Supplementary Materials for

Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain

Blue B. Lake1†, Rizi Ai2†, Gwendolyn E. Kaeser3,5†, Neeraj S. Salathia4†, Yun C. Yung3, Rui Liu1, Andre Wildberg2, Derek Gao1, Ho-Lim Fung1, Song Chen1, Raakhee Vijayaraghavan4, Julian Wong3, Allison Chen3, Xiaoyan Sheng3, Fiona Kaper4, Richard Shen4, Mostafa Ronaghi4, Jian-Bing Fan4*, Wei Wang2*, Jerold Chun3* and Kun Zhang1*

correspondence to: Kun Zhang (kzhang@bioeng.ucsd.edu); Jerold Chun (jchun@scripps.edu); Wei Wang (wei-wang@ucsd.edu); Jian-Bing Fan (jianbing_fan@yahoo.com)

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S16

Other Supplementary Materials for this manuscript includes the following:

Supplementary tables as Excel File
R Script as zipped folder: Clustering-and-Classification R Script, README.txt

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 OptiPrepTM 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 ¦ m 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 bulk controls, both 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 or 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/µl (10-17µm 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 12μM in the supplied Clontech Dilution buffer and added to Lysis Mix A for a final concentration of 4.2μM; where indicated (Table S1), PolydIdC (Sigma) in 0.4% NP-40 was added to Lysis Mix A for a final concentration of 33ng/μl. 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 μL 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/μL 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 μL 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 μL 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 log2(TPM) was calculated. A gene was considered expressed if 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 CV2 = variance/mean2 for each protein-coding gene across all samples within each cluster; (2) CV2 were then fitted to an inverse distribution and genes with CV2 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 log2(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 log2(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 (Log2(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 log2(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 (Log2(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 Log2(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 μm 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).

Fig. 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.

Fig. 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 μm. **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.

Fig. 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.

Fig. 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).

Fig. 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 [log10(counts+1) versus log10(concentration)] for all single nuclei libraries. **C.** Plots showing the frequency distributions of genes detected across single nuclei libraries using different gene count cutoffs.

Fig. 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.

Fig. S7

Single-nuclei RNA sequencing permits cell type identification. **A.** Cluster dendrogram of gene expression (Log2(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).

Fig. 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.

Fig. 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.

Fig. 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.

Fig. 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.

Fig. 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.

Fig. 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), *BHLEH22* (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.

Fig. 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).

Fig. S15

Subgroup variable genes. **A.** Plot of average expression and dispersion (binned log(Variance/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*).

Fig. 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.

Additional Data table S1 (separate file)
Summary of C1 experimental conditions and outcome

Additional Data table S2 (separate file)
Single nuclei library details and group/subgroup identifiers

Additional Data table S3 (separate file)
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.

Additional Data table S4 (separate file)
GO annotations for differentially expressed genes defining subgroup classifications (Table S3)

Additional Data table S5 (separate file)
Unique group specific genes and associated fraction of positive values using exon only derived TPM (see Fig. S9b)

Additional Data table S6 (separate file)
GO annotations for unique subgroup marker genes (Table S5)

Additional Data table S7 (separate file)
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)

Additional Data table S8 (separate file)
Genes differentially expressed between Ex2 and Ex3 subgroups

Additional Data table S9 (separate file)
GO annotations for genes differentially expressed between Ex2 and Ex3 subgroups (See Table S8)

Additional Data table S10 (separate file)
Genes differentially expressed between BA41/42 and BA17 brain regions within the Ex3 subgroup

Additional Data table S11 (separate file)
Allen Human Brain Atlas ISH citations (See Fig. 3D, Fig. S11, Fig. S12, Fig. S14)

Additional Data table S12 (separate file)
Variable genes identified within each subgroup

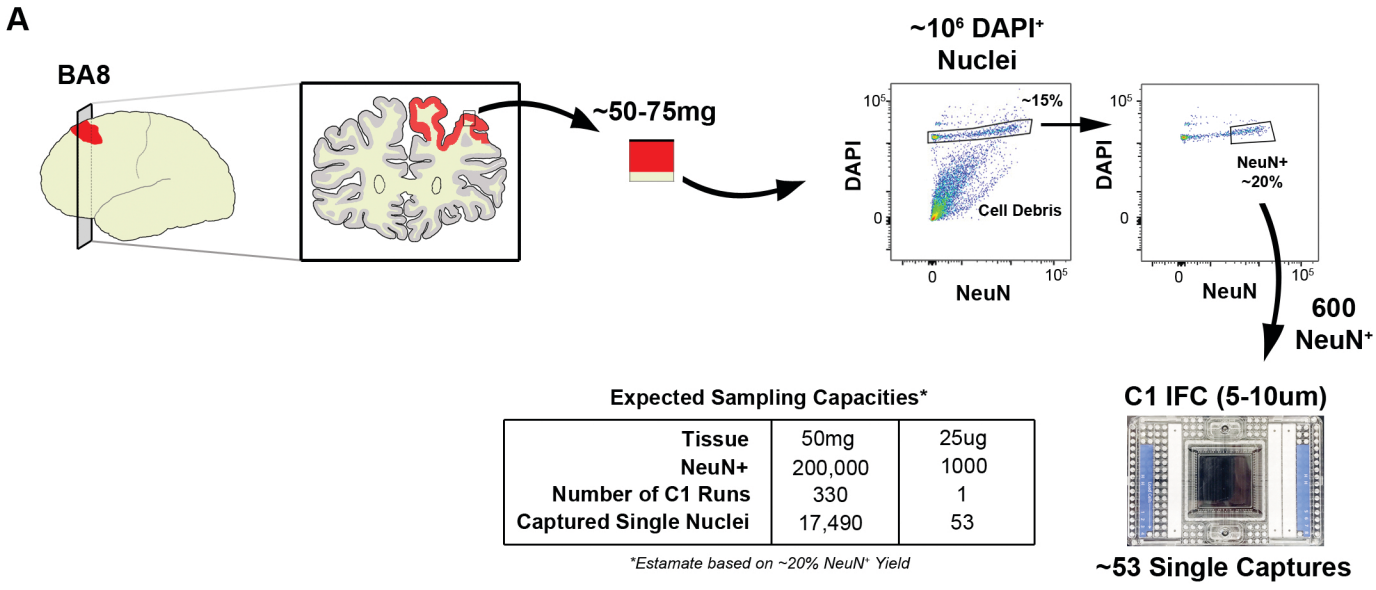
Additional Data table S13 (separate file)
Genes identified in each subgroup having 10-fold or more difference in expression level between at least two Brodmann Areas (BA)

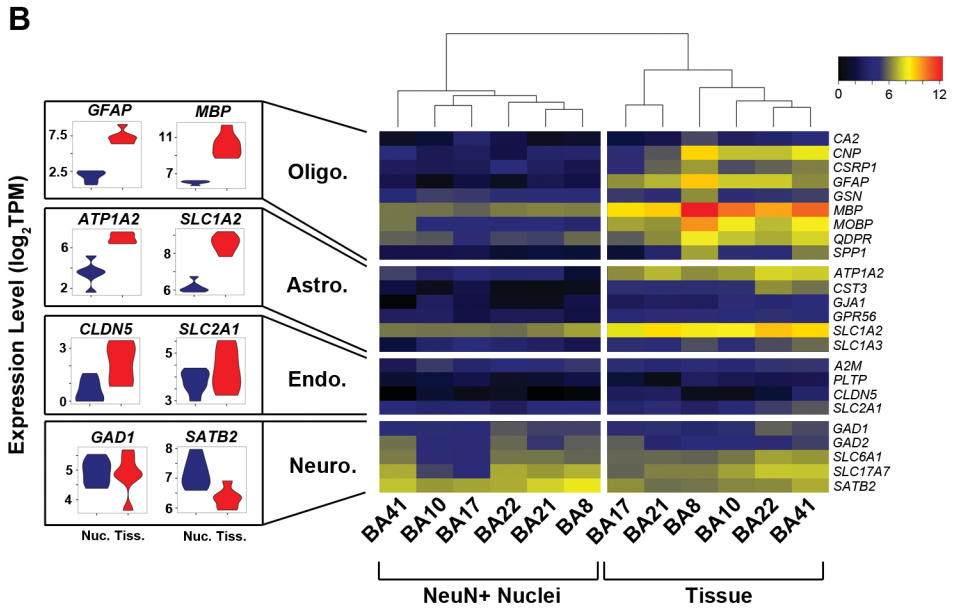
Additional Data table S14 (separate file)
Fraction of positive values for genes associated with neurotransmitter function or ion channels (see Fig. S16)

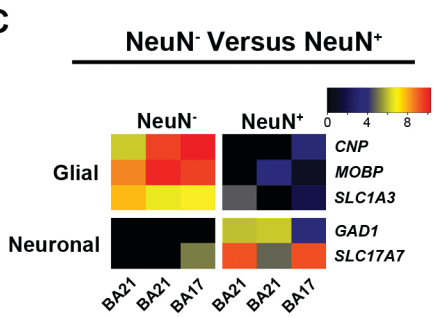
Additional Data table S15 (separate file)
Description of Imaging and methods used for RNAscope Validation

Additional Data table S16 (separate file)
Neuronal and glia marker genes (see Fig. S7)

Fig. S1


A
BA8
~50-75mg
~10⁶ DAPI⁺ Nuclei
DAPI
NeuN
~15%
Cell Debris
NeuN+
~20%
600
NeuN⁺
Expected Sampling Capacities*
Tissue 50mg 25ug
NeuN+ 200,000 1000
Number of C1 Runs 330 1
Captured Single Nuclei 17,490 53
*Estamate based on ~20% NeuN⁺ Yield
C1 IFC (5-10um)
~53 Single Captures


B
Expression Level (log2TPM)
GFAP
MBP
ATP1A2
SLC1A2
CLDN5
SLC2A1
GAD1
SATB2
Oligo.
Astro.
Endo.
Neuro.
Nuc. Tiss.
CA2
CNP
CSRP1
GFAP
GSN
MBP
MOBP
QDPR
SPP1
ATP1A2
CST3
GJA1
GPR56
SLC1A2
SLC1A3
A2M
PLTP
CLDN5
SLC2A1
GAD1
GAD2
SLC6A1
SLC17A7
SATB2
BA41 BA10 BA17 BA22 BA21 BA8 BA17 BA21 BA8 BA10 BA22 BA41
NeuN+ Nuclei
Tissue


C
NeuN⁻ Versus NeuN⁺
NeuN⁻
NeuN⁺
Glial
Neuronal
CNP
MOBP
SLC1A3
GAD1
SLC17A7
BA21 BA21 BA17 BA21 BA21 BA17

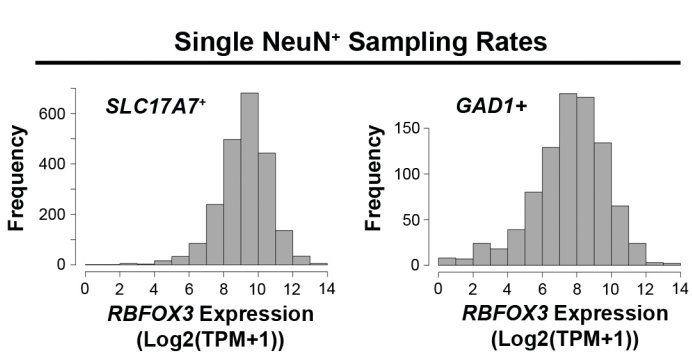
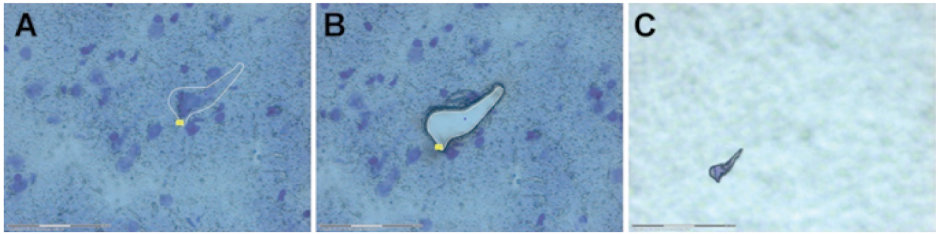

D
Single NeuN⁺ Sampling Rates
SLC17A7⁺
GAD1+
Frequency
RBFOX3 Expression
(Log2(TPM+1))

Fig. S2


A
B
C

D

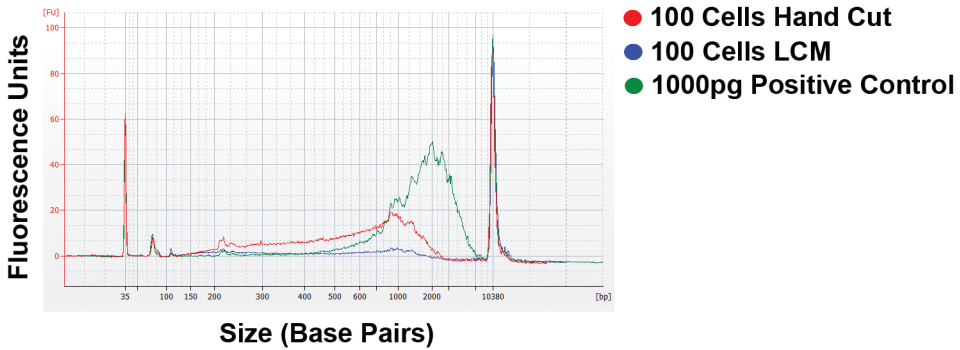

Fluorescence Units
[FU]
100
80
60
40
20
0
35
100
150
200
300
400
500
600
1000
2000
10380
[bp]
Size (Base Pairs)
100 Cells Hand Cut
100 Cells LCM
1000pg Positive Control

Fig. S3

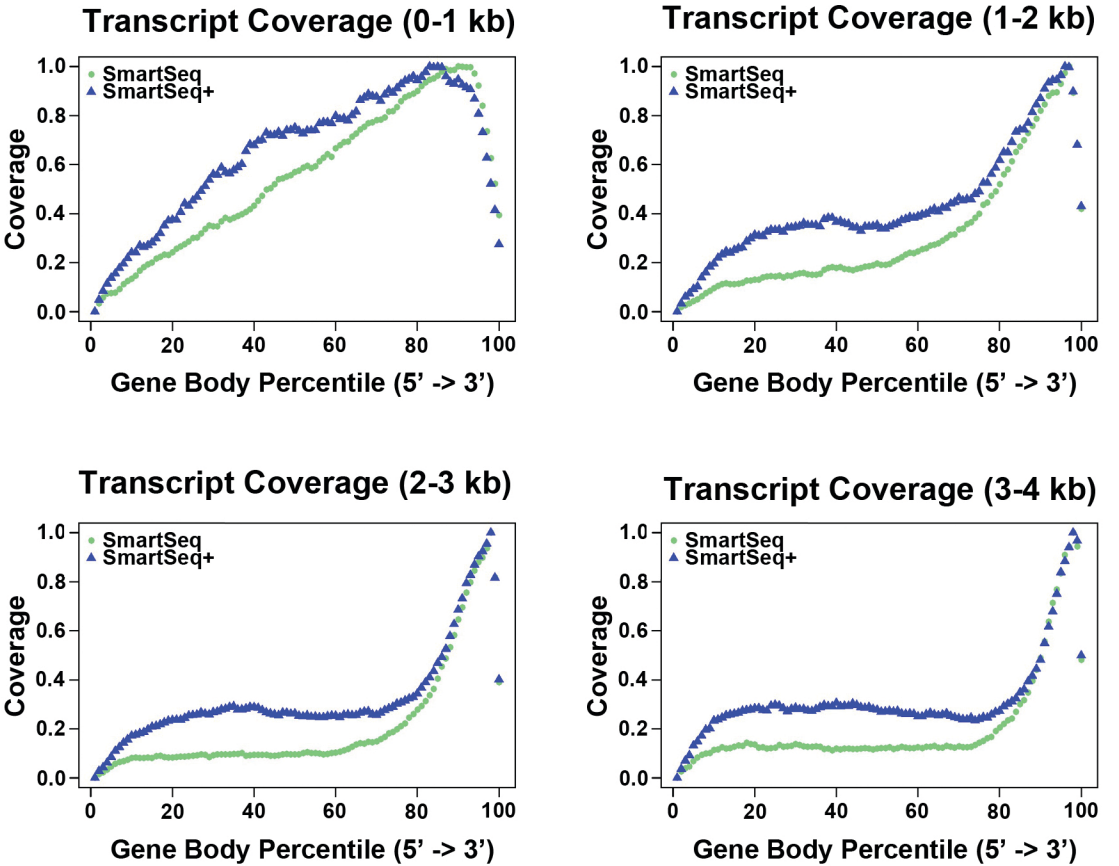

Transcript Coverage (0-1 kb)
Transcript Coverage (1-2 kb)
Transcript Coverage (2-3 kb)
Transcript Coverage (3-4 kb)
SmartSeq
SmartSeq+
Coverage
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Gene Body Percentile (5' -> 3')

Fig. S4

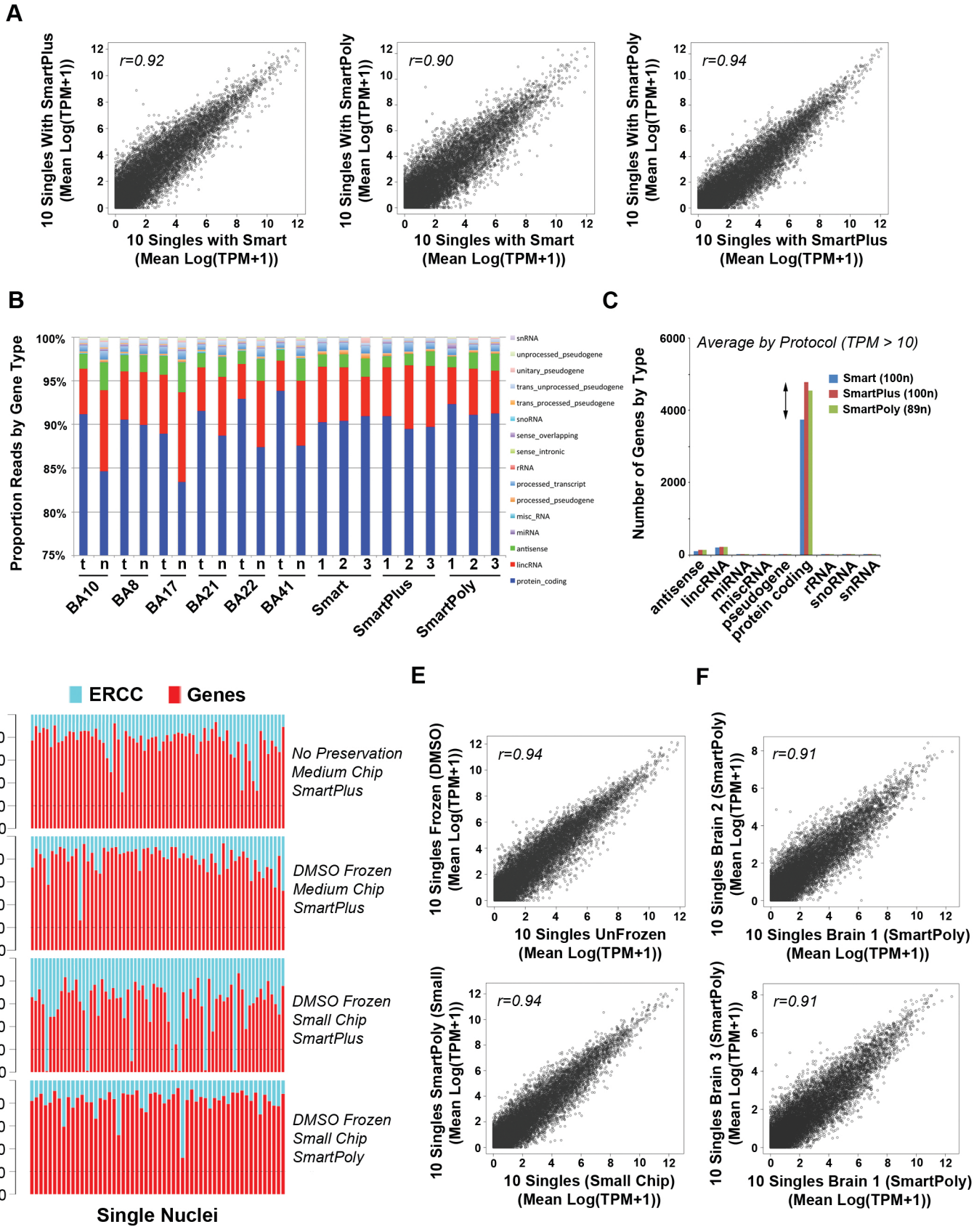

A
10 Singles With SmartPlus (Mean Log(TPM+1))
10 Singles with Smart (Mean Log(TPM+1))
r=0.92
10 Singles With SmartPoly (Mean Log(TPM+1))
10 Singles with Smart (Mean Log(TPM+1))
r=0.90
10 Singles With SmartPoly (Mean Log(TPM+1))
10 Singles with SmartPlus (Mean Log(TPM+1))
r=0.94
B
Proportion Reads by Gene Type
100%
95%
90%
85%
80%
75%
t n t n t n t n t n t n 1 2 3 1 2 3 1 2 3
BA10 BA8 BA17 BA21 BA22 BA41 Smart SmartPlus SmartPoly
snRNA
unprocessed_pseudogene
unitary_pseudogene
trans_unprocessed_pseudogene
trans_processed_pseudogene
snoRNA
sense_overlapping
sense_intronic
rRNA
processed_transcript
processed_pseudogene
misc_RNA
miRNA
antisense
lincRNA
protein_coding
C
Average by Protocol (TPM > 10)
Number of Genes by Type
6000
4000
2000
0
Smart (100n)
SmartPlus (100n)
SmartPoly (89n)
antisense lincRNA miRNA miscRNA pseudogene protein coding rRNA snoRNA snRNA
D
ERCC
Genes
Mapping Ratio (%)
No Preservation Medium Chip SmartPlus
DMSO Frozen Medium Chip SmartPlus
DMSO Frozen Small Chip SmartPlus
DMSO Frozen Small Chip SmartPoly
Single Nuclei
E
10 Singles Frozen (DMSO) (Mean Log(TPM+1))
10 Singles UnFrozen (Mean Log(TPM+1))
r=0.94
10 Singles SmartPoly (Small) (Mean Log(TPM+1))
10 Singles (Small Chip) (Mean Log(TPM+1))
r=0.94
F
10 Singles Brain 2 (SmartPoly) (Mean Log(TPM+1))
10 Singles Brain 1 (SmartPoly) (Mean Log(TPM+1))
r=0.91
10 Singles Brain 3 (SmartPoly) (Mean Log(TPM+1))
10 Singles Brain 1 (SmartPoly) (Mean Log(TPM+1))
r=0.91

Fig. S5

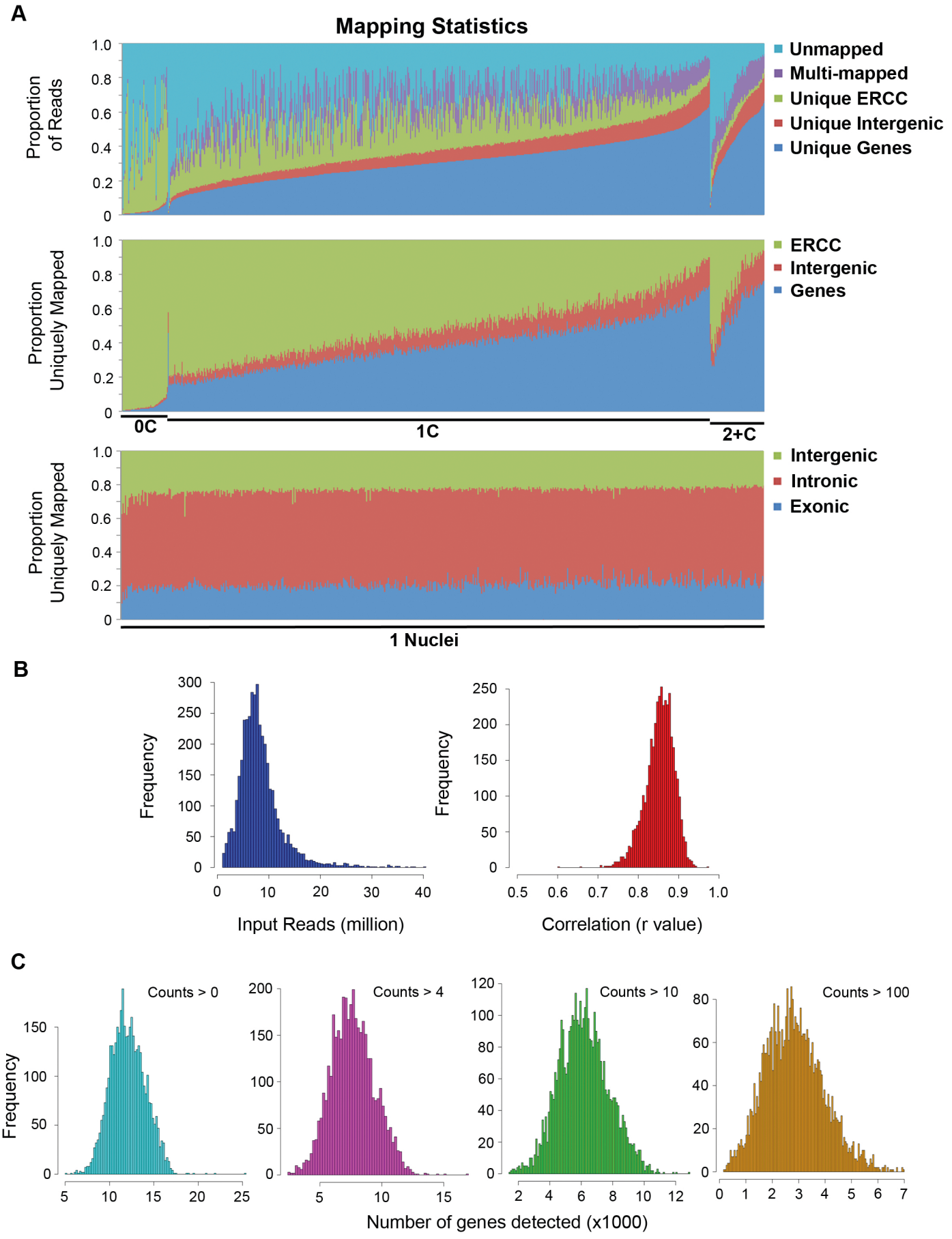

A
Mapping Statistics
Proportion of Reads
Unmapped
Multi-mapped
Unique ERCC
Unique Intergenic
Unique Genes
Proportion Uniquely Mapped
ERCC
Intergenic
Genes
0C
1C
2+C
Proportion Uniquely Mapped
Intergenic
Intronic
Exonic
1 Nuclei
B
Frequency
Input Reads (million)
Frequency
Correlation (r value)
C
Counts > 0
Counts > 4
Counts > 10
Counts > 100
Frequency
Number of genes detected (x1000)

Fig. S6

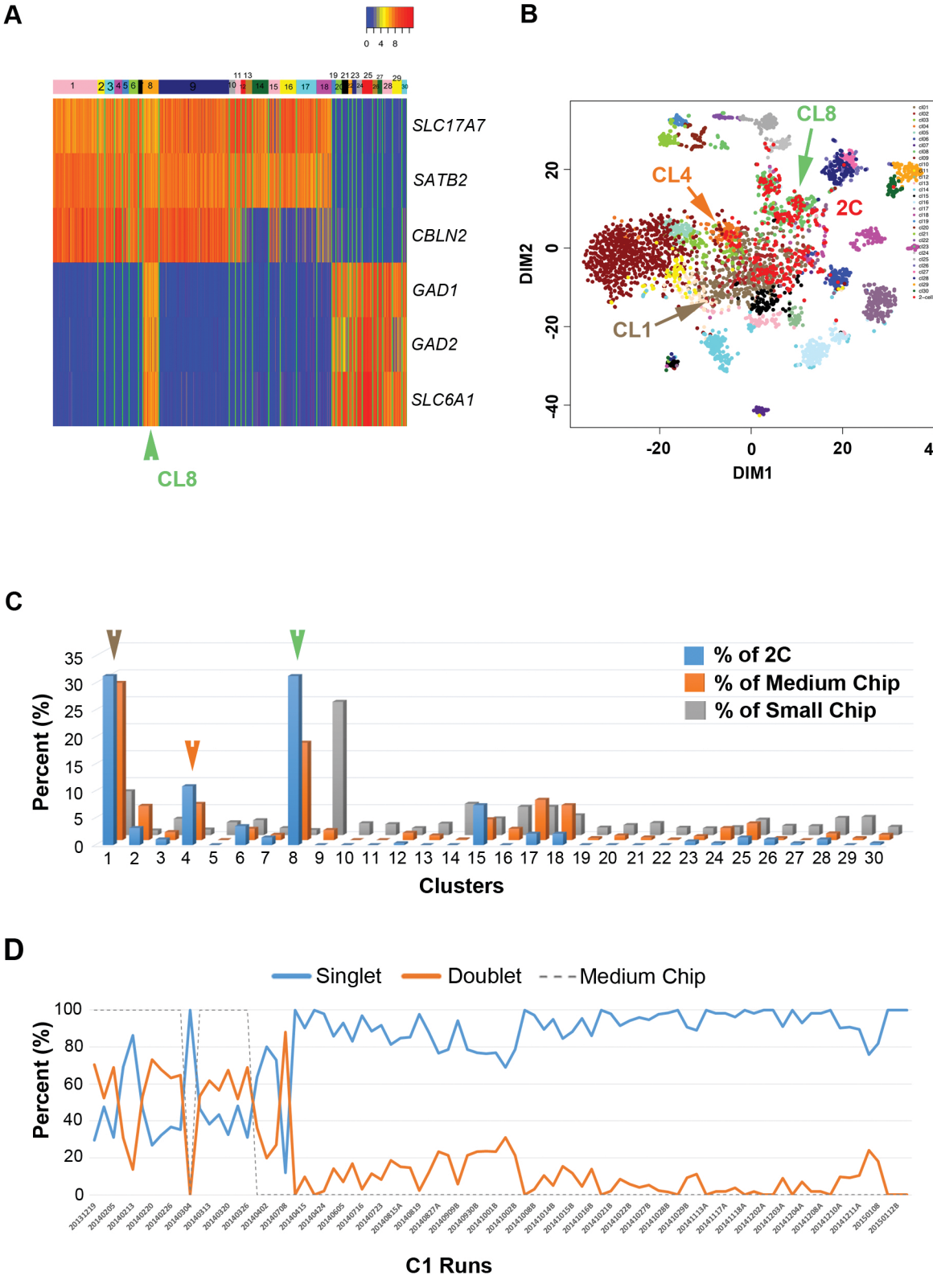

A
0 4 8
SLC17A7
SATB2
CBLN2
GAD1
GAD2
SLC6A1
CL8
B
CL8
CL4
2C
CL1
DIM2
DIM1
C
% of 2C
% of Medium Chip
% of Small Chip
Percent (%)
Clusters
D
Singlet
Doublet
Medium Chip
Percent (%)
C1 Runs

Fig. S7

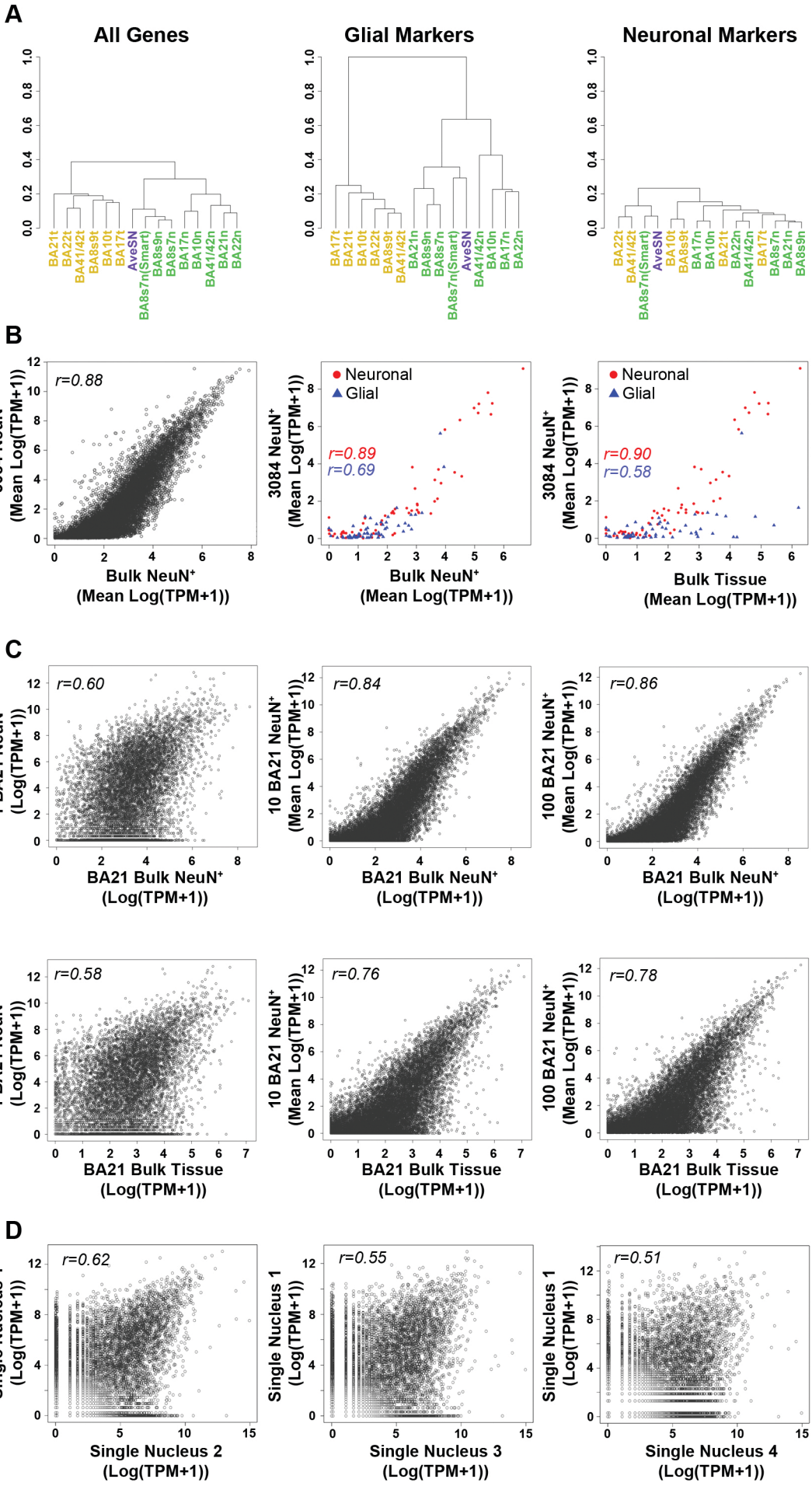
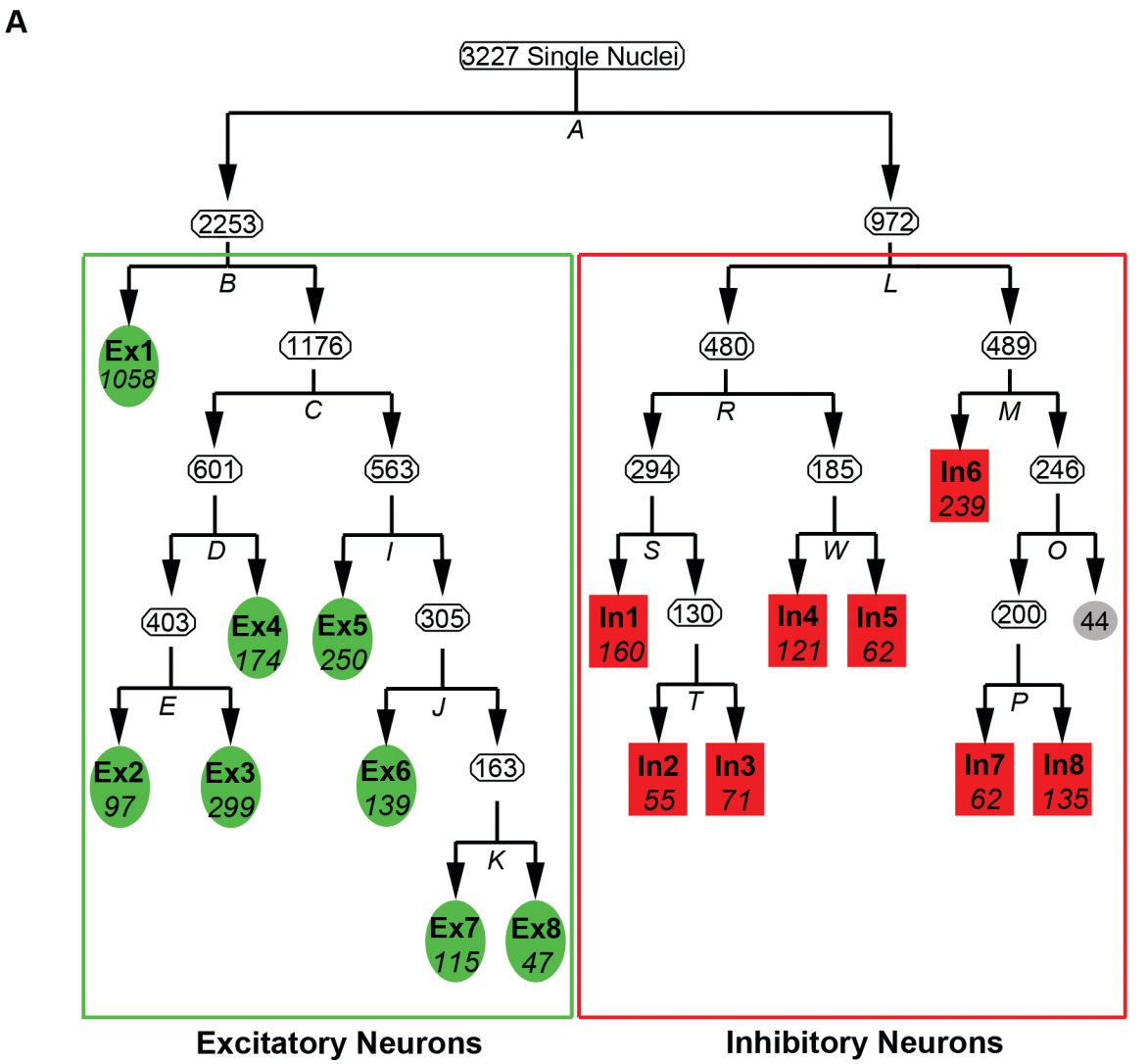
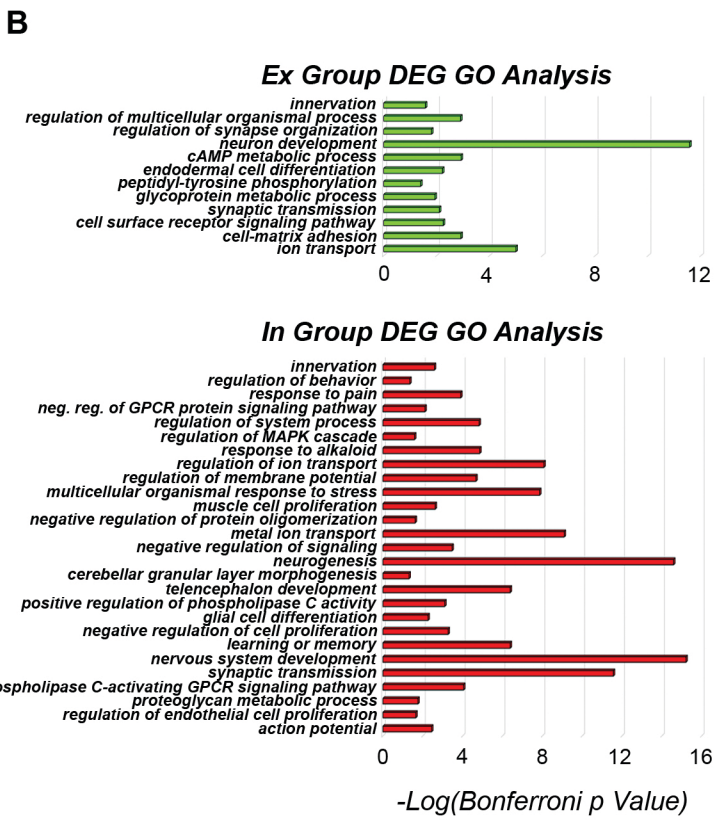

A
All Genes
Glial Markers
Neuronal Markers
BA21t
BA22t
BA41/42t
BA8s9t
BA10t
BA17t
AveSN
BA8s7n(Smart)
BA8s9n
BA8s7n
BA17n
BA10n
BA41/42n
BA21n
BA22n
B
3084 NeuN+ (Mean Log(TPM+1))
Bulk NeuN+ (Mean Log(TPM+1))
r=0.88
Neuronal
Glial
r=0.89
r=0.69
r=0.90
r=0.58
Bulk Tissue (Mean Log(TPM+1))
C
1 BA21 NeuN+ (Log(TPM+1))
10 BA21 NeuN+ (Mean Log(TPM+1))
100 BA21 NeuN+ (Mean Log(TPM+1))
BA21 Bulk NeuN+ (Log(TPM+1))
r=0.60
r=0.84
r=0.86
BA21 Bulk Tissue (Log(TPM+1))
r=0.58
r=0.76
r=0.78
D
Single Nucleus 1 (Log(TPM+1))
Single Nucleus 2 (Log(TPM+1))
Single Nucleus 3 (Log(TPM+1))
Single Nucleus 4 (Log(TPM+1))
r=0.62
r=0.55
r=0.51

Fig. S8


A
3227 Single Nuclei
A
2253
972
B
L
Ex1
1058
1176
480
489
C
R
M
601
563
294
185
In6
239
246
D
I
S
W
O
403
Ex4
174
Ex5
250
305
In1
160
130
In4
121
In5
62
200
44
E
J
T
P
Ex2
97
Ex3
299
Ex6
139
163
In2
55
In3
71
In7
62
In8
135
K
Ex7
115
Ex8
47
Excitatory Neurons
Inhibitory Neurons


B
Ex Group DEG GO Analysis
innervation
regulation of multicellular organismal process
regulation of synapse organization
neuron development
cAMP metabolic process
endodermal cell differentiation
peptidyl-tyrosine phosphorylation
glycoprotein metabolic process
synaptic transmission
cell surface receptor signaling pathway
cell-matrix adhesion
ion transport
0
4
8
12
In Group DEG GO Analysis
innervation
regulation of behavior
response to pain
neg. reg. of GPCR protein signaling pathway
regulation of system process
regulation of MAPK cascade
response to alkaloid
regulation of ion transport
regulation of membrane potential
multicellular organismal response to stress
muscle cell proliferation
negative regulation of protein oligomerization
metal ion transport
negative regulation of signaling
neurogenesis
cerebellar granular layer morphogenesis
telencephalon development
positive regulation of phospholipase C activity
glial cell differentiation
negative regulation of cell proliferation
learning or memory
nervous system development
synaptic transmission
phospholipase C-activating GPCR signaling pathway
proteoglycan metabolic process
regulation of endothelial cell proliferation
action potential
0
4
8
12
16
-Log(Bonferroni p Value)

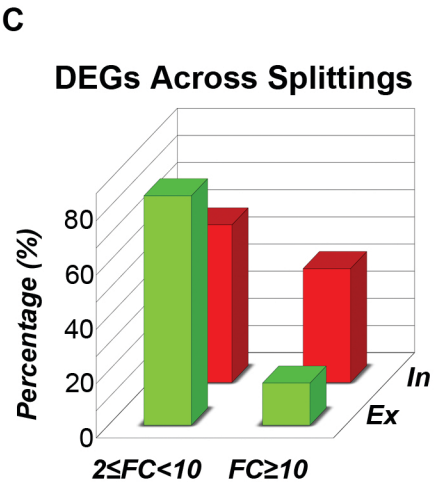

C
DEGs Across Splittings
Percentage (%)
0
20
40
60
80
In
Ex
2≤FC<10
FC≥10

Fig. S9

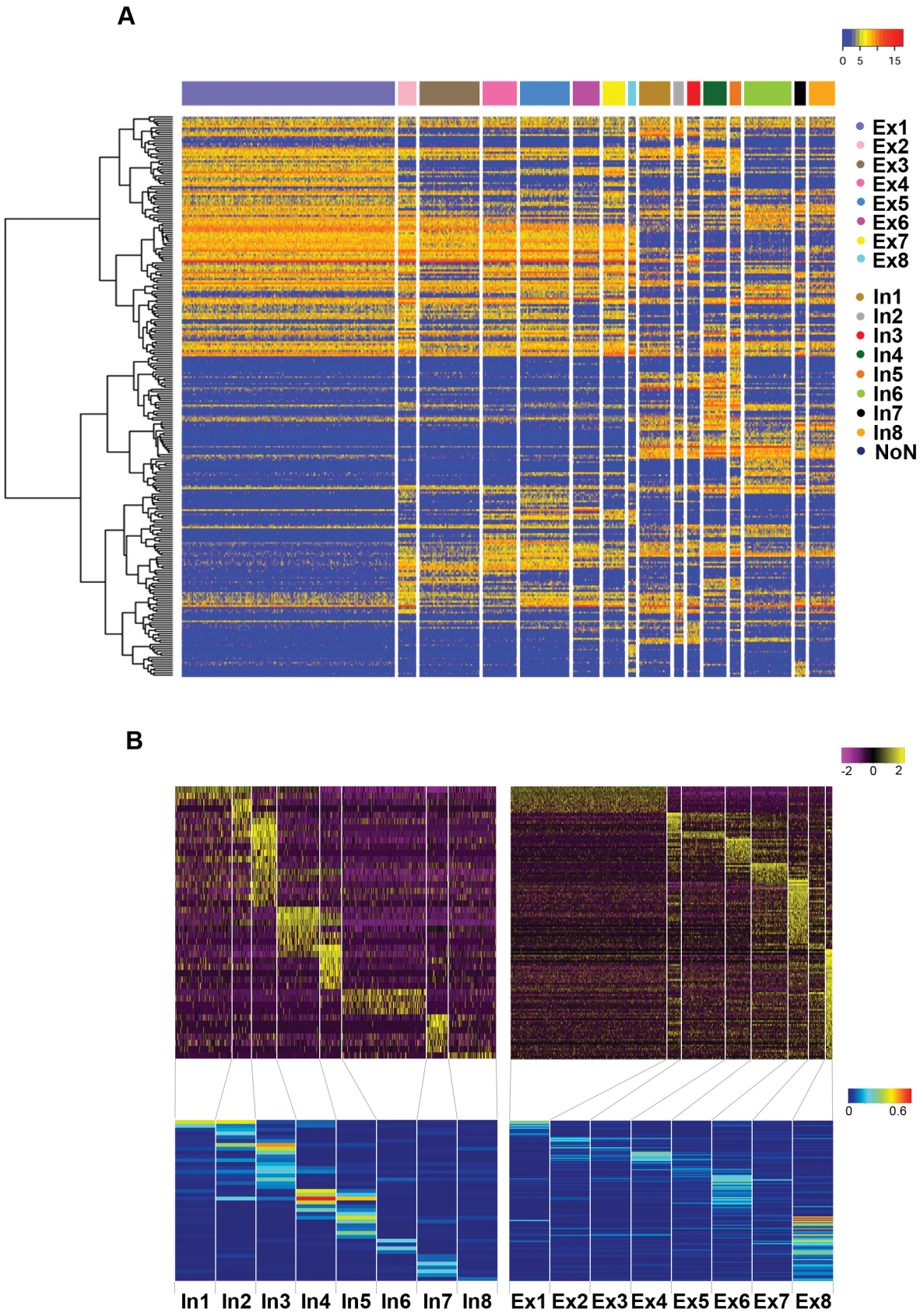
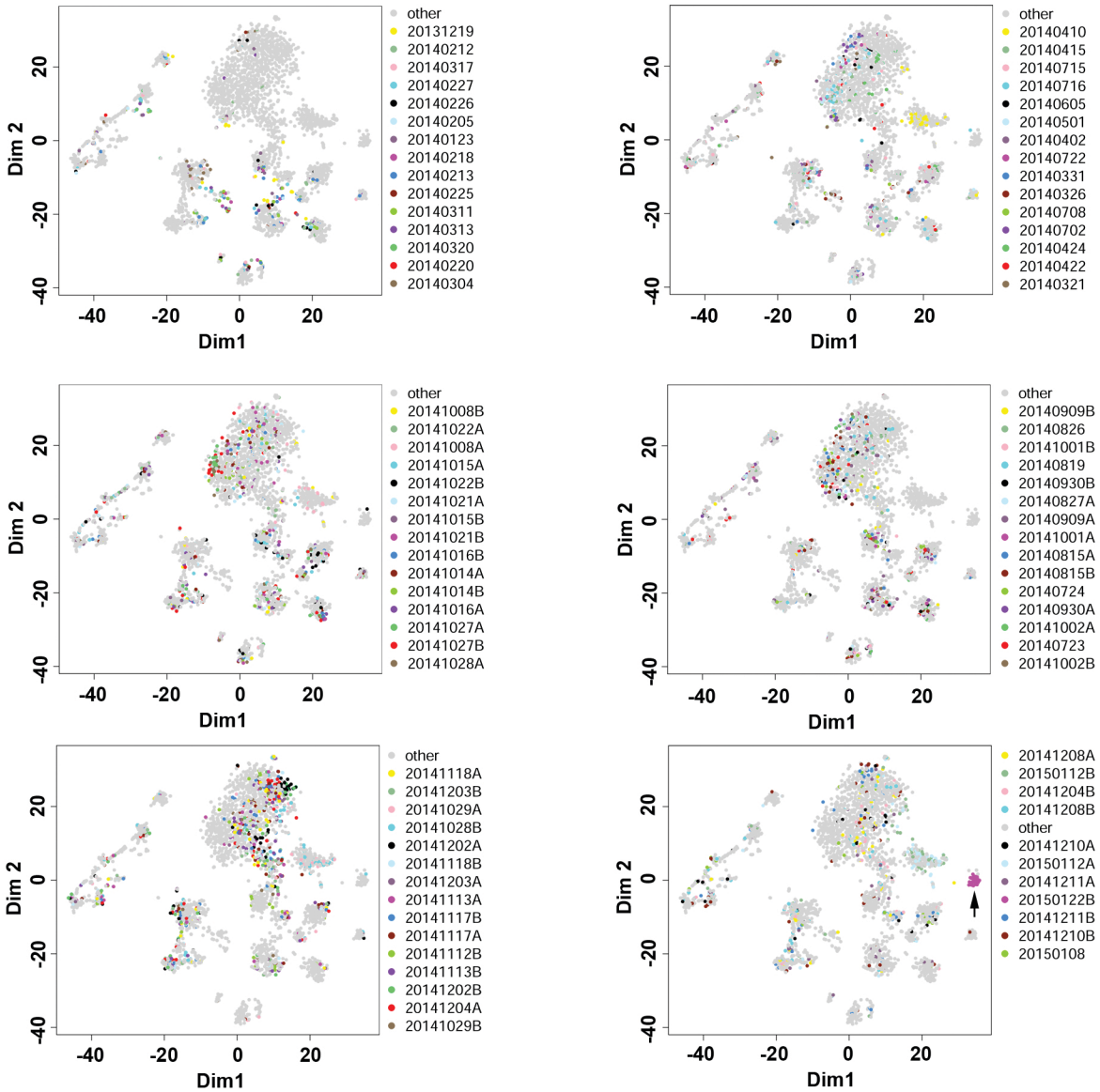

A
0 5 15
Ex1
Ex2
Ex3
Ex4
Ex5
Ex6
Ex7
Ex8
In1
In2
In3
In4
In5
In6
In7
In8
NoN
B
-2 0 2
0 0.6
In1 In2 In3 In4 In5 In6 In7 In8
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8

Fig. S10

A


Dim 2
Dim1
other
20131219
20140212
20140317
20140227
20140226
20140205
20140123
20140218
20140213
20140225
20140311
20140313
20140320
20140220
20140304
other
20140410
20140415
20140715
20140716
20140605
20140501
20140402
20140722
20140331
20140326
20140708
20140702
20140424
20140422
20140321
other
20141008B
20141022A
20141008A
20141015A
20141022B
20141021A
20141015B
20141021B
20141016B
20141014A
20141014B
20141016A
20141027A
20141027B
20141028A
other
20140909B
20140826
20141001B
20140819
20140930B
20140827A
20140909A
20141001A
20140815A
20140815B
20140724
20140930A
20141002A
20140723
20141002B
other
20141118A
20141203B
20141029A
20141028B
20141202A
20141118B
20141203A
20141113A
20141117B
20141117A
20141112B
20141113B
20141202B
20141204A
20141029B
20141208A
20150112B
20141204B
20141208B
other
20141210A
20150112A
20141211A
20150122B
20141211B
20141210B
20150108

B

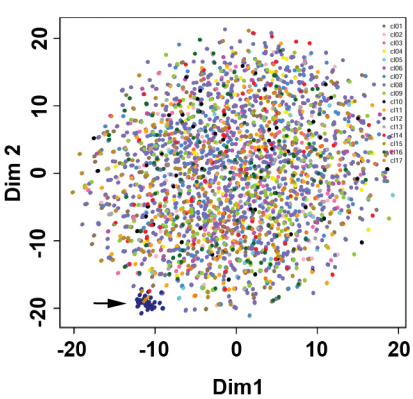

Dim 2
Dim1

Fig. S11

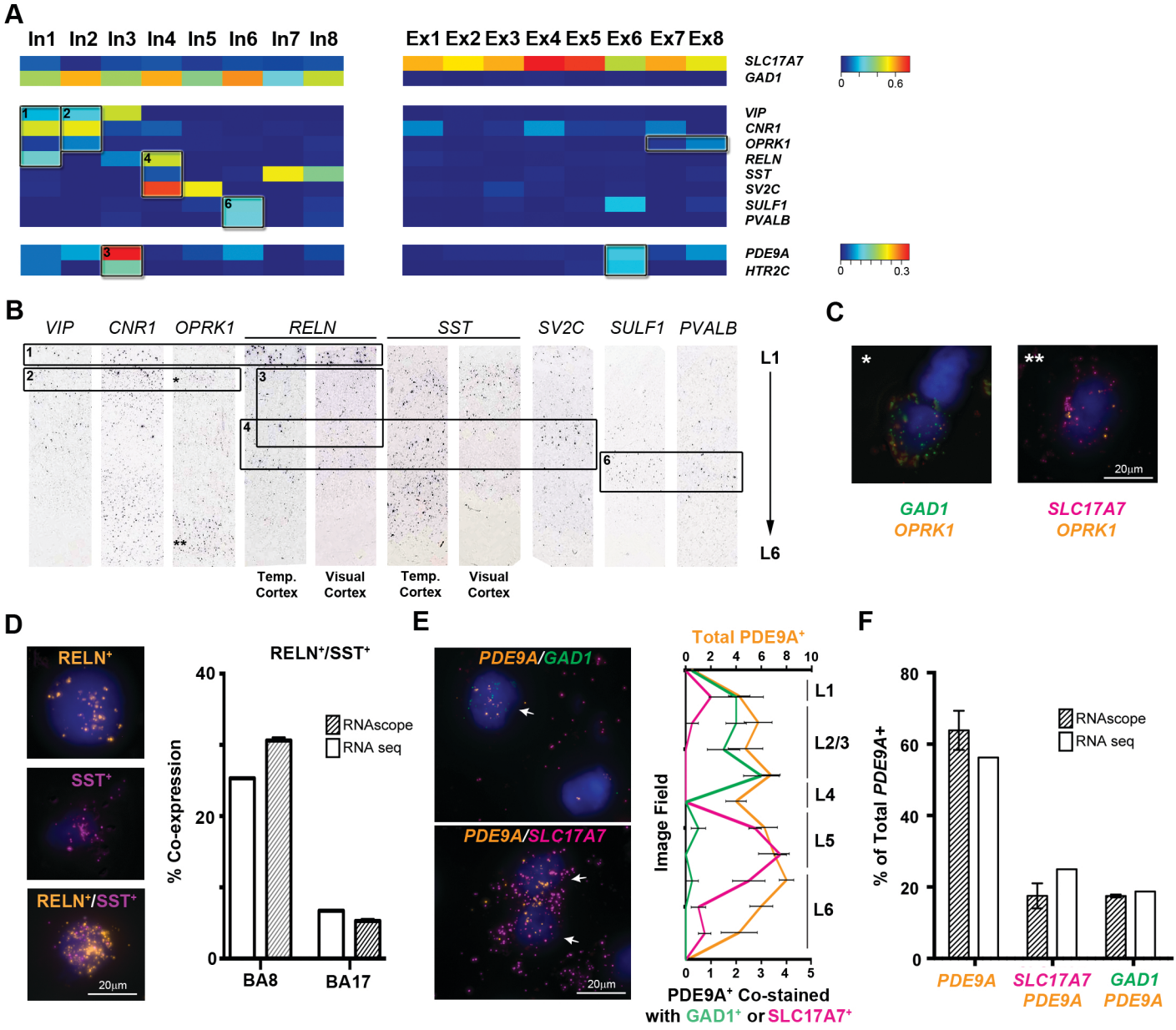

A
In1 In2 In3 In4 In5 In6 In7 In8
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8
SLC17A7
GAD1
0
0.6
VIP
CNR1
OPRK1
RELN
SST
SV2C
SULF1
PVALB
PDE9A
HTR2C
0
0.3
B
VIP
CNR1
OPRK1
RELN
SST
SV2C
SULF1
PVALB
L1
L6
Temp. Cortex
Visual Cortex
Temp. Cortex
Visual Cortex
C
GAD1
OPRK1
SLC17A7
OPRK1
20μm
D
RELN+
SST+
RELN+/SST+
20μm
RELN+/SST+
% Co-expression
RNAscope
RNA seq
BA8
BA17
E
PDE9A/GAD1
PDE9A/SLC17A7
20μm
Total PDE9A+
Image Field
L1
L2/3
L4
L5
L6
PDE9A+ Co-stained with GAD1+ or SLC17A7+
F
% of Total PDE9A+
RNAscope
RNA seq
PDE9A
SLC17A7 PDE9A
GAD1 PDE9A

Fig. S12

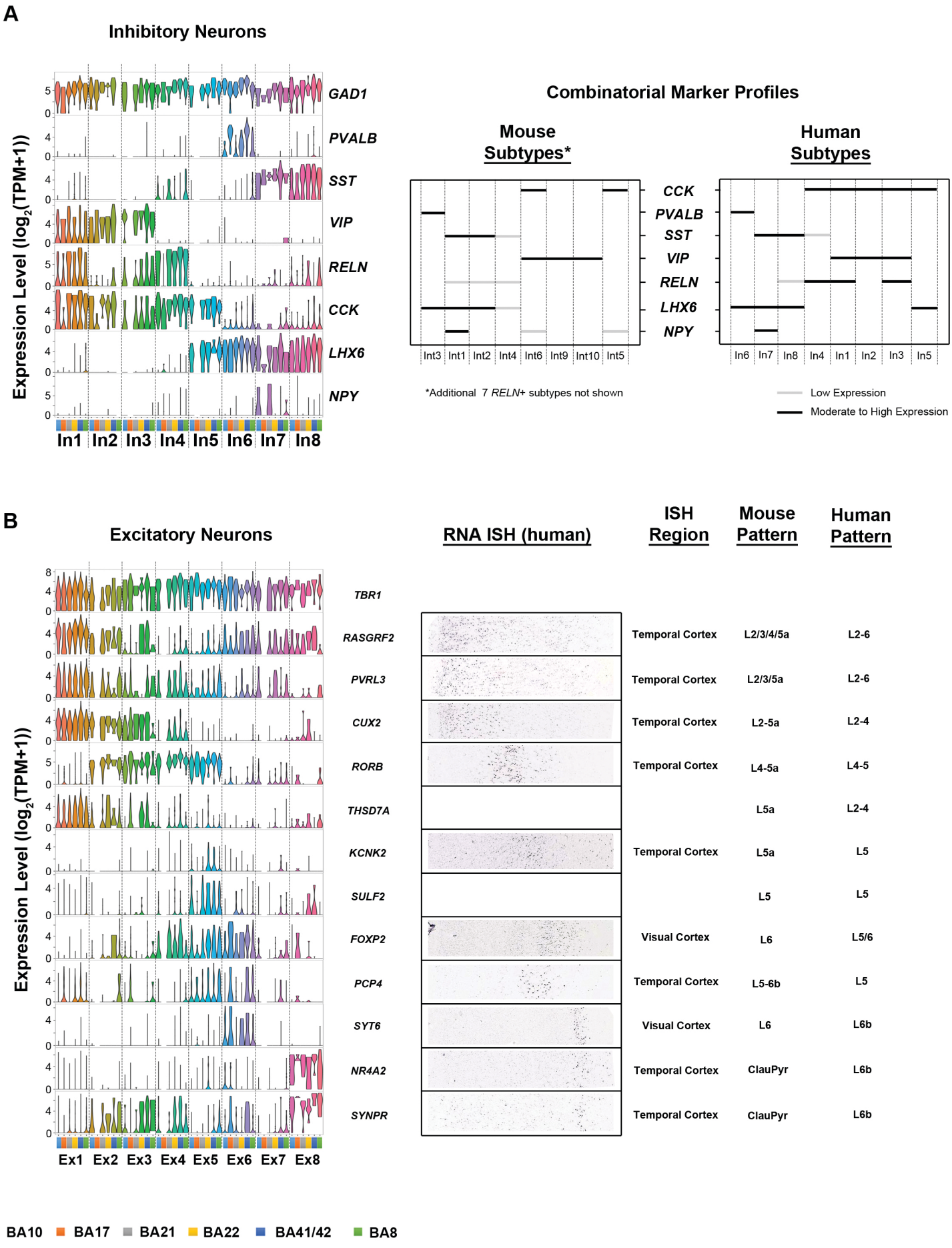

A
Inhibitory Neurons
Expression Level (log2(TPM+1))
GAD1
PVALB
SST
VIP
RELN
CCK
LHX6
NPY
In1 In2 In3 In4 In5 In6 In7 In8
Combinatorial Marker Profiles
Mouse Subtypes*
Human Subtypes
Int3 Int1 Int2 Int4 Int6 Int9 Int10 Int5
In6 In7 In8 In4 In1 In2 In3 In5
*Additional 7 RELN+ subtypes not shown
Low Expression
Moderate to High Expression
B
Excitatory Neurons
Expression Level (log2(TPM+1))
TBR1
RASGRF2
PVRL3
CUX2
RORB
THSD7A
KCNK2
SULF2
FOXP2
PCP4
SYT6
NR4A2
SYNPR
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8
RNA ISH (human)
ISH Region
Mouse Pattern
Human Pattern
Temporal Cortex L2/3/4/5a L2-6
Temporal Cortex L2/3/5a L2-6
Temporal Cortex L2-5a L2-4
Temporal Cortex L4-5a L4-5
L5a L2-4
Temporal Cortex L5a L5
L5 L5
Visual Cortex L6 L5/6
Temporal Cortex L5-6b L5
Visual Cortex L6 L6b
Temporal Cortex ClauPyr L6b
Temporal Cortex ClauPyr L6b
BA10 BA17 BA21 BA22 BA41/42 BA8

Fig. S13

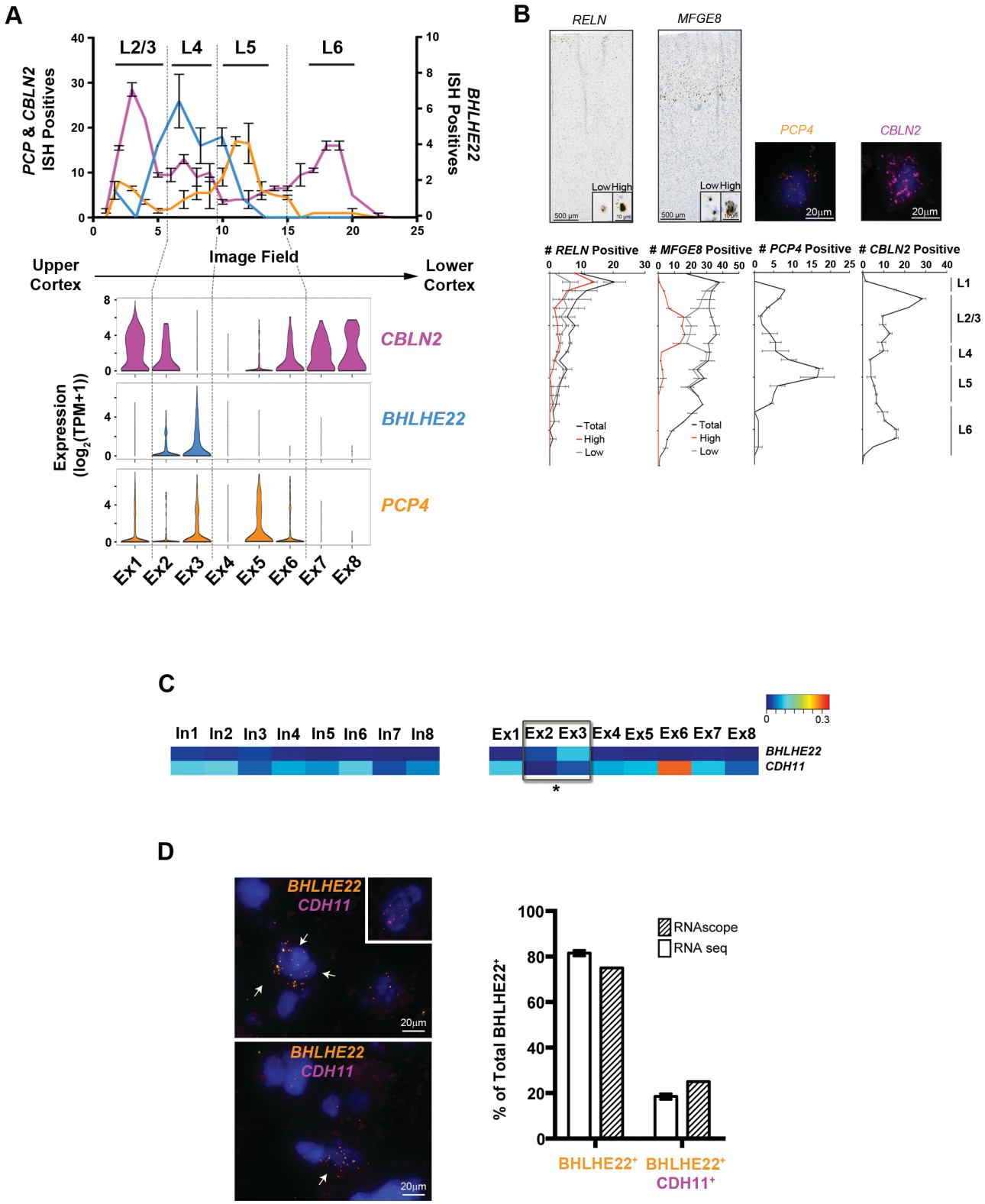

A
L2/3
L4
L5
L6
PCP & CBLN2 ISH Positives
BHLHE22 ISH Positives
Image Field
Upper Cortex
Lower Cortex
CBLN2
BHLHE22
PCP4
Expression (log2(TPM+1))
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8
B
RELN
MFGE8
PCP4
CBLN2
Low High
20μm
RELN Positive
MFGE8 Positive
PCP4 Positive
CBLN2 Positive
Total
High
Low
L1
L2/3
L4
L5
L6
C
In1 In2 In3 In4 In5 In6 In7 In8
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8
BHLHE22
CDH11
D
BHLHE22
CDH11
20μm
RNAscope
RNA seq
% of Total BHLHE22+
BHLHE22+
BHLHE22+ CDH11+

Fig. S14

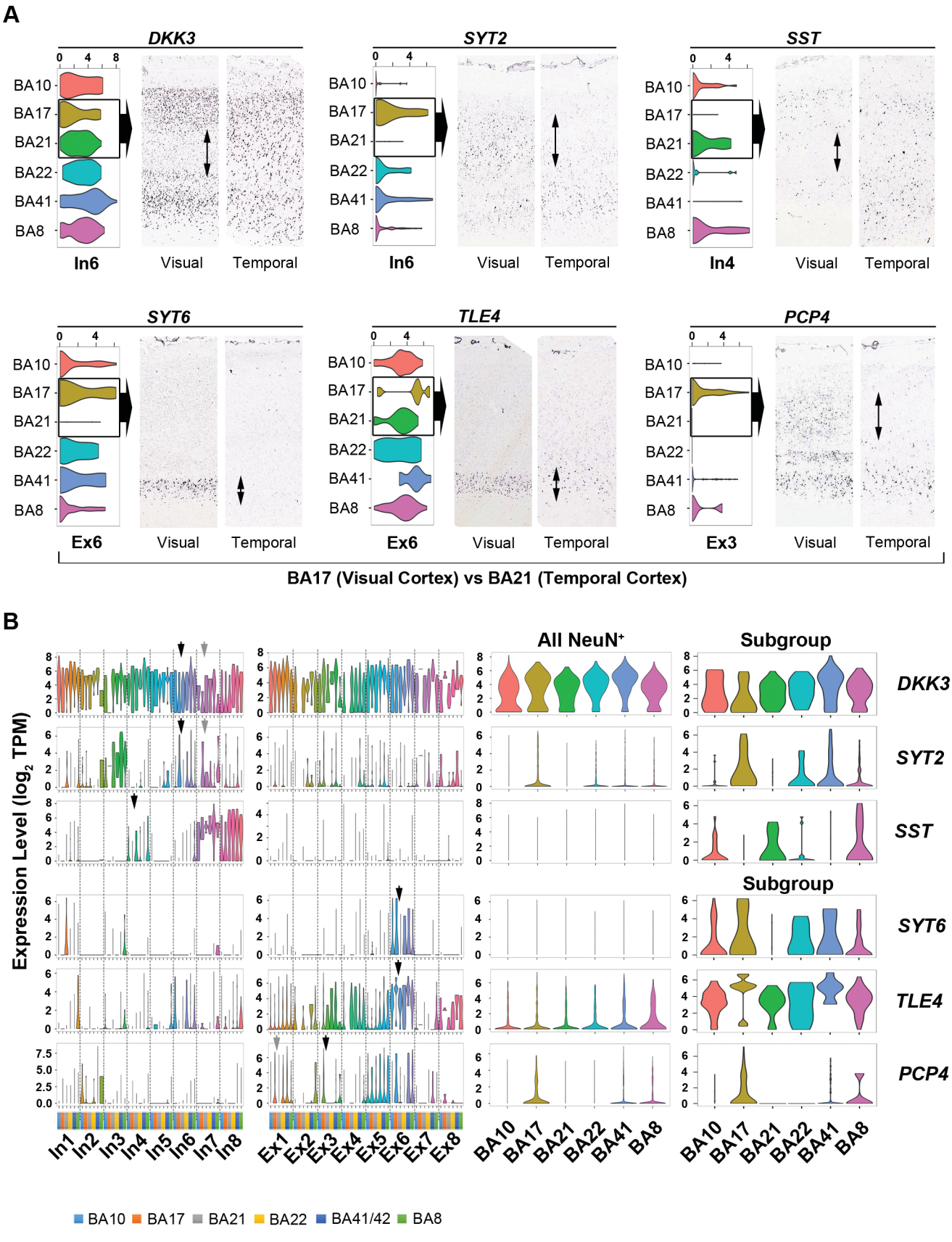

A
DKK3
SYT2
SST
SYT6
TLE4
PCP4
BA10
BA17
BA21
BA22
BA41
BA8
In6
In4
Ex6
Ex3
Visual
Temporal
BA17 (Visual Cortex) vs BA21 (Temporal Cortex)
B
All NeuN+
Subgroup
Expression Level (log2 TPM)
In1 In2 In3 In4 In5 In6 In7 In8
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8
BA10 BA17 BA21 BA22 BA41 BA8
BA10 BA17 BA21 BA22 BA41/42 BA8

Fig. S15

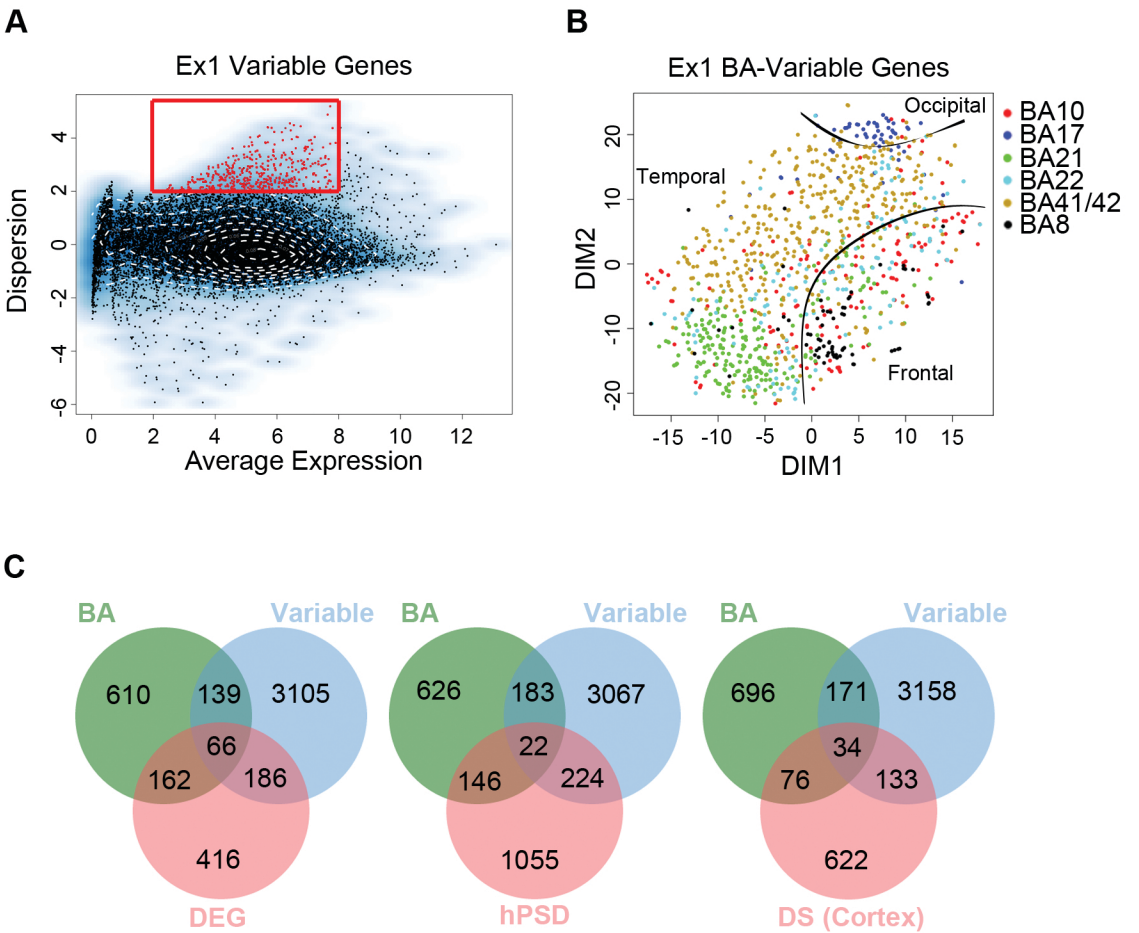
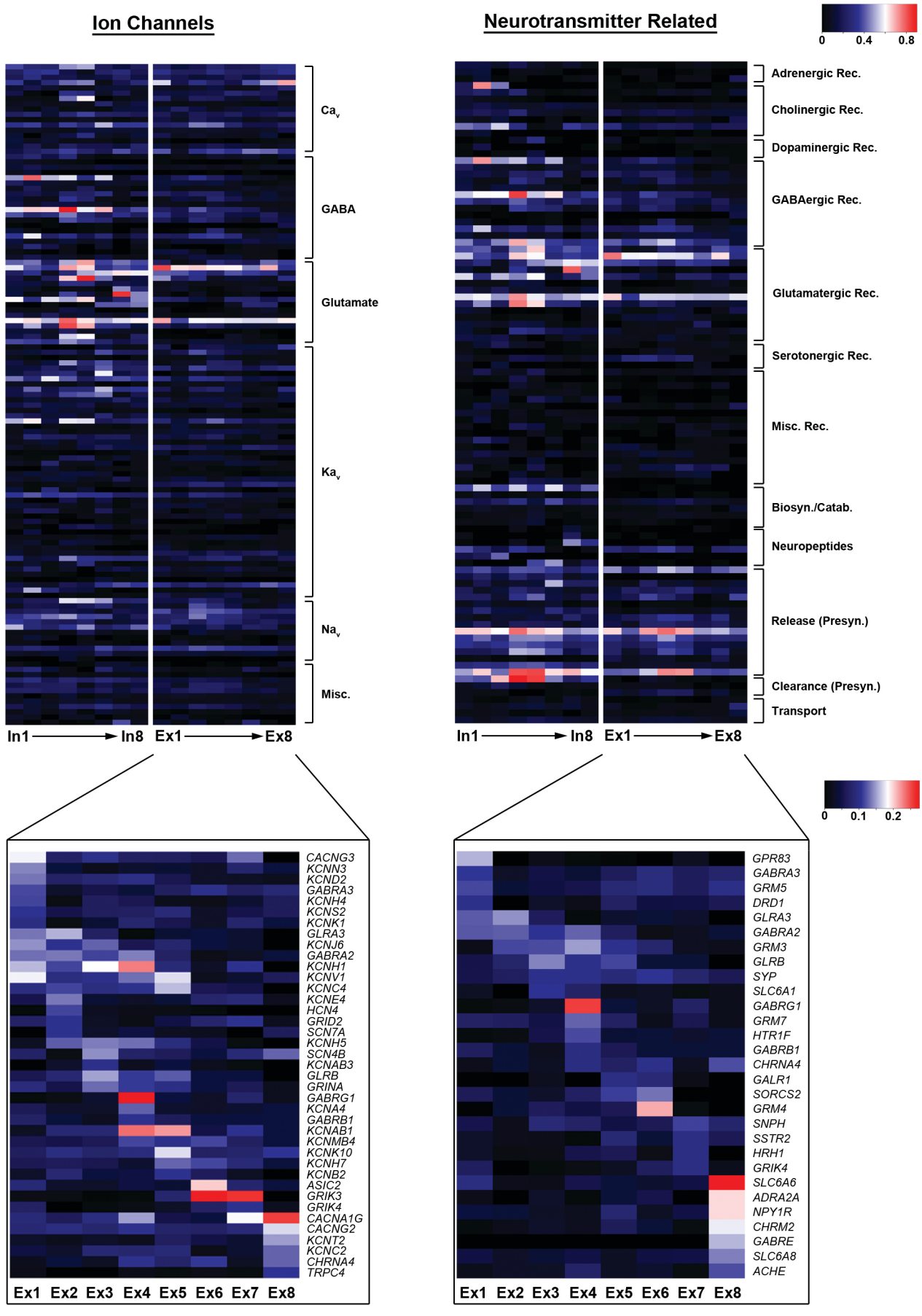

A
Ex1 Variable Genes
Dispersion
Average Expression
B
Ex1 BA-Variable Genes
Occipital
Temporal
Frontal
DIM1
DIM2
BA10
BA17
BA21
BA22
BA41/42
BA8
C
BA
Variable
610
139
3105
66
162
186
416
DEG
626
183
3067
22
146
224
1055
hPSD
696
171
3158
34
76
133
622
DS (Cortex)

Fig. S16


Ion Channels
Neurotransmitter Related
0
0.4
0.8
Ca$_v$
GABA
Glutamate
Ka$_v$
Na$_v$
Misc.
In1 → In8 Ex1 → Ex8
Adrenergic Rec.
Cholinergic Rec.
Dopaminergic Rec.
GABAergic Rec.
Glutamatergic Rec.
Serotonergic Rec.
Misc. Rec.
Biosyn./Catab.
Neuropeptides
Release (Presyn.)
Clearance (Presyn.)
Transport
In1 → In8 Ex1 → Ex8
0
0.1
0.2
CACNG3
KCNN3
KCND2
GABRA3
KCNH4
KCNS2
KCNK1
GLRA3
KCNJ6
GABRA2
KCNH1
KCNV1
KCNC4
KCNE4
HCN4
GRID2
SCN7A
KCNH5
SCN4B
KCNAB3
GLRB
GRINA
GABRG1
KCNA4
GABRB1
KCNAB1
KCNMB4
KCNK10
KCNH7
KCNB2
ASIC2
GRIK3
GRIK4
CACNA1G
CACNG2
KCNT2
KCNC2
CHRNA4
TRPC4
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8
GPR83
GABRA3
GRM5
DRD1
GLRA3
GABRA2
GRM3
GLRB
SYP
SLC6A1
GABRG1
GRM7
HTR1F
GABRB1
CHRNA4
GALR1
SORCS2
GRM4
SNPH
SSTR2
HRH1
GRIK4
SLC6A6
ADRA2A
NPY1R
CHRM2
GABRE
SLC6A8
ACHE
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8