Science Advances

Supplementary Materials for

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

Zhen Li*, William A. Tyler*, Ella Zeldich, Gabriel Santpere Baró, Mayumi Okamoto, Tianliuyun Gao, Mingfeng Li, Nenad Sestan*, Tarik F. Haydar*,

*Corresponding author. Email: thaydar@childrensnational.org (T.F.H.); nenad.sestan@yale.edu (N.S.)

The PDF file includes:

Figs. S1 to S9

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/45/eabd2068/DC1)

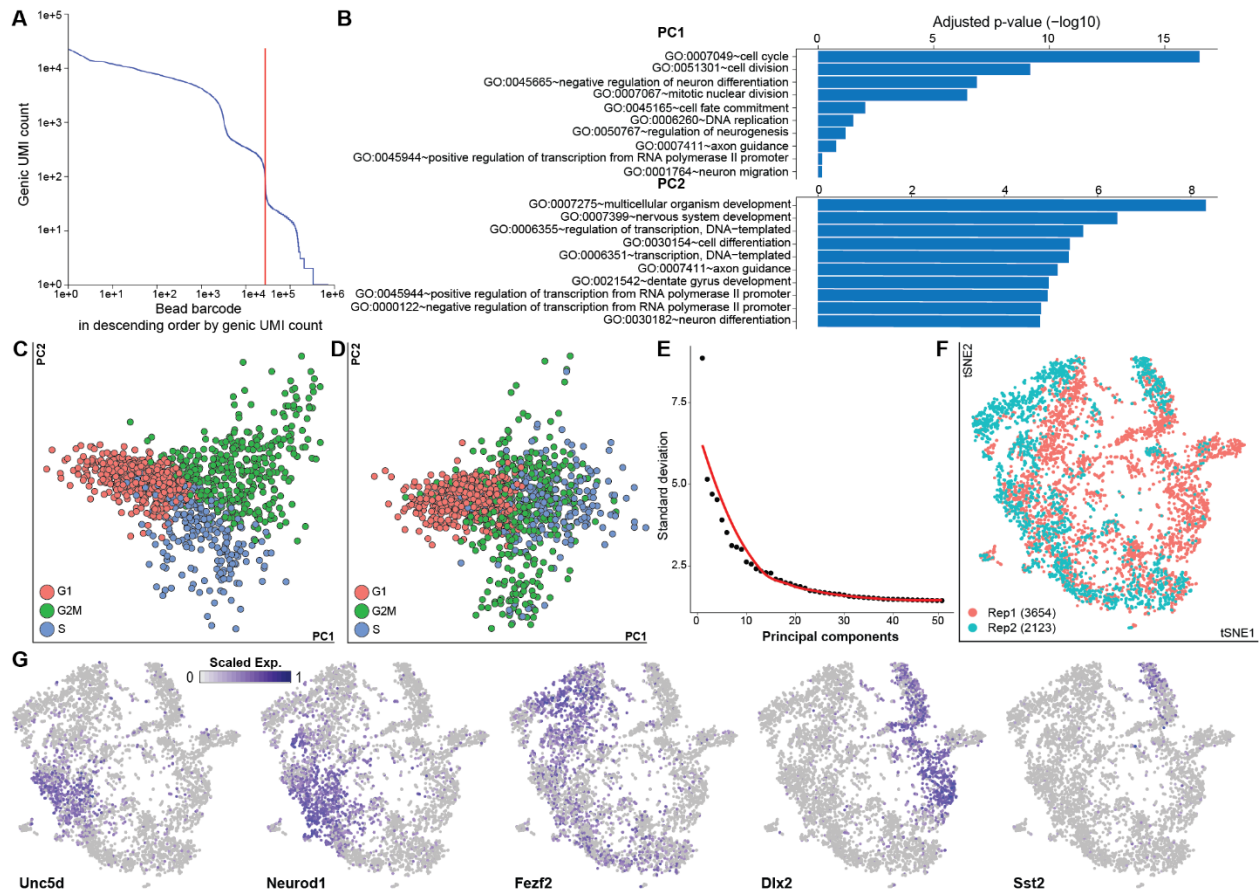Table S1

Supplementary Figures S1-S9

Figure S1 Single cell RNAseq quality control and preprocessing

(A)   Knee plot showing number of unique molecular identifier (UMI) per cell. Cells are arranged in descending order by the total number of UMIs. Cells to the left of the vertical red line are considered valid single cells and kept for downstream analysis. Both the horizontal and vertical axes are log scaled.

(B)   Gene ontology terms associated with top 100 genes related to principal component (PC) 1 and 2. Bar plot shows Bonferroni-corrected p-value (-log10).

(C-D) Scatterplot of first two principal components before cell cycle regression (D) and after cell cycle regression (E). The principal components are calculated with cell cycle genes.

(E)   Elbow plot of principal component analysis using highly variable genes in all mouse single cells. The first 50 principal components are shown. Red line shows local weighted scatterplot smoothing.

(F)   T-SNE plot as in Fig. 1B colored by technical replicates. The number of cells in each replicate is indicated in parentheses.

(G)   Feature plots of neuronal marker genes in t-SNE coordinates as in Fig. 1B. Color represent scaled normalized expression level (Scaled Exp.).

(H-J) Density distribution of *Eomes* expression levels in mouse radial glial cells (mRGCs) and intermediate progenitor cells (mIPCs) (G), as well as in mRGC1 and mRGC2 (H) and in

mIPC1-5 (I) separately. Only cells with non-zero *Eomes* mRNA expression (UMI > 0) are plotted. Horizontal axis shows log transformed UMI counts of *Eomes*.
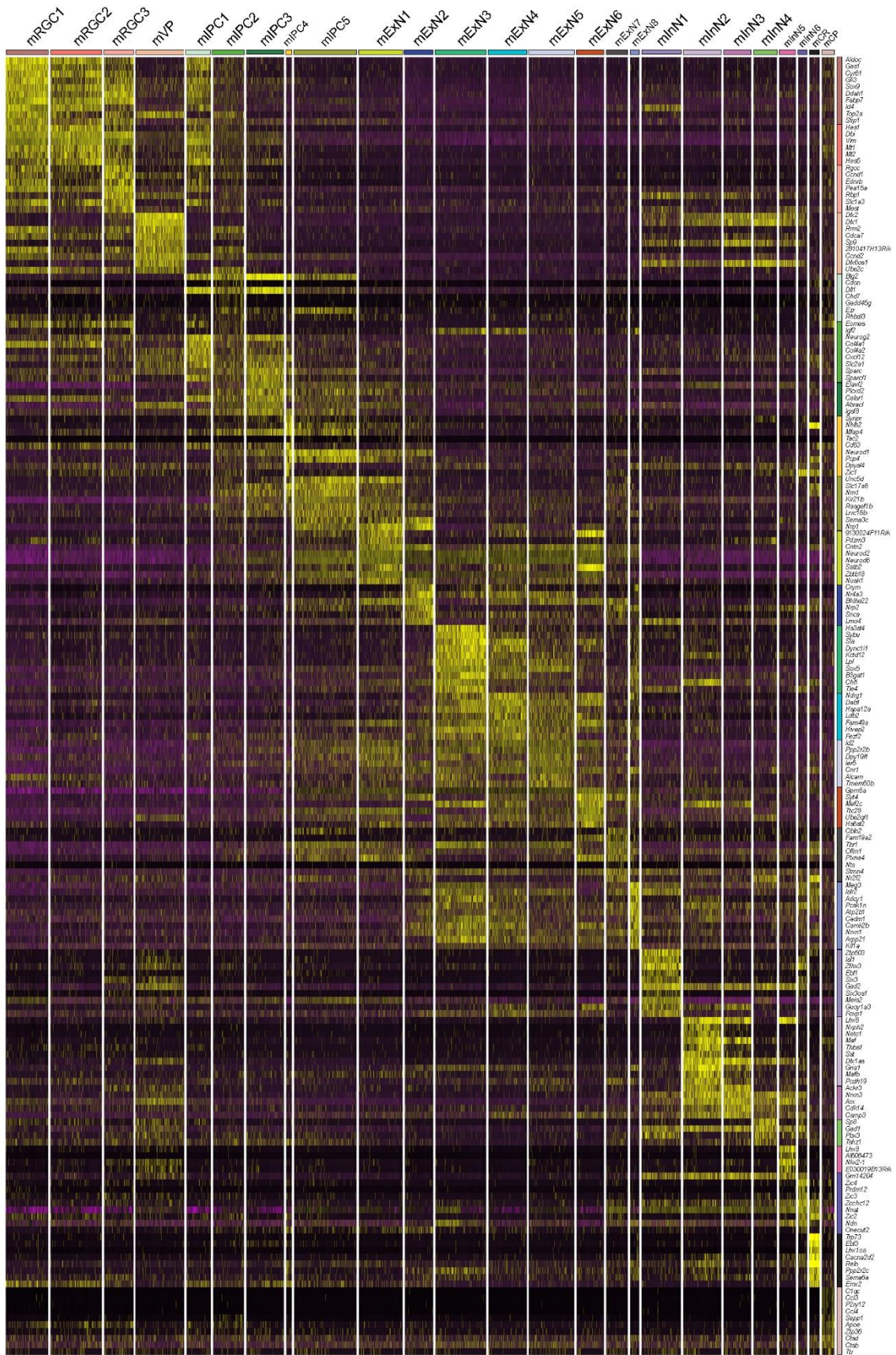
Figure S2 Heatmap of differentially expressed genes between mouse cell types. Colors represent scaled expression level of each gene. M, mouse; RGC, radial glial cell; VP, ventral progenitor; IPC, intermediate progenitor cell; ExN, excitatory neuron; InN, inhibitory neuron; CR, cajal-retzius cell; CP, choroid plexus.

Figure S3 Gene modules from weighted gene co-expression network analysis on mouse single cells. M, module; IPC, intermediate progenitor cell. InN, inhibitory neuron; CR, cajal-retzius cell; CP, choroid plexus.
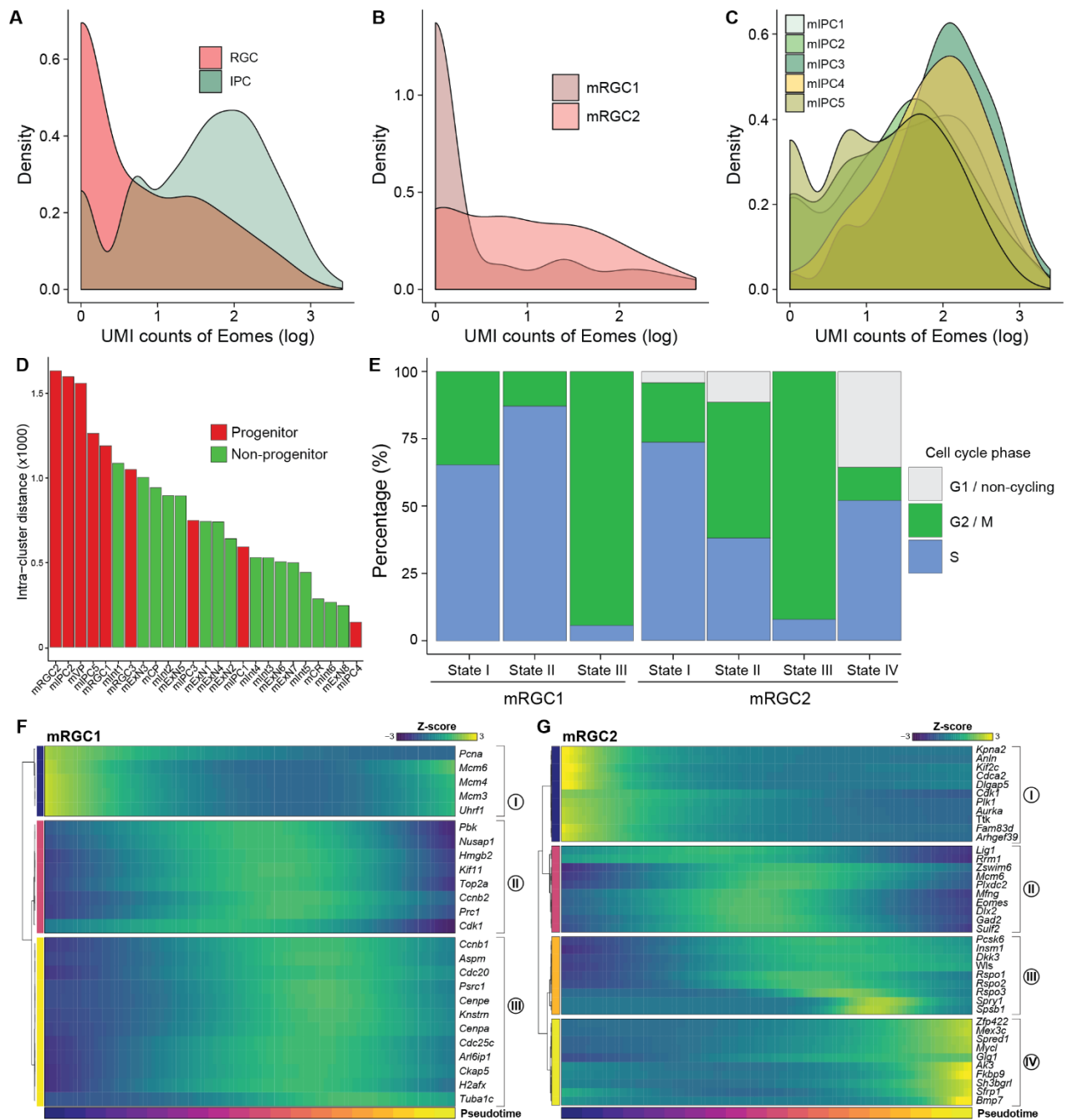
Figure S4 Pseudotime analysis of mouse radial glial cells.

(A-C)   Density distribution of *Eomes* expression levels in mouse radial glial cells (mRGCs) and intermediate progenitor cells (mIPCs) (G), as well as in mRGC1 and mRGC2 (H) and in mIPC1-5 (I) separately. Only cells with non-zero *Eomes* mRNA expression (UMI > 0) are plotted. Horizontal axis shows log transformed UMI counts of *Eomes*. M, mouse; RGC, radial glial cell; IPC, intermediate progenitor cell.

(D)   Bar plot showing inter-cluster distance of each mouse cell type. Progenitors are colored in red and differentiated cell types in green.

(E)   Bar plot showing percentages of cells in G1/non-cycling, G2/M or S cell cycle phase from each state of mRGC1 and 2 cell types. Colors represent cell cycle phases.

(F-G) Heatmap of genes that are highly associated with each of the identified cell states (circled Roman numerals) in cell type mRGC1 (F) and mRGC2 (G). Heatmap color represent z-score of gene expression levels. Dendrogram on the left shows clusters of genes based on expression pattern. Color bars on the left represent cell states as in Fig. 2A and B. Color bar at the bottom represent pseudotime.
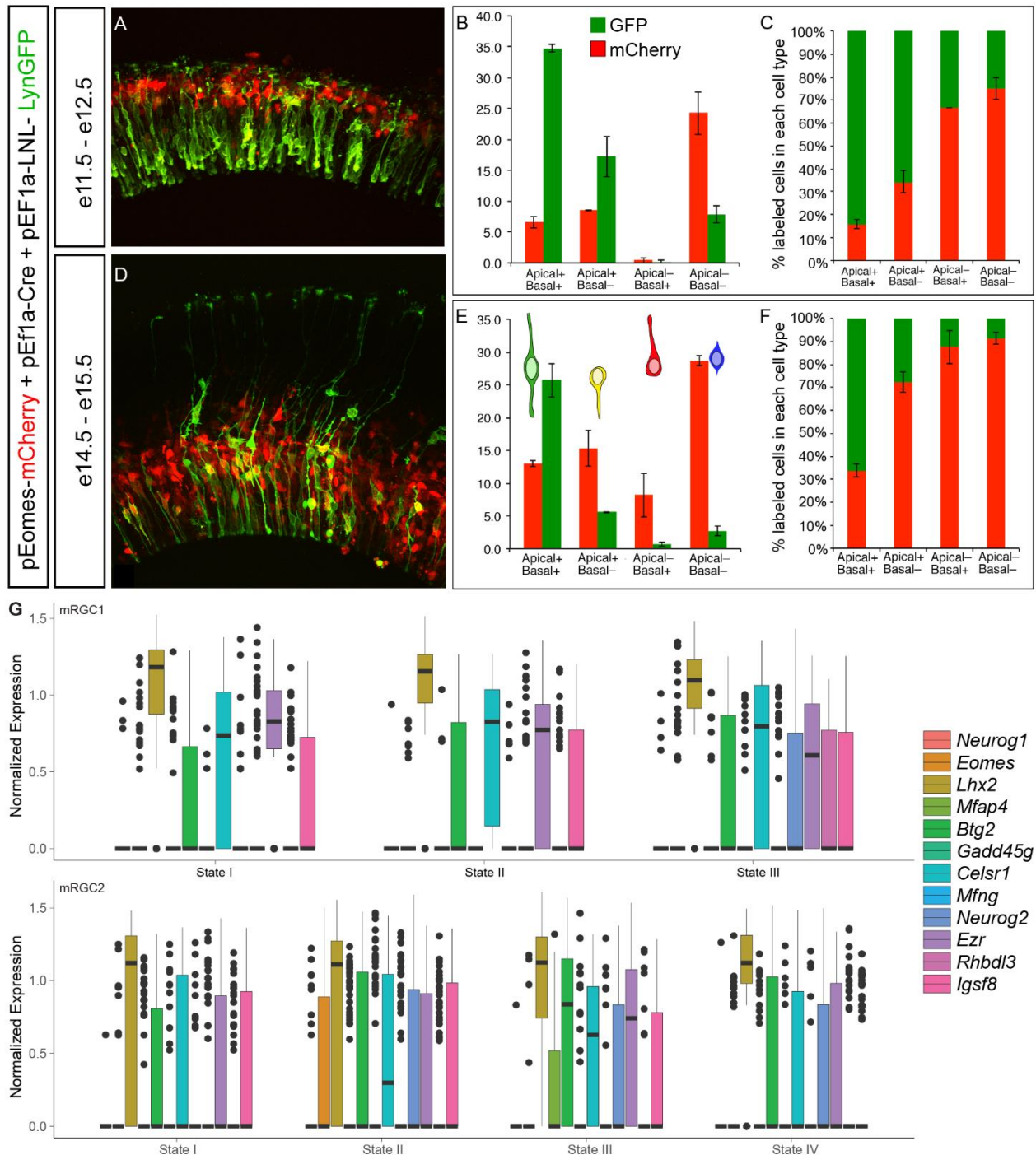
Figure S5 Morphology of *Eomes*-expressing ventricular zone and subventricular zone cells during mouse neurogenesis.

(A)   E11.5 neural precursors transfected for 24hrs pEomes-mCherry (red) and pEF1□-Cre+pEF1□-LNL-LynGFP (green).

(B)   Quantification of green (total) and red (Eomes-expressing) VZ and SVZ cells possessing apical and/or basal cell processes at e12.5 (morphologies illustrated in E; n=3).

(C)   Proportion of e12.5 VZ and SVZ cells with apical and/or basal cell processes by lineage (total green versus Eomes-expressing red).

(D)   e14.5 neural precursors transfected for 24hrs pEomes-mCherry (red) and pEF1□-Cre+pEF1□-LNL-LynGFP (green).

(E)   Quantification of green (total) and red (Eomes-expressing) VZ and SVZ cells possessing apical and/or basal cell processes at e15.5, n=3.

(F)   Proportion of e15.5 VZ and SVZ cells with apical and/or basal cell processes by lineage (total green versus Eomes-expressing red).

(G)   Bar plot showing expression of module 8 (M8) genes from weighted gene co-expression network analysis in each state of mRGC1 and 2. Colors represent different genes.
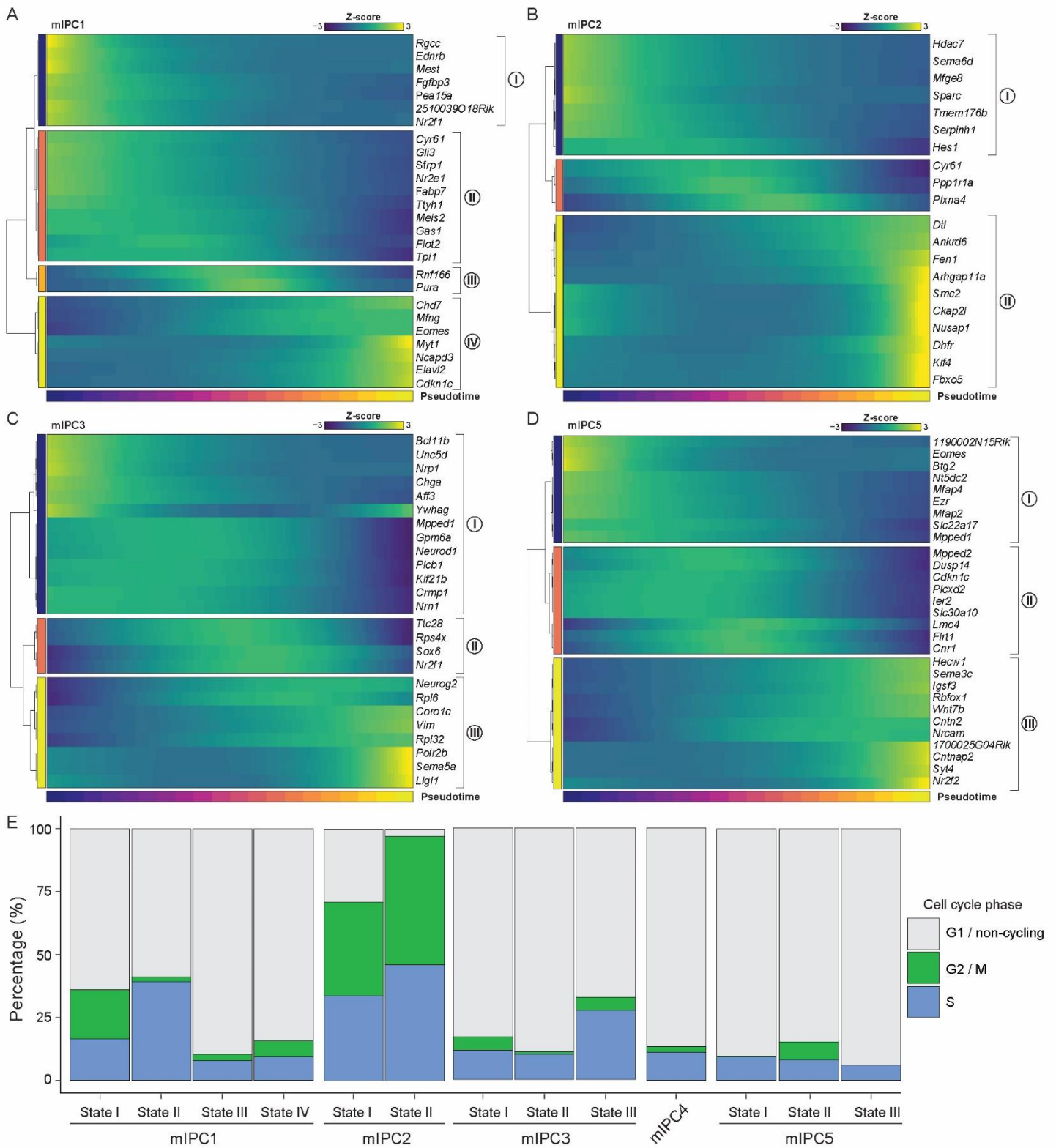
Figure S6 Pseudotime analysis of mouse intermediate progenitor cells

(A-D) Heatmap of genes that are highly associated with each of the identified cell states (circled Roman numerals) in cell type mIPC1, mIPC2, mIPC3 and mIPC5. Heatmap color represent z-score of normalized gene expression levels. Dendrogram shows clusters of genes based on expression pattern over pseudotime. Color bars on the left represent cell state as in Fig. 3F to I. Color bar at the bottom represent pseudotime.

(E) Bar plot showing percentages of cells in G1/non-cycling, G2/M or S cell cycle phase. Colors represent cell cycle phases.

A — pEomes-Flpe + pBlbp-FNF-Cre + pCAG-LNL-eGFP
eGFP
PH3

B — pEomes-Flpe + pHes1-FNF-Cre + pCAG-LNL-eGFP
eGFP
SOX2

B' — SOX2

C — pNeuroD4-Flpe + pSlc1a3-FNF-Cre + pCAG-LNL-eGFP
e14.5-15.5

D — pNeuroD4-Flpe + pHes1-FNF-Cre + pCAG-LNL-eGFP
e14.5-15.5

E, F — pEomes-Cre+ pCAG-LNL-eGFP; *Eomes* ISH; *NeuroD4* ISH
IZ

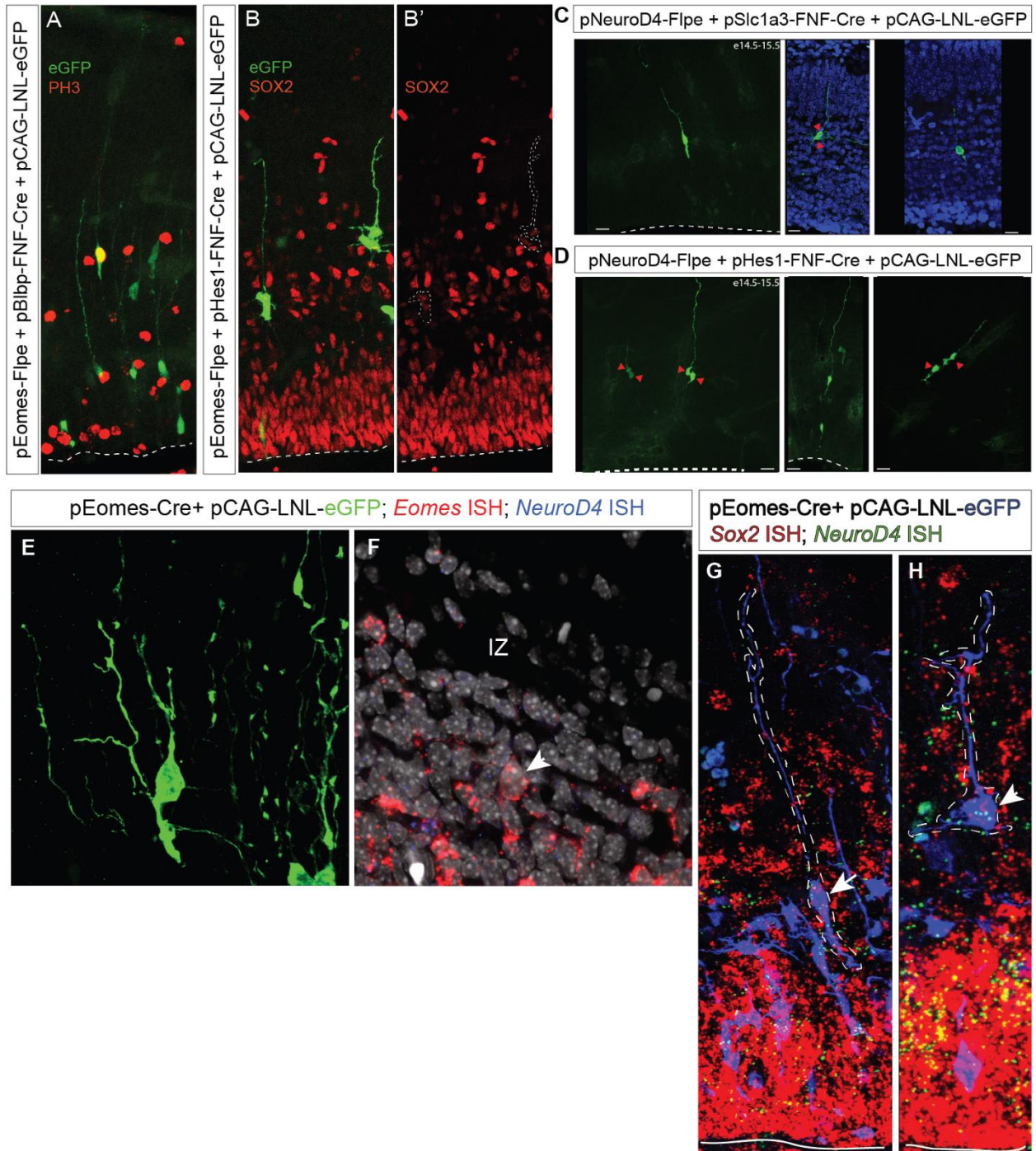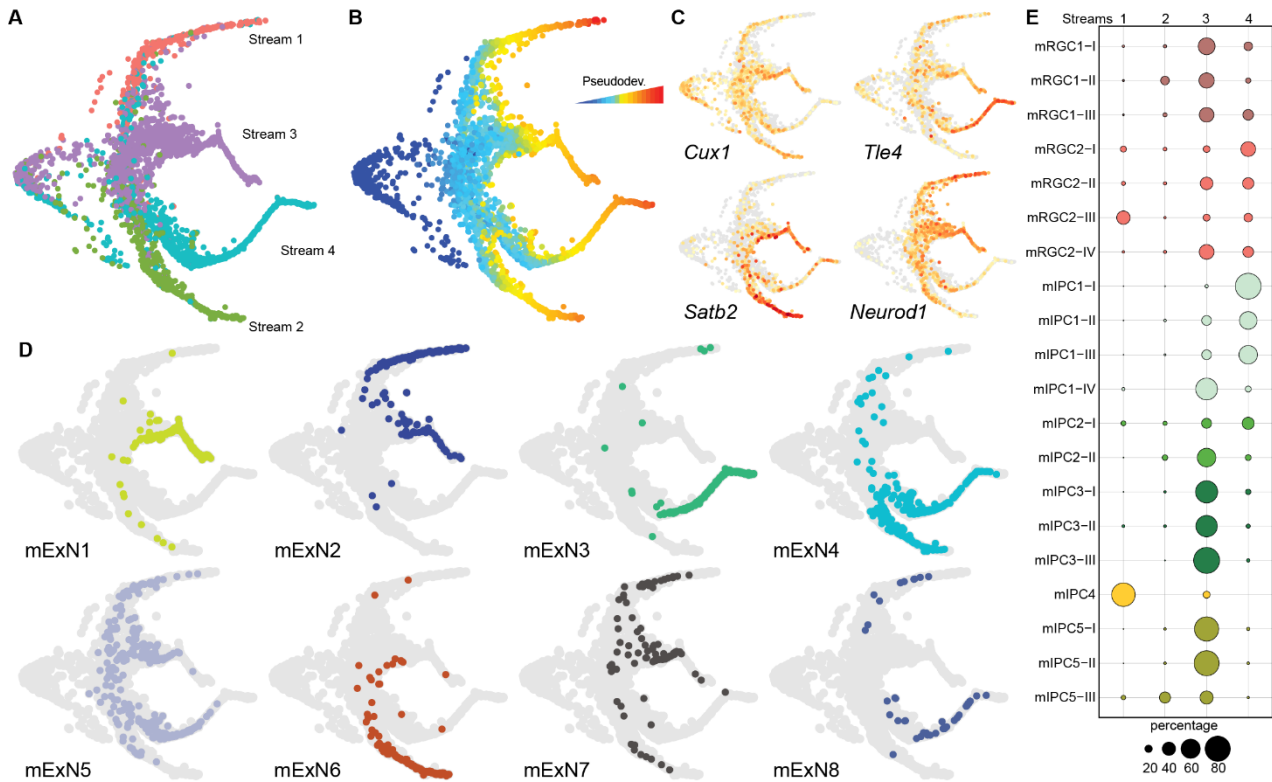G, H — pEomes-Cre+ pCAG-LNL-eGFP
Sox2 ISH; *NeuroD4* ISH

**Figure S7 Mouse basal radial glial cells are a proliferative population.**

(A)   Basal radial glial cells (bRGCs; green) from precursors co-expressing *Eomes* and *Blbp* (*Fabp7*) are pH3 positive (red).

(B)   bRGC and their putative daughter cells (green), derived from precursors co-expressing *Eomes* and *Hes1*, are Sox2+ (red).

(C,D) Precursors co-expressing *NeuroD4*/*Slc1a3* (green) 24hrs post IUE exhibit bRGC morphology and are found near putative daughter cells (red arrowheads).

(E)   bRGC cell (green) labeled with a Cre-conditional GFP reporter activated by *Eomes* promoter expression.

(F)   Same field in (A) stained with in situ probes (RNAScope) for *Eomes* and *NeuroD4*.

(G,H) In utero electroporation with pEomes-Cre + pCAG-LNL-eGFP (blue) co-labeled with RNAScope probes for *Sox2* (red) and *NeuroD4* (green). Arrows point to bRGCs. Ventricular surface is denoted by white dashed line.

A
Stream 1
Stream 3
Stream 4
Stream 2

B
Pseudodev.

C
Cux1   *Tle4*
Satb2   *Neurod1*

D
mExN1
mExN2
mExN3
mExN4
mExN5
mExN6
mExN7
mExN8

E

	Streams 1	2	3	4
mRGC1–I				
mRGC1–II				
mRGC1–III				
mRGC2–I				
mRGC2–II				
mRGC2–III				
mRGC2–IV				
mIPC1–I				
mIPC1–II				
mIPC1–III				
mIPC1–IV				
mIPC2–I				
mIPC2–II				
mIPC3–I				
mIPC3–II				
mIPC3–III				
mIPC4				
mIPC5–I				
mIPC5–II				
mIPC5–III				

percentage
20 40 60 80

F Type A: bipolar or unipolar ramified, no surface contact

or

G Type B: long unipolar, no surface contact

H Type C: unipolar, pial surface contact
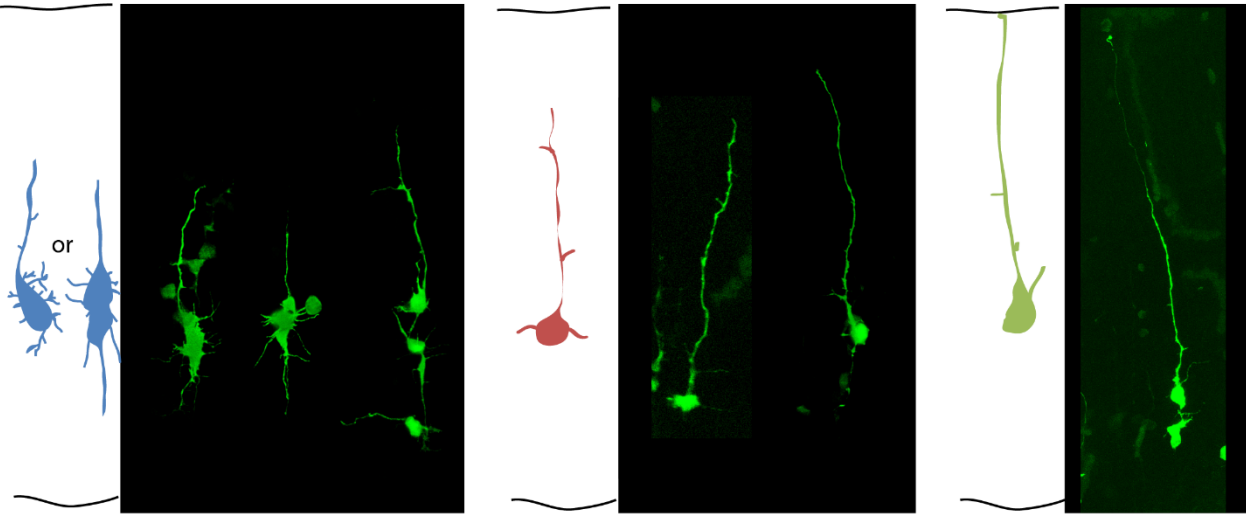
I pEomes-Flpe + pBlbp-FNF-Cre + CAG-LNL-eGFP
Type C / Type A / Type B

J pEomes-Flpe + pHes1-FNF-Cre + CAG-LNL-eGFP
Type C / Type A / Type B

K pEomes-Flpe + pSlc1a3-FNF-Cre + CAG-LNL-eGFP
Type C / Type A / Type B

L pNeuroD4-Flpe + pBlbp-FNF-Cre + CAG-SL
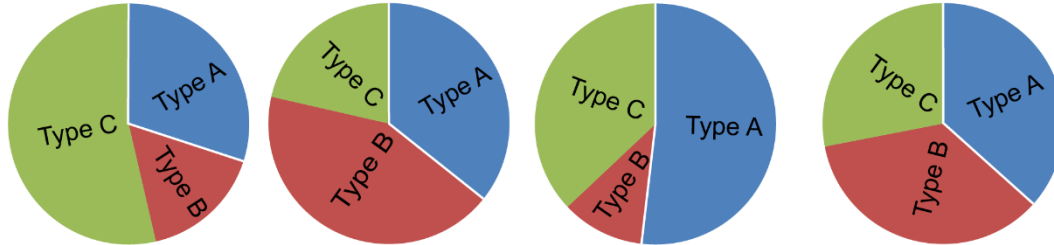Type C / Type A / Type B

Figure S8 Mouse neocortical excitatory neuron differentiation streams.

(A)   A scatter plot representation of the four mouse excitatory neuron streams. Cells from mRGC1-2, all mIPC and mExN cell types are included. Each dot represents a cell. Colors show different streams. RGC, radial glial cell; IPC, intermediate progenitor cell; ExN, excitatory neuron.

(B)   Same scatter plot as A colored by pseudodevelopment process.

(C)   Same scatter plot as A colored by normalized expression levels of *Cux1, Tle4, Satb2* and *Neurod1*.

(D)   Same scatter plot as A colored by mExN cell types. Same color scheme was used as in Fig. 1D.

(E)   Percentage of cells involved in each stream from all precursor states. Colors represent cell types (same color scheme as in Fig. 1D). Size of point represent percentage.

(F-H)   The three morphological subtypes of bRGCs: Type A cells (F) are bipolar with a basally or apically-directed process, neither of which extends to the pial or ventricular surface; Type B cells (G) are unipolar with a long basal process that does not contact the pial surface; Type C cells (H) are unipolar with a long basal process that touches the pial surface.

(I)   Pie chart illustrates distribution of bRGC morphologies with pEomes-Flpe + pBlbp-FNF-Cre + CAG-LNL-eGFP labeling strategy, n=4.

(J)   Pie chart illustrates distribution of bRGC morphologies with pEomes-Flpe + pHes1-FNF-Cre + CAG-LNL-eGFP labeling strategy, n=4.

(K)   Pie chart illustrates distribution of bRGC morphologies with pEomes-Flpe + pSlc1a3-FNF-Cre + CAG-LNL-eGFP labeling strategy, n=4.

(L)   Pie chart illustrates distribution of bRGC morphologies with pNeuroD4-Flpe + pBlbp-FNF-Cre + CAG-LNL-eGFP labeling strategy, n=4.
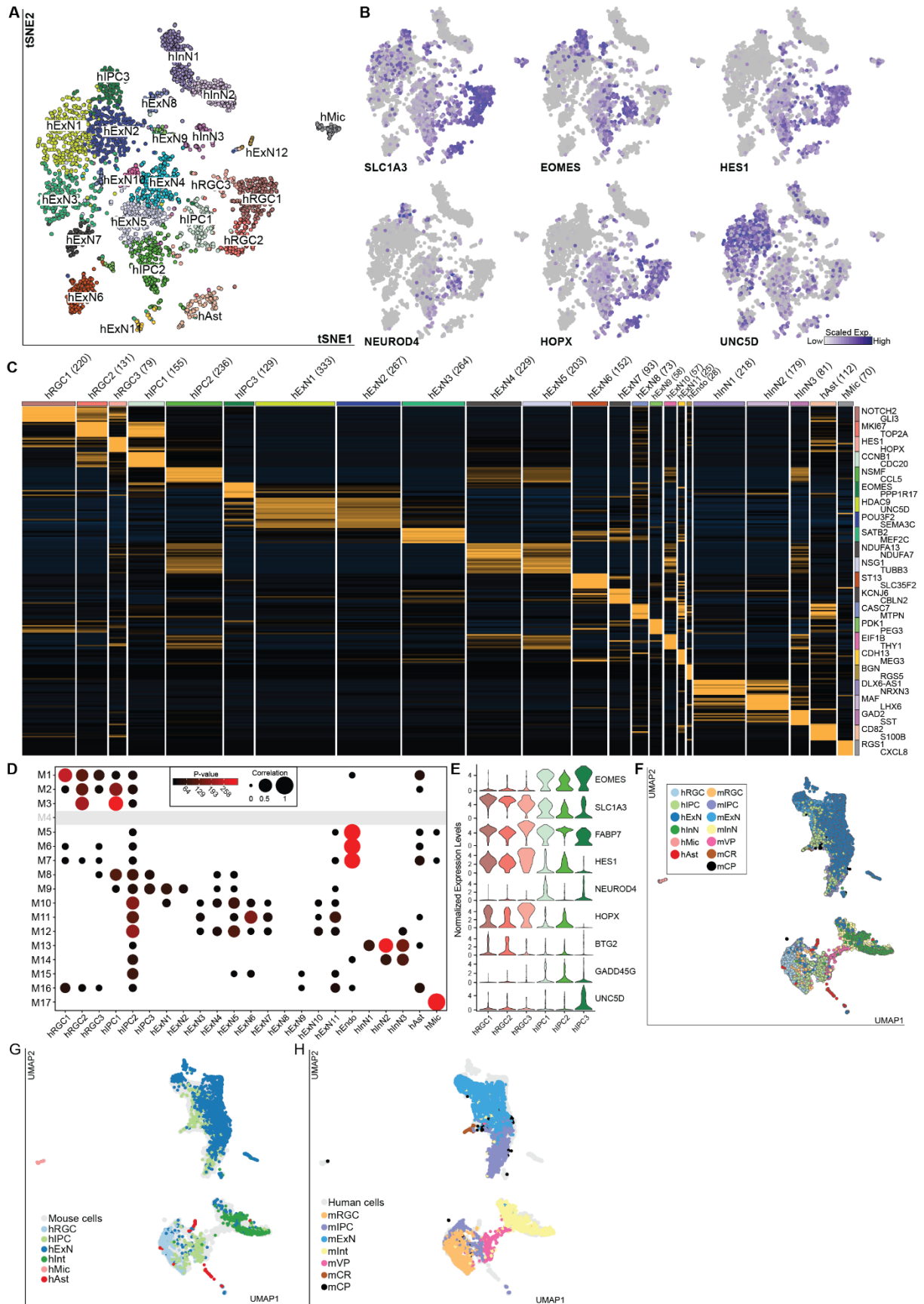
Figure S9 Single cell RNAseq analysis on human single cells at 12-20 postconceptional weeks.

(A)   T-SNE plot of human single cells. Colors represent cell types. RGC, radial glial cell; IPC, intermediate progenitor cell; ExN, excitatory neuron; InN, inhibitory neuron; Ast, astrocyte; Mic, microglia.

(B)   Feature plots of canonical marker gene expression. Colors represent scaled gene expression.

(C)   Heatmap of differentially expressed (DEX) genes. Colors on top of the heatmap represent cell types as in A. Two of the DEX genes from each cell type are listed on the right. The number of cells in each cell type is labeled in parentheses.

(D)   Gene expression modules from mouse weighted gene co-expression network analysis in human cell types. Size of the dots indicates level of correlation between a network and a cell type, whereas color represents level of significance. Module M4 which is mainly comprised by ribosomal proteins is excluded due to high technical variability among human datasets.

(E)   Violin plot of curated RGC, IPC and neuronal marker genes. Color represent cell types as in A.

(F)   UMAP plot of integrated mouse and human single cell RNAseq datasets. Colored dots present human and mouse cell types.

(G-H) UMAP plot as in F with cells from human (G) or mouse (H) species plotted separately. Colors represent same cell types as in F.