Supplementary Materials for

Single cell profiling of the developing mouse brain and spinal cord with split-pool barcoding

Alexander B. Rosenberg[1†*], Charles M. Roco[2†], Richard A. Muscat[1], Anna Kuchina[1], Paul Sample[1], Zizhen Yao[3], Lucas Gray[3], David J. Peeler[2], Sumit Mukherjee[1], Wei Chen[4], Suzie H. Pun[2], Drew L. Sellers[2,5], Bosiljka Tasic[3], Georg Seelig[1,4,6*]

Correspondence to: alex.b.rosenberg@gmail.com, gseelig@uw.edu

This PDF file includes:

Other Supplementary Materials for this manuscript includes the following:

Materials and Methods

Experimental Methods

Oligonucleotides

The sequences and modifications for all oligonucleotides used can be found in Table S1.

Cell Culture

HEK293 and Hela-S3 cells were cultured in DMEM + 10% FBS, while NIH/3T3 cells were cultured in DMEM + 10% FCS. Cells were rinsed twice with 1x PBS, then detached by incubating 2-5 min at room temperature with 1ml of TrypLE. Once cells were detached, they were added to 2mL of media with 10% FBS. In mouse-human species mixing experiments, cells were combined at the desired concentrations at this step.

Fixation

Cells were first centrifuged for 3 min at 500g at 4˚C. The pellet was resuspended in 1mL of cold PBS-RI, 1x PBS + 0.05U/μL RNase Inhibitor (Enzymatics). The cells were then passed through a 40 μm strainer into a 15 mL falcon tube. 3 mL of cold 1.33% formaldehyde solution (in 1x PBS) was then added to 1 mL of cells. Cells were fixed for 10 min before adding 160 μL of 5% Triton X-100. Cells were then permeabilized for 3 min and centrifuged at 500g for 3 min at 4˚C. Cells were resuspended in 500uL of PBS-RI before adding 500 μL of cold 100 mM Tris-HCL pH 8. In order to make the cells easier to pellet, 20 μL of 5% Triton-X100 was added. Then, cells were spun down at 500g for 3 min at 4˚C and resuspended in 300 μL of cold 0.5 X PBS-RI. Finally, cells were again passed through a 40 μm strainer into a new 1.5 mL tube. Cells were then counted on a hemocytometer or flow-cytometer and diluted to 1,000,000 cells/mL.

In-cell Reverse Transcription

The first round of barcoding occurs through an *in situ* reverse transcription (RT) reaction. Cells are split into up to 48 wells, each containing barcoded well-specific reverse transcription primers. Both random hexamer and anchored poly(dT)$_{15}$ barcoded RT primers were used in each well (Table S1). For each well, we added 4 μL of 5X RT Buffer, 0.625 μL of RNase-free water, 0.125 μL RNase Inhibitor (Enzymatics), 0.25 μL μL SuperaseIn RNase Inhibitor (Ambion), 1 μL of 10 mM dNTPs each (ThermoFisher), 2 μL of 25 μM random hexamer barcoded RT primer, 2uL of 25 μM poly(dT)$_{15}$ barcoded RT primer, 2 μL of Maxima H Minus Reverse Transcriptase (ThermoFisher), and 8 μL of cells in 0.5X PBS-RI. The plate incubated in a thermocycler for 10 min at 50˚C before cycling for three times at 8˚C for 12s, 15˚C for 45s, 20˚C for 45s, 30˚C for 30s, 42˚C for 2 min, and 50˚C for 3 min, followed by a final step at 50˚C for 5 min. RT reactions are pooled back together into a 15 mL falcon tube. After adding 9.6 μL of 10% Triton X-100, cells were centrifuged for 3 min at 500g at 4˚C. The supernatant was removed and cells

were resuspended in 2 mL of 1X NEB buffer 3.1 with 20 μL of Enzymatics RNase Inhibitor.

Preparing Oligonucleotides for Ligations

The second and third barcoding round consist of a ligation reaction. Each round uses a different set of 96 well barcoding plates (Table S1). Ligation rounds have a universal linker strand with partial complementarity to a second strand containing the unique well-specific barcode sequence added to each well. These strands were annealed together prior to cellular barcoding to create a DNA molecule with three distinct functional domains: a 5' overhang that is complementary to the 3' overhang present on the cDNA molecule (may originate from RT primer or previous barcoding round), a unique well-specific barcode sequence, and a 3' overhang complementary to the 5' overhang present on the DNA molecule to be subsequently ligated (Fig. S1A). For the third round barcodes, the 5' overhang also contains a unique molecular identifier (UMI), a universal PCR handle, and a biotin molecule. Linker strands and barcode strands (IDT) for the ligation rounds are added to RNase-free 96 well plates to a total volume of 10 μL/well with the following concentrations: round 2 plates contain 11 μM linker strand (BC_0215) and 12 μM barcodes and round 3 plates contain 13 μM linker strand (BC_0060) and 14 μM barcodes. Strands for ligation barcoding rounds are annealed by heating plates to 95°C for 2 minutes and cooling down to 20°C at a rate of -0.1°C per second.

Blocking strands are complementary to the 3' overhang present on the DNA barcodes used during ligation barcoding rounds. Blocking occurs after well-specific barcodes have hybridized and were ligated to cDNA molecules, but before all cells are pooled back together. Blocking ensures that unbound DNA barcodes cannot mislabel cDNA in future barcoding rounds. 10 μL of blocking strand solution was added to each of the 96 wells after each round of hybridization and ligation of DNA barcodes. Blocking strand solutions were prepared at a concentration of 26.4 μM (BC_0216) for round 2 and 30.8 μM (BC_0066) for round 3. Blocking strands for the first two rounds were in a 2.5X T4 DNA Ligase buffer (NEB) while the third round was in a 150 mM EDTA solution (to terminate ligase activity). Blocking strands were incubated with cells for 30 min at 37°C with gentle shaking (50 rpm).

In-cell Ligations

A 2.04 mL ligation mix was made containing 1,287.5 μL of RNase-free water, 500 μL 10X T4 Ligase buffer (NEB), 100 μL T4 DNA Ligase (400 U/μL, NEB), 40 μL RNase inhibitor (40 U/μL, Enzymatics), 12.5 μL SuperaseIn RNase Inhibitor (20 U/μL, Ambion), and 100 μL of 5% Triton-X100. This ligation mix and the 2 mL of cells in 1X NEB buffer 3.1were added to a basin and mixed thoroughly to make a total of 4.04 mL.

Using a multichannel pipet, 40 μL of cells in ligation mix were added to each of the 96 wells in the first-round barcoding plate. Each well already contained 10 μL of the appropriate DNA barcodes. The round 2 barcoding plate was incubated for 30 min at

37°C with gentle shaking (50 rpm) to allow hybridization and ligation to occur before adding blocking strands. Cells from all 96 wells were passed through a 40 μM strainer and combined into a single multichannel basin, where an additional 100 μL of T4 DNA Ligase was added. Subsequent steps in round 3 were identical to round 2, except that 50 μL of pooled cells were split and added to barcodes in round 2 (total volume of 60 μL/well). This adjustment was made to account for increased total volume during each split-pool round as well as pipetting errors.

Lysis and Sublibrary Generation

After the third round of barcoding, 70 μL of 10% Triton-X100 is added to the cell solution before spinning it down for 5 min at 1000G and 4˚C. We carefully aspirated the supernatant, leaving about 30 μL to avoid removing the pellet. We then resuspended the cells in 4 mL of wash buffer (4 mL of 1X PBS, 40 μL of 10% Triton X-100 and 10 μL of SUPERase In RNase Inhibitor) and spun down for 5 min at 1000G at 4˚C. We then aspirated the supernatant and resuspended in 50 μL of PBS-RI. After counting cells, we aliquoted them into sublibraries (in 1.7 mL tubes). The number of sublibraries generated will determine how many splits are made for the fourth round of barcoding. After adding the desired number of cells to each sublibrary, we brought the volume of each to 50 μL by adding 1x PBS, then added 50 μL of 2X lysis buffer (20 mM Tris (pH 8.0), 400 mM NaCl, 100 mM EDTA (pH 8.0), and 4.4% SDS) and 10 μL of proteinase K solution (20mg/mL). We incubated cells at 55˚C for 2 hours with shaking at 200 rpm to reverse formaldehyde crosslinks. Afterwards, we froze lysates at -80°C.

Purification of cDNA

We first prepared 40 μL Dynabeads MyOne Streptavidin C1 beads (ThermoFisher) per sublibrary by washing them 3x with 800 μL of 1X B&W buffer with 0.05% Tween-20 (refer to manufacturer's protocol for B&W buffer), before resuspending beads in 100 μL 2X B&W buffer (with 2 μL of SUPERase In Rnase Inhibitor) per sample.

To inhibit residual proteinase K activity, we added 5 μL of 100 μM PMSF to each thawed lysate and incubated at room temperature for 10 minutes. We then added 100 μL of resuspended Dynabeads MyOne Streptavidin C1 (ThermoFisher) magnetic beads to each lysate. We then allowed binding to occur for 60 min at room temperature (with agitation on a microtube foam insert). The beads were washed twice with 1X B&W buffer and once more with 10mM Tris containing 0.1% Tween-20 (with each wash including of 5 min of agitation after resuspension of beads).

Template Switch

Streptavidin beads with bound cDNA molecules were resuspended in a solution containing 44 μL of 5X Maxima RT buffer (ThermoFisher), 44 μL of 20% Ficoll PM-400 solution, 22 μL of 10 mM dNTPs each (ThermoFisher), 5.5 μL of RNase Inhibitor (Enzymatics), 11 μL of Maxima H Minus Reverse Transcriptase (ThermoFisher), and 5.5 μL of 100uM of a template switch primer (BC_0127). The template switch primer

contains two ribonucleic guanines followed by a locked nucleic acid guanine at the end of the primer (Exiquon). The beads were incubated at room temperature for 30 minutes and then at 42°C for 90 minutes with gentle shaking.

PCR

After washing beads once with 10 mM Tris and 0.1% Tween-20 solution and once with water, beads were resuspended into a solution containing 110 μL of 2X Kapa HiFi HotStart Master Mix (Kapa Biosystems), 8.8 μL of 10 μM stocks of primers BC_0062 and BC_0108, and 92.4 μL of water. PCR thermocycling was performed as follows: 95°C for 3 mins, then five cycles at 98°C for 20 seconds, 65°C for 45 seconds, 72°C for 3 minutes. After these five cycles, Dynabeads beads were removed from PCR solution and EvaGreen (Biotium) was added at a 1X concentration. Samples were again placed in a qPCR machine with the following thermocycling conditions: 95°C for 3 minutes, cycling at 98°C for 20 seconds, 65°C for 20 seconds, and then 72°C for 3 minutes, followed by a single 5 minutes at 72°C after cycling. Once the qPCR signal began to plateau, reactions were removed.

Tagmentation

PCR reactions were purified using a 0.8X ratio of SPRI beads (Kapa Pure Beads, Kapa Biosystems) and cDNA concentration was measured using a qubit. For tagmentation, a Nextera XT Library Prep Kit was used (Illumina). 600 pg of purified cDNA was diluted in water to a total volume of 5 μL. 10 μL of Nextera TD buffer and 5 μL of Amplicon Tagment enzyme were added to bring the total volume to 20 μL. After mixing by pipetting, the solution was incubated at 55°C for 5 minutes. A volume of 5 μL of neutralization buffer was added and the solution was mixed before incubation at room temperature for another 5 minutes. In this order, a 15 μL volume of Nextera PCR mix, 8 μL of water, and 1 μL of each primer (P5 primer: BC_0118, one indexed P7 primer: BC_0076-BC_0083) at a stock concentration of 10 μM was added to the mix, making a total volume of 50 μL. Using distinct, indexed PCR primers, this PCR reaction can be used to add a unique barcode to each sublibrary barcoded. PCR was then performed with the following cycling conditions: 95°C for 30 seconds, followed by 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds, and 72°C for 5 minutes after the 12 cycles. 40 μL of this PCR reaction was removed and purified with a 0.7X ratio of SPRI beads to generate an Illumina-compatible sequencing library.

Illumina Sequencing

Libraries were sequenced on MiSeq or NextSeq systems (Illumina) using 150 nucleotide (nt) kits and paired-end sequencing. Read 1 (66 nt) covered the transcript sequences. Read 2 (94 nt) covered the UMI and UBC barcode combinations. The index read (6 nt), serving as the fourth barcode, covered the sublibrary indices introduced after tagmentation.

Mouse Brain and Spine Nuclei Extraction

Brain and spinal cord tissue was harvested from two postnatal mouse pups (P2 and P11) that had been exsanguinated by transcardial saline perfusion. The mouse strain used was C57BL/6 x DBA/2. All animal procedures were done using protocols approved by the Institutional Animal Care and Use Committee at the University of Washington.

Nuclei extraction protocol was adapted from Krishnaswami *et al* (*42*). Briefly, a NIM1 buffer was made consisting of 250 mM sucrose, 25 mM KCl, 5 mM $MgCl_2$, and 10 mM Tris (pH=8.0). A homogenization buffer was made consisting of 4.845 mL of NIM1 buffer, 5 μL of 1 mM DTT, 50 μL of Enzymatics RNase Inibitor (40U/μL), 50 μL of SuperaseIn RNase Inhibtor (20U/μL), and 50 μL of 10% Triton-X100.

A 1 mL dounce homogenizer (Wheaton, cat. no. 357538) was used for nuclei extraction. After adding mouse brain and spinal cord tissue, 700 μL of homogenization buffer was added to the douncer. Then 5 strokes of loose pestle followed by 10-15 strokes of tight pestle were performed. Homogenization buffer was added up to a volume of 1 mL. The homogenate was filtered with a 40um strainer into 5mL Eppendorf tubes and then spun down for 4 minutes at 600g at 4˚C. After removing supernatant, the pellet was resuspended in 1 mL of 1X PBS-RI. Then 10 μL of BSA was added and solution was spun down again for 4 min at 600g at 4˚C. Nuclei were then passed through a 40 μm strainer once more before being counted.

Computational Methods

Alignment and generation of cell-gene matrices

To simplify analysis, we first removed any dephased reads in our library (last 6 bases of read did not match the expected sequence). Reads were then filtered based on quality score in the UMI region. Any read with >1 low-quality base (phred <=10) were discarded. Reads with more than one mismatch in any of the three 8 nt cell barcodes were also discarded. The cDNA reads (Read 1) were then mapped to either a combined mm10-hg19 genome or the mm10 genome using STAR (*44*). The aligned reads in the resulting bam file were mapped to exons and genes using TagReadWithGeneExon from the drop-seq tools (*4*). We only considered the primary alignments. Reads that mapped to a gene, but no exon, were considered intronic. Reads mapping to no gene were considered intergenic. We then used Starcode (*45*) to collapse UMIs of aligned reads that were within 1 nt mismatch of another UMI, assuming the two aligned reads were also from the same UBC. Each original barcoded cDNA molecule is amplified before tagmentation and subsequent PCR, so a single UMI-UBC combination can have several distinct cDNA reads corresponding to different parts of the transcript. Occasionally STAR will map these different reads to different genes. As a result, we chose the most frequently assigned gene as the mapping for the given UMI-UBC combination. We then generated a matrix of gene counts for each cell (N x K matrix, with N cells and K genes). For each gene, both intronic and exonic UMI counts were used.

Selecting high quality transcriptomes from the mouse CNS experiment

We discarded any transcriptomes with >1% reads mapping to mt-RNA, to ensure that all of our transcriptomes originated from nuclei. Transcriptomes with fewer than 250 expressed genes or greater than 2,500 expressed genes were also discarded. This resulted in retention of 163,069 transcriptomes. After clustering (see below), cells in putative doublet clusters were filtered as well, yielding 156,049 transcriptomes used for downstream analysis.

Hierarchical clustering of nuclei from the mouse CNS

Cells that passed the QC were clustered using an iterative clustering pipeline described in previous studies (*7*, *46*), with adaption for sparse datasets with large numbers of cells. Briefly, cells were clustered in an iterative top down approach. Each clustering iteration consists of three key steps: high variance gene selection, dimensionality reduction, and clustering. To choose high variance genes, we first fitted a loess regression curve between average scaled gene counts and dispersion (variance divided by mean). The regression residuals are then fitted by a normal distribution based on 25% and 75% quantiles to calculate p-values and adjusted p-value. High variance genes with adjusted p-values smaller than 0.1 were used to compute principle components. The proportion of variance for all PCs were converted to Z-values, and PCs with Z-values greater than two were selected for clustering. The Jaccard-Louvain algorithm (*47*) was then used for clustering, which first computes the k-nearest-neighbors (k=15) for each cell based on reduced PCs, then constructs the cell-cell similarity matrix with the Jaccard index based on the number of shared neighbors between every cell, and finally performs clustering using the Louvain algorithm. To make sure the resulting clusters all had distinguishable transcriptomic signatures, we calculated the differentially expressed genes (DEG) for every pair of clusters. A pair of clusters are separable if the deScore, defined as sum of -log10 (adjusted p-value) for all DEG (fold change >2 and adjusted p-value < 0.01, present in >40% cells in foreground, and foreground vs background enrichment ratio is greater than 3.3) is greater than 150 (every gene can contribute maximal score of 20). We merged the nearest pair of clusters that did not pass the above criterion iteratively until all clusters were separable.

We then iteratively applied the same steps described above to each cluster identified from the first clustering iteration. This iterative process was repeated until no further partitions were found. It was possible that clusters derived from different parent clusters could be similar to each other. Therefore, we computed pairwise DEG again, but reduced the threshold deScore threshold to 80 to prevent over agglomeration. This resulted in 98 clusters.

Clusters consisting of less than 40 cells were discarded (5 clusters consisting of 141 total cells).

Putative doublet clusters (clusters in which many transcriptomes were generated from doublets between two different cell types) were identified by searching for co-expression of known markers of different cell types (*e.g.* the neuronal marker *Meg3* (*48*) and the

oligodendrocyte marker *Mbp* (*9*)). This resulted in the identification of 12 clusters that were likely generated from doublet transcriptomes. These 12 clusters, consisting of 6,878 cells, were then discarded from further analysis.

Finally, we applied a more stringent test of differential expression between clusters. Using previously described criteria (*4*), we merged pairs of clusters with less than 10 differentially expressed genes (>1 natural log difference between clusters and expressed in >20% of cells in one of the two clusters). This procedure resulted in 8 clusters merging into other clusters, yielding 73 final clusters (156,049 nuclei).

PCA and t-distributed Stochastic Neighbor Embedding

We first normalized the matrix of UMI counts. For each cell, we divided the UMI counts by the total number of UMIs per cell. We subtracted the mean from each gene and then divided by the standard deviation of each gene.

We selected a subset of genes on which to perform PCA (for the mouse CNS analysis we selected genes with at least 4 UMI counts in 10 or more transcriptomes). PCA was performed on the normalized matrix using TruncatedSVD from the python package scikit-learn (*49*). For the mouse CNS analysis, we retained the first 100 PCA components and then performed t-distributed Stochastic Neighbor Embedding using a Matlab implementation of the barnes-hut t-SNE algorithm (*50*).

Lineage analysis

While previous work has identified developmental trajectories using scRNA-seq, this has mostly been confined to *in vitro* differentiation experiments. Finding differentiation trajectories in our dataset first required grouping clusters together that might form a putative lineage. To do this we followed the following procedure:

1. *Find clusters near one another in the original t-SNE embedding (Fig. 2A) that seem to form elongated structures.* In a developmental lineage, we expect cells at the start of the lineage to have substantially different transcriptomes than those at the end of the lineage, leading to this "stretching" of the lineage in the t-SNE, with intermediate cells connecting these early and late cells.
2. *Confirm that transcriptomes from the P2 mouse and P11 mouse segregate towards opposite ends of the putative lineage in the t-SNE embedding (Fig. S6).* This gives us additional confidence that the primary variance in gene expression across the putative lineage does in fact correspond to developmental stages rather than some other factor (*e.g.* regional origins).
3. *Re-embed the clusters in the putative lineage with PCA and t-SNE.* For each putative lineage, we redid PCA and t-SNE with just transcriptomes in the clusters forming the putative lineage. We did this to ensure the ordering of transcriptomes was driven only by expression of relevant genes (*e.g.* so neuronal PCA components do not drive oligodendrocyte ordering). This analysis resulted in the vast majority of re-embeded transcriptomes forming one connected group, with a small number of cells forming other distinct, new clusters. We used the density-

based DBSCAN algorithm (*51*) to identify this main lineage and to discard the transcriptomes from these smaller clusters.

4. *Measure gene expression along pseudotime curve.* Each of the lineages we analyzed in this work again formed elongated connected structures when we re-embedded transcriptomes using t-SNE. We then determined the order of transcriptomes through the t-SNE, by projecting them onto a manually drawn curve spanning the entire embedding. The moving average of gene expression was then calculated across these ordered transcriptomes.
5. *Confirm the identity of lineages using known gene markers from literature.* For each lineage, we further validated it using previously characterized marker genes from literature in addition to *in situ* hybridization data from the Allen Brain Institute (*25*).

Generating composite ISH maps for each cluster

We downloaded *in situ* hybridization (ISH) data collected by the Allen Brain Institute from the Kharchenko lab for P4 and P14 mice (*25*). The data consists of 2,187 gene measurements compiled into a 3D (P4: 50 x 43 x 77, P14: 50 x 40 x 68) map of expression "energies" corresponding to ISH staining intensities in each voxel. We used the Allen brain structure annotations to mask any voxels outside annotated brain structures. For voxels with missing data for a given gene, we set the voxel energies to the mean energy for that gene.

We then used the Allen ISH data to create a composite map of differentially expressed genes for each cluster. For each cluster, the top 5 enriched genes (with ISH data) were determined using differential gene expression described above. We normalized the expression of each gene by dividing the intensity in each voxel by the average intensity across all the voxels. To generate a composite map, we then averaged the intensity across all 5 genes within each voxel. To visualize the 3D map in a single image, we summed across sagittal slice numbers 7-24 (out of 50 slices) of the 3D map. Only slices 7-24 were used because many genes did not contain data in the other slices. Genes used to generate these maps can be found in Table S6 and S7.

Re-clustering spinal cord transcriptomes

In our original clustering, over 60% of transcriptomes from the spinal cord clustered into a large unresolved cluster (Fig. S19). Given that we had ~6x more brain nuclei relative to spinal cord nuclei, it is not surprising that the majority of PCA components selected for clustering describe variance in gene expression in the brain rather than the spine. PCA components explaining expression differences in the spinal cord may have been filtered out as "not significant" (Z-value<2).

Therefore, we reasoned that we might be able to distinguish more cell types in the spinal cord if we re-clustered only the spinal cord transcriptomes. For this clustering, we chose to use the Monocle 2 package (*52*). As suggested in the Monocle manual, we first selected high-variance genes (with high dispersion), before performing PCA. We then

used the first 50 components as input to t-SNE. Clusters were then identified using the density peak clustering option in Monocle. As previously, we removed clusters with less than 40 nuclei and then merged pairs of clusters with less than 10 differentially expressed genes (>1 natural log difference between clusters and expressed in >20% of cells in one of the two clusters). We also removed one putative doublet cluster based on co-expression of VLMC markers (*Col1a1*) (*52*) and neuronal markers (*Meg3*) (*48*). This led to the identification of 44 different of cell types in the spinal cord.

Inferring spatial origin of spinal cord nuclei

To infer the spatial origin of each spinal cord cluster, we used annotated P4 ISH maps from the Allen Spinal Cord Atlas (*41*). Each gene has been manually annotated with 11 binary values to describe different expression patterns (Laminae 1-3, Laminae 4-6, Laminae 7-8, Laminae 9, Intermediolateral Column, Gray Matter, White Matter, Central Canal, Ventral-dorsal Midline in Gray Matter, Radially Arrayed in White Matter, and Vascular-like in Gray and White Matter).

We then used these data to create a composite map of differentially expressed genes for each cluster. For each cluster, the top 10 enriched genes (with spinal cord ISH data) were determined using differential gene expression described above. We then plotted the fraction of these 10 genes with expression in each laminae/region in the spinal cord (see Fig. S20 for all neuronal clusters).

Comparing the sensitivity of SPLiT-seq to droplet-based methods

To compare the sensitivity of SPLiT-seq to droplet-based approaches, we measured the number of UMIs and genes detected in mouse NIH/3T3 cells for SPLiT-seq, Drop-seq, and 10x Genomics (Chromium v2 chemistry) as a function of raw sequencing reads per cell. For Drop-seq, we analyzed the 100 STAMP dataset collected in Macosko et. al. (*4*) (SRA: SRR1748412). For 10x Genomics, we analyzed the 100 cell dataset available on their website (https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.1.0/hgmm_100).

The same pipeline was used to process each sample, with the only modifications made to account for the changes in cell barcode and UMI lengths. In the first step, reads with >1 base with quality score less than phred 10 in the UMI were discarded. Reads were then aligned with STAR defaults to a combined mouse-human genome. We fixed cell barcodes with an edit distance of <=1 for all three methods. UMIs that were <=1 edit distance and corresponded to both the same cell barcode and gene were then collapsed. When we generated digital count matrices, we included intronic reads for 10x Genomics and SPLiT-seq, but excluded them from DropSeq because they led to a substantial increase in species impurity. We then subsampled each dataset between 5,000-50,000 raw reads per NIH/3T3 cell and recorded the number of UMIs as well as genes detected per cell (Fig. S3).

Supplementary Text

Barcode Collisions

Barcode collisions result from two scenarios: a mouse and human cell form a physical doublet and remain stuck to each other for the entirety of split-pool barcoding experiment, or a mouse and human cell happen to be barcoded with the same combination of barcodes by chance. The first issue can in principle be addressed by FACS sorting cells before barcoding, the second by increasing the number of barcode combinations (either by adding additional barcoding rounds or by switching to 384-well plates for any or all of the barcoding rounds).

Cluster Identification of Mouse Brain and Spinal Cord Cell Types

The 54 neuronal clusters deriving from the brain were characterized using a number of known markers. Three neuronal clusters (clusters 1-3) were determined to be mitral/tufted cells (*Tbx21+*, *Deptor+*) (*53*, *54*), types of projection neurons specific to the olfactory bulb. Consistent with previous work (*55*), medium spiny neurons (cluster 4) expressed markers specific to the striatum (*Rarb*, *Drd2*). More than 22,000 cortical pyramidal neurons clustered into 14 types (clusters 5-18), with nearly all expressing the pan-excitatory cortical marker *Satb2* (*56*). Using known markers (*7*, *8*), we were able to further assign most cortical types to specific layers (Fig. 3C).

We assigned cluster 19 to the rostral midbrain based on unique expression of *Tfap2d*, a gene known to be required in midbrain development (*57*). We identified one excitatory (cluster 20, *Slc17a6*$^{+}$) and one inhibitory neuron (cluster 21, *Gad1/2*$^{+}$) type originating from the thalamus (both *Fign+*). Eight different types of neurons expressed markers specific to the cerebellum (clusters 22-29). Expression of *Pax2* was found in two inhibitory interneuron types from the medulla (clusters 30 and 31), consistent with ISH data (*25*). Nigral dopaminergic neurons (cluster 32) were identified based on specific expression of the dopamine transporter *Slc6a3* (*35*), whose expression is restricted to the substantia nigra of the basal ganglia (*25*). Nine types of pyramidal cells and granular cells were inferred to have originated from the hippocampus (clusters 33-41). Excitatory neurons from the spinal cord (clusters 42 and 43) were marked by specific expression of *Pde11a*. We found eight types of migrating GABAergic interneurons (clusters 44-51, *Gad1/2+*), based on expression of members of the *Dlx* family (*58*). Cajal-Retzius cells (cluster 52) expressed *Trp73*, which is expressed in the P4 hippocampus and the marginal zone of the cortex, confirming their known distribution (*59*). Clusters 53 and 54 contained pan-neuronal markers, but could not be assigned to a specific cell type. After inspection of sample distribution for the unresolved clusters (Fig. S19), it was found that these clusters represented neurons originating from the spinal cord. Re-embedding of these cells resulted in substantial increase in resolution (Fig. 5).

There were 19 non-neuronal clusters composed of 27,096 individual transcriptomes. We identified 6 types of oligodendrocytes (clusters 55-60) and one OPC cluster (cluster 61), which together formed a lineage (Fig. 2D-E, Fig. S7, Fig. S8). Immune cells expressed the pan-immune marker *Dock2*, but *Ly86* expression was restricted to microglia (cluster

63), whereas *Mrc1* expression occurred only in macrophages (cluster 62, Fig. S5) (*20*, *21*). Both types of vascular cells expressed *Rgs5*, with endothelial cells (cluster 64) marked by distinct expression of *Flt1* and *Kdr* (*10*) and smooth muscle cells (cluster 65) marked by expression of *Abcc9* and *Pdgfrb* (Fig. S5) (*60*). We identified two vascular and leptomeningeal cell (VLMC) subtypes (clusters 67 and 66), both expressing previously characterized markers *Col1a1* and *Pdgfra* (*12*). Cluster 67 VLMCs expressed *Slc47a1* and *Slc47a2*, while cluster 66 VLMCs specifically expressed *Slc6a13* (Fig. S5).

Astrocytes were the most abundant non-neuronal cell type, accounting for 50% of all non-neuronal nuclei (n=13,481). Among the four astrocyte types (all *Aldh1l1*+), only Bergman glia (cluster 71) expressed *Gria1* (Fig. 2C) (*22*). Cluster 70 astrocytes—found only in the spinal cord—expressed *Gfap* highly, while cluster 69 astrocytes—found almost exclusively in the brain—expressed *Prdm16*. Cluster 68 astrocytes—found in both the brain and spinal cord—were defined by specific expression of *Slc7a10*. Ependymal cells (cluster 72) uniquely expressed many previously characterized markers (*11*) such as *Foxj1* and *Dnah1/2/5/9/10/11*. Olfactory ensheathing cells (OEC, cluster 73), a type of Schwann cell specific to the olfactory bulb (*26*), were identified by unique expression of *Mybpc1,* a gene expressed specifically in the outer layers of the olfactory bulb (*27*).

Among the 30 neuronal clusters found in the spinal cord, 28 neuronal types were identified using markers from previous literature. Clusters highly expressing *Gad1/2* were determined to be GABAergic neurons. Subsets of these GABAergic neurons included glycinergic neurons, marked by *Slc6a5*, and cerebrospinal fluid-contacting neurons, marked by *Pkd2l1* and *Pkd1l2* (*55*). Clusters expressing *Slc17a6* (VGlut1) were identified as glutamatergic neurons. A subset of these glutamatergic neurons also expressed *Slc17a8* (VGlut3). The two cholinergic motor neurons (alpha and gamma) were identified with *Chat* expression (*56*).

Cost Analysis of SPLiT-seq

An itemized cost analysis of SPLiT-seq was conducted (Table S11) for an experiment using two sublibraries (884,000 barcode combinations: 48 x 96 x 96 x 2), which makes it possible to sequence 44,000 cells with an expected 5% barcode collision rate at a cost of $0.02 per cell. If six sublibraries are used (2.65 million barcode combinations: 48 x 96 x 96 x 6), more than 132,000 cells could be processed at a cost of $0.01 per cell. The majority of costs derive from reverse transcription and ligation enzymes. The price per cell drops dramatically with scale because experimental costs do not increase linearly with cell numbers. Adding additional sublibraries does marginally increase costs, largely due to the use of more reverse transcriptase, polymerase, and Nextera reagents.

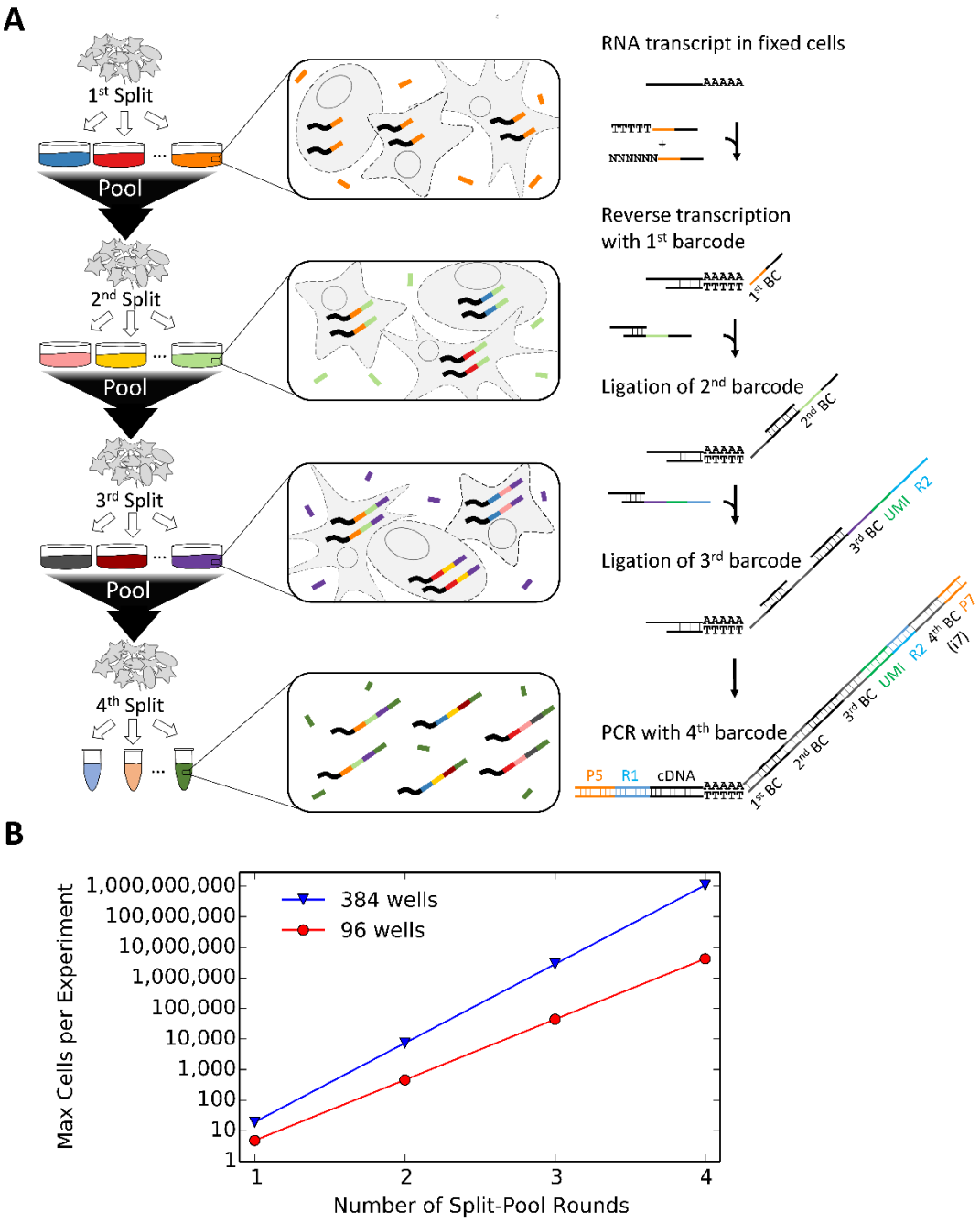

Fig. S1. Molecular diagram and exponential scalability of SPLiT-seq. (A) Fixed and permeabilized cells are randomly split into wells that each contain reverse transcription primers with a well-specific barcode. *In situ* reverse transcription converts RNA to cDNA while appending the well-specific barcode. Cells are then pooled and again randomly split into a second set of wells, each containing a unique well-specific barcode. These barcodes are hybridized and ligated to the 5'-end of the barcoded reverse transcription primer to add a second round of barcoding. The cells are pooled back together and a subsequent split-ligate-pool round can be performed. After the last round of ligation,

cDNA molecules contain a cell-specific combination of barcodes, a unique molecular identifier, and a universal PCR handle on the 5'-end. A fourth barcoding round is performed during the PCR step of library preparation. **(B)** Exponential scalability of SPLiT-seq with number of split-pool rounds. The maximum number of cells was calculated with the assumption that the number of barcode combinations must be twenty times greater than the number of cells.

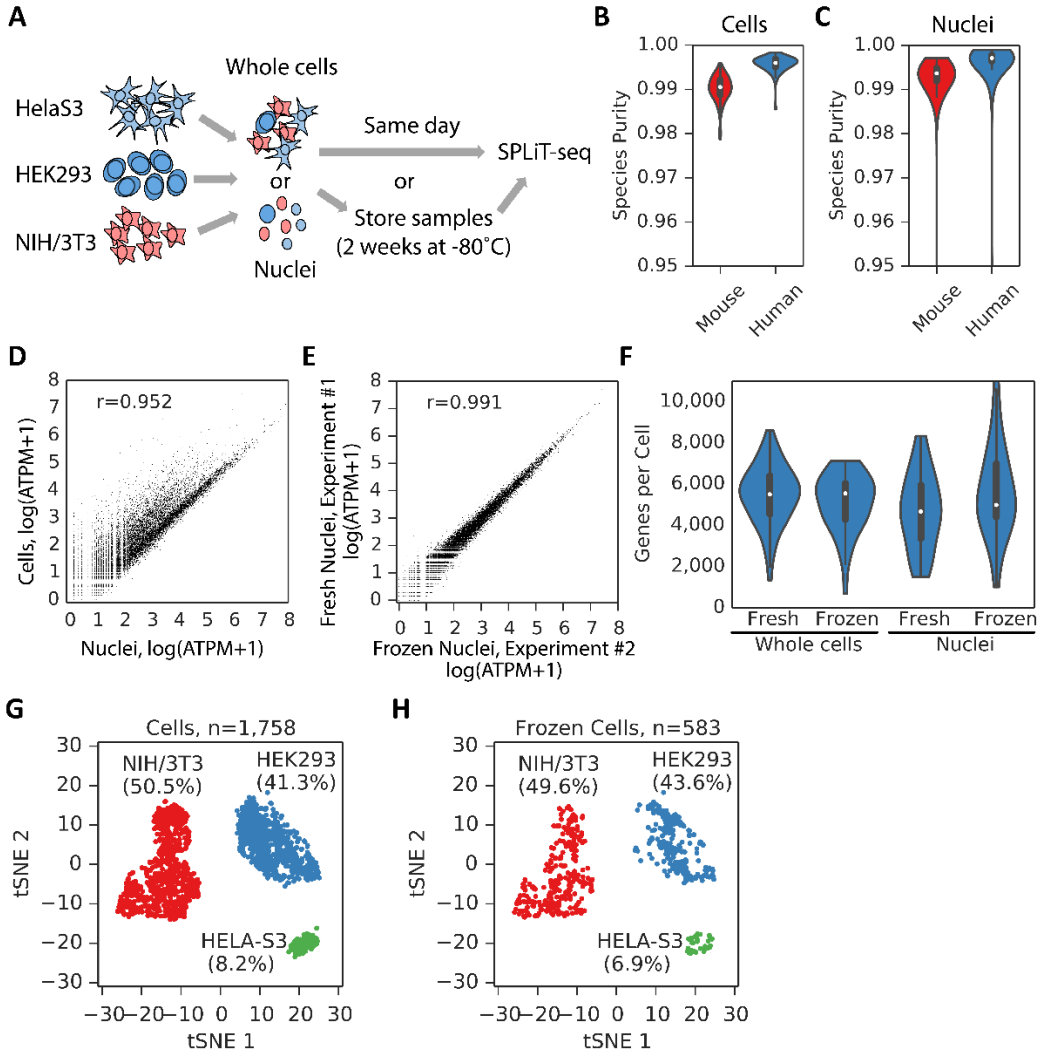

Fig S2. SPLiT-seq performance evaluation: species purity, gene expression correlation, gene detection, and cell preservation. (A) Cells or nuclei of different origin (e.g. mouse and human cell lines) were mixed and profiled with SPLiT-seq. Cells were either processed immediately or after two weeks at -80˚C. **(B)** Fraction of reads mapping to the correct species for mouse and human cells. **(C)** Fraction of reads mapping to the correct species for mouse and human nuclei. **(D)** Gene expression in nuclei and whole cells is highly correlated (Pearson-r: 0.952). Average gene expression across all cells (log average transcripts per million) is plotted for each experiment. **(E)** Gene expression from frozen and stored nuclei is highly correlated to nuclei processed immediately (Pearson-r: 0.991). **(F)** Gene counts from mixing experiments performed with fresh and frozen whole cells and nuclei. Median gene counts for fresh cells: 5,498; frozen cells: 5,540; nuclei: 4,663; frozen nuclei: 4,982**. (G)** Storing cells at -80C for two weeks does not affect cell type identification. Immediately processed cells as well as frozen and stored cells were clustered together using t-SNE. Immediately processed cells and frozen cells cluster according to cell type rather than batch/processing method. Similar proportions of cells in each cluster are maintained for frozen cells.

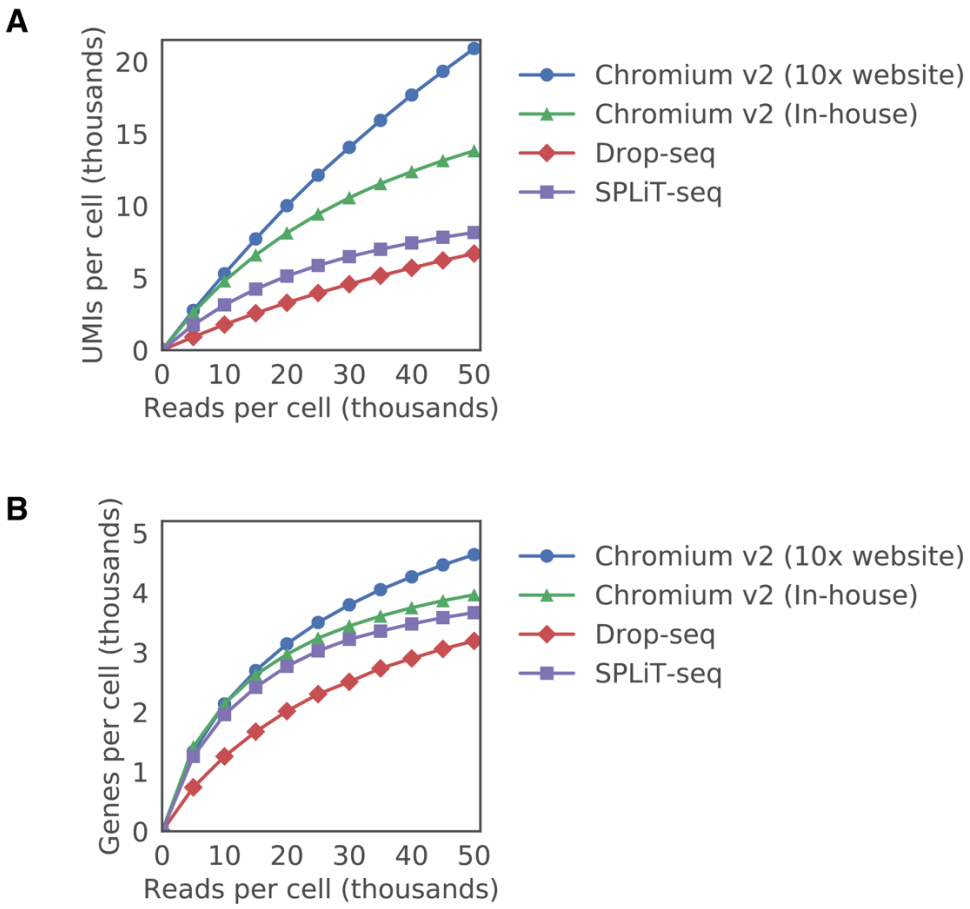

Fig. S3. Downsampling comparison of SPLiT-seq to other scRNA-seq methods. Median UMIs **(A)** or genes **(B)** detected per mouse cell (NIH/3T3) are shown as a function of raw sequencing reads. The reads for all cells were down-sampled from 50,000 to 5,000 in increments of 5,000. We compared 10x Genomics data collected in-house (at the Allen Institute for Brain Science) as well as the best available dataset on the 10x Genomics website with respect to detected UMIs per cell. Drop-seq data was taken from Macosko *et al*, while data used for SPLiT-seq was taken from the species-mixing experiment presented in Fig. 1.

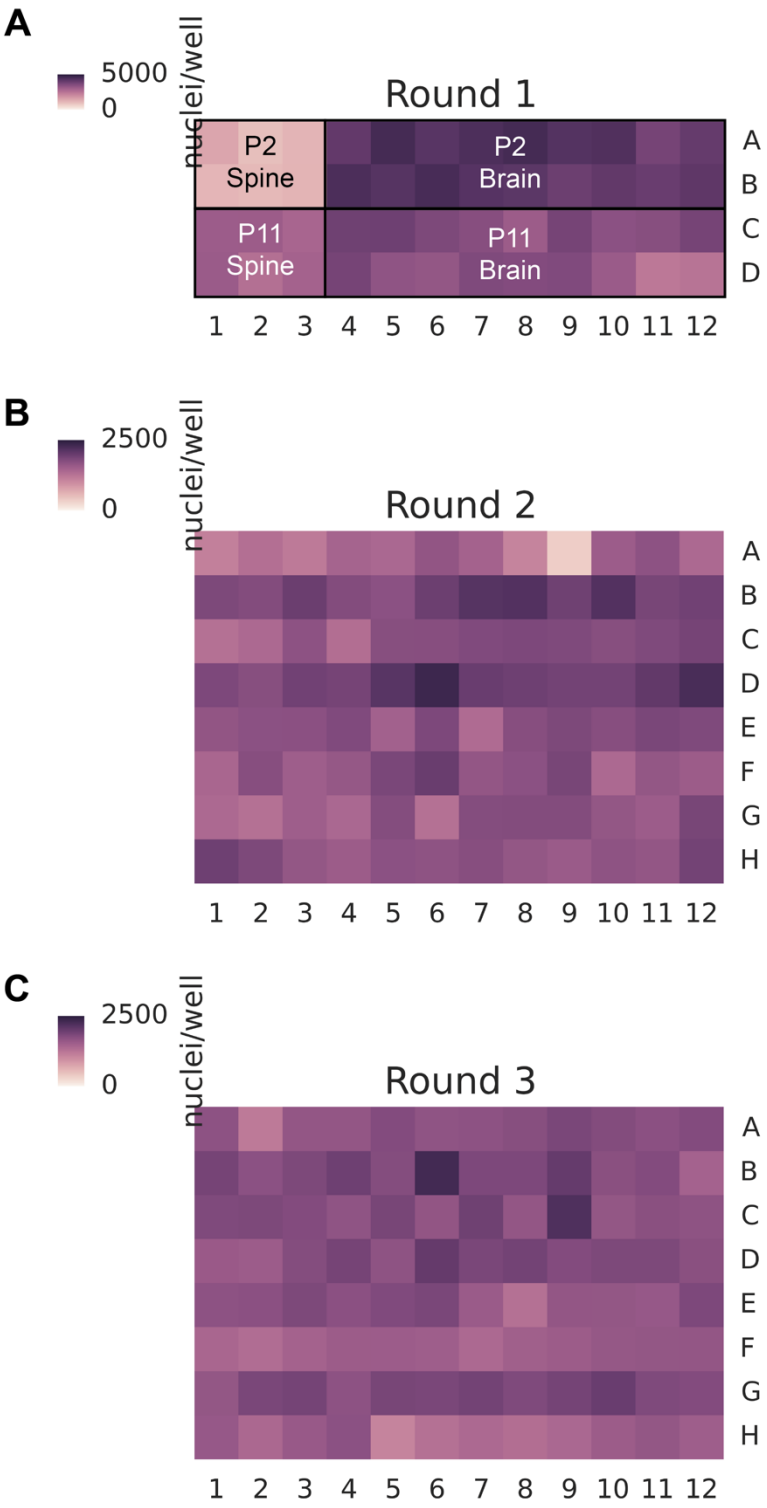

Fig. S4. Number of nuclei in each well during three rounds of barcoding. Despite pipetting cells by hand, most wells contain approximately equal numbers of nuclei. Dissociation of the P2 spinal cord resulted in fewer cells than the other samples, explaining the lower number of nuclei in the corresponding first round wells.

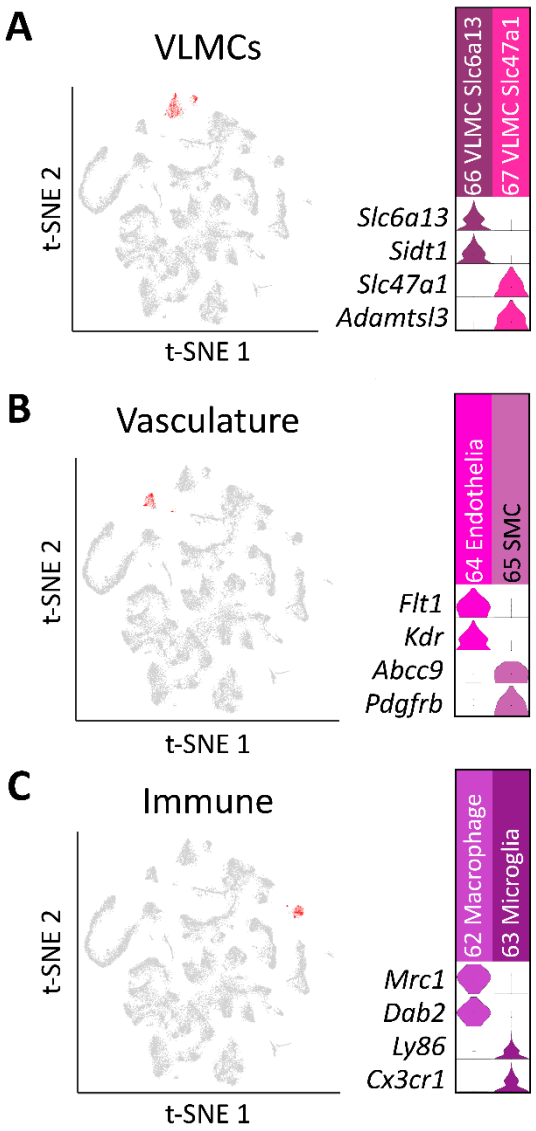

Fig. S5. Differences in VLMC, vasculature, and immune cell types. (A) During postnatal development, one VLMC subtype was found to differentially express *Slc6a13* and *Sidt1* whereas another subtype was found to differentially express *Slc47a1* and *Adamtsl3*. **(B)** Endothelia were found to differentially express *Flt1* and *Kdr* whereas smooth muscle cells were found to differentially express *Abcc9* and *Pdgfrb*. **(C)** Macrophages differentially express *Mrc1* and *Dab2* whereas microglia differentially express *Ly86* and *Cx3cr1*.

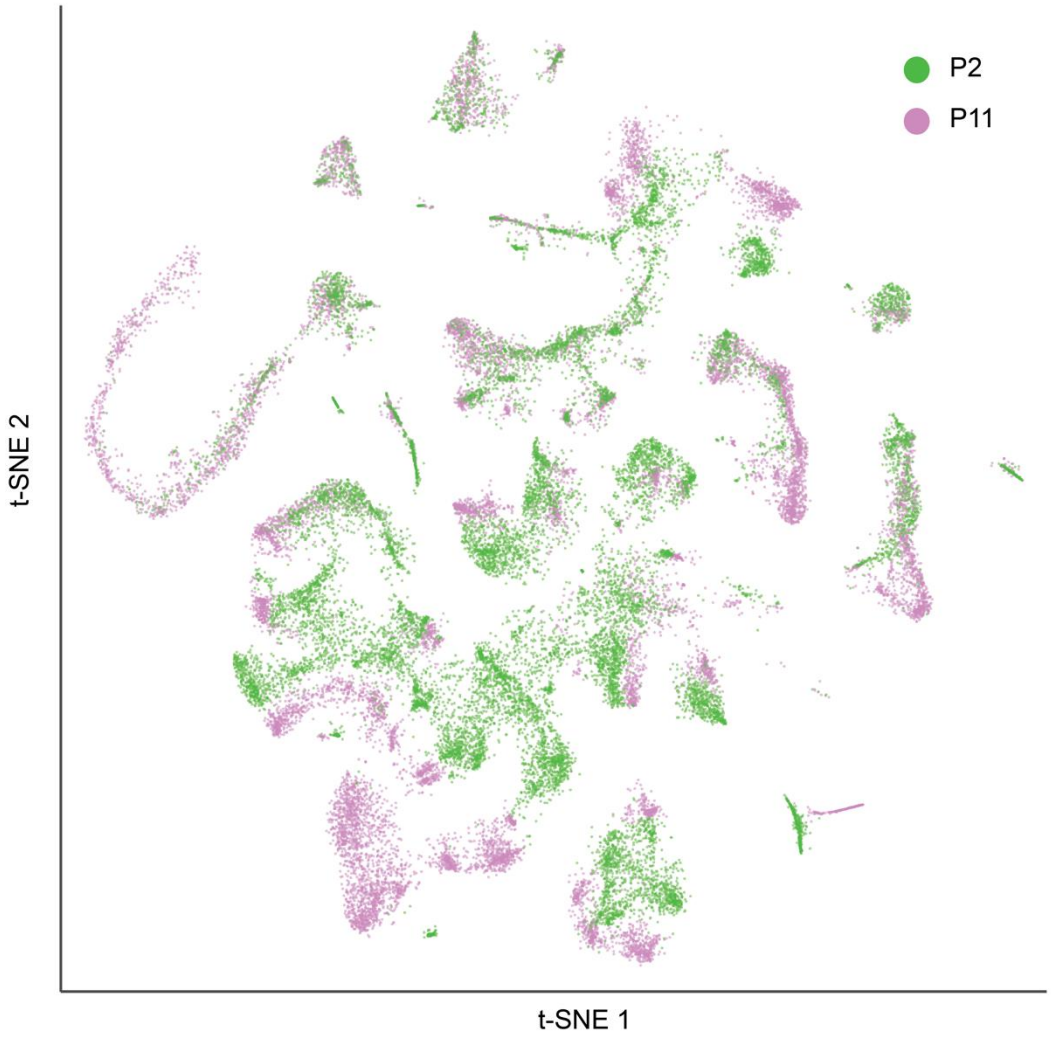

Fig. S6. Distribution of P2 and P11 transcriptomes projected with t-SNE.

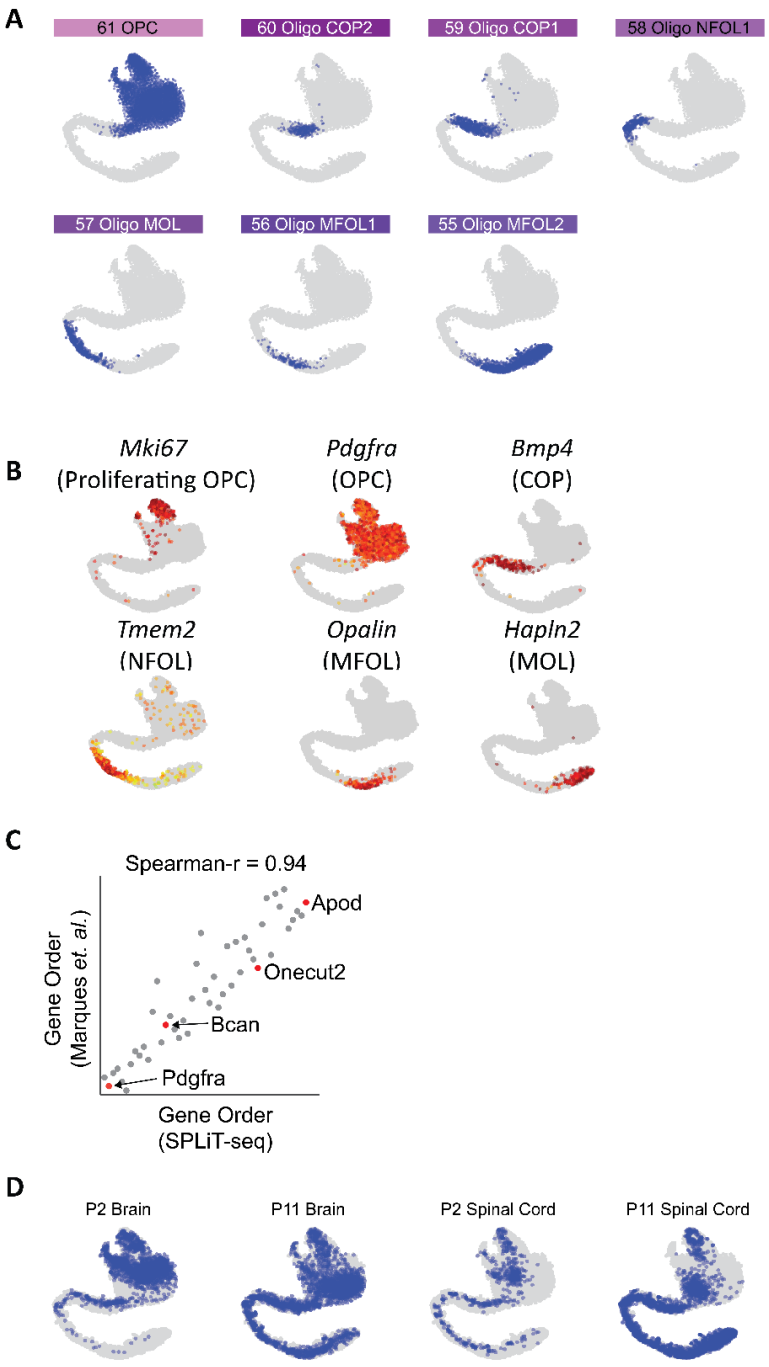

Fig. S7. Oligodendrocyte lineage. (A) Single-nucleus transcriptomes from seven clusters within the oligodendrocyte lineage were re-embedded with t-SNE. **(B)** Gene markers overlaid on the re-embedded t-SNE show proliferative markers like *Mki67* on one end with mature markers like *Hapln2* on the other end **(C)** Comparison of gene ordering between our oligodendrocyte lineage and that in Marques *et. al*. **(D)** Distribution

of P2/P11 brain and spinal cord single-nucleus transcriptomes within the oligodendrocyte lineage.

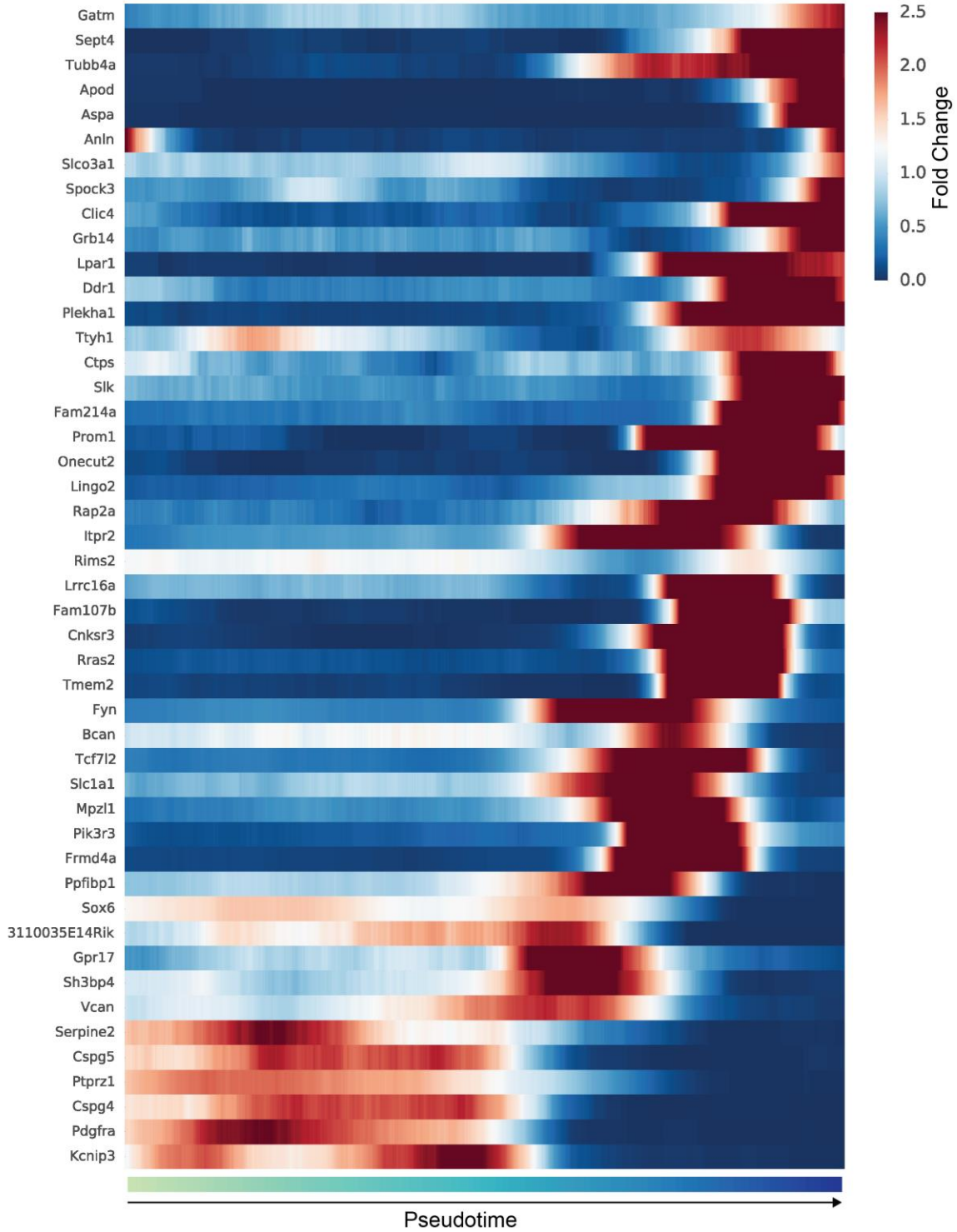

Fig. S8. Gene expression in oligodendrocyte lineage. Genes are chosen from Marques *et. al.*(*9*). Fold change is calculated relative to mean gene expression in the entire lineage.

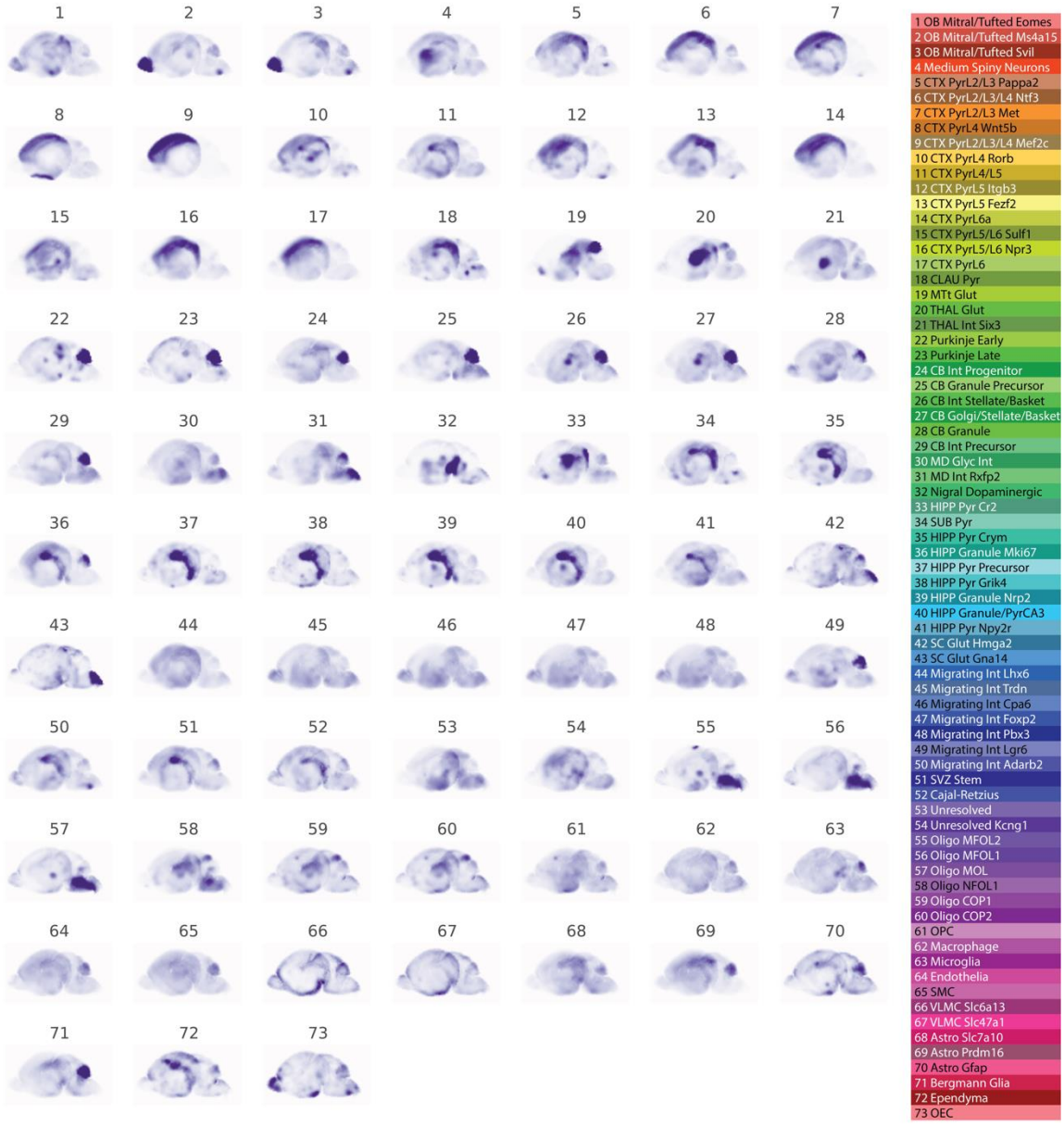

Fig. S9. Composite P4 ISH maps generated from the Allen Developing Brain Atlas. For each cluster, we selected the 5 genes most enriched in that cluster. We then averaged P4 ISH data for these genes and plotted the cumulative ISH signal across sagittal slices.

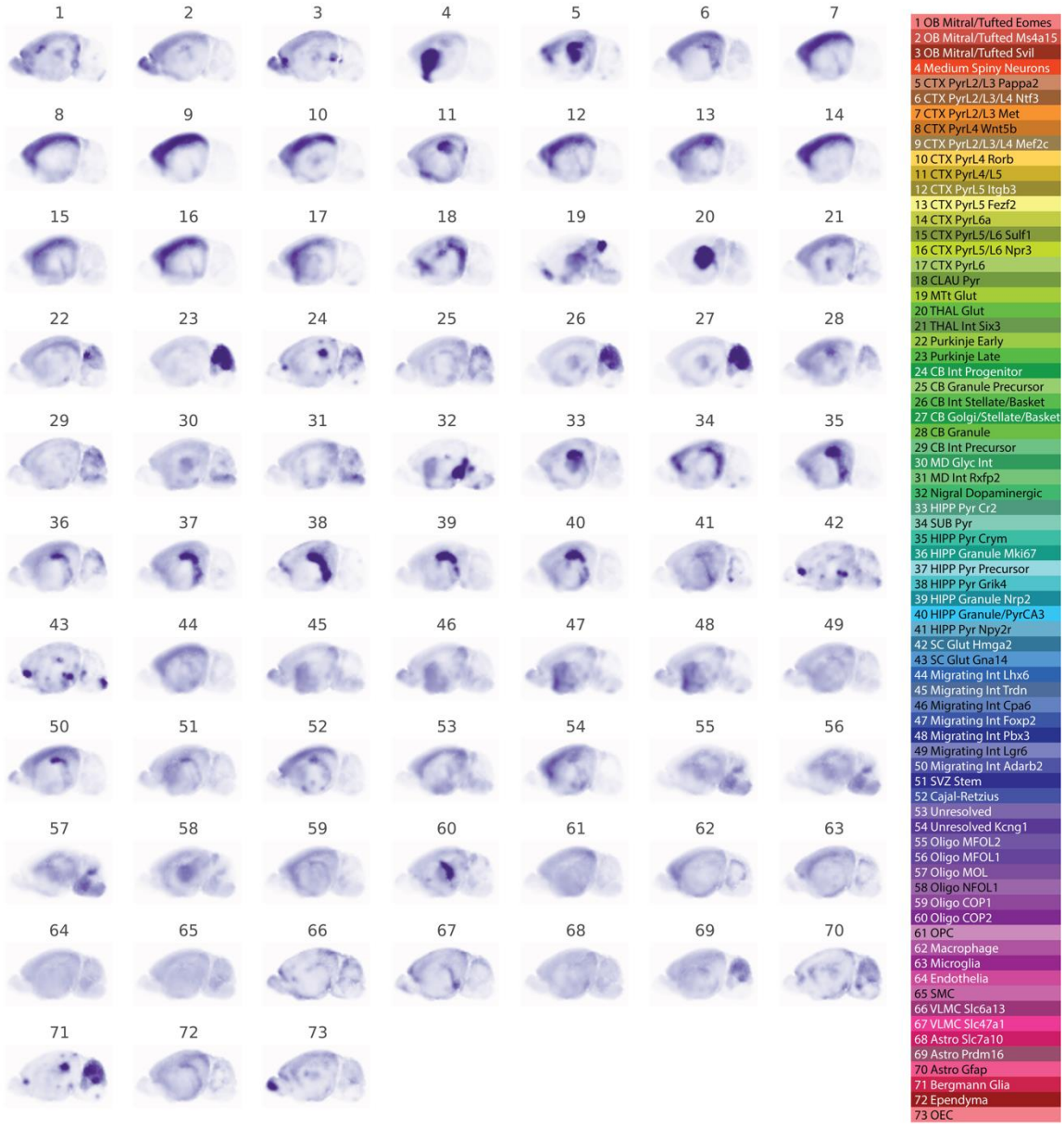

Fig. S10. Composite P14 ISH maps generated from the Allen Developing Brain Atlas. For each cluster, we selected the 5 genes most enriched in that cluster. We then averaged P14 ISH data for these genes and plotted the cumulative ISH signal across sagittal slices.

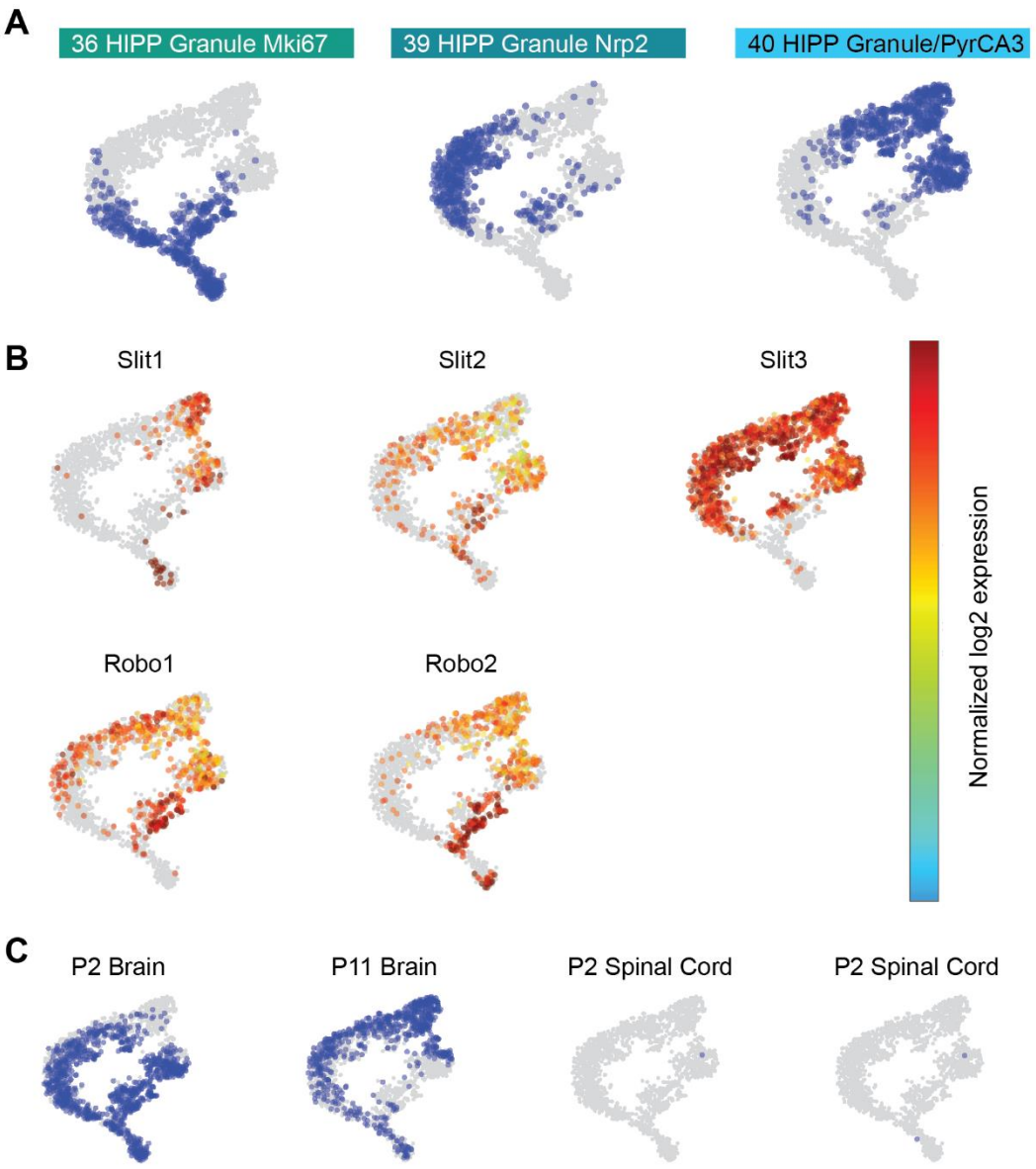

Fig. S11. Branching hippocampal neuronal lineage. (A) Three hippocampal clusters were re-embedded with t-SNE. The original clusters are overlaid over the resulting t-SNE. **(B)** Dynamics of *Slit1/2/3* and *Robo1/2* across pseudotime. **(C)** Distribution of P2/P11 brain and spinal cord single-nucleus transcriptomes within the lineage.

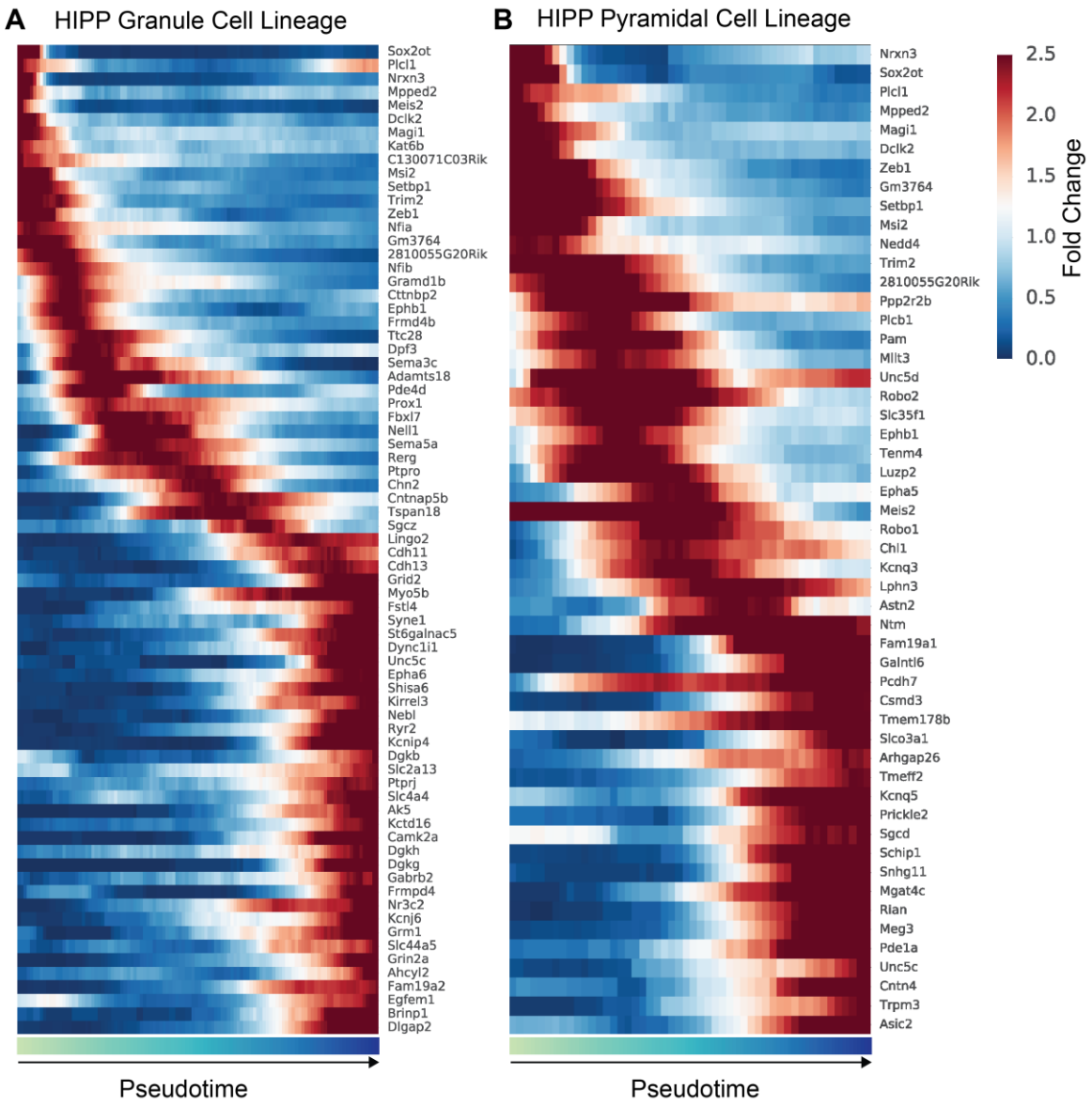

Fig. S12. Gene expression in branching hippocampal neuronal lineage. (A) Genes with differential expression across pseudotime in the granule cell lineage. **(B)** Genes with differential expression across pesudotime in the pyramidal cell lineage. Fold change is calculated relative to mean gene expression in the entire lineage.

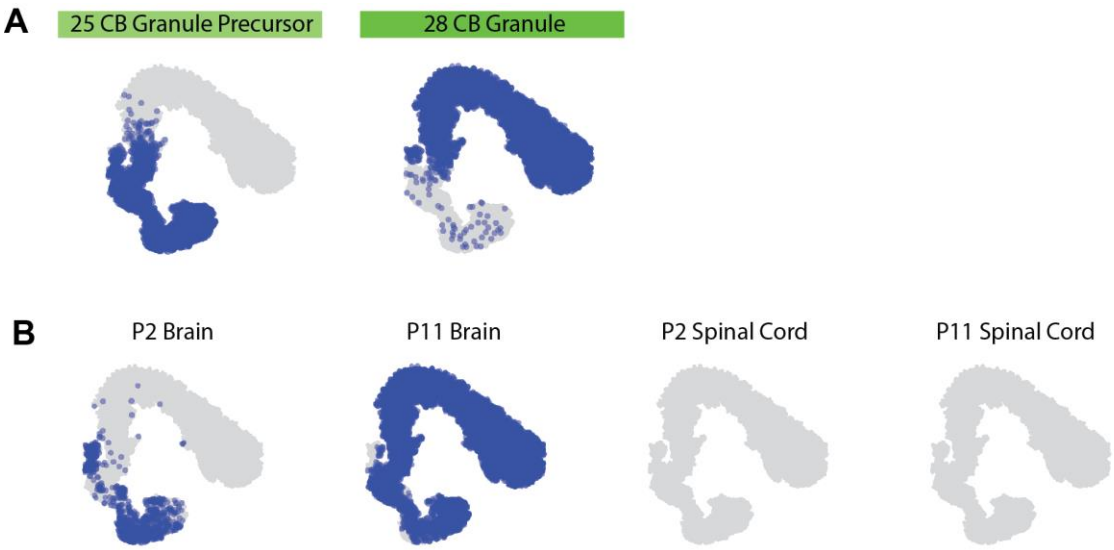

Fig. S13. (A) Cerebellar granule cell lineage. Transcriptomes from two cerebellar granule clusters were re-embedded with t-SNE. **(B)** Distribution of P2/P11 brain and spinal cord cells within the lineage.

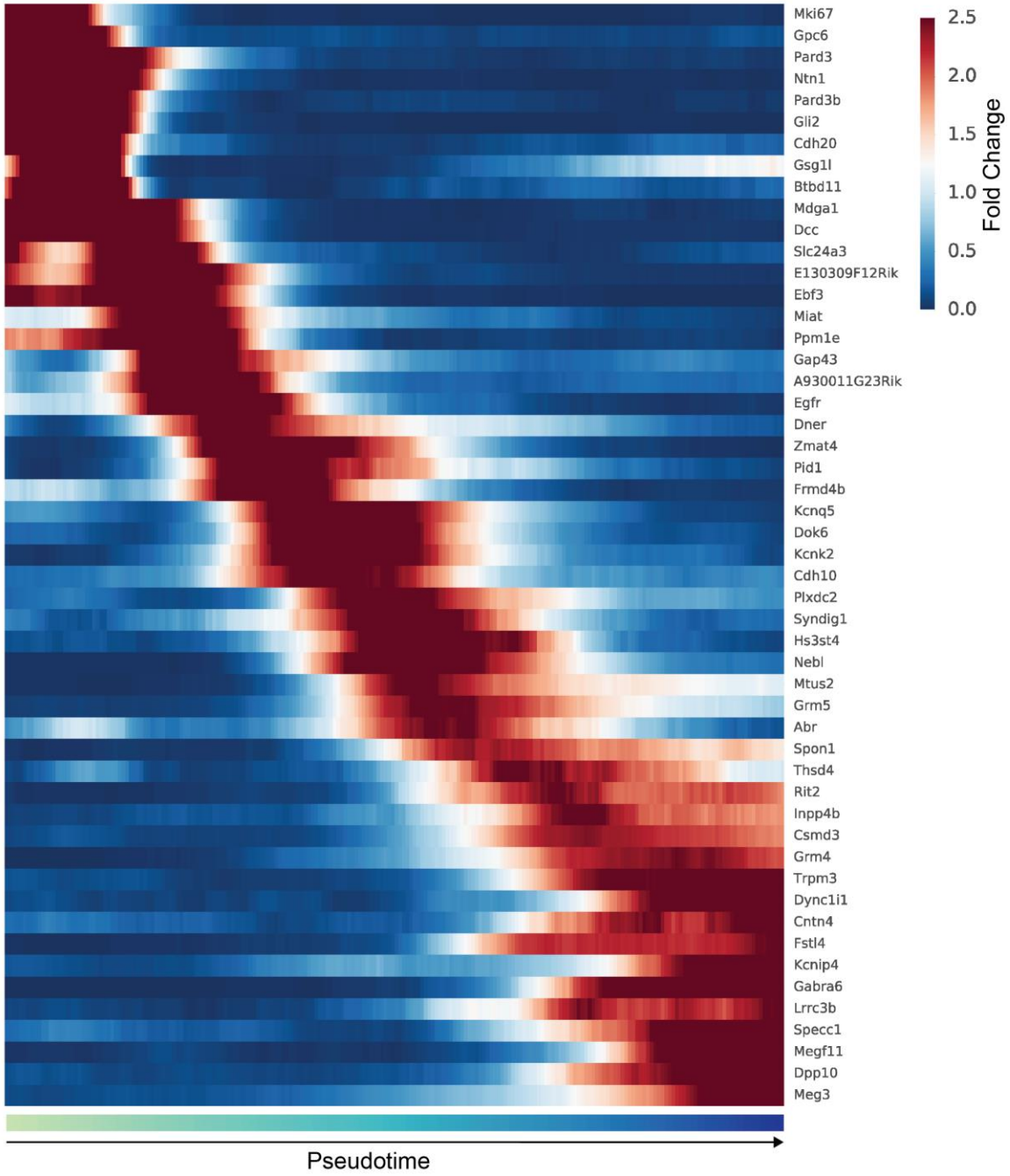

Fig. S14. Genes with differential expression across pseudotime in the cerebellar granule cell lineage.

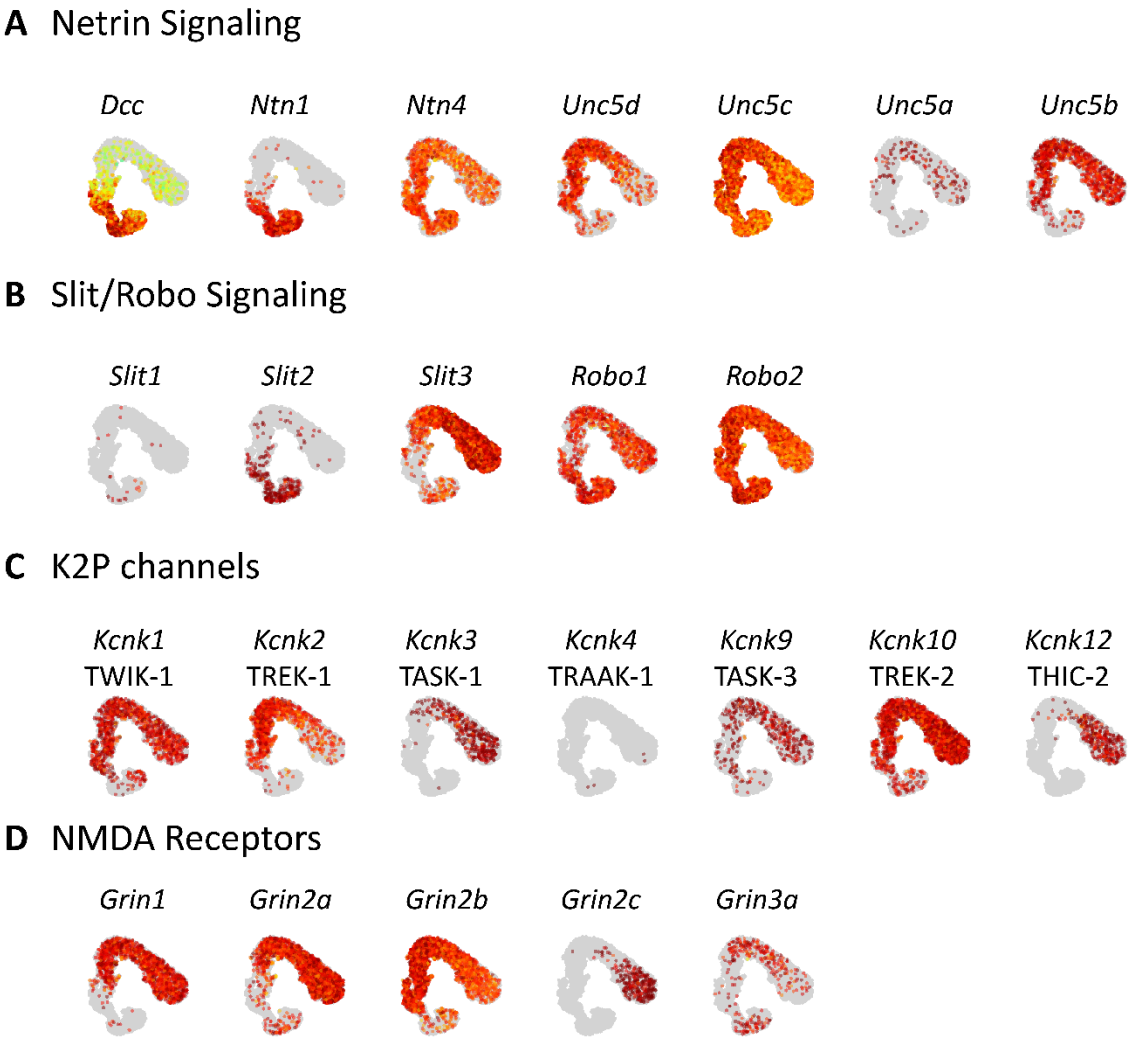

Fig. S15. Pathways relevant to cerebellar granule cell migration and development. Cerebellar granule cell lineage t-SNE overlaid with expression of genes contributing to **(A)** netrin signaling, **(B)** *Slit/Robo* signaling, **(C)** two-pore domain potassium (K2P) channels, and **(D)** N-methyl-D-aspartate (NMDA) receptors

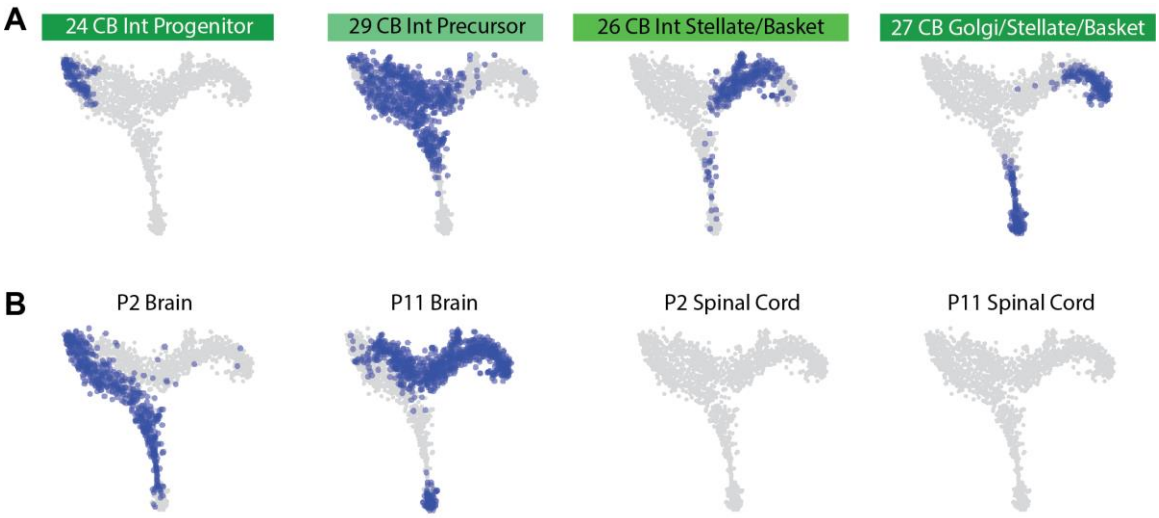

Fig. S16. (A) Cerebellar interneuron cell lineage. Four clusters were re-embedded with t-SNE. **(B)** Distribution of P2/P11 brain and spinal cord cells within the lineage.

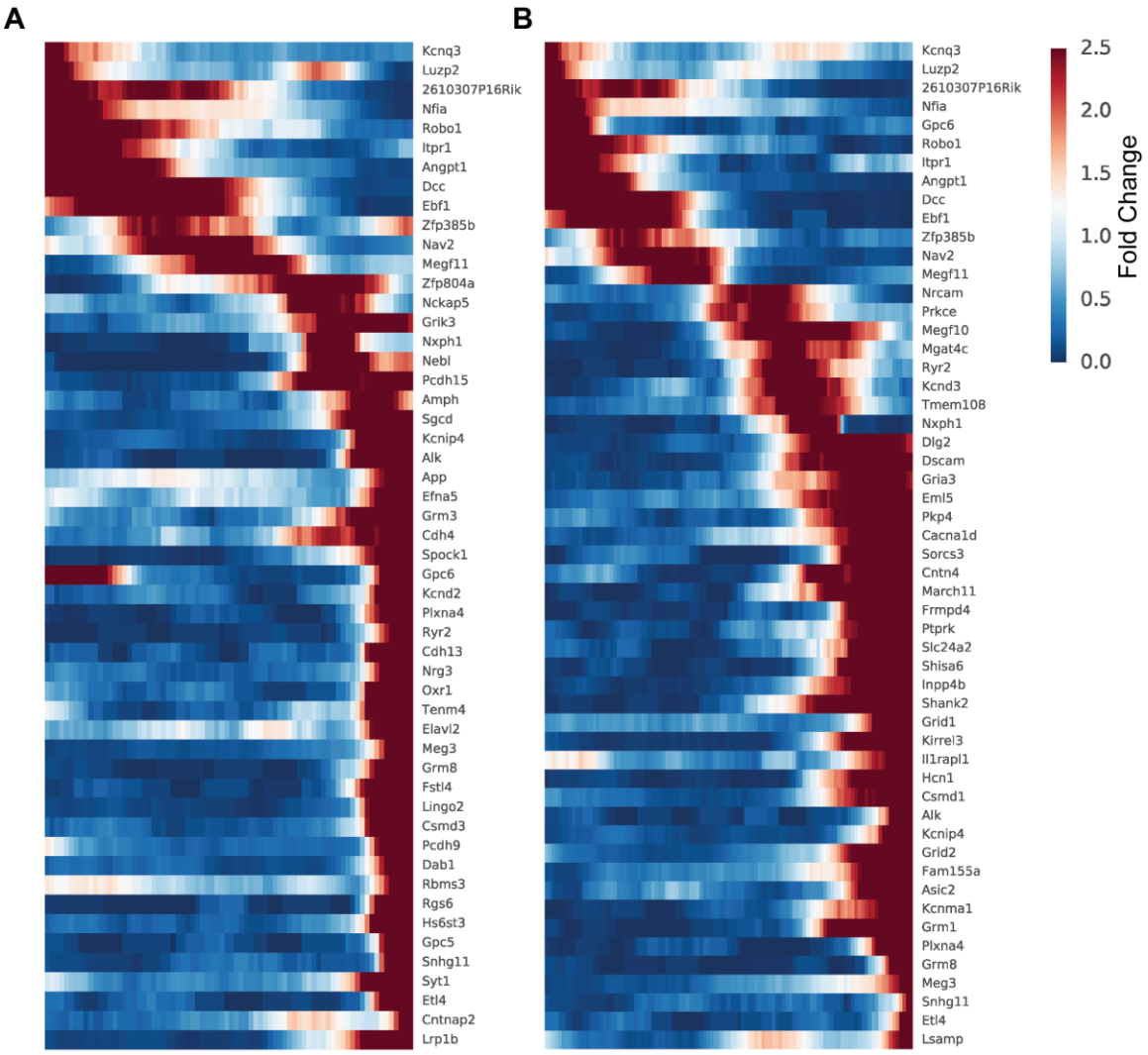

Fig. S17. Gene expression in branching cerebellar interneuron lineage. (A) Genes with differential expression across pseudotime in the stellate/basket cell lineage. **(B)** Genes with differential expression across pesudotime in the Golgi cell lineage.

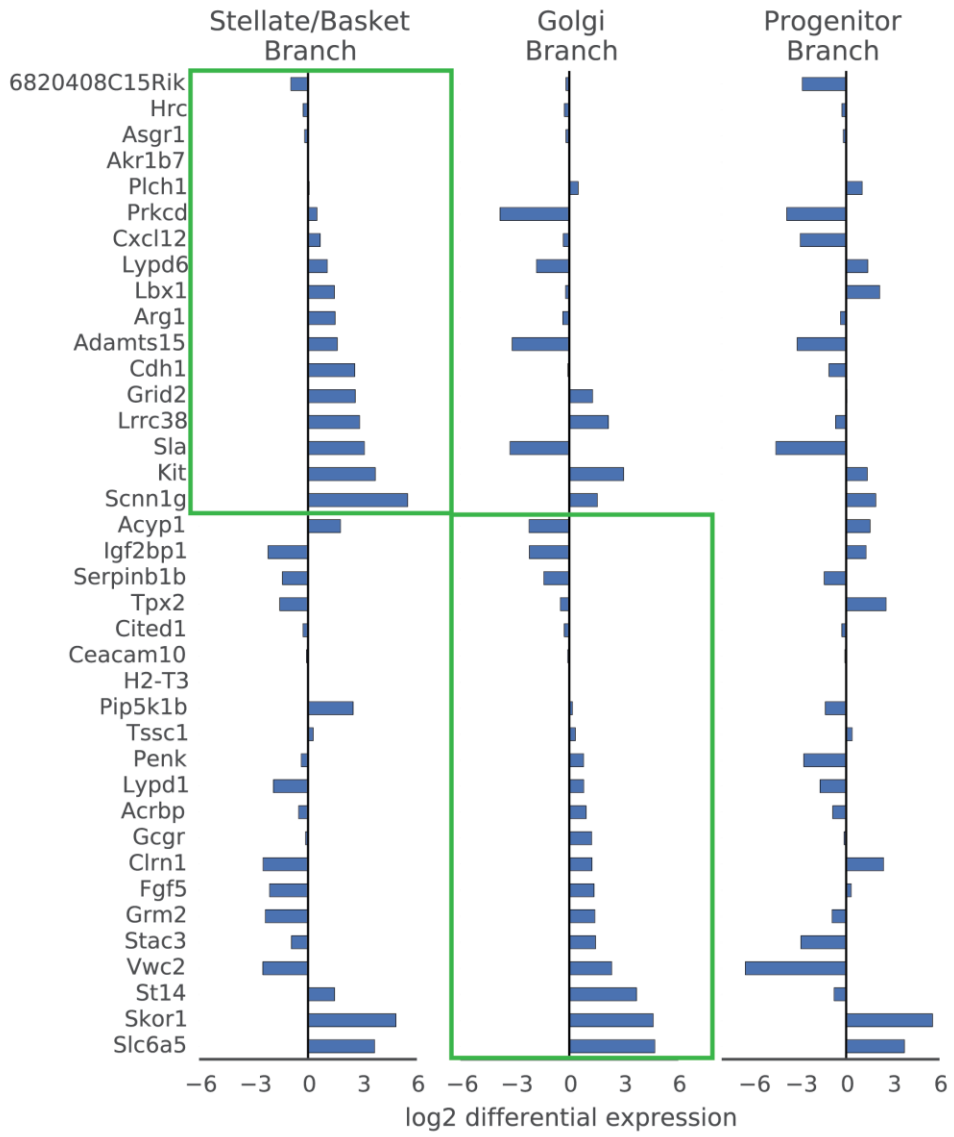

Fig. S18. Validating stellate/basket and Golgi cell identities of cerebellar interneuron lineage branches. Previously characterized marker genes of Golgi and stellate/basket cells (*55*) are plotted for each branch of the cerebellar interneuron lineage.

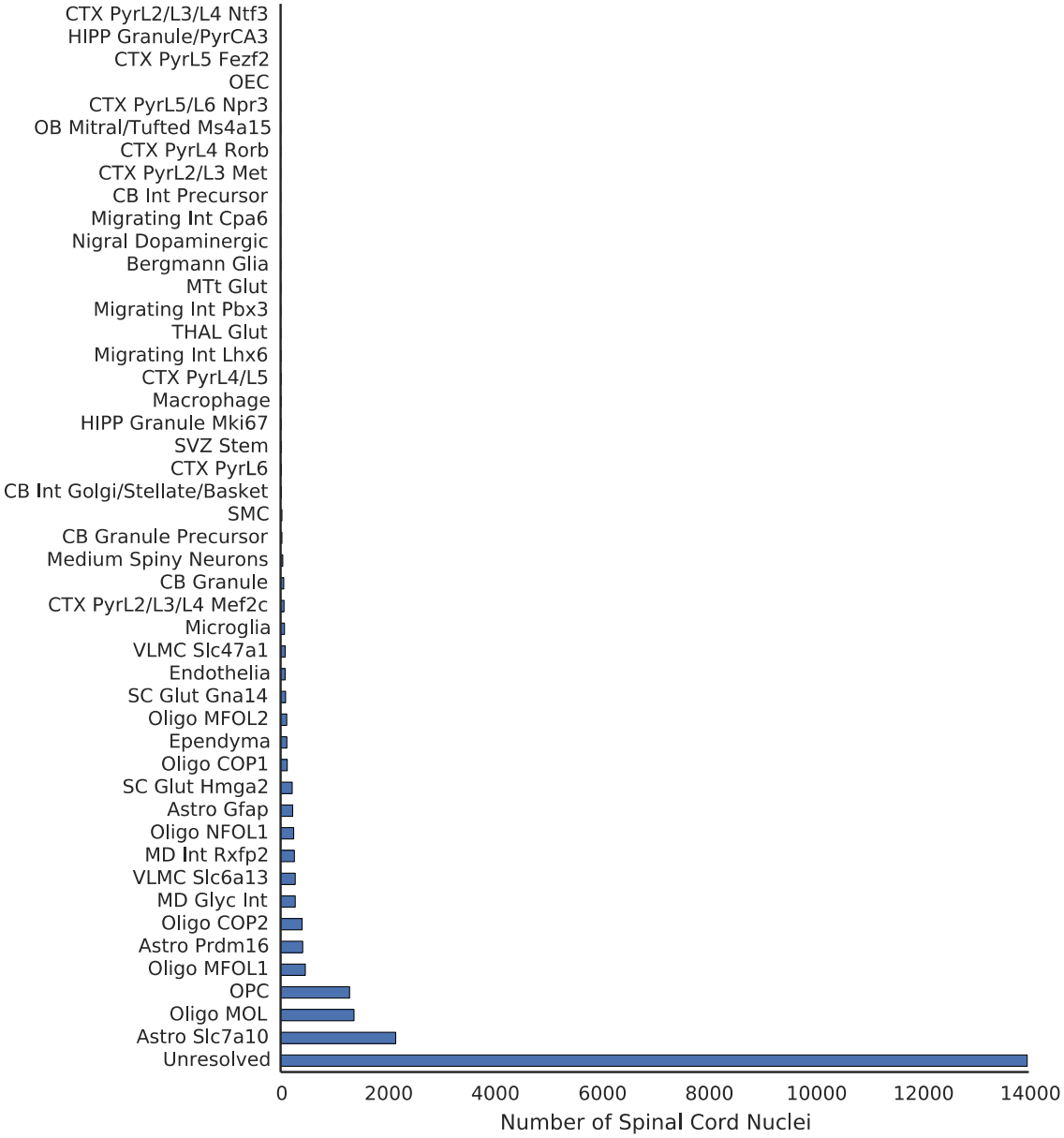

Fig. S19. Number of nuclei in each cluster from the spinal cord. Over 60% of spinal cord nuclei clustered into the unresolved cluster, leading us to re-cluster the spinal cord nuclei without brain nuclei.

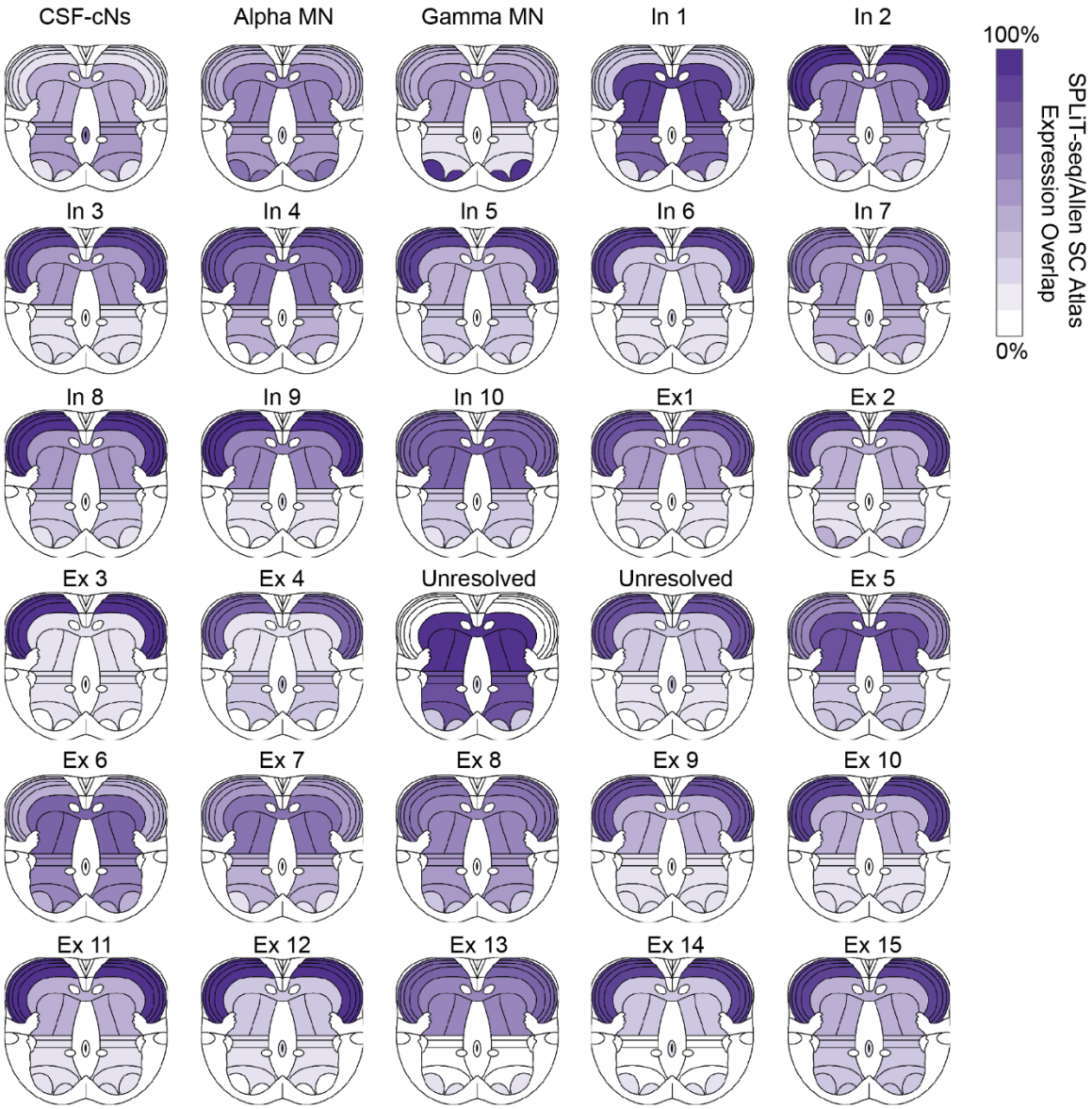

Fig. S20. Inferred spatial origin of all neuronal clusters within the spinal cord. Inferred spatial origin of neuronal clusters within the spinal cord. We analyzed the Allen Spinal Cord Atlas expression patterns of the top ten enriched genes in each cluster. Dark purple indicates expression of all ten genes in the given region, while white indicates none of the ten genes were expressed in the given region.

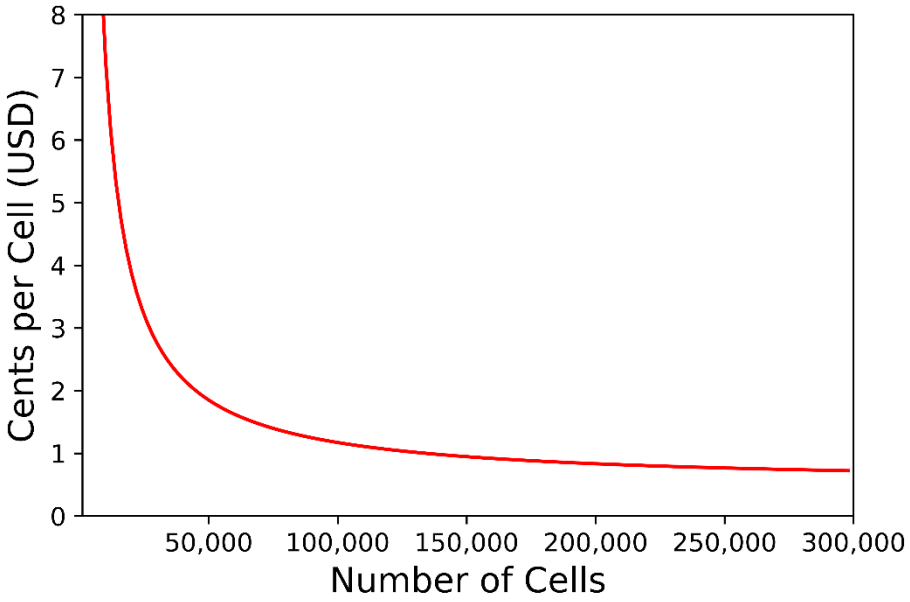

Fig. S21. Cost of library preparation per cell for SPLiT-seq. As more cells are processed, costs drop below 1 cent per cell, making SPLiT-seq a cost-effective platform to profile large numbers of cells. This analysis does not include Illumina sequencing cost.

	Total Cells	Human Cells	Mouse Cells	Mixed Cells	Fraction Human Cells	Fraction Mouse Cells	Fraction Mixed Cells	Mean Species Purity - Human	Mean Species Purity - Mouse	Median UMIs/UBC - Human	Median UMIs/UBC - Mouse	Median Genes/UBC - Human	Median Genes/UBC - Mouse	UMI Duplication
Whole Cells														
Fresh, Sample 1	1758	868	888	2	0.4937	0.5051	0.0011	0.9963	0.9919	9146.5	6808.5	4183	3253.5	66.78%
Fresh, Sample 2	168	80	88	0	0.4762	0.5238	0	0.9958	0.9901	15365	12243	5498	4497	94.54%
Frozen, Sample 1	583	293	289	1	0.5026	0.4957	0.0017	0.9953	0.9921	8363	6702	4046	3231	74.29%
Frozen, Sample 2	94	43	50	1	0.4574	0.5319	0.0106	0.9944	0.9883	15078	10951.5	5540	4319	93.43%

	Total Nuclei	Human Nuclei	Mouse Nuclei	Mixed Nuclei	Fraction Human Nuclei	Fraction Mouse Nuclei	Fraction Mixed Nuclei	Mean Species Purity - Human	Mean Species Purity - Mouse	Median UMIs/UBC - Human	Median UMIs/UBC - Mouse	Median Genes/UBC - Human	Median Genes/UBC - Mouse	UMI Duplication
Nuclei														
Fresh, Sample 1	1488	695	757	36	0.4671	0.5087	0.0242	0.9961	0.9925	9716	4847	4140	2566	66.78%
Fresh, Sample 1, Filtered	471	407	42	22	0.8641	N/A	N/A	0.9955	N/A	8193	N/A	3822	N/A	66.78%
Fresh, Sample 2	144	66	75	3	0.4583	0.5208	0.0208	0.9959	0.9922	15652	8467	5417.5	3607	94.54%
Fresh, Sample 2, Filtered	39	32	6	1	0.8205	N/A	N/A	0.9958	N/A	12113	N/A	4663	N/A	94.54%
Frozen, Sample 1	585	252	319	14	0.4308	0.5453	0.0239	0.9960	0.9930	12564.5	6162	4819	2998	74.29%
Frozen, Sample 1, Filtered	159	140	14	5	0.8805	N/A	N/A	0.9949	N/A	10489.5	N/A	4365.5	N/A	74.29%
Frozen, Sample 2	109	52	54	3	0.4771	0.4954	0.0275	0.9943	0.9911	16815	9883	5411	4020.5	93.43%
Frozen, Sample 2, Filtered	40	35	3	2	0.8750	N/A	N/A	0.9930	N/A	13636	N/A	4982	N/A	93.43%

Note 1: Two different samples were processed for each of the 4 conditions (fresh whole cells, frozen/stored whole cell, fresh nuclei, and frozen/stored nuclei). In all cases, sample 1 consists of more cells with lower sequencing depth and sample 2 consists of less cells with higher sequencing depth.
Note 2: Nuclei samples were filtered to contain less than 1% mitochondrial reads. In some cases, nuclei extraction on cell lines was inefficient, yielding a mixture of whole cells and nuclei. For this reason, statistics for both unfiltered and filtered nuclei are shown.

Table S1. Summary of species-mixing experiments with fresh/frozen whole cells/nuclei. Metrics of species-mixing experiments performed on SPLiT-seq. “Fresh” indicates that cells were harvested and directly processed using the SPLiT-seq workflow. “Frozen”

indicates that cells were harvested, fixed and stored for 2 weeks at -80˚C before continuing with the SPLiT-seq workflow. Nuclei samples underwent a computational filtering step where any uniquely barcoded nuclei containing > 1% mitochondrial reads was removed (“Filtered” samples). Due to insufficient nuclei extraction from NIH/3T3 (mouse) cells, metrics for mouse samples in the filtered nuclei datasets have been excluded.

	Total Nuclei	Mean Genes/nucleus	Mean UMIs/nucleus	Median Genes/nucleus	Median UMIs/nucleus	Total Reads	Raw reads/nucleus	UMI Duplication Rate
Small Library	131	2,729.92	11,701.03	2,055	4,943	36,332,629	277,348	94.60%
Large Library	163,069	797.96	1,347.19	677	1,022	2,456,381,771	15,063	58.20%

Table S2. Summary of snRNA-seq on mouse central nervous system.

Cluster Number	Cluster Name	Class	Number Cells	Marker Genes from Literature	Reference
1	Olfactory Bulb Mitral and Tufted - Eomes	Neuron	117	*Tbx21*	(*53*, *54*)
2	Olfactory Bulb Mitral and Tufted - Ms4a15	Neuron	271	*Tbx21*	(*53*, *54*)
3	Olfactory Bulb Mitral and Tufted - Svil	Neuron	89	*Tbx21*	(*53*, *54*)
4	Striatal Medium Spiny Neurons	Neuron	6106	*Drd2*, *Ppp1r1b*	(*55*, *61*)
5	Cortex Layer2/Layer 3 Pyramidal - Satb1	Neuron	69	*Ntf3*	(*62*)
6	Cortex Layer 2/Layer3/Layer 4 Pyramidal - Ntf3	Neuron	1446	*Rasgrf2*	(*8*)
7	Cortex Layer 2/Layer 3 Pyramidal Met	Neuron	1880	*Rasgrf2*, *Pvrl3*, *Cux2*	(*8*, *63*)
8	Cortex Layer 4 Pyramidal - Wnt5b	Neuron	255	*Slc17a6*, *Satb2*, *Sema3c*	(*8*)
9	Cortex Layer 2/Layer3/Layer 4 Pyramidal - Mef2c	Neuron	8332	*Rasgrf2*, *Pvrl3*, *Cux2*, *Rorb*	(*8*)
10	Cortex Layer 4 Pyramidal - Rorb	Neuron	779	*Thsd7a*, *Rorb*, *Cux2*, *Pvrl3*, *Rasgrf2*	(*8*, *63*)
11	Cortex Layer 4/Layer 5 Pyramidal	Neuron	2831	*Thsd7a*	(*8*)
12	Cortex Layer 5 Pyramidal - Itbg3	Neuron	198	*Rorb*, *Thsd7a*, *Sulf2*, *Kcnk2*,*Grik3*, *Etv1*	(*8*, *63*)
13	Cortex Layer 5 Pyramidal - Fezf2	Neuron	352	*Kcnk2*, *Grik3*, *Foxp2*, *Tle4*, *Tmem200a*, *Glra2*, *Etv1*	(*8*, *63*)
14	Cortex Layer 6a Pyramidal	Neuron	1498	*Grik3*	(*63*)
15	Cortex Layer 5/Layer 6 Pyramidal - Sulf1	Neuron	493	*Sulf2*,*Grik3*, *Tle4*, *Htr1f*, *Sulf1*	(*8*, *63*)
16	Cortex Layer 5/Layer 6 Pyramidal - Npr3	Neuron	550	*Grik3*, *Tle4*, *Rxfp1*	(*8*, *63*)
17	Cortext Layer 6 Pyramidal - Htr1f	Neuron	3479	*Syt6*, *Grik3*, *Foxp2*, *Tle4*, *Htr1f*	(*8*, *63*)
18	Claustrum Pyramidal	Neuron	140	*Nr4a2*	(*8*)
19	Mesencephalic Tectum Glutamatergic	Neuron	736	*Tfap2d*, *Slc17a6*	(*57*)
20	Thalamic Gluatamatergic	Neuron	2627	*Lef1*, *Tcf7l2*,*Cacna1g*,*Slc17a6*,*Wnt3*	(*64*, *65*)

21	Thalamic Interneuron	Neuron	202	*Six3, Gad1, Gad2*	(*66*)
22	Purkinje Early	Neuron	208	*Pcp2, Pde1c*	(*55*)
23	Purkinje Late	Neuron	171	*Pcp2, Slc9a3*	(*55*)
24	Cerebellar Interneuron Progenitor	Neuron	160	*Pax3, Mki67*	(*36*, *55*)
25	Cerebellar Granule Precursor	Neuron	4364	*Gli2*	(*67*)
26	Cerebellar Interneuron - Stellate and Basket	Neuron	459	*Rora*	(*36*, *55*)
27	Cerebellar Interneuron - Golgi, Stellate and Basket	Neuron	420	*Tfap2b*	(*36*, *55*)
28	Cerebellar Granule	Neuron	10996	*Gabra6*	(*68*)
29	Cerebellar Interneuron Precursor	Neuron	851	*Pax2*	(*36*, *55*)
30	Medulla Glycinergic Interneuron - Rxfp2	Neuron	329	*Stac, Slc6a5, Glra1*	(*69*)
31	Medulla Interneuron - Rxfp2	Neuron	282	*Stac*	(*69*)
32	Nigral Dopaminergic	Neuron	47	*Slc6a3*	(*70*)
33	Hippocampal Pyramidal - Cr2	Neuron	232	*Cr2*	(*71*)
34	Subiculum Pyramidal	Neuron	78	*Ntm, Rxfp1, Nr4a2*	(*72*)
35	Hippocampal Pyramidal - Crym	Neuron	1260	*Crym*	(*73*)
36	Hippocampus Granule Progenitor Mki67	Neuron	657	*Prox1, Mki67*	(*26*)
37	Hippocampal Pyramidal Precursor	Neuron	315	*Nrp1, Zbtb20*	(*74*, *75*)
38	Hippocampal Pyramidal - Grik4	Neuron	625	*Grik4, Slc17a7*	(*76*)
39	Hippocampal Granule Precursor - Nrp2	Neuron	515	*Prox1, Nrp2*	(*26*)
40	Hippocampal Granule/Pyramidal CA3	Neuron	772	*Prox1, Slc17a7*	(*26*)
41	Hippocampal Pyramidal - Npy2r	Neuron	117	*Npy2r, Slc17a7*	(*77*)
42	Spinal Cord Glutamatergic - Hmga2	Neuron	231	*Slc17a6, Slc17a8*	(*78*)
43	Spinal Cord Glutamatergic Gna14	Neuron	95	*Slc17a8, Gna14*	(*78*)

44	Migrating Interneuron - Lhx6	Neuron	3907	*Dlx family, Lhx6*	(*79*)
45	Migrating Interneuron - Trdn	Neuron	246	*Dlx family, Trdn*	(*79*)
46	Migrating Interneuron - Cpa6	Neuron	2166	*Dlx family, Cpa6*	(*79*)
47	Migrating Interneuron - Foxp2	Neuron	701	*Dlx family, Foxp2*	(*79*)
48	Migrating Interneuron - Pbx3	Neuron	1835	*Dlx family, Pbx3*	(*79*)
49	Migrating Interneuron - Lgr6	Neuron	484	*Dlx family, Lgr6*	(*79*)
50	Migrating Interneuron - Adarb2	Neuron	47	*Dlx family, Adarb*	(*79*)
51	Subventricular Zone Stem Cell	Neuron	182	*Gfap, Vim, Nes, Dlx family*	(*80*)
52	Cajal-Retzius	Neurons	133	*Trp73, Reln*	(*81*)
53	Unresolved	Neuron	36469		
54	Unresolved	Neuron	90		
55	Oligodendroctye Myelinating 2	Oligo -dendrocyte	721	*Opalin*	(*9*)
56	Oligodendroctye Myelinating 1	Oligo -dendrocyte	1781	*Opalin*	(*9*)
57	Oligodendrocyte Mature	Oligo -dendrocyte	191	*Hapln2*	(*9*)
58	Oligodendroctye Newly Formed 1	Oligo -dendrocyte	467	*Tmem2*	(*9*)
59	Committed Oligodendroctye Precursor Cells 1	Oligo -dendrocyte	811	*Gpr17*	(*9*)
60	Committed Oligodendroctye Precursor Cells 2	Oligo -dendrocyte	323	*Bcas1*	(*9*)
61	Oligodendroctye Precursor Cells	OPC	5793	*Pdgfra*	(*9*)
62	Perivivascular Macrophage	Immune	63	*Dab2*	(*82*)
63	Microglia	Immune	558	*Tgfbr1*	(*83*)
64	Endothelia	Vasc -ulature	561	*Flt1, Kdr*	(*7*)
65	Smooth Muscle Cells	Vasc -ulature	98	*Abcc9, Pdgfrb*	(*60*)
66	Vascular and Leptomeningeal Cells 2	VLMC	1223	*Slc6a13*	(*9*)
67	Vascular and Leptomeningeal Cells 1	VLMC	251	*Slc47a1, Slc47a2*	(*9*)
68	Astrocyte - Slc7a10	Astrocyte	3569	*Slc7a10*	(*84*)
69	Astrocyte - Prdm16	Astrocyte	8103	*Prdm16*	(*85*)
70	Astrocyte - Gfap	Astrocyte	282	*Gfap*	(*86*)
71	Bergman Glia	Astrocyte	1527	*Gria1*	(*87*)

72	Ependyma	Ependyma	518	*Dnah1/2/5/9/10/11*	(*8*)
73	Olfactory Ensheathing Cells (OEC)	Schwann Cells	256	*Lama4, Col5a2, Runx1*	(*88*, *89*)

Table S3. Table of assigned cell types and marker genes from literature

Table S4. Top 50 differentially expressed genes in each cluster from the joint brain and spinal cord clustering. Differential expression is calculated as log2($TPM_{CLUSTER}$+1)/ log2($TPM_{\sim CLUSTER}$+1), where $TPM_{\sim CLUSTER}$ is the average TPM for all the cells not in the cluster of interest. We only include genes expressed in at least 20% of the transcriptomes in a cluster.

Table S5. Average expression for each cluster from the joint brain and spinal cord clustering. All values are listed as TPM+1.

Table S6. Genes used to generate P4 sagittal composite ISH maps. Top ten differentially expressed genes from each cluster that were also available in the Allen ISH database for a postnatal day 4 mouse were used.

Table S7. Genes used to generate P14 sagittal composite ISH maps. Top ten differentially expressed genes from each cluster that were also available in the Allen ISH database for a postnatal day 4 mouse were used.

Cluster Number	Cluster Name	Number Cells	Marker Genes from Literature	Ref
1	Ependymal	139	*Dnah1/2/5/9/10/11*	(*8*)
2	Unassigned	58		
3	Unassigned	44		
4	Astroctye - Unassigned	116		
5	Astrocyte - Gfap	394	*Aldh1l1*	(*90*)
6	Astrocyte - Slc7a10	1230	*Aldh1l1*	(*90*)
7	Astro - Svep1	986	*Aldh1l1*	(*90*)
8	Endothelial	95	*Aldh1l1*	(*90*)
9	VLMC	444	*Col1a2*	(*9*)
10	Microglia	91	*Tgfbr1*	(*83*)
11	Oligodendrocyte Mature	1436	*Mog*	(*8*)
12	Oligodendroctye Myelinating	489	*Tmem2*	(*8*)
13	OPC	1213	*Pdgfra*	(*8*)
14	Committed OPC	835	*Bcas1*	(*8*)
15	Cerebrospinal Fluid-Contacting Neurons (CSF-cNs)	51	*Pkd2l1, Pkd1l2*	(*38*)
16	Alpha motor neurons	100	*Chat, Esrrg-*	(*39, 40*)
17	Gamma motor neurons	77	*Chat, Esrrg, Esrrb, Htr1d*	(*39, 40*)
18	Inhibitory 1	46	*Gad1, Gad2*	(*91*)
19	Inhibitory 2	365	*Gad1, Gad2*	(*91*)
20	Inhibitory 3	59	*Gad1, Gad2*	(*91*)
21	Inhibitory 4	361	*Gad1, Gad2*	(*91*)
22	Inhibitory 5	397	*Gad1, Gad2*	(*91*)
23	Inhibitory 6	289	*Gad1, Gad2*	(*91*)
24	Inhibitory 7	220	*Gad1, Gad2*	(*91*)
25	Inhibitory 8	81	*Gad1, Gad2*	(*91*)
26	Inhibitory 9	321	*Gad1, Gad2*	(*91*).
27	Inhibitory 10	40	*Gad1, Gad2*	(*91*)
28	Excitatory 1	54	*Slc17a6*	(*92*)
29	Excitatory 2	450	*Slc17a6*	(*92*)
30	Excitatory 3	185	*Slc17a6*	(*92*)
31	Excitatory 4	365	*Slc17a6*	(*92*)
32	Unresolved	7634		
33	Unresolved	65		
34	Excitatory 5	53	*Slc17a6*	(*92*)
35	Excitatory 6	243	*Slc17a6*	(*92*)
36	Excitatory 7	41	*Slc17a6*	(*92*)
37	Excitatory 8	189	*Slc17a6*	(*92*)
38	Excitatory 9	389	*Slc17a6*	(*92*)
39	Excitatory 10	385	*Slc17a6*	(*92*)
40	Excitatory 11	85	*Slc17a6*	(*92*)
41	Excitatory 12	612	*Slc17a6*	(*92*)
42	Excitatory 13	166	*Slc17a6*	(*92*)
43	Excitatory 14	597	*Slc17a6*	(*92*)
44	Excitatory 15	144	*Slc17a6*	(*92*)

Table S8. Table of assigned spinal cord cell types and marker genes from literature

Table S9. Top 50 differentially expressed genes in each cluster from the spinal cord clustering. Differential expression is calculated as $\log2(TPM_{CLUSTER}+1)/ \log2(TPM_{\sim CLUSTER}+1)$, where $TPM_{\sim CLUSTER}$ is the average TPM for all the cells not in the cluster of interest. We only include genes expressed in at least 20% of the transcriptomes in a cluster.

Table S10. Average expression for each cluster from the spinal cord clustering. All values are listed as TPM+1.

Items	Supplier	Item Code	Cost Per Experiment (USD)
Maxima H Minus	ThermoFisher	EP0753	386.28
RNase Inhibitor	Enzymatics, Ambion	Y9240L, AM2696	69.20
T4 DNA Ligase	New England Biolabs	M0202L	307.20
Kapa Pure Beads	Kapa Biosystems	KK8002	12.00
Dynabeads MyOne C1	ThermoFisher	65002	1.70
Nextera XT DNA Preparation Kit	Illumina	FC-131-1096	28.90
Kapa Hotstart HiFi ReadyMix	Kapa Biosystems	KK2602	22.26
Proteinase K	ThermoFisher	EO0491	3.44
dNTPs	ThermoFisher	R0192	5.03
Oligonucleotides	Integrated DNA Technologies, Exiqon	N/A	37.05
Total			873.07

Table S11. Itemized cost breakdown of SPLiT-seq

Table S12. List of all oligonucleotide sequences used