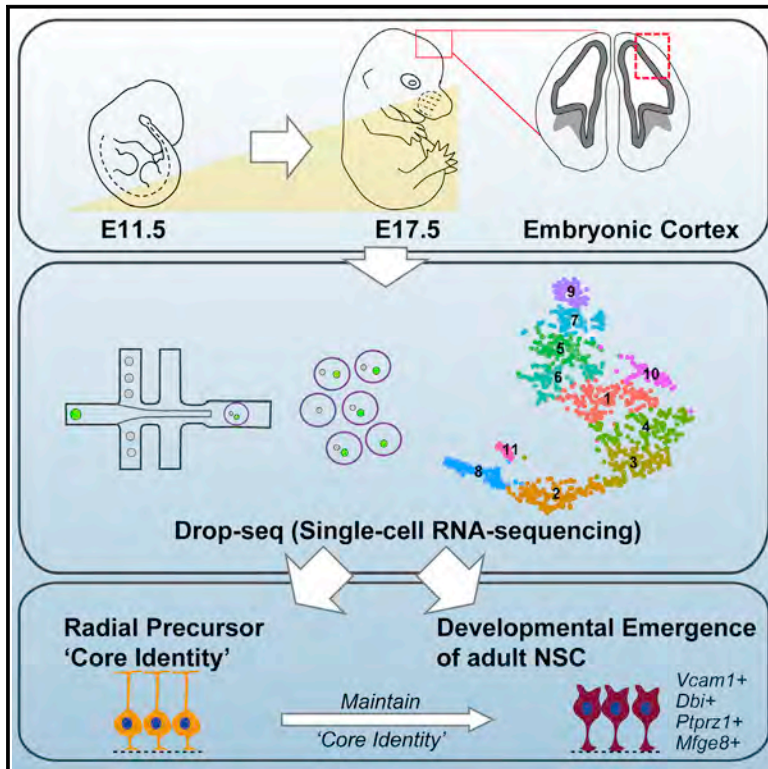# Cell Reports

Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling

Graphical Abstract

Authors

Scott A. Yuzwa, Michael J. Borrett, Brendan T. Innes, ..., David R. Kaplan, Gary D. Bader, Freda D. Miller

Correspondence

fredam@sickkids.ca

In Brief

Yuzwa et al. use single-cell RNA-seq to define the transcriptional identity of precursor cells in the embryonic mouse cortex and, in so doing, characterize the developmental emergence of adult forebrain neural stem cells.

Highlights

- Single-cell RNA-seq identifies transcriptional profiles of cortical precursors

- Radial precursors (RPs) express a core identity reinforced throughout development

- RPs adopt a non-proliferative state during late neurogenesis

- RPs share a transcriptional core identity with adult B1 neural stem cells

Data and Software Availability

GSE107122

CellPress

Developmental Emergence
of Adult Neural Stem Cells as Revealed
by Single-Cell Transcriptional Profiling

Scott A. Yuzwa,[1,7] Michael J. Borrett,[1,3,7] Brendan T. Innes,[4,5] Anastassia Voronova,[1] Troy Ketela,[2] David R. Kaplan,[1,3,5] Gary D. Bader,[4,5] and Freda D. Miller[1,3,5,6,8,*]

[1]Program in Neuroscience and Mental Health, Hospital for Sick Children, Toronto, ON M5G 1L7, Canada
[2]Princess Margaret Genomics Centre, University Health Network, Toronto, ON M5G 2M9, Canada
[3]Institute of Medical Science
[4]The Donnelly Centre
[5]Department of Molecular Genetics
[6]Department of Physiology
University of Toronto, Toronto, ON M5G 1A8, Canada
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: fredam@sickkids.ca
 https://doi.org/10.1016/j.celrep.2017.12.017

SUMMARY

Adult neural stem cells (NSCs) derive from embryonic precursors, but little is known about how or when this occurs. We have addressed this issue using single-cell RNA sequencing at multiple developmental time points to analyze the embryonic murine cortex, one source of adult forebrain NSCs. We computationally identify all major cortical cell types, including the embryonic radial precursors (RPs) that generate adult NSCs. We define the initial emergence of RPs from neuroepithelial stem cells at E11.5. We show that, by E13.5, RPs express a transcriptional identity that is maintained and reinforced throughout their transition to a non-proliferative state between E15.5 and E17.5. These slowly proliferating late embryonic RPs share a core transcriptional phenotype with quiescent adult forebrain NSCs. Together, these findings support a model wherein cortical RPs maintain a core transcriptional identity from embryogenesis through to adulthood and wherein the transition to a quiescent adult NSC occurs during late neurogenesis.

INTRODUCTION

Neural stem cells (NSCs) reside within two well-characterized niches in the adult mammalian brain, the subgranular zone of the hippocampus and the ventricular-subventricular zone (V-SVZ) of the forebrain lateral ventricles. NSCs in both niches generate neurons and, to varying degrees, glial cells, and these progeny are important for a variety of functions, including olfactory and spatial learning and injury responses. Nonetheless, we know little about the developmental genesis of adult NSCs. To address this issue, we have focused on

V-SVZ NSCs, which arise developmentally from embryonic radial glial precursors (RPs) within the ganglionic eminences and, to a lesser degree, the cortex (Young et al., 2007; Gallagher et al., 2013; Kohwi et al., 2007; Willaime-Morawek et al., 2006). In this niche, the quiescent adult NSCs (also called B1 NSCs) retain characteristics of their embryonic RP parents, including expression of markers like Vcam1, Glast, and Sox2 (Anthony et al., 2004; Hu et al., 2017; Kriegstein and Alvarez-Buylla, 2009); extension of primary cilia into the lateral ventricles; and the potential to make neurons, astrocytes, and oligodendrocytes. There are, nonetheless, differences. As one example, embryonic cortical RPs make only excitatory neurons, whereas cortically derived adult V-SVZ NSCs make olfactory bulb interneurons (Fuentealba et al., 2015; Merkle et al., 2007).

Here we have focused on the murine cortex to gain insights into the transition from embryonic RPs to adult NSCs. The embryonic cortex is initially comprised of neuroepithelial stem cells that divide symmetrically to expand their numbers and then transit to RPs around embryonic day 11 (E11) (Haubensak et al., 2004). Cortical RPs then divide asymmetrically from E11 to E17 to generate neurons directly or indirectly via transit-amplifying cells called intermediate progenitor (IP) cells (Englund et al., 2005; Noctor et al., 2004) and subsequently transition to making glial cells postnatally. So when do cortical RPs generate adult V-SVZ NSCs? Recent lineage-tracing studies have shown that adult V-SVZ NSCs emerge from Vcam1-positive RPs during embryonic neurogenesis and that some RPs make both embryonic cortical neurons and adult V-SVZ NSCs (Fuentealba et al., 2015; Furutachi et al., 2015; Hu et al., 2017). Moreover, embryonic perturbations that increase or deplete cortical RPs during neurogenesis cause corresponding alterations in adult V-SVZ NSCs (Gallagher et al., 2013; Hu et al., 2017; Yang et al., 2016). Thus, the genesis of adult NSCs from embryonic cortical RPs is apparently determined during mid- to late neurogenesis. However, we still do not understand what the parental RPs look like at the molecular

level and/or how similar they are to their adult NSC progeny. Here we have addressed these questions using single-cell RNA sequencing (scRNA-seq) to transcriptionally profile embryonic cortical RPs, provide evidence that cortical RPs maintain a core transcriptional identity from embryogenesis to adulthood, and determine that the transition to a quiescent adult NSC occurs during late embryogenesis.

RESULTS

Analysis of the Transcriptome of Individual Cortical Cells Across Embryonic Development

We utilized a droplet-based, high-throughput, scRNA-seq technology, Drop-seq (Macosko et al., 2015), to profile dissected and dissociated E11.5, E13.5, E15.5, and E17.5 murine cortical cells (Figure 1A). We used fluorescence-activated cell sorting (FACS) to remove non-viable, propidium iodide-positive cells, ran the live cells through a Drop-seq apparatus, and then sequenced the resultant single-cell transcriptomes to an average depth of more than 50,000 reads per cell (Figure 1B; for cDNA bioanalyzer traces, see Figure S1A). In total, we collected single-cell transcriptomes from 2,000 cells each at E11.5, E13.5, and E17.5 and from 5,000 cells at E15.5, as confirmed by knee plots of the cumulative fraction of total reads contributed by each cell (Figure S1B). The average gene and transcript numbers per cell were similar at all developmental time points (genes per cell ± SD and transcripts per cell ± S.D: 1,490 ± 840 and 2,750 ± 2,340 (E11.5), 1,270 ± 720 and 2,350 ± 1,780 (E13.5), 1,460 ± 720 and 2,640 ± 2,020 [E15.5], 1,190 ± 511 and 1,940 ± 1,140 [E17.5]) (Figure 1C). As a control for cell sorting, we compared E13.5 cortical cells that were or were not sorted by FACS after dissociation. In both cases, similar numbers of genes and transcripts per cell were detected (Figure S1C), and similar proportions of major cell types were identified (Figure S1D).

We analyzed these single-cell transcriptomes with a modified analysis pipeline we developed that incorporated extensive low-level data quality analysis with visualization and clustering methods using evidence-based parameter selection (see Figure 1D and Experimental Procedures for details). We initially removed predicted cell doublets, predicted blood cells (Lun et al., 2016), and cells with high relative mitochondrial gene transcripts, likely dying cells. This left 1,828 (E11.5), 1,805 (E13.5), 4,515 (E15.5), and 1,824 (E17.5) single-cell transcriptomes (9,972 total). We then removed genes seen in fewer than three cells and normalized the data at each individual time point (see Experimental Procedures). Genes with high variance were used to compute principal components as input to visualize cells in two dimensions using t-distributed stochastic neighbor embedding (t-SNE). To identify cell types, we then used the graph-based clustering method from the Seurat R package with a range of resolution parameters until the number of differentially expressed genes between the two most similar clusters reached a lower limit of 10 genes (p < 0.01 family-wise error rate [FWER]). This iterative approach enabled unbiased selection of resolution parameters for clustering. Subsequent analysis with cell-type-specific markers (below) provided biological validation for this approach.

This analysis resulted in identification of 13 cell type clusters at E11.5, 16 at E13.5, and 15 each at E15.5 and E17.5 (Figures 1E–1H). Because we wanted to analyze only cortically derived cells, we selected clusters that contained cells expressing the cortical transcription factor gene *Emx1*, which is not expressed in ganglionic eminence (GE) cells, meninges, vasculature, or microglia (Gorski et al., 2002; Gulisano et al., 1996; Simeone et al., 1992), and that did not express the GE transcription factor genes *Dlx1*, *Dlx2*, or *Dlx5* (Eisenstat et al., 1999) or the interneuron genes *Gad1* or *Gad2* because cells positive for these genes derive from the GE (Figures 1E–1H; Figure S1E). The *Emx1*-positive clusters and the *Dlx/Gad*-positive clusters were mutually exclusive, validating this filtration step. We also removed the small clusters containing cells expressing *Aif1*, a microglial gene (Imai et al., 1996), or collagen genes characteristic of mesenchymal cells. We then reanalyzed the cells within the cortically derived clusters using the same spectral clustering and t-SNE approach. This resulted in 8 cell clusters at E11.5, 11 at E13.5, 16 at E15.5, and 9 at E17.5 (Figure 2A).

A Continuous Rather than Sharply Demarcated Transcriptional Transition from RPs to Cortical Neurons

To identify cells in the different cortical clusters, we asked about marker genes, including *Sox2* for neural precursors (Hutton and Pevny, 2011), *Pax6* for neuroepithelial stem cells and RPs (Götz et al., 1998), *Eomes/Tbr2* for IPs (Englund et al., 2005), *Tubb3* (βIII-tubulin) for neurons (Menezes and Luskin, 1994), and *Tbr1*, expressed in many early-born cortical neurons (Hevner et al., 2001). We overlaid cells with the mRNAs for these genes, showing relative expression levels as a color gradient, with yellow indicating no detectable mRNA and blue indicating the highest mRNA level detected (Figures 2B–2E). This analysis identified clusters corresponding to apical precursor (AP)/RPs, IPs, and neurons at all time points (Figure 2A; Table S1; details below). AP/RPs decreased from 63% at E11.5 to 9% at E15.5 and E17.5, neurons increased from 16% at E11.5 to 88% at E17.5, whereas IPs decreased from 27% at E13.5 to 3% at E17.5 (Figure 2A; Figure S2E), consistent with known developmental trajectories.

The eight cell clusters at E11.5 were organized into two larger clusters. One of these (comprised of clusters 1, 3, 4, 6, and 8) contained cells expressing *Sox2* (66%) and the AP markers *Pax6*, *Hes1*, and *Hes5* mRNAs (44%, 37%, and 41%, respectively) (Figures 2A and 2B; Figure S2A). Only scattered cells (12%) expressed the RP marker *Slc1a3/Glast* (Figure S2A), suggesting that these were neuroepithelial stem cells just starting to adopt an RP phenotype, consistent with previous work (Anthony et al., 2004; Hartfuss et al., 2001). By contrast, the second large cluster (comprised of clusters 2, 5, and 7) contained cells expressing little or no *Sox2* mRNA (16%) but high levels of the IP marker *Eomes/Tbr2* (56%) and the neuronal genes *Tubb3* (97%) and *Tbr1* (59%) (Figure 2B). Most *Eomes/Tbr2*-positive cells were in clusters 7 and 5, and many of these expressed other IP genes, including *Gadd45g*, *Ngn1*, and *Ngn2* (42%, 29%, and 44%, respectively) (Figure S2A). Most *Tbr1*-positive cells were instead in clusters 2 and 5 (73%), and some cells in these clusters expressed other early-born cortical neuron genes, such as *Foxp2* (19%, Figure S2A) and *Reln* (8%).
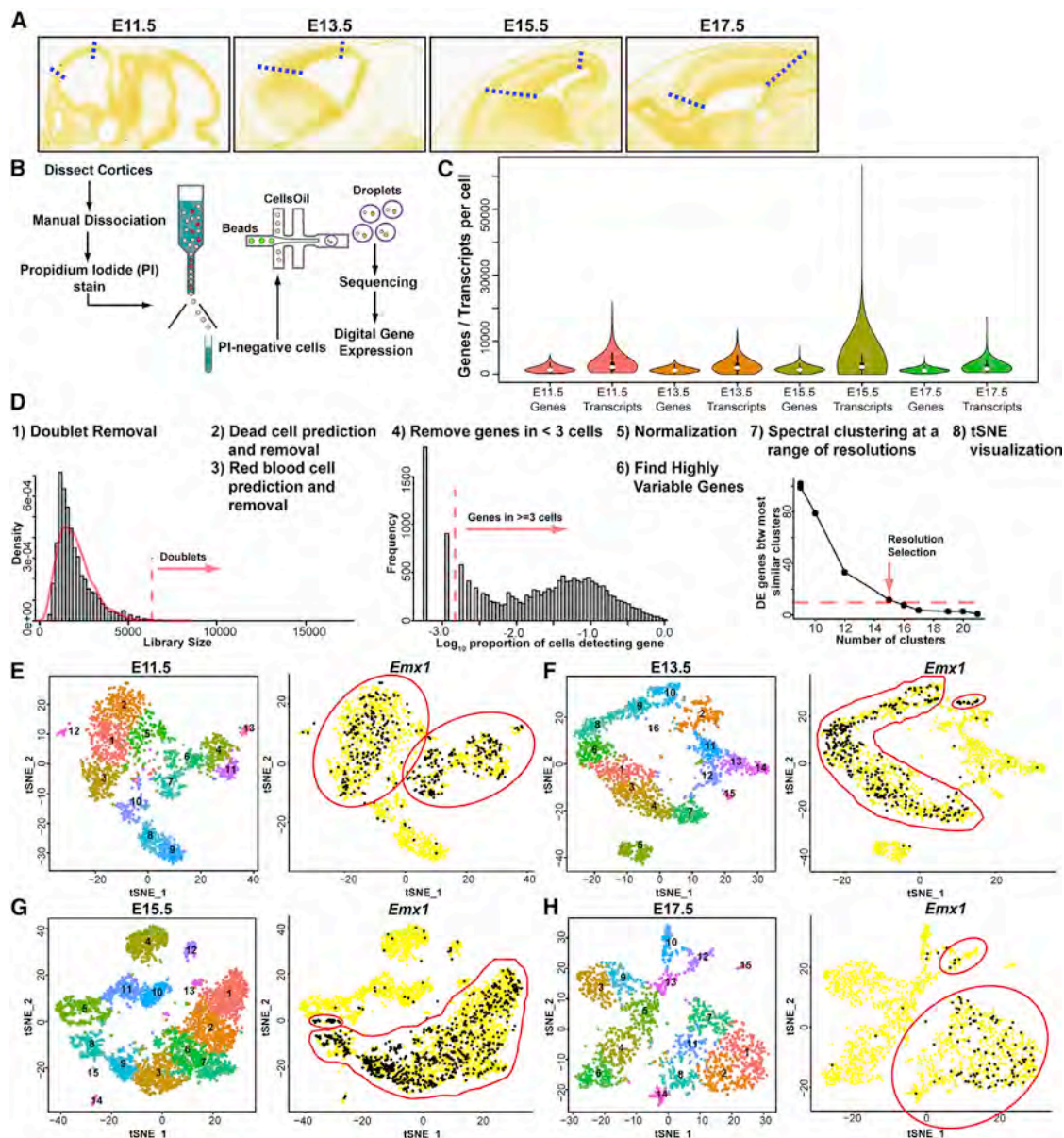
Figure 1. scRNA-Seq Characterization of the Embryonic Cortex across Neurogenesis

(A) Images of sagittal murine brain sections (from the Allen Developing Mouse Brain Atlas), indicating the dissected cortical regions (demarcated by hatched lines).

(B) Schematic of cell preparation and Drop-seq workflow.

(C) Violin plots of the number of genes and transcripts detected per cell at each embryonic age. White dots and black lines indicate median and maximum and minimum values.

(D) Summary of the computational pipeline used to analyze Drop-seq data.

(E–H) t-distributed stochastic neighbor embedding (t-SNE) visualizations of cell clusters identified using the computational pipeline at E11.5 (E), E13.5 (F), E15.5 (G) and E17.5 (H). Left: the distinctly colored numbered cell clusters. Right: *Emx1* mRNA expression overlaid on the single cells (black and yellow, detectable and non-detectable *Emx1* expression, respectively). Positive clusters (circled in red) contained few/no *Dlx*-or *Gad*-positive cells (Figure S1D) and, thus, were defined as cortical.

See also Figure S1.

Intriguingly, this analysis defined clear cell type clusters but also identified considerable transcriptional overlap. For example, many cluster 5 cells expressed *Eomes/Tbr2* and *Tbr1* mRNAs (49% double-positive), whereas some *Foxp2*-positive, *Tbr1*-positive neurons in cluster 2 also expressed *Pax6* mRNA (18%).

By E13.5, there were fewer precursors (21% RPs and 27% IPs) and more neurons (52%) (Figure S2E). *Sox2*, *Pax6*, *Hes1*, *Hes5*, and *Slc1a3/Glast* mRNAs were expressed in many/most cells in clusters 2 and 8 (67%, 51%, 36%, 49%, and 31%, respectively) (Figure 2C; Figure S2B), consistent with their identification as
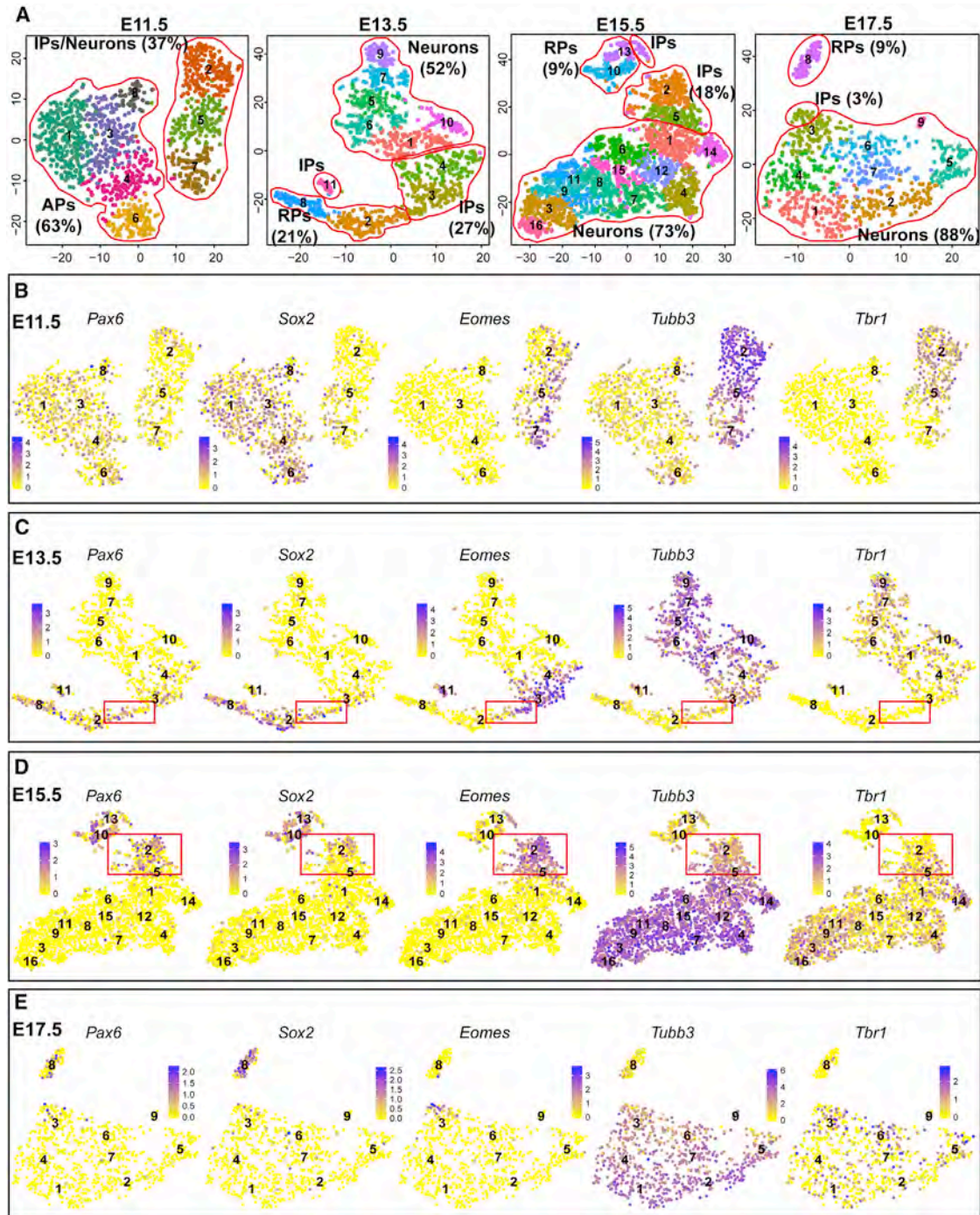
Figure 2. Identification of Cortical Precursors and Neurons within E11.5–E17.5 Cortical Drop-seq Data

(A) t-SNE visualizations of cortically derived cell clusters at each embryonic age. Colors and associated numbers represent individual clusters. Circled in red are clusters containing APs, RPs, IPs, and neurons. Also shown are the percentages of each cell type.

(B–E) t-SNE visualization of E11.5 (B), E13.5 (C), E15.5 (D), and E17.5 (E) scRNA-seq data overlaid with expression of the cell-type-specific markers *Pax6*, *Sox2*, *Eomes*, *Tubb3*, and *Tbr1*. Cells are color-coded according to expression level, ranging from not detected (yellow) to the highest detected levels (blue), according to the adjacent color key. Cluster numbers are as shown in (A). The boxed regions in (C) and (D) are shown at higher resolution in Figures 3A and 3B. See also Figure S2.
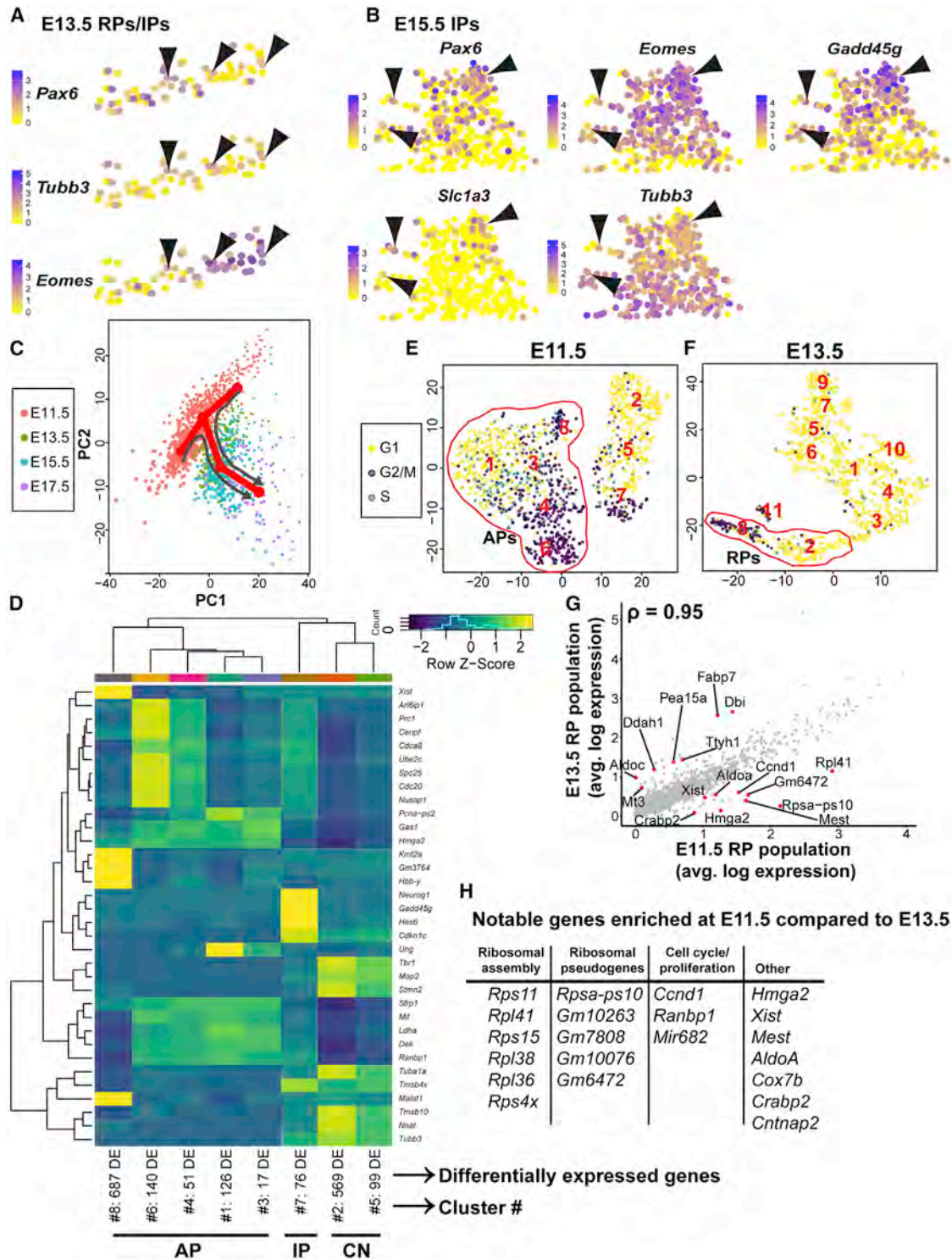
Figure 3. Characterization of a Developmental Transition in AP Cells between E11.5 and E13.5

(A and B) Higher-resolution t-SNE visualizations of E13.5 RPs/IPs (A) or E15.5 IPs (B) from the boxed regions in Figures 2C and 2D, respectively, showing overlaid expression of *Pax6*, *Eomes*/*Tbr2*, and *Tubb3* (A) or *Pax6*, *Slc1a3*/*Glast*, *Eomes*/*Tbr2*, *Gadd45g*, and *Tubb3* (B). Gene expression levels are color-coded from not detected (yellow) to the highest detectable expression levels (blue). Arrowheads indicate cells that express all three (A) or five (B) mRNAs.

(C) PCA of combined AP/RPs from E11.5, E13.5, E15.5, and E17.5 scRNA-seq data. Cells of different embryonic ages are color-coded. Red lines show *k*-means cluster centers and the minimum spanning tree connecting cluster centers computed using the Waterfall method (Shin et al., 2015). One path begins

(legend continued on next page)

RPs. The IP gene *Eomes/Tbr2* was instead expressed in 78% of cells in clusters 3, 4, and 11 (Figure 2C). These cells also expressed the IP genes *Gadd45g*, *Ngn1*, and *Ngn2* mRNAs (57%, 13%, and 62%) (Figure S2B). Finally, the neuronal genes *Tbr1* and *Tubb3* were expressed in most cells in clusters 1, 5, 6, 7, 9, and 10 (58% and 97%, respectively) and in many cells in cluster 4. These same clusters contained cells expressing the cortical neuron transcription factor genes *Satb2* (19%) and *Bhlhe22* (31%) (Figure S2B). However, the expression of cell type identity genes was not sharply demarcated in individual cells. This point is best exemplified by considering the IPs. 14% of cells in clusters 3 and 4 co-expressed *Pax6*, *Eomes/Tbr2*, and *Tubb3* mRNAs, markers of RPs, IPs, and neurons, respectively (Figures 2C and 3A), and 17% of cluster 11 cells co-expressed *Pax6*, *Eomes/Tbr2*, *Tubb3*, and *Tbr1* mRNAs (Figure 2C).

The E15.5 data were similar to E13.5, although RPs and IPs were further decreased to 9% and 18%, and neurons proportionately increased to 73% (Figure 2A; Figure S2E). RPs were present in clusters 10 and 13, with many/most cells expressing *Sox2*, *Pax6*, *Hes1*, *Hes5*, and *Slc1a3/Glast* mRNAs (66%, 52%, 35%, 32%, and 45%, respectively) (Figure 2D; Figure S2C). The *Eomes/Tbr2*-positive IPs were largely in clusters 2 and 5 (80%), and some of these expressed *Ngn1* (8%) or *Ngn2* (56%) mRNAs (Figure 2D; Figure S2C). The other clusters contained neurons, with almost all cells (99%) expressing *Tubb3* mRNA and 50% *Tbr1* mRNA. Subsets of these neurons expressed the neuronal transcription factor genes *Satb2* (27%) and *Bhlhe22* (38%). In addition, the two neuronal clusters closest to the IPs, clusters 1 and 14, contained cells expressing the neuronal migration genes *Sema6d* (18%) and *Unc5d* (34%) (Figure S2C), suggesting that they were just-born migratory neurons. As seen at earlier ages, there was transcriptional overlap that was most evident in the precursor cells; many *Eomes/Tbr2*-positive, *Gadd45g*-positive IPs in cluster 2 co-expressed *Pax6* (28%), *Slc1a3/Glast* (9%), or *Tubb3* (45%) mRNAs, and 3%–4% of cells expressed all 5 mRNAs (Figures 2D and 3B).

By E17.5, cluster 8 contained the remaining RPs (9% of total cells) that expressed *Sox2*, *Pax6*, *Hes1*, *Hes5*, and *Slc1a3/Glast* mRNAs (73%, 25%, 31%, 58%, and 77%, respectively) (Figure 2E; Figure S2D). A few cells (3%) at the tip of cluster 3 were *Eomes/Tbr2*-positive IPs, but the remaining cells (88%) were various types of cortical neurons expressing genes like *Tubb3*, *Tbr1*, *Satb2*, and *Bhlhe22* (97%, 28%, 47%, and 49% of the neurons, respectively) (Figure 2E; Figure S2D).

These data validated the cell analysis pipeline by clearly distinguishing the predicted cell types and also identified a surprising degree of transcriptional overlap, with many cells co-expressing genes for multiple cell types.

A Developmental Transition in Apical Cortical Precursors between E11.5 and E13.5

RPs sequentially generate deep-layer neurons, superficial-layer neurons, and then glial cells. To ask whether RPs varied over time to attain this timed cell genesis, we performed a developmental trajectory analysis. We combined transcriptomes from E11.5–E17.5 AP/RP cell clusters, normalized them together, and used the Waterfall method to perform principal-component analysis (PCA), *k*-means clustering, and construction of a minimum spanning tree through the *k*-means cluster centers (Shin et al., 2015). This analysis (Figure 3C) demonstrated a single linear trajectory from E13.5 to E17.5, with RPs ordered in the first two dimensions of PCA space largely accordingly to embryonic age, with some overlap that might reflect spatial heterogeneity in cortical development (O'Leary and Sahara, 2008). Thus, RPs apparently developed in a homogeneous way rather than branching into differentially biased RP subpopulations. By contrast, the E11.5 precursors occupied a more diffuse PCA space with more *k*-means clusters. Although this might be partially due to some heterogeneity in the number of transcripts detected per cell (see point sizes in Figure 3C), it might also reflect cellular heterogeneity as cells transition from neuroepithelial stem cells to RPs around E11 (Anthony et al., 2004; Haubensak et al., 2004).

To better characterize the E11.5 cells, we performed differential gene expression, comparing each individual cluster with the average of all other clusters (Figure 3D). Hierarchical clustering analysis confirmed that APs in clusters 1, 3, 4, 6, and 8 (Figures 2A and 2B) were very similar to each other and distinct from other cells. The top five differentially expressed genes in these clusters were almost all associated with cell cycle stage. For example, clusters 1 and 3 were enriched for the G1/S phase mRNAs *Pcna* and *Ung*, with 57% and 33% positive cells, as identified by t-SNE overlays. Clusters 4 and 6 were differentially enriched for the mitosis-associated genes *Prc1*, *Cenpf*, *Ube2c*, *Cdc20*, *Nusap1*, and *Tpx2* (74%, 80%, 88%, 69%, 74%, and 80% positive cells). No further differences were found between APs and other cells when we regressed the effects of the cell cycle and re-clustered the E11.5 data (Figures S3A–S3D).

Additional analysis using Cyclone (Scialdone et al., 2015) to compute cell cycle scores (Figure 3E) confirmed the differential gene expression findings and predicted that 62% of E11.5 APs were in G2/M or S phases, with cells in clusters 4, 6, and 8 in G2/M and some cells in clusters 1 and 3 in S phase (Figure 3E). By comparison, at E13.5, only 36% of RPs in clusters 2 and 8

near (−11, −2) and a second near (11, 12), and both end near (20, −10). Similar results were obtained with both. Point size is proportional to transcript counts per cell.

(D) Heatmap of the top five differentially upregulated genes (p < 0.01, FWER) per E11.5 cluster relative to all other clusters. Gene expression is represented by *Z* scores, and dendrograms represent hierarchical clustering of cell clusters (columns) and genes (rows). Below each column are cluster numbers (from Figure 2A), total differentially expressed (DE) gene numbers, and cell types. CN, cortical neuron.

(E and F) Cyclone analysis of E11.5 (E) and E13.5 (F) scRNA-seq data, visualized by t-SNE. Cells are color-coded for predicted cell cycle phases (yellow, G1; light blue, S; purple, G2/M). AP and RP clusters are outlined in red, and cluster numbers are shown.

(G and H) Pearson correlation analysis of the average expression of each gene detected by scRNA-seq in E11.5 AP and E13.5 RP clusters. The red dots in (G) and the list in (H) denote genes with low correlations/high average differences between these two populations.

See also Figure S3.

were predicted to be in G2/M and S phase (Figure 3F). These numbers coincide well with cell cycle analyses performed using *in vivo* nucleotide labeling (Takahashi et al., 1995).

The Cyclone analyses identified cell cycle status as one major difference between E11.5 and E13.5 precursors. We defined additional differences using correlation analysis of average gene expression values to compare the *Pax6*-positive, *Sox2*-positive precursor clusters at these two ages (Figure 3G). Although the E11.5 and E13.5 AP/RP populations were very similar (rho = 0.95), some genes showed low correlation/high differential expression. Genes expressed at higher average levels at E11.5 were associated with ribosomal biogenesis and cell cycle control (Figures 3G and 3H). By contrast, E13.5 precursors expressed higher levels of genes associated with an RP phenotype, *Fabp7/Blbp* and *Aldoc* (Figure 3G) and *Dbi*, *Ddah1*, *Pea15a*, *Ttyh1*, and *Mt3*, which subsequent analyses showed were also RP markers (see below). Thus, between E11.5 and E13.5, APs decreased their proliferation and adopted an RP transcriptional identity.

Characterization of RPs and IPs from E13.5 to E17.5
The trajectory analysis (Figure 3C) suggested that a relatively homogeneous RP population progressed transcriptionally from E13.5 to E17.5. To define this progression, we initially performed differential gene expression analysis combined with t-SNE overlays. This approach showed that, at E13.5 (Figures 4A and 4B), the two RP clusters (2 and 8) differed predominantly in cell cycle status, with cluster 8 enriched in the proliferation genes *Nusap1*, *Top2a*, *Cdca8*, and *Mki67* (60%–77% positive cells), consistent with the Cyclone analysis (Figure 3F).

A similar cell cycle-dependent segregation was seen for the *Eomes/Tbr2*-positive IP clusters (3, 4, and 11), with cluster 11 enriched for *Nusap1* and *Cdca8* (83% and 72% cells), consistent with the Cyclone analysis (Figure 3F). By contrast, the non-proliferative clusters 3 and 4 defined two distinct types of IPs (Figure 4B). Cluster 3 was enriched for *Ngn2*, *Gadd45g*, and *Btg2/Tis21* (86%, 78%, and 56% positive cells), whereas cluster 4 was enriched for *Neurod1*, *Pam*, and *Slc17a6* (64%, 60%, and 44% positive cells). These likely correspond to previously identified apical and basal IPs, respectively (Kawaguchi et al., 2008). This analysis also identified *Sstr2* (somatostatin receptor 2) as a potential IP marker because it was enriched in *Eomes/Tbr2*-positive IPs at all ages (Figure S3E).

Results at E15.5 (Figures 4A and 4C) were similar to E13.5. One RP cluster (13), but not the other (cluster 10), was proliferative, as indicated by Cyclone analysis (Figure 6D) and enrichment for *Mki67*, *Top2a*, and *Nusap1* (68%–78% positive cells) (Figure 4C). Both RP clusters were enriched relative to other clusters for the RP markers *Fabp7/Blbp*, *Vim/Vimentin*, and *Aldoc* (90%, 86%, and 65% positive cells, respectively), as well as for *Dbi*, a marker for adult V-SVZ neural precursors (96% cells) (Alfonso et al., 2012), and for *Ttyh1* (69%). Of the two IP clusters, cluster 2 was enriched in the apical IP markers *Ngn2* and *Gadd45g* (>70% positive cells), and cluster 5 was enriched in the basal IP marker *Neurod1* (79%) (Figure 4C). Cyclone analysis indicated that both clusters were likely in G1 but identified a small population of proliferating IPs expressing *Eomes/Tbr2*, *Gadd45g* (75%), and *Ngn2* (55%) as a side branch of RP cluster 13 (Fig-

ure 4A). All IP clusters were enriched for *Sstr2* and for another potential IP marker defined by the differential gene expression analysis, *Mfap4*, with 56% of all IPs positive for these genes (Figure 4C; Figures S3E and S3F).

At E17.5 (Figures 4A and 4D), RP cluster 8 was enriched for *Slc1a3/Glast*, *Fabp7/Blbp*, and *Dbi* (77%, 97%, and 95% positive cells, respectively) and for another potential RP marker, *Ednrb* (endothelin receptor B) (71%). The few remaining *Eomes/Tbr2*-positive IPs at the tip of cluster 3 expressed *Mfap4* (42%) and *Sstr2* (58%) (Figures S3E and S3F).

The remaining clusters at all ages were comprised of neurons (Figures 4A–4D) appropriately enriched for cytoskeleton-associated genes like *Stmn2* and *Stmn4* (Stathmins 2 and 4), *Gap43*, *Mapt* (Tau), *Tuba1* and *Ina*, and other neuronal genes like *Gria2* (an α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid [AMPA] receptor subunit) and *Syt4* (Synaptotagmin 4). Some clusters were also enriched for genes associated with different cortical neuron subpopulations like *Fezf2*, *Satb2*, *Sox5*, and *Mef2c*.

This analysis distinguished proliferative versus non-proliferative RPs and IPs, defined two transcriptionally distinct IP populations, and identified potential markers for RPs and IPs.

Identification of an Embryonic RP Core Transcriptional Identity
We next asked how RPs progressed transcriptionally over time by comparing the genes that were significantly differentially enriched (p < 0.01, FWER) in RPs relative to all other clusters at E13.5, E15.5, and E17.5 (Figure 5A; Table S2). Of 739 mRNAs in total, 98 were enriched at all three ages (Figure 5A; Table S3). Overlays showed that 65 of these were highly enriched in RPs throughout embryogenesis (examples are shown in Figure 5B and Figure S4A), with some expressed at lower levels in IPs and scattered neurons (summarized in Table S3). The remaining 33 genes were expressed at high relative levels in RPs but also in many/most other cortical cells.

These 65 mRNAs fell into four different groups. One group included known markers of cortical RPs, validating our analysis. This group included *Vcam1*, *Fabp7/Blbp*, *Pax6*, *Slc1a3/Glast*, *Sox9*, *Sox2*, *Vim*, *Tlx/Nr2e1*, *Mdk*, *Aldoc*, *Hopx*, *Hes1*, *Hes5*, *Sox21*, *Id4*, *Nde1*, and *Nes* (Nestin) mRNAs. A second group included genes known to be enriched in postnatal or adult neural precursors but not well-characterized in embryonic cortical RPs, including *Lfng*, *Hmgb2*, *Pbk*, and *Dbi*. A third group consisted of proliferation genes like *Rgcc*, *Mcm3*, *Ccnd1*, *Cdca8*, *Cdk1*, *Gmnn*, *Cenpq*, and *Nek6*. The final and largest group included less well-characterized genes within the neural precursor context, including *Ednrb*, *Mfge8*, *Ttyh1*, *Metrn*, *Pdpn*, *Pon2*, *Sparc*, *Pea15a*, *Ddah1*, *Sfrp1*, *Ckb*, *Gas1*, *Mt1*, *Mt2*, *Mt3*, *Zfp36l1*, *Ppap2b*, *Gsta4*, *Psat1*, *Phgdh*, *Tead2*, *Efhd2*, *Gpx8*, *Rcn1*, *Kbtbd11*, *Fgfbp3*, *Gm11627*, *Serpinh1*, *Cyr61*, *Cd63*, *Myo10*, *2810417H13Rik*, *Acadl*, *Oat*, *Magt1*, and *Asrgl1* (examples are shown in Figure 5B and Figure S4A).

A similar overlay analysis for 47 mRNAs enriched only at E15.5 and E17.5 (Figure 5A) identified 25 additional mRNAs that were highly enriched in RPs (examples are shown in Figure S4B). Many were highly specific for RPs, but some were also expressed in IPs and/or at low levels in neurons (summary in
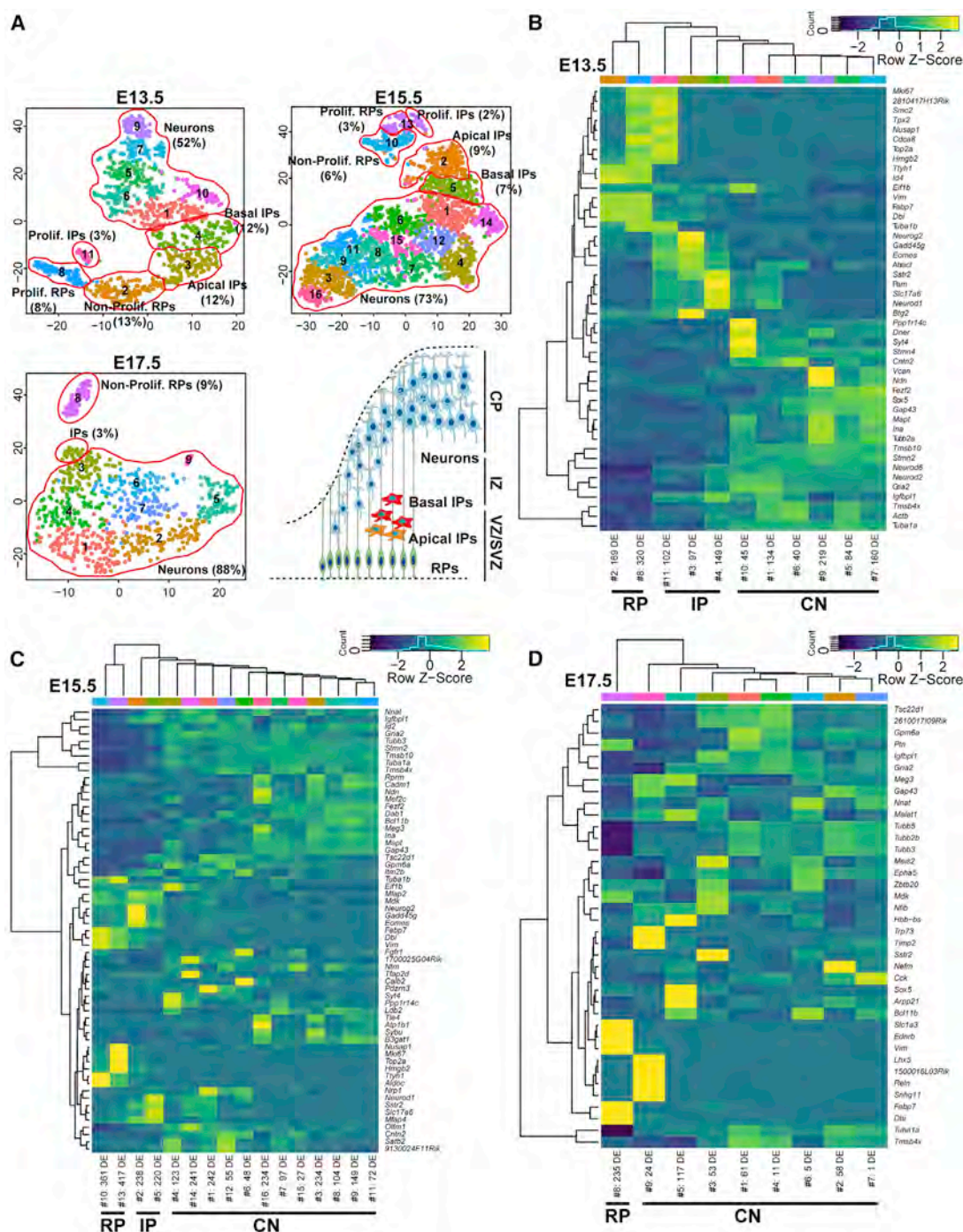
Figure 4. Characterization of Transcriptional Profiles of Cortical Precursors from E13.5 to E17.5

(A) t-SNE visualizations of cortically derived E13.5–E17.5 cell clusters, denoting identified cortical cell types (outlined in red). Colors and associated numbers represent individual clusters. The percentages of each cell type are also shown. Bottom right: a schematic of the embryonic cortex.

(B–D) Heatmaps of the top five differentially upregulated genes (p < 0.01, FWER) relative to all other clusters at E13.5 (B), E15.5 (C), and E17.5 (D). Gene expression is represented by *Z* scores, and dendrograms represent hierarchical clustering of cell clusters (columns) and genes (rows). Below each column are cluster numbers (as in A), total DE gene numbers, and cell types. CN, cortical neuron.

See also Figure S4.
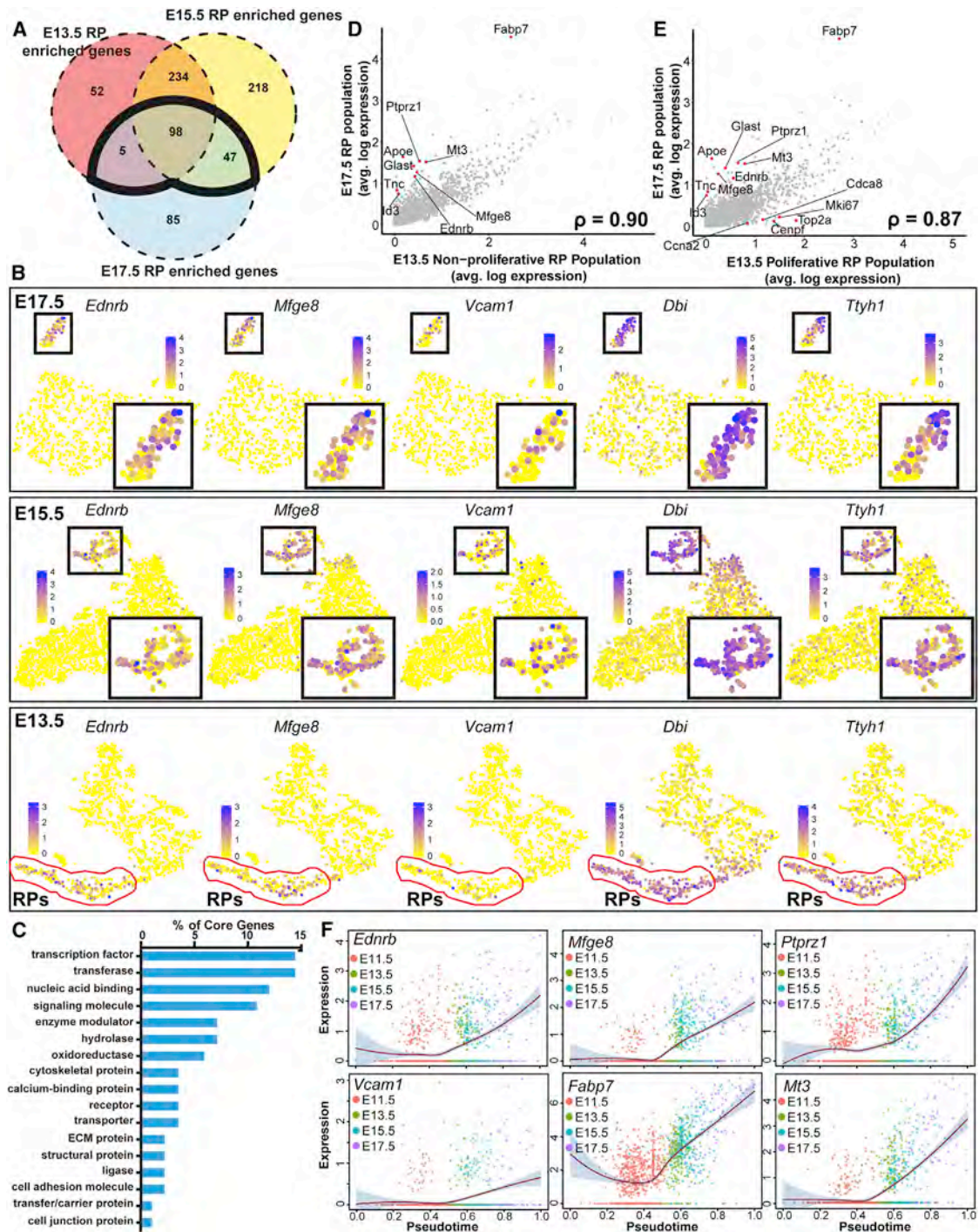
Figure 5. RPs Share a Core Transcriptional Identity throughout Embryogenesis

(A) Venn diagram showing the overlap in genes significantly upregulated in RP clusters relative to all other cell clusters (p < 0.01 FWER) at E13.5, E15.5, and E17.5.

(B) t-SNE visualizations of E13.5, E15.5, and E17.5 scRNA-seq data overlaid with expression of *Ednrb*, *Mfge8*, *Vcam1*, *Dbi*, and *Ttyh1* mRNAs. Cells are color-coded according to expression levels, ranging from not detected (yellow) to the highest detected levels (blue), according to the adjacent color key. Boxed regions highlight RP clusters and are shown at higher resolution at the bottom right. For E13.5, RP clusters are outlined in red.

(C) Functional classification of the 90 core identity genes using PANTHER.

(D and E) Pearson correlation analysis of the average expression of each gene detected by scRNA-seq in E17.5 RPs versus E13.5 non-proliferative (D) (cluster 2 in Figure 4A) or proliferative (E) (cluster 8 in Figure 4A) RP clusters. Red dots denote genes with low correlations/high average differences.

(legend continued on next page)

Table S3). Some of these—*Tnc*, *Id3*, *Id1*, *Tfap2c* and *Ptprz1*—are known to be enriched in cortical RPs, and two, *Apoe* and *Aldh1l1*, are known adult neural precursor markers. However, most have no known association with neural precursors, including *Ccdc80*, *Slc9a3r1*, *Bcan*, *Vit*, *Acss1*, *Mlc1*, *Acsbg1*, *Gng12*, *Atp1a2*, *Clu*, *Tgfb2*, *Acaa2*, *Pnp*, *Nrarp*, *Rhoc*, *Veph1*, *Nim1*, and *Rcn3*.

These analyses identified 90 genes highly enriched in and/or specific to RPs that we have termed the embryonic RP core identity genes. Protein annotation through evolutionary relationship (PANTHER) classification identified transcription factors as one major protein class encoded by these genes (Figure 5C), including ten transcription factors (*Hopx*, *Hes5*, *Sox2*, *Hes1*, *Sox21*, *Sox9*, *Tfap2c*, *Nr2e1*, *Tead2*, and *Pax6*) and four transcriptional regulators (*Id3*, *Id1*, *Hmgb2*, and *Id4*).

The RP Transcriptional Identity Is Reinforced throughout Neurogenesis

These data identified a stable, core transcriptional identity for embryonic RPs but also showed that some core genes were enriched in RPs at E15.5 and E17.5 but not at E13.5. To find out whether the RP core transcriptional identity was upregulated over developmental time, we compared E17.5 RPs with E13.5 non-proliferative and proliferative RPs (clusters 2 and 8, respectively; Figure 4A). Correlation analysis (Figure 5D) demonstrated that, although E17.5 and E13.5 non-proliferative RPs were very similar (rho = 0.90), there was increased E17.5 expression of the core RP genes *Fabp7*, *Apoe*, *Slc1a3/Glast*, *Mfge8*, *Ednrb*, *Ptprz1*, *Mt3*, *Tnc*, and *Id3*. We therefore directly compared these two populations for expression of all 90 core RP genes. 32 core RP mRNAs were upregulated by at least 1.8-fold at E17.5 (*Fabp7/Blbp*, *Apoe Mfge8*, *Slc1a3/Glast*, *Ptprz1*, *Tnc*, *Id3*, *Mt3*, *Ednrb*, *Sparc*, *Atp1a2*, *Bcan*, *Clu*, *Mlc1*, *Fgfbp3*, *Slc9a3r1*, *Aldh1l1*, *Asrgl1*, *Gng12*, *Vcam1*, *Nim1*, *Acad1*, *Vit*, *Acsbg1*, *Metrn*, *Pnp*, *Ccdc80*, *Tgfb2*, *Rhoc*, *Acaa2*, *Id1*, and *Veph1*) (Table S3). The comparison between E17.5 RPs and proliferative E13.5 RPs led to similar conclusions (Figure 5E), with variance between the two due to increased E17.5 expression of RP core genes and increased E13.5 expression of cell cycle-associated genes like *Cenpf*, *Top21*, *Cdca8*, and *Mki67* (Figure 5E).

These analyses used average gene expression to define upregulation of 32 RP core genes from E13.5 to E17.5. We asked about this on the single cell level by ordering each single cell in pseudotime along the developmental trajectory (Figure 3C) in a path beginning at the *k*-means center near (−11, −2) and ending near (20, −11). We then plotted expression of the RP core genes *Ednrb*, *Mfge8*, *Ptprz1*, *Mt3*, *Fabp7/Blbp*, and *Vcam1* in pseudotime. Expression of all of these genes was increased in individual RPs from E13.5 to E17.5 (Figure 5F), demonstrating a reinforcement of the RP transcriptional identity over this time frame. Consistent with this, expression of these RP core genes was generally lower in individual E11.5 APs.

To ensure that changes in the cell cycle did not influence these conclusions, we regressed out the cell cycle genes and reanalyzed a combined dataset including RPs from E13.5, E15.5, and E17.5 using our unsupervised clustering approach (Figure S5). Eight clusters were identified (Figure S5A), with 5 (1, 2, 3, 4, and 6) forming a larger cluster that included most E13.5 and E15.5 RPs (Figure S5B), supporting the idea that these cells were all relatively similar. This analysis also identified 3 more divergent clusters (5, 7, and 8), with the largest (5), containing most of the E17.5 RPs (Figure S5B). Differential gene expression analysis showed that core identity genes were upregulated in this cluster (*Apoe*, *Fabp7*, *Id3*, *Mt3*, *Tnc*, *Slc1a3/Glast*, *Ptprz1*, *Sparc*, *Ednrb*, and *Dbi*), consistent with our other analyses. The other two divergent clusters contained only 10–11 cells each from E15.5 and E17.5. Differential gene expression analysis indicated that cluster 7 specifically expressed the oligodendrocyte precursor genes *Olig1*, *Pdgfra* (Figure S5E), and *Olig2* and cluster 8 genes such as *Sulf1* and *Rspo3* that likely mark the earliest-born astrocytes or ependymal cells (Figure S5F).

The RP Core mRNAs Are Enriched in Apical RPs *In Vivo*

We next validated the RP core identity genes *in vivo*. Initially we performed qRT-PCR analysis of the E12.5–E17.5 cortex for 6 RP core mRNAs that were predicted to be upregulated between E13.5 and E17.5: *Aldoc*, *Ednrb*, *Mfge8*, *Mt3*, *Ptprz1*, and *Vcam1*. For comparison, we examined *Ttyh1*, *Pax6*, and *Hes5*, RP core mRNAs that were not increased from E13.5–E17.5 (Table S4). Because the proportion of RPs decreases over this time frame, we normalized the qPCR data to levels of *Sox2* mRNA, a pan-precursor marker expressed at similar levels in RPs from E13.5 to E17.5 (Table S4; Figure S6A). As predicted, *Pax6*, *Hes5*, and *Ttyh1* mRNA levels were maintained or decreased relative to *Sox2* mRNA from E12.5 to E17.5, whereas the other RP core mRNAs increased robustly (Figure 6A).

We next confirmed that *Ptprz1*, *Dbi*, *Mfge8*, *Ttyh1*, *Ednrb*, and *Mt3* mRNAs were expressed in apical Sox2-positive RPs in the E17.5 cortex using single molecular fluorescence *in situ* hybridization (FISH) (Figure 6B). Immunostaining showed that the Dbi, Ptprz1, and Mfge8 proteins were also expressed in Sox2-positive cortical apical RPs (Figure 6C). These validation studies indicate that the scRNA-seq analysis accurately predicted RP gene expression *in vivo*.

RPs Largely Transition to a Non-Proliferative State between E15.5 and E17.5

The Cyclone analysis predicted one additional striking change in RPs during late embryogenesis. In particular, it predicted that 26% and 7% of E15.5 RPs were in G2/M and S phases of the cell cycle (33% total) and that, by E17.5, this was reduced to 9% in G2/M and none in S phase (Figure 6D). To validate this observation, we injected pregnant mothers with 5-ethynyl-2′-deoxyuridine (EdU) on gestational days 15.5 or 17.5 and immunostained cortical sections from their embryos 6 hr later (Figure 6E;

(F) Gene expression dynamics of *Ednrb*, *Mfge8*, *Ptprz1*, *Vcam1*, *Fabp7*, and *Mt3* in the RP developmental trajectory analysis, as computed using the Waterfall method. The cells in the trajectory shown in Figure 3C were ordered in pseudotime and fit using local polynomial regression fitting (red line) with 95% confidence intervals (gray shade).

See also Figure S5.
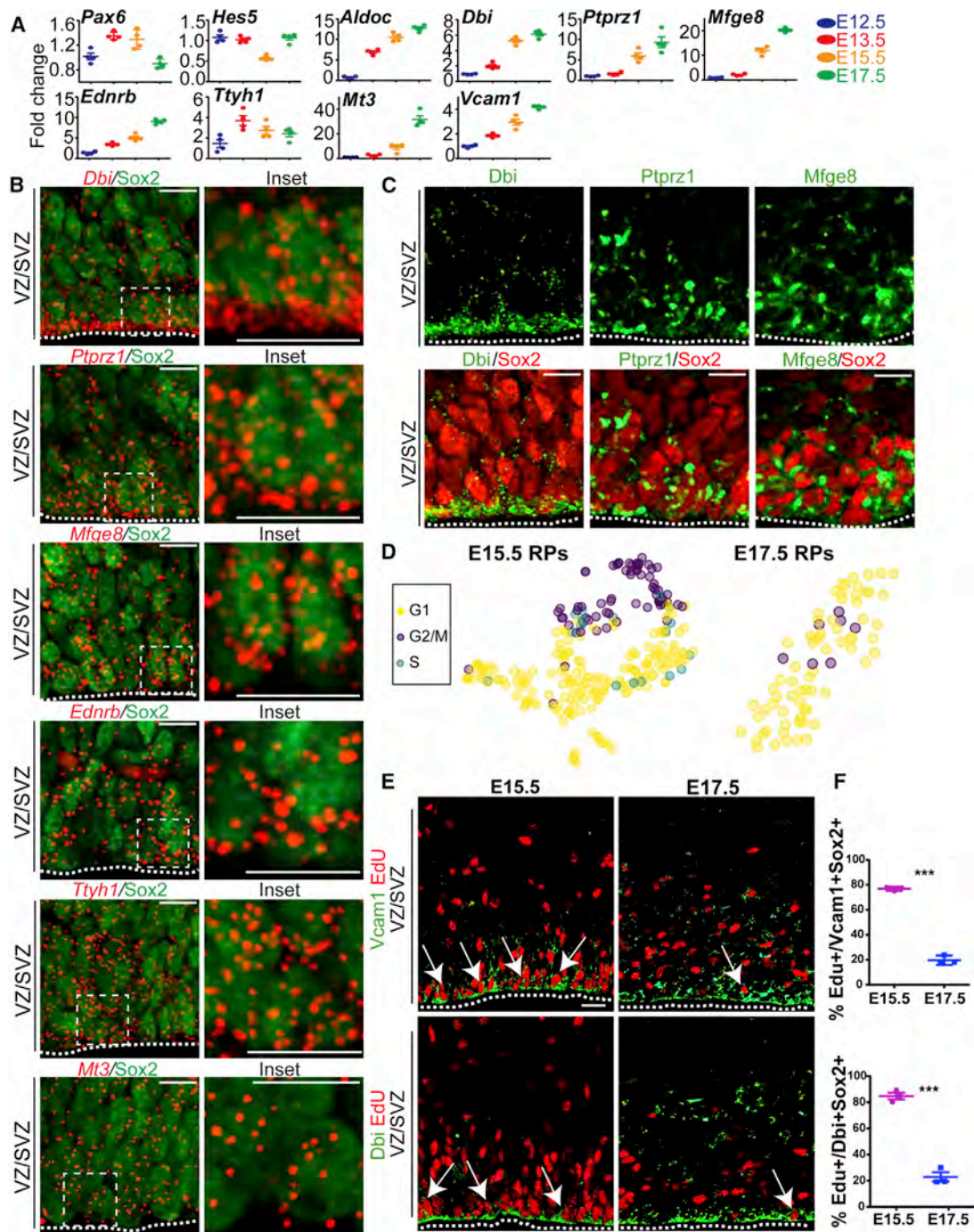
Figure 6. RPs Express Core Identity mRNAs *In Vivo* and Transition to a Non-proliferative State between E15.5 and E17.5

(A) qRT-PCR analysis of *Pax6*, *Hes5*, *Aldoc*, *Dbi*, *Ptprz1*, *Mfge8*, *Ednrb*, *Ttyh1*, *Mt3*, and *Vcam1* mRNAs in E12.5–E17.5 cortices. mRNA levels were normalized to *Sox 2* mRNA levels in the same samples and are expressed as fold change relative to E12.5 (set as one). Data are shown as scatterplots and indicate mean ± SEM. n = 4 independent samples per time point.

(B) Confocal z stack images of E17.5 coronal cortical sections analyzed by FISH for *Dbi*, *Ptprz1*, *Mfge8*, *Ednrb*, *Ttyh1*, and *Mt3* mRNAs (red dots) and by immunostaining for Sox2 (green). Images show the VZ/SVZ adjacent to the lateral ventricle (dotted line). Hatched boxes are shown at higher magnification at the right. Scale bars represent 10 μm.

(C) Confocal images of E17.5 coronal cortical sections immunostained for Sox2 (red) and Dbi, Ptprz1, or Mfge8 protein (green in all cases). Images show the VZ/SVZ adjacent to the lateral ventricle (indicated by the dotted line). Scale bars represent 10 μm.

Figure S6B). At E15.5, almost 80% of Vcam1-positive, Sox2-positive RPs were EdU-positive, but at E17.5, this was reduced to about 20% (Figure 6F). We obtained similar results when we triple-labeled sections for Sox2, EdU, and Dbi (Figures 6E and 6F). Thus, between E15.5 and E17.5, RPs upregulate their core identity genes and exit the cell cycle.

Embryonic Cortical RPs Share Transcriptional Similarities with Adult Forebrain Quiescent NSCs

One reason why RPs might exit the cell cycle at E17.5 is to become quiescent V-SVZ NSCs. Consistent with this idea, our analysis showed that embryonic cortical RPs expressed well-characterized markers of adult V-SVZ neural precursors, including Vcam1, Fabp7/Blbp, Slc1a3/Glast, Sox9, Sox2, Vim, Tnc, Tlx/Nr2e1, Aldoc, Hes1, Hes5, Apoe, Aldh1l1, and Dbi. To test this idea further, we asked whether embryonic RP core genes were expressed in adult V-SVZ precursors. Specifically, we performed FISH for Mfge8, Ednrb, Ttyh1, Mt3, and Ptprz1 mRNAs, combined with immunostaining for Sox2 and Vcam1, which is expressed in quiescent adult V-SVZ B1 NSCs (Codega et al., 2014; Kokovay et al., 2012), and analyzed a region of the V-SVZ that derives from cortical RPs (Figure 7A; Gallagher et al., 2013; Kohwi et al., 2007; Willaime-Morawek et al., 2006). As positive controls, we also analyzed Dbi and Aldoc mRNAs. This analysis showed that, as predicted, Dbi mRNA was enriched in Vcam1-positive, Sox2-positive V-SVZ cells (Figure 7B), whereas Aldoc mRNA was expressed in apical neural precursors and in non-apical Vcam1-negative cells, potentially astrocytes (Figure 7B). Importantly, Mfge8, Ednrb, Ttyh1, Mt3, and Ptprz1 were also all enriched in Sox2-positive, Vcam1-positive V-SVZ cells (Figure 7B). All of these mRNAs were also expressed in Vcam1-positive V-SVZ cells in regions known to derive from the GE rather than the cortex (Figures S7A and S7B).

We extended this analysis of similarities between embryonic cortical RPs and adult V-SVZ precursors using a previously published microarray dataset comparing quiescent with proliferative (activated) adult V-SVZ NSCs (Codega et al., 2014). We directly compared the 1572 genes that were enriched 2-fold in quiescent V-SVZ NSCs with the 235 genes that were statistically enriched in the E17.5 RP scRNA-seq data (Table S2). We found that 97 of the 235 E17.5 RP-enriched genes were also enriched in quiescent adult V-SVZ B1 NSCs (Figure 7C; Table S5). Among these 97 shared genes were 42 embryonic RP core identity genes (Figure 7C; Table S5) and 2 genes that were enriched in E17.5 and E13.5 but not E15.5 RPs, Fgfr3 and Nfe2l2 (Table S3).

Of the remaining 53 shared mRNAs, overlay analysis showed that 29 were very specific for RPs in the E17.5 scRNA-seq data (Table S5). For example, Scrg1, Hepacam, Fxyd1, Slo1c1, Slc38a3, S1pr1, Gja1, Htra1, S100a1, and Aqp4 mRNAs (Figure 7D) were detectable in less than 0.5% of cells outside of

E17.5 RP cluster 8. 27 of these 29 genes (all except Aqp4 and Fxyd1) were also detectably expressed at E15.5 (Figure S7C). Of these, 10 (Slc38a3, Etv4, Gja1, S100a1, Hepacam, Emp2, Htra1, Scrg1, Slco1c1, Serpine2, and S1pr1) were expressed in less than 2.5% of cells in the non-RP clusters at both E15.5 and E17.5. Four of the remaining 17 mRNAs (Mgst1, Myo6, B2m, and Tnfaip8) were expressed in both RPs and IPs at E15.5, whereas the others (Cnp, Il18, Sfxn5, Zfyve21, Paqr7, Tmem47, Plxnc1, S100a16, Slc4a4, Slc15a2, Smpdl3a, and Timp3) were expressed in RPs and scattered neurons. Thus, 73 genes enriched in adult V-SVZ quiescent NSCs were highly enriched in E17.5 cortical RPs, and almost all of these (71) were also expressed in E15.5 RPs, indicating that the embryonic RP transcriptional identity is at least partially maintained in quiescent adult V-SVZ NSCs.

DISCUSSION

Here we have used scRNA-seq to obtain a global overview at the single-cell level of embryonic murine cortical development. We have developed a methodology to analyze these data and provide evidence for three major biological conclusions. First, we show that developmental transitions between cortical RPs, IPs, and neurons do not occur in a stepwise fashion but via a transcriptional continuum. Second, our data indicate that RPs arise from neuroepithelial stem cells around E11.5, when they adopt a core RP transcriptional identity they maintain and reinforce throughout the neurogenic period. Third, we show that RPs become slowly proliferating at the end of neurogenesis and that these relatively quiescent embryonic RPs share transcriptional similarities with adult V-SVZ quiescent NSCs. Thus, we propose that RPs share a core transcriptional identity throughout life and that the transition to quiescent adult NSCs occurs during late embryogenesis.

The high-throughput scRNA-seq analysis presented here identified different embryonic cortical cell types, including apical and basal IPs; distinguished RPs and IPs as a function of cell cycle status; and identified many potential markers for IPs and RPs. The pipeline also clustered cells with regard to developmental transition state so that apical IPs were closer to RPs, basal IPs closer to neurons, and newborn migratory neurons adjacent to IPs. We did not analyze embryonic cortical neurons in detail, but our analysis indicated that they were clustered in part on the basis of developmental stage and in part based on expression of hallmark transcription factors like Bhlhe22 or Satb2. Our data therefore validate scRNA-seq as a way to examine cells through developmental time. Nonetheless, Drop-seq has limitations. First, the coverage of any individual cell's transcriptome is only partial, and, thus, the inability to detect an mRNA, particularly lower abundance

---

(D) Cyclone analysis of E15.5 and E17.5 scRNA-seq data, visualized by t-SNE, showing RP clusters 10 and 13 (E15.5) and 8 (E17.5) (Figure 4A). Cells are color-coded for predicted cell cycle phases (yellow, G1; light blue, S; purple, G2/M).
(E and F) E15.5 or E17.5 embryos were exposed to maternally injected EdU, and 6 hr later, coronal cortical sections were immunostained for EdU (red, E), Sox2 (not shown) and Vcam1 (green, top, E), or Dbi (green, bottom, E) and quantified for the proportion of Sox2-positive, Vcam1-positive, or Dbi-positive cells that were also EdU-positive (F). Images show the VZ/SVZ adjacent to the lateral ventricles (hatched white lines). Arrows denote double-labeled cells. ***p < 0.001; n = 3 brains each, 3 sections per brain. Scale bars indicate 20 μm (E). Error bars denote SEM (F).
See also Figure S6.
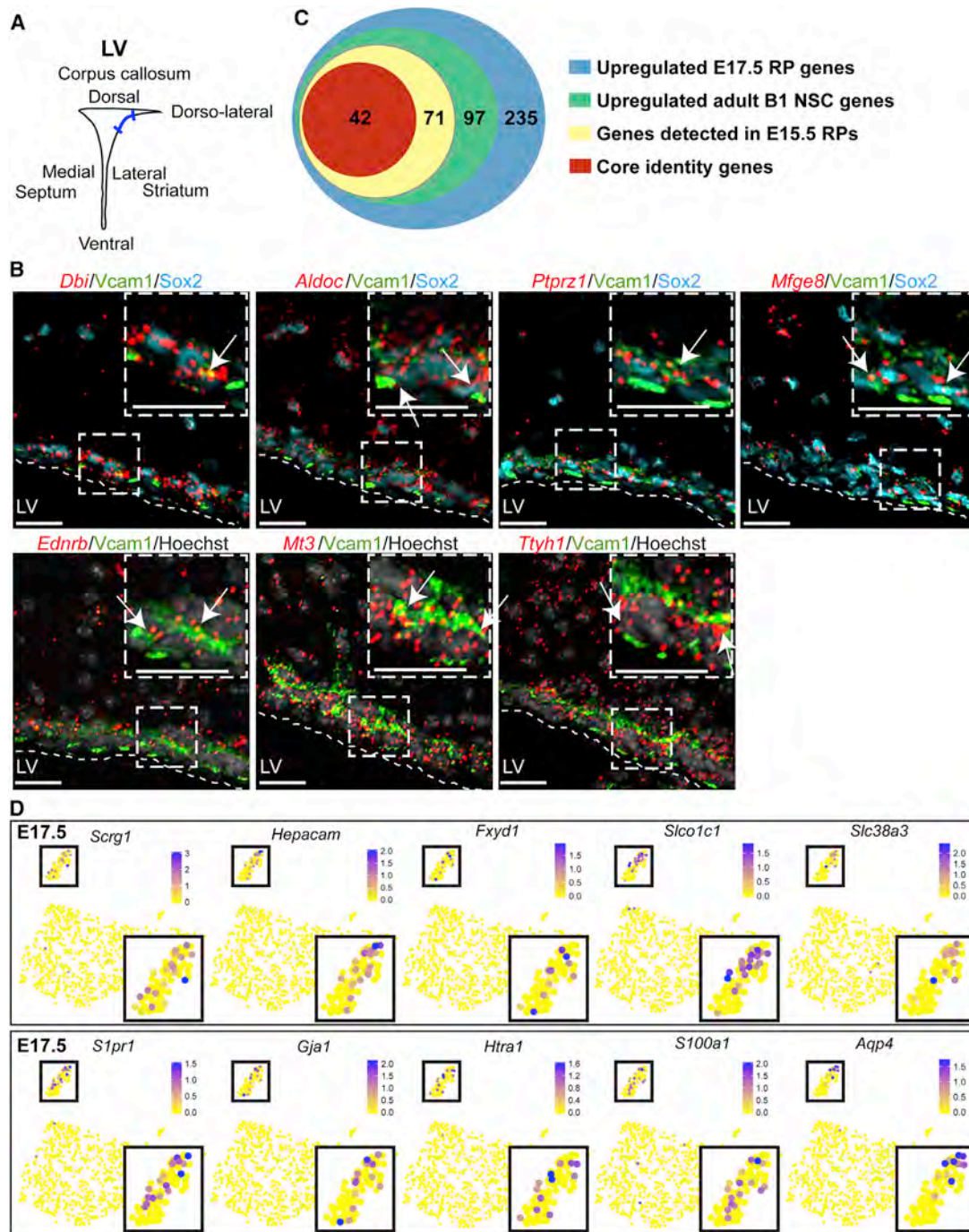
Figure 7. Adult Forebrain Quiescent NSCs and Embryonic RPs Share Transcriptional Similarities

(A) Schematic showing the cortically derived region (blue) of the adult mouse V-SVZ that was analyzed.

(B) Confocal z stack images of coronal sections through the V-SVZ region shown in (A), analyzed by FISH for *Dbi*, *Aldoc*, *Ptprz1*, *Mfge8*, *Ednrb*, *Mt3*, or *Ttyh1* mRNAs (red dots) and immunostaining for Vcam1 (green) and Sox2 (blue, top). Bottom: Hoechst 33258 counterstaining (light gray). Hatched white lines denote borders of the lateral ventricle (LV). Boxed regions are shown at higher magnification in the top right corners, and arrows indicate triple- or double-labeled cells, as relevant. Scale bars represent 23 μm.

(C) Venn diagram showing the overlap between the 235 E17.5 RP-enriched genes (p < 0.01 FWER, blue oval) and genes upregulated more than 2-fold (p < 0.05 false discovery rate [FDR]) in quiescent adult V-SVZ B1 NSCs, as identified by Codega et al. (2014; GEO: GSE54653). Of the 97 overlapping genes (green oval), 71 were detected in both E15.5 and E17.5 RPs (yellow oval), and of those, 42 were RP core identity genes (red oval).

(legend continued on next page)

mRNAs, does not definitively establish that the mRNA is not expressed. Second, there may be ambient RNA within the droplets. Here we have sorted out dead and dying cells prior to Drop-seq to minimize this issue, but it might nonetheless confound our analysis to some degree.

Our conclusion that there is a transcriptional continuum rather than stepwise acquisition of a new cell state is consistent with a recent study of human hematopoietic stem cells (Moussy et al., 2017). This transcriptional continuum was perhaps best exemplified by IPs that expressed RP genes like *Pax6* and *Slc1a3/Glast* and neuronal genes like *Tubb3*. Why then are these genes considered to be cell-type-specific markers? We propose that the proteins are more specific than the mRNAs and that this reflects an important role for post-transcriptional mechanisms in regulating embryonic corticogenesis. As one example of such a mechanism, we previously provided evidence that RPs are transcriptionally primed to differentiate and that they are maintained in an undifferentiated RP state by translational repression (Amadei et al., 2015; Yang et al., 2014).

One major result of this study is the identification of a core embryonic RP transcriptional identity that is in place at E13.5 and that persists throughout embryogenesis. Although some of these core identity genes were previously known, such as *Pax6*, *Hes1*, *Sox9*, and *Slc1a3/Glast*, many of them have not previously been characterized within a neural precursor context. The ability to identify a large cohort of previously undescribed gene markers highlights a major advantage of scRNA-seq. However, identification of an embryonic RP core identity does not imply that RPs are transcriptionally identical from E13.5 to E17.5. Indeed, we show that the levels of 32 RP core mRNAs increase over this time frame and that, of 739 genes differentially enriched in RPs at E13.5, E15.5, or E17.5, 355 of them were enriched at only one embryonic age. Nonetheless, even when not enriched, the large majority of these genes were still expressed in RPs at the other time points, consistent with the very high correlations between RPs at E13.5 and E17.5. Thus, the fundamental RP cellular identity was similar throughout this developmental window.

Our final major conclusion involves the emergence of a slowly proliferating E17.5 RP population that shares transcriptional similarities with quiescent adult V-SVZ NSCs. Previous studies have shown that the same cortical RPs that make embryonic neurons also persist, at least in part, to populate the V-SVZ as adult NSCs (Fuentealba et al., 2015; Furutachi et al., 2015). Here we define a molecular identity for these embryonic cortical RPs and provide evidence that they maintain this transcriptional identity as they transition to quiescent NSCs that ultimately persist into adulthood.

EXPERIMENTAL PROCEDURES

Animals
All animal use was approved by the Animal Care Committee of the Hospital for Sick Children in accordance with the Canadian Council of Animal Care policies. CD1 mice, purchased from Charles River Laboratories, were used for all embryonic experiments, whereas C57BL/6 mice (The Jackson Laboratory) were used for the adult analyses. For the EdU analyses, E15.5 and E17.5 pregnant CD1 mice were injected with 225 μL of 20 mg/mL EdU intraperitoneally, and the brains of their embryonic progeny was dissected 6 hr later. For all studies, mice of either sex were used.

Single-Cell Isolation, and Drop-seq
For single cell isolation, E13.5, E15.5, and E17.5 embryos were collected from CD1 timed-pregnant mice (Charles River Laboratories), and the age of individual embryos was confirmed by measuring crown-rump lengths. In total, 12, 16, 13, and 8 embryos were analyzed at E11.5, E13.5, E15.5, and E17.5, respectively. Cortices from embryonic brains were dissected and mechanically dissociated in cortical precursor medium (Neurobasal medium [Invitrogen] supplemented with 2% B27 [Invitrogen] and 0.5 mM L-glutamine [Invitrogen]). Following dissociation, cells were filtered through a 70-μm cell strainer (BD Biosciences), counted with a hemocytometer, and pelleted by centrifugation at 1,200 rpm for 5 minutes (Eppendorf, 5804R). The cell pellet was then re-suspended in a solution containing 0.25% BSA in 1× Hank's balanced salt solution (HBSS) at a concentration of 5,000,000 cells/mL. Propidium iodide (Abcam) was added to the cell suspension at a final concentration of 1 μg/mL, and cells were filtered a second time through a 70-μm cell strainer. Viable (propidium iodide [PI]-negative) cells were subsequently sorted using a MoFlo Astrios (Beckman Coulter) cell sorter. Following cell sorting, cells were pelleted at 1,200 rpm for 10 min (Eppendorf, 5804R), re-suspended in 0.01% BSA (dissolved in 1× PBS, Invitrogen), counted with a hemocytometer, and adjusted to a final concentration of 140,000 cells/mL. Following cell sorting, cells were typically more than 90% trypan blue (Invitrogen)-negative. Drop-seq was then carried out at the Princess Margaret Genomics Facility (Toronto, ON) exactly as described previously (Macosko et al., 2015), except that flow rates of 3000 μL/hr (cells and beads) and 13,000 μL/hr (oil) were used, along with a bead concentration of 167,000 beads/mL. In our hands, these conditions yielded droplets of an average size of ∼110 μm.

Cells isolated from E11.5 embryonic cortices were treated identically except that they were not sorted for PI-negativity because of the low cell number and the high viability of ∼80%, as judged by trypan blue exclusion. For experiments investigating the effect of sorting, E13.5 single cells were isolated as described above, and one-half of the cell preparation was sorted whereas the other half was not.

Following Drop-seq droplet collection, cDNA amplification and sequencing library preparation were carried out exactly as described previously (Macosko et al., 2015), and the libraries were sequenced on an Illumina NextSeq500 at the Princess Margaret Genomics Facility (Toronto, ON). For Drop-seq data from E11.5, E13.5, and E17.5 embryos, the libraries from one batch of droplets were sequenced individually on a single NextSeq500 flow cell, whereas, at E15.5, libraries from two separate droplet batches (prepared from the same batch of cells) were sequenced together on two separate NextSeq500 flow cells, and the data were combined.

Drop-seq Data Sequencing Read Alignment
FASTQ sequencing reads were processed, aligned to the mouse genome (mm10), and converted to digital gene expression matrices using the Drop-seq tools (version 1.12, http://mccarrolllab.com/dropseq/) with settings as described in the Drop-seq Alignment Cookbook (version 1.2, January 2016, http://mccarrolllab.com/dropseq/), written by Jim Nemesh in the McCarroll laboratory. The number of cell barcodes per embryonic age was identified by calculating the cumulative fraction of reads attributable to each individual cell barcode and arranging these in decreasing order in R as described in the Drop-seq Alignment Cookbook. Plotting the cumulative fractions of reads in decreasing order results in knee plots, where the inflection points in the graphs are attributable to the number of cell barcodes for which data were

---

(D) t-SNE visualizations of the expression of *Scrg1*, *Hepacam*, *Fxyd1*, *Slco1c1*, *Slc38a3*, *S1pr1*, *Gja1*, *Htra1*, *S100a1*, and *Aqp4* mRNAs in the E17.5 scRNA-seq data. Cells are color-coded for relative expression levels, from not detected (yellow) to the highest detectable levels (blue), according to the adjacent color key. The boxed regions highlight the RP clusters and are shown at higher resolution in the bottom right corners.
See also Figure S7.

collected. For E11.5, E13.5, and E17.5, the number of cell barcodes was set at 2,000. For the E15.5 data, the number of cell barcodes was set at 2,000 for the first batch and 3,000 for the second batch, and, after the raw digital gene expression matrices were determined, these barcodes were merged into one combined 5,000-cell barcode matrix.

Drop-seq Data Analysis Pipeline

To analyze the Drop-seq data, we assembled a modified analysis pipeline that combines low-level data analysis using the computational work-flow described by Lun et al. (2016) with shared nearest neighbor (SNN)-Cliq-inspired clustering and t-SNE projection using the Seurat R package (http://satijalab.org/seurat/). Data were then visualized in a custom-written Shiny script in R. The full details of this modified analysis pipeline can be found at https://github.com/BaderLab/scRNAseqPipeline. Briefly, our modified analysis pipeline begins with a number of low-level data analysis steps implemented in the workflow described by Lun et al. (2016). First, cell doublets were filtered by fitting the library size of each cell to a negative binomial distribution and removing large libraries ($p \leq 0.001$). Second, cell contaminant filtering (red blood cells) was carried out by removing cells that were outliers in a correlation between library size and gene detection rate in each cell. Third, cells were filtered for mitochondrial gene content by removing cells that were at least 4 median average deviations higher than the median mitochondrial transcript proportion at each time point using the scran package. Fourth, the cell cycle phase of each cell was predicted using the Cyclone method (Scialdone et al., 2015), and then low-abundance genes were removed by removing any gene that was found (unique molecular identifier expression value > 0) in less than 3 cells. Last, the data were normalized according to Lun et al. (2016) using the default implementation of their pool and deconvolute normalization algorithm in the scran package. In brief, hierarchical clustering on a distance metric derived from Spearman's correlation was performed to subset the data into more homogeneous groups. Within each group, cell-wise scaling factors were determined, and then normalization was performed between groups. Scaling factors per cell were determined by pooling random subsets of cells, summing their library sizes, and comparing with average library size across all cells in the group. This is iteratively performed, and the cell-wise scaling factors can be deconvolved from the set of pool-wise scaling factors. This method is robust to the sparsity of the data and respects the assumption of minimal differential gene expression common to most normalization methods. Following low-level data analysis and normalization, data were transferred to a Seurat object where highly variable genes were used to first carry out PCA. Significant principal components were then used to iteratively carry out SNN-Cliq-inspired clustering (using Seurat) with increasing resolution until the number of differentially expressed genes (calculated by the Seurat FindMarkers function, $p < 0.01$ family-wise error rate, Holm's method) between the most similar clusters reached a minimum of 10 genes. t-SNE projections, cell cycle annotations, and differentially expressed genes (heatmaps) were produced in R using our custom-written Shiny script. Most t-SNE gene expression overlays were carried out using the FeaturePlot function in Seurat.

Following cell clustering and visualization by t-SNE projection, we annotated each identified cluster based on the expression of known marker genes for each population. Following cell clustering annotation at each embryonic age, we selected all barcodes from clusters containing cells expressing the cortical transcription factor *Emx1* and that had few or no cells expressing the GE transcription factors *Dlx1*, *Dlx2*, or *Dlx5* or the interneuron genes *Gad1* or *Gad2*. We also removed the small cell clusters that contained cells expressing genes characteristic of microglia (*Aif1*) or mesenchymal cells (*Col* genes). We extracted these barcodes from the raw digital gene expression matrices for each embryonic age and then carried out our analysis pipeline on these cortex-only populations. Digital gene expression matrices and raw sequence data have been deposited into GEO: GSE107122.

Other Computational Methods

Trajectory inference analysis was carried out by selecting all of the barcodes in the E11.5 AP clusters (clusters 1, 3, 4, and 6) and the RP clusters at E13.5 (clusters 2 and 8), E15.5 (clusters 10 and 13), and E17.5 (cluster 8), as defined in Figure 2A. We extracted these cell barcodes from each raw digital gene expression matrix and merged these selected barcodes together for all four

ages to yield one raw digital gene expression matrix for the AP populations. We then analyzed this dataset using our analysis pipeline. Following the normalization step, PCA analysis was performed on all genes and all cells in the matrix. k-means cluster, minimum-spanning tree (MST) calculation, and pseudotime ordering of cells was performed using the Waterfall method (Shin et al., 2015) with functions implemented in R. Point sizes in the PCA plot were adjusted proportionally according to the number of transcripts detected per cell by computing the sum of the number of transcripts for all genes per cell and dividing by 2,500. Two possible trajectory paths are possible through the MST, with one path beginning near (-11, -2) and ending near (20, -11) and a second path beginning near (11, 12) and ending near (20, -11). We analyzed both but ultimately focused on the path beginning near (-11, -2) and ending near (20, -11) because we found that, qualitatively, both paths led to similar conclusions. To analyze single genes across this trajectory, we ordered cells in pseudotime and plotted cells based on their expression in pseudotime, and then expression curves were fit using local polynomial regression fitting with 95% confidence intervals using the Waterfall method functions implemented in R.

The combined analysis of E13.5, E15.5, and E17.5 RP populations was performed by combining the same RP populations described above for the Waterfall trajectory analysis except that the E11.5 data were omitted. Following normalization, cell cycle regression was performed as described by Lun et al. (2016) using the removeBatchEffect function implemented in the limma package in R using the S and G2/M cell cycle scores (calculated by the Cyclone method) as covariates. Following cell cycle regression, the data were then clustered exactly as described under Drop-seq Data Analysis Pipeline.

Correlation analysis was carried out by averaging the expression of each gene across all cells in individual clusters and then carrying out Pearson correlation analysis using the cor.test function in R. Venn diagrams comparing differentially expressed genes in RP clusters with all other clusters at each embryonic age of E13.5, E15.5, and E17.5 were generated using the VennDiagram package in R. Last we used the VennDiagram package described above to compare differentially expressed genes in RP clusters at E17.5 (Table S2) to genes with unique gene symbols that were upregulated in quiescent adult V-SVZ NSCs versus proliferating aNSCs (>2-fold upregulation, $p < 0.05$ false discovery rate), as shown in Table S1 of Codega et al. (2014) (GEO: GSE54653).

PANTHER classification was performed using the 90 RP core identity genes symbols as input gene IDs for functional classification implemented at pantherdb.org.

Tissue Preparation, Immunostaining, and EdU Analysis

For embryonic tissue, brains were dissected from pregnant CD1 mice at E17.5, fixed in 4% paraformaldehyde (PFA) for 24 hr, transferred to 30% sucrose for 24–48 hr, embedded in optimum cutting temperature (O.C.T.) compound (Tissue-Tek), and sectioned coronally at 16 μm either immediately or after storage at $-80°$C. Adult brains were dissected from 8-week-old C57BL/6 mice (The Jackson Laboratory) following whole-animal perfusion with 30 mL of PBS and 30 mL 4% PFA. Adult brains were post-fixed, cryoprotected, and sectioned as the embryonic brains.

For immunostaining, frozen sections were dried for 30 min at 37°C, rehydrated in PBS for 5 min, and blocked and permeabilized in a 5% BSA solution containing 0.3% Triton X-100. Sections were incubated in primary antibody in 2.5% BSA overnight at 4°C in a humidified chamber. Sections were then washed three times in PBS, incubated in fluorescently labeled secondary antibody (Invitrogen) at 1:1,000 dilution in PBS for 1 hr at room temperature, washed three times in PBS, counterstained with 0.5 μg/mL Hoechst 33258 (Sigma-Aldrich) for 5 min at room temperature, washed twice in PBS, and mounted on glass slides using PermaFluor (Thermo Scientific). For the EdU experiments, following incubation in secondary antibody, sections were treated with 4% PFA for 20 min and washed three times with PBS, and EdU was detected using the Molecular Probes Click-It EdU reaction kit (Invitrogen) as described by the manufacturer. Following the Click-It reaction, sections were washed three times in PBS, counterstained with 1 μg/mL Hoechst for 30 min, washed twice in PBS, and mounted as described above. The following primary antibodies were used: goat anti-Sox2 (Santa Cruz

Biotechnology, 1:250), rabbit anti-Sox2 (Cell Signaling Technology, 1:500), rabbit anti-Vcam1 (Abcam, 1:200), rabbit anti-Dbi (Aviva Systems Biology, 1:100), rabbit anti-Ptprz1 (Abcam, 1:200), and goat anti-Mfge8 (R&D Systems; 1:200).

Single-Molecule FISH

Sections for FISH were prepared as for immunostaining with RNase-free conditions and reagents and analyzed using the RNAscope Multiplex Fluorescent Assay Kit (Advanced Cell Diagnostics). Sections were dried for 10–20 min at 37°C, rehydrated in PBS for 5 min, and washed in 50%, 70%, and 100% ethanol (EtOH) sequentially for 5, 5, and 2 × 5 min, respectively. After drying, sections were permeabilized for 10 min at 37°C using a 1:10 dilution of RNAscope Pretreatment-4 protease solution (Advanced Cell Diagnostics), washed, and maintained in PBS until probe addition. Probes were pre-heated to 40°C for 10 min and added to sections in probe solutions (Advanced Cell Diagnostics) for 2 hr at 40°C. Probes were used to target *Vcam1* (catalog no. 438641, NM_011693.3), *Aldoc* (catalog no. 429531-C3, NM_009657.3), *Dbi* (catalog no. 502601, NM_007830.4), *Ptprz1* (catalog no. 460991, NM_001081306.1), *Mfge8* (catalog no. 408771, NM_001045489.1), *Ttyh1* (catalog no. 504051-C3, NM_001001454.4), *Mt3* (catalog no. 504061-C3, NM_013603.2), and *Ednrb* (catalog no. 473801, NM_007904.4). Sections were subsequently washed (always with wash buffer for 1 min each) four times, incubated in RNAscope AMP-1 solution for 30 min at 40°C, washed four times, incubated in RNAscope AMP-2 solution for 15 min at 40°C, washed four times, incubated in RNAscope AMP-3 solution for 30 min at 40°C, washed four times, incubated in RNAscope AMP-4-FL solution for 15 min at 40°C, washed four times, and incubated in RNAscope DAPI solution for 1 min at room temperature (RT). For concomitant immunostaining, the DAPI was eliminated, sections were instead incubated in 5% BSA blocking buffer at RT for 30 min, and immunostaining was performed as described above.

Imaging and Microscopy

Images of immunostaining or FISH were collected using a Quorom spinning disk confocal microscope system or a Zeiss Axio Imager M2 system with an X-Cite 120 LED light source and a C11440 Hamamatsu camera. For FISH, z stacks of confocal images were taken with an optical slice thickness of 0.25 μm, and projected z-stacked images are shown.

qRT-PCR

qRT-PCR was performed as described in Voronova et al. (2017) from dissected cortices at ages E12.5, E13.5, E15.5, and E17.5 using the following primers: *Mfge8* 5′-CAGCAACTATGATAGCAAGCCC-3′ (forward) and 5′-CCT GCGTCATCACACCTGATA-3′ (reverse); *Mt3* 5′-ACCTGCCCCTGTCCTA CTG-3′ (forward) and 5′-CCTTGGCACACTTCTCACATC-3′ (reverse); *Aldoc* 5′-AGAAGGAGTTGTCGGATATTGCT-3′ (forward) and 5′-TTCTCCACCCCAA TTTTGGCTC-3′ (reverse); *Dbi* 5′-CAAGCTACTGTGGGCGATGTA-3′ (forward) and 5′- CACATAGGTCTTCATGGCACTTT-3′ (reverse); *Ptprz1* 5′-GGAGAAGA ACAGAACATCGTCC-3′(forward) and 5′-TCATTGCTCTGGTAATAGCCCA-3′ (reverse); *Ednrb* 5′-CGTGTTCGTGCTAGGCATCAT-3′ (forward) and 5′-GCG ATCAAGATATTGGGACCAT-3′ (reverse); *Hes5* 5′-GTCTCCACGATGATCCT TAAA-3′ (forward) and 5′-CTGTGTTCTCCCACATGAC-3′ (reverse); *Sox2* 5′-GCGGAGTGGAAACTTTTGTCC-3′ (forward) and 5′-CGGGAAGCGT GTACTTATCCTT-3′ (reverse); *Pax6* 5′-TACCAGTGTCTACCAGCCAAT-3′ (forward) and 5′- TGCACGAGTATGAGGAGGTCT-3′ (reverse); *Vcam1* 5′-GCC CACTAAACGCGAAGGT-3′ (forward) and 5′- ACTGGGTAAATGTCTGGA GCC-3′ (reverse); *Ttyh1* 5′-CTAGAGGGCCTACTGTTCCTG-3′ (forward) and 5′- TCGAAGACTTCTGTCCCTATTCT-3′ (reverse); *Rps18* 5′-AGTTCCAG CACATTTTGCGAG-3′ (forward) and 5′-TCATCCTCCGTGAGTTCTCCA-3′ (reverse).

Quantification and Statistical Analysis

Quantification of EdU-positive cells was performed by counting Vcam1 or Dbi and EdU double-positive cells in the ventricular zone/subventricular zone (VZ/SVZ) regions in anatomically matched areas of the cortex. Three distinct regions in the dorsal ventral axis of each section (dorsal region, medial region, and ventral region) were counted in each of three sections/mouse, and the

average of the three regions on these sections was determined for each mouse. Statistical significance was tested using unpaired two-tailed t tests, with p < 0.05 considered to be statistically significant. The statistics used for the computational analyses are described in the relevant sections.

DATA AND SOFTWARE AVAILABILITY

The accession number for the scRNA-seq data reported in this paper is GEO: GSE107122.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.12.017.

AUTHOR CONTRIBUTIONS

S.A.Y., M.J.B., and T.K. collected the scRNA-seq data. S.A.Y. and M.J.B. analyzed the scRNA-seq data and performed most biological experiments. B.T.I. designed and implemented the scRNA-seq data analysis pipeline and assisted with data analysis. A.V. performed qRT-PCR and assisted with FISH and immunostaining experiments. D.R.K., G.D.B., and F.D.M. participated in data interpretation, directed the study, and co-wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Alfonso, J., Le Magueresse, C., Zuccotti, A., Khodosevich, K., and Monyer, H. (2012). Diazepam binding inhibitor promotes progenitor proliferation in the postnatal SVZ by reducing GABA signaling. Cell Stem Cell *10*, 76–87.

Amadei, G., Zander, M.A., Yang, G., Dumelie, J.G., Vessey, J.P., Lipshitz, H.D., Smibert, C.A., Kaplan, D.R., and Miller, F.D. (2015). A Smaug2-based translational repression complex determines the balance between precursor maintenance versus differentiation during mammalian neurogenesis. J. Neurosci. *35*, 15666–15681.

Anthony, T.E., Klein, C., Fishell, G., and Heintz, N. (2004). Radial glia serve as neuronal progenitors in all regions of the central nervous system. Neuron *41*, 881–890.

Codega, P., Silva-Vargas, V., Paul, A., Maldonado-Soto, A.R., Deleo, A.M., Pastrana, E., and Doetsch, F. (2014). Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche. Neuron *82*, 545–559.

Eisenstat, D.D., Liu, J.K., Mione, M., Zhong, W., Yu, G., Anderson, S.A., Ghattas, I., Puelles, L., and Rubenstein, J.L. (1999). DLX-1, DLX-2, and DLX-5 expression define distinct stages of basal forebrain differentiation. J. Comp. Neurol. *414*, 217–237.

Englund, C., Fink, A., Lau, C., Pham, D., Daza, R.A., Bulfone, A., Kowalczyk, T., and Hevner, R.F. (2005). Pax6, Tbr2, and Tbr1 are expressed sequentially by

radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. J. Neurosci. 25, 247–251.

Fuentealba, L.C., Rompani, S.B., Parraguez, J.I., Obernier, K., Romero, R., Cepko, C.L., and Alvarez-Buylla, A. (2015). Embryonic origin of postnatal neural stem cells. Cell 161, 1644–1655.

Furutachi, S., Miya, H., Watanabe, T., Kawai, H., Yamasaki, N., Harada, Y., Imayoshi, I., Nelson, M., Nakayama, K.I., Hirabayashi, Y., and Gotoh, Y. (2015). Slowly dividing neural progenitors are an embryonic origin of adult neural stem cells. Nat. Neurosci. 18, 657–665.

Gallagher, D., Norman, A.A., Woodard, C.L., Yang, G., Gauthier-Fisher, A., Fujitani, M., Vessey, J.P., Cancino, G.I., Sachewsky, N., Woltjen, K., et al. (2013). Transient maternal IL-6 mediates long-lasting changes in neural stem cell pools by deregulating an endogenous self-renewal pathway. Cell Stem Cell 13, 564–576.

Gorski, J.A., Talley, T., Qiu, M., Puelles, L., Rubenstein, J.L., and Jones, K.R. (2002). Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. J. Neurosci. 22, 6309–6314.

Götz, M., Stoykova, A., and Gruss, P. (1998). Pax6 controls radial glia differentiation in the cerebral cortex. Neuron 21, 1031–1044.

Gulisano, M., Broccoli, V., Pardini, C., and Boncinelli, E. (1996). Emx1 and Emx2 show different patterns of expression during proliferation and differentiation of the developing cerebral cortex in the mouse. Eur. J. Neurosci. 8, 1037–1050.

Hartfuss, E., Galli, R., Heins, N., and Götz, M. (2001). Characterization of CNS precursor subtypes and radial glia. Dev. Biol. 229, 15–30.

Haubensak, W., Attardo, A., Denk, W., and Huttner, W.B. (2004). Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. Proc. Natl. Acad. Sci. USA 101, 3196–3201.

Hevner, R.F., Shi, L., Justice, N., Hsueh, Y., Sheng, M., Smiga, S., Bulfone, A., Goffinet, A.M., Campagnoni, A.T., and Rubenstein, J.L. (2001). Tbr1 regulates differentiation of the preplate and layer 6. Neuron 29, 353–366.

Hu, X.L., Chen, G., Zhang, S., Zheng, J., Wu, J., Bai, Q.R., Wang, Y., Li, J., Wang, H., Feng, H., et al. (2017). Persistent expression of VCAM1 in radial glial cells is required for the embryonic origin of postnatal neural stem cells. Neuron 95, 309–325.e6.

Hutton, S.R., and Pevny, L.H. (2011). SOX2 expression levels distinguish between neural progenitor populations of the developing dorsal telencephalon. Dev. Biol. 352, 40–47.

Imai, Y., Ibata, I., Ito, D., Ohsawa, K., and Kohsaka, S. (1996). A novel gene iba1 in the major histocompatibility complex class III region encoding an EF hand protein expressed in a monocytic lineage. Biochem. Biophys. Res. Commun. 224, 855–862.

Kawaguchi, A., Ikawa, T., Kasukawa, T., Ueda, H.R., Kurimoto, K., Saitou, M., and Matsuzaki, F. (2008). Single-cell gene profiling defines differential progenitor subclasses in mammalian neurogenesis. Development 135, 3113–3124.

Kohwi, M., Petryniak, M.A., Long, J.E., Ekker, M., Obata, K., Yanagawa, Y., Rubenstein, J.L.R., and Alvarez-Buylla, A. (2007). A subpopulation of olfactory bulb GABAergic interneurons is derived from Emx1- and Dlx5/6-expressing progenitors. J. Neurosci. 27, 6878–6891.

Kokovay, E., Wang, Y., Kusek, G., Wurster, R., Lederman, P., Lowry, N., Shen, Q., and Temple, S. (2012). VCAM1 is essential to maintain the structure of the SVZ niche and acts as an environmental sensor to regulate SVZ lineage progression. Cell Stem Cell 11, 220–230.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. Annu. Rev. Neurosci. 32, 149–184.

Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. 5, 2122.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202–1214.

Menezes, J.R., and Luskin, M.B. (1994). Expression of neuron-specific tubulin defines a novel population in the proliferative layers of the developing telencephalon. J. Neurosci. 14, 5399–5416.

Merkle, F.T., Mirzadeh, Z., and Alvarez-Buylla, A. (2007). Mosaic organization of neural stem cells in the adult brain. Science 317, 381–384.

Moussy, A., Cosette, J., Parmentier, R., da Silva, C., Corre, G., Richard, A., Gandrillon, O., Stockholm, D., and Páldi, A. (2017). Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. PLoS Biol. 15, e2001867.

Noctor, S.C., Martínez-Cerdeño, V., Ivic, L., and Kriegstein, A.R. (2004). Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. Nat. Neurosci. 7, 136–144.

O'Leary, D.D., and Sahara, S. (2008). Genetic regulation of arealization of the neocortex. Curr. Opin. Neurobiol. 18, 90–100.

Scialdone, A., Natarajan, K.N., Saraiva, L.R., Proserpio, V., Teichmann, S.A., Stegle, O., Marioni, J.C., and Buettner, F. (2015). Computational assignment of cell-cycle stage from single-cell transcriptome data. Methods 85, 54–61.

Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Song, J., Bonaguidi, M.A., Enikolopov, G., Nauen, D.W., Christian, K.M., Ming, G.L., and Song, H. (2015). Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. Cell Stem Cell 17, 360–372.

Simeone, A., Gulisano, M., Acampora, D., Stornaiuolo, A., Rambaldi, M., and Boncinelli, E. (1992). Two vertebrate homeobox genes related to the Drosophila empty spiracles gene are expressed in the embryonic cerebral cortex. EMBO J. 11, 2541–2550.

Takahashi, T., Nowakowski, R.S., and Caviness, V.S., Jr. (1995). The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall. J. Neurosci. 15, 6046–6057.

Voronova, A., Yuzwa, S.A., Wang, B.S., Zahr, S., Syal, C., Wang, J., Kaplan, D.R., and Miller, F.D. (2017). Migrating interneurons secrete fractalkine to promote oligodendrocyte formation in the developing mammalian brain. Neuron 94, 500–516.e9.

Willaime-Morawek, S., Seaberg, R.M., Batista, C., Labbé, E., Attisano, L., Gorski, J.A., Jones, K.R., Kam, A., Morshead, C.M., and van der Kooy, D. (2006). Embryonic cortical neural stem cells migrate ventrally and persist as postnatal striatal stem cells. J. Cell Biol. 175, 159–168.

Yang, G., Smibert, C.A., Kaplan, D.R., and Miller, F.D. (2014). An eIF4E1/4E-T complex determines the genesis of neurons from precursors by translationally repressing a proneurogenic transcription program. Neuron 84, 723–739.

Yang, G., Cancino, G.I., Zahr, S.K., Guskjolen, A., Voronova, A., Gallagher, D., Frankland, P.W., Kaplan, D.R., and Miller, F.D. (2016). A Glo1-Methylglyoxal pathway that is perturbed in maternal diabetes regulates embryonic and adult neural stem cell pools in murine offspring. Cell Rep. 17, 1022–1036.

Young, K.M., Fogarty, M., Kessaris, N., and Richardson, W.D. (2007). Subventricular zone stem cells are heterogeneous with respect to their embryonic origins and neurogenic fates in the adult olfactory bulb. J. Neurosci. 27, 8286–8296.

Supplemental Information

Developmental Emergence

of Adult Neural Stem Cells as Revealed

by Single-Cell Transcriptional Profiling

Scott A. Yuzwa, Michael J. Borrett, Brendan T. Innes, Anastassia Voronova, Troy Ketela, David R. Kaplan, Gary D. Bader, and Freda D. Miller
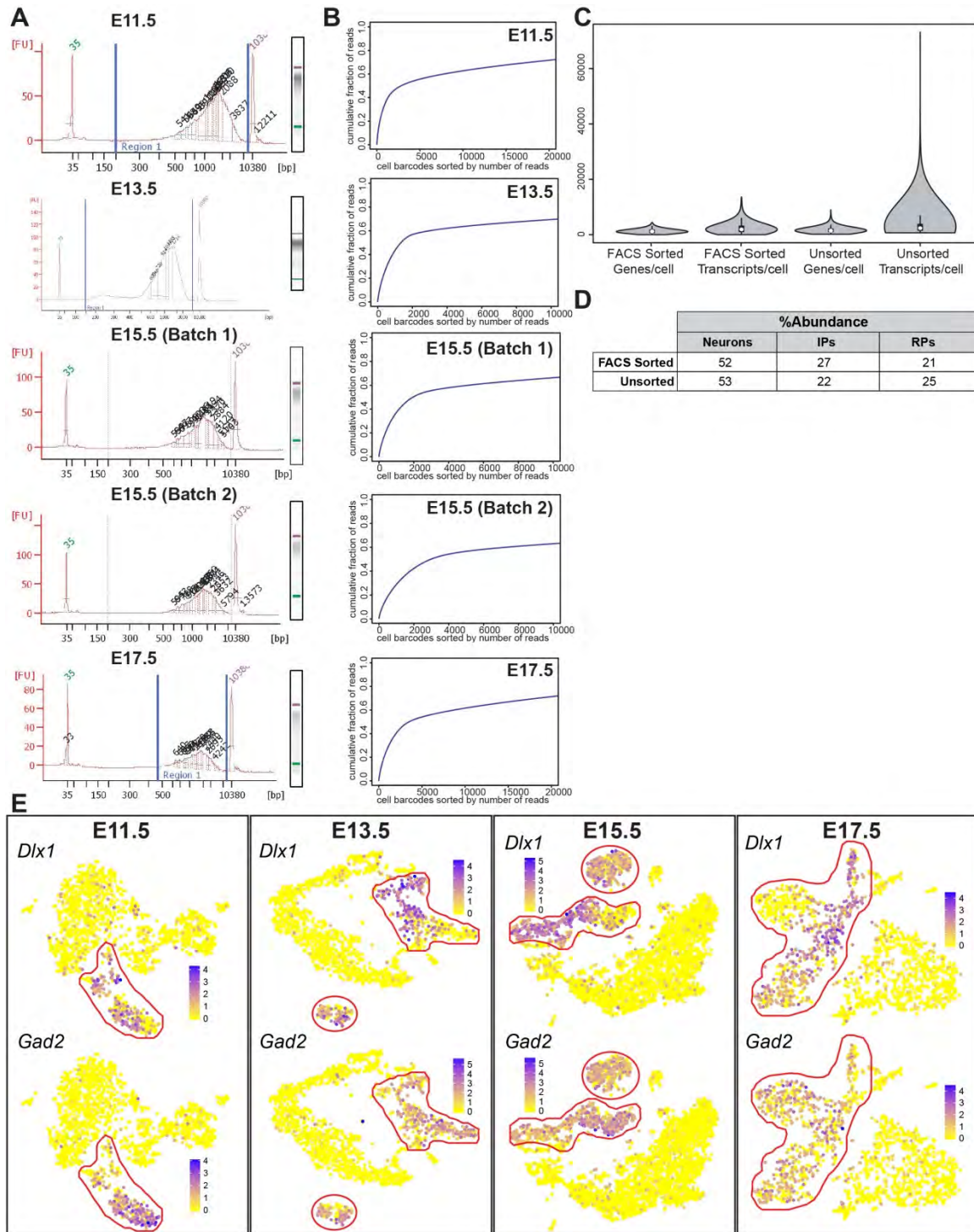
Supplemental Figure S1: *Drop-seq data quality characterization and identification of non-cortical cell populations.* (**Related to Fig. 1).** (**A**) BioAnalyzer traces of amplified cDNA from each embryonic age. Note that at E15.5 two batches of Drop-seq data were collected from the same sample of isolated cells. (**B**) 'Knee-plots' of the cumulative fraction of reads for each cell-barcode in decreasing order for each embryonic age. The inflection point of the 'knee-plot' was used to select the number of cell-barcodes for further analysis (2000 cell-barcodes at each of

E11.5, E13.5, E17.5, 2000 cell barcodes for E15.5 Batch 1 and 3000 cell-barcodes for E15.5 Batch 2). (**C**) Violin plots comparing the number of detected genes and transcripts per cell with cells that were isolated at E13.5 and either FAC sorted or not (unsorted). (**D**) Calculated proportions of RPs, IPs and neurons from both unsorted and FAC-sorted Drop-Seq runs at E13.5. (**E**) t-SNE visualizations of the expression of the GE marker gene *Dlx1* and the GE-derived interneuron marker *Gad2* in the E11.5, E13.5, E15.5 and E17.5 scRNAseq data. Cells in the visualizations are colour coded according to the expression of each gene in each individual cell ranging from not detected (in yellow) to the highest level of detection (in blue), according to the color key included in each individual panel. The clusters where most cells were positive (outlined in red) were removed from further analysis.
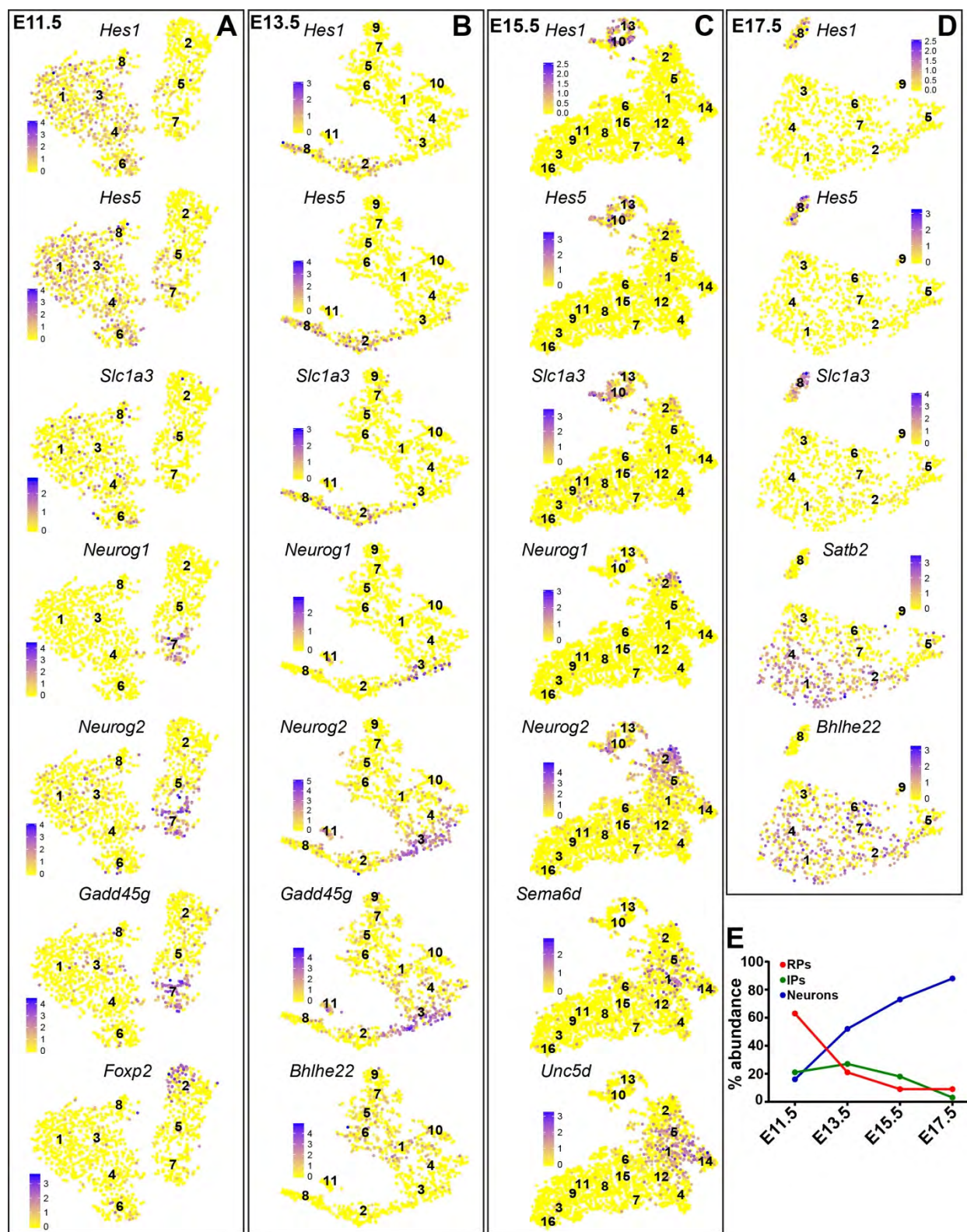
Supplemental Figure S2: *Expression of known marker genes in single cells of the embryonic cortex from E11.5 to E17.5* (**Related to Fig. 2**). (**A-D**) t-SNE visualizations of marker gene expression in scRNAseq data from E11.5 (**A**), E13.5 (**B**), E15.5 (**C**) and E17.5 (**D**). Shown are the RP marker genes *Hes1, Hes5* and *Slc1a3/Glast*, the IP

marker genes *Neurog1*, *Neurog2* and *Gadd45g*, the migratory neuron marker genes *Sema6d* and *Unc5d*, and the cortical neuron subtype marker genes *Foxp2*, *Bhlhe22* and *Satb2*. In all cases, cells are colour coded according to the expression of each gene in each individual cell ranging from not detected (in yellow) to the highest level of detection (in blue), according to the color key included in each individual panel. The cluster numbers as defined in Fig. 2A are overlaid for clarity. (**E**) Plot of the calculated percent abundance of RPs, IPs and neurons at all timepoints (E11.5, E13.5, E15.5, and E17.5).
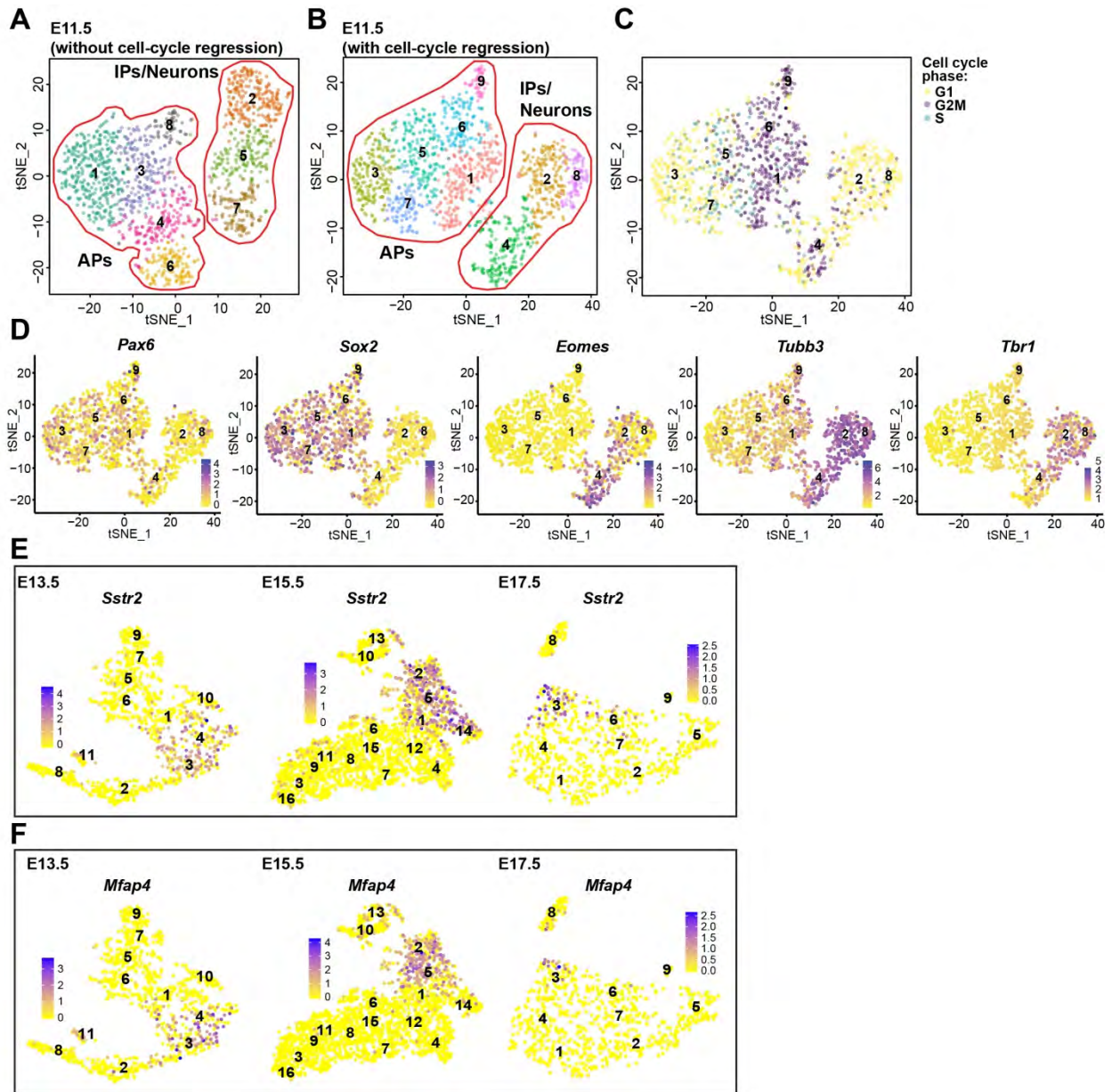
Supplemental Figure S3: *Cell cycle regression of E11.5 RPs and expression of new potential IP marker genes in single cells of the embryonic cortex during neurogenesis.* (**Related to Fig. 3 and 4**). (**A,B**) t-SNE visualization of E11.5 cell clusters with no cell cycle regression (**A**) and following cell cycle regression using computed Cyclone (Scialdone et al., 2015) cell cycle scores (**B**). (**C**) Cyclone analysis of E11.5 cell cycle regressed scRNAseq data, as visualized by t-SNE. Cells are colour coded according to their cell-cycle phases, with cells in yellow predicted to be in G1, those in light blue in S-phase, and those in dark blue/purple in G2/M phase. (**D**) t-SNE visualizations of cell cycle regressed scRNAseq data from E11.5 with overlaid expression of well characterized marker genes for RPs/APs (*Pax6, Sox2*), IPs (*Eomes/Tbr2*), newborn neurons (*Tubb3*) and a subpopulation of early-born cortical neurons (*Tbr1*). Cells in the t-SNE visualizations are colour coded according to the expression of each gene in each individual cell from not detected (in yellow) to the highest level of detection (in blue) according to the color key

included in each individual panel. (**E,F**) t-SNE visualizations of the potential IP marker genes *Sstr2* (**E**) and *Mfap4* (**F**) in scRNAseq data from E13.5 to E17.5. Cells in the t-SNE visualizations are colour coded according to the expression of each gene in each individual cell from not detected (in yellow) to the highest level of detection (in blue) according to the color key included in each individual panel. The cluster numbers as defined in Fig. 2A are overlaid for clarity.
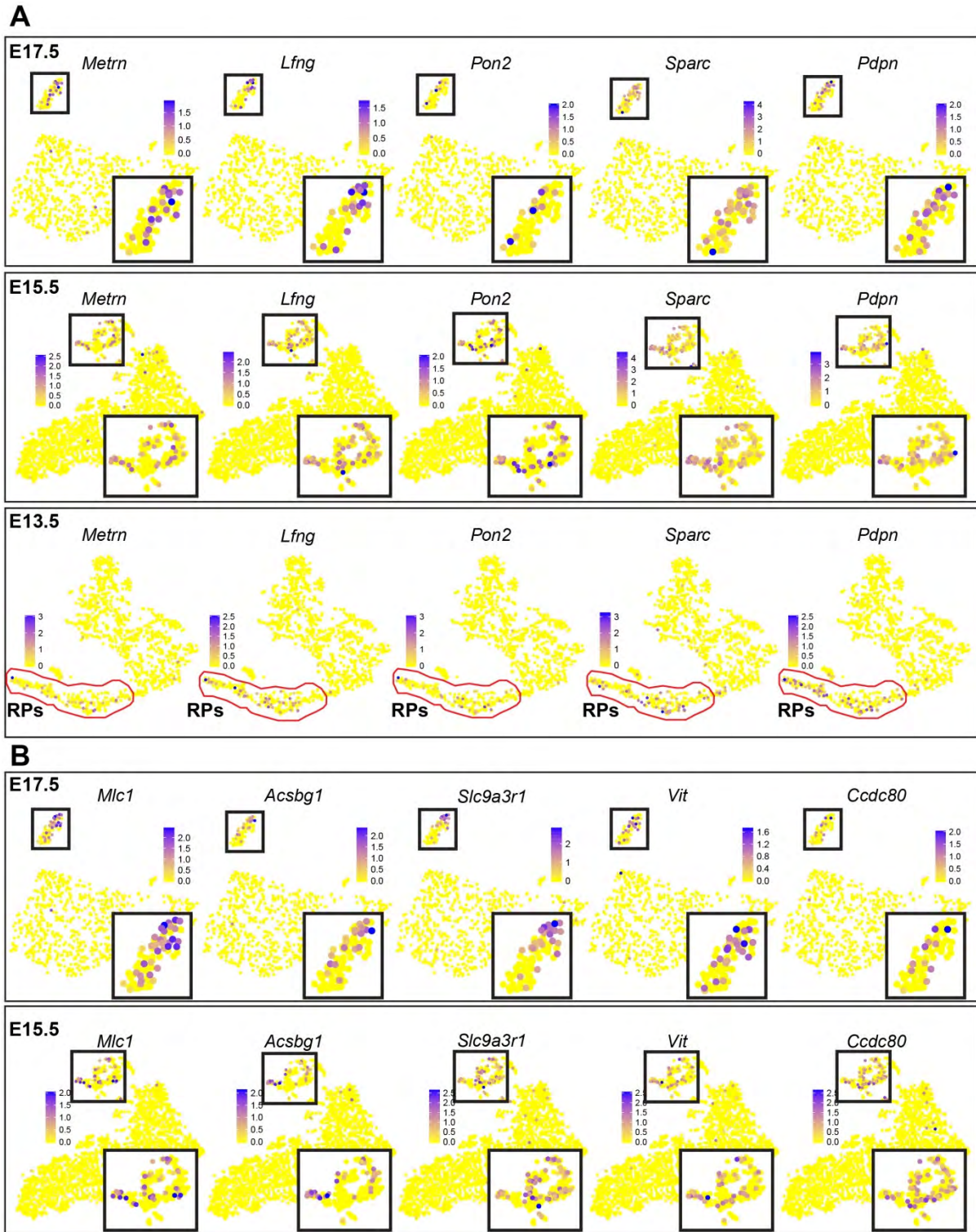
Supplemental Figure S4: *Expression of embryonic RP core identity genes.* (**Related to Fig. 5**). (**A**) t-SNE visualizations of some of the embryonic RP core identity genes, *Metrn*, *Lfng*, *Pon2*, *Sparc* and *Pdpn,* that were differentially enriched in the RP clusters versus all other cell clusters at E13.5, E15.5 and E17.5. For E15.5 and E17.5, the boxes identify the RP cell clusters, which are shown at higher resolution in the bottom right corner of

each panel. For the E13.5 data, the RP cell clusters are outlined in red. (**B**) t-SNE visualizations of some of the RP core identity genes, *Mlc1*, *Acsbg1*, *Slc9a3r1*, *Vit* and *Ccdc80,* that were differentially enriched in the RP clusters versus all other cell clusters at E15.5 and E17.5, but not E13.5. The boxes identify the RP cell clusters, which are shown at higher resolution in the bottom right corner of each panel. In all cases, cells in the t-SNE visualizations are colour coded according to the expression of each gene in each individual cell from not detected (in yellow) to the highest level of detection (in blue) according to the color key included in each individual panel.
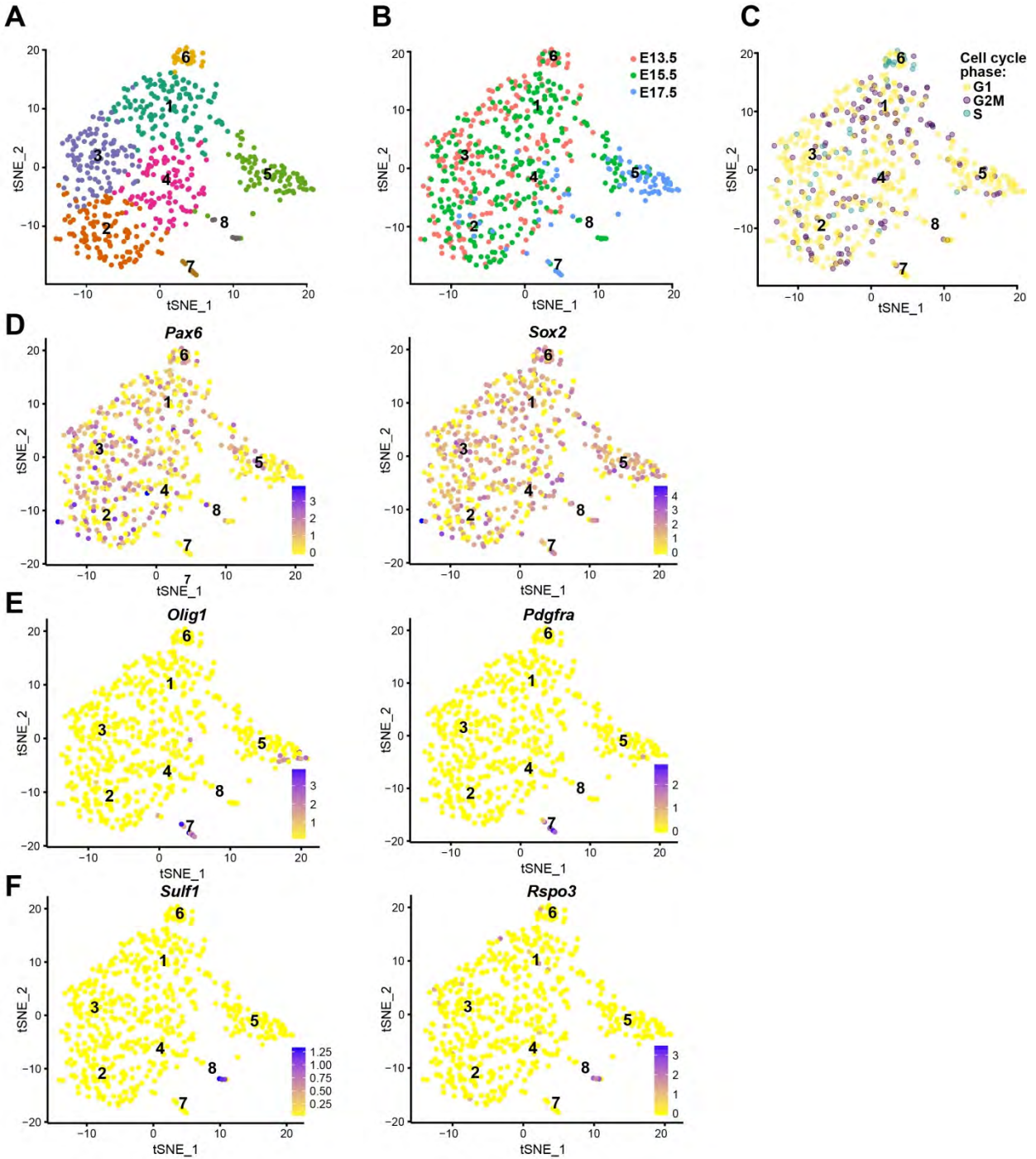
**Supplemental Figure S5**: *Combined analysis of E13.5, E15.5 and E17.5 RP populations after cell cycle regression* (**Related to Fig. 5**). (**A**) t-SNE visualizations of cell clusters from combined E13.5, E15.5 and E17.5 RP populations following cell cycle regression using computed Cyclone (Scialdone et al., 2015) cell cycle scores, with colors and the associated numbers representing individual cell clusters. (**B**) t-SNE visualizations of cell clusters from combined E13.5, E15.5 and E17.5 RP populations, with colors representing the embryonic age at the time of data collection. (**C**) Cyclone cell-cycle analysis of combined E13.5, E15.5 and E17.5 RP clusters, as visualized by t-SNE.

Cells are colour coded according to the predicted cell-cycle phases, with cells in yellow predicted to be in G1, those in light blue in S-phase, and those in dark blue/purple in G2/M phase. (**D-F**) t-SNE visualizations of the RP marker genes *Pax6* and *Sox2* (**D**), the oligodendrocyte progenitor cell marker genes *Olig1* and *Pdgfra* (**E**) and genes specific to cluster 8, *Sulf1* and *Rspo3* (**F**). In all cases, cells in the t-SNE visualizations are colour coded according to the expression of each gene in each individual cell from not detected (in yellow) to the highest level of detection (in blue) according to the color key included in each individual panel.
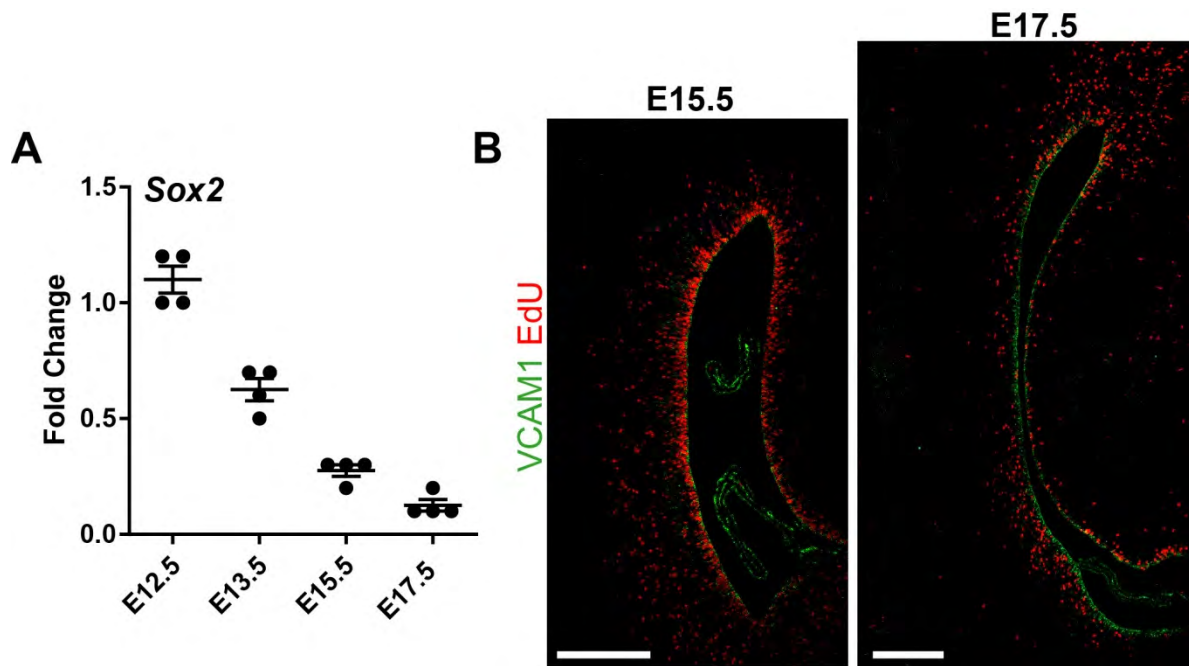
Supplemental Figure S6: *Analysis of expression of Sox2 mRNA and RP proliferation in the embryonic cortex.* **(Related to Fig. 6).** (**A**) qRT-PCR analysis of *Sox2* mRNA in E12.5, E13.5, E15.5 and E17.5 cortices. *Sox2* expression was measured relative to the reference gene *Rps18* and is shown as a scatter-plot, indicating the mean +/- SEM. n = 4 independent cortical samples at each time point. (**B**) E15.5 or E17.5 embryos were exposed to maternally-injected EdU, and coronal cortical sections from these embryos were immunostained 6 hours later for EdU (red) and Vcam1 (green). Shown are low-magnification confocal images of the cortex (Ctx) adjacent to the lateral ventricles (LV). Scale bars = 200 µm.
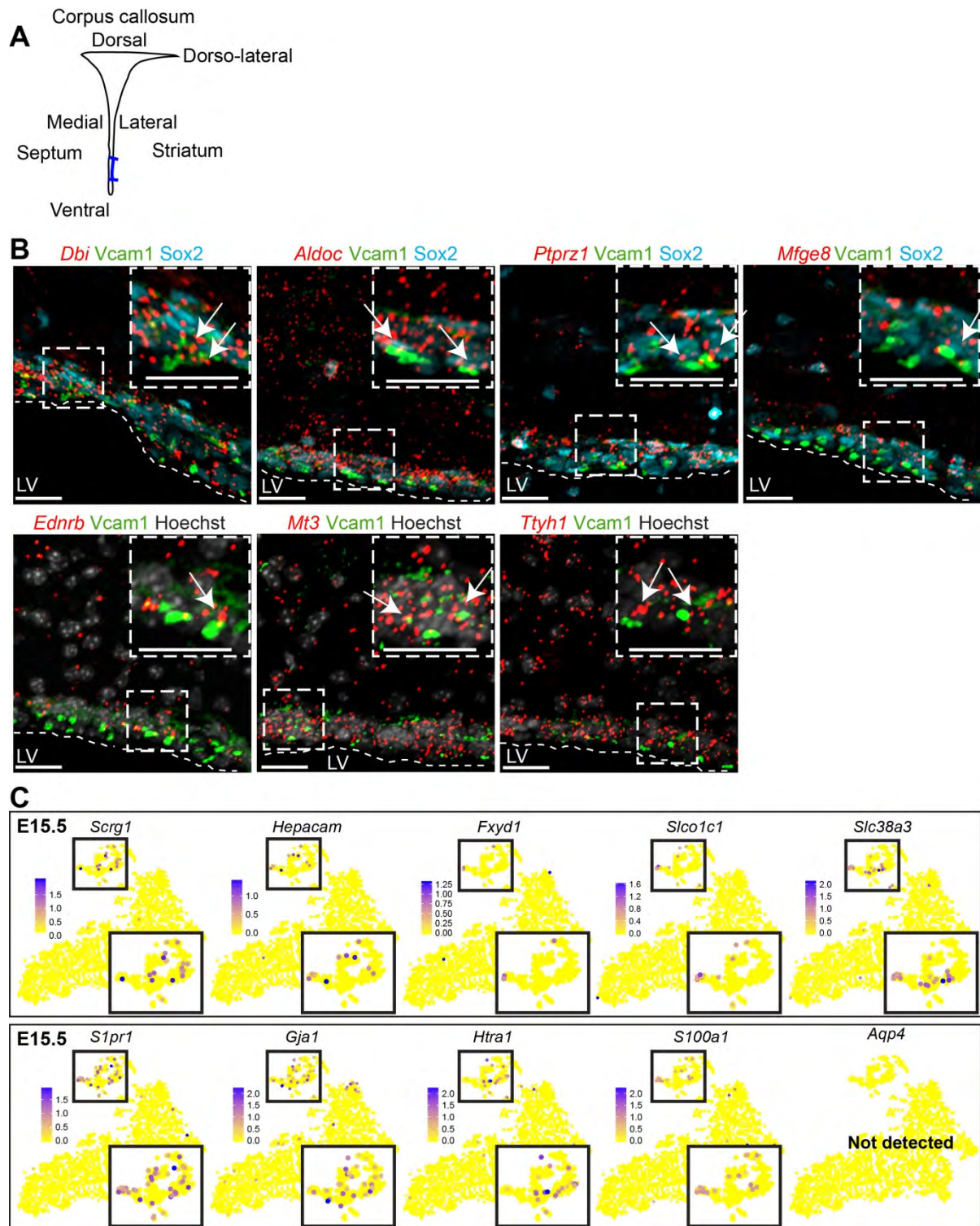
Supplemental Figure S7: *Expression of genes shared between quiescent adult forebrain NSCs and embryonic RPs.* **(Related to Fig. 7).** (**A**) Schematic diagram of the V-SVZ area of the adult mouse forebrain, highlighting in blue the GE-derived region that was analyzed by FISH. (**B**) Confocal z-stack images of coronal sections through the adult GE-derived V-SVZ analyzed by single-molecule FISH for *Dbi*, *Aldoc*, *Ptprz1*, *Mfge8*, *Ednrb*, *Mt3* and *Ttyh1*

mRNAs (all in red) together with immunostaining for Vcam1 (in green). Sections in the top panels were also immunostained for Sox2 (in blue) while those in the bottom panels were counterstained with Hoechst 33258 (in light grey). All images show the region highlighted in (A), with the borders of the lateral ventricles indicated by the hatched white lines. The boxed regions are shown at higher magnification in the top right-hand corners of each panel, and arrows denote triple- or double-labelled cells as relevant. Scale bars represent 23 µm. (**C**) t-SNE visualizations of the expression of *Scrg1*, *Hepacam*, *Fxyd1*, *Slco1c1*, *Slc38a3, S1pr1, Gja1, Htra1, S100a1,* and *Aqp4* mRNAs in the E15.5 scRNAseq data. Cells in the t-SNE visualizations are colour coded according to the expression of each gene in each individual cell from not detected (in yellow) to the highest level of detection (in blue) according to the color key included in each individual panel. The boxed regions highlight the RP clusters and are shown at higher resolution in the bottom right corner of each panel.

E11.5

cluster #	cell type	% abundance	marker genes
1	Apical precursors	23	Sox2, Pax6, Hes1, Hes5, Pcna, Ung
2	Newborn neurons	16	Eomes, Tubb3, Tbr1, Foxp2, Reln
3	Apical precursors	14.5	Sox2, Pax6, Hes1, Hes5, Pcna, Ung
4	Apical precursors	12	Sox2, Pax6, Hes1, Hes5, Prc1, Cenpf, Ube2c
5	IPs/Newborn neurons	11	Eomes, Tubb3, Tbr1
6	Apical precursors	10	Sox2, Pax6, Hes1, Hes5, Prc1, Cenpf, Ube2c
7	IPs	9.5	Eomes, Tubb3, Gadd45g, Ngn1, Ngn2
8	Apical precursors	4	Sox2, Pax6, Hes1, Hes5...

E13.5

cluster #	cell type	% abundance	marker genes
1	Excitatory neurons	13.5	Tbr1, Tubb3, Satb2, Bhlhe22
2	Non-proliferative RPs	12.5	Sox2, Pax6, Hes1, Hes5, Slc1a3
3	Apical IPs	12	Eomes, Gadd45g, Ngn1, Ngn2, Btg2, Sstr2, Mfap4
4	Basal IPs	12	Eomes, Neurod1, Pam, Slc17a6, Sstr2, Mfap4
5	Excitatory neurons	10	Tbr1, Tubb3, Bhlhe22
6	Excitatory neurons	9	Tbr1, Tubb3, Satb2, Bhlhe22
7	Excitatory neurons	8.5	Tbr1, Tubb3, Bhlhe22
8	Proliferative RPs	8	Sox2, Pax6, Hes1, Hes5, Slc1a3, Mki67
9	Excitatory neurons	6	Tbr1, Tubb3, Bhlhe22
10	Excitatory neurons	5.5	Tbr1, Tubb3, Satb2, Bhlhe22
11	Proliferative IPs	3	Eomes, Gadd45g, Ngn1, Ngn2, Mki67, Sstr2, Mfap4

E15.5

cluster #	cell type	% abundance	marker genes
1	Excitatory neurons	9.5	Tbr1, Tubb3, Satb2, Bhlhe22, Unc5d, Sema6d
2	Apical IPs	9	Eomes, Ngn1, Ngn2, Gadd45g, Sstr2, Mfap4
3	Excitatory neurons	8.5	Tbr1, Tubb3
4	Excitatory neurons	7.5	Tbr1, Tubb3, Satb2, Bhlhe22
5	Basal IPs	7	Eomes, Neurod1, Sstr2, Mfap4
6	Excitatory neurons	7	Tbr1, Tubb3, Satb2, Bhlhe22
7	Excitatory neurons	6	Tbr1, Tubb3, Satb2, Bhlhe22
8	Excitatory neurons	6	Tbr1, Tubb3, Bhlhe22
9	Excitatory neurons	6	Tbr1, Tubb3, Bhlhe22
10	Non-proliferative RPs	6	Sox2, Pax6, Hes1, Hes5, Slc1a3
11	Excitatory neurons	5	Tbr1, Tubb3, Bhlhe22
12	Excitatory neurons	5	Tbr1, Tubb3, Satb2, Bhlhe22
13	Proliferative RPs/IPs	5	Sox2, Pax6, Hes1, Hes5, Slc1a3, Mki67, Eomes
14	Excitatory neurons	4.5	Tbr1, Tubb3, Bhlhe22, Sema6d, Unc5d
15	Excitatory neurons	4	Tbr1, Tubb3, Satb2, Bhlhe22
16	Excitatory neurons	4	Tbr1, Tubb3, Satb2

E17.5

cluster #	cell type	% abundance	marker genes
1	Excitatory neurons	19	Tbr1, Tubb3, Satb2, Bhlhe22
2	Excitatory neurons	14	Tubb3, Satb2, Bhlhe22
3	IPs/Excitatory neurons	14	Eomes, Sstr2, Mfap4, Unc5d, Sema3c, Tubb3, Tbr1
4	Excitatory neurons	13	Tbr1, Tubb3, Satb2, Bhlhe22
5	Excitatory neurons	10	Tbr1, Tubb3, Bhlhe22
6	Excitatory neurons	10	Tbr1, Tubb3, Bhlhe22
7	Excitatory neurons	9.5	Tbr1, Tubb3, Satb2, Bhlhe22
8	Non-proliferative RPs	9	Sox2, Pax6, Hes1, Hes5, Slc1a3
9	Cajal-Retzius neurons	1.5	Tbr1, Tubb3, Reln

Supplemental Table 1. *Summary tables of t-SNE cluster plots shown in Figure 2A.* (**Related to Fig. 2**). Shown are annotation tables for the t-SNE visualizations of the cortically-derived cell clusters from each embryonic age.

Annotations of each cluster are shown along with the relative abundance of each cluster and specific marker genes for each cluster.