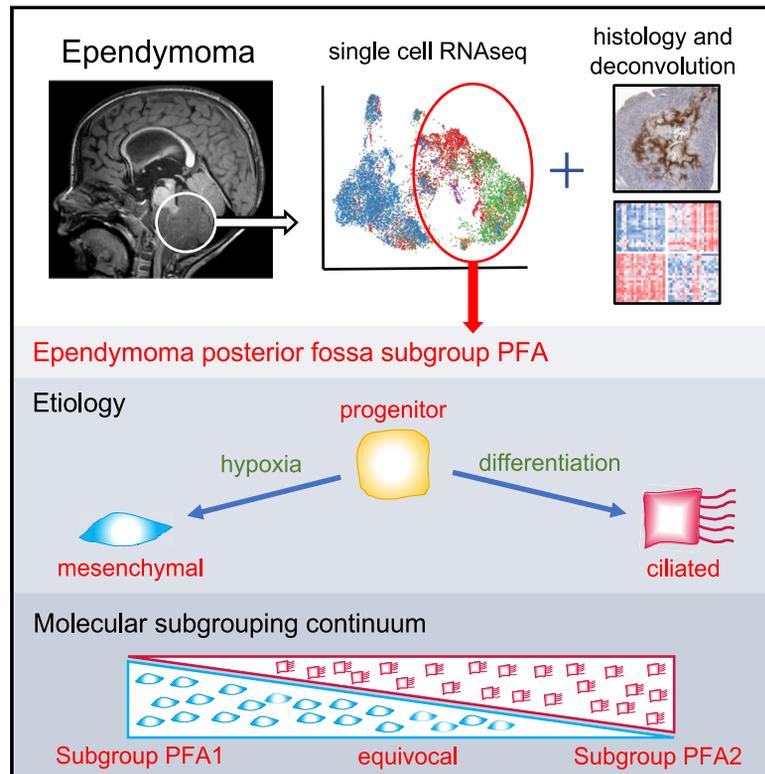


# Cell Reports

## Single-Cell RNA Sequencing of Childhood Ependymoma Reveals Neoplastic Cell Subpopulations That Impact Molecular Classification and Etiology

### Graphical Abstract



### Authors

Austin E. Gillen, Kent A. Riemondy, Vladimir Amani, ..., Jay R. Hesselberth, Nicholas K. Foreman, Andrew M. Donson

### Correspondence

andrew.donson@cuanschutz.edu

### In Brief

Gillen et al. use scRNA-seq to reveal diverse neoplastic subpopulations that underlie the bulk-tumor molecular subgrouping schema in childhood ependymoma. They identify potential divergent lineage trajectories that are driven by either developmental or hypoxic stimuli. This resource facilitates a more accurate interpretation of complex childhood ependymoma biological data.

### Highlights

- Characterization and interactive browser of ependymoma single cells
- Identification of progenitor, ependyma-differentiated, and mesenchymal cell types
- Subpopulation proportions influence bulk-tumor transcriptomic molecular subgrouping
- Progenitor and mesenchymal subpopulations co-localize in peri-necrotic zones



## Resource

# Single-Cell RNA Sequencing of Childhood Ependymoma Reveals Neoplastic Cell Subpopulations That Impact Molecular Classification and Etiology

Austin E. Gillen,<sup>1</sup> Kent A. Riemondy,<sup>1</sup> Vladimir Amani,<sup>2,3</sup> Andrea M. Griesinger,<sup>2,3</sup> Ahmed Gilani,<sup>4</sup> Sujatha Venkataraman,<sup>2,3</sup> Krishna Madhavan,<sup>2,3</sup> Eric Prince,<sup>2,5</sup> Bridget Sanford,<sup>3</sup> Todd C. Hankinson,<sup>2,5</sup> Michael H. Handler,<sup>2,5</sup> Rajeev Vibhakar,<sup>2,3</sup> Ken L. Jones,<sup>3</sup> Siddhartha Mitra,<sup>2,3</sup> Jay R. Hesselberth,<sup>1</sup> Nicholas K. Foreman,<sup>2,3,5,6,7</sup> and Andrew M. Donson<sup>2,3,6,7,8,\*</sup>

<sup>1</sup>RNA Biosciences Initiative, University of Colorado Denver, Aurora, CO 80045, USA

<sup>2</sup>Morgan Adams Foundation Pediatric Brain Tumor Research Program, Children's Hospital Colorado, Aurora, CO 80045, USA

<sup>3</sup>Department of Pediatrics, University of Colorado Denver, Aurora, CO 80045, USA

<sup>4</sup>Department of Pathology, University of Colorado Denver, Aurora, CO 80045, USA

<sup>5</sup>Department of Neurosurgery, University of Colorado Denver, Aurora, CO 80045, USA

<sup>6</sup>Senior author

<sup>7</sup>These authors contributed equally

<sup>8</sup>Lead Contact

\*Correspondence: [andrew.donson@cuanschutz.edu](mailto:andrew.donson@cuanschutz.edu)

<https://doi.org/10.1016/j.celrep.2020.108023>

## SUMMARY

Ependymoma (EPN) is a brain tumor commonly presenting in childhood that remains fatal in most children. Intra-tumoral cellular heterogeneity in bulk-tumor samples significantly confounds our understanding of EPN biology, impeding development of effective therapy. We, therefore, use single-cell RNA sequencing, histology, and deconvolution to catalog cellular heterogeneity of the major childhood EPN subgroups. Analysis of PFA subgroup EPN reveals evidence of an undifferentiated progenitor subpopulation that either differentiates into subpopulations with ependymal cell characteristics or transitions into a mesenchymal subpopulation. Histological analysis reveals that progenitor and mesenchymal subpopulations co-localize in perinecrotic zones. In conflict with current classification paradigms, relative PFA subpopulation proportions are shown to determine bulk-tumor-assigned subgroups. We provide an interactive online resource that facilitates exploration of the EPN single-cell dataset. This atlas of EPN cellular heterogeneity increases understanding of EPN biology.

## INTRODUCTION

Ependymoma (EPN) is a brain tumor commonly presenting in childhood, which remains fatal in most children. Our understanding of the biology underlying this tumor has greatly expanded in the era of genomics, transcriptomics, and methylomics. Initial transcriptomic studies identified supratentorial (ST) and posterior fossa (PF) EPN as distinct biological entities (Taylor et al., 2005). Further studies have identified two major groups within PF EPN, based on transcriptomic (Hoffman et al., 2014; Wani et al., 2012; Witt et al., 2011) and methylomic studies (Mack et al., 2014; Pajtler et al., 2015), termed PF group A (PFA) and PF group B (PFB). Two ST EPN subgroups have also been identified, those with a C11ORF95-RELA gene fusion (RELA) and a rarer subgroup with YAP-MAMLD1 fusions (YAP) (Parker et al., 2014; Pajtler et al., 2015). Bulk-tumor molecular studies have identified distinct cellular processes associated with each of these groups. PFB tumors have been shown to occur in adults and older children and to harbor cilia-associated programs (Pajtler et al., 2015; Witt et al., 2011). On the other hand, PFA commonly arise in younger children, who have a worse prog-

nosis than those with PFB and have been shown to harbor biological processes associated with inflammation (Griesinger et al., 2015; Witt et al., 2011). PFA have been shown to be epigenetically distinct from PFB, harboring a CpG island methylator phenotype (CIMP) (Mack et al., 2014). PFA has been further subdivided into two major subgroups: PFA1 and PFA2 (Pajtler et al., 2018).

Classification of tumors into subgroups and inference of subgroup-specific biological programs based on analysis of bulk-tumor samples is, however, susceptible to misinterpretation. This is apparent from single-cell transcriptomic analysis of glioblastoma multiforme (GBM), which revealed considerable cellular heterogeneity within individual patient samples, including cellular subpopulations with transcriptional programs corresponding to GBM subgroups previously identified by bulk-tumor profiling (Nefitel et al., 2019; Patel et al., 2014). Despite the presence of multiple subpopulations within an individual tumor specimen, the dominant cellular subpopulation dictates bulk-tumor subgrouping and apparent biological phenotype. We, therefore, hypothesized that PFA tumors comprise cellular subpopulations corresponding to both PFA1 and PFA2 transcriptional programs.



A few studies have investigated tumor-cell heterogeneity in EPN. These include identification of a putative EPN cancer stem cell (Taylor et al., 2005) shown to occupy perivascular niches (Calabrese et al., 2007). A recent study used single-cell RNA sequencing (scRNA-seq) to analyze childhood cerebellar tumors, including EPN, and matched these with mouse cerebellar developmental cell lineages (Vladoiu et al., 2019), identifying divergent populations of cells in PFA that resembled prenatal gliogenic progenitors. Immune-cell subpopulations within the EPN tumor microenvironment are associated with clinical outcome and specific molecular subgroups (Donson et al., 2009; Hoffman et al., 2014). To more comprehensively catalog the cellular heterogeneity of EPN, the present study analyzed samples from the spectrum of childhood EPN subgroups using scRNA-seq analysis.

## RESULTS

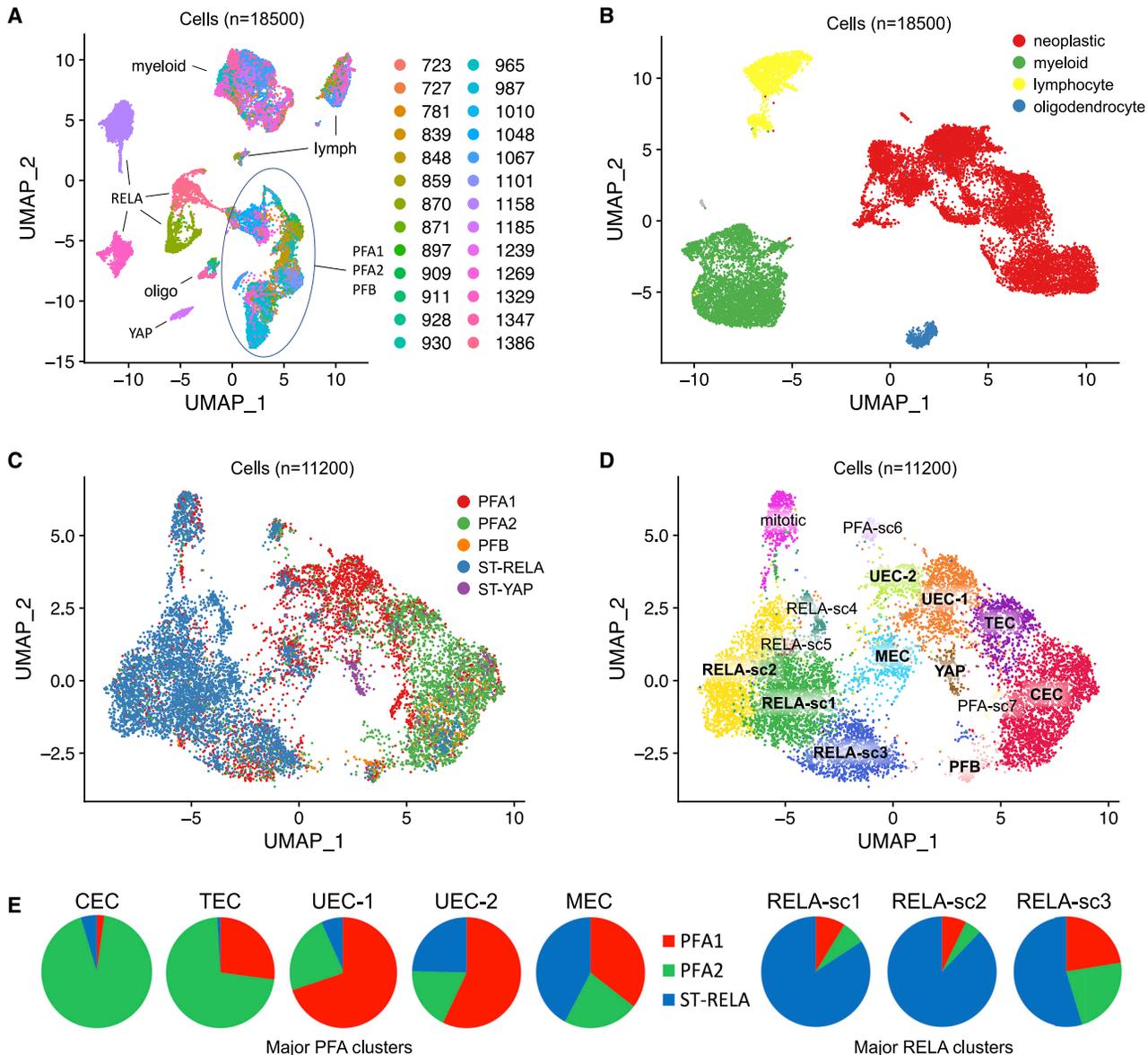
### scRNA-Seq Analysis of Childhood EPN Reveals Classification-Subgroup-Specific Neoplastic Clusters

scRNA-seq was performed on 26 childhood EPN patient samples that had been disaggregated at the time of surgery and viably banked at our institution during a 10-year period. Most of these EPN samples were classified as PFA ( $n = 19$ ) and RELA ( $n = 5$ ) based on bulk-tumor methylome profiling (Table S1). The cohort also included one specimen from the PFB subgroup, which is uncommon in childhood, and a single YAP subgroup sample, which is a rarer variant of ST EPNs. The 19 PFA samples were further assigned to more recently described PFA1 ( $n = 12$ ) and PFA2 ( $n = 7$ ) subgroups based on methylome analysis (courtesy of Dr. Martin Sill, Heidelberg, Germany). All samples were from initial presentation, apart from three of the five RELA samples that were from first recurrences. The Chromium drop sequencing (drop-seq) platform (10x Genomics) was used to generate transcriptomes at the single-cell level. The goal of the study was to capture approximately 2,000 cells per sample, a sufficient number of cells to provide a broad view of the cellular heterogeneity of EPN. Cell Ranger (10x Genomics) and Seurat analyses were used to normalize and filter cells, resulting in 18,500 cells (700+ per sample) that passed quality controls (Figure S1). The reduced number of cells passing quality-control thresholds is likely a result of the relatively abrasive mechanical disaggregation process and cryopreservation that was required to obtain this large EPN sample series. Cell gene-expression matrices were projected as 2D uniform manifold approximation and projection (UMAP) plots, revealing multiple clusters of cells that were either unique to, or shared between, samples (Figures 1A and S2A–S2C). PFA1/PFA2 samples clustered together in an intermixed fashion. In contrast, RELA samples clustered largely according to individual sample of origin, a pattern that has also been seen in high-grade glioma and medulloblastoma scRNA-seq studies (Hovestadt et al., 2019; Neftel et al., 2019; Patel et al., 2014). We applied Harmony alignment (Korsunsky et al., 2018) to the entire EPN dataset to identify potential commonalities in RELA and to correct for inter-sample variations from experimental or sequencing batch effects across the entire dataset. Harmony alignment identified multiple clusters (Figures 1B and S2A–S2C), a number of which

revealed gene expression profiles consistent with subpopulations of non-neoplastic cells: myeloid cells (*SPP1*, *C1QC*, *APOC1*, and *TYROBP*), lymphocytes (*CD3E*, *IL32*, *LCK*, and *GZMA*), and oligodendrocytes (*APC*, *OLIG1*, *OLIG2*, and *CSPG2*). The remaining closely grouped clusters harbored gene-expression profiles indicative of neoplastic EPN cells, such as *TNC* and *AQP4* (Araki et al., 2016; Wang and Owlver, 2011). To further confirm neoplastic versus non-neoplastic cluster designations, we performed inferred copy number variation (CNV) analysis (inferCNV). This showed that non-neoplastic cells portrayed balanced chromosomes, whereas neoplastic cells from samples that showed CNVs by bulk-tumor-methylation analysis (Table S1) revealed a high level of concordance in chromosomal gains and losses (Figure S3).

For the present study, we focused on neoplastic cells ( $n = 11,200$ ) to determine the extent of cellular heterogeneity and to identify discrete neoplastic subpopulations. Neoplastic cells were separated from non-neoplastic cells and re-clustered using Harmony, again identifying cell clusters derived from multiple PFA1/PFA2 patient samples but, now, also showing clusters derived from multiple RELA patient samples (Figure S2D). Cells revealed a predominant subgroup-specific separation, most notably in the form of two main groups of clusters corresponding to either PFA (PFA1 and PFA2) or RELA samples (Figure 1C). A degree of overlap between PFA and RELA cluster groups was observed, with a portion of cells (19.1%) from PFA samples falling within the RELA clusters group and vice versa (11.6%). To determine whether cells in non-corresponding subgroup clusters were a result of over-alignment, ST and PF samples were re-analyzed individually using Harmony. Cells that had previously fallen into non-corresponding subgroup clusters formed were shown to cluster discretely from most cells (Figure S2E). This result further supports the existence of a portion of non-corresponding subgroup cells in both PFA and RELA and provides verification that Harmony alignment analysis did not over-correct clusters. Smaller subgroup-specific clusters largely comprise either PFB or ST-YAP were also observed (Figure 1C).

Seurat's graph-based clustering approach identified 15 distinct neoplastic clusters, consisting of seven clusters in the PFA cluster group, five in the RELA cluster group, and one each corresponding to the single PFB and YAP samples (Figure 1D). Based on subsequent molecular characterization of scRNA-seq subpopulations, the main five PFA clusters were named ciliated EPN cells (CECs), transportive EPN cells (TECs), undifferentiated EPN cells -1 and -2 (UEC-1s and UEC-2s), and mesenchymal EPN cells (MECs). The three main RELA clusters, which were less well defined are referred to as RELA-sc1, -sc2, and -sc3. Major PFA and RELA scRNA-seq subpopulations predominantly comprised cells from samples that were classified according to their corresponding molecular subgroups (Figure 1E). The remaining minor PFA and RELA clusters (PFA-sc6 and PFA-sc7; RELA-sc4 and RELA-sc5), were excluded from further analysis because all were less than 2% of total cells, and PFA-sc7 and RELA-sc4 were predominantly derived from individual patient samples. A cluster composed of cells distinguished by mitosis-related gene expression, corresponding to cycling cells, was also revealed and comprised cells from most samples.



**Figure 1. scRNA-Seq Analysis of Childhood EPN Reveals Classification Subgroup-Specific Neoplastic Clusters**

(A) Unaligned UMAP projection of single-cell expression data of 26 EPN patient samples reveals neoplastic clusters and non-neoplastic lymphocyte (lymph), myeloid, and oligodendrocyte (oligo) clusters.

(B) Harmony alignment showing collapsed neoplastic clusters and discrete non-neoplastic clusters.

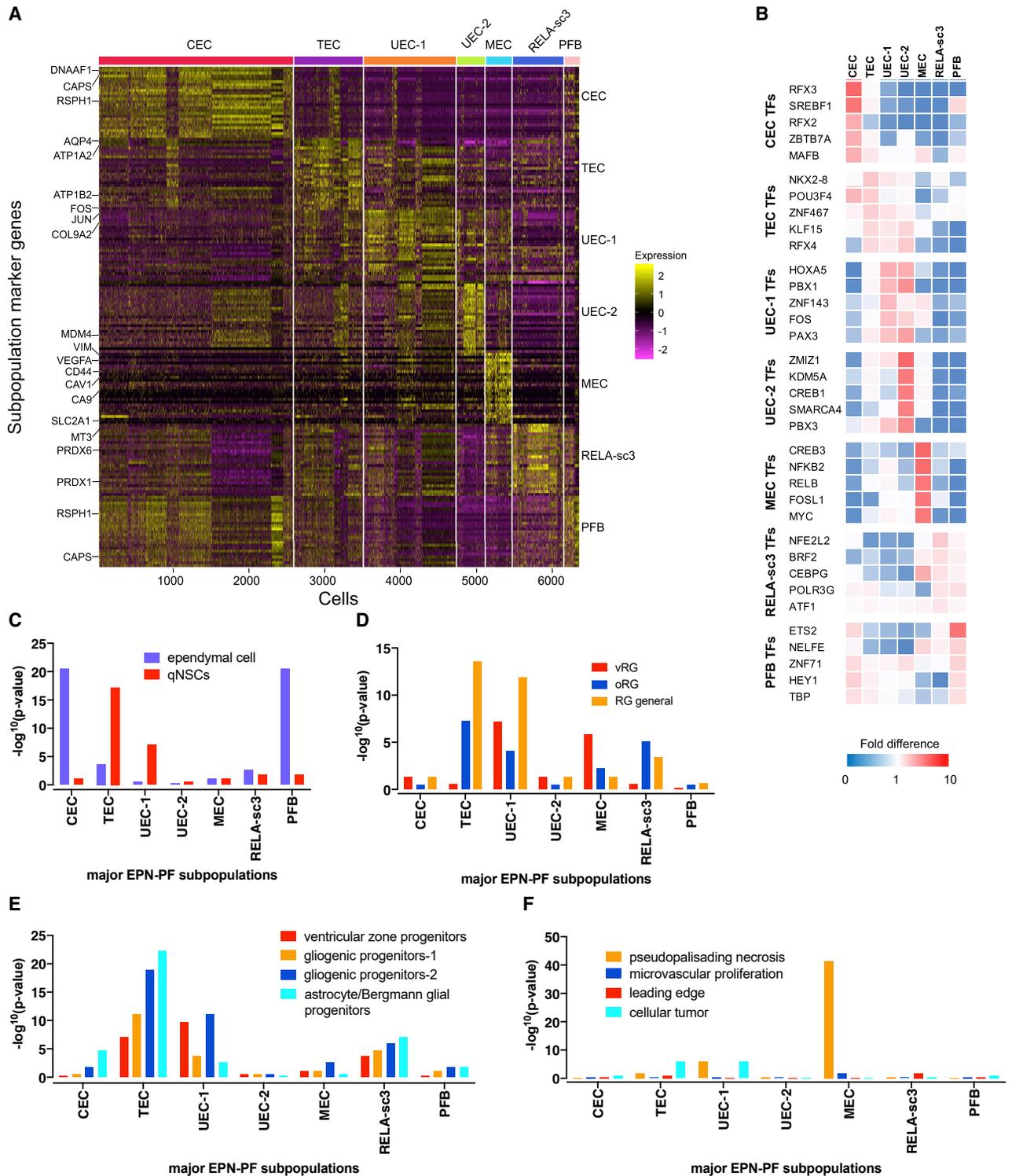
(C and D) Harmony re-alignment of neoplastic cells, colored by (C) classification subgroup and identified clusters (D). See also Figures S2.

(E) Subgroup proportions in major PFA and RELA clusters.

### PFA Conceals Multiple, Discrete, Intra-tumoral Neoplastic Subpopulations

Neoplastic-cluster subpopulations were characterized by examination of individual marker genes, geneset enrichment analysis, and inferred transcription-factor regulatory network activation. Marker genes for each of the major PFA and RELA subpopulations were generated by differential gene-expression analysis (Figure 2A; Table S2). Looking first at the major PFA subpopulations, the most abundant PFA subpopulation CECs was distinguished by cilia-associated gene expression, including

*DNAAF1*, *RSPH1*, and *CAPS* (Figure 2A; Table S2), which we had previously identified as an EPN-specific marker (Amari et al., 2017). Multiple cilia-related gene ontology (GO) terms were significantly enriched in the CEC subpopulation (Table S3). We inferred active transcription factors in individual neoplastic cell transcriptomes using SCENIC (single-cell regulatory network inference and clustering) (Aibar et al., 2017) and identified discrete transcription factor activity profiles for each major PF and ST subpopulation (Figure 2B; Table S4). Several CEC-specific transcription factors were identified, most notably



**Figure 2. PF EPN Conceals Multiple Discrete, Intra-tumoral Neoplastic Subpopulations**

(A) Heatmap of gene upregulated in major PF subpopulations with key subpopulation signature genes indicated (full list in Table S2).

(B) Heatmap of the most significant neoplastic subpopulation-specific transcription factor (TF) regulatory networks identified using SCENIC (single-cell regulatory network inference and clustering) analysis and ranked by fold difference (full list in Table S4).

(legend continued on next page)

RFX2 and RFX3, which are critical for normal ependymal development (Baas et al., 2006; Chung et al., 2012). *RFX2* was also previously identified as an EPN-associated super-enhancer gene in a study that performed bulk-tumor mapping of active chromatin landscapes (Mack et al., 2018). The abundant cilia-related gene expression in CECs recapitulates the multi-ciliated phenotype of normal ependymal cells responsible for coordinated movement of cerebrospinal fluid in the ventricles. CEC markers were also specifically enriched with ependymal cell-specific genes (fold change [FC] = 30.2) (Figure 2C), which have been defined by scRNA-seq analysis of normal ependymal cells and quiescent neural stem cells in a transgenic mouse model (Shah et al., 2018).

Similar to CECs, the TEC subpopulation revealed molecular features of ependymal cell differentiation, namely expression of transporter function-related marker genes, such as aquaporins (*AQP4* and *AQP1*), which have previously been documented in EPN (Wang and Oowler, 2011), and ion transporters, including ATPase Na<sup>+</sup>/K<sup>+</sup> transporters (*ATP1A2* and *ATP1B2*) (Figure 2A; Table S2). The “ion transmembrane transport” GOTERM geneset was, accordingly, highly enriched in TECs (Table S3). The transportive phenotype in this subpopulation recapitulates another function of normal ependymal cells, namely transport of molecules between the ventricle lumen and the parenchyma of the brain. Collectively, these data suggest that TECs and CECs represent differentiated PFA subpopulations.

Subpopulations UEC-1 and UEC-2 lacked ependymal-function phenotypes, as indicated by negligible cilia- or transporter-related geneset enrichment (Table S3) and were, therefore, termed “undifferentiated ependymoma cells-1” (UEC-1) and UEC-2, respectively. UEC-2 and UEC-1 showed enrichment for GOTERMs related to transcriptional regulation and osteoblast differentiation (Table S3). UEC-1 was distinguished from UEC-2 by enrichment of “somatic stem cell population maintenance” GOTERM genes (including *HES1*, *PBX1*, and *SOX9*). Conversely, UEC-2 was distinguished by RNA splicing-related GOTERM enrichment. UEC-1 showed upregulated gene expression and transcription factor activity in early response genes *FOS* and *JUN* (Figures 2A and 2B; Tables S2 and S4). These established oncogenes, with roles in immediate early response to trauma in normal ependyma (Katano et al., 1998), have not previously been associated with EPN biology. The apparent undifferentiated phenotype in UEC-1s and UEC-2s prompted us to examine potential overlap of these populations with neurodevelopmental cell lineages. UEC-1 and UEC-2 marker genes and activated transcription factors included *PAX3*, which is expressed during early neurodevelopment, notably being expressed in the embryonic ventricular lining but being absent from ependyma at later stages of development (Goulding et al., 1991). *PAX6*, expressed predominantly by UEC-2, is a marker of radial glia differentiation (Götz et al., 1998; Suter et al., 2009), providing a link to previous studies that identified radial glia-like cells as putative cancer stem cells in EPN (Taylor et al., 2005). PF subpopulation marker genes

were, therefore, compared with scRNA-seq transcriptional profiles of gestational human neocortical radial glia, which provided us with a close approximation of cerebellar radial glia. The scRNAs radial glia dataset included profiles of ventricular (vRG) and outer subventricular zone (oRG) radial glia subpopulations, which are located in distinct niches and have distinct neurodevelopmental roles (Pollen et al., 2015). General radial glia genes were significantly enriched in UEC-1 (FC = 27.7) and TEC (FC = 30.5) but not UEC-2 subpopulations (Figure 2D), suggesting that the former are related to the radial glia lineage. UEC-1s were distinguished from TECs by exclusive enrichment of vRG-specific genes (FC = 18.8). In contrast, TECs had a greater enrichment of oRG-specific genes (FC = 12.1) than UEC-1s had (FC = 8.1). Previously defined EPN and neural cancer stem cell markers *CD133* (*PROM1*), *nestin* (*NES*), *EMX2*, *NOTCH1*, *SOX2*, and *HES1* were examined, but none, apart from *HES1*, were strongly restricted to UEC-1s (Figure S4A). Recent mapping of mouse cerebellar developmental lineages by scRNA-seq provided a comprehensive dataset that was used to predict putative cells of origin in childhood cerebellar tumors, showing that PFA most closely mirrored mouse gliogenic progenitors, but with subpopulations resembling roof-plate-like stem cells (Vladoiu et al., 2019). We compared our PF subpopulations to these mouse cerebellar lineages and found that, among these subpopulations, UEC-1s and TECs showed the greatest overlap with mouse cerebellar developmental lineages, most notably ventricular zone progenitors, gliogenic progenitors-1, gliogenic progenitors-2, and astrocyte/Bergmann glia progenitors, but not roof-plate-like stem cells (Figure 2E; Table S5A). UEC-1s showed a greater enrichment of mouse ventricular zone progenitors (FC = 17.7) than TECs did (FC = 14.2). In a similar pattern to oRG, gliogenic progenitor-2 and gliogenic progenitor-1 genes were significantly enriched in TECs (FC = 28.4 and FC = 19.5, respectively) but to a lesser extent in UEC-1s (FC = 19.5 and FC = 8.9, respectively). This also matches the enrichment pattern of mouse ependymal layer quiescent neural stem cell genes that were greater in TECs (FC = 26.6) than UEC-1s (FC = 14.2) (Figure 2C). TECs, and to a much lesser extent UECs, were enriched for astrocyte/Bergmann glia progenitor genes (FC = 31.9).

Vladoiu et al. (2019) used pseudotime trajectory analysis of mouse cerebellar progenitor populations to demonstrate that ventricular zone progenitors give rise to gliogenic precursors and then astrocyte/Bergmann glia progenitors. UEC-1s may, therefore, recapitulate this process as the PFA progenitor subpopulation, which gives rise to TECs. Collectively, UEC-1-specific enrichment of human vRG and mouse ventricular zone progenitor genes suggest a potential common ventricular zone cell of origin for PFA tumors. Indeed, ventricular radial glia and ventricular zone progenitors may represent the same population of cells, because midneurogenesis radial glia has been shown to constitute most ventricular zone progenitors in most CNS regions (Hartfuss et al., 2001). UEC-2s exhibited no significant enrichment of radial glia or gliogenic progenitor genes,

(C–F) Neoplastic subpopulation enrichments (hypergeometric test) of genesets from (C) normal ependymal layer cells genesets from scRNA-seq characterization (Abbreviation: qNSC, quiescent neural stem cell); (D) human radial glia from scRNA-seq characterization (Abbreviations: vRG, ventricular radial glia; oRG, outer radial glia); (E) mouse cerebellar developmental cell lineages isolated by scRNA-seq (full dataset in Table S5A); and (F) GBM intratumoral anatomic structures.

suggesting that despite molecular similarities to UEC-1, this subpopulation has a more-differentiated phenotype.

MEC was characterized by stress-response gene-expression programs, namely angiogenesis, hypoxia, and glycolysis, including genes such as *VEGFA*, *CA9*, *CHI3L1*, and *CD44* (Figure 2A; Table S2). These genes have previously been described as markers of a mesenchymal phenotype in EPN (Wani et al., 2012). The nuclear factor  $\kappa$ B (NF- $\kappa$ B2) transcription factor activity inferred in MECs (Figure 2B) is consistent with a response to hypoxia and other cellular stress. MECs were enriched for GBM pseudopalisading necrosis genes (<http://glioblastoma.alleninstitute.org>) (FC = 49.7) (Figure 2F). Collectively, these data suggest that MECs are undergoing epithelial-mesenchymal transition (EMT), a phenotype that has recently been described in PFA (Malgulwar et al., 2018) as a response to hypoxic stress, which is also inferred by the presence of necrosis commonly seen in PFA tumors.

PFA tumors were also shown to harbor a portion of cells that fell into RELA clusters, predominantly RELA-sc3 and to a lesser extent RELA-sc1 and RELA-sc2 (Figure 1E). These major RELA subpopulations are described in more detail in the following section. Being the most abundant RELA subpopulation in the context of PFA, RELA-sc3s were designated as a major PFA subpopulation. Comparison of RELA-sc3 with PFA subpopulations revealed that RELA-sc3 shared similarities with TEC with respect to radial glial cells and mouse cerebellar lineages, showing enrichment of genes associated with oRGs (FC = 9.4) and astrocyte/Bergmann glia progenitors (FC = 14.2) (Figures 2D and 2E).

We compared our PFA subpopulations to existing PFA scRNA-seq data (n = 4) that was available from the recent study by Vladoiu et al. (2019). Their study had identified heterogeneous neoplastic cell subpopulations in each sample; two of which showed significant overlap with the major PFA neoplastic subpopulations identified in this study (Figure S4B).

### RELA Tumors Harbor Less-Diverse Intra-tumoral Subpopulations Than PFAs

Four major RELA subpopulations, RELA-sc1, RELA-sc2, RELA-sc3, and MEC, emerged from Harmony analysis of the five RELA scRNA-seq samples, with all but one containing all four subpopulations (Figure S5A). RELA samples showed less intra-tumoral heterogeneity than PFAs showed, and RELA subpopulations showed less-distinctive transcriptomic profiles than PFA subpopulations. Three of the five RELA samples were obtained from first recurrences and, therefore, were previously treated with radiation, but those samples did not cluster discretely from those RELAs obtained at initial presentation. The differential gene expression heatmap for the main RELA subpopulations (Figure S5B) demonstrated overlapping expression of genes across all three major RELA subpopulations, apart from MECs, such as nervous system developmental gene *HES4*, which was ubiquitously expressed (Table S2). RELA-sc1 and -sc2 shared enrichments for “skeletal system development” and associated gene ontologies, including such genes as *COL9A3* and *COMP*, and “negative regulation of Wnt signaling,” including genes *NOTUM* and *WIF1* (Table S3). Additionally RELA-sc1 and -sc2 showed overlap in inferred transcription-factor activity, notably

RELA (Figure S5C). The predominant geneset enrichment distinguishing RELA-sc1, the most abundant RELA subpopulation, was nervous system development genes, such as *HES4*, *MDK*, *HES6*, and *NEUROD2* (Table S3). RELA-sc2 marker genes were enriched for Wnt signaling pathway genes, including activating Wnt ligands *WNT3A* and *WNT7B*. The RELA-sc3 subpopulation was distinguished by enrichment of ontologies related to oxidative stress response with upregulation of *PRDX1*, *PRDX2*, *SOD1*, and *MT3*. Because of the less-distinctive transcriptomic characteristics of these RELA subpopulations, they were not renamed based on inferred phenotypes, and analysis of these subpopulations was not pursued as fully as that of the PFA subpopulations.

Significant overlap of RELA-sample-derived cells with predominantly PFA subpopulation MECs was also observed (Figure 1E). In the context of RELA tumors, MECs likely are also likely present as a result of hypoxic stress because necrotic zones exist in RELA as well as in PFA tumors.

### PFB and YAP Tumor Subpopulations Recapitulate Previous Findings with Enrichment in the Markers of Ependymal Differentiation and Quiescent Neural Stem Cells, Respectively

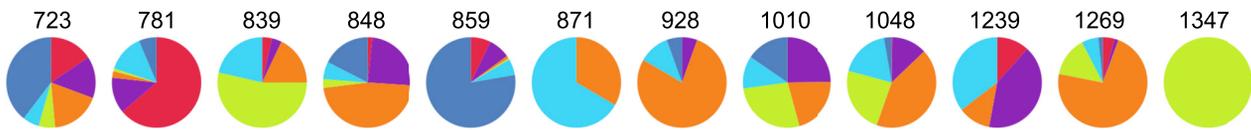
Similar to CECs, PFB marker genes were also enriched for cilia-associated genes (Tables S2 and S3), including *CAPS* and *RSPH1*, and showed mouse ependymal-cell gene enrichment (FC = 30.2) (Figure 2C). This is consistent with the cilia gene expression previously described in PFBs, a marker of differentiation that underlies their more indolent clinical behavior (Witt et al., 2011).

The YAP subpopulation showed upregulation of consensus YAP marker *ARL4D* (Figure S5B; Table S2) (Pajtler et al., 2015) and transcription factors distinct from RELA samples, such as *FOXG1* (Figure S5C; Table S4), which is specifically activated in the forebrain, a common site of YAP presentation (Andreiulo et al., 2019). YAP subpopulation markers were also significantly enriched for quiescent neural stem cell (qNSC) (FC = 8.9) and RG (FC = 19.4) genesets (Figures S5D and S5E).

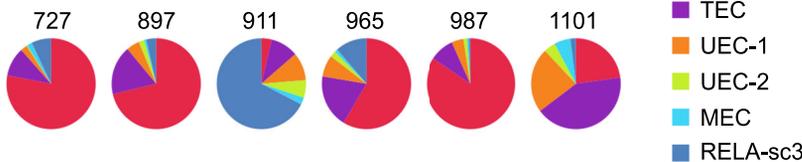
### PFA Neoplastic Subpopulations Are Transcriptomically Related to Classification Subgroups

Each individual sample of the PFA sample was shown to comprise at least two of the six major PFA subpopulations (18 of 19 PFAs) with only a single PFA sample being composed of a single subpopulation (Figure 3A). Over half of the PFA samples (12 of 19) contained all six major PFA subpopulations (6 of 12 PFA1s and 6 of 7 PFA2s). CEC subpopulations were predominant in PFA2 samples (45% of neoplastic cells), significantly higher (FC = 5.6, p = 0.0021) than in PFA1 samples (7.9% of neoplastic cells) (Figures 3A and 3B). This is consistent with previous data showing cilia-related ontologies as a hallmark of the PFA2 subgroup and that of PFB as well (Pajtler et al., 2015, 2018; Wani et al., 2012; Witt et al., 2011). Accordingly, published PFB marker genes were significantly enriched (FC = 24) in CECs (Figure 3C). PFA1 and PFA2 subgroup marker genesets were generated by dichotomizing our primary PFA transcriptomic dataset based on methylome subgroup assignment, showing clearly that CEC subpopulation markers genes are significantly

**A PFA1 subgroup samples**



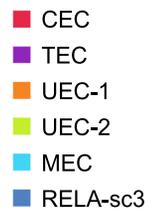
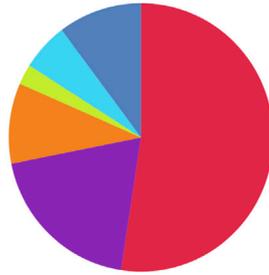
**PFA2 subgroup samples**



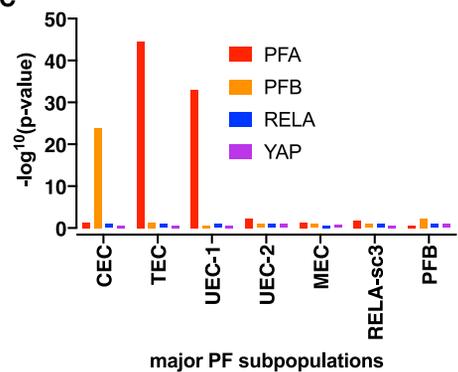
**B PFA1**



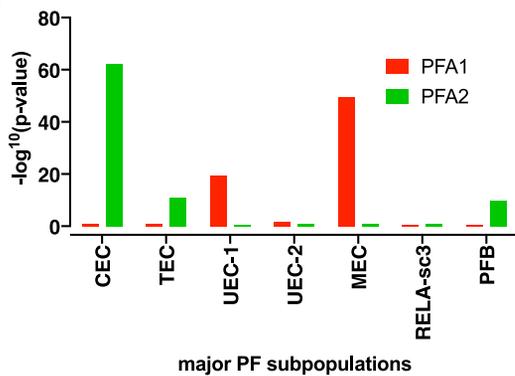
**PFA2**



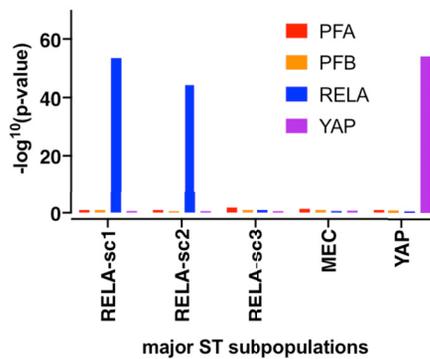
**C**



**D**



**E**



**Figure 3. PFA Neoplastic Subpopulations Are Transcriptomic Related to Classification Subgroups**

(A) Fraction of major PFA subpopulations within each PFA scRNA-seq sample, grouped according to major molecular subgroups PFA1 and PFA2. (B) Average PFA subpopulation fractions in major PFA molecular subgroups. (C and D) PF EPN subpopulation enrichment (hypergeometric test) of genesets from (C) consensus EPN molecular subgroups PFA, PFB, RELA, and YAP; and (D) PFA1 and PFA2 molecular subgroups. (E) Consensus EPN molecular subgroup geneset enrichment in ST EPN subpopulations.

enriched for PFA2 subgroup genes (FC = 25)(Figure 3D). TECs constituted approximately 12% of PFA neoplastic cells and were highly enriched for PFA signature genes (FC = 20) (Figure 3C), notably those genes putatively associated with ependymal transport (*AQP4*, *AQP1*, and *ATP1A2*). TECs were ubiquitously distributed across PFA samples, being slightly more abundant in PFA2s (FC = 1.6) (Figures 3A and 3B) and, accord-

ingly, showed a moderate enrichment of PFA2-specific genes (FC = 8.0) (Figure 3D). Similar to TECs, UEC-1s showed a significant enrichment of PFA signature genes (FC = 17) (Figure 3C). However, they differed from TECs in that they were more abundant (FC = 2.4, p = not significant [n.s.]) in PFA1s (19.4%) than in PFA2s (8.35%) (Figure 3B) and were significantly enriched for PFA1 subgroup-specific genes (FC = 12) (Figure 3D). UEC-2s

were not significantly associated with any EPN subgroups tested. MEC subpopulation cells were more abundant (FC = 2.0,  $p = n.s.$ ) in PFA1s (9.6%) than in PFA2s (4.9%) (Figures 3A and 3B). Significantly, MECs showed a higher enrichment of PFA1 marker genes (FC = 21) than any other EPN subpopulation (Figure 3D). Shared MEC and PFA1 genes included hypoxia (*CA9* and *ICAM1*)-, glycolysis (*LDHA* and *SLC2A1*)-, mesenchyme (*CHI3L1* and 2, *CD44*, *CAV1*, and *TGFB1*)-, and inflammation (*CCL2* and *CXCL2*)-related genes. The finding that NF- $\kappa$ B2 TF activity, previously documented in PFA1 (Griesinger et al., 2017), is upregulated in MECs (Figure 2B) also suggest that this subpopulation underlies the bulk-tumor PFA1 phenotype.

In the context of PF samples, RELA-sc3 cells were similarly abundant in both PFA1s (10.1%) and PFA2s (8.7%). The single PFB subpopulation showed moderate PFB subgroup geneset enrichment (FC = 3.3) (Figure 3C). ST EPN subpopulations showed enrichment with corresponding RELA- and YAP-subgroup-signature genesets, except RELA-sc3, which showed no RELA-subgroup-geneset enrichment (Figure 3E).

### Histological Characterization of PFA Subpopulations Identifies Co-localization of UEC and MEC in Perivascular and Perinecrotic Zones

We sought to further explore the biology of PFA scRNA-seq subpopulations using immunohistochemistry (IHC) to both verify their distribution between samples across PFA1 and PFA2 subgroups and to map their histological spatial relationships. In this study, we focused on those PFA subpopulations for which distinct biological roles had been inferred from transcriptomic profiles, namely CEC, TEC, UEC-1, and MEC. Subpopulation-specific antibodies were developed for use in a cohort formalin-fixed, paraffin-embedded (FFPE) material from 31 primary patient PFA samples (Table S1). CAPS and *FOXJ1* were identified as CEC-specific markers based on differential expression analysis (Figures 2A and S6; Table S2). Cytoplasmic CAPS and nuclear *FOXJ1* protein-expressing cells were colocalized in sequential sections (Figure S7A), staining true ependymal rosette cells in some cases, but most commonly being expressed by small clusters of cells scattered throughout the parenchyma. CAPS and *FOXJ1* were expressed in most PFA samples tested (26 of 31 and 27 for 31, respectively), and IHC staining scores for these two CEC markers were positively correlated across the cohort, further supporting their subpopulation specificity (Table S5B).

In general, TEC and UEC-1 subpopulation markers showed less subpopulation specificity than did CEC and MEC markers (Figure S6). TEC marker *VIPR2* was expressed in cells of the tumor parenchyma, with decreased expression in true ependymal rosettes (Figure S7B). *VIPR2* was expressed to a varying degree in all samples, but showed no significant correlation with any other subpopulation marker across the PFA FFPE cohort (Table S5B).

UEC-1 markers *FOS* and *COL9A2* were expressed in patches of cells, often in perivascular or perinecrotic areas, which were colocalized in some cases, but not in others (Figures 4A, 4B, and S7C) and showed significant correlation of IHC staining scores across the PFA histology cohort (Table S5B). *FOS* expression was observed in all 31 PFA samples, apart from

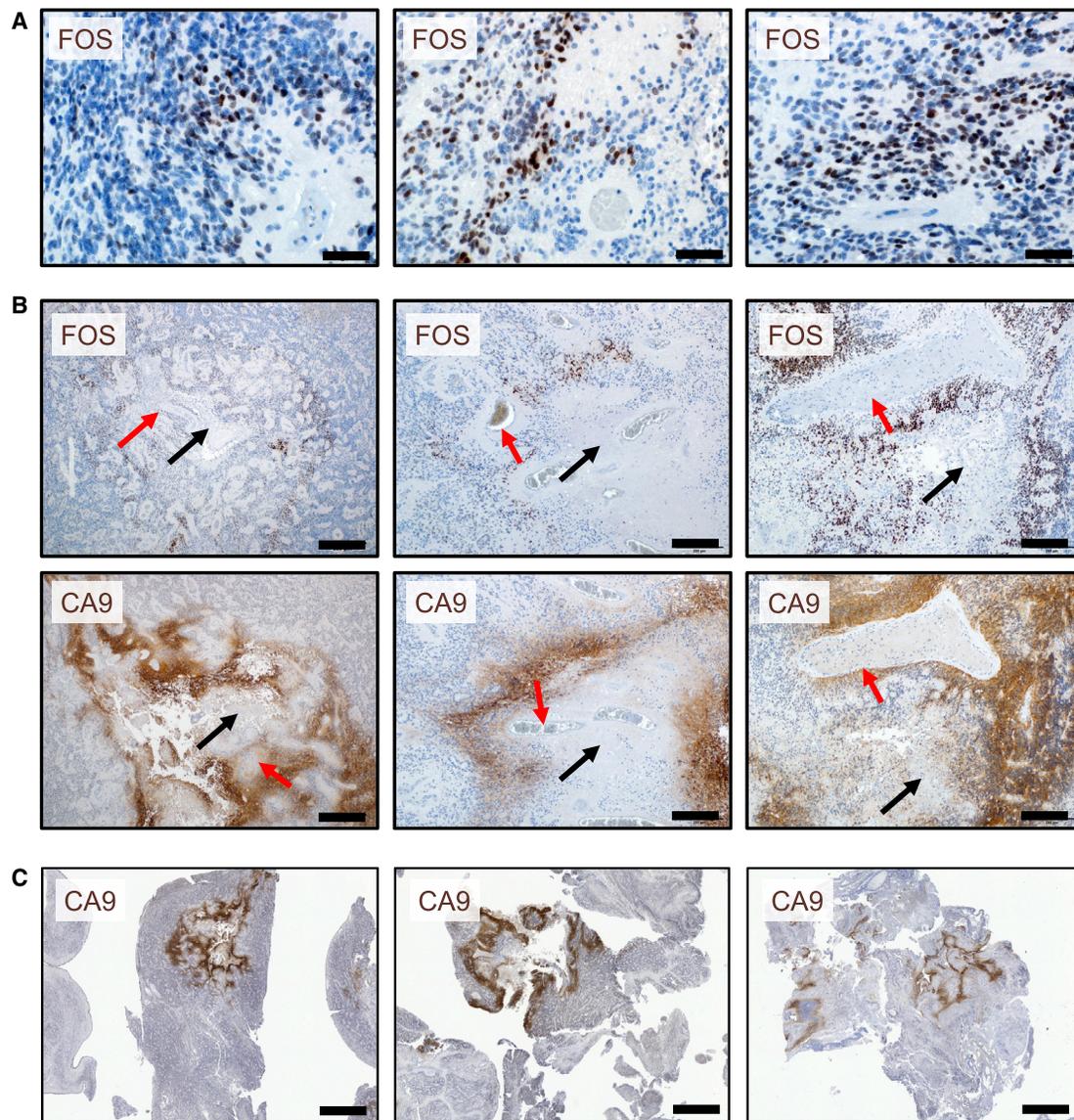
*COL9A2*, which was absent in ~50% of the cohort. Localization of UEC-1 in perivascular zones is consistent with previous studies that identified a perivascular niche for putative EPN cancer stem cells (Calabrese et al., 2007; Taylor et al., 2005), further supporting UEC-1 as a potential progenitor subpopulation in PFA.

MEC-specific markers *CA9*, *CAV1*, and *SLC2A1* were predominantly co-localized in sequential sections, staining large swathes of cells (Figures 4B, 4C, and S7D). MEC-specific marker expression was commonly seen in perinecrotic regions, consistent with our observation that MEC markers are enriched for genes upregulated in GBM pseudopalisading necrosis zones (Figure 2F) and with enriched stress-response-related gene expression in this subpopulation. Across our PFA histology sample cohort, all three MEC markers were positively correlated, *CA9* and *SLC2A1* showing the strongest correlation (Table S5B). MEC markers were detected in most PFA samples tested, being undetectable in only two PFA2 samples. *CA9* expression has previously been investigated in EPN and, consistent with our study, was shown, along with VEGF mRNA, to be significantly co-localized in perinecrotic areas and areas with aberrant vascular patterning (Preusser et al., 2005).

Colocalization of MEC with UEC-1 was shown in most samples in zones of necrosis and aberrant vascularization (Figure 4B). Previous studies had explored the relationship between EPN cancer stem cells and vascular endothelial cells, providing evidence that the latter secrete factors that maintain the cancer stem cell population. We show here that the UEC-1 niche is largely constituted by MECs, which may, therefore, also contribute significantly to the maintenance and renewal of the putative cancer stem cell population in PFAs. Using a large-scale map of ligand-receptor pairs (Ramilowski et al., 2015), we investigated possible interactions between UEC-1 and MEC. This revealed a predominance of extracellular matrix (ECM) receptor signaling between both cell types (Table S5D). ECM receptor signaling can control a variety of cellular activities, including neural stem cell maintenance (Xu et al., 2014).

### PFA Subpopulations Underlie Bulk-Tumor Subgroup Assignment and Are Associated with Disparate Clinical Outcomes

Bulk-tumor transcriptome data (Affymetrix HG-U133plus2) from 121 primary PFA EPN samples, which combined samples from our institution with a large publicly available transcriptomic dataset (Pajtlar et al., 2018), were deconvoluted using CIBERSORT (Newman et al., 2015) to estimate the fraction of each of the major PFA neoplastic subpopulations. The relationship of subpopulation proportions to PFA classification subgroups was assessed in samples that were assigned to PFA1 or PFA2 subgroups based on bulk-tumor methylome analysis (Table S1). As hypothesized from geneset-enrichment analysis of scRNA-seq subpopulation markers (Figure 3C), MECs and CECs were the prevalent subpopulations in PFA1s and PFA2s, respectively (Figure 5A). Additionally, deconvoluted TEC and UEC-1 fractions were significantly more prevalent in PFA2s and PFA1s, respectively (Figure 5A). IHC subpopulation marker expression was concordant with deconvoluted subpopulation



**Figure 4. Histological Characterization of PFA Subpopulations Identifies Co-localization of UEC and MEC in Perivascular and Perinecrotic Zones**

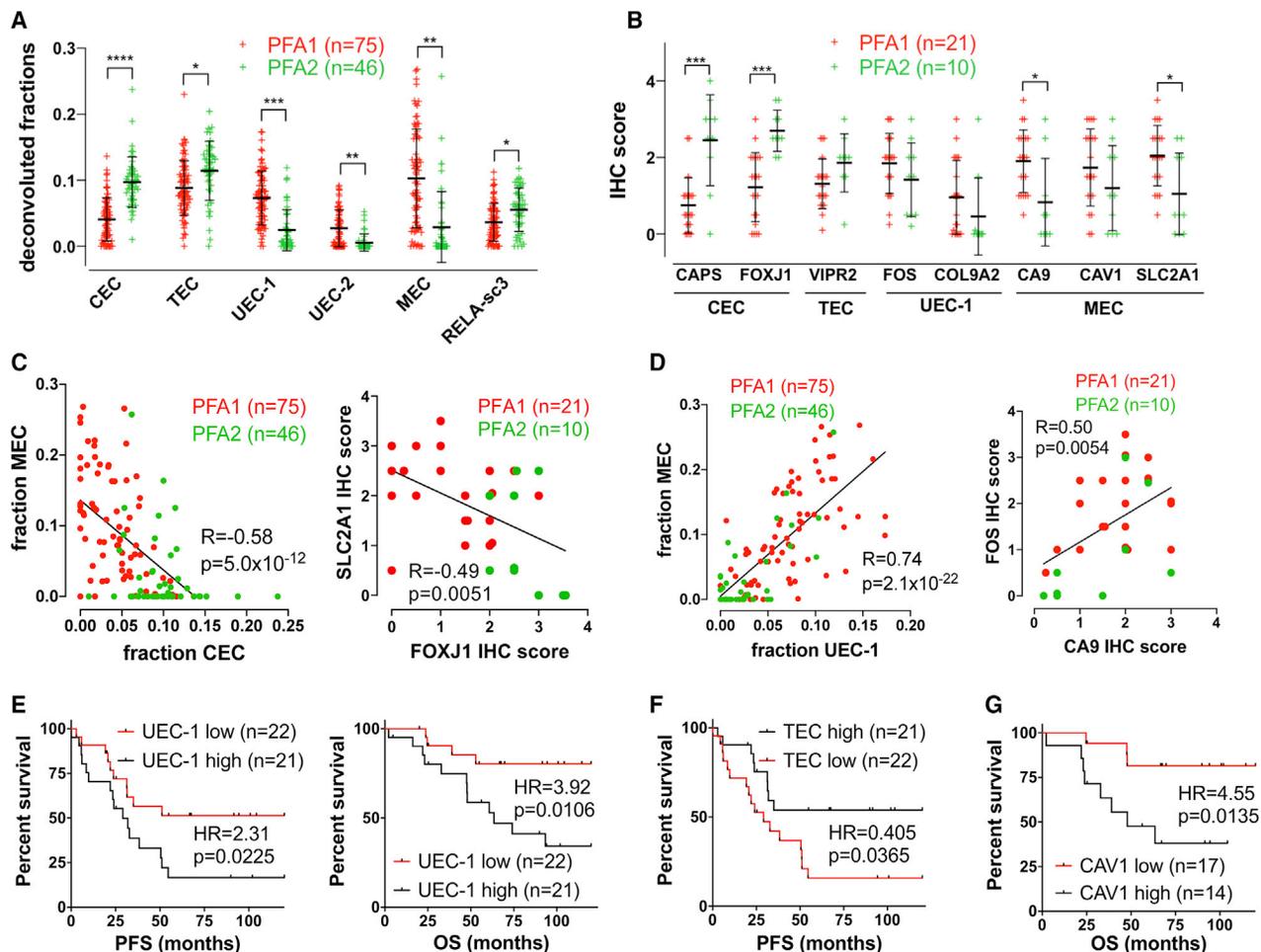
(A) Representative perivascular IHC staining pattern of UEC-1 marker FOS (brown) in three PFA patient samples (scale bars: 50  $\mu$ m).

(B) Perinecrotic colocalization of UEC marker FOS (brown) and MEC marker CA9 (brown) in sequential sections from three PFA patient samples (black arrows, necrosis; red arrows, aberrant vascularization). Scale bars: 500  $\mu$ m (left panel) and 200  $\mu$ m (right panels).

(C) Low-magnification images demonstrating grossly localized distribution of MEC marker CA9 (brown) in three PFA patient samples (scale bar: 2mm). See also Figures S6 and S7.

fractions, showing greater expression of CEC and TEC markers in PFA2s and greater expression of UEC-1s and MECs in PFA1s (Figure 5B). Pajtler et al. (Pajtler et al., 2018) recently identified subtypes within PFA1 and PFA2 that demonstrated variable clinical characteristics, which we examined for PFA subpopulation associations. CEC subpopulation fractions showed the greatest variability among subtypes, being significantly higher in PFA1c than in other PFA1 subtypes, in particular PFA1a and PFA1e (Figure S8A). Conversely, UEC-1 and MEC deconvoluted fractions were significantly higher in PFA1a than in PFA1c.

The relationship between deconvoluted subpopulation fractions was examined, identifying a significant negative correlation between CEC and MEC fractions (Figure 5C; Table S5C), and concordantly, CEC and MEC IHC marker correlations were also negatively correlated (Figure 5C; Table S5C). Of note, CEC and MEC subpopulations occupied predominantly non-overlapping histological zones (Figures S7A and S7D). Other subpopulations fractions were significantly correlated; the strongest of which was the positive correlation between UEC-1 and MEC (Figure 5D). This finding, which was corroborated by IHC



**Figure 5. PFA Subpopulations Underlie Bulk-Tumor Subgroup Assignment and Are Associated with Variable Clinical Outcomes**

(A and B) Differential subpopulation fractions (mean, error bars = SD) between PFA1 and PFA2 (A) inferred by deconvolution of PFA bulk-tumor transcriptome profiles (n = 121); \*\*\*\*p < 10<sup>-13</sup>, \*\*\*p < 10<sup>-9</sup>, \*\*p < 10<sup>-7</sup>, \*p < 0.005 (see also Figure S8); and (B) scored by subpopulation IHC markers in FFPE samples (n = 31); \*\*\*p < 10<sup>-4</sup>, \*p < 0.01.

(C and D) Subpopulation fraction correlation between (C) CEC and MEC, and (D) UEC-1 and MEC, based on deconvoluted fractions (left) and subpopulation IHC marker scoring (right). See Tables S5B and S5C for full subpopulation correlation matrices.

(E and F) (E) Kaplan-Meier progression-free survival (PFS) and overall survival (OS) analysis of PFA with higher- versus lower-than-median, deconvoluted UEC-1 fractions. (F) Kaplan-Meier PFS association for higher- versus lower-than-median, deconvoluted TEC fraction.

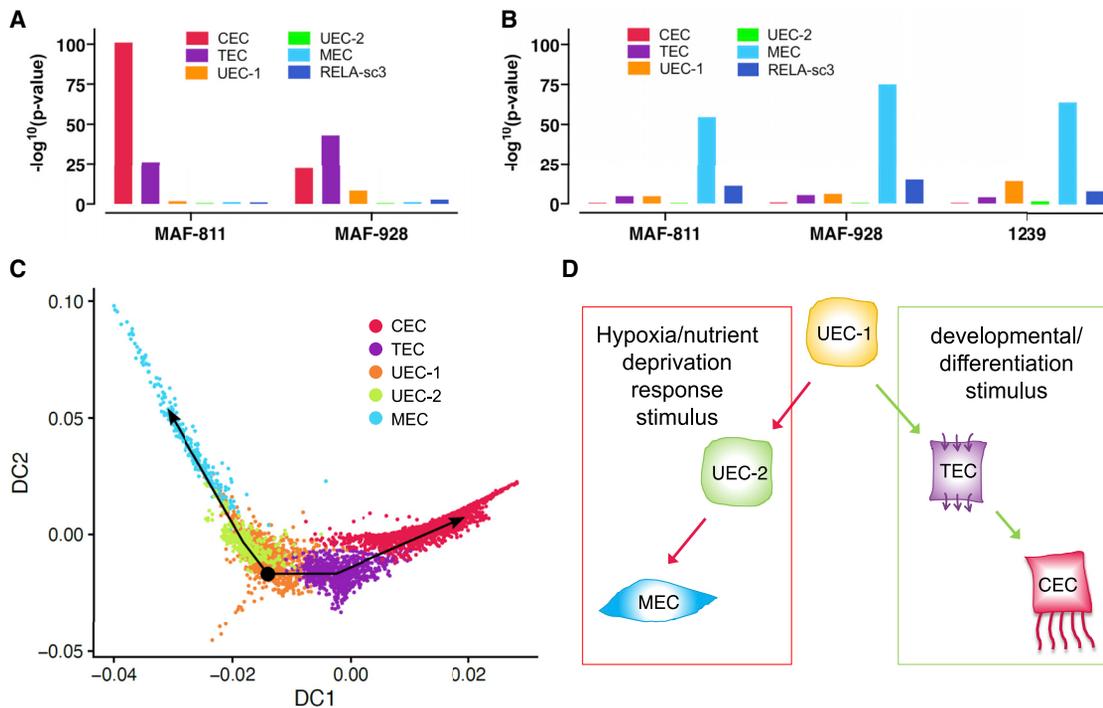
(G) Kaplan-Meier overall survival associations for higher- versus lower-than-median MEC marker CAV1 IHC score. See Table S5E for full outcome analyses.

(Figure 5D), also revealed a common perivascular, perinecrotic localization for both UEC-1 and MEC (Figure 4B). A significant negative correlation of the TEC deconvoluted fraction with both UEC-1 and MEC fractions was also identified (Table S5C).

The dependence of bulk tumor subgroup assignment on underlying MEC and CEC subgroup proportions explains the difficulty in subgroup assignment for samples in which MEC and CEC proportions are closely balanced. For example, in our PFA cohort, we observed non-concordance in methylome and transcriptome-based PFA1/PFA2 classification in 20% (9 of 44) of samples tested (Table S1) in samples that contained balanced proportions of MECs and CECs (Figure S8B). The difficulty in molecular subgroup assignment is further complicated by the heterogeneous intra-tumoral localization of subpopulations. This is particularly evident for MECs that are distributed

throughout the tumor in a grossly localized fashion (Figure 4C). In these cases, sampling error is likely to lead to unreliable subgroup assignment, being dependent on whether RNA/DNA is extracted from zones with high or low MEC abundance.

To further elucidate the biology of PFA subpopulations, we examined their association with clinical outcome, using both deconvoluted subpopulation fractions and IHC scores in the clinically annotated sample cohort from our institution (Table S1). Progression-free survival (PFS) and overall survival (OS) were determined using Kaplan-Meier curve analysis and a log-rank t test, with subjects divided according to clinical variables and higher or lower than median deconvoluted subpopulation fractions in the transcriptomic PFA cohort (n = 43; median follow-up: 5.6 years). Consistent with previous studies (Pajtler et al., 2018), our transcriptomic cohort demonstrated no difference in



**Figure 6. *In vitro* Perturbations and Pseudotime Analysis Identifies Developmental and Microenvironmental Stress Associated with PFA Subpopulation Lineage Trajectories**

(A and B) Neoplastic subpopulation signature enrichment (hypergeometric test p value) of genes upregulated (A) in PFA cell lines MAF-811 and MAF-928 maintained as non-adherent 3D neurospheres as compared with adherent monolayer cultures, and (B) in PFA cell lines MAF-811 and MAF-928 and short-term culture from sample 1,239 in response to hypoxia. See also Figure S8C.

(C) Pseudotime analysis of PFA neoplastic cells with Destiny and Slingshot reveals divergent trajectories starting with UEC-1 and either (1) transitioning into MECs via UEC-2s, or (2) differentiating into TECs and then CECs. Abbreviation: DC, diffusion component. See also Figure S9.

(D) Schematic of hypothetical divergent PFA subpopulation trajectories either as a response to stress (e.g., hypoxia), or recapitulating normal developmental processes in response to developmental cues.

PFA or OS between PFA1 and PFA2 based on methylation subgroup assignment (Table S5E). Deconvoluted subpopulation fractions did, however, show significant outcome associations. In particular, UEC-1 was significantly ( $p < 0.05$ ) associated with outcome, with higher-than-median fractions of UEC-1s being associated with higher and lower risk of recurrence and death (Figure 5E; Table S5E). Conversely, a higher-than-median fraction of TECs was associated with a significantly longer PFS (Figure 5F) and a longer OS, but this did not reach significance by log-rank t test (Table S5E). A higher-than-median fraction of CECs and MECs, respectively, was associated with longer and shorter survival, but neither of those reached significance (Table S5E). These deconvoluted subpopulation fraction outcome trends were reiterated in the smaller IHC study cohort ( $n = 31$ ; median follow-up: 5.6 years) (Table S5E), with MEC marker CAV1 being significantly associated with shorter OS (Figure 5G).

### ***In vitro* Perturbations and Pseudotime Analysis Identifies Developmental and Microenvironmental Stress Associated PFA Subpopulation Lineage Trajectories**

Molecular characterization of PFA subpopulations revealed evidence of lineage differentiation, which we further explored by

perturbation of EPN cell lines, short-term culture and pseudotime lineage trajectory analysis of scRNA-seq data. We recently established PFA cell lines (MAF-811 and MFA-928), which, when cultured as monolayers, showed increased MECs and decreased CECs and TECs compared with primary tumor as measured by deconvolution of transcriptomic profiles (Figure S8C). Our previous study of these cell lines demonstrated that cells cultured as 3D structures harbored cilia-related gene expression and archetypal EPN cellular architecture that was absent in cells grown as monolayers (Amani et al., 2017). Reanalysis of these data demonstrated a significant enrichment of CEC and TEC signature genes in those genes upregulated in 3D cultures versus monolayers in both cell lines (Figure 6A), suggesting that PFA differentiation is mediated, in part, by architectural developmental cues. Gene expression in MECs suggests a stress-response program, possibly because of hypoxic conditions found in peri-necrotic tumor zones in PFA. We modeled these conditions in PFA *in vitro* by culturing cell lines MAF-811 and MAF-928 under hypoxic conditions (2% O<sub>2</sub>, 72 h), demonstrating a striking upregulation of MEC signature genes in both cell lines (Figure 6B). Both cell lines were established from recurrent tumor samples and harbor a chromosome arm 1q gain. To exclude the possibility that these results were related to these

factors, we repeated this experiment with a short-term culture from a primary sample with no 1q aberrations (1,239), which showed a consistent hypoxia response (Figure 6B). Collectively, functional analyses demonstrate the adaptive capacity of PFA cells that likely underlies their discrete subpopulation characteristics.

PFA scRNA-seq data were subject to unsupervised lineage trajectory analysis using Slingshot (Street et al., 2018) to identify trajectories and order cells on a diffusion map produced by Destiny (Angerer et al., 2015). We identified a divergent cellular trajectory, in which UEC-1 developed into either (1) TECs and then CECs, or (2) transitioned into UEC-2s then MECs (Figure 6C). Trajectory gene expression showed a continuous gradual change across cell states for both trajectories (Figure S9)

Based on these *in vitro* and scRNA-seq analyses, we hypothesize that divergent lineage trajectories exist in PFA (Figure 6D). One branch represents a differentiating trajectories; the most predominant of which is a recapitulation of normal ependymal development, namely the development of TEC and CEC. This PFA developmental trajectory, whereby UEC-1 develop into more-differentiated progeny, conforms to the classic hierarchical lineage described by the cancer stem cell model, in which a stem cell subpopulation that is associated with an aggressive clinical phenotype differentiates into progressively less-aggressive progeny presumably in response to developmental or differentiation stimuli. Such stimuli include cellular architectural organization cues, a process that we recapitulated *in vitro* (Figure 6A), potentially through planar cell-polarity signaling that has been implicated as a requirement for ciliogenesis (Kim et al., 2010). This PFA-specific cancer stem cell trajectory is further supported by the selective enrichment of vRG and ventricular progenitor-specific transcripts in UEC-1 (Figures 2D and 2E). Given the favorable clinical behavior of EPN PF with a low proportion of UEC-1 and a higher proportion of TEC (Figure 5E), differentiation, agent treatment may, therefore, prove beneficial. In support of this therapeutic strategy, we recently demonstrated PFA-selective efficacy of differentiating agent retinoid-analogs *in vitro* (Donson et al., 2018). In the second-lineage trajectory, UEC-1s undergo EMT to become MECs, a process that is likely driven by microenvironmental stress stimuli, such as hypoxia and/or nutrient deprivation (Figure 6D). Consequent MEC-mediated vasculogenesis may then generate a perivascular niche that supports the putative PFA cancer stem-cell subpopulation UEC-1, a hypothesis that is also supported by histological colocalization of these subpopulations (Figure 4B). Functional testing of these hypotheses will be required to more definitively characterize the putative PFA populations revealed by the present, largely molecular, study.

#### Resource: Childhood EPN Single-Cell Browser

We created a browsable web interface that integrates the Harmony EPN scRNA-seq dataset with the University of California, Santa Cruz (UCSC) Cell Browser format (<http://pneurooncellatlas.org>). This interactive resource allows users to study the cellular restriction and expression level of transcripts of interest at the single-cell level in childhood EPN. In addition to raw gene expression values, we include imputed gene-expression values that were generated using adaptively thresholded

low-rank approximation (ALRA) (<https://www.biorxiv.org/content/10.1101/397588v1>).

#### DISCUSSION

In the present study, we used scRNA-seq to describe intra- and inter-tumoral cellular heterogeneity in a well-characterized cohort of childhood ST and PF EPNs (n = 26). PFA tumors in particular appear to conceal multiple conserved and well defined subpopulations corresponding to normal ependymal functional states, neurodevelopmental states and stress-response mechanisms.

Two PFA subpopulations, CEC and TEC, recapitulated normal ependymal phenotypes of cilia-function and cellular transport respectively, evidence of differentiation processes that, consistent with cancers in general, convey a favorable clinical outcome. A recent scRNA-seq study that mapped mouse cerebellar developmental lineages and compared these with childhood cerebellar tumors, did not identify differentiated subpopulations in PFA, postulating the existence of a differentiation block in these tumors (Vladoiu et al., 2019). Our study contradicts their hypothesis through identification of differentiated PFA subpopulations. We believe this discrepancy is due to the absence of ependymal cell types in their mouse cerebellar developmental lineage comparators.

Our study identified an undifferentiated PFA subpopulation, UEC-1, which is associated with a more-aggressive clinical course. These characteristics, combined with pseudotime cellular lineage analysis and *in vitro* functional assays, lead us to hypothesize that UEC-1s are the PFA CSC population (Figure 6D). Consistent with earlier EPN CSC studies, UEC-1s reside in perinecrotic and perivascular zones (Calabrese et al., 2007). We now show that this CSC niche is largely composed of MECs that may contribute to CSC maintenance and/or expansion, an important area for further exploration. Comparison of PFA subpopulations to scRNA-seq-defined neurodevelopmental lineages, namely of human radial glia (Pollen et al., 2015), mouse cerebellum (Vladoiu et al., 2019), and mouse ependymal layer (Shah et al., 2018), revealed significant overlap of the UEC-1 subpopulation with VZ lineages, namely, human vRG and mouse VZ progenitors. Vladoiu et al. (2019) combined deconvolution of bulk-tumor PFA transcriptomes and scRNA-seq analysis of four PFA samples with mouse cerebellar developmental lineage markers, identifying gliogenic progenitors as a putative PFA cell of origin. Our study refines their findings, identifying enrichment of gliogenic progenitor genes in UEC-1s but to a greater extent in the more-differentiated TEC subpopulation, in addition to identifying greater ventricular zone progenitor gene enrichment in UEC-1.

The molecular characteristics of MECs underlie many of the phenotypes that had previously been identified as hallmarks of PFA and PFA1 subgroup tumors. The original study to define PFAs showed that this subgroup was distinguished from PFBs by enrichment of gene ontologies, such as angiogenesis, VEGF signaling, tissue invasion, and response to hypoxia (Witt et al., 2011). Conversely, the PFBs were distinguished by cilia-related ontologies. A study contemporary to that one identified a subgroup analogous to PFA that was defined by mesenchymal

characteristics (Wani et al., 2012). Subsequent studies have focused on the NF- $\kappa$ B2-driven inflammatory phenotype in PFA (Griesinger et al., 2015; Griesinger et al., 2017). It is now apparent that these functions that were previously ascribed to PFA1 are specifically restricted to the MEC subpopulation of PFA. The transition of UEC-1s to MECs suggested by pseudotime lineage analysis recapitulates EMT, an essential process in embryogenesis and wound healing, which, in cancer, results in tissue invasion and pluripotency. All of the MEC-associated phenotypes above are widely recognized to contribute to poor clinical outcomes, and not surprisingly, MEC was associated with shorter OS in our present study.

We demonstrated that the predominance of either MEC or CEC subpopulations determined bulk-tumor subgroup assignment to PFA1 or PFA2, respectively, but that no subpopulations was unique to either subgroup. This finding potentially explains the difficulty in accurately classifying PFA to either subgroup, as demonstrated by non-concordance to methylome and transcriptome-based subgroup assignment in a portion (20%) of samples. Indeed, these data suggest that the existence of discrete PFA1 and PFA2 molecular subgroups is a misconception. Rather PFA tumors are a continuum of varying ratios of PFA1-like (MEC) and PFA2-like (CEC) subpopulations, with a potentially common underlying biology and cell or origin. This is an important consideration given that subgroup-specific targeted therapies are being suggested based on the theory that PFA1 and PFA2 are distinct. Thus, in addition to a misinterpretation of the cancer biology, neoplastic cell heterogeneity also confounds bulk-tumor-based subgrouping of tumors beyond a certain threshold. The results of the present study show that this threshold has been reached at the PFA1/PFA2 subgroup level, as defined by transcriptomic analysis. Whether this caveat extends to subgrouping based on methylation profiling is of critical importance in future PFA studies.

PFA contributes overwhelmingly to the high relapse rate in EPN, which is fatal in more than 50% of children and has not seen any therapeutic improvements in more than 30 years (Ramswamy et al., 2016). Most children with PFA undergo complete tumor resection and receive radiation when in a state of minimal disease. In spite of that, the delayed relapse rate is high, with 70% relapsing by 10 years (Marinoff et al., 2017). This suggests that relapse occurs from a subpopulation of resistant cells. Our study identified two PFA subpopulations that were associated with a more-aggressive clinical course, namely UEC-1 and MEC. These subpopulations, therefore, represent novel candidate drivers of relapse in PFA, which could be specifically targeted for therapeutic gain. Further functional and pre-clinical studies are, therefore, required to more definitively demonstrate the proposed progenitor-cell characteristics and therapeutic resistance properties of these subpopulations.

Based on the results of recent studies in high-grade glioma (Nefitel et al., 2019; Patel et al., 2014), we had anticipated identifying PFA subpopulations corresponding to bulk PFA1 and PFA2 subgroup profiling (MEC and CEC). However, scRNA-seq also revealed the existence of additional conserved subpopulations (notably TEC and UEC-1). These subpopulations harbor distinct and clinically relevant biologies and, yet, were not apparent from bulk-tumor transcriptomic analyses, obscured by the overpow-

ering transcriptomic profiles of MEC and CEC. This surprising finding further underscores the power of single-cell analyses in oncology research and the critical need to address cancer-cell heterogeneity in our goal to provide effective therapy for this devastating childhood tumor.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Tissue harvest for single cell suspension, single-cell RNA library preparation and sequencing
  - ScRNaseq data analysis
  - Bulk-tumor methylome and transcriptome analyses
  - Deconvolution
  - Immunohistochemistry
  - Survival studies
  - Perturbation studies of PFA EPN *in vitro*
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108023>.

## ACKNOWLEDGMENTS

A.E.G. is supported as an informatics fellow of the RNA Bioscience Initiative, University of Colorado School of Medicine. Bioinformatic support for this work was supported by the RNA Bioscience Initiative, University of Colorado School of Medicine. This study was supported by the Tanner Seebaum Foundation, the Morgan Adams Foundation, the National Institutes of Health (R01 CA237608-01), and the Cancer League of Colorado (research grant AWD 173796-SV). The University of Colorado Denver Genomics and Microarray, and Histology Shared Resources are supported by the University of Colorado's NIH/NCI Cancer Center (P30CA046934). The authors appreciate the contribution to this research made by E. Erin Smith, HTL (ASCP) CMQIHC; Allison Quador, HTL (ASCP) CM; and Jessica Arnold HTL (ASCP) CM of the University of Colorado Denver Histology Shared Resource. We would like to thank Dr. Martin Sill at German Cancer Research Center DKFZ, Division of Biostatistics, for assistance with methylation-based brain tumor subgroup classification.

## AUTHOR CONTRIBUTIONS

A.M.D. performed experiments with assistance from A.G., S.V., K.M., E.P., V.A., A.M.G., T.C.H., and M.H.H. A.E.G. performed bioinformatic analyses with assistance from K.A.R., B.S., and K.L.J. made additional bioinformatic contributions. A.G. assisted with neuropathology concerns. N.K.F., A.E.G., and A.M.D. were responsible for the design of the study. All authors assisted with manuscript preparation.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 11, 2019  
Revised: June 16, 2020  
Accepted: July 21, 2020  
Published: August 11, 2020

**REFERENCES**

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086.

Amani, V., Donson, A.M., Lummus, S.C., Prince, E.W., Griesinger, A.M., Witt, D.A., Hankinson, T.C., Handler, M.H., Dorris, K., Vibhakhar, R., et al. (2017). Characterization of 2 Novel Ependymoma Cell Lines With Chromosome 1q Gain Derived From Posterior Fossa Tumors of Childhood. *J. Neuropathol. Exp. Neurol.* **76**, 595–604.

Andrieuolo, F., Varlet, P., Tauziède-Espariat, A., Jünger, S.T., Dörner, E., Dreschmann, V., Kuchelmeister, K., Waha, A., Haberler, C., Slavc, I., et al. (2019). Childhood supratentorial ependymomas with YAP1-MAML1 fusion: an entity with characteristic clinical, radiological, cytogenetic and histopathological features. *Brain Pathol.* **29**, 205–216.

Angerer, P., Haghverdi, L., Büttner, M., Theis, F., Marr, C., and Büttner, F. (2015). Destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243.

Araki, A., Chocholous, M., Gojo, J., Dorfer, C., Czech, T., Heinzl, H., Dieckmann, K., Ambros, I.M., Ambros, P.F., Slavc, I., and Haberler, C. (2016). Chromosome 1q gain and tenascin-C expression are candidate markers to define different risk groups in pediatric posterior fossa ependymoma. *Acta Neuropathol. Commun.* **4**, 88.

Baas, D., Meinel, A., Benadiba, C., Bonnafé, E., Meinel, O., Reith, W., and Durand, B. (2006). A deficiency in RFX3 causes hydrocephalus associated with abnormal differentiation of ependymal cells. *Eur. J. Neurosci.* **24**, 1020–1030.

Calabrese, C., Poppleton, H., Kocak, M., Hogg, T.L., Fuller, C., Hamner, B., Oh, E.Y., Gaber, M.W., Finklestein, D., Allen, M., et al. (2007). A perivascular niche for brain tumor stem cells. *Cancer Cell* **11**, 69–82.

Chung, M.I., Peyrot, S.M., LeBoeuf, S., Park, T.J., McGary, K.L., Marcotte, E.M., and Wallingford, J.B. (2012). RFX2 is broadly required for ciliogenesis during vertebrate development. *Dev. Biol.* **363**, 155–165.

Donson, A.M., Birks, D.K., Barton, V.N., Wei, Q., Kleinschmidt-Demasters, B.K., Handler, M.H., Waziri, A.E., Wang, M., and Foreman, N.K. (2009). Immune gene and cell enrichment is associated with a good prognosis in ependymoma. *J. Immunol.* **183**, 7428–7440.

Donson, A.M., Amani, V., Warner, E.A., Griesinger, A.M., Witt, D.A., Levy, J.M.M., Hoffman, L.M., Hankinson, T.C., Handler, M.H., Vibhakhar, R., et al. (2018). Identification of FDA-approved oncology drugs with selective potency in high-risk childhood ependymoma. *Mol. Cancer Ther.* **17**, 1984–1994.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278.

Götz, M., Stoykova, A., and Gruss, P. (1998). Pax6 controls radial glia differentiation in the cerebral cortex. *Neuron* **21**, 1031–1044.

Goulding, M.D., Chalepakis, G., Deutsch, U., Erselius, J.R., and Gruss, P. (1991). Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J.* **10**, 1135–1147.

Griesinger, A.M., Birks, D.K., Donson, A.M., Amani, V., Hoffman, L.M., Waziri, A., Wang, M., Handler, M.H., and Foreman, N.K. (2013). Characterization of distinct immunophenotypes across pediatric brain tumor types. *J. Immunol.* **191**, 4880–4888.

Griesinger, A.M., Josephson, R.J., Donson, A.M., Mulcahy Levy, J.M., Amani, V., Birks, D.K., Hoffman, L.M., Furtek, S.L., Reigan, P., Handler, M.H., et al. (2015). Interleukin-6/STAT3 Pathway Signaling Drives an Inflammatory Phenotype in Group A Ependymoma. *Cancer Immunol. Res.* **3**, 1165–1174.

Griesinger, A.M., Witt, D.A., Grob, S.T., Georgio Westover, S.R., Donson, A.M., Sanford, B., Mulcahy Levy, J.M., Wong, R., Moreira, D.C., DeSisto, J.A., et al. (2017). NF- $\kappa$ B upregulation through epigenetic silencing of LDOC1 drives tumor biology and specific immunophenotype in Group A ependymoma. *Neuro-oncol.* **19**, 1350–1360.

Hartfuss, E., Galli, R., Heins, N., and Götz, M. (2001). Characterization of CNS precursor subtypes and radial glia. *Dev. Biol.* **229**, 15–30.

Hoffman, L.M., Donson, A.M., Nakachi, I., Griesinger, A.M., Birks, D.K., Amani, V., Hemenway, M.S., Liu, A.K., Wang, M., Hankinson, T.C., et al. (2014). Molecular sub-group-specific immunophenotypic changes are associated with outcome in recurrent posterior fossa ependymoma. *Acta Neuropathol.* **127**, 731–745.

Hovestadt, V., Smith, K.S., Bihannic, L., Filbin, M.G., Shaw, M.L., Baumgartner, A., DeWitt, J.C., Groves, A., Mayr, L., Weisman, H.R., et al. (2019). Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74–79.

Katano, H., Masago, A., Harada, S., Iwata, A., and Yamada, K. (1998). Differential induction of immediate early gene mRNAs following cryogenic and impact trauma with/without craniotomy in rats. *Brain Res.* **800**, 69–77.

Kim, S.K., Shindo, A., Park, T.J., Oh, E.C., Ghosh, S., Gray, R.S., Lewis, R.A., Johnson, C.A., Attie-Bittach, T., Katsanis, N., and Wallingford, J.B. (2010). Planar cell polarity acts through septins to control collective cell movement and ciliogenesis. *Science* **329**, 1337–1340.

Korsunsky, I., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2018). Fast, sensitive, and accurate integration of single cell data with Harmony. *bioRxiv*. <https://doi.org/10.1101/461954>.

Mack, S.C., Witt, H., Piro, R.M., Gu, L., Zuyderduyn, S., Stütz, A.M., Wang, X., Gallo, M., Garzia, L., Zayne, K., et al. (2014). Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445–450.

Mack, S.C., Pajtker, K.W., Chavez, L., Okonechnikov, K., Bertrand, K.C., Wang, X., Erkek, S., Federation, A., Song, A., Lee, C., et al. (2018). Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling. *Nature* **553**, 101–105.

Malgulwar, P.B., Nambirajan, A., Pathak, P., Rajeshwari, M., Suri, V., Sarkar, C., Singh, M., and Sharma, M.C. (2018). Epithelial-to-mesenchymal transition-related transcription factors are up-regulated in ependymomas and correlate with a poor prognosis. *Hum. Pathol.* **82**, 149–157.

Marinoff, A.E., Ma, C., Guo, D., Snuderl, M., Wright, K.D., Manley, P.E., Al-Sayegh, H., Sinai, C.E., Ullrich, N.J., Marcus, K., et al. (2017). Rethinking childhood ependymoma: a retrospective, multi-center analysis reveals poor long-term overall survival. *J. Neurooncol.* **135**, 201–211.

Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849.e821.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457.

Pajtker, K.W., Witt, H., Sill, M., Jones, D.T., Hovestadt, V., Kratochwil, F., Wani, K., Tatevossian, R., Punchihewa, C., Johann, P., et al. (2015). Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer Cell* **27**, 728–743.

Pajtker, K.W., Wen, J., Sill, M., Lin, T., Orisme, W., Tang, B., Hübner, J.M., Ramaswamy, V., Jia, S., Dalton, J.D., et al. (2018). Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. *Acta Neuropathol.* **136**, 211–226.

Parker, M., Mhankuma, K.M., Punchihewa, C., Weinlich, R., Dalton, J.D., Li, Y., Lee, R., Tatevossian, R.G., Phoenix, T.N., Thiruvankatam, R., et al. (2014). C11orf95-RELA fusions drive oncogenic NF- $\kappa$ B signalling in ependymoma. *Nature* **506**, 451–455.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014).

Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401.

Pollen, A.A., Nowakowski, T.J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C.R., Shuga, J., Liu, S.J., Oldham, M.C., Diaz, A., et al. (2015). Molecular identity of human outer radial glia during cortical development. *Cell* 163, 55–67.

Preusser, M., Wolfsberger, S., Haberler, C., Breitschopf, H., Czech, T., Slavc, I., Harris, A.L., Acker, T., Budka, H., and Hainfellner, J.A. (2005). Vascularization and expression of hypoxia-related tissue factors in intracranial ependymoma and their impact on patient survival. *Acta Neuropathol.* 109, 211–216.

Ramaswamy, V., Hielscher, T., Mack, S.C., Lassaletta, A., Lin, T., Pajtler, K.W., Jones, D.T., Luu, B., Cavalli, F.M., Aldape, K., et al. (2016). Therapeutic impact of cytoreductive surgery and irradiation of posterior fossa ependymoma in the molecular era: a retrospective multicohort analysis. *J. Clin. Oncol.* 34, 2468–2477.

Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Sattagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., and Forrest, A.R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* 6, 7866.

Shah, P.T., Stratton, J.A., Stykel, M.G., Abbasi, S., Sharma, S., Mayr, K.A., Koblinger, K., Whelan, P.J., and Biernaskie, J. (2018). Single-cell transcriptomics and fate mapping of ependymal cells reveals an absence of neural stem cell function. *Cell* 173, 1045–1057.e9.

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19, 477.

Suter, D.M., Tirefort, D., Julien, S., and Krause, K.H. (2009). A Sox1 to Pax6 switch drives neuroectoderm to radial glia progression during differentiation of mouse embryonic stem cells. *Stem Cells* 27, 49–58.

Taylor, M.D., Poppleton, H., Fuller, C., Su, X., Liu, Y., Jensen, P., Magdaleno, S., Dalton, J., Calabrese, C., Board, J., et al. (2005). Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell* 8, 323–335.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.

Vladoiu, M.C., El-Hamamy, I., Donovan, L.K., Farooq, H., Holgado, B.L., Sundaravadanam, Y., Ramaswamy, V., Hendrikse, L.D., Kumar, S., Mack, S.C., et al. (2019). Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* 572, 67–73.

Wang, D., and Owler, B.K. (2011). Expression of AQP1 and AQP4 in paediatric brain tumours. *J. Clin. Neurosci.* 18, 122–127.

Wani, K., Armstrong, T.S., Vera-Bolanos, E., Raghunathan, A., Ellison, D., Gilbertson, R., Vaillant, B., Goldman, S., Packer, R.J., Fouladi, M., et al.; Collaborative Ependymoma Research Network (2012). A prognostic gene expression signature in infratentorial ependymoma. *Acta Neuropathol.* 123, 727–738.

Witt, H., Mack, S.C., Ryzhova, M., Bender, S., Sill, M., Isserlin, R., Benner, A., Hielscher, T., Milde, T., Remke, M., et al. (2011). Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma. *Cancer Cell* 20, 143–157.

Wohnhaas, C.T., Leparo, G.G., Fernandez-Albert, F., Kind, D., Gantner, F., Vollet, C., Hildebrandt, T., and Baum, P. (2019). DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. *Sci. Rep.* 9, 10699.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26, 873–881.

Xu, J.C., Xiao, M.F., Jakovcevski, I., Sivukhina, E., Hargus, G., Cui, Y.F., Irintchev, A., Schachner, M., and Bernreuther, C. (2014). The extracellular matrix glycoprotein tenascin-R regulates neurogenesis during development and in the adult dentate gyrus of mice. *J. Cell Sci.* 127, 641–652.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
CA9 rabbit Pab	Novus	NB100417; RRID:AB_10003398
CAPS rabbit Pab	Novus	NBP1-91746; RRID:AB_11016241
COL9A2	ThermoFisher	PA5-63286; RRID:AB_2640030
FOS	Origene	TA806833; RRID:AB_2628246
FOXJ1	Sigma	HPA005714; RRID:AB_1078902
SLC2A1	Proteintech	21829-1-AP; RRID:AB_10837075
CAV1	ThermoFisher	PA5-32297; RRID:AB_2549770
VIPR2 (VPAC2)	ThermoFisher	PF3-114; RRID:AB_2216680
<b>Biological Samples</b>		
Human surgical samples (refer to <a href="#">Table S1</a> )	University of Colorado / Morgan Adams Foundation tumor bank	Refer to <a href="#">Table S1</a>
<b>Critical Commercial Assays</b>		
Allprep RNA/DNA Mini Kit	QIAGEN	Cat #80204
Chromium Single Cell V2 and V3 Chemistry Library Kit, Gel Bead & Multiplex Kit and Chip Kit	10x Genomics	Cats #120237 #120262 #120236
<b>Deposited Data</b>		
scRNaseq and Affymetrix microarray data superseries		GEO: GSE126025
EPN scRNaseq browsable webresource		Database: <a href="http://pneuroonccellatlas.org">pneuroonccellatlas.org</a>
<b>Experimental Models: Cell Lines</b>		
MAF-811	University of Colorado / Morgan Adams Foundation	N/A
MAF-928	University of Colorado / Morgan Adams Foundation	N/A
1239	University of Colorado / Morgan Adams Foundation	N/A
<b>Software and Algorithms</b>		
Bioconductor R (version 3.5.3)	Roswell Park Comprehensive Cancer Center	<a href="https://www.bioconductor.org/">https://www.bioconductor.org/</a>
CellRanger (version 2.1.1)	10x Genomics	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome">https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome</a>
Seurat (version 2.0)	Satija laboratory	<a href="https://satijalab.org/seurat">https://satijalab.org/seurat</a>
DAVID (version 6.8)	NCI	<a href="https://david.ncicrf.gov/">https://david.ncicrf.gov/</a>
SCENIC	( <a href="#">Aibar et al., 2017</a> )	N/A
Slingshot	( <a href="#">Street et al., 2018</a> )	N/A
Destiny	( <a href="#">Angerer et al., 2105</a> )	N/A
MolecularNeuropathology.org version 12	German Cancer Research Center (DKFZ)	<a href="https://www.moleculareuropathology.org/mnp">https://www.moleculareuropathology.org/mnp</a>
NMF	Broad Institute	<a href="https://cloud.genepattern.org">https://cloud.genepattern.org</a>
gSNAP	( <a href="#">Wu and Nacu, 2010</a> )	N/A
Cufflinks	( <a href="#">Trapnell et al., 2012</a> )	N/A
Illustrator 23.0.3	Adobe	N/A
InferCNV	Broad Institute	<a href="https://github.com/broadinstitute/inferCNV">https://github.com/broadinstitute/inferCNV</a>
Prism 8.0.1	GraphPad	N/A

## RESOURCE AVAILABILITY

### Lead Contact

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Andrew Donson ([andrew.donson@cuanschutz.edu](mailto:andrew.donson@cuanschutz.edu)).

### Materials Availability

This study did not generate any new unique reagents

### Data and Code Availability

No unpublished custom code, software or algorithm was used in this study. The accession number for the scRNaseq and transcriptome data reported in this paper is Gene Expression Omnibus (GEO) database: GSE126025 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126025>). Data is also available in the browsable webresource of the full EPN scRNaseq neoplastic subpopulation dataset is available at the Pediatric Neuro-oncology Cell Atlas (Database: [pneurooncellatlas.org](https://pneurooncellatlas.org)).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

Surgical material was collected at our institution at the time of surgery with consent (COMIRB 95-500). Samples were snap frozen for bulk-tumor methylome (n = 52) and transcriptome analysis (n = 69) and a subset of these were collected into PBS for scRNaseq analysis (n = 26) (Table S1). ScRNaseq samples were rapidly dissociated into single cells using a mechanical process as described previously (Griesinger et al., 2013), viably cryopreserved and banked at < 80°C for later use. In this way we were able to batch samples and thus limit experimental variance without compromising sample quality, as DMSO cryopreservation has been shown to match and in some cases exceed the quality of fresh samples with respect to scRNaseq analysis (Wohnhaas et al., 2019).

EPN cell lines MAF-811 and MAF-928 were established as previously described (Amani et al., 2017). Cell lines were validated by DNA fingerprinting through the University of Colorado Molecular Biology Service Center utilizing the STR DNA Profiling PowerPlex-16 HS Kit (DC2101, Promega). Short term culture 1239 was established from disaggregated primary tumor that had been viably cryopreserved, and maintained as for EPN cell lines using Optimem media supplemented with 15% fetal bovine serum.

## METHOD DETAILS

### Tissue harvest for single cell suspension, single-cell RNA library preparation and sequencing

For scRNaseq, samples were thawed in batches and flow sorted (Astrios EQ) to obtain viable single cells based on propidium iodide (PI) exclusion. With the study goal of performing scRNaseq on 2,000 cells per sample, we utilized a Chromium Controller in combination with Chromium Single Cell V2 and V3 Chemistry Library Kits, Gel Bead & Multiplex Kit and Chip Kit (10X Genomics). This approach involves the isolation of single cells into microfluidic droplets containing oligonucleotide-covered gel beads that capture and barcode the transcripts. Transcripts were converted to cDNA, barcoded and sequenced on Illumina HiSeq4000 and Nova-Seq6000 sequencers to obtain approximately 50 thousand reads per cell.

### ScRNaseq data analysis

Raw sequencing reads were demultiplexed, mapped to the human reference genome (build GRCh38) and gene-expression matrices were generated using Cell Ranger (version 3.0.1). The resulting count matrices were further filtered in Seurat 3.1.0 (<https://satijalab.org/seurat/>) to remove cell barcodes with less than 200 genes, more than 30% of UMIs derived from mitochondrial genes, or more than 70,000 UMIs (to exclude putative doublets). This filtering resulted in 18500 single cells across all samples (Figure S1). After normalization, these cells were clustered using the Seurat workflow based on dimensionality reduction by PCA using the 1,602 most variable genes. We applied Harmony (Korsunsky et al., 2018) alignment (theta = 2) to correct for inter-sample variation due to experimental or sequencing batch effects. After assessment of clustering using a variety of dimensions, we used the default setting of the first 20 harmony dimensions to cluster the data and perform dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP). Differential expression and marker gene identification was performed using MAST (Finak et al., 2015). After exclusion of non-neoplastic cell populations the cells were re-clustered and re-projected to refine neoplastic subpopulation-specific differential expression gene lists that were further filtered to remove ribosomal protein genes.

Chromosomal CNVs of single cells from scRNaseq were inferred on the basis of average relative expression in variable genomic windows using InferCNV (<https://github.com/broadinstitute/inferCNV>). Cells classified as non-neoplastic were used to define a baseline of normal karyotype such that their average copy number value was subtracted from all cells.

Neoplastic subpopulations were characterized by direct examination of neoplastic-subpopulation specific gene lists, enrichment of comparator genesets and inference of regulatory gene networks. DAVID (Database for Annotation, Visualization, and Integrated Discovery: <https://david.ncifcrf.gov/>; version 6.8) was used to measure enrichment of GOterm Biological Process Direct genesets in subpopulation signatures, providing gene ontology mappings directly annotated by the source database. Subpopulation signature genesets included the top 250 differentially expressed subpopulation markers ranked by adjusted p value and filtered

to remove ribosomal genes (Table S6). Enrichment of select comparator genesets (Table S6) was performed by hypergeometric testing. Normal mouse ependymal cell and qNSC genesets, human radial glia subpopulation genesets, and mouse cerebellar developmental lineage genesets were obtained from published scRNaseq datasets (Pollen et al., 2015; Shah et al., 2018; Vladoiu et al., 2019). GBM intra-tumoral anatomic structure transcriptomic genesets were obtained from the Ivy Glioblastoma Atlas Project [<http://glioblastoma.alleninstitute.org/>].

EPN molecular subgroup PFA, PFB, RELA and YAP signature genesets (Table S6) were obtained by differential expression analysis of published EPN transcriptomic dataset GSE64415 (Pajtlér et al., 2015). Comparator genesets for classification subgroups PFA1 and PFA2 (Table S6) were obtained by differential gene expression analysis of our subgroup-classified bulk-tumor transcriptomic data based on methylation profile classification (Table S1).

Single-cell regulatory network inference and clustering (SCENIC) was used to identify gene regulatory networks at the single cell level (Aibar et al., 2017). Slingshot lineage analysis (Street et al., 2018) and Destiny diffusion mapping (Angerer et al., 2015) were utilized for pseudotime modeling due to their ability to identify multiple trajectories.

Putative cell-to-cell communication between neoplastic subpopulations were identified based on overexpression of a receptor by one cell type and overexpression of an interacting ligand by the other cell type. Overexpression of the receptor by a particular cell type is defined as an incoming signal, and the expression of the ligand as an outgoing signal. Ligand-receptor pairs used in this analysis were defined by Ramiłowski et al. (2015). The product of ligand and receptor subpopulation-specific fold change was used to rank communications between MEC and CEC.

### Bulk-tumor methylome and transcriptome analyses

DNA and RNA were extracted from snap frozen EPN-PF surgical tumor samples (QIAGEN, Allprep DNA/RNA mini kit). Forty nine surgical tumor samples were from initial presentation and 3 samples (all RELA) from first recurrence (Table S1). Methylome analysis of DNA from presentation samples (n = 52) was performed using the Illumina 850K methylation array. Resulting IDAT files were uploaded to [MolecularNeuropathology.org](https://www.molecularneuropathology.org/mnp) (<https://www.molecularneuropathology.org/mnp>) which provided subgrouping into RELA (n = 4), RELA-like (n = 1), YAP (n = 1), PFA (n = 44) and PFB (n = 2) molecular subgroups and chromosomal copy number variants (CNVs) (Table S1). ST subgroup assignment was confirmed by IGV analysis of RNaseq (Illumina NovaSeq6000) profiles of ST samples, with both RELA and RELA-like samples exhibiting C11ORF95-RELA fusions, and a YAP-MAMLD1 fusion in the YAP sample. Methylation profiles of PFA samples were used to assign PFA samples to PFA1 (n = 27) and PFA2 (n = 17) and subtypes within these subgroups using the same approach as used by the study initially defining these subtypes (Pajtlér et al., 2018) (courtesy of Dr. Martin Sill, DKFZ, Heidelberg, Germany). Transcriptomic analysis was performed at our institution using the Human Genome U133plus2 Array (Affymetrix) platform as described previously (Griesinger et al., 2017). Microarray CEL datafiles were background corrected and normalized using the guanine cytosine robust multi array average (gcRMA) algorithm to generate a gene expression matrix of log<sub>2</sub> expression values. To reduce error associated with multiple testing, a filtered gene expression matrix was created containing the highest expressed probe across all samples for each gene that possessed multiple probe sets. This was further filtered to remove probe sets that were expressed below a threshold level that denoted absence of expression in any sample. PFA samples were also assigned to PFA1 and PFA2 subgroups based on transcriptomic profiles using non-negative matrix factorization (NMF; <https://cloud.genepattern.org>) as described previously (Griesinger et al., 2017). PFA transcriptomic profiles segregated into 2 groups corresponding to PFA1 (n = 22) and PFA2 (n = 22) (Table S1). Our microarray data were merged with primary PFA EPN transcriptomic datasets from The German Cancer Research Center DKFZ, Heidelberg, Germany (Accession number GSE64415; n = 39) and St Jude Children's Research Hospital, Memphis, USA (Accession number GSE100240; n = 39) with full methylation-based subgroup annotation (Courtesy of Dr. Marcel Kool, Heidelberg, Germany).

### Deconvolution

CIBERSORT was used to perform deconvolution of bulk-tumor PFA transcriptomes of 44 patient samples, using scRNaseq subpopulation signature genes (Newman et al., 2015). The combined transcriptomic dataset (n = 121) was used unlogged and used as the mixture file. A signature gene input matrix was generated from log<sub>2</sub> values of normalized raw scRNaseq expression data averaged for each neoplastic and non-neoplastic cluster (Table S6). CIBERSORT was run using these datasets with 100 permutations.

### Immunohistochemistry

Immunohistochemistry was performed on 5- $\mu$ m formalin-fixed, paraffin-embedded tumor tissue sections using a Ventana autostainer. All immunostained sections were counterstained with hematoxylin. Neuropathological review of staining and blinded IHC scoring was then performed (A.G., A.M.D.). Antibody details are provided in Table S5F.

### Survival studies

Survival data were obtained with ethics committee approval (COM-IRB 09-0906) (Table S1). All patients were treated at our institution with standard treatment of gross total resection and radiotherapy (radiation withheld until > 1 year of age for those patients less than one year old). Survival analyses were performed using Prism (GraphPad) software, with outcome censored at 10 years. Hazard ratios

(HRs) for progression free survival (PFS) and overall survival (OS) were estimated using log-rank (Mantel–Cox) analysis of high versus low (i) expression of each subpopulation as defined by median deconvoluted subpopulation fractions, and (ii) IHC as defined by median score.

#### **Perturbation studies of PFA EPN *in vitro***

Microarray gene expression data from EPN PFA cell lines MAF-811 and MAF-928 that had been cultured as monolayers or 3D structures was obtained from a prior study by our laboratory (Accession number GSE86574) (Amani et al., 2017). We identified the top 250 most differentially overexpressed genes in 3D compared to monolayer-cultured cells (Table S6) which were then subject to hypergeometric analysis, as described above, to quantify enrichment of EPN neoplastic subpopulation-specific genesets (Table S6). For hypoxia treatment, PFA EPN cell lines MAF-811, MAF-928, and a short term culture established from sample 1239 were cultured in Optimem/15% FBS in as monolayer either (1) in hypoxic conditions (2% O<sub>2</sub>), or (2) in normoxia for 72 hours. RNA was then harvested from cells using an Allprep RNA/DNA isolation kit (QIAGEN) and subjected to RNaseq as follows. Quality of extracted mRNA was assessed using DNA Analysis ScreenTape (Aligent Technologies) to ensure sufficient quality for library preparation. Following quality assessment, cDNA libraries were generated using the Universal Plus mRNA-Seq Sample Prep Kit (Nugen). Libraries were then sequenced using the Illumina NovaSeq6000 platform with paired-end reads (2x151). On average 70 million reads were collected for each sample. The resulting sequences were filtered and trimmed to remove low-quality bases (Phred score < 15). The remaining sequences were analyzed using a custom computational pipeline consisting of the open-source gSNAP, Cufflinks. R was used for alignment and discovery of differential gene expression. Reads were mapped to the human genome (GRCh38) by gSNAP. Gene expression, as fragments per kilobase per million (FPKM), was derived by Cufflinks. Genes differentially expressed between hypoxia and normoxia were ranked by difference in FPKM. The top 250 upregulated genes in hypoxia in each cell line (Table S6) were subject to hypergeometric analysis to quantify enrichment of EPN neoplastic subpopulation-specific genesets.

#### **QUANTIFICATION AND STATISTICAL ANALYSIS**

Statistical analyses were performed using R bioinformatics, Prism (GraphPad), and Excel (Microsoft) software. Details of statistical tests performed are included in figure legends. For all tests, statistical significance was defined as  $p < 0.05$ .