bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GANGLIOGLIOMA DEEP TRANSCRIPTOMICS REVEALS PRIMITIVE NEUROECTODERM NEURAL PRECURSOR-LIKE POPULATION

Joshua A Regal[1], Maria E Guerra García[1], Vaibhav Jain[2], Vidyalakshmi Chandramohan[3], David M Ashley[3], Simon G Gregory[2], Eric M Thompson[3], Giselle Y López[3,4], Zachary J Reitman[1,3,4*]

[1] Department of Radiation Oncology,
[2] Duke Molecular Physiology Institute,
[3] Department of Neurosurgery,
[4] Department of Pathology,
Duke University, Durham, NC 27710, USA.
* To whom correspondence should be addressed. Electronic address: zjr@duke.edu.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Gangliogliomas are brain tumors composed of neuron-like and macroglia-like components that occur in children and young adults. Gangliogliomas are often characterized by a rare population of immature astrocyte-appearing cells expressing CD34, a marker expressed in the neuroectoderm (neural precursor cells) during embryogenesis. New insights are needed to refine tumor classification and to identify therapeutic approaches. We evaluated five gangliogliomas with single nucleus RNA-seq, cellular indexing of transcriptomes and epitopes by sequencing, and/or spatially-resolved RNA-seq. We uncovered a population of CD34-rich neoplastic cells with mixed neuroectodermal, immature astrocyte, and neuronal markers. Gene regulatory network interrogation in these neuroectoderm-like cells revealed control of transcriptional programming by TCF7L2/MEIS1-PAX6 and SOX2, similar to that found during neuroectodermal/neural development. Developmental trajectory analyses place neuroectoderm-like tumor cells as precursor cells that give rise to neuron-like and macroglia-like neoplastic cells. Spatially-resolved transcriptomics revealed a neuroectoderm-like tumor cell niche with relative lack of vascular and immune cells. We used these high resolution results to deconvolute clinically-annotated transcriptomic data, confirming that CD34+ cell-associated gene programs associate with gangliogliomas compared to other glial brain tumors. Together, these deep transcriptomic approaches characterized a ganglioglioma cellular hierarchy - confirming CD34+ neuroectoderm-like tumor precursor cells, controlling transcription programs, cell signaling, and associated immune cell state. These findings may guide ganglioglioma tumor classification, diagnosis, and therapeutic investigations.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Glioneuronal tumors (GNTs) are brain tumors that are composed of both neoplastic neuron-like and macroglial-like components[1]. GNTs account for approximately 2% of all primary brain tumors and most often occur in children and young adults[1]. GNTs include gangliоglioma (GG)[1]. The majority of GNTs are slow-growing tumors that do not result in death of the patients[1]. Nevertheless, GNTs can be associated with significant morbidity due to associated seizure disorders, risks of surgical resection, and tumor recurrence[1]. The standard of care therapy is surgical resection[1]. However, new treatment strategies are needed for patients with anaplastic, recurrent, or progressive GNTs[1]. Genomic alterations in gangliоglioma are minimal[2–4]. Nonetheless, at least 90% of GGs have driver alterations in the BRAF/MAPK pathway, with about half of GGs carrying BRAF V600E[2–4]. Additional recurrent alterations include chromosome 7 gains with *BRAF*, *KIAA1549*, and *EGFR* copy number gains as well as homozygous *CDKN2A/B* loss[2–4]. BRAF/MAPK targeted therapies are undergoing clinical testing for BRAF-mutant glioneuronal and other brain tumors, with early evidence suggesting clinically-significant anti-tumor activity[5,6]. However, further understanding of how to molecularly target GNTs is needed. Additionally, immunotherapeutic advances have resulted in significant breakthroughs for many malignancies, but immunotherapies have not yet improved outcomes for GNTs[1,7–9]. Better understanding of tumor microenvironment may help to identify opportunities for effective immunotherapy.

Opportunities exist to further understand GNT development and maintenance and thereby improve upon GNT treatment. There is now decades of work supporting the existence of cancer stem cells for myriad cancer types, and putative primitive progenitor cell types have been identified in other low grade brain tumors, such as pilocytic astrocytoma[10,11]. Understanding cellular hierarchy is important for understanding tumorigenesis, tumor maintenance, and tumor treatment, but little is understood about the potential for or nature of GNT stem cells. ScRNA-seq approaches have revealed that many subtypes of brain tumors contain tumor cells that transcriptionally resemble normal brain progenitor cell subtypes[12,13]. However, this approach has not yet been used to investigate GNTs. Many pediatric low grade gliomas such as GGs contain tumor cells that express CD34, (an endothelial and hematopoietic stem cell marker but also) a transient marker of neuroectodermal neural precursor cells during neural development[14–16]. The precise nature of these cells is not well understood. However, because of their primitive neuroectodermal neural precursor cell marker expression and histologic appearance, it is tempting to hypothesize that GNT CD34+ cells are neoplastic stem or precursor cells that transcriptionally resemble normal neuroectodermal neural precursor cells.

In addition to the therapeutic challenges, classification of GNTs is a major diagnostic challenge with resultant impacts on understanding of prognosis and medical decision making[1]. Brain tumor classification has evolved enormously over the last several years with the gradual incorporation of tumor genetic features leading to significant improvements in tumor classification and medical decision making[17,18]. Tumor heterogeneity and associated under-sampling are notorious issues with brain tumors, resulting in significant rates of under-grading and under-treatment but can sometimes be compensated for to a large extent by accounting for genetic factors[19]. Many types of pediatric brain tumors are now being understood as separate entities, with implications for prognosis and medical decision making[17,18].

Single nucleus (or cell) RNA-sequencing (snRNA-seq), cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), and spatial transcriptomics (stRNA-seq) have been carried out on a number of brain tumor subtypes to provide insights to guide

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the development of new therapeutic strategies and new classification schemes for the tumors[10,12,13,20–24]. However, these high-resolution transcriptomic techniques have not yet been published for GNTs. As outlined above, a deep understanding of GNT cellular composition is needed to generate hypotheses for therapeutic interventions that may benefit select patients with GNT and to provide insights that may refine the GNT pathologic classification. We hypothesize that the GNT CD34+ cells transcriptionally resemble embryonic neuroectodermal (neural precursor) cells that may serve as progenitors for the other neuron-like and macroglia-like tumor cells. We additionally hypothesize that a detailed definition of GNT immune cell composition and activation status will further the understanding of the utility of immunotherapeutic targeting for GNT. Consequently, we endeavored to characterize a major GNT subtype (ganglioglioma) using snRNA-seq, CITE-seq, and stRNA-seq.

Results

Neoplastic neuron-like, macroglia-like, and *CD34*-associated neuroectoderm-like components in GNTs

To elucidate normal and neoplastic cell states in GNTs, single nucleus RNA-seq was carried out on 4 GG tumors (**sTable 1**). This yielded high-quality RNA-seq profiles from 34,907 nuclei. Nuclei were subjected to non-linear dimensional reduction and clustered based on k-nearest neighbor search[25–28] yielding 30 clusters (**Figure 1A**). Nuclei largely segregated by tumor-of-origin except for select normal-appearing clusters (**sFigure 1A-G**). Normal and putative neoplastic cells were initially identified by cell type markers (**Figure 1B and sFigure 1H**).

Normal brain, immune cell, and vascular clusters expressed classic cell type markers (**Figure 1B and sFigure 1H**). Normal clusters included macroglia comprising astrocytes (*GFAP*-, *AQP4*-, and *ALDH1L1*-expressing, cluster 23), oligodendrocytes (*MBP*-, *MOG*-, and *CNP*-expressing, cluster 20), and oligodendrocyte precursor cells (OPCs, *GPR17*-, *CSPG4*-, *OLIG1*-, and *OLIG2*-expressing, cluster 21)[29–31]. Cluster 15 was composed of neurons defined by *RBFOX3* expression including *CUX2*-, *NRGN*-, *SLC17A7*-, and/or *TBR1*-expressing excitatory neuron subclusters and *GAD1*-, *SLC32A1*-, *SST*-, and/or *VIP*-expressing inhibitory neuron subclusters[29,30,32]. Immune cell clusters included a microglial supercluster (*PTPRC*-, *CD14*-, and *P2RY12*-expressing, clusters 1, 5, 11, and 17) and lymphocyte clusters (*PTPRC*-expressing but not *CD14*-expressing, clusters 7 and 28 with cluster 7 nuclei expressing T cell marker (CD3, CD4, and CD8)-encoding transcripts)[29,30]. We further identified an endothelial cluster (*VWF*-, *PECAM1*-, and *CD34*-expressing, cluster 12) and vascular leptomeningeal cells (VLMCs, cluster 22, *DCN*-, *COL1A1*-, *COL1A2*-expressing)[29,30,33,34]. After identifying brain stromal cells with high confidence, we turned our attention to the remaining putative neoplastic cell clusters.

The remaining clusters contained abnormal appearing macroglia- and neuronal-like cells (cluster 0, 2-4, 6, 8-10, 13, 14, 16, 18, 19, 24-27, and 29). These clusters contained abnormal combinations of markers (despite doublet removal during preprocessing, **Figure 1B and sFigure 1H**). For instance, cluster 6 cells tended to express *CD34* (an endothelial, hematopoietic stem, and neuroectodermal neural precursor cell marker), *RBFOX3* (normally a pan-neuron-specific marker), and *AQP4* (normally an astrocyte-specific marker)[14,29,30,33,34]. Among these putative neoplastic cells, cluster 6 was the only *CD34*-rich cluster. In terms of cell type proportions, 22932/34907=66% of cells were neoplastic-appearing. Of these neoplastic cells, 694/22932=3.0% were in cluster 6, 259/22932=1.1% were *CD34*+, and

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

168/22932=0.7% were cluster 6, *CD34*+ cells. Substantially smaller proportions of *CD34*+ cells were present among clusters 24 (5% *CD34*+), 3 (1.6% *CD34*+), 19 (1.2% *CD34*+), 14 (0.9% *CD34*+), 10 (0.5% *CD34*+), 8 (0.4% *CD34*+), 13 (0.2% *CD34*+), and 0 (0.02% *CD34*+). Hence, analysis of initial cell markers identified the majority of ganglioglioma cells as neoplastic-appearing with a rare *CD34*-rich population possibly representing tumor neuroectoderm neural precursor cell-like stem cells.

Preliminary normal and neoplastic cell identification was confirmed by analysis of top differentially expressed genes (**Supplementary Information** and **sTable 2**), inference of copy number variation, and Seurat-based label transfer. Inference of copy number variations using inferCNV[35] identified clonal del1p and subclonal del14 exclusively in the putative neoplastic cell clusters in tumor 4 (**sFigure 1I**), providing support that these clusters represented neoplastic cells. Note that most gangliogliomas do not carry large CNVs[3,36], so the absence of such alterations for snRNA-seq cells from the other tumors does not rule out their assignment as neoplastic. We next mapped snRNA-seq profiles to transcriptional atlases from developing brain[37–41], adult brain[32,42,43], primary brain tumors[10,12,44], and tumor-associated immune cells[45–48]. This mapping reinforced and refined stromal cell type assignments (examples in **Figure 1C** and **sFigure 1J-L**). Nuclei of the suspected neoplastic clusters generally had much more mixed and uncertain mapping, consistent with these representing abnormal, neoplastic cells. Interestingly, the *CD34*-rich cluster 6 nuclei were confidently identified as immature astrocytes where represented in reference atlases. For instance, cluster 6 mapped confidently to first-/second-trimester (i.e. immature) astrocytes in a developmental brain atlas (**Figure 1C**)[37], and cluster 6 was confidently identified as protoplasmic/immature astrocyte by mapping to a glioblastoma atlas (**sFigure 1L**)[44]. Other neoplastic-appearing clusters tended to map to either neurons or macroglia at different developmental stages. For instance, clusters 0, 4, 8, 9, 10, 18, and 19 mapped to first-/second-trimester neurons while cluster 13 mapped to developing OPCs and clusters 2, 16, and 29 mapped more closely to ventricular radial glia (vRG; **Figure 1C**)[37]. Thus, snRNA-seq results support the presence of large neuron-like and macroglia-like populations in addition to a smaller neuroectoderm neural precursor cell-like compartment among neoplastic GG cells.

GG *CD34*-associated neuroectodermal neural precursor-like cells exhibit stem cell states

Many cancers, including gliomas, are composed of a cellular hierarchy containing stem-like cells capable of self-renewal, differentiation, tumorigenecity, tumor progression, or resilience in the face of anti-tumor therapy and ultimately tumor regrowth[20,21,35,49]. Understanding these cells is of special interest because their identification could be used for better tumor classification, prognostication, and therapeutic targeting. Cell typing and mapping above identified a neoplastic, *CD34*-rich cluster 6. Such CD34+ cells have previously been hypothesized to represent ganglioglioma tumor precursor/stem cells, but their transcriptomic profile and location in the tumor cell hierarchy was not previously well understood.

Given the working hypothesis that gangliogliomas arise from neuroectodermal neural precursor-like cells, neoplastic cells were next interrogated for individual neuroectodermal markers in addition to *CD34*[50]. Interestingly, cluster 6 identity and *CD34* expression were each associated with expression of neuroectodermal neural precursor cell markers *PAX6*, *SOX2*, and *MSI1* (**Figure 2A**)[51–53]. Hence, analysis of cell markers supported our hypothesis that neoplastic *CD34*+ cells resemble primitive neuroectoderm neural progenitor cells.

We next sought to identify where the neuroectoderm neural precursor cell-like cluster 6 cells fall within a putative GG neoplastic cell developmental hierarchy. First, we used

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CellRank/CytoTRACE to infer pseudotime[54,55]. Analysis of our neoplastic-appearing ganglioglioma nuclei in this manner assigned cluster 6 nuclei or *CD34*+ nuclei (and particularly cluster 6, *CD34*+ nuclei) the earliest pseudotimes (**Figure 2B-E**). We next used SCENT to calculate signaling entropy rates (SR), which closely approximate stemness[56]. This method largely recapitulated the results with CellRank/CytoTRACE (**Figure 2F-I**). Finally, we used RNA velocity to infer temporal relationships between cell states[57,58]. Multiple RNA velocity methods consistently agreed that cluster 6 and/or *CD34*+ neoplastic cells appeared to be the most primordial of neoplastic cells (**Figure 2J-M**). Therefore, several orthogonal methods showed that neoplastic cells represented in cluster 6 and/or *CD34*+ appeared particularly potent (i.e. stem cell-like), consistent with these representing primitive, neoplastic stem/precursor cells.

To further pin-down stem cell-like states and possibly the cell type of origin for ganglioglioma, we combined analysis of neuroectodermal neural precursor markers and cellular hierarchy results. Co-expression of *CD34*, *PAX6*, *SOX2*, and *MSI1* (after stringent counter selection against vascular cells using *PECAM1*, *VWF*, *DCN*, and *COL1A2*), was particularly strongly associated with ganglioglioma neoplastic cell stemness (**sFigure 2A**). To evaluate the timing of disappearance of such cell states during normal development, we tested existing single cell brain atlases for the presence of *CD34+PAX6+SOX2+MSI1+PECAM1-VWF-DCN-COL1A2-* cells. We first noted that prenatal brain reference atlases we probed all contained small but appreciable numbers of these cells: 14/2394 in Zhong *et al*.[41], 10/4129 in Nowakowski *et al*.[37], 7/25161 in van Bruggen *et al*.[39], and 9/48215 in Eze *et al*.[38] Considering data drop out, the actual potent precursor cells of the intended type are likely somewhat more prevalent. In contrast, adult brain reference atlases had no *CD34+PAX6+SOX2+MSI1+PECAM1-VWF-DCN-COL1A2-* cells detected: E.g. 0/47432 in the Allen Institute multiple cortical areas atlas[32] and 0/78886 cells in Nagy *et al*.[43] Interestingly, a very large, recently published brain single cell atlas spanning the third trimester to 40 years of age helped to fill the gaps in developmental data during childhood and adolescence[40]; 0 out of 154748 cells in the Herring *et al*. atlas were *CD34+PAX6+SOX2+MSI1+PECAM1-VWF-DCN-COL1A2-.* A potential confounder is the possible spatial variability of this cell type. However, when limiting our analysis to the pre-frontal cortex (due to there being robust data spanning development for this region), we still found an apparent disappearance of *CD34+PAX6+SOX2+MSI1+PECAM1-VWF-DCN-COL1A2-* cells during fetal development. Prenatal pre-frontal cortex *CD34+PAX6+SOX2+MSI1+PECAM1-VWF-DCN-COL1A2-* cells were found by Nowakowski *et al*. (2 cells out of 1076)[37] and Zhong *et al*. (14 cells out of 2394)[41]. No such cells were detected in the pre-frontal cortex third trimester to 40 years old (0/154748)[40] or a purely adult pre-frontal cortex atlas (0/78886)[43]. Hence, these results overall support the disappearance of the neuroectodermal neural precursor cells that ganglioglioma stem-like cells most resemble during normal brain development (fetal development for the pre-frontal cortex) and suggest that the ganglioglioma cell of origin may in fact arise during fetal brain development.

Ligand-receptor analysis reveals PTN-PTPRZ1, FGF family, and PDGF family communication among GG neoplastic cells

Intercellular communication pathways were interrogated in an unbiased manner using CellChat[59] (**sFigure 3A-B, sTables 3-5**). This analysis suggested neoplastic tumor cell PTPRZ1 targeting by PTN, the latter largely produced by the neoplastic cells (**Figure 3A and sFigure 3C-D**). 57% of neoplastic cluster-neoplastic cluster pairs exhibited a PTN-PTPRZ1 interaction (p-value <0.01). Every neoplastic cluster had at least one cluster by which it was

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

targeted (p-value<0.01). Cluster 6 was particularly active in signaling, as a source for 17/18 neoplastic clusters (p<0.01, all except cluster 18) and as a target from 10/18 neoplastic clusters (p<0.01). PTPRZ1 is a receptor tyrosine phosphoprotein phosphatase[60]. PTN binding antagonizes PTPRZ1, which has been shown to have pleiotropic effects including activation of MAPK and AKT signaling axes[60]. A mouse model of BRAF-V600E neuroectodermal tumors required a second hit with increased AKT/mTOR signaling to produce ganglioglioma-like tumors[61]. These results identify a neoplastic-cell-to-neoplastic-cell interaction that is known to drive the signaling axes needed for gangliogliomagenesis as a major feature of GG in general, and of GG cluster 6 stem/neuroectoderm neural precursor-like cells in particular.

We also localized neoplastic cell compartments that take part in FGF and PDGF family ligand-receptor interactions. FGFR alteration is associated with a substantial minority of pediatric low grade gliomas[2–4]. PDGF(R) aberrant hyperactivity is potently oncogenic, more typically associated with high grade gliomas[62]. FGF1-FGFR1 interactions occurred between most clusters of neoplastic cells or OPCs (p<0.01 for 37% of possible neoplastic cluster-neoplastic cluster interactions, cluster 6 as a source p<0.01 for 10/18 neoplastic clusters including itself, cluster 6 as a target p<0.01 for 12/18 neoplastic clusters; **Figure 3B** and **sTables 4-5**). Other FGF-FGFR family members exhibited highly significant interactions between multiple neoplastic clusters (**sFigure 3E-I** and **sTables 4-5**). The PDGF family pathway was also a top hit (**sTable 3**). For instance, we identified interactions between PDGFA-PDGFRB in cluster 6 (p<0.01 for cluster 6 targeting by 5/18 neoplastic clusters, including itself, with no other significant neoplastic cluster-neoplastic cluster interaction (p=1)) as well as endothelial cluster 12 and VLMC cluster 22 (**Figure 3C** and **sTables 4-5**). We also detected PDGFB-PDGFRB and PDGFD-PDGFRB targeting directed at cluster 6 and vascular cells (**sFigure 3J-K** and **sTables 4-5**). These results pinpoint GG neoplastic cell states that take part in FGF and PDGF signaling interactions, and reveal that such processes are especially active in neuroectoderm neural progenitor cell-like cluster 6 cells.

BRAF and AKT pathway signaling in neoplastic cells

We next sought to identify signaling programs that may function downstream of the ligand-receptor interactions to drive growth and maintenance of the neoplastic cells. To ascertain important cellular signaling pathways in ganglioglioma, neural cells were subject to gene set enrichment analysis (GSEA) using clusterProfiler for gene ontology (GO), KEGG, and Wiki pathways[63,64].

To uncover neoplastic cluster-associated pathway suppression or activation, neoplastic-appearing ganglioglioma clusters were compared to normal-appearing neural clusters (clusters 15, 20, 21, and 23). Consistently among the most suppressed pathways (in neoplastic cells) were various pathways responsible for oxidative phosphorylation or ribosomal function, whether comparing by individual cluster or using neoplastic-appearing cells altogether (**Figure 4A and sFigure 4A-F**). Oxidative metabolism suppression is consistent with the Warburg effect typical of neoplastic cells and offers further support for our identification of cells as neoplastic or stromal[65].

In order to ascertain the pathways significant to *CD34*+ neoplastic cells in particular, we compared neoplastic *CD34*+ cells to neoplastic *CD34*- cells. The most *CD34*+ cell-activated pathways were those involving oxidative phosphorylation and ribosomal machinery (**Figure 4B and sFigure 4G-L**). Unsurprisingly given neoplastic *CD34*+ cells' promiscuous use of signaling nodes, they had very little in the way of pathway suppression (recall the high signaling entropy rate discussed above, **Figure 4B**, and **sFigure 4G-L**). Nearly identical results were obtained when comparing cluster 6 *CD34*+ cells to the remaining neoplastic-

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

appearing cells. Close inspection of KEGG pathways of interest showed especially brisk expression of the PKB/AKT pathway machinery from FGF through downstream AKT effectors in neoplastic *CD34*+ versus neoplastic *CD34*- cells (**sFigure 4K**). Additionally, there was generally higher expression of BRAF pathway machinery encoding transcripts in *CD34*+ cells, including BRAF itself (**sFigure 4K**). Pluripotency machinery encoding transcripts were generally greatly enriched in the *CD34*+ cells, including the transcription factor *SOX2*, a SOX B1 member which contributes to embryonic, neural stem, and progenitor cell self-renewal and pluripotency (very early; **sFigure 4L**)[53]. These results were overall congruous with the preceding analysis. Particularly interesting is the persistent theme of signaling via AKT in addition to BRAF, particularly in the most stem-like of the neoplastic cells, the *CD34*+ and/or cluster 6 cells. For the AKT pathway, there may very well be contribution by autocrine signaling via FGFs and/or PTN.

Neoplastic transcriptional programs and putative drivers

We next characterized transcription factors that may account for compartmental expression of the gene programs identified above. To do so, gene regulatory networks were reconstructed in an unbiased manner using SCENIC[66]. Top candidate transcription factors for both cluster 6 and *CD34*-expressing neoplastic cells included paired box 6 (PAX6) and its transcriptional activators myeloid ectopic viral integration site homeobox 1 (MEIS1) and transcription factor 7 like 2 (TCF7L2; **sFigure 5A-D**)[67,68]. In the healthy adult brain, these transcription factors are not normally coexpressed[29,30]. More interestingly, PAX6 is a neuroectodermal neural precursor cell marker, and the MEIS1-PAX6 cascade appears critical to early neuroectodermal cell fate determination[38,51,67]. In addition, TCF7L2 coordinates PAX6 activation in neural cells and modulates MYC as part of the WNT pathway[68]. *PAX6*, *MEIS1*, and *TCF7L2* concentrations as well as SCENIC-calculated regulon scores correlated strongly with ganglioglioma neoplastic cell stemness by CellRank/CytoTRACE pseudotime or SCENT signaling entropy rate (**Figure 5A** and **sFigure 5E-F**). SOX2 also appeared as a factor with significantly increased activity (and transcript expression) among certain neoplastic cells, particularly those most stem-like (including among cluster 6) and/or *CD34*-expression (**Figure 5B**). SOX2 contributes to embryonic, neural stem, and progenitor cell self-renewal and pluripotency (very early, including in the neuroectoderm) and is not normally appreciably expressed in the adult brain[29,30,53]. These observations implicate PAX6, MEIS1, TCF7L2, and SOX2 programs in the more primitive, neuroectoderm neural precursor-like (cluster 6 and/or *CD34*+) tumor cells, and these programs are possibly controlling tumor cell potency/stemness/differentiation.

Immune landscape

Myeloid cells

In healthy brain, myeloid cells are predominantly resident microglia; however, others have established the presence of bone marrow derived macrophages associated with brain tumors[22]. To determine the nature of myeloid cells in gangliogliomas, we first used markers of *a priori* interest to further decipher cell states. Ganglioglioma clusters 1, 5, 11, and 17 were *PTPRC*-, *ITGAM*-, *CD14*-, and *P2RY12*-expressing consistent with these representing microglia (**Figure 6A** and recall **Figure 1B** and **sFigure 1H**)[29,30,69]. Myeloid cells lacked appreciable expression of the classical pro-inflammatory markers *IL1A*, *IL6*, *TNF*, and *CD40* (all myeloid cells *IL6*-; otherwise **Figure 6A**)[45]. In terms of anti-inflammatory activity, they did not express appreciable levels of *IL10*, but there was diffuse expression of *TGFB1* (**Figure**

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6A)[45]. Moreover, when looking at the broader cellular context, lymphocytes from clusters 7 and 28 expressed significant *TGFB1* and tumor clusters 3, 8, and 14 expressed substantial *CSF1*, which promote immunosuppressive and pro-inflammatory properties in macrophages, respectively (**Supplementary Information** and **sTable 2**)[45]. Clusters 1, 5, 11, and 17 all appeared to express significant amounts of *CD163* (except cluster 11) and *MSR1* (encodes CD204) and variable *MRC1* (encodes CD206), classic markers of immunosuppressive/tissue reparative/tumor promoting M2 macrophages (**Figure 6A**)[45]. In terms of classic pro-inflammatory M1 markers, myeloid cells expressed *TLR2* highly, *CD86*, variable *HLA-DR* (most in cluster 17), and very little *NOS2* (**Figure 6A**)[45]. This co-expression of M1 and M2 markers appears consistent with prior observations suggesting that the M1 versus M2 dichotomy does not seem to hold on the single cell level in tumor microenvironments[45]. Instead, M1 and M2 traits tend to be strongly positively co-varying in individual tumor-associated cells[22,45]. More recently, with the help of single cell typing of tumor-associated immune cells, it has been suggested that C1Q and SPP1 status best differentiate different types of tumor-associated macrophages and that these markers are also associated with prognosis[70,71]. C1Q expression appears to be associated with T lymphocyte recruitment and activation whereas SPP1 expression appears to be associated with tumor growth and metastasis. *C1Q* expression was variable in ganglioglioma myeloid cells, with high concentrations in cluster 17 and consistently elevated concentrations in cluster 11 (**Figure 6A**). *C1Q*+ cells were not distinct from *SPP1*+ cells, with *SPP1* expression high in cluster 11 and parts of clusters 17 and 5 (**Figure 6A**). These results overall support the presence of tumor-associated, aberrantly activated microglia.

We attempted to use differential gene expression analysis, multi-gene signature scoring, and k-nearest neighbors mapping to reference atlases to further type these cells. Top microglial cluster-overexpressed genes were enriched for complement-encoding transcripts (particularly C1Qs) and MHC class II-encoding transcripts, consistent with their expected roles in the complement cascade and professional antigen presentation, respectively (**sFigure 6A**). To confirm ganglioglioma myeloid cells represented microglia, Seurat-based label transfer was performed[28]. Using annotated tumor reference atlases differentiating microglia from macrophages[10,44] for label transfer, the bulk of myeloid cells were consistently more strongly identified with an overall microglial rather than macrophage signature (recall **sFigure 1K-L**). When ganglioglioma-associated myeloid cells were subjected to scoring (by UCell) based on established M1 and M2 signatures[45], the two scores appeared positively correlated both within each cluster and among cells of the supercluster, consistent with the aberrant activation that was found at the individual marker level and further supporting the lack of this dichotomy in the context of some tumors (**sFigure 6B-D**). Overall, differential gene expression analysis, multi-gene signature scoring, and k-nearest neighbors mapping to reference atlases confirmed ganglioglioma-associated microglial cell identity and refined their cell states.

Ganglioglioma-associated myeloid cellular hierarchy was evaluated with CellRank/CytoTRACE, SCENT signaling entropy rate, and scVelo RNA velocity[54–57]. These generally identified cluster 11 and 17 members of greater potency/stemness and therefore candidates for tumor-associated myeloid precursor states (e.g., **Figure 6B**).

Myeloid cell interactions were explored. Interestingly, myeloid cell-expressed concentrations of transcripts encoding lymphocyte co-stimulatory and co-inhibitory ligands appeared positively correlated throughout the supercluster (**Figure 6C**). Among the top ranked pathways in unbiased ligand-receptor analysis by CellChat and CellPhoneDB[59,72], microglia participated in significant antigen presentation, including as senders and receivers of the MHC class II pathway, senders for the MHC class I pathways (clusters 5, 11, 17), and

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

receivers for the APP-CD74 signaling pathway (MHC class II antigen processing, recall **sFigure 3A-B** and **sTable 3-5**). On the other hand, microglia also participated heavily in TGFB signaling, as senders and receivers, and they were receivers for CSF1-CSF1R (from neoplastic clusters 3, 14, and 8), consistent with a classically M2-promoting milieu and activities (**Figure 6D,** recall **sFigure 3A-B** and **sTable 3-5**)[45]. Hence, ganglioglioma-associated microglial cell interactions appeared along a spectrum of aberrant activation, consistent with what was ascertained about myeloid cell states earlier.

To ascertain important cellular signaling pathways in ganglioglioma-associated myeloid cells, microglia were subject to gene set enrichment analysis (GSEA) using clusterProfiler[63]. Among Wikipathways, the TYROBP causal network in microglia was among the top 2 activated pathways for each microglial cluster (**Figure 6E**). The microglia pathogen phagocytosis pathway was also among the top 2 activated pathways for clusters 5 and 17. Top activated GO pathways for clusters 11 and 17 were pathways related to the ribosome and translation whereas those for clusters 1 and 5 included B cell activation, regulation of immune response, leukocyte differentiation, lymphocyte activation, and innate immune response among other immune cell activation/differentiation pathways. These findings overall support the assignment of these cells as microglia and show a spectrum of pathway utilization among the microglia particularly as it pertains to modulation of immune response.

Lymphocytes

Understanding tumor-infiltrating lymphocytes has become of great basic, translational, and clinical interest of late[73]. To explore ganglioglioma tumor-infiltrating lymphocytes, we started by interrogating for lymphocyte markers. Ganglioglioma cluster 7 was largely *PTPRC*-, *CD2*-, and *CD3*-expressing (**Figure 7A**), consistent with T lymphocyte identity[29,30]. Cytotoxic T lymphocyte (CTL) activity is central to anti-tumor immune response[73]. Cells represented in the upper and right side of cluster 7 UMAP expressed the classic CD2+CD3+CD4-CD8+ cytotoxic T lymphocyte markers including granzymes and perforin (**Figure 7A**), consistent with these cells representing cytotoxic T lymphocytes[29,30]. On the other hand, the remaining cluster 7 cells expressed *CD4* but not *CD8* or other cytotoxic T lymphocyte-specific markers (**Figure 7A**), consistent with these representing helper T lymphocytes[29,30]. Cluster 28 was *PTPRC*-expressing but negative for *CD2* and *CD3* (**Figure 7A**). On the other hand, this cluster had expression of B cell markers *CD19*, *CD20* (*MS4A1*), *CD22*, and immunoglobulin components (**Figure 7A**), consistent with B lymphocyte identity[29,30]. Hence, interrogation for lymphocyte markers revealed two ganglioglioma-infiltrating lymphocyte clusters, T lymphocytes in cluster 7 and B lymphocytes in cluster 28. This was further supported by top cluster-specific differentially expressed transcripts (**sFigure 7A** and recall **sTable 2**) and Seurat-based label transfer (**sFigure 7B**).

To fine tune our understanding of tumor-infiltrating lymphocyte cell states, we explored cell markers further. The cluster 7 CTL subpopulation expressed a mixture of co-inhibitory receptor-encoding transcripts *CTLA4*, *PD1*, *TIGIT*, *LAG3*, *TIM3*, and *CD160* and co-stimulatory receptor-encoding RNAs *CD28*, *ICOS*, and *CD40LG* (**Figure 7B**)[29,30,45]. They had higher expression of exhaustion-related transcription factor-encoding RNAs *EOMES* and *TOX* than the precursor or pre-exhaustion factors *TBX21* and *HNF1A* (**Figure 7B**)[29,30,45]. These observations altogether suggest some degree of CTL dysfunction and exhaustion.

In addition to interrogation of individual markers, cell states were determined by signature score (rank-based, by UCell) based on subtype signatures previously evaluated in a single cell context[45]. Similar to what was observed for myeloid cell M1 versus M2 state, ganglioglioma-associated T lymphocyte M1 polarizing signatures and M2 polarizing signatures[45] were actually positively correlated (**sFigure 7C-D**). Signature scoring also

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

identified a G2/M subpopulation of lymphocytes in cluster 7 (**sFigure 7E**)[45]. UCell scores were significantly higher for the Azizi *et al*. CD8 T cell activation, pro-inflammatory, and cytolytic effector pathway signatures[45] for those cells previously identified as enriched for CTL markers (**sFigure 7E**), confirming this assessment. In sum, TIL signature-scoring confirmed and refined preliminary cell state assignments based on individual markers.

We next turned to lymphocyte intercellular interactions because they constitute important determinants of T lymphocyte anti-tumor and anti-inflammatory roles. First, we analyzed *a priori* CTL ligand-receptor interactions of interest. As was expected, MHC class I component-encoding expression was particularly robust among microglia and only subpopulations of neoplastic cells (**Figure 7C**)[29,30,45]. Some transcripts encoding co-stimulatory or co-inhibitory ligands, *CD86*, *LILRB4*, *CD58*, and *HAVCR2* were robustly expressed across the microglial supercluster but other such ligands were generally scarce (**Figure 7C**)[29,30,45]. In unbiased ligand-receptor interaction nomination by CellChat[59], the lymphocyte clusters were among the least interactive cluster in terms of both incoming and outgoing signals (**sFigure 7F** and recall **sFigure 3A-B**).

In terms of cellular signaling, GSEA[63] revealed cluster 7 top activated pathways to include T cell activation (from GO); T cell receptor signaling (from KEGG); and TCR signaling, modulators of TCR signaling and T cell activation (WP; **Figure 7D and sFigure 7G-H**). Cluster 28 top activated pathways included B cell receptor signaling, B cell differentiation, B cell activation, antigen-mediated signaling, immune response-regulating, lymphocyte activation, regulation of immune response, and regulation of lymphocyte activation (GO); BCR signaling and NF-κB signaling (KEGG); and BCR signaling (WP; **Figure 7D and sFigure 7G, I**). Unbiased transcription factor nomination by SCENIC[66] concurred with the spectrum of T cell activation among cluster 7 and B cell activation among cluster 28 (**sFigure 7J-K**). Altogether, these findings further confirmed preliminary lymphocyte typing and state.

Validation with CITE-seq

To validate the above findings, fresh ganglioglioma tissue was subjected to CITE-seq. We analyzed n=3725 cells from three tumors with CITE-seq (recall **sTable 1**). These findings reiterated much of the results from snRNA-seq, but due to the smaller number of cells used for CITE-seq, the latter assays appeared insensitive to some of the more rare phenomena observed in the snRNA-seq data (please see **Supplementary Information**, **sFigure 8,** and **sTable 6** for more information). To correlate RNA profiles with the presence of expressed epitopes, eight epitopes were simultaneously probed, including CD34. CITE-seq analysis confirmed a strong correlation between *CD34* expression and expression of the CD34 epitope commonly used for diagnostic purposes ($R=0.37$, $p<2.2e-16$ by Pearson correlation test, **sFigure 8S).** These results confirm the snRNA-seq results using an orthogonal method which includes protein-level data (CITE-seq).

Spatial transcriptomics identified neuroectoderm neural precursor-like cells deep in neoplastic niches

To determine spatial context of tumor cells, four GGs analyzed by snRNA-seq above were subjected to spatially-resolved transcriptomic profiling (stRNA-seq). H&E slides were manually neuropathologist (GYL)-annotated with exclusion of spots with significant artifact, folded tissue, tissue degradation, or blood products. StRNA-seq yielded a mean of n=1791 high-quality RNA-seq profiles per tumor in 55 µm spots (**sFigure 9A-D**; note that with this spot size, there are multiple cells per spot). Two of the four slides show appreciable

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

histological variation. Tumor 1 had distinct (prominently or minimally/mildly) myxoid, hypercellular, hypocellular, and vascular areas making up 10 distinct annotations (**sFigure 9C**). Tumor 4 had (prominently) myxoid, non-myxoid, generic tumor, and vascular areas making up 4 distinct annotations (**sFigure 9D**). KNN clustering correlated with these distinctions (**sFigure 9E-F**). Final clusters were determined based on the combination of histopathological annotation and kNN clustering (**Figure 8A-B**).

To characterize these distinct tumor regions, top cluster-specific differentially expressed genes were determined, and cluster-variable pathway activation was probed by clusterProfiler[63]. Several cancer-relevant patterns emerged (**Figure 8C, sFigure 9G,** and **sTable 7**). There was a clearly inverse correlation between activation of immune cell response/effector function pathways and activation of neuronal- (or macroglial-) function related pathways. For instance, in tumor 1, clusters (prominently) myxoid 1, (prominently) myxoid 2, and hypocellular 3 were each characterized by suppression of various immune response pathways but with activation of neuronal morphogenesis pathways ((prominently) myxoid 1), GO astrocyte projection ((prominently) myxoid 2), or neuron-related pathways (hypocellular 3, **Figure 8C**). In contrast, in tumor 1, clusters mildly myxoid, hypocellular 1U2, and hypercellular 1U2 each showed activation of numerous immune response pathways with associated suppression of cell morphogenesis (mildly myxoid), synaptic pathways (hypocellular 1U2), other neuron-related pathways (hypercellular 1U2), or neural crest differentiation (hypercellular 1U2, **Figure 8C**). Hence, tumor cellular composition appeared quite heterogenous on the molecular level with pro-inflammatory - and presumably anti-tumor - environments and neoplastic glioneuronal cell areas that appear to be relatively privileged from immune regulation. There also appeared to be an inverse correlation between activation of neuron-related pathways and vascular genesis-related pathways including the Wikipathways VEGFA-VEGFR2 pathway. For each of tumor 1 and tumor 4, there were dispersed spots that co-clustered (into tumor 1 hypercellular 3 and tumor 4 dispersed) wherein there was significant activation in vascular genesis-related pathways and Wikipathways VEGFA-VEGFR2 pathway with associated activation of inflammatory pathways, PI3K-AKT(-mTOR) pathways, and suppression of neuron-related pathways (**Figure 8C** and not shown). In contrast, tumor 1 hypocellular 3 showed suppression of Wikipathways VEGFA-VEGFR2 pathway in the context of significant activation of multiple neuron-related pathways (**Figure 8C** and not shown). Other cluster-specific differences also emerged. For example, tumor 1 (prominently) myxoid 1 activation of Wikipathways glioblastoma signaling and Wikipathways WNT signaling pathway and pluripotency; tumor 1 (prominently) myxoid 2 activation of GO zinc homeostasis and suppression of Wikipathways oxidative phosphorylation; tumor 1 hypocellular 1U2 suppression of various (m)RNA metabolic pathways (**Figure 8C** and not shown).

Given the multiple cells per spot, stRNA-seq spot cell composition was determined by decomposition/deconvolution by cell2location[74] and UCell signature scoring[75] using ganglioglioma snRNA-seq and CITE-seq data as parallel references. Cell composition appeared to vary considerably, yet appropriately, within slides and also between samples with spots estimated to be composed of typically 20-80% neoplastic cells (**Figure 8D-E**). Substantial proportions of each of microglia and macroglia were found as were smaller proportions of neurons, lymphocytes, and vascular cells, as expected. Cell2location tissue regions derived from cell abundance kNN generally correlated with histopathological annotations and up front kNN clustering (**sFigure 9H-I**), further supporting the model-calculated cell abundances. For slides from two of the tumors, methods concurred on the presence of clusters of significant areas of neuroectoderm neural precursor-like cell signature (**Figure 8F-G** and **sFigure 9J-K**).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In order to better understand important intercellular interactions, particularly with these precursor-like cells, we tested for co-localization of different cell types to any given spot. When comparing all four tumor samples, themes emerged. Interestingly, neuroectoderm neural precursor cell-like cell abundance was highly inversely correlated with the presence of vascular cells, microglia, or lymphocytes (**Figure 8H-J**). This suggests these cells occupy a hypovascular, perhaps relatively hypoxic environment. Also, they appear to be relatively spared of immune targeting. In contrast, neuroectoderm neural precursor cell-like cell abundance appeared strongly correlated with cell abundance of other neoplastic clusters (not shown). To further investigate co-localization, cell2location-calculated cell abundance was subjected to non-negative matrix factorization (NMF). Interestingly, tumor 4 neuroectoderm neural precursor cell-like cells tended to occupy discrete locations, with these cells separating out at a low number of factors compared to the number of reference clusters (not shown). In contrast, at least a subset of tumor 1 neuroectoderm neural precursor cell-like cells appeared to co-localize with particular other neoplastic cell types (**sFigure 9L-N**). These observations appear consistent with these neuroectoderm neural precursor cell-like tumor cells residing deeply within neoplastic niches.

We next sought to determine the signaling and transcriptional program context of neuroectodermal neural precursor cell-like tumor cell niches. Interestingly, there was visually apparent co-localization of tumor neuroectoderm-like cells with high *PTPRZ1*, *PTN*, and *FGFR3* (**Figure 8K**), identified as important for neuroectoderm-like neoplastic cell signaling as above. Cell type-specific expression was also estimated by cell2location. This identified tumor neuroectoderm-like cell-specific expression of important neuroectoderm neural precursor like-cell markers including *CD34*, *PAX6*, *SOX2*, *MSI1*, *MEIS1*, and *TCF7L2*, further validating the nature of the snRNA-seq cluster 6 neuroectodermal cells and the identification of these cells *in situ* despite their being a small minority of tissue cells analyzed spatially (**Figure 8K**). Interestingly, there was significant neuroectoderm cell-specific *PTN-PTPRZ1*, *FGF1/2-FGFR3*, and *PDGFA/D-PDGFRB* expression (**Figure 8K**). This suggests a substantial autocrine mechanism for maintenance of the ganglioglioma stem/progenitor cell compartment.

To further explore the spatial context of ganglioglioma cells, top spatially variable genes, agnostic to cell type assignment, were uncovered using the mark-variogram method within Seurat[28]. Interesting patterns emerged. For example, for tumor 1, metallothioneins dominated top spatially-variable features, including *M1E* (#4), *MT1G* (#6), *MT3* (#7), *MT1X* (#10), *MT1M* (#13), and *MT2A* (#24) among the top 25 (**Figure 8L** and not shown). Interestingly, their expression appeared correlated with neuroectoderm neural precursor cell-like cell abundance, with high *MT1E*, *MT1G*, *MT1X*, and *MT1M* each exclusive to areas of neuroectoderm-like cells (**Figure 8L** and not shown). In contrast, *CD74* (antigen-presenting cell marker) and *SPP1* (oligodendrocyte and microglia marker[29,30]) were top-25 spatially-variable transcripts that were largely excluded from neuroectoderm neural precursor cell-like cell areas (**Figure 8L**), once again consistent with the findings above suggesting neuroectoderm neural precursor cell-like cells to occupy niches with relatively little immune cell occupancy. *FTL* and *TPT1* had somewhat different distributions but were similarly inversely associated with neuroectoderm neural precursor cell-like cell abundance (**Figure 8L**).

Neuroectoderm neural precursor cell-like programs associate with GG bulk RNA-seq profiles

We next turned to translation of these findings to clinically-annotated bulk pediatric low grade glioma data sets (n=151 patients for Bergthold *et al*. of whom n=122 had clinical

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

outcomes annotations and n=81 patients for the US National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) with the diagnosis of GG or PLGG)[76,77].

To determine the degree of representation of the single nucleus/CITE-seq ganglioglioma cluster cells in the bulk data, deconvolution of the bulk data was performed using deconvolution/decomposition algorithms[78] and UCell signature scoring[75]. Despite differences in the nature of the bulk datasets (e.g. 6k gene array vs. bulk RNA-seq), the methods tested, and the cells present in reference snRNA-seq and CITE-seq datasets, common themes emerged. Bulk-sequenced tumors were identified as composed of ~40-70% tumor cells with microglial, macroglial, and vascular minority populations identified (**Figure 9A-B**).

In comparing cell composition by diagnosis, other low grade gliomas - pilocytic astrocytoma in particular - actually appeared to have greater components of the OPC-like cell states (**Figure 9C-D**). Conversely, gangliogliomas appeared to have a greater neuron component than other low grade gliomas and pilocytic astrocytoma in particular (**Figure 9C-D**). SnRNA-seq neuroectoderm neural precursor cell-like abundance appeared lowest among pilocytic astrocytomas and at least trended somewhat higher for other histologies, including ganglioglioma (where it was still rare, with median 2% of GG cells from the Bergthold *et al*. dataset and 1% of GG cells from the CPTAC dataset[76,77], similar to expected, **Figure 9C-D**).

Diagnostic and prognostic genetic signatures

We next used the insights from the above deep transcriptomic analyses to nominate prognostic features for low grade glioma patients. For Bergthold *et al*. patients, clinical annotations included diagnosis, extent of resection, self-organizing matrix cluster, age, receipt of chemotherapy, BRAF status, event free survival (EFS), and death[76]. Resection status was found to be important with gross total resection associated with significantly better EFS whereas EFS was numerically similar to one another after biopsy, subtotal resection, and near total resection (**Figure 10A-B** and **sFigure 10A-B**). For CPTAC, clinical annotations included diagnosis (we excluded diagnoses other than ganglioglioma and other low grade gliomas), extent of resection, BRAF status, tumor grade, EFS, and death[77]. Resection status was once again strongly associated with EFS (**sFigure 10C-H**).

We next turned our attention to prognostic genetic markers. *CDKN2A* (and adjacent *CDKN2B*) loss has been previously identified in a small minority of gangliogliomas as well as in other pediatric CNS tumors, such as pleomorphic xanthroastrocytoma (where *CDKN2A/B* loss is typical)[2–4]. The prognostic significance of *CDKN2A/B* loss in ganglioglioma is unclear though there may be some association with more adverse histopathological features and poorer prognosis[2–4]. In contrast, we found *CDKN2A* expression was among the top event-associated transcripts in both data sets (Bergthold *et al*. #16/6100 and CPTAC #23/14827 by average log2-fold-change). When limiting the diagnosis to ganglioglioma, *CDKN2A* was not prognostic (Bergthold *et al*. rank #3201 and CPTAC rank #2840). Instead the association of *CDKN2A* expression with EFS events appeared to be associated with other diagnoses, mostly pilocytic astrocytomas.

The remaining top low grade glioma event-associated transcripts in common between Bergthold *et al*. and CPTAC were overwhelmingly associated with inflammation and/or the extracellular matrix. Particularly interesting was a possible clinical event-associated enrichment for components of the cytosolic DNA sensing pathways (represented in KEGG and Wikipathways, **sFigure 10I-J**), including the cGAS/STING pathway, which has apparent and/or potential roles in tumorigenesis, prognosis, and therapeutic approach[79].

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To interrogate for prognostic multi-gene-signatures, we started with gene signatures of curated stemness-associated factors from the snRNA-seq data and unsupervised machine learning. From this approach, multiple 3-5 gene, curated signatures were of borderline significance, though none were robustly significant by both Kaplan-Meier and Cox proportional hazards modeling using both gene signature z- and Mann Whitney U-test-scoring (arbitrarily using score quarterization for the Kaplan-Meier method). For example, the four-gene signature *CD34*, *SOX2*, *CD99*, and *CTSC* scores appeared at least trending towards association with worse EFS in this regard (**Figure 10C-D** and **sFigure 10K-L**).

To develop more sophisticated prognostic genetic signatures and models, we employed supervised machine learning. A model was trained on Bergthold *et al*. (in mlr3/mlr3proba: top 2.5% clinical or genetic features selected by surv.rfsrc (fast unified random forests for survival, regression, and classification) embedded method, surv.aorsf (accelerated oblique random survival forests) learner selected (**sFigure 10M**) with subsequent sequential feature reduction (**sFigure 10N**) and learner hyperparameterization yielding tuned model with Harrell's concordance-index of 0.87) and validated on CPTAC, yielding a 14-feature signature (resection status combined with expression of 13 transcripts: *UBE2M*, *ELAVL1*, *FRY*, *IGFBP4*, *STX4*, *PSMC6*, *MTX2*, *NFE2L3*, *C1QB*, *ELANE*, *SMARCD2*, *ZNF117*, and *ABL1*). This model predicted CPTAC outcomes with Gonen and Heller's concordance-index 0.89. Hence, we were able to nominate two promising multi-gene prognostic signatures: a curated 4-gene signature and a 13-gene signature (in combination with resection status) from supervised machine learning.

Discussion

Deep transcriptomics has recently led to advances in our understanding of many cancers, including brain tumors[10,35,80]. However, such techniques have not yet been published for GNTs, and application of spatial transcriptomics approaches to brain tumors is in its infancy. Here, we applied deep transcriptomic approaches including spatial transcriptomics to gain insight into important questions regarding GNT biology. Our data supports our hypothesis that GNT CD34+ cells represent neuroectoderm neural precursor cell-like tumor precursor cells. Moreover, we found evidence of perturbations to mTOR/AKT signaling via PDGF/PDGFR and/or FGF/FGFR agonism in combination with PTN antagonism of the otherwise antagonistic and highly expressed PTPRZ1. The combination of this AKT perturbation with MAPK activation (which was present via BRAF V600E in all tumors studied presently) has been found to be sufficient for glioneuronal tumorigenesis in animal models[61]. Our data suggests such a mechanism for gangliogliomagenesis with CD34+ GNT cells as the tumor stem cells. Additionally, we identified important transcriptional regulators for primitive tumor cells including TCF7L2/MEIS1-PAX6 and SOX transcriptional cascades. Interestingly, though the neuroectoderm neural precursor-like cells appeared to occupy a hypovascular niche, they (at least among ganglioglioma neoplastic cells) appeared to be especially high users of oxidative phosphorylation pathways. It is possible this is related to an adaptation, e.g. via upregulation of mitochondrial activity given that the master regulator of mitochondrial biogenesis PPARGC1A was nominated as a neuroectoderm neural precursor-like cell associated transcription factor in an unbiased fashion by SCENIC.

This study reiterated neoplastic and stromal cell states common to myriad brain tumors. These include neoplastic OPC-like cells identified in primary brain tumors as diverse as (adult) glioblastoma[12], diffuse midline glioma[80], pilocytic astrocytoma[10], and now in glioneuronal tumors. Interestingly, these cells have in the past been hypothesized to be tumor

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

precursor cells for other primary brain tumor types. However, subsequent study has not always been consistent with this hypothesis[81]. Within our study, we saw this effect. By the CITE-seq data alone with 1.8k neoplastic cells, the most primitive neoplastic cluster was an OPC-like cluster. However, with increase in cell sampling by an order of magnitude (i.e. to what was obtained for snRNA-seq), we were able to identify a rarer, much more primitive neuroectoderm neural precursor-like population that also fits with the hypothesized tumor stem cell population. This data points to the limitations of undersampling in this context or - conversely - the significant insights possible with sufficient sampling. Given the alignment of our results with *a priori* data and hypotheses, our sampling may very well be adequate to confidently identify the most important characteristics of ganglioglioma stem cells.

Our major results appear generalizable. For instance, though CD34+ neuroectoderm neural precursor-like cells were only found from tumor 5 in the snRNA-seq data, we were able to uncover nests of these cells in tumors 1 and 4 by stRNA-seq and transcriptionally-similar cell states in the other tumors. Hence, we are inclined to understand sampling bias to be the primary factor affecting whether we observe this population in any given tumor.

In addition to neuroectoderm neural precursor-like and OPC-like neoplastic cells, we were able to identify myriad stromal cell types in the context of ganglioglioma, including oligodendrocytes, OPCs, astrocytes, various inhibitory and excitatory neurons, endothelial cells, VLMCs, microglia, and lymphocytes and their associated cell states including significant interactions with the neoplastic cells. We characterized the immune cells extensively. We weighed in on the controversy regarding tumor-associated myeloid cell M1-M2 polarization[45,70]. Our data supports the model of aberrant co-activation of these programs rather than polarization between them. Furthermore, we found evidence that this aberrancy may be directly related to T lymphocyte dysfunction and T lymphocyte stimulation of M1 and M2 characteristics simultaneously. Our data also appears to further support recent observations that tumor-associated immune cells appear to fall along a spectrum rather than as discrete states[45]. Additionally, we found that even low grade tumors with tumor-infiltrating lymphocytes may include potentially suppressible regulatory components (e.g. T regs) and potentially anti-tumor cytotoxic T lymphocytes. Altogether, this data suggests a possible role for immune therapies such as immune check point inhibitors. However, we did uncover CTL exhaustion and dysfunction as potential limiting factors to be overcome.

Another theme of general interest that we uncovered was neoplastic PTPRZ1 antagonism. We found this among diverse primary brain tumors tested including glioblastoma, pilocytic astrocytoma, and ganglioglioma. Interestingly, PTPRZ1 appeared in our SCENIC SOX regulons, and SOX transcription factors are often hyperactive in primary brain tumors[82]. PTPRZ1 overexpression on its own would be expected to abrogate AKT/mTOR signaling and cancer progression[60]. These tumors appear to evolve to overcome this antagonism by overexpression of PTN. This suggests antagonism of PTPRZ1 antagonism by PTN as a possible therapeutic approach for many primary brain tumor types. FGF/FGFR and/or PDGF/PDGFR antagonism may also be important in the context of ganglioglioma, either in combination with PTN-PTPRZ1 antagonism or on their own.

Finally, we translated these findings into existing clinically-annotated bulk transcriptomic data sets to identify different cell state compositions therein as well as expound upon potentially clinically useful genetic markers and multi-gene signatures.

Acknowledgments

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The work was supported by funds from the Botha family and by developmental funds of the Duke Cancer Institute to ZJR as part of the NIH P30CA014236 Cancer Center Support Grant and by Fund to Retain Clinician Scientists funds to ZJR from the Doris Duke Foundation. ZJR is supported by career development funds from a K08CA2560450, the Pediatric Brain Tumor Foundation, St. Baldrick's Foundation, Emily Beazley's Kures for Kids Fund, and ChadTough Defeat DIPG. GYL is supported by the National Cancer Center for Advancing Translational Sciences of the National Institutes of Health under Award Number 1KL2TR002554. The authors would like to thank members of the Duke Brain Tumor Center Biorepository and Database (Diane Satterfield, Merrie Burnett, and Elizabeth Thomas) and the Duke Molecular Physiology Institute (Karen Abramson and Emily Hocke) for assistance with this project.

Competing interests

ZJR receives royalties for patents managed by Duke Office of Licensing and Ventures that have been licensed to Genetron Health, and honoraria for lectures to Eisai Pharmaceuticals and Oakstone Publishing Group. EMT is a scientific advisor for Oncohereos Biosciences.

Data availability

Seurat objects of CITE-seq, snRNA-seq, and stRNA-seq count matrices and associated annotations are uploaded on Zenodo (zenodo.org). Additional original data is available upon request.

Cited Literature

1. Krauze, A. V. Glioneuronal Tumors: Insights into a Rare Tumor Entity. in (ed. Debinski, W.) (2021). doi:10.36255/exonpublications.gliomas.2021.chapter13.
2. Pekmezci, M., Villanueva-Meyer, J. E., Goode, B., Van Ziffle, J., Onodera, C., Grenert, J. P., Bastian, B. C., Chamyan, G., Maher, O. M., Khatib, Z., Kleinschmidt-DeMasters, B. K., Samuel, D., Mueller, S., Banerjee, A., Clarke, J. L., Cooney, T., Torkildson, J., Gupta, N., Theodosopoulos, P., *et al.* The genetic landscape of ganglioglioma. *Acta Neuropathol. Commun.* **6**, 47 (2018).
3. Dyke, J., Calapre, L., Beasley, A., Gray, E., Allcock, R. & Bentel, J. Application of multiplex ligation-dependent probe amplification (MLPA) and low pass whole genome sequencing (LP-WGS) to the classification / characterisation of low grade glioneuronal tumours. *Pathol. Res. Pract.* **229**, 153724 (2022).
4. Ryall, S., Zapotocky, M., Fukuoka, K., Ellison, D. W., Tabori, U. & Hawkins Correspondence, C. Integrated Molecular and Clinical Analysis of 1,000 Pediatric Low-Grade Gliomas. (2020) doi:10.1016/j.ccell.2020.03.011.
5. Kaley, T., Touat, M., Subbiah, V., Hollebecque, A., Rodon, J., Lockhart, A. C., Keedy, V., Bielle, F., Hofheinz, R. D., Joly, F., Blay, J. Y., Chau, I., Puzanov, I., Raje, N. S., Wolf, J., DeAngelis, L. M., Makrutzki, M., Riehl, T., Pitcher, B., *et al.* BRAF inhibition in BRAFV600-mutant gliomas: Results from the VE-BASKET study. *J. Clin. Oncol.* **36**, 3477–3484 (2018).
6. Marks, A. M., Bindra, R. S., DiLuna, M. L., Huttner, A., Jairam, V., Kahle, K. T. & Kieran, M. W. Response to the BRAF/MEK inhibitors dabrafenib/trametinib in an adolescent with a BRAF V600E mutated anaplastic ganglioglioma intolerant to vemurafenib.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pediatr. Blood Cancer **65**, 4–7 (2018).

7. Wolchok, J. D., Kluger, H., Callahan, M. K., Postow, M. A., Rizvi, N. A., Lesokhin, A. M., Segal, N. H., Ariyan, C. E., Gordon, R.-A., Reed, K., Burke, M. M., Caldwell, A., Kronenberg, S. A., Agunwamba, B. U., Zhang, X., Lowy, I., Inzunza, H. D., Feely, W., Horak, C. E., *et al.* Nivolumab plus Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **369**, 122–133 (2013).
8. Antonia, S. J., Villegas, A., Daniel, D., Vicente, D., Murakami, S., Hui, R., Yokoi, T., Chiappori, A., Lee, K. H., de Wit, M., Cho, B. C., Bourhaba, M., Quantin, X., Tokito, T., Mekhail, T., Planchard, D., Kim, Y.-C., Karapetis, C. S., Hiret, S., *et al.* Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **377**, 1919–1929 (2017).
9. Garon, E. B., Rizvi, N. A., Hui, R., Leighl, N., Balmanoukian, A. S., Eder, J. P., Patnaik, A., Aggarwal, C., Gubens, M., Horn, L., Carcereny, E., Ahn, M.-J., Felip, E., Lee, J.-S., Hellmann, M. D., Hamid, O., Goldman, J. W., Soria, J.-C., Dolled-Filhart, M., *et al.* Pembrolizumab for the Treatment of Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).
10. Reitman, Z. J., Paolella, B. R., Bergthold, G., Pelton, K., Becker, S., Jones, R., Sinai, C. E., Malkin, H., Huang, Y., Grimmet, L., Herbert, Z. T., Sun, Y., Weatherbee, J. L., Alberta, J. A., Daley, J. F., Rozenblatt-Rosen, O., Condurat, A. L., Qian, K., Khadka, P., *et al.* Mitogenic and progenitor gene programmes in single pilocytic astrocytoma cells. *Nat. Commun.* **10**, (2019).
11. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).
12. Darmanis, S., Sloan, S. A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., Zhang, Y., Neff, N., Kowarsky, M., Caneda, C., Li, G., Chang, S. D., Connolly, I. D., Li, Y., Barres, B. A., Gephart, M. H. & Quake, S. R. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep.* **21**, 1399–1410 (2017).
13. Bhaduri, A., Di Lullo, E., Jung, D., Müller, S., Crouch, E. E., Espinosa, C. S., Ozawa, T., Alvarado, B., Spatazza, J., Cadwell, C. R., Wilkins, G., Velmeshev, D., Liu, S. J., Malatesta, M., Andrews, M. G., Mostajo-Radji, M. A., Huang, E. J., Nowakowski, T. J., Lim, D. A., *et al.* Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* **26**, 48-63.e6 (2020).
14. Blümcke, I., Giencke, K., Wardelmann, E., Beyenburg, S., Kral, T., Sarioglu, N., Pietsch, T., Wolf, H. K., Schramm, J., Elger, C. E. & Wiestler, O. D. The CD34 epitope is expressed in neoplastic and malformative lesions associated with chronic, focal epilepsies. *Acta Neuropathol.* **97**, 481–490 (1999).
15. Deb, P., Sharma, M. C., Tripathi, M., Sarat Chandra, P., Gupta, A. & Sarkar, C. Expression of CD34 as a novel marker for glioneuronal lesions associated with chronic intractable epilepsy. *Neuropathol. Appl. Neurobiol.* **32**, 461–468 (2006).
16. Fauser, S., Becker, A., Schulze-Bonhage, A., Hildebrandt, M., Tuxhorn, I., Pannek, H. W., Lahl, R., Schramm, J. & Blümcke, I. CD34-immunoreactive balloon cells in cortical malformations. *Acta Neuropathol.* **108**, 272–278 (2004).
17. Bale, T. A. & Rosenblum, M. K. The 2021 WHO Classification of Tumors of the Central Nervous System: An update on pediatric low-grade gliomas and glioneuronal tumors. *Brain Pathol.* 1–10 (2022) doi:10.1111/bpa.13060.
18. Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., Von Deimling, A. & Ellison, D. W. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro. Oncol.* **23**, 1231–1251 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

19. Zhang, Y., Lucas, C. G., Young, J. S., Morshed, R. A., Mccoy, L., Ann, N., Bush, O., Taylor, J. W., Daras, M., Butowski, N. A., Villanueva-meyer, J. E., Cha, S., Wrensch, M., Wiencke, J. K., Lee, J. C., Pekmezci, M., Phillips, J. J., Perry, A., Bollen, A. W., *et al.* glioblastoma leads to improved clinical outcomes. 1–14 (2022).
20. 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., Dewitt, J., Gritsch, S., Perez, E. M., Gonzalez Castro, L. N., Lan, X., Druck, N., Rodman, C., Dionne, D., Kaplan, A., *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).
21. Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., Fisher, J. M., Rodman, C., Mount, C., Filbin, M. G., Neftel, C., Desai, N., Nyman, J., Izar, B., Luo, C. C., Francis, J. M., Patel, A. A., Onozato, M. L., Riggi, N., *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
22. Müller, S., Kohanbash, G., Liu, S. J., Alvarado, B., Carrera, D., Bhaduri, A., Watchmaker, P. B., Yagnik, G., Di Lullo, E., Malatesta, M., Amankulor, N. M., Kriegstein, A. R., Lim, D. A., Aghi, M., Okada, H. & Diaz, A. Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol.* **18**, 1–14 (2017).
23. Pombo Antunes, A. R., Scheyltjens, I., Lodi, F., Messiaen, J., Antoranz, A., Duerinck, J., Kancheva, D., Martens, L., De Vlaminck, K., Van Hove, H., Kjølner Hansen, S. S., Bosisio, F. M., Van der Borght, K., De Vleeschouwer, S., Sciot, R., Bouwens, L., Verfaillie, M., Vandamme, N., Vandenbroucke, R. E., *et al.* Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals macrophage competition and specialization. *Nat. Neurosci.* **24**, 595–610 (2021).
24. Ravi, V. M., Neidert, N., Will, P., Joseph, K., Maier, J. P., Kückelhaus, J., Vollmer, L., Goeldner, J. M., Behringer, S. P., Scherer, F., Boerries, M., Follo, M., Weiss, T., Delev, D., Kernbach, J., Franco, P., Schallner, N., Dierks, C., Carro, M. S., *et al.* T-cell dysfunction in the glioblastoma microenvironment is mediated by myeloid cells releasing interleukin-10. *Nat. Commun.* **13**, (2022).
25. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
26. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
27. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
28. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
29. Human Protein Atlas proteinatlas.org.
30. Ponten, F., Jirstrom, K. & Uhlen, M. The Human Protein Atlas--a tool for pathology. *J. Pathol.* **216**, 387–393 (2008).
31. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardinoglu, A., *et al. A single-cell type transcriptomics map of human tissues*. *Sci. Adv* vol. 7 www.proteinatlas.org (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

32. Dataset: Allen Institute for Brain Science (2021). Allen Cell Types Database -- Human Multiple Cortical Areas [dataset]. Available from celltypes.brain-map.org/rnaseq.
33. HuBMAP Azimuth https://azimuth.hubmapconsortium.org/references/.
34. Consortium, H., Human, N. I. H. & Atlas, B. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
35. 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., Louis, D. N., Rozenblatt-Rosen, O., Suvà, M. L., Regev, A. & Bernstein, B. E. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344file:///**, 1396–1401 (2014).
36. Pekmezci, M., Villanueva-Meyer, J. E., Goode, B., Van Ziffle, J., Onodera, C., Grenert, J. P., Bastian, B. C., Chamyan, G., Maher, O. M., Khatib, Z., Kleinschmidt-DeMasters, B. K., Samuel, D., Mueller, S., Banerjee, A., Clarke, J. L., Cooney, T., Torkildson, J., Gupta, N., Theodosopoulos, P., *et al.* The genetic landscape of ganglioglioma. *Acta Neuropathol. Commun.* **6**, 47 (2018).
37. Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S. J., Velmeshev, D., Ounadjela, J. R., Shuga, J., Wang, X., Lim, D. A., West, J. A., Leyrat, A. A., Kent, W. J. & Kriegstein, A. R. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (80-.).* **358**, 1318–1323 (2017).
38. Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J. & Kriegstein, A. R. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.* **24**, 584–594 (2021).
39. van Bruggen, D., Pohl, F., Langseth, C. M., Kukanja, P., Lee, H., Albiach, A. M., Kabbe, M., Meijer, M., Linnarsson, S., Hilscher, M. M., Nilsson, M., Sundström, E. & Castelo-Branco, G. Developmental landscape of human forebrain at a single-cell level identifies early waves of oligodendrogenesis. *Dev. Cell* **57**, 1421-1436.e5 (2022).
40. Herring, C. A., Simmons, R. K., Freytag, S., Poppe, D., Moffet, J. J. D., Pflueger, J., Buckberry, S., Vargas-Landin, D. B., Clément, O., Echeverría, E. G., Sutton, G. J., Alvarez-Franco, A., Hou, R., Pflueger, C., McDonald, K., Polo, J. M., Forrest, A. R. R., Nowak, A. K., Voineagu, I., *et al.* Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428-4447.e28 (2022).
41. Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., Xu, X., Tang, F., Zhang, J., Qiao, J. & Wang, X. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
42. Bakken, T. E., Jorstad, N. L., Hu, Q., Lake, B. B., Tian, W., Kalmbach, B. E., Crow, M., Hodge, R. D., Krienen, F. M., Sorensen, S. A., Eggermont, J., Yao, Z., Aevermann, B. D., Aldridge, A. I., Bartlett, A., Bertagnolli, D., Casper, T., Castanon, R. G., Crichton, K., *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
43. Nagy, C., Maitra, M., Tanti, A., Suderman, M., Théroux, J. F., Davoli, M. A., Perlman, K., Yerko, V., Wang, Y. C., Tripathy, S. J., Pavlidis, P., Mechawar, N., Ragoussis, J. & Turecki, G. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat. Neurosci.* **23**, 771–781 (2020).
44. Wang, L., Babikir, H., Müller, S., Yagnik, G., Shamardani, K., Catalan, F., Kohanbash, G., Alvarado, B., Di Lullo, E., Kriegstein, A., Shah, S., Wadhwa, H., Chang, S. M., Phillips, J. J., Aghi, M. K. & Diaz, A. A. The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. *Cancer Discov.* **9**, 1708–1719 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45. Azizi, E., Carr, A. J., Plitas, G., Cornish, A. E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V., Setty, M., Choi, K., Fromme, R. M., Dao, P., McKenney, P. T., Wasti, R. C., Kadaveru, K., Mazutis, L., Rudensky, A. Y. & Pe'er, D. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* **174**, 1293-1308.e36 (2018).
46. Nieto, P., Elosua-Bayes, M., Trincado, J. L., Marchese, D., Massoni-Badosa, R., Salvany, M., Henriques, A., Nieto, J., Aguilar-Fernández, S., Mereu, E., Moutinho, C., Ruiz, S., Lorden, P., Chin, V. T., Kaczorowski, D., Chan, C. L., Gallagher, R., Chou, A., Planas-Rigol, E., *et al.* A single-cell tumor immune atlas for precision oncology. *Genome Res.* **31**, 1913–1926 (2021).
47. Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., Qin, S., Zhang, L., Ouyang, H., Du, P., Jiang, L., Zhang, B., Yang, Y., Wang, X., Ren, X., Bei, J. X., Hu, X., Bu, Z., Ji, J., *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792-809.e23 (2021).
48. Zheng, L., Qin, S., Si, W., Wang, A., Xing, B., Gao, R., Ren, X., Wang, L., Wu, X., Zhang, J., Wu, N., Zhang, N., Zheng, H., Ouyang, H., Chen, K., Bu, Z., Hu, X., Ji, J. & Zhang, Z. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science (80-.).* **374**, (2021).
49. Piper, K., DePledge, L., Karsy, M. & Cobbs, C. Glioma Stem Cells as Immunotherapeutic Targets: Advancements and Challenges. *Front. Oncol.* **11**, 1–13 (2021).
50. Vallier, L., Reynolds, D. & Pedersen, R. A. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275**, 403–421 (2004).
51. Zhang, X., Huang, C. T., Chen, J., Pankratz, M. T., Xi, J., Li, J., Yang, Y., LaVaute, T. M., Li, X. J., Ayala, M., Bondarenko, G. I., Du, Z. W., Jin, Y., Golos, T. G. & Zhang, S. C. Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90–100 (2010).
52. Kawase, S., Kuwako, K., Imai, T., Renault-Mihara, F., Yaguchi, K., Itohara, S. & Okano, H. Regulatory factor X transcription factors control musashi1 transcription in mouse neural stem/progenitor cells. *Stem Cells Dev.* **23**, 2250–2261 (2014).
53. Stevanovic, M., Drakulic, D., Lazic, A., Ninkovic, D. S., Schwirtlich, M. & Mojsin, M. SOX Transcription Factors as Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and Adult Neurogenesis. *Front. Mol. Neurosci.* **14**, 1–24 (2021).
54. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H. B., Pe'er, D. & Theis, F. J. *CellRank for directed single-cell fate mapping*. *Nature Methods* (2022). doi:10.1038/s41592-021-01346-6.
55. Gulati, G. S., Sikandar, S. S., Wesche, D. J., Manjunath, A., Bharadwaj, A., Berger, M. J., Ilagan, F., Kuo, A. H., Hsieh, R. W., Cai, S., Zabala, M., Scheeren, F. A., Lobo, N. A., Qian, D., Yu, F. B., Dirbas, F. M., Clarke, M. F. & Newman, A. M. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science (80-.).* **367**, 405–411 (2020).
56. Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat. Commun.* **8**, 1–15 (2017).
57. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
58. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C. & Shendure, J. The single-

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

59. Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., Myung, P., Plikus, M. V. & Nie, Q. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1–20 (2021).
60. Xia, Z., Ouyang, D., Li, Q., Li, M., Zou, Q., Li, L., Yi, W. & Zhou, E. The expression, functions, interactions and prognostic values of PTPRZ1: A review and bioinformatic analysis. *J. Cancer* **10**, 1663–1674 (2019).
61. Cases-Cunillera, S., van Loo, K. M. J., Pitsch, J., Quatraccioni, A., Sivalingam, S., Salomoni, P., Borger, V., Dietrich, D., Schoch, S. & Becker, A. J. Heterogeneity and excitability of BRAF V600E -induced tumors is determined by Akt/mTOR-signaling state and Trp53 -loss . *Neuro. Oncol.* 1–14 (2021) doi:10.1093/neuonc/noab268.
62. Puget, S., Philippe, C., Bax, D. A., Job, B., Varlet, P., Andreiuolo, F., Carvalho, D., Reis, R., Guerrini-rousseau, L., Roujeau, T., Dessen, P., Richon, C., Lazar, V., Teuff, G. Le & Sainte-rose, C. Mesenchymal Transition and PDGFRA Amplification / Mutation Are Key Distinct Oncogenic Events in Pediatric Diffuse Intrinsic Pontine Gliomas. **7**, (2012).
63. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X. & Yu, G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov.* **2**, 100141 (2021).
64. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
65. Warburg, O. The metabolism of carcinoma cells. *J. Cancer Res.* **9**, 148-163. (1925).
66. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., Verbeiren, T., De Maeyer, D., Reumers, J., Saeys, Y. & Aerts, S. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
67. Owa, T., Taya, S., Miyashita, S., Yamashita, M., Adachi, T., Yamada, K., Yokoyama, M., Aida, S., Nishioka, T., Inoue, Y. U., Goitsuka, R., Nakamura, T., Inoue, T., Kaibuchi, K. & Hoshino, M. Meis1 coordinates cerebellar granule cell development by regulating pax6 transcription, BMP signaling and atoh1 degradation. *J. Neurosci.* **38**, 1277–1294 (2018).
68. Pavlaki, I., Shapiro, M., Pisignano, G., Jones, S. M. E., Telenius, J., Muñoz-Descalzo, S., Williams, R. J., Hughes, J. R. & Vance, K. W. Chromatin interaction maps identify Wnt responsive cis-regulatory elements coordinating Paupar-Pax6 expression in neuronal cells. *PLOS Genet.* **18**, e1010230 (2022).
69. Albert, G. & Besson, C. Microglia and Neuroinflammation□: What Place for P2RY12□? (2021).
70. Li, X., Zhang, Q., Chen, G. & Luo, D. Multi-Omics Analysis Showed the Clinical Value of Gene Signatures of C1QC+ and SPP1+ TAMs in Cervical Cancer. *Front. Immunol.* **12**, 1–12 (2021).
71. Zhang, L., Li, Z., Skrzypczynska, K. M., Fang, Q., Zhang, W., O'Brien, S. A., He, Y., Wang, L., Zhang, Q., Kim, A., Gao, R., Orf, J., Wang, T., Sawant, D., Kang, J., Bhatt, D., Lu, D., Li, C. M., Rapaport, A. S., *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **181**, 442-459.e29 (2020).
72. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

73. Paijens, S. T., Vledder, A., de Bruyn, M. & Nijman, H. W. Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell. Mol. Immunol.* **18**, 842–859 (2021).
74. Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H. W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., Jain, M. S., Park, J. S., Ramona, L., Tuck, E., Arutyunyan, A., Vento-Tormo, R., Gerstung, M., James, L., Stegle, O., *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
75. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
76. Bergthold, G., Bandopadhayay, P., Hoshida, Y., Ramkissoon, S., Ramkissoon, L., Rich, B., Maire, C. L., Paolella, B. R., Schumacher, S. E., Tabak, B., Ferrer-Luna, R., Ozek, M., Sav, A., Santagata, S., Wen, P. Y., Goumnerova, L. C., Ligon, A. H., Stiles, C., Segal, R., *et al.* Expression profiles of 151 pediatric low-grade gliomas reveal molecular differences associated with location and histological subtype. *Neuro. Oncol.* **17**, 1486–1496 (2015).
77. Petralia, F., Tignor, N., Reva, B., Koptyra, M., Chowdhury, S., Rykunov, D., Krek, A., Ma, W., Zhu, Y., Ji, J., Calinawan, A., Whiteaker, J. R., Colaprico, A., Stathias, V., Omelchenko, T., Song, X., Raman, P., Guo, Y., Brown, M. A., *et al.* Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. *Cell* **183**, 1962-1985.e31 (2020).
78. Pfister S, Kuettel V, Ferrero E (2022). granulator: Rapid benchmarking of methods for *in silico* deconvolution of bulk RNA-seq data. R package version 1.6.0, H. com/xanibas/granulator. No Title.
79. Härtlova, A., Erttmann, S. F., Raffi, F. A. M., Schmalz, A. M., Resch, U., Anugula, S., Lienenklaus, S., Nilsson, L. M., Kröger, A., Nilsson, J. A., Ek, T., Weiss, S. & Gekara, N. O. DNA Damage Primes the Type I Interferon System via the Cytosolic DNA Sensor STING to Promote Anti-Microbial Innate Immunity. *Immunity* **42**, 332–343 (2015).
80. Filbin, M. G., Tirosh, I., Hovestadt, V., Shaw, M. L., Escalante, L. E., Mathewson, N. D., Neftel, C., Frank, N., Pelton, K., Hebert, C. M., Haberler, C., Yizhak, K., Gojo, J., Egervari, K., Mount, C., van Galen, P., Bonal, D. M., Nguyen, Q.-D., Beck, A., *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331–335 (2018).
81. Tomita, Y., Shimazu, Y., Somasundaram, A., Tanaka, Y., Takata, N., Ishi, Y., Gadd, S., Hashizume, R., Angione, A., Pinero, G., Hambardzumyan, D., Brat, D. J., Hoeman, C. M. & Becher, O. J. A novel mouse model of diffuse midline glioma initiated in neonatal oligodendrocyte progenitor cells highlights cell-of-origin dependent effects of H3K27M. *Glia* **70**, 1681–1698 (2022).
82. Stevanovic, M., Kovacevic-Grujicic, N., Mojsin, M., Milivojevic, M., Drakulic, D., Stevanovic, M. & Stevanovic, M. SOX transcription factors and glioma stem cells: Choosing between stemness and differentiation. *World J. Stem Cells* **13**, 1417–1445 (2021).
83. Waldron, L., Oh, S., Abdelnabi, J., Al-Dulaimi, R., Aggarwal, A., Ramos, M., Davis, S. & Riester, M. HGNChelper: Identification and correction of invalid gene symbols for human and mouse. *F1000Research* **9**, 1–16 (2020).
84. Gao, R., Bai, S., Henderson, Y. C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A., Wang, F., Shaitelman, S. F., Wang, J. R., Chen, K., Moulder, S., Lai, S. Y. & Navin, N. E. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608 (2021).
85. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V.,

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., van Bruggen, D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

86. Luo, W. & Brouwer, C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
87. Patil, I. Visualizations with statistical details: The 'ggstatsplot' approach. *J. Open Source Softw.* **6**, 3167 (2021).
88. Terry M. Therneau & Patricia M. Grambsch. *Modeling Survival Data: Extending the {C}ox Model*. (Springer, 2000).
89. Alboukadel Kassambara, Marcin Kosinski, P. B. survminer: Survival Analysis and Visualization.
90. Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L. & Bischl, B. {mlr3}: A modern object-oriented machine learning framework in {R}. *J. Open Source Softw.* (2019) doi:10.21105/joss.01903.
91. Sonabend, R., Király, F. J., Bender, A., Bischl, B. & Lang, M. mlr3proba: An R Package for Machine Learning in Survival Analysis. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab039.

Methods

Reference data sets

Reference single-cell data was obtained via UCSC cell browser (Nowakowski *et al*. developing brain[37], Eze *et al*. developing brain[38], Darmanis *et al*. glioblastoma[12], Wang/Muller *et al*. glioblastoma[44], Allen human cortex[32], Allen human M1 cortex[32]), Broad Institute single cell portal (Tirosh *et al*. oligodendroglioma[21], Filbin *et al*. diffuse midline glioma[80], Reitman *et al*. pilocytic astrocytoma[10]), and GEO GSE104276 (Zhong *et al*. developing prefrontal cortex[41]), GSE168408 (Herring *et al*. developing prefrontal cortex[40]), GSE144136 (Nagy *et al*. adult prefrontal cortex[43]), GSE156728 (Zheng *et al*. T cells[48]), GSE154763 (Cheng *et al*. myeloid cells[47]), and GSE114724 (Azizi *et al*. immune cells[45]). Bulk transcriptomic data and associated annotations were obtained from CPTAC[77] via pedcbioportal.kidsfirstdrc.org and, for Bergthold *et al*., from GEO GSE60898[76].

Patients

Ganglioglioma tumor samples for deep transcriptomics were obtained per Duke University Health System Institutional Review Board (IRB) protocol Pro00072150. The methods were carried out in accordance with the approved guidelines, with written informed consent obtained from all subjects or their guardians where appropriate. Patient characteristics are summarized in **supplementary table 1**.

SnRNA-seq

Frozen banked tumor tissue was obtained from the Duke Brain Tumor Center Biorepository and Database. For quality assurance, RNA integrity number was checked with goal RIN>7. Nuclei were isolated. Briefly, 50 mg tissue was minced to ~0.5 mm cubes, transferred to lysis

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

solution (10 mM Tris-HCl pH 7.4 (Sigma), 5 mM NaCl (Sigma), 3 mM $MgCl_2$ (VWR), 0.1% NP-40 substitute (Sigma), and 0.5% RNasin Plus (Promega, aq)), incubated on ice 5 minutes, and incubated with tituration 10-15 times every 30 seconds for 10 minutes on ice. Residual debris was removed by 70 mcm filter (VWR). Nuclei were centrifuged 300g x5 minutes at 4°C, rinsed x2 with resuspension buffer (1% BSA (ThermoFisher), 0.5% RNasin Plus, and 1X PBS pH 7.4 (Corning, aq)), resuspended in OptiPrep solution (25 mM KCl (ThermoFisher), 5 mM $MgCl_2$ (ThermoFisher), 20mM Tris-HCl pH 7.8 (ThermoFisher), 50% OptiPrep Density Gradient Medium (Sigma), and 100 mM sucrose (Sigma, aq)), pelleted 10000g x10 minutes at 4°C, resuspended in resuspension buffer, and assessed for intact nuclei. Otherwise, snRNA-seq was performed using 3′ v3 Single Cell technology according to the manufacturer's protocol (10X Genomics, San Diego, CA).

CITE-seq

Fresh tumor tissue was obtained for CITE-seq. Tumor cells were dissociated, washed, and resuspended. Antibodies for CITE-seq were anti-CD34 (clone 581), anti-CD31 (clone WM59), anti-CD45RA (clone HI100), anti-CD3 (clone UCHT1), anti-CD8A (clone RPAT8), anti-CD4 (clone RPAT4), anti-CD14 (clone 63D3), and anti-CD19 (clone HIB19) TotalSeq-A antibodies (BioLegend, San Diego, CA). Cells from one tumor were used to optimize preparation by adding variable antibody concentrations to 1 million cells in 50 mcl, washing with 200 mcl FACS buffer, resuspending in 200 mcl FACS buffer, and analyzing by flow cytometry. Otherwise, cells were used for CITE-seq using 3′ v3 Single Cell Immune Profiling technology according to the manufacturer's protocol (10X Genomics, San Diego, CA).

Spatial transcriptomics

Frozen banked tumor tissue was obtained from the Duke Brain Tumor Center Biorepository and Database. StRNA-seq was performed using Spatial 3' v1 technology according to the manufacturer's protocol (10X Genomics, San Diego, CA).

SnRNA-seq and CITE-seq data preprocessing and initial processing

Raw sequencing data was processed into unique molecular identifier count (UMI) matrices using CellRanger (v6.1.2 for snRNA-seq and v3.1.0 for CITE-seq, human genome build GRCh38, cellranger mkfastq -> cellranger count, 10X Genomics). Processing beyond this point was carried out in Ubuntu 20.04 using R 4.2.1 or python version 3.x (depending upon python packages in use). Within Seurat v4[28], snRNA-seq log-normalization, scaling, SCTransformation, PCA, and UMAP (based on PCA dims 1-50) was performed. For the projection shown, SCTransform was run based on all cells, 10k variable features, regressing out based on percent mitochondrial RNA, and otherwise defaults. For CITE-seq, gene symbols were updated with HGNChelper v0.8.1[83], and this was followed by log-normalization, scaling, PCA, and UMAP (based on PCA dims 1-50). Neighbors and clusters were also found within Seurat. Gene symbols were updated by HGNChelper v0.8.1 for analyses sensitive to antiquated gene symbols.

StRNA-seq data preprocessing and initial processing

Raw sequencing data was processed into unique molecular identifier count (UMI) matrices using SpaceRanger v1.3.0 (human genome build GRCh38, spaceranger mkfastq ->

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spaceranger count, 10X Genomics). The graph-based clusters shown were also calculated with spaceranger count. Slides were manually annotated (GYL). Empty spots were excluded as were those with excessive artifact (folded tissue and excessive blood product in particular). Graph-based clusters were then regrouped based on manual annotation for final clusters. Processing beyond this point was carried out in Ubuntu 20.04 using R 4.2.1 or python version 3.x (depending upon python packages in use). Gene symbols were updated by HGNChelper v0.8.1.

Initial cell typing

SnRNA-seq and CITE-seq cells were initially typed based expression of *a priori* markers of interest, favoring unbiased atlases where possible, such as the Human Protein Atlas[29,30] and HuBMAP[33,34]. Markers are outlined in the results. Additionally, cells were typed using Seurat v4-base label transfer to published reference atlases listed under data above[28]. Cells were also typed by UCell v2.0.1 signature scoring[75].

Inference of copy number alterations

Neoplastic-appearing cells from merged snRNA-seq or merged CITE-seq data were analyzed by inferCNV v1.12.0 in parallel[35]. Oligodendroglioma and associated oligodendrocytes were used as positive and negative controls, respectively[21]. Tumor 3 snRNA-seq data was excluded due to low complexity. SnRNA-seq and CITE-seq stromal cells were also used as controls. For run, cutoff=0.1 was used. Data was subjected to apply_median_filtering with window_size=7. Runs were otherwise by defaults. Data was displayed using ComplexHeatmap v2.13.2. Results were verified with CopyKAT v1.1.0[84].

Cellular hierarchy

From CellRanger outputs, counts were obtained by velocyto v0.17.17 run10x and otherwise using the defaults[85]. Data from different tumors was then merged and then subset by cell lineage (e.g. neoplastic, microglia, lymphoid, vascular) based on cell typing for RNA velocity analysis by scVelo v0.2.4[57]. The results shown were obtained using the package defaults, with modes described in the text. SCENT v1.0.3 signaling entropy rate results shown were calculated using the original counts and defaults with the net17Jan16.m used as the protein-protein interaction network adjacency matrix[56]. CellRank v1.5.1 pseudotimes were calculated using the CytoTRACE kernel[54,55] based on the original counts after merging samples and subsetting by lineage (neoplastic, myeloid, and lymphoid). Tumor 3 snRNA-seq data was excluded due to low complexity resulting in aberrantly high calculated pseudotime. Defaults used unless specified otherwise.

Ligand-receptor analysis

Significant snRNA-seq and CITE-seq ligand-receptor interactions were nominated in an unbiased manner among merged snRNA-seq or merged CITE-seq data using CellChat v1.5.0 and CellPhoneDB v3.1.0[59,72]. Defaults were used for the results shown.

Gene set enrichment analysis (GSEA)

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Average log2-fold-change was calculated within Seurat v4 based on the comparisons described in the text[28]. Gene lists were sorted by descending log2-fold-change and clusterProfiler v4.4.4 gseGO run with p-value cutoff=0.05 with Bonferroni correction[63]. GseKEGG and gseWP were performed analogously. Defaults were otherwise used. KEGG pathways of interest were visualized by Pathview v1.36.1[86].

Gene regulatory network inference

Transcription factors were nominated among snRNA-seq and CITE-seq by cluster or by neoplastic *CD34*-status using pySCENIC v0.12.0 defaults[66]. Transcription factors/regulons of interest were further curated by Wilcoxon rank sum test comparison using AUC log2-fold-change≥0.1. For individual clusters this was performed relative to all other clusters. For each neoplastic cluster, it was also performed relative to the union of stromal neural clusters. For neoplastic *CD34*+ cells, it was performed relative to neoplastic *CD34*- cells.

StRNA-seq data deconvolution

StRNA-seq spots were deconvoluted using cell2location v0.1[74]. For deconvolution, mitochondrial, ribosomal, X, and Y transcripts were excluded. Each of snRNA-seq and CITE-seq data sets was used as reference in turn. N cells per spot was 12 and detection_alpha=20. Training max epochs=1000 for references and =30000 for spatial data. Deconvolution, NMF, and cell type-specific gene expression were otherwise calculated using recommended/default settings.

Top spatially variable genes

Top spatially variable genes were calculated for each stRNA-seq sample in Seurat v4 using the mark-variogram method and defaults otherwise[28].

Bulk transcriptomic data deconvolution

Bulk data deconvolution was performed with granulator v1.1.0 using defaults[78]. Dtangle results shown are the results from using defaults within granulator, subsequently scaled to total composition of one for each sample. Statistical analysis comparing subpopulations was performed using the ggstatsplot v0.9.4.900 package and otherwise as described in the text[87].

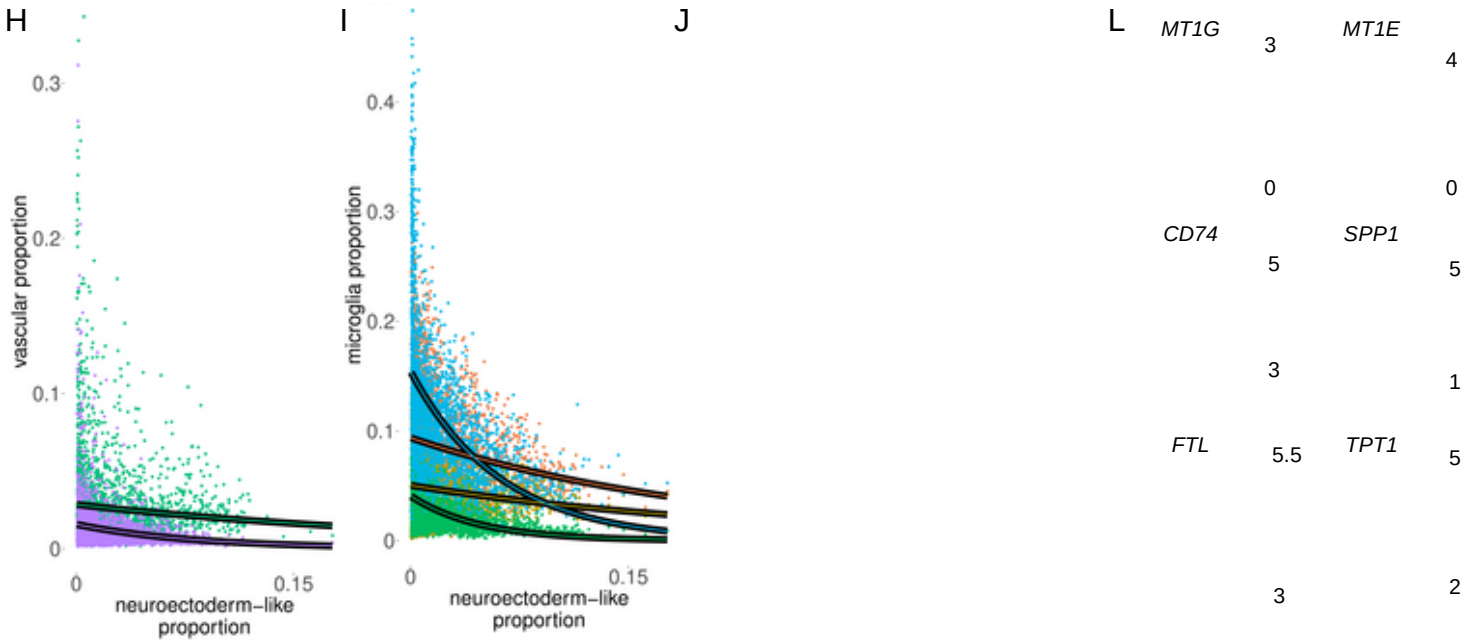
Clinical data analysis

Bergthold *et al*.[76] and CPTAC[77] bulk transcriptomic data was filtered for those with annotation of outcomes (event free survival in particular). Kaplan-Meier curves were calculated survfit from survival package v3.4.0[88] and plotted with log-rank p-values using ggsurvplot from the survminer library v0.4.9[89]. Correlograms of available clinical annotations were created by calculation of bias-corrected Cramer's V for nominal vs. nominal variables, Spearman for numeric vs. numeric variables, and ANOVA for nominal vs. numeric variables. Multi-gene signature z-scores were calculated by multiplying n gene z-scores and taking the nth root. Multi-gene Mann Whitney U-test scores were calculated using UCell[75]. Prognostic models were developed by supervised machine learning in mlr3 using mlr3proba[90,91]. Bergthold *et al*. relative-count expression combined with clinical annotations was subjected to importance scoring using surv.rfsrc learner, the top 2.5% features were selected, and surv.aorsf learner

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

selected with subsequent sequential feature reduction and learner algorithm hyperparameterization yielding tuned model using a 14-feature signature (resection status combined with 13 transcripts: *UBE2M*, *ELAVL1*, *FRY*, *IGFBP4*, *STX4*, *PSMC6*, *MTX2*, *NFE2L3*, *C1QB*, *ELANE*, *SMARCD2*, *ZNF117*, and *ABL1*). Surv.aorsf final parameters were n_tree=10000, n_split=5, n_retry=24, mtry_ratio=0.8111, control_type=cph, and split_min_stat=11.2. The model was validated on CPTAC FPKM expression combined with clinical annotations.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



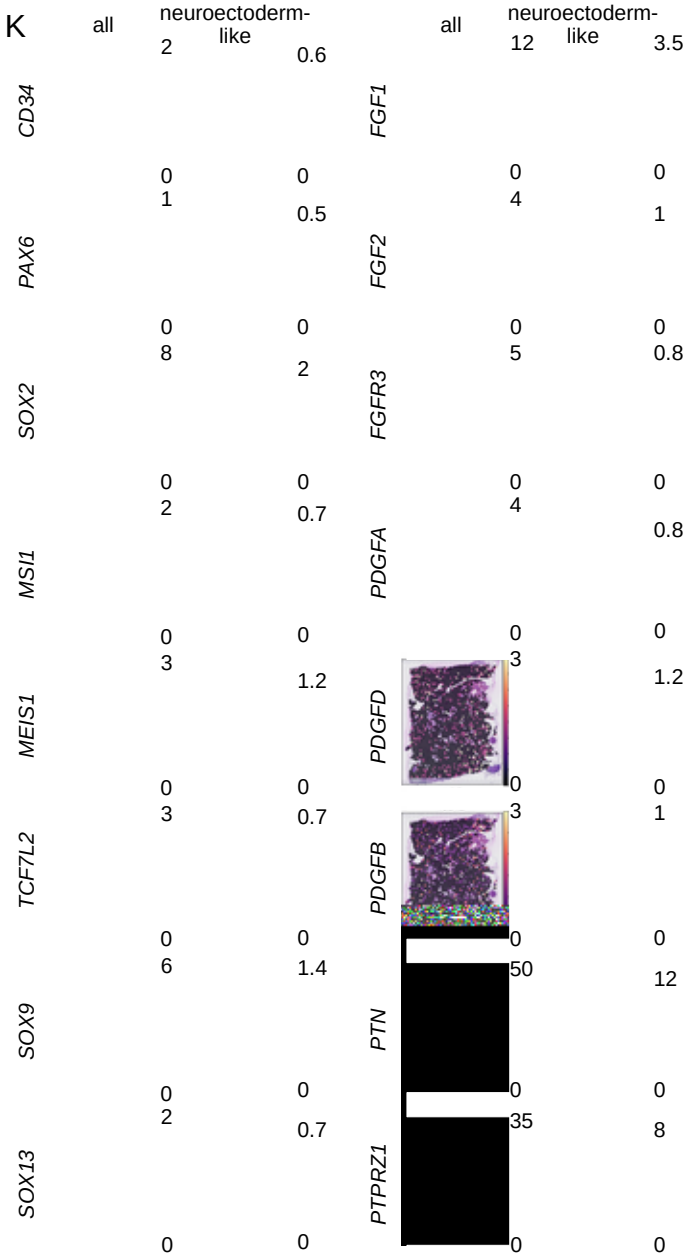


Figure 8: Ganglioglioma spatial transcriptomics shown for two samples from tumor 1 and tumor 4. (A-B) Spatial feature plots of final clusters for tumor 1 (A) and tumor 4 (B). Note 'U' in the final cluster name is used to denote union of multiple Space Ranger-generated graph-based (kNN) clusters. (C) GO term enrichment for select tumor 1 clusters. (D-E) Cell type proportions per spot as determined by cell2location with snRNA-seq as reference for tumor 1 (D) and tumor 4 (E). OPC=oligodendrocyte precursor cell. VLMC=vascular leptomeningeal cell. (F) Tumor 1 cell2location-calculated neuroectoderm neural precursor cell-like cell abundance per spot. (G) Analogous to (F) but for tumor 4. (H-J) Inverse correlation of neuroectoderm neural precursor-like cell proportion and that of vascular (H), microglial (I), and lymphoid (J) cells, shown for tumor 1 spot cell proportions calculated by cell2location with snRNA-seq clusters as reference. Points and curves are colored according to the cell type whose proportion is being compared to the proportion of neuroectoderm neural precursor-like cells. Curves (outlined in black to enhance visibility) represent non-linear (exponential) least-squares fit for each reference cell type. (K) Overall (left) and cell type-specific (right) expression of neuroectoderm neural precursor cell-like cell markers as well as important transcription factors and ligand-receptor pairs based on snRNA/CITE-seq analysis. (L) Feature plots of log1p (base 2) [RNA] for examples of top spatially-variable features from tumor 1, as determined by mark-variogram method in Seurat.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

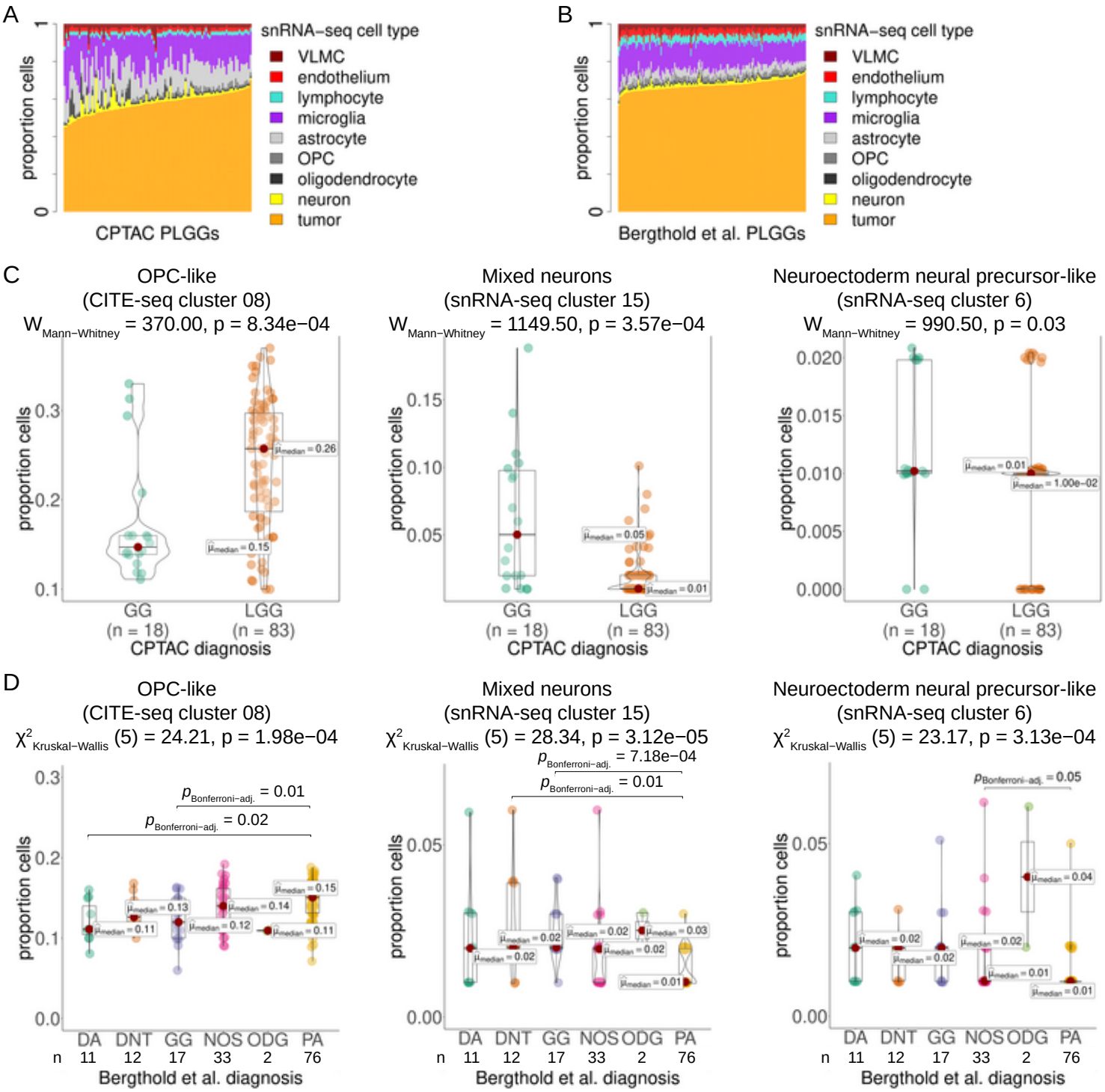


Figure 9: Pediatric low grade glioma bulk transcriptomic data deconvolution in light of snRNA/CITE-seq results.
(A) CPTAC pediatric low grade glioma and ganglioglioma bulk RNA-seq deconvolution example using dtangle method with snRNA-seq data cell type reference.
(B) Bergthold *et al*. low grade glioma bulk RNA-seq deconvolution analogous to (A).
(C) Violin plots with CPTAC tumor OPC-like, mixed neuron, and neuroectoderm-like cell composition, by diagnosis. Medians marked with red dots and labeled. Statistical comparisons by Mann Whitney with p-values shown.
(D) Analogous to (C) for Bergthold *et al*. tumors except for statistical testing, Kruskal−Wallis was used initially, followed by Dunn test with Bonferroni adjustment as needed.
OPC=oligodendrocyte precursor cell, DA=diffuse astrocytoma, GG=ganglioglioma, ODG=oligodendroglioma, DNT=dysplastic neuroepithelial tumor, NOS or LGG=low-grade glioma, not otherwise specified. PA=pilocytic astrocytoma.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.17.520880; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

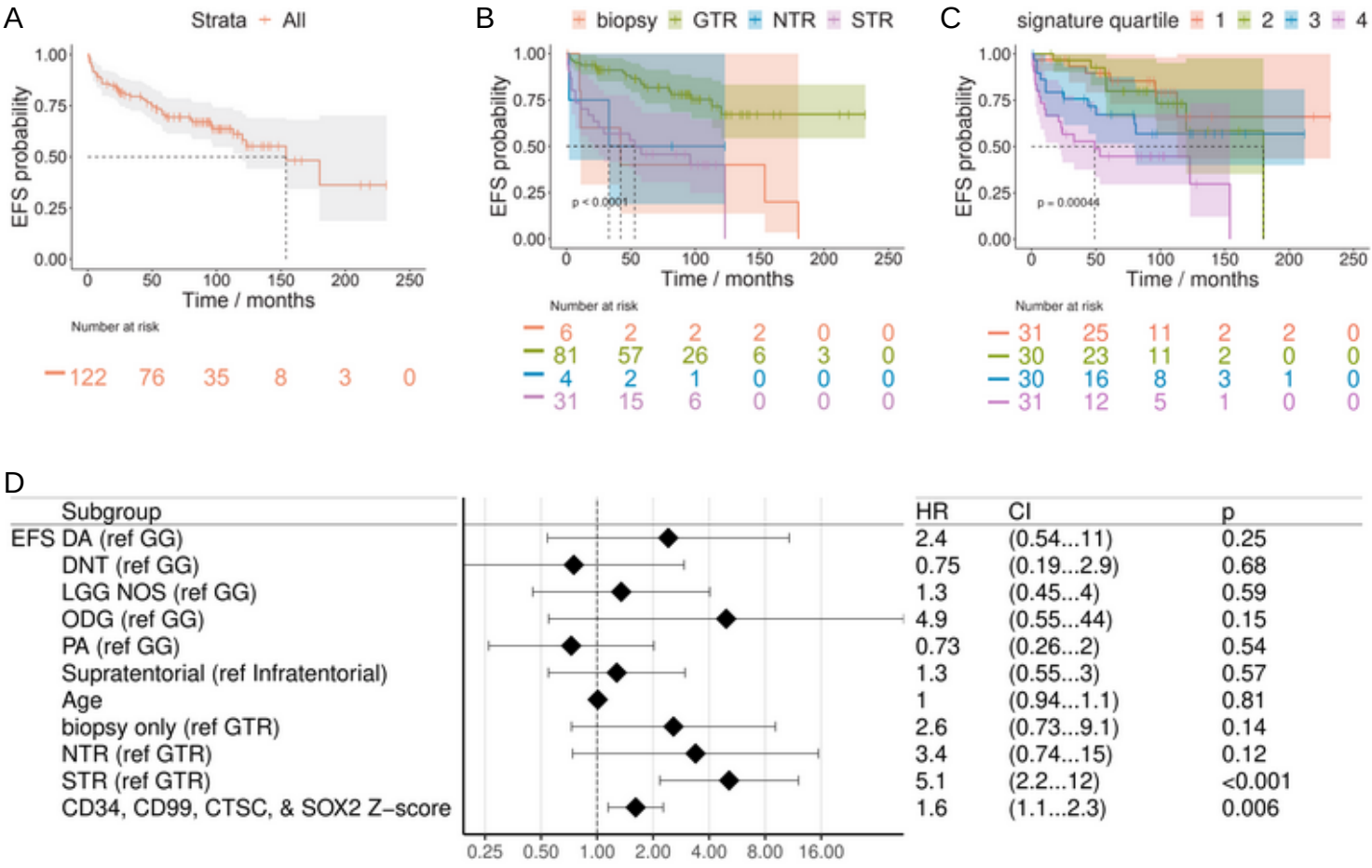


Figure 10: Low grade glioma outcomes analysis.
(A-B) Kaplan-Meier EFS (95% CI shaded, dotted line for median) for Bergthold *et al*. patients overall (A) and by resection status (B).
(C) Kaplan-Meier EFS (95% CI shaded, dotted line for median) for Bergthold *et al*. patients by UCell score quartile for the gene signature of CD34, SOX2, CD99, and CTSC.
(D) Forest plot of COXPH EFS HR for Bergthold *et al*. patients with continuous variables kept continuous. Includes HR for Z-score for the gene signature of CD34, SOX2, CD99, and CTSC.
EFS=event-free survival, NR=none reported, GTR=gross total resection, NTR=near total resection, G/NTR=gross or near total resection, STR=subtotal resection, WT=wild-type. DA=diffuse astrocytoma, GG=ganglioglioma, ODG=oligodendroglioma, DNT=dysplastic neuroepithelial tumor, NOS or LGG=low-grade glioma, not otherwise specified. PA=pilocytic astrocytoma.