Re-convolving the compositional landscape of primary and recurrent glioblastoma reveals prognostic and targetable tissue states

Osama Al-Dalahmah[1,2], Michael Argenziano[3], Aayushi Mahrajan[3], Julia Furnari[3], Fahad Paryani[5], Deborah Boyett[3], Akshay Save[3], Fatima Khan[1], Juncheng Li[1], Hong Lu[1], Yu Sun[1], John F. Tuddenham[5], Alexander R. Goldberg[1], Athanassios Dovas[1], Matei A. Banu[3], Tejaswi Sudhakar[3], Erin Bush[4], Andrew B. Lassman[2,5], Jeffrey N. Bruce[2,3], Peter A. Sims[2,4], Vilas Menon[5], Peter Canoll[1,2]

[1]Department of Pathology and Cell Biology, Columbia University Vagelos College of Physicians and Surgeons (VP&S), New York, NY, USA
[2]Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY, USA
[3]Department of Neurological Surgery, Columbia University Vagelos College of Physicians and Surgeons (VP&S), New York, NY, USA
[4]Department of Systems Biology, Columbia University Vagelos College of Physicians and Surgeons (VP&S), New York, NY, USA
[5]Department of Neurology, Columbia University Vagelos College of Physicians and Surgeons (VP&S), New York, NY, USA

Running title
Re-convolving the compositional landscape of GBM

Corresponding Authors:

Osama Al-Dalahmah, Email: oa2298@cumc.columbia.edu

Vilas Menon, Email: vm2545@cumc.columbia.edu

Peter Canoll, Email: pc561@cumc.columbia.edu

Abstract

Glioblastoma is an aggressive diffusely infiltrating neoplasm that spreads beyond surgical resection margins, where it intermingles with non-neoplastic brain cells. This complex microenvironment harbouring infiltrating glioma and non-neoplastic brain cells is the origin of tumor recurrence. Thus, understanding the cellular and molecular features of the glioma microenvironment is therapeutically and prognostically important. Here, we used single-nucleus RNA sequencing (snRNAseq) of primary and recurrent glioma to define three compositional 'tissue-states' rooted in cohabitation of cell-types and transcriptional states. These comprise a state featuring A) abundant normal brain microenvironment cells, B) reactive/inflamed brain microenvironment, and C) cellular/proliferative tumor. All these states exhibited variable degrees of infiltration by glioma cells. Spatial transcriptomics confirmed that the cell-types and transcriptomics states which compositionally cohabitate tissue states also colocalize in space. Tissue states are clinically significant because they correlate with radiographic, histopathologic, and prognostic features. We found that enrichment of tissue state B signature correlated with shorter survival. Importantly, we found that our compositionally defined tissue states are enriched in distinct metabolic pathways. One such pathway is fatty acid biosynthesis, which was enriched in tissue state B – a state enriched in recurrent glioblastoma and composed of astrocyte-like/mesenchymal glioma cells, reactive astrocytes resembling those seen in neurodegeneration, and monocyte-like myeloid cells. We showed that treating acute slices of GBM with a fatty acid synthesis inhibitor is sufficient to deplete the transcriptional signature of tissue state B. Our findings define a novel compositional approach to the glioma-infiltrated tissue which allows us to discover prognostic and targetable features, paving the way to new mechanistic and therapeutic discoveries.

Introduction

Glioblastoma (GBM) is the most malignant glial tumor of the brain and is refractory to current treatment. Although gross surgical resection of the visible tumor is sometimes feasible, glioma cells infiltrate the brain beyond the resection margins. While many studies have characterized the transcriptional and genomic features of GBM cells and glioma associated microglia/myeloid cells, a comprehensive analysis of other cells in the GBM microenvironment, and the patterns of cohabitation of different cell types is lacking. Previous studies have shown that the composition of glioma infiltrated samples varies from cellular tumor comprised of GBM and myeloid cells, to minimally infiltrated GBM margin tissue composed largely of non-neoplastic brain microenvironment cells, including neurons and glia [1, 2]. This is the microenvironment into which tumor cells migrate and proliferate, leading to recurrence, and is also the target of adjuvant therapy. Thus, understanding the cellular milieu of the tumor microenvironment at presentation and recurrence, including both neoplastic and non-neoplastic cells, is vital for advancing the management of GBM. Our goal is to determine patterns of cellular composition and transcriptional states in primary and recurrent GBM, including both neoplastic glioma cells and non-neoplastic brain cells.

Early studies used bulk RNA sequencing approaches to understand GBM states in MRI-localized samples from contrast-enhancing (CE) and non-contrast enhancing (NCE) margins [3-6]. More resolution is attained using single cell RNAseq (scRNAseq) approaches, which are being increasingly used to understand heterogeneity in gliomas. Several studies have employed scRNAseq from freshly resected surgical samples to explore the heterogeneity of GBM [2, 7-12]. These studies have significantly advanced our understanding of the heterogeneity and pathology of glioma. However, application of whole cell scRNAseq is faced with practical challenges related to the limitations of acquiring and processing freshly resected glioma tissue and the technical incompatibility with banked frozen glioma tissue. Moreover, scRNAseq is limited in sampling non-neoplastic cells of the microenvironment like neurons and astrocytes, which are major constituents of the tumor-margins [2, 8-11], in part because of cell-type survivability/selection bias during tissue dissociation. Thus, while advances have been made in defining the genetic alterations in recurrent glioma [13, 14] and features of myeloid cells [15], comprehensive analyses of cellular composition and diversity of cellular phenotypes in primary and recurrent gliomas remain a challenge.

Here, we circumvented these limitations of scRNAseq by using single nucleus RNAseq (snRNAseq), which allowed us to analyze frozen tissue, and inclusively sample cells of the microenvironment from primary and recurrent glioma. We sampled glioma-infiltrated tissue, from cellular tumor to minimally infiltrated surrounding brain tissue at the single cell level. Transcriptional analysis of copy number variations (CNVs) provided a metric to distinguish neoplastic (CNVpos) and non-neoplastic (CNVneg) nuclei, and unbiased clustering revealed that primary and recurrent tumors harbor CNVpos glioma cells with similar transcriptional states. Conversely, the microenvironment of primary and recurrent glioma displayed distinct cell-type specific states and different compositional landscapes. Leveraging information from the snRNAseq-derived compositional make-up of glioma-infiltrated samples defines three generalizable “tissue states” defined by cellular cohabitation in the glioma-infiltrated tissue, with each tissue state showing enrichment for specific gene signatures that can be identified in more accessible bulk RNAseq samples. This cohabitation of cell-types/transcriptional states was confirmed using spatial transcriptomics. We demonstrate that tissue states are prognostically relevant and display metabolic dependencies that can be pharmacologically targeted.

RESULTS

Transcriptional analysis of the glioma microenvironment reveals prognostically significant subpopulations of non-neoplastic astrocytes

Given the importance of glioma microenvironment in tumor progression, we decided to investigate the implications of microenvironmental states on the prognosis of GBM. To achieve that, we first identified neoplastic and non-neoplastic nuclei based on chromosomal copy number variation (CNV) inference (Supplementary results). Given the repertoire of transcriptional states glioma cells that have been previously described ([2, 7-10, 12]), we confirmed that our CNV positive (CNVpos) neoplastic nuclei from primary and post-treatment recurrence GBM recapitulate known transcriptional states. We provide this data in the supplementary results including discussion of glioma states in primary and recurrent glioma (**Figures S1, S3**), CNV analysis of primary and recurrent glioma samples (**Figures S2, S4**), localization studies of glioma states in the tissue (**Figure S5**), survival analysis of glioma signatures (**Figure S6**), and details on other low grade glioma and epilepsy samples included in this study (**Figures S7-8**). We focused on the non-neoplastic CNV negative (CNVneg) nuclei of the glioma microenvironment and combined in our analysis nuclei from primary and recurrent glioma, as well as nuclei from low-grade glioma (LGG) and epilepsy, to include a spectrum of neurological diseases with alterations to non-neoplastic cells in the brain microenvironment. The clinical data on the samples, QC metrics, and number of nuclei per lineage/cluster is provided in **supplementary table-1.** Our CNVneg nuclei datasets included 16831 nuclei: 6929 from primary glioma, 6008 from post-treatment recurrent glioma, 2875 from epilepsy, and 1019 from LGG. We projected these nuclei in UMAP space and assigned cell lineages as shown in **Figure-1A**. The expression of a select number of marker genes per lineage is shown in **Figure-1B**. We present the results on myeloid lineage nuclei in the supplementary results (**Figure-S10**), which demonstrates that monocyte-derived tumor-associated macrophages (TAMs) were enriched in recurrent glioma, while microglia-derived TAMs were enriched in primary glioma, consistent with a previous report[15].

We focused on astrocytes, which are key elements of the glioma microenvironment and are not well represented in glioma single-cell RNAseq datasets [2, 7-12, 16]. A recent paper implicated GBM-associated astrocytes in promoting an immunosuppressive microenvironment [17]. Moreover, the distinction between tumor-astrocytes and reactive-astrocytes is of major diagnostic importance in neuropathology. Thus, we analyzed astrocytes (707 nuclei – 284 from primary glioma, 254 from recurrent glioma, 45 from LGG, and 121 from epilepsy) in isolation from other cell types, reduced their dimensions, and clustered them into three states; Ast1 – protoplasmic astrocytes, Ast2 – reactive astrocytes with expression of oligodendroglial and neuronal genes, and Ast3 – reactive astrocytes with inflammatory gene expression (**Figure-1C and supplementary table 4**). The astrocytes are projected by disease condition in **Figure 1D**. Expression of select markers of these astrocytes states is shown in **Figure 1E**. A more in depth analysis of glioma-microenvironment astrocytes, including discussion of the sub-clusters, sub-cluster markers, differential distribution between conditions, and prognostic relevance are provided in the supplementary results (**Figure-S9A-F and supplementary table-5**). Notably, a random forest classifier trained to identify astrocytes from Huntington disease versus controls[18], when applied to our glioma data classified the majority of Ast3 cells as Huntington disease like, and the majority of Ast1 as control like (**Figure-1F).** Since astrocytes and glioma shared gene signatures (for example, CLU and LGALS3 expression), we performed differential gene expression analysis between primary and recurrent glioma non-neoplastic astrocytes and all CNVpos glioma nuclei and identified 1620 genes were higher in astrocytes compared to glioma and 3380 were higher glioma compared to astrocytes. Examples of genes higher in non-neoplastic astrocyte include genes associated with Alzheimer's disease (CLU, APOE)[19, 20], metallothionein genes (MT1H, MT1G, MT1M, MT1F, MT1E, MT1X, MT2A, and MT3 – increased in reactive astrocytes [18]), Synuclein genes (SNCA, SNCB, and SNCG), WIF1, CHI3L2 (associated with poor prognosis in glioma[21]), ALDOC, ALDOA, AQP4, carbonic anhydrases CA2 and CA11, and CXCL14, a cytokine implicated in promoting glioma invasion [22] (**Supplementary Table-4)**. Conversely, genes higher in CNVpos glioma include EGFR, PTPRZ1, NOVA1, CD24, Nestin (NES), SOX5, and SOX4. We used KEGG pathway enrichment analysis to query the function of these genes (**Figure-1G-H**). Further analysis of the differentially expressed genes showed that several KEGG pathways were enriched in genes higher in non-neoplastic astrocytes (**Figure-1G)**, with some relating to neurodegeneration such as Parkinson disease, and prion disease. Notably, these signatures are highly enriched in oxidative phosphorylation genes (**supplementary Table-4**), which is dysregulated in neurodegenerative diseases [23]. Moreover, other metabolic pathways enriched in astrocyte DGE included metabolism of fatty acids, glycolysis, TCA cycle, and ferroptosis. Conversely, KEGG pathways increased in CNVpos tumor-astrocytes were largely related to DNA replication, cancer-related pathways including ErbB and MAPK signaling, DNA replication and mismatch repair (**Figure-1H** and **Supplementary Table-4)**.

Re-convolution of snRNAseq identifies three tissue states based on cellular composition of glioma and its microenvironment

Given the heterogeneity of cellular states of glioma and non-neoplastic cells in the glioma microenvironment, we hypothesized that the transcriptional landscape of GBM is determined by patterns of cohabitation of specific types and transcriptional states of neoplastic and non-neoplastic cells. To test this hypothesis, we first asked if specific glioma, or brain microenvironment lineages were differentially abundant or deplete across primary and recurrent glioma using a regression model [24] to test for differential abundance (**Figure-2A**). The results showed that for CNVpos cells, gl_Mes2 were significantly more abundant in recurrent glioma, while gl_PN1 were more abundant in primary glioma, (Benjamini-Hochberg adjusted p values (q-value) 3.99e-2 and 1.318e-5 respectively). For the CNVneg cells in the glioma microenvironment, OPCs were significantly more abundant in primary glioma (q-value 1.085e-03). These results show that patterns of cellular composition vary in primary and recurrent glioma, and likely contribute to determining the transcriptional landscape of glioma.

To uncover patterns of ‘tissue states” with correlated cell states/lineages, we took advantage of the relatively unbiased sampling of cellular composition in the brain tumor microenvironment provided by snRNAseq. We approximated the cellular composition of each surgical sample by recombining the cells from all the distinct cell populations, as identified by snRNA-seq, to create a compositional matrix containing the abundance of all cell types across all samples (**Supplementary Table-1**). The cellular composition matrix includes three astrocytic clusters (Ast1-3 – see methods), five immune-cell states (Myel1, moTAM, mgTAM, prTAM, and T cells - see supplementary results and methods), neurons, oligodendrocytes, endothelial cells, OPCs, and glioma cells. We then used principal component analysis of the resulting cellular composition matrix and identified the compositional features that account for the variance across the samples (**Figure-2B**). We used the glioma states as supplementary quantitative variables [25] – the coordinates of which can be predicted from the other variables inputted into the PCA analysis. The results showed that the relative abundance of CNVpos glioma cells versus CNVneg non-neoplastic cells (neurons, oligodendrocytes, OPCs) is the major feature of the first principal component, and the abundance of reactive astrocytes (Ast3), macrophage-like myeloid cells (moTAM), and T cells is the major feature of the second principal component. Notably, the abundance of a specific subpopulation of mesenchymal glioma cells (gl_Mes2) was also highly correlated with the second principal component (PC2). These finding indicate that specific subpopulations of neoplastic and non-neoplastic cells tend to co-inhabit glioma samples. Additionally, PC2 was positively correlated with the recurrent glioma condition (p.value= 0.027). To assess if the cohabitation of cell types and transcriptional states is prognostically relevant, we used the IDH-WT GBM TCGA and CGGA survival datasets and performed a log-rank test on samples with positive versus negative PC2 signature enrichment and found that positive enrichment is significantly associated with poor survival (**Figure-2C**). This data show that glioma infiltrated tissue shows patterns of cellular composition driven by co-habitation of specific cell-types and transcriptional states and reveal prognostically-relevant gene signatures that span across both neoplastic and non-neoplastic cell states that co-inhibit tissue samples.

As a second approach to characterize the patterns of co-habitation of cell types and transcriptional states in GBM, we analyzed nine samples of IDH-WT GBM infiltrated brain tissue using spatial transcriptomics (ST - **Supplementary table-1**, **Figure 2D-E** and **Figure S11**). This analysis highlighted the spatial patterns of colocalization of high enrichment scores for specific glioma states and non-neoplastic cell types, such as astrocyte-like/mesenchymal glioma (gl_Mes2) and reactive astrocytes Ast3 (**Figure 2D**). Extending this analysis, we determined the enrichment scores of all cell-type/transcriptional state gene signatures measured across all nine

GBM-infiltrated ST samples, correlated them using Pearson correlation, and then clustered the cell-types/transcriptional states on the Euclidian distance matrix derived from the correlation scores using hierarchical clustering (**Figure 2E** – see methods). Three main clusters were apparent: 1- which showed high correlation between astrocyte-like/mesenchymal glioma gl_Mes1 and gl_Mes2, reactive astrocytes Ast3, endothelial cells, T-cells, and monocyte-like myeloid cells moTAM; 2- with high correlation between OPCs and proliferative glioma gl_Pro1 and gl_Pro2; and 3- with high correlation between neurons, proneural glioma gl_PN1 and gl_PN2, mgTAM, protoplasmic astrocytes Ast1. These findings are consistent with the results from snRNAseq, and provide additional evidence to support cohabitation of cell-types/transcriptional states.

Driven by the above findings, we clustered the snRNAseq samples into 3 distinct “tissue-states” based on the approximated cellular compositions described above; **tissue-state A** samples are predominantly composed of non-neoplastic brain cells, including neurons oligodendrocytes, and OPCs**, tissue-state B** samples are enriched in reactive astrocytes, myeloid/macrophages, and T-cells, and **tissue-state C** samples are predominantly composed of CNVpos glioma cells (**Figure 3A-B**). To generate a gene signature for each tissue state, we combined the snRNAseq for all nuclei in each sample and performed differential gene expression analysis between tissue-state clusters, using the re-convolved expression profile of each sample as a biological replicate. This analysis identified the top-differentially expressed genes unique to each tissue state (**Supplementary Table-7**). To assess the generalizability of the three tissue-states, we performed single sample GSEA analysis for the tissue state gene signatures using a dataset of bulk RNAseq analysis performed on 91 primary and recurrent MRI-localized samples from 39 patients. We found that these samples separated into 3 compositional clusters based on their enrichment score for snRNAseq-defined “tissue-states” (**Figure 3C**). We refer to the compositional clusters and tissue-states interchangeably henceforth. These three tissue-states are further demonstrated by projecting the RNA-expression levels for canonical markers of the predominant cell types for each tissue-state in **Figure 3E** showing RBFOX3 (neuronal marker) in tissue-state A, CD68 (myeloid marker) in tissue-state B, and MKI67 (proliferation marker) in tissue-state C. Sox2 (pan-glioma marker) was widely distributed across the samples, indicating variable degrees of tumor infiltration across samples in all three tissue-states (**Figure 3E**). Further analysis revealed that these tissue-state gene signatures are enriched for specific biologically relevant functional ontologies. For example, Tissue-state A is enriched of genes involved in synaptic transmission, tissue-state B is enriched for genes associated with oxidative stress and inflammation, and tissue state C is associated with cell proliferation (**Figure 3D**). To further validate these findings, we quantified total cellularity and the IHC labeling indices SOX2, NeuN, CD68, and Ki67 in 45 recurrent and primary glioma samples (**Figure 3F**) and found that total cellularity was highest in cluster C, which also had the highest abundance of SOX2+ and Ki67+ cells, while cluster A had the highest abundance of NeuN+ cells, and Cluster B had the highest abundance of CD68+ cells. While Clusters A and B resemble normal and reactive brain tissue, the SOX2 and Ki67 labeling indices indicate that these clusters comprise samples with variable levels of glioma infiltration.

To substantiate the clinical relevance of investigating glioma tissue in terms of tissue states, we investigated whether the enrichment of tissue state signatures correlated with survival in the TCGA-CGGA IDH-WT glioblastoma dataset. Given that tissue state B was enriched for the gene signatures of Ast3, moTAM, T-cells, and gl_Mes2 (**Figure-3G**), and considering our findings in Figure-2C, we expected it to be associated with increased risk of death in survival cohorts. As expected, enrichment of Cluster B gene signature in the IDH-WT TCGA and CGGA datasets was associated with a significant increase in the hazard of death in cox proportional hazard regression model, with covariates controlled for including age, gender, and MGMT methylation status (**Figure 3H**). To further establish the clinical relevance of taking a tissue state approach in investigating glioblastoma transcriptomics, we asked if the tissue state B differentially enriched in primary versus recurrent glioblastoma status. This question was especially relevant given that compositional cluster B was largely composed of recurrent glioma samples (**Figure-3A**). We thus asked if that signature is positively enriched in RNAseq profiles from previously published paired primary and recurrent glioblastoma samples [26] (**Figure-3I**). The results showed significant enrichment is tissue state B signatures at recurrence. Together, the results show that tissue state B signature is prognostic and enriched during GBM recurrence. As a third way to establish the clinical relevance of tissue states, we determined if these signatures are localized to specific radiographic regions of glioblastoma. Our results showed that while control samples were most enriched in tissue state A signature, GBM-infiltrated samples displayed spatially defined patterns that varied according to the disease condition (primary vs post-treatment recurrence - **Figure S12C**). For example, contrast-enhancing primary GBM samples were most enriched in tissue state C signature, while contrast-enhancing recurrent samples were most enriched in tissue state B signature. Glioma-margin samples (FLAIR abnormal) from primary GBM showed a range of enrichment scores for each of the three tissue state signatures. In contrast, most FLAIR-abnormal samples from recurrent GBM showed enrichment of tissue state A or B signatures, but very few showed enrichment for the tissue state C signature (see supplementary results - **Figure S12**).

Glioma-associated tissue states are targetable and associated with distinct metabolic states

Given the distinct cohabitation patterns that drive tissue states, we hypothesized that these patterns of cellular cohabitation are associated with metabolic dependencies. To test this hypothesis, we investigated whether metabolic pathways are differentially enriched in genes differentially expressed between bulk RNAseq samples of the three tissue states. Unbiased analysis of enrichment of KEGG pathways in genes differentially expressed between compositional clusters/tissue-states revealed that they exhibit enrichment of multiple unique and specific pathways (**Figure 4A**). Interestingly, several of the tissue state-enriched pathways were metabolic pathways. Tissue-state A showed highest enrichment for oxidative phosphorylation and beta-glutamate metabolism, tissue-state C was most enriched for pyrimidine, folate, and branched chain amino acid metabolism, and tissue-state B showed highest enrichment of fatty acid and lipid metabolism (**Figure 4B**). We focused on fatty acid biosynthesis genes, a tissue state B enriched pathway, and projected the average normalized expression per lineage as a heatmap in **Figure 4C**. We found that genes in this pathway were distributed across multiple cell types, suggesting that the metabolic status of a tissue can have distinct, but functionally related effects on different cell types in that tissue. Notably, fatty acid synthase (FASN), a rate-limiting enzyme in fatty acid synthesis{Garcia Corrales, 2021 #2471}, was most highly expressed in astrocytes and glioma cells **(Figure 4C).** FASN inhibition has been shown to kill glioma cells [27], however, the impact of FASN blockade on the glioma microenvironment is yet to be fully explored. Defining the effects of FASN blockade on the glioma microenvironment is important because fatty acid metabolism is a physiologic pathway that involves interactions between multiple cell types that reside in the same habitat. In non-neoplastic brain tissue, fatty acids are synthesized by astrocytes and are distributed to other cells including neurons and oligodendrocytes [28], where they drive physiologic and cellular functions like neuronal maturation, membrane synthesis [29], and neuroprotection [30]. We thus hypothesized blocking fatty acid synthesis pathway would interfere with the cells that make up tissue state B and/or their interactions, and therefore would lead to depletion of tissue state B signature in glioblastoma infiltrated brain. To test this hypothesis, we treated astrocytes and explants of human IDH-WT glioblastoma with the FASN inhibitor Cerulenin (5mg/ml) and measured gene expression using the plate-seq RNAseq (**Figure 4D**). Astrocytes treated with Cerulenin exhibited numerous differentially expressed genes compared with DMSO controls (**Supplementary table-8, Figure 4E**). Genes increased in treated astrocytes were enriched in KEGG and Reactome pathways involved in mTOR signaling, ferroptosis, and unfolded protein response, while those decreased in treated astrocytes were enriched in pathways involved in cell cycle. Cerulenin treatment did not alter astrocyte viability (data not shown). We then treated IDH-WT glioblastoma explants with DMSO or Cerulenin **(Supplementary table-8)** and measured gene expression (**Figure 4F**). We found that genes increased in Cerulenin treated astrocytes were significantly enriched in Cerulenin treated IDH-WT glioblastoma explants, and that the tissue state B signature was depleted (negatively enriched). Overall, these results demonstrate that tissue-states exhibit enrichment of metabolic pathways, which can be targeted leveraging compositional information and metabolic dependencies.

DISCUSSION

In this work, we investigated the landscape cellular composition and transcriptional states of glioma and its microenvironment in primary and post-treatment recurrent IDH-WT GBM using snRNAseq and spatial transcriptomics. Understanding heterogeneity of GBM is important for guiding treatment and meeting the challenge of recurrence. Recent studies revealed a diversity of glioma states that resemble cell lineages found during development and adulthood [2, 9-12, 16]. Our study provides a comprehensive analysis of the GBM microenvironment, including non-neoplastic cell types. Using a compositional approach rooted in relatively unbiased sampling of different GBM microenvironment cell types, we discovered that specific cell types/transcriptional states colocalize in “tissue states”. Leveraging insight into correlated cellular states and lineages that co-inhabit tissue samples, we identified gene signatures that classify primary and recurrent GBM tissue into three tissue states: (A) normal brain, (B) reactive/inflammatory tissue, and (C) cellular/proliferative tumor. The tissue states exhibited variable levels of infiltrated by glioma cells**.** The patterns of co-habitation in the tissue state model are further supported by spatial transcriptomics data, which highlights the differential distribution of specific astrocyte states. Importantly, we discovered that enrichment tissue state B, a reactive state that harbors a reactive astrocyte state (Ast3) resembling neurodegenerative astrocytes, was associated with increased risk of death. We show that gene signatures for these tissue states can also be identified in more accessible bulk RNAseq samples and correlate with immunohistochemical profiles. Significantly, we found that tissues states were transcriptionally enriched in distinct metabolic pathways, and that targeting fatty acid synthesis, a pathway enriched tissue state B, resulted in depletion of that signature in *ex vivo* GBM slice cultures. The therapeutic implications of our findings help expand the target of therapy from targeting one gene or one cell type, to targeting to tissue states comprising cell populations that co-inhabit the tissue under defined metabolic constraints.

Our analysis of the cellular phenotypes in the glioma microenvironment revealed that subpopulations of non-neoplastic astrocytes show enrichment for abnormal transcriptional signatures that are also seen in the context of neurodegenerative diseases. In contrast to CNVpos neoplastic astrocytes, which express high levels of proliferation and glioma genes, a subpopulation of non-neoplastic astrocytes (Ast3) displayed a reactive signature reminiscent of astrocytes described in neurodegenerative diseases like Huntington disease, Parkinson disease and Alzheimer's disease cortex [18, 31]. An unbiased random forest classifier also demonstrated that most Ast3 cells were classified as similar to Huntington disease astrocytes. It is notable that the non-neoplastic glia in the microenvironment of glioma exhibited neurodegenerative signatures even though most of the cases we profiled were less than 65 years old - the age after which neurodegenerative diseases become more prevalent. Our findings thus uncover commonalities in the astrocytic response to glioma and neurodegeneration. Given that GBM is a disease of the elderly population, one can envisage that future astrocyte modifying therapies developed for neurodegenerative disease may have a role in therapeutically modifying the microenvironment of GBM.

Our analysis of glioma states (supplementary results) revealed they can be grouped into three main categories based on commonalities with previously described glioma-state signatures: astrocyte-like/mesenchymal (gl_Mes1 and gl_Mes2), progenitor-like/proneural (gl_PN1 and gl_PN2), and proliferative (gl_Pro1 and gl_Pro2). We found that the astrocyte-like/mesenchymal signatures (gl_Mes1 in primary GBM and gl_Mes2 in recurrent GBM) portended poor survival in the IDH-WT GBM survival series (TCGA and CGGA), consistent with previous studies [6]. Moreover, GBM recurrence was correlated with increased gl_Mes2 and reduced gl_PN1 abundance in our dataset. This is consistent with results from bulk RNAseq on paired primary-recurrent glioma pairs, where the authors found enrichment of the mesenchymal subtype, and depletion of the classical subtype [6]. Analysis of paired primary and recurrent GBM samples from Wang et al. 2021 [26] revealed that tissue state B was enriched at recurrence. That being stated, our dataset is based on largely unpaired snRNAseq from primary and recurrent GBM and is of relatively smaller size, which may preclude us from further interpretation.

One of the main findings highlighted by our analysis of cellular composition is that specific cell types are correlated with each other both compositionally and spatially, indicating that they co-inhabit the same tissue-states. Cohabitation between cell types and transcriptional states was reflected in enrichment of distinct metabolic pathways. For example, tissue state B was enriched in the glutathione pathway, which determines a cell's sensitivity to ferroptosis-inducing drugs [32], and in fatty acid metabolism, which has been implicated in glioma survival, stemness and progression [27, 33, 34]. We found that fatty acid metabolism genes were distributed among different cell types in the brain, however, FASN, the rate-limiting enzyme in fatty acid synthesis {Garcia Corrales, 2021 #2471} was most highly expressed in astrocytes and glioma cells. Astrocytes play key roles in lipid metabolism; for example, in synthesizing fatty acids necessary for neuronal membranes {Tabernero, 2001 #2491} and catabolizing fatty acids released by neurons during excitotoxicity {Ioannou, 2019 #2493}. We showed that blocking FASN effectively depleted tissue state B signature from treated GBM slices. This may be explained by either a change in the composition of the GBM slices, given that FASN inhibition may lead to glioma cell death [34]{Grube, 2014 #2418}, a change of gene expression of the cells that reside in the slices, or both. The latter is likely the case, given that GBM slices treated with FASN inhibitor showed a positive enrichment for the gene signature of astrocytes treated with FASN inhibitor, and negative enrichment for tissue state B. These finding are clinically relevant, given FASN is a promising target against glioblastoma [36], and highlight how the tissue-state approach can provide new insights into the effects of targeted therapies on the GBM microenvironment.

Methods

Human subjects and glioma tissue

Frozen primary untreated GBM tissue was acquired from the Bartoli brain tumor bank at Columbia University Medical Center. All diagnoses were rendered by specialized neuropathologists. Study protocols were approved by Columbia University Medical Center Institutional Review Board. All clinical samples were de-identified prior to analysis. Analyses were carried out in alignment with the principles outlined in the WMA Declaration of Helsinki and the Department of Health and Human services Belmont Report. Informed written consent was provided by all patients. The demographics of the cases used are provided in **Supplementary Table 1**.

Extraction of nuclei and snRNAseq procedure

Nuclei were isolated from frozen surgical resection specimen as described in Al-Dalahmah O et al. 2020. Briefly, the frozen tissue samples were dissected from fresh frozen tissue or frozen OCT-embedded tissue blocks to yield tissue measuring in general from 5 x 2 x 1 mm to 10 x 6 x 3 mm. The tissue was homogenized using a dounce homogenizer in ice-cold 30% sucrose 0.1% Triton-X 100 based homogenization buffer. 10-15 strokes of the loose dounce pestle were followed by 10-15 strokes of the tight dounce pestle on ice. Mixing using a P1000 pipette followed before filtration through a BD Falcon 40um filters. Filtration was repeated after a 10-minute spin at 1000g at 4c. A cleanup step followed using a density gradient step as described in [37]. The nuclear pellet was suspended in 1% BSA in PBS resuspension buffer containing RNAse inhibitors. A final filtration step using 20um Flowmi ™ filters followed before dilution to 700-1200 nuclei per ul in resuspension buffer. The nuclear suspensions were processed by the Chromium Controller (10x Genomics) using single Cell 3' Reagent Kit v2 or v3 (Chromium Single Cell 3' Library & Gel Bead Kit v2, catalog number: 120237; Chromium Single Cell A Chip Kit, 48 runs, catalog number: 120236; 10x Genomics).

Sequencing and raw data analysis

Sequencing of the resultant libraries was done on Illumina NOVAseq 6000 platformV4 150bp paired end reads. Alignment was done using the CellRanger pipeline (10X Genomics) to GRCh38.p12 (refdata-cellranger-GRCh38-1.2.0 file provided by 10x genomics). Count matrices were generated from BAM files using default parameters of the DropEst pipeline (Petukhov V et al. 2018). Filtering and QC was done using the scater package (3). Nuclei with percent exonic reads from all reads in the range of 25-75% were included. Nuclei with percent mitochondrial reads aligning to mitochondria genes of more than 19% were excluded. Genes were filtered by keeping features with >10 counts per row in at least in 31 cells. Further filtering of low quality cells was done to include cells with at least 400 detected genes and 10,000 reads.

Single Nuclei RNAseq analysis

Sequencing and analysis of raw data

Sequencing of the resultant libraries was done on Illumina NOVAseq 6000 platformV4 150bp paired end reads. We used 10X chromium v2 chemistry for samples PO1 and PO2, and v3 chemistry for samples PA1, PA2, and P3. Read alignment was done using the CellRanger pipeline (v3.1 - 10X genomics) to reference GRCh38.p12 (refdata-cellranger-GRCh38-1.2.0 file provided by 10x genomics). Count matrices were generated from BAM files using default parameters of the DropEst pipeline [38].

Data-cleanup

Filtering and QC was done using the scater package [39, 40]. Nuclei with percent exonic reads from all reads in the range of 25-73% were included. Nuclei with percent mitochondrial reads aligning to mitochondria genes of more than 15% were excluded. Genes were filtered by keeping features with >10 counts per row in at least in 31 cells. The count matrix of each sample was normalized by first running the quickcluster function, then estimating size-factors by calling scran::computeSumFactors() function with default options and clusters set to clusters identified by calling quickcluster function. scater::normalize() function was then used to generated normalized counts. Doublet identification was done using scran::doubletCells function with default options, and cells with doublet score of NMADs > 3 were excluded as we described previously [18].

Combining multiple datasets from different sequencing batches

To control sequencing and technical batches, we utilized canonical correlation analysis in Seurat [41] accounting for batch and mitochondrial read percentage for CNVneg nuclei. For CNVpos nuclei, we accounted for case and mitochondrial read percentage.

Pre-Clustering and clustering of nuclei

Pre-clustering of nuclei was done in Seurat using the shared nearest neighbor smart local moving algorithm. PCA reduction was used as the reduction in the FindNeighbors() step. Pre-cluster identity determination was done using geneset enrichment analysis of lineage markers [18] and by inspecting cluster markers generated by scran::findmarkers(direction=”up”) function. Microglia +/- oligodendrocytes were used as negative control cell for InferCNV pipeline (below). Once CNVneg cells were verified, cells from all cases were aligned using Seurat and clustered. Clusters with mixed identities based on enrichment of multiple lineage genes were sub-clustered iteratively until all “pre-clusters” showed pure identities. Only then do we combine the pre-clusters of the same lineage into lineages (Astrocytes, neurons, oligodendrocytes, myeloid, endothelial). For subclustering of astrocytes and myeloid cells, we analyzed the nuclei in isolation of other lineages, and re-aligned them in Seurat, and reduced the dimensions before subclustering. For CNVpos nuclei, unbiased clusters were combined into glioma states/lineages based on similarity in marker expression and enrichment for known gene sets described in Figures 1D and 2C.

Count normalization

Raw counts were normalized in Seurat using the sctransform function SCT() function with default settings and controlling for percent mitochondrial gene expression [42].

Copy number variation analysis of snRNAseq

To detect putative neoplastic tumor cells, we used combination of marker expression and large scale copy number variation inference as per the InferCNV R package [43]. We used the default parameter as described in the package documentation. As a control population, we used microglia and oligodendrocytes from case PO2_1. Iteratively, CNVneg clusters including Oligodendrocytes and Neurons were identified and added as control cells. Different gene window sizes were tested (50, 100, 200) and yield similar results. We then applied an orthogonal approach to label putative neoplastic cells based on previous approaches described in [2, 44]. Briefly, Log2+1 counts were averaged across chromosomes for each nucleus. A principal component analysis (PCA) was performed on autosomal chromosomes in factominer R package [25]. Chromosomes with high correlation with PC2 were the same as the ones shown in the detected by inferCNV() with the exception is PO1, where no neoplastic cells were detected by InferCNV or CONICS. A malignancy score was calculated by dividing the log2 gained chromosome counts over the sum of those that are lost (selection was limited to three chromosomes or less). The scores were then z-neoplastic per sample. To identify putative neoplastic nuclei in this method, we next performed K means clustering of the scaled malignancy scores in R using the kmeans function and centers argument set to 2. PO1 does not show bimodal malignancy score distribution and the results of kmeans clustering were not considered. For case PA3, only a minority of nuclei had malignancy scores > 2 standard deviations above the mean. Therefore, these cells were identified using outlier detection in a normal distribution as done in the getOutliers(, method = "I", rho=c(0.1,3))$iRight) in R. getOutliers is part of extremevalues R package https://github.com/markvanderloo/extremevalues. Only the consensus nuclei that were identified as CNV positive in both approaches were considered for analysis. Less than 7.0% of the nuclei were called alternately by the two methods and were excluded from the analysis. Identification of CNVpos nuclei in recurrent glioma samples was conducted through a combination of InferCNV and identification of clusters with high expression of tumor markers SOX2 and PTPRZ1.

Survival Analysis

Survival analysis was performed using the survfit() function in the survival package in R [45, 46], using the binarized enrichment of each of the gene sets as the covariate in the formula. For cox proportional hazards, the function coxph() was used in R, and the covariates are indicated in the main figures.

Correlation Analysis

Correlation between deconvolution of values of microglial as well as astrocytic clusters, and glioma proportions was done using Pearson correlation (function cor() in R). The heatmap was generated using the corrplot package in R.

Identification of glioma state and lineage top gene markers

The lineage specific genes were determined using scater::findmarkers(…, direction =”up”) function on the top-level lineages (Neurons, astrocytes, microglia, undetermined, oligodendrocytes, OPC, and endothelial cells). The glioma-state specific genes were determined using scater::findmarkers(…, direction =”up”) function on the neoplastic glioma states only. To select specific lineage/glioma state markers, we further filtered the top markers generated above by selecting the genes with positive log-fold change values in 90% or more of the cluster-to-cluster comparisons. The top 150 genes were selected and are provided in **Supplementary Tables 2-4** for primary glioma, recurrent glioma, and non-neoplastic lineages, respectively.

Principal component analysis

PCA analysis was done in factorminer R package[25]. A matrix of snRNAseq sample x cell type/cluster was used as input. The composition of each snRNAseq sample astrocyte and myeloid subclusters was consolidated as follows: Astrocytes subclusters 0, 1, and 8 were combined as Ast1 (baseline/protoplasmic); subclusters 2, 3, 4, and 5 as Ast2, and subclusters 6 and 7 as Ast3. Myeloid subclusters 0 and 9 were combined as Myel1 (baseline), subclusters 1 and 3 – moTAM (monocyte derived TAMs); subclusters 2, 4, 6, and 8 as mgTAM (microglia-derived TAMs); subcluster 5 was kept as prTAM (proliferative TAMs); and subcluster 7 was kept as T-cells. The all CNV positive cells were consolidated as CNVpos, and the glioma states were included as supplementary quantitative variables. Condition (primary versus recurrent) was used as a supplementary qualitative variable.

Random forest classifier

Classification of astrocytes as control versus Huntington disease like was done by training a random forest classifier based on the HD vs control astrocytes dataset previously published [18]. The classifier was trained on normalized log2+1 gene expression values (15516 genes) from

1064 astrocytic nuclei using the package caret[47] in R. The training dataset was centered and scaled. Three-fold cross-validation followed. The classifier achieved a 97.5% accuracy. The classification.

Acquisition of Tissue and Preparation of Acute Slice Cultures

Primary GBM tissue from two separate surgeries, TB 6571 (3 blocks of tissue) and TB 6579 (2 blocks of tissue) (see supplemental table for related clinical information), performed at Columbia University Medical Center/New York Presbyterian Hospital were retrieved fresh from the operating room in a sterile specimen cup and transported back to the laboratory on ice. Primary GBM acute slice cultures were prepared exactly as described previously [48]. Slices were treated as described previously with either DMSO or 5µg/mL Cerulenin for 18 hours prior to preservation and RNA extraction.

Bulk RNAseq using Plate-Seq

RNA extraction was done using the RNeasy Mini Kit (Qiagen cat# 74106). RNAseq was on spatially localized biopsies was performed using Plate-seq as described [49]. 75bp paired end sequencing was performed on Illumina NextSeq platform and read alignment was done using STAR [50] to the human genome (hg19, annotation: UCSC known genes), and analysis was done as previously described [3]. FPKM values were used in GSEA analysis. The count matrix for the TCGA GBM dataset was downloaded using the GDCquery tool in R. The Chinese Glioma Genome Atlas (CGGA) RNAseq datasets [51, 52] was downloaded from (http://www.cgga.org.cn/download.jsp). The counts were normalized using the vst() function in deseq2 R package [53]. IDH-WT only samples were kept from both datasets (TCGA: 139 samples, CGGA: 179 samples) and used for downstream analysis.

For acute slice-culture PLATE-seq analysis, slices were then transferred to OCT and frozen into blocks. Tissue from each slice was isolated for RNA extraction by the Columbia Molecular Pathology Core using QiaSymphony extraction method. Total RNA was quantified using Nanodrop measurements, and 150ng of RNA from each slice/condition was loaded into a well of a 96 well plate. Pooled library amplification for transcriptome expression (PLATE-Seq) was then performed on the 96 well plate as previously described [49]. FASTQ files were demultiplexed and aligned to reference genome and transcript counts were normalized via DESeq2.

Geneset enrichment analysis and Gene Ontology Analysis

The average normalized counts per gene per cluster was calculated. The resultant cluster-wise count matrix was used as input to the GSVA pipeline [54]. Gene sets used for various tests are provided in the supplementary material (**Supplementary Table-1**). The options used for performing the GSVA pipeline are as follows: method= ssgsea, kcdf="Gaussian", mx.diff=TRUE. Heat maps were generated using the heatmap.2 in R function from the package gplots (R Package) and scores z-scaled were indicated. Ontology enrichment analysis in gProfiler with default settings [55]. For GSEA of the combined TCGA and CGGA dataset, the enrichment was performed using method = “ssgsea” option on normalized counts, which normalizes the enrichment scores for each gene set per sample. GSEA in **Figure-3** was conducted using pre-ranked GSEA and was performed as described in Subramanian et al 2005 with 1000 permutations [56]. Log-2-fold-change between cluster B and the remaining clusters was used to rank the genes for the analysis, and marker genes from each sub-lineage was used for the gene sets. For GSEA in Figure-4, preranked GSEA (based on log2foldchange) was performed using tissue cluster gene sets and the genes significantly upregulated after astrocyte treatment with cerulenin.

Spatial transcriptomics

Spatial transcriptomics was conducted using 10X™ Visium Spatial Gene Expression Slide & Reagent Kit, 16 rxns (PN-1000184), according to the protocol detailed in document CG000239_RevD available in 10X demonstrated protocols. 10 micron-thick tissue sections were mounted on the ST slides and stained for nuclei – DAPI among other antigens using a rapid immunofluorescence protocol described in document CG000312_RevB available in 10X demonstrated protocols. Imaging of whole slides was done at 20X magnification on a Leica Aperio Versa scanner or a Leica DMI6 thunder tissue imager. After imaging, the slides were de-cover-slipped and the tissue permeabilized for 11 minutes (which was empirically determined to yield best results based on the Visium Spatial Tissue Optimization Slide & Reagent Kit PN-1000193 as detailed in the protocol provided in document CG000238_RevD available in 10X demonstrated protocols). The remaining steps were conducted according to the manufacturer's protocol. The libraries were sequenced on multiple Illumina Nextseq 550 (paired end dual-indexed sequencing) flowcells to achieve the recommended number reads per ST spot. The spatial transcriptomic (ST) samples were prepared using 10X genomics Cell Ranger (version 6.1.2) and Space Ranger (version 1.2.1) software. Raw tiff images of the tissue were labeled with Cell Ranger which generated a json file for Space Ranger to use during alignment. Labeled spots from Cell Ranger were inputted into the loupe-alignment argument in Space Ranger along with its respective tiff image file, FASTQ reads, and slide numbers. The reference genome used for alignment was built using the Space Ranger function spaceranger mkgtf with GRCh38 as the assembly and Ensemble 91 for the transcript annotations. All other parameters to generating the counts data for ST were set to its default setting. After the alignment was complete, counts data were pre-processed using Seurat's SCTransform (version 4.06) function to account for technical variates while retaining biological variance in the tissues. To increase robustness of downstream analysis, counts from ~20 neighboring ST spots were summed into 10 ST regions that are approximately 1 mm in diameter before use for gsea of lineage/cell type signatures. The ST regions were determined using an unbiased k-means clustering method on the actual distance matrix between the ST points, with k set to 10. ssGSEA (GSVA package) analysis was performed on the summed counts. The process of tissue shattering was repeated 100 times and the average GSEA scores per 1 mm ST region was used for correlation (Pearson) analysis as indicated above.

Differential gene expression analysis

EdgeR glmQLFTest was used and the top 3000 differentially expressed genes with an FDR cutoff of 25% [57] were extracted. Only datapoints with adjusted p-values less than 0.05 were used in downstream analysis. For plate-seq data differential gene expression analysis between treatment and control was performed adjusting for tissue block and patient using the Deseq2 pipeline [53]. For astrocyte cultures, differential gene expression analysis between treatment and control was performed adjusting for astrocyte passage and cell culture batch using the Deseq2 pipeline.

Immunohistochemistry, histology, and in situ hybridization

Standard chromogenic Immunohistochemistry was done as described previously [18]. Paraffin-embedded formalin-fixed tissue sections or fresh frozen sections briefly fixed in 4% PFA, for 10 min (40 C) in 4% PFA in PBS. Paraffin sections after deparaffinization were treated with antigen unmasking solution according manufacture recommendations (Vector Laboratories, Burlingame, CA). The following antibodies and dilutions were used SOX2 (1:200, Mouse monoclonal, Abcam, Ab218520), KI67 (1:500, rat monoclonal polyclonal, Thermo Scientific, 14-5698-80), CD68 (1:200, mouse monoclonal, Abcam cat# ab955), NeuN (1:1000, mouse monoclonal, Millipore, MAB377). For fluorescent IHC, secondary antibody conjugated to fluorophores: anti-mouse Alexa

Fluor 488 and 594 and anti-rabbit Alexa Fluor 488 and 594; goat or donkey (1:300, ThermoFisher Scientific, Eugene, OR) were applied for 1 hr at room temperature. In situ hybridization was done using RNAscope™ multiplex fluorescent v2 (ACDbio cat no 323100) per the manufacturer's protocol in 5-micron paraffin-embedded, formalin-fixed tissue sections. We used predesigned probes for PTPRZ1, CLU, TOP2A, NOVA1, MEG3, and SOX2 from ACDbio; cat# 584781, 584771, 470321, 400871, 584801, and 400871, respectively. Fluorescent images were taken on a Zeiss 810 Axio confocal microscope at 40X. Brightfield fluorescent images were taken on an Aperio LSM™ slide scanner at 20X and 40X.

Quantification of ISH

For quantification of in situ hybridization images we used the positive cell detection function in Qupath v0.2.3 [58]. We only quantified signal contained in DAPI-positive nuclei. First, DAPI positive nuclei were detected using the cell detection tool. Next, subcellular detection function was employed to segment puncta per each of the three probe channels. A random forest classifier was used to classify nuclei to be positive or negative, with a minimum of two puncta per channel to classify a nucleus as positive for the probe. Infiltrated cortex and cellular tumor core were annotated by a neuropathologist.

Cell Culture

Human Astrocytes (ScienCell cat #1800) were cultured in Astrocyte culture medium (ScienCell cat# 1801), 2% fetal bovine serum (ScienCell cat #0010), 1% astrocyte growth supplement (ScienCell cat# 1852) and 1% penicillin/streptomycin (ScienCell cat # 0503). The cells were maintained as adherent cultures on poly-L-Lysine coated tissue culture plates. The cells were passaged at 70-90% confluence and treated at passage numbers 5-7. DMSO or Cerulunin Sigma cat#C2389 at 5ug/ml was used to treat the cells for 18hours as indicated.

Statistical testing

Statistical comparisons were done using one-way ANOVA (or Kruskal Wallis test) and Tuckey post-hoc comparison in R. Statistical testing for RNAseq application is reported in the main text or respective methods section. Differential abundance analysis was done employing a moderated regression model in ANCOMBC with default parameters, assigned Condition (primary vs post-treatment recurrence) and CNVpos proportions in the design formula, and as described by the authors ([24]). One tailed paired t-tests were done to compare the core and margin percentages of the same case (Figure 3). A one sample t-test was conducted to determine if the percentage of TOP2A+ that were CLU+ was less than 50%.

Data availability

Data for spatial transcriptomic can be queried using an interactive web app: https://vmenon.shinyapps.io/gbm_expression/. All other data will be deposited as in GEO repositories prior to publication.

Acknowledgment

The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This research was funded by NIH/NINDS R01NS103473 (PC, PAS, JNB), The William Rhodes and Louise Tilzer-Rhodes Center for Glioblastoma at New York-Presbyterian Hospital (OA, VM, PC), Herbert Irving Comprehensive Cancer Center pilot research grant (OA), and the NIH/NCI Cancer Center Support Grant P30CA013696 (OA, ABL, JNB, PAS, PC).

This research was funded by NIH/NINDS R01NS103473 (PC, PAS, JNB), William Rhodes and Louise Tilzer-Rhodes Center for Glioblastoma (RCG) Collaborative Research Initiative grant (OA, VM, PC), HICCC pilot research grant (OA), and in part through the NIH/NCI Cancer Center Support Grant P30CA013696. This research was supported by the Genomics and High Throughput Screening Shared Resource and the Digital Computational Pathology Laboratory in the Department of Pathology and Cell Biology at Columbia University Irving Medical Center. We thank Dr. Claudia Deoge for help with RNAscope. We thank the immunohistochemistry core in the Department of Pathology and Cell Biology at Columbia University Irving Medical Center for help with CD14 and GFAP-SOX2 immunostains. We thank Al-Fardthakh© for support with aspects of bulk RNAseq data analysis.

Author contribution

OA, JNB, PAS, VM, and PC designed the study; OA, MA, DB, AS, AM, FK, JFT, ARG, AT, MAB, JF, TS, EB, JNB and PC conducted the experiments, including tissue procurement, tissue processing, and tissue analysis. OA, MA, PAS, VM, and PC analyzed the data. OA, and PC wrote the manuscript and all authors edited, read, and approved the final manuscript.

Supplementary Results

Single nucleus RNAseq reveals proliferative, astrocyte-like/mesenchymal, and progenitor-like/proneural states in both primary and recurrent GBM

Radiographically, GBM typically has a CE core surrounded by a non-enhancing infiltrated brain that is highlighted by FLAIR-signal abnormality by MRI (**Figure-S1A**). The histopathological features of the resected tumor can vary from highly cellular tumor with vascular proliferation to less cellular infiltrated brain. These features are shown in **Figure-S1H**, demonstrating samples with a cellular GBM core (red star in **Figure-1A, Figure-S2H PA1, PA2, PA3, and PO2_1**) and others with overlying cortex (green star in **Figure-S1A, Figure-S2H PO2_2 and PO1**), which we use below.

To explore the heterogeneity of primary GBM, we analyzed several banked surgical samples using snRNAseq as shown in (**Figure-S1A**). A total of 8 samples from 7 patients were selected for analysis (**Supplementary Table-1**). Neuropathological assessment of tumor cellularity ranged from cellular tumor with hallmarks of GBM, to reactive brain parenchyma with few atypical cells. This assessment was made on Hematoxylin and Eosin (H&E) stained formalin fixed paraffin embedded sections adjacent to or frozen cryosections of the frozen tissue analyzed by snRNAseq (**Figure-S2H)**. We isolated nuclei from the frozen tissue and subjected them to snRNAseq followed by downstream analyses including clustering, differential gene expression analysis, cluster marker detection, and gene set enrichment analysis (GSEA) as outlined (**Figure-S1A**). 15189 nuclei passed our QC (**Supplementary Table-1**). To distinguish putative glioma cells from non-neoplastic cells, we employed an established approach that infers large scale copy number alterations/variations (CNV) from RNA expression profiles [43]. Chromosomal heat maps showing putative neoplastic nuclei are shown in **Supplementary Figure-S2A-G**. Next, we also applied a second method to label nuclei based on a "malignancy score", which we have previously shown to be a robust metric to distinguish glioma cells from non-neoplastic cells [2, 10], and the consensus nuclei designated by both methods was used for downstream analysis. Nuclei with no consensus CNV status were excluded (4.7%). Uniform manifold approximation and projection (UMAP) plots from individual cases labeled by transformation status are shown in **Figure-S1B**.

We identified 7954 putative neoplastic nuclei with inferred large scale chromosomal CNV (CNVpos/glioma nuclei). Glioma nuclei showed multiple chromosomal alterations including gains of chromosome 7 and losses of chromosome 10 (**Figure-S1**). Having identified neoplastic and non-neoplastic nuclei, we aligned the datasets from multiple samples and performed clustering analyses separately on CNVpos (glioma) nuclei from all cases using shared nearest neighbor and the smart local moving algorithm [59]. A UMAP plot is shown for all primary glioma nuclei non-neoplastic nuclei color-coded by glioma state/lineage **Figure-S1C**. This approach identified 6 distinct clusters: these resembled progenitors (oligodendrocyte-progenitors (gl_PN1 - proneural) and neural-progenitors (gl_PN2 - proneural), astrocytes (gl_Mes1 and gl_Mes2 - mesenchymal), and proliferative cells (gl_Pro1 and gl_Pro2).

The identity of the glioma states is akin to previously described glioma states, as demonstrated by the enrichment of several gene lists from [3, 9, 60, 61] **– (Figure-S1D, supplementary Table-2)**. For example, gl_Pro1 and gl_Pro2 showed enrichment in gene sets specific for cell-cycle phases [61], with gl_Pro1 showing highest enrichment of G2/M genes (Gobin_G1) and gl_Pro2 showing highest enrichment of G1/S phase genes and DNA repair related genes (Gobin_G3). Clusters gl_PN1 showed enrichment of the Verhaak's proneural, and OPC signature genes, while gl_PN2 showed enrichment of NPC signature genes. Finally, gl_Mes1 showed enrichment of astrocyte-like signatures and Verhaak's classical signature while cluster gl_Mes2 showed enrichment of several gene sets related to reactive astrocytes, and Verhaak's mesenchymal signature [3, 7, 60]. Our clustering is consistent with that described in Neftel et al. 2019 [9] and Wang et al. 2019 [7], and the states we describe are compatible with those in Yuan et al 2018 [2]. To further clarify the cellular phenotypes represented in our glioma clusters, we measured the enrichment of the major biologic process and molecular function gene ontologies (GO) in the glioma state top gene markers (see methods)**.** GO enrichment analysis demonstrated enrichment of GO's relating to locomotion, neurogenesis, neuronal migration, and cell projection in gl_PN1 markers genes; Notch signaling, neuron development, and GABA reuptake differentiation, and synaptic signaling in gl_PN2 genes; response to organic substances, ion homeostasis, and Signaling by tyrosine kinases in gl_Mes1 genes; response to cytokines, interferon gamma, and leukocyte activation and immune response in gl_Mes2 genes; mitosis and nuclear division in gl_Pro1, and S-phase, DNA replication, and DNA repair in gl_Pro2(**Figure-S1F** and **Supplementary Table-2**). The identities of the clusters can also be appreciated by examining select gene markers **Figure-S1E and Supplementary Table-2.** gl_Pro1 expressed cell-cycle genes TOP2A, CENPF, and AURKB. gl_Pro2 showed highest expression of DNA damage/repair including FANCI, HELLS, and XRCC2. gl_PN2 showed high levels of CD24, MEG3, and SOX4. gl_Mes1 showed high levels of protoplasmic astrocyte genes including SLC1A3, LIFR, ATP1A2, C1orf61, and NTM, while gl_Mes2 showed highest expression levels for reactive astrocyte and immune genes including CLU, VIM, and SAT1. While our glioma states resemble those described in the literature, less is known about whether glioma cells assume similar states in the recurrent setting. Therefore, we bridged this gap by directly analyzing recurrent IDH-WT glioma samples using the same approach we used for primary GBM samples.

To define the states of IDH-WT glioma in the post-treatment recurrence setting, we analyzed 8 cases of post-recurrent IDH-WT glioma using snRNAseq (**Figure-S3A**). We identified 8908 neoplastic nuclei harboring large-scale CNV (**Supplementary Figure-S4**). Of the eight cases, two were paired recurrences from the primary samples (TB5124 – recurrent of TB4916, and TB5053 – recurrent of TB4718, see respective section on comparing paired samples below). We treated recurrent gliomas similarly to the treatment naïve primary tumors and clustered all neoplastic nuclei together. Like primary gliomas, we found that recurrent glioma clusters can be assigned two proneural, two mesenchymal, and two proliferative states (**Figure-S3B**). The gene markers of the recurrent glioma states are enriched for similar ontologies to those seen for primary glioma states (**Figure-S3C** and **Supplementary Table-3**), showed similar patterns of enrichment for the previously presented gene sets in **Figure-S1D** (Figure**-S3D**), and displayed comparable gene marker expression (**Figure-S3F**). These results demonstrate that post-treatment recurrent glioma states closely resemble states observed in the primary pre-treatment setting. Indeed, Pearson correlation analysis demonstrates that corresponding states were positively correlated (**Figure-S3E**). The correlation patterns reveal that gl_Mes1 and gl_Mes2 are positively correlated with each in the primary and recurrent settings. This is also seen with gl_PN1 and gl_PN2, as well as gl_Pro1 and gl_Pro2. We therefore contend that a view of primary and recurrent glioma states may benefit from simplification and embrace a viewpoint that primary and recurrent glioma states can be classified as progenitor-like/proneural (gl_PN1 and gl_PN2), astrocyte-like/mesenchymal (gl_Mes1 and gl_Mes2), and proliferative (gl_Pro2 and gl_Pro1) states. A select set of markers of both primary and recurrent GBM states is provided in **Figure S4I**. Assigning cell cycle scores using Seurat cell-cycle score assignment reveals that gl_Pro1 has the majority of cells in G2M phase, whilst gl_Pro2 has the majority of cells in S phase **Figure S4I**. Integration of both primary and recurrent glioma nuclei shows cells from primary and recurrent samples overlap in the UMAP space, and that this overlap is seen for all 6 GBM states (**Figure S5G**).

While the transcriptional signatures of glioma are relatively well defined, the spatial distribution of these glioma states is less well understood. Given the marked difference in cellular composition between the cortex and the deeper (typically more heavily infiltrated) white matter, and the highly cellular tumor core, we asked if these different anatomic regions harbor distinct glioma states. In other words, we posited that the cellular microenvironment of glioma influences glioma states. Specifically, we hypothesized that we would find more glioma cells that resemble astrocytes (astrocyte-like/mesenchymal glioma) or neurons (progenitor-like - specifically gl_PN2) in the cortical margins. To address this question, we examined the expression of select combinations of glioma state transcripts using in situ hybridization (ISH) across the cellular tumor and the infiltrated cortical margin. We used probes to detect PTPRZ1 (high in glioma), CLU (high in astrocytes and astrocyte-like/mesenchymal glioma), SOX2 (high glioma), NOVA1 (high in progenitor-like/proneural glioma), and MEG3 (high in neurons and progenitor-like/proneural glioma - gl_PN2) in the cellular core and overlying infiltrated cortical margin in 5 cases of primary GBM (**Figure-S5A, C**). We found that significantly higher proportion of PTPRZ1+ glioma cells co-expressed CLU in the cortex versus the core (**Figure-S5B**). Similarly, we found that significantly higher proportion of SOX2+NOVA1+MEG3+ glioma cells in the cortex versus the core (**Figure-3D**). These findings indicate that the different glioma states have distinct distributions throughout the landscape of glioma and suggest that local tissue cellular composition and perhaps other microenvironmental influences can affect glioma states. We note that astrocyte-like/mesenchymal glioma states were negatively correlated with proliferative states. Consistent with this result, our ISH findings demonstrated a significantly smaller proportion of CLU+ cells that co-expressed TOP2A (mean=31.71388837%, Standard deviation = 15.73850618, one-tailed t-test p= 0.000249641, n=5, **Figure S5E-F**).

Comparison between primary and recurrent glioma pairs

Not surprisingly, the recurrent tumors did not show identical chromosomal CNVs with their primary counterparts. While TB5014 retained the CNV of the TB4916 (gain of 7, loss of 10 and 14) and acquired additional alterations including gains in chromosomes 19 and 20 (**Figures S2F and S4A)**, TB5053 showed a complex gains and losses across multiple chromosomes (**Figures S2G and S4B)**.

In the main text, we note that gl_PN1 is depleted from our recurrent GBM samples (**Figure 2A**). This is consistent with the literature [6], since the Verhaak classical subtype resembles our gl_PN1, which showed positive enrichment scores of the Verhaak's classical gene set. Of the non-neoplastic cell types, OPCs were depleted in recurrence. This may be explained by the fact the OPCs are the proliferative cell type in the brain and glioma treatment with chemotherapy and radiotherapy depletes proliferative cells, as have been previously demonstrated [62].

Survival analysis of glioblastoma state signatures

To determine if enrichment of our glioma states has prognostic relevance we performed GSEA of the top gene markers for each glioma state, both primary and recurrent gene sets (**Supplementary Table 2 and 3**), using the TCGA-GBM [63] and the Chinese Glioma Genome Atlas (CGGA) databases [51, 52]. 318 IDH-WT samples (TCGA - 139, CGGA - 179) were used for survival analysis. Kaplan-Meyer survival curves are shown on **Figure-S6A-B** for primary and recurrent glioma gene sets, respectively. The astrocyte-like/mesenchymal signatures Mes1 in primary and Mes2 in recurrent glioma significantly predicted poor survival. These two gene sets share 46 genes including CD44, and CHI3L1, which may explain the effect of primary gl_Mes1 and recurrent gl_Mes2 with survival.

Analysis of low-grade glioma and epilepsy cases

To sample states of myeloid cells and astrocytes across different disease states, we chose to analyze the microenvironment of low-grade glioma (LGG) and temporal lobe epilepsy. We conducted snRNAseq on 6 cases: two IDH-mutant oligodendroglioma (TB3652 & TB3926), one IDH-mutant astrocytoma (TB4100), and three temporal lobe epilepsies (TB4189, TB4437, & TB4957). We identified 970, 1154, 1036 nuclei for LGG cases TB3652, TB3926, and TB4100, respectively. We identified CNVpos nuclei using a combination of chromosomal CNV, clustering, and tumor marker expression as shown in **Figure S7**. Cases TB3652 and TB3926 had typical chromosome 1p and 19q codeletions (**Figure S7A, D**), and harbored 817 and 942 CNVpos nuclei, respectively (**Figure S7B, E**). The tumor nuclei expressed tumor markers SOX2, EGFR, and PTPRZ1, and/or OPC markers DSCAM and TNR; myeloid cells expressed a CD74, C3, ITGAX (CD11c), ITM2B, and/or HLA-B; while oligodendrocytes expressed MBP and MOG (**Figure S7C, F**). 382 CNVpos nuclei were found in case TB4100, which did not harbor CNVs across most cells, and CNVpos nuclei were identified by clustering and marker expression as noted above. Of the epilepsy cases, we identified 2558, 179, and 138 nuclei in cases TB4189, TB4437, and TB4957, respectively. **Figure S8A-C** show marker expression in cases TB4437, TB4189, and TB4957, where markers of astrocytes (GFAP, AQP4, SLC1A2, SLC1A3), neurons (RBFOX3, MEG3, GAD1, and SLC17A6), myeloid cells (CD74, ITGAX, C3, ITM2B), oligodendrocytes (MBP, MOG, OPALIN, and CNP), and OPCs (DSCAM, TNR, SOX2, and PDGFRA). The CNVneg nuclei from all LGG and epilepsy cases were combined with those from primary and recurrent IDH-WT GBM and were analyzed as presented in the section below (myeloid cells) and main text (astrocytes).

Astrocyte subclusters and prognostic relevance

Based on unbiased shared nearest neighbor clustering of all astrocytes, we identified multiple sub-clusters that share in gene expression. These are projected in the tSNE space in **figure S9A**.

Select markers are shown in violin plots and gene expression heatmap in **figure S9B-C**, respectively. The markers of the subclusters based on gene-wise Wilcox test are provided in **supplementary table 5**. Generally, all sub-clusters showed expression of canonical astrocyte gene GFAP, as well as variable levels of expression of other astrocyte genes including AQP4, S100B, GPC5, SLC1A2, SLC4A4, and ATP1B2. Examples of gene expression are show in violin plots in (**Figure-S9B**). Compared to other sub-clusters, sub-clusters 1 and 8 showed significantly higher expression of many of the protoplasmic genes such as SLC1A2 and GLUL, whereas sub-clusters 6 and 7 showed significantly higher expression of reactive genes including CLU, LGALS3, and VIM. One sub-cluster was characterized by expression of ribosomal genes (sub-cluster 3). While an over-representation of ribosomal genes may be a sign of low-coverage, we filtered cells with less than 10,000 reads and 400 detected genes and thus interpret this ribosomal gene signature as of potential biologic relevance - future studies will investigate that further. Another sub-cluster had a hypoxic signature characterized by expression of VEGFA and HILPDA (sub-cluster 5). Sub-clusters 2, 4, and 5 were each characterized by expression of genes associated with other lineages including MEG3 and PDFGRA (Sub-cluster 5), GRIA1, TNR (Sub-cluster 4), and PLP1 as well as SNAP25 (Sub-cluster 2). We previously showed that reactive astrocytes can mis-express non-astrocyte lineage genes in neurodegenerative setting [18]. Sub-cluster 0 did not show many differentially increased genes and is thus considered a baseline sub-cluster (**Figure-S9B-C)**.

Astrocyte sub-clusters were differentially distributed between primary glioma, recurrent glioma, LGG, and epilepsy (**Figure-S9D)**. Specifically, sub-clusters 2 and 3 were mostly represented in primary glioma, sub-clusters 6 and 7 in recurrent glioma, sub-cluster 1 in epilepsy, and sub-cluster 4 in LGG. Moreover, sub-cluster 0 was represented in all glioma conditions, mostly in primary and recurrent gliomas and least in epilepsy, while sub-cluster 8 was represented in recurrent glioma and epilepsy (**Figure-S9D**).

Because sub-cluster 6 shows significantly higher expression of several genes previously associated with poor glioma outcomes, including CD44, CHI3L1, LGALS3, CLU and APOE [64-66], we asked if the gene signature for sub-cluster 6 (**Supplementary Table-5**) has prognostic relevance in survival data associated with publicly available datasets. We performed GSEA of the top gene markers for astrocyte sub-clusters, using the TCGA-GBM [63] and the Chinese Glioma Genome Atlas (CGGA) databases [51, 52]. 318 IDH-WT samples (TCGA - 139, CGGA - 179) were used for survival analysis. Enrichment of the gene signature for astrocyte sub-cluster 6 is significantly associated with poor survival (p=0.007 negative log rank test - **Figure-S9E**) on univariate analysis by the Kaplan-Meier method. To investigate the relationship between survival and enrichment of each astrocyte sub-cluster top gene markers, we determined the hazard ratio of death given enrichment of each gene set using the cox proportional hazard regression model, accounting for age as a covariate (**Figure-S9F**). Only positive enrichment of sub-cluster 6 genes was significantly associated with increased hazard ratio of death by more than 1.5 folds (confidence interval 1.08-2.20, p.value 0.015 **Figure-S9F**), an effect independent of age.

Based on the overlap of gene expression between astrocyte clusters, and the resemblance to known astrocyte phenotypes we derived three gene sets (**Supplementary Table-5**), which represent three major astrocyte states (protoplasmic, reactive-1, and reactive-2), and then re-clustered astrocyte nuclei using Ward D2 hierarchical clustering on the Manhattan distance of the enrichment scores (overlaid on the 3D tSNE plots in **Figure-S9G-I**), into a protoplasmic cluster (Ast1), and two reactive clusters (Ast2 and Ast3 – as described in the main text **Figure 1C**). These astrocyte states correspond very closely to those used in the compositional PCA (**Figure-S9J** - see methods).

Analysis of myeloid cell states

Myeloid cells have been implicated in modulating glioma migration, infiltration, and progression [67]. We identified 5925 nuclei we classified as myeloid cells. Unbiased clustering revealed 8 subclusters which we then used to assign the specific myeloid lineages. We merged clusters with similar enrichment scores of gene sets representing microglia-derived tumor-associated macrophages (mgTAM), monocyte-derived TAMs (moTAM), proliferative TAMs (prTAMs), and T-cells as described in [15] **Figure S10A** - see methods (section on PCA analysis). The enrichment of these gene sets in the final myeloid states is provided in **Figure S10E**. A subset of myeloid cells showed mixed enrichment scores across mgTAM, moTAM, and dendritic cells, and were considered baseline (referred to as Myel1). Overall, we classified 2678, 1346, 1364, 360, and 177 nuclei as Myel1, moTAM, mgTAM, prTAM, and T-cells, respectively, and these are shown in 3D tSNE space in **Figure S10A**. Myel1 state showed higher expression of SAT1, CEBPD, and GLUL (**Figure S10C**-top row). moTAM showed highest expression of CD163, MS4A4E, NHSL1, FMN1, and MSR1 (**Figure S10C**-2nd row). mgTAM showed highest expression of SORL1, RIN3, ITGAX, HS3ST4, and FRMD4A (**Figure S10C**-3rd row). prTAMs showed highest expression of CST3, MEF2A, DBI, PLXDC2, and DOCK4 (**Figure S10C**-4th row). Finally, T-cells showed highest expression of CD2, CD247, CD96, FYN, and SKAP1 (**Figure S10C**-5th row). Different myeloid states were accounted for different conditions (**Figure S10B).** While Myel1 was present in Epilepsy, primary and recurrent GBM, mgTAM was the main state found in LGG, but was also in primary and recurrent GBM. moTAM, T-cells, and prTAM were found in primarily in recurrent GBM (**Figure S10D)**. The gene-wise DGE between myeloid states and the myeloid state markers are provided in **Supplementary Table-6**.

The spatial landscape of glioma associated tissue-states in primary and recurrent GBM

To understand the spatial landscape of primary and recurrent glioma, we mapped the distribution of our “tissue-state” signatures in space in primary and recurrent GBM. First, we tested one of our cases that we utilized for snRNAseq (PO2) and took 48 localized biopsies that we analyzed using plate-seq [49]. Immunofluorescence of frozen sections taken prior to analysis revealed a cellular DAPI-dense glioma core and a NeuN rich cortical margin (**Figure S12A**). We conducted GSEA analysis of our tissue-state signatures in the RNAseq data from the localized biopsies and mapped that against the location of the biopsies (**Figure S12B**). Tissue-state C signature was highest in the core, compared to tissue-state A signature, which was highest in the cortical margin. Tissue-state B signature showed a more patchy distribution with foci of enrichment in both the core and margin. Interestingly, the intermediate region between the core and cortex, showed mixed enrichment across all three tissue-states. This data highlights the anatomic localization of tissue-state signatures and underscores the heterogeneous patterns in the intermediate non-cortical “margin” region. Next, to assess the generalizability of these results, we conducted the same enrichment analysis on a dataset of MRI-localized primary and recurrent GBM samples [3] (**Figure S12C**). As expected, control brain samples showed high enrichment of tissue-state A signature while the contrast-enhancing samples in primary GBM showed high enrichment of tissue-state C signature. Notably, the contrast enhancing samples from post-treatment recurrent GBM showed highest enrichment of tissue states B and C, indicating a mixture of recurrent tumor and treatment effect. The FLAIR+ primary GBM samples showed a wide distribution of enrichment of all three tissue types while the FLAIR+ recurrent GBM samples showed enrichment of tissue-states A and B, with only a few samples showing enrichment of tissue state C. This data indicates that the non-enhancing margins of recurrent GBM samples predominantly represent reactive/gliotic brain tissue with relatively low levels of tumor infiltration, whereas the non-enhancing margins of primary GBM can contain a wider range of pathological features, including regions of abundant glioma infiltration.

References

1. Quail, D.F. and J.A. Joyce, *The Microenvironmental Landscape of Brain Tumors.* Cancer Cell, 2017. **31**(3): p. 326-341.
2. Yuan, J., et al., *Single-cell transcriptome analysis of lineage diversity in high-grade glioma.* Genome Med, 2018. **10**(1): p. 57.
3. Gill, B.J., et al., *MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma.* Proc Natl Acad Sci U S A, 2014. **111**(34): p. 12550-5.
4. Puchalski, R.B., et al., *An anatomic transcriptional atlas of human glioblastoma.* Science, 2018. **360**(6389): p. 660-663.
5. Jin, X., et al., *Targeting glioma stem cells through combined BMI1 and EZH2 inhibition.* Nat Med, 2017. **23**(11): p. 1352-1361.
6. Wang, Q., et al., *Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment.* Cancer Cell, 2017. **32**(1): p. 42-56 e6.
7. Wang, L., et al., *The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation.* Cancer Discov, 2019. **9**(12): p. 1708-1719.
8. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014. **344**(6190): p. 1396-401.
9. Neftel, C., et al., *An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma.* Cell, 2019. **178**(4): p. 835-849 e21.
10. Levitin, H.M., et al., *De novo gene signature identification from single-cell RNA-seq with hierarchical Poisson factorization.* Mol Syst Biol, 2019. **15**(2): p. e8557.
11. Darmanis, S., et al., *Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma.* Cell Rep, 2017. **21**(5): p. 1399-1410.
12. Bhaduri, A., et al., *Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma.* Cell Stem Cell, 2020. **26**(1): p. 48-63.e6.
13. Lee, J.K., et al., *Spatiotemporal genomic architecture informs precision oncology in glioblastoma.* Nat Genet, 2017. **49**(4): p. 594-599.
14. Barthel, F.P., et al., *Longitudinal molecular trajectories of diffuse glioma in adults.* Nature, 2019. **576**(7785): p. 112-120.
15. Pombo Antunes, A.R., et al., *Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals macrophage competition and specialization.* Nat Neurosci, 2021. **24**(4): p. 595-610.
16. Weng, Q., et al., *Single-Cell Transcriptomics Uncovers Glial Progenitor Diversity and Cell Fate Determinants during Development and Gliomagenesis.* Cell Stem Cell, 2019. **24**(5): p. 707-723 e8.

17. Henrik Heiland, D., et al., *Tumor-associated reactive astrocytes aid the evolution of immunosuppressive environment in glioblastoma.* Nat Commun, 2019. **10**(1): p. 2541.
18. Al-Dalahmah, O., et al., *Single-nucleus RNA-seq identifies Huntington disease astrocyte states.* Acta Neuropathol Commun, 2020. **8**(1): p. 19.
19. Wojtas, A.M., et al., *Astrocyte-derived clusterin suppresses amyloid formation in vivo.* Mol Neurodegener, 2020. **15**(1): p. 71.
20. Schellenberg, G.D. and T.J. Montine, *The genetics and neuropathology of Alzheimer's disease.* Acta Neuropathol, 2012. **124**(3): p. 305-23.
21. Liu, L., et al., *CHI3L2 Is a Novel Prognostic Biomarker and Correlated With Immune Infiltrates in Gliomas.* Front Oncol, 2021. **11**: p. 611038.
22. Fazi, B., et al., *The Expression of the Chemokine CXCL14 Correlates with Several Aggressive Aspects of Glioblastoma and Promotes Key Properties of Glioblastoma Cells.* Int J Mol Sci, 2019. **20**(10).
23. Cenini, G., A. Lloret, and R. Cascella, *Oxidative Stress in Neurodegenerative Diseases: From a Mitochondrial Point of View.* Oxid Med Cell Longev, 2019. **2019**: p. 2105607.
24. Lin, H. and S.D. Peddada, *Analysis of compositions of microbiomes with bias correction.* Nat Commun, 2020. **11**(1): p. 3514.
25. Lê, S., J. Josse, and F. Husson, *FactoMineR: A Package for Multivariate Analysis.* Journal of Statistical Software, 2008. **25**(1): p. 1-18.
26. Wang, L., et al., *The evolution of alternative splicing in glioblastoma under therapy.* Genome Biol, 2021. **22**(1): p. 48.
27. Grube, S., et al., *Overexpression of fatty acid synthase in human gliomas correlates with the WHO tumor grade and inhibition with Orlistat reduces cell viability and triggers apoptosis.* J Neurooncol, 2014. **118**(2): p. 277-287.
28. Garcia Corrales, A.V., et al., *Fatty Acid Synthesis in Glial Cells of the CNS.* Int J Mol Sci, 2021. **22**(15).
29. Tabernero, A., et al., *Neuronal differentiation is triggered by oleic acid synthesized and released by astrocytes.* J Neurochem, 2001. **79**(3): p. 606-16.
30. Wei, H., et al., *De novo Lipogenesis in Astrocytes Promotes the Repair of Blood-Brain Barrier after Transient Cerebral Ischemia Through Interleukin-33.* Neuroscience, 2022. **481**: p. 85-98.
31. Mathys, H., et al., *Single-cell transcriptomic analysis of Alzheimer's disease.* Nature, 2019. **570**(7761): p. 332-337.
32. Stockwell, B.R., et al., *Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease.* Cell, 2017. **171**(2): p. 273-285.
33. Yasumoto, Y., et al., *Inhibition of Fatty Acid Synthase Decreases Expression of Stemness Markers in Glioma Stem Cells.* PLoS One, 2016. **11**(1): p. e0147717.
34. Zhao, W., et al., *Fatty acid synthase: a novel target for antiglioma therapy.* Br J Cancer, 2006. **95**(7): p. 869-78.
35. Lee, J.A., et al., *Lipid metabolism in astrocytic structure and function.* Semin Cell Dev Biol, 2021. **112**: p. 123-136.
36. Zhou, Y., et al., *Inhibition of fatty acid synthase suppresses neovascularization via regulating the expression of VEGF-A in glioma.* J Cancer Res Clin Oncol, 2016. **142**(12): p. 2447-2459.
37. Krishnaswami, S.R., et al., *Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons.* Nat Protoc, 2016. **11**(3): p. 499-524.
38. Petukhov, V., et al., *dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments.* Genome Biol, 2018. **19**(1): p. 78.
39. McCarthy, D.J., et al., *Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R.* Bioinformatics, 2017. **33**(8): p. 1179-1186.

40. Lun, A.T., D.J. McCarthy, and J.C. Marioni, *A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor.* F1000Res, 2016. **5**: p. 2122.
41. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.
42. Hafemeister, C. and R. Satija, *Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.* Genome Biol, 2019. **20**(1): p. 296.
43. Tickle, T., et al., *inferCNV of the Trinity CTAT Project.* 2019.
44. Levitin, H.M., J. Yuan, and P.A. Sims, *Single-Cell Transcriptomic Analysis of Tumor Heterogeneity.* Trends Cancer, 2018. **4**(4): p. 264-268.
45. Therneau, T.M. and P.M. Grambsch, *Modeling Survival Data: Extending the Cox Model*. 2000, New York: Springer.
46. Therneau, T.M., *A Package for Survival Analysis in R*. 2020.
47. Kuhn, M., *Building Predictive Models in R Using the caret Package.* {Journal of Statistical Software, Articles, 2008. **28**.
48. Zhao, W., et al., *Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq.* Genome Med, 2021. **13**(1): p. 82.
49. Bush, E.C., et al., *PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens.* Nat Commun, 2017. **8**(1): p. 105.
50. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
51. Zhao, Z., et al., *Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas.* Sci Data, 2017. **4**: p. 170024.
52. Liu, X., et al., *A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas.* Neuroimage Clin, 2018. **20**: p. 1070-1077.
53. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
54. Hanzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data.* BMC Bioinformatics, 2013. **14**: p. 7.
55. Reimand, J., et al., *g:Profiler-a web server for functional interpretation of gene lists (2016 update).* Nucleic Acids Res, 2016. **44**(W1): p. W83-9.
56. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
57. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
58. Bankhead, P., et al., *QuPath: Open source software for digital pathology image analysis.* Sci Rep, 2017. **7**(1): p. 16878.
59. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nat Biotechnol, 2018. **36**(5): p. 411-420.
60. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.* Cancer Cell, 2010. **17**(1): p. 98-110.
61. Gobin, M., et al., *A DNA Repair and Cell-Cycle Gene Expression Signature in Primary and Recurrent Glioblastoma: Prognostic Value and Clinical Implications.* Cancer Res, 2019. **79**(6): p. 1226-1238.
62. Lopez, K.A., et al., *Convection-enhanced delivery of topotecan into a PDGF-driven model of glioblastoma prolongs survival and ablates both tumor-initiating cells and recruited glial progenitors.* Cancer Res, 2011. **71**(11): p. 3963-71.

63. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma.* Cell, 2013. **155**(2): p. 462-77.
64. Pelloski, C.E., et al., *YKL-40 expression is associated with poorer response to radiation and shorter overall survival in glioblastoma.* Clin Cancer Res, 2005. **11**(9): p. 3326-34.
65. Wu, G., et al., *Expression of CD44 and the survival in glioma: a meta-analysis.* Biosci Rep, 2020. **40**(4).
66. He, X., et al., *Increased LGALS3 expression independently predicts shorter overall survival in patients with the proneural subtype of glioblastoma.* Cancer Med, 2019. **8**(5): p. 2031-2040.
67. Markovic, D.S., et al., *Microglia stimulate the invasiveness of glioma cells by increasing the activity of metalloprotease-2.* J Neuropathol Exp Neurol, 2005. **64**(9): p. 754-62.

Figure legends

Figure 1: Astrocytes in the glioma microenvironment are heterogeneous and prognostically relevant

A) Uniform-manifold approximation and projection (UMAP) graphs showing putative non-neoplastic (CNVneg) from primary glioma, recurrent glioma, low grade glioma (LGG) - and epilepsy (see supplementary data for the analysis of LGG and epilepsy cases). The nuclei are color-coded by lineage (Oligodendrocytes, oligodendrocyte-precursor cells (OPC), neurons, astrocytes, myeloid cells, and endothelial cells). **B**) Dot plots showing normalized expression of select lineage genes (row) in the lineage from A (columns). The size of each circle corresponds to the proportion of the lineage that expresses a given gene. **C**) Three-dimensional tSNE plots showing all astrocyte nuclei color-coded by astrocyte state (Ast1 – protoplasmic astrocytes, Ast2 – reactive astrocytes with misexpression of non-astrocyte lineage genes, and Ast3 – reactive astrocytes with expression of inflammatory genes. **D**) Three-dimensional tSNE plots showing all astrocyte nuclei color-coded by disease condition. **E**) Gene expression dot plots showing select gene marker expression for the astrocyte states. (**F**) Classification of astrocytes states into Huntington disease (HD) like or control (Con) like based on Al-Dalahmah et al. 2020 human astrocyte dataset, using a random forest classifier. The heatmap indicates the proportions of astrocytes classified into each class scaled by astrocytes state (column). **G-H**) Active subnetwork enrichment analysis of KEGG pathways in genes differentially expressed in CNVneg glioma-associated astrocytes compared to CNVpos glioma cells in primary and recurrent IDH-WT glioma. Fold enrichment is represented on the x-axis and the pathways in the y-axis. The pathways are clustered to denote shared genes driving enrichment. The size of the circle per pathway denotes the number of enriched genes, and the negative log10 of the adjusted p.value is represented by color. Pathways enriched in genes significantly higher in astrocytes compared to glioma cells are shown in **G** and include neurodegenerative diseases and oxidative phosphorylation, metabolism including fatty acid metabolism. Pathways enriched in genes significantly higher in glioma cells compared to astrocytes are shown on the **H** and include DNA replication, splicing, and ErbB signaling.

Figure 2: snRNAseq and spatial transcriptomics identify patterns of co-habitation that correlate with survival

A) Bar plots demonstrating the fractional composition of each one of 16 samples analyzed by snRNAseq (8 primary IDH-WT glioma from 7 patients, one case was divided to core and overlying cortex, and 8 recurrent IDH-WT glioblastoma) The first row of bar plots represent the fraction of neoplastic (CNVpos) and non-neoplastic (CNVneg) nuclei. The middle row represents the fraction of the non-neoplastic nuclei contributed by neurons, oligodendrocytes, OPCs, astrocytes, myeloid cells, endothelial cells, and astrocytes. The red arrowhead in the OPC legend box indicate that OPCs were significantly reduced in recurrent GBM as determined by differential abundance analysis. The bottom row represents the fraction of the neoplastic nuclei contributed by proneural/progenitor-like glioma (gl_PN1, gl_PN2), astrocyte-like/mesenchymal glioma (gl_Mes1, gl_Mes2), and proliferative glioma (gl_Pro1, and gl_Pro2). The description of glioma states is provided in the supplementary results. The red arrowhead in the gl_Mes2 legend box and the black star in gl_PN1 indicate that gl_Mes2 and gl_PN1 were significantly increased and reduced in recurrent GBM, respectively, as determined by differential abundance analysis. **B**) Principal component analysis of the fractional composition matrix of 19 samples encompassing eight primary and eight recurrent gliomas plus three epilepsy samples. The tissue composition matrix consists of the percentage of nuclei per each tissue state. Immune cell states are: mgTAMs (microglia-derived Tumor-associated macrophages), moTAM (monocyte-derived TAMs), prTAM (proliferative TAM), Myel1 (baseline myeloid cells), and T cells. Astrocyte states include baseline (protoplasmic) astrocytes (Ast1), reactive CD44+ astrocytes (Ast3), and reactive astrocytes with expression of non-astrocyte genes (Ast2) – see text and supplementary results for additional description of these cell states. CNVpos represents the total percentage of all tumor states per sample. Individual tumor states were not used in PCA calculation, rather they were used as used supplementary quantitative variables and their coordinates were predicted from the PCA analysis – see methods. **C**) Kaplan-Meier survival plot graphing survival in the combined TCGA and CGGA RNAseq datasets. The samples were classified based on enrichment of gene signatures of microenvironment states correlated with PC2 into positive or negative enrichment. Statistical significance was computed using the log rank test. **D**) Representative plots showing enrichment scores of gene signatures of astrocyte state and select glioma states projected in space in a GMB-infiltrated sample analyzed by spatial transcriptomics. **E**) Heatmap showing the Pearson Correlation of cell-type and cell-state transcriptional signatures in nine GBM-infiltrated samples analyzed by spatial transcriptomics. Hierarchical clustering of the distance matrix derived from the correlations showed three clusters.

Figure 3: Tissue composition analysis defines "tissue states" recapitulated in validation bulk RNAseq dataset glioma and microenvironmental states better predict prognosis

A) Dendrogram of hierarchically clustered glioma and epilepsy samples based on Minkowski sample distance analysis drawn from the fractional composition matrix (see Figure 4A). Three clusters were identified and are color-coded on the dendrogram in black (Tissue-state C), red (Tissue-state B), and green (Tissue-state A). The condition (primary, recurrent and epilepsy) is indicated in the top bar below. The proportion of neoplastic nuclei (CNVprop) is indicated in the bottom bar. **B**) Three-dimensional scatter plot showing the samples in **A** projected in the first three principal component loadings – see figure 4B for PCA analysis. The samples are color-coded by cluster designation as in A. **C**) Bulk RNAseq samples from 92 primary and recurrent IDH-WT glioblastoma samples projected in the space represented by the enrichment of the three gene signatures characteristic of the tissue states in B. The samples were clustered on the Minkowski distance of the enrichment scores into three clusters A-C and are color-coded as such. **D**) Gene ontology term analysis of the genes uniquely and differentially expressed in each of the clusters in **C**. The bar plots are color coded as per the clusters in C. KEGG, REACTOME, or Biological Process GO pathways are shown in the y-axis. Negative log10 of the adjusted p-value is shown on the x-axis. **E**) Normalized expression of select genes characteristic of each of the clusters projected onto the compositional-signature enrichment score space shown in C. Red denotes high expression, and grey denotes low expression. NeuN (RBFOX3) is highest in the samples of Cluster A. CD68 is highest in the samples of cluster B. SOX2 is highest in the samples of clusters B and C. MKI67 is highest in the samples of cluster C. **F**) Quantification of histological cellularity analysis (far left) as well as immunohistochemistry labeling indices of (from left to right) SOX2, KI67, CD68, and NeuN. The labeling index is shown on the y-axis. Note that the y-axis for the cellularity graph is total cellularity normalized to the most cellular sample. The sample clusters are labeled (A-C) as in **C**. p values were calculated using Kruskal-Wallis test and are indicated on the graphs. N=8 for cluster A, 25 for cluster B, and 12 for cluster C. **G**) Pre-ranked Gene Set Enrichment Analysis (GSEA) comparing tissue state B bulk RNAseq samples with tissue states A & C samples for 4 sub-lineages: Ast3, moTAM, gl_Mes2, and T-cells. Marker genes for each sub-lineage were used as the gene set for each analysis. Normalized Enrichment Score (NES) is displayed, along with p-values and FDR-adjusted q-values. **H**) Cox proportional hazard ratio of survival in the combined TCGA and CGGA IDH-WT GBM dataset given enrichment of each of the tissue state signatures. Age, gender, and MGMT status are included as co-variates in the model. The p values are shown on the left, bars indicate confidence intervals (also noted on the right). Enrichment of each geneset was categorized as negative or positive. **I**) Boxplots of the tissue state B normalized enrichment scores in the Wang, L. et al. 2021 paired primary and recurrent GBM dataset. Paired samples are denoted by connected points. Paired t-test – one-tailed. The p value is indicated.

Figure 4: Metabolic pathways drive targetable tissue state signatures

A) Heatmap displaying scaled enrichment scores for all KEGG pathways across all PLATE-seq samples. The heatmap is grouped by tissue state (cluster A, B, C), annotated by the horizontal bar at the top. Hierarchical clustering was performed on the rows (pathways), demonstrating cluster-specific metabolic programs. **B**) Bar plot displaying scaled ssGSEA scores for select KEGG metabolic programs from A. Bar plots represent mean scaled ssGSEA score ± standard error for each of the three clusters for a given pathway. **C**) Representative example showing a heatmap displaying mean lineage-specific scaled normalized expression of genes in the GO: Biological Process - Fatty Acid Biosynthesis gene set – which was most enriched in tissue state B. Note the expression of the rate-limiting enzyme FASN is highest in astrocytes and glioma cells. **D**) Scheme of *in vitro* and *ex vivo* FASN perturbation studies. **E**) Volcano plot showing the log2 fold change (x-axis) and log10 p value (y-axis) of differentially expressed genes in astrocytes treated with Cerulenin (5mg/ml) versus control – Upper panel. Lower panel shows KEGG and Reactome pathway enrichment analysis with the terms indicated on the y-axis, and the log10 p value on the x-axis. The sign of the log10 p value indicates the direction of change (i.e. negative = reduced in Cerulenin treatment). **F**) Volcano plot showing the log2 fold change (x-axis) and log10 p value (y-axis) of differentially expressed genes in GBM slice cultures treated with Cerulenin (5mg/ml) versus control – Upper panel. Lower panel indicate GSEA plots of pre-ranked enrichment of the genes increased in astrocytes treated with Cerulenin (left) and tissue state B signature (right). The normalized enrichment scores (NES), p value (p), and adjusted p value (q) are indicated.

Figure S1: snRNAseq-derived transcriptional states and lineage of putative neoplastic nuclei from primary IDH-wildtype GBM samples

A) Outline of Analytic Design: T 2/FLAIR and post-contrast T1 MRI sequences of a glioblastoma showing the classic radiological appearance of a glioblastoma (Case PO2); with a ring enhancing mass (red star) with surrounding increased FLAIR signal (green star). The tumor was resected and banked (frozen). Nuclei are extracted from frozen tissue and are subjected to droplet based single nuclei RNA sequencing using the 10X chromium platform. The resultant barcoded cDNA is then sequenced and analyzed. Analyses performed include identification of putative neoplastic cells by identifying cells with inferred copy number variations (CNV), clustering, differential gene expression (DGE), and gene set enrichment analysis (GSEA). Scale bars = 50 um. **B)** Uniform-manifold approximation and projection (UMAP) graphs showing putative neoplastic (CNVpos) and non-neoplastic (CNVneg) nuclei from the seven primary IDH-wildtype glioma cases selected for analysis indicated by subpanels b1-b7. **C**) UMAP plot showing all putative CNVpos (**C**) nuclei from the seven primary glioma cases aligned and projected in shared UMAP spaces. The nuclei are color-coded by glioma state: Oligodendrocyte-progenitor-like (proneural - gl_PN1), Neural-progenitor-like (proneural - gl_PN2), Mesenchymal/astrocyte like (gl_Mes1 and gl_Mes2), and proliferative (gl_Pro1 & gl_Pro2). **D**) Geneset enrichment analysis (GSEA) of selected genesets from Verhaak et al. 2009 (v), Gobin M et al 2019, Gill et al 2014, Wang et al. 2019 (W), and Neftel et al. 2019 (N) showing enrichment of genes specific for states described in the literature in our described glioma states. **E**) Gene ontology (GO) term enrichment analysis (KEGG and REACTOME pathways and biological process GO) of the major terms enriched in glioma state top gene markers. The bars represent the negative log10 of the false discovery rate adjusted p.value, and are color-coded as in **C**.

Figure S2: Histopathologic characterization of glioma cases, Identifying neoplastic nuclei

Large scale chromosomal copy number alterations were inferred from RNA expression using InferCNV R package (see methods for details). The heat maps show gains (red) and losses (blue) in case PA1 (**A**), PA2 (**B**), PA3 (**C**), PO1 (**D**), PO2 (two samples – core and margin) (**E**), TB4916 (**F**), and TB4718 (**G**). Representative Hematoxylin and Eosin-stained section of the brain tissue used for single nuclei RNAseq of the first five cases (**H**). Some cases showed clear infiltration with glioma cells PA1, PA2, PA3, and PO2_c, PO2_2. Cases PO1 and PO2_m showed no clear evidence of cellular tumor.

Figure S3 snRNAseq-derived transcriptional states and lineage of putative neoplastic nuclei from post-treatment recurrent IDH-wildtype GBM samples

A) Uniform-manifold approximation and projection (UMAP) graphs showing putative neoplastic (CNVpos) and non-neoplastic (CNVneg) nuclei from the eight post-treatment recurrent IDH-wildtype glioblastoma cases. **B**) UMAP plot showing all putative CNVpos nuclei from the eight recurrent glioma cases aligned and projected in shared UMAP spaces. The nuclei are color-coded by glioma state: Oligodendrocyte-progenitor-like (proneural - gl_PN1), Neural-progenitor-like (proneural - gl_PN2), Mesenchymal/astrocyte like (gl_Mes1 and gl_Mes2), and proliferative (gl_Pro1 & gl_Pro2). **C**) Gene ontology (GO) term enrichment analysis (KEGG and REACTOME pathways and biological process GO) of the major terms enriched in glioma state top gene markers. The bars represent the negative log10 of the false discovery rate adjusted p.value and are color-coded as in B. **D**) Geneset enrichment analysis (GSEA) of selected genesets from Verhaak et al. 2009, Gobin M et al 2019, Gill et al 2014, and Neftel et al. 2019 showing enrichment of genes specific for states described in the literature in our described glioma states. **E**) Correlation heatmap between glioma states in primary and post-treatment recurrent GBM based on expression on glioma state marker genes. The size and color of the circles denote the strength of correlation. **F**) Gene expression dot plots showing select gene marker expression in glioma states.

Figure S4: CNV analysis of recurrent glioma samples

Large scale chromosomal copy number alterations were inferred from RNA expression using InferCNV R package. The heat maps show gains (red) and losses (blue) in case TB5014 (**A**),

TB5053 (**B**), TB3864 (**C**), TB4898 (**D**), TB8762 (**E**), TB4416 (**F**), and TB4027 (**G**), and TB3966 (**H**). **I**) Dotplot showing expression of select set of markers of both primary and recurrent glioma states. The proportion of each glioma state in cell cycle phases as determined by Seurat cell-cycle scoring is shown on the bottom.

Figure S5: The spatial landscape of glioma states across the cellular tumor and cortical

A) Confocal images showing optical sections of in situ hybridization for PTPRZ1 and CLU in the core (upper row) and cortex (lower row). The pial surface is outlined (lower row). High-power images of the insets show that PTPRZ1+ CLU+ cells (arrows) are more abundant in the cortex, while PTPRZ1+CLU- (arrowheads) are more numerous in the core. scale bars = 20 µm. M.V: Microvascular proliferation **B)** Quantification of PTPRZ1 and CLU expression across the core (orange boxplot) and cortex (green boxplot). The data is shown as boxplots, with the bar indicating the median. Paired t-test, N=5. The p value is indicated. **C**) Confocal images showing optical sections of in situ hybridization for NOVA, SOX2, and MEG3 in the core (upper row) and cortex (lower row). The pial surface is outlined (lower row). High-power images of the insets show that NOVA1+SOX2+MEG3+ cells (arrows) are more abundant in the cortex, while MEG3- cells (arrowheads) are more numerous in the core. scale bars = 20 µm. **C**) Quantification of MEG3+NOVA1+SOX2+ cells as a proportion of all tumor cells (SOX2+ and/or NOVA1+) across the core (orange boxplot) and cortex (green boxplot). The data is shown as boxplots, with the bar indicating the median. Paired t-test, N=5. The p value is indicated. **E)** Confocal images showing optical sections of in situ hybridization for TOP2A and CLU in the GBM infiltrated tissue. Arrows indicate CLU+TOP2A+ cells, and arrowheads indicate CLU+TOP2A- cells. scale bar = 20 µm. **F)** Quantification of TOP2A and CLU expression. The percentage of TOP2A+CLU+/CLU+ cells is shown as a boxplot. One-sample t-test, N=5. *=p value < 0.001. G) Integration of primary and recurrent GBM CNVpos nuclei color-coded by glioma state and condition

Figure S6: Prognostic relevance of glioma-state signatures

Kaplan-Meyer plots showing the survival curves of IDH-WT glioblastoma cases from the combined TCGA and CGGA dataset stratified by enrichment of each of the glioma-state genesets of primary (**A**) and recurrent (**B**) glioma datasets. The p values are shown.

Figure S7: Analysis of Low-grade glioma samples using single nucleus RNAseq

Large scale chromosomal copy number alterations were inferred from RNA expression of cases TB3652 (**A**), TB3926 (**D**) – both IDH1-mutant oligodendrogliomas, and TB4100 (**G**) – IDH-mutant astrocytoma. Uniform manifold approximation and projection (UMAP) plots of the three cases are shown in panels **B**, **E**, and **H**, color-coded by copy number alteration status. Gene expression UMAPs showing markers of tumor cells (PTPRZ1, EGFR, SOX2, TNR, and DSCAM), immune cells (CD74, C3, HLA-B, ITGAX, ITM2B), and oligodendrocytes (MBP, MOG) of cases TB3652, TB3926, and TB4100 in panels **C, F,** and **I**, respectively.

Figure S8: Analysis of Epilepsy samples using single nucleus RNAseq

A-C) Uniform-manifold approximation and projection (UMAP) graphs plots showing normalized gene expression of select lineage markers for cases TB4437 (**A**), TB4189 (**B**), TB4957 (**C**). The markers include astrocyte markers (GFAP, AQP4, SLC1A2, and SLC1A3), neuron makers (RBFOX3, MEG3, GAD1, SLC17A6), myeloid markers (CD74, C3, ITGAX, ITM2B), oligodendrocyte markers (MBP, MOG, OPALIN, CNP), and OPC markers (PDGFRA, DSCAM, TNR, and SOX2).

Figure S9: Related to Figure 1. Astrocytes sub-clusters

A) tSNE plot (three-dimensional) showing astrocytes color-coded by sub-cluster. **B**) Violin plots showing expression of select astrocyte markers in the unbiased sub-clusters. Note all sub-clusters show expression of GFAP. **C**) Scaled gene expression heatmap showing select gene marker expression for the unbiased astrocyte sub-clusters. Sub-clusters 1 and 8 show high expression of protoplasmic astrocyte genes, while sub-clusters 6 and 7 show high expression of reactive genes. Sub-cluster 0 has very few differentially increased genes and is considered baseline. **D**) Heatmap showing the proportion of nuclei in each astrocyte sub-cluster (columns) contributed by disease condition (row). The proportions are scaled by columns. **E**) Kaplan-Meier survival plot graphing survival in the combined TCGA and CGGA RNAseq datasets. The samples were classified based on enrichment of Astrocyte sub-cluster 6 signature genes into positive or negative enrichment. Statistical significance was computed using the log rank test. **F**) Cox proportional hazard ratio of survival in the combined TCGA and CGGA dataset given enrichment of each of the astrocyte cluster specific signature. Age is also included in the model. The p values are shown on the left, bars indicate confidence intervals (also noted on the right). Enrichment of each geneset was categorized as negative or positive. **G-I**) Geneset enrichment scores used for astrocyte state detection (clustering). **J**) The scaled proportion of the astrocytes nuclei used for compositional analysis (c-ast1, c-ast2, c-ast3 - columns) that were clustered into the three astrocyte states (Ast1, Ast2, and Ast3 – as per clustering on gene-set enrichment scores). Scaling was done by column.

Figure S10: The transcriptional landscape of microglia in glioma

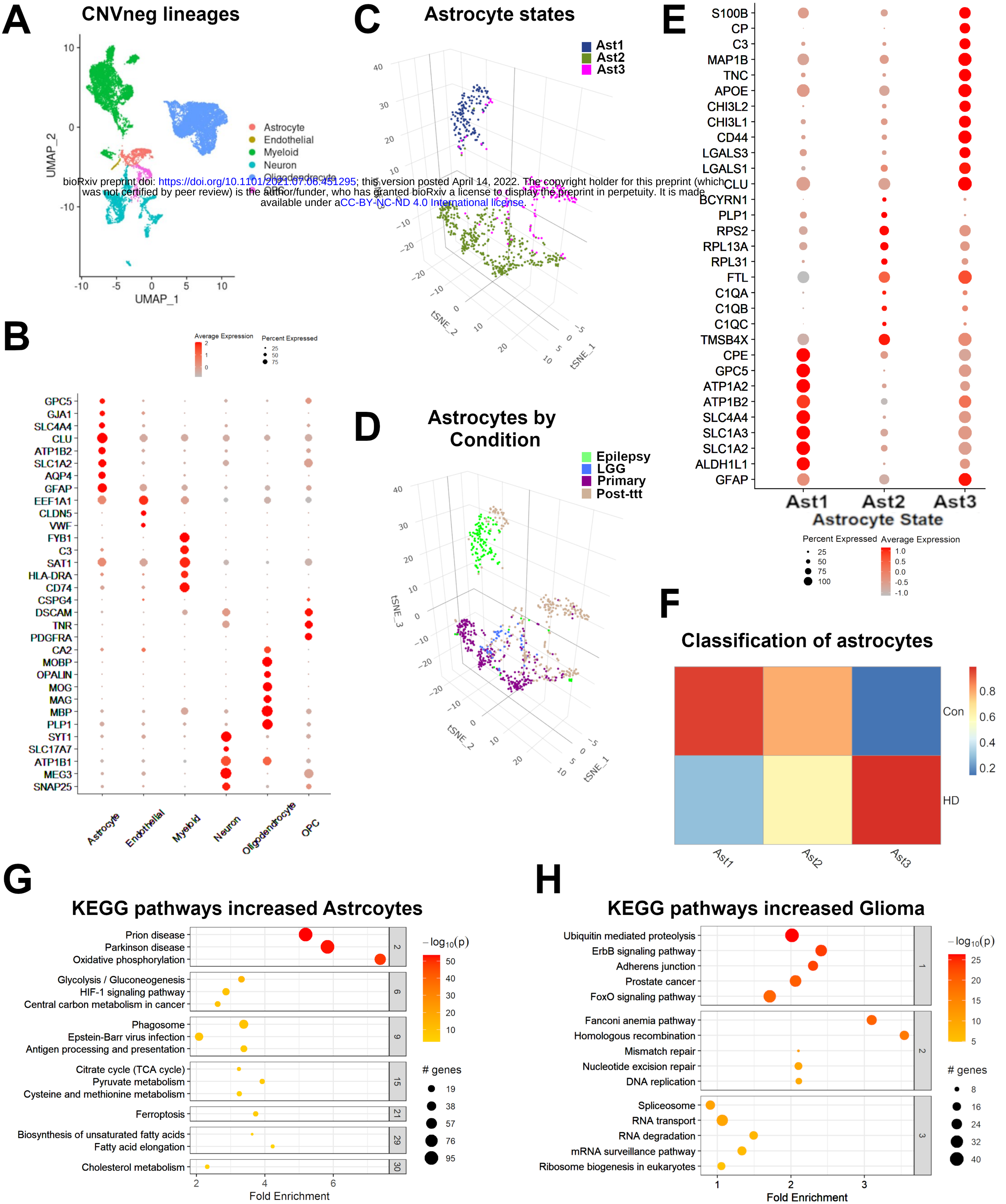
A) Uniform-manifold approximation and projection (UMAP) graphs plots showing all myeloid nuclei from color-coded by cluster (**B**) and condition (primary glioma, recurrent glioma, low grade glioma (LGG), and epilepsy (**C**). Gene expression violin plots showing select gene marker expression for the immune cell clusters from top to bottom; Myel1, mgTAM, moTAM, prTAM, and T cells. **D**) Heatmap showing the proportion of nuclei in each cluster (columns) contributed by condition (rows). **E**) Heatmap showing the scaled enrichment scores of gene sets derived from Movahedi et al 2021 in the nuclei pooled from each myeloid cluster.

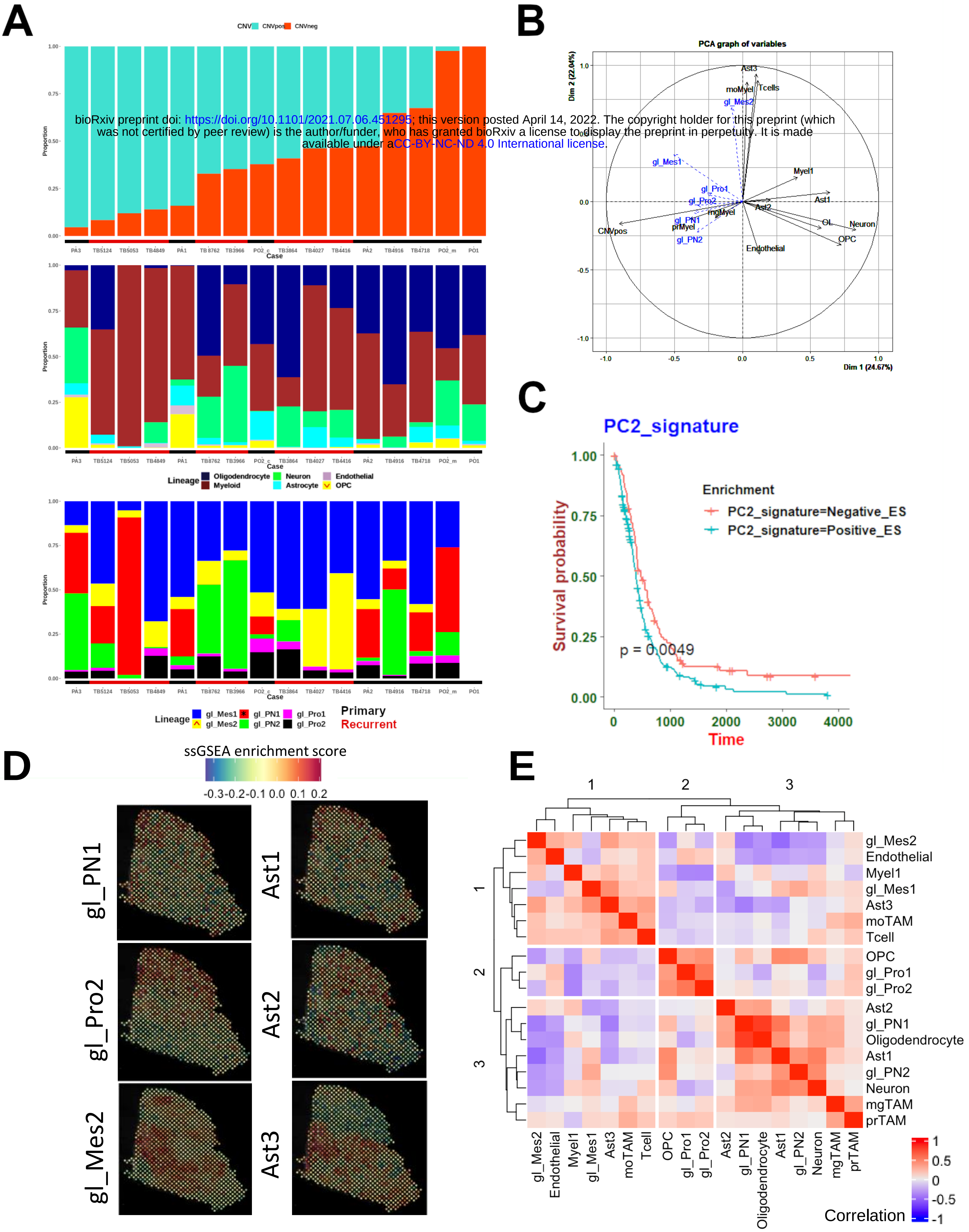
Figure S11: Spatial transcriptomics samples

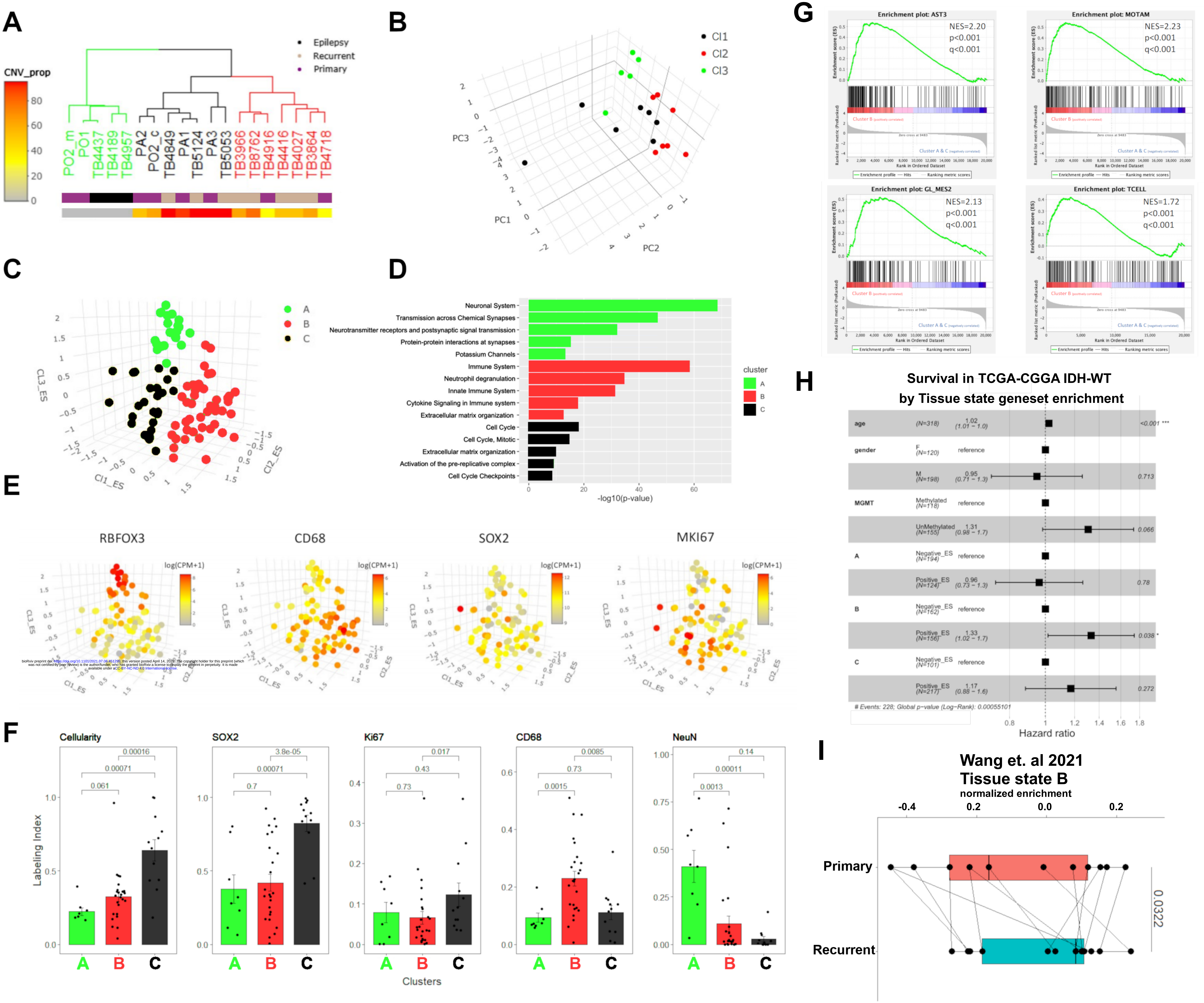
A) DAPI images of each of the 9 spatial transcriptomics (ST) capture regions, with 1mm scale bar. **B**) Transcript counts per spot projected on ST object. **C**) Representative iteration of k=10 segmentation of each ST object, used for the correlation analysis in Figure 2E. The scale bars are indicated.

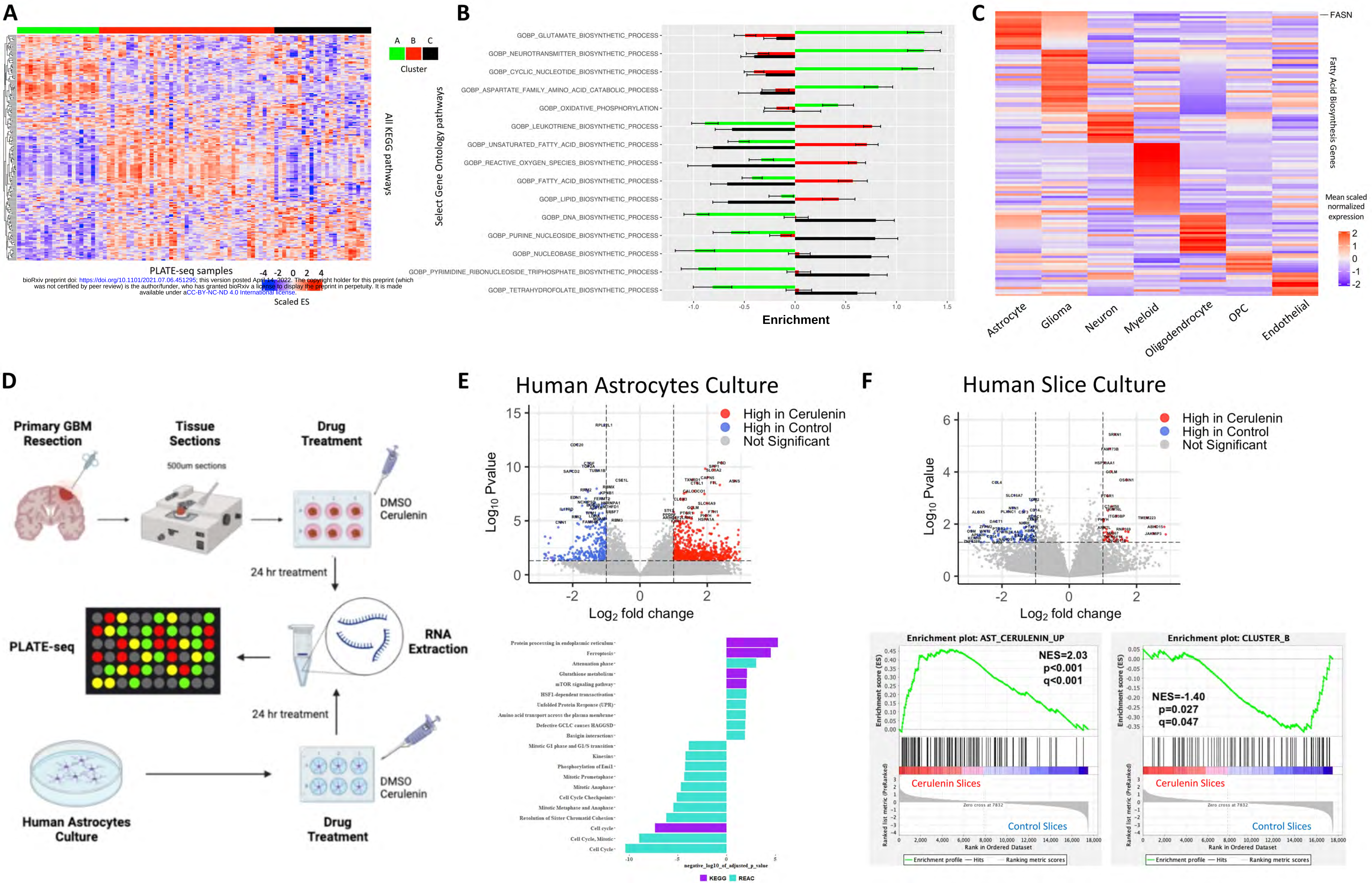
Figure S12: The spatial landscape of glioma margins

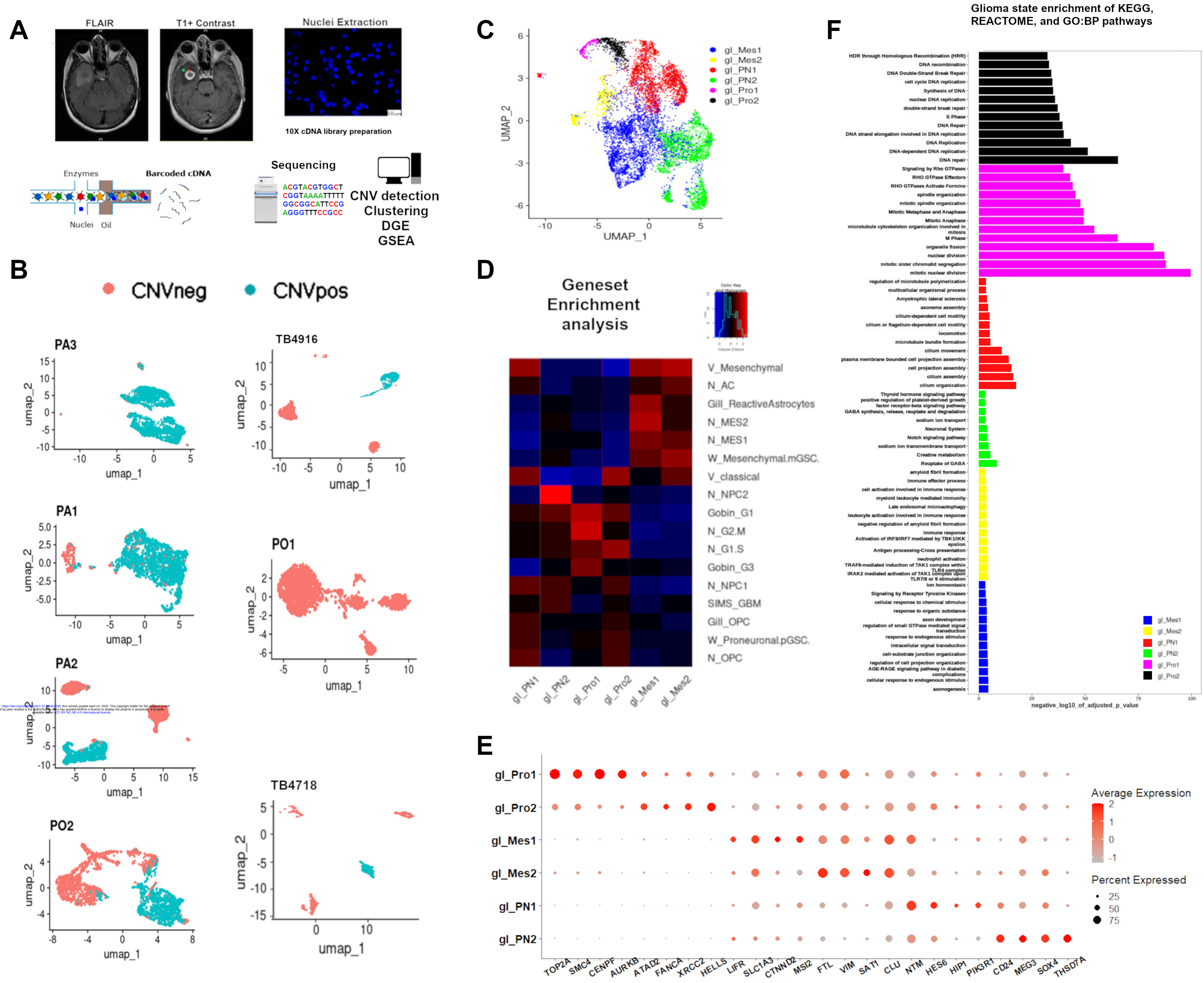
A) Outline of spatial transcriptomic analysis of infiltrating GBM. DAPI (left) and NeuN (right) immunostains of frozen sections from case PO2, for which snRNAseq was done. Each circle represents a biopsy on which bulk RNAseq was done. After the biopsies were taken, the specimen was bisected along the dashed white line (y-axis) and subjected to snRNASeq. **B**) Enrichment analysis of each of the spatially mapped biopsies using the genesets of the three compositional clusters (see text for details) displaying normalized single sample GSEA enrichment scores for the tumor cluster (C - upper panel), the tumor-reactive cluster (B– middle panel), and the normal brain cluster (A – lower panel). The enrichment scores are coded by color and size. The normalized RNA data for the spatial biopsy map is available in an interactive web interface at https://vmenon.shinyapps.io/gbm_expression/. **C**) Enrichment of the tissue state genesets as in B, applied to the Gill et al. 2014 MRI localized biopsy dataset. The normalized enrichment scores for each tissue state are shown in histograms for the control brain, contrast enhancing (Contrast_E) samples, and FLAIR-abnormal margin samples from primary and recurrent samples. The lighter segments of the histograms indicate samples with negative enrichment.

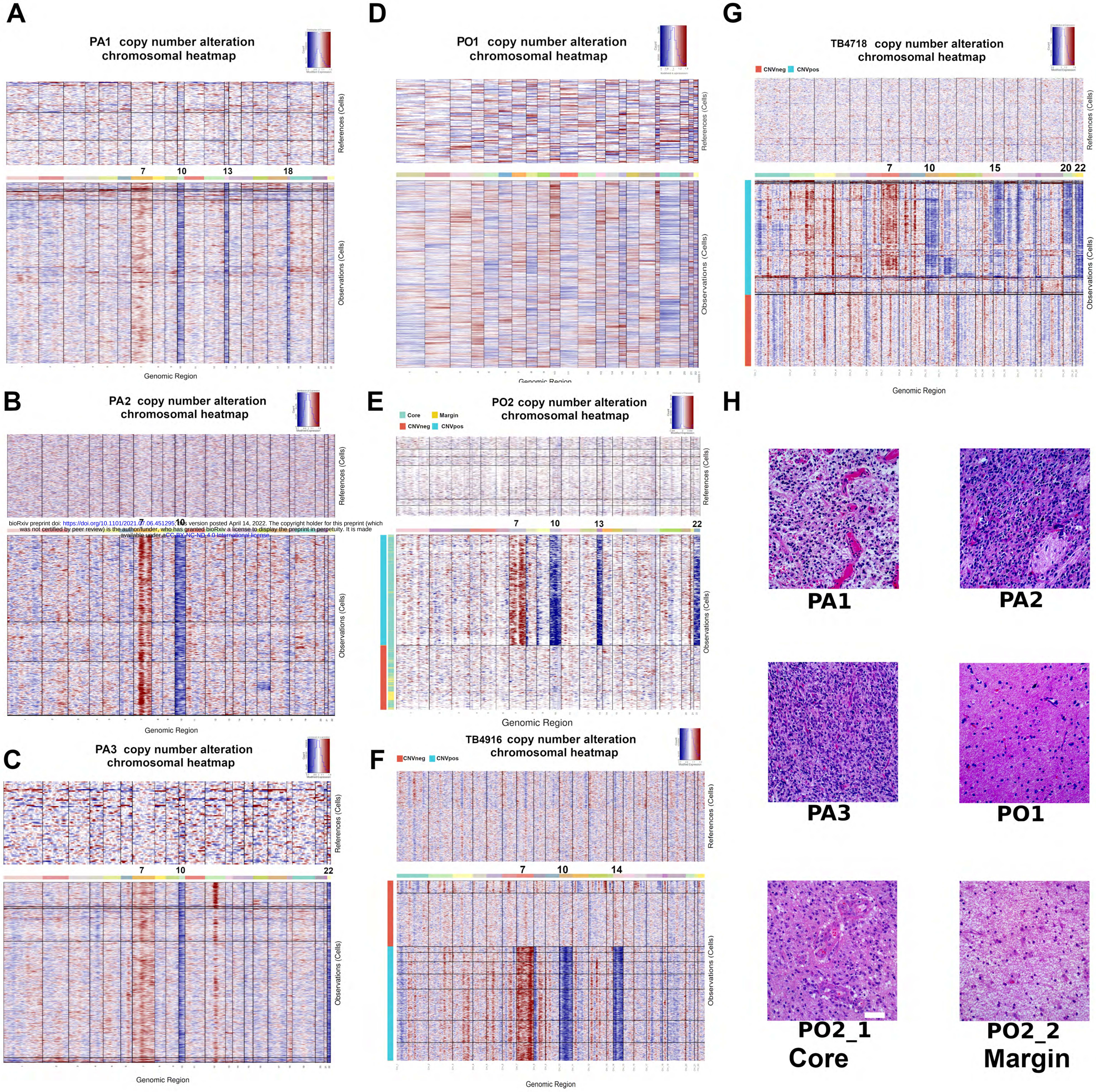

A
CNVneg lineages
UMAP_2
UMAP_1
10
0
-10
-10
-5
0
5
10
Astrocyte
Endothelial
Myeloid
Neuron
Oligodendrocyte
OPC
C
Astrocyte states
Ast1
Ast2
Ast3
tSNE_1
tSNE_2
tSNE_3
E
S100B
CP
C3
MAP1B
TNC
APOE
CHI3L2
CHI3L1
CD44
LGALS3
LGALS1
CLU
BCYRN1
PLP1
RPS2
RPL13A
RPL31
FTL
C1QA
C1QB
C1QC
TMSB4X
CPE
GPC5
ATP1A2
ATP1B2
SLC4A4
SLC1A3
SLC1A2
ALDH1L1
GFAP
Ast1
Ast2
Ast3
Astrocyte State
Percent Expressed
25
50
75
100
Average Expression
1.0
0.5
0.0
-0.5
-1.0
B
Average Expression
2
1
0
Percent Expressed
25
50
75
GPC5
GJA1
SLC4A4
CLU
ATP1B2
SLC1A2
AQP4
GFAP
EEF1A1
CLDN5
VWF
FYB1
C3
SAT1
HLA-DRA
CD74
CSPG4
DSCAM
TNR
PDGFRA
CA2
MOBP
OPALIN
MOG
MAG
MBP
PLP1
SYT1
SLC17A7
ATP1B1
MEG3
SNAP25
Astrocyte
Endothelial
Myeloid
Neuron
Oligodendrocyte
OPC
D
Astrocytes by Condition
Epilepsy
LGG
Primary
Post-ttt
tSNE_1
tSNE_2
tSNE_3
F
Classification of astrocytes
Con
HD
Ast1
Ast2
Ast3
0.8
0.6
0.4
0.2
G
KEGG pathways increased Astrcoytes
Prion disease
Parkinson disease
Oxidative phosphorylation
Glycolysis / Gluconeogenesis
HIF-1 signaling pathway
Central carbon metabolism in cancer
Phagosome
Epstein-Barr virus infection
Antigen processing and presentation
Citrate cycle (TCA cycle)
Pyruvate metabolism
Cysteine and methionine metabolism
Ferroptosis
Biosynthesis of unsaturated fatty acids
Fatty acid elongation
Cholesterol metabolism
2
6
9
15
21
29
30
Fold Enrichment
2
4
6
$-\log_{10}(p)$
50
40
30
20
10
genes
19
38
57
76
95
H
KEGG pathways increased Glioma
Ubiquitin mediated proteolysis
ErbB signaling pathway
Adherens junction
Prostate cancer
FoxO signaling pathway
Fanconi anemia pathway
Homologous recombination
Mismatch repair
Nucleotide excision repair
DNA replication
Spliceosome
RNA transport
RNA degradation
mRNA surveillance pathway
Ribosome biogenesis in eukaryotes
1
2
3
Fold Enrichment
1
2
3
$-\log_{10}(p)$
25
20
15
10
5
genes
8
16
24
32
40

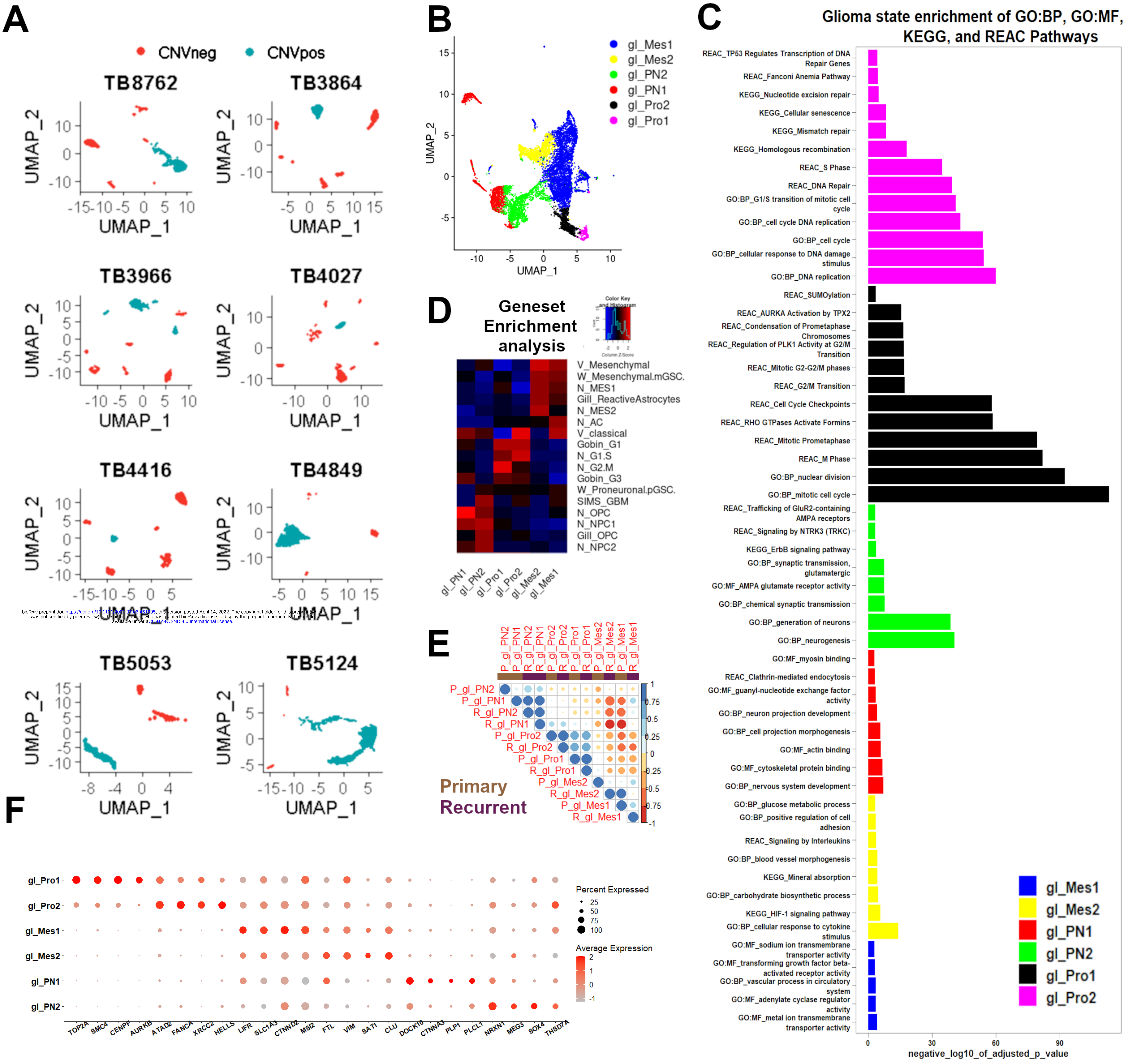

A
CNV CNVpos CNVneg
Proportion
1.00 0.75 0.50 0.25 0.00
PA3 TB5124 TB5053 TB4849 PA1 TB8762 TB3966 PO2_c TB3864 TB4027 TB4416 PA2 TB4916 TB4718 PO2_m PO1
Case
Lineage
Oligodendrocyte
Myeloid
Neuron
Astrocyte
Endothelial
OPC
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
Primary
Recurrent
B
PCA graph of variables
Dim 2 (22.04%)
Dim 1 (24.67%)
1.0 0.5 0.0 -0.5 -1.0
Ast3
moMyel
Tcells
gl_Mes2
gl_Mes1
Myel1
gl_Pro1
gl_Pro2
Ast1
Ast2
mgMyel
gl_PN1
prMyel
gl_PN2
CNVpos
OL
Neuron
OPC
Endothelial
C
PC2_signature
Survival probability
1.00 0.75 0.50 0.25 0.00
Enrichment
PC2_signature=Negative_ES
PC2_signature=Positive_ES
p = 0.0049
0 1000 2000 3000 4000
Time
D
ssGSEA enrichment score
-0.3 -0.2 -0.1 0.0 0.1 0.2
gl_PN1
gl_Pro2
gl_Mes2
Ast1
Ast2
Ast3
E
1 2 3
gl_Mes2
Endothelial
Myel1
gl_Mes1
Ast3
moTAM
Tcell
OPC
gl_Pro1
gl_Pro2
Ast2
gl_PN1
Oligodendrocyte
Ast1
gl_PN2
Neuron
mgTAM
prTAM
Correlation
1 0.5 0 -0.5 -1

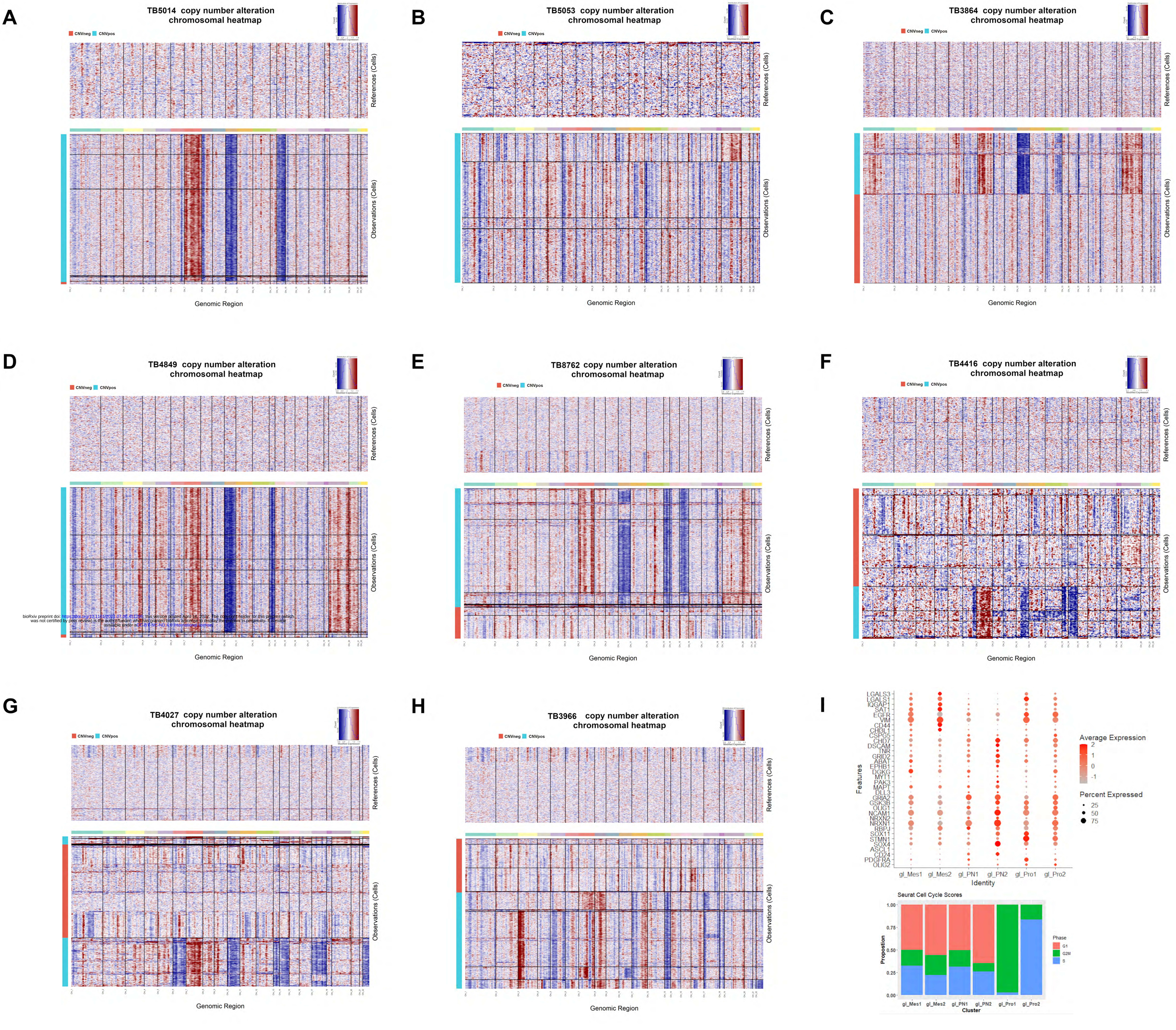

A
CNV_prop
80
60
40
20
0
Epilepsy
Recurrent
Primary
PO2_m
PO1
TB4437
TB4189
TB4957
PA2
PO2_c
TB4849
PA1
TB5124
PA3
TB5053
TB3966
TB8762
TB4916
TB4416
TB4027
TB3864
TB4718
B
Cl1
Cl2
Cl3
PC1
PC2
PC3
C
A
B
C
Cl1_ES
Cl2_ES
CL3_ES
D
Neuronal System
Transmission across Chemical Synapses
Neurotransmitter receptors and postsynaptic signal transmission
Protein-protein interactions at synapses
Potassium Channels
Immune System
Neutrophil degranulation
Innate Immune System
Cytokine Signaling in Immune system
Extracellular matrix organization
Cell Cycle
Cell Cycle, Mitotic
Extracellular matrix organization
Activation of the pre-replicative complex
Cell Cycle Checkpoints
0
20
40
60
-log10(p-value)
cluster
A
B
C
E
RBFOX3
CD68
SOX2
MKI67
log(CPM+1)
Cl1_ES
Cl2_ES
CL3_ES
F
Cellularity
SOX2
Ki67
CD68
NeuN
Labeling Index
0.061
0.00071
0.00016
0.7
0.00071
3.8e-05
0.73
0.43
0.017
0.0015
0.73
0.0085
0.0013
0.00011
0.14
A
B
C
Clusters
G
Enrichment plot: AST3
NES=2.20
p<0.001
q<0.001
Enrichment plot: MOTAM
NES=2.23
p<0.001
q<0.001
Enrichment plot: GL_MES2
NES=2.13
p<0.001
q<0.001
Enrichment plot: TCELL
NES=1.72
p<0.001
q<0.001
Enrichment score (ES)
Ranked list metric (PreRanked)
Cluster B (positively correlated)
Cluster A & C (negatively correlated)
Zero cross at 9483
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
H
Survival in TCGA-CGGA IDH-WT
by Tissue state geneset enrichment
age (N=318) 1.02 (1.01 − 1.0) <0.001 ***
gender F (N=120) reference
M (N=198) 0.95 (0.71 − 1.3) 0.713
MGMT Methylated (N=118) reference
UnMethylated (N=155) 1.31 (0.98 − 1.7) 0.066
A Negative_ES (N=194) reference
Positive_ES (N=124) 0.96 (0.73 − 1.3) 0.78
B Negative_ES (N=162) reference
Positive_ES (N=156) 1.33 (1.02 − 1.7) 0.038 *
C Negative_ES (N=101) reference
Positive_ES (N=217) 1.17 (0.88 − 1.6) 0.272
Events: 228; Global p-value (Log-Rank): 0.00055101
Hazard ratio
I
Wang et. al 2021
Tissue state B
normalized enrichment
-0.4
0.2
0.0
0.2
Primary
Recurrent
0.0322


A
A
B
C
Cluster
All KEGG pathways
PLATE-seq samples
-4 -2 0 2 4
Scaled ES
B
GOBP_GLUTAMATE_BIOSYNTHETIC_PROCESS
GOBP_NEUROTRANSMITTER_BIOSYNTHETIC_PROCESS
GOBP_CYCLIC_NUCLEOTIDE_BIOSYNTHETIC_PROCESS
GOBP_ASPARTATE_FAMILY_AMINO_ACID_CATABOLIC_PROCESS
GOBP_OXIDATIVE_PHOSPHORYLATION
GOBP_LEUKOTRIENE_BIOSYNTHETIC_PROCESS
GOBP_UNSATURATED_FATTY_ACID_BIOSYNTHETIC_PROCESS
GOBP_REACTIVE_OXYGEN_SPECIES_BIOSYNTHETIC_PROCESS
GOBP_FATTY_ACID_BIOSYNTHETIC_PROCESS
GOBP_LIPID_BIOSYNTHETIC_PROCESS
GOBP_DNA_BIOSYNTHETIC_PROCESS
GOBP_PURINE_NUCLEOSIDE_BIOSYNTHETIC_PROCESS
GOBP_NUCLEOBASE_BIOSYNTHETIC_PROCESS
GOBP_PYRIMIDINE_RIBONUCLEOSIDE_TRIPHOSPHATE_BIOSYNTHETIC_PROCESS
GOBP_TETRAHYDROFOLATE_BIOSYNTHETIC_PROCESS
Select Gene Ontology pathways
-1.0 -0.5 0.0 0.5 1.0 1.5
Enrichment
C
FASN
Fatty Acid Biosynthesis Genes
Mean scaled normalized expression
2 1 0 -1 -2
Astrocyte
Glioma
Neuron
Myeloid
Oligodendrocyte
OPC
Endothelial
D
Primary GBM Resection
Tissue Sections
500um sections
Drug Treatment
DMSO
Cerulenin
24 hr treatment
RNA Extraction
PLATE-seq
24 hr treatment
Human Astrocytes Culture
DMSO
Cerulenin
Drug Treatment
E
Human Astrocytes Culture
High in Cerulenin
High in Control
Not Significant
Log_{10} Pvalue
0 5 10 15
-2 0 2
Log_2 fold change
RPL22L1
CDC20
TOP2A
SAPCD2
TUBA1B
CSE1L
RRM2
EDN1
CNN1
PGD
SRP1
SLC3A2
CAPN5
TXNRD1
CTSL1
FTL
ASNS
CALCOCO1
CLCN3
SLC6A9
FTH1
PHYH
HSPA1A
Protein processing in endoplasmic reticulum
Ferroptosis
Attenuation phase
Glutathione metabolism
mTOR signaling pathway
HSF1-dependent transactivation
Unfolded Protein Response (UPR)
Amino acid transport across the plasma membrane
Defective GCLC causes HAGGSD
Basigin interactions
Mitotic G1 phase and G1/S transition
Kinesins
Phosphorylation of Emi1
Mitotic Prometaphase
Mitotic Anaphase
Cell Cycle Checkpoints
Mitotic Metaphase and Anaphase
Resolution of Sister Chromatid Cohesion
Cell cycle
Cell Cycle, Mitotic
Cell Cycle
-10 -5 0 5
negative_log10_of_adjusted_p_value
KEGG
REAC
F
Human Slice Culture
High in Cerulenin
High in Control
Not Significant
Log_{10} Pvalue
0 2 4 6
-2 0 2
Log_2 fold change
SRXN1
FAM173B
HSPA1A
GCLM
OSGIN1
TMEM223
ABHD15
JAKMIP3
COL4
SLC6A7
ALOX5
Enrichment plot: AST_CERULENIN_UP
NES=2.03
p<0.001
q<0.001
Enrichment score (ES)
Ranked list metric (PreRanked)
Cerulenin Slices
Zero cross at 7832
Control Slices
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
Enrichment plot: CLUSTER_B
NES=-1.40
p=0.027
q=0.047
Enrichment score (ES)
Ranked list metric (PreRanked)
Cerulenin Slices
Zero cross at 7832
Control Slices
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

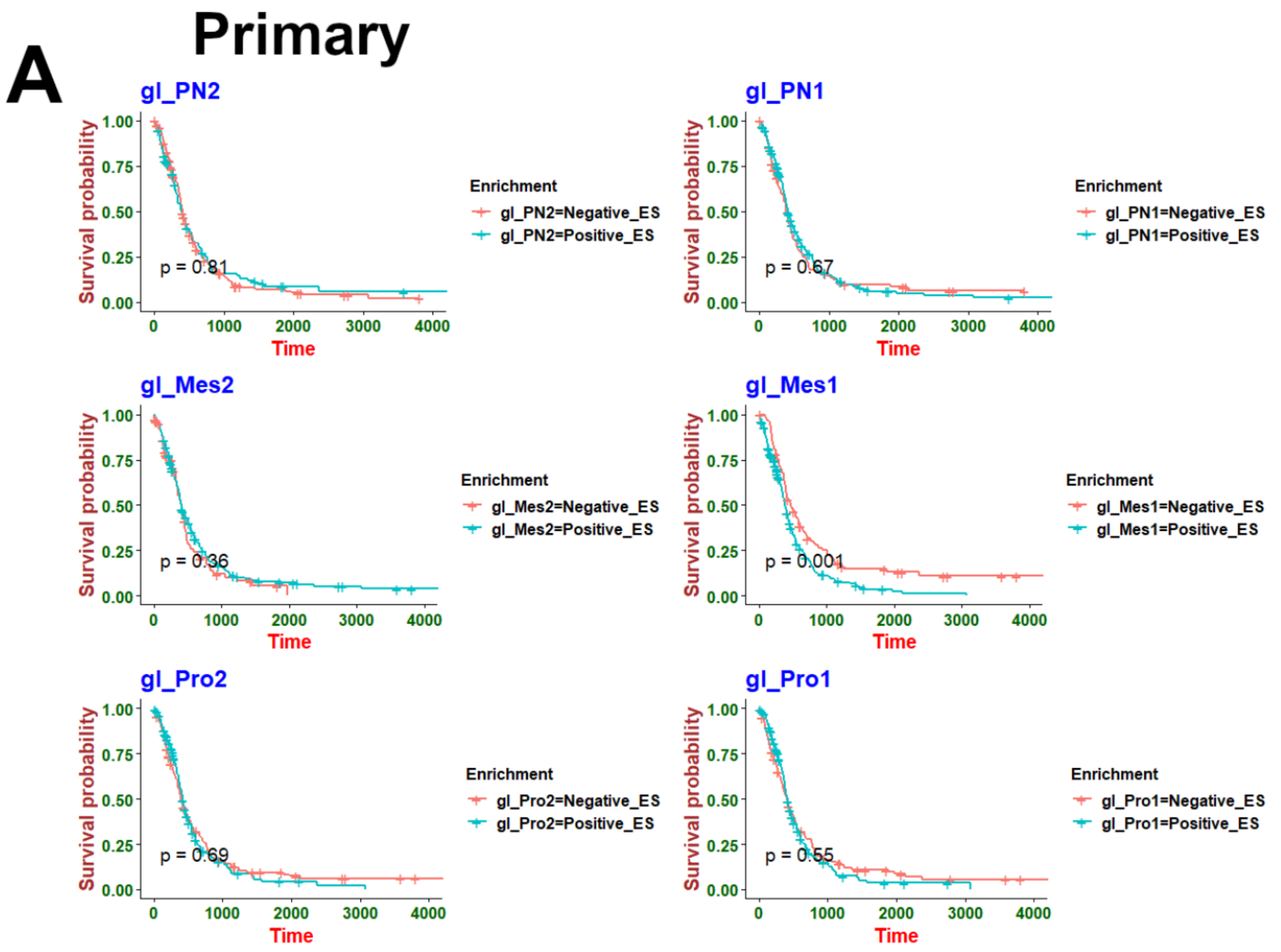

A
FLAIR
T1+ Contrast
Nuclei Extraction
50um
10X cDNA library preparation
Enzymes
Barcoded cDNA
Nuclei
Oil
Sequencing
ACGTACGTGGCT
CGGTAAAATTTTT
GGCGGCATTCCG
AGGGTTTCCGCC
CNV detection
Clustering
DGE
GSEA
B
CNVneg
CNVpos
PA3
TB4916
PA1
PO1
PA2
PO2
TB4718
umap_1
umap_2
C
UMAP_1
UMAP_2
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
D
Geneset Enrichment analysis
Color Key and Histogram
Column Z-Score
V_Mesenchymal
N_AC
Gill_ReactiveAstrocytes
N_MES2
N_MES1
W_Mesenchymal.mGSC.
V_classical
N_NPC2
Gobin_G1
N_G2.M
N_G1.S
Gobin_G3
N_NPC1
SIMS_GBM
Gill_OPC
W_Proneuronal.pGSC.
N_OPC
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
gl_Mes1
gl_Mes2
E
gl_Pro1
gl_Pro2
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
TOP2A
SMC4
CENPF
AURKB
ATAD2
FANCA
XRCC2
HELLS
LIFR
SLC1A3
CTNND2
MSI2
FTL
VIM
SAT1
CLU
NTM
HES6
HIP1
PIK3R1
CD24
MEG3
SOX4
THSD7A
Average Expression
2
1
0
-1
Percent Expressed
25
50
75
F
Glioma state enrichment of KEGG, REACTOME, and GO:BP pathways
HDR through Homologous Recombination (HRR)
DNA recombination
DNA Double-Strand Break Repair
cell cycle DNA replication
Synthesis of DNA
nuclear DNA replication
double-strand break repair
S Phase
DNA Repair
DNA strand elongation involved in DNA replication
DNA Replication
DNA-dependent DNA replication
DNA repair
Signaling by Rho GTPases
RHO GTPase Effectors
RHO GTPases Activate Formins
spindle organization
mitotic spindle organization
Mitotic Metaphase and Anaphase
Mitotic Anaphase
microtubule cytoskeleton organization involved in mitosis
M Phase
organelle fission
nuclear division
mitotic sister chromatid segregation
mitotic nuclear division
regulation of microtubule polymerization
multicellular organismal process
Amyotrophic lateral sclerosis
axoneme assembly
cilium-dependent cell motility
cilium or flagellum-dependent cell motility
locomotion
microtubule bundle formation
cilium movement
plasma membrane bounded cell projection assembly
cell projection assembly
cilium assembly
cilium organization
Thyroid hormone signaling pathway
positive regulation of platelet-derived growth factor receptor-beta signaling pathway
GABA synthesis, release, reuptake and degradation
sodium ion transport
Neuronal System
Notch signaling pathway
sodium ion transmembrane transport
Creatine metabolism
Reuptake of GABA
amyloid fibril formation
immune effector process
cell activation involved in immune response
myeloid leukocyte mediated immunity
Late endosomal microautophagy
leukocyte activation involved in immune response
negative regulation of amyloid fibril formation
immune response
Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon
Antigen processing-Cross presentation
neutrophil activation
TRAF6-mediated induction of TAK1 complex within TLR4 complex
IRAK2 mediated activation of TAK1 complex upon TLR7/8 or 9 stimulation
Ion homeostasis
Signaling by Receptor Tyrosine Kinases
cellular response to chemical stimulus
response to organic substance
axon development
regulation of small GTPase mediated signal transduction
response to endogenous stimulus
intracellular signal transduction
cell-substrate junction organization
regulation of cell projection organization
AGE-RAGE signaling pathway in diabetic complications
cellular response to endogenous stimulus
axonogenesis
0
25
50
75
100
negative_log10_of_adjusted_p_value
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2

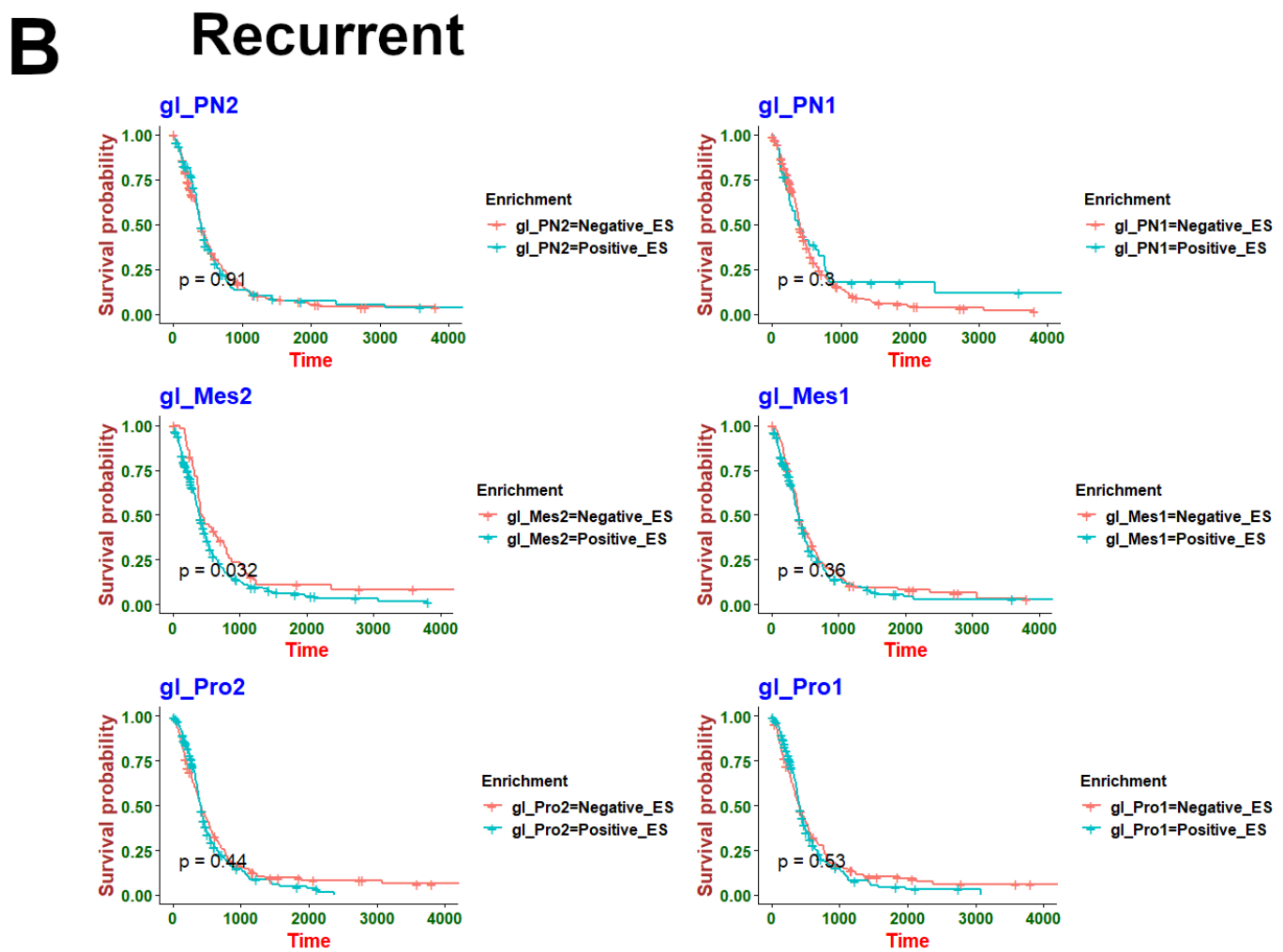

A
PA1 copy number alteration chromosomal heatmap
References (Cells)
7
10
13
18
Observations (Cells)
Genomic Region
B
PA2 copy number alteration chromosomal heatmap
References (Cells)
7
10
Observations (Cells)
Genomic Region
C
PA3 copy number alteration chromosomal heatmap
References (Cells)
7
10
22
Observations (Cells)
Genomic Region
D
PO1 copy number alteration chromosomal heatmap
References (Cells)
Observations (Cells)
Genomic Region
E
Core
Margin
CNVneg
CNVpos
PO2 copy number alteration chromosomal heatmap
References (Cells)
7
10
13
22
Observations (Cells)
Genomic Region
F
TB4916 copy number alteration chromosomal heatmap
CNVneg
CNVpos
References (Cells)
7
10
14
Observations (Cells)
Genomic Region
G
TB4718 copy number alteration chromosomal heatmap
CNVneg
CNVpos
References (Cells)
7
10
15
20
22
Observations (Cells)
Genomic Region
H
PA1
PA2
PA3
PO1
PO2_1
Core
PO2_2
Margin

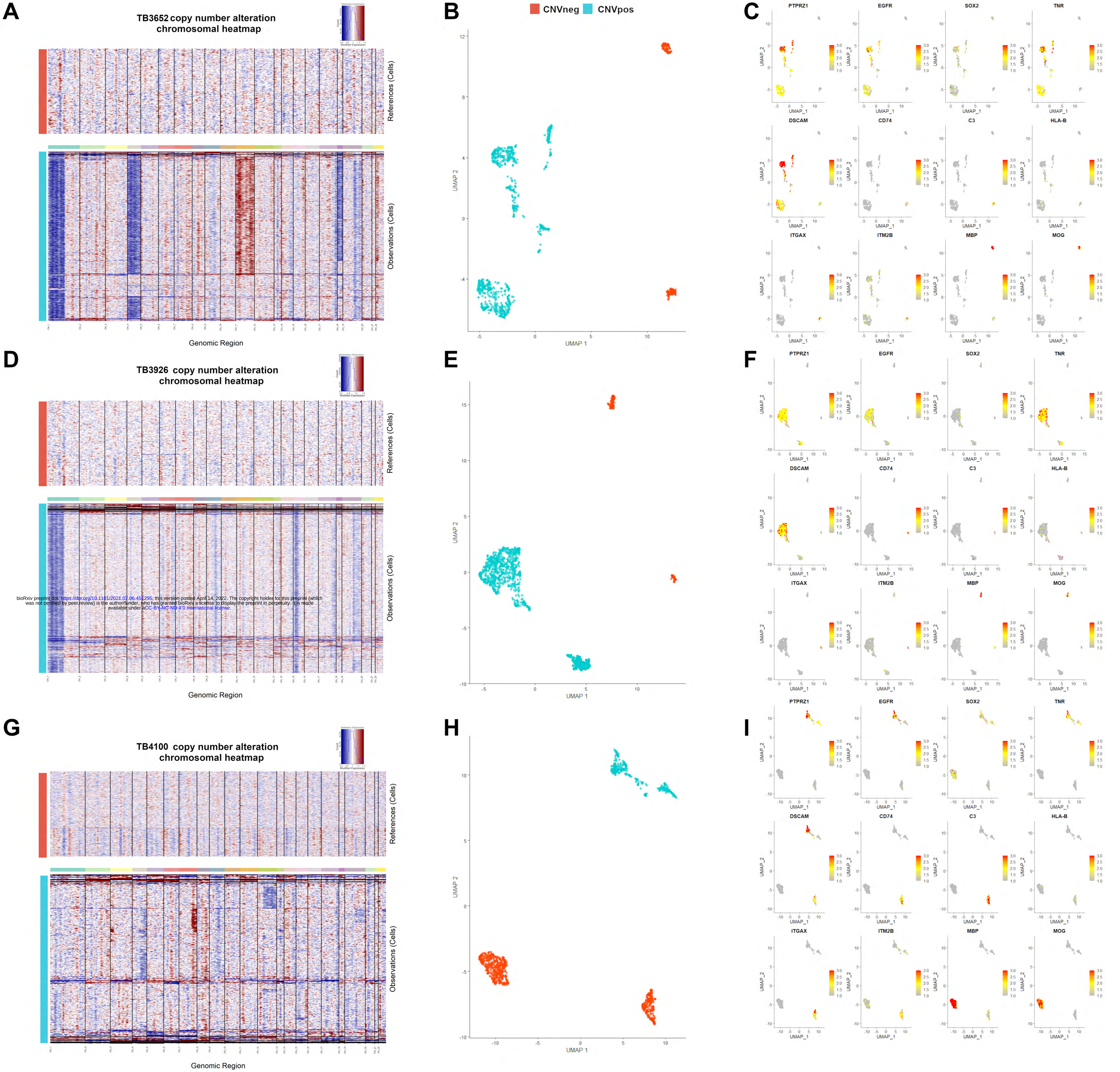

A
CNVneg
CNVpos
TB8762
TB3864
TB3966
TB4027
TB4416
TB4849
TB5053
TB5124
UMAP_1
UMAP_2
B
gl_Mes1
gl_Mes2
gl_PN2
gl_PN1
gl_Pro2
gl_Pro1
UMAP_1
UMAP_2
C
Glioma state enrichment of GO:BP, GO:MF, KEGG, and REAC Pathways
REAC_TP53 Regulates Transcription of DNA Repair Genes
REAC_Fanconi Anemia Pathway
KEGG_Nucleotide excision repair
KEGG_Cellular senescence
KEGG_Mismatch repair
KEGG_Homologous recombination
REAC_S Phase
REAC_DNA Repair
GO:BP_G1/S transition of mitotic cell cycle
GO:BP_cell cycle DNA replication
GO:BP_cell cycle
GO:BP_cellular response to DNA damage stimulus
GO:BP_DNA replication
REAC_SUMOylation
REAC_AURKA Activation by TPX2
REAC_Condensation of Prometaphase Chromosomes
REAC_Regulation of PLK1 Activity at G2/M Transition
REAC_Mitotic G2-G2/M phases
REAC_G2/M Transition
REAC_Cell Cycle Checkpoints
REAC_RHO GTPases Activate Formins
REAC_Mitotic Prometaphase
REAC_M Phase
GO:BP_nuclear division
GO:BP_mitotic cell cycle
REAC_Trafficking of GluR2-containing AMPA receptors
REAC_Signaling by NTRK3 (TRKC)
KEGG_ErbB signaling pathway
GO:BP_synaptic transmission, glutamatergic
GO:MF_AMPA glutamate receptor activity
GO:BP_chemical synaptic transmission
GO:BP_generation of neurons
GO:BP_neurogenesis
GO:MF_myosin binding
REAC_Clathrin-mediated endocytosis
GO:MF_guanyl-nucleotide exchange factor activity
GO:BP_neuron projection development
GO:BP_cell projection morphogenesis
GO:MF_actin binding
GO:MF_cytoskeletal protein binding
GO:BP_nervous system development
GO:BP_glucose metabolic process
GO:BP_positive regulation of cell adhesion
REAC_Signaling by Interleukins
GO:BP_blood vessel morphogenesis
KEGG_Mineral absorption
GO:BP_carbohydrate biosynthetic process
KEGG_HIF-1 signaling pathway
GO:BP_cellular response to cytokine stimulus
GO:MF_sodium ion transmembrane transporter activity
GO:MF_transforming growth factor beta-activated receptor activity
GO:BP_vascular process in circulatory system
GO:MF_adenylate cyclase regulator activity
GO:MF_metal ion transmembrane transporter activity
0
30
60
90
negative_log10_of_adjusted_p_value
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
D
Geneset Enrichment analysis
Color Key and Histogram
Column Z-Score
V_Mesenchymal
W_Mesenchymal.mGSC.
N_MES1
Gill_ReactiveAstrocytes
N_MES2
N_AC
V_classical
Gobin_G1
N_G1.S
N_G2.M
Gobin_G3
W_Proneuronal.pGSC.
SIMS_GBM
N_OPC
N_NPC1
Gill_OPC
N_NPC2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
gl_Mes2
gl_Mes1
E
P_gl_PN2
P_gl_PN1
R_gl_PN2
R_gl_PN1
P_gl_Pro2
R_gl_Pro2
P_gl_Pro1
R_gl_Pro1
P_gl_Mes2
R_gl_Mes2
P_gl_Mes1
R_gl_Mes1
Primary
Recurrent
F
gl_Pro1
gl_Pro2
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
TOP2A
SMC4
CENPF
AURKB
ATAD2
FANCA
XRCC2
HELLS
LIFR
SLC1A3
CTNND2
MSI2
FTL
VIM
SAT1
CLU
DOCK10
CTNNA3
PLP1
PLCL1
NRXN1
MEG3
SOX4
THSD7A
Percent Expressed
25
50
75
100
Average Expression


A
TB5014 copy number alteration chromosomal heatmap
B
TB5053 copy number alteration chromosomal heatmap
C
TB3864 copy number alteration chromosomal heatmap
D
TB4849 copy number alteration chromosomal heatmap
E
TB8762 copy number alteration chromosomal heatmap
F
TB4416 copy number alteration chromosomal heatmap
G
TB4027 copy number alteration chromosomal heatmap
H
TB3966 copy number alteration chromosomal heatmap
CNVneg
CNVpos
References (Cells)
Observations (Cells)
Genomic Region
I
Features
LGALS3
LGALS1
IQGAP1
SAT1
EGFR
VIM
CD44
CHI3L1
CSPG5
CHD7
DSCAM
TNR
GRID2
ABAT
EPHB1
DGKG
MYT1
PAK3
MAPT
DLL3
GRIA2
GSK3B
OLIG1
NCAM1
NRXN2
NRXN1
RBPJ
SOX11
STMN1
SOX4
ASCL1
CD24
PDGFRA
OLIG2
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
Identity
Average Expression
2
1
0
-1
Percent Expressed
25
50
75
Seurat Cell Cycle Scores
Propostion
1.00
0.75
0.50
0.25
0.00
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
Cluster
Phase
G1
G2M
S

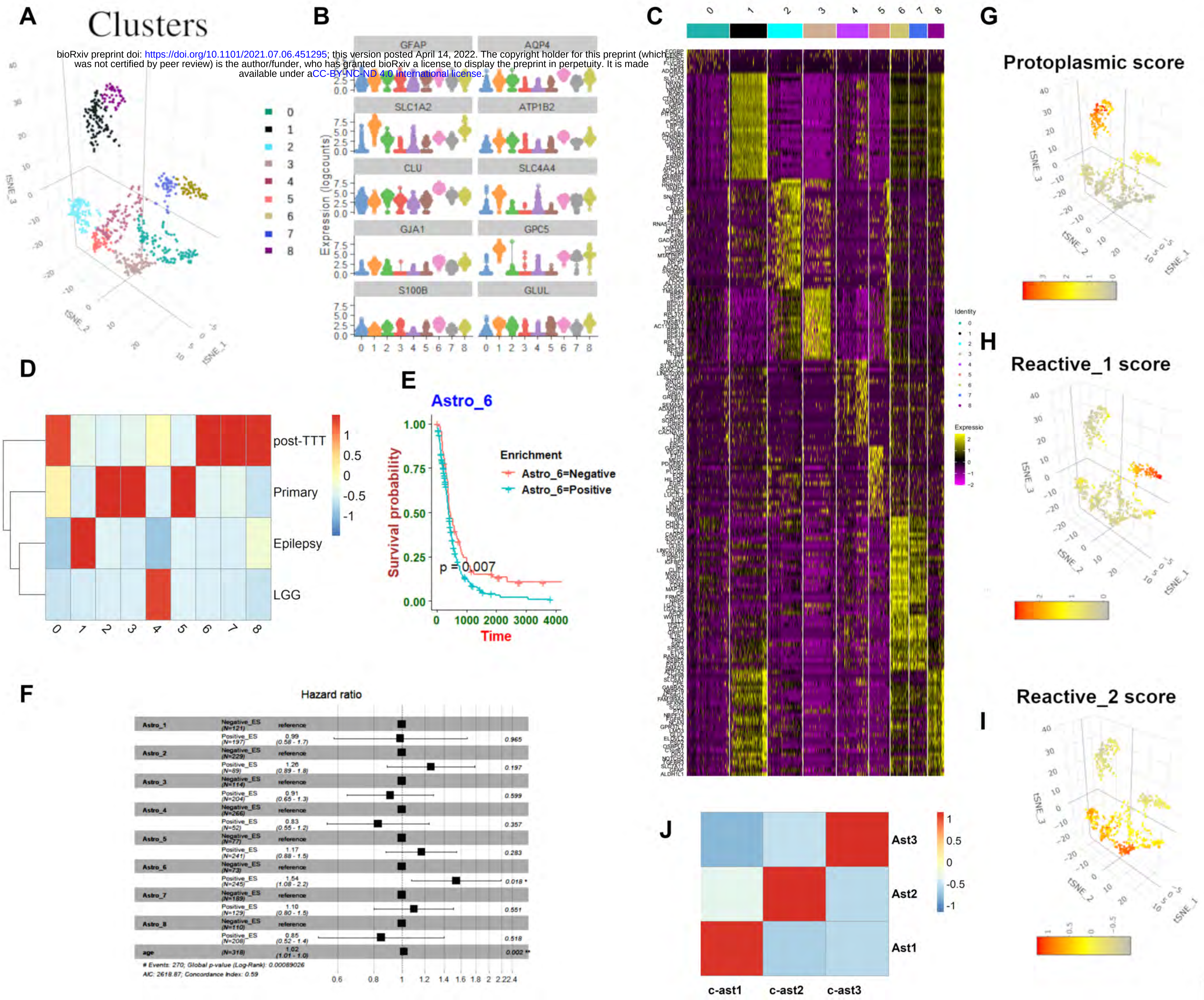

A
PTPRZ1 CLU DAPI
MV
Core
Cortex
Pia
B
0.0215
PTPRZ+CLU+/PTPRZ+ %
60
40
20
Core
Cortex
Region
C
NOVA1 MEG3 SOX2 DAPI
Pia
D
0.0423
SOX2+NOVA1+MEG3+/SOX2+ AND/OR NOVA1 %
20
15
10
5
Core
Cortex
Region
E
DAPI TOP2A CLU
F
*
TOP2A+CLU+/CLU+ %
75
50
25
CLU+TOP2A+
G
Lineage
gl_Mes1
gl_Mes2
gl_PN1
gl_PN2
gl_Pro1
gl_Pro2
5.0
2.5
0.0
-2.5
-5.0
UMAP_2
-4
0
4
8
UMAP_1
Condition
Primary
Recurrent
5.0
2.5
0.0
-2.5
-5.0
UMAP_2
-4
0
4
8
UMAP_1

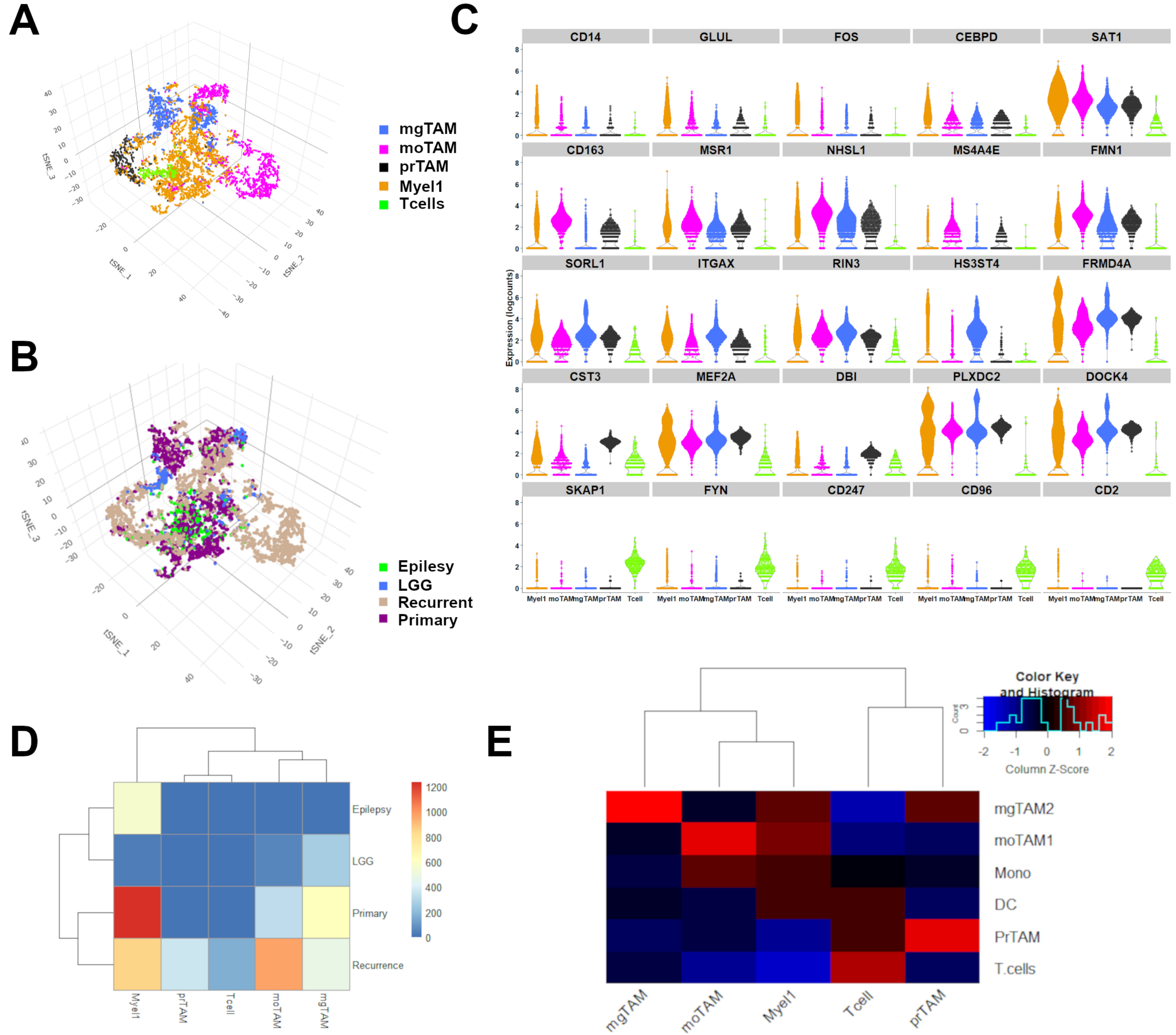

A
Primary
gl_PN2
p = 0.81
Enrichment
gl_PN2=Negative_ES
gl_PN2=Positive_ES
gl_PN1
p = 0.67
gl_PN1=Negative_ES
gl_PN1=Positive_ES
gl_Mes2
p = 0.36
gl_Mes2=Negative_ES
gl_Mes2=Positive_ES
gl_Mes1
p = 0.001
gl_Mes1=Negative_ES
gl_Mes1=Positive_ES
gl_Pro2
p = 0.69
gl_Pro2=Negative_ES
gl_Pro2=Positive_ES
gl_Pro1
p = 0.55
gl_Pro1=Negative_ES
gl_Pro1=Positive_ES
Survival probability
1.00
0.75
0.50
0.25
0.00
Time
0
1000
2000
3000
4000

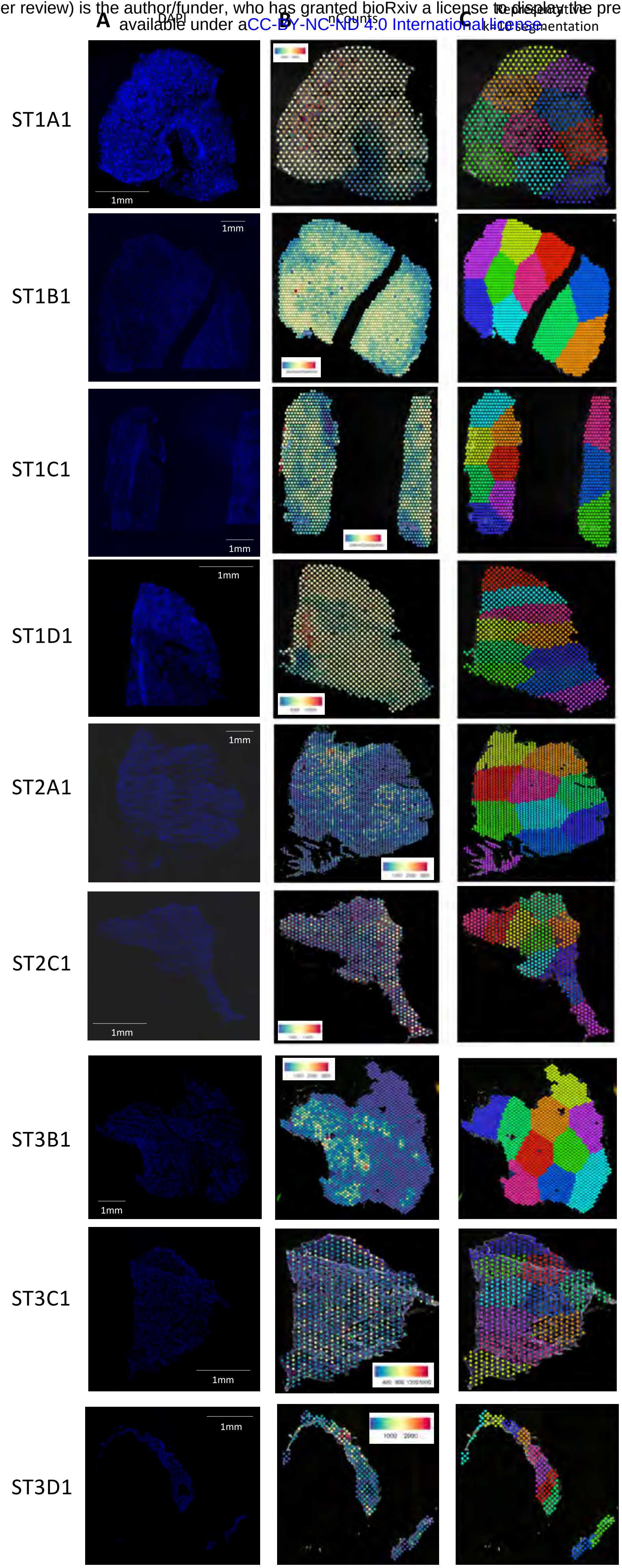

B
Recurrent
gl_PN2
p = 0.91
Enrichment
gl_PN2=Negative_ES
gl_PN2=Positive_ES
gl_PN1
p = 0.3
gl_PN1=Negative_ES
gl_PN1=Positive_ES
gl_Mes2
p = 0.032
gl_Mes2=Negative_ES
gl_Mes2=Positive_ES
gl_Mes1
p = 0.36
gl_Mes1=Negative_ES
gl_Mes1=Positive_ES
gl_Pro2
p = 0.44
gl_Pro2=Negative_ES
gl_Pro2=Positive_ES
gl_Pro1
p = 0.53
gl_Pro1=Negative_ES
gl_Pro1=Positive_ES
Survival probability
1.00
0.75
0.50
0.25
0.00
Time
0
1000
2000
3000
4000

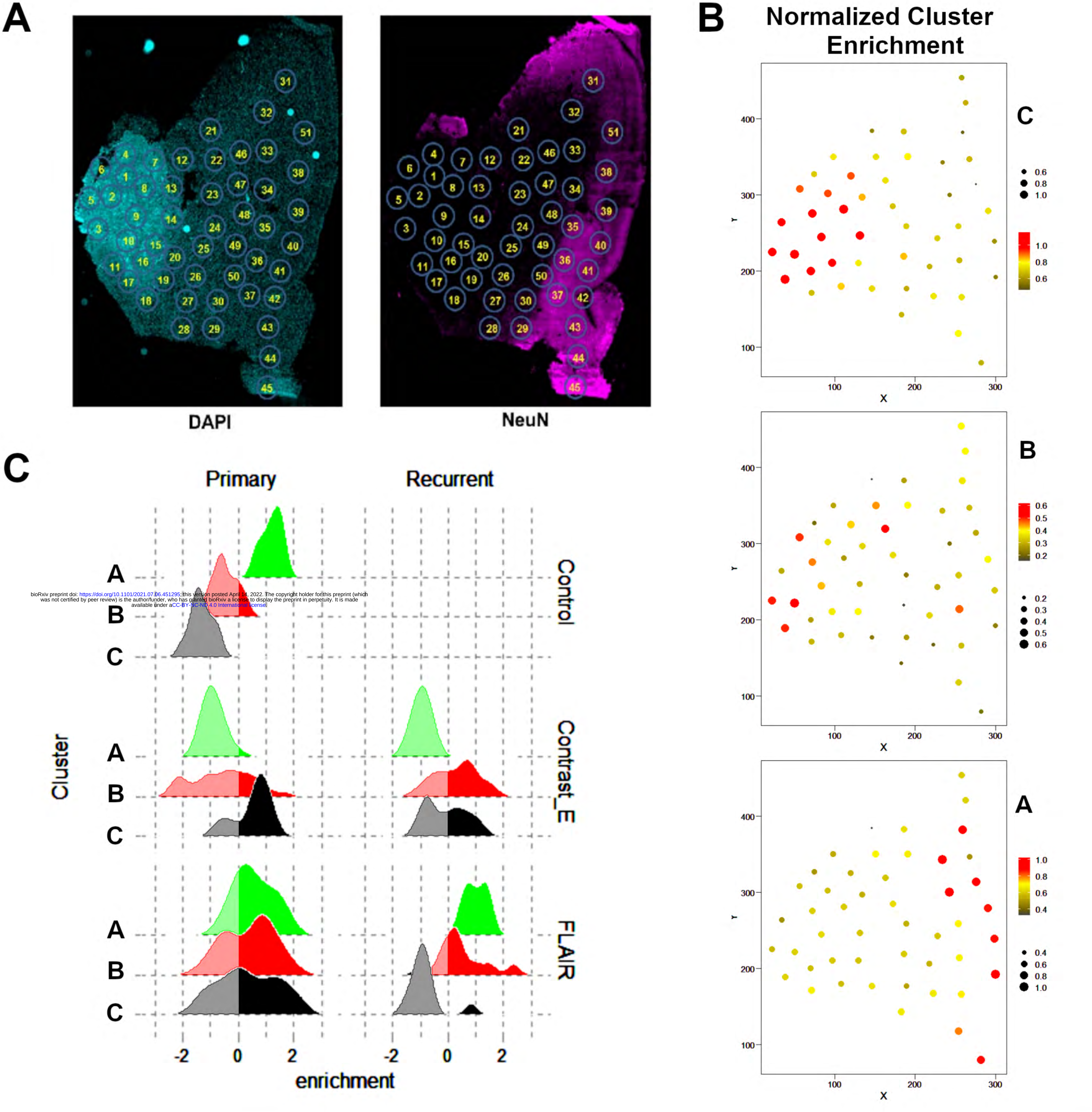

A
TB3652 copy number alteration chromosomal heatmap
References (Cells)
Observations (Cells)
Genomic Region
B
CNVneg
CNVpos
UMAP 1
UMAP 2
C
PTPRZ1
EGFR
SOX2
TNR
DSCAM
CD74
C3
HLA-B
ITGAX
ITM2B
MBP
MOG
UMAP_1
UMAP_2
D
TB3926 copy number alteration chromosomal heatmap
References (Cells)
Observations (Cells)
Genomic Region
E
UMAP 1
UMAP 2
F
PTPRZ1
EGFR
SOX2
TNR
DSCAM
CD74
C3
HLA-B
ITGAX
ITM2B
MBP
MOG
G
TB4100 copy number alteration chromosomal heatmap
References (Cells)
Observations (Cells)
Genomic Region
H
UMAP 1
UMAP 2
I
PTPRZ1
EGFR
SOX2
TNR
DSCAM
CD74
C3
HLA-B
ITGAX
ITM2B
MBP
MOG


A
GFAP
AQP4
SLC1A2
SLC1A3
RBFOX3
MEG3
GAD1
SLC17A6
CD74
C3
ITGAX
ITM2B
MBP
MOG
OPALIN
CNP
PDGFRA
DSCAM
TNR
SOX2
B
GFAP
AQP4
SLC1A2
SLC1A3
RBFOX3
MEG3
GAD1
SLC17A6
CD74
C3
ITGAX
ITM2B
MBP
MOG
OPALIN
CNP
PDGFRA
DSCAM
TNR
SOX2
C
GFAP
AQP4
SLC1A2
SLC1A3
RBFOX3
MEG3
GAD1
SLC17A6
CD74
C3
ITGAX
ITM2B
MBP
MOG
OPALIN
CNP
PDGFRA
DSCAM
TNR
SOX2
UMAP_1
UMAP_2
−15 −10 −5 0 5 10 15
3.0
2.5
2.0
1.5
1.0


A
Clusters
tSNE_1
tSNE_2
tSNE_3
0
1
2
3
4
5
6
7
8
B
GFAP
AQP4
SLC1A2
ATP1B2
CLU
SLC4A4
GJA1
GPC5
S100B
GLUL
Expression (logcounts)
0.0
2.5
5.0
7.5
0 1 2 3 4 5 6 7 8
C
0 1 2 3 4 5 6 7 8
Identity
Expressio
2
1
0
-1
-2
D
post-TTT
Primary
Epilepsy
LGG
1
0.5
0
-0.5
-1
0 1 2 3 4 5 6 7 8
E
Astro_6
Enrichment
Astro_6=Negative
Astro_6=Positive
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.007
0 1000 2000 3000 4000
Time
F
Hazard ratio
Astro_1 Negative_ES (N=121) reference
Positive_ES (N=197) 0.99 (0.58 - 1.7) 0.965
Astro_2 Negative_ES (N=229) reference
Positive_ES (N=89) 1.26 (0.89 - 1.8) 0.197
Astro_3 Negative_ES (N=114) reference
Positive_ES (N=204) 0.91 (0.65 - 1.3) 0.599
Astro_4 Negative_ES (N=266) reference
Positive_ES (N=52) 0.83 (0.55 - 1.2) 0.357
Astro_5 Negative_ES (N=77) reference
Positive_ES (N=241) 1.17 (0.88 - 1.5) 0.283
Astro_6 Negative_ES (N=73) reference
Positive_ES (N=245) 1.54 (1.08 - 2.2) 0.018 *
Astro_7 Negative_ES (N=189) reference
Positive_ES (N=129) 1.10 (0.80 - 1.5) 0.551
Astro_8 Negative_ES (N=110) reference
Positive_ES (N=208) 0.85 (0.52 - 1.4) 0.518
age (N=318) 1.02 (1.01 - 1.0) 0.002 **
Events: 270; Global p-value (Log-Rank): 0.00089026
AIC: 2618.87; Concordance Index: 0.59
0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4
G
Protoplasmic score
H
Reactive_1 score
I
Reactive_2 score
J
Ast3
Ast2
Ast1
c-ast1
c-ast2
c-ast3
1
0.5
0
-0.5
-1


A
tSNE_1
tSNE_2
tSNE_3
mgTAM
moTAM
prTAM
Myel1
Tcells
B
Epilesy
LGG
Recurrent
Primary
C
Expression (logcounts)
CD14
GLUL
FOS
CEBPD
SAT1
CD163
MSR1
NHSL1
MS4A4E
FMN1
SORL1
ITGAX
RIN3
HS3ST4
FRMD4A
CST3
MEF2A
DBI
PLXDC2
DOCK4
SKAP1
FYN
CD247
CD96
CD2
Myel1
moTAM
mgTAM
prTAM
Tcell
D
Epilepsy
LGG
Primary
Recurrence
Myel1
prTAM
Tcell
moTAM
mgTAM
1200
1000
800
600
400
200
0
E
Color Key and Histogram
Count
Column Z-Score
-2
-1
0
1
2
mgTAM2
moTAM1
Mono
DC
PrTAM
T.cells
mgTAM
moTAM
Myel1
Tcell
prTAM


A
DAPI
B
nCounts
C
Representative
k=10 segmentation
ST1A1
ST1B1
ST1C1
ST1D1
ST2A1
ST2C1
ST3B1
ST3C1
ST3D1
1mm


A
DAPI
NeuN
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
B
Normalized Cluster Enrichment
C
B
A
X
Y
100
200
300
400
0.2
0.3
0.4
0.5
0.6
0.8
1.0
C
Primary
Recurrent
Control
Contrast_E
FLAIR
Cluster
A
B
C
-2
0
2
enrichment