PREPRINT ONLY

Early stage differentiation of glia in human and mouse

Frazel PW[1*], Labib D[2], Brosh R[3], Boeke JD[3,4,5,6], Fossati V[2], Liddelow SA[1,7,8,9*]

[1] Neuroscience Institute, NYU Grossman School of Medicine, New York, NY., USA
[2] New York Stem Cell Foundation, New York, NY., USA
[3] Institute for Systems Genomics, NYU Grossman School of Medicine, New York, NY., USA
[4] Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY., USA
[5] Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY., USA
[6] Department of Pathology, NYU Langone Health, New York, NY., USA
[7] Department of Neuroscience and Physiology, NYU School of Medicine, New York, NY., 10016, USA
[8] Department of Ophthalmology, NYU School of Medicine, New York, NY., 10016, USA
[9] Parekh Center for Interdisciplinary Neurology, NYU Grossman School of Medicine, New York, NY., USA

* Correspondance should be addressed to paul.frazel@nyulangone.org, shane.liddelow@nyulangone.org

Central nervous system macroglia (astrocytes and oligodendrocytes) are required for normal brain development and function, and are among the last cells to emerge during neurodevelopment. Many questions remain about their emergence in the brain and spinal cord, including how early glial fates are specified during development or differentiation, and similarly when subtypes of glia are specified. Here, we used single-cell RNA sequencing (scRNAseq) to analyze ~90,000 cells across multiple timepoints during the differentiation of astrocytes and oligodendrocytes from human induced pluripotent stem cells and mouse embryonic stem cells. Using time series analysis of gene expression, we uncovered multiple genes involved in fate specification of glial subtypes in both species. We examined gene expression changes during intermediate states of glial specification, and were able to identify genes that were correlated with the choice between neuron versus glia in both species. Using our scRNAseq data we optimized previous mouse astrocyte differentiation protocols by highlighting and removing non-required transition states and decreasing the overall protocol from 3 weeks to less than 12 days. Our data will be useful for researchers interested in optimizing glial differentiations in either species, and provide a window into human glial differentiation, which is difficult to study given its lateness in development.

INTRODUCTION

Macroglia like astrocytes and oligodendrocytes are among the last cells to emerge during the self-organization of the mammalian brain. Understanding their development is critical to understand the complex heterogeneity of glia[1], and for understanding how developmental processes go awry in diseases like cancer[2,3], neurodevelopmental disorders, and other conditions linked to reactive states of glia[4]. Given that small subgroups of glia respond to insults very differently than other cells[5,6], understanding developmental processes that may inform this specialization is critical[7]. Single cell RNA sequencing (scRNAseq) has enabled identification of rare populations of cells, impossible to observe with previous methods like bulk sequencing, that play key roles in development and other processes. Continued use of scRNAseq has produced massive, developmental cell atlases of mouse neurodevelopment[8–10], multiple atlases of earlier human developmental stages[11,12], and has been used to track heterogeneity during the differentiation of stem cells[13,14]. These analyses have identified genes underpinning key moments in neurodevelopment, such as the emergence of the earliest astrocytes during the so-called “gliogenic switch”[9,15]. In mice, this switch to astrocyte production seems to be controlled by a combination of various transcription factors including Nfia[16,17], Nfib[18,19], Sox9[20–22], and signaling pathways like Notch[23,24] (reviewed elsewhere[25–28]). Studies have begun to uncover regional specifications to this process[29–31] in line with a recent focus on identifying brain-area specific astrocyte subtypes[32–35].

In humans, much less is understood about the mechanisms directing gliogenesis[27,36] and neurogenesis more broadly. Even though we understand less about human glial development, there are protocols for human astrocyte differentiation that have been verified at the scRNAseq level by comparing gene expression to acutely isolated human astrocytes[37]. Conversely, there are many mouse astrocyte differentiation protocols[38–41], but these have not widely compared the differentiated cells to acutely purified bona fide primary cells via scRNAseq. This is an important gold standard for a differentiation protocol, as the ideal differentiated version of a cell type would be indistinguishable from the primary cell type at the transcriptomic, proteomic, and functional levels. While the first scRNAseq analyses of mouse glial development have begun to appear[41,42], no one has yet produced such an analysis for the differentiation of astrocytes from mouse stem

cells. Some groups have analyzed differentiating astrocytes using bulkRNAseq[38,43], but no one has compared astrocyte gene expression at the scRNAseq level between fully matured primary mouse astrocytes and astrocytes during differentiation in vitro to validate these methods.

Thus, little is known about heterogeneity among different groups of cells within a single dish of astrocytes, as they respond to growth factors. This is especially interesting for astrocytes, as they are some of the last cells to emerge in the brain, and thus pass through more intermediate stages in comparison to earlier-differentiating neurons. It seems like that these early stages are important for determining mature astrocyte subtype identity after differentiation. Recent work demonstrated that single neuroblast cells in the mouse can produce astrocytes of either very similar or very different mature subtypes, suggesting that the state of astrocyte progenitors can indeed markedly influence downstream astrocyte subtype. Further, single cell sequencing data has proved helpful in other differentiation protocols: to improve the efficiency of a differentiation[13], understand heterogeneity of differentiated cells[44], and see how various intermediate stages are kept or skipped in a differentiation compared to in normal development[45]. Thus, scRNAseq analysis of differentiating astrocytes would be useful to better understand how the differentiation process works, and how it is similar or different to regular development at key intermediate steps, particularly for gliogenesis.

Here, we present how to use scRNAseq to understand intermediate changes in astrocyte differentiation in both human and mouse model systems. We analyzed ~43,000 cells from 3 timepoints from a human astrocyte differentiation previously verified at the scRNAseq level[37], and compared it with a dataset of ~46,000 cells from 8 timepoints across a mouse astrocyte differentiation[39]. We computationally identify intermediate and precursor cells in both mouse and human differentiations, and subsequently identified genes likely to play roles in fate specification for both species. Our human dataset gives insight into late stages of development that would otherwise be untenable for technical and ethical reasons, underscoring the importance of studying late developmental process like gliogenesis using in vitro models. Finally, understanding the fate specification of glia during differentiation is important for efforts by our group and others to use these models to understand gene expression of glia-expressed genes related to neurodegeneration.

RESULTS

Analysis of intermediate stages during in vitro human astrocyte development

As validated protocols for the differentiation of human induced pluripotent stem cells (iPSCs) already exist, we first started by producing a time-series scRNAseq dataset that we could then use to anchor our mouse astrocyte differentiation protocol. To better understand astrogenesis in human iPSC cultures, we chose two intermediate timepoints from a previously published[37] astrocyte differentiation protocol (Fig 1A, Fig S1A) to analyze with scRNAseq using 10X Genomics. We analyzed a total of 56,446 cells from 6 separate samples: 33,152 cells from the two new timepoints for this study at 30 and 50 days of differentiation, with two independent replicates for each timepoint, and 23,294 from a previous study (73 days)[37]. Following quality-control processing via the Muscat46 software package 43,506 cells remained (Fig S1B). We used Harmony[47] to integrate our data (Fig S1E), and then visualized the output using a force-directed graphical dimensional reduction (FLE[48]; Fig 1B,C). Using this approach, we identified mature neurons, astrocytes, and oligodendrocytes based on expression of canonical cell-type marker genes (Fig S3A-C).

To begin to explore intermediate stages of astrocyte development, we computed RNA velocities[49] for each cell using the ScVelo[50] Python software package (Fig 1D). RNA velocity streamlines clearly depict flow of the velocity vectors over time towards the mature clusters (Fig 1D arrowheads; neuron, astrocyte, oligodendrocyte). This temporal ordering is further reinforced by velocity pseudotime analysis (Fig S2A), and is consistent with our timepoint labeling (Fig 1B). Thus, combining these methods let us begin to visualize intermediate states during in vitro gliogenesis.

We next analyzed the data using Waddington Optimal Transport (WOT), an algorithm designed for time-series analysis of scRNAseq data[13]. Implementation of WOT within the CellRank software package[51] uncovered 4 stable macrostates (Fig S2E), corresponding to two different clusters of mature astrocytes (clusters 0, 2), and one cluster each of neurons and oligodendrocytes (clusters 6 and 3 respectively). Inspection of the WOT coarse transition matrix confirmed the extreme stability of these 4 macrostates (Fig S2G,H; see Methods). Using the CellRank package, we calculated the probability for each cell to enter one of the four macrostates, and noted that the probability maps for the two astrocyte clusters were very similar (Fig S2F). For simplicity, we focused our subsequent analysis on astrocytes from cluster 0, which represent the majority of the astrocytes detected using these methods.

We next plotted the probability for each cell to enter one of the three macrostates (astrocytes, oligodendrocytes, or neurons). We find that most of the intermediate differentiating cells are likely to become astrocytes or neurons, and a smaller minority will become oligodendrocytes. Further, the probability maps suggest that astrocytes and neurons share a common set of early intermediate cells that are likely to enter either state (Fig 1E,F, orange arrowhead). These early neural/astrocytic progenitors, which mainly belong to the day 30 timepoint (Fig 1B), are also clearly distinct from those cells that are likely to become oligodendrocytes (Fig 1G). Consistent with the observation that these cells share macrostate likelihoods, these progenitors also belong to the same cluster based solely on gene expression, cluster 9 from the Louvain clustering (Fig 1C). Examination of positively enriched genes expressed by this cluster revealed enrichment for markers of neurogenesis like *HES6*[52] and *ASCL1*[53]. This cluster did not score highly on cell cycle/proliferation scores (Fig S4B), further suggesting its role as a very early progen-

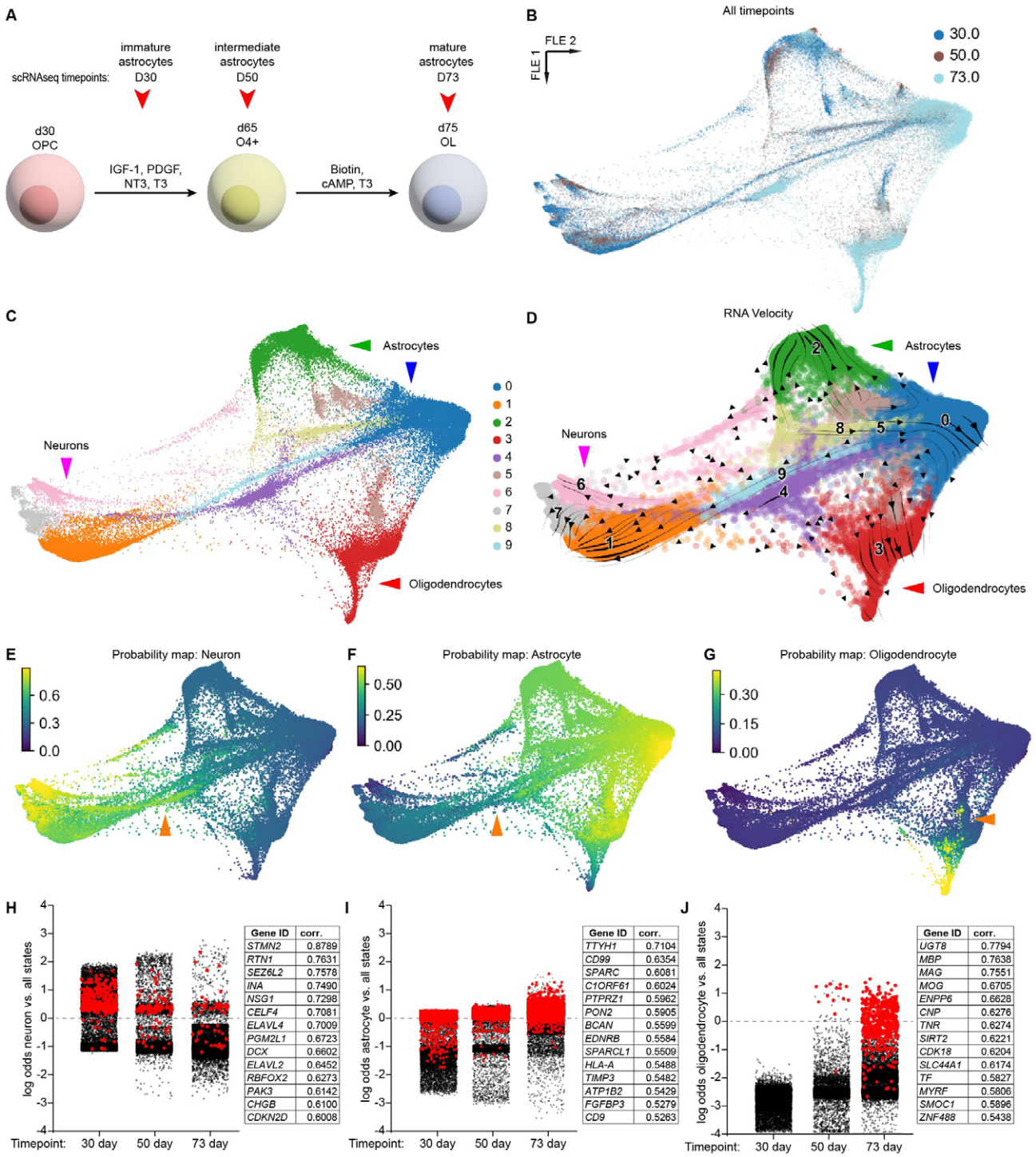


Figure 1 | scRNAseq of human glial differentiation uncovers lineage-specific precursors and transient states. **A.** Differentiation schematic. **B.** Dimensional reduction of gene expression for all human data (43,506 cells) shown via force-directed graph (FLE). Each timepoint represents two independent replicates. **C.** FLE graph with cells clustered by gene expression using the Louvain algorithm (resolution: 0.3); mature cell types are identified with arrowheads. **D.** FLE graph with RNA velocity streamlines overlain, to depict flow of cell states over time based on splicing ratios for each gene. **E-G.** Probability maps for each mature cell type based on Waddington Optimal Transport (WOT) algorithm, scale represents the probability for each cell to eventually enter a given state. Orange arrowheads in E-F represent cells likely to become astrocytes or neurons; orange arrowhead in G represents cells likely to become oligodendrocytes. **H-J.** Log odds plots for each mature cell type based on WOT analysis. For every cell from each timepoint, the log odds to become a given cell type versus all other are plotted. In H, red cells express *DCX* > 3 (normalized expression); in I, red cells express *TTYH1* > 3; in J, red cells express *MOG* > 3.

itor (i.e. temporally prior to the return to proliferation that defines early astrogenesis[54]). Consistent with this classification, cells from the proliferative cluster (Fig 1C, cluster 5) are also clearly likely to become either astrocytes or oligodendrocytes based on probability maps (Fig 1E-G), but are not likely to become neurons. Finally, these cells also express *NHLH1*, an understudied transcription factor recently denoted as a neuroblast marker in mouse neurogenesis[9]. In our data, cells that express *NHLH1* are fated to become either neurons or astrocytes, in contrast to the recent finding in mouse[9], where *Nhlh1* expression was exclusive to neural lineages. This could represent 1) true species differences, or 2) a difference resulting from in vitro differentiation versus in utero development.

Based on RNA velocity and fate probability maps, cells from cluster 5 are likely to form both astrocytes and oligodendrocytes, suggesting that this may represent an early group of cells that are committed to gliogenesis. In addition to cell cycle genes like *CENPF* and *MKI67*, positively enriched genes for this cluster included *HMGB2*, known to be critical for astrogenesis in mice[55]; published bulk RNA datasets report enrichment for *HMGB2* in acutely purified fetal human astrocytes[56]. We next investigated cells within this cluster that were more likely to become astrocyte versus oligodendrocytes, and similarly attempted to identify oligodendrocyte precursor (OPC) cells. Accordingly, we re-clustered the data at a higher resolution (1.2 versus 0.3) to see if cluster 5 would splinter into separate astrocyte- fated and oligodendrocyte-fated clusters (Fig S4A). The resulting cluster (cluster 13) showed clear *OLIG1* enrichment (Fig S3C) although this subcluster initially clustered with *OLIG1*$^{-}$ cells. Interestingly, this cluster is *OLIG1*$^{+}$, but it has low probability of reaching the oligodendrocyte fate; it is also positive for *OLIG2*, which is known to specify oligodendrocyte fate in mouse when its expression is maintained[27]. Thus, these observations are consistent with identifying cluster 13 as a putative pre-OPC state in human cells, perhaps like the one recently hypothesized within an in-depth scRNAseq analysis of mouse neurogenesis[9]. Finally, this cluster also expresses *PDGFRA* (Fig S4C), a known oligodendrocyte lineage marker. The two clusters directly adjacent to cluster 13 (4 and 6) also have high probability of becoming oligodendrocytes, and express both *EGFR* and *DLL3* (Fig S4C), consistent with identifying them as OPCs.

Discovery of genes correlated with fate decisions

To further investigate intermediate stages in glial development, we computed the genes most highly correlated with fate probabilities for a given macrostate. For each cell in the dataset, we computed the odds to reach a given macrostate versus the odds of differentiating into other cell types (Fig 1H-J). Consistent with the order of emergence of cell types during brain development, the overall odds for cells to become neurons decreased over time (Fig 1H), and the inverse was true for astrocytes and oligodendrocytes (Fig 1I-J). This is an interesting overlap between the order of emergence of brain cell types during both differentiation and development.

To determine how gene expression impacts the likelihood of differentiation into neurons, we labeled cells highly expressing *DCX* (normalized gene expression > 3) in red to trace how decreasing *DCX* expression over the differentiation correlates with decreasing odds to become neuronal (Fig 1H). For astrocytes, we found that the gene most correlated with probability to become an astrocyte was *TTYH1*, which we visualized by coloring in red cells with high *TTYH1* expression (normalized expression > 3, Fig 1I). *TTYH1* has not previously been implicated in astrocyte differentiation to our knowledge, although a recent report suggested a role for the gene in murine neural stem cell differentiation[57], and other groups have linked it to the formation of tumor microtubes in glioma[58]. Published bulk RNAseq datasets on purified primary astrocytes report *TTYH1* expression is restricted to astrocytes, and increases with age[56]. Finally, we colored the log odds graph red for cells highly expressing *MBP* (normalized expression > 3), which unsurprisingly marked the cells most likely to become oligodendrocytes at 50 days, and then marked many more cells at 73 days (Fig 1J), further bolstering a model where oligodendrocytes develop later in the differentiation. Thus, we uncovered a list of genes that are correlated with cells likely to choose one fate over the other, which boosts our understanding of the genes and pathways involved during intermediate stages of human glial differentiation.

Optimization of mouse astrocyte differentiation protocol

Given our discovery of key driver genes and heterogeneity in human in vitro gliogenesis, we were next interested in how this process occurs during the differentiation of astrocytes from mouse embryonic stem cells (mESCs). Much work over the past decade has implicated individual transcription factors[27] and signaling molecules in mouse astrocyte differentiation from mESCs, however no one has yet examined this process using scRNAseq. Moreover, a recent report suggested that current murine astrocyte differentiation protocols are unable to produce true, mature astrocytes[41], in contrast to the thoroughly verified[37] human differentiation protocol we analyzed above. Other research groups have also revealed a similar lack of maturity in astrocytes differentiated without cues from neurons[24,38]. Thus, a more granular understanding of the molecular mechanisms underpinning astrocyte differentiation would be useful to the field, as many groups work to refine and improve differentiation protocols to overcome this maturity barrier.

To study the mechanisms of astrocyte differentiation from mESCs, we optimized a serum-free[59,60] astrocyte differentiation protocol[39] to produce immature astrocytes (Fig 2A; Fig S5A). We altered the protocol by: 1) removing CNTF as a growth factor due to recent reports that it induces a reactive state in astrocytes[38]; 2) coating plates with poly-d-lysine and laminin during later protocol stages; and 3) removing retinoic acid and smoothened agonist from the early stages to avoid induction of spinal cord identity. We focused our scRNAseq analysis of these cells on earlier timepoints in our protocol, to better understand early decision points in the acquisition of

astrocyte identity.

To ensure that these cells eventually reach maturity, we stimulated them with CNTF and tested for *Gfap* expression, as mature astrocytes will upregulate Gfap in response to reactive signals[5,59]. We found robust *Gfap* expression in cells after 12 days total in the differentiation: 4 days of embryoid body (EB) expansion, 4 days of adherent differentiation, then 4 days of differentiation with CNTF (Fig 2B). This was significantly faster than the onset of Gfap expression at ~21 days noted in the original differentiation paper[39]. We confirmed this result by performing GFAP immunostaining in another mESC cell line (Fig S5G). This *Gfap* expression time course is also broadly comparable to a recent differentiation protocol, where *Gfap* expression begins at low levels (~5% of cells) at day 14, but >95% of cells express *Gfap* by day 35 in the presence of CNTF[38]. This high level of *Gfap* confirmed our decision to not use CNTF during our actual differentiation due to concerns of producing reactive cells[5,59]. Further, CNTF is a known activator of the IL-6/STAT signaling pathway[61], and this pathway has recently been proven to drive astrocyte reactivity during inflammation and in several mouse models of neurodegeneration[5].

Analysis of early stages of mouse astrocyte differentiation

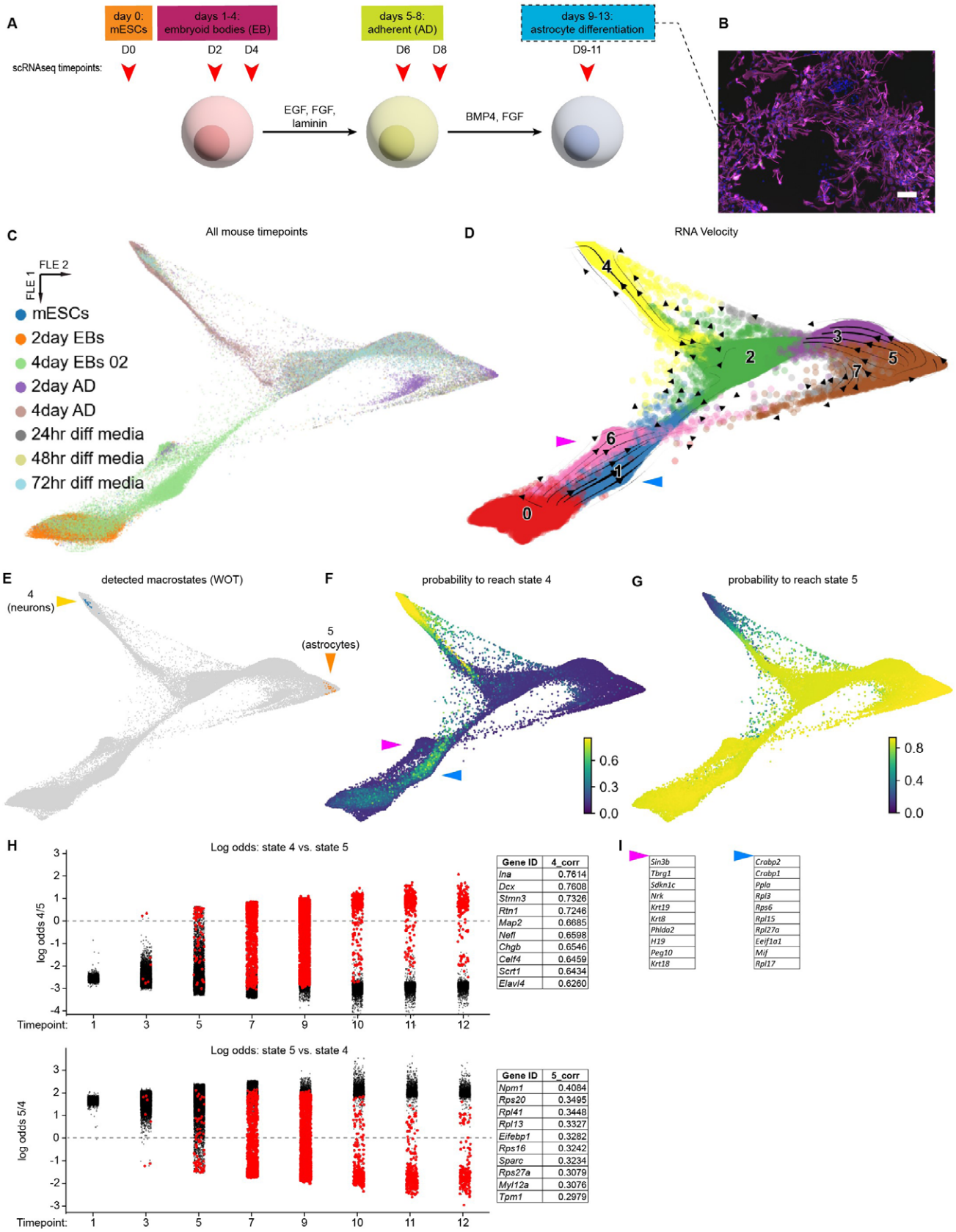
To better understand the granular gene expression changes in early astrocyte development we analyzed gene expression during the first 12 days of astrocyte differentiation via scRNAseq, examining a total of 53,370 mouse cells after preprocessing with Muscat for quality control (Fig S5E). Our dataset spans 9 timepoints (Fig 2A), from mESCs to embryoid bodies, through the intermediate adherent differentiation (AD) stage, and finally through exposure to FGF1/BMP4 (see Fig 2A). We integrated samples using Harmony, and corrected for batch effects (Fig S5F). RNA velocity analysis demonstrated a clear flow away from the mESC and EB stages towards the intermediate stages (Fig 2C). Similar to the velocity results in the human dataset (Fig 1D), the velocity vectors for many of the intermediate stages were incoherent, which is a known aspect of velocity analysis[62]. Further, even though confidence in velocity was high (Fig S6B) and the velocity streamlines aligned with known timepoints (Fig 2C), the calculated velocity pseudotime was inaccurate (Fig S7A). Calculating latent time based on the dynamic model of RNA velocity also did not produce an accurate pseudotime (Fig S6C). Accordingly, we next analyzed this data with the WOT time-series algorithm. The algorithm detected two macrostates that were consistent with the known temporal order of the samples: one neuronal (state 4) and one non-neuronal (state 5; Fig 2E). State 4 was confidently assigned as neuronal given expression of *Stmn2*, *Dcx*, and *Tubb3* (Fig S7A), while state 5 is likely an early astrocytic progenitor based on enrichment for genes like *Vim*, *Sparcl1*, and *Clu* (Fig S7B). Identification of two macrostates by the WOT algorithm was unsurprising, as this mouse astrocyte differentiation is not expected to produce a third cell type (oligodendrocytes), in contrast to the human protocol (see Fig 1 and Discussion).

We next calculated the probability for each cell to enter the two macrostates (Fig 2F,G) in addition to genes correlated with probability to enter each state (Fig 2H, I). Consistent with the overall arc of the differentiation towards astrocyte production, most cells in the dataset had a low chance of entering the neuronal state, and instead had high odds to reach the non-neuronal state. To investigate this further, we plotted the log odds to reach one state versus the other in (Fig 2H, I), and colored cells red if they expressed *Dcx* above a normalized threshold of 1. Unsurprisingly, while most cells were *Dcx*$^+$ early in the differentiation, in later cells *Dcx* expression was correlated with higher odds to become neural state over nonneuronal, mirroring our analysis of human iPSC neuronal cell differentiation above. Consistent with these observations, *Dcx* was one of the top 10 genes most correlated with odds to reach state 4 versus 5, as seen in the list of genes representative of this microstate (Fig 2H). Finally, we also noted enrichment for *Nhlh1* and *Nhlh2* in the group of cells likely to reach the neural state (Fig S7C), which are transcription factors recently identified as specific markers for a neuroblast population that yields exclusively neural daughter cells[9]. Thus, based on assessment of these genes, similar genes are expressed during the beginning of astrocyte differentiation and early mouse neurodevelopment.

Analysis of mechanisms of heterogeneity in murine astrocyte differentiation

Most glia differentiation protocols in mouse and human generate other cell types at varying levels; thus, understanding how this heterogeneity occurs would enable the field to optimize protocols to minimize the production of undesirable cell types. Other groups have recently begun to address heterogeneity in neuron differentiations via scRNAseq[14], but to our knowledge no groups have examined this diversity at the single-cell level during astrocyte/glial differentiation. We noted a subpopulation of cells from the EB stage that had high odds to reach the neural state (Fig 2F, teal arrowhead) immediately adjacent to an EB population with low odds to reach the neural state (Fig 2I, pink arrowhead). This observation could represent an unexpected early specification of neural versus glial fate as was reported recently using in vivo fate mapping of CNS cell development.

To investigate this possibility, we compared gene expression between both clusters to better understand their fate differences (Fig 2J,K). First, there are no differences in normalized gene counts (Fig S6D) or apoptotic score (Fig S7E) between these clusters, so we examined genes differentially enriched in each cluster. The cluster with high probability for neuronal fate (teal) showed enrichment for *Crabp1* and *Crabp2* (Fig 2I), cellular retinoic acid (RA) binding proteins that have been implicated in neurogenesis[63]. While the original astrocyte protocol[39] included RA, we did not include it in any of our differentiation media. The presence of *Crabp1/2* may be explained by RA production by cells during the differentiation, analogous to production of RA during nervous system development in vivo[64]. It seems somewhat unlikely that only a small subset of cells from a large group would re-



Gene ID	4_corr
Ina	0.7614
Dcx	0.7608
Stmn3	0.7326
Rtn1	0.7246
Map2	0.6685
Nefl	0.6598
Chgb	0.6546
Celf4	0.6459
Scrt1	0.6434
Elavl4	0.6260

Gene ID	5_corr
Npm1	0.4084
Rps20	0.3495
Rpl41	0.3448
Rpl13	0.3327
Eifebp1	0.3282
Rps16	0.3242
Sparc	0.3234
Rps27a	0.3079
Myl12a	0.3076
Tpm1	0.2979

Sin3b	Crabp2
Tbrg1	Crabp1
Sdkn1c	Ppia
Nrk	Rpl3
Krt19	Rps6
Krt8	Rpl15
Phlda2	Rpl27a
H19	Eeif1a1
Peg10	Mif
Krt18	Rpl17

(opposite page)
Figure 2 | scRNAseq of mouse glial differentiation uncovers lineage-specific precursors and transient states. **A.** Mouse differentiation schematic. **B.** Micrograph of differentiated astrocytes displaying robust GFAP staining after 12 days of differentiation. **C.** Force-directed graph (FLE) graph of all mouse timepoints (53,370 cells total). **D.** FLE graph with RNA velocity streamlines overlain, to depict flow of cell states over time based on splicing ratios for each gene. Cells were clustered with the Louvain algorithm (resolution: 0.7). **E.** Two macrostates were detected with the Waddington Optimal Transport (WOT) algorithm: neurons and early astrocytes. **F,G.** Probability maps for each mature cell type based on the WOT algorithm, scale represents the probability for each cell to eventually enter a given state. In panel F, pink arrowhead marks cells unlikely to become neurons; blue arrowhead represents cells likely to become neurons. **H.** Log odds plots for each mature cell type based on WOT analysis. For every cell from each timepoint, the log odds to become a given cell type versus all other are plotted. In H, red cells express *DCX* > 1. **I.** List of top 10 positively enriched genes in each of the clusters denoted by arrowheads in panel F.

spond to RA, which is generally a potent morphogen[64], so the *Crabp1/2* expression could also be unrelated to the actual presence of RA.

We sought to further investigate this potential early lineage specification by examining the EB and AD stages on their own. When we examined reclustered data from only the 4 timepoints from those two stages (Fig 3A,B; Fig S8), we could clearly identify the *Crabp1/2*$^+$ cluster, in addition to a *Sin3b*$^+$ cluster (Fig 3C). Based on RNA velocity for this analysis, it does not seem like the *Sin3b*$^+$ cluster is an early glial progenitor (Fig 3C). This is further supported by the WOT fate maps produced for this subset of the data (Fig 3E,F), which also did not replicate the separation of fates seen in our mouse data. Instead of two differentially committed precursor populations, these two clusters could represent leading and lagging cells progressing through the differentiation, or subtypes of cells differentially exposed to local RA signaling dependent on physical location within the differentiation dish.

Analysis of transient stages in astrocyte differentiation

We next sought to better understand the gene expression changes immediately preceding the neuronal "off-ramp", which was clearly visible again in the subset of the data. The most enriched gene in the putative off-ramp cluster was *Cdk1nc*, an inhibitor of mitosis. This is consistent with the exit from cell cycle that is critical for specification of neurons, and is further consistent with enrichment in this cluster for *Srrm4*, a critical regulator of neuron-specific alternative splicing[65]. Further, these cells specifically express a variety of neuronal genes such as *Dcx*, *Syt1*, and *Stmn2* (Fig S6A). We then searched gene expression data for indications of cell signaling pathway activation, as a way to understand what drives certain subsets of cells to become neurons. We first found enrichment for the Notch ligand *Dll3* in this cluster, which was implicated as a marker for early gliogenesis in a recent study[9]. However, these authors note that *Egfr* and *Dll3* co-expression was a marker for gliogenesis, and we did not observe *Egfr* expression in this cluster. Thus, Notch pathway activation via *Dll3* may not be sufficient on its own to direct gliogenesis, but may also require Egf signaling.

After examining genes that enable the neuronal off-ramp, we investigated how cells change upon exposure to astrocyte growth factors during the next step of the differentiation. To do so, we subset the data with both days of AD and our three timepoints of exposure to FGF1/BMP4 (Fig 3H,I). This analysis suggests a similar lineage pattern to that discussed above, as there are two clear paths: 1) neuronal lineage; and 2) astrocyte lineage (Fig 3J). The astrocyte lineage is composed of two clusters, 0 and 1. WOT analysis identified cluster 1 as a macrostate (Fig 3K), in addition to again recovering the neuronal state (cluster 3; Fig 3L). Cluster 1 expressed a host of genes implicated in astrocyte development, including the RNA-binding protein *Qk*[66], *Hmgb2*, and *Slc1a3* (GLAST), in addition to cell cycle genes. This cluster also expressed *Igfbp2*, a molecule not previously implicated in astrocyte differentiation or development to our knowledge, but one implicated in nervous system development[67]. Cluster 0 expressed *Tgln* and *Tgln2*, genes that have been shown to promote migration and slow proliferation in other developmental systems[68]. These are also two well-known TGFβ target genes[69], consistent with exposure to BMP4 during the differentiation, which is a major member of the TGFβ signaling family[70].

Comparison with published datasets

We next compared our mouse differentiation data with a recent scRNAseq dataset of early astrocytes acutely purified from 3 day old mice using the cell surface marker ACSA-2[41]. We integrated our data from the last five timepoints (AD and FGF1/BMP4 exposure; Fig S9) with the 3 replicates of P3 ACSA-2$^+$ astrocytes (Fig 4A,B). RNA velocity analysis of this integrated datasets highlights flow from our putative pre-astrocyte cluster (cluster 0, positive for *Sox9*, *Slc1a3*; Fig 4F,G) towards some of the cells from the P3 dataset (teal arrowhead in Fig 4C). Particularly, this cluster (cluster 0) expresses high levels of *Mki67* (Fig 4E), which is consistent with the identification of *Mki67*$^+$ cells as astrocytic precursors in the original P3 dataset[41]. This analysis suggests that ACSA-2 sorting could potentially be used to purify astrocytic precursors, although it appears that *Atp1b2*, the gene encoding ACSA-2, is simultaneously expressed at low levels in our neuronal clusters and intermediate astrocyte clusters (Fig 4D). Conversely, given the low expression of *Atp1b2* in the most mature astrocytes from our dataset, this may show that enrichment for 'early' astrocytes via ACSA-2 could miss certain populations of early astrocytes, in line with the recent

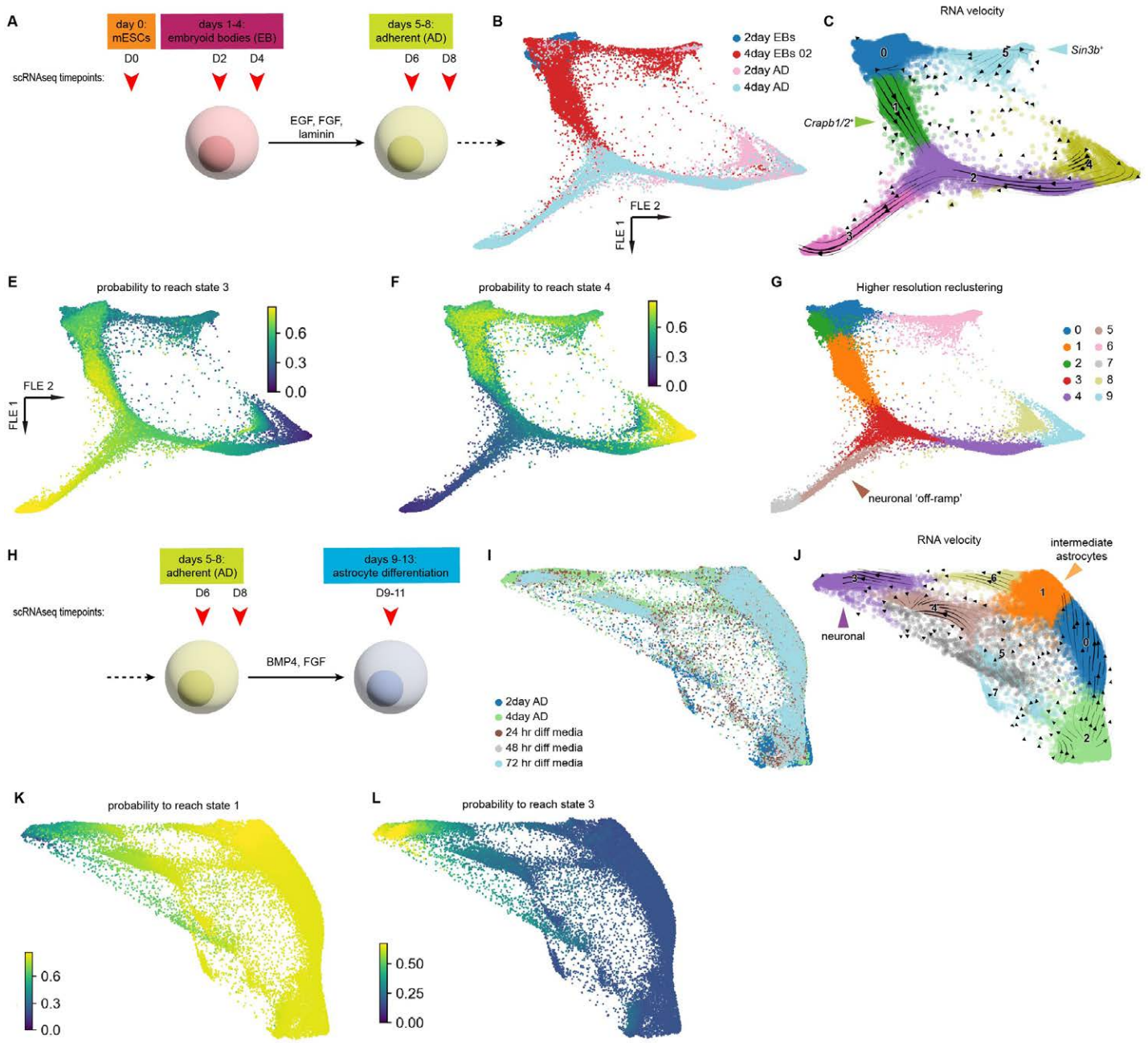


Figure 3 | Analysis of subsets of early mouse astrocyte differentiation. **A.** Schematic of mouse data subset for panels A-G. **B.** Force-directed graph (FLE) plot of reclustered, subset data colored by batch. **C.** FLE graph with RNA velocity streamlines overlain, to depict flow of cell states over time based on splicing ratios for each gene. Arrowheads depict potential early-specified precursor groups, labeled by marker gene. **D.** FLE graph with RNA velocity streamlines overlain, to depict flow of cell states over time based on splicing ratios for each gene. Cells were clustered with the Louvain algorithm (resolution: 0.7). **E,F.** Probability maps for each mature cell type based on the WOT algorithm, scale represents the probability for each cell to eventually enter a given state (state 3 [neuron] or 4 [intermediate astrocyte]). There appear to be no major differences in fate probability between the *Crabp*⁺ and *Sin3b*⁺ clusters. **G.** Reclustering of data from C at higher resolution identifies the earliest cells from the neuronal "off-ramp" path (brown arrowhead). **H.** Schematic of mouse data subset for panels H-L. **I.** FLE plot of reclustered, subset data colored by batch. **J.** FLE graph with RNA velocity streamlines overlain. Neuronal (purple) and intermediate astrocyte (orange) fates depicted with arrowheads. **K-L.** Probability maps for each mature cell type based on the WOT algorithm, scale represents the probability for each cell to eventually enter a given state (state 1 [intermediate astrocyte] or 3 [neuron]).

identification of *Atp1b2*/ACSA-2 negative glial progenitors in the murine cerebellum[71].

To investigate alternative candidates for purification of early astrocytes, we examined expression of *Slc1a3*/GLAST, which was the original astrocyte cell surface marker (ACSA-1; Fig 4G). Expression of this gene slowly ramped up during development, and was specific for cells that avoided the neuronal off-ramp (red arrowhead in Fig 4C). Thus, enrichment of early astrocyte precursors via ACSA-1 instead of ACSA-2 may allow access to earlier precursors, although this should be validated in vivo. Finally, we assessed cells for expression of *Gfap*, a traditional astrocyte marker gene (Fig 4H). While we did not observe any *Gfap* expression within our differentiating mouse cells, we did observe heterogenous *Gfap* expression within the putatively pure P3 astrocyte population. These observations are consistent with the known variability of *Gfap* expression across astrocytes4, and with the lack of *Gfap* expression during the majority of mouse astrocyte differentiation[38].

We did not attempt to analyze these integrated data via WOT, given that it is unclear where to temporally locate the P3 data within our 11-day differentiation time series. Overall, this comparison to published data strengthens our earlier arguments about the astrocytic fate of our mitotic, late differentiation cluster. Our data also reveal potential drawbacks to using the popular ACSA-2 antibody for sorting immature astrocytes, given that *Atp1b2* (ACSA-2) expression occurs later in our differentiation than Slc1a3 expression. Thus *Slc1a3*/GLAST, another popular astrocyte sorting antigen, may allow recovery of earlier mouse astrocyte precursors for in vivo experiments.

Detection of transcription factor activity via BITFAM algorithm

Given our discovery of putative specified progenitors, we next sought to find transcription factor (TF) activity that differed between groups of cells. We opted to use BITFAM[72], a recent algorithm that improves significantly on past algorithms for TF activity detection in both accuracy and speed. Applying BITFAM to our human scRNAseq data revealed multiple TFs with highly cluster-specific activity (Fig 5A). For example, in our astrocytes_0 cluster we observed high activity of *ONECUT2*, which is in contrast to a recent report that *ONECUT2* drives neuronal characteristics in human cells when overexpressed[73]. However, that study examined human 'iNeurons', which are differentiated via overexpression of *NEUROG2*, and thus not directly comparable to the use of growth factors in our differentiation protocol. We also noted activity of *POU2F2* in our cluster 2 astrocytes, which was recently identified as a potential driver of glioma formation from human astrocytes[74], in addition to *PAX2*, a gene known to play key roles in astrogenesis in mice[75,76]. Interestingly, we identified *ARID3B* activity in one of our neuron clusters, which is a TF understudied in the brain, although one report showed increased expression in fetal compared to adult mouse brain[77]. *ID1* was identified as highly enriched in two populations of glial progenitors, fitting with its recent identification as an astrocyte-subtype marker in mice[78]. We report evidence from BITFAM for *ID3* activity specifically in our oligodendrocyte cluster, even though gene expression of this TF was higher in astrocytes, in line with observations from a recent scRNAseq analysis of human iPSC-derived oligodendrocytes[79]. Further, we also noted *NFXL1* activity in the oligodendrocyte cluster, which is an underexplored TF recently associated with a familial speech disorder[80] and localized to the brain[81], but otherwise not connected to oligodendrocyte development to our knowledge.

These inferred transcription factor activity results suggest future TFs to study during both glial differentiation and glial development processes in humans. These data may also provide initial evidence for a mechanism of a familial speech disorder (*NFXL1* loss in oligodendrocytes). Further, we note the inferred specific activity of a putative astrocyte subtype marker (*ID1*) in an early subset of astrocytes, demonstrating that some subtypes of human astrocytes may have expression of their subtype-specific TF early in development. We further note the difficulty of performing these sorts of actual in vivo lineage tracing and sequencing experiments, given the late development of astrocytes and related scarcity of tissue.

We next applied BITFAM to a subset of our mouse differentiation data, and similarly noted clear cluster-specific TF activity (Fig 5B). We saw similar activity in clusters 0 (intermediate astrocyte from differentiation) and 2 (P3 astrocytes from primary data), further bolstering our identification of our intermediate astrocyte cells in the differentiation. Interestingly, we detected *Sox9* activity in early neural precursors (cluster 1), consistent with a recent observation of Sox9 expression during the radial glia stage[22]. However, we did not detect *Sox9* expression in intermediate astrocytes (cluster 0) or any other cluster, even though Sox9 is potentially a more general marker for astrocytes in the brain[21]. Thus, *Sox9* expression may restart once astrocytes have matured farther than the stages analyzed here. We also noted Nkx6-1 activity in cluster 4, a cluster of cells that seem to precede the intermediate astrocytes of cluster 0 (based on velocity arrows in Fig 4C). This TF was recently identified as a putative region-specific TF[30] but has not been implicated in glial development yet to our knowledge. Thus, in both human and mouse datasets, we detected activity of a putative astrocyte-subtype specific TF early in development (*ID1* in humans, *Nkx6-1* in mouse), suggesting a role for these genes in driving subtype specification during intermediate stages of glia development.

DISCUSSION

Here, we report the use of complementary bioinformatic analyses of scRNAseq data to parse the temporal order of glial development in both and mouse human glial differentiation protocols. We provide a clear set of hypotheses about progenitor cells and the order of differentiation to test in future lineage tracing experiments, both in vitro and in vivo. These data will also help researchers better understand the progression of gene expression during astrocyte lineage commitment.

Our dataset consists mainly of scRNAseq data from mul-

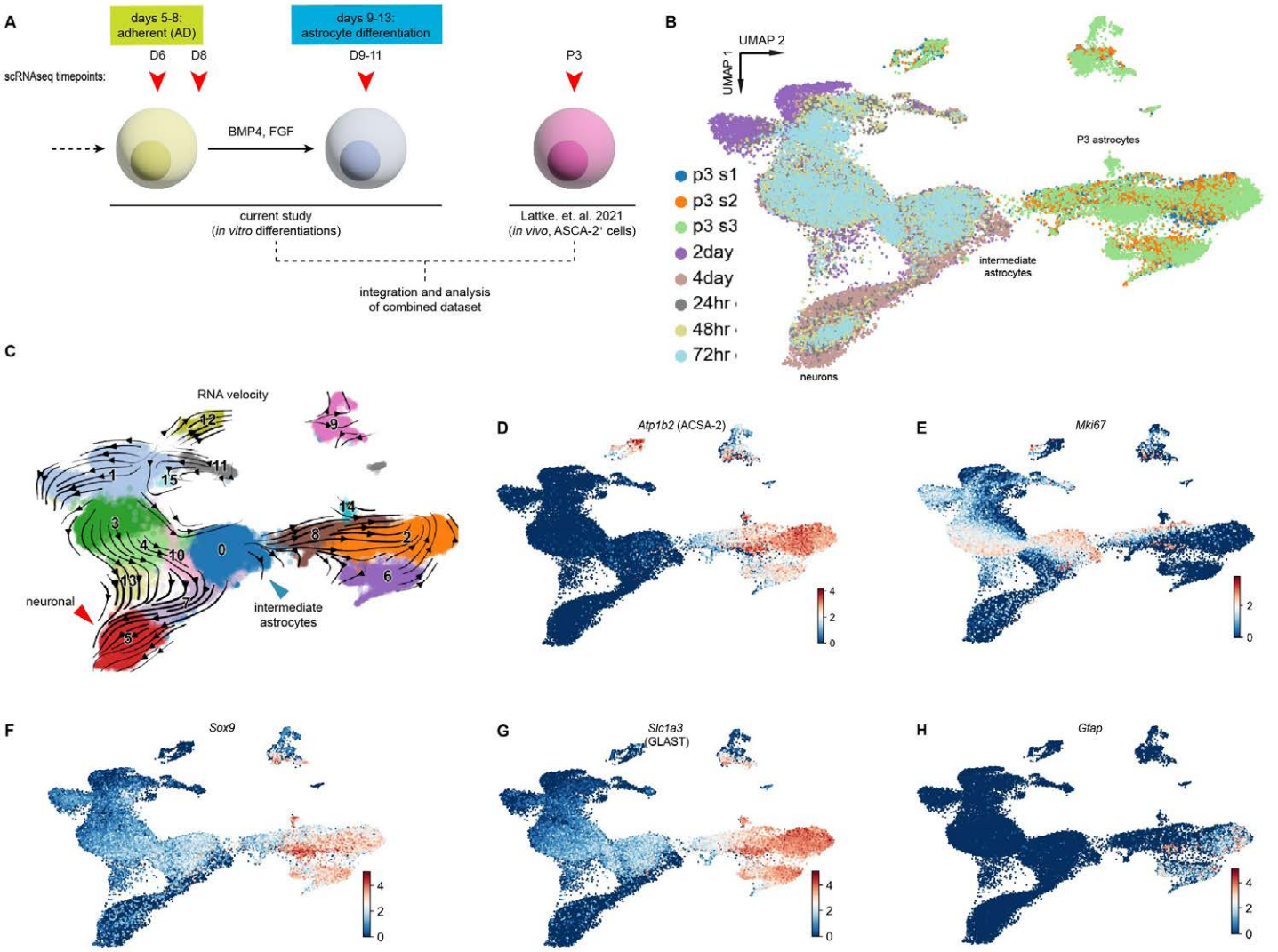
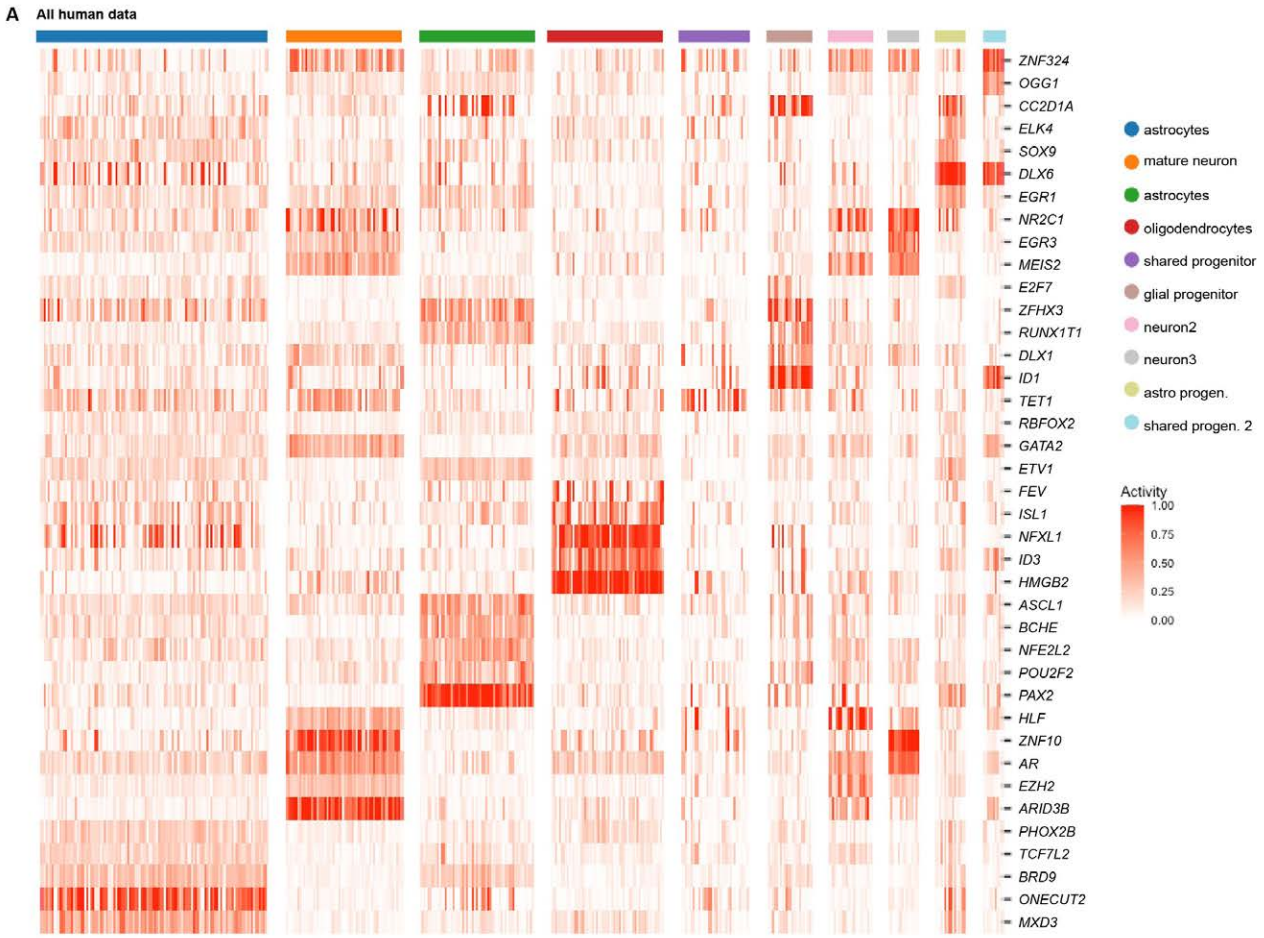
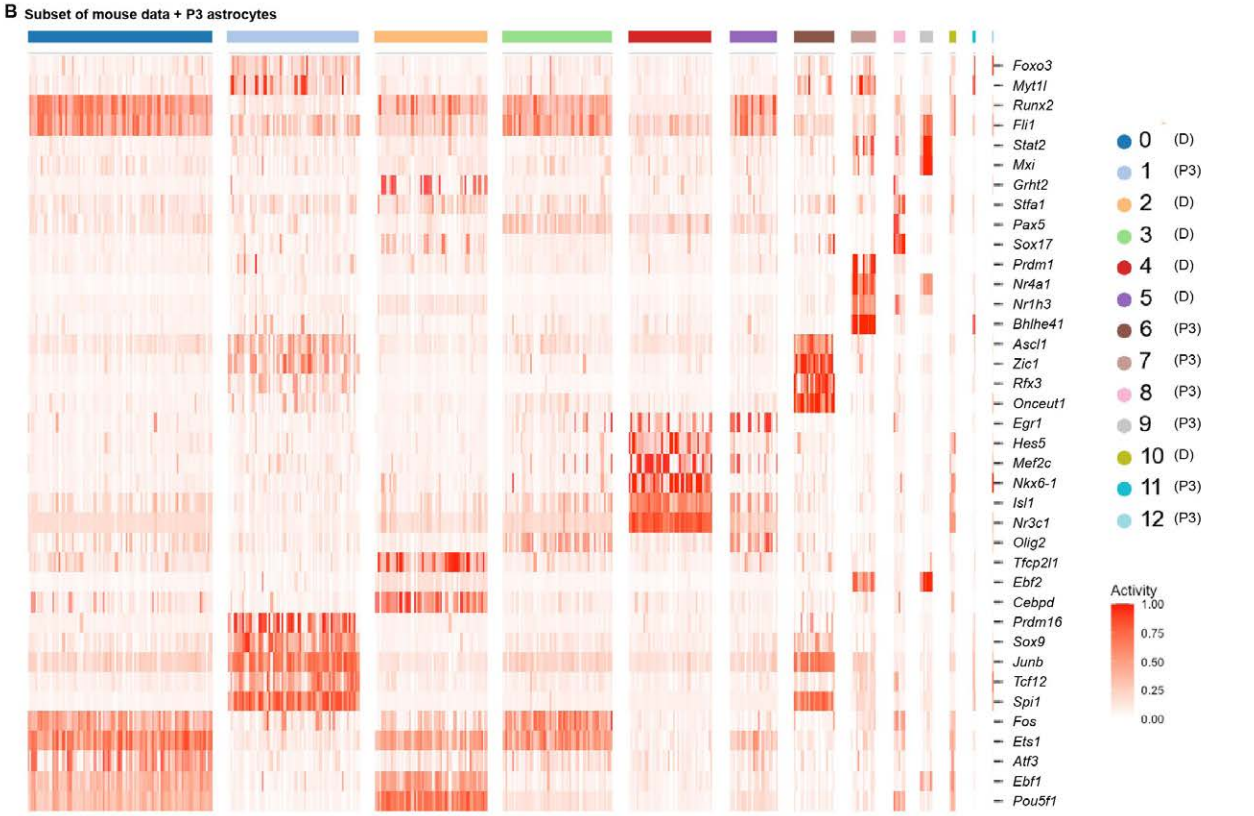


Figure 4 | Integration of mouse astrocyte differentiation in vitro with published in vivo scRNAseq data. **A.** Overview schematic of later mouse differentiation timepoints integrated with published postnatal day 3 (P3) mouse astrocyte data. **B.** Uniform Manifold Approximation and Projection (UMAP) plot of all integrated cells. Differentiation data (adherent differentiation were integrated with primary astrocyte data with the Harmony algorithm. **C.** RNA velocity streamlines are overlain on UMAP with Louvain clusters colored for the integrated data (clustering resolution: 1.0). **D.** Feature plots of *Atp1b2*, the gene encoding for the surface protein ACSA-2 used to enrich for astrocytes in the original P3 dataset. Clear heterogeneity of *Atp1b2* expression is visible across both datasets. **E,** Feature plot of *Mki67* expression depicts likely mitotic cells across both datasets. **F.** Feature plots for *Sox9* expression, an early astrocyte precursor marker. **G.** Feature plot for *Slc1a3* expression shows medium levels of expression in astrocyte-like differentiating cells, but not in the neuronal 'off-ramp', suggesting its potential use as a sorting marker for early astrocytes. **H.** Feature plot for *Gfap* shows its heterogeneous expression within the P3 data.

tiple independent differentiations at different time points. Single cell RNAseq allows the capture of rare or transient subtypes of cells that may be cryptic drivers of macrostate changes during complicated biological processes like differentiation[82]. Bioinformatic analysis of our rich scRNAseq dataset (~96,000 cells across human and mouse cells) helped us identify likely glial-lineage-committed precursors, and explore the genes involved in specification of fate within neurodevelopment in mice and human. While we cannot assert causality based solely upon our current data, we hope that other researchers will find the data useful to prioritize study of various genes during neurodevelopment and gliodevelopment. This is particularly the case for our human data, given the constraints on the study of later-occurring human developmental processes like glial specification. Although gene expression is clearly a useful tool for analyzing the progression of glial differentiation, transcription factors (TFs) are ultimately the key determinants of cell fate choices. Thus, we applied the BITFAM algorithm to our datasets, as it uses a dataset of published ChIP-seq data for hundreds of TFs as


A All human data
ZNF324
OGG1
CC2D1A
ELK4
SOX9
DLX6
EGR1
NR2C1
EGR3
MEIS2
E2F7
ZFHX3
RUNX1T1
DLX1
ID1
TET1
RBFOX2
GATA2
ETV1
FEV
ISL1
NFXL1
ID3
HMGB2
ASCL1
BCHE
NFE2L2
POU2F2
PAX2
HLF
ZNF10
AR
EZH2
ARID3B
PHOX2B
TCF7L2
BRD9
ONECUT2
MXD3
astrocytes
mature neuron
astrocytes
oligodendrocytes
shared progenitor
glial progenitor
neuron2
neuron3
astro progen.
shared progen. 2
Activity
1.00
0.75
0.50
0.25
0.00


B Subset of mouse data + P3 astrocytes
Foxo3
Myt1l
Runx2
Fli1
Stat2
Mxi
Grht2
Stfa1
Pax5
Sox17
Prdm1
Nr4a1
Nr1h3
Bhlhe41
Ascl1
Zic1
Rfx3
Onceut1
Egr1
Hes5
Mef2c
Nkx6-1
Isl1
Nr3c1
Olig2
Tfcp2l1
Ebf2
Cebpd
Prdm16
Sox9
Junb
Tcf12
Spi1
Fos
Ets1
Atf3
Ebf1
Pou5f1
0 (D)
1 (P3)
2 (D)
3 (D)
4 (D)
5 (D)
6 (P3)
7 (P3)
8 (P3)
9 (P3)
10 (D)
11 (P3)
12 (P3)
Activity
1.00
0.75
0.50
0.25
0.00

(previous page)

Figure 5 | Figure 5. Inferred transcription factor (TF) activity in differentiating human glia detected using BITFAM. **A.** Heatmap of BITFAM detected activities for each cluster from Figure 1C (all human data timepoints). Clusters are numbered and labeled on the x-axis, and detected TFs are on the y-axis. **B.** Heatmap of BITFAM detected activities for each cluster from Figure 4C (later mouse timepoints with P3 primary astrocyte data). Clusters are numbered on the x-axis, and detected TFs are on the y-axis. P3 data obtained from Latkke et al. 2021 (Ref 31).

priors in a Bayesian model that attempts to detect TF activity in new cell types. Given the lack of robust datasets of glial TF activity or glial epigenetic data in either species, this algorithm has a major advantage, as it does not require extant cell-type specific ChIP-seq data to operate.

We analyzed a human glial differentiation with scRNAseq analysis of gene expression and inferred TF activity via the BITFAM algorithm. Using the WOT algorithm, we calculated final, stable macrostates for the differentiation and calculated probabilities for each intermediate cell to reach one of the states. This approach revealed that astrocytes and neurons likely share common progenitor cells in our differentiation, and that oligodendrocytes are produced after astrocytes, in line with the progression of development in vivo. Further, we were able to identify genes highly correlated with high probabilities for cells to enter one of the macrostates. At least some of these genes are likely involved in causally driving fate specification, although some of the genes are likely downstream of causal drivers like TFs. We suggest that *TTYH1* is a strong candidate for a driver of astrocyte fate specification, given its high correlation with astrocyte probability and its purported ability to signal through the Notch pathway[57], thought to be important for the acquisition of astrocyte identity[27]. Further, past work has localized *TTYH1* to cytoskeletal processes in glioma cells, and shown that its presence leads to tighter networks of tumor cells; perhaps the same phenomenon may occur normally during astrocyte development.

In our mouse differentiation data, we found an early "off-ramp" fate pathway for a subset of cells that will become neurons, consistent with past reports of transient progression through neuronal fate during astrocyte differentiation[38]. However, unlike past groups using bulk RNAseq, we were able to observe heterogeneity of cell states at the single cell level; by taking this approach, we importantly highlight that all differentiating cells do not necessarily commit fully to the neuronal fate prior to gliogenesis. Moreover, there is clear heterogeneity across cells regarding degree of commitment to the neuronal fate, based on WOT fate maps. This finding will likely prove fruitful for researchers attempting to optimize glial differentiation, as it suggests that there is no need for cells to undergo neuronal specification prior to the acquisition of astrocyte fate. To understand this fate specification further, we also found genes associated with cells that avoid the neuronal fate. Interestingly, our analysis suggests that during the early stages of astrocyte fate commitment (modeled by BMP4/FGF1 exposure in the differentiation), there are not dramatic day-to-day gene changes, given the transcriptional similarity of cells across the 3 sequential 24-hour samples in our dataset. Further, we found evidence for heterogenous expression of the popular astrocyte sorting marker ASCA2 during differentiation, suggesting caution when designing experiments using this marker.

In many cases, we observed similar gene expression in both our human and mouse astrocytes, such as early expression of *Dcx*/*DCX*, a well-established key transcription factor for neurogenesis. Similarly, we noted similar expression of putative astrocyte precursor genes like *Hmgb2*/*HMGB2* and *Slc1a3*/*SLC1A3* across both species. We also saw some divergence, for example with the gene *Nhlh1*/*NHLH1*, which is suggested to define neuron-committed progenitors in mouse, but we saw clearly expressed in human astrocyte progenitors. This divergence in gene expression between species is perhaps unsurprising, given that many of the genes that are highly expressed in human versus nonhuman primate brain are expressed mainly by glia[83]. We noted the formation of early neurons in both protocols, followed by mitotic activity prior to the beginnings of glial specification, consistent with past reports[54]. We mainly examined early timepoints in mouse, and our human cells are not aged for as long as other papers, but we believe that focusing on early differentiation is still warranted, given that human organoid-derived astrocytes eventually show decreases in functional capacity after extensive time in culture[84].

One may wonder why multiple robust human astrocyte differentiation protocols exist, while a bona fide mouse protocol does not yet exist. Similarly, there are robust differentiation protocols for the floor plate using human[85] but not mouse cells. This phenomenon may be a result of the speed of mouse development, which causes cells to change states rapidly, making it harder to recapitulate this speed in a dish with only daily media changes. Alternatively, in the case of mouse astrocytes, the differentiation difficulties may reflect the more general difficulties of culturing mouse macroglia in a physiologic, homeostatic, normal environment, which has only been achieved in the past decade or so[4]. To this end, the absence of *Gfap* expression during early mouse differentiation is a major difference between in vivo and in vitro progression of astrocyte differentiation first noted by others[38] and replicated in our dataset. This is a striking difference, given the historical emphasis on the study of *Gfap*$^+$ radial glia as the earliest mammalian neural progenitors[15].

We hope that this dataset, which to our knowledge is the first scRNAseq time series analysis of glial differentiation in either rodents or humans, will provide a valuable resource for researchers interested in studying these processes in either species. The genes identified as key drivers of specific lineages may also guide groups interested in further optimizing differentiations, either to improve purity of differentiation protocols (e.g., minimizing the production of cells from the non-desired lineage) or to help researchers begin to differ-

entiate specific cell types of astrocytes, a burgeoning area of interest.

METHODS

Cell culture and differentiation

Human iPSC culture

Culture of hESC was performed following the previously published protocol[37]. Human iPSC line 51121-01-MR-017 (52 year old female) was maintained in mTeSR1 medium and Geltrex-coated plates[86]. Astrocyte differentiation was performed following the protocol described in Fig S1A and previously[37]. For collection of 10X samples, cultures at day 29 and day 50 were digested to a single cell suspension using a papain solution. For the final time point corresponding to mature astrocytes we used transcriptomic data from the same iPSC line generated in our previous study.

In brief, iPSCs were reprogrammed from skin fibroblasts of a female healthy individual (age 52 at collection), using the NYSCF Global Stem Cell Array®. We maintained human iPSCs in mTeSR1 in a 37˚C incubator with 5% CO_2. To prepare for induction, we performed enzymatic digestion using Accutase, counted and plated $1.5x10^5$ iPSCs per well on geltrex-coated 6-well plate and fed with maintenance medium with 10 μM Y27632. For astrocyte differentiation, we followed the protocol previously described[37,87]. At day 20, we plated 50 neurospheres per well into poly-ornithine/laminin coated 6-well plates. Spheres attached to the matrices within a few hours and progenitors migrated out and spread across the surface area of the well. At days 29 and 50, we collected single cells using papain digestion and prepared cells for scRNA-seq analysis.

Mouse ESC culture and differentiation

mESC experiments were performed on two separate mESC lines: A17iCre and BL6/CAST ΔPiga. Cells from both lines were maintained in 80/20 media for expansion and freezing, in a 37˚C incubator with 5% CO_2. This media contains 80% 2i mESC media, and 20% general mESC media. Specifically, this 80/20 media contains 40% Advanced DMEM/F12 (Gibco), 40% Neurobasal (Gibco), 10% Knockout DMEM (Gibco), in addition to 2.5% fetal bovine serum (Gibco), 0.75x N2 nutrient mix (Thermo), 0.75x B27 nutrient mix (Thermo), 0.75x L-Glutamine (2mM; Gibco), 1X beta-mercaptoethanol (Gibco), 1x non-essential amino acids (Gibco), 1x nucleotides (Millipore). This media also contained leukemia inhibitory factor (LIF) at 1000 units/mL (Millipore), 3 uM CHIR99201 (Biovision), and 1 μM PD0325901 (Sigma) to maintain pluripotentcy during the mESC stage.

Differentiation was performed with a protocol derived from[39]. On day 1 of the differentiation, mESCs were trypsinized with TrypLE for 2 min, quenched with 80/20 media, then replated in low-adherence plates (Corning) at densities of 1.5 million cells per 10cm dishes in AK media (50% Neurobasal (Gibco), 50% Advanced DMEM/F12 (Gibco), 1X L-Glutamine (Gibco), 13% Knockout Serum Replacement (Gibco)). At days 3 and 5 of the differentiation, embryoid bodies were collected in a 50ml tube and allowed to settle via gravity for 20 min. AK media was refreshed and cells were replated back into the same low-adherence 10 cm dishes. At day 6, embryoid bodies were collected via gravity, then trypsinized in 4mL TrypLE (Gibco) for 5 min with intermittent swirling by hand. TrypLE was quenched with AK media, and cells were filtered through a pre-wet 70 μM nylon cell filter. Filtered cells were replated on poly-d-lysine (PDL)- and laminin- coated 10cm tissue-culture dishes at 1 million cells/dish, and murine EGF (Peprotech, 20 ng/mL), murine FGF1 (Peprotech, 10 ng/mL), and murine laminin (Sigma, 1 μg/mL) were added to AK media for the adherent differentiation phase. Media was replaced every other day until day 11, when cells were switched to differentiation media, AK with FGF1 (Peprotech, 10 μg/mL) and BMP4 (Peprotech, 10 μg/mL). Cells were replated once more onto PDL+laminin dishes upon reaching 70% confluency (~day 15), and maintained in differentiation media (AK plus FGF1, BMP4) replaced every other day.

10X data collection & sequencing

For collection of 10X scRNAseq samples, media was replaced 3 hours prior to 10X processing for all cultures. Mouse cells were trypsinized briefly with TrypLE, then quenched with appropriate media. Cells were triturated with a 10 mL serological pipette, filtered with a 40 μM filter, and diluted to 1000-1200 cells/μl. Human samples were dissociated according to a modified papain-based protocol (Worthington), then filtered through a 40 μM filter before counting and dilution for scRNAseq.

17,400 cells were loaded onto 10X chip G (mouse samples) or chip B (human samples), and run with v3.1 3' chemistry (mouse samples) or v3.0 3' chemistry (human samples). Library preparation was performed according to manufacturer's instructions, and sequenced using a NOVASeq6000 sequencer (Illumina).

Bioinformatic analysis

10X data processing and quality control

All analyses were executed locally on Macbook laptop, except for Cell Ranger, which was performed on a Cray supercomputer cluster. Fastq files from sequencing were demultiplexed, then used as input for Cell Ranger (v6.0.1). Filtered count matrices from Cell Ranger were then used as input to the Muscat quality control pipeline in R[46]. After doublet detection using scds (v1.1.2) and calculation of the mitochondrial gene contribution using scater (v1.13.23) within the SingleCellExperiment library (v1.7.11), cells were filtered using the median absolute deviation (MAD) module where cells containing a greater than 2.5 MAD in feature counts, number of expressed features and percentage of mitochondrial genes were discarded. Between 65 and 85 percent of cells were kept post-filtering, see Figs S1B, S4E for details and numbers for each sample.

Barcodes of cells that passed Muscat quality control were used to filter the .loom file produced by the velocyto python package[49], which analyzes the .bam file output of CellRanger for splicing to prepare for RNA velocity analysis. This filtered .loom file was used as input to the Scanpy[88] pipeline, imple-

mented as part of the scvelo pipeline[50]. All samples were normalized and filtered according to standard Scanpy parameters: genes were filtered to only keep genes with number of spliced counts >20, and then the top 2000 most variable genes were kept by the pp.filter_and_normalize command; principal component analysis (PCA) was run using pp.pca command, then data were integrated along the 'batch' variable via implementation of the Harmony algorithm in the scanpy external API (pp.harmony_integrate); neighbors were found via pp.neighbors using 40 PCs, 40 neighbors, and using X_pca_harmony; cells were clustered using the Louvain algorithm implemented in scvelo (scv.tl.louvain) at resolution values ranging from 0.3 to 1.6, then dimensionality reduction was performed via scv.tl.umap or using a force-directed graph (sc.tl.draw_graph); and finally, marker genes (i.e., top-positively-enriched genes) for each Louvain cluster were found using sc.tl.rank_genes_groups with a Wilcoxon rank-sum algorithm.

RNA velocity analysis

The .loom file from velocyto was used to compute RNA velocities for each cell according to standard parameters for the software. ScVelo produces both stochastic and dynamic models of RNA velocity, which we compared by computing a consistency score for each cell, for each modeling method, as recommended by the authors. Pseudotime was then computed based on RNA velocity results, and latent time was inferred via dynamic velocity results.

Waddington Optimal Transport (WOT) analysis

The WOT algorithm[13] was used via its implementation in CellRank. Batch names were changed to sequential numbers, and then a WOT kernel was initialized via the WOTKernel command. Growth rates were calculated by compute_initial_growth_rates, which uses the scanpy growth score. A transition matrix was calculated partially using these growth rates as a basis via compute_transition_matrix with 3 growth iterations. We found that the transition matrix was nearly identical whether or not we included growth rates. We then computed a connectivity kernel using the ConnectivityKernel command, which calculates a diffusion pseudotime score for each cell. These kernels were combined at weight of 0.9 for WOT, 0.1 for the connectivity kernel, as recommended by the original authors. We saw minimal differences between our outcomes whether or not we included the 0.1 weighted connectivity kernel. We then used a Generalized Perron Cluster-Cluster Analysis (GPCCA) estimator to define macrostates, i.e. cell states with metastability. We then used a Schur decomposition to compute a coarse transition matrix that enabled us to compare the stability and likely temporal order of the various macrostates. We selected macrostates for further analysis based on prior knowledge of mature cell types (humans: oligodendrocyte, astrocyte, neuron; mouse: astrocyte-precursor, neuron), and then produced absorption probability plots/fate maps for these macrostates with the plot_absorption_probabilities command. Finally, we computed genes correlated with high probability to become a certain macrostate via the compute_lineage_drivers command. In some figures, we colored cells red based on normalized expression of a gene of interest as explained in the text/figure legend.

BITFAM algorithm

We exported the count matrices from our Scanpy objects for use with the BITFAM algorithm (v1.0) implemented in R, using standard settings. BITFAM results were first visualized as a t-SNE using the Rtsne R package (v0.15). Finally, cluster IDs were exported from Scanpy and used to create a heatmap of inferred TF activity for each cluster, following sample code from the BITFAM package.

Analysis of Lattke et. al. 2021

Raw .fastq files were downloaded from P3 ACSA-2 purified astrocytes published in[41], NCBI GEO #GSE152223. The fastq files for the raw data were analyzed via CellRanger v6.0.1, and then count matrices were imported into muscat and processed as per above. We went back to the raw data for this integration so that we could produce a .loom file with splice calling, which was not done in the original analysis. We then processed this .loom file for RNA velocity as per above, filtered with Muscat barcodes, and then we combined Lattke et. al. 2021 (Ref 41) data with our mouse data via Harmony as per above.

Immunohistochemistry

Cultures were fixed in 4% paraformaldehyde (Electron Microscopy Services) for 15 min at room temperature (RT), then plates were washed 3x with phosphate-buffered saline (PBS) before blocking in PBS + 0.1% Triton-X and 3% goat serum for 1 hour at RT. Primary monoclonal antibodies to GFAP (ThermoFIsher #13-0300, 1:500) were applied in blocking solution overnight at 4C with rotation. One well per plate was processed without primary antibody as a negative control. The next day, plates were washed quickly 3x with PBS, then washed 3x more with 10 min rotation. Secondary antibodies (Invitrogen, 1:2000) with 4′,6-diamidino-2-phenylindole (DAPI; ThermoFisher, 1:5000) were applied in blocking solution for 1 hour at RT with rotation. Plates were washed again as with the primary antibody, then imaged using a Keyence BZ-X700 fluorescent microscope. Channels were separated to make figures using Fiji/ImageJ software.

Statistics

Clustering was calculated by the Louvain algorithm. Marker genes for each cluster were calculated using the Wilcoxon rank-sum algorithm implemented in Scanpy. Statistical tests used in the WOT and BITFAM algorithms are discussed in those sections of methods.

Reproducibility

For human data, we separately loaded single-cell suspensions from two adjacent wells of a 6-well plate in individual 10X lanes, processed each lane as a separate library, then combined the lanes for analysis bioinformatically after check-

ing for any batch effect; the 30-day and 50-day timepoints were from two separate cultures differentiated on separate dates. For mouse data, EB timepoints were collected from two independent cultures first, to check for replicability/batch effects. Once convinced of minimal batch effects, we then collected all remaining time points on the same day and processed each condition in its own 10x lane. Accordingly, each time point is from an independent differentiation culture (i.e., each timepoint began from frozen mESC vials on a different day, staggered for collection on the same day).

Acknowledgements

The computational requirements for this work were supported in part by the NYU Langone High Performance Computing (HPC) Core's resources and personnel. Funding for this work was provided by the Cure Alzheimer's Fund, MD Anderson Neurodegeneration Consortium, Anonymous Donors, the Blas Frangione Foundation, and The Alzheimer's Association (SAL); the Dark Matter Project funded by an NIH Center for Excellence in Genome Science – Grant IDs 5RM1HG009491, 3RM1HG009491-03S1, 3RM-1HG009491-03S2 (JDB); National Institute of Neurological Disorders and Stroke (NINDS) grant 1R21NS111186 (VF); NINDS T32 (Dasen) 5T32NS086750 (PWF).

Competing Interests

SAL is an academic founder and sits on the SAB of AstronauTx Ltd. JDB is a Founder and Director of CDI Labs Inc., a Founder of and consultant to Neochromosome Inc., a Founder, SAB member of and consultant to ReOpen Diagnostics LLC, and serves or served on the Scientific Advisory Board of the following: Sangamo Inc., Modern Meadow Inc., Rome Therapeutics Inc., Sample6 Inc., Tessera Therapeutics Inc., and the Wyss Institute.

REFERENCES

1. Clarke BE, Taha DM, Tyzack GE, Patani R (2021) Regionally encoded functional heterogeneity of astrocytes in health and disease: A perspective. *Glia* 69:20–27.
2. Laug D, Glasgow SM, Deneen B (2018) A glial blueprint for gliomagenesis. *Nat Rev Neurosci* 19:393–403.
3. Shen Z, et al. (2021) Distinct progenitor behavior underlying neocortical gliogenesis related to tumorigenesis. *Cell Rep* 34.
4. Han RT, Kim RD, Molofsky AV, Liddelow SA (2021) Astrocyte-immune cell interactions in physiology and pathology. *Immunity* 54:211–224.
5. Hasel P, Rose IVL, Sadick JS, Kim RD, Liddelow SA (2021) Neuroinflammatory astrocyte subtypes in the mouse brain. *Nat Neurosci* 24:1475–1487.
6. Guttenplan KA, et al. (2021) Neurotoxic reactive astrocytes induce cell death via saturated lipids. *Nature* 599:102–107.
7. Khakh BS, Deneen B (2019) The Emerging Nature of Astrocyte Diversity. *Annu Rev Neurosci* 42:187–207.
8. Loo L, et al. (2021) Single-cell transcriptomic analysis of mouse neocortical development. *Nat Commun* 10:1–11.
9. La Manno G, et al. (2021) Molecular architecture of the developing mouse brain. *Nature* 596:92–96.
10. Di Bella DJ, et al. (2021) Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* 595:554–559.
11. Trevino AE, et al. (2021) Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* 184:5053-5069.
12. Eze UC, Bhaduri A, Haeussler M, Nowakowski TJ, Kriegstein AR (2021) Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat Neurosci* 24:584–594.
13. Schiebinger G, et al. (2019) Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* 176:928-943.
14. Lin HC, et al. (2021) NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Reports* 16:2118–2127.
15. Bayraktar OA, Fuentealba LC, Alvarez-Buylla A, Rowitch DH (2015) Astrocyte development and heterogeneity. *Cold Spring Harb Perspect Biol* 7:1–16.
16. Deneen B, et al. (2006) The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron* 52:953–968.
17. Kang P, et al. (2012) Sox9 and NFIA coordinate a transcriptional regulatory cascade during the initiation of gliogenesis. *Neuron* 74: 79–94.
18. Tsuyama J, et al. (2015) MicroRNA-153 regulates the acquisition of gliogenic competence by neural stem cells. *Stem Cell Reports* 5:365–377.
19. Matuzelski E, et al. (2017) Transcriptional regulation of Nfix by NFIB drives astrocytic maturation within the developing spinal cord. *Dev Biol* 432:286–297.
20. Vong KI, Leung CKY, Behringer RR, Kwan KM (2015) Sox9 is critical for suppression of neurogenesis but not initiation of gliogenesis in the cerebellum. *Mol Brain* 8, 0–17.
21. Sun W, et al. (2017) SOX9 is an astrocyte-specific nuclear marker in the adult brain outside the neurogenic regions. *J Neurosci* 37:4493–4507.
22. Fabra-Beser J, et al. (2021) Differential expression levels of Sox9 in early neocortical radial glial cells regulate the decision between stem cell maintenance and differentiation. *J Neurosci* 41:6969–6986.
23. De Melo J, Clark BS, Blackshaw S (2016) Multiple intrinsic factors act in concert with Lhx2 to direct retinal gliogenesis. *Sci Rep* 6:1–12.
24. Hasel P, et al. (2017) Neurons and neuronal activity control gene expression in astrocytes to regulate their development and metabolism. *Nat Commun* 8.
25. Sloan SA, Barres BA (2014) Mechanisms of astrocyte development and their contributions to neurodevelopmental disorders. *Curr Opin Neurobiol* 27:75–81.
26. Zuchero JB, Barres BA (2015) Glia in mammalian development and disease. *Development* 142:3805–3809.
27. Sardar D, Cheng Y-T, Szewczyk LM, Deneen B, Molof-

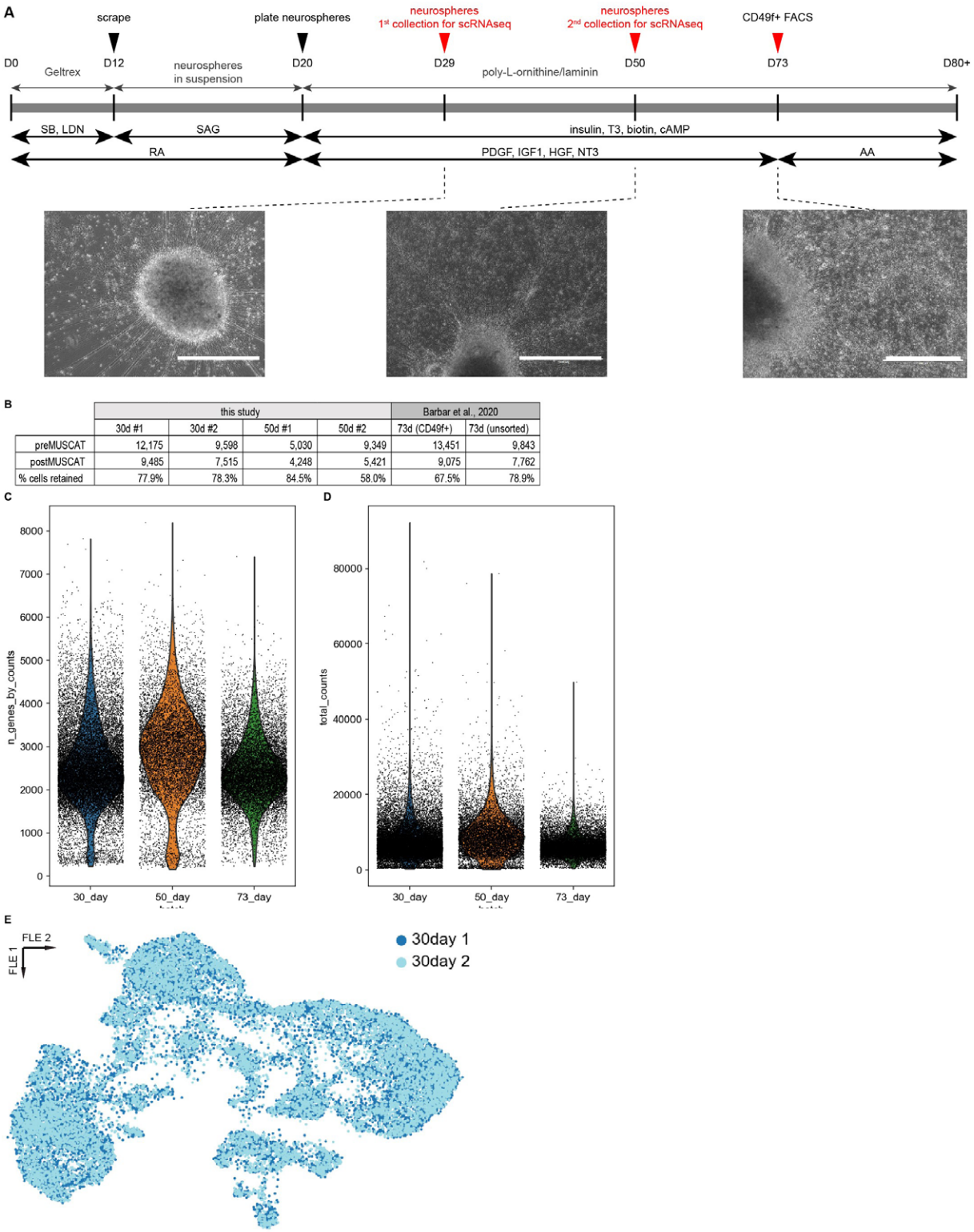
sky AV (2020) Mechanisms of astrocyte development. In: Patterning and cell type specification in the developing CNS and PNS (Elsevier Inc., 2020). Eds. Rubernstein J, Rakic P. Academic Press.
28. Gallo V, Deneen B (2014) Glial development: The crossroads of regeneration and repair in the CNS. *Neuron* 83:283–308.
29. Clavreul S, et al. (2019) Cortical astrocytes develop in a plastic manner at both clonal and cellular levels. *Nat Commun* 10:1–14.
30. Lozzi B, Huang TW, Sardar D, Huang AYS, Deneen B (2020) Regionally distinct astrocytes display unique transcription factor profiles in the adult brain. *Front Neurosci* 14:1–14.
31. Welle A, et al. (2021) Epigenetic control of region-specific transcriptional programs in mouse cerebellar and cortical astrocytes. *Glia* 69:2160–2177.
32. Batiuk, MY, et al. (2020) Identification of region-specific astrocyte subtypes at single cell resolution. *Nat Commun* 11:1–15.
33. Ohlig S, et al. (2021) Molecular diversity of diencephalic astrocytes reveals adult astrogenesis regulated by Smad4. *EMBO J* 40:1–22.
34. Herrero-Navarro Á, et al. (2021) Astrocytes and neurons share region-specific transcriptional signatures that confer regional identity to neuronal reprogramming. *Sci Adv* 7.
35. Bayraktar OA, et al. (2020) Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. *Nat Neurosci* 23:500–509.
36. Molofsky AV, Deneen B. (2015) Astrocyte development: A guide for the perplexed. *Glia* 63:1320–1329.
37. Barbar L, et al. (2020) CD49f is a novel marker of functional and reactive human iPSC-derived astrocytes. *Neuron* 10:436-453.
38. Sardar D, et al. (2021) Mapping astrocyte transcriptional signatures in response to neuroactive compounds. *Int J Mol Sci* 22:3975.
39. Thompson RE, et al. (2017) Different mixed astrocyte populations derived from embryonic stem cells have variable neuronal growth support capacities. *Stem Cells Dev* 26:1597–1611.
40. Benveniste RJ, Keller G, Germano I (2005) Embryonic stem cell-derived astrocytes expressing drug-inducible transgenes: Differentiation and transplantion into the mouse brain. *J Neurosurg* 103:115–123.
41. Lattke M, et al. (2021) Extensive transcriptional and chromatin changes underlie astrocyte maturation in vivo and in culture. *Nat Commun* 12.
42. Li X, et al. (2021) Decoding cortical glial cell development. *Neurosci Bull* 37:440–460.
43. Tiwari N, et al. (2018) Stage-specific transcription factors drive astrogliogenesis by remodeling gene regulatory landscapes. *Cell Stem Cell* 23:557-571..
44. Li J, et al. (2021) Conservation and divergence of vulnerability and responses to stressors between human and mouse astrocytes. *Nat Commun* 12:1–20.
45. Briggs JA, et al. (2017) Mouse embryonic stem cells can differentiate via multiple paths to the same state. *eLife* 6:1–23.
46. Crowell HL, et al. (2020) Muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat Commun* 11:1–12.
47. Korsunsky I, et al. (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 16:1289–1296.
48. Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 9:1–12.
49. La Manno G, et al. (2018) RNA velocity of single cells. *Nature* 560:494–498.
50. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. (202) Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 38:1408–1414.
51. Lange M, et al. CellRank for directed single-cell fate mapping. *bioRxiv* 2020.10.19.345983.
52. Gratton M-O, et al. (2003) Hes6 promotes cortical neurogenesis and inhibits Hes1 transcription repression activity by multiple mechanisms. *Mol Cell Biol* 23:6922–6935.
53. Park NI, et al. (2017) ASCL1 reorganizes chromatin to direct neuronal fate and suppress tumorigenicity of glioblastoma stem cells. *Cell Stem Cell* 21:209-224.
54. Ge WP, Miyawaki A, Gage FH, Jan YN, Jan LY (2012) Local generation of glia is a major astrocyte source in postnatal cortex. *Nature* 484:376–380.
55. Bronstein R, Kyle J, Abraham AB, Tsirka SE (2017) Neurogenic to gliogenic fate transition perturbed by loss of HMGB2. *Front Mol Neurosci* 10:1–10.
56. Zhang Y, et al. (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* 34:11929–11947.
57. Kim J, et al. (2018) Ttyh1 regulates embryonic neural stem cell properties by enhancing the Notch signaling pathway. *EMBO Rep* 19:1–12.
58. Jung E, et al. (2017) Tweety-homolog 1 drives brain colonization of gliomas. *J Neurosci* 37:6837–6850.
59. Zamanian JL, et al. (2012) Genomic analysis of reactive astrogliosis. *J Neurosci* 32:6391–6410.
60. Liddelow SA (2019) Modern approaches to investigating non-neuronal aspects of Alzheimer's disease. *FASEB J* 33:1528–1535.
61. Levison SW, Jiang FJ, Stoltzfus OK, Ducceschi MH (2000) IL-6-type cytokines enhance epidermal growth factor-stimulated astrocyte proliferation. *Glia* 32:328–337.
62. Bergen V, Soldatov RA, Kharchenko PV, Theis FJ (2021) RNA velocity—current challenges and future perspectives. *Mol Syst Biol* 17:1–9.
63. Lin YL, Persaud SD, Nhieu J, Wei LN (2017) Cellular retinoic acid-binding protein 1 modulates stem cell proliferation to affect learning and memory in male mice. *Endocrinology* 158:3004–3014.

64. Maden M (2007) Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat. *Rev Neurosci* 8:755–765.
65. Quesnel-Vallières M, Irimia M, Cordes SP, Blencowe BJ (2015) Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev* 29:746–759.
66. Sakers K, et al. (2021) Loss of Quaking RNA binding protein disrupts the expression of genes associated with astrocyte maturation in mouse brain. *Nat Commun*. 12:1–14.
67. Khan S (2019) IGFBP-2 signaling in the brain: From brain development to higher order brain functions. *Front Endocrinol (Lausanne)* 10:1–14.
68. Elsafadi M, et al. (2016) Transgelin is a TGFβ-inducible gene that regulates osteoblastic and adipogenic differentiation of human skeletal stem cells through actin cytoskeleston organization. *Cell Death Dis* 7:1–14.
69. Yu H, et al. (2008) Transgelin is a direct target of TGF-beta/Smad3-dependent epithelial cell migration in lung fibrosis. *FASEB J* 22:1778–1789.
70. Cole AE, Murray SS, Xiao J (2016) Bone morphogenetic protein 4 signalling in neural stem and progenitor cells during development and after Injury. *Stem Cells Int* 2016:9260592.
71. Kantzer CG, et al. (2021) ACSA-2 and GLAST classify subpopulations of multipotent and glial-restricted cerebellar precursors. *J Neurosci Res* 99:2228–2249.
72. Gao S, Dai Y, Rehman J (2021) A bayesian inference transcription factor activity model for the analysis of single-cell transcriptomes. *Genome Res* 31:1296–1311.
73. Van Der Raadt J, Van Gestel SHC, Kasri NN, Albers CA (2019) ONECUT transcription factors induce neuronal characteristics and remodel chromatin accessibility. *Nucleic Acids Res* 47:5587–5602.
74. Yang R, et al. (2021) POU2F2 regulates glycolytic reprogramming and glioblastoma progression via PDPK1-dependent activation of PI3K/AKT/mTOR pathway. *Cell Death Dis* 12.
75. Mi H, Barres BA (1999) Purification and characterization of astrocyte precursor cells in the developing rat optic nerve. *J Neurosci* 19:1049–1061.
76. O'Sullivan ML, et al. (2017) Astrocytes follow ganglion cell axons to establish an angiogenic template during retinal development. *Glia* 65:1697–1716.
77. Samyesudhas SJ, Roy L, Cowden Dahl KD (2014) Differential expression of ARID3B in normal adult tissue and carcinomas. *Gene* 543:174–180.
78. Lanjakornsiripan D, et al. (2018) Layer-specific morphological and molecular differences in neocortical astrocytes and their dependence on neuronal layers. *Nat Commun* 9.
79. Chamling X, et al. (2021) Single-cell transcriptomic reveals molecular diversity and developmental heterogeneity of human stem cell-derived oligodendrocyte lineage cells. *Nat Commun* 12:1–20.
80. Villanueva P, et al. (2015) Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. *PLoS Genet* 11:1–24.
81. Nudel R (2016) An investigation of NFXL1, a gene implicated in a study of specific language impairment. *J Neurodev Disord* 8:1–5.
82. Stumpf PS, et al. (2017) Stem cell differentiation as a non-Markov stochastic process. *Cell Syst* 5:268-282.
83. Zhang Y, Barres BA (2013). A smarter mouse with human astrocytes. *BioEssays* 35:876–880.
84. Sloan SA, et al. (2017) Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron* 95:779-790.
85. Kriks S, et al. (2012) Floor plate-derived dopamine neurons from hESCs efficiently engraft in animal models of PD. *Nature* 480:547–551.
86. Marotta D, Rao C, Fossati V (2021) Human induced pluripotent stem cell (iPSC) handling protocols: Maintenance, expansion, and cryopreservation. *Methods Mol Biol* doi: 10.1007/7651_2021_358. Online ahead of print.
87. Barbar L, Rusielewicz T, Zimmer M, Kalpana K, Fossati V (2020) Isolation of human CD49f+ astrocytes and in vitro iPSC-based neurotoxicity assays. *STAR Protoc* 1, 100172.
88. Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. 19:15.

SUPPLEMENTARY INFORMATION

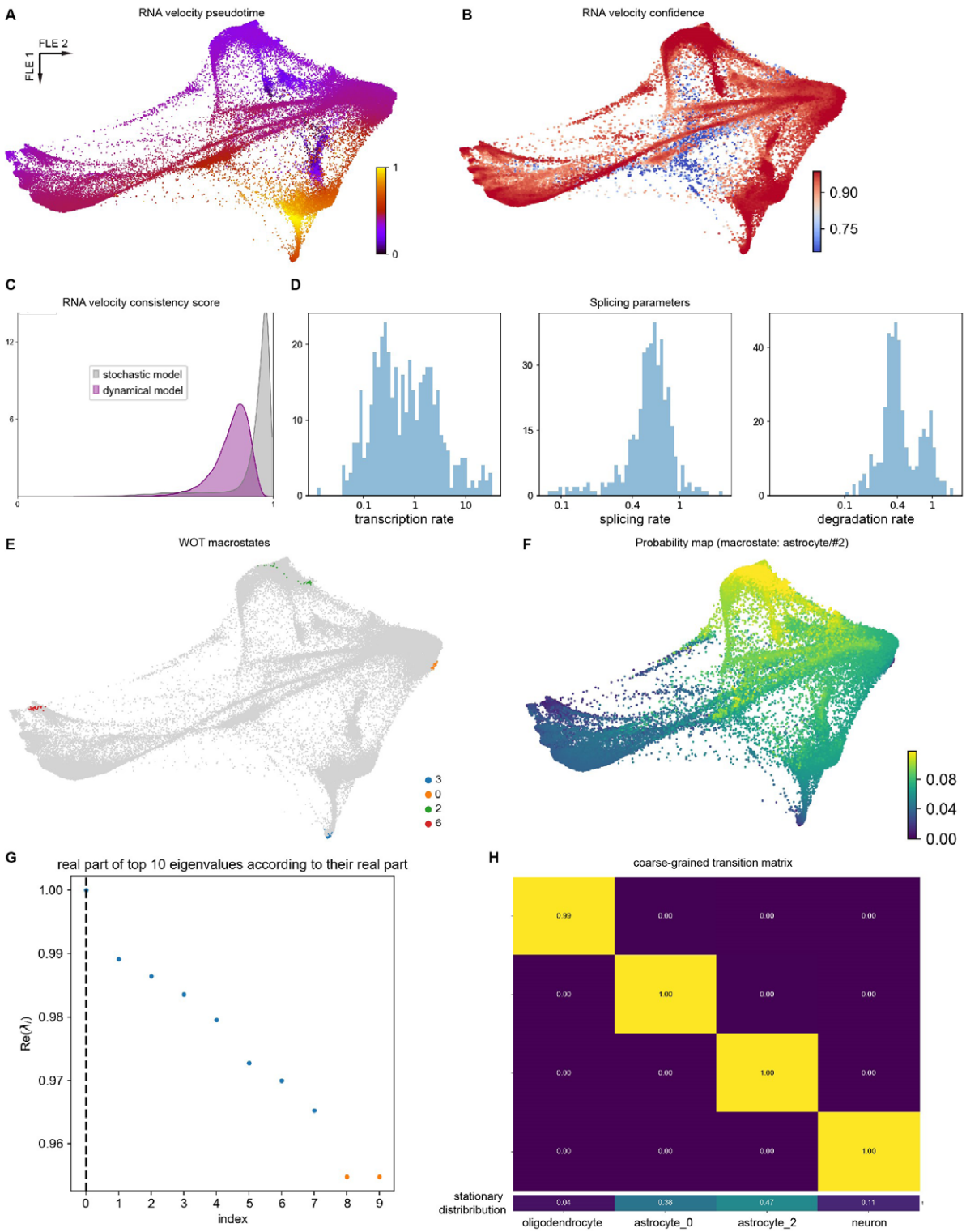
Supplementary Figures

Fig S1 | Differentiation overview and scRNAseq analysis metrics.
Fig S2 | RNA velocity and Waddington Optimal Transport (WOT) analysis.
Fig S3 | Cell type feature plots.
Fig S4 | Reclustering and analysis of transient states in human glial differentiation.
Fig S5 | Differentiation overview and scRNAseq analysis metrics.
Fig S6 | RNA velocity and Waddington Optimal Transport (WOT) analysis of mouse differentiation data.
Fig S7 | Cell type feature plots.
Fig S8 | RNA velocity and Waddington Optimal Transport (WOT) analysis of subsets of mouse differentiation data.
Fig S9 | RNA velocity and Waddington Optimal Transport (WOT) analysis of subset mouse differentiation data integrated with previously published P3 primary mouse astrocyte data.

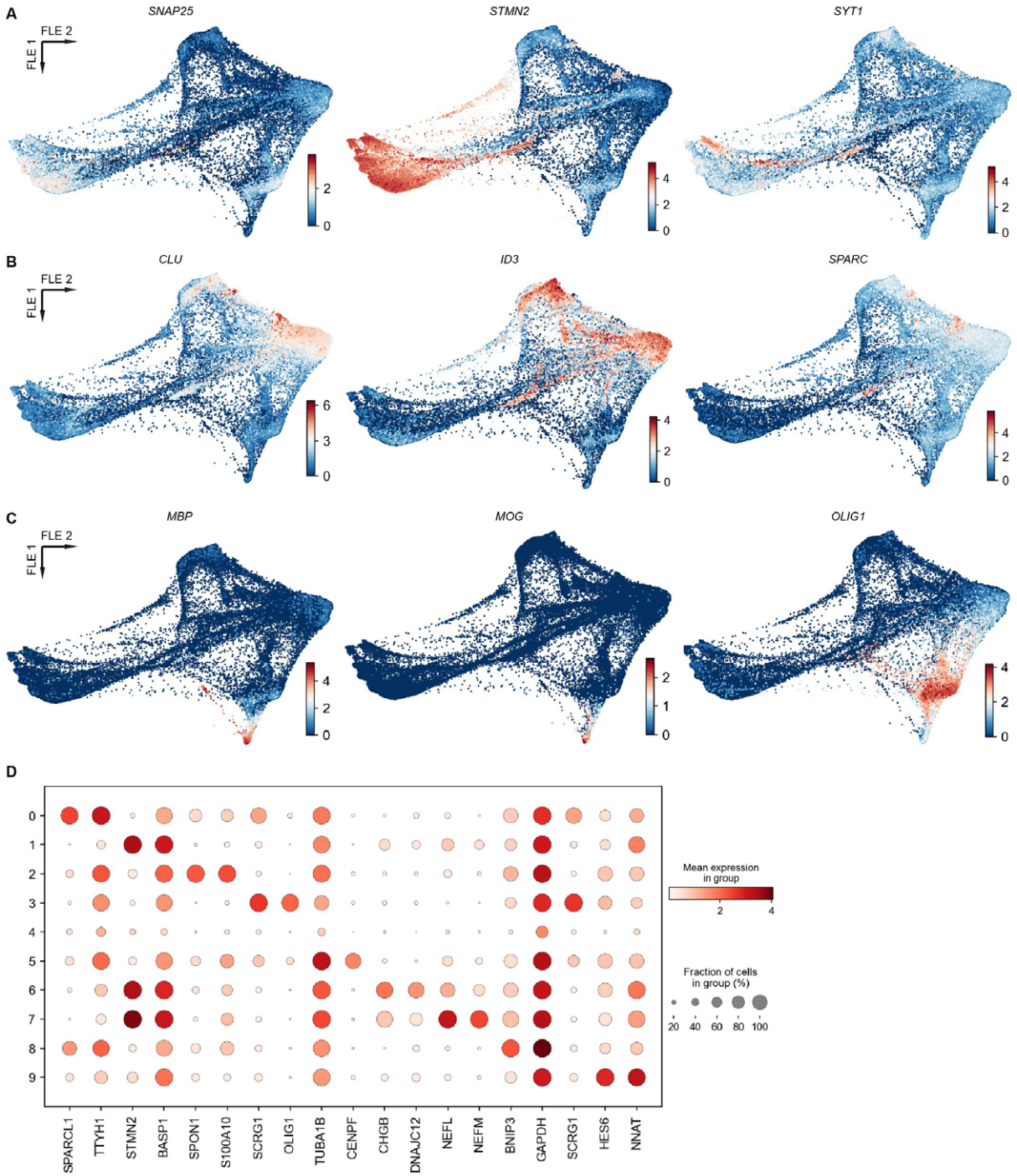


	this study				Barbar et al., 2020	
	30d #1	30d #2	50d #1	50d #2	73d (CD49f+)	73d (unsorted)
preMUSCAT	12,175	9,598	5,030	9,349	13,451	9,843
postMUSCAT	9,485	7,515	4,248	5,421	9,075	7,762
% cells retained	77.9%	78.3%	84.5%	58.0%	67.5%	78.9%

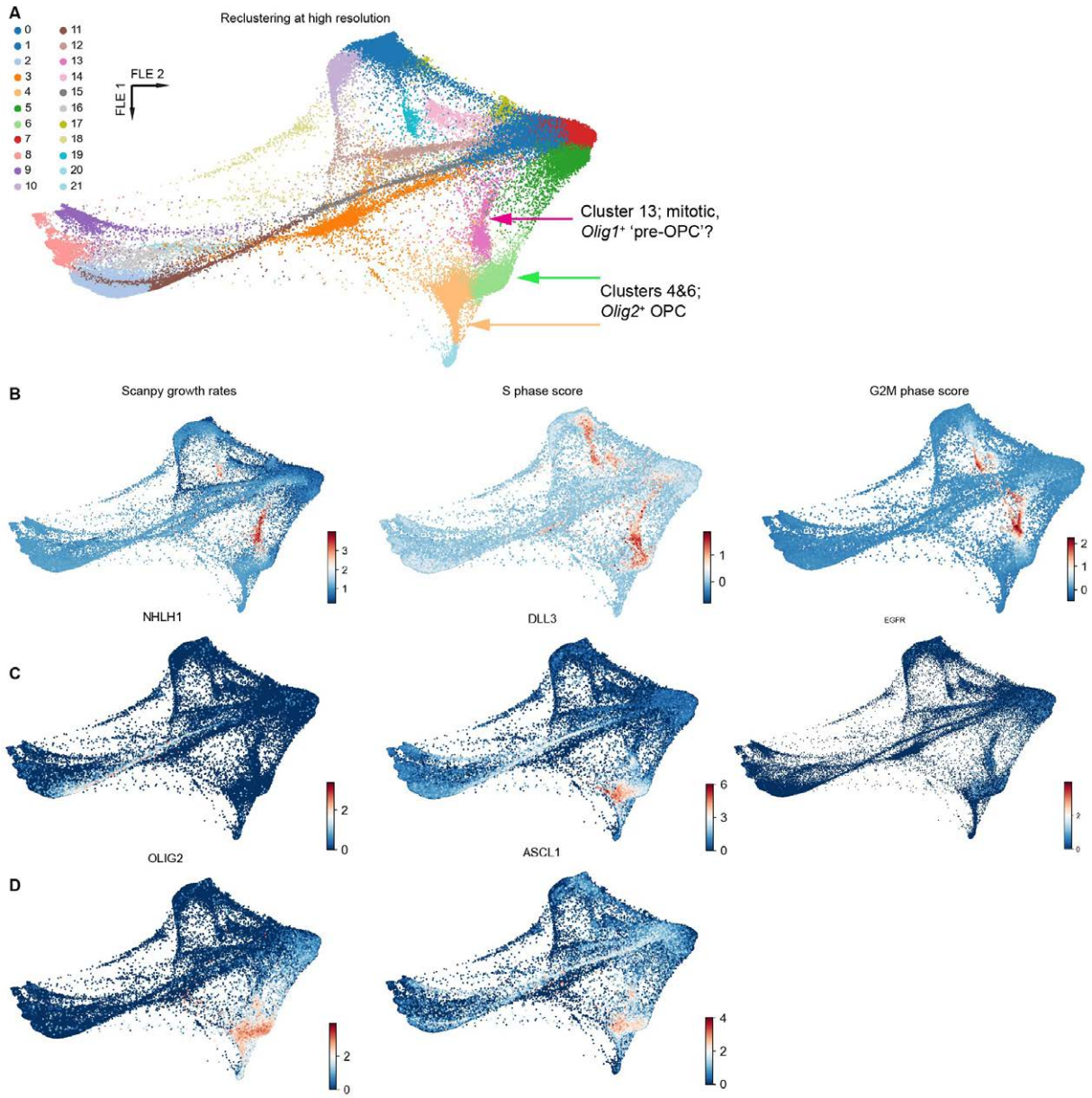
Supplementary Figure 1 | Differentiation overview and scRNAseq analysis metrics. **A.** Detailed human differentiation overview, with phase contrast micrographs of cells at the timepoints analyzed via scRNAseq. **B.** Table of quality control processing statistics for the newly produced data in this study and for the reanalyzed data from37. **C.** Number of genes identified for each cell and **D.** number of unique molecular identifier data for each cell. **E.** Harmony integration of two independently harvested and processed samples from the same human timepoint (30day) shows no evidence of batch effect.



Supplementary Figure 2 | RNA velocity and Waddington Optimal Transport (WOT) analysis. **A.** Pseudotime calculated based on the stochastic model of RNA velocity. **B.** Confidence in RNA velocity calculated for each cell based on local coherence of velocity vectors[50]. **C.** Velocity consistency scores for each cell calculated for each of the two RNA velocity models employed above. **D.** Splicing parameters calculated using the dynamic model of RNA velocity show roughly normal distributions across all cells. **E.** Four macrostates detected using the WOT algorithm. 0: astrocytes, 2: astrocytes, 3: oligodendrocytes, 6: neurons. **F.** Probability map for the astrocytes_2 fate shows (compare with astrocytes_0 in Fig 1F). **G.** Plot of eigenvalues derived from the Schur decomposition of the WOT coarse transition matrix. "Elbows" are visible after the first eigenvalue, and at the 5th eigenvalue. **H.** Plot of coarse-grained transition matrix depicts high probability to stay in one of the 4 macrostates upon entering the state (yellow boxes).



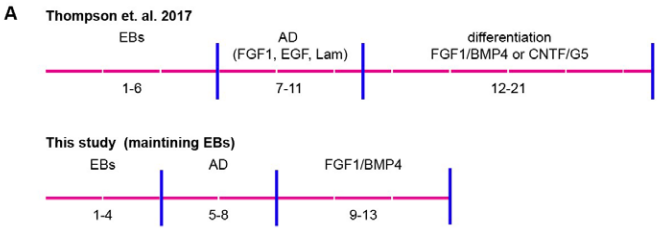
Supplementary Figure 3 | Cell type feature plots. **A.** Feature maps for neuron cell-type marker genes (*STMN2*, *SYT1*, *SNAP25*). **B.** Feature maps for astrocyte cell-type marker genes (*CLU*, *SPARC*, *ID3*). **C.** Feature maps for oligodendrocyte cell-type marker genes (*MOG*, *MBP*, *OLIG1*).

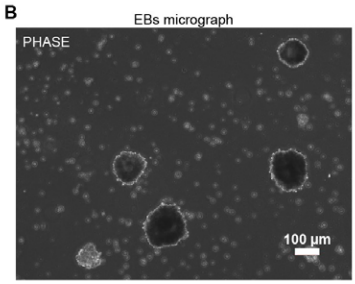


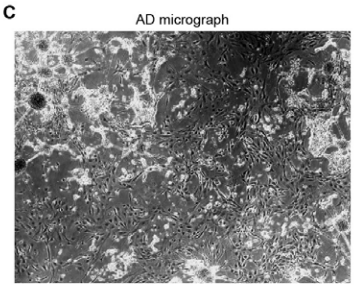
Supplementary Figure 4 | Reclustering and analysis of transient states in human glial differentiation. **A.** FLE reduction of the same data from Fig 1, reclustered at higher resolution (1.2). **B.** Feature maps for mitotic scores from Scanpy[88] ("Scanpy growth rates") and from scVelo[50] ("S phase score" and "G2M phase score"). **C.** Feature maps for genes enriched in shared neuron/astrocyte precursors (*NHLH1*, *DLL3*, *EGFR*). **D.** Feature maps for genes enriched in astrocyte/oligodendrocyte precursors (*OLIG2*, *ASCL1*, *DCX*).

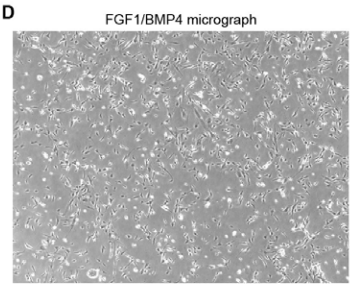
(opposite page)

Supplementary Figure 5 | Differentiation overview and scRNAseq analysis metrics. **A.** Mouse differentiation overview, with comparison across protocols. **B-D.** Micrographs of cells during the embryoid body (EB), adherent differentiation (AD), and astrocyte growth factor stages. **E.** Table of quality control processing statistics for the mouse scRNAseq data this study. **F.** Harmony integration of two independently harvested and processed samples from the same mouse timepoint (2 day EB) shows no evidence of batch effect. **G.** GFAP immunostaining from differentiations in two different mouse embryonic stem cell lines (mESCs). The A17iCRE mESC line was differentiated for 12 days using ciliary neurotrophic factor (CNTF; see text). The BL6/CAST dPIGA mESC line was differentiated for 12 days using ciliary neurotrophic factor (CNTF). **H.** Number of genes identified for each cell and H) number of unique molecular identifier data for each cell from each timepoint.





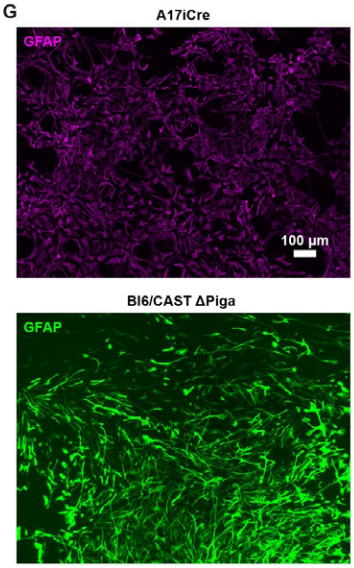


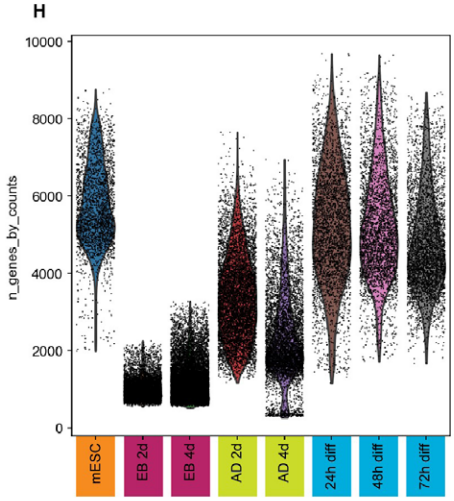


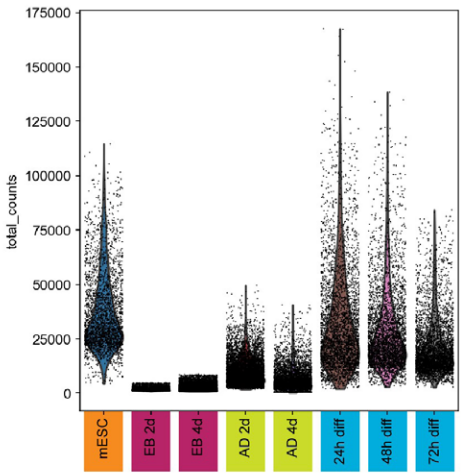
E

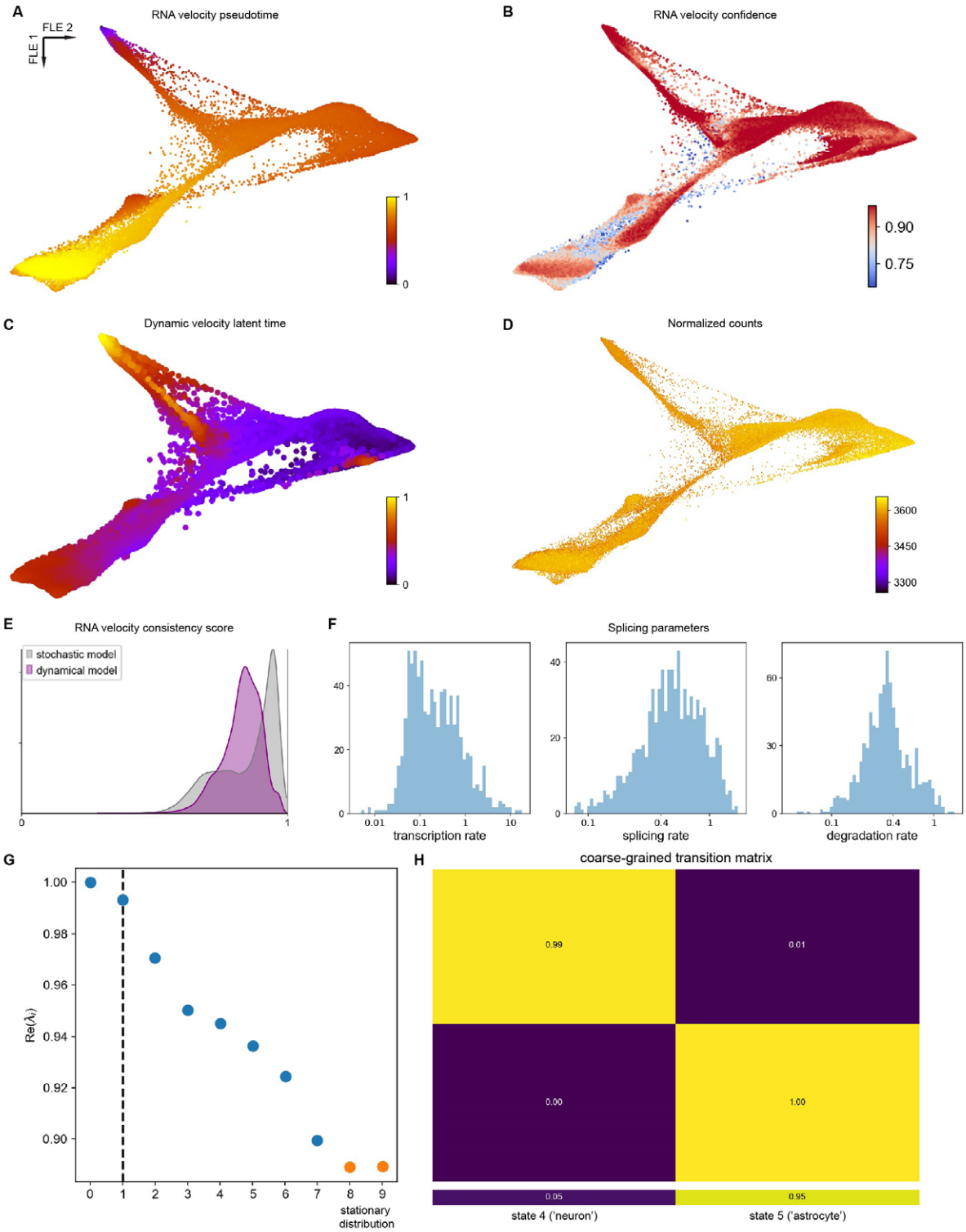
Muscat quality control and filtering - final cell numbers

	mESC	EB 2d	EB 4d	AD 2d	AD 4d	24h diff	48h diff	72h diff
PRE Muscat	4,073	17,729	16,866	8,625	9,233	4,922	5,371	5,899
POST Muscat	2,641	11,580	14,257	7,009	6,360	3,478	3,796	4,249
% cells retained	64.8%	65.3%	84.5%	81.3%	68.9%	70.7%	70.7%	72.0%

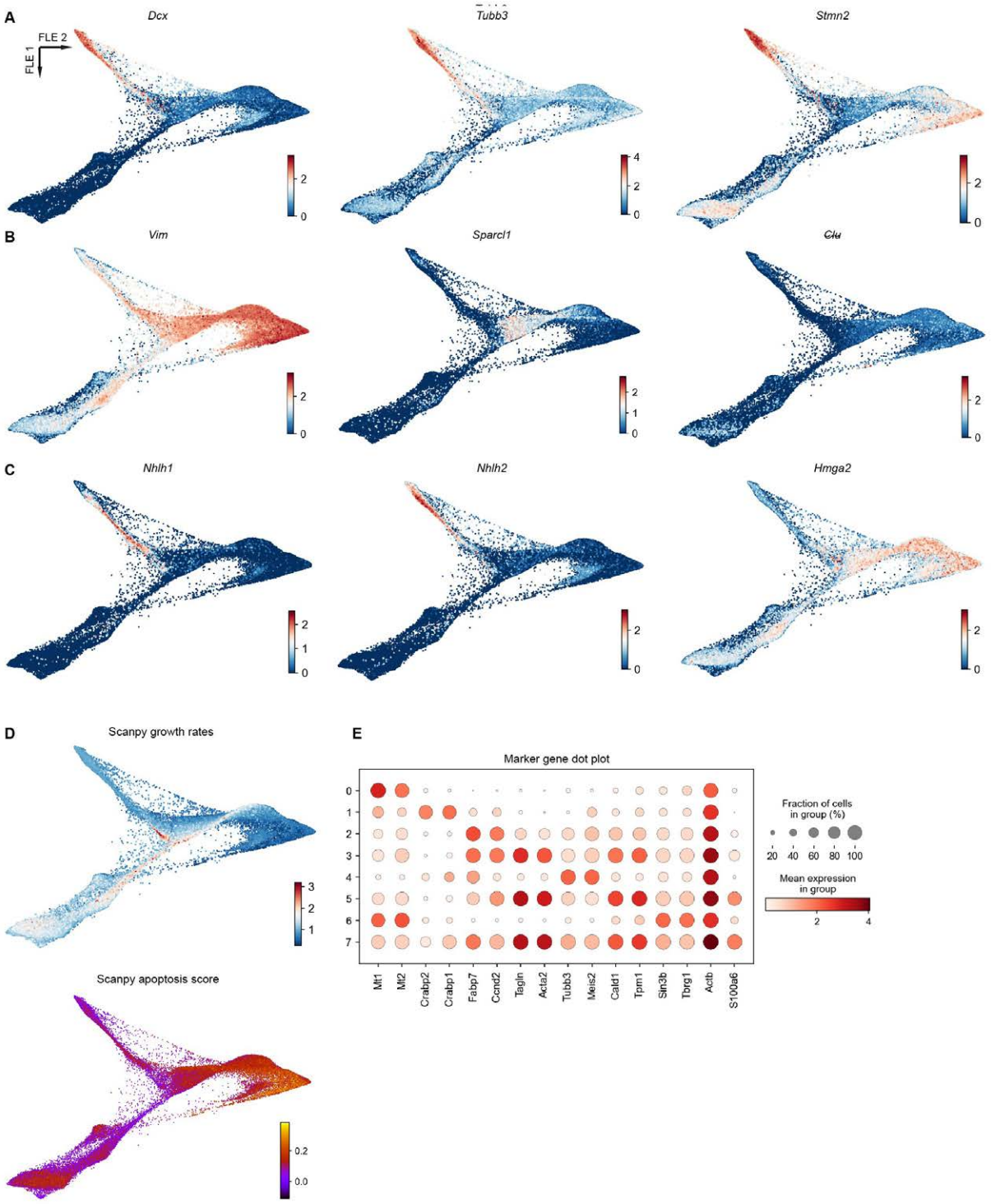




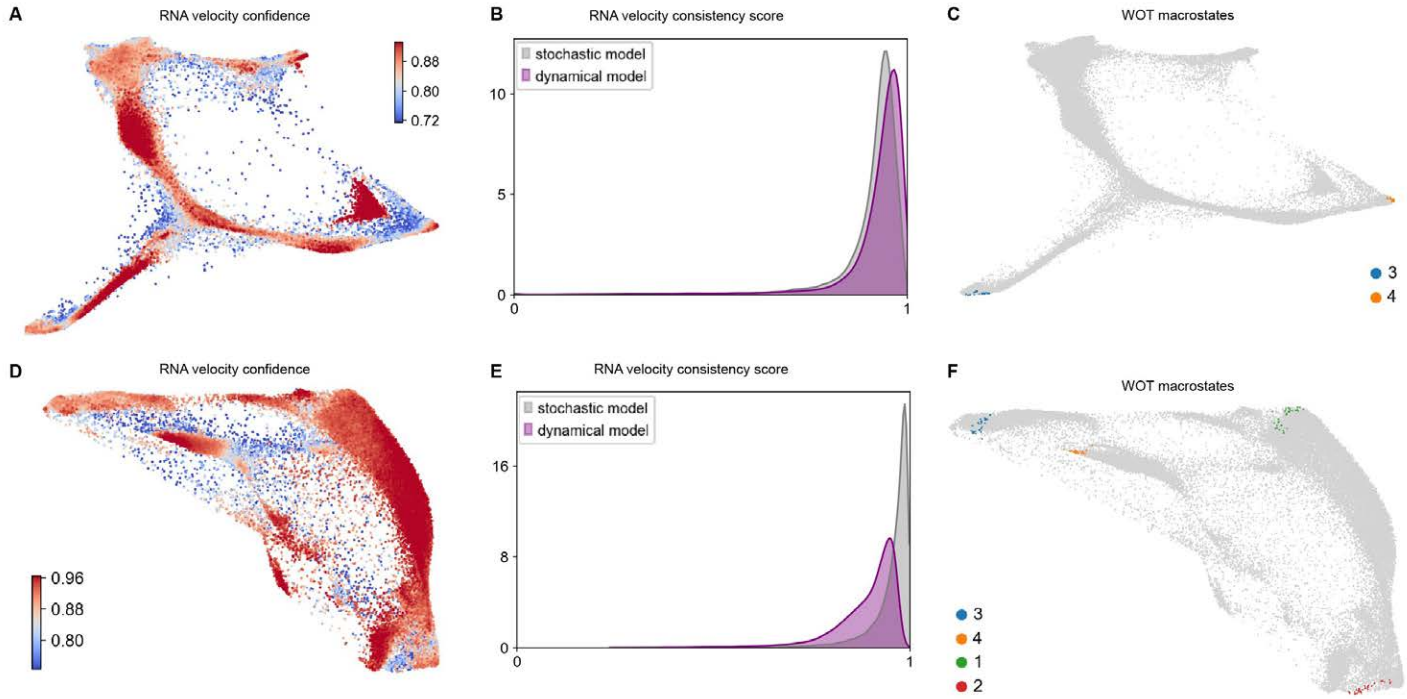




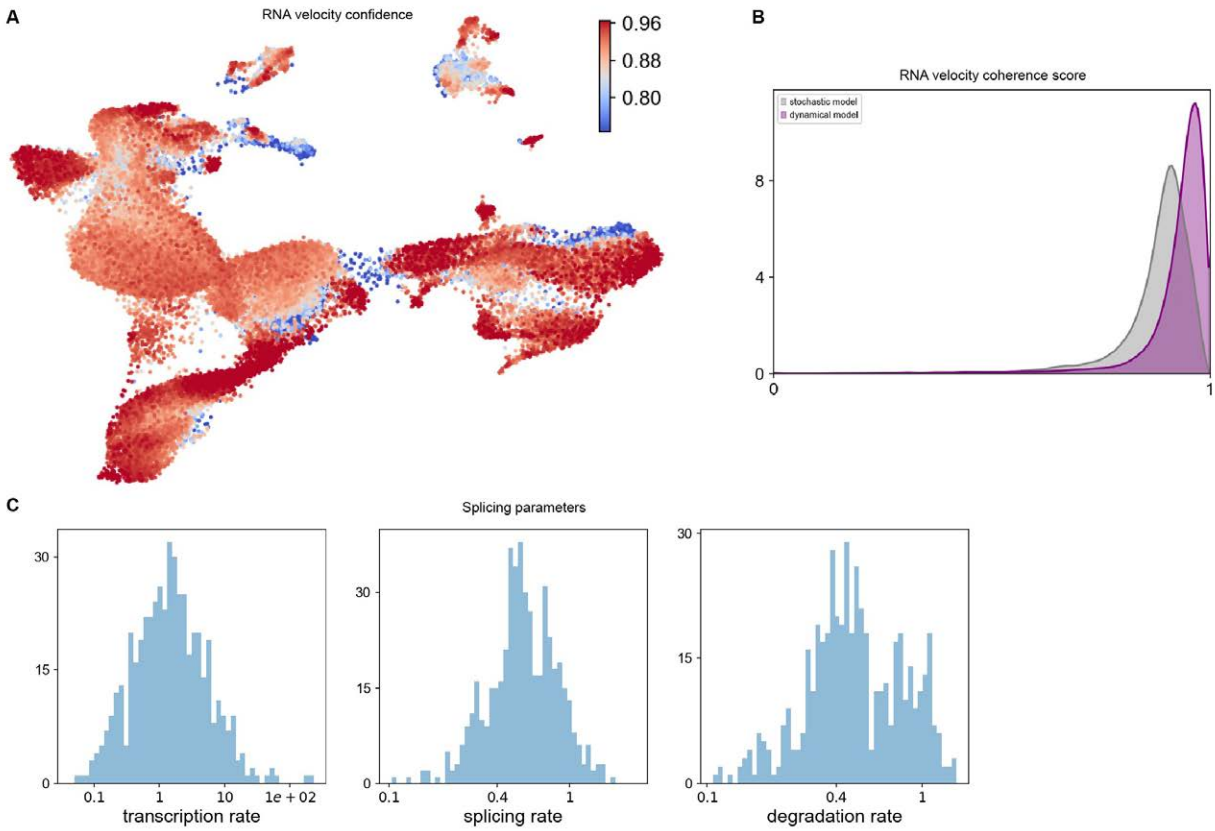
Supplementary Figure 6 | RNA velocity and Waddington Optimal Transport (WOT) analysis of mouse differentiation data. **A.** Pseudotime calculated based on the stochastic model of RNA velocity. **B.** Confidence in RNA velocity calculated for each cell based on local coherence of velocity vectors. **C.** Latent time (similar to pseudotime) calculated based on the dynamic model of RNA velocity. **D.** Feature plot of normalized counts for each cell shows minimal differences across cells/timepoints. **E.** Consistency scores for each cell calculated for each of the two RNA velocity models employed above. **F.** Splicing parameters calculated using the dynamic model of RNA velocity show roughly normal distributions across all cells. **G.** Plot of eigenvalues derived from the Schur decomposition of the WOT coarse transition matrix. "Elbows" are visible after the second and third eigenvalue. **H.** Plot of coarse-grained transition matrix depicts high probability to stay in one of the 2 macrostates upon entering the state (yellow boxes).



Supplementary Figure 7 | Cell type feature plots. **A.** Feature maps for neuron cell-type marker genes (Dcx, Tubb3, Syt1). **B.** Feature maps for astrocyte cell-type marker genes (*Vim*, *Sparcl1*, *Clu*). C. Feature maps for putative astrocyte precursor genes (*Nhlh1*, *Nhlh2*, *Hmga2*). **D.** Scanpy growth and apoptosis scores were calculated for each cell.



Supplementary Figure 8 | RNA velocity and Waddington Optimal Transport (WOT) analysis of subsets of mouse differentiation data. **A.** RNA velocity confidence, **B.** RNA velocity consistency score, and **C.** WOT macrostates identified for the subset of data shown in Fig 3A-G (embryoid body stages + adherent differentiation stages). **D.** RNA velocity confidence, **E.** RNA velocity consistency score and **F.** WOT macrostates identified for the subset of data shown in Fig 3H-L (adherent differentiation stages + FGF1/BMP4 stages).



Supplementary Figure 9 | RNA velocity and Waddington Optimal Transport (WOT) analysis of subset mouse differentiation data integrated with previously published P3 primary mouse astrocyte data. **A.** RNA velocity confidence **B.** RNA velocity consistency score, and **C.** Inferred splicing parameters for the subset of data shown in Fig 4B (adherent differentiation stages + FGF1/BMP4 stages integrated with P3 primary astrocyte data[41]).