Supplementary Text and Figures

Supplementary Experimental Procedures

NSC isolation from adult mouse brains

For single cell RNA-seq library generation of *in vivo* cells, four 3-month old male GFAP-GFP mice (Jax cat #003257) were euthanized, and brains were immediately harvested, and placed in a buffered solution containing glucose and PIPES buffer (20mM PIPES, 4.5mM KCl, 120mM NaCl, 0.5% glucose, 10µg/mL gentamicin (ThermoFischer), 1:100 Penicillin-Streptomycin-Glutamine (ThermoFischer)). As described in (Codega et al., 2014), the SVZ from each hemisphere was micro-dissected by first performing a mid sagittal cut and then removing the midbrain structures to reveal the lateral wall of the lateral ventricles. The subventricular zone was subsequently sheared off of the lateral wall of the ventricle and placed in a PIPES buffered solution (20mM PIPES, 4.5mM KCl, 120mM NaCl, 0.5% glucose, 10µg/mL gentamicin (ThermoFischer), 1:100 Pennicilin-Streptomycin-Glutamine (ThermoFischer)). Following a brief spin, the SVZ was dissociated with enzymatic digestion with Papain for 10 minutes (Worthington Biochemical #LS003118) at a concentration of 14U/mL. The dissociated SVZ was then titrated in a solution containing 0.7mg/mL ovomucoid, and 0.5 mg/mL DNAse-I in DMEM/F12. The dissociated SVZ was then centrifuged through 22% Percoll in PBS to remove myelin debris. Following centrifugation through Percoll solution, cells were washed 1x with FACS buffer (HBSS, 1% BSA, 1% Glucose). Antibody staining was carried out in FACS buffer at the following dilutions: Prom1-Biotin (eBioscience Cat.#13-1331-80 [1:300]), EGF-AlexaFluor 647 (Life Technologies Cat. #E35351 [1:300]), CD24-PacBlue (eBioscience Cat.#48-0242-80 [1:400]), CD31-PE (eBioscience Cat.#12-0311-81 [1:50]), CD45-BV605 (Biolegend Cat.#110737 [1:50]), Strep-PECy7 (eBioscience Cat.#25-4517-82 [1:500]). FACS sorting was performed on a BD FACS Aria II sorter, using a 100µm nozzle at 13.1 PSI. Cell gates were defined as follows (Beckervordersandforth et al., 2010; Codega et al., 2014; Garcia et al., 2004; Zhuo et al., 1997):
Astrocytes: (GFAP-GFP)$^{+}$ PROM1^{-}CD31^{-}CD24^{-}CD45^{-}
qNSCs: (GFAP-GFP)$^{+}$PROM1^{+}EGFR^{-}CD31^{-}CD24^{-}CD45^{-}
aNSCs: (GFAP-GFP)$^{+}$PROM1^{+}EGFR^{+}CD31^{-}CD24^{-}CD45^{-}
NPCs: (GFAP-GFP)$^{-}$EGFR^{+}CD31^{-}CD24^{-}CD45^{-}
Endothelial Cells: (GFAP-GFP)$^{-}$CD31^{-}
Cells were sorted into a catching medium of DMEM/F12 with B27 (1:50), B27 supplement (ThermoFisher, no Vitamin A, 1:50), N2 supplement (ThermoFisher, 1:100), 15mM HEPES buffer, 0.6% glucose, Penicillin-Streptomycin-Glutamine (Life Technologies, 1:100), and Insulin-Transferrin-Selenium (Life Technologies, 1:1000). Following sorting, cells were spun down at 300xg at 4C. Media was aspirated and cells were resuspended in catching media at a concentration of 300 cells/µL, and each population was subsequently separately loaded on to a C1 chip as described below.

Isolation, culture, and preparation of single cell libraries from neurospheres

Neurospheres were prepared by first sorting aNSCs based on the protocol described above, and allowing them to proliferate in Neurobasal-A (ThermoFisher) supplemented with B27-supplement (ThermoFisher, no Vitamin A, 1:50), Penicillin-Streptomycin-Glutamine (Life Technologies, 1:100), 20ng/mL of EGF (Peprotech), 20ng/mL of bFGF (Peprotech). Cells were grown for three passages in culture in Neurobasal-A (ThermoFisher) supplemented with B27-supplement (ThermoFisher, no Vitamin A, 1:50), Penicillin-Streptomycin-Glutamine (Life Technologies, 1:100), 20ng/mL of EGF (Peprotech), 20ng/mL of bFGF (Peprotech). Immediately prior to single cell sequencing, cells were dissociated for 5 minutes in Accutase (EMD Millipore) and washed once in PBS (ThermoFisher). Following resuspension in Neurobasal-A (ThermoFisher) supplemented with B27-supplement (ThermoFisher, no Vitamin A, 1:50), Penicillin-Streptomycin-Glutamine (Life Technologies, 1:100), 20ng/mL of EGF (Peprotech), 20ng/mL of bFGF (Peprotech) at 300 cells/µL, and they were subsequently loaded onto the Fluidigm C1 Single-Cell Auto Prep System, as described below.

Single cell RNA-seq library preparation

A 300 cell/µL cell solution was mixed at a 7:3 ratio with the Fluidigm C1 Suspension reagent and this solution was loaded onto a small size (5-10µm) Fluidigm C1 Single-Cell Auto Prep chip for all *in vivo* single cells studied and medium size (10-17µm) Fluidigm C1 Single-Cell Auto Prep chip for *in vitro* cultured neurosphere derived single cells. Note that medium size Fluidigm C1 chips can include some undetected doublets (http://info.fluidigm.com/FY16Q2-C1WhitePaperUpdate_LP.html). However, medium chips were only used for the *in vitro* cultured neurosphere cells, and undetected doublets would not alter the results. Fluidigm C1 chips can also be prone to cell size biases. However, the populations of cells loaded onto each chip were highly selected by FACS, and did not exhibit drastic changes in cell sizes. Live/dead staining was performed using the Fluidigm Live/Dead Cell Staining Solution as described in the Fluidigm C1 mRNA seq protocol (https://www.fluidigm.com/documents) and imaged using a Leica DMI4000B microscope. Following imaging, reverse transcription was performed directly on the chip using the SMARTer chemistry from Clontech, and PCR was also performed on the chip using the Advantage PCR kit (SMARTer Ultra Low RNA Kit for the Fluidigm C1, Clontech #634832). Resulting cDNA was transferred from the chip to a 96 well-plate and a subset of representative samples were analyzed by bioanalyzer to verify cDNA quality. A quarter of the cDNA for each library was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Cat.# P11496) and verified to be within a range of 0.1-0.5ng/µL (or diluted when necessary with the C1 DNA dilution buffer). Sequencing libraries were prepared directly in a 96-well plate using the Nextera XT Library Preparation Kit (Illumina Cat. # FC-131-1024). Each library (corresponding to a single cell) was individually barcoded using the Nextera XT 96-Sample Index Kit (Illumina Cat. # FC-131-1002), and all 96 bar-coded libraries from each chip were pooled into single multiplexed libraries. The DNA concentration of multiplexed libraries was measured using BioAnalyzer. These multiplexed libraries were sequenced using either the Illumina MiSeq (Illumina) or HiSeq2000 (Illumina) at a concentration of 2 pM. The details on the preparation and sequencing of each of the individual lanes can be found in Table S1.

Read alignment

Reads were trimmed using trim_galore v0.3.7 and aligned to the mm9 genome using STAR 2.3.0. Reads were counted using HTSeq v0.6.1 using the transcriptome annotation (UCSC – mm9). Following alignment, low quality cells were excluded using the following criteria: cells must be live on chip, >20,000 reads mapping to transcriptome, >500 genes detected, <30% ERCC spike-in contribution to sequencing library (when applicable).

Normalization of read counts and generation of PCA plots

Genes were considered detected if they were detected at a level of 10 counts in at least 5 cells. This detected gene set was determined for each subset of cells used for analysis (e.g. after the outlying cells were removed it was recalculated). This accounted for unexpressed genes, and for the presence of a particularly rare subset of cells. Across-sample normalization of all cells was conducted using TMM normalization for all high quality cells as implemented in the R/Bioconductor EdgeR package v3.10.5, followed by FPKM normalization. All PCAs presented in this manuscript were conducted using log2 transformed FPKM values using the PCA_int() function and the first two principal components of variance.

Population RNA-seq on *in vivo* populations isolated from the SVZ

For *in vivo* populations of astrocytes, qNSCs, aNSCs, NPCs, and endothelial cells, ~400 cells were sorted and population-level RNA-sequencing libraries were prepared using the SMARTer Universal Low Input RNA Kit (Clontech Cat. #634938). Libraries were sequenced on HiSeq. Reads were trimmed using trim_galore (v0.3.7) and aligned to the mm9 genome using STAR 2.3.0. Reads were counted using HT-seq (v0.6.1) using the transcriptome annotation (UCSC – mm9). Further details will be available in (Leeman et al, submitted).

Clustering of single cell RNA-seq together with aggregated single cell and population data

Single cell aggregated datasets were generated by summing all counts for all cells defined by their sorting identity (astrocytes, qNSCs, aNSCs, NPCs). Only genes with >1 count in at least one of the single cell aggregated or population-level datasets were included in the analysis. Then, one single cell from each of the cell types (astrocytes, qNSCs, aNSCs, NPCs) was selected at random. PCA and spearman clustering (using the R function varclust() from the R package Hmisc v3.17-0) were conducted using log2 transformed FPKM values.

Identification and exclusion of oligodendrocyte-like cells in putative NSC populations

PCA on our single cell RNA-seq data was conducted using all detected genes for all high quality cells. A clear group of cells consisting primarily of qNSCs and NPCs separated from the remainder of the sorted cells. The oligodendrocyte-like cells separated from the bulk of the NSCs on the 2nd principal component in Figure 1B. Thus, we selected all genes with PC loadings < -0.05 in the second principal component and performed pathway enrichment using GoRilla (http://cbl-gorilla.cs.technion.ac.il/) with all detected genes as the background that revealed strong enrichments for genes associated with oligodendrocytes and myelination.

Putative oligodendrocyte-like cells were excluded from all subsequent analyses based on their clustering relative to the NSC lineage in Figure 1B, and the tight cluster of NSCs containing the vast majority of qNSCs and aNSCs was used for subsequent analyses. A limited number of outlying astrocytes (5) and NPCs (2) that did not cluster with the NSC lineage in the global PCA were also excluded from all subsequent analyses at this stage.

Ordering of cells with Monocle using all detected genes and determination of 'cell-cycle-low' and 'cell-cycle-high' aNSCs

Monocle (Trapnell et al., 2014) ordering was conducted using all qNSC, aNSC, and NPC cells using all detected genes. Resulting pseudotimes were linearly scaled between 0 and 100 and were plotted using the plot_genes_in_pseudotime() function as provided in the Monocle R package v1.2.0 (Trapnell et al., 2014) (http://bioconductor.org/packages/release/bioc/html/monocle.html). Genes known to be regulated through the activation and differentiation of NSCs were plotted with respect to pseudotime. aNSCs and NPCs separated by the expression of cell cycle genes. Cell-cycle-low and cell-cycle-high aNSCs were defined by splitting aNSCs at the pseudotime at which the predominant expression of mitotically associated cyclins and cyclin dependent kinases (*Cdk1*, *Ccna2*, Figure 2B) is first observed in this ordering approach.

Intracellular FACS staining for the cell cycle marker Ki67

Populations of qNSCs, aNSCs, and NPCs were isolated from GFAP-GFP transgenic mice as described above. Cells were fixed in 1.6% PFA, and were permeabilized with ice-cold methanol. Populations were stained with antibodies to Ki67 (Clone SolA15 – eBioscienes) and Sox2 antibodies. Samples were read using an LSRII FACS analyzer (BD Biosciences). Percentages of Ki67- cells were compared for all Sox2 positive cells in a given population.

Construction of machine learning model and determination of consensus-ordering genes

We carried out a four-way classification between the following groups that correspond to key states/subpopulations: qNSCs, 'cell-cycle low' aNSCs, 'cell-cycle high' aNSCs, and NPCs. Classification was carried out by implementing a stochastic gradient boosted classification model using the R CRAN package GBM v2.1.1. Briefly, 20 single cells from each group (training set) were randomly selected and subjected to GBM modeling as implemented by the Caret package v.6.0-58 in R (http://topepo.github.io/caret/index.html) (.interaction.depth = 5, .n.trees = 2500, .shrinkage = 0.001, .n.minobsinnode = 5). Accuracy of the model was tested on cells that were not used for the training set. The GBM classification was bootstrapped by repeatedly sampling 20 cells from each group and building an independent model. In total, 100 GBM models were built. Following construction of the models, the top 100 features from each of the 100 models were obtained as determined by the relative.influence() function in the GBM package. A consensus set of ordering genes defining the transition of cells between groups: (qNSCs, 'cell-cycle low' aNSCs, 'cell-cycle high' aNSCs, and NPCs) was built using genes that were in the top 100 most important features of at least half of the classification models, or in the top 100 most important features of at least 25% of the models.

Ordering cells using consensus-ordering genes

Monocle ordering was repeated for all qNSC, aNSC, and NPC cells using the set of consensus-ordering genes (Table S7A) identified by machine learning. The expression of genes of interest was plotted with respect to pseudotime. The resulting pseudotime expression spectrum was divided according to the expression of genes of interest. The approach used to divide the pseudotime expression spectrum is enumerated below:

qNSC-like to aNSC-early – Earliest pseudotime at which *Rpl4*, *Rpl32*, and *Egfr* are predominantly expressed.

aNSC-early to aNSC-mid – Earliest pseudotime at which *Ccna2*, *Cdk1*, and *Ccnb2* are predominantly expressed.

aNSC-mid to aNSC-late– Earliest pseudotime at which *Dlx1* and *Dlx2* are predominantly expressed.

aNSC-late to NPC-like – Earliest pseudotime at which *Nrxn3*, *Dlx6as1*, and *Dcx* are predominantly expressed.

Differential expression between the putative groups was conducted using the R package SCDE v1.2.1 (Kharchenko et al., 2014) and genes were ranked by Z-score for differential expression between groups. Pathway enrichment was performed on ranked lists using GSEA, using GO Biological Process and lists related neuroepithelial cell identity (Lein et al., 2007) lists.

Calculation of correlation between individual genes and generation of carpet plots

To assess the correlation and mutual exclusivity of the expression of particular genes in the aNSC-mid and aNSC-late populations, pairwise correlations between individual genes were calculated. Briefly, all cells that had been classified as aNSC-mid or aNSC-late cells in Figure 3 were used, and the Spearman correlation between genes of interest (*Atp1a2*, *Ntsr2*, *Gja1*, *Jag1*, *Fgfr3*, *Dlx1*, *Dlx2*) was calculated. The correlation (spearman rho) between individual genes was represented in a heatmap.

Diffusion mapping

Diffusion mapping was performed using the Destiny R package (Haghverdi et al., 2015) using. Most variable genes were selected using the Seurat R package (Satija et al., 2015), using the 2500 most variable detected genes, as calculated by the function logVarDivMean().

Validation of correlation modules in aNSC-mid and aNSC-late cells using HiSeq data

To validate that the mutually exclusive expression of markers of astrocytes/self-renewal and markers of neurogenesis in aNSC-mid and aNSC-late cells, the carpet plot with genes of interest (*Atp1a2*, *Ntsr2*, *Gja1*, *Jag1*, *Fgfr3*, *Dlx1*, *Dlx2*) was repeated using the data from the subset of aNSC-mid and aNSC-late cells that were sequenced as HiSeq. The Spearman correlation between genes of interest (*Atp1a2*, *Ntsr2*, *Gja1*, *Jag1*, *Fgfr3*, *Dlx1*, *Dlx2*) was calculated and correlation (spearman rho) between individual genes was represented in a heatmap.

Identification of additional genes associated with correlation modules observed in aNSC-mid and aNSC-late cells

To identify additional genes that were correlated with genes enriched in aNSC-mid cells and genes enriched in aNSC-late cells, all genes were ranked by the sum of their spearman correlation with genes associated with aNSC-mid cells (*Atp1a2*, *Ntsr2*, *Gja1*, *Jag1*, *Fgfr3),* or genes associated with aNSC-late cells (*Dlx1*, *Dlx2*). The top 10 genes correlated with aNSC-mid genes and the top 10 genes associated with aNSC-late cells were plotted in a carpet plot as described above.

Sorting subpopulations of aNSCs based on level of GFAP-GFP expression

Using transgenic mice expressing GFP under the GFAP promoter (GFAP-GFP), we divided the cells that were positive for GFAP-GFP into three groups based on the level of GFP fluorescence. aNSCs (PROM1$^+$EGFR$^+$) were sorted from each of these groups, and were negatively selected for markers of endothelial cells (CD31) and markers of hematopoietic cells (CD45). Additionally, NPCs (GFAP-GFP$^-$EGFR$^+$) were also sorted. Briefly, SVZ cells were isolated and pooled from 2 GFAP-GFP mice as described above. 4 groups of cells (GFAP-high aNSCs, GFAP-mid aNSCs, GFAP-low aNSCs, and NPCs) were sorted and subjected to downstream analyses. For this, 100 cells were directly sorted from each cell population directly into Cells-to-cDNA II lysis buffer (Ambion). In some cases, due to low numbers of aNSCs in this gate, fewer than 100 GFAP-high aNSCs were sorted for each replicate. However, all expression values were ultimately normalized to β-Actin expression. cDNA preparation was conducted as described in the Cells-to-cDNA II protocol (Ambion). Briefly, cells were incubated at 75°C for 10 minutes to stop lysis. Subsequently, the lysed cells were incubated with DNAse at 37°C for 15 minutes followed by heat inactivation at 75°C for 5 minutes to remove genomic DNA. Reverse transcription was performed at 42°C for 1 hour followed by heat inactivation at 92°C for 10 minutes. RT-qPCR primers were designed to span exons, and RT-qPCR reactions were conducted using iTaq qPCR Master Mix (Bio-Rad). In RT-qPCR reactions, primers were used at a concentration of 1μM, and template cDNA was used at a 1:10 dilution. RT-qPCR measurements were performed on a CFX96 Real-Time System (Bio-Rad), and thermo-cycling parameters used were those recommended in the iTaq Master Mix documentation (Bio-Rad). For quantification, Ct values were normalized to B-Actin for all samples. Four independent experiments were conducted on different days, in each experiment multiple replicates of cells were sorted and the mean gene expression of all replicates in each experiment was used for plotting and statistical analysis. No reverse transcriptase controls were run for each sample, and were verified to exhibit minimal or no amplification. RT-qPCR experiments confirmed that *GFP* transcript levels are indeed correlated with GFP fluorescence level (Figure S4H).

Projection of neurosphere single cells on PCAs from *in vivo* NSCs

High quality single neurosphere (NS) cells were normalized together with the *in vivo* single NSCs using TMM Normalization (as implemented by EdgeR) followed by calculation of FPKM. To assess the identity of NS single cells with respect to the activation and differentiation of *in vivo* NSCs, PCA was performed with all qNSCs, aNSCs, and NPCs using log2 transformed FPKM values for the consensus-ordering genes (Table S7A). NS single cells were projected on to the resulting PC space using PCA_predict() (Figure 5B).

Differential expression between *in vitro* neurosphere single cells and *in vivo* NSCs. Differential expression between all *in vivo* aNSCs and NPCs (cells that were deemed aNSC-early, aNSC-mid, aNSC-late, or NPC in Figure 3) and all *in vitro* neurosphere single cells was conducted using the R package SCDE v1.2.1 (Kharchenko et al., 2014). Genes were ranked by Z-score for differential expression between groups. Pathway enrichment was performed on ranked lists using GSEA, using the GSEA Hallmark pathways.

Generation of violin plots

Violin plots were generated by grouping cells according to their identities as determined by ordering all qNSCs, aNSCs, and NPCs by Monocle using the consensus ordering genes (Table S7A) (qNSC-like, aNSC-early, aNSC-mid, aNSC-late, NPC-like). NS cells were also included in these plots. Normalization of reads was conducted using all *in vivo* and NS cells using TMM normalization as implemented in the EdgeR package. Log2 transformed FPKM values are plotted for each cell.

Alignment and normalization of single cell NSC data from the Llorens-Bobadilla study

Raw FASTQ files for each of the single cells profiled in (Llorens-Bobadilla et al., 2015) were downloaded from the NCBI database. Reads from multiple sequencing lanes were aggregated and mapped to the mm9 genome using the same annotation and mapping parameters we used for the cells profiled in this study. We excluded cells exhibiting an oligodendrocyte-like molecular signature, as defined by their clustering with the oligodendrocytes identified in this study when the global transcriptomes were projected onto the principal component space generated by PCA analysis of the cells isolated in this study (data not shown). We also did not use the neuroblasts that were isolated by Llorens-Bobadilla as PSA-NCAM+ in any of our analyses.

Determination of 2500 most variable genes in *in vivo* NSCs

Most variable genes were selected using the Seurat R package (Satija et al., 2015), using the 2500 most variable detected genes, as calculated by the function logVarDivMean().

PCA projections of data from the Llorens-Bobadilla study

Because of large batch effects between the Llorens-Bobadilla study and our own (caused by different cDNA preparation chemistry, vastly different sequencing depths), conducting PCA for all cells together was inadequate. Thus we preprocessed and normalized the datasets separately and relied on the projection of the data from (Llorens-Bobadilla, et al., 2015) on to the principal component spaces generated from our data. Briefly we generated PCA spaces for all Astrocytes, qNSCs, aNSCs, and NPCs from our dataset using the 2500 most variable genes as defined by our dataset, or the consensus-ordering genes (Table S7A). Transcriptomes of single cells from the Llorens-Bobadilla study were subsequently projected onto these principal component spaces.

Monocle ordering of cells from the Llorens-Bobadilla study

Cells from (Llorens-Bobadilla, et al., 2015) were subjected to monocle ordering using the FPKM values for the consensus-ordering genes (Table S7A) identified through machine learning. Pseudotime values were linearly scaled between 0 and 100 and genes of interest were plotted with respect to pseudotime.

Calculating average pseudotime of expression (APE)

To rank genes by their propensity for expression early or late in a pseudotime continuum, genes were ranked by what we will henceforth refer to as Average Pseudotime of Expression (APE). The APE was calculated for each detected gene as the mean pseudotime (determined as indicated in the each figure) of all cells expressing each detected gene. For all pseudotime expression heatmaps, genes were vertically ordered by their APE as calculated for all qNSCs, aNSCs, and NPCs when ordered by Monocle using the consensus-ordering genes as determined by machine learning (Table S7A). For the APE correlation plots, genes were ranked by their APE for each single cell study, using pseudotime values obtained as indicated in the corresponding figure.

Generation of pseudotime expression heatmaps for cells from our study

To compare the global expression dynamics along the activation and differentiation trajectory of various single cell studies, we generated heatmaps representing the expression of all genes with respect to pseudotime for the various single cell studies analyzed in this study. First, we generated a pseudotime expression heatmap for our dataset. Briefly, the pseudotime continuum obtained by ordering all qNSCs, aNSCs, and NPCs by Monocle using the consensus-ordering genes obtained by machine learning (Table S7A) was split into 10 intervals, each containing an equal number of cells, and the mean expression value for each gene and pseudotime interval was calculated. Expression values were log2 transformed and linearly normalized between 0 and 1. The genes were vertically ordered by their APE as calculated by the pseudotime spectrum generated when qNSCs, aNSCs, and NPCs presented in the current study are ordered by Monocle using the consensus-ordering genes (Table S7A).

Generating pseudotime expression heatmaps for the Llorens-Bobadilla study

All NSCs and TAPs from Llorens-Bobadilla, et al. (Llorens-Bobadilla et al., 2015) were ordered by Monocle using the consensus-ordering genes (Table S7A). The resulting pseudotime continuum was divided into 10 intervals, each containing an equal number of cells, and the mean expression value for each gene and pseudotime interval was calculated. The genes were vertically ordered by their APE as calculated using qNSCs, aNSCs, and NPCs in our study when ordered by Monocle using the consensus-ordering genes (Table S7A). To ensure that the alignment and normalization approach was not contributing to the similarities between the pseudotime expression heatmaps, an identical analysis was conducted with FPKM values provided by the authors.

Generation of pseudotime expression heatmaps for activating hippocampal NSCs

We generated pseudotime expression heatmaps using hippocampal NSCs (Shin et al., 2015). FPKMs and pseudotime values (calculated by the Waterfall algorithm) used were those provided by the author. The pseudotime continuum provided by the authors was divided into 10 intervals, each containing an equal number of cells, and the mean expression value for each gene and pseudotime interval was calculated. The genes were vertically ordered by their APE as calculated using qNSCs, aNSCs, and NPCs in our study when ordered by Monocle using the consensus-ordering genes (Table S7A).

We also performed Monocle ordering on the single hippocampal cells using the consensus-ordering genes (Table S7A) to determine if this affected the appearance of the pseudotime expression heatmap. Briefly, all single hippocampal cells were ordered with Monocle using the consensus-ordering genes (Table S7A). The resulting pseudotime continuum was divided into 10 intervals, each containing an equal number of cells, and the mean expression value for each gene and pseudotime interval was calculated. The genes were vertically ordered by their APE as calculated using qNSCs, aNSCs, and NPCs in our study when ordered by Monocle using the consensus-ordering genes (Table S7A).

Generation of pseudotime expression heatmaps for differentiating human myoblasts

To determine the specificity of the pseudotime ordering spectra generated by ordering single NSCs, we also generated pseudotime expression heatmaps using differentiating myoblasts (Trapnell, et al., 2015). FPKMs and pseudotime values used were those provided by the author (determined by Monocle to characterize process of differentiation). We inverted the pseudotime provided by the authors (to align the proliferative and non-proliferative cells with the representation for qNSCs and aNSCs). The pseudotime continuum provided by the authors was divided into 10 intervals, each containing an equal number of cells, and the mean expression value for each gene and each pseudotime interval was calculated. The genes were vertically ordered by their APE as calculated using qNSCs, aNSCs, and NPCs in our study when ordered by Monocle using the consensus-ordering genes (Table S7A).

We also performed Monocle ordering on the differentiating myoblasts using the consensus-ordering genes (Table S7A) to determine if this affected the appearance of the pseudotime expression heatmap (Figure S7C). Briefly all single hippocampal cells were ordered with Monocle using the consensus-ordering genes (Table S7A). The resulting pseudotime continuum was divided into 10 intervals, each containing an equal number of cells, and the mean expression value for each gene and each pseudotime interval was calculated. The genes were vertically ordered by their APE as calculated using qNSCs, aNSCs, and NPCs in our study when ordered by Monocle using the consensus-ordering genes (Table S7A).

Average pseudotime of expression (APE) gene rank correlation

To generate correlation plots comparing the APE gene rankings for each of the datasets, APE (defined above) was calculated for all genes detected in our dataset when all single qNSCs, aNSCs, and NPCs were ordered by the consensus-ordering genes (Table S7A). We then independently ranked all genes by APE using the expression for single cells from either (1) (Llorens-Bobadilla et al.,(Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015) (Llorens-Bobadilla et al., 2015) (Llorens-Bobadilla et al., 2015) (Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015)(Llorens-Bobadilla et al., 2015) 2015) (2) (Shin et al., 2015) from hippocampal NSCs or (3) (Trapnell et al., 2014) from differentiating. Psuedotime values used to calculate APE values for these three studies was either provided by the authors, or calculated by ordering by Monocle using the consensus-ordering genes obtained through machine learning (Table S7A) as indicated in the associated figure. Genes were omitted from this analysis if they were not detected in any of the single cells in the particular dataset being analyzed, were not present in the annotations provided by the authors (for myoblasts from (Trapnell et al., 2014)), or if they were not considered detected (10 counts in 5 cells) in our dataset. We then used the smoothScatter() default R function to plot the ranks of each gene by their APE in our dataset (when all qNSCs, aNSCs, and NPCs are ordered by Monocle using the consensus-ordering genes) relative to their APE ranks in each of the datasets enumerated above.

Supplementary Table Legends

Table S1 – Related to Figure 1 – Sequencing parameters for all single cells
Number of total reads, uniquely mapping reads, and percent uniquely mapping reads for each single cell sequenced. A selection of aNSCs was sequenced both on MiSeq and two lanes of HiSeq. For these cells "_hiseq_run1" or "_hiseq_run2" are appended to the end of the cell name. The designator 1g indicates that the cell was a live single cell on the C1 chip. Cell subgroup refers to cell subgroup identity as defined in Figure 3, except for "Astrocytes", which refers to the FACS-sorting identity and "Oligodendrocyte-like cell" and "Outlier", which refers to cellular identity as defined in Figure 1.

Table S2 – Related to Figure 1 – Raw counts for all live single cells profiled
Raw counts for each single cell for all genes. A selection of aNSCs was sequenced both on MiSeq and two lanes of HiSeq. For these cells "_hiseq_run1" or "_hiseq_run2" are appended to the end of the cell name.

Table S3 – Related to Figure 1 – Normalized counts for *in vivo* single cells sequenced by Illumina MiSeq with outlying cells removed
All cells sequenced on MiSeq with contaminating oligodendrocytes and outlying single cells removed. All detected genes (>10 counts in ≥5 single cells), normalized by TMM normalization as implemented by EdgeR.

Table S4 – Related to Figure 1 – Normalized counts for *in vivo* NSCs and *in vitro* neurosphere derived single cells
All cells sequenced on MiSeq including oligodendrocytes and outlying cells as defined in Figure 1, as well as single neurosphere derived cells. All detected genes (>10 counts in ≥5 single cells), normalized by TMM normalization as implemented by EdgeR.

Table S5 – Related to Figure 2 – Differential expression of all detected genes between astrocytes and qNSCs
All detected genes ranked by Z-score for differential expression between astrocytes and qNSCs. No genes reach statistical significance. Values reported are the most likely fold expression difference, Z-score, and adjusted Z-score as calculated by SCDE (Kharchenko et al., 2014).

Table S6 – Related to Figure 2 – Complete list of top 100 genes from machine learning models.
The 100 most important genes from each machine learning model generated in Figure 2.

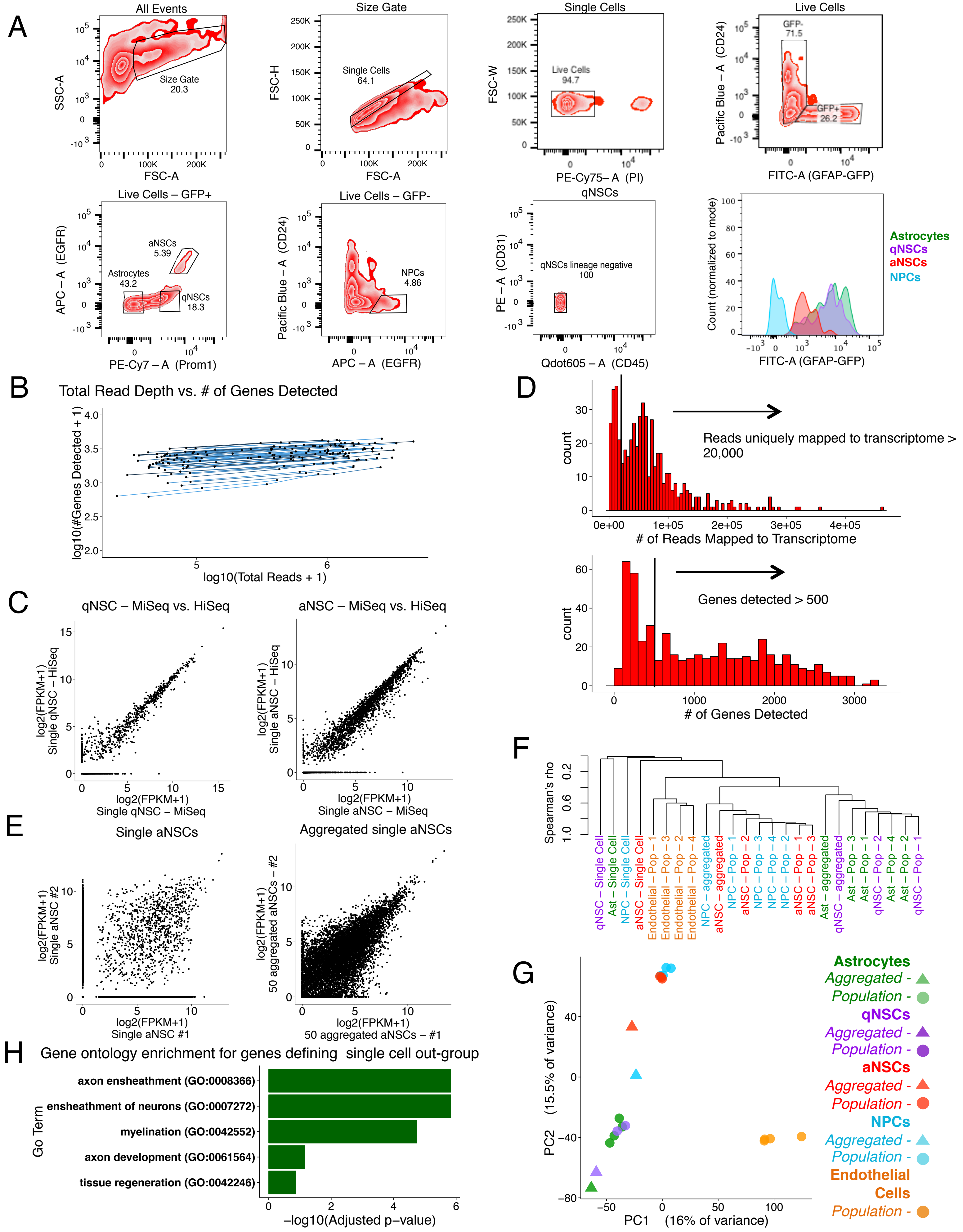
Table S7 – Related to Figure 2 – Consensus-ordering gene lists defined by machine learning
(A) Consensus-ordering genes defined by presence in the top 100 most important features in at least 50% of the models.
(B) Less stringent consensus-ordering genes defined by presence in the top 100 most important features in at least 25% of the models.

Table S8 – Related to Figure 3 – Differential expression of all detected genes between intermediate NSC states
All detected genes ranked by Z-score for differential expression between the intermediate states identified along the course of activation and differentiation described in Figure 3. Values reported are the most likely fold expression difference, Z-score, and adjusted Z-score as calculated by SCDE (Kharchenko et al., 2014). The title of the excel tab indicates the cellular state transition for which the differential expression parameters are reported.

Table S9 – Related to Figure 5 – Differential expression of all detected genes between *in vivo* NSCs and *in vitro* cultured neurosphere single cells.
All detected genes ranked by Z-score for differential expression between *in vitro* neurosphere single cells and *in vivo* cells defined as aNSCs or NPCs in Figure 3. Values reported are the most likely fold expression difference, Z-score, and adjusted Z-score as calculated by SCDE (Kharchenko et al., 2014).

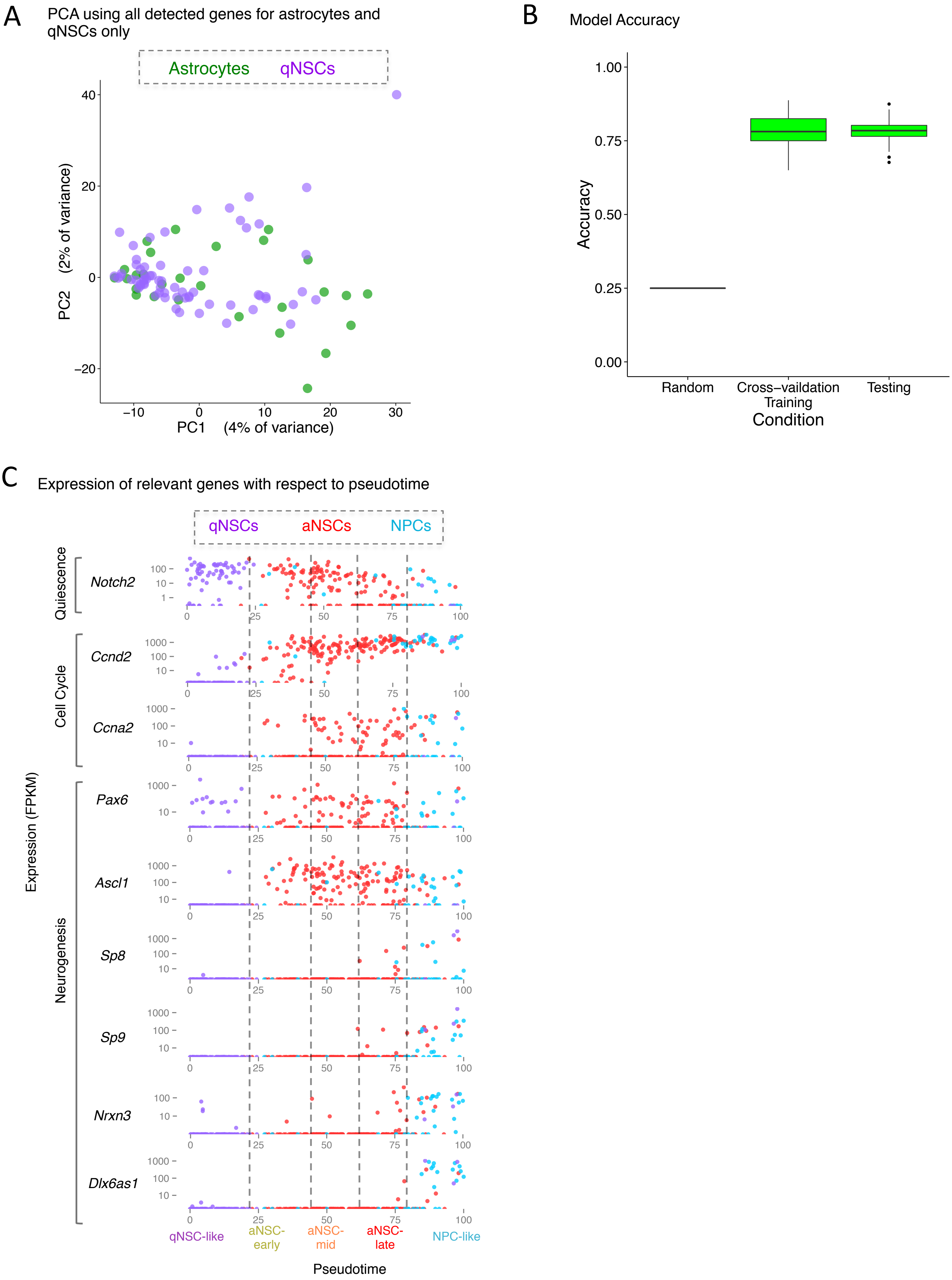
Supplementary Figures

Figure S1


A
All Events
SSC-A
FSC-A
Size Gate
20.3
Size Gate
FSC-H
Single Cells
64.1
Single Cells
FSC-W
Live Cells
94.7
PE-Cy75– A (PI)
Live Cells
Pacific Blue – A (CD24)
GFP-
71.5
GFP+
26.2
FITC-A (GFAP-GFP)
Live Cells – GFP+
APC – A (EGFR)
aNSCs
5.39
Astrocytes
43.2
qNSCs
18.3
PE-Cy7 – A (Prom1)
Live Cells – GFP-
Pacific Blue – A (CD24)
NPCs
4.86
APC – A (EGFR)
qNSCs
PE – A (CD31)
qNSCs lineage negative
100
Qdot605 – A (CD45)
Count (normalized to mode)
FITC-A (GFAP-GFP)
Astrocytes
qNSCs
aNSCs
NPCs
B
Total Read Depth vs. # of Genes Detected
log10(#Genes Detected + 1)
log10(Total Reads + 1)
C
qNSC – MiSeq vs. HiSeq
log2(FPKM+1)
Single qNSC – HiSeq
log2(FPKM+1)
Single qNSC – MiSeq
aNSC – MiSeq vs. HiSeq
log2(FPKM+1)
Single aNSC – HiSeq
log2(FPKM+1)
Single aNSC – MiSeq
D
count
Reads uniquely mapped to transcriptome > 20,000
# of Reads Mapped to Transcriptome
count
Genes detected > 500
# of Genes Detected
E
Single aNSCs
log2(FPKM+1)
Single aNSC #2
log2(FPKM+1)
Single aNSC #1
Aggregated single aNSCs
log2(FPKM+1)
50 aggregated aNSCs – #2
log2(FPKM+1)
50 aggregated aNSCs – #1
F
Spearman's rho
qNSC – Single Cell
Ast – Single Cell
NPC – Single Cell
aNSC – Single Cell
Endothelial – Pop – 1
Endothelial – Pop – 3
Endothelial – Pop – 2
Endothelial – Pop – 4
NPC – aggregated
aNSC – aggregated
NPC – Pop – 1
aNSC – Pop – 2
NPC – Pop – 3
NPC – Pop – 4
NPC – Pop – 2
aNSC – Pop – 1
aNSC – Pop – 3
Ast – aggregated
qNSC – aggregated
Ast – Pop – 3
Ast – Pop – 1
qNSC – Pop – 2
Ast – Pop – 4
Ast – Pop – 2
qNSC – Pop – 1
G
PC2 (15.5% of variance)
PC1 (16% of variance)
Astrocytes
Aggregated -
Population -
qNSCs
Aggregated -
Population -
aNSCs
Aggregated -
Population -
NPCs
Aggregated -
Population -
Endothelial Cells
Population -
H
Gene ontology enrichment for genes defining single cell out-group
Go Term
axon ensheathment (GO:0008366)
ensheathment of neurons (GO:0007272)
myelination (GO:0042552)
axon development (GO:0061564)
tissue regeneration (GO:0042246)
−log10(Adjusted p−value)

Supplementary Figure 1 – Related to Figure 1. Quality control and analysis for single cell RNA-seq of 329 cells from four populations of FACS-purified cells from the SVZs of adult mice
(A) Zebra-plots showing FACS gates for the isolation of astrocytes, qNSCs, aNSCs, and NPCs from the SVZs of adult mice. Plot title indicates parent gate. Negative gating for CD45 (hematopoietic cells) and CD31 (endothelial cells) is shown for astrocytes and qNSCs, but was also performed for aNSCs and NPCs.
(B) A large fraction of transcriptome complexity is captured in low-coverage sequencing of single cells. Number of genes detected in aNSCs plotted as a function of total reads obtained for a single library subjected to sequencing on Illumina MiSeq and subsequently two lanes of Illumina HiSeq2000. Points represent the number of genes detected and total reads for the aggregated counts from (one MiSeq run), (one MiSeq + one HiSeq run), and (one MiSeq + two HiSeq runs) respectively from left to right. Lines represent one single cell library.
(C) *(Left)* Scatter plot of TMM-normalized expression values of each detected gene (>10 counts in ≥5 high-quality single cells) for a representative qNSC sequenced on MiSeq or two lanes of HiSeq. *(Right)* Scatter plot of TMM-normalized expression values of each detected gene (>10 counts in ≥5 high-quality single cells) for a representative aNSC sequenced on MiSeq or two lanes of HiSeq.
(D) Filtering of live single cells based on coverage and gene detection. Histograms show reads mapped and genes detected for each live single cell profiled. Vertical lines indicate cutoffs above which cells were used for analysis (>20,000 reads mapped to transcriptome, and >500 unique genes detected).
(E) *(Left)* Scatter plot of TMM-normalized expression values of each detected gene (>10 counts in ≥5 high-quality single cells) for two randomly selected aNSCs (Pearson correlation = 0.60, p-value < 2.2e-16). *(Right)* scatter plot of TMM-normalized expression values of each detected gene (>10 counts in ≥5 high-quality single cells) for two randomly selected populations of 50 single aNSCs, aggregated by summing the counts for each gene across cells (Pearson correlation = 0.93, p-value < 2.2e-16).
(F,G) RNA-seq data from aggregated single cells for each of the four populations (astrocytes, qNSCs, aNSCs, and NPCs) cluster with RNA-seq data previously obtained from these populations.
(F) Hierarchal clustering using log2(FPKM+1) expression values of all detected genes; branch position indicates Spearman rho. Scale indicated on the top left. Single Cell: randomly selected single cells. Aggregated: Aggregated single cell RNA-seq. Pop: Population RNA-seq (with number indicating replicate); Ast: astrocytes. Endothelial cell RNA-seq data provide an outgroup.
(G) Principal component analysis (PCA). The first two principal components (PC) are plotted. Performed using log2(FPKM+1) expression values of all detected genes. Key on right. Aggregated: Aggregated single cell RNA-seq. Population: Population RNA-seq.
(H) Outlying single cells in Figure 1B exhibit an oligodendrocyte-like gene signature. Bar plot showing significance for gene set enrichments [-log10(Adjusted P-Value)] calculated by GoRilla for all genes with PC loadings < -0.05 on PC2 in Figure 1B.

Figure S2


A
PCA using all detected genes for astrocytes and qNSCs only
Astrocytes
qNSCs
40
20
0
−20
PC2 (2% of variance)
−10
0
10
20
30
PC1 (4% of variance)
B
Model Accuracy
1.00
0.75
0.50
0.25
0.00
Accuracy
Random
Cross−vaildation Training
Testing
Condition
C
Expression of relevant genes with respect to pseudotime
qNSCs
aNSCs
NPCs
Quiescence
Notch2
Cell Cycle
Ccnd2
Ccna2
Neurogenesis
Pax6
Ascl1
Sp8
Sp9
Nrxn3
Dlx6as1
Expression (FPKM)
1000
100
10
1
0
25
50
75
100
qNSC-like
aNSC-early
aNSC-mid
aNSC-late
NPC-like
Pseudotime

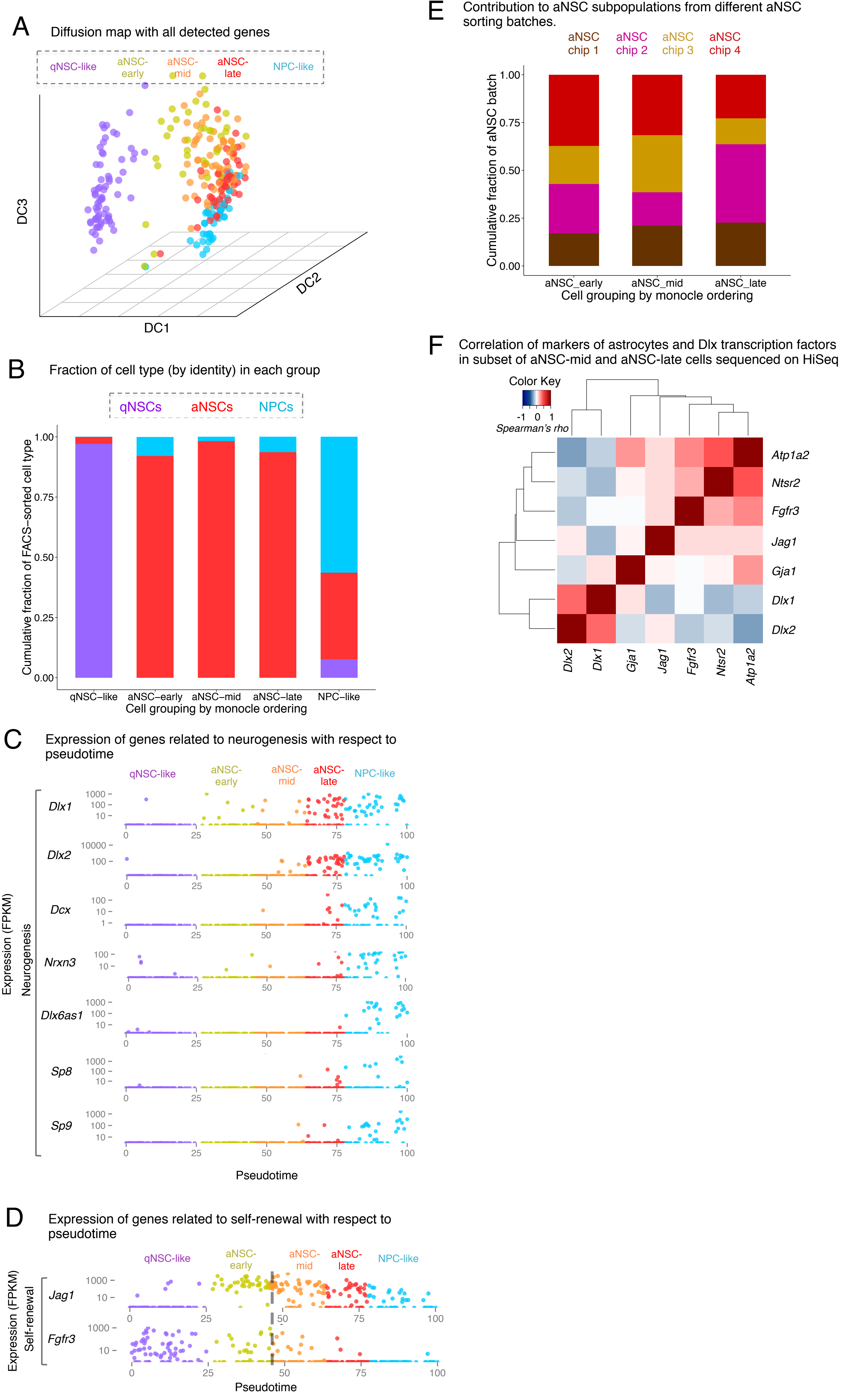
Supplementary Figure 2 – Related to Figure 2. Ordering of single cells from populations of qNSCs, aNSCs, and NPCs reveals transcriptional dynamics through activation differentiation and suggests heterogeneity and intermediary states

(A) PCA on all Astrocytes and qNSCs. Performed using log2(FPKM+1) expression values of all detected genes (>10 counts in ≥5 cells).

(B) Accuracy of machine learning models (GBM) for the resampled training set (used for construction of model), and testing set (not used for construction of model), as compared to the theoretical accuracy if groups were assigned randomly.

(C) Monocle ordering using consensus-ordering genes from machine learning models suggests intermediary states along the processes of activation and differentiation. Expression (FPKM) of key genes related to quiescence (*Notch2*), cell cycle (*Ccnd2*, *Ccna2*), neuronal differentiation (*Ascl1*, *Pax6, Sp8*, *Sp9*, *Nrxn3*, *Dlx6as1*) in each cell is plotted with respect to pseudotime produced by Monocle when all qNSCs, aNSCs, and NPCs are ordered using the consensus-ordering genes from machine learning models (Table S7A). Cells are color-coded by their FACS-sorting identity (indicated on top). Bottom: name of the intermediary states defined by Monocle ordering using consensus-ordering genes from machine learning (qNSC-like, aNSC early, aNSC mid, aNSC late, and NPC-like).

Figure S3



A Diffusion map with all detected genes

B Fraction of cell type (by identity) in each group

C Expression of genes related to neurogenesis with respect to pseudotime

D Expression of genes related to self-renewal with respect to pseudotime

E Contribution to aNSC subpopulations from different aNSC sorting batches.

F Correlation of markers of astrocytes and Dlx transcription factors in subset of aNSC-mid and aNSC-late cells sequenced on HiSeq

Supplementary Figure 3 – Related to Figure 3. Activated NSCs can be divided into specific subpopulations, defined by the expression of markers, along the spectrum of activation and differentiation

(A) Diffusion map using all detected genes in the dataset for all qNSCs, aNSCs, and NPCs. Cells are colored by the identity of the intermediate states defined from Figure 2G: qNSC-like, aNSC-early, aNSC-mid, aNSC-late, and NPC-like.

(B) Cumulative fraction of each sorted population (qNSC, aNSC, or NPC), in each of the intermediate states defined in Figure 2G. Bar indicates intermediate state defined in Figure 2G, color indicates FACS sorting identity (qNSC, aNSC, NPC).

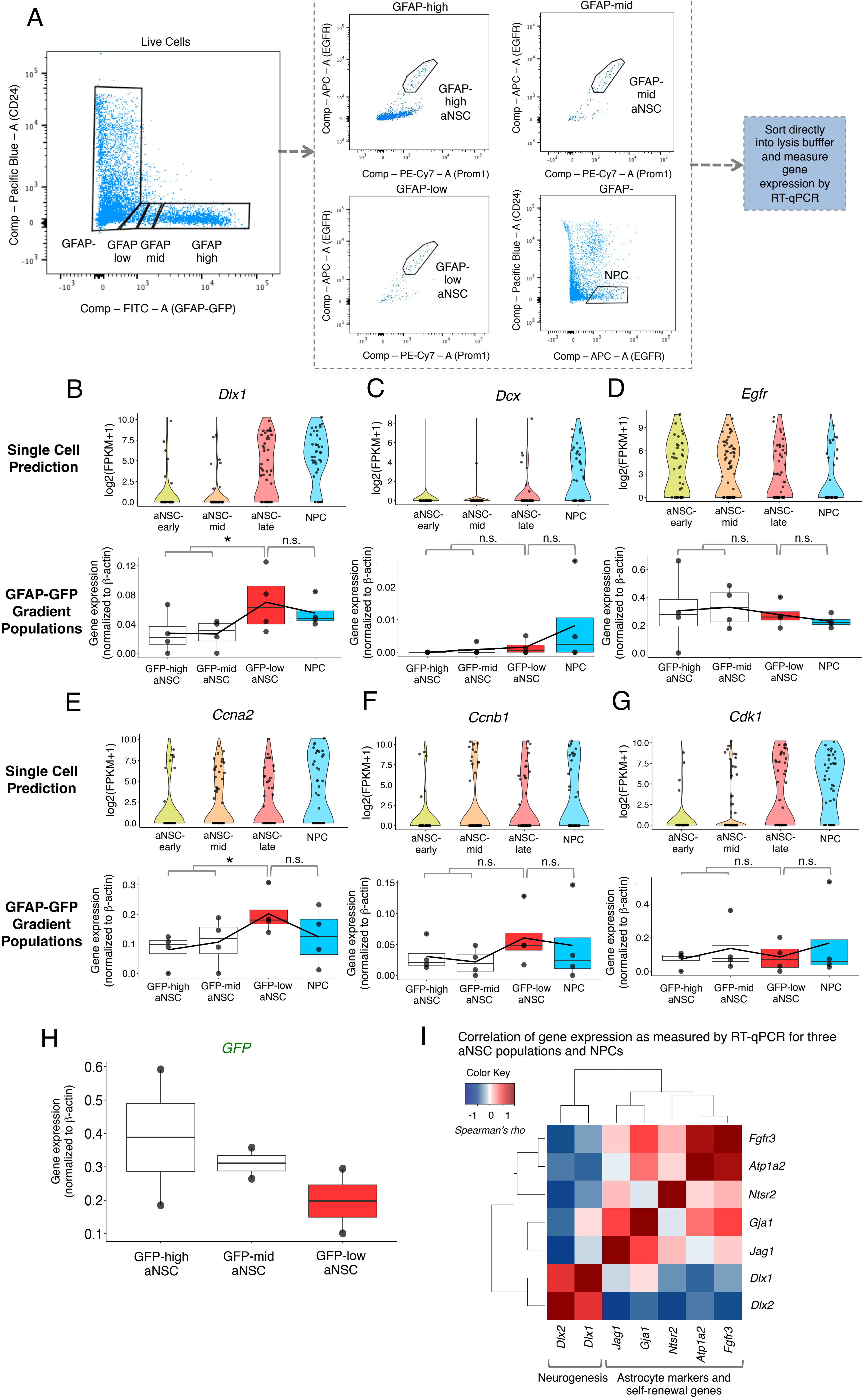
(C) Expression (FPKM) of regulators of neurogenesis (*Dlx1*, *Dlx2*, *Dcx*, *Nrxn3*, *Dlx6as1*, *Sp8*, *Sp9*) in each cell plotted as a function of pseudotime generated by Monocle when all qNSCs, aNSCs, and NPCs are ordered using the consensus-ordering genes identified by machine learning (Table S7A). Cells are colored by the identity of the intermediate states defined from Figure 2G: qNSC-like, aNSC-early, aNSC-mid, aNSC-late, and NPC-like.

(D) Expression (FPKM) of select regulators of self-renewal (*Jag1*, *Fgfr3*) in each cell plotted as a function of pseudotime generated by Monocle when all qNSCs, aNSCs, and NPCs are ordered using the consensus-ordering genes identified by machine learning (Table S7A). Cells are colored by the identity of the intermediate states defined from Figure 2G: qNSC-like, aNSC-early, aNSC-mid, aNSC-late, and NPC-like.

(E) aNSC batches do not define the distinctions between subpopulations of aNSCs. Cumulative fraction of each chip of aNSCs (Chips 1-4), in each of the intermediate aNSC states defined in Figure 2G. Bar indicates intermediate state defined in Figure 2G, color indicates chip number.

(F) Markers of astrocytes (*Atp1a2*, *Ntsr2*, *Gja1*) and mediators of self-renewal (*Jag1*, *Fgfr3*) are correlated with each other and are anticorrelated with early markers of neuronal differentiation (*Dlx1*, *Dlx2*) in aNSC-mid and aNSC-late cells sequenced to greater depth on HiSeq. Carpet plot showing correlation (Spearman's rho) between individual genes in the subset aNSC-mid and aNSC-late cells sequenced on HiSeq. Color of box indicates correlation (Spearman's rho) between a given gene pair (scale on upper left).

Figure S4


A
Live Cells
Comp – Pacific Blue – A (CD24)
Comp – FITC – A (GFAP-GFP)
GFAP-
GFAP low
GFAP mid
GFAP high
GFAP-high
GFAP-high aNSC
GFAP-mid
GFAP-mid aNSC
GFAP-low
GFAP-low aNSC
GFAP-
NPC
Comp – APC – A (EGFR)
Comp – PE-Cy7 – A (Prom1)
Comp – Pacific Blue – A (CD24)
Comp – APC – A (EGFR)
Sort directly into lysis bufffer and measure gene expression by RT-qPCR
B
Dlx1
C
Dcx
D
Egfr
Single Cell Prediction
log2(FPKM+1)
aNSC-early
aNSC-mid
aNSC-late
NPC
GFAP-GFP Gradient Populations
Gene expression (normalized to β-actin)
*
n.s.
GFP-high aNSC
GFP-mid aNSC
GFP-low aNSC
NPC
E
Ccna2
F
Ccnb1
G
Cdk1
Single Cell Prediction
GFAP-GFP Gradient Populations
H
GFP
Gene expression (normalized to β-actin)
GFP-high aNSC
GFP-mid aNSC
GFP-low aNSC
I
Correlation of gene expression as measured by RT-qPCR for three aNSC populations and NPCs
Color Key
-1 0 1
Spearman's rho
Fgfr3
Atp1a2
Ntsr2
Gja1
Jag1
Dlx1
Dlx2
Dlx2
Dlx1
Jag1
Gja1
Ntsr2
Atp1a2
Fgfr3
Neurogenesis
Astrocyte markers and self-renewal genes

Supplementary Figure 4 – Related to Figure 4. Dividing aNSCs by level of GFAP-GFP expression can enrich for the aNSC-late subpopulation.

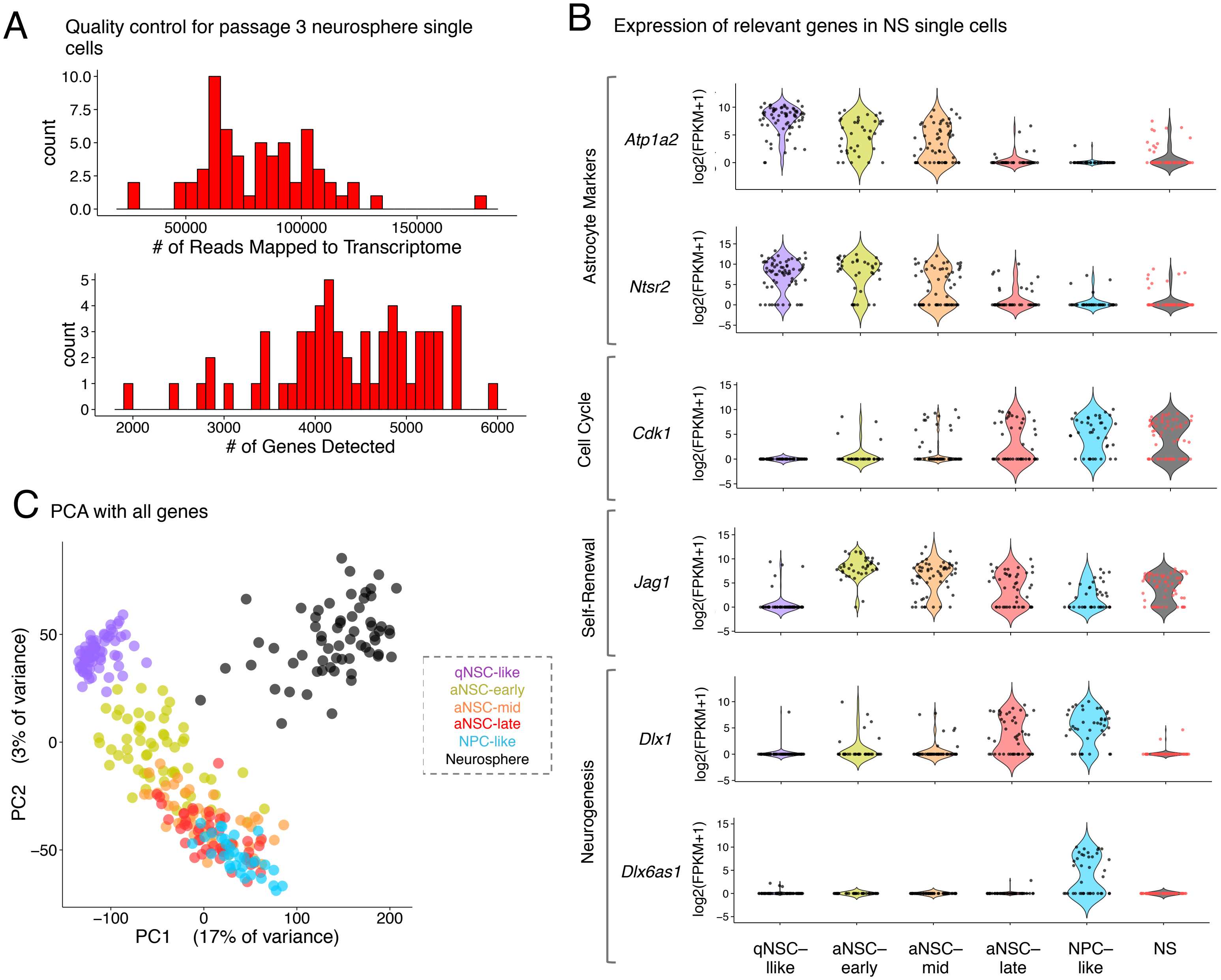
(A) FACS sorting scheme to isolate subpopulations of aNSCs expressing various levels of GFAP-GFP. All gates are negatively selected for cells expressing CD31 (endothelial) and CD45 (hematopoietic) lineage markers.

(B-G) Patterns of gene expression in subpopulations of NSCs sorted by level of GFAP-GFP fluorescence are similar to those predicted by single cell analyses. *(upper)* Gene expression in single cells grouped by molecular subtype as defined in Figure 3. Gene expression expressed as log2(FPKM+1). *(lower)* Gene expression measured by RT-qPCR in subpopulations of aNSCs divided by their level of GFAP-GFP expression (GFAP-GFP high aNSC, GFAP-GFP mid aNSC, GFAP-GFP low aNSC) and NPCs. Gene expression was normalized to the expression of β-actin. Results from independent experiments are shown as individual dots overlaid with a boxplot in which the center bar represents the median and the extent of the box represents the first and third quartile. Whiskers extend to data points within 1.5 times the inter-quartile range. Line represents mean expression for each cell population. The subpopulation of aNSCs with low levels of GFP and the population of NPC with no GFP were labeled in red and blue, respectively, because the markers validate these populations. The populations with high and mid levels of GFP were left white because the cell cycle markers (shown in Supplementary Figure) did not validate these subpopulations. Tests for significance were performed between the GFAP-GFP high and GFAP-GFP mid aNSCs versus the GFAP-GFP low aNSCs, and between the GFAP-GFP low aNSCs and NPCs. P-value was obtained by a one-sided Wilcoxon signed-rank test. * $p \leq 0.05$

(H) Gene expression of *GFP* transcript measured by RT-qPCR in subpopulations of aNSCs divided by their level of GFAP-GFP expression (GFAP-GFP high aNSC, GFAP-GFP mid aNSC, GFAP-GFP low aNSC) and NPCs. Gene expression was normalized to the expression of β-actin.

(I) Correlation between expression of key markers of NSCs and neurogenesis in aNSC populations divided by GFAP-GFP fluorescence resembles that observed in single aNSCs and NPCs. Correlation between markers of astrocytes (*Atp1a2*, *Ntsr2*, *Gja1*), mediators of self-renewal (*Jag1*, *Fgfr3*), and early markers of neuronal differentiation (*Dlx1*, *Dlx2*) in aNSC subpopulations with varying levels of GFAP-GFP fluorescence and NPCs. Carpet plot showing correlation (Spearman's rho) between individual genes in all aNSC-subpopulations and NPCs. Color of box indicates correlation (Spearman's rho) between a given gene pair (scale on upper left).

Figure S5


A
Quality control for passage 3 neurosphere single cells
count
10.0
7.5
5.0
2.5
0.0
50000
100000
150000
# of Reads Mapped to Transcriptome
count
5
4
3
2
1
0
2000
3000
4000
5000
6000
# of Genes Detected
B
Expression of relevant genes in NS single cells
Astrocyte Markers
Atp1a2
Ntsr2
Cell Cycle
Cdk1
Self-Renewal
Jag1
Neurogenesis
Dlx1
Dlx6as1
log2(FPKM+1)
15
10
5
0
−5
qNSC–
llike
aNSC–
early
aNSC–
mid
aNSC–
late
NPC–
like
NS
C
PCA with all genes
PC2 (3% of variance)
50
0
−50
−100
0
100
200
PC1 (17% of variance)
qNSC-like
aNSC-early
aNSC-mid
aNSC-late
NPC-like
Neurosphere

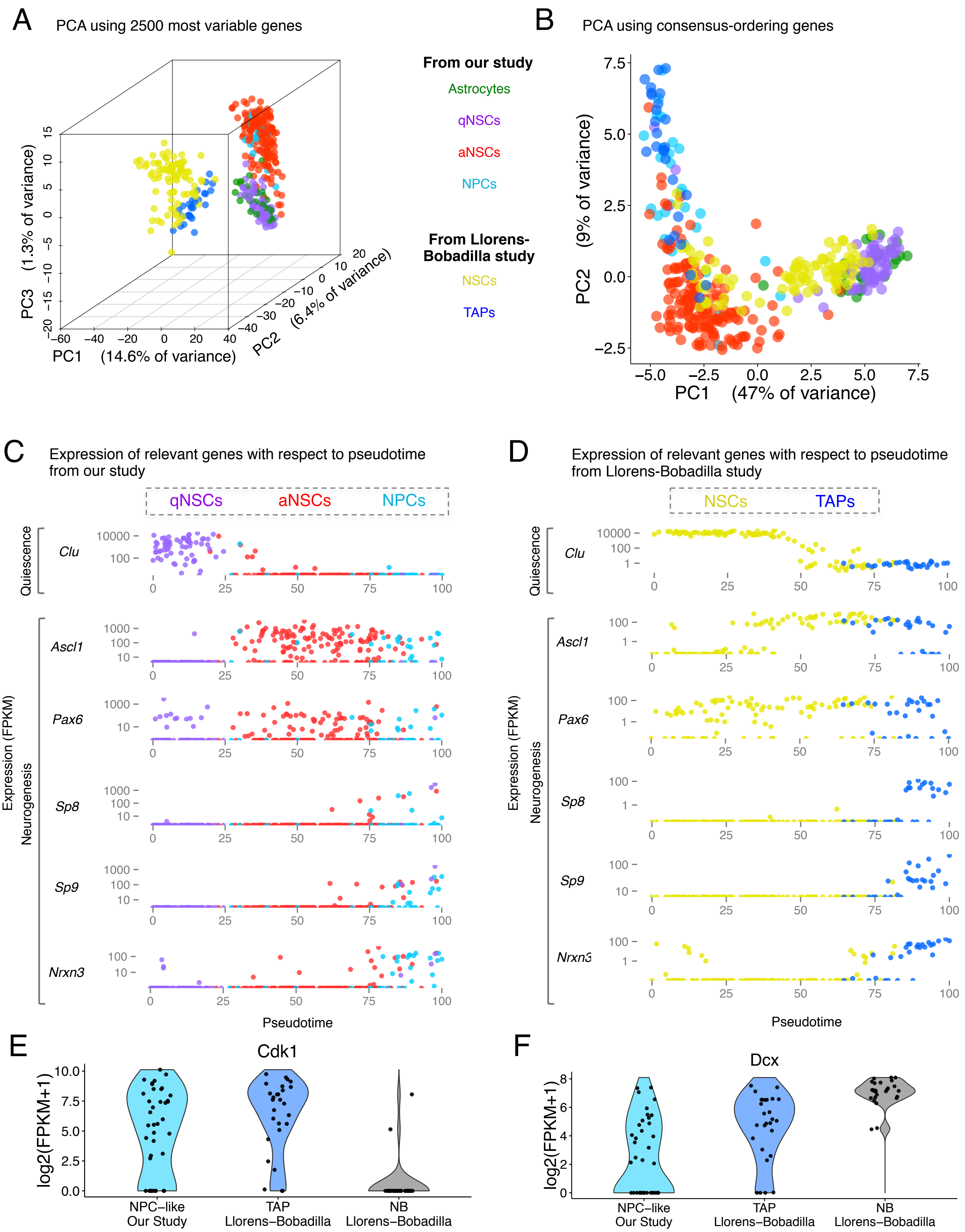
Supplementary Figure 5 – Related to Figure 5. In the spectrum of NSC activation and differentiation *in vivo, in vitro* cultured NSCs resemble aNSCs and exhibit a signature of inflammation.

(A) Histograms indicating number of reads mapping to transcriptome and unique genes detected for all neurosphere (NS) derived single cells. No live single cells had <20,000 reads mapping to transcriptome or <500 unique genes detected.

(B) Expression of genes associated with astrocytic identity, cell cycle, self-renewal, and neurogenesis in *in vitro* NS single cells and *in vivo* NSCs. Violin plots showing gene expression in the cellular states described in Figure 3 (qNSC-like, aNSC early, aNSC mid, aNSC late, and NPC-like) as well as in NS single cells.

(C) PCA with qNSCs, aNSCs, NPCs, and NS single cells using expression [log2(FPKM+1)] of all detected genes. Cells are colored by identity as described in Figure 3 (qNSC-like, aNSC-early, aNSC-mid, aNSC-late, and NPC-like), and NS single cells are represented in black.

Figure S6


A
PCA using 2500 most variable genes
PC1 (14.6% of variance)
PC2 (6.4% of variance)
PC3 (1.3% of variance)
−60 −40 −20 0 20 40
−40 −30 −20 −10 0 10 20
−20 −15 −10 −5 0 5 10 15
From our study
Astrocytes
qNSCs
aNSCs
NPCs
From Llorens-Bobadilla study
NSCs
TAPs
B
PCA using consensus-ordering genes
PC1 (47% of variance)
PC2 (9% of variance)
−5.0 −2.5 0.0 2.5 5.0 7.5
−2.5 0.0 2.5 5.0 7.5
C
Expression of relevant genes with respect to pseudotime from our study
qNSCs aNSCs NPCs
Quiescence
Clu
Neurogenesis
Ascl1
Pax6
Sp8
Sp9
Nrxn3
Expression (FPKM)
10000 1000 100 10
0 25 50 75 100
Pseudotime
D
Expression of relevant genes with respect to pseudotime from Llorens-Bobadilla study
NSCs TAPs
Quiescence
Clu
Neurogenesis
Ascl1
Pax6
Sp8
Sp9
Nrxn3
Expression (FPKM)
10000 100 10 1
0 25 50 75 100
Pseudotime
E
Cdk1
log2(FPKM+1)
0.0 2.5 5.0 7.5 10.0
NPC−like Our Study
TAP Llorens−Bobadilla
NB Llorens−Bobadilla
F
Dcx
log2(FPKM+1)
0 2 4 6 8
NPC−like Our Study
TAP Llorens−Bobadilla
NB Llorens−Bobadilla

Supplementary Figure 6 – Related to Figure 6. Transcriptomic comparison resolves single cell identities of NSCs and progeny from the SVZ, isolated using divergent sorting schemes

(A) NSCs and progeny from our study and the Llorens-Bobadilla study cluster independently by global PCA. PCA on all qNSCs, aNSCs, and NPCs from our study, and all NSCs and TAPs from the Llorens-Bobadilla study, using the expression [log2(FPKM+1)] of the 2500 most variable genes in our cells (qNSCs, aNSCs, and NPCs). Cells colored by FACS sorting identity indicated on right.

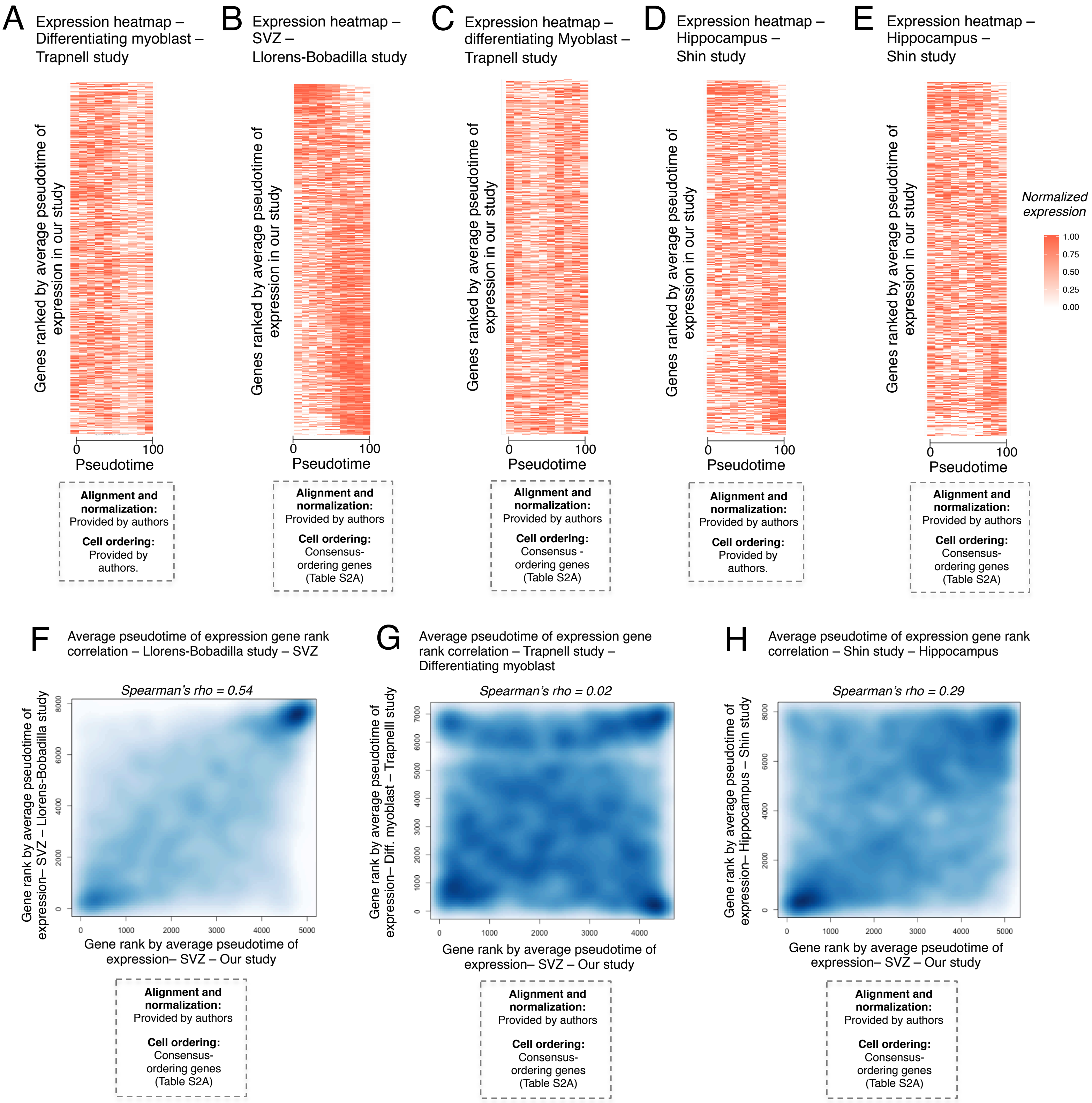
(B) NSCs and progeny from our study and the Llorens-Bobadilla study exhibit similar clustering behavior by PCA projection using genes that define the activation and differentiation of the NSC lineage. PCA on all qNSCs, aNSCs, and NPCs from our study, using the expression [log2(FPKM+1)] of the consensus-ordering genes identified by machine learning (Table S7A). All NSCs and TAPs from the Llorens-Bobadilla study are projected on to the resulting principal component space. Cells colored by FACS sorting identity indicated on left.

(C) Expression (FPKM) of key markers of activation and differentiation in each cell plotted as a function of pseudotime generated by Monocle ordering using the consensus-ordering genes identified by machine learning (Table S7A) for all qNSCs, aNSCs, and NPCs from our study. Cells colored by FACS sorting identity indicated on top.

(D) Expression (FPKM) of key markers of activation and differentiation in each cell plotted as a function of pseudotime generated by Monocle ordering using the consensus-ordering genes identified by machine learning (Table S7A) for the NSCs and TAPs analyzed in (Llorens-Bobadilla et al., 2015). Cells colored by FACS sorting identity indicated on top.

(E-F) Expression [log2(FPKM+1)] of *(E) Cdk1* and *(F) Dcx* in NPCs from our study, TAPs from Llorens-Bobadilla, et al. 2015, and neuroblasts from Llorens-Bobadilla, et al. 2015.

Figure S7



A Expression heatmap – Differentiating myoblast – Trapnell study

B Expression heatmap – SVZ – Llorens-Bobadilla study

C Expression heatmap – differentiating Myoblast – Trapnell study

D Expression heatmap – Hippocampus – Shin study

E Expression heatmap – Hippocampus – Shin study

F Average pseudotime of expression gene rank correlation – Llorens-Bobadilla study – SVZ

G Average pseudotime of expression gene rank correlation – Trapnell study – Differentiating myoblast

H Average pseudotime of expression gene rank correlation – Shin study – Hippocampus

A
B
C
D
E
Genes ranked by average pseudotime of expression in our study
Pseudotime
0
100
Normalized expression
1.00
0.75
0.50
0.25
0.00
Alignment and normalization: Provided by authors
Cell ordering: Provided by authors.
Cell ordering: Consensus-ordering genes (Table S2A)
Cell ordering: Consensus - ordering genes (Table S2A)
F
G
H
Spearman's rho = 0.54
Spearman's rho = 0.02
Spearman's rho = 0.29
Gene rank by average pseudotime of expression– SVZ – Llorens-Bobadilla study
Gene rank by average pseudotime of expression– Diff. myoblast – Trapnell study
Gene rank by average pseudotime of expression– Hippocampus – Shin study
Gene rank by average pseudotime of expression– SVZ – Our study
0
1000
2000
3000
4000
5000
6000
7000
8000

Supplementary Figure 7 – Related to Figure 7. Correlation of global pseudotime-dependent gene expression in various single cell datasets.

(A-E) Heatmaps representing the expression of all detected genes ranked by average pseudotime of expression (APE, see Supplemental Experimental Methods) defined by our study. Expression of each gene plotted as a function of pseudotime. Expression [log2(FPKM+1)] is linearly scaled between 0 (white) and 1 (red) for each gene.

(A) Expression data from the Trapnell study (differentiating myoblasts), pseudotime defined by authors using Monocle ordering.

(B) Expression (FPKM values provided by authors) from the Llorens-Bobadilla study (NSCs and TAPs), pseudotime defined by Monocle ordering using consensus-ordering genes identified by machine learning (Table S7A).

(C) Expression from the Trapnell study (differentiating myoblasts), pseudotime defined by Monocle ordering using consensus-ordering genes identified by machine learning (Table S7A).

(D) Expression data from the Shin study (hippocampal NSCs), pseudotime defined by authors using Waterfall.

(E) Expression from the Shin study (hippocampal NSCs), pseudotime defined by Monocle ordering using consensus-ordering genes identified by machine learning (Table S7A).

(F) Smooth scatter plot representing gene rankings by average pseudotime of expression (APE) in *(x-axis)* qNSCs, aNSCs, and NPCs from the current study ordered by Monocle using the consensus-ordering genes identified by machine learning (Table S7A) and *(y-axis)* NSCs and TAPs from the Llorens-Bobadilla study (FPKMs provided by author) ordered by Monocle using the consensus-ordering genes identified by machine learning (Table S7A) (Spearman's rho = 0.54, p-value < $2.2e^{-16}$). Intensity of blue shading indicates density of points.

(G) Smooth scatter plot representing gene rankings by average pseudotime of expression (APE) in *(x-axis)* qNSCs, aNSCs, and NPCs from the current study ordered by Monocle using the consensus-ordering genes identified by machine learning (Table S7A) and *(y-axis)* differentiating myoblasts from the Trapnell study ordered by Monocle using the consensus-ordering genes identified by machine learning (Table S7A) (Spearman's rho = 0.02, p-value = $2.714e^{-05}$). Intensity of blue shading indicates density of points.

(H) Smooth scatter plot representing gene rankings by average pseudotime of expression (APE) in *(x-axis)* qNSCs, aNSCs, and NPCs from the current study ordered by Monocle using the consensus-ordering genes identified by machine learning (Table S7A) and *(y-axis)* hippocampal NSCs from the Shin study ordered by Monocle using the consensus-ordering genes identified by machine learning (Table S7A) (Spearman's rho = 0.29, p-value < $2.2e^{-16}$). Intensity of blue shading indicates density of points.