of single-cell dissociation followed by the capture, amplification, and sequencing of single-cell transcriptomes may distort the abundance of brain cell types. To test for distortions in our tissue preparation or analysis, we compared the abundances of molecularly-defined cell populations in our frontal cortex dataset to intact tissue using immunohistochemistry and *in situ* hybridization **(Figure S8D,E** and **Methods).** We found that neurons were over-represented relative to non-neurons in the Drop-seq dataset (Drop-seq: 0.76 ± 02 mean ± sem; tissue: 0.51 ± 03). We believe this effect is partially explained by cell-inclusion thresholds used for analysis, in which some real single-cell libraries are not included due to their small size; neurons, which had over three-fold higher transcript counts than non-neurons, are relatively protected from this effect (neurons: 5,039 ± 15 mean ± sem; non-neurons: 1,696 ± 9, **Figure S8F** and **Methods**). Compared to GABAergic interneurons, glutamatergic neurons exhibited two-fold greater abundance than expected from tissue (ISH: 5.1:1; Drop-seq: 11:1)(Sahara et al., 2012), an effect that again could be driven by larger libraries (glutamatergic: 5,298.916 ±

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

16, mean ± sem; GABAergic: 2,626 ± 21) (**Figure S8H**). Comparing three major subtypes of GABAergic interneurons, we observed that *Vip*⁺ cells were over-represented and *Sst*⁺ and *Pvalb*⁺ cells were underrepresented in the data (*Vip*⁺ ISH: 16%; Dropseq: 35%; *Pvalb*⁺ ISH: 31%; Dropseq: 25%; *Sst*⁺ ISH: 28%; Dropseq, 22%). Since greater *Vip*⁺ interneuron abundance cannot be explained by higher transcript counts (*Pvalb*⁺: 2,996 ± 61; *Sst*⁺: 2,758 ± 53; *Vip*⁺: 2,236 ± 32), these data may suggest that *Pvalb*/*Sst* interneurons were preferentially depleted during the preparation. This distortion could also contribute to the increased fraction of glutamatergic/GABAergic neurons that we observe. We conclude that our data exhibit modest, quantitative skews in cellular representation that are primarily driven by differences in transcript abundance and cell viability.

To identify molecular specializations across cortical regions, we performed semi-supervised ICA on cells from the frontal and posterior cortex together, separately analyzing cells of the various cell classes: excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes/polydendrocytes, microglia/macrophages, and endothelial tip and stalk cells. To identify transcriptional signals that differed between regions, we first calculated a "region skew score" for each of the biological ICs (based on the tendency of high-scoring cells to come from one region or the other, relative to low-scoring cells) (**Figure 8A, Methods**). Glutamatergic excitatory neurons generated more such regionally specialized ICs than GABAergic interneurons and non-neuronal cell classes did. Consistent with this, we identified strong regional differences in cellular abundance (a normalized ratio of greater than 3:1, P < 0.05, Barnard's Test, Bonferroni corrected) for seven subclusters of excitatory neurons and none of the other cell classes (**Figure 8B,C and Figure S9A)**. To visualize the spatial distributions of the excitatory-neuron populations corresponding to these subclusters, we examined the ISH patterns of markers for each of these populations; the ISH patterns consistently confirmed asymmetric expression across the cortical mantle (**Figure 8C and Figure S9B**). To determine whether molecular differences, in addition to differences in cell-type representation, contributed to frontal versus posterior specialization, we performed differential expression analysis comparing cells (within the same subcluster) from the two cortical regions. These analyses revealed far more differentially expressed genes in comparisons of excitatory

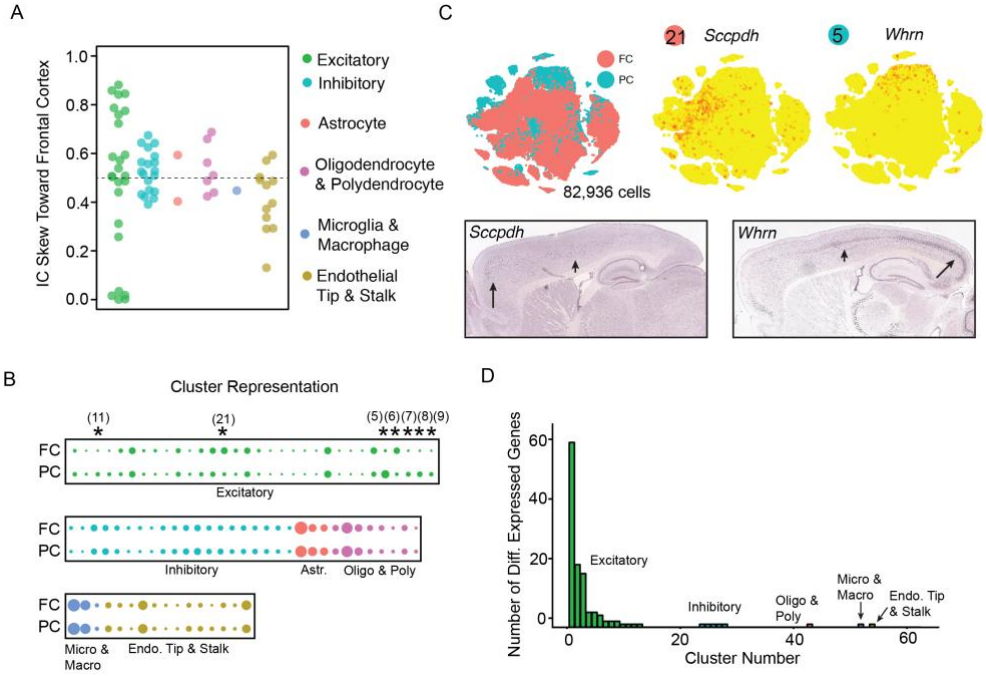


Figure 8. Cortical regions remain transcriptionally specialized via excitatory glutamatergic neurons
(A) Beeswarm plot showing the relative contribution of frontal (FC) vs posterior (PC) cells to top-loading cells for each of the biological ICs in six separate cell-class analyses (glutamatergic excitatory neurons, n=82,936 cells; GABAergic inhibitory neuron, n=7,783; astrocyte, n=7,782, oligodendrocytes/polydendrocytes, n=3,505; microglia/macrophage, n=1,027; and endothelial stalk/tip cells, n=3,578). For each analysis, equivalent cell classes from the frontal and posterior datasets were jointly reanalyzed. The normalized proportion of cells in FC and PC is plotted. The IC Skew score is 1 if only FC cells contribute and 0 if only PC cells contribute. Equal contribution is 0.5 (dotted line).

(B) Plot comparing the relative number of FC vs PC cells within each subcluster following clustering of the biological ICs for each cell class analysis. Circle size denotes the fractional representation. Subclusters with greater than 3:1 compositional skew are indicated with an asterisk and their label is shown (P < 0.05, Barnard's test, Methods). Subclusters plots for each cell class are shown in Figure S9.
(C) Subcluster analysis of glutamatergic excitatory neurons from FC and PC illustrating that excitatory neuron populations are transcriptionally specialized by cortical region. Top left, tSNE plot of excitatory neurons color-coded by region. Top right, expression tSNE plots for Sccpdh and Whrn, genes enriched in subclusters 21 and 5 that are disproportionately composed of FC or PC cells respectively. Bottom, ISH stain (Allen Institute Mouse Brain Atlas) for Sccpdh and Whrn, which exhibit skewed expression in the FC or PC, respectively. High expression, long arrow; Medium expression, short arrow.
(D) Barplot summarizing the number of genes differentially expressed across FC and PC cells within each subcluster across all cell class experiments (> 2-fold change, P < .05, Bonferroni corrected).

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

populations than in comparisons of GABAergic interneurons or non-neuronal cell classes (**Figure 8D**). These data suggest that molecular specializations across cortical areas are largely driven by differences involving excitatory projection neurons. This result is consistent with independent single-cell analyses in mouse (Tasic et al., 2017) and with theories of cortical specialization in humans and other primates (Krienen et al., 2016).

Resolving Neuron Types within the Basal Ganglia

Neuronal types and subtypes in many brain regions are incompletely identified or characterized, limiting their study in neural circuits, behavior, and disease. Our data provide an opportunity to more systematically identify neuronal types in many regions. From the 9 regions, we identified 323 neuronal subclusters which arose from 368 biological ICs (**Figure S4**).

Here, we illustrate the identification of neuronal types by focusing on the basal ganglia, a collection of richly interconnected subcortical areas. While neuron types in the striatum have received most attention, decades of *in vivo* and *in vitro* electrophysiological recordings in other basal ganglia nuclei have documented extensive neuronal diversity – yet the anatomical complexity of these regions has made identifying neuron types challenging and laborious (Bolam et al., 2000; DeLong et al., 1984). We turned to the globus pallidus externus (GPe) and substantia nigra reticulata (SNr), two highly disease-relevant (Albin et al., 1989) regions in which neuron types are not yet well classified (**Figure 1 & Figure S1**).

Like many functionally specialized nuclei in the rodent brain, the GPe and SNr are anatomically small structures, with 10^3–10^5 total neurons (Oorschot, 1996), as compared to the 10^6–10^7 found in larger regions like the striatum, cortex, and hippocampus (Abusaad et al., 1999; Schüz and Palm, 1989). This poses a challenge for Drop-seq, which samples only about 5% of the total cells in a cell suspension (Macosko et al., 2015). We addressed this issue by both optimizing dissociation (to maximize cell yield, **Methods**) and sampling larger volumes of tissue that encompassed multiple adjacent structures. (We note, however, that as the number of included regions goes up, so too does the chance for biases in neuronal representations (because of differences in library size and sensitivity to the digest), as well as the complexity and ambiguity of the post-hoc anatomical analysis.) The GPe and SNr were included in the GP/NB and SN/VTA analyses, respectively (**Figure S1**).

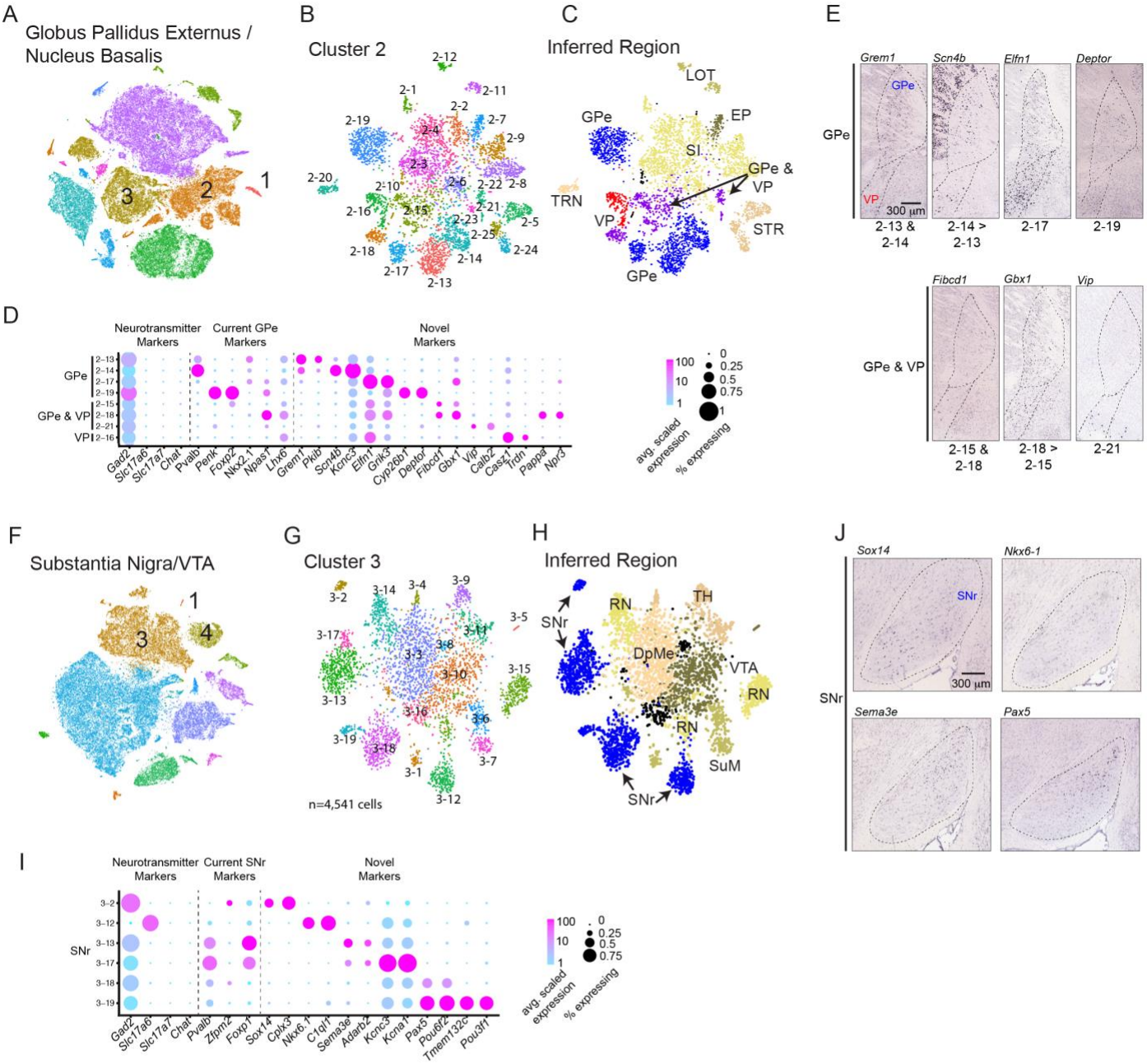
To identify GPe neuron populations, we leveraged the Allen Institute's digital ISH atlas to spatially map expression of markers of global clusters and subclusters to the GPe or surrounding regions from the GP/NB dataset (**Figure 9, Figure S1**). GPe neurons were present in cluster 2 (n=11,103 cells), one of three neuron clusters we identified. Cluster 1 corresponded to cholinergic neurons from the NB and GPe (n=437 cells) and cluster 3 to neurons of the adjacent striatum and basolateral amygdala (n=9,847 cells; **Figure 9A**).

To determine which of the 25 subclusters within cluster 2 were intrinsic to the GPe or the adjacent ventral pallidum (VP), we identified markers from our data and evaluated their spatial expression using ISH. We found 8 candidate subclusters intrinsic to the GPe and/or VP, of which 4 were exclusive to the GPe (**Figure 9C-E**). The other 17 subclusters mapped to surrounding regions, including the thalamic reticular nucleus (TRN), the substantia innominata (SI), and the lateral olfactory tract (LOT) (**Figure S10**).

Modern approaches to classify GPe neurons rely on iterative mapping of anatomical, electrophysiological, and behavioral response properties onto single molecular markers (e.g., (Mallet et al., 2012; Saunders et al., 2016)); such data provide essential insights into neuronal diversity, but have trouble conclusively identifying distinct populations. To associate subclusters with putative GPe neuron types, we compared the expression of published GPe markers to pairs of markers that we identified for each subcluster (**Figure 9D**). Of markers described previously, only *Pvalb* and *Penk/Foxp2* were selectively expressed in particular GPe subclusters (Kita, 1994; Voorn et al., 1999). This correspondence suggests subcluster 2-14 represents the fast-spiking "prototypical" population, while 2-19 represents the slow-firing "arkypallidal" population (Abdi et al., 2015; Mallet et al., 2012). Interestingly, 2-13 is transcriptionally similar to 2-14, sharing markers such as *Grem1*, but is distinguished by its stronger expression of genes like *Scn4b* and *Kcnc3* and those from the "Neurofilament program" (described above) (**Figure S6A**). Further experiments will be necessary to determine whether subclusters 2-14 and 2-13 represent anatomically distinct subpopulations or transient states. The fourth GPe subcluster (2-17) is enriched for markers *Elfn1/Grik3*, but to our knowledge, has not been characterized. Other published markers (e.g., *Nkx2.1*, *Npas1,* and *Lhx6*) show mild, reciprocal enrichments across multiple subclusters, illustrating how markers can fail to capture the true population structure of neurons (Abdi et al., 2015; Hernandez et al., 2015; Mastro et al., 2014).

To our surprise, several subclusters (2-15, 2-21, and 2-18) expressed markers found in both the VP and bordering GPe (**Figure 9D,E**). This border-spanning ISH pattern suggests that some neuron types may be found in regions with differing behavioral functions and connectivity (Gittis et al., 2014; Kita, 2007; Smith et al., 2009). Conceptually, this may be analogous to distinct

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 9. Transcription-based identification of known and novel neuron type distinctions within the basal ganglia.
(A-E) Globus pallidus externus (GPe).
(F-J) Substantia nigra reticulata (SNr).
(K-O) Dopaminergic vs acetylcholinergic neuromodulatory neuron populations.
(A) Color-coded global clusters (n=11) for globus pallidus externus/nucleus basalis dataset displayed on a tSNE plot. Clusters 1, 2, and 3 are neuronal.
(B) Subcluster structure within cluster 2.
(C) Subclusters color-coded by candidate anatomical regions, inferred by ISH expression patterns of selective marker genes (Figure S10). Populations are intrinsic to the GPe, ventral pallidum (VP), substantia innomatita (SI), adjacent striatum (STR), lateral olfactory tract (LOT), rostral entopeduncular nucleus (EP) and the thalamic reticular nucleus (TRN) consistent with dissections (Figure S1).
(D) Dot plot illustrating the expression patterns of neurotransmitter marker genes, neuron type markers from the literature, and novel markers identified in this analysis. Dot diameter represents the fraction of cells within a subcluster where a transcript was counted. Colors represent average single-cell scaled expression value (out of 100K UMIs, log10).
(E) ISH experiments with subsets of novel marker genes illustrating the expression within the anatomical regions of the GPe and/or VP (sagittal sections). Region borders are approximated by a dotted line (Allen Institute Mouse Atlas).
(F) Color-coded global clusters (n=14) for substantia nigra/VTA dataset displayed on a tSNE plot. Clusters 1, 3, and 4 are neuronal.
(G) Subcluster structure within cluster 3.
(H) Subclusters color-coded by candidate anatomical regions, inferred by ISH expression patterns of selective marker genes (Figure S10). Populations are intrinsic to the SNr, ventral tagmental area (VTA), red nucleus (RN), supramammillary nucleus (SuM), thalamus (TH), and deep mesencephalic nucleus (DpMe).
(I) Dot plot as in (D). Marker genes for neurotransmitters, current SNr markers, and novel markers identified in this study.
(J) ISH experiments with subsets of novel marker genes illustrating the expression within the anatomical regions of SNr (sagittal sections). Region borders are approximated by a dotted line (Allen Institute Mouse Atlas).
(K) Subcluster structure within Th+/Ddc+ dopaminergic cluster 3 from the SN/VTA dataset.
(L-M) Example cluster 3 ICs that encode spatial signals within the SNc/VTA.
(L) IC cell loadings displayed on tSNE plot.
(K). IC gene-loadings. Top ten genes shown at right.
(L) ISH experiments (sagittal sections) for Lpl (IC 10, top) and Aldh1a1 (IC 12, bottom). IC 10 identifies the dorsal VTA, while IC 12 identifies the ventral VTA and SNc (Allen Institute Mouse Atlas).
(O-P) Minimal heterogeneity identified within Chat+/Slc5a7+ cholinergic cluster 1 from the GP/NB dataset.
(O) Plot of IC 4 cell-loadings. Based on IC 4, cells are assigned as subcluster 1-1 or 1-2.
(P) IC 4 gene-loading plot. Top ten loading genes at right suggest a Neurofilament-type signal.

regions of the cerebrum harboring similar interneurons (Harwell et al., 2015). These shared subclusters may explain a neuron type that is synaptically incorporated into the GPe, but exhibits VP-like axonal projections (Chen et al., 2015b; Saunders et al., 2015).

To identify the neuron populations intrinsic to the SNr, we followed a similar procedure using the SN/VTA dataset (**Figure 9F, Figure S1**). Of the 4 global neuron clusters identified, one (cluster 3, n=10,049 cells) contained neurons of the SNr as well as the surrounding regions (**Figure 9H**). The other three clusters (clusters 1, 2, and 4, respectively containing 73, 297, and 1,841 cells) contained hippocampal, thalamic, and dopaminergic neurons, respectively.

Of the 19 subclusters in cluster 3, we mapped 6 to the SNr (**Figure 9G,H** and **Figure S10B**). Two sets of subclusters shared highly selective markers (3-18/3-19: *Sema3a*, *Adarb2*; 3-17/3-13: *Pax5*, *Pou6f2*), suggesting they are closely related. Each pair is distinguished by genes whose expression might imply a state or subtype distinction, like ion channels (*Kcnc3*, *Kcna1*), transmembrane proteins (*Tmem132c*), and transcription factors (*Pou3f1*) (**Figure 9I**). Of the remaining 2 SNr subclusters, 3-12 expressed *Slc17a6*$^{+}$, presumably corresponding to the glutamatergic projection from the SNr to the thalamus (Antal et al., 2014). Subcluster 3-2 was *Gad2*$^{+}$/*Pvalb*$^{-}$, expressed the developmental marker *Zfpm2* (Lahti et al., 2016), and likely represents a third GABAergic type marked by *Sox14* and *Cplx3* expression (**Figure 9I**). To our knowledge, this is the first comprehensive molecular description of adult SNr neurons.

The GPe and SNr each abut small populations of neuromodulatory neurons that have profound effects on circuit physiology and behavior through the widespread release of acetylcholine (ACh) and dopamine (DA), respectively. Molecular specializations are known to exist in the DA system that are, at least in part, associated with the anatomical location of DA neuron subtypes (Lammel et al., 2008; Poulin et al., 2014). It is unclear whether the ACh-releasing population exhibits similar heterogeneity and whether this diversity has a spatial component.

Subcluster analyses of the dopaminergic (SN/VTA, cluster 4, 919 cells) and cholinergic (GP/NB, cluster 1, n=218 cells) clusters revealed that the sampled dopaminergic neurons were indeed more heterogeneous than the cholinergic neurons (DA: 6 biological ICs, 9 subclusters; ACh: 2 biological ICs, 2 subclusters)(**Figure 9A,F**). As expected, aspects of DA transcriptional diversity were related to spatial positioning (Poulin et al., 2014) – for example, delineating the dorsal (IC 10) from ventral (IC 12) VTA as illustrated by ISH stains for the weighted genes *Lpl*

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and *Aldh1a1* (**Figure 9L-N**). In contrast, the major signal in cholinergic neurons (IC 4) appears to be Neurofilament-like, with no spatial component (**Figure 9O,P**). Thus, DA neurons of the SNc/VTA appear to be regionalized, whereas cholinergic neurons of the GPe/NB are not. Sampling more cholinergic neurons, especially from other areas of the basal forebrain (Zaborszky et al., 2013), could reveal additional signals. Spatial transcriptomic technologies that measure single cells could also reveal anatomical relationships missed by ISH analysis on thin slices of tissue (Chen et al., 2015a; Shah et al., 2016).

Molecular Specializations of Striatal Principal Neurons

Spiny projection neurons (SPNs) represent about 95% of the neurons in the rodent striatum. Decades of anatomical and functional work has sought to establish how molecularly defined subtypes of SPNs contribute to circuit function, behavior, and disease (Albin et al., 1989; Kozorovitskiy et al., 2012; Kravitz et al., 2010). Two principal categories have been used to distinguish SPN subsets. The first – based on divergent axonal projections and receptors for dopamine – assigns SPNs to the "direct" (dSPN) and "indirect" (iSPN) pathways; dSPNs and iSPNs are similarly numerous. The second – based traditionally on processing limbic versus sensory/motor information – groups SPNs into two spatial compartments within the striatum, the so-called "patch" and "matrix" compartments (Gerfen, 1992; Graybiel and Ragsdale, 1978). Both dSPNs and iSPNs are present in the patch and matrix.

The direct- and indirect-pathway SPN types were readily identified by the expression of the pan-SPN marker *Ppp1r1b* along with the iSPN marker *Adora2a* or the dSPN marker *Drd1* (**Figure 10A,B**). Two large clusters were highly enriched for *Ppp1r1b*/*Drd1* and *Ppp1r1b*/*Adora2a*, suggesting they corresponded to dSPNs (Cluster 10, n=30,835 cells) and iSPNs (Cluster 11, n=25,305 cells), respectively. There were 68 differentially expressed genes between the two, including both known and previously undescribed markers (**Figure 11A** and **Table S2**).

To identify cells from the patch and matrix compartments among dSPNs and iSPNs, we inspected transcriptional signals identified from SPN subclustering (dSPN Cluster 10: n=9 biological ICs; iSPN Cluster 11: n=10 biological ICs, **Figure S11**). Each analysis identified a single candidate "patch IC" whose most strongly contributing genes included known patch markers such as *Tac1* and *Pdyn*. Approximately 10% of iSPNs and dSPNs exhibited this patch transcriptional phenotype (dSPN: IC 15, subcluster 10-2; iSPN: IC 20 in subcluster 11-2; **Figure 10C,D**).

To appreciate the extent to which the patch/matrix distinction affects the functional specializations of dSPNs and iSPNs, we compared gene loadings between the patch-encoding dSPN IC 15 and iSPN IC 20 (**Figure 10E**). As expected, we observed both classic (*Tac1*) and undescribed (*Tshz1*) pan-patch markers (**Figure 10F**). We found that the proton-gated Na channel *Asic4* is enriched in iSPN patch cells, while Ca-binding protein *Necab1* is enriched in dSPN patch cells (**Figure 10G**). While studies of SPNs have focused independently on the patch/matrix or pathway axes, our data suggest a more complex pattern of specialization between these systems: the transcriptional features endowed by a patch habitat are not identical for dSPNs and iSPNs, and some of these differences appear to remove pathway expression differences found in the matrix dSPNs and iSPNs. For example, among matrix SPNs, *Asic4* was selectively expressed in the dSPNs but not iSPNs; among patch SPNs, however, *Asic4* was expressed in both dSPNs and iSPNs. Thus *Asic4* helps tailor iSPNs but not dSPNs to the patch habitat.

"Eccentric" SPNs: A novel, third axis of SPN diversity

Surprisingly, about 4% of SPNs (*Ppp1r1b*$^+$) were observed in a third, smaller cluster that also expressed *Adora2a* and *Drd1* (cluster 13: n=2,744 cells; 4.5% of *Ppp1r1b*$^+$ neurons; **Figure 10A,B**). All three biological replicates contributed to this distinct population, suggesting it was not due to a dissection artifact. These SPNs had a transcriptional phenotype distinct from other SPNs: some 110 genes differed significantly in expression between this cluster and the canonical dSPNs+iSPNs (fold ratio > 2 and $P < 10^{-100}$ by binomTest (Robinson et al., 2010)), more than the number of genes that distinguish dSPNs and iSPNs from each other (n=68) (**Figure 11A, Table S3**), Several genes were selectively expressed in cluster 13 with very little expression in the rest of the striatum (e.g., *Casz1, Otof, Cacng5, and Pcdh8*) (**Figure 11B**). Due to their expression of canonical SPN genes (e.g., *Ppp1r1b*) yet large transcriptional divergence from canonical SPNs, we call this population "eccentric" SPNs (eSPNs).

To determine where eSPNs are located, we performed triple smFISH labeling with *Ppp1r1b* and pairs of selective eSPN markers (**Figure 11C,D**). Candidate eSPNs (triple-positive cells) clearly localized to the striatum; like dSPNs and iSPNs, eSPNs exhibited no obvious spatial organization within the striatum and were intermixed with other SPNs. We conclude that eSPNs are an intrinsic striatal neuron population. Our Drop-seq data account for all known striatal interneuron types at approximately the expected proportions (3.9% of total neurons)(Tepper and Bolam, 2004), suggesting

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

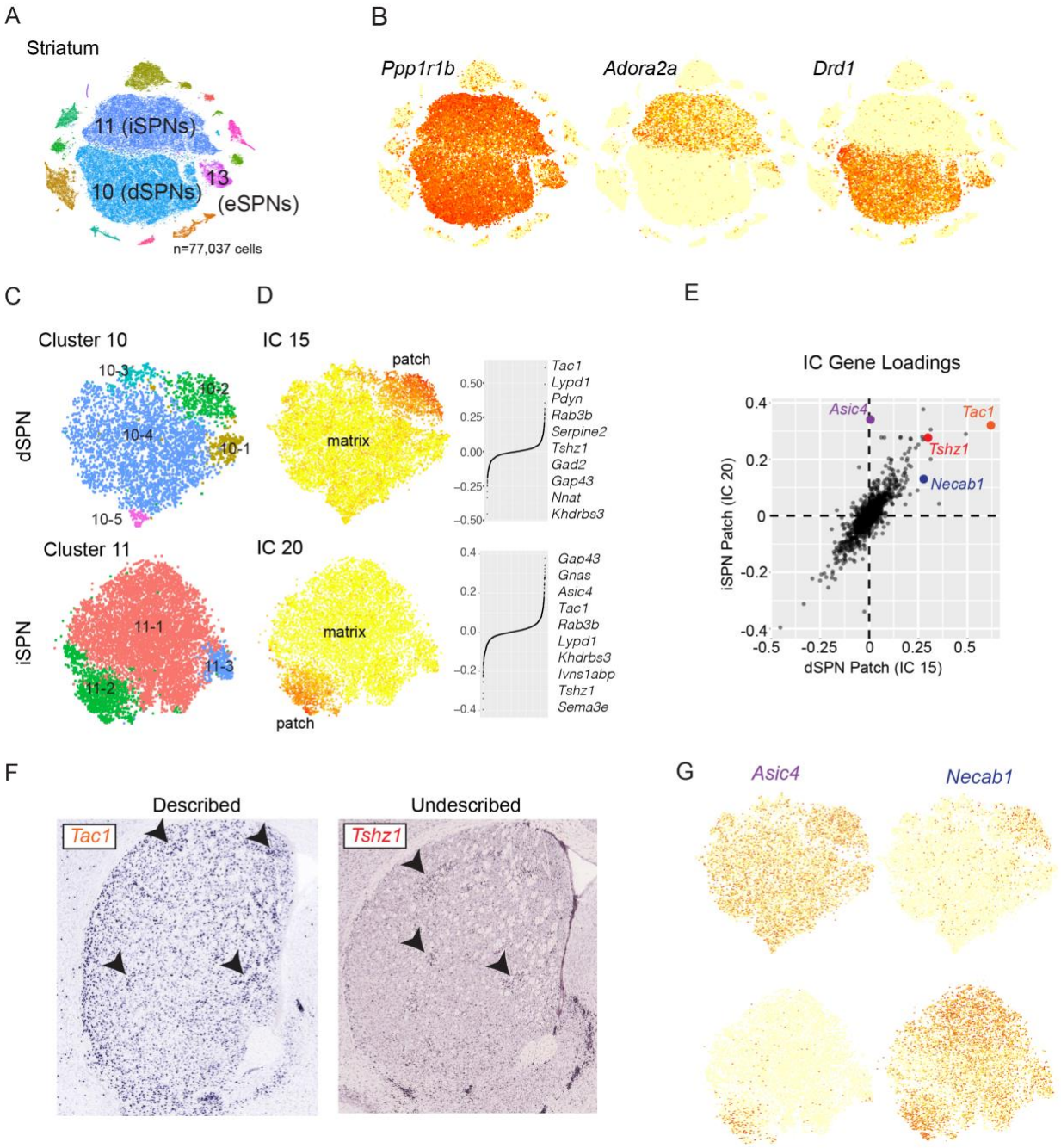


Figure 10. Molecular specializations within spiny projection neuron patch and matrix spatial compartments

(A) Color-coded global clusters (n=15) for striatum dataset displayed on a tSNE plot. Clusters 10, 11, and 13 are presumed SPNs, corresponding to dSPNs, iSPNs, and eSPNs (described in Figure 8).

(B) Expression tSNE plot of pan-SPN marker Ppp1r1b, direct pathway SPN (dSPN) marker Drd1, and indirect pathway SPN (iSPN) marker Adora2a. Cluster 13 expresses SPN markers. Ppp1r1b+ cells within Cluster 13 are eccentric SPNs (eSPN).

(C-E) Identification and comparison of patch iSPN and dSPN transcriptomes.

(C) Color-coded subcluster assignments for striatum cluster 10 (dSPNs) and cluster 11 (iSPNs) displayed on tSNE plot.

(D) ICs encoding patch-specific transcriptional signals in dSPNs (IC 15) and iSPNs (IC 20). Left, IC cell-loadings on tSNE plots. Right, gene-loadings with top 10 genes displayed.

(E) Identification of shared and distinct dSPN and iSPN patch-enriched genes. Scatterplot comparing gene loadings between patch ICs for dSPNs (IC 15) and iSPNs (IC 20). Genes with high loading in both ICs are hypothesized to be enriched in both patch iSPNs and dSPNs. This group includes described (Tac1) and previously undescribed (Tshz1) genes enriched in patch dSPNs and iSPNs. Genes that load strongly onto either IC are candidates for patch molecular specialization by SPN pathway. For example, Asic4 is enriched in patch iSPNs, while Necab1 is enriched in patch dSPNs.

(F) ISH stains for Tac1 and Tshz1 on coronal sections of striatum (Allen Mouse Brain Atlas). Arrowheads point to example patches.

(G) Expression tSNE plots for Asic4 and Necab illustrating patch enrichment exclusive to iSPNs and dSPNs, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

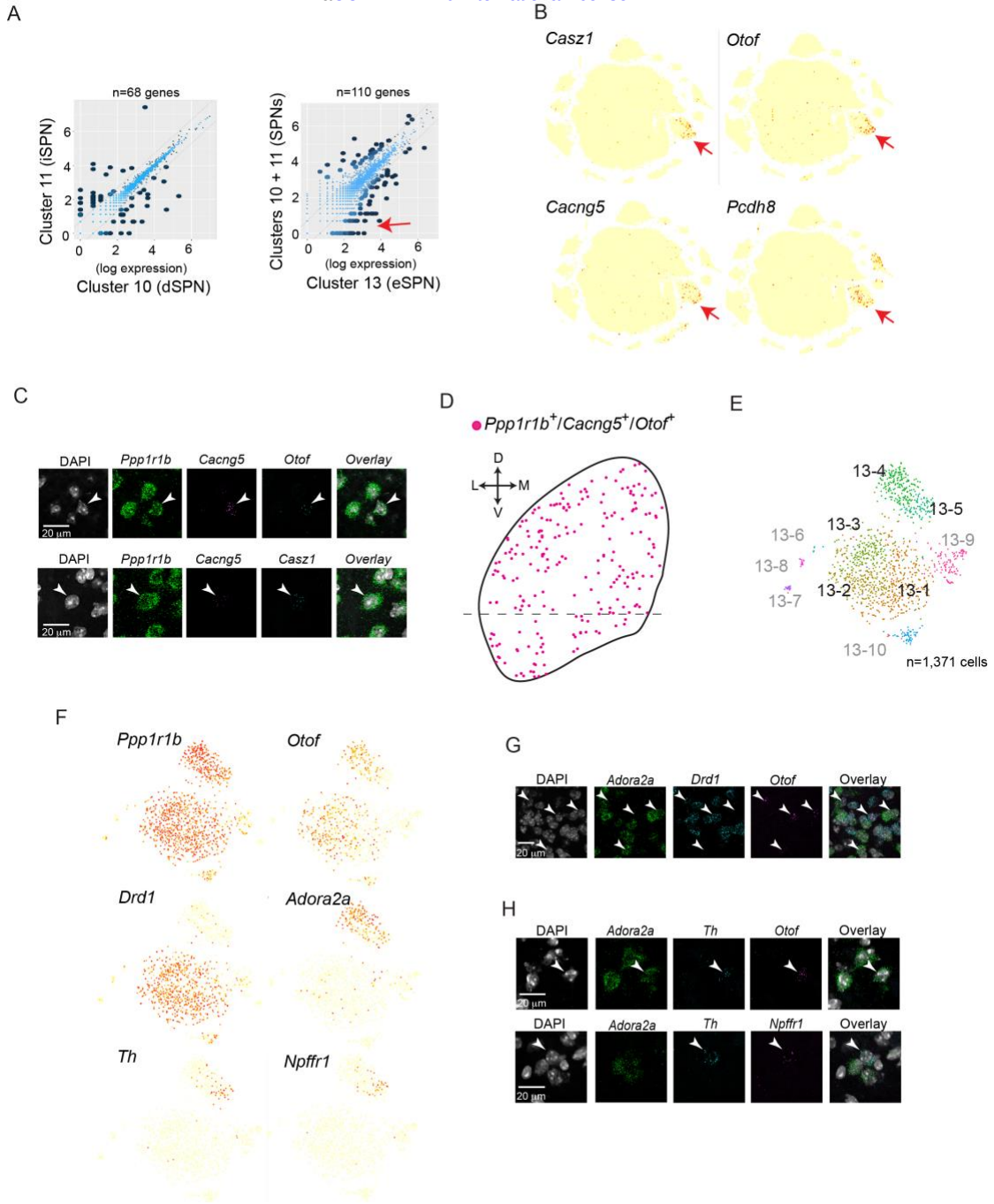


Figure 11. Eccentric spiny projection neurons represent a third axis of SPN diversity

(A) Genome-wide mean expression comparisons between SPN populations (log-normal scale). Left, cluster 10 vs cluster 11 (iSPN vs dSPNs). Right, cluster 13 vs clusters 10 and 11 (eSPNs vs d/iSPNs). Differentially expressed genes are shown with larger, dark dots (>2 fold difference and $P<10^{-100}$, binomTest (Robinson et al., 2010) and total number listed above each plot. Red arrow indicates genes with selective expression in eSPNs vs d/iSPNs.

(B) Expression tSNE plot of n=4 genes (*Casz1*, *Otof*, *Cacng5,* and *Pcdh8*) enriched in eSPNs vs d/iSPNs (red arrow in C). Across all global clusters, genes are highly enriched in cluster 13 (red arrows).

(C-D) eSPNs are anatomically dispersed throughout the striatum.

(C) Single confocal planes from smFISH experiments validating co-expression of pan-SPN (*Ppp1r1b*) and highly-selective eSPN markers (*Cacng5*, *Otof,* and *Casz1*) in dorsal striatum. Top, *Ppp1r1b*, *Cacng5,* and *Otof*. Bottom, *Ppp1r1b*, *Cacng5*, and *Casz1*. Arrowhead indicates triple-positive cells.

(D) Locations of triple positive *Ppp1r1b*, *Cacng5,* and *Otof* cells on a schematic of coronal striatum. Only dorsal striatum was used in the Drop-seq analysis (approximated by the dotted line). D, dorsal; V, ventral; L, lateral; M, medial.

(E) Color-code subclusters from cluster 13. Subclusters 13-1, 13-2, 13-3, 13-4, and 13-5 correspond to eSPNs (83% of cluster 13 cells, black labels). The anatomical identity of other subclusters (17% of cluster 13 cells, gray labels) is described in **Figure S12**. Non-eSPN subclusters were excluded from the differential expression analysis described in (A).

Figure 11 continued…

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(F) Expression tSNE plot of pan-SPN (*Ppp1r1b*), pan-eSPN (*Otof*), dSPN (*Drd1*), iSPN (*Adora2a*), subcluster 13-5 (*Th, Npffr1*) markers.
(G-H) Single confocal planes from smFISH experiments validating co-expression of markers in dorsal striatum. Arrowhead indicates triple-positive cells.
(G) Co-expression of eSPN marker (*Otof*) with iSPN (*Adora2a*) and dSPN (*Drd1*) markers.
(H) Co-expression of eSPN subcluster 13-5 markers. Triple-positive cells observed in dorsal striatum are indicated with white arrowheads. Top, *Adora2a*, *Th*, *Otof*. Bottom, *Adora2a*, *Th*, *Npffr1*.

by exclusion that the subclusters we identify as eSPNs do not correspond to an interneuron type but rather are SPNs.

To establish whether the eSPNs harbor additional molecular diversity, we examined the five subclusters in which pan-SPN (*Ppp1r1b*) and pan-eSPN (*Otof*) markers were expressed (**Figure S12A-C**). These subclusters were divided into two major groups of eSPNs (**Figure 11F. Table S4**), separated by a gene set that included markers used to distinguish canonical iSPNs and dSPNs, such as *Drd1* and *Adora2a*, which we confirmed were expressed in eSPNs with smFISH (**Figure 11G**). eSPN expression of markers typically associated with canonical SPNs suggests eSPNs have been molecularly "camouflaged," including in studies using mice that have employed *Drd1* and *Adora2a* driven–transgenes to label and manipulate dSPNs or iSPNs (Cui et al., 2014; Kozorovitskiy et al., 2012; Kravitz et al., 2010; Oldenburg and Sabatini, 2015). Despite sharing markers, *Adora2a*$^+$ eSPNs and *Drd1*$^+$ eSPNs are distinguished from their canonical SPN counterparts by expression levels of many genes (*Adora2a*$^+$ SPNs: 35 genes; *Drd1*$^+$ SPNs: 96 genes; **Figure S12D, Table S5,6**).

In addition to this major distinction, the Drop-seq data predict additional eSPN diversity, including an ultra-rare eSPN *Adora2a*$^+$ subtype (13-5) that accounts for just 0.3% of all SPNs (n=88 cells). To validate the presence of this ultra-rare type and support the full spectrum of eSPN subcluster diversity implicated by Drop-seq, we performed two triple smFISH experiments for markers of subcluster 13-5. We confirmed instances in the striatum where subcluster 13-5 markers *Th* and *Npffr1* were co-expressed with each other and with the pan-eSPN marker *Otof* (**Figure 11H**). One clue about the anatomical identity of eSPNs comes from this small Th$^+$ population, as spiny Th$^+$ principal cells have been observed in striatum with similar spatial arrangement to eSPNs and appear to be dynamically regulated by dopamine (Darmopil et al., 2008).

We conclude that 1) the eSPN vs SPN distinction represents a third axis of SPN diversity, orthogonal to the dSPN/iSPN and patch/matrix SPNs distinctions; 2) eSPNs harbor rare, additional molecular diversity; and 3) by using markers thought to exclusively distinguish iSPNs from dSPNS, functional studies have lumped eSPNs in with canonical SPNs, in spite of their considerable transcriptional divergence. These results highlight the utility of unbiased, high-throughput single-cell methods for defining neuronal populations, even in brain regions and cell populations that have been extensively studied.

DISCUSSION

The mammalian brain is composed of synaptically interconnected regions, each of which contains a complex mosaic of spatially intermixed cell classes and types. Since Cajal and Golgi, single-cell analyses of cell morphology, cell membranes, and synaptic properties have formed the foundation of our understanding of how structure relates to function in neural circuits. The advent of high-throughput single-cell molecular techniques – such as genome-wide transcript counting with Drop-seq – allow newly systematic approaches to cataloging the essential units of the brain mosaic. Here, we analyzed RNA expression in more than 690K individual cells sampled from nine different regions of the adult mouse brain, encompassing all cell classes. We deployed a novel ICA-based method to disentangle technical effects from endogenous biological signals and highlighted several ways in which such data can identify novel brain cell types, ascertain cell states, and clarify the molecular basis of regionalization across brain circuits and cell classes.

While single-cell profiling provides an exciting opportunity to aid the interrogations of neural circuits, it also presents numerous challenges that we sought to address here. The representations of cells in single-cell datasets can bear an uncertain relationship to their representation in intact complex tissues: we observed quantitative distortions in the representations of cell classes and cell types, driven by library size and cell viability. Such representational issues will be important to consider thoughtfully in future studies in which quantitative cell representation is a key outcome variable, for example, in studies of disease states.

The clusters of cells that are derived from computational analyses cannot be reflexively equated with cell "types". We identified many patterns of RNA expression that corresponded to cell types, but also many that appear to correspond to cell states or to reflect spatial locations in ways that are continuously varying rather than categorical. The examples described above highlight diverse sources of transcriptional variation among individual cells. Comparisons of expression patterns between cell populations can reflect combinations of these influences, and it will be

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

important to be able to parse these component contributions. We developed an analytical approach to help make signal identification – and its relationship to subclusters – more transparent. This approach allowed us to identify a ubiquitous transcriptional program we believe is enacted to maintain axon and presynaptic function, to different degrees both within and across neuron types. We were also able to resolve signals from striatal SPNs representing differences in pathway ("direct" versus "indirect"), spatial arrangement ("patch" versus "matrix"), and eccentricity—a novel molecular SPN distinction we identified in our dataset. The analytical approach we used was critical for being able to recognize such distinctions and be confident of their biological origin.

The size and complexity of single-cell datasets can limit their utilization by researchers in neuroscience or genetics. To facilitate the utilization of such data throughout neurobiology, we developed an interactive web-based software (DropViz, http://dropviz.org/) that facilitates access and dynamic, responsive exploration of the atlas data.

An exciting direction will be to identify functional and anatomical characteristics that correspond to these transcriptional signatures. We hope that high-volume single-cell gene expression profiles, and the patterns present in such profiles, can function as a *lingua franca* for discussing cellular diversity of the adult mouse brain.

Author Contributions

A.S., E.M. and S.A.M. designed the study and wrote the paper. A.S. developed the tissue preparation protocols. A.S., E.M. and M.G. performed Drop-seq and sequenced libraries. A.S, E.M, J.N, N.K., A.W. and S.A.M. developed the ICA analysis pipeline. A.S. and E.M. analyzed the data. D.K. developed DropViz software with help from S.B. M.B. and E.B. performed smFISH experiments. F.K. performed stereological count experiments. S.W. performed the nicotinic receptor analysis. A.G. assisted with cortex regionalization analysis.

Acknowledgments

This work was supported by the Broad Institute's Stanley Center for Psychiatric Research and by a Helen Hay Whitney Fellowship to A.S. We thank Rahul Satija and Andrew Butler for advice on single-cell analysis. M.B. was supported by T32GM007753, T32MH020017, HST Idea^2, The Sackler Scholar Programme in Psychobiology. We thank Dr. Christina Usher for assistance with manuscript preparation and Dr. Gordon Fishell for comments on manuscript drafts.

References

Abbott, N.J., Rönnbäck, L., and Hansson, E. (2006). Astrocyte–endothelial interactions at the blood–brain barrier. Nat Rev Neurosci *7*, 41–53.

Abdi, A., Mallet, N., Mohamed, F.Y., Sharott, A., Dodson, P.D., Nakamura, K.C., Suri, S., Avery, S.V., Larvin, J.T., Garas, F.N., et al. (2015). Prototypic and Arkypallidal Neurons in the Dopamine-Intact External Globus Pallidus. J Neurosci *35*, 6667–6688.

Abusaad, I., MacKay, D., Zhao, J., Stanford, P., Collier, D.A., and Everall, I.P. (1999). Stereological estimation of the total number of neurons in the murine hippocampus using the optical disector. J Comp Neurol *408*, 560–566.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1873.e21.

Albin, R.L., Young, A.B., and Penney, J.B. (1989). The functional anatomy of basal ganglia disorders. Trends Neurosci *12*, 366–375.

Andermann, M.L., and Lowell, B.B. (2017). Toward a Wiring Diagram Understanding of Appetite Control. Neuron *95*, 757–778.

Anderson, D.J. (2012). Optogenetics, sex, and violence in the brain: implications for psychiatry. Bio Psych *71*, 1081–1089.

Antal, M., Beneduce, B.M., and Regehr, W.G. (2014). The Substantia Nigra Conveys Target-Dependent Excitatory and Inhibitory Outputs from the Basal Ganglia to the Thalamus. J Neurosci *34*, 8032–8042.

Bading, H. (2013). Nuclear calcium signalling in the regulation of brain function. Nat Rev Neurosci *14*, 593–608.

Baroncelli, L., Molinaro, A., Cacciante, F., Alessandrì, M.G., Napoli, D., Putignano, E., Tola, J., Leuzzi, V., Cioni, G., and Pizzorusso, T. (2017). A mouse model for creatine transporter deficiency reveals early onset cognitive impairment and neuropathology associated with brain aging. Hum Mol Genet *25*, 4186–4200.

Bolam, J.P., Hanley, J.J., Booth, P.A., and Bevan, M.D. (2000). Synaptic organisation of the basal ganglia. J Anat *196*, 527–542.

Bondjers, C. (2006). Microarray analysis of blood microvessels from PDGF-B and PDGF-R mutant mice identifies novel markers for brain pericytes. The FASEB Journal *20*, 1703–1705.

Boström, K., Zebboudj, A.F., Yao, Y., Lin, T.S., and Torres, A. (2004). Matrix GLA Protein Stimulates VEGF Expression through Increased Transforming Growth Factor-β1 Activity in Endothelial Cells. J Biol Chem *279*, 52904–52913.

Buckner, R.L., and Krienen, F.M. (2013). The evolution of distributed association networks in the human brain. Trends Cogn Sci *17*, 648–665.

Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M.J., Resch, J.M., McCarroll, S.A., et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. Nat Neuro *20*, 484–496.

Carpenter, A.E., Genome, T.J., Lamprect, M., Clark, C., Kang, I.H., Friman, O., Guertin, D., Chang, J.H., Moffat, J., Golland, P., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Gen Bio *7*, R100.

Carter, B.C., and Bean, B.P. (2009). Sodium Entry during Action Potentials of Mammalian Neurons: Incomplete Inactivation and Reduced Metabolic Efficiency in Fast-Spiking Neurons. Neuron *64*, 898–909.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015a). Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090–aaa6090.

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Chen, M.C., Ferrari, L., Sacchet, M.D., Foland-Ross, L.C., Qiu, M.-H., Gotlib, I.H., Fuller, P.M., Arrigoni, E., and Lu, J. (2015b). Identification of a direct GABAergic pallidocortical pathway in rodents. Eur J Neurosci *41*, 748–759.

Chen, Y., Yu, F.H., Sharp, E.M., Beacham, D., Scheuer, T., and Catterall, W.A. (2008). Functional properties and differential neuromodulation of Na(v)1.6 channels. Mol Cell Neurosci *38*, 607–615.

Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics *31*, 545–554.

Chung, W.-S., Welsh, C.A., Ben A Barres, and Stevens, B. (2015). Do glia drive synaptic and cognitive impairment in disease? Nat Neuro. *18*, 1539–1545.

Contois, L.W., Nugent, D.P., Caron, J.M., Cretu, A., Tweedie, E., Akalu, A., Liebes, L., Friesel, R., Rosen, C., Vary, C., et al. (2012). Insulin-like growth factor binding protein-4 differentially inhibits growth factor-induced angiogenesis. J Biol Chem *287*, 1779–1789.

Cui, G., Jun, S.B., Jin, X., Pham, M.D., Vogel, S.S., Lovinger, D.M., and Costa, R.M. (2014). Concurrent activation of striatal direct and indirect pathways during action initiation. Nature *494*, 238–242.

Cui, Q., Pitt, J.E., Pamukcu, A., Poulin, J.-F., Mabrouk, O.S., Fiske, M.P., Fan, I.B., Augustine, E.C., Young, K.A., Kennedy, R.T., et al. (2016). Blunted mGluR Activation Disinhibits Striatopallidal Transmission in Parkinsonian Mice. Cell Reports *17*, 2431–2444.

Darmopil, S., Muñetón-Gómez, V.C., de Ceballos, M.L., Bernson, M., and Moratalla, R. (2008). Tyrosine hydroxylase cells appearing in the mouse striatum after dopamine denervation are likely to be projection neurones regulated by L-DOPA. Eur J Neurosci *27*, 580–592.

DeLong, M.R., Georgopoulos, A.P., Crutcher, M.D., Mitchell, S.J., Richardson, R.T., and Alexander, G.E. (1984). Functional organization of the basal ganglia: contributions of single-cell recording studies. Ciba Found Symp *107*, 64–82.

Fitch, M.J., Campagnolo, L., Kuhnert, F., and Stuhlmann, H. (2004). Egfl7, a novel epidermal growth factor-domain gene expressed in endothelial cells. Dev. Dyn. *230*, 316–324.

Fu, Y., Lai, Y., Wang, Q., Liu, X., He, W., Zhang, H., Fan, C., and Yang, G. (2013). Overexpression of clusterin promotes angiogenesis via the vascular endothelial growth factor in primary ovarian cancer. Mol Med Rep *7*, 1726–1732.

Gerfen, C.R. (1992). The neostriatal mosaic: multiple levels of compartmental organization in the basal ganglia. Annu. Rev. Neurosci. *15*, 285–320.

Gierahn, T.M., Wadsworth, M.H., Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Neuro. *14*, 395–398.

Gittis, A.H., Berke, J.D., Bevan, M.D., Chan, C.S., Mallet, N., Morrow, M.M., and Schmidt, R. (2014). New roles for the external globus pallidus in basal ganglia circuits and behavior. J Neurosci *34*, 15178–15183.

Goldberg, E.M., Clark, B.D., Zagha, E., Nahmani, M., Erisir, A., and Rudy, B. (2008). K+ Channels at the Axon Initial Segment Dampen Near-Threshold Excitability of Neocortical Fast-Spiking GABAergic Interneurons. Neuron *58*, 387–400.

Gotti, C., Zoli, M., and Clementi, F. (2006). Brain nicotinic acetylcholine receptors: native subtypes and their relevance. Trends in Pharmacological Sciences *27*, 482–491.

Graybiel, A.M., and Ragsdale, C.W. (1978). Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining. Proc Natl Acad Sci *75*, 5723–5726.

Guadall, A., Orriols, M., Rodríguez-Calvo, R., Calvayrac, O., Crespo, J., Aledo, R., Martínez-González, J., and Rodríguez, C. (2011). Fibulin-5 Is Up-regulated by Hypoxia in Endothelial Cells through a Hypoxia-inducible Factor-1 (HIF-1α)-dependent Mechanism. J Biol Chem *286*, 7093–7103.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., and Regev, A. (2016). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science *353*, 925–928.

Harwell, C.C., Fuentealba, L.C., Gonzalez-Cerrillo, A., Parker, P.R.L., Gertz, C.C., Mazzola, E., Garcia, M.T., Alvarez-Buylla, A., Cepko, C.L., and Kriegstein, A.R. (2015). Wide Dispersion and Diversity of Clonally Related Inhibitory Interneurons. Neuron *87*, 999–1007.

He, L., Vanlandewijck, M., Raschperger, E., Mäe, M.A., Jung, B., Lebouvier, T., Ando, K., Hofmann, J., Keller, A., and Betsholtz, C. (2016). Analysis of the brain mural cell transcriptome. Sci. Rep. *6*, 1–13.

Hernandez, V.M., Hegeman, D.J., Cui, Q., Kelver, D.A., Fiske, M.P., Glajch, K.E., Pitt, J.E., Huang, T.Y., Justice, N.J., and Chan, C.S. (2015). Parvalbumin+ Neurons and Npas1+ Neurons Are Distinct Neuron Classes in the Mouse External Globus Pallidus. J Neurosci *35*, 11830–11847.

Hochgerner, H., Zeisel, A., nnerberg, P.L.X., and Linnarsson, S. (2018). Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. Nat Neuro. *163*, 1–15.

Hrvatin, S., Hochbaum, D.R., Nagy, M.A., Cicconet, M., Robertson, K., Cheadle, L., Zilionis, R., Ratner, A., Borges-Monroy, R., Klein, A.M., et al. (2017). Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. Nat Neuro. 1–19.

Hu, H., Gan, J., and Jonas, P. (2014). Fast-spiking, parvalbumin+ GABAergic interneurons: From cellular design to microcircuit function. Science *345*, 1255263–1255263.

Hughes, S., and Chan-Ling, T. (2004). Characterization of Smooth Muscle Cell and Pericyte Differentiation in the Rat Retina In Vivo. Invest. Ophthalmol. Vis. Sci. *45*, 2795–12.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. *10*, 626–634.

Karnik, S.K., Brooke, B.S., Bayes-Genis, A., Sorensen, L., Wythe, J.D., Schwartz, R.S., Keating, M.T., and Li, D.Y. A critical role for elastin signaling in vascular morphogenesis and disease. Development *130*, 411–423.

Kita, H. (1994). Parvalbumin-immunopositive neurons in rat globus pallidus: a light and electron microscopic study. Brain Res *657*, 31–41.

Kita, H. (2007). Globus pallidus external segment. Prog Brain Res *160*, 111–133.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell *161*, 1187–1201.

Kozorovitskiy, Y., Saunders, A., Johnson, C.A., Lowell, B.B., and Sabatini, B.L. (2012). Recurrent network activity drives striatal

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

synaptogenesis. Nature *485*, 646–650.

Krabbe, S., Gründemann, J., and Lüthi, A. (2017). Amygdala Inhibitory Circuits Regulate Associative Fear Conditioning. Bio Psych 1–10.

Kravitz, A.V., Freeze, B.S., Parker, P.R.L., Kay, K., Thwin, M.T., Deisseroth, K., and Kreitzer, A.C. (2010). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. Nature *466*, 622–626.

Krienen, F.M., Yeo, B.T.T., Ge, T., Buckner, R.L., and Sherwood, C.C. (2016). Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. Proc Natl Acad Sci *113*, E469–E478.

Lahti, L., Haugas, M., Tikker, L., Airavaara, M., Voutilainen, M.H., Anttila, J., Kumar, S., Inkinen, C., Salminen, M., and Partanen, J. (2016). Differentiation and molecular heterogeneity of inhibitory and excitatory neurons associated with midbrain dopaminergic nuclei. Development *143*, 516–529.

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., and Roeper, J. (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron *57*, 760–773.

Lange, H., Thörner, G., Hopf, A., and Schröder, K.F. (1976). Morphometric studies of the neuropathological changes in choreatic diseases. J. Neurol. Sci. *28*, 401–425.

Letzkus, J.J., Wolff, S.B.E., and Lüthi, A. (2015). Disinhibition, a Circuit Mechanism for Associative Learning and Memory. Neuron *88*, 264–276.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Mallet, N., Micklem, B.R., Henny, P., Brown, M.T., Williams, C., Bolam, J.P., Nakamura, K.C., and Magill, P.J. (2012). Dichotomous Organization of the External Globus Pallidus. Neuron *74*, 1075–1086.

Marder, E., and Goaillard, J.-M. (2006). Variability, compensation and homeostasis in neuron and network function. Nat Rev Neurosci *7*, 563–574.

Mastro, K.J., Bouchard, R.S., Holt, H.A.K., and Gittis, A.H. (2014). Transgenic Mouse Lines Subdivide External Segment of the Globus Pallidus (GPe) Neurons and Reveal Distinct GPe Output Pathways. J Neurosci *34*, 2087–2099.

Mayer, C., Jaglin, X.H., Cobbs, L.V., Bandler, R.C., Streicher, C., Cepko, C.L., Hippenmeyer, S., and Fishell, G. (2015). Clonally Related Forebrain Interneurons Disperse Broadly across Both Functional Areas and Structural Boundaries. Neuron *87*, 989–998.

Ming, G.-L., and Song, H. (2011). Adult Neurogenesis in the Mammalian Brain: Significant Answers and Significant Questions. Neuron *70*, 687–702.

Moser, E.I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. Annu. Rev. Neurosci. *31*, 69–89.

Nehls, V., and Drenckhahn, D. (1991). Heterogeneity of microvascular pericytes for smooth muscle type alpha-actin. J Cell Biol *113*, 147–154.

Okaty, B.W., Miller, M.N., Sugino, K., Hempel, C.M., and Nelson, S.B. (2009). Transcriptional and Electrophysiological Maturation of Neocortical Fast-Spiking GABAergic Interneurons. J Neurosci *29*, 7040–7052.

Oldenburg, I.A., and Sabatini, B.L. (2015). Antagonistic but Not Symmetric Regulation of Primary Motor Cortex by Basal Ganglia Direct and Indirect Pathways. Neuron *86*, 1174–1181.

Oorschot, D.E. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: a stereological study using the cavalieri and optical disector methods. J Comp Neurol *366*, 580–599.

O'Hare, J.K., Ade, K.K., Sukharnikova, T., Van Hooser, S.D., Palmeri, M.L., Yin, H.H., and Calakos, N. (2016). Pathway-Specific Striatal Substrates for Habitual Behavior. Neuron *89*, 472–479.

Poulin, J.-F., Zou, J., Drouin-Ouellet, J., Kim, K.-Y.A., Cicchetti, F., and Awatramani, R.B. (2014). Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. Cell Reports *9*, 930–943.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Gray, L., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *12*, eaam8999–13.

Rudy, B., and McBain, C.J. (2001). Kv3 channels: voltage-gated K+ channels designed for high-frequency repetitive firing. Trends Neurosci *24*, 517–526.

Sahara, S., Yanagawa, Y., O'Leary, D.D.M., and Stevens, C.F. (2012). The Fraction of Cortical GABAergic Neurons Is Constant from Near the Start of Cortical Neurogenesis to Adulthood. J Neurosci *32*, 4755–4761.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat Biotechnol *33*, 495–502.

Saunders, A., Huang, K.W., and Sabatini, B.L. (2016). Globus Pallidus Externus Neurons Expressing parvalbumin Interconnect the Subthalamic Nucleus and Striatal Interneurons. PLoS ONE *11*, e0149798–20.

Saunders, A., Oldenburg, I.A., Berezovskii, V.K., Johnson, C.A., Kingery, N.D., Elliott, H.L., Xie, T., Gerfen, C.R., and Sabatini, B.L. (2015). A direct GABAergic output from the basal ganglia to frontal cortex. Nature *7*, 85–89.

Schüz, A., and Palm, G. (1989). Density of neurons and synapses in the cerebral cortex of the mouse. J Comp Neurol *286*, 442–455.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron *92*, 342–357.

Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M., et al. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell *166*, 1308–1323.e1330.

Smith, K.S., Tindell, A.J., Aldridge, J.W., and Berridge, K.C. (2009). Ventral pallidum roles in reward and motivation. Behav Brain Res *196*, 155–167.

Sternson, S.M. (2013). Hypothalamic survival circuits: blueprints for purposive behaviors. Neuron *77*, 810–824.

Sweeney, M.D., Ayyadurai, S., and Zlokovic, B.V. (2016). Pericytes of the neurovascular unit: key functions and signaling pathways. Nat Neuro *19*, 771–783.

Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. Nature *541*, 331–338.

bioRxiv preprint doi: https://doi.org/10.1101/299081; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.-K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016a). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neuro 1–18.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.-K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016b). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neuro *19*, 335–346.

Tasic, B., Yao, Z., Smith, K.A., Graybuck, L., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2017). Shared and distinct transcriptomic cell types across neocortical areas. bioRxiv 1–37.

Tepper, J.M., and Bolam, J.P. (2004). Functional diversity and specificity of neostriatal interneurons. Curr Opin Neurobiol *14*, 685–692.

Trost, A., Lange, S., Schroedl, F., Bruckner, D., Motloch, K.A., Bogner, B., Kaser-Eichberger, A., Strohmaier, C., Runge, C., Aigner, L., et al. (2016). Brain and Retinal Pericytes: Origin, Function and Role. Front Cell Neurosci *10*, 3171–13.

Trost, A., Schroedl, F., Lange, S., Rivera, F.J., Tempfer, H., Korntner, S., Stolt, C.C., Wegner, M., Bogner, B., Kaser-Eichberger, A., et al. (2013). Neural Crest Origin of Retinal and Choroidal Pericytes. Invest. Ophthalmol. Vis. Sci. *54*, 7910–7912.

van de Kamp, J.M., Betsalel, O.T., Mercimek-Mahmutoglu, S., Abulhoul, L., Grünewald, S., Anselm, I., Azzouz, H., Bratkovic, D., de Brouwer, A., Hamel, B., et al. (2013). Phenotype and genotype in 101 males with X-linked creatine transporter deficiency. J Med Genet *50*, 463–472.

Voorn, P., van de Witte, S., Tjon, G., and Jonker, A.J. (1999). Expression of enkephalin in pallido-striatal neurons. Ann N Y Acad Sci *877*, 671–675.

Wallace, M.L., Saunders, A., Huang, K.W., Philson, A.C., Goldman, M., Macosko, E.Z., McCarroll, S.A., and Sabatini, B.L. (2017). Genetically Distinct Parallel Pathways in the Entopeduncular Nucleus for Limbic and Sensorimotor Output of the Basal Ganglia. Neuron *94*, 138–152.e5.

Waltman, L., and van Eck, N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J. B *86*, 75–14.

Wang, L.P., Li, F., Wang, D., Xie, K., Wang, D., Shen, X., and Tsien, J.Z. (2011). NMDA receptors in dopaminergic neurons are crucial for habit learning. Neuron *72*, 1055–1066.

Wälchli, T., Mateos, J.M., Weinman, O., Babic, D., Regli, L., Hoerstrup, S.P., Gerhardt, H., Schwab, M.E., and Vogel, J. (2015a). Quantitative assessment of angiogenesis, perfused blood vessels and endothelial tip cells in the postnatal mouse brain. Nat Protoc *10*, 53–74.

Wälchli, T., Wacker, A., Frei, K., Regli, L., Schwab, M.E., Hoerstrup, S.P., Gerhardt, H., and Engelhardt, B. (2015b). Wiring the Vascular Network with Neural Cues: A CNS Perspective. Neuron *87*, 271–296.

Wu, Y.E., Pan, L., Zuo, Y., Li, X., and Hong, W. (2017). Detecting Activated Cell Populations Using Single-Cell RNA-Seq. Neuron *96*, 313–328.e317.

Yuan, A., Rao, M.V., Veeranna, and Nixon, R.A. (2012). Neurofilaments at a glance. J Cell Sci *125*, 3257–3263.

Zaborszky, L., Csordas, A., Mosca, K., Kim, J., Gielow, M.R., Vadasz, C., and Nadasdy, Z. (2013). Neurons in the Basal Forebrain Project to the Cortex in a Complex Topographic Organization that Reflects Corticocortical Connectivity Patterns: An Experimental Study Based on Retrograde Tracing and 3D Reconstruction. Cerebral Cortex.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications *8*, 1–12.

Zoli, M., Pistillo, F., and Gotti, C. (2015). Diversity of native nicotinic receptor subtypes in mammalian brain. Neuropharmacology *96*, 302–311.