NEUROSCIENCE

Transcriptional priming as a conserved mechanism of lineage diversification in the developing mouse and human neocortex

Zhen Li[1,2]*, William A. Tyler[2,3]*, Ella Zeldich[3], Gabriel Santpere Baró[1,4], Mayumi Okamoto[3,5], Tianliuyun Gao[1], Mingfeng Li[1], Nenad Sestan[1,6†], Tarik F. Haydar[2,3†]

How the rich variety of neurons in the nervous system arises from neural stem cells is not well understood. Using single-cell RNA-sequencing and in vivo confirmation, we uncover previously unrecognized neural stem and progenitor cell diversity within the fetal mouse and human neocortex, including multiple types of radial glia and intermediate progenitors. We also observed that transcriptional priming underlies the diversification of a subset of ventricular radial glial cells in both species; genetic fate mapping confirms that the primed radial glial cells generate specific types of basal progenitors and neurons. The different precursor lineages therefore diversify streams of cell production in the developing murine and human neocortex. These data show that transcriptional priming is likely a conserved mechanism of mammalian neural precursor lineage specialization.

INTRODUCTION

The cerebral cortex is the seat of higher-order cognition, motor control, and social behavior. It emerges early during embryonic development from a simple epithelial sheet in the prosencephalon and expands into a complex six-layered amalgam of neural cells and circuits, with cell identity, morphology, and function consolidated both by laminar position and regional localization. At least 55 excitatory and 60 inhibitory transcriptomically defined neuron cell types (ExN and InN, respectively) have recently been reported in two regions of the adult mouse neocortex (1), and it is probable that even more cell types exist in other cortical areas and in other mammalian species. This repertoire of neurons is produced during neurogenesis from germinal zone stem and progenitor cells and is essential for the normal development of cognitive, sensory, and motor functions. Alterations in neurogenesis are known to lead to numerous neurodevelopmental and neuropsychiatric disorders (2). Crucially, although the importance of neuron diversity in the neocortex is well recognized, the fundamental mechanisms underlying production of this neuronal variety from the comparatively homogeneous stem and progenitor cells is not currently understood.

All excitatory neocortical neurons are born from two broad classes of neural stem and progenitor cells that reside in the dorsal ventricular and subventricular zones during embryonic development. Apical or ventricular radial glial cells (aRGCs or vRGCs) have been identified as the neural stem cells of the neocortex, because they alone exhibit multipotency and the ability to undergo self-renewing asymmetric cell divisions. Daughter cells born from aRGCs are fated to either become neurons (direct neurogenesis) or to generate the second class of precursors, the intermediate progenitor cells (IPCs), which in turn undergo limited rounds of cell division to amplify neuronal output (indirect neurogenesis) (3, 4). In recent years, subgroups of IPCs have been characterized on the basis of marker gene expression, morphology, cell cycle (CC) dynamics, and the location of their mitosis. These include at least three known IPC types: apical IPCs (aIPCs) (5–8), basal IPCs (bIPCs) (9, 10), and basal or outer radial glial cells (bRGCs or oRGCs) (11–13). Despite the identification of these major apical and basal precursor groups, major deficiencies remain in our understanding of these cells and their roles in generating the extensive neuronal variation that arises during brain development. For example, our recent data indicate that different precursor types can contemporaneously produce excitatory neurons with distinct properties (14, 15). Thus, if aRGCs, aIPCs, bIPCs, and bRGCs are each comprised by distinct subtypes of cells, it could provide a mechanism for generating a wide number of neuron types with specific functions and roles in the cortical circuitry.

Using single-cell droplet capture, we conducted a high-throughput gene expression analysis of neocortical cells in mouse and identified many groups of stem and progenitor cells with distinct transcriptional profiles. Comparison of these results to published human single-cell RNA-sequencing (scRNA-seq) datasets not only (2, 16–18) revealed remarkable similarities across species but also highlighted an increase in bRGC diversity in human neocortex. We observed multiple cell types with "mixed identity" transcriptional profiles, that is, coexpression of genes typically thought to define separate types of stem and progenitor cells. In vivo intersectional fate mapping and in situ gene expression experiments revealing the identity of these cells in both mouse and human brain indicate that transcriptional priming in aRGC subgroups is a primary mechanism used to generate progenitor heterogeneity during neurogenesis. Last, state-of-the-art bioinformatics approaches indicate that, as a population, the neural precursor lineages simultaneously produce multiple streams of cortical neurons. These data, describing the shared and divergent

[1]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, USA. [2]Center for Neuroscience Research, Children's Research Institute, Children's National Hospital, Washington, DC, USA. [3]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA. [4]Neurogenomics group, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), DCEXS, Universitat Pompeu Fabra, Barcelona, Spain. [5]Department of Anatomy and Cell Biology, Nagoya University Graduate School of Medicine, Aichi 466-8550, Japan. [6]Departments of Genetics, of Psychiatry and of Comparative Medicine, Program in Cellular Neuroscience, Neurodegeneration and Repair, Child Study Center, Yale School of Medicine, New Haven, CT, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: thaydar@childrensnational.org (T.F.H.); nenad.sestan@yale.edu (N.S.)

features between rodents and primates, provide a new picture of how neural precursor heterogeneity is leveraged to influence cortical size and neuronal diversity in a species-specific manner.

RESULTS

Single-cell transcriptomic classification identifies neural precursor diversity

We used the ddSEQ Single-Cell Isolator (Bio-Rad and Illumina) to capture cells from the developing wall of the cerebral neocortex at embryonic day (E) 15.5, when excitatory neurons destined for the upper layers are generated (Fig. 1A). After applying stringent quality control measures, 5777 cells from multiple litters (N = 8) and two technical replicates were subject to downstream analyses (fig. S1A). Principal components analysis (PCA) of highly variable genes (HVGs) and subsequent gene ontology analysis revealed that the first two principal components (PCs) were related to CC/cell division and neuron differentiation (fig. S1B). To minimize the effect of CC on cell type classification, we next regressed out the variance related to CC (fig. S1, C and D). Using the first 33 PCs (fig. S1E), we then performed t-stochastic neighbor embedding (t-SNE) analysis (19) and Louvain-Jaccard clustering. Cells from the two replicates mixed well, indicating negligible technical variation (fig. S1F). The clustering analysis resulted in the identification of 25 clusters of cells with distinct transcriptional profiles. On the basis of the expression of canonical marker genes, two types of dorsal telencephalic mouse radial glial cells (mRGCs), five types of mouse IPCs (mIPCs), eight types of mouse excitatory neurons (mExN), and six types of mouse inhibitory interneurons (mInN) were identified (Fig. 1B). In addition, we also identified one group of ventral radial glial cells (mRGC3; inadvertently included due to microdissection procedures), ventral progenitor cells (VPs), Cajal-Retzius cells (CRs), and cells from the choroid plexus (CPs). In general, these clusters expressed cell type selective markers including *Pax6*, *Hes1*, and *Sox2* (RGCs); *Dlx1* and *Sp9* (VPs); *Reln* (CRs); and *Ttr* (CPs), as well as layer-specific neuronal markers including *Satb2* (layers 2 to 4) and *Sox5* and *Fezf2* (layers 5 and 6) (Fig. 1, C and D; figs. S1G and S2; and table S1). Further confirming the putative identities of many of these clusters, weighted gene coexpression network analysis (WGCNA) (20) indicated the concerted expression of numerous gene modules likely to play key functional and cell type–specific roles (Fig. 1E and fig. S3). For example, genes in module 1 (M1), such as *Nes*, *Pax6*, *Gli3*, *Hes1*, *Notch1*, *Notch2*, and *Fgfr3*, have been associated with apical progenitor populations, and the expression of M1 genes was enriched in mRGC1/2/3 and mIPC1/2 (Fig. 1F). In contrast, genes in M8 were enriched in mIPCs and included the canonical IPC marker gene *Tbr2/Eomes* and several genes, such as *Neurog1* and *Neurog2*, which have been associated with IPCs (Fig. 1F). This module also included *Mfng* and *Mfap4*, genes encoding extracellular matrix proteins involved in cell adhesion or intercellular interactions whose role in IPCs has not been extensively characterized. WGCNA also identified core gene expression networks expressed in mExN subtypes (M10, M11, and M12) and in mInN subtypes (M12, M13, and M14). These core gene networks likely play fundamental roles in establishing and maintaining the identity and function of the corresponding cell types.

Although the WGCNA and marker gene expression were sufficient to discriminate and provisionally uncover the cellular identity within the dataset, several cell types were found to be simultaneously associated with mixed molecular signatures. For example, both mRGC2 and mIPC1 expressed apical progenitor markers, including *Sox2*, *Hes1*, and *Fabp7*, as well as basal progenitor markers such as *Eomes*. Although these "mixed character cells" may represent transitional phases between apical and basal progenitor types, we were struck by the remarkably high proportion of such cells within their corresponding cell types (fig. S4, A to C). Similarly, some clusters were highly correlated for the expression of genes in WGCNA modules, indicative of alternate cell identities. For example, although M8 included several genes known to be associated with mIPCs, mRGC2 was also enriched for the expression of genes in M8 (Fig. 1E). These results raised the possibility that such mixed signatures may be distinct cellular states contributing to differentiation or lineage diversification.

Mouse aRGCs show early signs of lineage diversification

To distinguish whether mixed character progenitors represent transient and transitional or stable and distinctive cell states, we first assessed the diversity of cells within each cluster by intracluster distance (fig. S4D). We found that mRGC cell types generally showed higher transcriptome complexity, demonstrated by a high level of intracluster distance, compared with other cell types; mRGC2, which contained cells exhibiting a mixed RGC/IPC expression profile, was the most diverse. This observation prompted us to investigate the substructure within each progenitor cell type using pseudotime analysis (21), which identified multiple states within mRGC1 and mRGC2. We identified three states in mRGC1 that were related to CC progression (Fig. 2A and fig. S4E), demonstrated by gene expression patterns of CC-related genes along pseudotime (fig. S4F). However, in mRGC2, we found four states that appeared to reflect distinctive lineage commitments, as evidenced by the expression of lineage-selective marker genes (Fig. 2B). In particular, state II in mRGC2 was enriched for *Eomes* expression as compared with all other states in both mRGC1 and mRGC2 (Fig. 2C). Differential expression (DEX) analysis between the four mRGC2 states confirmed that state II was enriched with bIPC genes, including *Eomes* and *Mfng* (fig. S4G). Notably, although mRGC2 state II exhibited higher *Eomes* expression relative to other mRGCs, its expression of *Eomes* and other IPC markers was significantly lower than found in mIPC cell types.

To determine whether some aRGCs in the mouse neocortex may express *Eomes* and to test whether this expression reflects lineage identity, we used in utero electroporation (IUE) to label precursors at E11.5 and E14.5 with plasmids expressing mCherry under the control of the *Eomes* promoter along with a plasmid expressing Lyn–green fluorescent protein (GFP) from the constitutive *EF1α* promoter. After classifying precursor types based on morphological properties, these experiments showed that 16% of bipolar apical precursors (presumed aRGCs) and 32% of the unipolar apical precursors express the pEomes construct at E12.5 (fig. S5, A to C). The percentages of bipolar and unipolar apical precursors increased to 35 and 72% at E14.5, respectively (fig. S5, D to F). To also assess this question with fate mapping constructs, we performed IUE with a plasmid expressing membrane-tagged *Lyn*-GFP governed by Cre recombination driven by the *Eomes* promoter into the E14.5 developing neocortical wall. Twenty-four hours later, this labeling method elucidated multiple classes of progenitors expressing the *Eomes* promoter construct, identified by morphological and anatomical properties as aRGCs, aIPCs, bIPCs, and bRGCs (Fig. 2D) (22). To
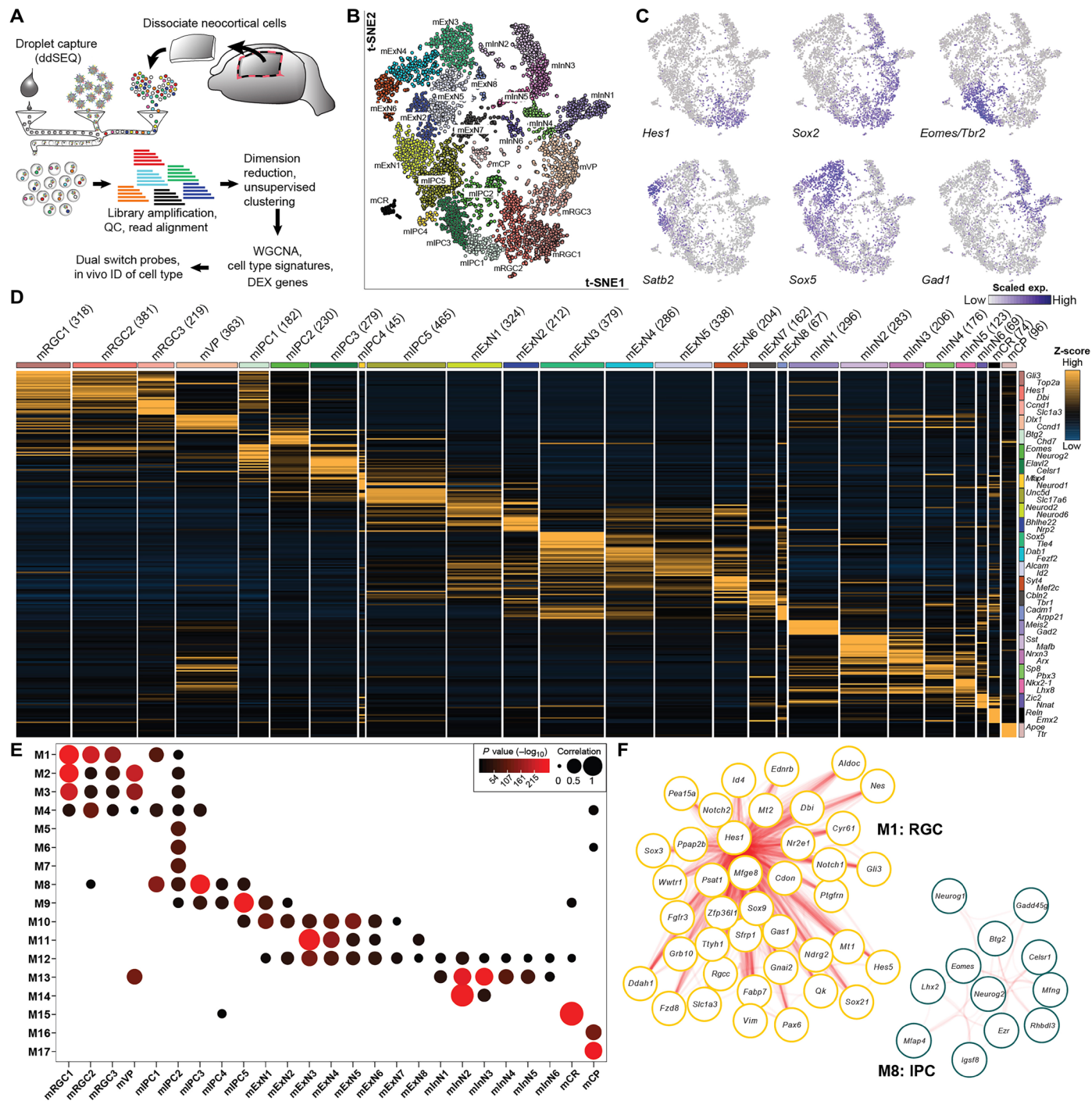
Fig. 1. Single-cell RNA-seq analysis of E15.5 mouse cortex. (**A**) Schematics of experimental design. Dissociated neocortical cells from embryonic day 15.5 (E15.5) mouse brain were captured by ddSEQ method. (**B**) t-SNE plot of single cells from E15.5 mouse cortex. Colors represent cell types. VP, ventral progenitor; CR, Cajal-Retzius cell; CP, choroid plexus. (**C**) Feature plots of canonical marker gene expression. Heatmap represents normalized level of gene expression. The number of cells in each cell type is indicated in parentheses (**D**) Heatmap of differentially expressed (DEX) genes between cell types. Colors represent cell types as in (B). Two of the DEX genes from each cell type are listed on the right. (**E**) Weighted gene correlation networks of all mouse cell types found in the current dataset. Seventeen coexpression modules are identified. Size of the dots indicates level of correlation between network and cell type, whereas colors represent level of significance (Bonferroni-corrected P value). (**F**) Genes in module 1 (M1: RGC) and module 8 (M8: IPC) are shown.
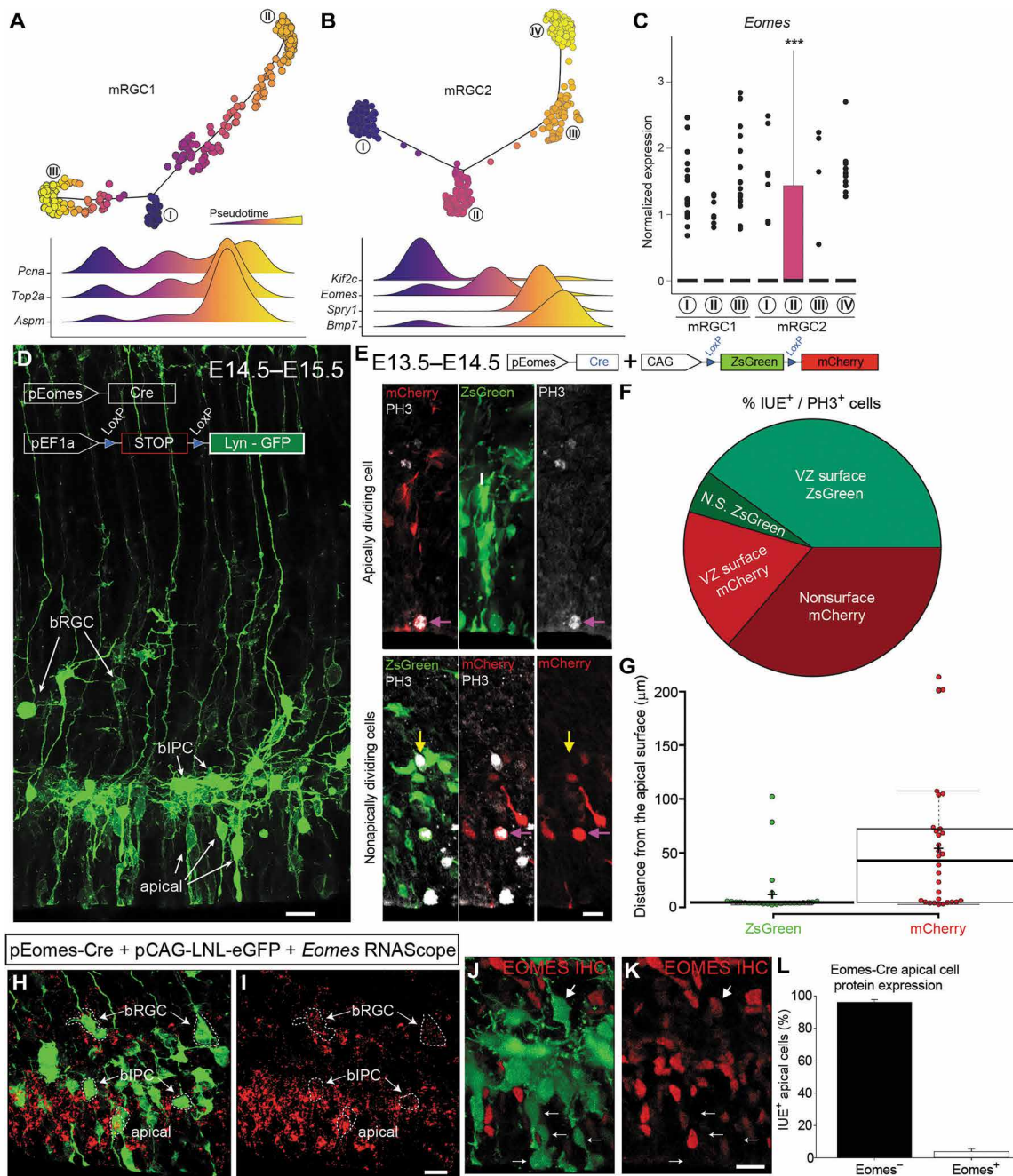
Fig. 2. Dynamic cell states are present among mouse radial glia cells. (**A** and **B**) Pseudotime trajectory of mRGC1 (A) and mRGC2 (B). Color indicates pseudotime progression. Cell states are indicated with circled Roman numerals. Genes showing strong association with pseudotime, and cell states are shown at the bottom of each panel. (**C**) Boxplot of *Eomes* expression levels in each cell state (circled Roman numerals) of mRGC cell types. Asterisks indicate statistical significance (Fisher's exact test) compared with any other cell state. (**D**) Eomes-Cre IUE-based fate mapping demonstrates multiple cell morphotypes including aRGCs, bIPCs, and bRGCs. (**E**) IUE of Eomes-Cre with dual-color StopLight reporter using PH3 to isolate mitotic cells. A subpopulation of Eomes-Cre–expressing cells divides at the VZ surface while non–Tbr2-Cre–expressing cells primarily divide at the VZ surface. (**F**) Location of PH3[+] divisions by Eomes-Cre fate map lineage. (**G**) Proportion of precursors dividing at the surface of the lateral ventricle or subapically differs by lineage; 36.7% of mitotic cells expressing Eomes-Cre divide at the ventricular surface. Mann-Whitney *U* test, *n* = 3, *P* < 0.001. (**H** to **J**) Cells with aRGC morphology expressing Eomes-Cre plasmid do not express EOMES protein. (**K** and **L**) Precursors expressing Eomes-Cre plasmid express *Eomes* mRNA. Scale bars, 20 μm.

confirm expression of *Eomes* by aRGCs, we next electroporated E13.5 brains with pEomes-Cre and a conditional dual-color StopLight plasmid that expresses mCherry after Cre-mediated recombination (Fig. 2E), followed 24 hours later by immunohistochemical labeling for phosphorylated histone H3, a marker for mitotic cells. Consistent with the prominence and rapid cycling of aRGCs, the majority of pEomes-Cre–negative cells (ZsGreen-positive) were located at apical positions near the ventricle. In contrast, and in agreement with the canonical view of *Eomes* expression by basal progenitors, most (63.3%) pEomes-Cre (mCherry)–expressing cells divided at basal

positions away from the ventricle (Fig. 2, F and G). However, we also observed that approximately 36.7% of the cells expressing mCherry following Eomes-Cre–mediated recombination were found dividing at the ventricular zone (VZ) surface, the preferential location for aRGC mitoses, most likely representing cells of the mRGC2 cluster. To confirm that the pEomes-Cre plasmid faithfully reports endogenous *Eomes* expression, we used single molecular fluorescent RNA in situ hybridization and detected *Eomes* mRNA in cells transfected with pEomes-Cre and a Cre-conditional enhanced GFP (eGFP) reporter plasmid (Fig. 2, H and I). However, immunolabeling using an antibody against the EOMES protein failed to detect expression of EOMES in the eGFP-expressing aRGCs (Fig. 2, J to L). Thus, using morphological and gene expression tools, we revealed a pEomes-Cre$^+$ apical precursor cell population that expresses *Eomes* mRNA but not the EOMES protein.

Transcriptional priming may drive lineage diversification in the mouse neocortex

The detection of *Eomes* mRNA but not EOMES protein in certain mRGCs based on in silico and in vivo data, suggests that transcriptional priming, a phenomenon whereby mRNA for proteins that will be expressed in progeny is present but not translated in the parent cell (*23*), may contribute to key features of the developing mouse brain. To assess whether this phenomenon is widespread among progenitor cell populations in the developing mouse brain, we correlated mRNA and protein expression in several cell types and states exhibiting mixed character gene expression signatures similar to those we observed in mRGC2. To identify candidate genes/proteins potentially subject to transcriptional priming, we focused on the genes in the IPC module (M8) from the WGCNA analysis (Fig. 1F). Because genes from M8 are likely important for the establishment of IPC identity, it is reasonable that other genes from the module, in addition to *Eomes*, exhibit transcriptional priming in RGCs. We observed multiple M8 genes expressed in one or multiple states in RGCs (fig. S5G). In particular, *Igsf8* was expressed by all states, except state IV, in mRGC2, whereas *Mfap4* was only expressed by state III in mRGC2. Using immunohistochemistry, we found that IGSF8 and MFAP4 are widely expressed at the protein level in the mouse ventricular zone and are therefore not candidates for transcriptional priming (not shown). Applying this same approach to mIPC1, mIPC2, mIPC3, and mIPC5 also identified multiple states in each of these mIPC cell types characterized by expression of *Eomes* and many other genes previously attributed to bIPCs, although *Eomes* expression in these IPC groups was at least 10-fold greater than that found in mRGC2 (fig. S6).

Because the pEomes-Cre fate mapping approach labeled multiple morphotypes (Fig. 2D), we hypothesized that an in vivo approach restricting labeling to cells that coexpress both *Eomes* and apical marker genes such as *Hes1*, *Fabp7*, or *Slc1a3* could specifically highlight the cells with mixed identity, including those within the mRGC2 and mIPC1 profiles. To this end, we designed an intersectional approach combining FLP recombinase (Flpe) driven by the *Eomes* promoter, an Flpe conditional plasmid expressing Cre under the control of one of the apical marker genes (e.g., *Hes1*, *Slc1a3*, or *Fabp7*), and a Cre-conditional eGFP reporter (Fig. 3A). Fifteen hours following IUE with these intersectional fate mapping plasmids at E14.5, most of the labeled cells resembled aRGCs and expressed SOX2 protein (Fig. 3, B and B$_2$), demonstrating the presence of a subpopulation of aRGCs expressing the *Eomes* transcript as pre-

dicted by the expression profile of the mRGC2 cluster. In addition, a few cells with bRGC morphology were also present among the GFP-labeled cohort, and these bRGCs expressed EOMES protein (Fig. 3B1) as well as SOX2 (Fig. 3, D and E). By 24 hours elapsed time, the SOX2$^+$/EOMES$^-$ aRGCs in the VZ were joined by a much larger SOX2$^+$/EOMES$^+$ bRGC population in the subventricular zone (SVZ) (Fig. 3, C, C1, C2, and C3), consistent with the contemporaneous expression of aRGC and IPC genes (including *Eomes*) in mIPC1 (fig. S6A). Many of the cells with bRGC morphology were located adjacent to a second eGFP$^+$ cell, suggestive of a recent cell division and the generation of a daughter cell within the 24-hour period Fig. 3 (D and E). Immunostaining for PH3 and SOX2 confirmed the proliferative status of eGFP$^+$ bRGCs and indicated that they express SOX2 as in primate and carnivore brain (fig. S7A). The bRGCs as well as their daughter cells expressed the SOX2 protein (fig. S7, B and B′), suggesting that they produce daughter cells that retain molecular aspects of apical progenitor identity despite their distinctive morphology and localization within deeper regions of the neocortical wall. Together, these results indicate that a subgroup of aRGCs expresses translationally blocked *Eomes* mRNA transcripts and that this lineage of aRGCs generates proliferative bRGCs.

In addition to mIPC1, our bioinformatics analysis also revealed other IPC groups with mixed apical and basal gene expression (Fig. 3, F to I). Specifically, mIPC3 strongly expressed both *Eomes* and apical markers like *Fabp7*, but was further defined by the expression of *NeuroD4* (Fig. 3J). A previous single-cell study bioinformatically identified a subset of cells that coexpress *FABP7* and *NEUROD4* as bRGCs in human and ferret neocortex, but a cognate population was absent in the mouse (*24*). To determine whether these *NeuroD4*$^+$ mouse progenitors align with the bRGC morphotype, we fate mapped cells in the neocortical wall at E14.5 using pEomes and pNeuroD4 plasmids driving mRFP and GFP reporters, respectively (Fig. 4A). Cotransfection of these four constructs highlighted the overall *Eomes*-expressing cell population with RFP and identified that a subset of these cells (30%) also expressed *NeuroD4* (GFP$^+$). The *NeuroD4*$^+$ subset was comprised entirely by cells with bIPC and bRGC morphology, the latter of which were often found closely opposed to a presumed daughter cell (Fig. 4, B and C, and fig. S7, C and D). Immunostaining showed that the *NeuroD4* lineage bIPCs are EOMES$^+$/SOX2$^-$, whereas the *NeuroD4*-expressing bRGCs are EOMES$^+$/SOX2$^+$ (Fig. 4D). To determine whether the *NeuroD4*$^+$ bRGCs are the same population of bRGCs found in the *Eomes*/VZ gene intersectional cohort in Fig. 3, we quantified the proportion of pNeuroD4-cre–expressing bRGCs in the total bRGC population. Using the dual-color StopLight reporter driven by the ubiquitous chicken beta-actin promoter, we identified the total bRGC population with morphological criteria and found that only 40% of the *Eomes*-expressing bRGCs also expressed *NeuroD4* (Fig. 4, E to I). We then used fluorescence mRNA labeling along with intersectional fate mapping to confirm the presence of *Eomes*$^+$/*NeuroD4*$^+$ and *Eomes*$^+$/*NeuroD4*$^-$ bRGCs in the mouse neocortical wall (fig. S7E to H). These are the first data to indicate multiple different classes of bRGCs in the mouse neocortex.

## Precursor diversity and neuron production streams

Our published work demonstrates that neurons born from distinct precursor groups can express specific electrophysiological and morphological properties, even when they are generated on the same day and migrate to the same neocortical layer (*14,15*). We therefore
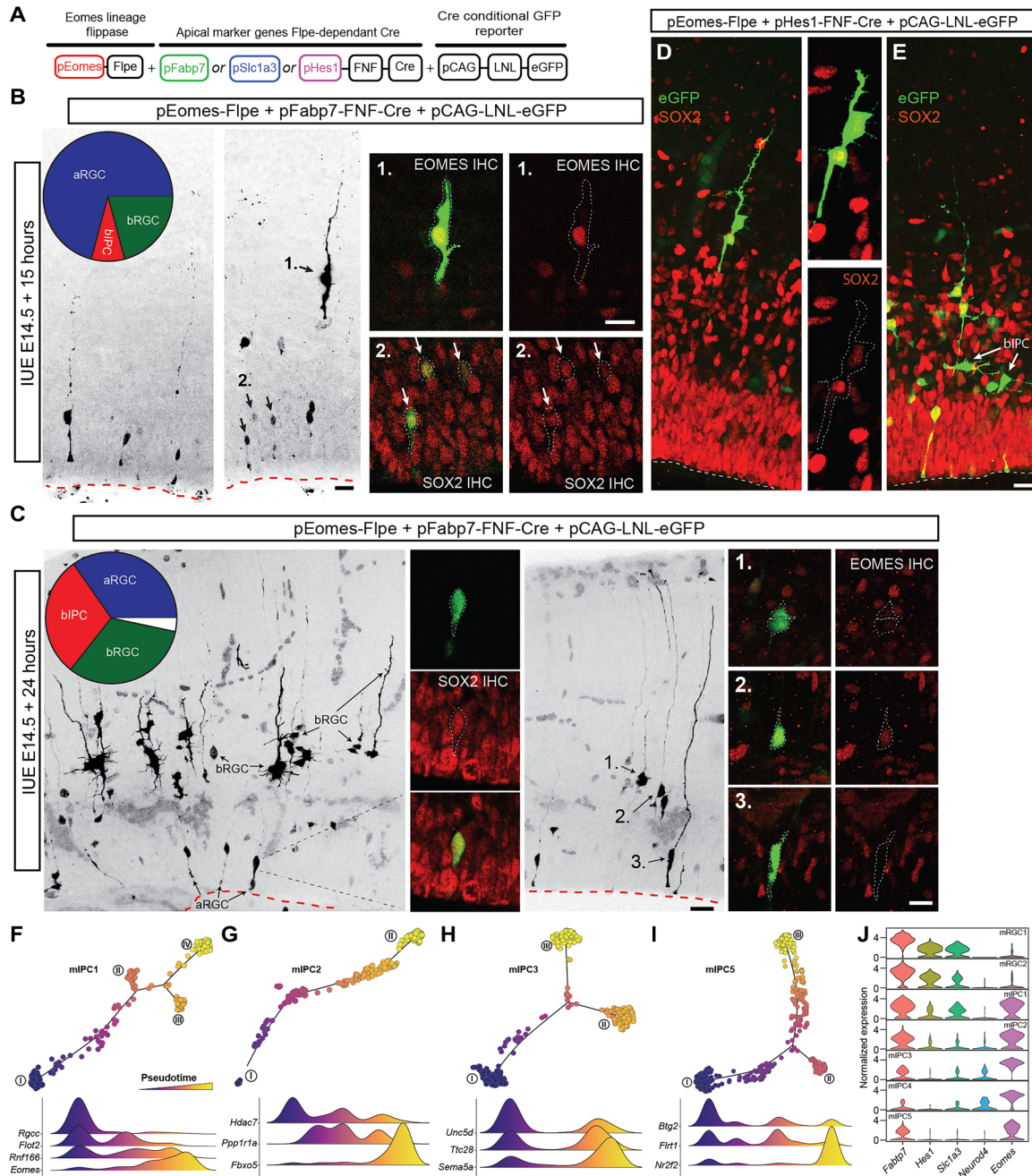
Fig. 3. Dual switch fate mapping to reveal identity of mixed expression cell types. (**A**) Intersectional (dual switch) genetic fate mapping strategy for in vivo labeling of cell types. The Flpe-conditional Cre construct used either *Fapb7*, or *Slc1a3*, or *Hes1* promoters. (**B**) IUE to label cells coexpressing *Eomes* and *Fabp7* at E14.5 with 15-hour survival. Proportion of each morphological cell type represented in pie chart. Eomes⁺ bRGCs (1) and Sox2⁺ aRGCs (2) were present. (**C**) IUE to label cells coexpressing *Eomes* and *Fabp7* at E14.5 with 24-hour survival. Proportion of each morphological cell type represented in the pie chart. Transfectants include Eomes⁺ bRGCs (1 and 2) and Sox2⁺/Eomes⁻ aRGCs (3). (**D**) IUE to label cells coexpressing *Eomes* and *Hes1* at E14.5 with 24-hour survival, demonstrating that bRGCs and their daughter cells express SOX2. (**E**) IUE to label cells coexpressing *Eomes* and *Hes1* at E14.5 with 24-hour survival, demonstrating expression of SOX2 in aRGC and bRGC (white arrowheads) but not bIPCs (white arrows). (**F** to **I**) Pseudotime trajectories of mIPC1 (F), mIPC2 (G), mIPC3 (H), and mIPC5 (I). Color indicates pseudotime progression. Cell states are indicated with circled Roman numerals. Genes showing strong association with pseudotime and cell states are shown at the bottom of each panel. (**J**) Violin plot of canonical marker genes for RGCs and IPCs expressed by mIPCs. Colors represent different genes. Vertical axis shows normalized gene expression levels. Scale bars, 20 μm.

sought to determine how the precursor diversity found in this single-cell study correlates with the multiple subgroups of excitatory neurons found after cell capture and analysis (e.g., mEx-1 through mEx-8). To do this, we focused on the mouse single-cell data and

used novel trajectory reconstruction methods to resolve the pseudo-developmental process from progenitors to highly differentiated excitatory neurons. Through this process, we were able to establish four well-separated streams emanating from the precursor cell types
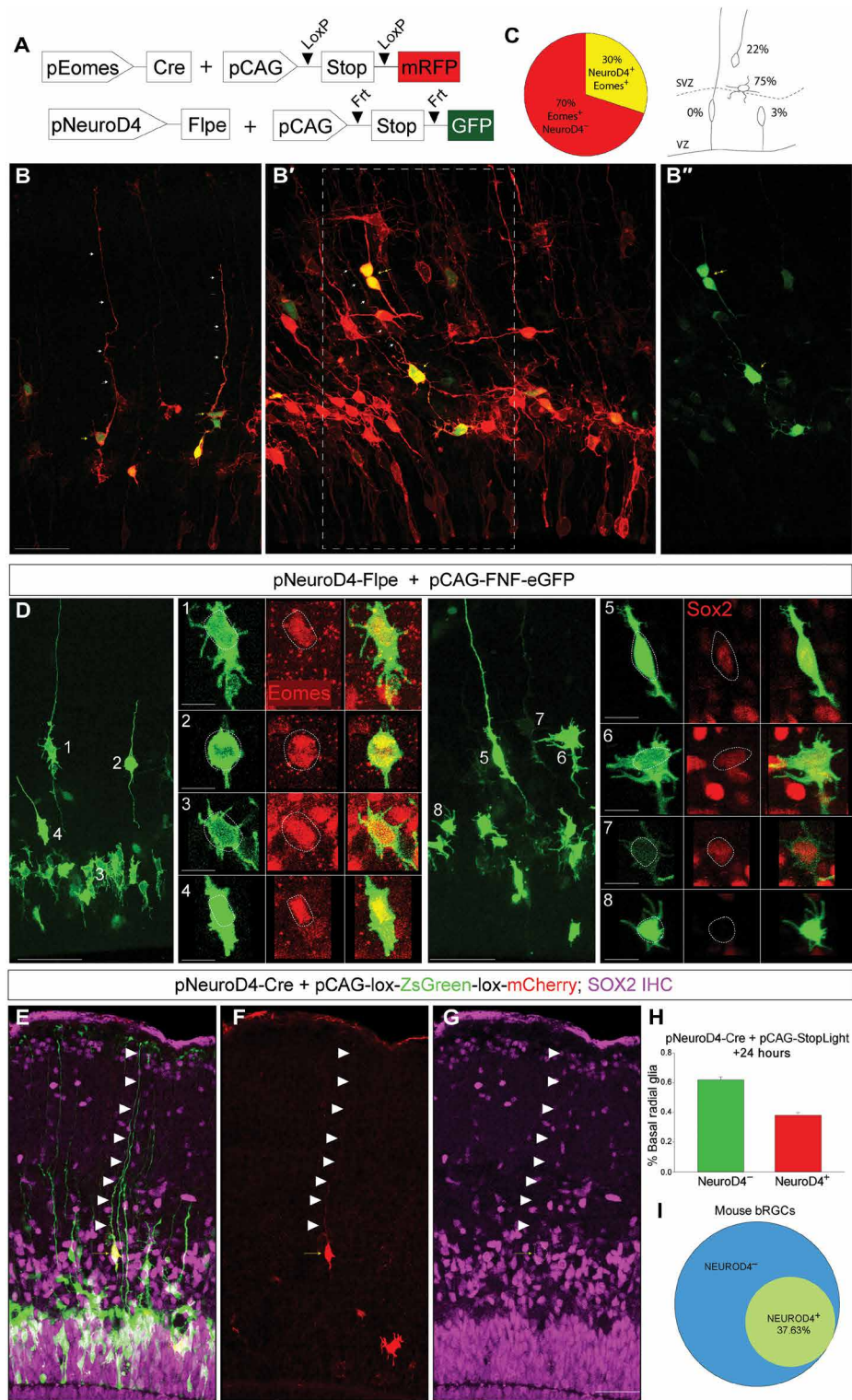
Fig. 4. A subset of cortical precursors express *Eomes/NeuroD4* mRNA and exhibit bRGC morphology. (**A**) Fate mapping constructs used to elucidate identity of cells expressing both *Eomes* and *NeuroD4* via IUE in E14.5 mouse neocortex. (**B**) Quantification and morphologies of Eomes⁺/NeuroD4⁻ and Eomes⁺/NeuroD4⁺ cells 24 hours after IUE. (**C**, **B′**, **B″**) Eomes⁺/NeuroD4⁺ cells exhibit bIPC and bRGC morphology. Cell colors as in (A). White arrows indicate radial processes, and yellow arrows indicate cell bodies. (**D**) NeuroD4-expressing cells in mouse neocortical wall 24 hours after IUE. Insets show numbered cells and their expression of Eomes and Sox2. White dashed lines show location of cell bodies. (**E** to **G**) NeuroD4 StopLight fate mapping with IUE on E14.5 followed by 24-hour survival demonstrates that NeuroD4⁺ bRGCs are Sox2⁺. White arrows indicate radial processes, and yellow arrows indicate cell bodies. (**H** and **I**) Quantification of bRGCs 24 hours after IUE with NeuroD4 fate mapping approach. NeuroD4⁺ bRGCs represent 38% of entire bRGC population. Scale bars, 50 μm.

(fig. S8, A and B). The excitatory neurons in these streams expressed genes identifying them as either superficial (streams 1, 2, and 3) or deep (stream 4) excitatory neurons (fig. S8C). Stream 1 also displayed characteristics of immature neurons as shown by the expression of Neurod1. We then plotted the ExN groups onto these streams and found that mExN cell types exhibited highly specific locations within these streams (fig. S8D). Some mExN groups were primarily restricted to one stream (e.g., mExN1, mExN3, and mExN6), while other groups (mExN2, mExN4, and mExN7) were present in two streams, and cells from the mExN5 cluster were present in all four streams. These computational results support previous studies indicating that excitatory neuron identity is varied, as measured by gene expression profiling, and that excitatory neuron types potentially resolve into different lineage streams produced by neocortical precursors.

To determine whether specific precursor cell types may be lineally correlated with the streams and the ExN contained therein, we quantified the percentage of each precursor cell state (i.e., mRGC and mIPC cell types) in the four streams (fig. S8E). In general, multiple precursor types are found along any given stream trajectory, but the contributions of each precursor cell state to the streams are distinctive. For example, a majority of cells from state III in mRGC2 and mIPC4 contribute to stream 1, the stream that showed immature neuron characteristics. A high percentage of cells from mRGC1-II and mIPC5-III contribute significantly to stream 2 (superficial excitatory neurons). All states from mRGC1 and mIPC3 contribute predominantly to stream 3 (deep excitatory neuron). State I of mRGC2 as well as mIPC1 states I, II, and III contribute almost exclusively to stream 4 (superficial excitatory neurons). These analyses suggest that multiple different precursor types and states may underlie each stream of excitatory neurons and that particular admixtures of precursor cell types cooperate to produce specific lineage streams during neocortical neurogenesis. These data, coupled with our recent publications demonstrating that different precursor lineages produce ExN with specific properties (14, 15), suggest that the precursor heterogeneity identified in the current study, while subtle and dynamic, may be an important driving factor for excitatory neuron diversity and circuit complexity.

The morphology of bRGCs has been previously shown to be quite variable (25), and we noticed this as well. To quantify bRGC morphology, we conducted three-dimensional (3D) image analysis to determine whether bRGC shape differed between subtypes or across labeling procedures. We scored cells as belonging to one of three categories (fig. S8, F to H): type A, characterizing cells having many small filopodial projections along with short apical or basally directed main processes; type B, unipolar bRGCs with one relatively unbranched basally directed process that terminates before reaching the pia; and type C, unipolar bRGCs that project to the pial surface. We found that all dual switch labeling strategies yielded bRGCs with all three types of morphology, suggesting that these variations in shape may be a general property of bRGCs, perhaps relating to transitory phases of their maturation or proliferation state. We did find differences in proportions of bRGC type, though, suggesting that bRGC diversity may be correlated with signaling from the basal lamina and with cytoskeletal complexity during cell production. For example, the pEomes-Flpe + pFabp7-FNF-Cre population had a higher proportion of pia-touching type C bRGC, whereas the pEomes-Flpe + pSlc1a3-FNF-Cre population was overrepresented by short bRGCs of type A morphology (fig. S8, I to L). Together, these results confirmed the separation of cell types elucidated by the

bioinformatics approaches and that *Eomes*-expressing bRGCs consist of multiple subgroups. We next sought to determine how similar these newly elucidated mouse cell types are to those found during human neocortical development, especially because primate bRGCs have not been previously described as part of the *EOMES* lineage.

Conserved and divergent features of mouse and human neocortical stem and progenitor cells

To compare transcriptomic features of mouse and human neocortical stem and progenitor cells, we created a human cell database by combining multiple published human scRNA-seq datasets (2, 16–18) of 12 to 20 postconceptional week (PCW) neocortex into one containing cell number comparable to our mouse dataset. We confirmed that the combined human data also contained all major cell types in the developing human neocortex (fig. S9, A to C) and then used several methods to conduct a cross-species comparison. First, we correlated the WGCNA modules identified from the mouse single-cell analysis to human single cells and found a similar correlation pattern, suggesting that core transcriptomic networks are shared by the same cell types across species (fig. S9D). This approach also identified a human IPC cell type with mixed character gene expression. Specifically, hIPC1 was highly correlated with both RGC (M1) and IPC (M8) modules (fig. S9D). Marker gene expression profiling confirmed the coexpression of RGC markers (i.e., *SLC1A3*, *FABP7*, and *HES1*) and IPC markers (i.e., *EOMES*) in hIPC1 (fig. S9E). Next, we used an established method (17) to integrate the human single-cell dataset with our mouse dataset and showed that all major cell types were well integrated between the two species (Fig. 5A and fig. S9, F to H). Focusing on RGCs and IPCs, we observed that certain human and mouse cell types overlap in UMAP (uniform manifold approximation and projection) space (Fig. 5, B and C). Using Meta-Neighbor analysis (26), we identified pairs of human and mouse neural precursor cell types that were highly similar to each other after integration (Fig. 5D). DEX analysis was then performed separately for the human or mouse precursor types, and the intersection between the human and mouse results was used to identify common genes by cell type using stringent criteria (Materials and Methods; table S1). These genes were then used to conduct enrichment analysis for the shared human and mouse precursor cell types in the context of human developmental and neurodegenerative disorders (Fig. 5F). We found that some pairs of human and mouse cell types were enriched for genes associated with neurodevelopmental or neurodegenerative diseases. For example, genes specific to and shared by hIPC2 and mIPC3 were significantly enriched for low IQ and schizophrenia.

In general, a continual transcriptomic profile shift emerged, with apical RGCs (i.e., mRGC1, hRGC2, and mRGC2) at one end and more differentiated IPCs (i.e., hIPC3, mIPC4, and mIPC5) at the other end, with combinations of canonical marker genes clearly demarcating the two ends of this range (Fig. 5F). Specifically, RGCs expressed HES1, ID4, CYR61, FOS, and TUBA1B at high levels consistently, whereas IPCs expressed EOMES, NEUROD1, ELAVL2, and ELAVL4. We observed similar mixed signatures in hRGC3, mIPC1, hIPC1, mIPC2, and hIPC2. For example, hRGC3, which is likely a human bRGC cell type based on *HOPX* expression (fig. S9E), also expressed IPC genes such as NEUROD6 and ELAVL4 at relatively high levels compared with other RGCs. Clusters mIPC1 and hIPC1, on the other hand, showed an undoubtable IPC identity with high levels of EOMES, but also expressed a panel of canonical RGC genes including ID4, FABP7, and CKB. We also observed that
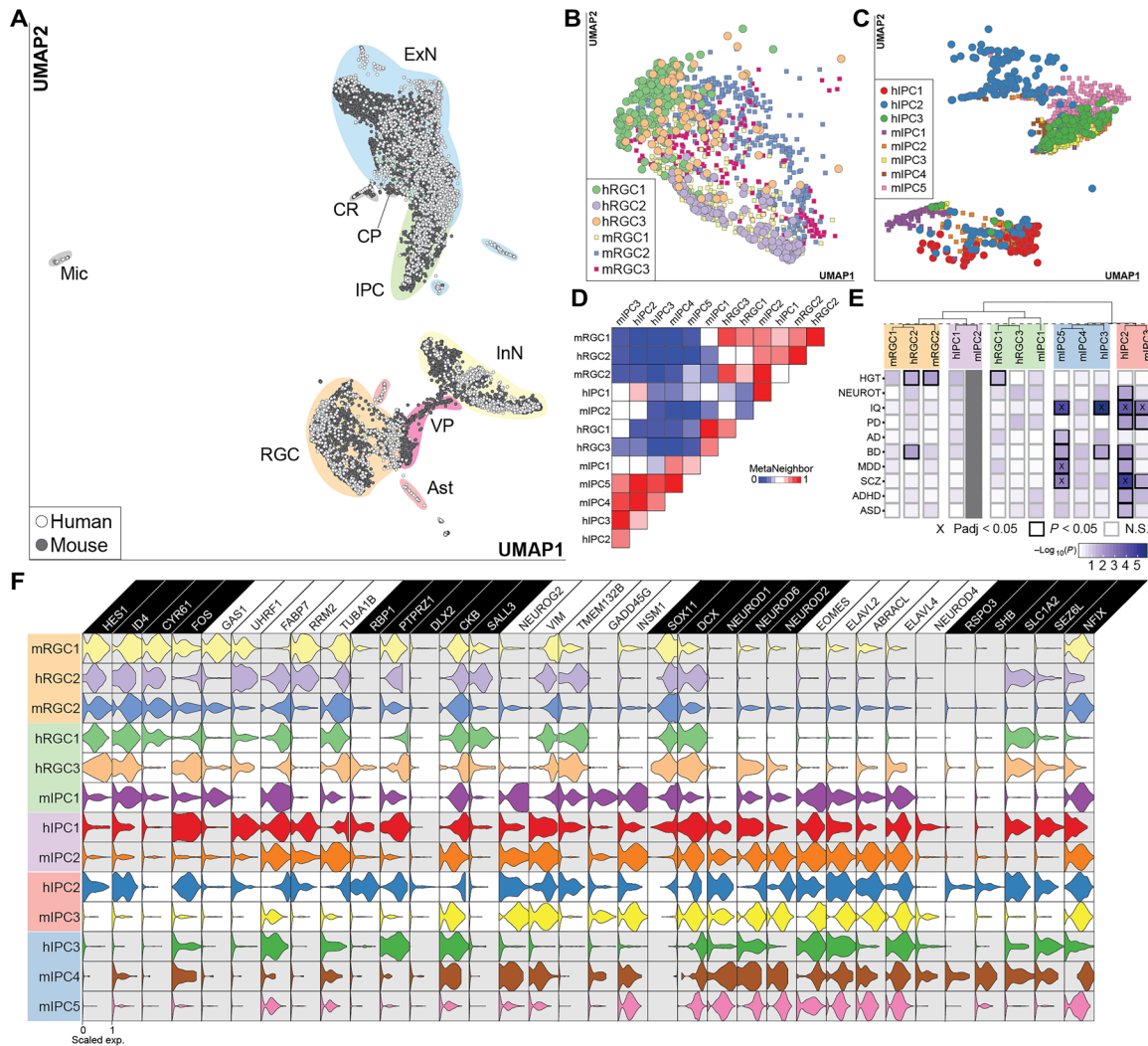
Fig. 5. Integration of human and mouse single-cell dataset reveals conserved and divergent progenitor cell types. (**A**) UMAP plot of integrated human (*4*, *22–24*) and mouse datasets (open and gray circles, respectively). Colors represent different major cell types. (**B** and **C**) UMAP plots of integrated human (h) and mouse (m) RGC (B) and IPC (C) single-cell data. Human single cells are represented by circles, whereas mouse single cells by square. Colors represent cell types. (**D**) Heatmap plot of MetaNeighbor analysis. Colors represent AUROC score. (**E**) Enrichment analysis for genes associated with human neurodevelopmental and neurodegenerative disorders in human and mouse progenitor cell types. Clustering based on MetaNeighbor analysis is shown on top (colors represent clusters). Dashed line indicates clustering threshold. Heatmap color represents unadjusted *P* value. Significant enrichment at unadjusted *P* < 0.05 is indicated by box, and adjusted *P value* (Padj) < 0.05 by "x." N.S., not significant. (**F**) Violin plot of common DEX genes in both human and mouse progenitors. Colors are the same as in (B), (C), and (E). Horizontal axis indicates scaled expression level (Scaled exp.).

hIPC1 and hIPC3 expressed *NEUROD4*, perhaps highlighting a common bRGC precursor identity with mIPC3.

These gene expression results indicate that the human neocortex also contains multiple subtypes of bRGCs (i.e., expressing *EOMES*/*NEUROD4* and *EOMES*/*SOX2*). To confirm this in vivo, we used multiplex fluorescence in situ hybridization to localize cells expressing *NEUROD4*, *EOMES* and *SOX2* in sections of the human neocortex at 12 weeks of gestation. Because of fixation and RNA degradation artifacts in two of the three samples in our archive, we were only able to obtain results from one brain. Nevertheless, the mRNA localization we report for *EOMES* and *SOX2* below matches the Allen Developing Human Brain Atlas and other previous publications. We found that *EOMES* and *NEUROD4* are largely coexpressed in cells of the inner SVZ (iSVZ), although a substantial

number of cells express only *EOMES* or *NEUROD4* in this zone (Fig. 6, A to C). In contrast, the number of positively stained cells fell to 9.1% of the overall population in the outer SVZ (oSVZ) (Fig. 6, D and E). Whereas most of these labeled cells in the oSVZ expressed *EOMES* only (56%), 27.7% also expressed *NEUROD4*, while fewer cells (16.3%) expressed *NEUROD4* alone (Fig. 6E). We noted that a substantial proportion (45.4%) of cells expressed *NEUROD4* or *EOMES* mRNA in the human VZ as well (Fig. 6, F and G).

Mapping of *EOMES* and *SOX2* led to similar results. A significant number of iSVZ and oSVZ cells coexpressed *EOMES* and *SOX2*, and the *SOX2*[+]/*EOMES*[+] population was greater in the iSVZ than in the oSVZ, where it comprised 9.6% of the labeled cells (Fig. 6, H to K). We also identified a large population of *SOX2*[+]/*EOMES*[−] cells in the oSVZ that represented the largest population (73.1%) of
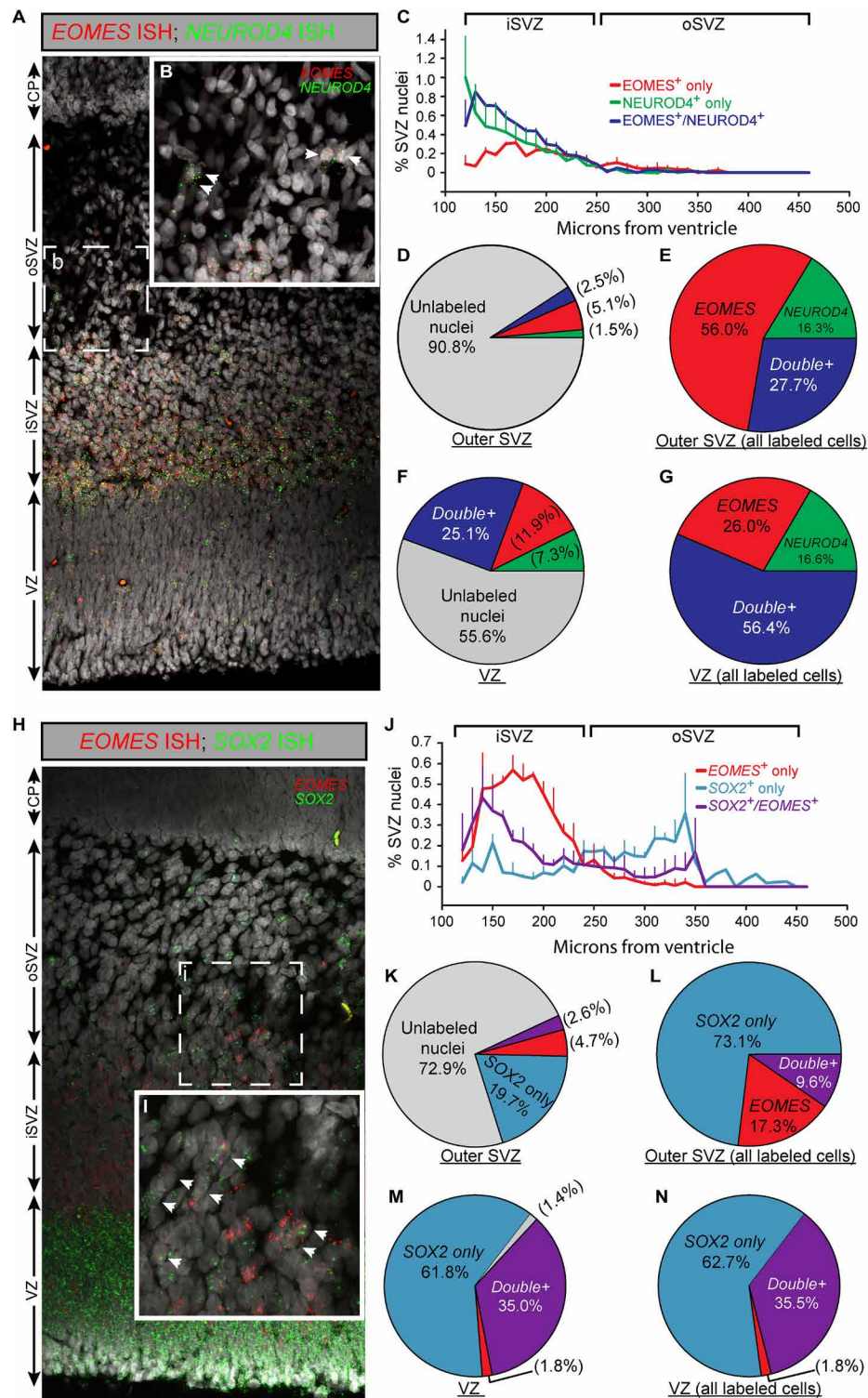
Fig. 6. Molecular phenotyping of human basal RGCs at 12 gestation weeks. (**A** and **B**) Single-molecule multiplex mRNA hybridization (RNAScope) for *NEUROD4* (green) and *EOMES* (red) in human 12wg neocortical wall. Boxed inset (B) shows larger magnification. Arrows indicate *NEUROD4*+/*EOMES*+ cells. (**C**) Quantification of *EOMES*+ (red), *NEUROD4*+ (green), and *EOMES*+/*NEUROD4*+ (blue) cells from the top of the ventricular zone (VZ) to the top of the OSVZ. (**D** and **F**) Quantification of all un-labeled (gray), *EOMES*+ (red), *NEUROD4*+ (green), and *EOMES*+/*NEUROD4*+ (blue) cells in OSVZ (D) and VZ (E). (**E** and **G**) Distribution of RNAScope+ cells expressing *EOMES*, *NEUROD4*, or both markers in OSVZ (E) and VZ (G); colors as in (C). (**H** and **I**) RNAScope for *SOX2* (green) and *EOMES* (red) in human 12wg neocortical wall. Boxed inset (i) shows larger magnification. Arrows indicate *SOX2*+/*EOMES*+ cells. (**J**) Quantification of *EOMES*+ (red), *SOX2*+ (green), and *EOMES*+/*SOX2*+ (purple) cells from the top of the VZ to the top of the OSVZ. (**K** and **M**) Quantification of all unlabeled (gray), *EOMES*+ (red), *SOX2*+ (green), and *EOMES*+/*SOX2* (purple) cells in OSVZ (K) and VZ (M). (**L** and **N**) Distribution of *EOMES*+, *SOX2*+, or *EOMES*+/*SOX2*+ cells in OSVZ (L) and VZ (N); colors as in (J).

labeled cells in this zone (Fig. 6L). As expected, the VZ contained a high number of *SOX2*⁺ cells and a very small number of *EOMES*-only cells. Unexpectedly, though, we found that 36% of the human VZ cells coexpressed *SOX2* and *EOMES*, a finding that further supports the single-cell gene expression evidence for aRGC heterogeneity and identifying the presence of aRGC cell types with expression features, indicating transcriptional priming (Fig. 6, M and N). Together, these in vivo human data indicate substantial similarities with the mouse bRGC results, showing that *NEUROD4/EOMES* and *SOX2/EOMES* double-positive bRGCs exist in both species and that their parent cell types are present in the VZ. Furthermore, adding to this cell diversity, our RNA labeling studies indicate that the largest human oSVZ population is a *SOX2*-only population of oRG (Fig. 6, K and L) that has been previously identified (*27–29*). This *SOX2*-only oRG cell type is not observed in the mouse neocortical SVZ.

DISCUSSION
Cellular imaging studies indicate that neocortical aRGCs can divide to generate neurons directly and can also produce other precursor types; how one group of cells accomplishes these varied tasks is unknown. Our bioinformatics and in vivo findings demonstrate a much larger variety of neural precursor cell types than previously recognized, indicating multiple types of specialized aRGCs and IPCs during neurogenesis. We showed that the dynamic transcriptomic states of specific aRGCs may be indicative—perhaps even instructive—to the route of differentiation that an individual RGC may undertake. In addition, in vivo validation experiments demonstrate that bRGC cells and their precursors share novel transcriptional profiles in the mouse and human neocortex, as well as species-specific profiles, that contribute to separate lineages in the developing brain. Here, we describe two subtypes of dorsal neocortical aRGCs in mouse and three aRGC subtypes in human. Two of these RGC types in each species (mRGC1, mRGC2, hRGC1, and hRGC2) exhibit remarkably shared properties, and one (hRGC3) is likely the *SOX2*-only bRGC precursor present in human but not mouse. We also identify three similar IPC subtypes in human and mouse as well as two mouse IPC types that do not appear to have human counterparts.

Crucially, the novel cellular diversity we describe is not represented exclusively by morphological or anatomical characteristics but rather by newfound mixed transcriptional profiles. Although previous scRNA-seq studies have categorized cell populations based on morphological type (i.e., aRGC versus bIPC versus bRGC) and the expression of cardinal transcription factor genes (*24*, *27*), we found several clusters of cells that coexpressed gene sets previously regarded as exclusive for distinct progenitor populations. Moreover, our in vivo fate mapping and RNAScope analyses revealed that these mixed marker clusters are comprised by cells with multiple morphotypes, including subsets of aRGCs, bIPCs, and bRGCs. This suggests that transcriptional profiles may reflect discrete lineages of progenitors more accurately than morphological classes. It is noteworthy that the differences we found between states within any precursor cell type are subtle, both in terms of the number of DEX genes and the magnitude of differences in expression levels. Therefore, we suspect that the states are dynamic or cyclic in nature. However, the identification of developmental streams and the differential involvement of precursor cells, at the level of their cell states, suggest that the differences we observed may be instructive for neuronal differentiation and the establishment of neuronal diversity. While still

preliminary, this observation sheds new light on the complexity of neuronal precursor populations and invites further investigation.

Transcriptional priming may underlie some aspects of the diversification of cell types and the complexity of precursor dynamics both within and between species. Regulatory mechanisms driving precursor diversity have long been known to include temporal maturation and epigenetic modification (*23*, *30*, *31*). Recently, transcriptional priming, or the accumulation of untranslated mRNAs preceding the staged expression of protein, has been linked to specification of neuronal subtypes of daughter cells (*23*). Here, we show that transcriptional priming of the transcription factor Eomes may be a driver of precursor and lineage diversity, consistent with another report describing changes in neuronal identity and localization in response to deletion of the *Eomes* gene (*32*). We observed, in particular, that subgroups of apical precursors within mouse and human VZ express *Eomes* mRNA but not protein. Fate mapping with the Eomes regulatory sequence supported this conclusion as we observed multiple precursor morphotypes, including aRGCs, which lacked immunoreactivity for the Eomes protein. This suggests that the *Eomes* transcript itself contains regulatory motifs that are not found in the mRNAs encoding our fluorescent reporters and that these properties allow for repression of Eomes translation in these precursors. miRNA-92b has been shown to bind to and regulate *Eomes* mRNA (*33*); its activity may be important in regulating apical-to-basal precursor transition within this lineage. Together, these results suggest that transcriptional profile diversity is seeded in the broader population of aRGC stem cells and is then further amplified by the progression of individual neuron-producing lineages. This may be comparable to, and potentially allows inferences to be drawn from, similar mechanisms underlying lineage diversification in the development of organ systems other than the brain (*34*).

bRGCs and neocortical growth
While several studies have noted small numbers of cells in the mouse neocortex that resemble bRGCs (*12*, *13*, *28*, *35*), they are thought to have lower proliferative capacity and unique gene expression profiles compared with those found in species with convoluted or gyrencephalic brains, such as ferret and human. These findings have prompted the theory that bRGCs have enabled the large expansion in cortical surface area during carnivore and primate evolution (*11*). However, here, we have identified and developed specific labeling tools for multiple mouse bRGCs as well as the aRGC group that precedes them. The presence of these subtypes of bRGCs in both the mouse and human neocortex provides a new facet to the theory that bRGC underlie gyrencephalization and cortical expansion. The bRGC subtypes we identified express similar gene and protein expression patterns in both species, clearly indicating their conservation during mammalian evolution. These cell types found in both human and mouse should be called bRGCs and not oRGs because the mouse neocortex does not contain an oSVZ and because most of these cells are found in the human iSVZ. However, our study also confirms the presence of a human-specific type of bRGC (*SOX2*-only, hRGC3) that is present in the oSVZ and has been previously named the oRG cell (*11*, *27*). The presence of this oRG cell type in the human brain may lead to the expansion of the human neocortex.

A new framework for neuronal specification
The demonstration of multiple progenitor cell types and states supports a model whereby progenitor diversity yields neuronal diversity.

In general, while a brain with more progenitor cells has a larger growth capacity, a brain with greater progenitor diversity may have important additional qualities. Formation of the neocortex by multiple groups of dividing progenitors provides a varied landscape from which individual neuron properties can be germinated; our scRNA-seq and in vivo fate mapping studies now confirm that such a varied landscape is present in both mouse and human neocortex. Our data indicate the existence of coherent molecular signatures that constitute a thread tying cells within the same lineage together, even though individual cells within the lineage express various levels of such "threading genes" along the path of differentiation and may be found in various morphological states. This lineage diversity is likely to be integral to cortical complexity. For example, several reports have indicated that the morphology and action potential firing properties of the eventually produced neurons can vary, even in the same cortical layer (i.e., neurons born at the same developmental time) (5, 36). These differences appear to correlate with the properties of the neuron's parent cell type; we recently showed in mouse brain that progenitor lineage directly contributes to the intralaminar diversity of neurons both in the somatosensory and the frontal cortex (14, 15).

Overall, the bioinformatics results presented here identify unique expression profiles of progenitor diversity. These profiles then served as guides for in vivo fate mapping experiments that, in turn, reinforced and clarified the bioinformatics findings. The confirmed molecular properties can now be used to track the developmental roles of the multiple progenitor types undergoing neurogenesis in the fetal brain. Building information from fate maps of particular precursor lineages into a system-level analysis of neocortical circuitry and function will greatly elucidate how the neocortex is generated and how it is altered in neurodevelopmental disorders.

MATERIALS AND METHODS
Animals
Cohorts of timed pregnant CD-1 IGS mice (#022) were obtained from Charles River at E9.5 or E10.5 stage and were maintained at the Boston University Laboratory Animal Science Center with a 12-hour light/dark cycle in conventional housing cages until surgery.

Single-cell capture and library preparation
The ddSEQ Single-Cell Isolator (Bio-Rad and Illumina) was used for single-cell capture. Briefly, freshly dissected mouse brain tissue was transferred into a tube with 37°C prewarmed trypsin solution. The mix was pipetted with a wide-bore pipette tip for 10 times, and then the tube was incubated in a 37°C water bath for 30 min. The mix was gently pipetted 10 to 20 times every 10 min during incubation. After incubation, the cell suspension was repipetted again until homogenized. Last, the well-dissociated cell suspension was centrifuged at 300g for 3 min, washed, and resuspended in Dulbecco's phosphate-buffered saline (DPBS) to about 1000 cells/μl. The cells were loaded onto the four wells of the ddSEQ Single-Cell Isolator following the manufacturer's protocol to generate cDNA and sequencing libraries. The capture experiment was conducted in two separate technical replicates.

Library sequencing
cDNA and sequencing library concentration was quantified with Quant-iT PicoGreen (Invitrogen, P7589). All sequencing libraries were assessed for quality by Agilent Bioanalyzer using high-sensitivity double stranded DNA (dsDNA) assay. Library was sequenced on an Illumina NextSeq 500 platform with the pair-end mode following the manufacturer's instructions.

Read alignment and quality control
About 305 million (M) were generated, and FastQC was used to assess the quality of reads. Reads with average quality less than 30 were removed. We used SureCell RNA Single-Cell v1.1.0 (Illumina) to align all the reads in FastQC files to the mouse genome reference (mm10). Over 270M reads (89.58%) were aligned, with 260M that had valid barcode (85.6%). Among the aligned reads, 56.15% mapped to unique genes, whereas only 0.17% mapped to mitochondrial chromosome. We then assessed the distribution of unique molecular identifiers (UMI) in each cell as knee plot (fig. S1A) and removed barcodes with low UMI counts, with 16,681 barcodes remaining. We further removed low-quality barcodes with less than 200 total UMI counts, and 5777 cells passed quality control. UMI counts were normalized by NormalizeData function with log transformation using natural log as base.

Removal of CC effect
The removal of CC effect was performed similarly as described before (20). Briefly, to minimize the effect of CC in the identification of progenitor cell types, we sought to remove CC from our data through regression. Briefly, we used a published list of CC genes (37) and calculated G_1/S and G_2/M phase scores for each cell using function CellCycleScoring from R package Seurat (38). Then, we calculated the difference between G_1/S phase score and G_2/M phase score. This result was used to perform regression on all cells in our dataset with Seurat. Using this approach, CC differences among dividing cells were regressed out, while signals segregating cycling and noncycling cells were maintained.

Defining HVGs
To define HVGs, we calculated the mean of logged expression values using Seurat function FindVariableGenes and plotted it against variance to mean expression level ratio (VMR) for each gene. Genes with log-transformed mean expression level between 1 and 8.5 and VMR above were considered as HVGs.

Dimension reduction and clustering
We used PCA and t-distributed SNE (19) as our main dimension reduction approaches. PCA was performed with RunPCA function (Seurat) using HVGs. Following PCA, we conducted JACKSTRAW analysis with 100 iterations to identify statistically significant ($P < 0.01$) PCs that were driving systematic variation. We used t-SNE to present data in 2D coordinates, generated by RunTSNE function in Seurat. Significant PCs identified by JACKSTRAW analysis were used as input. Perplexity was set to 30. t-SNE plots were generated using R package ggplot2 (39). Clustering was done with the Luvain-Jaccard algorithm using t-SNE coordinates by FindClusters function from Seurat with default setting.

WGCNA and modular characterization
WGCNA (20) was performed using R package WGCNA. The UMI counts from all cells were used to generate correlation matrix with bicorrelation algorithm. Next, pickSoftThreshold function was used to analyze the network topology with 3 as soft-threshold power. Minimum size of modules was set to 10 genes. Module was identified using the "tree" method with deepSplit. For each module, WGCNA generated an eigengene to represent modular features.

Network edge and node information of each module were exported using exportNetworkToCytoscape function and was visualized with Cytoscape software.

Applying mouse WGCNA modules to human single cells

The human authologs of the mouse genes from each module were selected. The expression of the selected genes was used in PCA, and the first PC was used as the module eigengenes.

Pseudotime reconstruction

We used R package Monocle3 alpha to reconstruct pseudotime on each of the analyzed progenitor cell type separately (40), following standard procedure with customized parameters. Briefly, we first calculated the dispersion of each gene and calculated an estimated dispersion by the mean-variance model using dispersionTable function. Only genes with dispersion greater than the estimated value and mean UMI greater than 0.1 were kept for further analysis. Then, preprocessing was conducted where the expression levels were log normalized with a residual model using number of UMI as the independent variable, followed by PCA (number of dimensions set to 30). UMAP was used to reduce dimensions further to two, with Minkowski metric. The number of neighbors was determined empirically based on the number of single cells in each cell type (ranging between 5 and 20). To identify states within each cell type, Louvain-Jaccard clustering was conducted. Ridge plots as in Figs. 2C and 3B were generated with ggridges package. Plot_pseudotime_heatmap function was used to visualize gene expression levels across pseudotime as in fig. S5 (C and D) and fig. S6 (A to D), with number of clusters set to 3 and a natural spline function with degree of freedom equal to 2.

Data integration and metaneighbor analysis

The integration of human and mouse datasets was conducted following recommended steps (41). We used the top 2000 most variable features from each of the mouse and human single-cell datasets to find integration anchors. UMAP analysis was conducted with Minkowski metric. MetaNeighbor analysis was then conducted on the integrated data, using average expression levels of cell types from either species. Enrichment analysis for genes associated with human developmental and neurodegenerative disorders in human and mouse progenitor cell types was performed as previously described using genes specific to each human and mouse precursor cell type pairs (20). To identify genes specific to each pair of precursor cell types between human and mouse, we first conducted DEX analysis between cell types within same species to find genes specific to each human or mouse cell type. Then, we intersected the lists of DEX genes from mouse and human cell types of the same pair and regarded the intersected list as genes specific to the pair.

Pseudodifferentiation lineage analysis

To identify neuronal differentiation lineages, we applied STREAM (v0.4.0) analysis pipeline (42) to a subset of our single-cell transcriptome dataset containing mRGCs, mIPCs, and mExNs. Briefly, the top 20 PCs from the selected cells were used to create a diffusion map with diffusion scale parameter of the Gaussian kernel (Sigma-Aldrich) set to 1 and number of nearest neighbors set to square root of the number of cells (43). The first three eigenvectors of the diffusion map were passed to STREAM pipeline, and differentiation lineages were identified by seed_elastic_principal_graph function

(with number of initial nodes set to 10) followed by elastic_principal_graph function. To present the lineages, we used R package URD following recommended steps with minor adjustments based on the structure of the dataset (44). Briefly, mRGC1 was set as the root, and the diffusion map was flooded 1000 times to establish the pseudo-development axis. Tips of the diffusion map were identified from the final stage of pseudodevelopment. Biased random walks were then performed from each tip. Last, a tree graph was built using buildTree with cells per bin set to 25 and bins per window set to 8. A 2D representation of the resulted 3D graph from buildTree function was produced using R package rgl with a manually selected view point.

Data visualization

t-SNE plots were generated using TSNEPlot function from R package Seurat. Unless otherwise noted, all heatmaps were generated with R function heatmap.3. All other plots were generated using ggplot2.

In utero electroporation

IUE was performed as described previously (14) on E11.5, E13.5, and E14.5 timed pregnancies. Briefly, dams were anesthetized with ketamine/xylazine cocktail, and the uterine horns were exposed by a midline laparotomy. One to two microliters of plasmid, or plasmid combination, mixed with 0.1% fast green dye (Sigma-Aldrich) in phosphate buffer was injected into the lateral ventricles using a pulled glass micropipette and a picoinjector (PLI-100, Harvard Apparatus). Final plasmid or plasmid mixture concentration was between 3 and 6 μg/μl. The anode of a tweezertrode (1-mm diameter for E11.5, 3-mm diameter for E13.5, and 5 mm diameter for E14.5, Harvard Apparatus) was placed over the dorsal telencephalon above the uterine muscle, and four pulses (50 V for E11.5, 35 V for E13.5, and E14.5, 50-ms duration separated by 950-ms intervals) were applied with a BTX ECM830 square pulse generator (Harvard Apparatus). Following electroporation, the uterine horns were replaced into the abdomen, and the cavity was filled with warm 0.9% saline before suturing the abdominal muscle and skin separately. Dams were then placed into a clean cage for recovery and monitoring. These procedures were reviewed and approved by the Institutional Animal Care and Use Committee at Boston University School of Medicine.

Immunostaining and imaging

For embryonic studies, the heads of electroporated embryos were harvested 24 hours after IUE, fixed overnight in 4% paraformaldehyde (PFA), and cryoprotected in 30% sucrose for 24 to 48 hours, or the brains were removed and cut into 60-μm vibratome sections for morphometric analysis. Cryoprotected tissue was frozen in OCT compound in tissue molds with an ethanol/dry ice bath. Frozen tissue was cut into 20 μm sections using an HM560 Cryostar cryostat and mounted and dried on to superfrost slides (Thermo Fisher Scientific). For all IUE studies, we used $n = 4$ brains. Before immunostaining, antigen retrieval was performed by microwaving sections in sodium citrate buffer [10 mM (pH 6)] at 800 W for 1 min followed by 80 W for 10 min. Sections were then blocked in diluent [5% goat serum, 0.3% Triton X-100, 1× phosphate-buffered saline (PBS)] for 1 hour at room temperature. Incubation with primary antibodies, anti-Sox2 (1:200; Santa Cruz), or anti-Tbr2 (1:250; Abcam), or anti-PH3 (1:300; Millipore) was performed overnight at 4°C. Following three 5-min washes in PBS, sections were incubated for 2 hours at room temperature in diluent containing the appropriate

secondary antibodies (1:250 for all). Sections were washed an additional three times for 5 min and mounted with Vectashield mounting medium containing 4′,6-diamidino-2-phenylindole (DAPI). Then, 40× Z-stack images (15 μm for cryosections, 20 μm for vibratome sections) were acquired using an upright Zeiss LSM 710 microscope at a minimum of 1024 × 1024 resolution, and positive cells were identified and counted using LSM image browser software. Counts for each experiment were then averaged.

RNAScope

In situ hybridization for mRNA expression in the fixed-frozen mouse and human tissues was performed using RNAscope Multiplex Fluorescent Reagent Kit V2 (catalog no. 323100, Advanced Cell Diagnostics Inc., Hayward, CA) according to the manufacturer's instructions. The commercially available human and mouse RNA probes were as follows: Hs-Eomes-C2 (429691-C2), Hs-Sox2 (400871), Mm-NeuroD4-O1-C2 (564191-C2), and Mm-Eomes-C2 (19078C), all from Advanced Cell Diagnostics Inc. The human probe Hs-NeuroD4-No-XMmRn (584701) was custom designed based on the NM_021191.3 cDNA sequence, targeting the region of 3180 to 3927 base pairs (bp) that does not overlap with other NeuroD genes.

Briefly, after pretreating of the section with hydrogen peroxide, followed by antigen retrieval (solution of 1× target retrieval, at 800 W for 1 min followed by 80 W for 10 min) and protease 3 pre-treatment, the probes were incubated with the brain slices and the fluorescent dye for each detection channel was assigned as recommended. The signal was amplified using the multiplex reagents following the instructions.

The integrity of the RNA was confirmed in each tissue section using the RNAscope 3-plex Positive Control Probe for housekeeping gene expression (catalog no. 320861 for human tissue and catalog no. 320881 for mouse tissue). To confirm signal specificity, RNAscope 3-plex Negative Control Probe (catalog no. 320871 for human and mouse tissues) was used. After the final amplification, the slides were treated with Hoechst 33342 (catalog no. H3570, Invitrogen, Carlsbad, CA) for 10 min and sealed with coverslips using ProLong Gold Antifade Reagent (catalog no. P36930, Invitrogen, Carlsbad, CA).

Z-stack 20× images (16 μm) were captured using a Zeiss LSM 710 microscope. Positive cells were identified and counted (four to five sections; $N = 1$ human brain) by two independent investigators using the LSM image browser software. The percentage of Eomes-, NeuroD4-, and Sox2-positive/double positive cells was assessed by normalizing the number to the total number of cells as marked by Hoechst nuclei staining. The laminar distribution analysis was performed using LSM image browser and Volocity software, and the results were plotted graphically using Sigma-Aldrich plot.

Statistical analysis

DEX analysis was conducted using Seurat function FindAllMarkers. Briefly, we took one group of cells and compared it with the rest of the cells using a Wilcoxon rank sum test. For any given comparison, we only considered genes that were expressed by at least 25% of cells in either population. Genes that exhibit Bonferroni-corrected P values under 0.01 were considered statistically significant.

To identify state-specific genes in mRGCs or mIPCs, we conducted principal graph test with nearest neighbor (k) parameter set to 10 (40). Genes expressed by more than 10% of cells in the state of interest and Moran's I greater than 0.1 were considered as state specific. To generate gene expression heatmap over pseudotime, we used the *plot_pseudotime_heatmap* function from Monocle 3 Alpha with top 10 state-specific genes as input. A natural spline model was used to describe gene expression as a function of pseudotime, with degree of freedom set to 2.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/45/eabd2068/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. R. D. Hodge, T. E. Bakken, J. A. Miller, K. A. Smith, E. R. Barkan, L. T. Graybuck, J. L. Close, B. Long, N. Johansen, O. Penn, Z. Yao, J. Eggermont, T. Höllt, B. P. Levi, S. I. Shehata, B. Aevermann, A. Beller, D. Bertagnolli, K. Brouner, T. Casper, C. Cobbs, R. Dalley, N. Dee, S.-L. Ding, R. G. Ellenbogen, O. Fong, E. Garren, J. Goldy, R. P. Gwinn, D. Hirschstein, C. D. Keene, M. Keshk, A. L. Ko, K. Lathia, A. Mahfouz, Z. Maltzer, M. M. Graw, T. N. Nguyen, J. Nyhus, J. G. Ojemann, A. Oldre, S. Parry, S. Reynolds, C. Rimorin, N. V. Shapovalova, S. Somasundaram, A. Szafer, E. R. Thomsen, M. Tieu, G. Quon, R. H. Scheuermann, R. Yuste, S. M. Sunkin, B. Lelieveldt, D. Feng, L. Ng, A. Bernard, M. Hawrylycz, J. W. Phillips, B. Tasic, H. Zeng, A. R. Jones, C. Koch, E. S. Lein, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).

2. M. Li, G. Santpere, Y. I. Kawasawa, O. V. Evgrafov, F. O. Gulden, S. Pochareddy, S. M. Sunkin, Z. Li, Y. Shin, Y. Zhu, A. M. M. Sousa, D. M. Werling, R. R. Kitchen, H. J. Kang, M. Pletikos, J. Choi, S. Muchnik, X. Xu, D. Wang, B. Lorente-Galdos, S. Liu, P. Giusti-Rodríguez, H. Won, C. A. de Leeuw, A. F. Pardiñas; BrainSpan Consortium; PsychENCODE Consortium; PsychENCODE Developmental Subgroup, M. Hu, F. Jin, Y. Li, M. J. Owen, M. C. O'Donovan, J. T. R. Walters, D. Posthuma, M. A. Reimers, P. Levitt, D. R. Weinberger, T. M. Hyde, J. E. Kleinman, D. H. Geschwind, M. J. Hawrylycz, M. W. State, S. J. Sanders, P. F. Sullivan, M. B. Gerstein, E. S. Lein, J. A. Knowles, N. Sestan, Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).

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

4. S. C. Noctor, A. C. Flint, T. A. Weissman, R. S. Dammerman, A. R. Kriegstein, Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**, 714–720 (2001).

5. T. J. Ellender, S. V. Avery, K. Mahfooz, J. Scaber, A. von Klemperer, S. L. Nixon, M. J. Buchan, J. J. van Rheede, A. Gatti, C. Waites, H. J. Pavlou, D. Sims, S. E. Newey, C. J. Akerman, Embryonic progenitor pools generate diversity in fine-scale excitatory cortical subnetworks. *Nat. Commun.* **10**, 5224 (2019).

6. J. S. Gal, Y. M. Morozov, A. E. Ayoub, M. Chatterjee, P. Rakic, T. F. Haydar, Molecular and morphological heterogeneity of neural precursors in the mouse neocortical proliferative zones. *J. Neurosci.* **26**, 1045–1056 (2006).

7. E. K. Stancik, I. Navarro-Quiroga, R. Sellke, T. F. Haydar, Heterogeneity in ventricular zone neural precursors contributes to neuronal fate diversity in the postnatal neocortex. *J. Neurosci.* **30**, 7028–7036 (2010).

8. W. A. Tyler, T. F. Haydar, Multiplex genetic fate mapping reveals a novel route of neocortical neurogenesis, which is altered in the Ts65Dn mouse model of Down syndrome. *J. Neurosci.* **33**, 5106–5119 (2013).

9. G. E. Elsen, R. D. Hodge, F. Bedogni, R. A. M. Daza, B. R. Nelson, N. Shiba, S. L. Reiner, R. F. Hevner, The protomap is propagated to cortical plate neurons through an Eomes-dependent intermediate map. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4081–4086 (2013).

10. B. R. Nelson, R. D. Hodge, F. Bedogni, R. F. Hevner, Dynamic interactions between intermediate neurogenic progenitors and radial glia in embryonic mouse neocortex: Potential role in Dll1-Notch signaling. *J. Neurosci.* **33**, 9122–9139 (2013).

11. D. V. Hansen, J. H. Lui, P. R. Parker, A. R. Kriegstein, Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).

12. A. Shitamukai, D. Konno, F. Matsuzaki, Oblique radial glial divisions in the developing mouse neocortex induce self-renewing progenitors outside the germinal zone that resemble primate outer subventricular zone progenitors. *J. Neurosci.* **31**, 3683–3695 (2011).

13. X. Wang, J.-W. Tsai, B. LaMonica, A. R. Kriegstein, A new subtype of progenitor cell in the mouse embryonic neocortex. *Nat. Neurosci.* **14**, 555–561 (2011).

14. W. A. Tyler, M. Medalla, T. Guillamon-Vivancos, J. I. Luebke, T. F. Haydar, Neural precursor lineages specify distinct neocortical pyramidal neuron types. *J. Neurosci.* **35**, 6142–6152 (2015).

15. T. Guillamon-Vivancos, W. A. Tyler, M. Medalla, W. W.-e. Chang, M. Okamoto, T. F. Haydar, J. I. Luebke, Distinct neocortical progenitor lineages fine-tune neuronal diversity in a layer-specific manner. *Cereb. Cortex* **29**, 1121–1138 (2019).

16. T. J. Nowakowski, A. Bhaduri, A. A. Pollen, B. Alvarado, M. A. Mostajo-Radji, E. D. Lullo, M. Haeussler, C. Sandoval-Espinosa, S. J. Liu, D. Velmeshev, J. R. Ounadjela, J. Shuga, X. Wang, D. A. Lim, J. A. West, A. A. Leyrat, W. J. Kent, A. R. Kriegstein, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

17. S. Zhong, S. Zhang, X. Fan, Q. Wu, L. Yan, J. Dong, H. Zhang, L. Li, L. Sun, N. Pan, X. Xu, F. Tang, J. Zhang, J. Qiao, X. Wang, A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).

18. X. Fan, J. Dong, S. Zhong, Y. Wei, Q. Wu, L. Yan, J. Yong, L. Sun, X. Wang, Y. Zhao, W. Wang, J. Yan, X. Wang, J. Qiao, F. Tang, Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* **28**, 730–745 (2018).

19. L. van der Maaten, G. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

20. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

21. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

22. A. Sessa, C. A. Mao, A. K. Hadjantonakis, W. H. Klein, V. Broccoli, Tbr2 directs conversion of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in the developing neocortex. *Neuron* **60**, 56–69 (2008).

23. S. K. Zahr, G. Yang, H. Kazan, M. J. Borrett, S. A. Yuzwa, A. Voronova, D. R. Kaplan, F. D. Miller, A translational repression complex in developing mammalian neural stem cells that regulates neuronal specification. *Neuron* **97**, 520–537 e526 (2018).

24. M. B. Johnson, Y. I. Kawasawa, C. E. Mason, Ž. Krsnik, G. Coppola, D. Bogdanović, D. H. Geschwind, S. M. Mane, M. W. State, N. Šestan, Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).

25. G.-A. Pilz, A. Shitamukai, I. Reillo, E. Pacary, J. Schwausch, R. Stahl, J. Ninkovic, H. J. Snippert, H. Clevers, L. Godinho, F. Guillemot, V. Borrell, F. Matsuzaki, M. Götz, Amplification of progenitors in the mammalian telencephalon includes a new radial glial cell type. *Nat. Commun.* **4**, 2125 (2013).

26. M. Crow, A. Paul, S. Ballouz, Z. J. Huang, J. Gillis, Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).

27. A. A. Pollen, T. J. Nowakowski, J. Chen, H. Retallack, C. Sandoval-Espinosa, C. R. Nicholas, J. Shuga, S. J. Liu, M. C. Oldham, A. Diaz, D. A. Lim, A. A. Leyrat, J. A. West, A. R. Kriegstein, Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).

28. J. Liu, W. Liu, L. Yang, Q. Wu, H. Zhang, A. Fang, L. Li, X. Xu, L. Sun, J. Zhang, F. Tang, X. Wang, The primate-specific gene TMEM14B marks outer radial glia cells and promotes cortical expansion and folding. *Cell Stem Cell* **21**, 635–649.e8 (2017).

29. F. K. Wong, J.-F. Fei, F. Mora-Bermúdez, E. Taverna, C. Haffner, J. Fu, K. Anastassiadis, A. F. Stewart, W. B. Huttner, Sustained Pax6 expression generates primate-like basal radial glia in developing mouse neocortex. *PLoS Biol.* **13**, e1002217 (2015).

30. L. Telley, G. Agirman, J. Prados, N. Amberg, S. Fièvre, P. Oberst, G. Bartolini, I. Vitali, C. Cadilhac, S. Hippenmeyer, L. Nguyen, A. Dayer, D. Jabaudon, Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364**, eaav2522 (2019).

31. K. J. Yoon, C. Vissers, G. L. Ming, H. Song, Epigenetics and epitranscriptomics in temporal patterning of cortical neural progenitor competence. *J. Cell Biol.* **217**, 1901–1914 (2018).

32. X. Lv, S.-Q. Ren, X.-J. Zhang, Z. Shen, T. Ghosh, A. Xianyu, P. Gao, Z. Li, S. Lin, Y. Yu, Q. Zhang, M. Groszer, S.-H. Shi, TBR2 coordinates neurogenesis expansion and precise microcircuit organization via Protocadherin 19 in the mammalian cortex. *Nat. Commun.* **10**, 3946 (2019).

33. T. J. Nowakowski, V. Fotaki, A. Pollock, T. Sun, T. Pratt, D. J. Price, MicroRNA-92b regulates the development of intermediate cortical progenitors in embryonic mouse brain. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7056–7061 (2013).

34. P. Laslo, C. J. Spooner, A. Warmflash, D. W. Lancki, H.-J. Lee, R. Sciammas, B. N. Gantner, A. R. Dinner, H. Singh, Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* **126**, 755–766 (2006).

35. S. Vaid, J. G. Camp, L. Hersemann, C. E. Oegema, A.-K. Heninger, S. Winkler, H. Brandl, M. Sarov, B. Treutlein, W. B. Huttner, T. Namba, A novel population of Hopx-dependent basal radial glial cells in the developing mouse neocortex. *Development* **145**, dev169276 (2018).

36. E. J. Kim, A. L. Juavinett, E. M. Kyubwa, M. W. Jacobs, E. M. Callaway, Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. *Neuron* **88**, 1253–1267 (2015).

37. D. Mi, Z. Li, L. Lim, M. Li, M. Moissidis, Y. Yang, T. Gao, T. X. Hu, T. Pratt, D. J. Price, N. Sestan, O. Marín, Early emergence of cortical interneuron diversity in the mouse embryo. *Science* **360**, 81–85 (2018).

38. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

39. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2016).

40. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

41. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

42. H. Chen, L. Albergante, J. Y. Hsu, C. A. Lareau, G. L. Bosco, J. Guan, S. Zhou, A. N. Gorban, D. E. Bauer, M. J. Aryee, D. M. Langenau, A. Zinovyev, J. D. Buenrostro, G.-C. Yuan, L. Pinello, Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat. Commun.* **10**, 1903 (2019).

43. P. Angerer, L. Haghverdi, M. Büttner, F. J. Theis, C. Marr, F. Buettner, Destiny: Diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2015).

44. J. A. Farrell, Y. Wang, S. J. Riesenfeld, K. Shekhar, A. Regev, A. F. Schier, Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).