

Supplementary Material for **Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain**

Lake B. Blue, Rizi Ai, Gwendolyn E. Kaeser, Neeraj S. Salathia, Yun C. Yung, Rui Liu, Andre Wildberg, Derek Gao, Ho-Lim Fung, Song Chen, Raakhee Vijayaraghavan, Julian Wong, Allison Chen, Xiaoyan Sheng, Fiona Kaper, Richard Shen, Mostafa Ronaghi, Jian-Bing Fan,* Wei Wang,* Jerold Chun,* Kun Zhang*

†Corresponding author. Email: kzhang@bioeng.ucsd.edu (K.Z.); jchun@scripps.edu (J.C.); wei-wang@ucsd.edu (W.W.); jianbing_fan@yahoo.com (J.-B.F.)

Published 24 June 2016, *Science* **352**, 1586 (2016)
DOI: 10.1126/science.aaf1204

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S16
Full Reference List

Other Supplementary Material for this manuscript includes the following: (available at www.sciencemag.org/content/352/6293/1586/suppl/DC1)

Tables S1 to S16 as a separate Excel file
Supplementary Files as a separate .zip file

Supplemental Materials:

Materials and Methods

Sample Origin and Nuclei Preparation

All human tissue protocols were approved by the Scripps Office for the Protection of Research Subjects (SOPRS) at The Scripps Research Institute (TSRI) and conform to National Institutes of Health guidelines. Patient 1568 was provided by the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland. Patient 1568 is a 51-year-old female with a post-mortem interval of 22 hours.

Nuclei were prepared using 1% Nonidet-P40 as described previously (10, 22) or nuclear extraction buffer (NEB) (0.32 M sucrose, 5 mM CaCl₂, 3 mM Mg(Ac)₂, 0.1 mM EDTA, 20 mM Tris-HCl (pH=8), and 0.1% Triton X-100) (24) with an OptiPrep™ gradient (Sigma, Optiprep Application Sheet S08) with modifications. Briefly, fresh frozen post-mortem brain tissue was cut and sampled using a razor blade or cryostat and placed in 1 ml of ice-cold NEB for 10 minutes. Nuclei were extracted in a glass dounce homogenizer with Teflon pestle using 10-12 up-and-down strokes in 1 ml of NEB. Following homogenization, samples were passed through a 50 µm filter (Sysmex Partec), incubated in 5 ml NEB for 10 minutes. Samples were spun for 5 minutes at 250-300 x g, the pellets washed in 3 ml PBS + 2 mM EGTA (PBSE), and resuspended in equal volumes of 10% iodixanol solution and Solution D (Sigma, Optiprep Application Sheet S08). Samples were layered onto a 10% and 35% iodixanol gradient, spun at 2500 x g for 20 minutes. A layer of nuclei was visible and collected, washed in PBSE supplemented with 1% fatty-acid free bovine serum albumin (FAF-BSA, Gemini), and blocked in PBSE + 1% FAFBSA for 5 minutes. Samples were incubated with primary antibody (rabbit anti-NeuN monoclonal, 1:1500; MABN140, Millipore), washed, and incubated with APC-conjugated secondary antibody (goat anti-rabbit F(ab')₂, 1:500; Jackson ImmunoResearch). Samples were counterstained with 4',6-diamidino-2-phenylindole (DAPI). High purity single DAPI+/APC+ neuronal nuclei were isolated by flow cytometry using either a FACSAria II (BD Biosciences) or MoFlo Astrios (Beckman Coulter). Nuclei were loaded directly onto a Fluidigm C1 chip (see below). Alternatively, for short term storage, sorted nuclei were supplemented with 50% glycerol and stored overnight at -20 °C before loading onto the C1. For long term storage, sorted nuclei were supplemented with 10% dimethyl sulfoxide (DMSO) and stored in small aliquots at -80 °C until C1 loading. For 50,000-100,000 identically sorted DAPI+/NeuN+ nuclei and whole tissue were used for RNA extraction using the ZR RNA MicroPrep™ Kit (Zymo Research) that included on-column DNase I treatment (per manufacturer's protocol).

Nuclei Loading, mRNA-Seq Library Preparation and Sequencing

For use on the Fluidigm C1 Single-Cell Auto Prep Array for mRNA Seq (Fluidigm, Cat# 100-5761), nuclei were either used directly after sorting, obtained from a glycerol stock stored at -20 °C and thawed rapidly from a DMSO frozen stock stored at -80 °C. Nuclei were loaded at ~120 nuclei/µl (5-10µm

capture sites, small chip) or ~160 nuclei/ μ l (10-17 μ m capture sites, medium chip) as per manufacturer protocols with the following modifications: FAF-BSA was added at 1% to C1 blocking buffer; C1 Suspension buffer was added in a ratio of 7 nuclei to 3 suspension RGT. In select experiments, where indicated (Table S1) C1 Cell Wash Buffer was replaced with C1 DNA Seq Cell Wash Buffer (Fluidigm, P/N 100-7158). For automatic image capture and processing, a plugin for Micro-Manager (25) was developed using Eclipse. The Olympus IX-81 inverted fluorescence microscope and Hamamatsu ORCA-R2 Digital CCD camera were controlled by this plugin to scan across all 96 individual chambers of a C1 chip, and optimally-focused images were automatically chosen to create an overlaid image of nuclei staining and brightfield for each chamber. Nuclei numbers were then counted and recorded manually from these images. cDNA synthesis was performed on chip using the SMARTer® Ultra™ Low RNA Kit for Fluidigm C1™ (Clontech, Cat# 634833), with the following modifications: ERCC RNA Spike-In Mix (Ambion, Cat# 4456740) was used at 1:40,000 dilution in the lysis reaction buffer; a tagged random primer (Integrated DNA Technologies (IDT); 5' AAGCAGTGGTATCAACGCAGAGTACNNNNNN 3') was diluted to 12 μ M in the supplied Clontech Dilution buffer and added to Lysis Mix A for a final concentration of 4.2 μ M; where indicated (Table S1), PolydIdC (Sigma) in 0.4% NP-40 was added to Lysis Mix A for a final concentration of 33ng/ μ l. Harvested cDNA was quantified using Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Cat# P7589). Bulk control libraries were prepared using 1 ng total RNA (or ~120 nuclei for Fig. S1C) using the C1 tube control protocol, but with above reaction modifications and with final PCR amplification reduced from 21 to 18 cycles.

For selective sequencing of libraries generated from only single nuclei, 6 μ L of cDNA per library (289 - 384 libraries per batch, including approximately eight zero capture control sites), were transferred to 96-well plates (Biorad, Cat# 9601), and normalized to 0.3 ng/ μ L in TE buffer (10 mM Tris, 1 mM EDTA pH 8.0), using the STARlet (Hamilton) liquid handling robot. Sequencing library preparation from normalized cDNA was performed as per the Fluidigm protocol (Cat # 100-7168 I1), with several modifications: 10 μ L PCR product from up to 96 uniquely barcoded samples of a single 96-well plate was pooled prior to bead-based purification and sequencing libraries were eluted in 22 μ L TE buffer after two rounds of purification. Purified sequencing libraries were diluted 1:10 in TE buffer and assayed for library size using High Sensitivity DNA Kits (Agilent, Cat# 5067-4626). Quant-iT picogreen (Invitrogen, Cat# P7581) with a lambda DNA standard (Invitrogen, Cat# P7589) were used to quantify pooled library yield. For sequencing, up to 96 samples were sequenced on 2 lanes of a HiSeq Paired-end Flow cell (V3) (Illumina), on a HiSeq 2000/2500 instrument (Illumina), with up to 384 samples being sequenced on all 8 lanes of the flow cell in a single experiment. Libraries were sequenced using 50 bp paired-end sequencing (2 x 50 bp) with dual index reads (2 x 8 bp). Raw sequence Fastq files were generated after sequencing runs using the BaseSpace Fastq generation algorithm (Illumina). Data are available as part of the NIH-supported Single Cell Analysis Program – Transcriptome (SCAP-T) Project (<http://www.scap-t.org>) with the dbGaP study accession phs000833.v3.p1.

RNA-seq data processing

An in-house SNS pipeline was established for automatic single-cell RNA-seq processing and analysis. Short reads were aligned to human hg38 reference genome using STAR (2.3.0) and assembled and quantified by HTSeq (0.5.4p5) with Gencode v20 as annotation. ERCC spike-ins were mapped at the

same time with human genome but quantified separately. A single-cell sample was excluded if the total number of input reads was less than one million or with less than 20% of human genome mapping rate of total uniquely mapped reads. Transcription levels were then converted to transcript per million mapped reads (TPM) and $\log_2(\text{TPM})$ was calculated. A gene was considered expressed if $\text{TPM} > 1$ and genes with more than 99% of missing values across all samples were removed. In total, 16,242 protein-coding genes in 4,039 single-cells were used for the following analyses.

Doublet Screening and removal

For doublet screening, we first performed "clustering and classification" method (below) on the 4,039 quality filtered data sets and then averaged gene expression within each cluster. Samples from multiple-nuclei data sets were assigned to the cluster with the highest Pearson correlation between them. Clusters showing more than 10% of multiple-nuclei data sets were considered as "doublets" and the associated datasets comprising these clusters were removed from further analyses.

Clustering and Classification

Single nuclei clusters were generated and verified by combined unsupervised hierarchical clustering and supervised classification. At each level, (1) gene expression variation was calculated as $\text{CV}^2 = \text{variance}/\text{mean}^2$ for each protein-coding gene across all samples within each cluster; (2) CV^2 were then fitted to an inverse distribution and genes with CV^2 beyond one standard error of the mean were selected as over-dispersed genes; (3) hierarchical clustering was performed using "Pearson correlation" as distance matrix and "ward.D2" as agglomeration method on over-dispersed genes. Two clusters at the first split were determined; (4) a 10-fold random forest feature selection was then performed to select significantly featured differentially expressed genes (DEGs) that separated the two clusters; (5) with selected features, random forest was then applied on the two clusters for verification. To adjust initial classifications, samples with internal vote probabilities > 0.6 for each class were selected as the training set and used to predict the rest of the samples; (6) 100 runs of 10-fold random forest cross validation were performed using selected genes. At each run, predicted classes were re-assigned to test set; (7) With two classes adjusted by cross validation, a random forest model was trained and samples with internal vote probabilities < 0.55 were discarded. (8). DEGs were identified between two verified classes; (9) repeated step 1-8 recursively on the newly formed classes until the random forest out-of-bag estimate of error rate of the specific cluster reached $> 10\%$ or sample size < 40 . As a result, a total of 17 unbiased major classes were identified to separate the 3,227 single neurons after removing doublets. 257 10-fold DEGs (Table S3) were collected along the "clustering and classification" process and dimension reduction plot was generated using the "Rtsne" package in R. The clustering and classification algorithm was provided in R script with instructions in Supplementary files.

Expression Analyses

Scatter plots and dendrogram (hclust method = ward.D2) plots comparing single, averaged and bulk expression values were generated using R software on $\log_2(\text{TPM}+1)$ expression values. Identification

of unique subgroup marker gene expression (Table S5) was performed separately for In or Ex groupings using Seurat software (Version 1.2) using the find_all_markers function (thresh.test = 2, test.use = "roc") on $\log_2(\text{TPM}+1)$ values generated from combined exon and intron reads. All genes with a power of 0.4 or more were used to generate corresponding heatmaps (Fig. 1C) using the Seurat software. Pie charts showing relative proportions of subtype populations (Table S7) across BA regions were generated using Microsoft Excel. All fraction of positive values for each interneuron and projection neuron group were calculated separately from exon-only TPM values using an established excel macro (2) and a threshold of expression of five percent of the calculated maximum. Heatmaps of fraction of positive values for select genes were generated using the R package gplots. All violin plots showing gene expression values (Fig. 2, 3, 4, S1, S12, S14) were calculated from exon only TPM values ($\log_2(\text{TPM}+1)$) using the Seurat software. For markers differentially regulated between Ex2 and Ex3 (Fig. 4C), the Seurat function find.markers (Ex2, Ex3 , thresh.use = 2, test.use = "roc") was used on $\log_2(\text{TPM}+1)$ values generated from combined exon and intron reads. Genes with a power of 0.4 or more were used to generate violin plots from exon only derived TPM values ($\log_2(\text{TPM}+1)$). For genes differentially expressed between BA17 and BA41/42 within subgroup Ex3 (Fig. 4D), the find_all_markers (thresh.test = 2, test.use = "roc") function was used only on these two sample groups and on $\log_2(\text{TPM}+1)$ values generated from combined exon and intron reads and the resultant heatmap generated using Seurat software. For genes variant by subgroup, each In and Ex subgroup were independently analysed using the mean.var.plot function (y.cutoff = 2, x.low.cutoff = 2, fxn.x = expMean, fxn.y = logVarDivMean) of the Seurat software. For genes differentially expressed between BA regions within each of the 16 clusters, variable genes were selected using one-way analysis of variance (ANOVA) with fold change > 10 and adjusted p value < 0.05. Venn diagrams were generated using jvenn (26). Gene ontology or GO analyses were performed either using the ToppFun function of the ToppGene suite (toppgene.cchmc.org) using default settings or using the ClueGO (v2.1.5) application in Cytoscape (v3.2.0) with the following settings: Biological Processes (02.02.2015), GO terms level: 3–8; GO term restriction: 2 genes and 4%; evidence code: all. Significance cutoff was set at a Bonferroni adjusted p value of 0.05.

RNA ISH and quantification

For RNA ISH (Fig. 3D, Fig. S11, Fig. S13), adjacent sections were thawed, fixed in neutral buffered formalin, and stained using the RNAscope Multiplex Fluorescence Kit or the RNAscope Brown Chromogenic Kit according to manufacturer's instructions (Advanced Cell Diagnostics, Cat# 322310 and Cat #320851) (27). After staining, the sections were tile-imaged on an Axio Imager 2 equipped with ApoTome2 (Carl Zeiss) and stitched. Images were taken across each of the layers, totaling approximately 24 images. Background lipofuscin autofluorescence was subtracted using ImageJ. RNAscope signals were manually quantified by at least two individuals, their results averaged, and the cell counts were graphed. Each section was imaged in two locations to ensure reproducibility. See Table S15 for detailed methods for each figure panel. For RNA ISH from the visual and temporal cortices (Fig. 3D, Fig. S11, Fig. S12, Fig. S14), representative images were obtained from the Allen Human Brain Atlas (<http://human.brain-map.org>) and corresponding links are provided in Table S11.

Laser capture microdissection (LCM)

Fresh frozen post-mortem cortical tissues were sectioned at thicknesses ranging from 10-20 μ m and mounted on nuclease-free PEN membrane slides (Carl Zeiss, Cat# 415190-9081-000), stained briefly in hematoxylin solution to visualize cells and layer architecture, dehydrated through 70%, 95%, and 100% ethanol (70% and 95% ethanol prepared with DEPC-treated water), and subjected to laser cutting and catapulting. Cutting and catapulting laser power was optimized for tissue thickness and objective magnification, and individual cells were catapulted into clear Adhesive Cap PCR tubes (Carl Zeiss, Cat# 415190-9191-000) or 96-well collector plate (Carl Zeiss, Cat# 415190-9151-000).

Supplementary Text

Single Nucleus RNA Sequencing Pipeline

Nuclei Preparation. Neuronal nuclei were isolated from non-neuronal (e.g. glial) nuclei using NeuN staining (Fig. S1). As expected, a clear separation of neuronal and non-neuronal cell types was found in NeuN positive and negative fractions, respectively (Fig. S1B, C). Furthermore, while single neurons showed a broad range in NeuN (or RBFOX3) expression, no bias in the sampling of these nuclei was observed (Fig. S1D). Therefore, NeuN sorting of cerebral cortical tissue provided an effective method for specifically analyzing neuronal nuclei.

SmartPlus Modifications. To improve efficiency of the SmartSeq method (28) for capture of nuclear RNA, we made several modifications that we refer to as SmartSeq Plus, for use in Fluidigm C1 microreactors (29). By including a random primer for cDNA synthesis, we achieved a more uniform transcript coverage compared to SmartSeq (Fig. S3), especially for long transcripts and protein coding transcripts, without affecting global gene expression values or introducing excessive ribosomal RNA reads (Fig. S4A-C). Presumably this improved coverage was due to a fraction of RNA molecules in post-mortem nuclei being either incompletely processed or fragmented and not accessible to poly-T priming. Additionally, we were able to enhance sampling depth by nuclei cryopreservation for multiple cDNA amplification experiments from a single nuclei preparation, without affecting gene mapping rates or gene expression values (Fig. S4D,E). To reduce sampling bias associated with nuclear size, we incorporated C1 chips with smaller capture sites. This permitted sampling of smaller nuclei exhibiting reduced gene mapping rates (Fig. S4D). To recover these nuclei, poly(dIdC) (30) was included with the SmartSeq Plus chemistry, leading to the efficient recovery of gene mapping rates without altering global gene expression values (Fig. S4D-E). To ensure consistency of our pipeline, we compared our single nuclei data sets from the frontal cortex (BA10) with data sets derived from two separate brains (BA9) obtained from a different repository (Fig. S4F). The high concordance of these data sets demonstrates the general applicability of our pipeline across patients or tissue sources.

Quality Control. In total, we processed 86 C1 chips with an average single capture rate of 53 percent, and sequenced 4,488 single nuclei to an average depth of 8.34M reads (Table S1, Fig. S5). ERCC spike-in RNA transcripts (31) confirmed high technical consistency across C1 runs (mean Pearson $r = 0.85$, Fig. S5) and were used to define quality cutoff parameters (see Supplemental Materials and Methods). As such, data sets that had fewer than 1 million reads and/or a gene mapping ratio to ERCC spike-in transcripts that was less than 20% were removed, leaving 4,039 quality filtered data sets with

an average detection of 6,159 genes having more than ten read counts (Fig. S5C). It is known, however, that a proportion of single C1 captures may actually represent multiple captures via masking of one cell by another at the capture site (32). Since the existence of such “doublets” would confound analyses, we performed an initial screen for their identification and removal guided by a set of known duplicate captures. Clusters showing high correlation to data generated from two nuclei were considered to also represent masked “doublets”, a conclusion supported by contradictory gene expression profiles (Fig. S6A) and centric positioning on multidimensional plots (Fig. S6B). These duplicates were prevalent in medium C1 chips (Fig. S6C,D), indicating that the larger capture sites promoted this technical artifact. Using these considerations, we eliminated prospective “duplicate” clusters from further analyses, thereby arriving at 3,227 single nucleus data sets passing quality filtering metrics (Tables S1-S2, Fig 1A).

Validation. To understand how well the transcriptome profile of neuronal nuclei represents whole cells, we compared data from single neuronal nuclei, pooled neuronal nuclei, and whole tissue from the same cortical samples (Fig. S7). Averaged single nuclei expression data showed high correlation with bulk nuclei, but less so with bulk tissue containing non-neuronal cell types (Fig. S7A, B). While single nuclei showed unique expression values, a higher correlation with the bulk samples could be achieved with as few as 10 single nuclei (Fig. S7C, D), underscoring both commonality of the global neuronal transcriptome and the ability of pooled neurons to mask their individual transcriptomes. Single neuronal nuclei showed weaker correlation with whole brain tissues from which they were derived, indicating possibly unique neuronal expression profiles. Consistent with this, single neuronal nuclei showed higher correlation with bulk samples on the basis of known neuronal, but not glial marker gene expression (33) (Table S16, Fig. S7B). Therefore, RNA sequencing data from isolated nuclei retained the ability to accurately predict their cellular identity.

Clustering. We developed a novel algorithm, “clustering-and-classification” (See Supplemental Materials and Methods) to identify neuronal subtypes. This combined method progressively splits single nuclei using unsupervised hierarchical clustering and repeatedly validates the clusters by a classification method random forest to ensure small out-of-bag error (Fig. S8A). Furthermore, the most informative genes that were selected in random forest at each round represent the markers of each cluster (Fig. S8B,C, Table S3). This method is easy to implement and combines the advantages of unsupervised and supervised learning. The “clustering-and-classification” R script is provided in Supplementary files with instructions.

Neuronal Subtypes. Neuronal subtypes were broadly defined on the basis of inhibitory or excitatory marker gene expression (Fig 1B, Table S3). To ensure consistency with prior gene expression studies, we chose to show fraction-of-positive values and expression-level plots from only exonic reads, even though intronic sequences clearly predict expression profiles that are supported by exon-only expression (Fig. S8B). Interestingly, excitatory neuronal subtypes showed less distinct expression differences compared with inhibitory neurons (Fig. S8C), suggesting greater transcriptional diversity of individual inhibitory neurons.

Inhibitory neuron or interneuron subtypes were distinguished on the basis of known marker genes associated with cortical layers, developmental origin, and interneuron classification (Fig. 2B, Fig. S11). Consistent with LGE/CGE-derived interneurons, upper cortical layer subtypes (*CXCL14*, *CHRNA7*, *CNR1*) were VIP+, RELN+, CR+ and showed positive expression of *SP8* and *NR2F2* (also known as *COUP-TFII*) (Fig. 2B, C). Alternatively, interneuron subgroups that were PV+, CB+, SOM+, nNOS+, and NPY+ showed MGE marker gene expression, including *LHX6* and the *LHX6*-target genes *SOX6*, *SATB1* and *MAFB* (34, 35) (Fig. 2B). These potential MGE-derived subgroups additionally expressed markers associated with localization to lower layers, such as the *SULF1*-expressing PV+ subtype (In6) localized to layers 4-5 (Fig. S11B). Furthermore, we were able to distinguish RELN+ populations originating from each of these developmental origins; the LGE/CGE-derived In1 population likely occupying layers 1/2 (e.g. *CXCL14*, *CHRNA7*, *CNR1*); and the In4 cluster potentially occupying deeper layer 3 (e.g. *SV2C*) (Fig. 2B, C, Fig. S11B). This latter population appears consistent with a previously identified SOM+ NR2F2+ population derived from the MGE (19) and shows specific absence of *RELN/SST* expression in BA17 (Fig. 2C, Fig. S11B, D). Furthermore, our dataset revealed three distinct VIP+ interneurons (In1-3). The latter population (In3), specifically expressed *PDE9A* (Fig. 2C), a potential intracellular mediator of the serotonin receptor HTR2C (36), and showed localization to layers 2/3 (Fig. S11B, E), highlighting a putative neocortical region-specific serotonergic signaling activity.

Excitatory neuron or projection neuron subtypes could be classified on the basis of their layer position within the cortex (Fig. 3). Subgroup Ex1 showing cortical projection neuron (CPN)-associated *CUX2* expression (16) in conjunction with layer 2/3 marker genes *LAMP5* and *GLRA3* (Fig. 3B,C), represents the most prevalent subtype in our datasets. We further identified distinct, putative subcortical projection neurons (SCPNs) associated with differential combinations of *RORB*, *FOXP2*, and the SCPN determinant *FEZF2* (16) (Fig. 3C). However, while Ex6 shows positivity for the layer 5 marker *HTR2C*, it additionally shows expression of layer 6 marker *TLE4* (Fig. 3B,C), indicating potential crossover of this subgroup between layers and indicating classifications that may not simply fit with expected layer positions. Layer 6 subgroups (Ex7-Ex8) showing *OPRK1* expression can further be resolved with *ADRA2A/NR4A2/CTGF* expression (Ex8) that is associated with layer 6b or white matter (17). Therefore, our dataset identifies layer specific excitatory neurons associated with pyramidal identity (Fig. 4A), having both expected and unexpected spatial marker expression.

In addition to these projection neuron subtypes, we detected two distinct *CUX2/RORB* expressing layer 4 granular neuron subtypes (Ex2, Ex3) (Fig. 3B,C), characterized by expression of a CNS specific transcription factor *BHLHE22* (Fig. 3C) and confirmed by RNA *in situ* hybridization (ISH) (Fig. S13A,D). Comparative distribution of *BHLHE22* positive neurons with those positive for the layer 2/3/6 marker *CBLN2* and the layer 5/6 marker *PCP4* (17) provided confirmation of the spatial order of excitatory neuron subtypes (Fig. S13A,B).

Neuronal subtypes identified in our data set revealed regional differences associated with the relative proportions of subtypes across BAs as well as BA-specific transcriptomic profiles within subtypes (Fig. 4). To further demonstrate the latter, we examined genes having known variability between the visual and temporal cortices from ISH studies (17) for transcriptomic differences that may be attributed to subtypes defined by our data set (Fig. 4E, Table S11). We found that the apparent downregulation of *DKK3* and upregulation of *SYT2* in the visual cortex can be attributed predominantly to the *SULF1* expressing layer 4 interneuron subgroup In6 (Fig. 2C, Fig. 4E, Fig. S12, Fig. S14). Further, the SOM+

population present in the temporal cortex, but not the visual cortex, can be attributed to the *SV2C* positive layer 3 subgroup, In4 (Fig. 3C, Fig. 4E, Fig. S12, Fig. S14). Excitatory subtypes also showed regional variation, with differential expression of the deep layer markers *SYT6* and *TLE4* contributing to the Ex6 subgroup, while expansion of *PCP4* into more superficial layers in the visual cortex contributed to Ex1 and Ex3 (Fig. 4E, Fig. S14).

Supplemental Figures:

Figure S1. Overview of single nuclei sampling methodology. **A.** Schematic of human brain at the level of BA8 showing approximate region sampled, typical tissue quantity processed for fluorescent activated cell sorting (FACS), the approximate proportion of NeuN+ nuclei obtained, the quantity of NeuN+ nuclei needed for a single C1 loading, and the average single nuclei capture rate. Expected sample scaling and minimal tissue needed for a single C1 experiment is summarized. **B.** Samples generated using pooled sorted NeuN+ nuclei from BA8, BA10, BA17, BA21, BA22 and BA41/42 as well as matching tissue sections were analyzed for expression of oligodendrocyte (Oligo.), astrocyte (Astro.), endothelial (Endo.) and neuronal (Neuro.) marker genes (17). Violin plots show expression values for associated nuclear (Nuc.) and tissue (Tiss.) sample groupings. **C.** Data sets from ~120 pooled nuclei derived from either BA21 or BA17 were used to confirm enrichment for neurons or glia in NeuN⁺ and NeuN⁻ sorts, respectively. **D.** Histograms showing the frequency of all single NeuN+ nuclei analyzed in this study binned by level of *RBFOX3* (NeuN) expression. Neuronal nuclei were distinguished on the basis of either *SLC17A7* (excitatory) or *GAD1* (inhibitory) marker gene expression.

Figure S2. Limited RNA recovery from laser capture microdissection (LCM) of post-mortem brain. **A.** Fresh-frozen BA8 cerebral cortex sections stained with hematoxylin were subjected to LCM; well-separated cells were manually outlined by software, **(B)** cut out of the tissue, and **(C)** projected into 96-well caps that were visually verified. Scale bar = 75 μ m. **D.** Bioanalyzer trace results showing total RNA yields extracted from 100 cells either hand cut or isolated using LCM from fresh frozen brain sections. Results were compared with 1000 pg control human brain reference RNA.

Figure S3. SmartSeq Plus provides more uniform transcript coverage. Exon read coverage of all transcripts from SmartSeq and SmartSeq Plus libraries prepared from RNA extracted from bulk sorted BA8 neuronal nuclei.

Figure S4. Protocol comparison and improved sampling depth. **A.** Scatter plots showing high Pearson correlation coefficients (r) for expression values from all protein coding genes averaged across 10 single BA8 nuclei libraries generated using SmartSeq, SmartSeq Plus (SmartPlus) and SmartSeq Plus containing PolydIdC (SmartPoly) protocols. **B.** Proportion of reads mapped to different gene types for bulk tissue (t) and bulk nuclei (n) datasets generated using the SmartSeq Plus protocol, as well as single nuclei (1-3) datasets generated using the indicated protocols. **C.** Total number of genes detected averaged over indicated number of single nuclei datasets that were generated using the different indicated protocols. Arrow indicates improved protein-coding gene detection with SmartSeq Plus compared to the standard SmartSeq protocol. **D.** Ratio of mapped reads that were to either ERCC or genes for a single preparation of BA8 nuclei, comparing libraries: from nuclei processed immediately after sorting (no preservation); after cryofreezing (frozen with DMSO) using a medium C1 chip; after cryofreezing (frozen with DMSO) using a small C1 chip; after cryofreezing using a small C1 chip and the SmartPoly protocol. **E.** Corresponding comparison of conditions using scatter plots from averaged expression values (10 single nuclei, all protein coding genes) that show high correlation between conditions. **F.** Comparison of SmartPoly protocol across different brains using scatter plots from averaged expression values (10 single nuclei, all protein coding genes) that show high correlation between brain samples from different repositories (Brain 1 = BA10 tissue from patient 1568, NICHD Brain and Tissue Bank; Brain 2 = BA9 tissue from PT-WZTO, Genotype-Tissue Expression (GTEx) Biobank; Brain 3 = BA9 tissue from PT-NPJ8, GTEx Biobank).

Figure S5. Mapping statistics. **A.** Top panel: Proportion of all reads that were: unmapped; multi-mapped; uniquely mapped to ERCC transcripts; uniquely mapped to reference genes (exon and intron regions); or uniquely mapped between genes (intergenic). Middle panel: relative proportion of reads uniquely mapped to: ERCC transcripts; intergenic regions; or reference genes. Bottom panel: the proportion of reads uniquely mapped to the genome that were associated with: intergenic regions; introns; or exons. Results are shown for all 0 capture (0C), single capture (1C) and multiple capture (2+C) nuclei libraries (top and middle panels) or single capture only (1 Nuclei, bottom panel). **B.** Plots showing the frequency distribution of total number of reads sequenced and ERCC Pearson correlation r values [$\log_{10}(\text{counts}+1)$ versus $\log_{10}(\text{concentration})$] for all single nuclei libraries. **C.** Plots showing the frequency distributions of genes detected across single nuclei libraries using different gene count cutoffs.

Figure S6. Doublet screening and filtering. **A.** Heatmap of expression for excitatory (*SLC17A7*, *SATB2*, *CBLN2*) and inhibitory (*GAD1*, *GAD2*, *SLC6A1*) marker genes across 30 groups of neuronal nuclei data sets generated from the first round of clustering and classification. Arrow indicates cluster CL8 showing co-expression of these marker types. **B.** Multidimensional plot showing the 30 identified clusters and known two capture (2C) data sets (DIM, Dimension). Clusters with high overlap of 2C data sets are indicated. **C.** The percentage of data sets contributing to each cluster was calculated separately for small C1 chips and medium C1 chips and compared with the proportion of 2C data sets associated with each cluster. Arrows indicate clusters showing the highest number of prospective “doublets”. **D.** Percentage of identified “doublets” and their association with use of medium C1 chips across successive C1 runs.

Figure S7. Single-nuclei RNA sequencing permits cell type identification. **A.** Cluster dendrogram of gene expression ($\log_2(\text{TPM}+1)$) using all protein coding genes, a subset of glial marker genes or a subset of neuronal marker genes (Table S16). Analyzed samples were generated using pooled sorted neuronal nuclei (n) from BA8 section 7 (s7n) or section 9 (s9n), BA10, BA17, BA21, BA22 and BA41/42 (BA41n) as well as matching tissue (t) sections using the SmartSeq Plus protocol or original SmartSeq protocol (BA8s7n(Smart)). Brain section numbers are assigned according to the University of Maryland Brain and Tissue Bank (Brain sectioning – Protocol Method 2). For comparison, single nuclei data sets from combined excitatory (Ex) and inhibitory (In) subtypes (n = 3084) were averaged (AveSN). **B.** Scatter plots comparing averaged single nucleus data and averaged bulk sorted nuclei or tissue data for protein coding, neuronal or glia marker genes. **C.** Scatter plots comparing single nucleus data, averaged 10 single nuclei data, or averaged 100 single nuclei data from BA21, with data from matched bulk nuclei or tissue (protein coding genes). **D.** Representative scatter plots comparing single nucleus data sets (protein coding genes). Associated Pearson r values are indicated (B-D).

Figure S8. Overview of subtype clustering and classification. **A.** Sample splitting at each step of the clustering and classification strategy showing: number of nuclei at each level; genes associated with each splitting (A-W, Table S3); final clusters associated with excitatory (Ex) and Inhibitory (In) neurons. Outlier cluster (n = 44, NoN) in the inhibitory branch is indicated by a dark circle. **B.** GO annotations associated with differentially expressed genes (DEGs) defining cluster splitting (fold change or FC ≥ 2) within excitatory or inhibitory subgroup branches (**A**) (Bonferroni adjusted p values < 0.05) (Table S4). **C.** Proportion of genes used for each branch of excitatory or inhibitory subgroup clustering (**A**) that were differentially expressed either between 2 and 10-fold or greater than 10 fold.

Figure S9. Differentially expressed subtype marker genes. **A.** Heatmap showing expression of 10-fold or more differentially expressed genes (Table S3) used for multidimensional plotting of neuronal subtypes (Fig. 1B, Fig. 4A, Fig. S10A). **B.** Heatmaps showing consistency in expression (top panel, TPM calculated from exon and intron reads) and corresponding fraction of positive values (bottom panel, TPM calculated from only exon reads) for unique marker genes (Table S5) identified for each neuronal subtype.

Figure S10. Neuronal subtypes do not show batch bias. **A.** Multidimensional plots as shown in Fig. 1B for neuronal subgroups indicating each experimental C1 run (Table S1). Arrow indicates an outlier cluster ($n = 44$, NoN, Table S2) derived from the 20150122B C1 run. **B.** Multidimensional plot using ERCC expression values showing all clusters identified in our analyses, demonstrating that unlike the outlier cluster indicated in **(A)** (arrow), our neuronal subtypes were identified independently of random batch specific expression differences.

Figure S11. Confirmation of subtype identity by RNA ISH. **A.** Fraction of positive heatmap of inhibitory (expressing *GAD1*) and excitatory (expressing *SLC17A7*) subtypes and a subset of representative combinatorial marker genes. **B.** Allen Human Brain Atlas ISH data (Table S11). Cortical stains are oriented from layer 1 (L1) to layer 6 (L6). **A and B.** Numbered boxes represent In subtype-specific combinations (e.g. 1 = In1) with spatial orientation in the cortex indicated through RNA ISH. Box 1 or In1 represents a layer 1 *VIP⁺CNR1⁺RELN⁺* subgroup. Box 2 or In2 represents a *VIP⁺CNR1⁺RELN* layer 2/3 subgroup that co-expresses *OPRK1*, as confirmed by neurons co-stained for *OPRK1* and *GAD1* in this region (**C**) (indicated by * in **B** and **C**). *OPRK1* also labels *SLC17A7* expressing layer 6/6b Ex7/Ex8 neurons (region indicated by ** in **B** and **C**). Box 4 or In4 represents a layer 3/4 *RELN⁺SV2C⁺* subgroup that co-expresses *SST* in BA8, BA10 and BA21, but not in BA17, BA22 or BA41/42 (Fig. 2C). Consistently, In4-associated *SST* expression can be found within the temporal (Temp.) cortex (BA21) but not the visual cortex (BA17) (**B**). **D.** RNAscope co-staining of *RELN* and *SST* in BA8 and BA17 showing co-positive cell distributions that are consistent with RNA-seq data. Box 6 or In6 represents a layer 4/5 subgroup co-expressing *SULF1* and *PVALB*. Box 3 or In3 represents a *VIP⁺ RELN⁺* subgroup positive for *PDE9A*, which is also specifically expressed in the layer 5 Ex6 subgroup and shows a consistent expression pattern with *HTR2C* (**A**). **E.** *PDE9A*-positive In3 and Ex6 expression was confirmed by RNAscope co-staining with *GAD1* (double positive restricted to layers 2/3) and *SLC17A7* (double positive restricted to layers 5/6). **F.** RNAscope counts consistent with RNA-seq data.

Figure S12. Subtype comparison with a recent mouse study on the somatosensory cortex (3). **A.** Violin plots showing core interneuron marker genes in In subtypes and across BA regions, indicating a similar proportion of VIP+, SOM+ (*SST*) and PV+ subtypes between mouse and human, with the exception of additional *RELN*+ subtypes likely associated with differences in sampling methods between studies. A schematic for combinatorial expression summarizes species-specific differences. Int = mouse subtypes identified (3); In = human subtypes. **B.** Violin plots of excitatory markers used to define mouse pyramidal subtypes showing associated expression in human Ex subtypes and across BA regions. A high concordance in the pattern of expression can be found using both human subtypes and Allen Human Brain Atlas ISH data (Table S11). Cortical images are oriented from left (pial layer) to right (white matter) and regions sampled are indicated. Associated species-specific similarities or differences in observed cortical layer identity are summarized for each marker gene, including a shift in SCPN marker *THSD7A* in mouse to CPN in human and an observed shift in a claustrum pyramidal

(clauPyr) neuronal subtype in mouse to layer 6b in human that may reflect evolutionary changes or differences in the regions examined between these studies.

Figure S13. Confirmation of layer identity in BA8 by RNA ISH. **A.** RNA *in situ* hybridization (ISH) analyses on BA8 cortical sections showing counts of positive cells for *CBLN2* (Layers 2/3/6), *BHLHE22* (Layer 4) and *PCP4* (Layer 5) in image fields spanning from the pial layer (upper cortex) through to the white matter (lower cortex). Violin plots are corresponding gene expression values for excitatory neuron subtypes across all BA regions. **B.** RNAscope technology was used to stain BA8 sections for the layer markers *RELN*, *MFGE8*, *PCP4* and *CBLN2*. Single-brown chromogenic or fluorescent staining (top) and average counts (bottom) are shown. Insets are of representative single positively stained neurons. Positive counts were derived from over 22 vertical sections spanning from pial surface and upper cortex (top) to lower cortex. Positive cells were counted by two independent observers over two independent regions. **C.** Fraction of positive heatmap showing the layer 4 specific marker *BHLHE22* in Ex2 and Ex3 subgroups (*) which have correspondingly low positivity of its negatively regulated target gene *CDH11* (37). **D.** Representative RNAscope images of *BHLHE22* positive cells used for counts shown in **(A)**, and which also show the expected absence of *CDH11*. Inset is a representative *CDH11* positive cell that is negative for *BHLHE22*. The proportion of *BHLHE22* positive cells having an absence or presence of *CDH11* expression is shown through RNAscope counts and is consistent with RNA-seq data.

Figure S14. Neuronal subtype heterogeneity between brain regions. **A.** Violin plots showing expression values (black boxes) for genes with known differential expression between BA17 (visual cortex) and BA21 (temporal cortex) (17). Stains are associated RNA ISH analyses of cortical sections oriented with pial layer at the top (Allen Human Brain Atlas, Table S11). Double arrows indicate regions associated with the indicated differential expression. **B.** Violin plots showing indicated marker gene expression values for: each inhibitory (In) and excitatory (Ex) subtype and brain region (bar colors, bottom); each brain region from combined neuronal (NeuN^+) data; each brain region from the specific subgroup showing BA17/BA21 expression differences (black arrows) associated with RNA ISH staining differences shown in **(A)**. Additional subtypes that may account for these RNA ISH differences are indicated (gray arrows).

Figure S15. Subgroup variable genes. **A.** Plot of average expression and dispersion (binned $\log(\text{Variance}/\text{mean})$) for subgroup Ex1, indicating genes that show variance (z-score cutoff = 2) across these single nuclei (Table S12). **B.** Multidimensional plot on Ex1 nuclei using genes identified as being differentially expressed between BA regions of this subgroup (10-fold cutoff, Table S13). Nuclei show a distribution consistent with their spatial origination (Occipital, Temporal, Frontal lobes). DIM = dimension. **C.** Venn diagrams showing overlap between: all subgroup-derived variable genes (**A**, Table S12); all differentially expressed genes between BA regions for each subgroup (**B**, Table S13); genes defining subgroup clustering (DEG, Table S3); genes associated with the human post-synaptic density (hPSD) (38); and genes within the top five percentile for stable expression differences across cortical parcels (DS (Cortex)) (7).

Figure S16. Subtype expression patterns of electrophysiological-relevant genes. Top panels: fraction of positive values for ion channel and neurotransmitter-related genes (see Table S14) are shown for In and Ex subtypes (minimum FOP value of 0.1 in at least one cluster). Bottom panels: Fraction of

positive values for select genes are shown for Ex subtypes, demonstrating potential for unique subtype-specific electrophysiological properties.

Supplemental Tables:

Table S1: Summary of C1 experimental conditions and outcome

Table S2: Single nuclei library details and group/subgroup identifiers

Table S3: Differentially expressed genes (fold change indicated as ≥ 2 and ≥ 10) underlying group cluster separation (**A-W**, See Fig. S8a), where "left" denotes genes upregulated in the left branch and "right" denotes genes upregulated in right branch.

Table S4. GO annotations for differentially expressed genes defining subgroup classifications (Table S3)

Table S5: Unique group specific genes and associated fraction of positive values using exon only derived TPM (see Fig. S9b)

Table S6. GO annotations for unique subgroup marker genes (Table S5)

Table S7: **A.** Distribution of brain regions amongst classification subgroups. **B.** Relative distribution of brain regions amongst classification groups on the basis of normalized input contributions (values are percentage of total within the group)

Table S8: Genes differentially expressed between Ex2 and Ex3 subgroups

Table S9. GO annotations for genes differentially expressed between Ex2 and Ex3 subgroups (See Table S8)

Table S10: Genes differentially expressed between BA41/42 and BA17 brain regions within the Ex3 subgroup

Table S11. Allen Human Brain Atlas ISH citations (See Fig. 3D, Fig. S11, Fig. S12, Fig. S14)

Table S12. Variable genes identified within each subgroup

Table S13. Genes identified in each subgroup having 10-fold or more difference in expression level between at least two Brodmann Areas (BA)

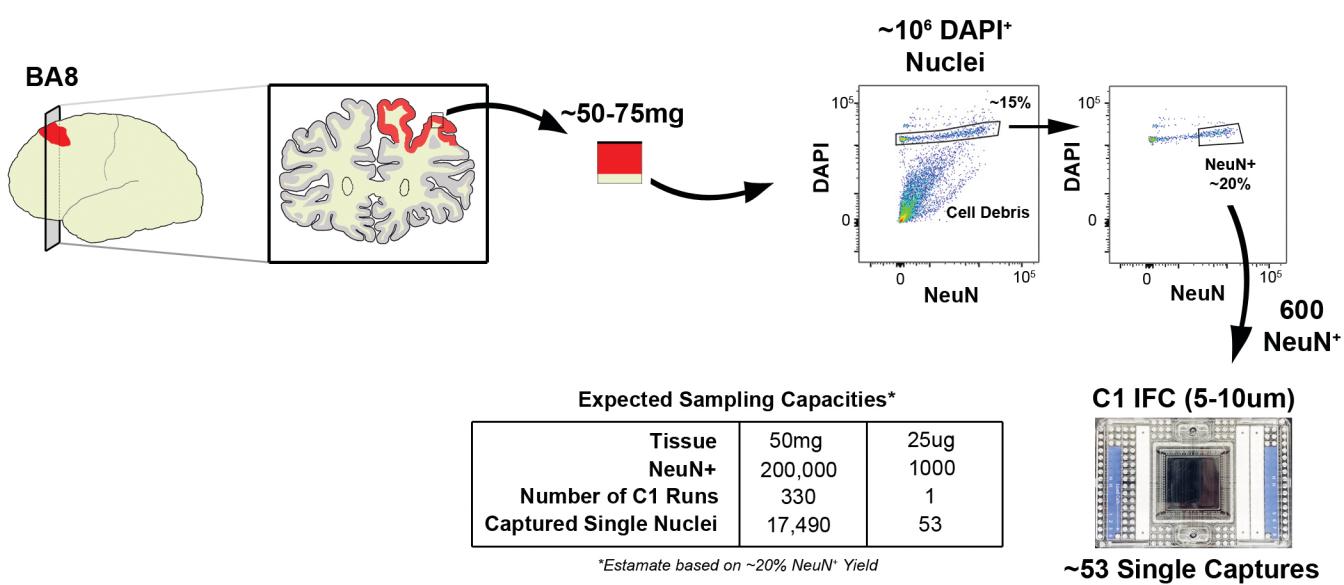
Table S14. Fraction of positive values for genes associated with neurotransmitter function or ion channels (see Fig. S16)

Table S15. Description of Imaging and methods used for RNAscope Validation

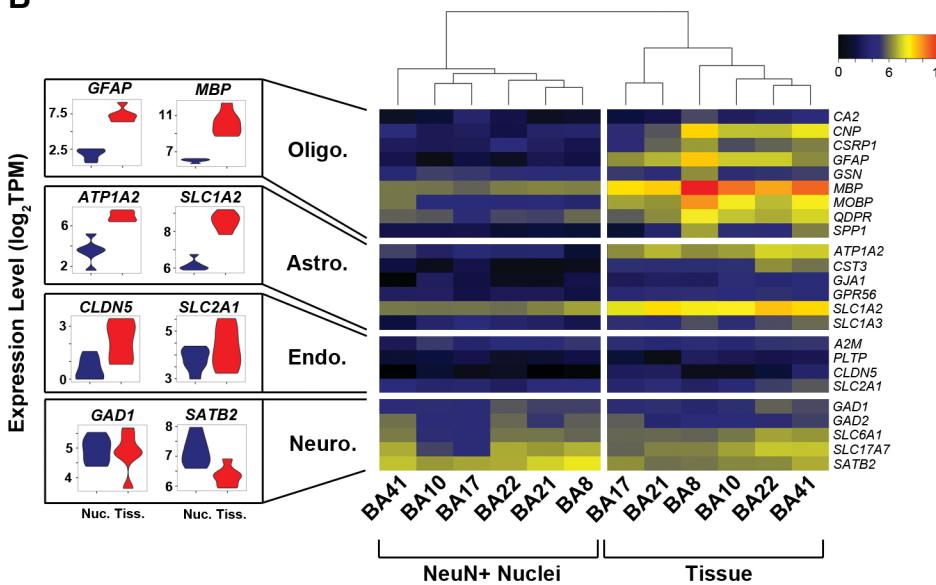
Table S16: Neuronal and glia marker genes (see Fig. S7)

Fig. S1

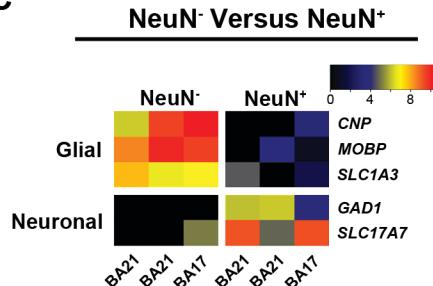
A



B



C



D

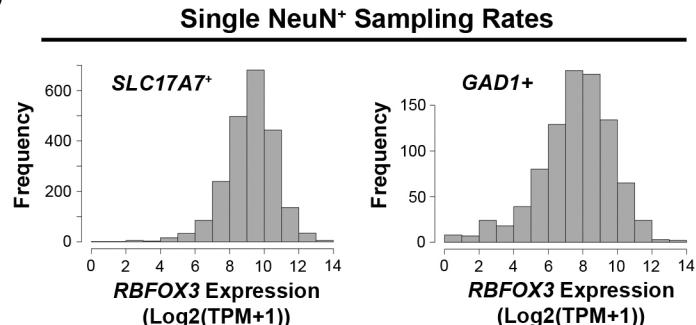


Fig. S2

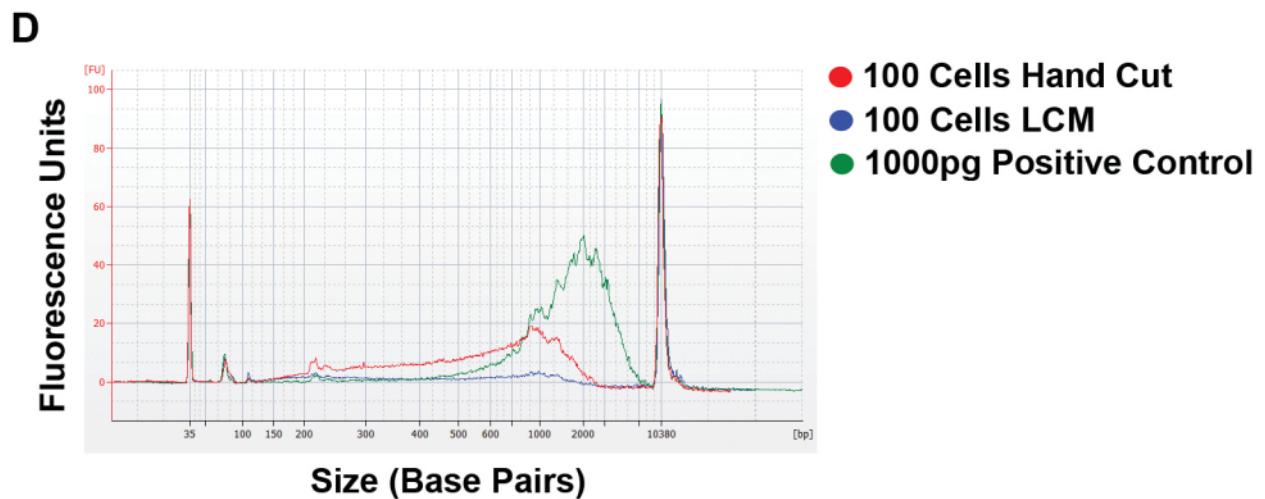
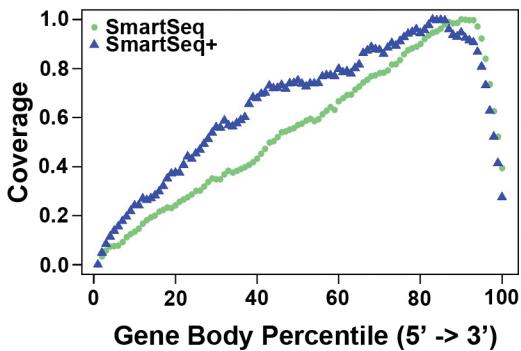
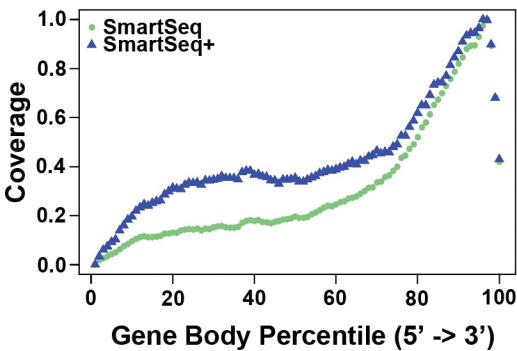


Fig. S3

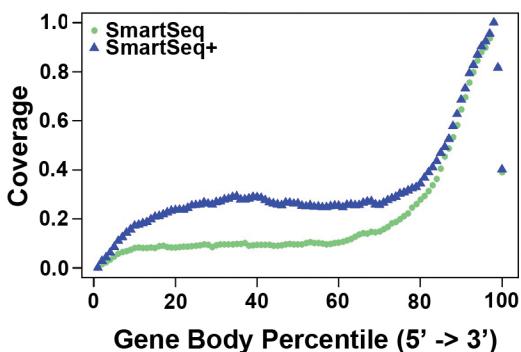
Transcript Coverage (0-1 kb)



Transcript Coverage (1-2 kb)



Transcript Coverage (2-3 kb)



Transcript Coverage (3-4 kb)

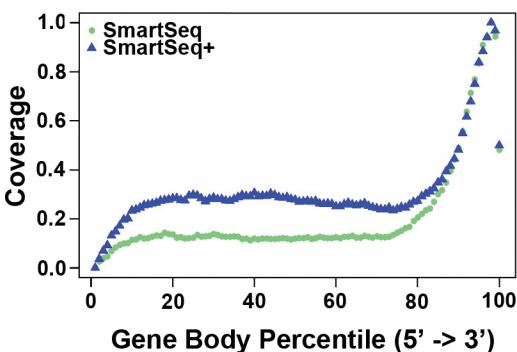
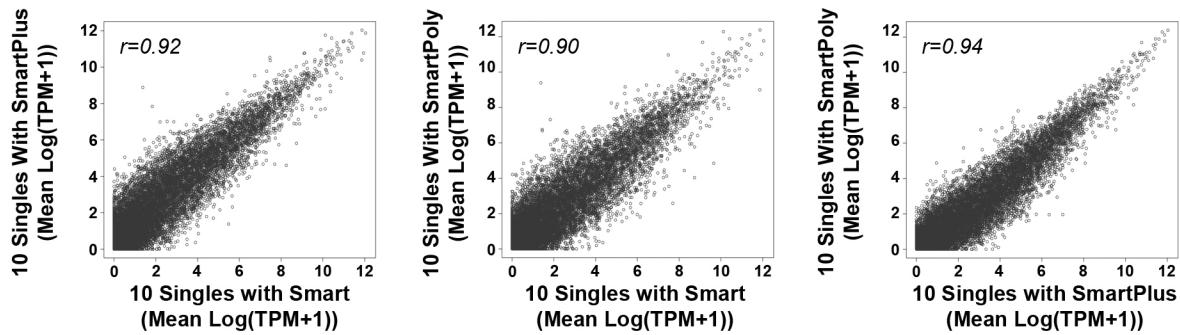
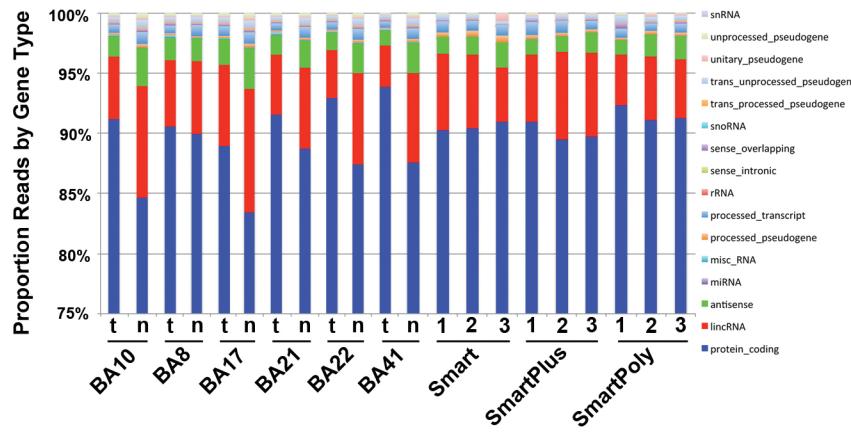


Fig. S4

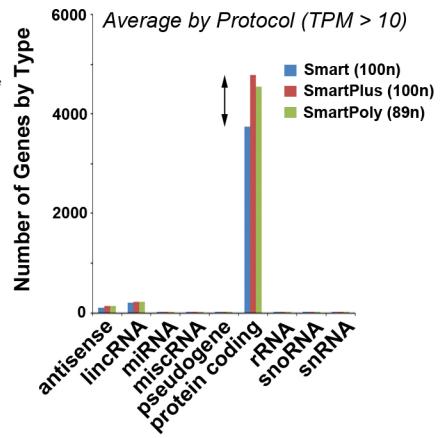
A



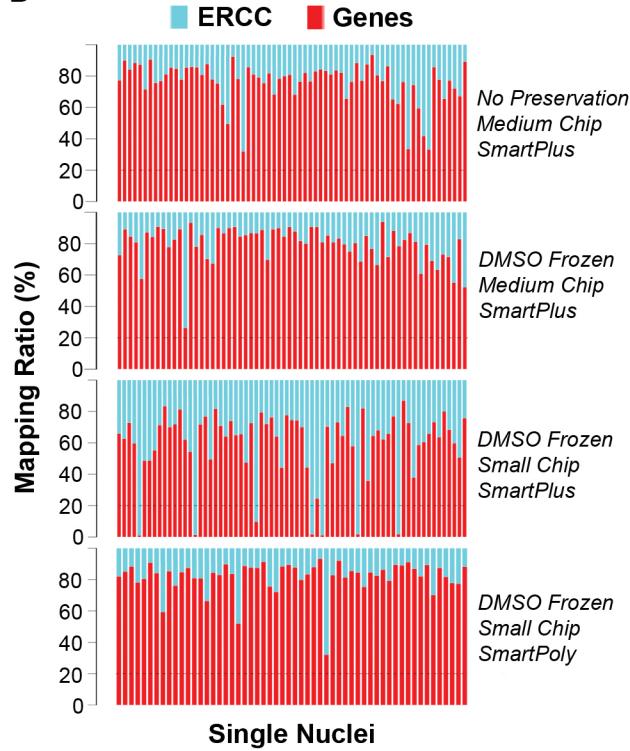
B



C



D



E

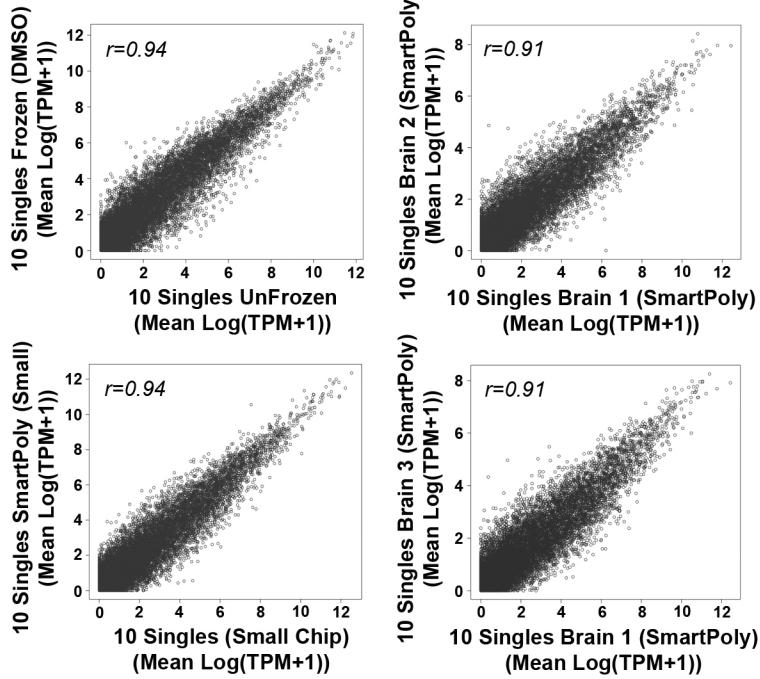


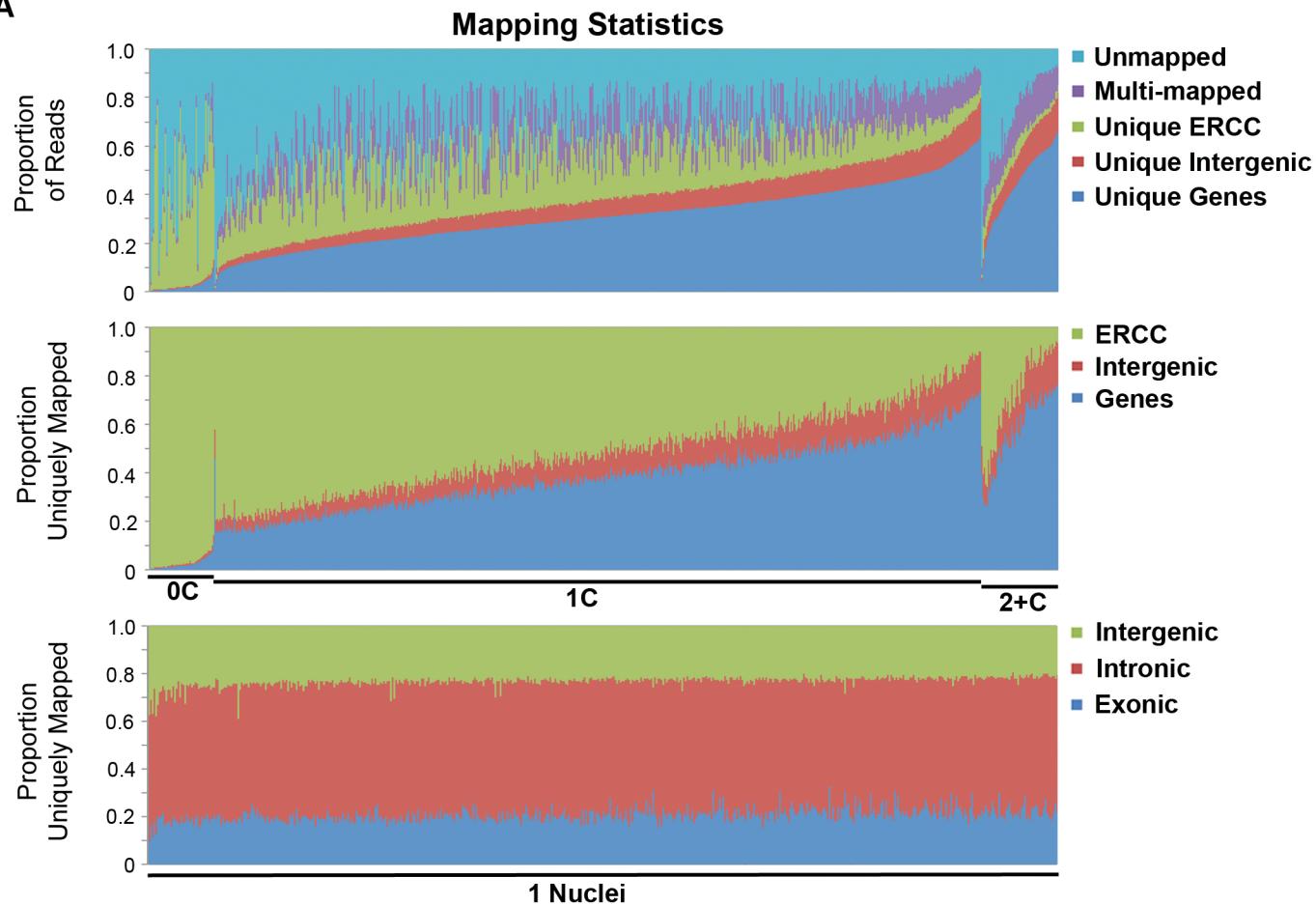
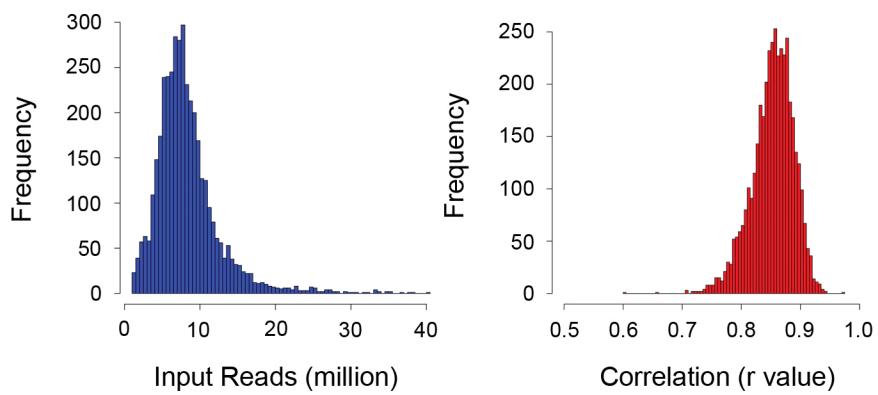
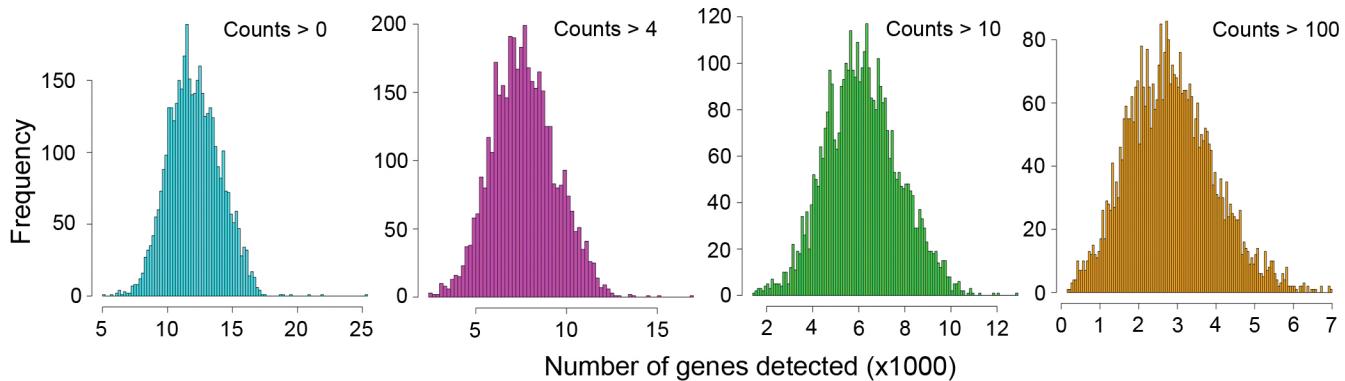
Fig. S5**A****B****C**

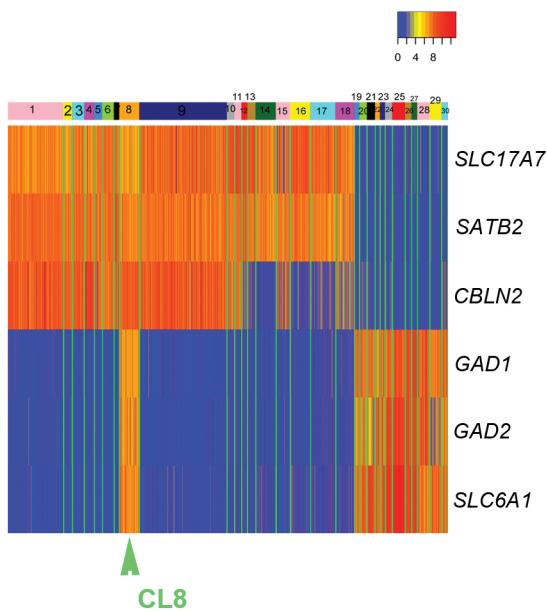
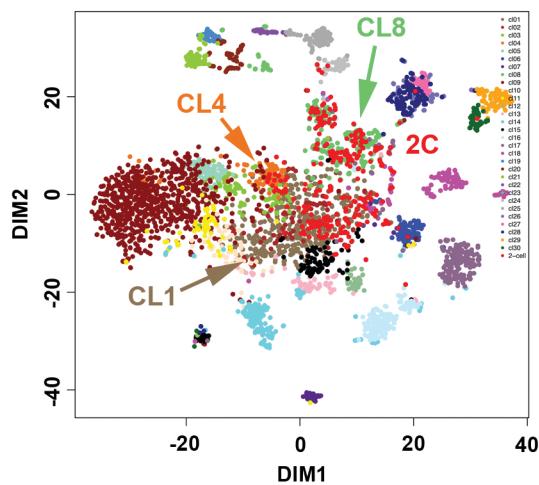
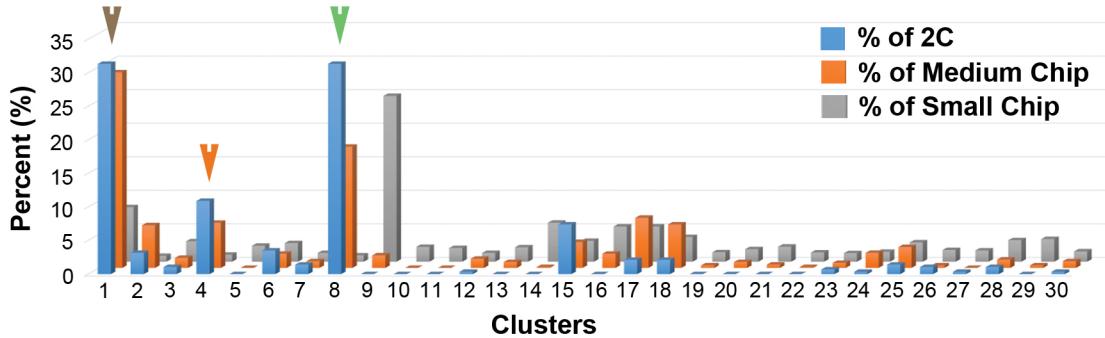
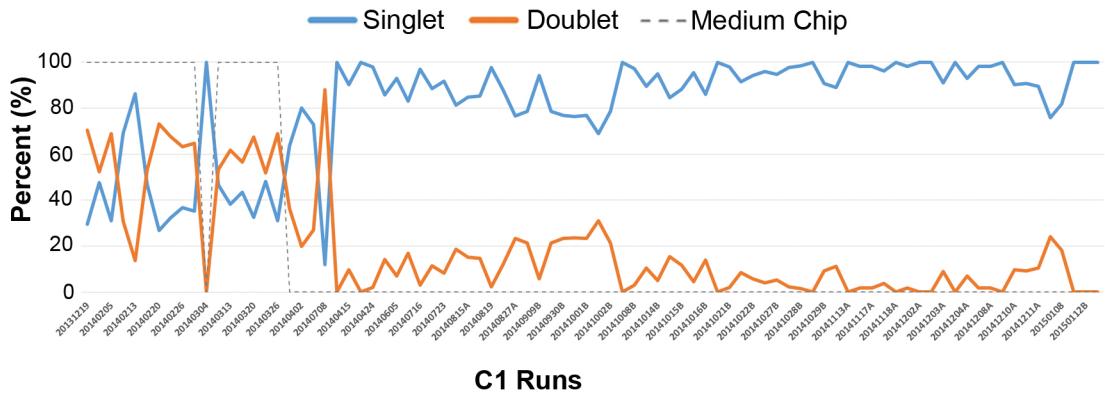
Fig. S6**A****B****C****D**

Fig. S7

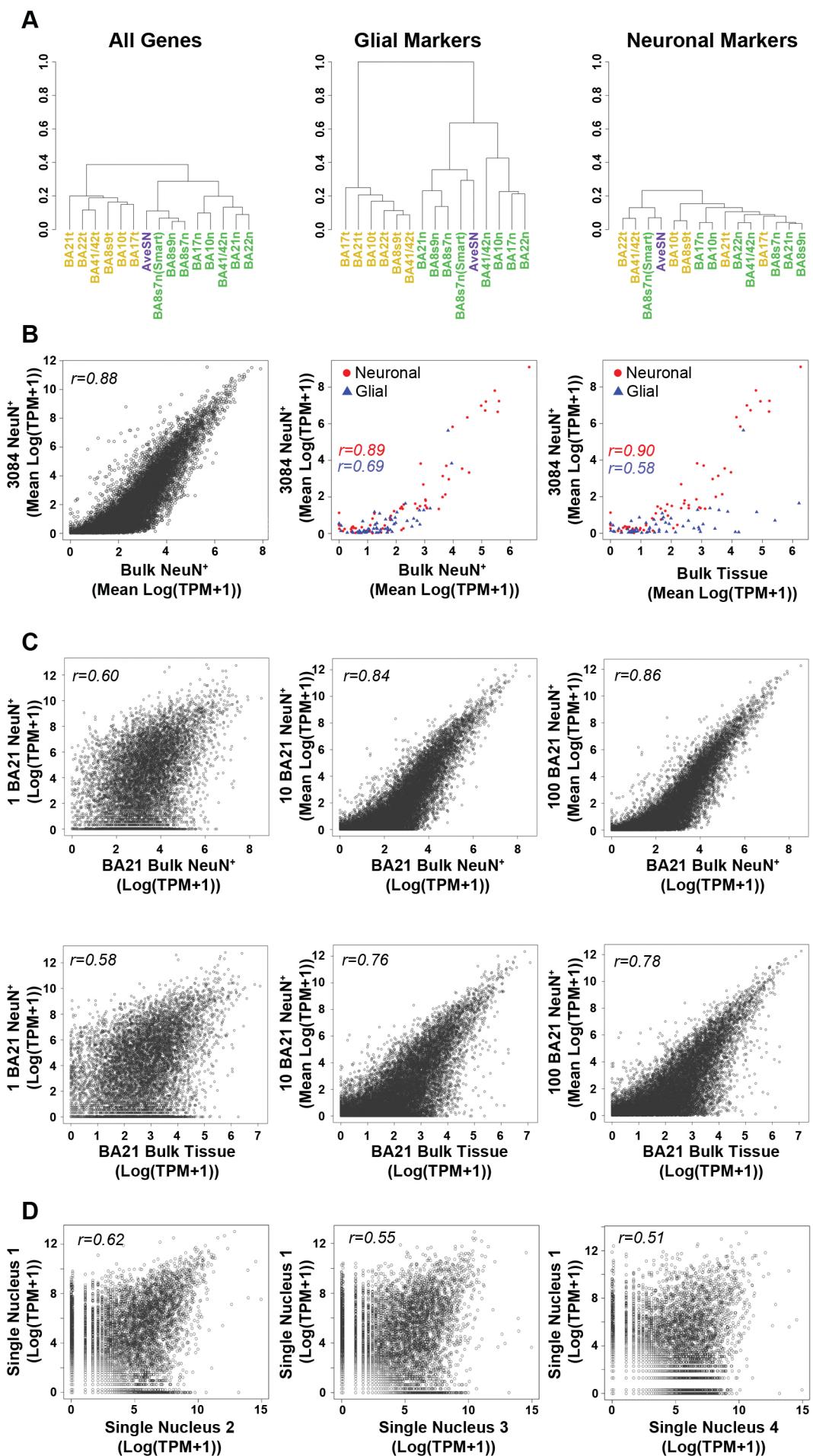
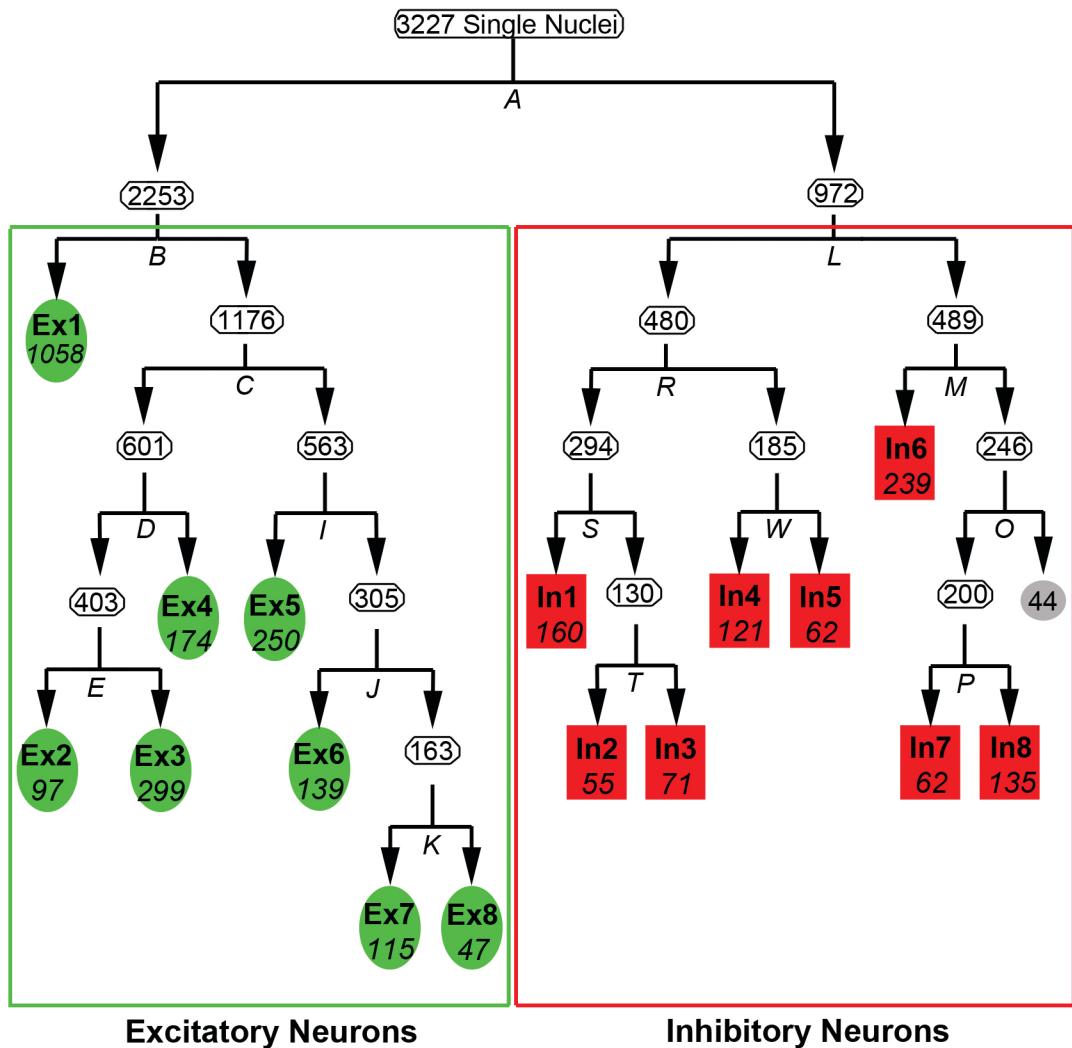
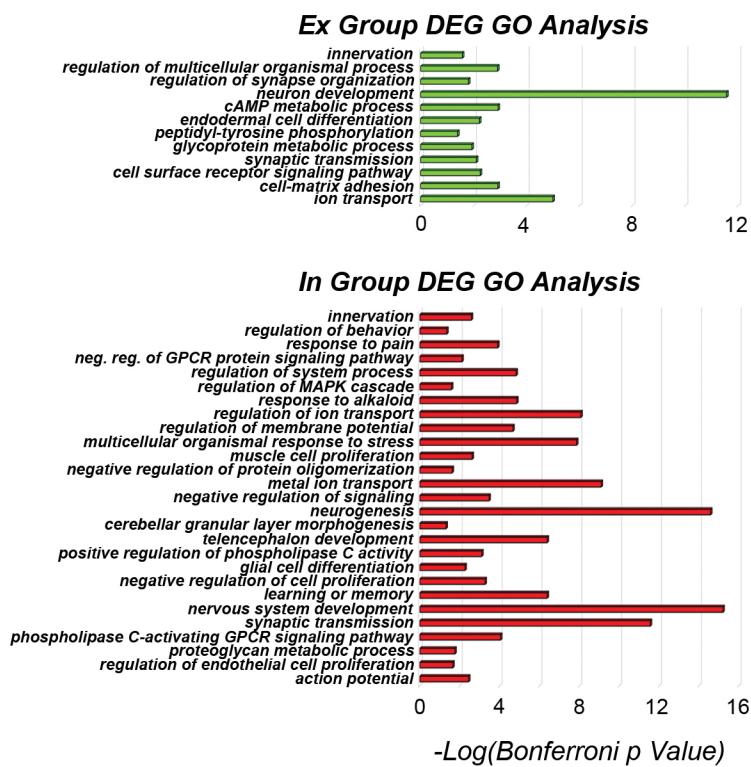


Fig. S8

A



B



C

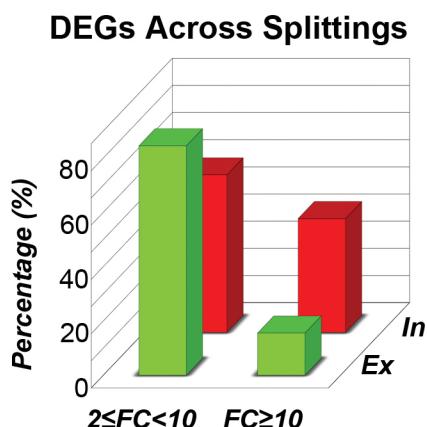


Fig. S9

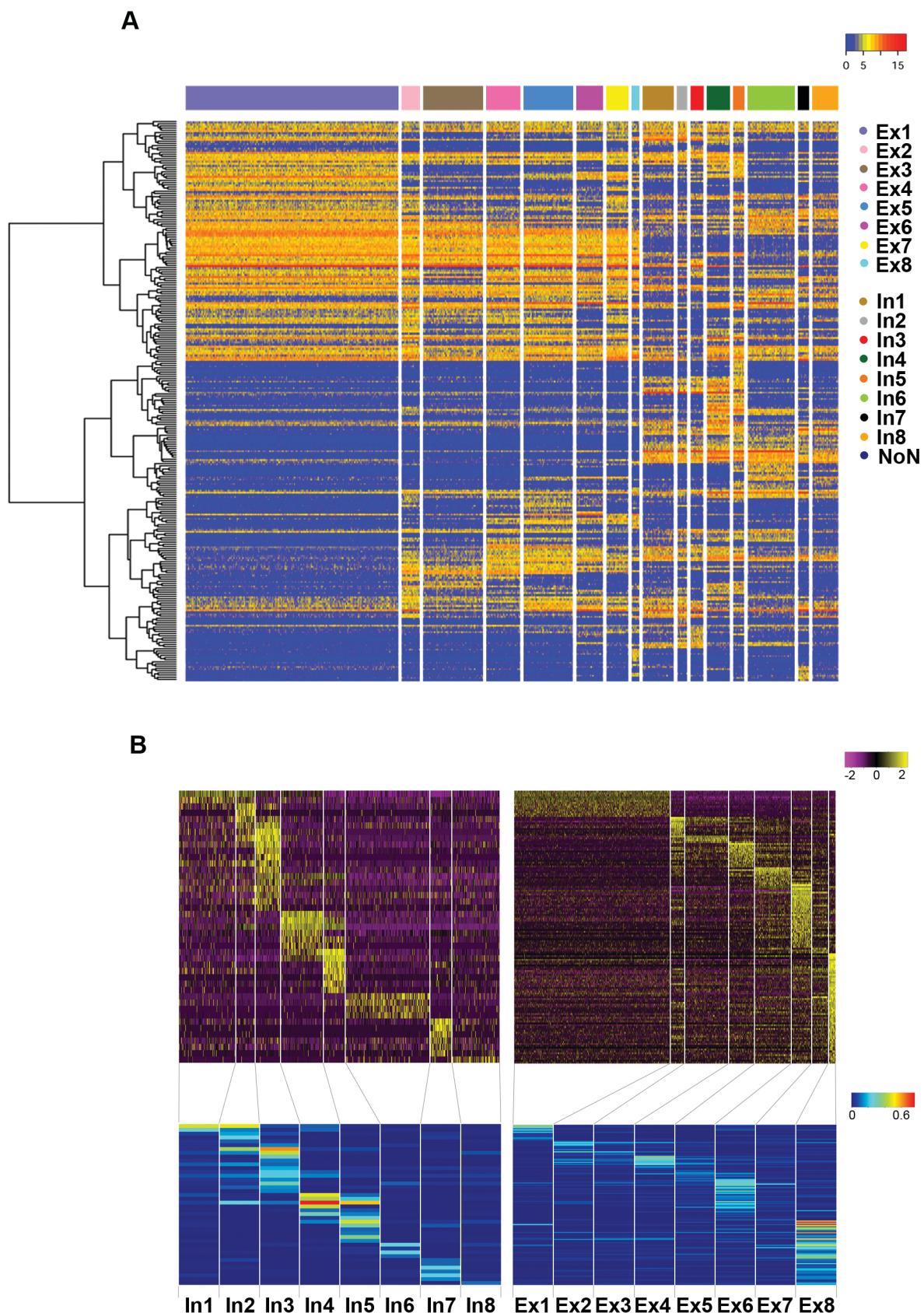
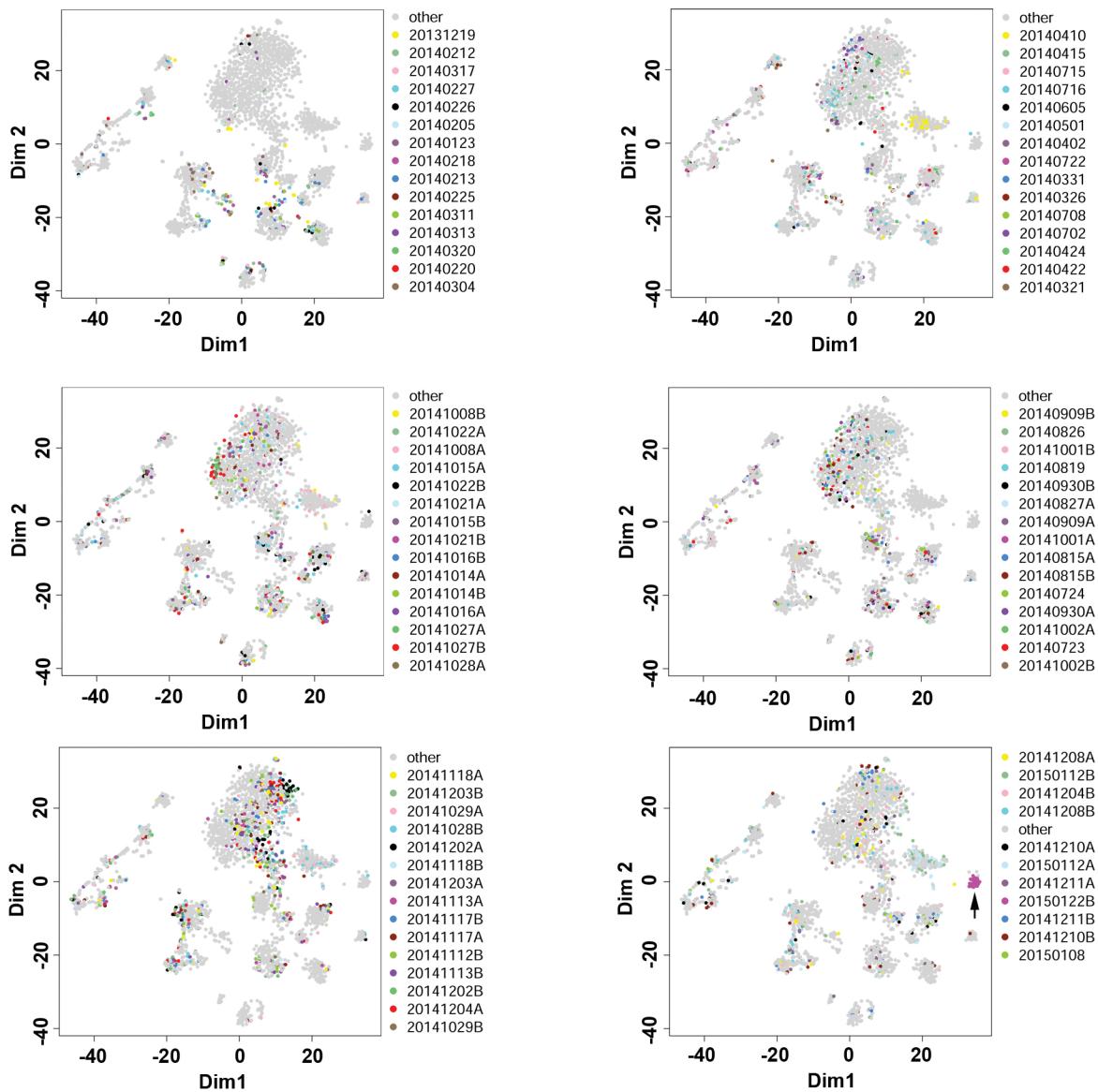


Fig. S10

A



B

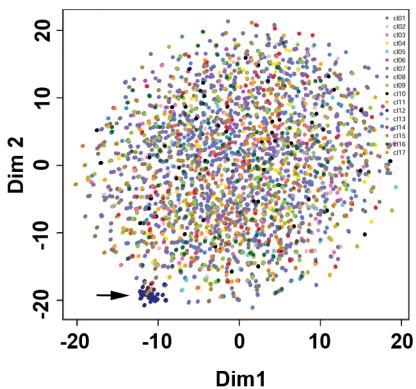


Fig. S11

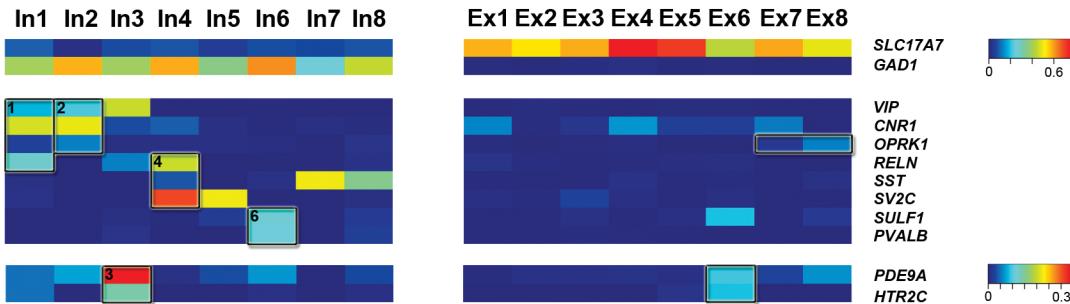
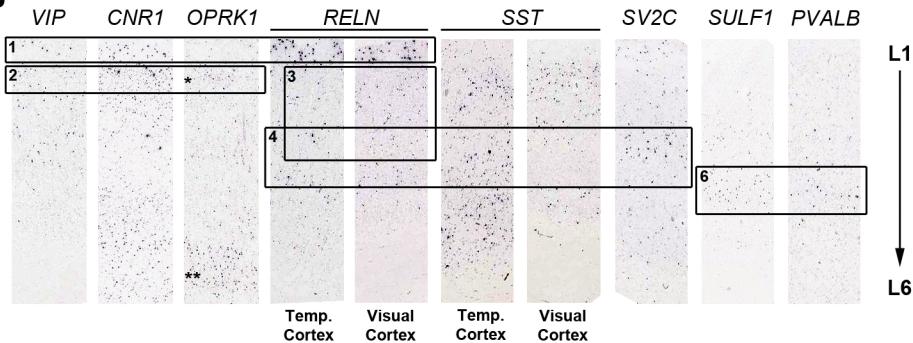
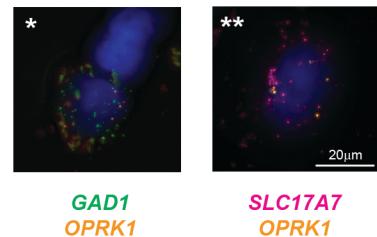
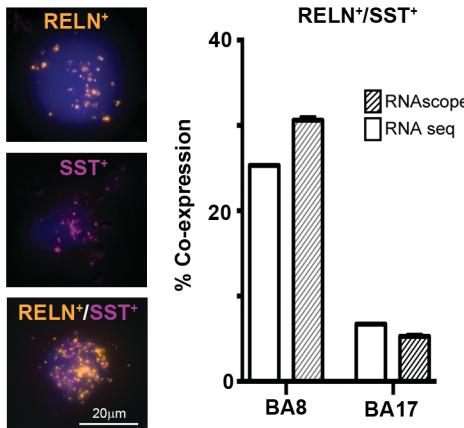
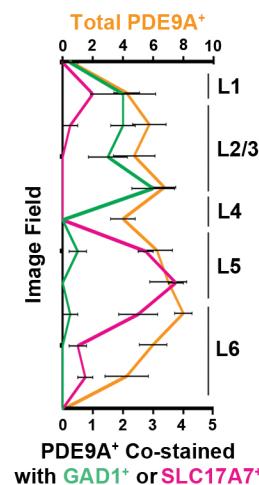
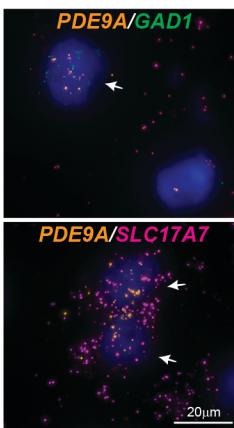
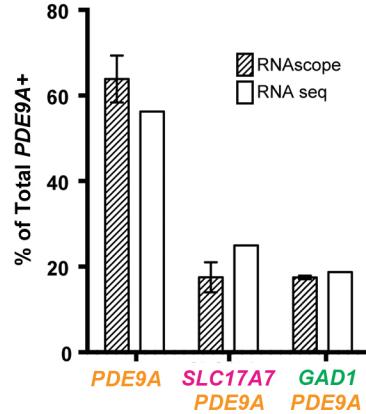
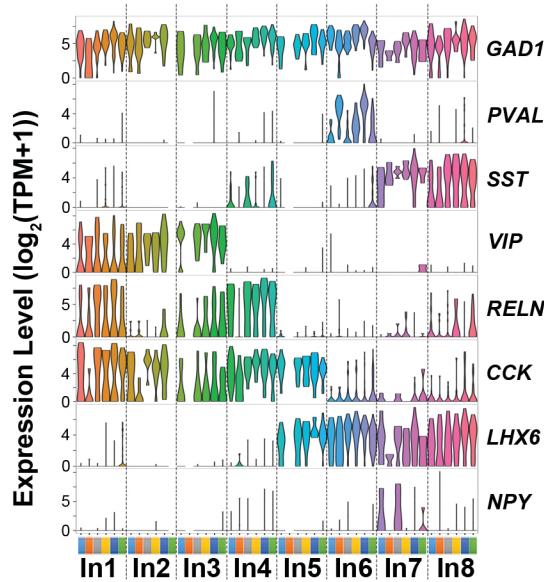
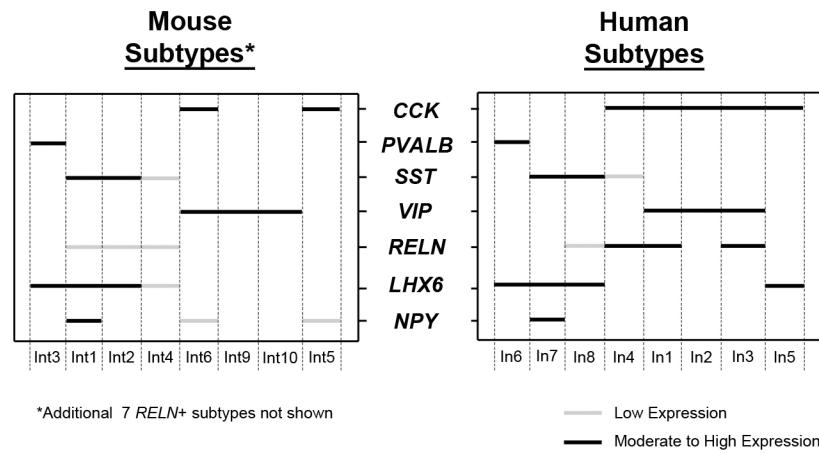
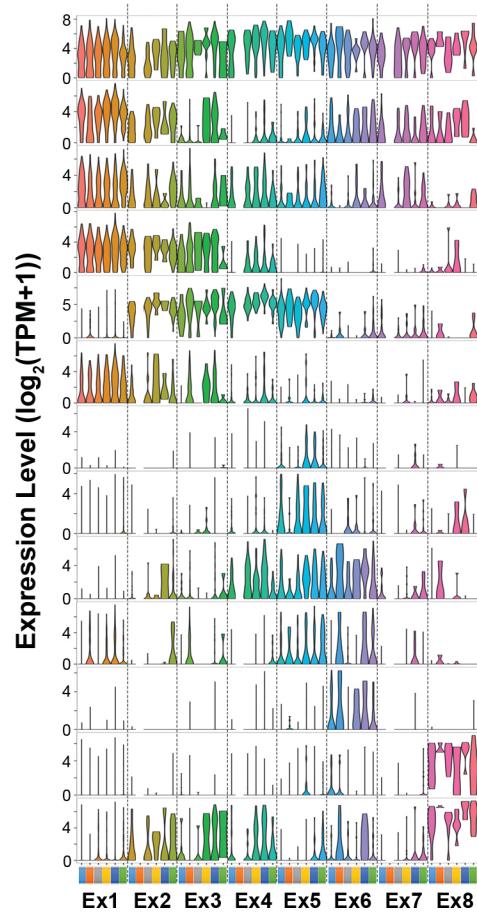
A

B

C

D

E

F


Fig. S12

A**Inhibitory Neurons****Combinatorial Marker Profiles****B****Excitatory Neurons****RNA ISH (human)**

	ISH Region	Mouse Pattern	Human Pattern
<i>TBR1</i>			
<i>RASGRF2</i>	Temporal Cortex	L2/3/4/5a	L2-6
<i>PVRL3</i>	Temporal Cortex	L2/3/5a	L2-6
<i>CUX2</i>	Temporal Cortex	L2-5a	L2-4
<i>RORB</i>	Temporal Cortex	L4-5a	L4-5
<i>THSD7A</i>		L5a	L2-4
<i>KCNK2</i>	Temporal Cortex	L5a	L5
<i>SULF2</i>		L5	L5
<i>FOXP2</i>	Visual Cortex	L6	L5/6
<i>PCP4</i>	Temporal Cortex	L5-6b	L5
<i>SYT6</i>	Visual Cortex	L6	L6b
<i>NR4A2</i>	Temporal Cortex	ClauPyr	L6b
<i>SYNPR</i>	Temporal Cortex	ClauPyr	L6b

Fig. S13

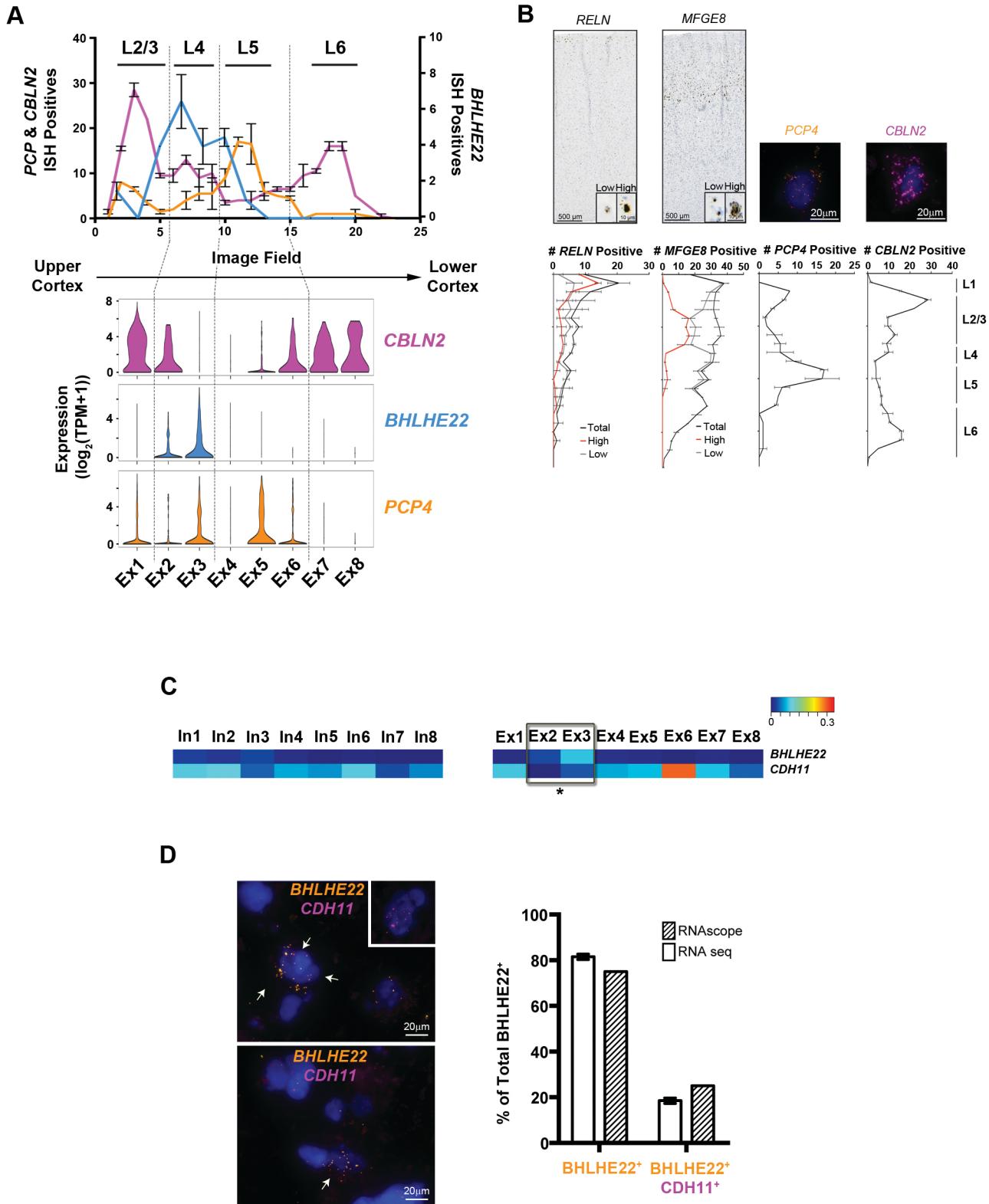
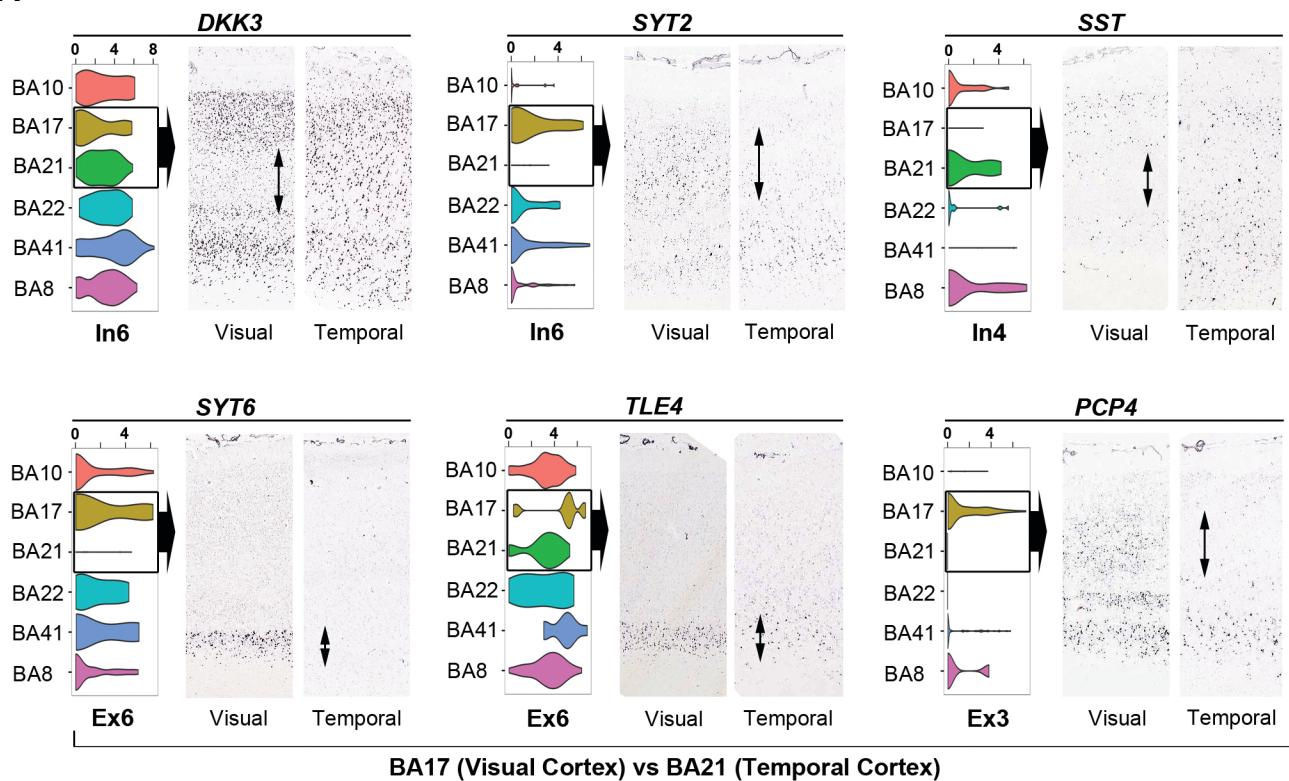


Fig. S14

A



B

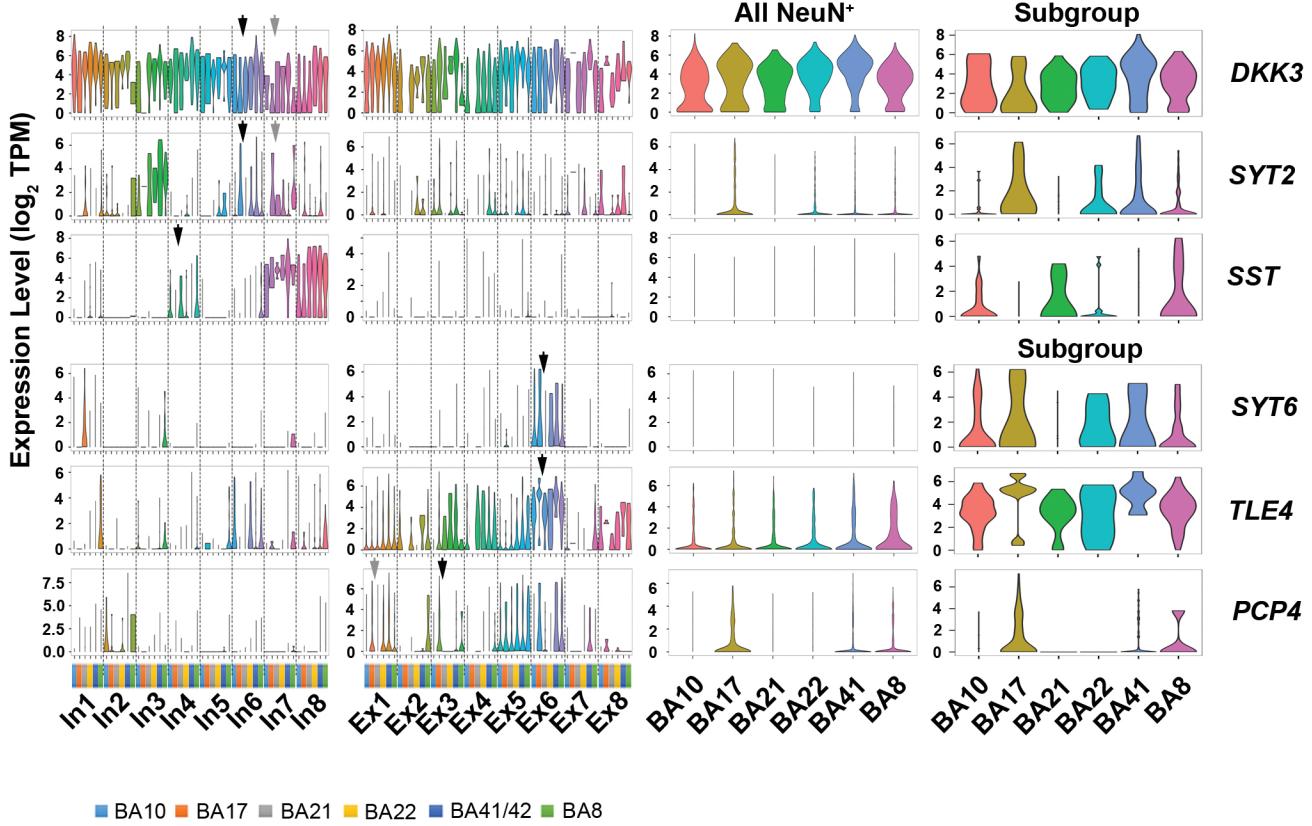


Fig. S15

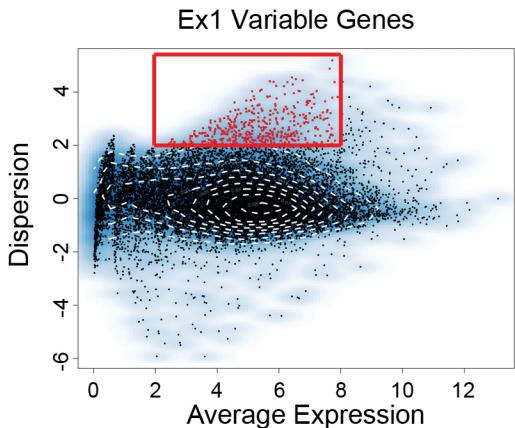
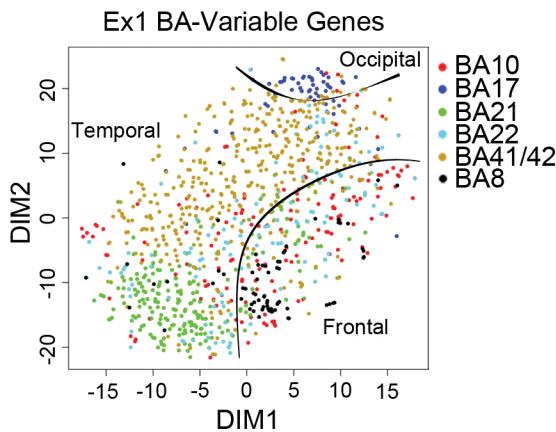
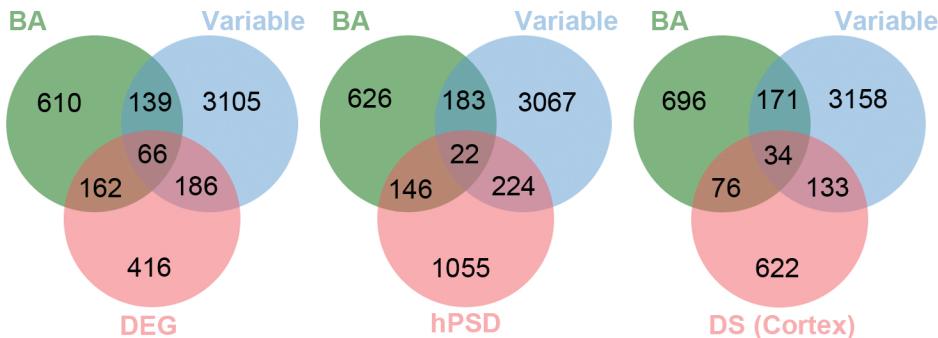
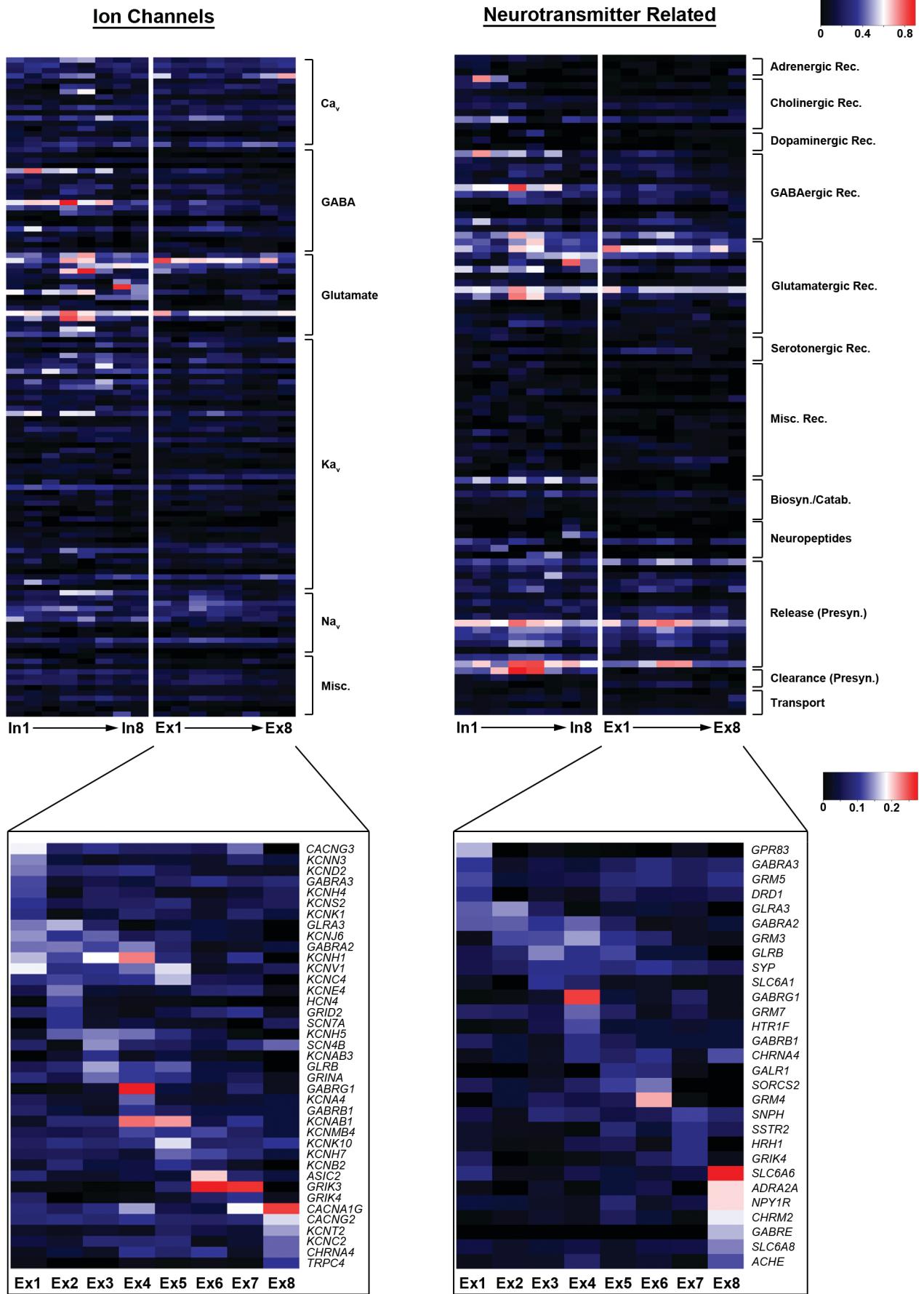
A**B****C**

Fig. S16



References and Notes

1. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015). [Medline](#) [doi:10.1016/j.cell.2015.05.002](https://doi.org/10.1016/j.cell.2015.05.002)
2. D. Usoskin, A. Furlan, S. Islam, H. Abdo, P. Lönnerberg, D. Lou, J. Hjerling-Leffler, J. Haeggström, O. Kharchenko, P. V. Kharchenko, S. Linnarsson, P. Ernfors, Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015). [Medline](#) [doi:10.1038/nn.3881](https://doi.org/10.1038/nn.3881)
3. A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, C. Rohy, G. Castelo-Branco, J. Hjerling-Leffler, S. Linnarsson, Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015). [Medline](#) [doi:10.1126/science.aaa1934](https://doi.org/10.1126/science.aaa1934)
4. M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-Bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, Z. L. Riley, C. Abajian, C. F. Beckmann, A. Bernard, D. Bertagnolli, A. F. Boe, P. M. Cartagena, M. M. Chakravarty, M. Chapin, J. Chong, R. A. Dalley, B. D. Daly, C. Dang, S. Datta, N. Dee, T. A. Dolbeare, V. Faber, D. Feng, D. R. Fowler, J. Goldy, B. W. Gregor, Z. Haradon, D. R. Haynor, J. G. Hohmann, S. Horvath, R. E. Howard, A. Jeromin, J. M. Jochim, M. Kinnunen, C. Lau, E. T. Lazarz, C. Lee, T. A. Lemon, L. Li, Y. Li, J. A. Morris, C. C. Overly, P. D. Parker, S. E. Parry, M. Reding, J. J. Royall, J. Schulkin, P. A. Sequeira, C. R. Slaughterbeck, S. C. Smith, A. J. Sodt, S. M. Sunkin, B. E. Swanson, M. P. Vawter, D. Williams, P. Wohnoutka, H. R. Zielke, D. H. Geschwind, P. R. Hof, S. M. Smith, C. Koch, S. G. Grant, A. R. Jones, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012). [Medline](#) [doi:10.1038/nature11405](https://doi.org/10.1038/nature11405)
5. H. J. Kang, Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, T. Guennel, Y. Shin, M. B. Johnson, Z. Krsnik, S. Mayer, S. Fertuzinhos, S. Umlauf, S. N. Lisgo, A. Vortmeyer, D. R. Weinberger, S. Mane, T. M. Hyde, A. Huttner, M. Reimers, J. E. Kleinman, N. Sestan, Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011). [Medline](#) [doi:10.1038/nature10523](https://doi.org/10.1038/nature10523)
6. M. B. Johnson, Y. I. Kawasawa, C. E. Mason, Z. Krsnik, G. Coppola, D. Bogdanović, D. H. Geschwind, S. M. Mane, M. W. State, N. Sestan, Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009). [Medline](#) [doi:10.1016/j.neuron.2009.03.027](https://doi.org/10.1016/j.neuron.2009.03.027)
7. M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C. K. Lee, A. Bernard, M. F. Glasser, D. L. Dierker, J. Menche, A. Szafer, F. Collman, P. Grange, K. A. Berman, S. Mihalas, Z. Yao, L. Stewart, A. L. Barabási, J. Schulkin, J. Phillips, L. Ng, C. Dang, D. R. Haynor, A. Jones, D. C. Van Essen, C. Koch, E. Lein, Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015). [Medline](#) [doi:10.1038/nn.4171](https://doi.org/10.1038/nn.4171)

8. S. Darmanis, S. A. Sloan, Y. Zhang, M. Enge, C. Caneda, L. M. Shuer, M. G. Hayden Gephart, B. A. Barres, S. R. Quake, A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015). [Medline](#) [doi:10.1073/pnas.1507125112](#)
9. D. M. Bushman, G. E. Kaeser, B. Siddoway, J. W. Westra, R. R. Rivera, S. K. Rehen, Y. C. Yung, J. Chun, Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *eLife* **4**, e05116 (2015). [Medline](#)
10. J. Gole, A. Gore, A. Richards, Y. J. Chiu, H. L. Fung, D. Bushman, H. I. Chiang, J. Chun, Y. H. Lo, K. Zhang, Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat. Biotechnol.* **31**, 1126–1132 (2013). [Medline](#) [doi:10.1038/nbt.2720](#)
11. R. V. Grindberg, J. L. Yee-Greenbaum, M. J. McConnell, M. Novotny, A. L. O'Shaughnessy, G. M. Lambert, M. J. Araúzo-Bravo, J. Lee, M. Fishman, G. E. Robbins, X. Lin, P. Venepally, J. H. Badger, D. W. Galbraith, F. H. Gage, R. S. Lasken, RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19802–19807 (2013). [Medline](#) [doi:10.1073/pnas.1319700110](#)
12. S. R. Krishnaswami, R. V. Grindberg, M. Novotny, P. Venepally, B. Lacar, K. Bhutani, S. B. Linker, S. Pham, J. A. Erwin, J. A. Miller, R. Hodge, J. K. McCarthy, M. Kelder, J. McCorrison, B. D. Avermann, F. D. Fuertes, R. H. Scheuermann, J. Lee, E. S. Lein, N. Schork, M. J. McConnell, F. H. Gage, R. S. Lasken, Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016). [Medline](#) [doi:10.1038/nprot.2016.015](#)
13. Materials and methods are available as supplementary materials on *Science Online*.
14. A. Graf, S. Krebs, V. Zakhartchenko, B. Schwalb, H. Blum, E. Wolf, Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4139–4144 (2014). [Medline](#) [doi:10.1073/pnas.1321569111](#)
15. D. Gaidatzis, L. Burger, M. Florescu, M. B. Stadler, Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat. Biotechnol.* **33**, 722–729 (2015). [Medline](#) [doi:10.1038/nbt.3269](#)
16. L. C. Greig, M. B. Woodworth, M. J. Galazo, H. Padmanabhan, J. D. Macklis, Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013). [Medline](#) [doi:10.1038/nrn3586](#)
17. H. Zeng, E. H. Shen, J. G. Hohmann, S. W. Oh, A. Bernard, J. J. Royall, K. J. Glattfelder, S. M. Sunkin, J. A. Morris, A. L. Guillozet-Bongaarts, K. A. Smith, A. J. Ebbert, B. Swanson, L. Kuan, D. T. Page, C. C. Overly, E. S. Lein, M. J. Hawrylycz, P. R. Hof, T. M. Hyde, J. E. Kleinman, A. R. Jones, Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483–496 (2012). [Medline](#) [doi:10.1016/j.cell.2012.02.052](#)
18. D. V. Hansen, J. H. Lui, P. Flandin, K. Yoshikawa, J. L. Rubenstein, A. Alvarez-Buylla, A. R. Kriegstein, Non-epithelial stem cells and cortical interneuron production in the human

ganglionic eminences. *Nat. Neurosci.* **16**, 1576–1587 (2013). [Medline](#)
[doi:10.1038/nn.3541](#)

19. T. Ma, C. Wang, L. Wang, X. Zhou, M. Tian, Q. Zhang, Y. Zhang, J. Li, Z. Liu, Y. Cai, F. Liu, Y. You, C. Chen, K. Campbell, H. Song, L. Ma, J. L. Rubenstein, Z. Yang, Subcortical origins of human and monkey neocortical interneurons. *Nat. Neurosci.* **16**, 1588–1597 (2013). [Medline](#) [doi:10.1038/nn.3536](#)
20. M. A. Lodato, M. B. Woodworth, S. Lee, G. D. Evrony, B. K. Mehta, A. Karger, S. Lee, T. W. Chittenden, A. M. D'Gama, X. Cai, L. J. Luquette, E. Lee, P. J. Park, C. A. Walsh, Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015). [Medline](#)
21. M. J. McConnell, M. R. Lindberg, K. J. Brennand, J. C. Piper, T. Voet, C. Cowing-Zitron, S. Shumilina, R. S. Lasken, J. R. Vermeesch, I. M. Hall, F. H. Gage, Mosaic copy number variation in human neurons. *Science* **342**, 632–637 (2013). [Medline](#)
[doi:10.1126/science.1243472](#)
22. S. K. Rehen, Y. C. Yung, M. P. McCreight, D. Kaushal, A. H. Yang, B. S. Almeida, M. A. Kingsbury, K. M. Cabral, M. J. McConnell, B. Anliker, M. Fontanoz, J. Chun, Constitutional aneuploidy in the normal human brain. *J. Neurosci.* **25**, 2176–2180 (2005). [Medline](#) [doi:10.1523/JNEUROSCI.4560-04.2005](#)
23. J. W. Westra, R. R. Rivera, D. M. Bushman, Y. C. Yung, S. E. Peterson, S. Barral, J. Chun, Neuronal DNA content variation (DCV) with regional and individual differences in the human brain. *J. Comp. Neurol.* **518**, 3981–4000 (2010). [Medline](#) [doi:10.1002/cne.22436](#)
24. K. L. Spalding, R. D. Bhardwaj, B. A. Buchholz, H. Druid, J. Frisén, Retrospective birth dating of cells in humans. *Cell* **122**, 133–143 (2005). [Medline](#)
[doi:10.1016/j.cell.2005.04.028](#)
25. A. D. Edelstein, M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, N. Stuurman, Advanced methods of microscope control using μManager software. *J. Biol. Methods* **1**, 10 (2014). [Medline](#) [doi:10.14440/jbm.2014.36](#)
26. P. Bardou, J. Mariette, F. Escudié, C. Djemiel, C. Klopp, jvenn: An interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293 (2014). [Medline](#) [doi:10.1186/1471-2105-15-293](#)
27. F. Wang, J. Flanagan, N. Su, L. C. Wang, S. Bui, A. Nielson, X. Wu, H. T. Vo, X. J. Ma, Y. Luo, RNAscope: A novel *in situ* RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012). [Medline](#)
[doi:10.1016/j.jmoldx.2011.08.002](#)
28. D. Ramsköld, S. Luo, Y. C. Wang, R. Li, Q. Deng, O. R. Faridani, G. A. Daniels, I. Khrebtukova, J. F. Loring, L. C. Laurent, G. P. Schroth, R. Sandberg, Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012). [Medline](#)
29. A. R. Wu, N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke, S. R. Quake, Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41–46 (2014). [Medline](#)
[doi:10.1038/nmeth.2694](#)

30. W. Zhou, R. V. Abruzzese, I. Polejaeva, S. Davis, S. Davis, W. Ji, Amplification of nanogram amounts of total RNA by the SMART-based PCR method for high-density oligonucleotide microarrays. *Clin. Chem.* **51**, 2354–2356 (2005). [Medline](#) [doi:10.1373/clinchem.2005.056721](#)
31. S. C. Baker, S. R. Bauer, R. P. Beyer, J. D. Brenton, B. Bromley, J. Burrill, H. Causton, M. P. Conley, R. Elespuru, M. Fero, C. Foy, J. Fuscoe, X. Gao, D. L. Gerhold, P. Gilles, F. Goodsaid, X. Guo, J. Hackett, R. D. Hockett, P. Ikonomi, R. A. Irizarry, E. S. Kawasaki, T. Kaysser-Kranich, K. Kerr, G. Kiser, W. H. Koch, K. Y. Lee, C. Liu, Z. L. Liu, A. Lucas, C. F. Manohar, G. Miyada, Z. Modrusan, H. Parkes, R. K. Puri, L. Reid, T. B. Ryder, M. Salit, R. R. Samaha, U. Scherf, T. J. Sendera, R. A. Setterquist, L. Shi, R. Shippy, J. V. Soriano, E. A. Wagar, J. A. Warrington, M. Williams, F. Wilmer, M. Wilson, P. K. Wolber, X. Wu, R. Zadro, External RNA Controls Consortium, The External RNA Controls Consortium: A progress report. *Nat. Methods* **2**, 731–734 (2005). [Medline](#) [doi:10.1038/nmeth1005-731](#)
32. J. Fan, N. Salathia, R. Liu, G. E. Kaeser, Y. C. Yung, J. L. Herman, F. Kaper, J. B. Fan, K. Zhang, J. Chun, P. V. Kharchenko, Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13**, 241–244 (2016). [Medline](#) [doi:10.1038/nmeth.3734](#)
33. Y. Zhang, K. Chen, S. A. Sloan, M. L. Bennett, A. R. Scholze, S. O'Keeffe, H. P. Phatnani, P. Guarnieri, C. Caneda, N. Ruderisch, S. Deng, S. A. Liddelow, C. Zhang, R. Daneman, T. Maniatis, B. A. Barres, J. Q. Wu, An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014). [Medline](#) [doi:10.1523/JNEUROSCI.1860-14.2014](#)
34. D. Vogt, R. F. Hunt, S. Mandal, M. Sandberg, S. N. Silberberg, T. Nagasawa, Z. Yang, S. C. Baraban, J. L. Rubenstein, Lhx6 directly regulates Arx and CXCR7 to determine cortical interneuron fate and laminar position. *Neuron* **82**, 350–364 (2014). [Medline](#) [doi:10.1016/j.neuron.2014.02.030](#)
35. G. L. McKinsey, S. Lindtner, B. Trzciński, A. Visel, L. A. Pennacchio, D. Huylebroeck, Y. Higashi, J. L. Rubenstein, Dlx1&2-dependent expression of Zfhx1b (Sip1, Zeb2) regulates the fate switch between cortical and striatal interneurons. *Neuron* **77**, 83–98 (2013). [Medline](#) [doi:10.1016/j.neuron.2012.11.035](#)
36. B. Schaerlinger, P. Hickel, N. Etienne, L. Guesnier, L. Maroteaux, Agonist actions of dihydroergotamine at 5-HT2B and 5-HT2C receptors and their possible relevance to antimigraine efficacy. *Br. J. Pharmacol.* **140**, 277–284 (2003). [Medline](#) [doi:10.1038/sj.bjp.0705437](#)
37. S. E. Ross, A. E. McCord, C. Jung, D. Atan, S. I. Mok, M. Hemberg, T. K. Kim, J. Salogiannis, L. Hu, S. Cohen, Y. Lin, D. Harrar, R. R. McInnes, M. E. Greenberg, Bhlhb5 and Prdm8 form a repressor complex involved in neuronal circuit assembly. *Neuron* **73**, 292–303 (2012). [Medline](#) [doi:10.1016/j.neuron.2011.09.035](#)
38. A. Bayés, L. N. van de Lagemaat, M. O. Collins, M. D. Croning, I. R. Whittle, J. S. Choudhary, S. G. Grant, Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14**, 19–21 (2011). [Medline](#) [doi:10.1038/nn.2719](#)