PNAS

Supplementary Information for

Specification of neuronal subtypes in the spiral ganglion begins prior to birth in the mouse

Tessa R. Sanders[1]* and Matthew W. Kelley[1]*

*Corresponding authors: Tessa R. Sanders and Matthew W. Kelley

Email: tessa.sanders@nih.gov and kelleymt@nidcd.nih.gov

This PDF file includes:

Supplementary methods

Figures S1 to S12

Tables S1 to S2

Other supplementary materials for this manuscript include the following:

Datasets S1 to S8

Supplementary methods

Animals: All mice were maintained within the Porter Neuroscience Research Center Shared Animal Facility. All animal care and housing was conducted in accordance with the NIH guidelines for animal use (Protocols 1254 and 1262). Randomly selected mice of either sex were used in all experiments. Timed-pregnant CD1 females were obtained from Charles River. $Ngn1^{creErt2}$ mice were obtained from Dr. Lisa Goodrich, Harvard Medical School(1) and Tac1Cre and $MAPT^{EGFP}$ mice were obtained from The Jackson Laboratory (Strain #021877(2) and strain # 004779 (3)), and were maintained on a mixed background.

Tissue collection: At E14, E16, and E18, pregnant CD1 females were euthanized and cochleae were dissected from 5 embryos. At P1, 5 pups were euthanized and cochleae were collected. At all ages spiral ganglia (SG) were grossly dissected out of the cochlea with some sensory epithelia and mesenchyme still attached. At E16 SG were separated into basal and apical portions by cutting through the middle turn. At E14, E16 and E18, three biological replicates were collected containing pooled tissue from separate litters. At P1, two biological replicates were collected.

Dissociation of SGNs at E14 and E16 : Pooled SG were dissociated for 14 minutes in 0.05% Trypsin-EDTA at 37˚C, with two vigorous triturations. Trypsin was neutralized with the addition of an equal volume of FBS.

Dissociation of SGNs at E18 and P1: Pooled SG were dissociated for 25 minutes in Papain (20U/mL) at 37˚C, with triturations every 5 mins. Papain was neutralized with the addition of Ovomucoid inhibitor.

Regardless of age of dissection, dissociated cells were passed through a 20µm strainer, spun down at 300 x g and then resuspended in PBS with 0.4mg/mL BSA and 0.2U/mL RNase inhibitor at a density of 1,100 cells/ml. Single cells from the dissociated preparations were captured on the 10X Genomics Chromium Controller, followed by lysis, conversion of mRNA to cDNA, and preparation of sequencing libraries using Chromium Single Cell 3' Reagents (v3.1) and the manufacturer's

suggested protocols. Libraries were sequenced on an Illumina NextSeq and were aligned to the Ensembl mouse MM10 assembly using Cell Ranger 3.0.1 analysis software (10X Genomics).

Initial data processing: Initial 'mean reads per cell', 'median genes per cell' and 'number of reads' data from the Cell Ranger output is available in Suppl data file 8. Further processing of individual datasets was carried out in Seurat (R package, v3.2)(4). Briefly, Cell Ranger outputs were imported into Seurat using the 'Read10x' function. Genes present in fewer than 3 cells, cells with fewer than 200 unique genes identified, and cells with more than 15,000 total RNA molecules detected were all removed from the analysis. Cells with higher than 5% mitochondrial or hemoglobin transcripts (*Hba-a1*, *Hba-a2*, *Hbb-bh1*, *Hbb-bs*, *Hbb-bt*) were also removed to avoid contamination with dead/dying cells, or red blood cells. Data were normalized, scaled and dimensional reduced using the standard Seurat workflow. In order to detect doublets within the dataset we used DoubletFinder (R package, v2.0.3) (5) with standard parameters, and the predicted doublet rate based on our cell loading (5%). Doublets were removed and we then moved forward with only the predicted 'singlet' cells which constituted 95% of cells in each dataset. Each dataset was then re-normalized, scaled and dimensionally reduced using standard Seurat workflow. UMI, and mitochondrial content scores were regressed-out using Seurat's "ScaleData" function. The top 20 principal components from Seurat's "RunPCA" function were used to define the dimensions of the UMAP non-linear dimensional reduction analysis which was then viewed on a 2D UMAP plot. An unsupervised clustering analysis was also carried out using the same number of principal components. In order to identify the neuronal clusters in the dataset we used a set of established neuronal marker genes (*Tubb3*, *Snhg11*, *Nefl*, *Nefm*, *Nefh*) and genes known to be expressed early in SGN development (*Pou4f1*, *Gata3*, *Neurod1*). Although neuronal cells made up only a small proportion of the total cells (10-15%), cells within these clusters always overlapped with one another and were well separated from other clusters in the dataset. We additionally generated the top 10 markers for all clusters in the dataset and confirmed putative SGN clusters had a coherent neuronal expression pattern. SGN clusters were then isolated from each individual dataset.

Merging datasets: At each individual timepoint isolated SGN clusters were merged using Seurat's standard integration workflow to account for batch effects. For the integration of all timepoints together, SGN datasets for each timepoint were merged before carrying out the standard Seurat workflow for PCA analysis. Harmony (R Package v1.0)(6) integration was then applied to this dimensionally reduced Seurat object to account for batch effects between the individual datasets. For all merged objects, clustering analysis was re-done. Clusters were checked for quality and were removed if their identity could not be determined (< 3 significant genes defining their expression, low or no expression of neuronal marker genes, and/or top 10 marker genes matching > 1 cell type (eg: a mixture of glial and neuronal genes)). The 'nFeature_RNA' (number of genes detected) and the batch origin of each cell was generated across each UMAP to ensure the dataset was not clustering based on either of these parameters after batch correction (Suppl fig 11).

Trajectory analysis: Slingshot (R package v1.8)(7) analysis was run using UMAP as the dimensionally reduced input. Slingshot uses a minimum spanning tree method to identify lineages of cells starting at the same origin and then diverging to unique endpoints. It then calculates pseudotime for all the cells on the lineage using principal curves. For all analyses, the start cluster was identified as the cluster containing the most immature SGNs defined by high expression of *Neurod1*. In the P1/E18 UMAP, E16 and whole combined datasets UMAP end clusters (Type 1A, B, C and Type 2) could be identified based on marker gene expression. For the E14 UMAP no end clusters were defined. In order to identify genes changing across pseudotime, tradeSeq (R package v1.4)(8) analysis was carried out on the E16/E18 Type 1A/2, E16 type B/C and E14 Slingshot trajectories with a Kknots value of 7 (E16/E18) or 8 (E14). For each dataset, tests were carried out to identify genes with gene expression patterns that were associated with pseudotime (associationTest), or had different expression patterns along the two lineages (patternTest), and genes that were differentially expressed at the endpoints of the lineages (diffEndTest). Additionally, in the E16/E18 Type 1A/2 dataset, we ran a test of genes differentially changing across the two

lineages as the Type 2 branch split from the Type 1A branch (earlyDETest between knots 3 and 7).

Monocle 3 analysis: Monocle 3 (R package) (9, 10) analysis was undertaken on the E14 dataset in order to test whether the same E14 dataset clusters were identified when using an alternative approach to Seurat. Raw gene counts for the E14 SGN dataset were extracted from the Seurat object. The standard Monocle 3 workflow was then applied to this raw data including preprocessing, removal of batch effects using mutual nearest neighbor alignment (11) and dimensional reduction and UMAP visualization of the data. A clustering analysis was undertaken and the top 10 marker genes for each cluster were calculated.

SCENIC regulon analysis: SCENIC (12) analysis was carried out as described in Kolla et al. (2020)(13). Briefly, SCENIC was run on datasets in which every 5-20 cells had been randomly pooled and gene expression averaged. Activity of a regulon within a specific cluster was assessed by both their regulon specificity score (RSS) and their AUCell scores, with only regulons significant in both tests considered active.

Immunohistochemistry and smFISH: Cochleae were collected at E14, E16, E18 and P1 and fixed in 4% PFA for 2 hours (immunohistochemistry) or overnight (smFISH). For sections, tissue was collected from either CD1 or $MAPT^{EGFP}$ animals (where quantification was required), and cryoprotected through a gradient of sucrose (5%, 10%, 15%, 20%, 25% and 30%). Samples were embedded in Tissue-Tek O.C.T compound and sectioned at 12 µm. For whole mounts, cochleae were dissected to expose the sensory epithelium and SG.

Immunohistochemistry: Sections and whole mounts were blocked for 1 hour in 10% normal donkey serum diluted in 0.5% triton-X in PBS. If the primary antibody was raised in mouse, then the blocking solution would also include 0.5% Unconjugated AffiniPure FAB fragment goat Anti-mouse

IgG (H L) (Jackson ImmunoResearchLabs). Primary antibody was added at the appreciate concentration (Suppl. Table 1) and incubated overnight. Tissue was then washed 3 times in PBS followed by incubation with AlexaFluor-conjugated secondary antibodies (1:500) diluted in blocking solution for 2 hours. Tissue was washed and cover slipped/mounted using Fluoromount-G. Antigen retrieval was carried out on POU4F1 and ESPN antibodies prior to initial blocking step by incubating slides in near boiling citrate buffer (10mM citic acid, 0.05% Tween20, pH6) for 5 minutes.

smFISH: For smFISH, gene specific probes were ordered from Advanced Cell Diagnostics (Suppl. Table 2) and the RNAscope® Fluorescent Multiplex Reagent Kit (320851) was run following manufacturer's instructions on sections. For slides where smFISH was being quantified an additional immunohistochemistry step labelling for GFP was added to the end of the protocol, using the standard immunohistochemistry protocol described above. For whole mounts RNAscope was run using a previously published protocol(14) followed by immunohistochemistry staining for ZSGreen using the standard protocol described above.

Cell count data analysis: For all cell counts, sections were imaged with identical confocal microscopy settings across all slides. Negative control slides were used to threshold out background in both immunohistochemistry and RNAScope experiments. Images were imported into ImageJ (15) software and using only the neuronal marker and DAPI signal cells were hand counted as a neuron if they colocalized both the marker and DAPI. Each identified neuron was then counted for presence or absence of additional markers via presence of signal above background levels. For RNAScope this was determined as >3 punta per cell using the negative control probe provided in the kit.

TH cell counts:

Whole mount: Whole mount TH immunohistochemistry was combined with genetic sparse labelling using *Neurog1*^{CreERT2}*R26R*$^{mT/mG}$ to quantify the proportion of TH expressing SGNs that were Type

1 or Type 2. TH+ neurons were identified using TH immunostaining of their cell bodies. The peripheral process of each TH+ neuron was then traced to its terminus either at the level of the inner hair cells (Type 1) or outer hair cells (Type 2) based on GFP labeling. Counts represent 20 TH+ neurons counted from pups across 4 litters.

Sections: The proportion of SGNs expressing TH across development was assessed by counting the total number of TH+ and TH- SGNs in a modiolar section through the cochlea at each time point using Neurofilament (NF) and TH immunolabelling. Counts were combined totals across all three turns of the SG for 5 pups from separate litters. Data presented as mean ± SEM. Statistical test is one-way ANOVA followed by Tukey's multiple comparison's test.

E18 SGN subgroup marker quantification: The proportion of SGNs expressing either Calb2/Lypd1 RNA, *Tle4/Lypd1* RNA or CALB1/CALB2 protein at E18 were assessed by counting the number of positive SGNs for each marker against the total number of SGNs labelled with either GFP (using $MAPT^{EGFP}$ tissue) or TUJ1 (using wildtype CD1 tissue). The proportion of POU4F1+ SGNs expressing ESPN was counted similarly, with the total number of POU4F1+/ESPN+ SGNs against the total number of POU4F1+ SGNS. Counts were taken from the basal turn in a mid-modiolar section from the middle of the confocal image stack. Counts were combined totals for 3 pups from separate litters. Data presented as mean ± SEM.

Data availability: Single-cell gene expression data have been deposited in the Gene Expression Omnibus data repository under accession code: GSE195500. Gene by cell expression matrix and data visualizations presented in this paper will be available through the gEAR Portal.

1. E. J. Koundakjian, J. L. Appler, L. V. Goodrich, Auditory neurons make stereotyped wiring decisions before maturation of their targets. *Journal of Neuroscience* **27**, 14078-14088 (2007).
2. J. A. Harris *et al.*, Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Frontiers in Neural Circuits* **8** (2014).
3. K. L. Tucker, M. Meyer, Y.-A. Barde, Neurotrophins are required for nerve growth during development. *Nature neuroscience* **4**, 29-37 (2001).
4. T. Stuart *et al.*, Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902. e1821 (2019).
5. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell systems* **8**, 329-337. e324 (2019).
6. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods* **16**, 1289-1296 (2019).
7. K. Street *et al.*, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics* **19**, 1-16 (2018).
8. K. Van den Berge *et al.*, Trajectory-based differential expression analysis for single-cell sequencing data. *Nature communications* **11**, 1-13 (2020).
9. J. Cao *et al.*, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
10. X. Qiu *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* **14**, 979-982 (2017).
11. L. Haghverdi, A. T. Lun, M. D. Morgan, J. C. Marioni, Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature biotechnology* **36**, 421-427 (2018).
12. S. Aibar *et al.*, SCENIC: single-cell regulatory network inference and clustering. *Nature methods* **14**, 1083-1086 (2017).
13. L. Kolla *et al.*, Characterization of the development of the mouse cochlear epithelium at the single cell level. *Nature communications* **11**, 1-16 (2020).
14. J. Kersigo *et al.*, A RNAscope whole mount approach that can be combined with immunofluorescence to quantify differential distribution of mRNA. *Cell and tissue research* **374**, 251-262 (2018).
15. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).

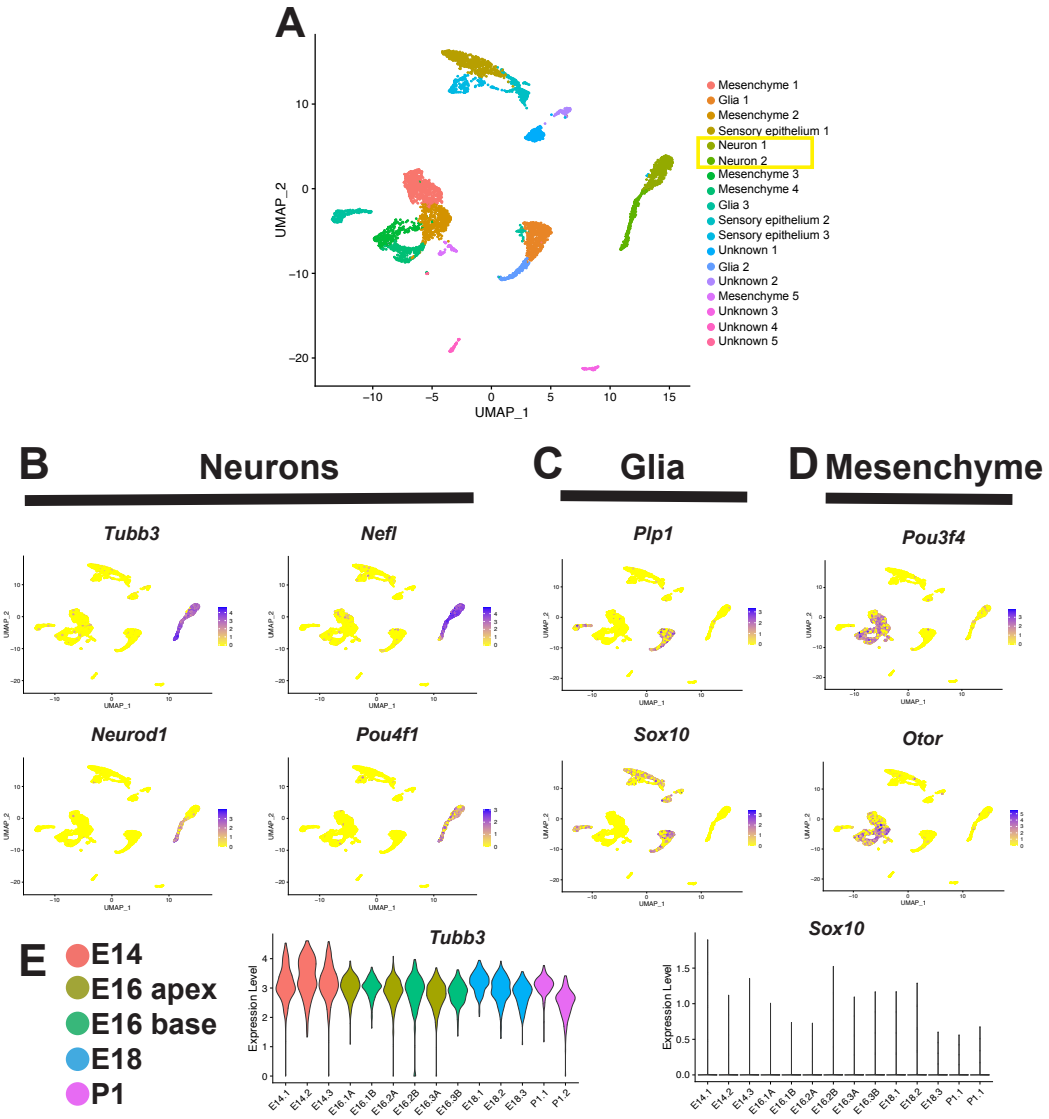


Fig. S1.

Identification of developing SGNs. **A.** UMAP displaying results from a single cell collection at E14 that includes 5,225 cells after initial QC and doublet removal. Legend indicates identities of clusters containing SGNs (yellow box), glia, mesenchyme, and sensory epithelium. Remaining cell clusters have not been identified. **B.** Feature plots of the same projection as in **A** displaying the expression of specific markers for SGNs (**B**), glia (**C**), and mesenchyme (**D**). **E.** Violin plots displaying the expression of the neuronal marker *Tubb3*, and the glial marker *Sox10* across the identified SGN clusters from each collection. The high expression of *Tubb3* and negligible expression of *Sox10* indicates the SGN clusters are free of glial contamination.

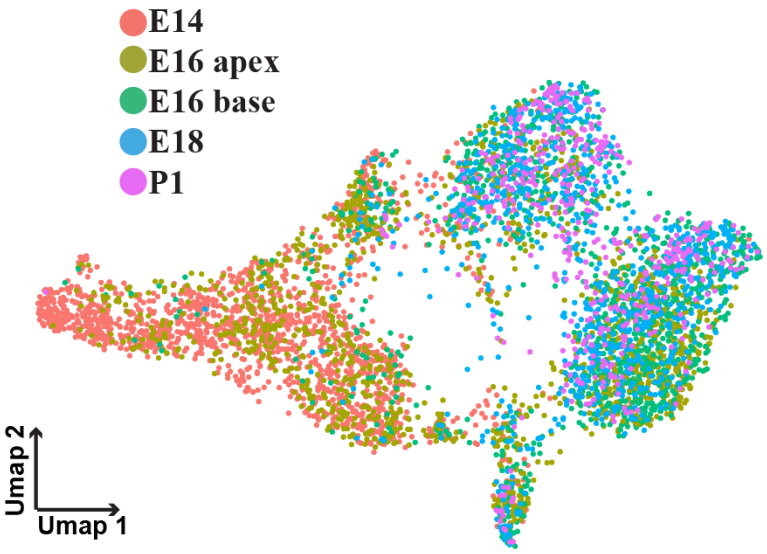


Fig. S2.

Combined SGN dataset. UMAP plot of 5,441 SGNs collected at E14, E16, E18 and P1 colored by the timepoint. Note that cells from each timepoint form a gradient from most immature to most mature along the UMAP 1 axis.

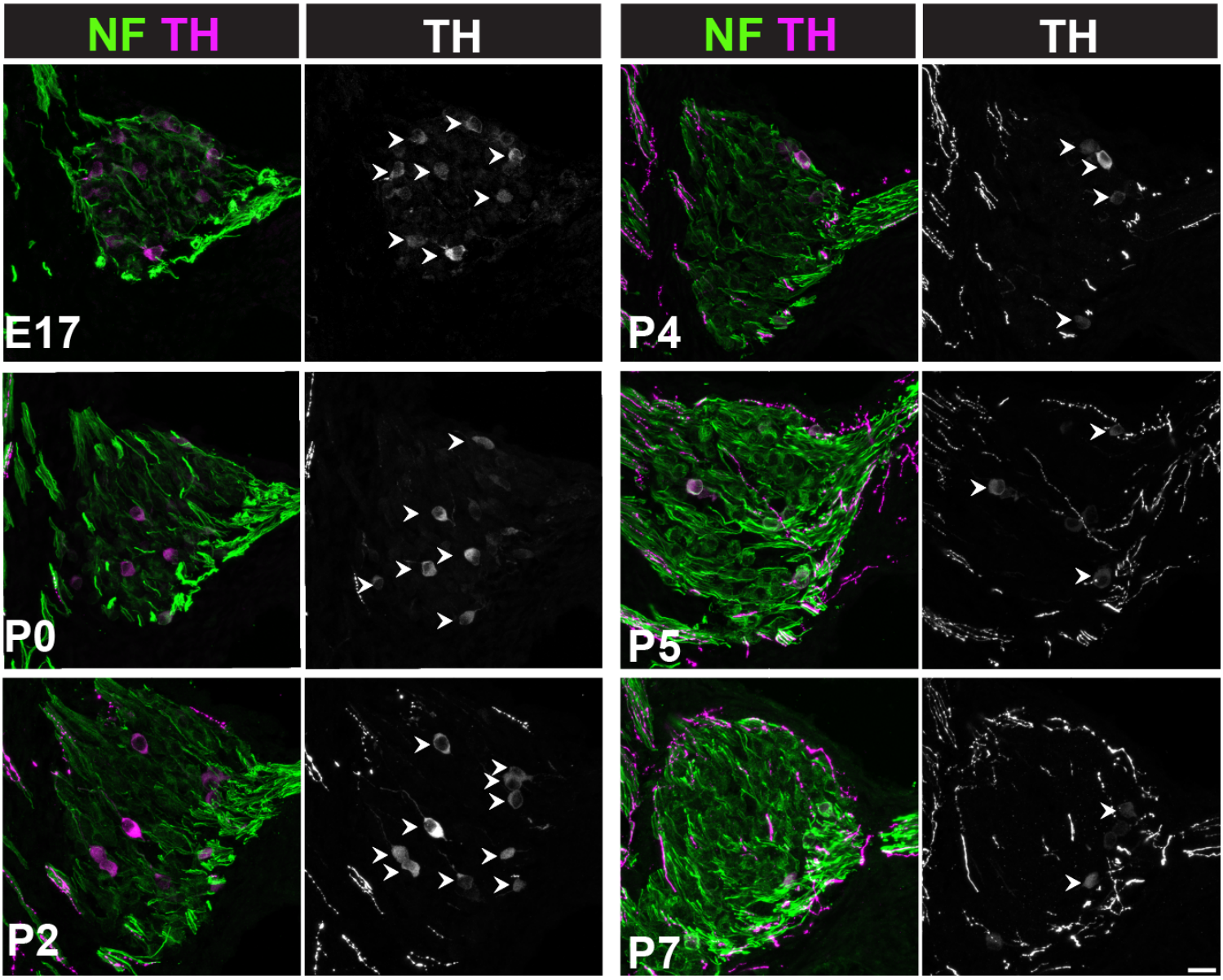


Fig. S3.
Timecourse of TH expression in the perinatal SG. Images of TH and Neurofilament immunolabeling in the mid SG from E17-P7. The number of TH positive SGNs drops after P4. Arrowheads point to TH positive cell bodies. Scale bar, 20µm.

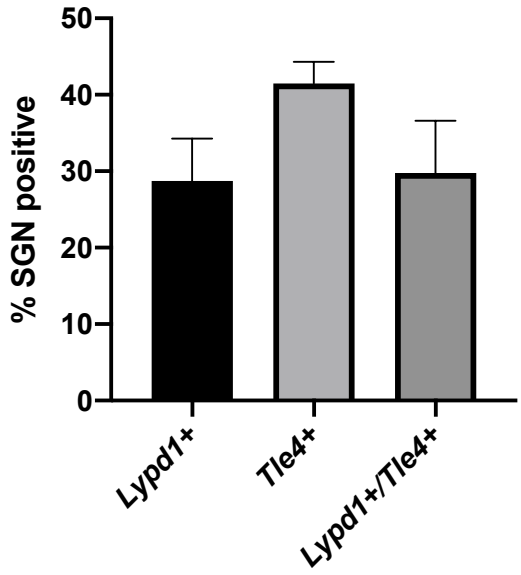


Fig. S4.

Proportion of Tle4+ and Lypd1+ SGNs at E18. Quantification of proportion of SGNs expressing each marker at E18. About 30% of SGNs express both Lypd1 and Tle4 to some degree while the remaining 70% express either Lypd1 only or Tle4 only. Data presented as mean ± SEM.

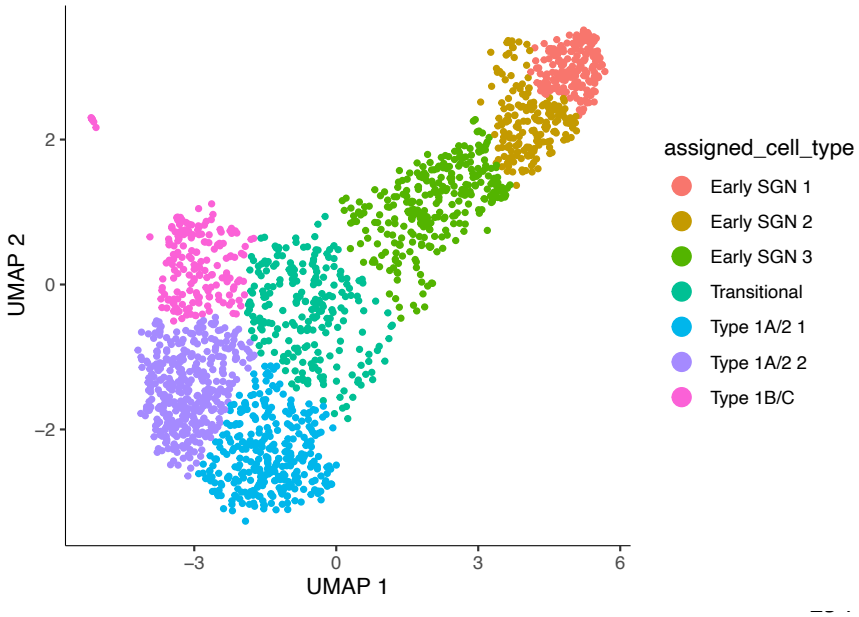
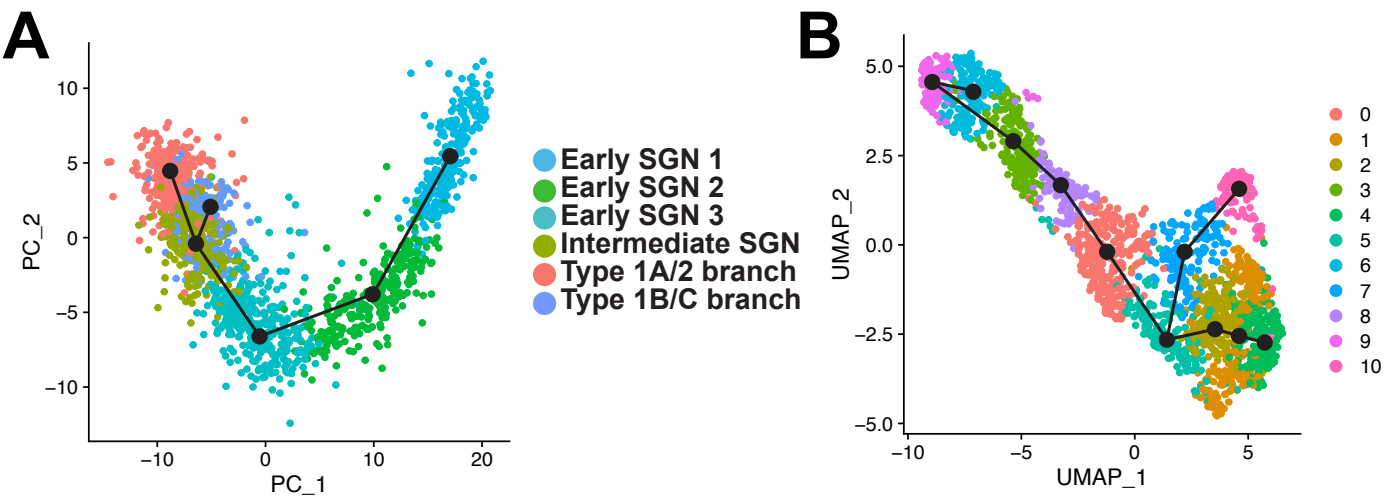


Fig. S5.

Monocle analysis at E14. Analysis of E14 dataset using Monocle reveals equivalent early SGN clusters.



299

300 **Fig. S6**

301 **Slingshot analysis at E14.** Modulation of Slingshot by (A) using PCA as an input, or (B)
302 over clustering the dataset, does not alter the finding of an E14 split between Type 1A/2
303 and Type 1B/C precursors.

304

305

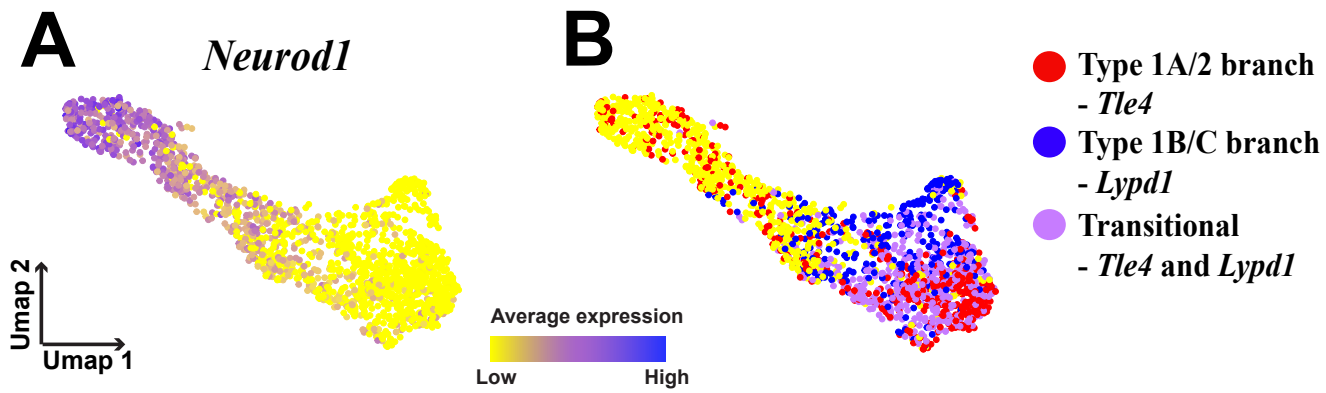


Fig. S7.
Gene expression defines clusters at E14. **a.** Feature plot of the E14 dataset showing *Neurod1* expression is highest in the Early SGN 1 cluster. **b.** Feature plot of the E14 dataset displaying binarized expression of *Tle4* and *Lypd1* in two spatially separated clusters in the more mature neurons in the dataset. Some neurons in a transitional phase are expressing both markers.

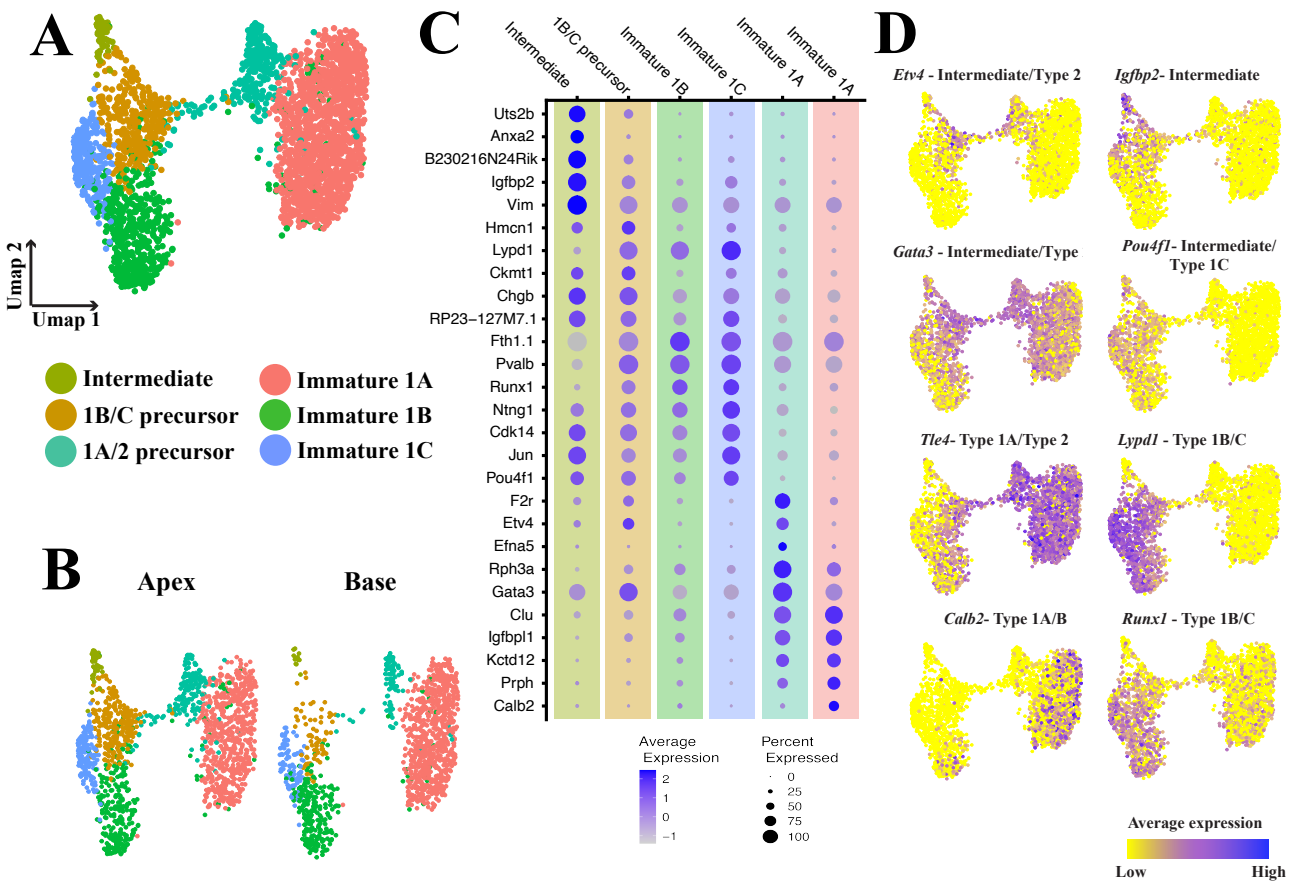


Fig. S8.

Characterization of the E16 dataset. **A**. UMAP of 2,702 SGNs collected from the base and apical portions of the SG with cluster identities indicated. **B**. UMAP in (**A**) split by SGNs originating in the apex and base. **c.** Dot plot showing the top 5 differentially expressed markers for each of the clusters identified in (**A**). **D.** Feature plots of gene markers for the clusters identified in (**A**).

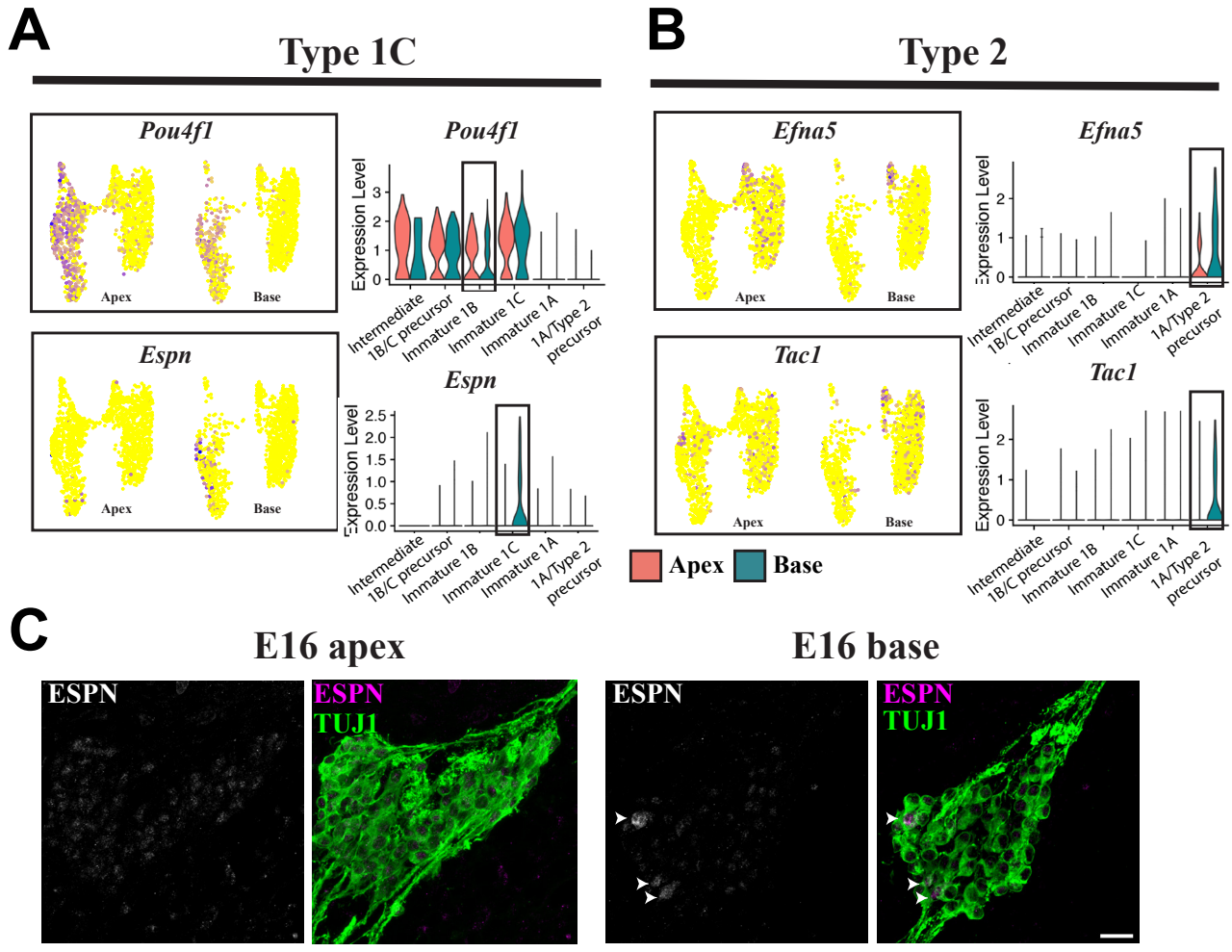


Fig. S9.

Type 1C and Type 2 SGNs are identifiable at E16. **A-B**. Feature plots (left) and violin plots (right) of E16 dataset from **Suppl. Fig 7** split by apex and base, displaying Type 1C **(A)** and Type 2 **(B)** marker gene expression. Boxes on violin plots indicate *Pou4f1* expression decreasing in the 1B cluster in the more mature basal neurons, and the other 3 genes all turning on in the basal neurons in either the Type 1C or Type 2 clusters. **C** immunohistochemistry confirms Type 1 C marker ESPN turns on expression in basal SGNs. Scale bar in all images indicates 20µm.

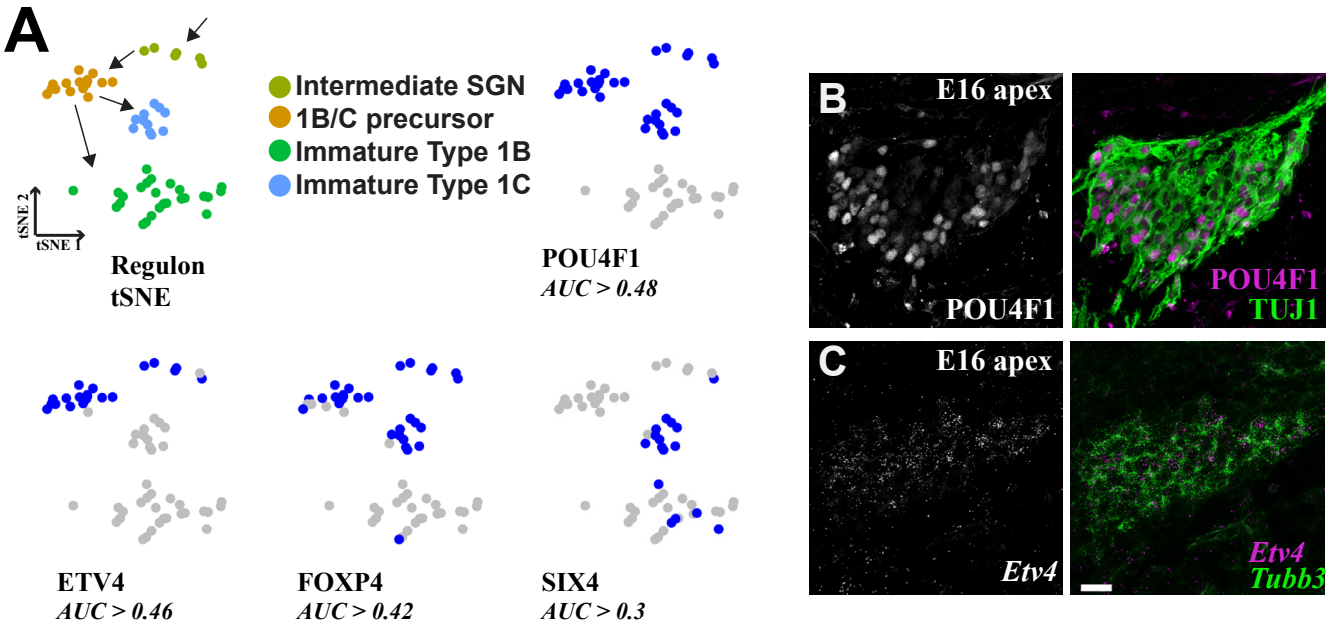


Fig. S10.

Regulon analysis in E16 1B/C dataset. A Top row: tSNE plot based on identification of transcriptional regulons using SCENIC. Clusters are colored based on their identification in **fig 7A**. Arrows indicate order of specification based on Slingshot analysis. Regulon tSNE plot of POU4F1 shows activity in early and 1C SGN clusters, blue dots indicate the regulon is active. Bottom Row: examples of other regulons which were both significantly restricted to specific clusters, and were predicted to be active based AUC thresholding (see methods). **B.** immunohistochemistry confirms POU4F1 expression in a subset of SGNs at E16. **C**. smFISH confirms *Etv4* expression in the SG at E16. Scale bar in **C**. (same as **B**.) indicates 20µm.

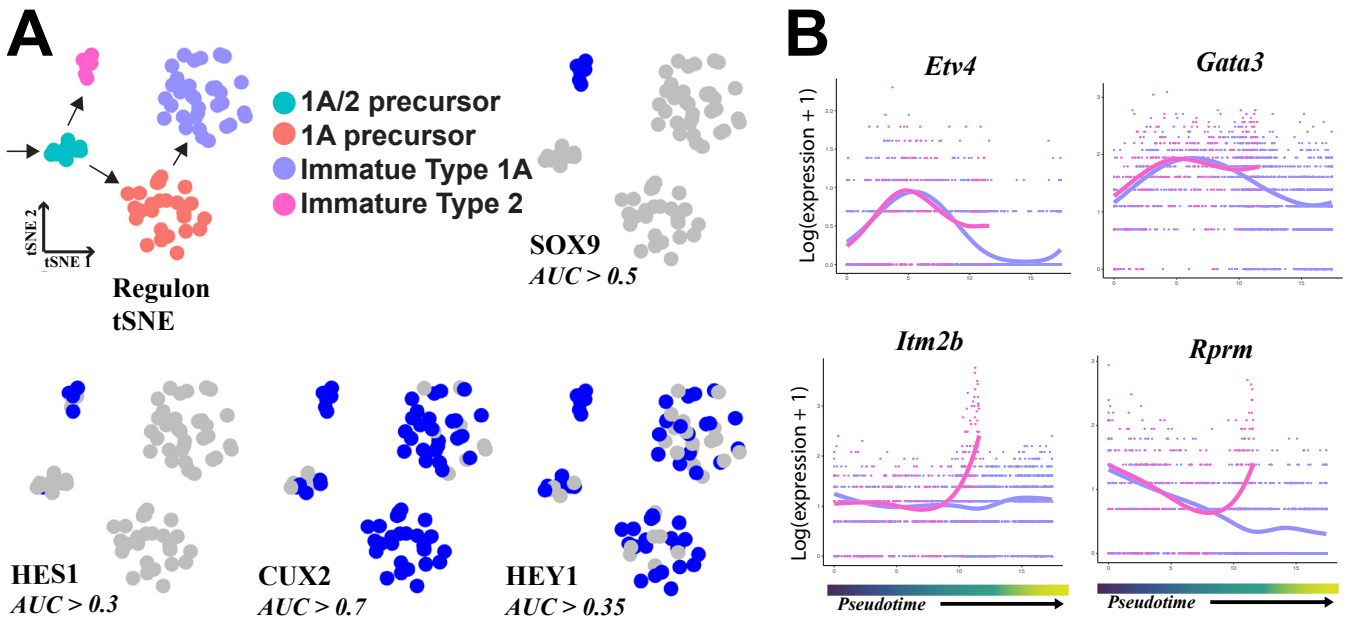


Fig. S11.

Regulon and tradeSeq analysis in E16/18 1A/2 dataset. a Top row: tSNE plot based on identification of transcriptional regulons using SCENIC. Clusters are colored based on their identification in **fig 7D**. Arrows indicate order of specification based on Slingshot analysis. Regulon tSNE plot of SOX9 shows specific activity in Type 2 cluster, blue dots indicate the regulon is active. Bottom Row: examples of other regulons which were both significantly restricted to specific clusters, and were predicted to be active based AUC thresholding (see methods). **B**. Plots of the expression of genes identified by tradeSeq analysis as differentially expressed along the two trajectories (top row) and as the two lineages split (bottom row).

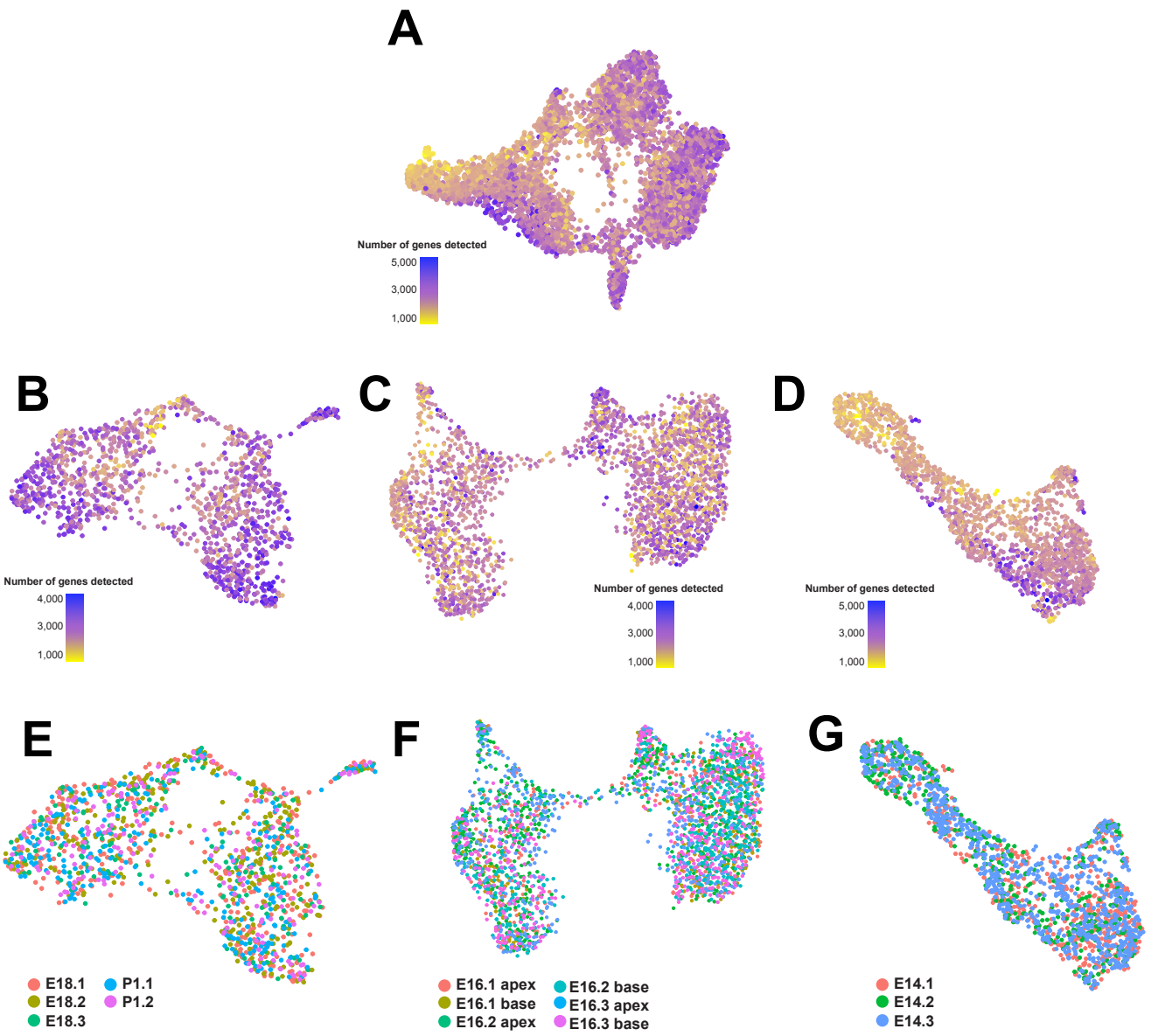


Fig. S12.
Dataset do not cluster based on gene detection rate and batch. UMAPs displaying the number of genes detected per cell and the batch identity of each cell in the whole combined dataset (**A**), E18/P1 combined dataset (**B,E**), E16 dataset (**C,F**), and E14 dataset (**D,G**).

384
385

386 **Table S1.**

387 **List of antibodies used for immunohistochemistry**

Antibody	Concentration	Source	Cat#
Rabbit anti-CALB1	1:200	Millipore	AB1778
Rabbit anti-TUJ1	1:1000	Sigma-Aldrich	T2200
Mouse anti-CALB2	1:500	Millipore/Chemicon	MAB1568
Rabbit anti-ESPN	1:200	Invitrogen	PA5-55941
Mouse anti-POU4F1	1:100	Millipore	MAB1585
Rabbit anti-ZSGreen	1:500	Takara Bio, USA	632474
Chicken anti-GFP	1:500	GeneTex	GTX26662
Rabbit anti-TH	1:500	Millipore	AB152
Chicken anti-Neurofilament L,M,H	1:1000	Aves Labs	NFL, NFM, NFH

388

Table S2.

List of probes used for single molecule fluorescent *in situ* hybridization

Probe	**Source**	**Catalog #**
Mm-Calb2	ACDBio	313641
Mm-Lypd1	ACDBio	318361
Mm-Efna5	ACDBio	316641
Mm-Tle4	ACDBio	417301
Mm-Etv4	ACDBio	458121
Mm-Tubb3	ACDBio	423391

Dataset S1. (separate file)
Description: Top 50 differentially expressed genes in each cluster for each dataset (combined, E14, E16, E18/P1, E18/P1 Type 1 vs Type 2).

Dataset S2. (separate file)
Description: Top 100 differentially expressed genes from tradeSeq analysis for the E14 dataset.

Dataset S3. (separate file)
Description: Top 100 differentially expressed genes from tradeSeq analysis for the E16 Type 1B/C lineage dataset.

Dataset S4. (separate file)
Description: Top 100 differentially expressed genes from tradeSeq analysis for the E16/E18 Type 1A/2 lineage dataset.

Dataset S5. (separate file)
Description: List of SCENIC regulons expressed for each cluster in the E14 dataset.

Dataset S6. (separate file)
Description: List of SCENIC regulons expressed for each cluster in the E16 dataset.

Dataset S7. (separate file)
Description: List of SCENIC regulons expressed for each cluster in the E16/E18 dataset

Dataset S8. (separate file)
Description: Initial 'mean reads per cell', 'median genes per cell' and 'number of reads' sequencing data from the Cell Ranger output