Heterogeneity of Human Neuroepithelial Cells and Early Radial Glia

Ugomma C. Eze[1,2]*, Aparna Bhaduri[1,2*#], Tomasz J. Nowakowski[1,3], Arnold R. Kriegstein[1,2#]

1. Department of Neurology, University of California, San Francisco (UCSF), San Francisco, CA, USA.
2. The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco (UCSF), San Francisco, CA, USA.
3. Department of Anatomy, University of California, San Francisco (UCSF), San Francisco, CA, USA.

*These authors contributed equally
[#]Correspondence should be addressed to Aparna.Bhaduri@ucsf.edu or Arnold.Kriegstein@ucsf.edu

Abstract: The human cortex is comprised of diverse cell types that emerge from an initially uniform neuroepithelium that, following neural tube closure, gives rise to radial glia, the neural stem cells of the cortex. Radial glia initially reside in the ventricular zone of the cortex, and contribute to cortical expansion which is particularly pronounced in the human compared to other mammals and non-human primates. To characterize the molecular signatures of cellular subtypes that may exist at the earliest stages of neurogenesis we performed single-cell RNA-sequencing across regions of the developing human brain. We observe similar progenitor programs across brain regions that each express region-specific transcription factors. In the telencephalon, we identify nine progenitor populations, suggesting more heterogeneity among neuroepithelial cells and radial glia than previously described, including a highly prevalent mesenchymal-like population that disappears after the onset of neurogenesis. Using velocity analysis, we find that genes implicated in various neurodevelopmental diseases drive the specification and maintenance of neuroepithelial cells and radial glia. Comparison of these progenitor populations to corresponding stages of mouse development identifies two progenitor clusters that are unique to the early stages of human cortical development. Organoid systems display a low fidelity to neuroepithelial and early radial glia cell types, but this improves as neurogenesis progresses. Overall, we provide a comprehensive molecular and spatial atlas of early stages of human brain and cortical development.

The human brain consists of billions of cells across several functionally interconnected structures that emerge from the neuroectoderm. Many of these structures are substantially expanded or distinct compared to other mammals[1], particularly the cerebral cortex, the outermost layer of the human brain responsible for perception and cognition. These differences emerge at developmental stages prior to birth[2], and thus exploring the cell types in the developing human brain is essential to better characterize how cell types across the brain are generated, how they may be impacted during the emergence of neurodevelopmental disorders, and how human neural stem cells can be directed to specific cell types for modeling or treatment purposes.

The brain exponentially increases in size after the neural tube closes[3,4]. Later in development, across brain regions, a series of similar neurogenic and gliogenic processes gives rise to the constituent cell types[5]. However, at the molecular level, the sequence of events that leads to the emergence of these progenitor cells early in development is much less well understood. In order to identify cell types and trajectories that lay the foundation for the development of the human brain, we performed single-cell RNA sequencing (scRNA-seq) using the droplet-based 10X Genomics Chromium platform on ten individuals during the first trimester of human development, spanning Carnegie Stages (CS) 12 to 22, which corresponds to gestational weeks (GW) 6 – 10[6]. In order to identify the major cell populations across brain regions and to compare them to one another, we sampled all available and identifiable structures, including the telencephalon, diencephalon, midbrain, hindbrain, cerebellum, ganglionic eminences, thalamus, hypothalamus, and cortex (SFig 1 – 2, STable 1). In total, we collected 289,000 cells passing quality control.

The diversity of brain structures is known to emerge as a result of segmentation events that generate the prosencephalon, mesencephalon, and rhombencephalon which are then further specified into the anatomical structures we dissected[7,8] (SFig 1 A-B). It has been proposed that there are core gene regulatory programs that enable the specification of these regions and subsequent development of relevant cells[9], and we sought to explore if our data could more comprehensively define these signatures and also identify cell type-specific similarities and differences in these nascent brain structures. Across brain regions, hierarchical analysis initially identified two major cell classes – progenitors and neurons. To identify the developmental region-specific gene signatures for each brain area, including the hindbrain, midbrain, thalamus, ganglionic

eminences, retina, and cortex, we performed differential gene expression across areas at each age (STable 2 – 4). Using the most developmentally-advanced samples at CS22, we generated data-driven regional signatures and explored when the emergence of the most structure-specific genes emerged. We observed that characteristic transcription factor expression (such as *HOX* genes [hindbrain][10]; *PAX7* [midbrain][11,12]; *GBX2* [thalamus][13]; *NKX2.1* [medial ganglionic eminence][14,15]; and *FOXG1* [cerebral cortex][16]) segregated these regions from one another as early as CS13. However, the cell type-defining gene expression programs were largely conserved across brain regions, resulting in subsets of progenitors or neurons that transcriptomically appear similar across brain regions (SFig 2C). At the earliest timepoint, CS12, the differences between brain regions were minimal and did not resemble the more advanced region-specific programs that were identifiable later in the first trimester (SFig 1 – 2).

The human cerebral cortex is more than three times expanded compared to our closest non-human primate relatives[17]. The cortex emerges from an initially pseudostratified neuroepithelium that gives rise to radial glia, the neural stem cells of the cortex[3]. Radial glia give rise to neurons, initially through direct neurogenesis, and then indirectly through transit-amplifying intermediate progenitor cells[18,19]. A number of subtypes of radial glia have been identified as the cortex matures[20,21], and their primary role in neurogenesis declines late in the second trimester, at which point they generate the glial populations of the cortex[22]. scRNA-seq has added substantially to our prior knowledge of these processes, particularly during stages of peak neurogenesis. However, the first trimester stage of cortical development has not been described at this level of molecular detail, and important questions remain about the timing of neurogenesis, the presumed uniformity of the neuroepithelium, and the signals that promote the transition to radial glia.

To better understand these early stages of cortical development, we focused on the 59,000 cells in our dataset that originated from the telencephalon (earliest dissections) and the cortex specifically (when it was identifiable for subdissection). Clustering this data revealed minimal batch effects, and segregated the samples based on early and late first trimester stages (Figure 1, SFig 3). Each of the 63 identified clusters could be assigned to a cell type identity of either neuroepithelial cells, radial glia, intermediate progenitor cells (IPCs), neurons, or mesenchymal-like cells. Additionally, we noted “other” cell populations of endothelial cells, microglia, and pericytes (STable 5). We observed that although neurogenesis ramps up by the end of the second trimester, even the earliest samples contained a small number of neurons (Fig 1A), presumably produced by direct neurogenesis. IPCs that give rise to neurons via indirect neurogenesis are nearly absent until later in the first trimester (after CS16). Additionally, we observed that the percentage of neuroepithelial cells declines over time as radial glia increase, and also that the mesenchymal cell population diminishes strikingly over the course of the first trimester.

Sub-clustering of the neuronal populations resulted in subtype clusters, some of which were strongly enriched in either younger samples (< CS16, presumed to result from direct neurogenesis) or older samples (> CS16) (SFig 4, STable 6). These included expected differences such as clusters marked by *NHLH1*, which was significantly higher in the older samples, and has been associated with newborn neurons derived from IPCs[23]. Other genes significantly enriched in older samples included *NEUROD6* and *BCL11B* which are associated with neuronal and deep layer identity[24,25], as well as *CALB2* which is expressed in migratory ventrally-derived interneurons and a subset of excitatory neurons[26]. We were surprised to note significantly higher expression of *MEF2C* in the younger samples, as it has primarily been described as a factor involved in synaptic maturation[27,28] and would be predicted to be activated at later timepoints, though it has also been identified as a regulator of early neuronal differentiation and layer formation[29]. The co-expression of these neuronal subgroups with *SATB2* suggests that there may exist functionally uncharacterized populations of very early-born neurons, likely via direct neurogenesis. We also observed that some clusters were marked by previously undescribed genes, including the younger sample-specific *LHX5-AS1* cluster. The remaining clusters consisted of expected early-born neuronal populations including Cajal-Retzius cells marked by *RELN*[30], and subplate cells marked by *TLE4* and *NR4A2*[31]. The subplate clusters were unexpectedly heterogeneous (SFig 4B) and contained marker genes not previously associated with subplate identity (STable 6), suggesting our analysis may provide additional characterization of early neuronal cell type gene expression patterns.

In addition to our transcriptional characterization of cell types across the first trimester of brain development, we also sought to describe the spatial organization of these populations across developmental timepoints. Thus, we performed immunostaining validation across five individuals during first trimester timepoints, and also in an early second trimester sample at GW14 for comparison. In these samples, we observed FOXG1 staining as early as CS16, but not earlier, as has been previously described[32] (SFig 5); this staining ensured that our identification of cortical cell types was in anatomically-relevant adjacent sections. In

concordance with our transcriptomic data, we observed prevalent SOX2 staining as early as CS13, and this staining persisted through GW14. Newborn neurons marked by DCX were identifiable as early as CS14, but markers of maturing neuronal identity such as BCL11B (CTIP2) emerged after CS16. As expected, IPCs identified by TBR2 staining did not emerge until CS18 (Fig 1B, SFig 6).

In order to characterize the lineage relationship between major cortical cell types, we performed RNA-velocity analysis[33] using the scVelo implementation[34]. This algorithm incorporates mRNA levels and inherent expression variability of individual cells to infer steady states that contribute to potential lineage relationships. Across cortical cell types, the most apparent lineage stream originated in progenitor cell populations from older samples, and followed a trajectory through IPCs to neurons (Fig 1C). This trajectory of neurogenesis is well described, and suggests that the stereotypical process of neuronal differentiation emerges after CS19. At earlier time points, there were local examples of radial glia giving rise to neurons, suggestive of direct neurogenesis, and also lineage relationships between the progenitors themselves.

The transition from neuroepithelial cells to radial glia has traditionally been characterized by the reorganization of tight junctions and the appearance of nestin RC2 immunoreactivity[35-37]. We sought to visualize this process and its transition across our samples. Using nestin as a marker of radial glia, ZO-1 as an indicator of neuroepithelial cells, and SOX2 as a label for all progenitor populations, we saw small numbers of nestin-positive radial glia at CS14, with substantial upregulation by CS22 that simultaneously corresponded to a waning of ZO-1 (Fig 2A). By GW14, ZO-1 staining dissipates and is largely constrained to presumed vascular structures. Across a timecourse of first trimester development, we find a progressive, and incremental, shift away from neuroepithelial identity (SFig 7). In order to better understand the heterogeneity and spatiotemporal trajectories of the neuroepithelial and radial glia progenitor populations within the dataset, we sub-clustered the progenitors. Because the mesenchymal population also expressed strong progenitor cell markers including *VIM* and *SOX2*, we removed all neurons, IPCs, and other cell populations (including microglia, pericytes, and endothelial cells) and subclustered the remaining cells. The resulting analysis yielded nine sub-clusters all marked by strong *VIM* and *SOX2* expression (Fig 2B). Each of the nine clusters contained some cells from all first trimester age samples, suggesting that the velocity and maturation trajectories were not purely a function of age (SFig 8, STable 7). Velocity analysis identified a strong gradient across the populations from cluster 2 towards clusters 1 and 6, and across all velocity analyses in this study, we characterized the genes that most strongly influence the velocity stream (and thus would be hypothesized to regulate cell fate transition the most). These drivers of velocity were enriched for genes that have been associated with neurodevelopmental or psychiatric disorders, including autism, cortical malformations, and schizophrenia (STable 8), indicating that very early events in cortical development may result in susceptibility or vulnerability to these disorders. More work at these early timepoints is required to fully characterize the developmental implications.

In order to explore the timing and spatial localization of each of the progenitor subpopulations, we performed immunostaining of the most specific gene markers for each cluster (SFigs 9 – 16). *NTRK3* was a highly specific marker of Progenitor Cluster 3, and intrigued us because it has been most commonly described as the TrkC receptor that enables survival of specific neuronal populations[38], and its broader neocortical expression has been described at later stages of mouse cortex development[39]. We observed that at early stages (CS14, CS16), *NTRK3* broadly co-localized with SOX2-positive progenitor cell populations (Fig 2C), but after CS18, it shifted from progenitor to neuronal expression. *DLK1* was a top marker gene for Progneitor Cluster 8 and also was highly enriched in early samples. We observed a small number of DLK1-positive cells that were lowly SOX2-positive at early stages such as CS16, after which its expression completely disappeared (Fig 2D, SFig 10). DLK1 is a non-canonical ligand of Notch that inhibits its function and impacts Notch1 receptor distribution[40,41], though it has previously been noted to not be widely expressed in the mammalian central nervous system during development[42]. DLK1 has been described to play a role in enabling postnatal neurogenesis in the mouse[43], and may play a similar regulatory role at the earliest neurogenic timepoints of human cortical development. Across progenitor populations, we describe known and new gene expression patterns that define the neuroepithelial and radial glia subpopulations, which suggests there is an expression gradient that transitions from early to more mature progenitor populations.

Both in our initial clustering and in the more focused progenitor analysis, a mesenchymal-like population labelled by the gene *LUM* was distinctly segregated from other progenitor populations. Lumican, the protein encoded by *LUM*, is an extracellular matrix protein that is widely present in mesenchymal tissues throughout the adult body[44], though it has been used to promote folding in cortical structures when added exogenously to media[45]. In our data, *LUM* expression is highly enriched in the mesenchymal cell population and diminishes substantially after CS15 (Fig 3A). Another marker of these clusters was *ALX1*, which also has one of the most correlated gene expression patterns to *LUM* (Fig 3B). ALX1 has been described as required in

knockout mice for the development of the forebrain mesenchyme[46], but has not been studied in humans. We observed prevalent LUM staining in our samples at or prior to CS16, with ALX1 at the edges of the forebrain, including several co-localized PAX6-positive cells (Fig 3 C-D). The ALX1 staining significantly diminishes later on in development. We hypothesize that the ALX1-positive cells may resemble previously described neural crest-derived cell populations that give rise to the meninges[47]; these cells may be secreting LUM as structural components that support the physical formation of the developing telencephalon prior to emergence of the radial glial scaffold that characterizes the peak stages of neurogenesis.

The molecular characterization of early human cortical development that we provide here enables the identification of major cell types and modeling of their trajectories, but ultimately mechanistically characterizing the events that maintain the neuroepithelial populations and that promote the transition to radial glia will require the use of *in vivo* or *in vitro* models. In order to evaluate the similarities and differences between these early developmental stages in human and rodent, we performed scRNA-seq of the mouse forebrain at Theiler Stages (TS) 14 and 16, which corresponds to embryonic day (E) 9 and 10. Although the onset of neurogenesis differs between human and mouse[48], we used immunostaining to verify that at these mouse stages, the forebrain was already expressing FOXG1 and undergoing neurogenesis, as marked by the presence of DCX (SFig 5). To compare the mouse data with human progenitor populations, we performed clustering and correlation analysis between the mouse clusters and our nine human progenitor populations. Using a correlative threshold, we observed that seven of the nine human populations had at least one corresponding cluster in the mouse single-cell data (Fig 4A). However, there were no high correlations for Progenitor Cluster 4 (marked most highly by *C1orf61*) or Progenitor Cluster 7 (marked most highly by *ID4*). To test the hypothesis that these populations were underrepresented in the mouse data, we used immunostaining to explore the expression of the major progenitor population markers in the mouse. Indeed, no expression of C1orf61 or ID4 was identifiable in the mouse forebrain at TS14 or 16 (Fig 4B, SFig 15, SFig 16).

C1orf61 (CROC-4) has been characterized to be widely expressed during late developmental stages and in the adult brain, regulating c-FOS signaling[49], but is otherwise uncharacterized. In contrast, ID4 has been identified in the developing mouse cortex at later stages (E15.5). ID4-deficient mice exhibit impaired brain size and mistimed neurogenesis[50,51], suggesting that earlier expression in human progenitors compared to mouse could indicate a potential mechanism by which early progenitors expand more rapidly in humans than in mice.

Despite the general conservation of cell types between the early development of the human and mouse cortex, the lack of specific populations at early developmental stages suggests that alternatives are required to fully model the earliest cell types involved in human cortical development. Cortical organoids are an attractive model because they can be generated from normal or patient-derived induced pluripotent stem cells (iPSCs) and are amenable to genetic modulation[52,53]. We previously generated an extensive catalog of single-cell sequencing across timepoints (week 3 to 24), four pluripotent stem cell lines, and three methods of generating cortical organoids[54]. Using the 108,593 cells derived from the week 3 and week 5 timepoints of this dataset, we explored the fidelity of the cell types in early human cortex to the organoid system. The preservation of excitatory neuron identity mirrored that of later developmental stages (~0.5, as we recently reported). However, the remaining primary cell types had much lower fidelity in the closest organoid population (average of 0.28) (Fig 4C). These data suggest that early cell types in organoids do not resemble their corresponding cortical progenitor counterparts, but that the fidelity of these cell types does improve at later stages of organoid culture. Interestingly, the two populations that were not well represented in mice at TS14 and 16 were identifiable in organoid cultures as evidenced by immunostaining for ID4 and C1orf61 (SFig 15, SFig 16). Moreover, the populations that were not present in the organoid, including the *LUM* and *ALX1* clusters, could be identified in the mice (SFig 13, SFig 14), indicating that depending on the key questions being investigated, either *in vivo* or *in vitro* models may be most appropriate for the study of neuroepithelial and early radial glia populations. Our description of early human cortical cell populations may also enable us to further refine *in vitro* culture systems to better reflect these neuroepithelial and radial glia populations. Together, the data we present here represent a characterization of major cell types across the first trimester of human brain development, and highlight the subpopulations of progenitor cells that form the basis for creating the human cortex.

Acknowledgments

The authors thank S. Wang, W. Walantus, M.G. Andrews, G. Wilkins, L. Subramanian, and members of the A.R.K. laboratory for providing resources, technical help and helpful discussions. Single-cell RNA sequencing data has been deposited at the NeMO archive under dbGAP restricted access. All primary tissue was obtained

from the Human Developmental Biology Resource (HDBR), with special thanks to S. Lisgo and M. Crosier. This study was supported by NIH award U01MH114825 A.R.K., and F32NS103266, K99NS111731, and L'Oreal For Women in Science Award through the AAAS to A.B. A.B. and A.R.K. designed the study and analysis. Experiments were performed by U.C.E., A.B., and T.J.N. Data analysis was performed by U.C.E., and A.B. The study was supervised by A.B. and A.R.K. This manuscript was prepared by A.B. and U.C.E. with input from all authors.

Methods

Sample Processing

Acquisition of all primary human tissue samples was approved by the UCSF Human Gamete, Embryo and Stem Cell Research Committee (GESCR, approval 10-03379 and 10-05113). All experiments were performed in accordance with protocol guidelines. Informed consent was obtained before sample collection and use for this study. First-trimester human samples were collected from elective pregnancy terminations through the Human Developmental Biology Resource (HDBR), staged using crown-rump length (CRL) and shipped overnight on ice in Rosewell Park Memorial Institute (RPMI) media or in 4% PFA. Fixed samples were used for downstream immunostaining and imaging. Mouse samples were sacrificed at E9 and E10, and staged using the somite number. Half of the mouse samples were fixed in 4% PFA and the remaining live samples were used for dissociation. Live samples were subdissected into identifiable regions and dissociated using papain (Worthington) with DNase. Samples were minced and incubated in 1ml activated papain for 15 min at 37°C, according to the manufacturer's instructions. Samples were then inverted for several times and incubated for an additional 15 min. The dissociated cells were centrifuged at 300 x *g* for 5 min and the papain was removed.

Single-cell RNA Sequencing

Single-cell capture was performed following the 10X v2 Chromium manufacturer's instructions. Each sample was its own batch. For each batch, 10,000 cells were targeted for capture and 12 cycles of amplification for each of the cDNA and library amplifications were performed. Libraries were sequenced according to the manufacturer's instructions on a NovaSeq S2 flow cell.

Single-cell RNA Analysis

Single-cell RNA-sequencing data was aligned to the GRCh38-0.1.2 reference genome, and cells were identified using CellRanger v2. Quality control removed cells with fewer than 500 genes per cell and cells with greater than 10% mitochondrial content. Clustering and batch effect was performed as has been previously described[55-57]. Briefly, Louvain-Jaccard clustering was performed, and cluster markers were identified using default parameters. Batch effects (as pertaining to day of capture) were removed by normalizing all cells in a batch to the most counts, and then multiplying by the median counts in that batch. Batch was regressed out in the space of variable genes during scaling. Clustering with batch effect correction was performed for each dissected area, and also without batch correction by each individual both across all brain regions and within cortex only.

Cluster markers were interpreted and assigned cluster identity by using known literature cell type annotations, or by associating progenitor and neuronal genes to other identifiable details. Progenitors were distinguished as radial glia if they expressed neurogenic genes, else they were determined to be neuroepithelial.

Subclustering within the neurons was performed by selecting and clustering the cells from neuronal clusters and repeating the clustering procedure with batch effect correction. Progenitor cells were subclustered by removing neuronal, IPC, and other (microglia, endothelial, and pericyte) clusters from the data. After subclustering, neuronal populations were identified and removed iteratively until 54 subclusters were identified and did not include neuronal populations. These 54 subclusters were recombined by correlating marker genes to one another, and then performing a dendrogram cut to generate the 9 clusters presented here.

Correlational analysis to mouse and human data was performed as we previously described[54]. Briefly, cluster markers were generated for each dataset individually, and gene scores integrating the specificity and fold enrichment for each marker was calculated. A matrix for every marker gene and gene score was generated across all clusters and used for correlations. This was performed in the space of all mouse forebrain clusters as compared to primary progenitor clusters, and in the space of all organoid clusters compared to the full set of primary clusters in our dataset.

RNA-velocity

Velocity estimates were calculated using the veloctyo.py and scvelo algorithms. Reads that passed quality control after clustering were used as input for the velocyto command line implementation. The human expressed repeat annotation file was retrieved from the UCSC genome browser. The genome annotation file used was provided via CellRanger. The output loom file was used as input to estimate velocity through scvelo. For each individual analysis, cells were filtered based on the following parameters: minimum total counts ≥ 200, minimum spliced counts ≥ 20 and minimum unspliced counts ≥ 10. For the combined cortical analysis, the processed loom files for each individual analysis were combined to generate a new UMI count matrix of 15,473 genes across 53,096 cells, for which the velocity embedding was estimated using the stochastic model. For the combined progenitor analysis, cells that were identified as progenitors were used to create the loom file. The loom files for each of the individuals were combined for a total count matrix of 14,207 genes across 30,562 cells for the velocity embedding using the same criteria. Each embedding was visualized using UMAP.

Immunostaining

Primary human and mouse samples and organoids were collected, fixed in 4% PFA, washed with 1x PBS and immersed in 30% sucrose in 1x PBS until saturated. Samples were embedded in cryomolds using 50% O.C.T. (Tissue-Tek) and 50% of 30% sucrose in 1x PBS and frozen at -80°C. All samples were sectioned at 16µm onto SuperFrost Plus microscope slides. Citrate antigen retrieval (Vector Labs, H-3300) was performed for 20 min at ~95-100°C. Slides were then washed with 1x PBS and blocked using 5% donkey serum, 2% gelatin, 0.1% Triton in 1x PBS for 90 min at RT. Primary antibody incubation occurred in blocking buffer overnight at 4°C, and washed 5x using 0.3% Triton + 10mM glycine. Secondary antibody incubation occurred in blocking buffer for 2-3h at RT using AlexaFluor secondaries (Thermo Fisher). Primary antibodies used: TRKC (R&D Systems, 1:200, AF373); ALX1 (Santa Cruz, 1:500, sc-130416); ID4 (Santa Cruz, 1:200, sc-365656); N-CADHERIN (Abcam, 1:300, ab18203); DLK1 (Abcam, 1:200, ab119930); DLK1 (Abcam, 1:100, ab21682); CROC-4 (Aviva Systems, 1:100, ARP34802_P050); LUM (Thermo Fisher, 1:50, MA5-29402); ZO-1 (Thermo Fisher, 1:100, 61-7300); SOX2 (R&D Systems, 1:100, AF2018); SOX2 (Santa Cruz, 1:250, sc-365823); CTIP2 (Abcam, 1:500, ab18465); KI67 (Thermo Fisher, 1:200, 14-5698); HOPX (Santa Cruz, 1:250, sc-398703); HOPX (Proteintech, 1:200, 11419-1-AP); TBR2 (Abcam, 1:250, ab23345); TBR2 (R&D Systems, 1:250, AF6166); NESTIN (Millipore Sigma, 1:200, MAB5326), DCX (Aves Labs, 1:500, DCX); NEUN (Millipore Sigma, 1:250, ABN91); PAX6 (BioLegend, 1:200, 901301); FOXG1 (Abcam, 1:1000, ab18259).

Imaging and Image Processing

Images were collected on the Leica SP8 inverted confocal microscope using a 40X oil-immersion objective. Because of the scarcity of first trimester primary tissue, only one sample per panel was imaged. For each imaging panel, the parameters (including the gain, offset, pinhole and laser power) for image acquisition was left constant for all samples. Images were later processed using FIJI Image J.

Supplementary Table Legends

Supplementary Table 1. Whole Brain Metadata Each cell that was used in this study is presented in this table with information about number of genes, number of UMIs, sample identity, sample age, cluster identity based upon whole brain clustering (though this analysis had some batch effect and was not used in this study), area, percent ribosomal content, area as annotated during actual dissection, and the cluster the cell belongs to based upon individual clusterings.

Supplementary Table 2. Whole Brain Clustering By Individual Areas Each of the 10 dissected areas was clustered across all samples with the same dissection; batch correction was used to mix the samples with one another. The resulting cluster markers as calculated by the Wilcoxon rank test are presented here for each area, including the p-value, the average $\log_2$(fold change) of the gene in the cluster compared to all other clusters, the percent of cells expressing the gene in the cluster (pct.1), the percent of cells expressing the gene in all other clusters (pct.2), and the adjusted p-value.

Supplementary Table 3. Whole Brain Clustering By Individual Each of the 9 individuals were clustered across all dissections. The resulting cluster markers as calculated by the Wilcoxon rank test are presented here for each area, including the p-value, the average $\log_2$(fold change) of the gene in the cluster compared to all other clusters, the percent of cells expressing the gene in the cluster (pct.1), the percent of cells expressing the gene in all other clusters (pct.2), and the adjusted p-value.

Supplementary Table 4. Differential Gene Expression Across Areas Within a Single Individual In each individual dataset, differential gene expression (Wilcoxon rank test) was performed across all dissected brain regions. The results from the oldest samples (CS22 and CS22_2) were used to identify the markers for each area and subsequently assessed across developmental time (SFig 1 – 2). The resulting cluster markers as calculated by the Wilcoxon rank test are presented here for each area, including the p-value, the average $\log_2$(fold change) of the gene in the cluster compared to all other clusters, the percent of cells expressing the gene in the cluster (pct.1), the percent of cells expressing the gene in all other clusters (pct.2), and the adjusted p-value.

Supplementary Table 5. Cortex Annotations Additional metadata similar to STable 1 is present for the cortex samples only. This also includes the progenitor cluster annotation.

Supplementary Table 6. Cortical Neuron Subclustering Neurons were subclustered and cluster markers were calculated. Neurons from early samples were similarly subclustered and presumed to result from direct neurogenesis. Within the neurons, differential expression was performed between older and younger neuronal populations across all subtypes. The resulting cluster markers for these three analysis are presented as calculated by the Wilcoxon rank test are presented here for each area, including the p-value, the average $\log_2$(fold change) of the gene in the cluster compared to all other clusters, the percent of cells expressing the gene in the cluster (pct.1), the percent of cells expressing the gene in all other clusters (pct.2), and the adjusted p-value.

Supplementary Table 7. Progenitor Subclustered Cluster Markers Progenitors were subclustered and cluster markers were calculated. The resulting cluster markers for these three analysis are presented as calculated by the Wilcoxon rank test are presented here for each area, including the p-value, the average $\log_2$(fold change) of the gene in the cluster compared to all other clusters, the percent of cells expressing the gene in the cluster (pct.1), the percent of cells expressing the gene in all other clusters (pct.2), and the adjusted p-value.

Supplementary Table 8. Velocity Genes and Disease Relevance Velocity genes were calculated using default parameters, and grouped by cluster. Genes were ranked based on velocity score, and compared across all cortical cells, progenitor cells and the cells for each individual sample. Genes that appear in ≥ 3 individual, cortical or progenitor samples were explored for their functional and clinical significance. Genes associated with neurodevelopmental disorders are listed in this table with their corresponding PMID references.

References

1 Herculano-Houzel, S. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences of the United States of America* **109 Suppl 1**, 10661-10668, doi:10.1073/pnas.1201895109 (2012).

2 Sakai, T. *et al.* Fetal brain development in chimpanzees versus humans. *Current biology : CB* **22**, R791-792, doi:10.1016/j.cub.2012.06.062 (2012).

3 Breunig, J. J., Haydar, T. F. & Rakic, P. Neural stem cells: historical perspective and future prospects. *Neuron* **70**, 614-625, doi:10.1016/j.neuron.2011.05.005 (2011).

4 Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol Rev* **20**, 327-348, doi:10.1007/s11065-010-9148-4 (2010).

5 Rowitch, D. H. & Kriegstein, A. R. Developmental genetics of vertebrate glial-cell specification. *Nature* **468**, 214-222, doi:10.1038/nature09611 (2010).

6 O'Rahilly, R. & Muller, F. Developmental stages in human embryos: revised and new measurements. *Cells, tissues, organs* **192**, 73-84, doi:10.1159/000289817 (2010).

7 Wilson, S. W. & Houart, C. Early steps in the development of the forebrain. *Dev Cell* **6**, 167-181, doi:10.1016/s1534-5807(04)00027-9 (2004).

8 Purves D, A. G., Fitzpatrick D, et al., editors. . in *Neuroscience. 2nd edition.* (Sinauer Associates, 2001).

9 Beccari, L., Marco-Ferreres, R. & Bovolenta, P. The logic of gene regulatory networks in early vertebrate forebrain patterning. *Mech Dev* **130**, 95-111, doi:10.1016/j.mod.2012.10.004 (2013).

10 Tumpel, S., Wiedemann, L. M. & Krumlauf, R. Hox genes and segmentation of the vertebrate hindbrain. *Curr Top Dev Biol* **88**, 103-137, doi:10.1016/S0070-2153(09)88004-6 (2009).

11 Agoston, Z., Li, N., Haslinger, A., Wizenmann, A. & Schulte, D. Genetic and physical interaction of Meis2, Pax3 and Pax7 during dorsal midbrain development. *BMC Dev Biol* **12**, 10, doi:10.1186/1471-213X-12-10 (2012).

12 Baizabal, J. M. & Covarrubias, L. The embryonic midbrain directs neuronal specification of embryonic stem cells at early stages of differentiation. *Developmental biology* **325**, 49-59, doi:10.1016/j.ydbio.2008.09.024 (2009).

13 Chen, L., Guo, Q. & Li, J. Y. Transcription factor Gbx2 acts cell-nonautonomously to regulate the formation of lineage-restriction boundaries of the thalamus. *Development* **136**, 1317-1326, doi:10.1242/dev.030510 (2009).

14 Du, T., Xu, Q., Ocbina, P. J. & Anderson, S. A. NKX2.1 specifies cortical interneuron fate by activating Lhx6. *Development* **135**, 1559-1567, doi:10.1242/dev.015123 (2008).

15 Gonzalez, A., Lopez, J. M., Sanchez-Camacho, C. & Marin, O. Regional expression of the homeobox gene NKX2-1 defines pallidal and interneuronal populations in the basal ganglia of amphibians. *Neuroscience* **114**, 567-575, doi:10.1016/s0306-4522(02)00326-3 (2002).

16 Tao, W. & Lai, E. Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork head gene family, in the developing rat brain. *Neuron* **8**, 957-966, doi:10.1016/0896-6273(92)90210-5 (1992).

17 Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18-36, doi:10.1016/j.cell.2011.06.030 (2011).

18 Gotz, M., Hartfuss, E. & Malatesta, P. Radial glial cells as neuronal precursors: a new perspective on the correlation of morphology and lineage restriction in the developing cerebral cortex of mice. *Brain Res Bull* **57**, 777-788, doi:10.1016/s0361-9230(01)00777-8 (2002).

19 Martinez-Cerdeno, V., Noctor, S. C. & Kriegstein, A. R. The role of intermediate progenitor cells in the evolutionary expansion of the cerebral cortex. *Cerebral cortex* **16 Suppl 1**, i152-161, doi:10.1093/cercor/bhk017 (2006).

20 Johnson, M. B. *et al.* Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nature neuroscience* **18**, 637-646, doi:10.1038/nn.3980 (2015).

21 Pollen, A. A. *et al.* Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55-67, doi:10.1016/j.cell.2015.09.004 (2015).

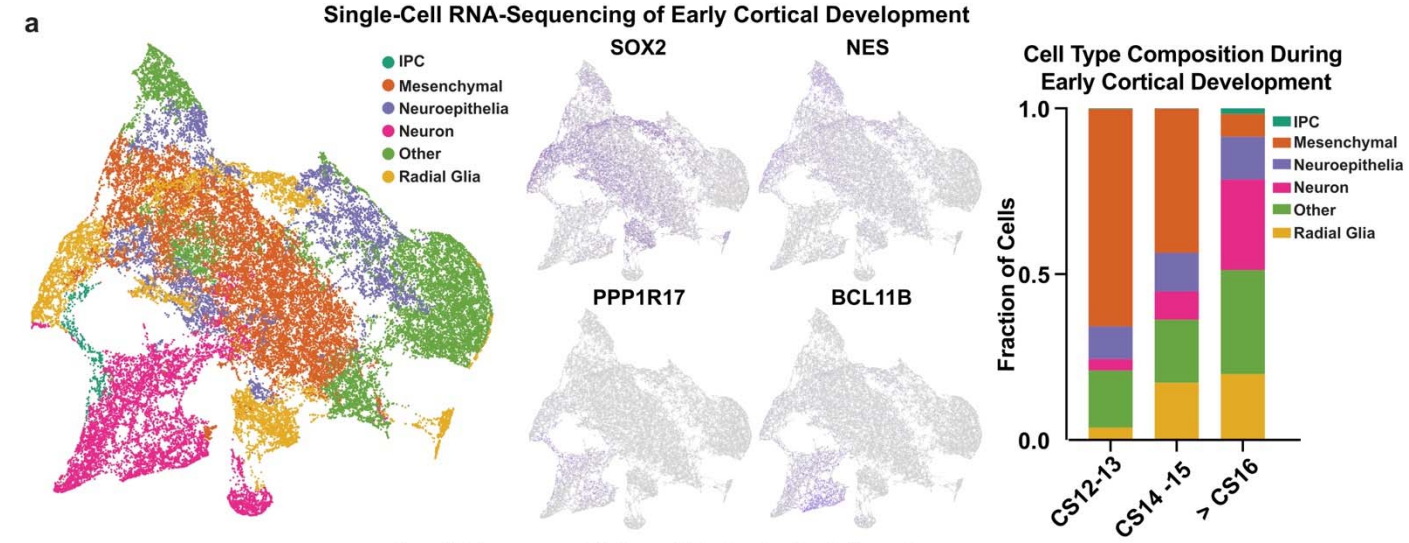
22 Miller, F. D. & Gauthier, A. S. Timing is everything: making neurons versus glia in the developing cortex. *Neuron* **54**, 357-369, doi:10.1016/j.neuron.2007.04.019 (2007).

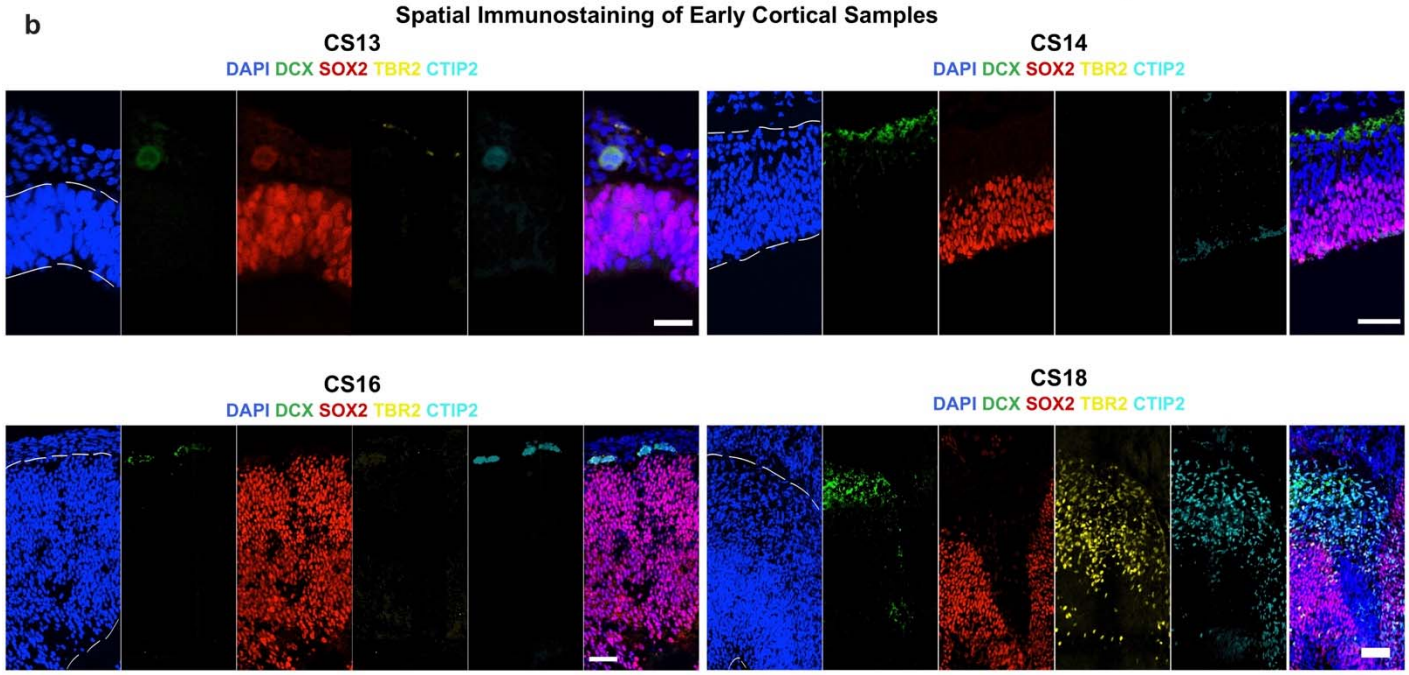
23 Li, C. M., Yan, R. T. & Wang, S. Z. Misexpression of a bHLH gene, cNSCL1, results in abnormal brain development. *Dev Dyn* **215**, 238-247, doi:10.1002/(SICI)1097-0177(199907)215:3<238::AID-AJA6>3.0.CO;2-F (1999).

24 Molyneaux, B. J., Arlotta, P., Hirata, T., Hibi, M. & Macklis, J. D. Fezl is required for the birth and specification of corticospinal motor neurons. *Neuron* **47**, 817-831, doi:10.1016/j.neuron.2005.08.030 (2005).
25 Uittenbogaard, M. & Chiaramello, A. Constitutive overexpression of the basic helix-loop-helix Nex1/MATH-2 transcription factor promotes neuronal differentiation of PC12 cells and neurite regeneration. *J Neurosci Res* **67**, 235-245, doi:10.1002/jnr.10119 (2002).
26 Conde, F., Lund, J. S., Jacobowitz, D. M., Baimbridge, K. G. & Lewis, D. A. Local circuit neurons immunoreactive for calretinin, calbindin D-28k or parvalbumin in monkey prefrontal cortex: distribution and morphology. *J Comp Neurol* **341**, 95-116, doi:10.1002/cne.903410109 (1994).
27 Barbosa, A. C. *et al.* MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9391-9396, doi:10.1073/pnas.0802679105 (2008).
28 Harrington, A. J. *et al.* MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife* **5**, doi:10.7554/eLife.20059 (2016).
29 Li, H. *et al.* Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9397-9402, doi:10.1073/pnas.0802876105 (2008).
30 Hirotsune, S. *et al.* The reeler gene encodes a protein with an EGF-like motif expressed by pioneer neurons. *Nat Genet* **10**, 77-83, doi:10.1038/ng0595-77 (1995).
31 Ozair, M. Z. *et al.* hPSC Modeling Reveals that Fate Selection of Cortical Deep Projection Neurons Occurs in the Subplate. *Cell Stem Cell* **23**, 60-73 e66, doi:10.1016/j.stem.2018.05.024 (2018).
32 Xuan, S. *et al.* Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres. *Neuron* **14**, 1141-1152, doi:10.1016/0896-6273(95)90262-7 (1995).
33 La Manno, G. *et al.* RNA velocity in single cells. *bioRxiv*, doi:10.1101/206052 (2017).
34 Bergen, V., Lange, M., Peidli, S., Wolf, F.A., Theis, F.J. . Generalizing RNA velocity to transient cell states through dynamical modeling. *Biorxiv*, doi:820936 (2019).
35 Aaku-Saraste, E., Hellwig, A. & Huttner, W. B. Loss of occludin and functional tight junctions, but not ZO-1, during neural tube closure--remodeling of the neuroepithelium prior to neurogenesis. *Developmental biology* **180**, 664-679, doi:10.1006/dbio.1996.0336 (1996).
36 Tarin, D. Ultrastructural features of neural induction in Xenopus laevis. *J Anat* **111**, 1-28 (1972).
37 Misson, J. P., Edwards, M. A., Yamamoto, M. & Caviness, V. S., Jr. Identification of radial glial cells within the developing murine central nervous system: studies based upon a new immunohistochemical marker. *Brain research. Developmental brain research* **44**, 95-108, doi:10.1016/0165-3806(88)90121-6 (1988).
38 Qian, X., Riccio, A., Zhang, Y. & Ginty, D. D. Identification and characterization of novel substrates of Trk receptors in developing neurons. *Neuron* **21**, 1017-1029, doi:10.1016/s0896-6273(00)80620-0 (1998).
39 Tessarollo, L. *et al.* trkC, a receptor for neurotrophin-3, is widely expressed in the developing nervous system and in non-neuronal tissues. *Development* **118**, 463-475 (1993).
40 Baladron, V. *et al.* dlk acts as a negative regulator of Notch1 activation through interactions with specific EGF-like repeats. *Exp Cell Res* **303**, 343-359, doi:10.1016/j.yexcr.2004.10.001 (2005).
41 Bray, S. J., Takada, S., Harrison, E., Shen, S. C. & Ferguson-Smith, A. C. The atypical mammalian ligand Delta-like homologue 1 (Dlk1) can regulate Notch signalling in Drosophila. *BMC Dev Biol* **8**, 11, doi:10.1186/1471-213X-8-11 (2008).
42 Miller, A. J. & Cole, S. E. Multiple Dlk1 splice variants are expressed during early mouse embryogenesis. *Int J Dev Biol* **58**, 65-70, doi:10.1387/ijdb.130316sc (2014).
43 Ferron, S. R. *et al.* Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* **475**, 381-385, doi:10.1038/nature10229 (2011).
44 Chakravarti, S. Functions of lumican and fibromodulin: lessons from knockout mice. *Glycoconj J* **19**, 287-293, doi:10.1023/A:1025348417078 (2002).
45 Long, K. R. *et al.* Extracellular Matrix Components HAPLN1, Lumican, and Collagen I Cause Hyaluronic Acid-Dependent Folding of the Developing Human Neocortex. *Neuron* **99**, 702-719 e706, doi:10.1016/j.neuron.2018.07.013 (2018).
46 Zhao, Q., Behringer, R. R. & de Crombrugghe, B. Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the Cart1 homeobox gene. *Nat Genet* **13**, 275-283, doi:10.1038/ng0796-275 (1996).

47 Choe, Y., Zarbalis, K. S. & Pleasure, S. J. Neural crest-derived mesenchymal cells require Wnt signaling for their development and drive invagination of the telencephalic midline. *PloS one* **9**, e86025, doi:10.1371/journal.pone.0086025 (2014).
48 Bystron, I., Rakic, P., Molnar, Z. & Blakemore, C. The first neurons of the human cerebral cortex. *Nature neuroscience* **9**, 880-886, doi:10.1038/nn1726 (2006).
49 Jeffrey, P. L. *et al.* CROC-4: a novel brain specific transcriptional activator of c-fos expressed from proliferation through to maturation of multiple neuronal cell types. *Molecular and cellular neurosciences* **16**, 185-196, doi:10.1006/mcne.2000.0866 (2000).
50 Yun, K., Mantani, A., Garel, S., Rubenstein, J. & Israel, M. A. Id4 regulates neural progenitor proliferation and differentiation in vivo. *Development* **131**, 5441-5448, doi:10.1242/dev.01430 (2004).
51 Bedford, L. *et al.* Id4 is required for the correct timing of neural differentiation. *Developmental biology* **280**, 386-395, doi:10.1016/j.ydbio.2005.02.001 (2005).
52 Kadoshima, T. *et al.* Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20284-20289, doi:10.1073/pnas.1315710110 (2013).
53 Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373-379, doi:10.1038/nature12517 (2013).
54 Bhaduri, A. *et al.* Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142-148, doi:10.1038/s41586-020-1962-0 (2020).
55 Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323, doi:10.1126/science.aap8809 (2017).
56 Peng, Y. R. *et al.* Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina. *Cell* **176**, 1222-1237 e1222, doi:10.1016/j.cell.2019.01.004 (2019).
57 Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166**, 1308-1323 e1330, doi:10.1016/j.cell.2016.07.054 (2016).

Figures and Legends





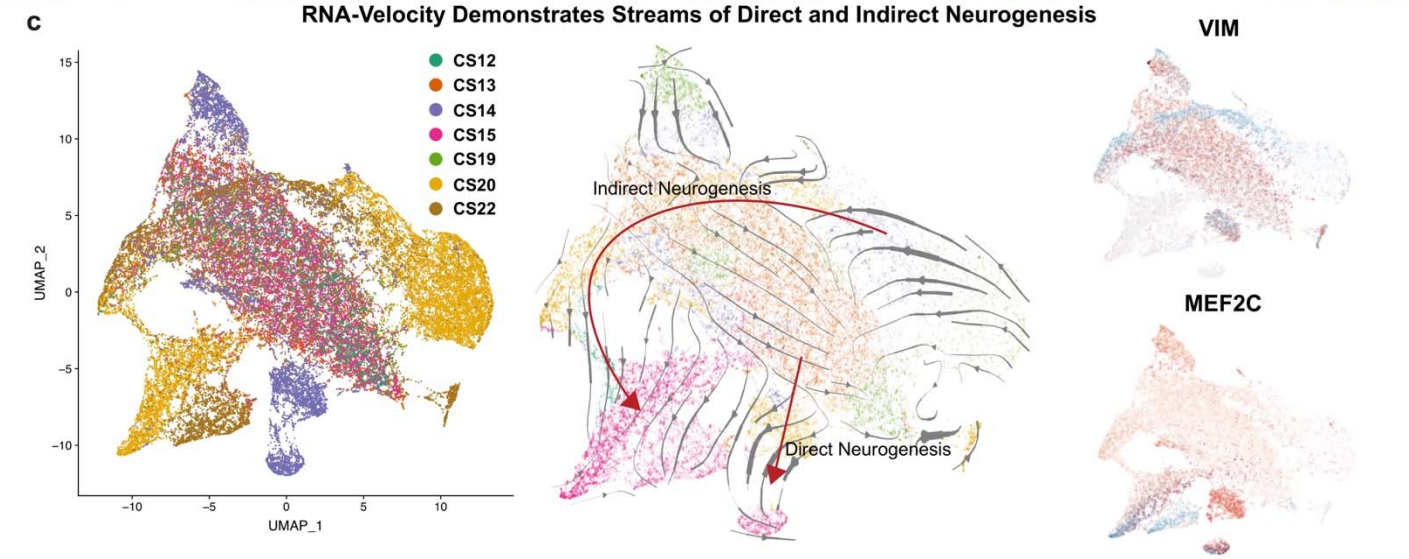


Figure 1. Cell Types in the Early Human Cortex a) UMAP plot of 58,145 telencephalon or cortical dissections colored by annotated cell type. Feature plots of markers of broad progenitors (*SOX2*), radial glia (*NES*), IPCs (*PPP1R17*), and neurons (*BCL11B*) are shown. Stacked bar chart shows cell type composition at earliest (CS12-13), middle (CS14-16), and late (CS19-22) first trimester. **b)** Immunostaining for major cell type markers across first trimester stages. Nuclei shown in blue (DAPI, dotted line demarcates cortical span), newborn neurons marked by doublecortin (DCX, green), progenitors by SOX2 (red), IPCs by TBR2 (yellow), and maturing neurons by CTIP2 (cyan). Scale bars are 50 μM, except CS13 which is 25 μM**. c)** UMAP colored by CS shows the segregation of samples by early and late first trimester stages. RNA-velocity trajectories are depicted by gray arrows in the middle UMAP plot with underlying color by cell types as annotated in (a). Thickness of the lined arrow indicates the differences in gene signature between cell types, and red arrows show predicted direct and indirect neurogenesis trajectories. The velocity plots on the right show the intensity of scored velocity for a progenitor gene, *VIM*, and a neuronal gene, *MEF2C*. Velocities highlight the distinction between progenitors and neurons in the clustering and velocity analyses.

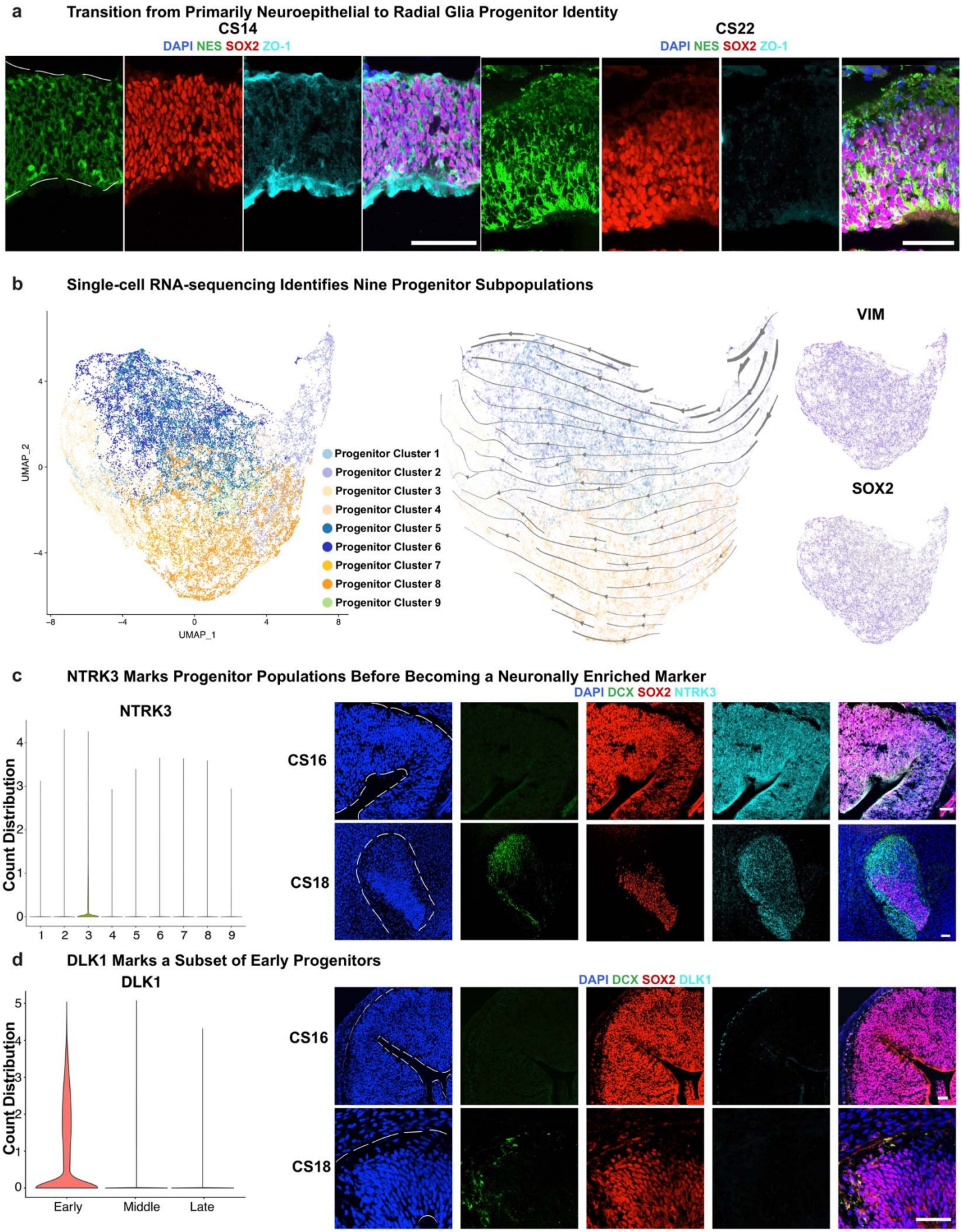


Figure 2. Early Progenitors Can be Divided into Nine Progenitor Subtypes a) Immunostaining of early first

trimester samples (including CS14, shown here) show strong staining for all progenitors (SOX2, red), as well as tight junctions (ZO-1, cyan), but limited staining for radial glia (nestin, NES, green). By CS22, NES staining increases substantially, ZO-1 decreases, and SOX2 expression is maintained. Scale bars are 50 μM. **b)** Left UMAP plot depicts subclustered progenitor cells, while the middle plot shows the velocity trajectory across progenitor subtypes. Feature plots on the right show high expression of *VIM* and *SOX2* marking all progenitor populations. **c)** Progenitor Cluster 3 is specifically and uniquely labeled by *NTRK3* as shown in the violin plot. Immunostaining for NTRK3 (cyan) shows early labeling of SOX2 (red) positive progenitors at CS16, but expression shifts to more closely coincide with newborn neurons marked by DCX (green) by CS18. Scale bars are 50 μM. **d)** DLK1 (cyan) is the top marker for Progenitor Cluster 8, and is exclusively expressed in early first trimester as shown in the violin plot on the left. Immunostaining for DLK1 at CS16 shows co-localization with low SOX2 (red) expressing cells at the boundary of the cortical edge. This staining disappears from the cortex entirely by CS18 when DCX (green) staining emerges.

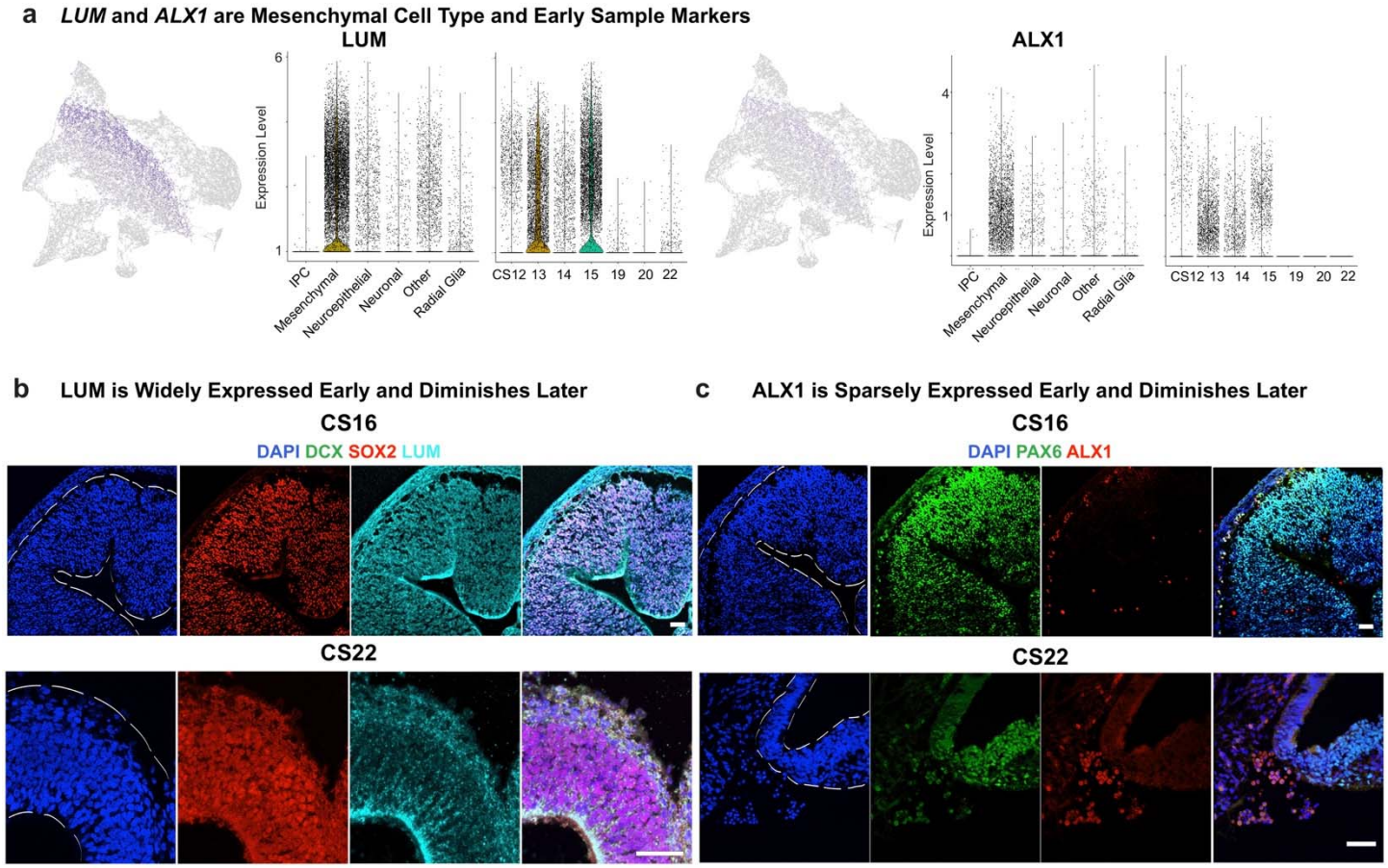


Figure 3. Single-cell RNA-sequencing Identifies Early Mesenchymal Cell Population a) In both progenitor and cortex clustering, *LUM* marks a separate population of cells, as shown by the feature plot on the left. *LUM* expression is highly specific to the mesenchymal cell type and is enriched in early samples. *ALX1* expression is highly correlated to *LUM* as shown in the right feature plot, and is similarly enriched in early, mesenchymal populations. **b)** Immunostaining for LUM (cyan) shows prevalent expression in and between progenitors marked by SOX2 (red) at CS16, but this expression dissipates by CS22 (bottom panels). Scale bars are 50 μM. **c)** Immunostaining for ALX1 (cyan) shows sparse expression in PAX6 (green) positive cells, that disappears from the cortex but is expressed in surrounding brain structures at CS22 (bottom panels). Scale bars are 50 μM.

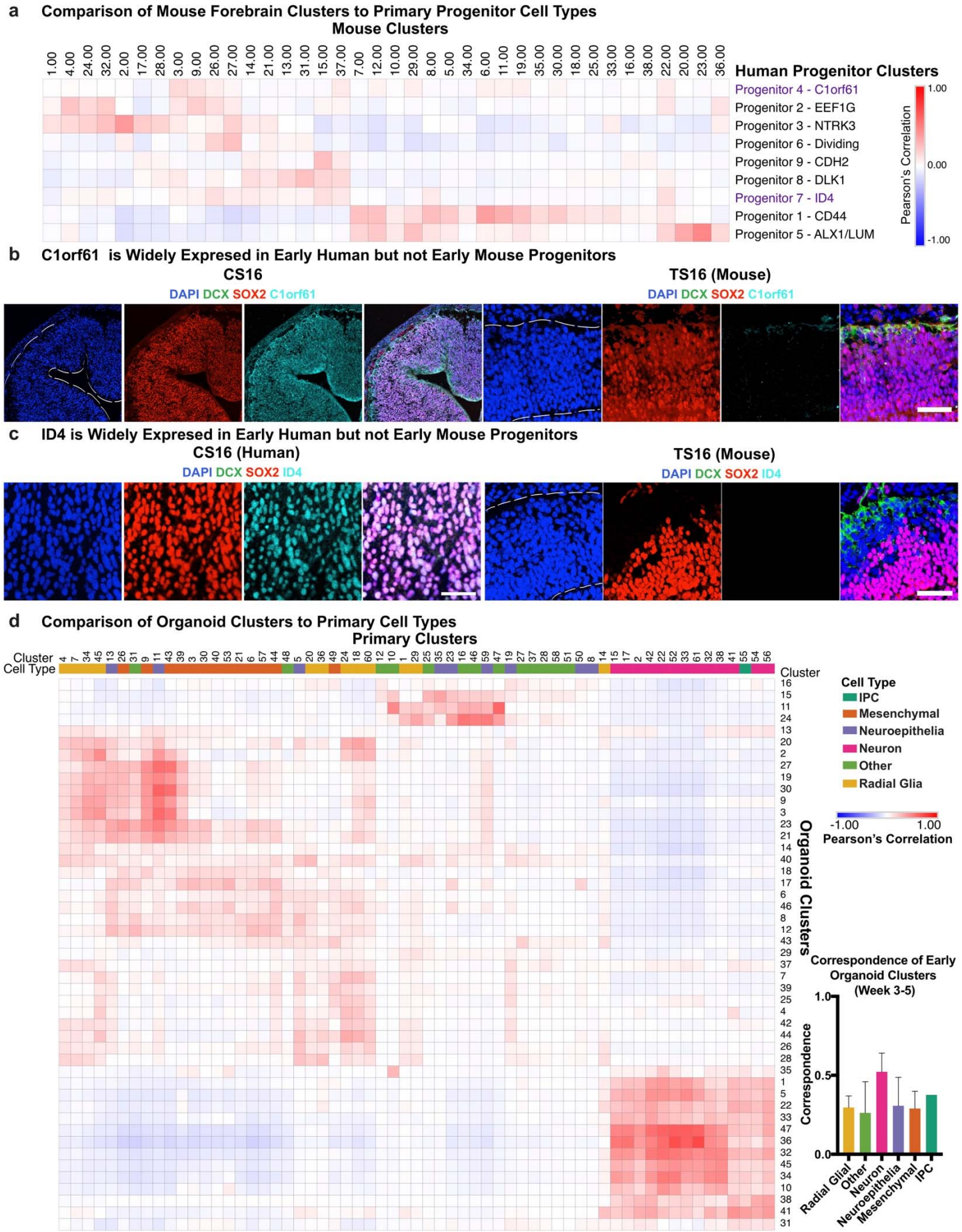

a
Comparison of Mouse Forebrain Clusters to Primary Progenitor Cell Types
Mouse Clusters
1.00 4.00 24.00 32.00 2.00 17.00 28.00 3.00 9.00 26.00 27.00 14.00 21.00 13.00 31.00 15.00 37.00 7.00 12.00 10.00 29.00 8.00 5.00 34.00 6.00 11.00 19.00 35.00 30.00 18.00 25.00 33.00 16.00 38.00 22.00 20.00 23.00 36.00
Human Progenitor Clusters
Progenitor 4 - C1orf61
Progenitor 2 - EEF1G
Progenitor 3 - NTRK3
Progenitor 6 - Dividing
Progenitor 9 - CDH2
Progenitor 8 - DLK1
Progenitor 7 - ID4
Progenitor 1 - CD44
Progenitor 5 - ALX1/LUM
Pearson's Correlation
1.00
0.00
-1.00
b
C1orf61 is Widely Expresed in Early Human but not Early Mouse Progenitors
CS16
DAPI DCX SOX2 C1orf61
TS16 (Mouse)
DAPI DCX SOX2 C1orf61
c
ID4 is Widely Expresed in Early Human but not Early Mouse Progenitors
CS16 (Human)
DAPI DCX SOX2 ID4
TS16 (Mouse)
DAPI DCX SOX2 ID4
d
Comparison of Organoid Clusters to Primary Cell Types
Primary Clusters
Cluster
Cell Type
4 7 34 45 13 26 31 9 11 43 39 3 30 40 53 21 6 57 44 48 5 20 36 49 24 18 60 12 10 1 29 25 35 23 16 46 59 47 19 27 37 28 58 51 50 8 14 15 17 2 42 22 52 33 61 32 38 41 55 54 56
Cluster
16 15 11 24 13 20 2 27 19 30 9 3 23 21 14 40 18 17 6 46 8 12 43 29 37 7 39 25 4 42 44 26 28 35 1 5 22 33 47 36 32 45 34 10 38 41 31
Organoid Clusters
Cell Type
IPC
Mesenchymal
Neuroepithelia
Neuron
Other
Radial Glia
-1.00
1.00
Pearson's Correlation
Correspondence of Early Organoid Clusters (Week 3-5)
Correspondence
1.0
0.5
0.0
Radial Glial
Other
Neuron
Neuroepithelia
Mesenchymal
IPC

Figure 4. *In Vivo* and *In Vitro* Systems Model Different Aspects of Early Cell Types a) Heatmap showing the comparison of mouse forebrain clusters from 16,053 cells to the primary progenitor populations identified in this dataset. Correlations are performed using Pearson correlations between cluster marker sets, and identify Progenitor Clusters 4 and 7 to not have a counterpart in mouse data. **b)** Immunostaining of human samples at CS16 shows broad expression of C1orf61 (cyan) in progenitor cells labeled by SOX2 (red). Parallel staining in TS16 mouse shows no C1orf61 expression. Scale bars are 50 μM. **c)** Immunostaining of human samples at CS16 shows broad expression of ID4 (cyan) in progenitor cells labeled by SOX2 (red). A parallel staining in TS16 mouse shows no ID4 expression. Scale bars are 50 μM. **d)** Heatmap showing correlation between organoid clusters from week 3 and 5 organoids (from Bhaduri et al, 2020) to primary cells from this study. Correlations are performed using Pearson correlations between cluster marker sets, and summary metrics across all primary clusters in each cell type category are shown in the bar plot (mean with standard deviation). Excitatory neurons are the most highly preserved cell type in organoids, but progenitors, including