

Supplemental Figure 1

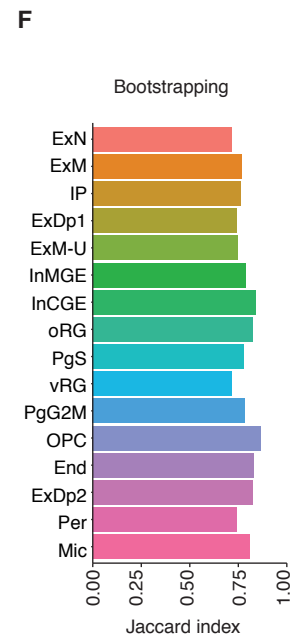
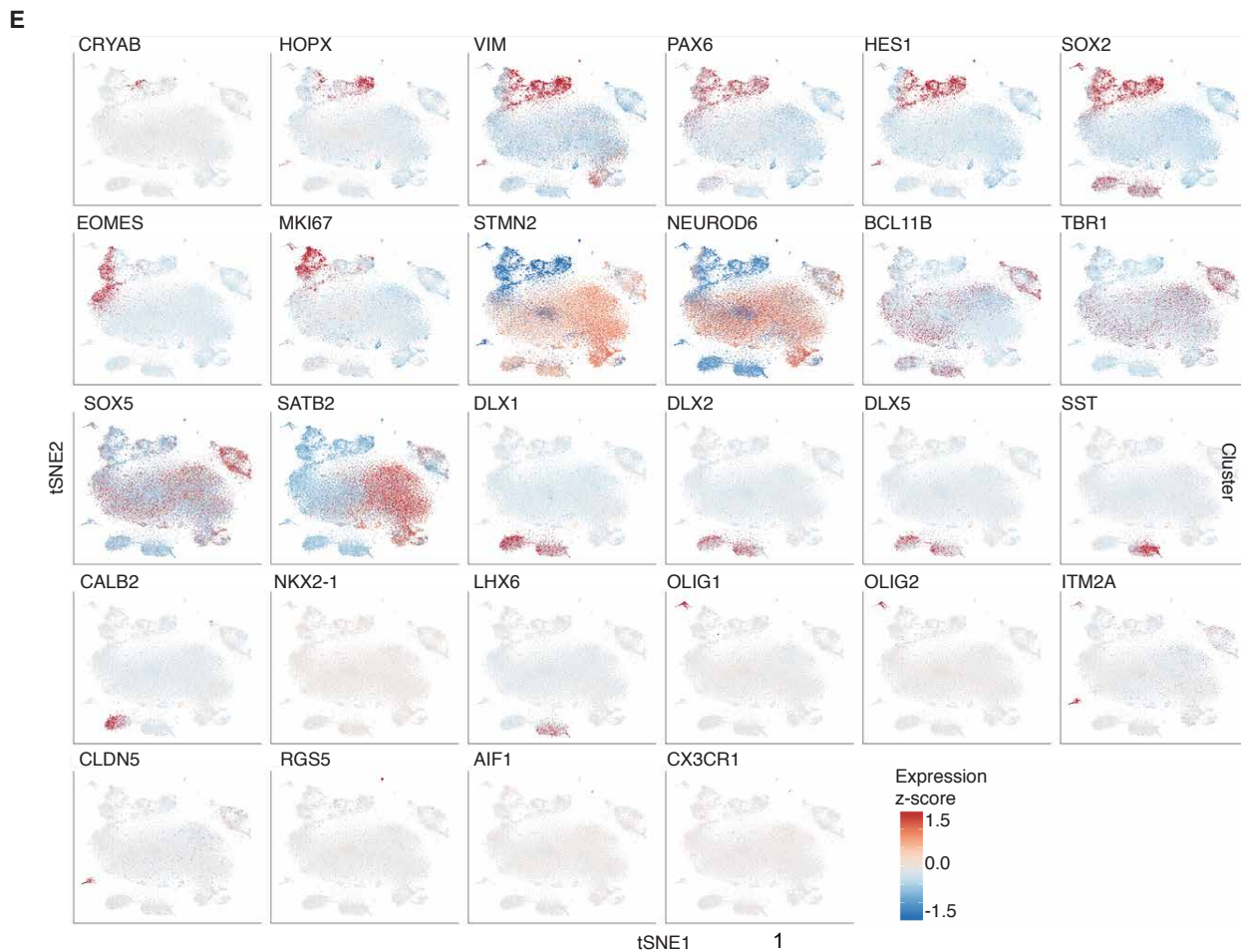
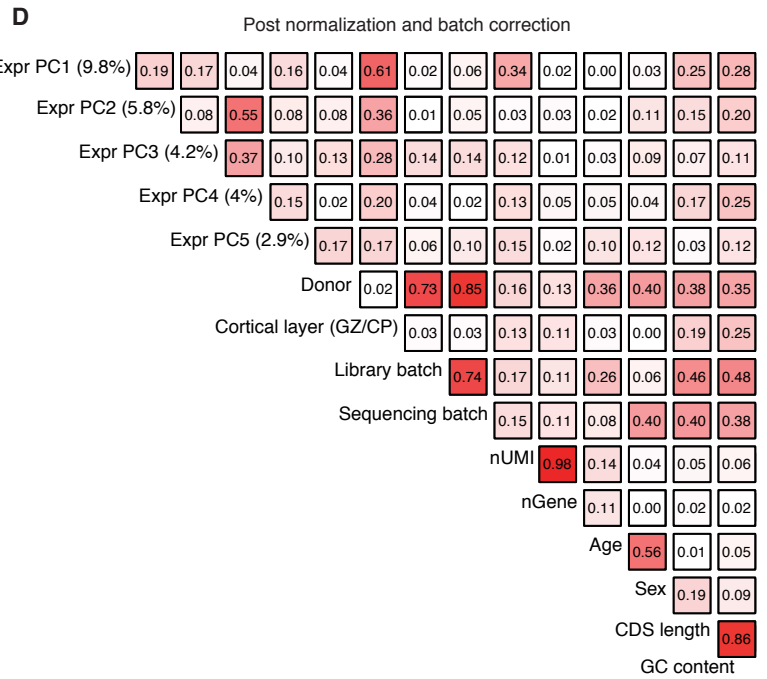
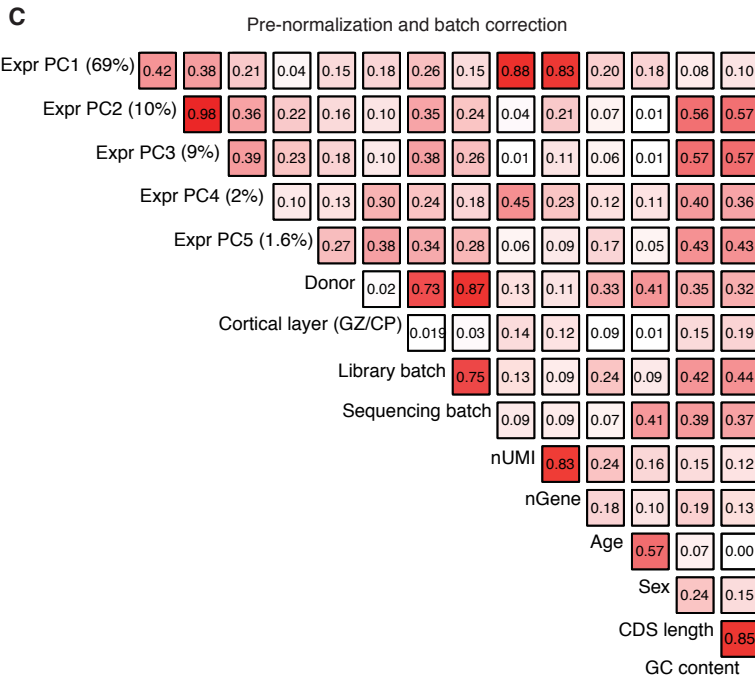
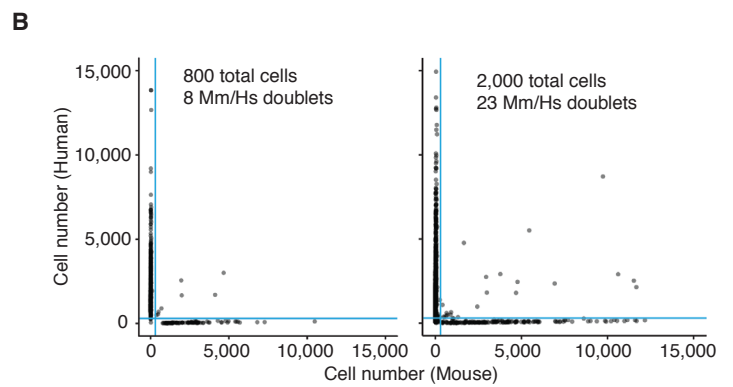
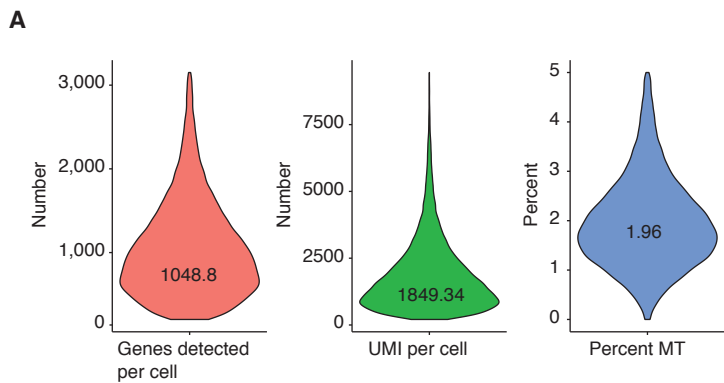


Figure S1 (related to Figure 1). Quality assessment of Drop-seq dataset, batch and technical covariate correction, and evaluation of cluster identity and reproducibility.

(A) scRNA-seq quality metrics of filtered dataset. Genes detected per cell, number of UMI per cell, and percentage of mitochondrial (MT) reads per cell. Value on violin plot is the mean. (B) Representative plots of barnyard experiments from independent cell isolations used to assess doublet rate. Mean doublet rate: 5.8%. (C) Correlation of technical and biological factors to the first five principal components of the raw counts expression matrix. (D) Correlation of technical and biological factors to the first five principal components of the genes used for clustering after normalization and batch correction. The first principal component correlates primarily with the cortical layer (GZ or CP) from which the cells were isolated. Color and number indicate absolute spearman correlation. (E) tSNE colored by expression of canonical cell type marker genes. vRG: CRYAB; oRG: HOPX; RG: VIM, PAX6, HES1, SOX2; IP: EOMES; Mitotic: MK167; Neuronal: STMN2, NEUROD6; Deep layer neuron: BCL11B, TBR1, SOX5; Callosal neuron: SATB2; Interneuron: DLX1, DLX2, DLX5, SST, CALB2, NKX2-1 (striatal), LHX6; OPC: OLIG1, OLIG2; Endothelial: ITM2A, CLDN5; Microglia: RGS5, AIF1, CX3CR1. (F) Jaccard index as a measure of cluster stability. Dataset was sampled, re-quality filtered, normalized, analyzed, and clustered over 100 iterations. Mean Jaccard index of 100 iterations of bootstrapping plotted.

Supplemental Figure 2

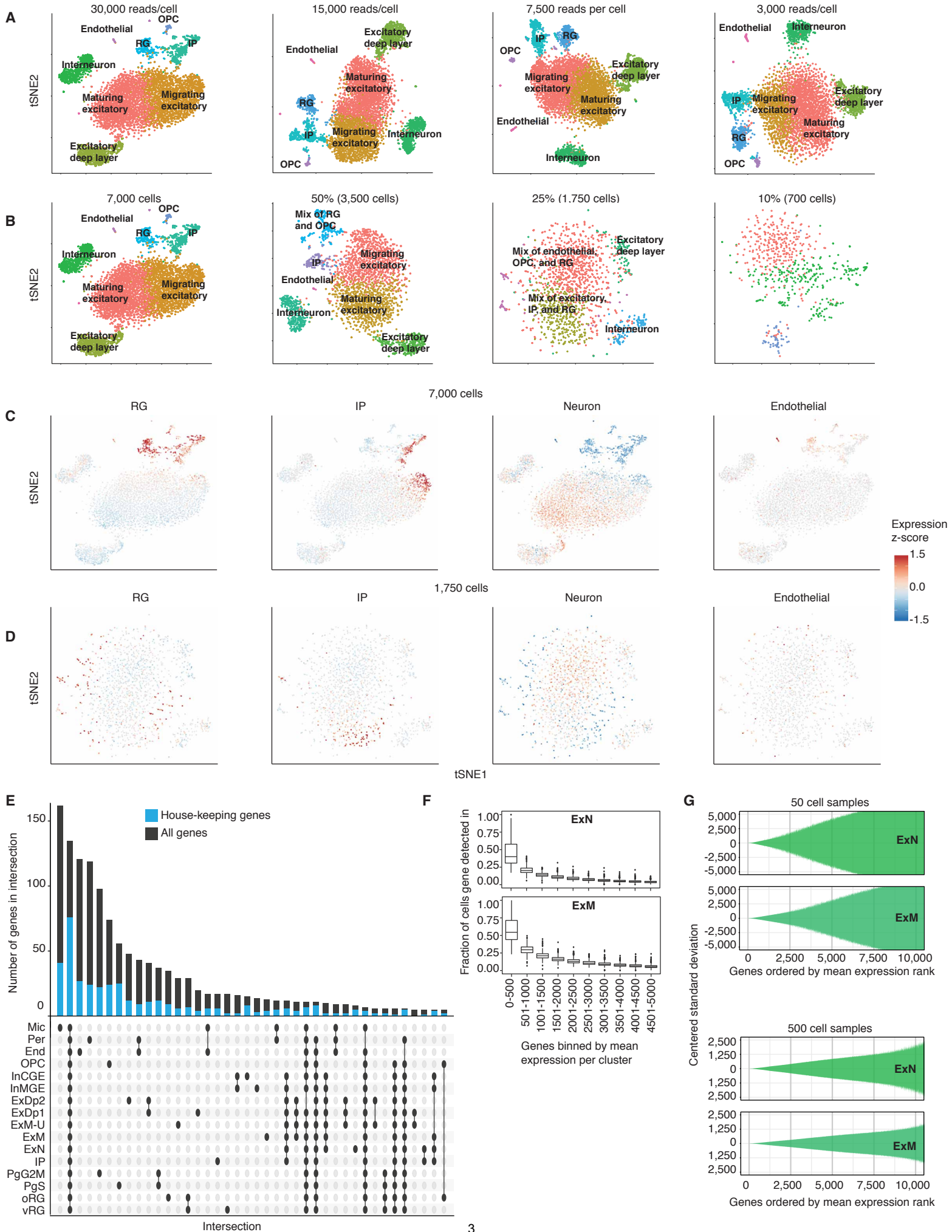


Figure S2 (related to Figure 1). Assessment of cell type detection and scRNA-seq cell type signatures.

(A to D) Influence of cell number and sequencing depth on cell type detection. Analysis performed with a subset of 7,000 cells sequenced to 30,000 reads per cell clustered using Seurat. Recognizing that with scRNA-seq there is an inherent practical tension between sequencing depth and the number of cells profiled, we explored the consequences of this tradeoff by performing down-sampling of sequencing depth and the number of cells analyzed to determine thresholds for cell type detection. (A) Sequencing depth is progressively down-sampled to assess the effect of sequencing depth on cell type detection. Separation of major cell types is maintained at 3,000 reads per cell. (B) Cell number is progressively down-sampled to assess the effect of cell number on cell type detection. Separation of oligodendrocyte precursors (OPCs) from RGs is lost by down-sampling to 3,500 cells, and separation of multiple major cell types is lost by down-sampling to 1,750 cells. (C and D) tSNE colored by mean expression of groups of canonical marker genes of major cell types. (C) 7,000 cell subset. (D) 1,750 cell subset. (E) Intersection of highly expressed genes by cluster. Among the genes most highly expressed in a given cluster, most are specific to a cell type, and not simply highly expressed house-keeping genes. The top 500 expressed genes by cluster are intersected. Bars show size of intersection. Black dots indicate clusters intersected. Blue bars show number of house-keeping genes in the intersected gene list. (F) Gene detection rate versus expression level by cluster. X-axis: genes are ranked by mean expression level by cluster and split into bins. Y-axis: mean gene detection rate of all genes in a bin across all cells in the cluster. (G) Stability of cluster gene expression signatures. Subsamples of 50 cells in the cluster. (G) Stability of cluster gene expression signatures. Subsamples of 50 cells showed high stability in ranking (see materials and methods) for the top 1,000 genes, whereas 500 cells showed very high rank stability for the top 4,000 genes, allowing us to empirically assess the completeness of cell type signatures with different sample sizes.

X-axis: genes are ranked by mean expression level by cluster, with 1 being the highest expressed. 50 cell or 500 cell samples were drawn from the cluster and mean expression rank was recalculated over 1,000 iterations. Y-axis: the standard deviation in mean expression ranking for each gene over the 1,000 iterations.

Supplemental Figure 3

	Drop-seq fetal cortex	Nowakowski <i>et al.</i>	Pollen <i>et al.</i>
Method	Drop-seq	Fluidigm	Fluidigm
Number of cells post QC	33,000	4,261	393
Reads per cell	52,000	2.2 million	2.5 million
Total reads	118 million	16 billion	992 million
Genes detected per cell	1,049	2,403	3,098

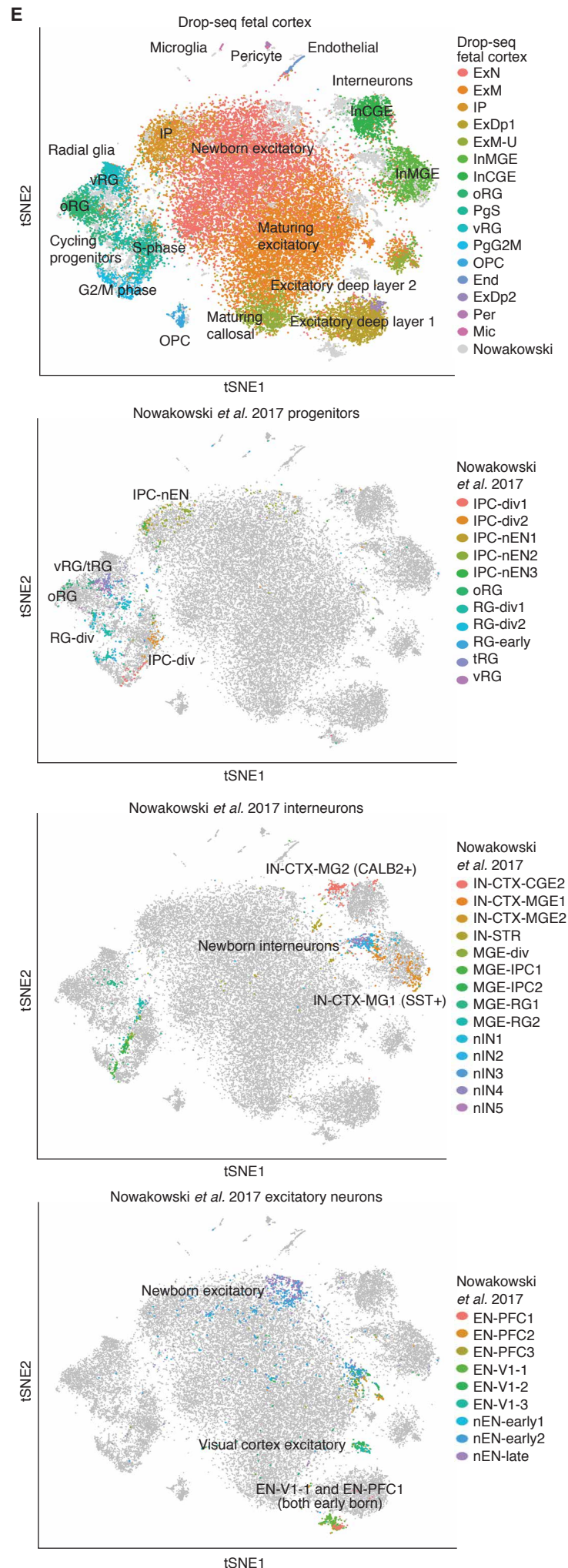
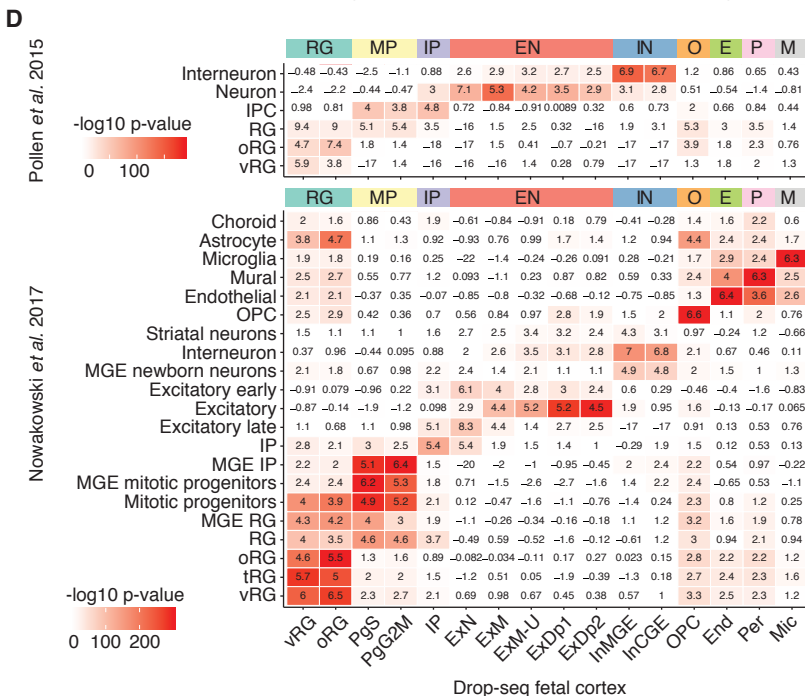
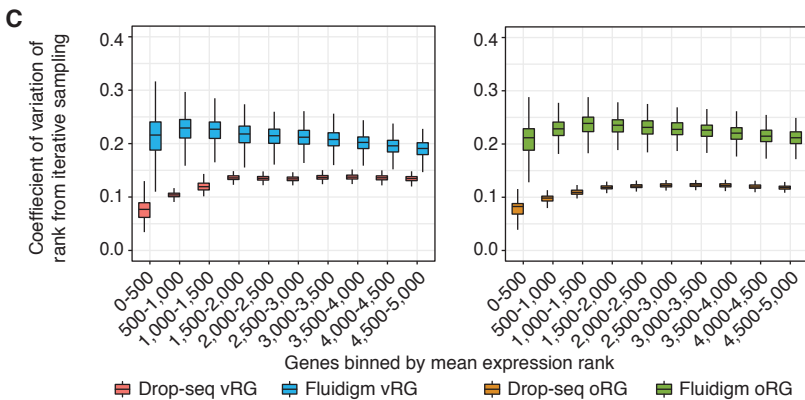
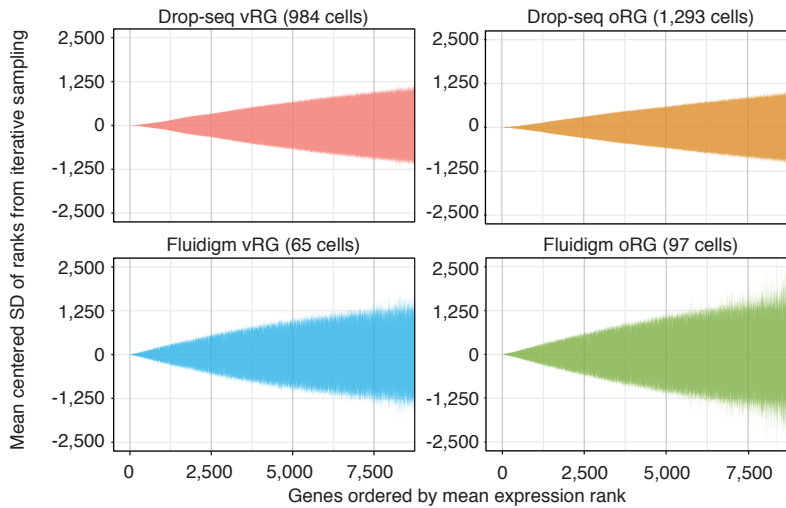
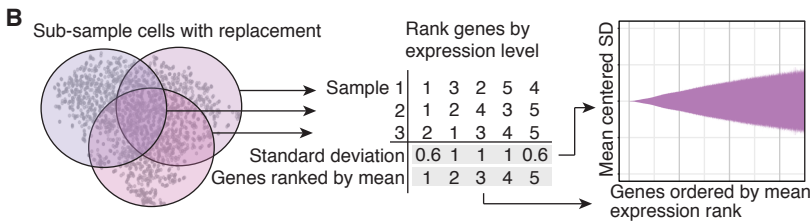


Figure S3 (related to Figure 1). Comparison and integration of scRNA-seq datasets.

(A) Metrics of developing brain scRNA-seq datasets profiled via Fluidigm C1 (Nowakowski et al., 2017; Pollen et al., 2015) or Drop-seq used to compare methods and assess the reproducibility of cell type signatures. vRG and oRG clusters were used to assess the tradeoff in using a higher-depth sequencing lower-throughput scRNA-seq method (Fluidigm C1) versus a lower-depth higher-throughput single-cell method (Drop-seq), with respect to the completeness of cell type gene expression signatures. (B and C) Stability of gene expression signatures from Drop-seq and Fluidigm C1 of vRG and oRG cell populations. Although Fluidigm C1 detected a greater number of genes on average from individual cells, the ability to leverage an order of magnitude more cells with Drop-seq provided more stable mRNA transcript profiles for a given cell type. (B) X-axis: genes are ranked by mean expression level for the indicated cell population and dataset, with 1 being the highest expressed. Y-axis: Samples equal to 75% of the cell population were drawn from the cluster and mean expression rank was recalculated over 1,000 iterations. The y-axis shows the standard deviation in mean expression ranking for each gene over the 1,000 iterations. (C) Coefficient of variation of the rankings from B. X-axis: genes are ranked by mean expression level for the indicated cell population and placed into bins. Y-axis: Samples equal to the cell population were drawn with replacement from the cluster and mean expression rank was recalculated over 1,000 iterations. The y-axis shows the coefficient of variation of the mean expression ranking for each gene over the 1,000 iterations. (D) Overlap of cell type enriched genes for fetal brain datasets. Sub-clusters were merged into major cell type for the Fluidigm C1 dataset (Nowakowski et al., 2017). Color indicates FDR-corrected $-\log_{10}$ p-value, number indicates \log_2 odds ratio. (E) Alignment of Drop-seq fetal cortex and Fluidigm C1 fetal brain datasets using Seurat CCA (Butler et al., 2018). Clusters are colored on separate panels by type (Drop-seq fetal cortex clusters, Fluidigm interneurons, Fluidigm progenitors, and Fluidigm excitatory neurons

from (Nowakowski et al., 2017)). For the alignment of excitatory neuron clusters, both clusters of visual cortex excitatory and EN-V1-1 and EN-PFC1 from (Nowakowski et al., 2017) group together slightly separate from other excitatory neurons. This was also the case in the original study (Nowakowski et al., 2017), and may be age related.

Supplemental Figure 4

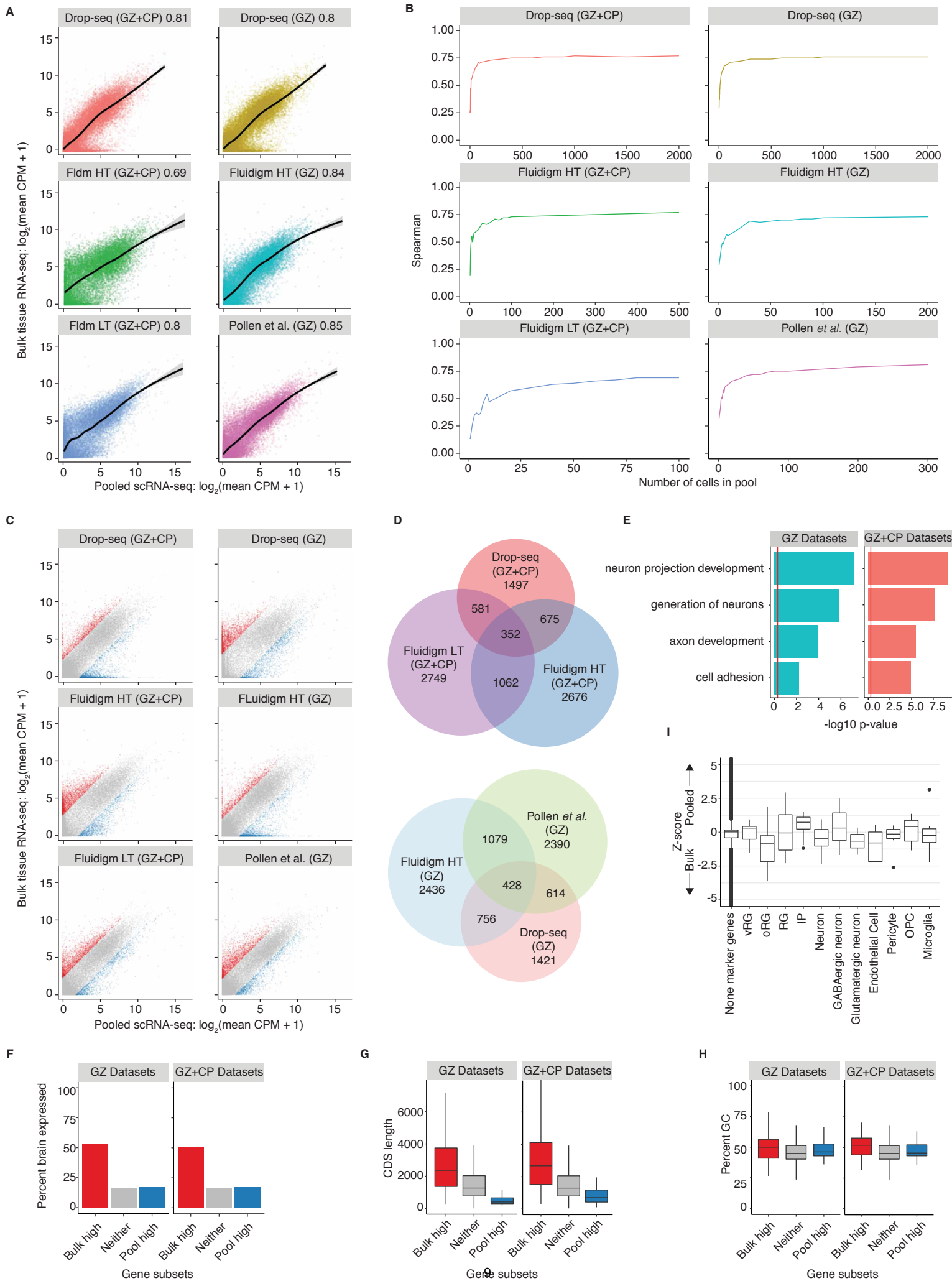
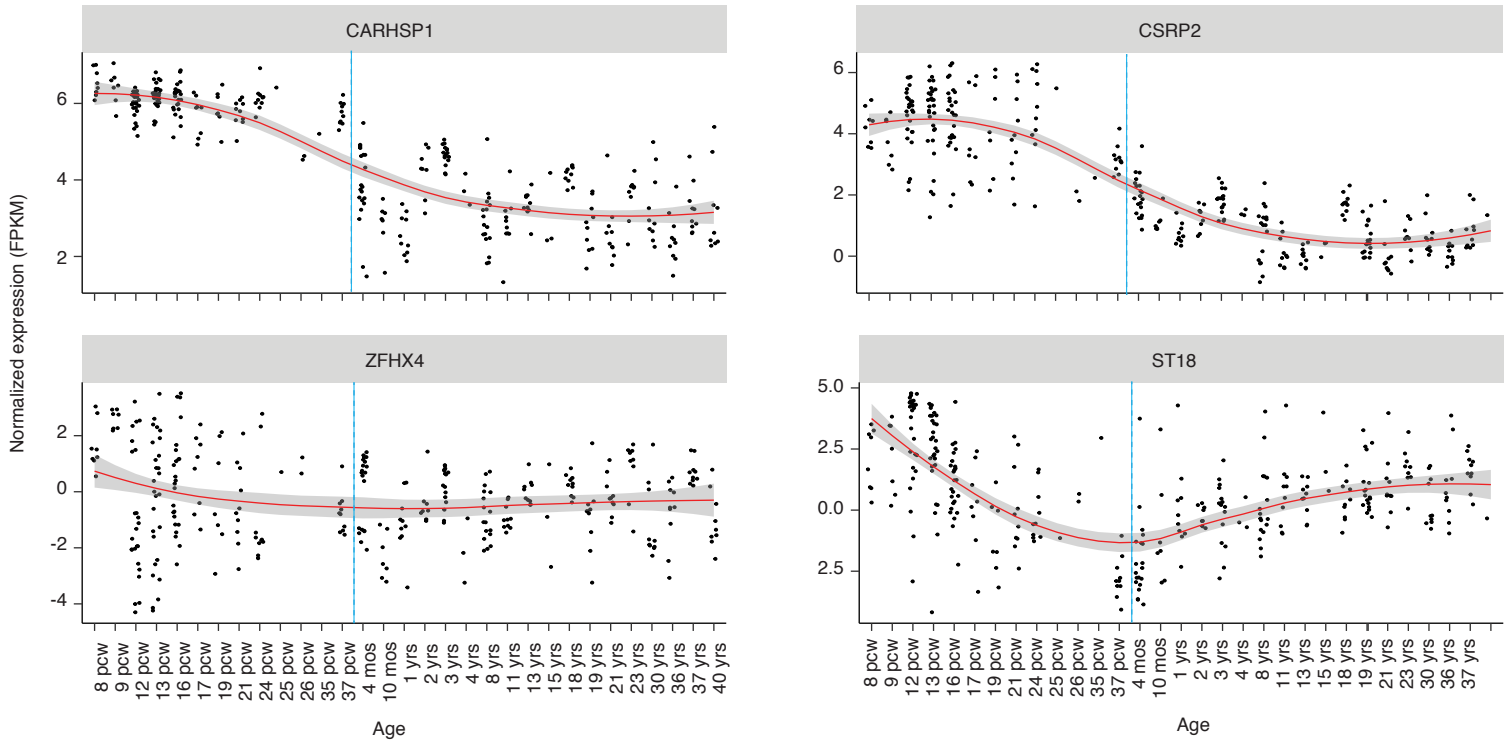


Figure S4 (related to Figure 1). Comparison of scRNA-seq to bulk tissue RNA-seq.

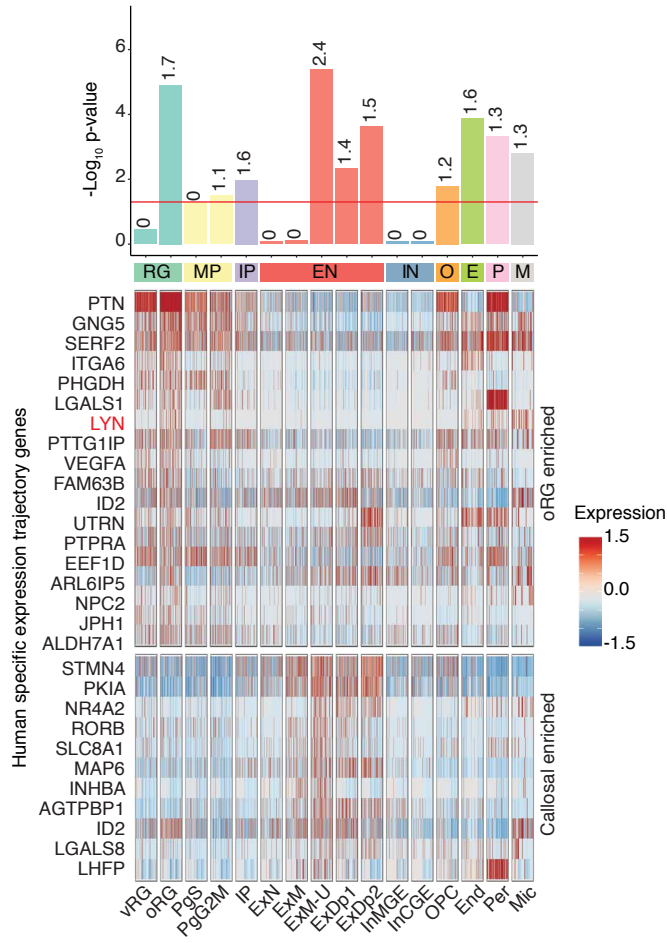
(A) scRNA-seq expression profiles were pooled by summing counts across groups of cells, and pooled expression profiles were compared to the mean expression profile of bulk tissue RNA-seq samples. Each point is a gene, the number in the grey bar is the spearman correlation. The black line is a LOESS curve, and grey band 95% confidence intervals. (B) scRNA-seq expression profiles were pooled by summing counts across groups of cells, and pooled expression profiles of different sizes were compared to the mean expression profile of bulk tissue RNA-seq samples. X-axis: number of cells in pooled scRNA-seq expression profile. Y-axis: correlation of pooled scRNA-seq expression profile to mean bulk tissue expression profile. (C) Identifying genes with higher or lower relative expression in scRNA-seq compared to bulk tissue. The differences between scRNA-seq methods and bulk tissue RNA-seq may be due to mapping differences between poly-A primed scRNA-seq datasets and rRNA depleted full length mRNA bulk tissue RNA-seq, or dissociation procedures common to all scRNA-seq methods. Mean expression for each gene of bulk tissue RNA-seq versus pooled scRNA-seq. scRNA-seq expression profiles were pooled by summing counts across groups of cells. Red: subset of genes with higher relative expression in bulk tissue RNA-seq. Blue: subset of genes with higher relative expression in pooled scRNA-seq. Higher relative expression was defined as two fold greater expression in pooled scRNA-seq versus bulk tissue RNA-seq or vice versa. (D) Intersection of subsets of genes with higher relative expression in bulk tissue RNA-seq compared to scRNA-seq for different datasets (red subset from C). Top: cells from GZ and CP; Bottom: cells from GZ. (E) Representative GO terms from GO analysis of shared genes that have higher relative expression in bulk tissue versus scRNA-seq datasets. Left: Shared genes from GZ samples analysis (352 genes); Right: Shared genes from GZ and CP samples analysis (428 genes). Red line indicates significance (FDR corrected p-value 0.05). (F) Percentage of genes enriched in brain tissue versus other tissues for gene

subsets that are over- or under-represented in scRNA-seq datasets versus bulk tissue RNA-seq. (G) Coding sequence (CDS) length of genes over- or under-represented in scRNA-seq datasets versus bulk tissue RNA-seq. (H) Percent GC content of genes over- or under-represented in scRNA-seq datasets versus bulk tissue RNA-seq. (I) Enrichment or depletion of groups of marker genes for major cortical cell types in Drop-seq relative to bulk tissue RNA-seq. A concern with scRNA-seq methods is bias in cell capture, due to cell size, fragility, adhesion or “stickiness”, or other factors that are intrinsically different between cell types, which will result in biased sampling of the cell population (Gawad et al., 2016). To gauge biases in cell capture with Drop-seq, we compared pooled scRNA-seq expression profiles to bulk tissue, with the reasoning that the relative expression levels of canonical cell type marker genes should be roughly equal if the cell populations captured in Drop-seq were equal to those in bulk tissue RNA-seq. Indeed, comparison of expression of canonical cell type marker genes showed similar expression levels compared to bulk tissue RNA-seq with no significant difference between bulk tissue and pooled scRNA-seq expression profiles for any of the groups of marker genes (Student’s T-test p-value 0.05). Boxplots: box indicates first and third quartiles; the whiskers extend from the box to the highest or lowest value that is within $1.5 * \text{inter-quartile range}$ of the box; and the line is the median.

A



B



C



D

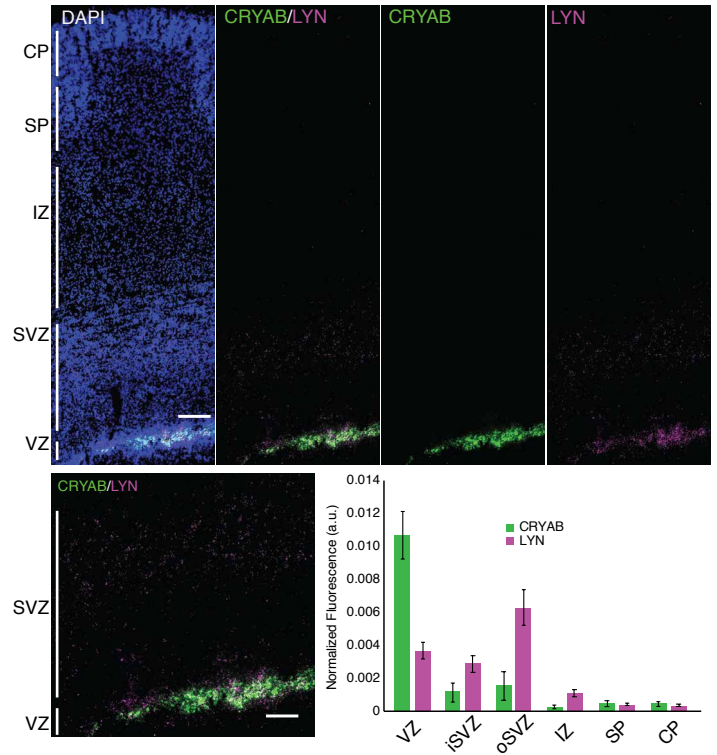
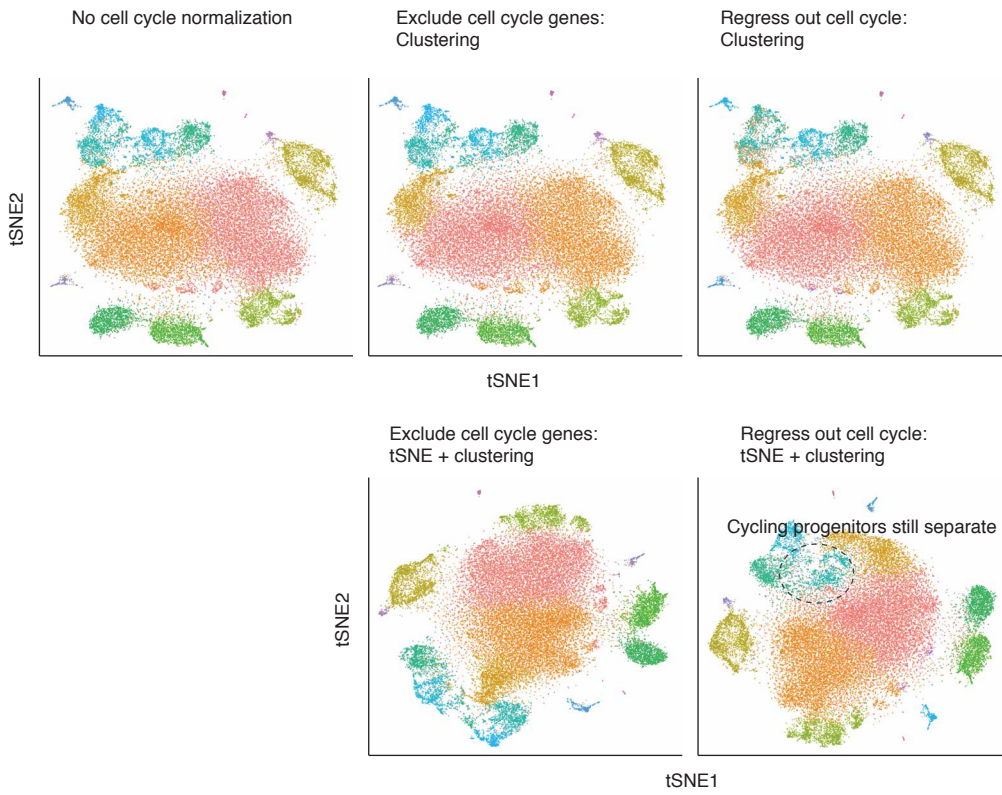


Figure S5 (related to Figure 2). Cell type enrichment of TFs.

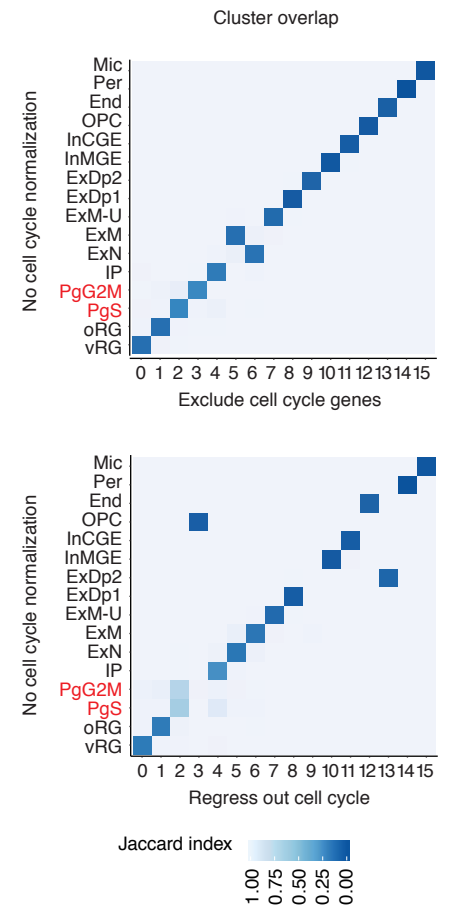
(A) Expression of cell type enriched factors of interest in bulk human cortex across development (www.brainspan.org). PCW = post-conception week, mos = months, yrs = years. Each point represents a sample, the red line a LOESS curve, and grey band 95% confidence intervals. (B to D) Genes with human specific developmental expression patterns. (B) Cell type enrichment of genes with human specific expression trajectories (hSET) in the cerebral cortex (Bakken et al., 2016). Heatmap: oRG (top) and ExM-U (bottom) enriched hSET genes. Red: genes previously unknown to be associated with oRG biology. Red vertical line indicates FDR-significance threshold (p-value 0.05). (C) Expression of oRG genes of interest in bulk tissue LCM laminae from developing cortex (Miller et al., 2014). (D) RNA FISH of fetal cortex probed with the hSET gene LYN showing restricted expression to the VZ and SVZ, and enrichment in the oSVZ. Quantification of normalized fluorescence intensity per layer for each set of probes (see materials and methods). Scale bar = 250 μ m (left) or 100 μ m (inset).

Supplemental Figure 6

A



B



C

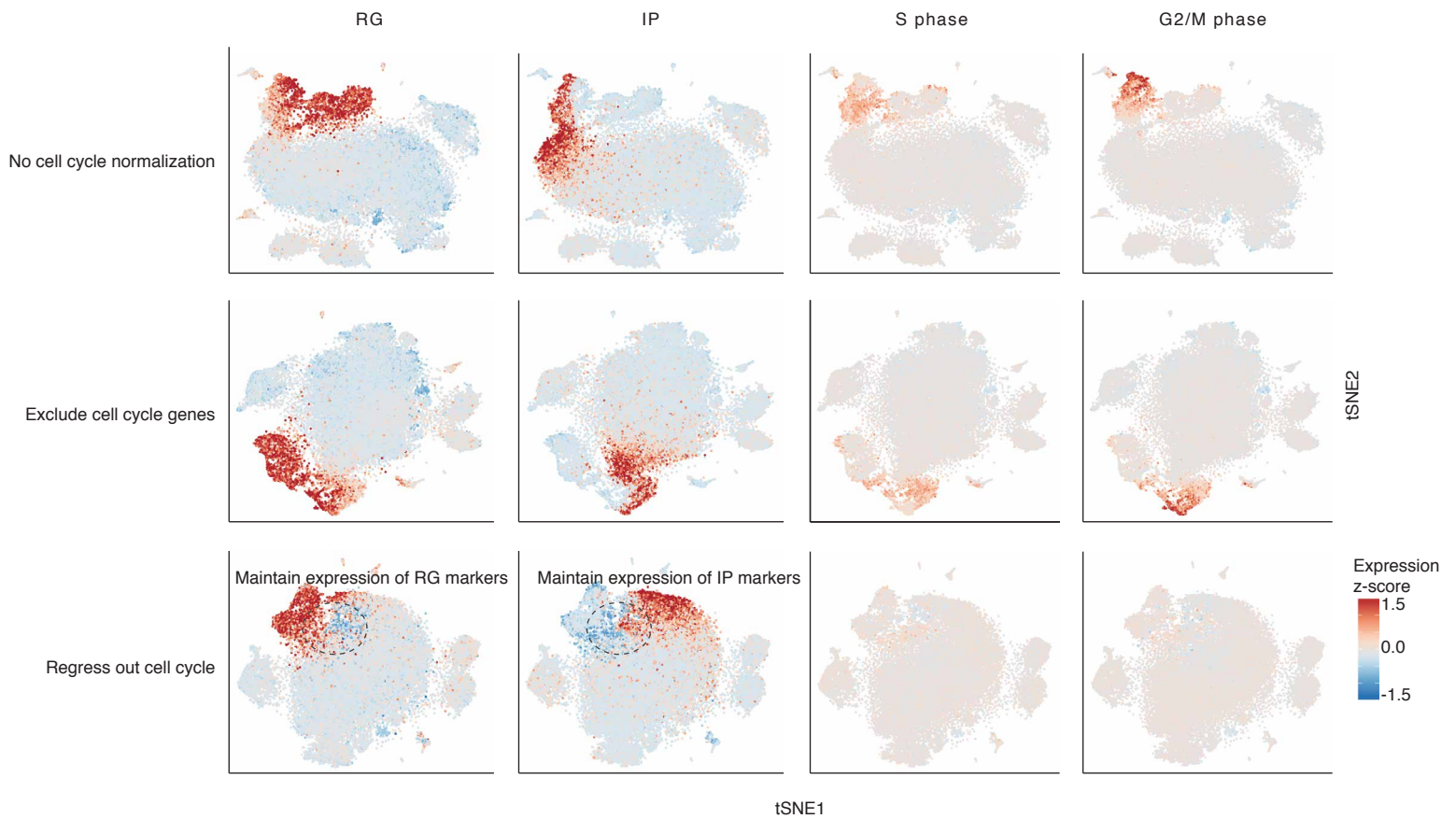
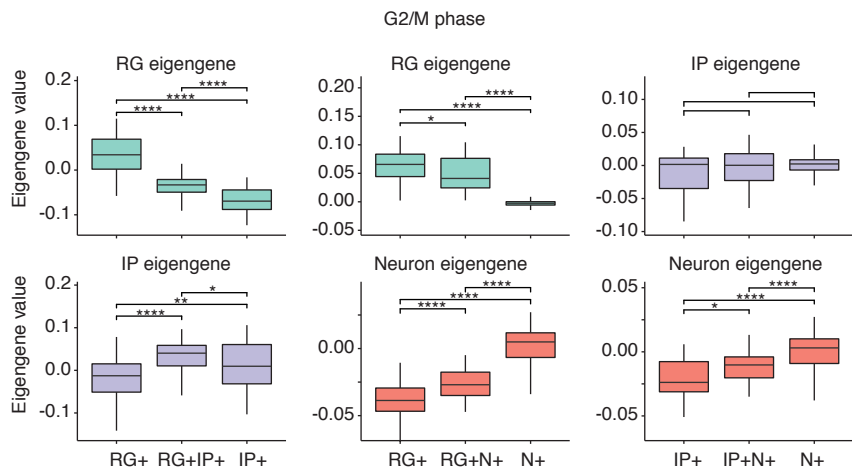


Figure S6 (related to Figure 6). Assessment of cell cycle normalization.

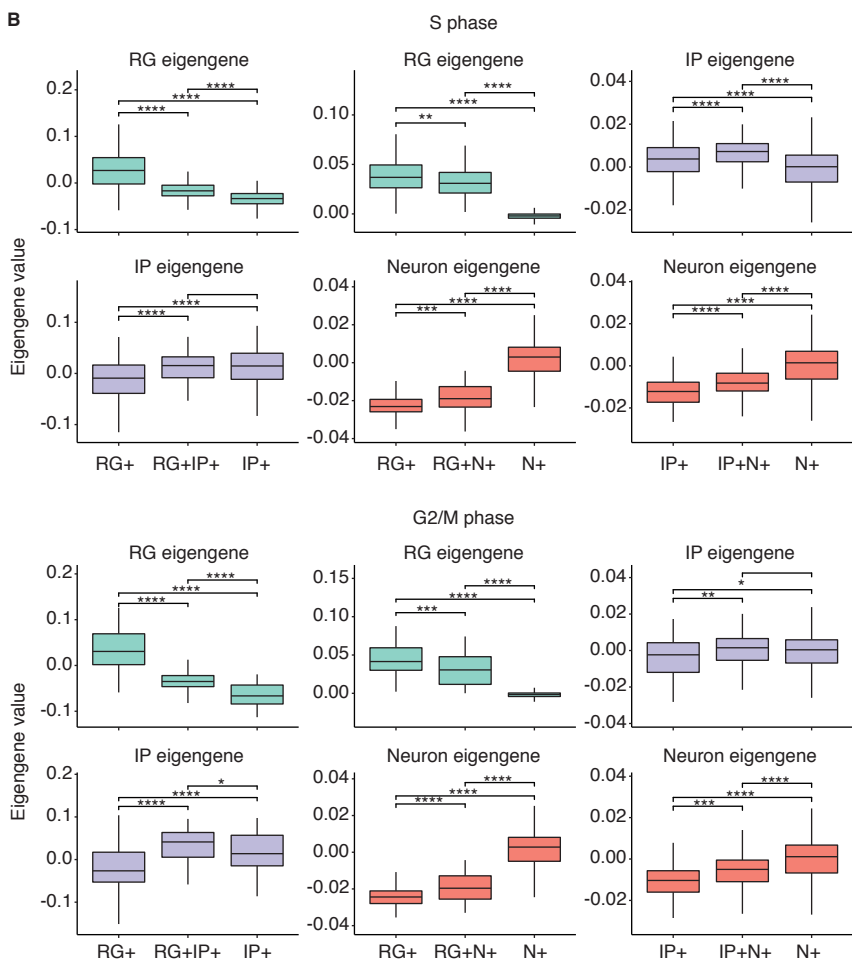
(A) tSNE and Seurat clustering before and after cell cycle normalization. Top row: tSNE with no cell cycle normalization and colored by clustering with different cell cycle normalizations. Bottom row: tSNE and clustering after cell cycle normalization. Normalizations: no cell cycle normalization, removal of cell cycle genes prior to analysis and clustering, or regressing out cell cycle using a score determined by cell cycle gene expression prior to analysis and clustering (see materials and methods). (B) Cluster membership overlaps with and without cell cycle normalization assessed by Jaccard index. Exclusion of cell cycle genes has minimal effect on cluster membership. Regressing out the cell cycle merges S-phase and G2/M phase progenitors (red labels) into one cluster. (C) tSNE before and after cell cycle normalization colored by mean expression of groups of cell type and cell cycle phase marker genes. Regressing out the cell cycle removes enrichment of expression of cell cycle genes in mitotic progenitors, but expression of RG or IP markers is maintained.

Supplemental Figure 7

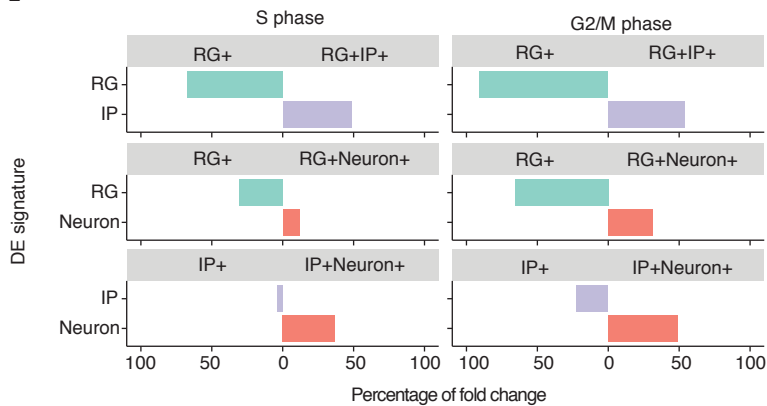
A



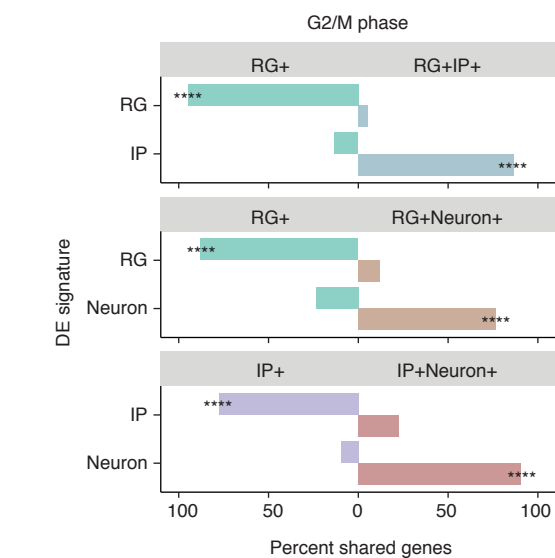
B



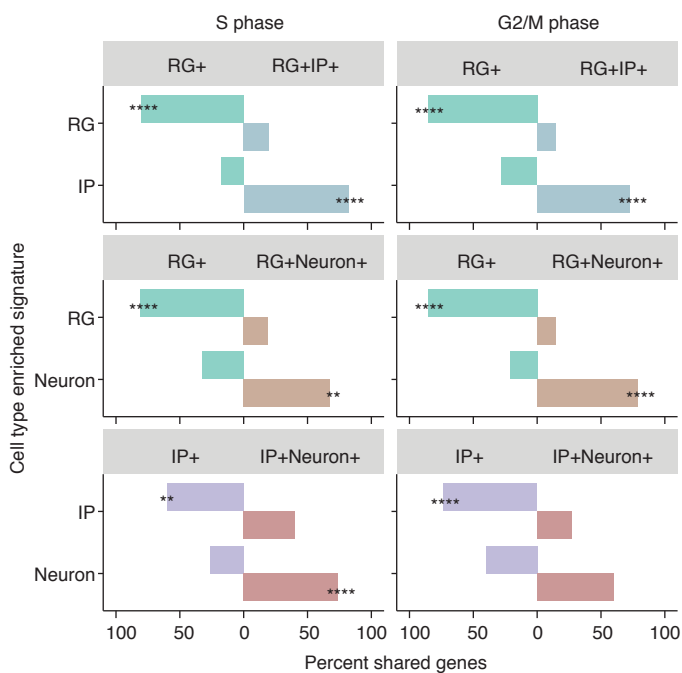
E



C



D



F

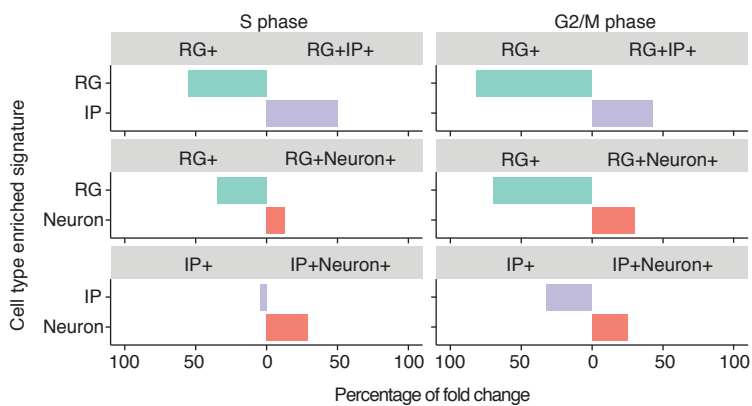


Figure S7 (related to Figure 6). Neurogenesis transition state transcriptomic signatures.

(A) Mixed cell type transcriptomic signatures of cells in G2/M phase corresponding to expression of markers from multiple cell types in the same cell. The eigengene of the indicated cell type gene signature is plotted across groups of cells that express the markers shown on the x-axis and are negative for other cell type markers. Cells were subset to cells from G2/M phase, RG, IP, or newborn neuron clusters. Eigengenes were derived from gene signatures of the indicated cell type (see materials and methods). (B) Mixed cell type transcriptomic signatures of cells in S-phase or G2/M phase corresponding to expression of markers from multiple cell types in the same cell. The eigengene of the indicated cell type gene signature is plotted across groups of cells that express the markers shown on the x-axis and are negative for other cell type markers. Cells were subset to cells from G2/M phase, RG, IP, or newborn neuron clusters. Eigengenes were derived from cell type enriched genes of the indicated cell type (see materials and methods). (C) Overlap of gene signatures from major cell types (y-axis), and genes differentially expressed between major cell types and mixed marker cells (grey bar). X-axis: percentage of genes differentially expressed between major cell types that are also differentially expressed between the corresponding major cell type and mixed marker cells. (D) Overlap of cell type enriched genes from major cell types (y-axis), and genes differentially expressed between major cell type and mixed marker cells (grey bar). X-axis: percentage of genes differentially expressed between major cell types that are also differentially expressed between the corresponding major cell type and mixed marker cells. (E and F) Relative magnitude of differential expression of cell type gene signatures between major cell types and transition state cells (grey bar). For example, RG signature differential expression in RG clusters versus IP clusters compared to RG signature differential expression in RG positive marker cells and RG and IP marker positive S-phase

cells. (A, C, E) Cell type gene signature: genes differentially expressed between major cell types, e.g. RG clusters versus IP clusters. (B, D, F) Cell type gene signature: cell type enriched genes. RG+: cells from vRG, oRG, or mitotic progenitor clusters positive for RG markers; IP+: cells from IP or mitotic progenitor clusters positive for IP markers; Neuron+: cells from newborn neuron cluster positive for neuron markers; RG+IP+, RG+Neuron+, IP+Neuron+: cells positive for both RG and IP markers, RG and Neuron markers, or IP and Neuron markers from S-phase or G2/M phase clusters as indicated. P-values: * <0.05, ** <0.01, *** <0.001, **** <0.0001.

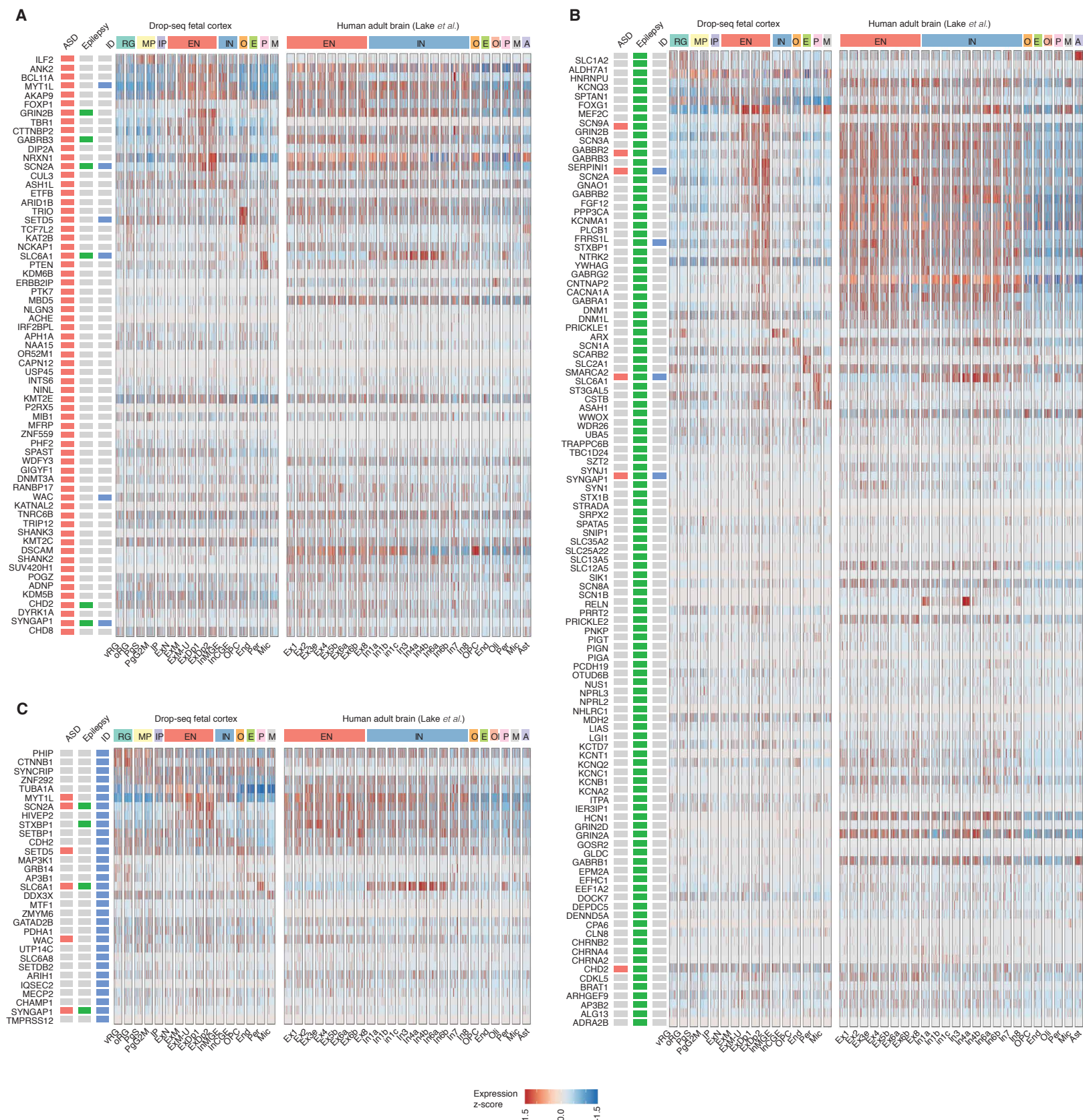


Figure S8 (related to Figure 7). Cell type expression of ASD, epilepsy, and ID risk genes.

(A to C) Cell type expression of ASD (A), epilepsy (B), and ID (C) risk genes in human fetal cortex and adult brain (Lake et al., 2018). Cells are ordered by cluster, clusters are split by columns. Colored boxes on the left indicate membership in ASD, epilepsy, or ID risk gene lists, grey denotes non-membership. RG: radial glia; MP: mitotic progenitor; IP: intermediate progenitor; EN: excitatory neuron; IN: interneuron; O: oligodendrocyte precursor; E: endothelial cell; P: pericyte; M: microglia; Ol: oligodendrocyte; A: astrocyte