

Supplementary Materials for

Early emergence of cortical interneuron diversity in the mouse embryo

Da Mi, Zhen Li, Lynette Lim, Mingfeng Li, Monika Moissidis, Yifei Yang, Tianliuyun Gao, Tim Xiaoming Hu, Thomas Pratt, David J. Price, Nenad Sestan,[†] & Oscar Marín[†]

[†]Correspondence to: oscar.marin@kcl.ac.uk (OM); nenad.sestan@yale.edu (NS)

This PDF file includes:

Materials and Methods
Figs. S1 to S24
Tables S1 to S4

Materials and Methods

Mice

All mouse colonies were maintained in accordance with protocols approved by UK Home Office project licenses and the Institutional Animal Care and Use Committee (IACUC) at Yale University. Experimental protocols were approved by King's College London and Yale University, School of Medicine welfare committees. Both male and female CD1 mice obtained from Charles River Laboratories were used in single-cell RNA sequencing experiments. *Nkx2-1-Cre* mice were generously provided by S. Anderson (University of Pennsylvania) (30). *Nkx2-1-Cre;Maf^{fl/fl};RCE* mice were generated by breeding *Nkx2-1-Cre;RCE* mice with *Maf^{fl/fl}* mice, which were generously provided by C. Birchmeier (Max Delbrück Center for Molecular Medicine) (31). Mice were housed in groups and kept on reverse light/dark cycle (12/12 h) regardless of genotype. Time-mated pregnant female mice were housed individually. The day on which the vaginal plug was found was considered as embryonic day (E) 0.5. The date of birth was considered as postnatal day (P) 0.

Dissection and single cell dissociation

Mouse embryos were isolated at E12.5 and E14.5 from timed pregnant CD1 female mice. Embryos were stored on ice in 1x PBS. Brains were dissected out, embedded in 1% Ultrapure Low Melting Point Agarose and cut in 100 μ m thick sections using a vibratome (Leica VT1200S). The dorsal MGE (dMGE), ventral MGE (vMGE) and CGE regions were dissected from these sections. Dissected tissue was then immediately transferred to warm (37°C) papain/protease/dispase (PPD) solution in a 15ml conical tube and incubated for 30 min at 37°C. Tissue was gently stirred by pipetting every 5 min during the incubation. Once tissue pieces were no longer visible, the suspension was centrifuged at 300g (0.3 RCF) for 3 min. After centrifugation, cells were re-suspended in 1x PBS and filtered through a 40 μ m cell strainer. Cell concentration was measured using a hemocytometer and adjusted accordingly.

Single-cell capture and cDNA synthesis

Cells were captured with small-sized (5-10 μ m) RNA-seq IFCs using the Fluidigm C1 system according to the manufacturer's instructions. Trypan Blue staining was used immediately before cell captures to assess viability. Cell viability was nearly 100% in all experiments. A concentration of 1,000–3,000 cells/ μ l was used for cell loading. After capture, C1 chips were examined visually and the number of cells at each capture site was recorded manually. Cells captured by C1 were subsequently processed through lysis, reverse transcription and PCR amplification to generate single-cell cDNA libraries using the Smarter Ultra Low Input RNA kit for Fluidigm (Clontech). cDNAs were harvested the next morning in about 3 μ l C1 harvest reagent and were transferred to a 96-cell PCR plate in 10 μ l C1 DNA dilution reagent.

Single-cell RNA-seq library preparation

cDNA concentration was quantified with Quant-iT™ PicoGreen® (Invitrogen™, P7589). Samples were diluted to 200 ng/ μ l with C1 harvest reagent. The resulting cDNA samples were indexed using Nextera XT Library Prep Kit (Illumina). We used a high sensitivity dsDNA assay in an Agilent Bioanalyzer to assess the quality of all sequencing libraries.

Sequencing, reads alignment and gene expression quantification

Libraries of single cells pooled from each C1 IFC were denatured using the Illumina protocol. Denatured libraries were diluted to 6 pM, followed by cluster generation on a single-end HiSeq flow cell using an Illumina cBOT, according to the manufacturer's instructions. The

HiSeq flow cell was run for 100 cycles using a single-read recipe according to the manufacturer's instructions. We used Illumina CASAVA to purify the low-quality and non-identified reads and FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to report the fundamental quality parameters (raw reads number, sequencing base quality score distribution and GC content). To avoid adapter contamination and higher error rates in reads boundary, we trimmed 12 nucleotides in 5'-end and 13 nucleotides in 3'-end, leaving 75 nucleotide long reads for sequence alignment. We employed STAR (v2.4.0) to uniquely align the filtered reads to mouse reference genome (mm9) with default parameters (32). We used RSEQTools and SAMtools to calculate gene expression (33, 34). We used HTSeq (v0.6.1) to calculate gene read count values for each annotation entry.

Quality control

In addition to FastQC, we implemented a series of quality control measures. First, we counted uniquely mapped reads per cell and used only cells with at least 50,000 unique reads mapped to coding sequences. Next, we checked exonic read distribution, read distribution across different chromosomes, GC content distribution and gene expression distribution. Any cell that was 3 standard deviations away from the mean for any of the above-mentioned metrics were removed. In total, 666 cells were removed and 2003 high quality single cells were passed on to downstream analysis. We also filtered gene expression profile of each cell. Any gene expressed by less than 10 cells at less than 5 counts per million (CPM) was removed. Pseudogenes, miRNA, rRNA, mitochondrial associated and ribosome related genes from further analysis. 13,907 genes were kept for downstream analysis. We used R package *Seurat* (35) to manage our dataset. Briefly, raw read counts were used to create Seurat object followed by log normalization using *NormalizeData* function with *scale.factor* parameter set to 1,000,000. We also reanalyzed an adult single cell dataset using the same procedure (8).

Removal of cell cycle effect

In order to minimize the effect of cell cycle (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 (36) and calculated G1/S and G2/M phase scores for each cell using function *CellCycleScoring* from R package *Seurat*. Then, we calculated the difference between G1/S phase score and G2/M phase score. This result was used to perform regression on all cells in our dataset with *Seurat*. Using this approach, cell cycle differences among dividing cells were regressed out while signals segregating cycling and non-cycling cells were maintained.

Defining highly variable genes

To define highly variable genes (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 0.5 and 8, and VMR between 0.5 and 5 were considered as highly variable genes.

Dimension reduction and clustering

We used principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) (37) as our main dimension reduction approaches. PCA was performed with *RunPCA* function (*Seurat*) using HVGs to analyze either all cells (as in Fig. 1) or neuronal cells (as in figs. S15 and S17). For the analysis of progenitor cells (combined or individual E12.5 and E14.5 datasets), we used a combined list of genes that were selected by random forest analysis (see random forest feature selection and classification section for detail) as features for ventricular zone (VZ) and subventricular zone (SVZ) identities, as well as differentially

expressed genes between dMGE, vMGE and CGE regions (Tables S2 and S3). Following PCA, we conducted JACKSTRAW analysis with 100 iterations to identify statistically significant (p -value < 0.01) principal components (PCs) that were driving systematic variation. We used t-SNE to present data in two-dimensional 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*. Clustering was done with Luvain-Jaccard algorithm using t-SNE coordinates by *FindClusters* function from *Seurat* with *resolution* parameter set to 1.

Differential expression analysis

Differential expression (DE) analyses were conducted using *Seurat* function *FindAllMarkers*. In brief, we took one group of cells and compared it with the rest of the cells, using a binomial model. For any given comparison, we only considered genes that were expressed by at least 25% of cells in either population. Genes that exhibit P values under 0.01 were considered statistically significant.

Random forest feature selection and classification

We used random forest (RF) feature selection and classification technique for a number of analyses in the current study. All the functions described here were from R package *randomForest* (38). The process can be broken down to the following steps:

1) We started with a list of genes as an initial set of identifiers for the particular biological process of interest to define a tentative identity of each sample in question by hierarchical clustering using R function *hclust* (with average linkage method) if identity was already defined by other metrics, then this step is omitted;

2) We assessed the importance of each HVG in defining identities using the *importance* function and ranked them in descending order (39). We then performed 10-fold cross validation of feature selection on the HVG using the *rfcv* function, with step size set to 0.75. The number of genes (n) that produces the least error is recorded. The top n genes with the highest importance were regarded as features and were used in classification analysis;

3) Lastly, we used the features found in previous step as input to build a RF model. The RF model was then used to generate a likelihood matrix of each identity for every cell in the dataset. We determined the threshold statistically by examine the distribution of maximum likelihood from each cell; a cell was assigned an identity if the likelihood was above population average. Cells that did not meet this criterion or had ties between more than one identities remain unassigned. After the initial assignment, we conducted permutation test, where the above assignment procedure was repeated 10,000 times with 90% of the cells chosen at random. Cells with false discovery rate (FDR) $< 5\%$ was kept for downstream analysis.

Defining progenitors and neurons

To define progenitor and neuronal populations in our dataset, we first curated a list of established progenitor and neuronal markers from literature (Table S1) (40). We then performed hierarchical clustering on all single cell samples and split them into tentative “progenitor” or “neuron” groups. Following hierarchical clustering, we conducted RF feature selection using the curated list of progenitor and neuronal marker expression matrix as input, and the tentative “progenitor” or “neuron” group identity of each sample as responses. Through this process we were able to identify a refined list of genes that were indicative of progenitor and neuronal identity, specific to our dataset. Finally, we conducted RF feature classification on all samples

using the expression of selected features as input and the identity with the majority votes (either “progenitor” or “neuron”) was assigned to each cell.

Defining ventricular zone and subventricular zone cells

We used a similar approach, as in defining progenitors and neurons, to define VZ and SVZ cell identities using a different list of markers (41) for hierarchical clustering and RF feature selection (Table S2).

MetaNeighbor analysis

MetaNeighbor analysis was performed using the R function *MetaNeighbor* with default settings (14). The AUROC (Area under the Receiver Operating Characteristic) scores produced by MetaNeighbor analysis indicate the degree of correlation between cell groups. A mean AUROC score of 0.7 or above typically suggests a good correlation, while mean AUROC score below 0.5 indicates no correlation. The results from the MetaNeighbor analysis were plotted as a heatmap using the R function *heatmap.3*.

Gene Ontology (GO) enrichment analysis

We performed GO enrichment analysis using the GO-PCA package in Python (42). A set of 5,031 GO terms for mouse was obtained using the Gene Ontology structure file (go-basic.obo downloaded from <http://geneontology.org>), with version “releases/2017-07-16” and UniProt Gene Ontology Annotation (GOA) gene association file for mouse (goa_mouse.gaf.gz downloaded from <ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/MOUSE>), version 156. We used DE genes as the inputs for GO-PCA. Correlation threshold was set as 0.3 and all other parameters were set as the default settings. Clustering of GO terms was done using *pheatmap* package with default settings.

Classification of embryonic neurons using information from adult cortical interneurons

To classify embryonic neurons, we utilized a publically available dataset of adult GABAergic interneurons (766 cells in total) (8). We first found all the shared HVGs in both embryonic and adult datasets. Then we performed RF feature selection within the shared HVGs that best represented each cell type for all 23 interneuron cell types defined by Tasic and colleagues (8), referred hereon as the adult cell type features. We then took two independent approaches to define embryonic neurons utilizing the adult cell type features:

1) For the first approach, we reanalyzed adult cells using the RF feature selection and classification method described earlier with the adult cell type features. Through this process we were able to reassign the identity of adult cells. Notably, we were not able to reestablish the same cell types for all the cells proposed in the original study mainly because we used an intersected gene list between embryonic and adult developmental periods. We then used the newly assigned cells to construct a RF model which we used to predict cell types for the embryonic cells through our RF classification pipeline; we added one criteria after the first cycle where we removed any cell identity with less than 5 cells assigned as these cell types tended to show poor accuracy. We were able to assign 244 embryonic neurons to 6 adult cell types.

2) For the second approach, we first conducted canonical correlation analysis (CCA) on embryonic and adult single cell datasets using the adult cell type features. Then we performed t-SNE analysis to reduce the dimension of embryonic and adult cell data onto the same two-dimensional space. Subsequently, we used the two t-SNE coordinates for adult cells to conduct *k*-nearest neighbors analysis (knn) of the 23 cell types and reassign cell identities for adult cells ($k = 30$) using the *knn.cv* function from R package *FNN*. Briefly, we calculated the average

distance (\bar{d}) and standard deviation (σ) of each cell with each of its 30 neighboring cells and removed any neighbor that was more than $\bar{d} + \sigma$ away. Among the remaining neighbors, we counted the identities represented by the neighbors. A cell was assigned the identity represented by the majority, and at least 10, of its neighbors; in case of ties, the cell remains unassigned. Through this process, we were able to confidently reestablish 422 adult cells to 11 cell types for the adult dataset. The cells unable to be assigned were removed from downstream analysis. Subsequently, we used the same knn approach on embryonic single cells, using adult cells as neighbors ($k = 5$) and assigned adult identities to embryonic cells. Finally, we repeated the process again with only the assigned embryonic cells as neighbors ($k = 5$) to find the core cells of each assigned cell type for most accurate representation. Through this process we were able to assign 247 embryonic cells to 11 cell types with high confidence.

Characterization of progenitor lineage identity

To infer lineage identities of progenitor clusters with high accuracy, we sought to firstly combine the results from neuronal lineage assignment using either RF or CCA method and found embryonic neurons that were assigned to the same identities by both methods, which are hereon referred to as consensus neurons. The consensus neurons we found belong to one PV lineage subtype (PV1) and two SST lineage subtypes (SST1 and SST2). We then conducted MetaNeighbor analysis on consensus neurons from PV1, SST1 and SST2 subtypes and progenitor clusters from E12.5 and E14.5 samples respectively. We used an AUROC score > 0.9 as a cutoff to infer the progenitor clusters that were most likely associated with any of the three lineage subtypes.

Data visualization

All violin plots and box plots were generated using *ggplot2*. TSNE plots were generated using *TSNEPlot* function from R package *Seurat*. Except otherwise noted, all heatmaps were generated with R function *heatmap.3* (43).

Cloning and virus making

The expression plasmids *rv::dio-Gfp* and *rv::dio-Maf-P2A-Gfp* were cloned by PCR into a conditionally expressing retroviral backbone (Addgene 87662) using *AscI* and *PacI* as cloning sites. Retroviruses were generated as described previously (44). In brief, HEK293FT cells were plated at 10^6 cells per well in six 15 cm plates. Upon 70-80% confluence, cells were transfected with retroviral constructs (*pRV-CAG-dio-mtdT*, *pCMVsv-g*, and *pCMV-GAG-pol*) using TransIt reagent (Mirus, Madison WI). Forty-eight hours post transfection, 120 ml of supernatant were concentrated by two rounds of ultracentrifugation and re-suspended in 80 μ l of PBS and stored in 5 μ l aliquots at -80°C . Typical virus titers range from $1-5 \times 10^8$ IU/ml.

In utero injections

Viral infection was performed on E12.5 or E14.5 embryos. In brief, pregnant mice were anesthetized with isoflurane (induction, 4%; surgery, 2.5%) along with intraperitoneal 0.1mg/kg of Buprenorphine as analgesic and 50mg/kg of Ritrodine as a muscle relaxant. The uterine horns were exposed by laparotomy and 400 nl of purified viruses were injected into the lateral ventricles of each embryo using a glass capillary needle (3.5'' Drummond 3-000-203Glx) coupled to a nanoinjector (World Precision Instruments). Littermates were injected with both control *Gfp* and *Maf* viruses. The uterine horns were then placed back in the abdominal cavity, and the abdominal wall was sutured (Ethicon coated Vicryl 4-0; V-4). The animals were placed in a 32°C recovering chamber for 1 hour post-surgery.

Histology

When *in utero* injected littermates, *Nkx2-1-Cre; Maf^{+/+};RCE* and *Nkx2-1-Cre; Maf^{fl/fl};RCE* littermates reached P21, they were perfused transcardially with 4% paraformaldehyde (PFA), and their brain removed, post-fixed for 3 hours at 4°C, washed 3 times in PBS over 24 hours, and cryopreserved overnight in sequential 15% and 30% sucrose solutions. Brains were sectioned coronally at 60 (mutant analysis) or 100 µm (viral infections) using a freezing microtome. Immunohistochemistry was performed on free-floating sections, blocked for 1 h in 10% horse serum, 2% Bovine Serum Albumin (BSA) and 0.25% Triton in PBS (PBS-T). The following primary antibodies in blocking solution at 4°C overnight: chicken anti-GFP (Aves Labs, 1:1000), rat anti-SST (EMD Millipore, 1:200) mouse anti-PV (Swant, 1:1000), rabbit anti-MAF (Bethyl Laboratories, 1:500). Sections were then washed three times for 10 minutes with 0.25% PBS-T, before being incubated in blocking solution with the following secondary antibodies: donkey anti-chicken 488 (1:200), goat anti-rat 555 (1:200), donkey anti-rabbit 555 (1:200), goat anti-mouse 647 (1:200), all from Life Technologies Alexa Fluor®. Sections were finally counterstained with DAPI (5 µM), before being mounted in series with Mowiol-Dabco mounting medium (SigmaAldrich).

For *in situ* hybridization, embryonic brains were dissected, post-fixed in 4% PFA for 3 hours at 4°C, washed in PBS overnight, and cryopreserved overnight in sequential 15% and 30% sucrose solutions. Brains were then sectioned coronally at 20 µm using a cryostat (Leica) and mounted on superfrost slides. Double fluorescent *in situ* hybridization for *Maf* was performed using RNAscope probes from Advanced Cell Diagnostics and the RNAscope Multiplex Fluorescent Reagent Kit V2 (Advanced Cell Diagnostics) following the manufacturer's recommendations. *In situ* hybridization images were obtained with a 63x objective (1.4 NA) with LSM800 Airyscan mode (Zeiss).

Neuronal quantification

Nkx2-1-Cre; Maf^{+/+};RCE and *Nkx2-1-Cre; Maf^{fl/fl};RCE* coronal sections were imaged using the 10x objective of an inverted confocal microscope (Leica TCS SP8). Images of cell bodies expressing GFP, PV and SST were analyzed using software written in Matlab (Mathworks). Cell bodies were segmented using disk morphological shape function, size, and intensity thresholding. Background and high-density noise were removed via filtering. For each brain, 10-12 images spanning rostral to caudal regions of the somatosensory cortex were taken. Data were statistically analyzed using One-way ANOVA with Tukey correction.

Coronal sections of P21 brains infected with *Gfp* or *Gfp-Maf* viruses were imaged using the 20x objective of an inverted confocal microscope (Leica TCS SP8). Images of cell bodies expressing GFP and PV were analyzed in the same way as described above, using custom written software in MatLab (Mathworks). The proportion of GFP+ cells co-expressing SST was quantified manually, using ImageJ software. Data were statistically analyzed using chi-squared test with post-hoc binomial test pairwise comparison with Bonferroni correction.

In all experiments, each animal was considered as a biological replicate.

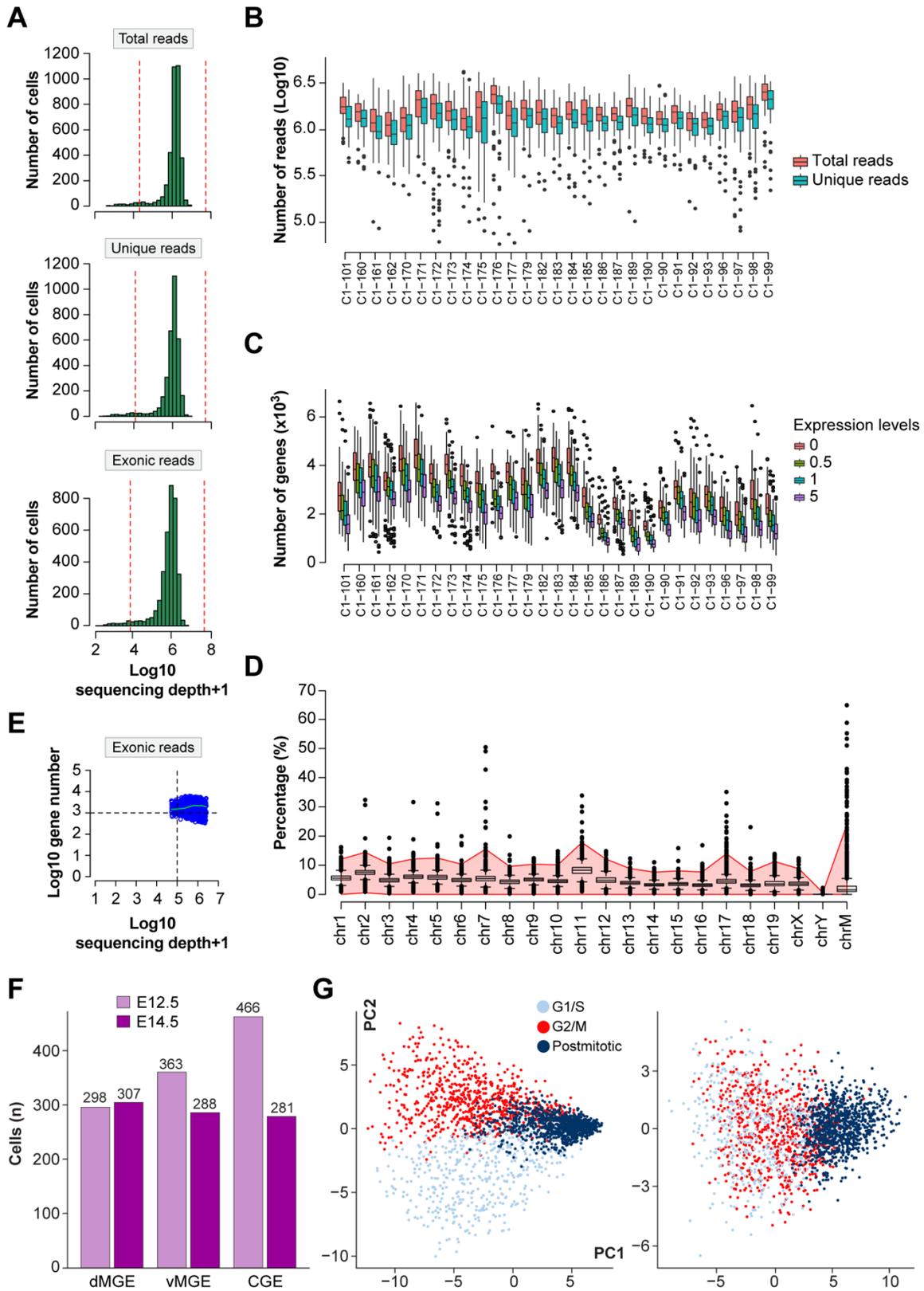
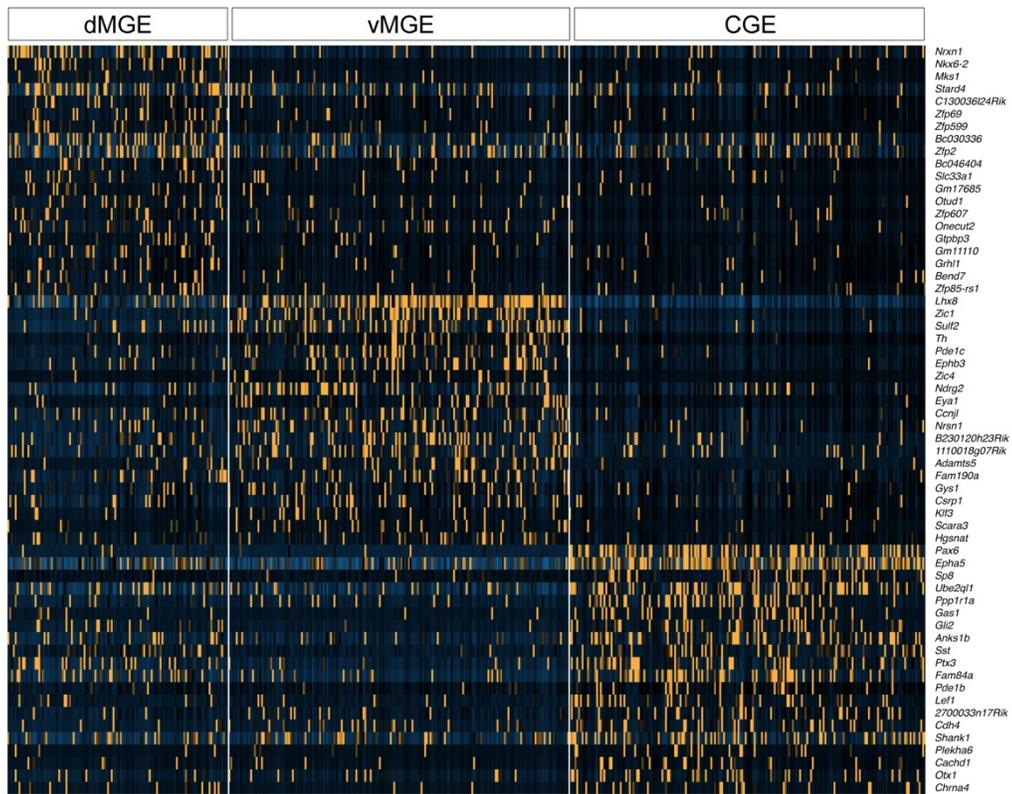
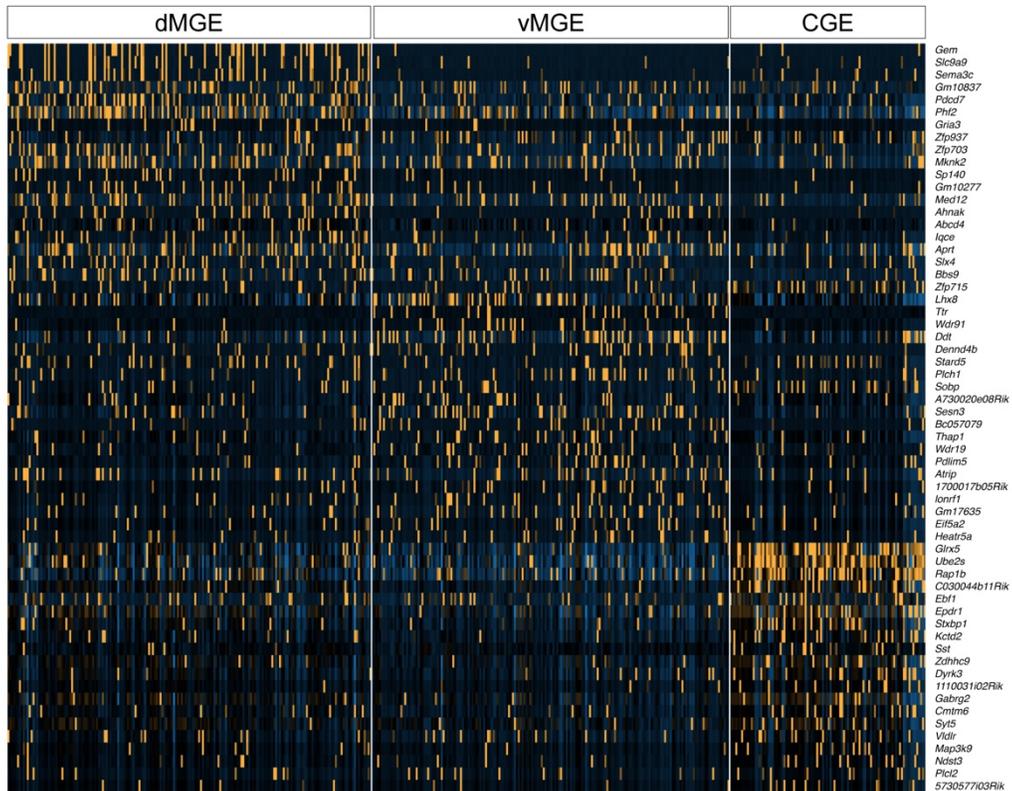


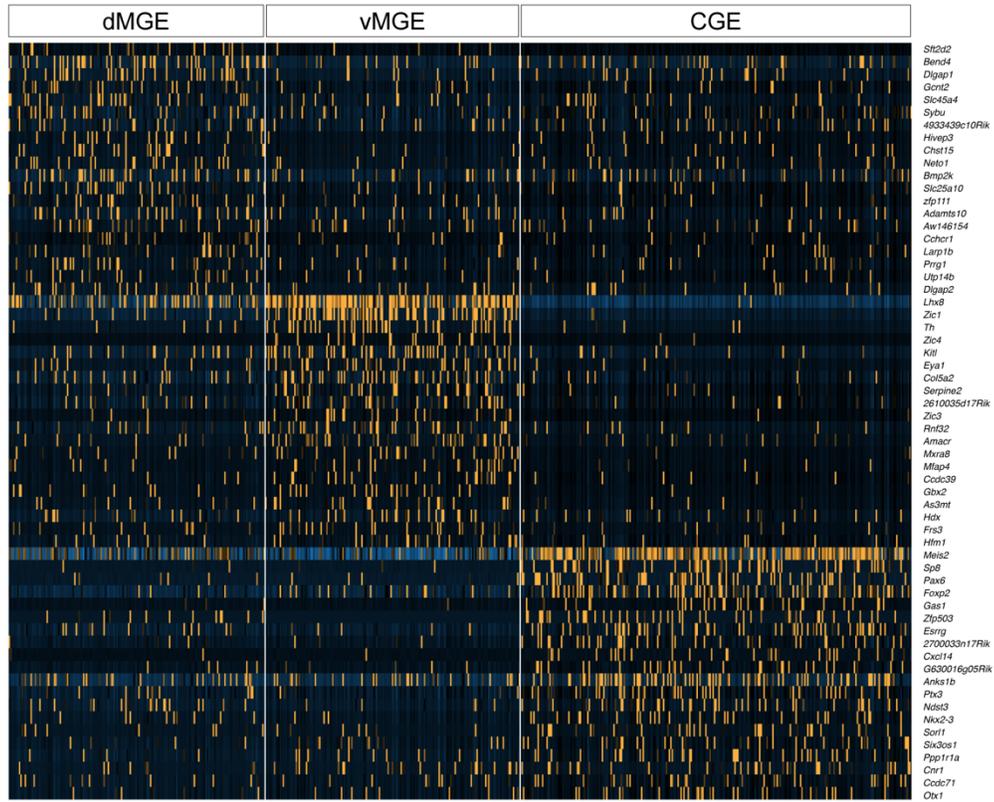
Fig. S1. Quality control in single-cell RNA-seq experiments. (A) Bar plots showing distribution of total reads, uniquely mapped reads (unique reads) and reads mapped to exons (exonic reads). Horizontal axes are log 10 transformed read counts, whereas vertical axes are number of cells. Dashed lines in red indicate three standard deviations from the mean (SDs). Cells beyond three standard deviations for any of the metrics were removed. (B) Box plot showing number of total reads and unique reads from each of the C1 single cell capture and RNA-seq runs (C1 runs). Vertical axis shows log 10 transformed read counts. (C) Box plot showing number of genes (transcriptome coverage) expressed at above 0, 0.5, 1 or 5 counts per million (CPM) in each of the C1 runs. Cells from runs that show low transcriptome coverage are removed from downstream analysis. (D) Box plot showing chromosome coverage from single cells that pass quality control (QC). Vertical axis indicates the percentage of total reads from one cell mapped to a given chromosome. All chromosomes are evenly covered in the current dataset. (E) Scatterplot showing gene coverage in each single cell passed QC. Horizontal axis shows log 10 transformed read counts, whereas vertical axis shows log 10 transformed transcriptome coverage. Among single cells passed QC, the transcriptome coverage is uniform. (F) Number of analyzed cells according to the region of origin and stage. (G) Principal component analysis (PCA) scatterplots before (left) and after (right) cell cycle regression. PCA was performed using a published list of cell cycle related genes (see Methods).

A**B**

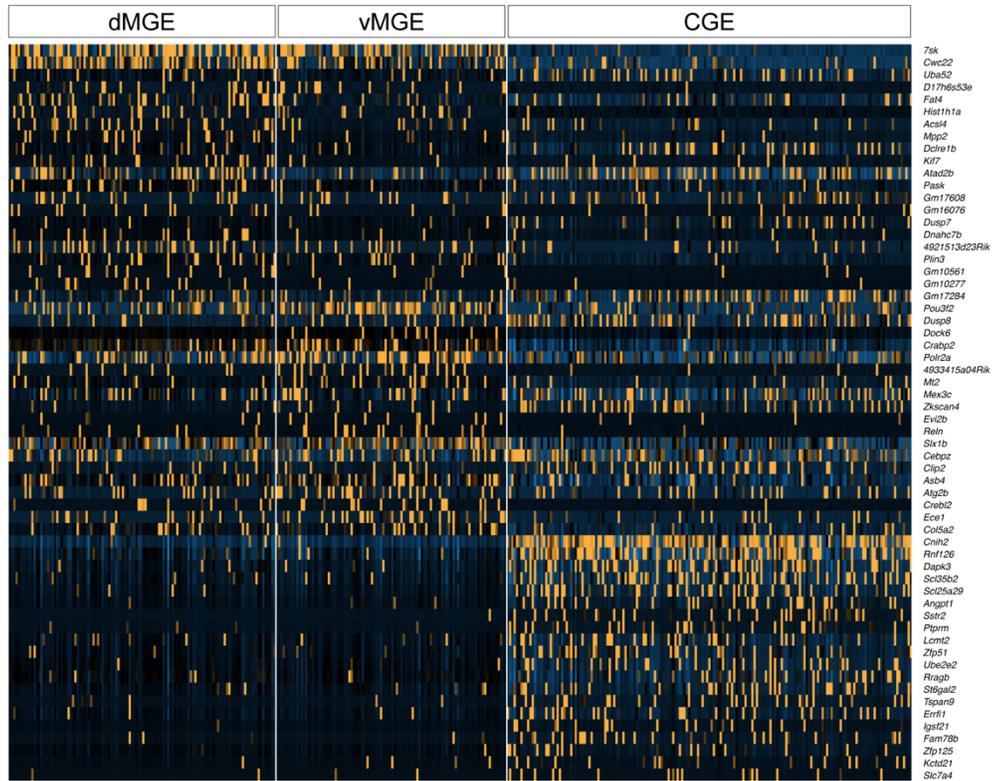
Z-score
-1 0 1

Fig. S2. Differential gene expression analysis on progenitor cells from dMGE, vMGE and CGE. (A) Heatmap showing the expression pattern of top 20 differentially expressed (DE) genes between E12.5 progenitors from dMGE, vMGE and CGE. Some of the DE genes are well-known markers for distinct progenitor domains in the mouse embryonic subpallium. For example, *Nkx6-2* is highly expressed in the ventricular zone of dMGE; *Lhx8* is enriched in the subventricular zone of vMGE, and *Pax6* is expressed in the progenitor domain of CGE. (B) Heatmap showing the expression pattern of top 20 differentially expressed (DE) genes between E14.5 progenitors from dMGE, vMGE and CGE.

A



B



Z-score
-1 0 1

Fig. S3. Differential gene expression analysis on neurons from dMGE, vMGE and CGE.

(A) Heatmap showing the expression pattern of top 20 differentially expressed (DE) genes between E12.5 neurons from dMGE, vMGE and CGE. A number of well-known region-specific markers for subventricular zone progenitor cells and newborn neurons, such as *Lhx8*, *Meis2* and *Pax6*, exhibit expression patterns that are consistent with previous studies on single genes. For example, *Neto1*, a lineage marker for SST+ interneurons, is enriched in dMGE neurons, consistent with the notion that dMGE is poised to produce SST+ interneurons, while *Sp8*, a lineage marker for VIP+ interneurons, is highly expressed in the CGE neurons, also in line with prior knowledge that CGE gives rise to VIP+ interneuron lineages. (B) Heatmap depicting the expression pattern of top 20 DE genes between E14.5 neurons from dMGE, vMGE and CGE.

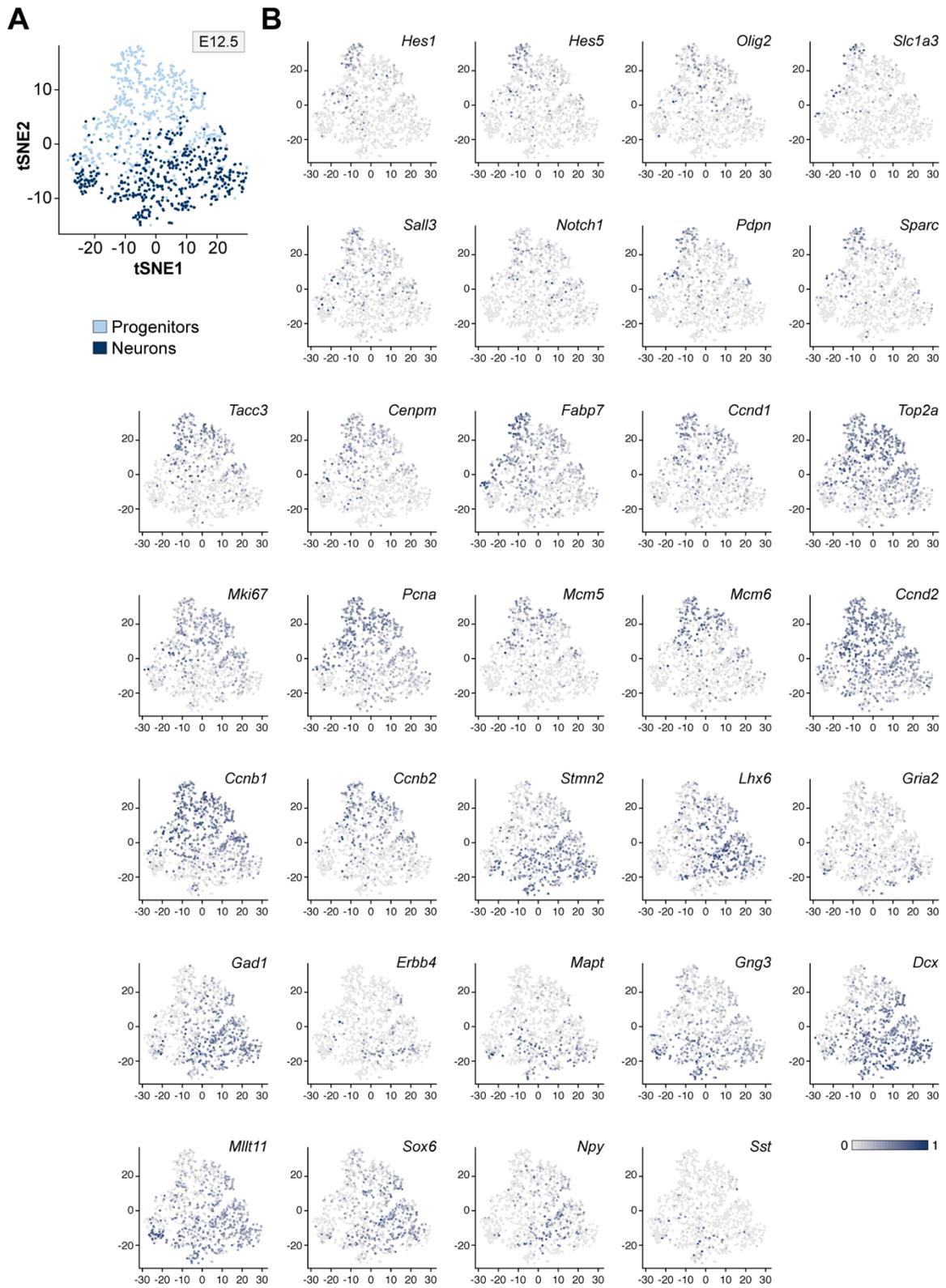


Fig. S4. Expression pattern of canonical progenitor and neuronal markers in single cells at E12.5. (A) Visualization of progenitor cells and neurons by t-SNE in the ganglionic eminences at E12.5. (B) The same t-SNE plots are used to illustrate gene expression in all cells at E12.5. Cells are colored according to gene expression levels (blue, high; grey, low). Among the marker genes, *Hes1*, *Hes5*, *Olig2*, *Slc1a3*, *Sall3*, *Notch1*, *Pdpn*, *Sparc*, *Tacc3*, *Fabp7* and *Top2a* are known to be expressed specifically by neural progenitor cells; *Mki67*, *Pcna*, *Cenpm*, *Mcm5*, *Mcm6*, *Ccnd1*, *Ccnd2*, *Ccnb1* and *Ccnb2* are canonical proliferating and cell cycle markers; and *Stmn2*, *Lhx6*, *Gria2*, *Gad1*, *ErbB4*, *Mapt*, *Gng3*, *Dcx*, *Mllt11*, *Sox6*, *Npy* and *Sst* are highly expressed in migrating postmitotic neurons.

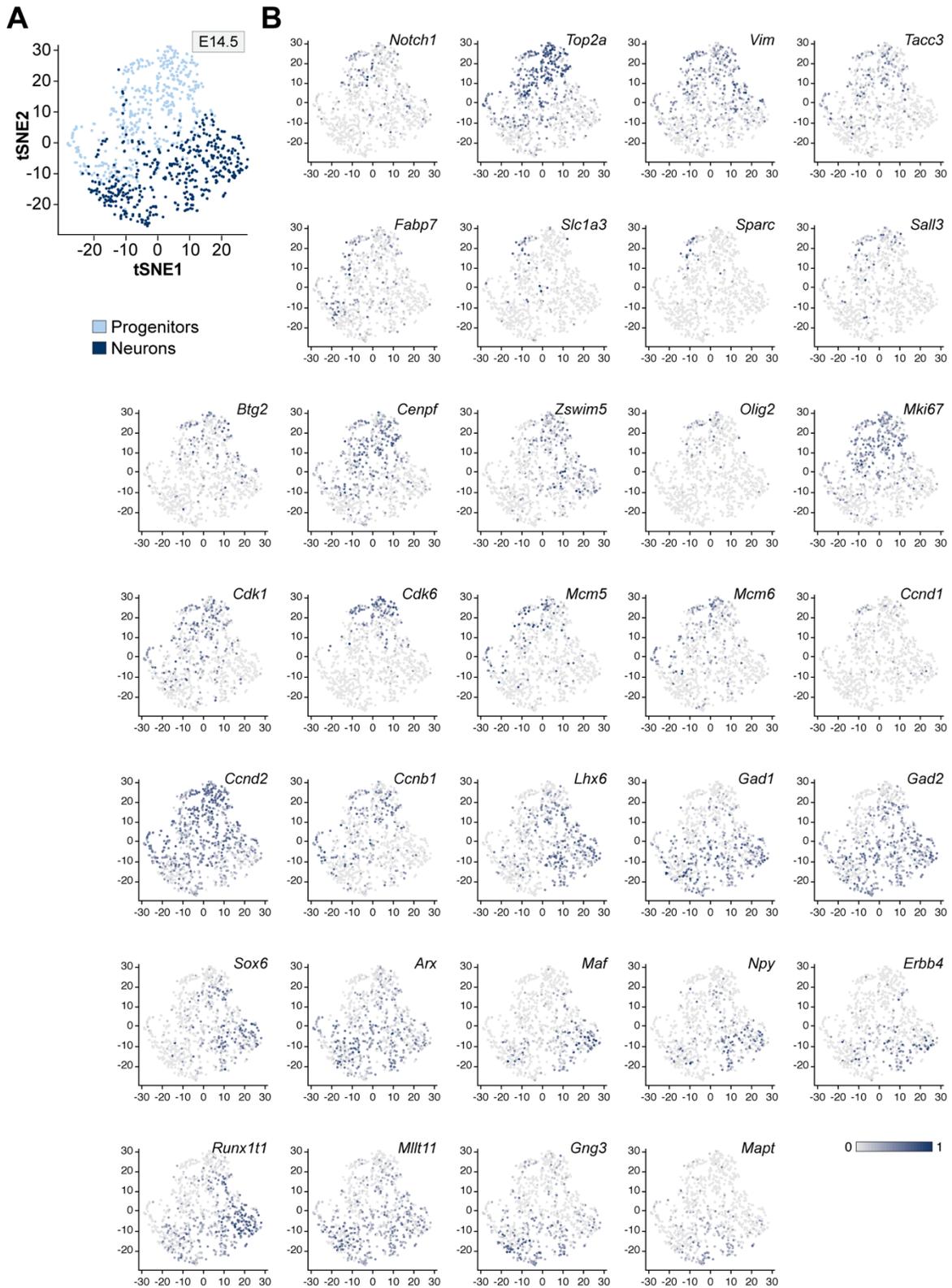


Fig. S5. Expression pattern of canonical progenitor and neuronal markers in single cells at E14.5. (A) Visualization of progenitor cells and neurons by t-SNE in the ganglionic eminences at E14.5. (B) The same t-SNE plots are used to illustrate gene expression in all cells at E12.5. Cells are colored according to gene expression levels (blue, high; grey, low). Among the marker genes, *Notch1*, *Top2a*, *Vim*, *Tacc3*, *Fabp7*, *Slc1a3*, *Sparc*, *Sall3*, *Btg2*, *Cenpf*, *Zswim5* and *Olig2* are canonical neural progenitor markers; *Mki67*, *Cdk1*, *Cdk6*, *Mcm5*, *Mcm6*, *Ccnd1*, *Ccnd2* and *Ccnb1* are proliferating and cell cycle markers; *Lhx6*, *Gad1*, *Gad2*, *Sox6*, *Arx*, *Maf*, *Npy*, *ErbB4*, *Runx1t1*, *Mllt11*, *Gng3* and *Mapt* are highly expressed in migrating postmitotic neurons.

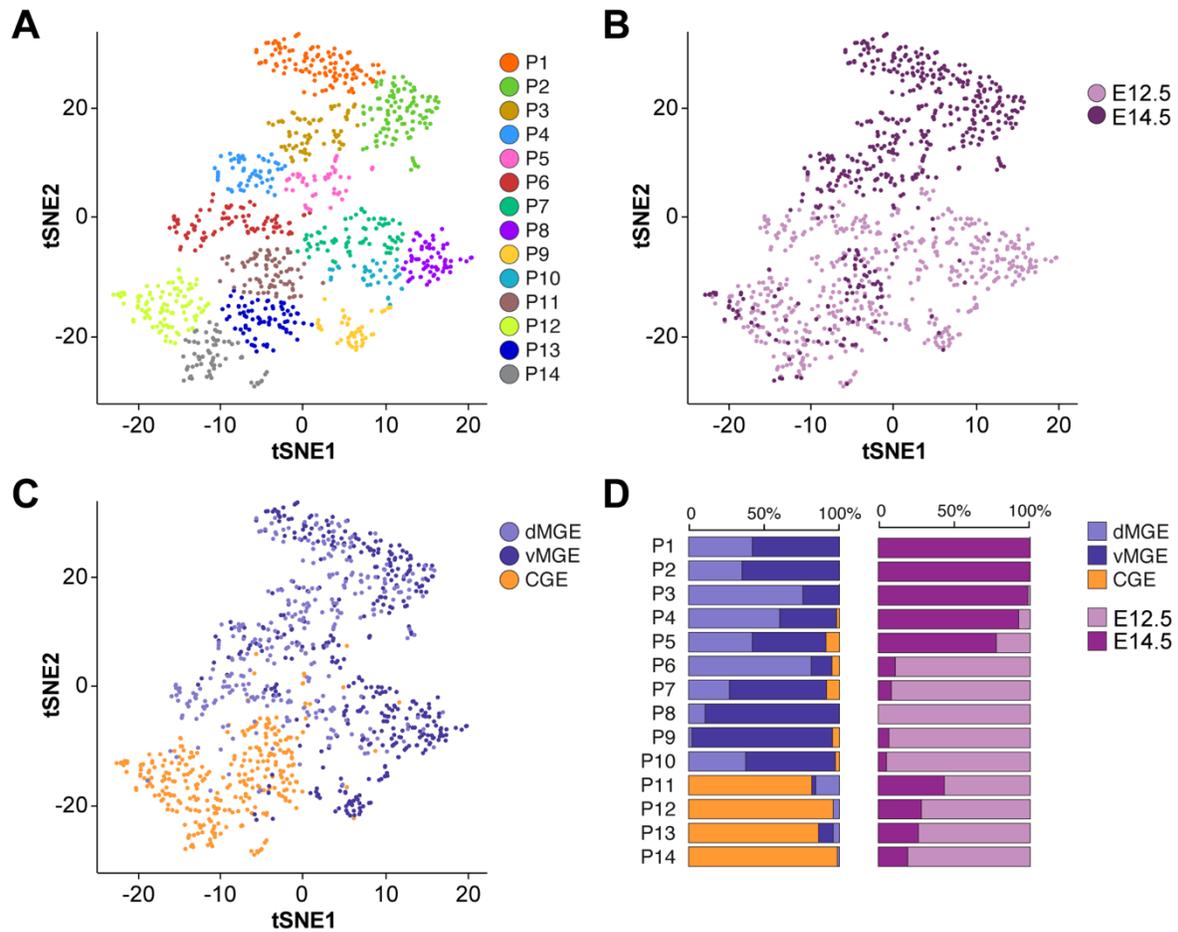


Fig. S6. Characterization of progenitor diversity at E12.5 and E14.5. (A) t-SNE plot depicting progenitor clusters by semi-supervised Luvain-Jaccard clustering with all progenitors. Progenitors from E12.5 and E14.5 datasets were jointly classified together into 14 clusters. Cells are colored by cluster identity. (B and C) The same t-SNE plot as in (A) highlights the stage (B) and (C) region of origin of single cells. Progenitors from E12.5 and E14.5 are largely separated in the t-SNE plot, except for CGE progenitors. This suggests that CGE progenitors are transcriptionally more homogeneous during embryonic development (E12.5 and E14.5) than MGE progenitors, which segregate according to the stage. (D) Histograms showing the percentage of cells from dMGE, vMGE or CGE (left) and from E12.5 or E14.5 samples (right) in each of the progenitor clusters.

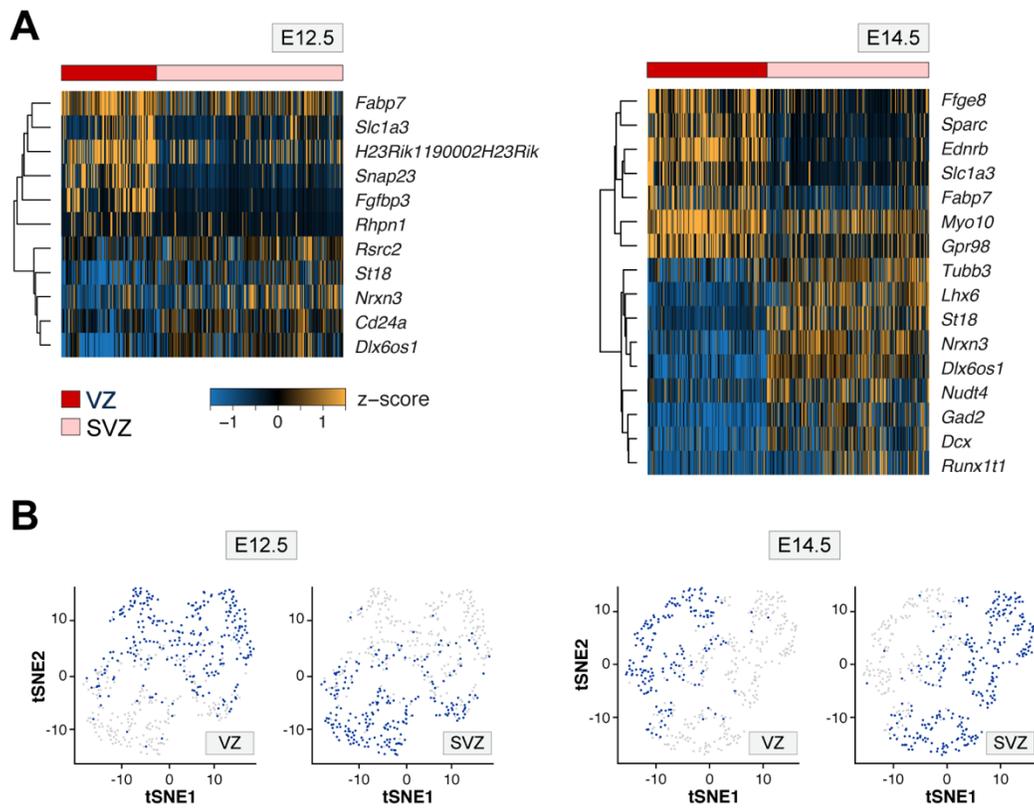


Fig. S7. Identification of VZ and SVZ progenitor cells. (A) Heatmaps illustrating the expression of genes selected by random forest analysis that best represent VZ and SVZ identities at E12.5 and E14.5. (B) VZ and SVZ identities at E12.5 and E14.5 are shown in the same t-SNE space as in Figure 2A.

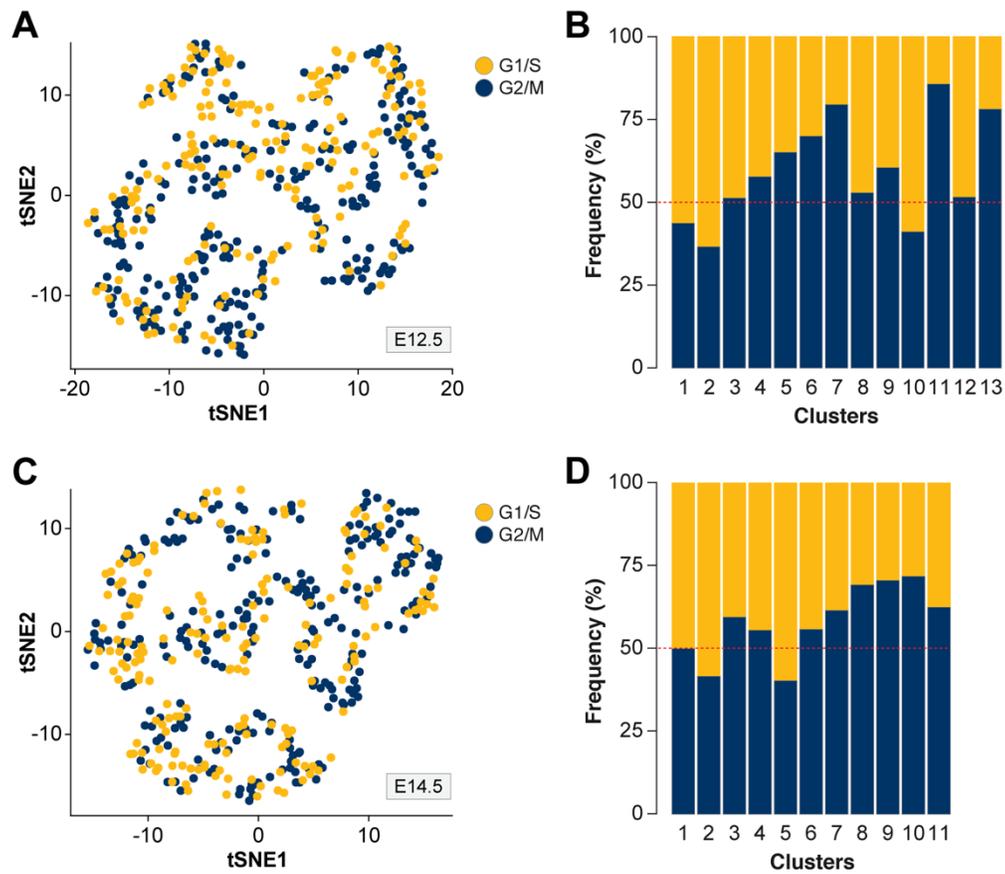


Fig. S8. Cell cycle status in progenitor cells. (A and C) t-SNE plots as in Fig. 2A illustrate cell cycle state in E12.5 (A) and E14.5 (C) progenitor cells. Cells are colored by cell cycle state based on a cell cycle score (see Methods). (B and D) Histograms show percentages of cells in either G1/S or G2/M phase in each of the E12.5 (B) and E14.5 (D) progenitor clusters.



Fig. S9. Expression patterns of differentially expressed genes in the subpallium. Sagittal sections through the E13.5 telencephalon showing mRNA expression for representative genes expressed in progenitor cells in the MGE and CGE. Images were obtained from the Allen Developing Mouse Brain Atlas from the Allen Institute for Brain Science.

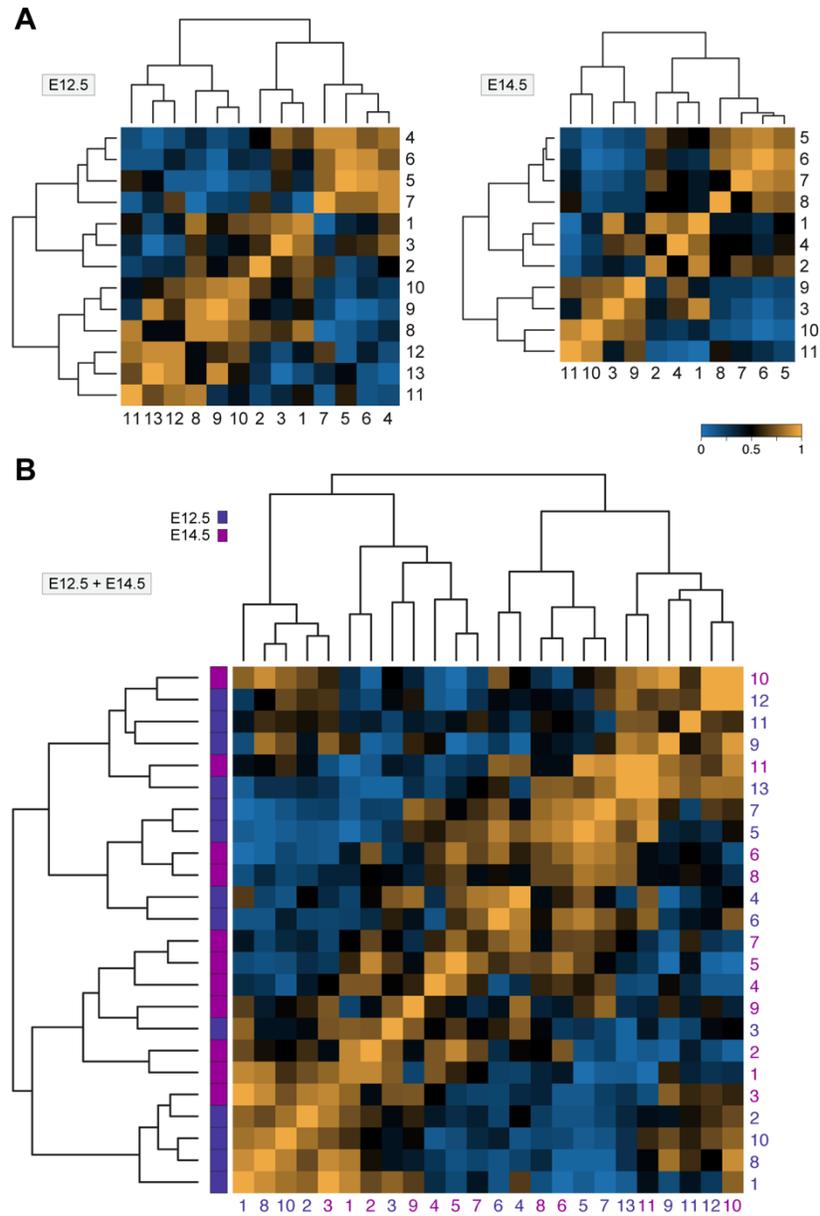


Fig. S10. MetaNeighbor analysis of E12.5 and E14.5 progenitor clusters. (A) Correlation analysis among E12.5 and E14.5 progenitor clusters. Heatmap of mean AUROC (Area under the Receiver Operating Characteristic) scores (produced by MetaNeighbor analysis, see Methods) indicating the degree of correlation between clusters. A mean AUROC score of 0.7 or above typically suggests a good correlation, while mean AUROC score below 0.5 indicates no correlation. Hierarchical clustering of mean AUROC score profiles recapitulate the biological properties of progenitor clusters. Notably, major branches represent the separation of VZ and SVZ identities. (B) Heatmap of mean AUROC scores showing correlations among E12.5 and E14.5 progenitor clusters. Hierarchical clustering of mean AUROC score profiles shows a separation of E12.5 and E14.5 clusters with only a few exceptions, suggesting that majority of the progenitor clusters are relatively distinct at E12.5 and E14.5.

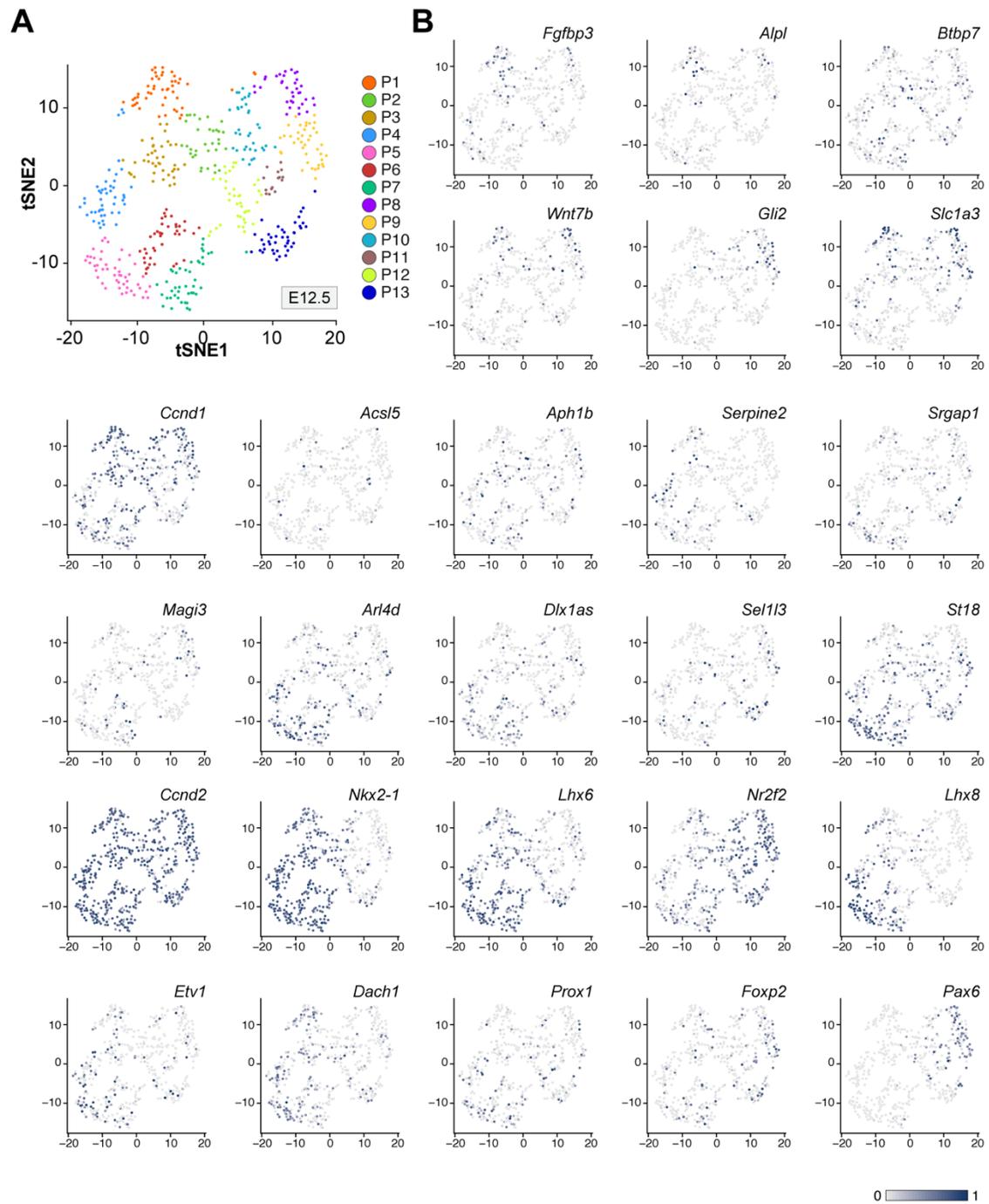


Fig. S11. Differential gene expression among E12.5 progenitor clusters. (A) Visualization of progenitor cells at E12.5 by t-SNE. **(B)** t-SNE plots show the expression of marker genes in progenitor cells among progenitor cell clusters at E12.5. Single cells are colored based on expression levels.

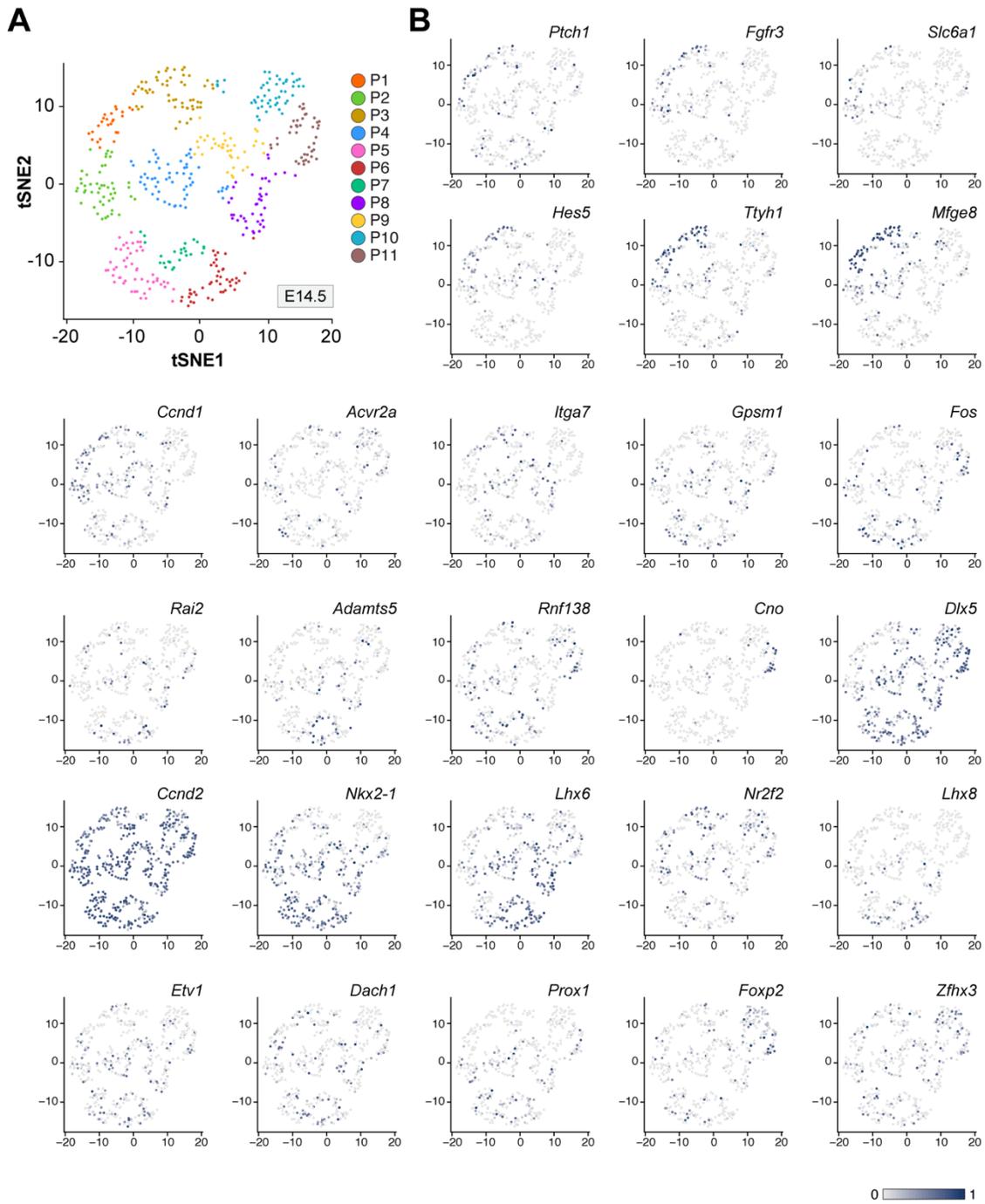


Fig. S12. Differential gene expression among E14.5 progenitor clusters. (A) Visualization of progenitor cells at E14.5 by t-SNE. **(B)** t-SNE plots show the expression of marker genes in progenitor cells among progenitor cell clusters at E14.5. Single cells are colored based on expression levels.

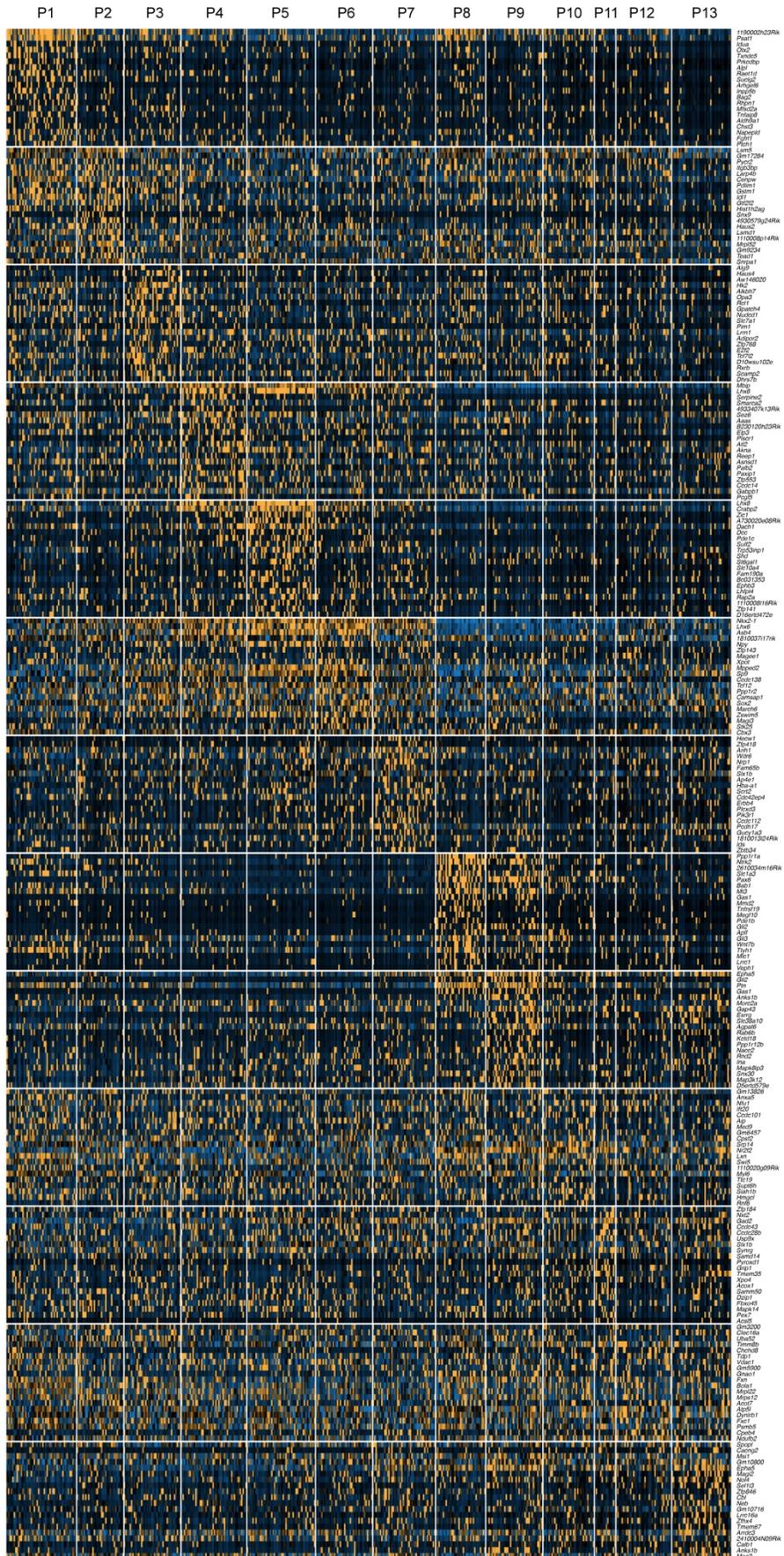


Fig. S13. Heatmap of top 20 differentially expressed genes among E12.5 progenitor clusters.

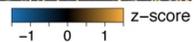
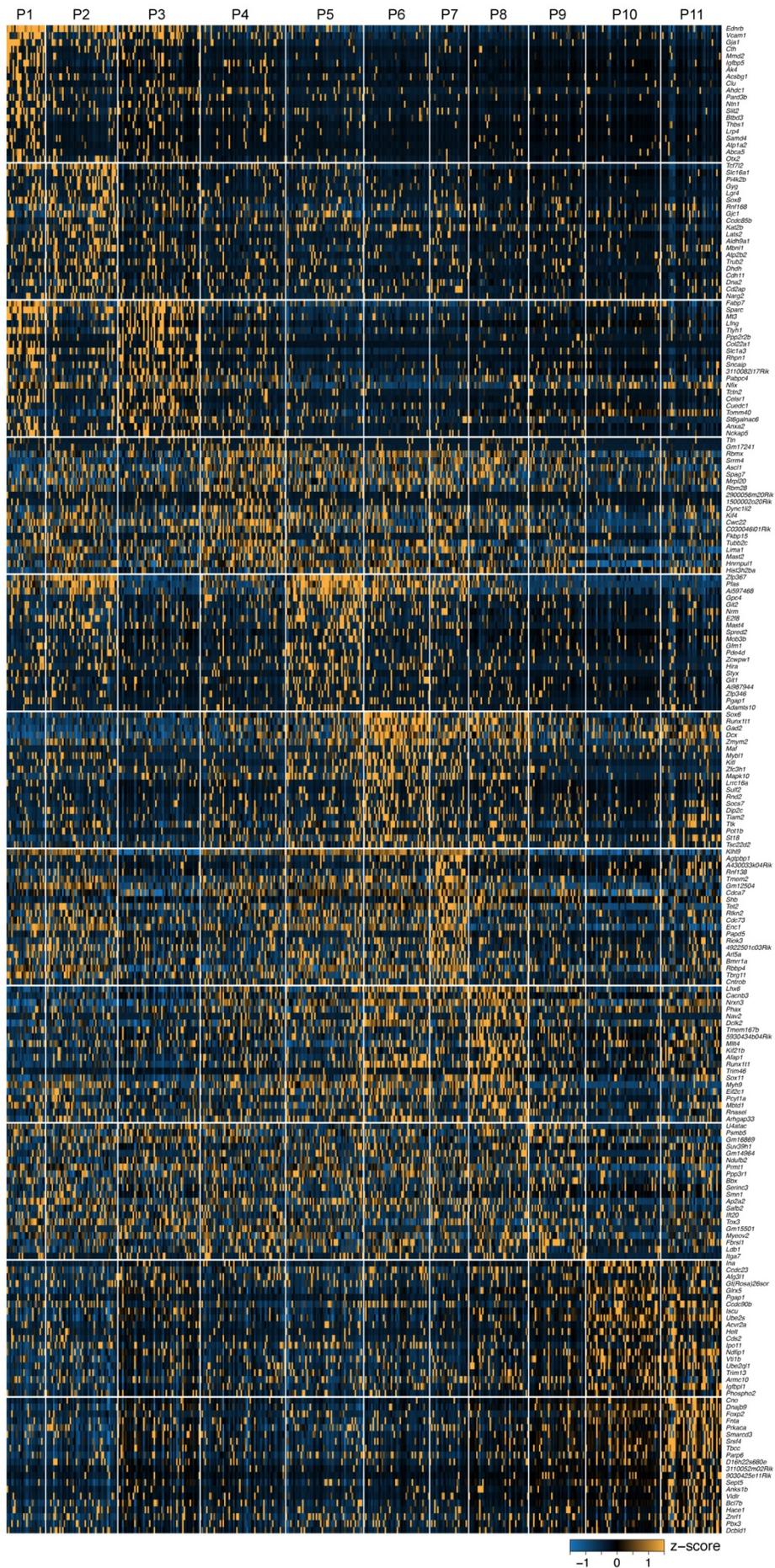


Fig. S14. Heatmap of top 20 differentially expressed genes among E14.5 progenitor clusters.

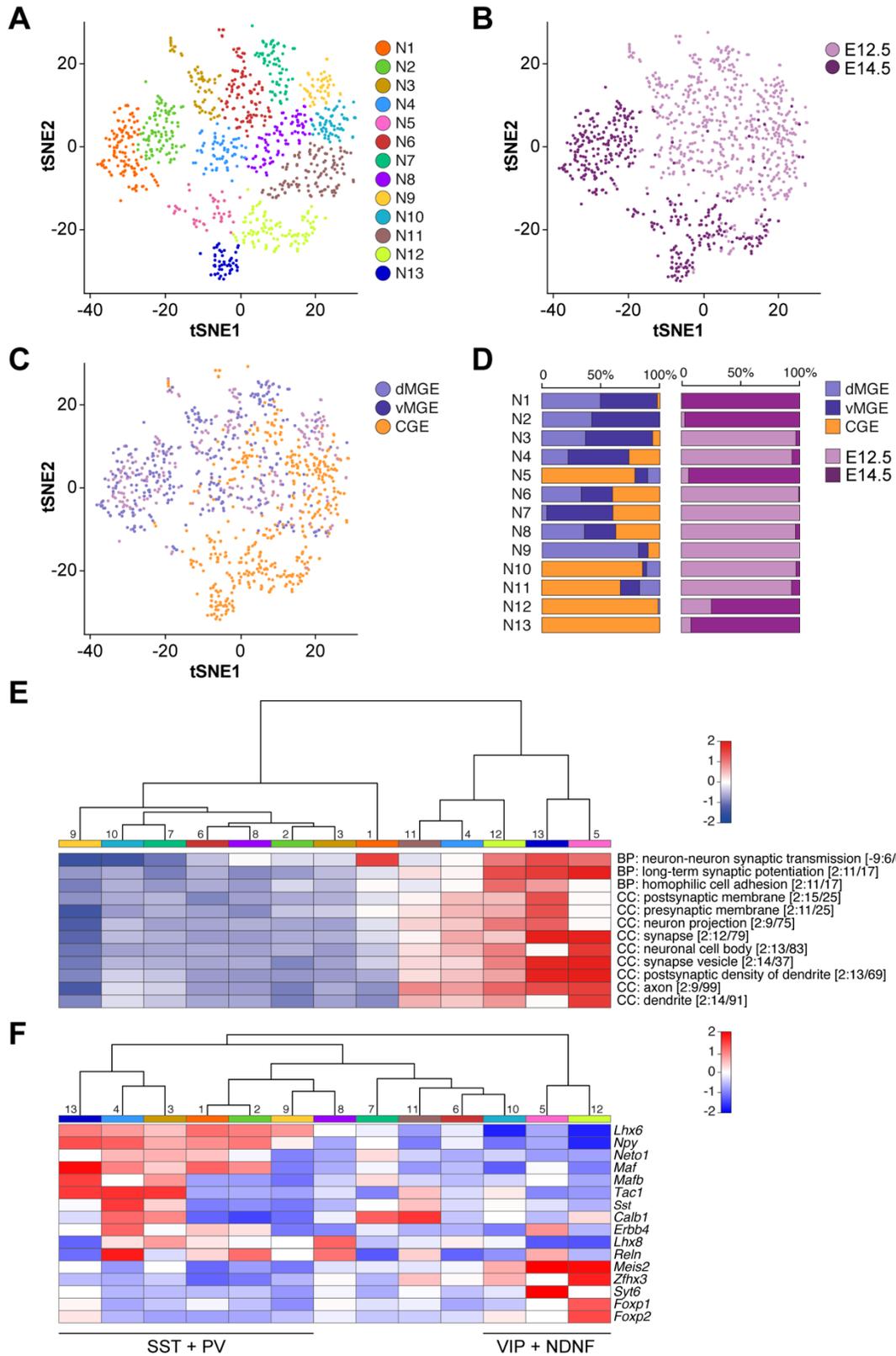
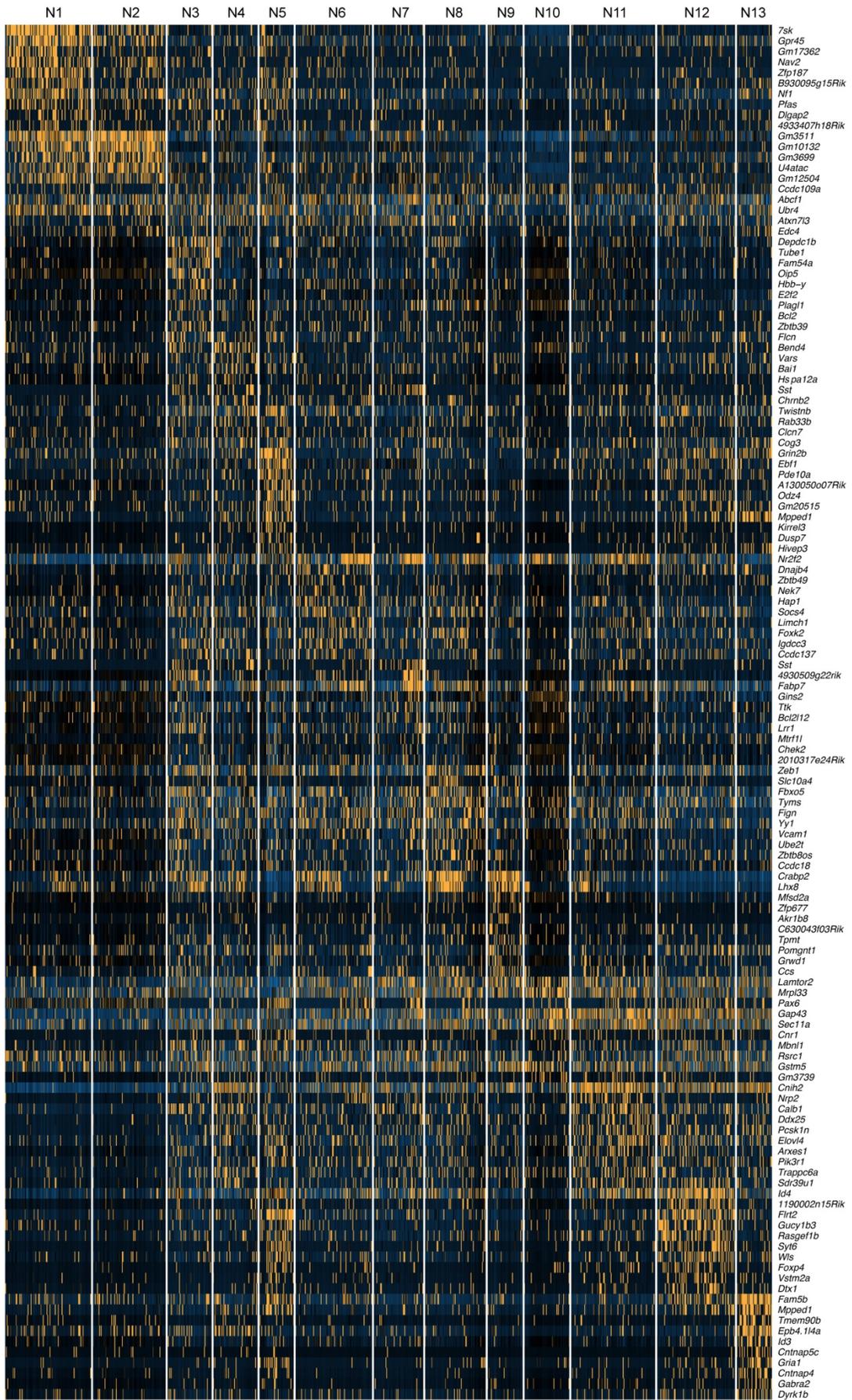


Fig. S15. Characterization of neuronal clusters. (A) t-SNE plot showing neuronal clusters by unsupervised Luvain-Jaccard clustering with all neurons. Neurons from E12.5 and E14.5 sample were jointly classified together into 13 clusters. Cells are colored by cluster identity. (B and C) The same t-SNE plot as in (A) highlights the stage (B) and (C) region of origin of single cells. Neurons from E12.5 and E14.5 are largely separated in the t-SNE plot. Similarly, MGE (purple) and CGE (orange) neurons are largely segregated in the t-SNE plot. These results reveal significant transcriptomic differences among neurons from both stages and regions. (D) Histograms showing the percentage of cells from dMGE, vMGE or CGE (left) and from E12.5 or E14.5 samples (right) in each of the neuronal clusters established through unsupervised clustering. (E) Heatmap showing gene ontology (GO) signatures for several aspects of neuronal functions, including synaptogenesis, neurotransmission, neuronal projection and development of axon and dendrite. The dendrogram generated by hierarchical clustering of GO enrichment profile suggests that the identified 13 neuronal clusters can also be broadly classified into two main groups that are distinguished by their maturation status. (F) Heatmap illustrating average expression of known interneuron lineage associated genes in the identified neuronal clusters. *Lhx6*, *Npy*, *Neto1*, *Maf*, *Mafb*, *Sst* and *Calb1* are enriched in MGE-derived interneuron lineages (SST+ and PV+ interneurons), while *Meis2*, *Zfhx3*, *Syt6*, *Foxp1* and *Foxp2* are enriched in CGE-derived interneuron lineages (VIP+ and NDNF+).



-1 0 1 Z-score

Fig. S16. Heatmap of top 10 differentially expressed genes between neuronal clusters.

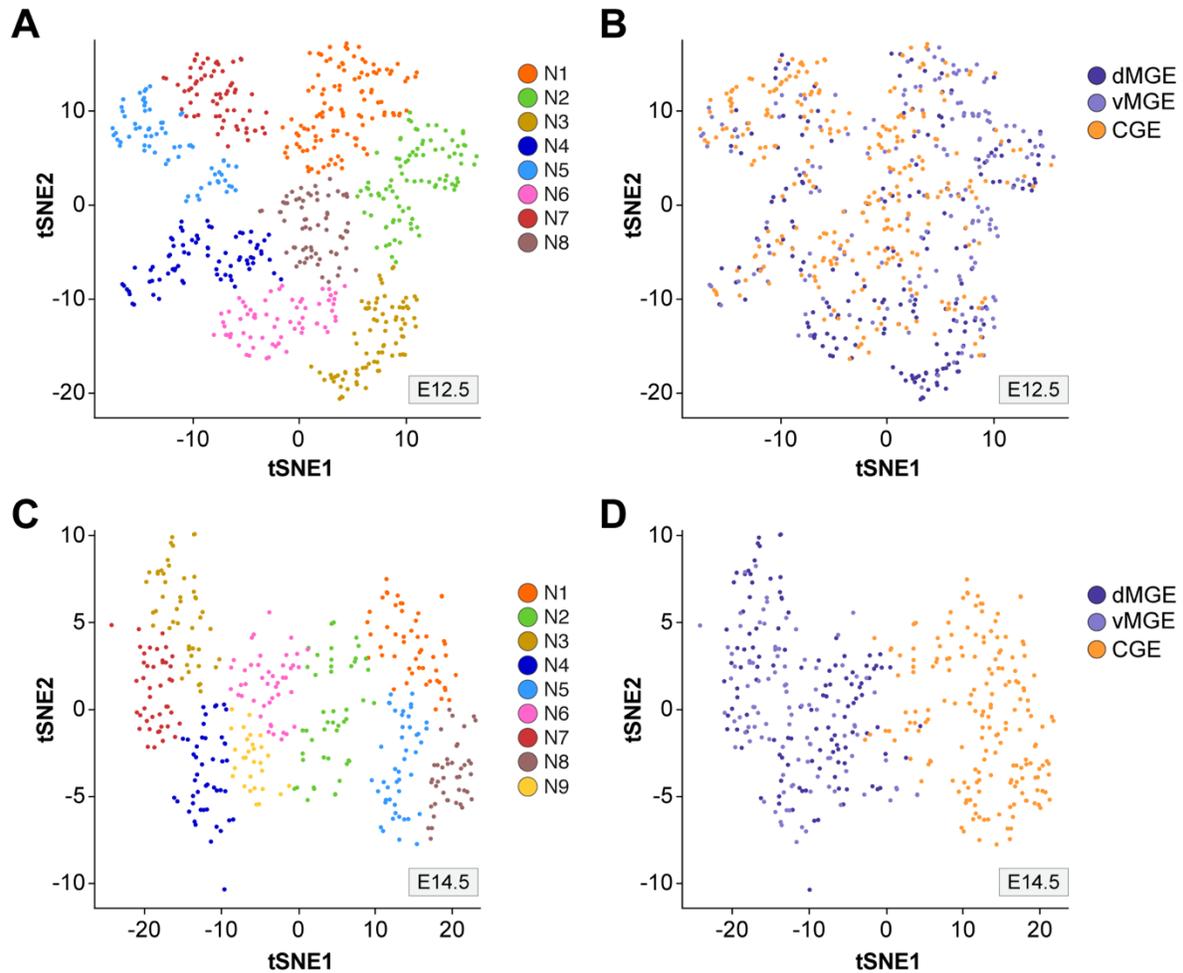


Fig. S17. Characterization of neuronal clusters. (A and C) t-SNE plots showing neuronal clusters by unsupervised Luvain-Jaccard clustering of separate E12.5 (A) and E14.5 (C) datasets. Cells are colored by cluster identity. Neurons from E12.5 and E14.5 were classified into 8 and 9 clusters, respectively. (B and D) The same t-SNE plots as in (A) and (C) highlight the region of origin of single cells for E12.5 (B) and E14.5 (D) neurons. Regional identity is more clearly defined at E14.5 than at E12.5. This may suggest that neurons are more transcriptionally homogeneous at early stages of development. Alternatively, CGE dissections may contain a heterogeneous pool of neurons, including many with molecular features similar to those of neurons generated in the MGE. This is consistent with the idea that the progenitor cells in the CGE are molecularly heterogeneous (see schema in Fig. 1A).

Fig. S18. Heatmap of top 40 differentially expressed genes among the 6 embryonic interneuron classes assigned through random forest analysis.

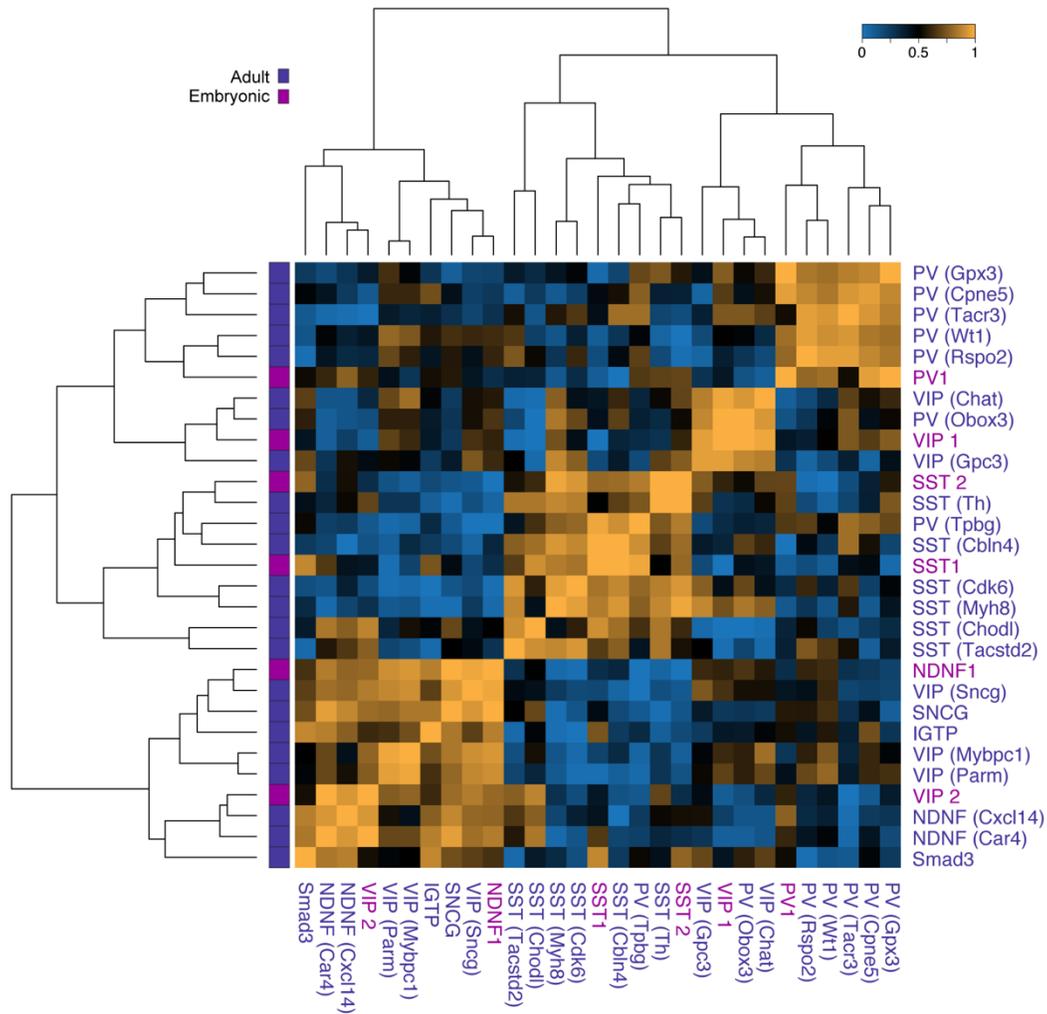


Fig. S19. MetaNeighbor analysis of 6 assigned embryonic interneuron cell types (random forest) and 23 adult interneuron types. Heatmap of mean AUROC scores (Area under the Receiver Operating Characteristic) indicates the degree of correlation among the 6 assigned embryonic classes and the 23 adult cortical interneuron cell types. The color bar indicates the origin (embryonic or adult) of each cluster. Hierarchical clustering indicates good correlation between assigned embryonic interneuron classes and the corresponding adult cortical interneuron cell types, with major branches representing PV, SST, VIP and NDNF lineages. The table shows mean AUROC scores between assigned embryonic interneuron classes and adult cortical interneuron cell types. A mean AUROC score of 0.7 or above typically suggests a good correlation, while mean AUROC score below 0.5 indicates no correlation.

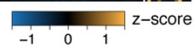
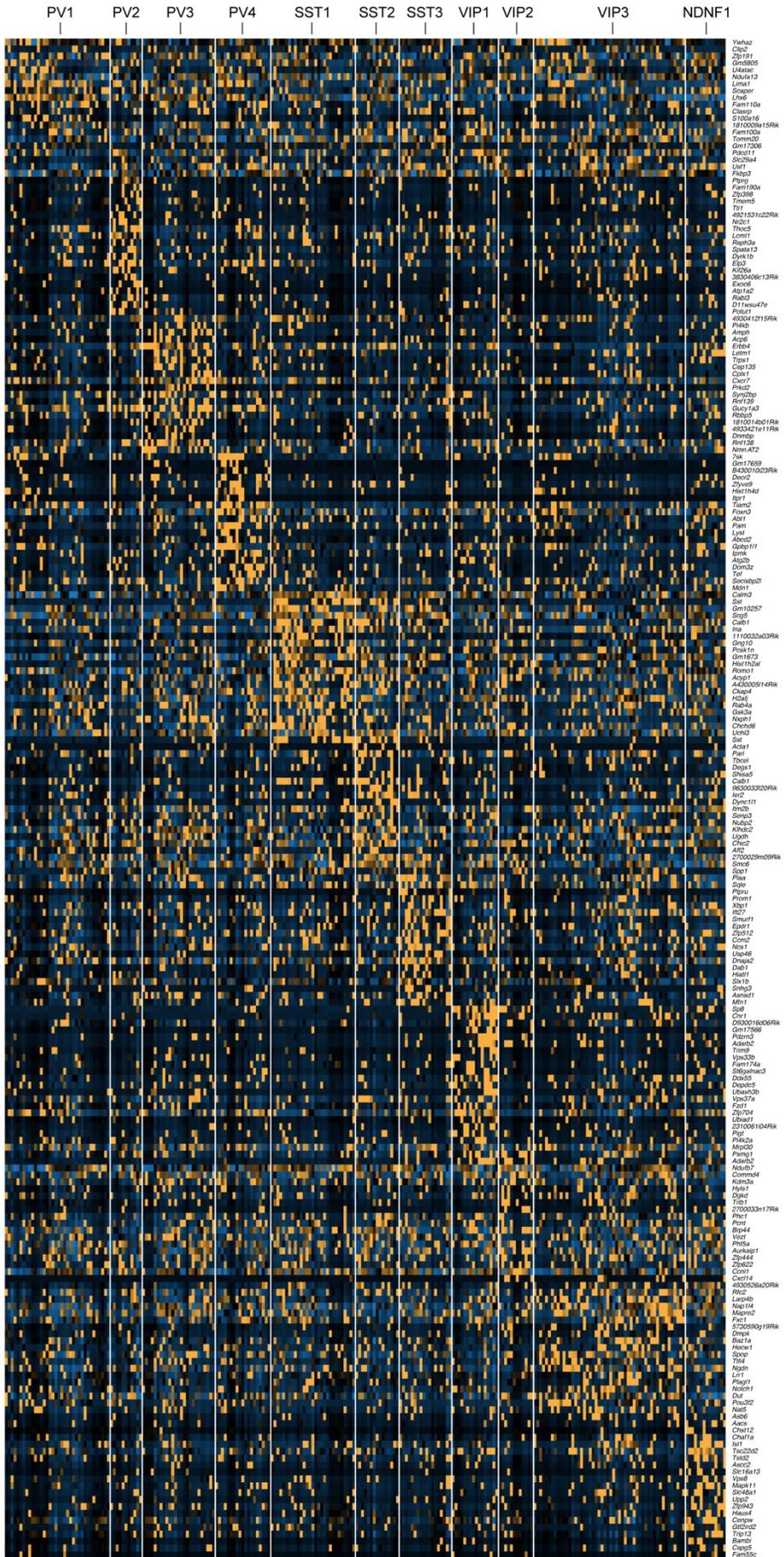
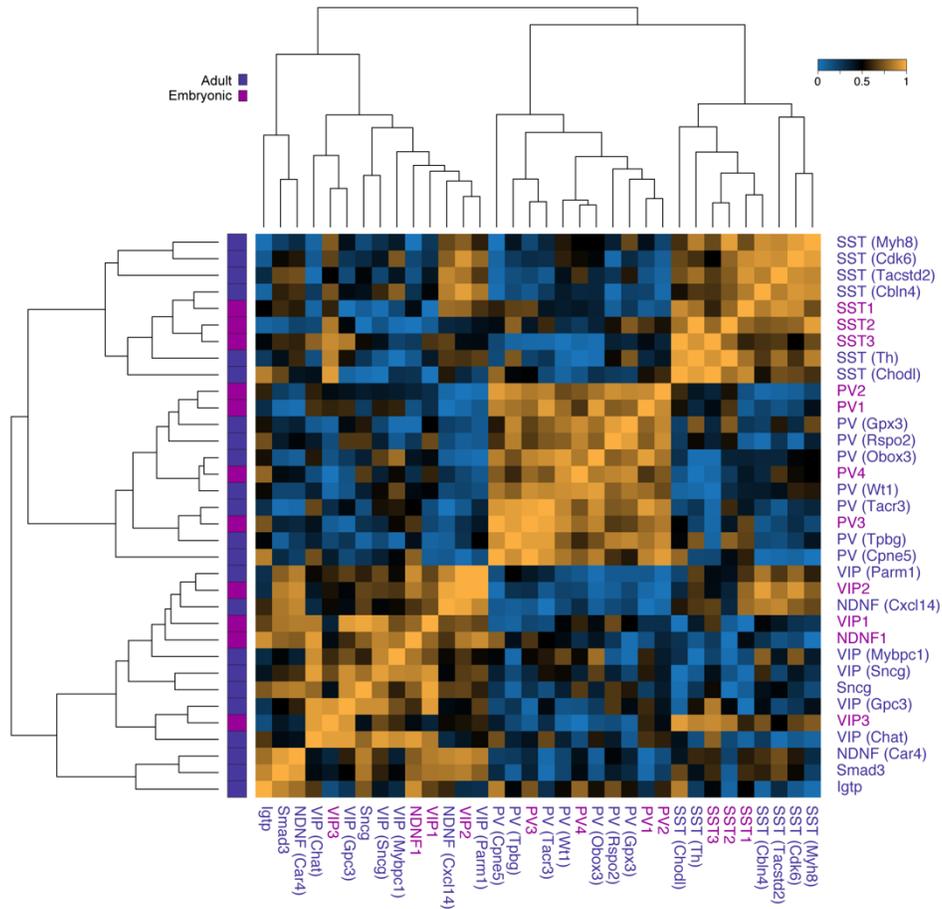


Fig. S20. Heatmap of top 20 differentially expressed genes among the 11 embryonic interneuron classes assigned through canonical correlation analysis and k-nearest neighbor.



Adult	CCA embryonic assignment	AUROC
PV (Gpx3)	PV1	0.79
PV (Rspo2)	PV2	0.91
PV (Tacr3)	PV3	0.98
PV (Wt1)	PV4	0.93
SST (Cbln4)	SST1	0.86
SST (Myh8)	SST2	0.95
SST (Chodl)	SST3	1.00
VIP (Chat)	VIP1	0.85
VIP (Parm1)	VIP2	1.00
VIP (Gpc3)	VIP3	0.93
NDNF (Car4)	NDNF1	0.76

Fig. S21. MetaNeighbor analysis of 11 assigned embryonic interneuron classes (canonical correlation analysis and k-nearest neighbor) and 23 adult interneuron types. Heatmap of mean AUROC scores (Area under the Receiver Operating Characteristic) indicates the degree of correlation among 11 assigned embryonic classes and the 23 adult cortical interneuron cell types. The color bar indicates the origin (embryonic or adult) of each cluster. Hierarchical clustering indicates good correlation between assigned embryonic interneuron classes and the corresponding adult cortical interneuron cell types, with major branches representing PV, SST, VIP and NDNF lineages. The table shows mean AUROC scores between assigned embryonic interneuron classes and adult cortical interneuron cell types. A mean AUROC score of 0.7 or above typically suggests a good correlation, while mean AUROC score below 0.5 indicates no correlation.

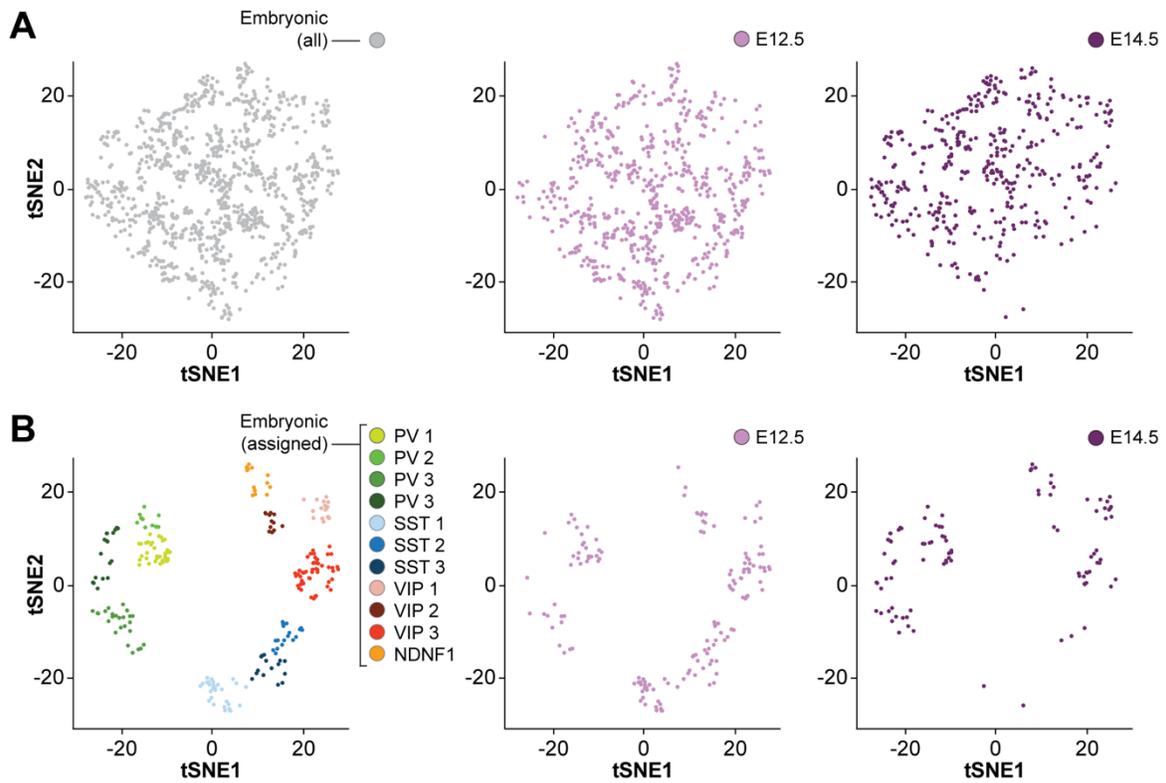


Fig. S22. Temporal analysis of embryonic neurons with assigned subtype identity. (A) All embryonic neurons are represented in the same t-SNE space as in Fig. 3B. Neurons are colored based on the stage (E12.5 or E14.5) of origin. (B) Assigned embryonic neurons are represented in the same t-SNE space as in Fig. 3C. Neurons are colored based on their stage (E12.5 or E14.5) of origin.

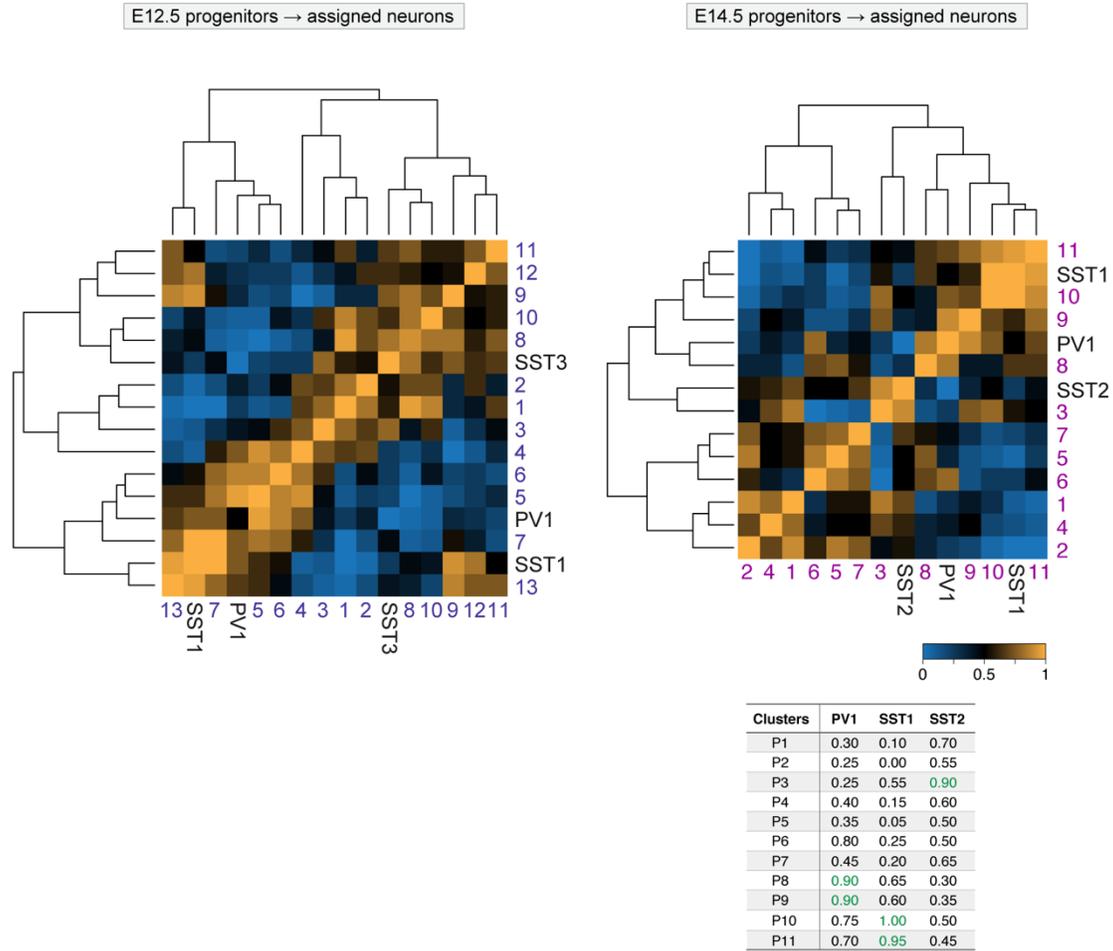


Fig. S23. MetaNeighbor analysis of progenitor clusters and assigned embryonic interneurons. Heatmaps of mean AUROC scores (Area under the Receiver Operating Characteristic) indicating the degree of correlation among E12.5 (left) and E14.5 (right) progenitor cell clusters and the 3 most conserved assigned embryonic interneuron cell types. Hierarchical clustering indicates correlations among progenitor clusters and with embryonic interneuron cell types. The table shows mean AUROC scores among E14.5 progenitor cell clusters and the 3 “consensus” embryonic interneuron classes (AUROC scores above 0.9). The corresponding table for E12.5 progenitor cell clusters and “consensus” embryonic interneurons is shown in Fig. 4A.

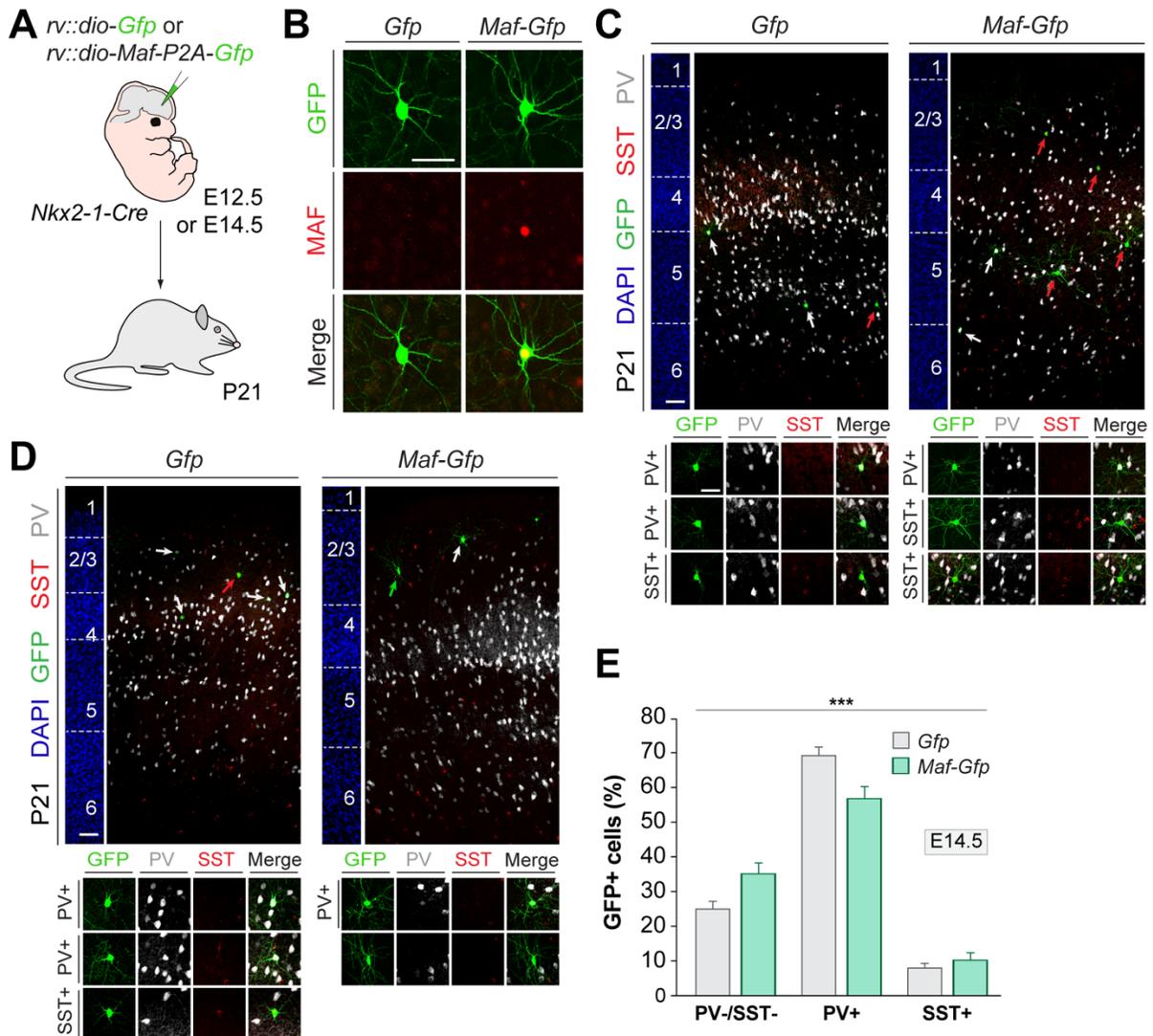


Figure S24. Maf is an early determinant of SST+ interneuron fate. (A) Schematic of experimental design. (B) Coronal sections through the P21 somatosensory cortex immunostained for GFP and MAF following viral infection with *Gfp* and *Maf-Gfp* expressing retroviruses in the MGE of *Nkx2-1-Cre* embryos at E12.5 or E14.5. (C and D) Coronal sections through the somatosensory cortex of P21 mice immunostained for GFP, PV and SST following viral infection in the MGE of *Nkx2-1-Cre* embryos at E12.5 (C) or E14.5 (D). Green, white and red arrows respectively point at PV-/SST-, PV+ and SST+ infected cells. Higher magnification images illustrate cortical interneuron fate changes following *Maf-P2A-Gfp* expression in MGE progenitor cells. (E) Quantification of the proportion of PV-/SST-, PV+ and SST+ interneurons in the cortex of P21 mice following viral infection with *Gfp* and *Maf-Gfp* expressing retroviruses in the MGE of *Nkx2-1-Cre* embryos at E14.5; $n = 5$ brains for each experimental condition; X^2 -square test, *** $p < 0.001$. Post-hoc binomial test pairwise comparison with Bonferroni correction; PV-/SST- vs PV+ *** $p < 0.001$, PV+ vs SST+ * $p < 0.05$. Scale bars equal $50 \mu\text{m}$ (B and small insets in C and D) and $100 \mu\text{m}$ (C and D).

Gene name	Identity
<i>Gfap</i>	Progenitor
<i>Slc1a3</i>	Progenitor
<i>Fabp7</i>	Progenitor
<i>Hes1</i>	Progenitor
<i>Hes5</i>	Progenitor
<i>Lhx2</i>	Progenitor
<i>Notch1</i>	Progenitor
<i>Notch2</i>	Progenitor
<i>Notch3</i>	Progenitor
<i>Vim</i>	Progenitor
<i>Sox2</i>	Progenitor
<i>Sall3</i>	Progenitor
<i>Sparc</i>	Progenitor
<i>Pdpn</i>	Progenitor
<i>Tubb3</i>	Neuron
<i>Mapt</i>	Neuron
<i>Dcx</i>	Neuron
<i>Nrxn3</i>	Neuron
<i>Stmn2</i>	Neuron
<i>Stmn3</i>	Neuron
<i>Gng3</i>	Neuron
<i>Scn3a</i>	Neuron
<i>Map2</i>	Neuron
<i>Gap43</i>	Neuron
<i>Tmem130</i>	Neuron
<i>L1cam</i>	Neuron
<i>Gria2</i>	Neuron

Table S1. Progenitor and neuronal markers.

Gene Name	Identity	Source	Gene Name	Identity	Source
<i>Fabp7</i>	VZ	Random forest features	<i>Esrrg</i>	CGE	Differential expression analysis
<i>Slc1a3</i>	VZ	Random forest features	<i>Foxp2</i>	CGE	Differential expression analysis
<i>1190002h23Rik</i>	VZ	Random forest features	<i>Gas1</i>	CGE	Differential expression analysis
<i>Snap23</i>	VZ	Random forest features	<i>Fktn</i>	CGE	Differential expression analysis
<i>Fgfbp3</i>	VZ	Random forest features	<i>Lix1</i>	CGE	Differential expression analysis
<i>Rhpn1</i>	VZ	Random forest features	<i>Gli2</i>	CGE	Differential expression analysis
<i>Rsrc2</i>	SVZ	Random forest features	<i>Ntrk2</i>	CGE	Differential expression analysis
<i>Nrxn3</i>	SVZ	Random forest features	<i>Anks1b</i>	CGE	Differential expression analysis
<i>St18</i>	SVZ	Random forest features	<i>Chmp2a</i>	CGE	Differential expression analysis
<i>Cd24a</i>	SVZ	Random forest features	<i>Runx1t1</i>	Neuron	Random forest features
<i>Dlx6os1</i>	SVZ	Random forest features	<i>Dcx</i>	Neuron	Random forest features
<i>Nrxn1</i>	dMGE	Differential expression analysis	<i>Gad2</i>	Neuron	Random forest features
<i>Jag1</i>	dMGE	Differential expression analysis	<i>Gad1</i>	Neuron	Random forest features
<i>Nkx6-2</i>	dMGE	Differential expression analysis	<i>Nrxn3</i>	Neuron	Random forest features
<i>Gm5069</i>	dMGE	Differential expression analysis	<i>Dlx60s1</i>	Neuron	Random forest features
<i>Npy</i>	dMGE	Differential expression analysis	<i>Gm14204</i>	Neuron	Random forest features
<i>Nek7</i>	dMGE	Differential expression analysis	<i>Lhx6</i>	Neuron	Random forest features
<i>8430410k20Rik</i>	dMGE	Differential expression analysis	<i>Mapk10</i>	Neuron	Random forest features
<i>Mks1</i>	dMGE	Differential expression analysis	<i>Maged1</i>	Neuron	Random forest features
<i>Nrp1</i>	dMGE	Differential expression analysis	<i>Mapt</i>	Neuron	Random forest features
<i>Ebf1</i>	dMGE	Differential expression analysis	<i>Erbp4</i>	Neuron	Random forest features
<i>Hspa12a</i>	dMGE	Differential expression analysis	<i>Gng2</i>	Neuron	Random forest features
<i>Wnk3-ps</i>	dMGE	Differential expression analysis	<i>Crmp1</i>	Neuron	Random forest features
<i>2310014h01Rik</i>	dMGE	Differential expression analysis	<i>Tubb3</i>	Neuron	Random forest features
<i>Mavs</i>	dMGE	Differential expression analysis	<i>Slain1</i>	Neuron	Random forest features
<i>Lgals1</i>	dMGE	Differential expression analysis	<i>Mllt11</i>	Neuron	Random forest features
<i>Zfp568</i>	dMGE	Differential expression analysis	<i>Gap43</i>	Neuron	Random forest features
<i>Ptdss2</i>	dMGE	Differential expression analysis	<i>Gabrg2</i>	Neuron	Random forest features
<i>Got2</i>	dMGE	Differential expression analysis	<i>2900011008Rik</i>	Neuron	Random forest features
<i>C130036l24Rik</i>	dMGE	Differential expression analysis	<i>Myo10</i>	Progenitor	Random forest features
<i>Lhx8</i>	vMGE	Differential expression analysis	<i>Gpr98</i>	Progenitor	Random forest features
<i>Nkx2-1</i>	vMGE	Differential expression analysis	<i>Ednrb</i>	Progenitor	Random forest features
<i>Zic1</i>	vMGE	Differential expression analysis	<i>Ddah1</i>	Progenitor	Random forest features
<i>Mbip</i>	vMGE	Differential expression analysis	<i>Slc1a3</i>	Progenitor	Random forest features
<i>Asb4</i>	vMGE	Differential expression analysis	<i>Fabp7</i>	Progenitor	Random forest features
<i>Sulf2</i>	vMGE	Differential expression analysis	<i>Ccne2</i>	Progenitor	Random forest features
<i>Sez6</i>	vMGE	Differential expression analysis	<i>Hells</i>	Progenitor	Random forest features
<i>Crabp2</i>	vMGE	Differential expression analysis	<i>Olig1</i>	Progenitor	Random forest features
<i>Lhx6</i>	vMGE	Differential expression analysis	<i>Hat1</i>	Progenitor	Random forest features
<i>Dach1</i>	vMGE	Differential expression analysis	<i>Mcm3</i>	Progenitor	Random forest features
<i>Serpine2</i>	vMGE	Differential expression analysis	<i>Mcm4</i>	Progenitor	Random forest features
<i>Th</i>	vMGE	Differential expression analysis	<i>Ccnd2</i>	Progenitor	Random forest features
<i>Cdkn1c</i>	vMGE	Differential expression analysis	<i>Nasp</i>	Progenitor	Random forest features
<i>Pde1c</i>	vMGE	Differential expression analysis	<i>Mcm6</i>	Progenitor	Random forest features
<i>Etv1</i>	vMGE	Differential expression analysis	<i>Tacc3</i>	Progenitor	Random forest features
<i>Spp1</i>	vMGE	Differential expression analysis	<i>Kif22</i>	Progenitor	Random forest features
<i>Lipg</i>	vMGE	Differential expression analysis	<i>Cdc6</i>	Progenitor	Random forest features
<i>Ephb3</i>	vMGE	Differential expression analysis	<i>Clsn</i>	Progenitor	Random forest features
<i>Zic4</i>	vMGE	Differential expression analysis	<i>Mki67</i>	Progenitor	Random forest features
<i>2610035d17Rik</i>	vMGE	Differential expression analysis	<i>Top2a</i>	Progenitor	Random forest features
<i>Pax6</i>	CGE	Differential expression analysis	<i>Rrm2</i>	Progenitor	Random forest features
<i>Nr2f2</i>	CGE	Differential expression analysis	<i>Psmc4</i>	Progenitor	Random forest features
<i>Epha5</i>	CGE	Differential expression analysis	<i>Ska1</i>	Progenitor	Random forest features
<i>Nr2f1</i>	CGE	Differential expression analysis	<i>Nde1</i>	Progenitor	Random forest features
<i>Sp8</i>	CGE	Differential expression analysis	<i>Aspm</i>	Progenitor	Random forest features
<i>Ube2q1</i>	CGE	Differential expression analysis	<i>Cenpe</i>	Progenitor	Random forest features
<i>Rbm28</i>	CGE	Differential expression analysis	<i>Pfas</i>	Progenitor	Random forest features
<i>Bax</i>	CGE	Differential expression analysis	<i>Zfp367</i>	Progenitor	Random forest features
<i>Fbxw8</i>	CGE	Differential expression analysis	<i>Faim2</i>	Progenitor	Random forest features
<i>Pak3</i>	CGE	Differential expression analysis	<i>Pip5k11</i>	Progenitor	Random forest features
<i>Ppp1r1a</i>	CGE	Differential expression analysis	<i>Frrmpd1</i>	Progenitor	Random forest features

Table S2. Progenitor cell selection markers (E12.5).

Gene Name	Identity	Source	Gene Name	Identity	Source
<i>Mfge8</i>	VZ	Random forest features	<i>Camk1</i>	CGE	Differential expression analysis
<i>Sparc</i>	VZ	Random forest features	<i>Vps37b</i>	CGE	Differential expression analysis
<i>Ednrb</i>	VZ	Random forest features	<i>1500011H22Rik</i>	CGE	Differential expression analysis
<i>Slc1a3</i>	VZ	Random forest features	<i>Cno</i>	CGE	Differential expression analysis
<i>Fabp7</i>	VZ	Random forest features	<i>Cnih2</i>	CGE	Differential expression analysis
<i>Myo10</i>	VZ	Random forest features	<i>Stard3nl</i>	CGE	Differential expression analysis
<i>Gpr98</i>	VZ	Random forest features	<i>Ntrk2</i>	CGE	Differential expression analysis
<i>Tubb3</i>	SVZ	Random forest features	<i>Kxne11</i>	CGE	Differential expression analysis
<i>Lhx6</i>	SVZ	Random forest features	<i>ler5l</i>	CGE	Differential expression analysis
<i>St18</i>	SVZ	Random forest features	<i>Foxp2</i>	CGE	Differential expression analysis
<i>Nrxn3</i>	SVZ	Random forest features	<i>Pid1</i>	CGE	Differential expression analysis
<i>Dlx6os1</i>	SVZ	Random forest features	<i>Runx1t1</i>	Neuron	Random forest features
<i>Nudt4</i>	SVZ	Random forest features	<i>Dcx</i>	Neuron	Random forest features
<i>Gad2</i>	SVZ	Random forest features	<i>Gad2</i>	Neuron	Random forest features
<i>Dcx</i>	SVZ	Random forest features	<i>Gad1</i>	Neuron	Random forest features
<i>Runx1t1</i>	SVZ	Random forest features	<i>Nrxn3</i>	Neuron	Random forest features
<i>Gm10837</i>	dMGE	Differential expression analysis	<i>Dlx6os1</i>	Neuron	Random forest features
<i>Gm10717</i>	dMGE	Differential expression analysis	<i>Gm14204</i>	Neuron	Random forest features
<i>Gm6984</i>	dMGE	Differential expression analysis	<i>Lhx6</i>	Neuron	Random forest features
<i>Gm10282</i>	dMGE	Differential expression analysis	<i>Mapk10</i>	Neuron	Random forest features
<i>Atp5g1</i>	dMGE	Differential expression analysis	<i>Maged1</i>	Neuron	Random forest features
<i>Gm10801</i>	dMGE	Differential expression analysis	<i>mapt</i>	Neuron	Random forest features
<i>Eid2</i>	dMGE	Differential expression analysis	<i>Erbp4</i>	Neuron	Random forest features
<i>Uqcrl1</i>	dMGE	Differential expression analysis	<i>Gng2</i>	Neuron	Random forest features
<i>Serf2</i>	dMGE	Differential expression analysis	<i>Crmp1</i>	Neuron	Random forest features
<i>Sfi1</i>	dMGE	Differential expression analysis	<i>Tubb3</i>	Neuron	Random forest features
<i>Gm8730</i>	dMGE	Differential expression analysis	<i>Slain1</i>	Neuron	Random forest features
<i>Gm9846</i>	dMGE	Differential expression analysis	<i>Mllt11</i>	Neuron	Random forest features
<i>Gm8759</i>	dMGE	Differential expression analysis	<i>Gap43</i>	Neuron	Random forest features
<i>Atp5g2</i>	dMGE	Differential expression analysis	<i>Gabrg2</i>	Neuron	Random forest features
<i>Gm10106</i>	dMGE	Differential expression analysis	<i>2900011O08Rik</i>	Neuron	Random forest features
<i>Gria3</i>	dMGE	Differential expression analysis	<i>Myo10</i>	Progenitor	Random forest features
<i>Cwc22</i>	dMGE	Differential expression analysis	<i>Gpr98</i>	Progenitor	Random forest features
<i>Sepw1</i>	dMGE	Differential expression analysis	<i>Ednrb</i>	Progenitor	Random forest features
<i>Gm10718</i>	dMGE	Differential expression analysis	<i>Ddah1</i>	Progenitor	Random forest features
<i>Cdk6</i>	vMGE	Differential expression analysis	<i>Slc1a3</i>	Progenitor	Random forest features
<i>Rbm15</i>	vMGE	Differential expression analysis	<i>Fabp7</i>	Progenitor	Random forest features
<i>Sept2</i>	vMGE	Differential expression analysis	<i>Ccne2</i>	Progenitor	Random forest features
<i>Gm3272</i>	vMGE	Differential expression analysis	<i>Hells</i>	Progenitor	Random forest features
<i>Set</i>	vMGE	Differential expression analysis	<i>Olig1</i>	Progenitor	Random forest features
<i>Poldip3</i>	vMGE	Differential expression analysis	<i>Hat1</i>	Progenitor	Random forest features
<i>Ptges3</i>	vMGE	Differential expression analysis	<i>Mcm3</i>	Progenitor	Random forest features
<i>Nkx2-1</i>	vMGE	Differential expression analysis	<i>Mcm4</i>	Progenitor	Random forest features
<i>Skil</i>	vMGE	Differential expression analysis	<i>Ccnd2</i>	Progenitor	Random forest features
<i>Sez6</i>	vMGE	Differential expression analysis	<i>Nasp</i>	Progenitor	Random forest features
<i>Rbm26</i>	vMGE	Differential expression analysis	<i>Mcm6</i>	Progenitor	Random forest features
<i>Lhx8</i>	vMGE	Differential expression analysis	<i>Tacc3</i>	Progenitor	Random forest features
<i>Zswim5</i>	vMGE	Differential expression analysis	<i>Kif22</i>	Progenitor	Random forest features
<i>Ttr</i>	vMGE	Differential expression analysis	<i>Cdc6</i>	Progenitor	Random forest features
<i>Map3k1</i>	vMGE	Differential expression analysis	<i>Clspn</i>	Progenitor	Random forest features
<i>Limd1</i>	vMGE	Differential expression analysis	<i>Mki67</i>	Progenitor	Random forest features
<i>Mt1</i>	vMGE	Differential expression analysis	<i>Top2a</i>	Progenitor	Random forest features
<i>Crabp2</i>	vMGE	Differential expression analysis	<i>Rrm2</i>	Progenitor	Random forest features
<i>Ai597468</i>	vMGE	Differential expression analysis	<i>Psmc4</i>	Progenitor	Random forest features
<i>Gm12940</i>	vMGE	Differential expression analysis	<i>Ska1</i>	Progenitor	Random forest features
<i>Jund</i>	CGE	Differential expression analysis	<i>Nde1</i>	Progenitor	Random forest features
<i>Ube2s</i>	CGE	Differential expression analysis	<i>Aspm</i>	Progenitor	Random forest features
<i>Glx5</i>	CGE	Differential expression analysis	<i>Cenpe</i>	Progenitor	Random forest features
<i>Med30</i>	CGE	Differential expression analysis	<i>Pfas</i>	Progenitor	Random forest features
<i>Dnajb9</i>	CGE	Differential expression analysis	<i>Zfp367</i>	Progenitor	Random forest features
<i>Rnf126</i>	CGE	Differential expression analysis	<i>Faim2</i>	Progenitor	Random forest features
<i>Gm8327</i>	CGE	Differential expression analysis	<i>Pip5k1l</i>	Progenitor	Random forest features
<i>Podxl2</i>	CGE	Differential expression analysis	<i>Frmpd1</i>	Progenitor	Random forest features
<i>Rap1b</i>	CGE	Differential expression analysis			

Table S3. Progenitor cell selection markers (E14.5).

Gene name	Identity	Gene name	Identity	Gene name	Identity
<i>Dkc1</i>	VZ	<i>E2f5</i>	VZ	<i>Ldha</i>	VZ
<i>Ckb</i>	VZ	<i>Mycn</i>	VZ	<i>Mki67</i>	VZ
<i>E2f1</i>	VZ	<i>Hist1h3e</i>	VZ	<i>Gm8096</i>	VZ
<i>Mcm6</i>	VZ	<i>Notch2</i>	VZ	<i>Phgdh</i>	VZ
<i>Tsen15</i>	VZ	<i>Sox2ot</i>	VZ	<i>Adk</i>	VZ
<i>Hist1h2ah</i>	VZ	<i>Snrnp25</i>	VZ	<i>Fam102b</i>	VZ
<i>Hist1h2ai</i>	VZ	<i>Tacc3</i>	VZ	<i>Polg</i>	VZ
<i>Hist1h2an</i>	VZ	<i>Fen1</i>	VZ	<i>Dtl</i>	VZ
<i>Hist1h2ad</i>	VZ	<i>Hist2h2ac</i>	VZ	<i>Clspn</i>	VZ
<i>Cdca5</i>	VZ	<i>Snap23</i>	VZ	<i>Crb2</i>	VZ
<i>Lsm3</i>	VZ	<i>Nr2f1</i>	VZ	<i>Idi1</i>	VZ
<i>Aen</i>	VZ	<i>Slc25a5</i>	VZ	<i>Ppat</i>	VZ
<i>Vim</i>	VZ	<i>Srm</i>	VZ	<i>Fbl</i>	VZ
<i>E2f2</i>	VZ	<i>Hmga2</i>	VZ	<i>Hspa8</i>	VZ
<i>Fgfr2</i>	VZ	<i>Vrk1</i>	VZ	<i>Rps2</i>	VZ
<i>Raver1</i>	VZ	<i>Mcm7</i>	VZ	<i>Ddah1</i>	VZ
<i>Pprc1</i>	VZ	<i>Ppia</i>	VZ	<i>Skp2</i>	VZ
<i>Lix1</i>	VZ	<i>Sox21</i>	VZ	<i>Gldc</i>	VZ
<i>Ltv1</i>	VZ	<i>Zfp238</i>	VZ	<i>119002H23Rik</i>	VZ
<i>Timm23</i>	VZ	<i>Hist1h4d</i>	VZ	<i>H2afx</i>	VZ
<i>2810417H13Rik</i>	VZ	<i>Hist1h4j</i>	VZ	<i>Ttyh1</i>	VZ
<i>Mthfd1</i>	VZ	<i>Hist1h4a</i>	VZ	<i>Lyar</i>	VZ
<i>Hmgn2</i>	VZ	<i>Hist1h4k</i>	VZ	<i>Slc1a3</i>	VZ
<i>Fgfr1</i>	VZ	<i>Hist1h4f</i>	VZ	<i>BC048355</i>	VZ
<i>Ascl1</i>	VZ	<i>Hist1h4b</i>	VZ	<i>Rrm1</i>	VZ
<i>Hist1h1e</i>	VZ	<i>Hdh2</i>	VZ	<i>Pbk</i>	VZ
<i>Racgap1</i>	VZ	<i>Anp32b</i>	VZ	<i>Mif</i>	VZ
<i>Ccnd1</i>	VZ	<i>Sox9</i>	VZ	<i>Gkap1</i>	VZ
<i>0610007L01Rik</i>	VZ	<i>Hist1h1d</i>	VZ	<i>Top2a</i>	VZ
<i>Pea15a</i>	VZ	<i>Cks2</i>	VZ	<i>Mcm2</i>	VZ
<i>Samd1</i>	VZ	<i>Rtkn2</i>	VZ	<i>Rpl12</i>	VZ
<i>Larp4</i>	VZ	<i>Hes5</i>	VZ	<i>Acss1</i>	VZ
<i>Pdpr</i>	VZ	<i>Ednrb</i>	VZ	<i>Hip1</i>	VZ
<i>Hist1h2af</i>	VZ	<i>Kif15</i>	VZ	<i>Nasp</i>	VZ
<i>Notch1</i>	VZ	<i>Nol8</i>	VZ	<i>Tmpo</i>	VZ
<i>Eif2a</i>	VZ	<i>Psmc3ip</i>	VZ	<i>Hprt</i>	VZ
<i>Dut</i>	VZ	<i>Pcna</i>	VZ	<i>Fancd2</i>	VZ
<i>Uhrf1</i>	VZ	<i>Hdac1</i>	VZ	<i>Acaa2</i>	VZ
<i>Myo10</i>	VZ	<i>Cdca3</i>	VZ	<i>Rps6ka6</i>	VZ
<i>Rrm2</i>	VZ	<i>Mcm5</i>	VZ	<i>Ilidr2</i>	VZ
<i>Usp39</i>	VZ	<i>Ethd2</i>	VZ	<i>Msh6</i>	VZ
<i>Msl3</i>	VZ	<i>C2cd3</i>	VZ	<i>Igfbp1</i>	VZ
<i>Pold1</i>	VZ	<i>Rfc4</i>	VZ	<i>Usp1</i>	VZ
<i>Siva1</i>	VZ	<i>Ckap2</i>	VZ	<i>Mcm3</i>	VZ
<i>Vamp3</i>	VZ	<i>Ran</i>	VZ	<i>Cdc73</i>	VZ
<i>Gcat</i>	VZ	<i>Gapdh</i>	VZ	<i>Tk1</i>	VZ
<i>Nes</i>	VZ	<i>Slbp</i>	VZ	<i>Hist1h2ak</i>	VZ
<i>Sgol1</i>	VZ	<i>Olig2</i>	VZ	<i>Fam136a</i>	VZ
<i>Kif11</i>	VZ	<i>Grb10</i>	VZ	<i>Rcc1</i>	VZ
<i>Lmo1</i>	VZ	<i>Gn3</i>	VZ	<i>Fgfr3</i>	VZ
<i>Nr2f6</i>	VZ	<i>Naa50</i>	VZ	<i>Dnajc19</i>	VZ
<i>Nr2e1</i>	VZ	<i>Gart</i>	VZ	<i>Hmgb2</i>	VZ
<i>Hist1h1b</i>	VZ	<i>Lhx2</i>	VZ	<i>Kif4</i>	VZ
<i>Mybl2</i>	VZ	<i>Lfng</i>	VZ	<i>St18</i>	SVZ
<i>Tuba1b</i>	VZ	<i>Sall3</i>	VZ	<i>Bcl11b</i>	SVZ
<i>Jun</i>	VZ	<i>Grwd1</i>	VZ	<i>Epha4</i>	SVZ
<i>Haus5</i>	VZ	<i>Id4</i>	VZ	<i>Top2b</i>	SVZ
<i>Tox3</i>	VZ	<i>Rnaseh2b</i>	VZ	<i>ORF19</i>	SVZ
<i>Itg2</i>	VZ	<i>Hist1h2bf</i>	VZ	<i>Sox11</i>	SVZ
<i>H2afz</i>	VZ	<i>Hist1h2bj</i>	VZ	<i>Arx</i>	SVZ
<i>Add3</i>	VZ	<i>Mcm4</i>	VZ	<i>Lhx6</i>	SVZ
<i>Nedd1</i>	VZ	<i>Hells</i>	VZ	<i>Nrxn3</i>	SVZ
<i>Gadd45g</i>	VZ	<i>Yap1</i>	VZ	<i>Dlx5</i>	SVZ
<i>Rad23a</i>	VZ	<i>Tbrg4</i>	VZ	<i>Pdcd4</i>	SVZ
<i>Hsp90aa1</i>	VZ	<i>Hist2h2ab</i>	VZ	<i>Rai2</i>	SVZ
<i>Nek6</i>	VZ	<i>Psat1</i>	VZ	<i>Pgm1</i>	SVZ
<i>Dek</i>	VZ	<i>Snhg1</i>	VZ	<i>Dlx1as</i>	SVZ
<i>Ccna2</i>	VZ	<i>Hist1h2aa</i>	VZ	<i>Sox6</i>	SVZ
<i>Nap111</i>	VZ	<i>Tcf19</i>	VZ	<i>Arl4d</i>	SVZ
<i>Wnt7b</i>	VZ	<i>Eed</i>	VZ	<i>Sp9</i>	SVZ
<i>Cdca8</i>	VZ	<i>Dbi</i>	VZ	<i>Dlg1</i>	SVZ
<i>Cenpq</i>	VZ	<i>Wdr76</i>	VZ	<i>Atg2b</i>	SVZ
<i>AI314976</i>	VZ	<i>lpo5</i>	VZ	<i>Fam48a</i>	SVZ
<i>Pa2g4</i>	VZ	<i>Fgfbp3</i>	VZ	<i>Tmsb4x</i>	SVZ
<i>Cad</i>	VZ	<i>Hspd1</i>	VZ	<i>Dlx1</i>	SVZ
<i>Hat1</i>	VZ	<i>Jam2</i>	VZ	<i>Dach2</i>	SVZ
<i>Mybbp1a</i>	VZ	<i>Nup37</i>	VZ	<i>Dclk2</i>	SVZ

Table S4. Ventricular zone and subventricular zone markers.