In Situ Single-Cell Spatial Profiling of Matrisome Gene Expression in Glioblastoma

Arpan De[1†], Santiago A. Forero[1†], Sumod Sebastian[1], John E. Morales[1], Marisol De La Fuente-Granada[1], Jason T. Huse[2], Leomar Y. Ballester[2], Jeffrey S. Weinberg[1], Frederick F. Lang[1], Kadir C. Akdemir[1], and Joseph H. McCarty[1*]

Departments of Neurosurgery[1] and Pathology[2], The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 1004, Houston, TX 77030, USA[1]

[†] These authors contributed equally to this work.

[*] Corresponding Author:

Joseph H. McCarty

Department of Neurosurgery, Unit 1004

University of Texas M. D. Anderson Cancer Center

1515 Holcombe Boulevard

Houston, TX 77030

jhmccarty@mdanderson.org

Running title: The extracellular matrix in malignant glioma

COI Disclosure: We have no conflicts of interest to disclose.

Abstract

Glioblastoma (GBM) is the most common and aggressive primary brain cancer in adults, characterized by robust infiltrative growth and invasion within the neural microenvironment. While various stromal cells, including astrocytes, microglia, and vascular cells in the microenvironment, communicate with GBM cells to promote growth and invasion, relatively less is known about another critical component of the tumor stroma: the extracellular matrix (ECM). The matrisome comprises hundreds of ECM genes and affiliated adhesion and signaling components; however, little is understood about (i) the repertoire of matrisome genes and proteins expressed in GBM, (ii) the stromal and cancer cells of origin that express different matrisome factors, and (iii) how these components are interconnected to drive tumor malignancy. Here, we have used in situ single-cell transcriptomics to analyze the spatial expression of nearly 350 genes that constitute the human brain matrisome. Our efforts have identified several differentially expressed matrisome genes involved in ECM adhesion and signaling between cancer and stromal cells. Additionally, matrisome genes associated with low-grade glioma pathogenesis and progression to high-grade GBM have been identified and spatially mapped. Collectively, these experiments reveal the single-cell origins of differentially expressed ECM biomarkers as well as their spatial expression patterns in brain tumors. These findings may lead to new therapeutic efforts to inhibit matrisome components to benefit patients with malignant brain cancer.

Introduction

Glioblastoma (GBM) is a highly proliferative and invasive primary brain cancer, with little improvement in overall survival over the past few decades [1]. Nearly 100% of patients will succumb to the disease within three years after diagnosis; hence, understanding the cellular and molecular events that contribute to GBM malignancy may lead to new therapeutic strategies to treat or prevent the pathogenesis and progression of the disease. Therapeutic interventions to treat brain tumors have often focused on targeting cancer cells; however, efforts to understand the critical roles of the complex brain stroma in tumor growth and progression have received less attention. Stromal cells in the GBM microenvironment, including reactive astrocytes, microglia and infiltrating immune cells, and angiogenic endothelial cells, play roles in promoting tumor malignancy [2,3]. GBM cells and their stromal partners also produce extracellular matrix (ECM) components that are a major element of the tumor microenvironment. While there have been elegant cellular and molecular studies that have profiled the diverse types of cancer and stromal cell populations in GBM [4,5], relatively less is known about the composition of the various ECM factors in the tumor stroma, and how these factors actively promote cancer initiation, progression and therapy resistance.

ECM adhesion and signaling events are coupled to growth and progression of GBM ; for example, stem-like GBM cells localize to ECM-rich vascular niches where they likely exploit local cues for growth and survival benefits [6]. As GBM cells proliferate they remodel ECM components in vascular basement membranes to locally activate growth factor and cytokine signaling events, leading to pathological angiogenesis and edema due to the breakdown of the blood-brain barrier (BBB) [7]. Invasive GBM cells also deposit new ECM factors in the interstitial matrices around neurons and perivascular spaces to form classic “secondary structures of Scherer” [8]. In addition, patients treated with FDA-approved therapeutics such as

Bevacizumab/Avastin, which blocks vascular endothelial growth factor (VEGF) signaling, invariably develop resistance with recurrent tumors often displaying aggressive patterns of perivascular and perineural invasion [9] that are associated with alterations in cell-ECM adhesion and signaling pathways in the brain microenvironment [10]. Furthermore, matrisome components likely promote therapy resistance; for example, the dense ECM scaffolding may serve as a barrier to impede the penetration of therapeutic agents [11].

There are more than 1,000 genes that encode ECM proteins and interacting components that comprise the human matrisome [12]. The 'core matrisome' includes over 250 genes encoding ECM glycoproteins, proteoglycans, and collagens deposited by cells in basement membranes and interstitial matrices. Additionally, more than 750 matrisome genes encode ECM-associated proteins including secreted growth factors and cytokines as well as affiliated factors such as extracellular enzymes that remodel the ECM and cell surface receptors that mediate adhesion and signaling responses [13]. We understand surprisingly little about which matrisome genes are expressed in the GBM microenvironment or how matrisome proteins facilitate contact and communication between cancer cells and stromal cells in the tumor microenvironment. Hence, new approaches are needed to fully reveal the composition, spatial expression patterns and levels of matrisome components, as well as how the various ECM components act in interconnected scaffolds to regulate tumor malignancy.

In situ single-cell spatial transcriptomics integrates the high-resolution spatial context of tissue histomorphology with single-cell gene expression [14]. This approach allows for quantitative analyses of RNA expression at the single-cell level by mapping their spatial origin within the tissue [15]. Single-cell transcriptomics also allows analysis of heterogeneity within a tumor microenvironment, identifying different cancer cell populations and their spatial relationships with stromal cells [16]. In particular, Xenium spatial single-cell transcript profiling (10X Genomics) utilizes *in situ* hybridization techniques to directly detect and quantify RNA transcripts within

their native tissue environment [17]. This method relies on highly specific probe sets designed to hybridize to target RNA sequences within intact tissue. Probes are designed with unique molecular identifiers and are coupled to fluorescent labels that allow for visualization of transcripts within cells, providing both quantitative measures and spatial context of gene expression.

In this study, we have utilized a high-throughput single-cell spatial transcriptomics platform to profile the expression of nearly 350 matrisome genes in multiple human brain tumor tissue samples. This platform allows for the detailed mapping of ECM gene expression in the tumor microenvironment *in situ*, revealing differential expression and spatial enrichment of matrisome genes in brain cancer cells and various stromal cell types. Collectively, these efforts reveal several novel ECM targets and their specific cells of origin for potential therapeutic intervention to benefit patients with GBM.

Results

We initially performed spatial transcriptome mapping at single cell resolution in a fixed human GBM sample (~1.5 cm^2 area) containing mostly viable tissue. Custom-designed fluorophore-tagged DNA probes recognizing 478 gene products encoding core ECM factors (n=148), ECM-affiliated factors (n=167), integrins (n=23), and other selected markers enriched in tumor and/or brain stromal cell types (n=140) were used to map gene expression. Single cell clustering by uniform manifold approximation and projection (UMAP) analyses identified multiple cell types in the GBM sample based on differential gene expression profiles (Fig. 1A-C). The cancer cell population could be resolved into four subclusters based on expression profiles of matrisome markers, tumor-enriched markers, as well as cell proliferation markers (Fig. 1D). Differentially expressed proliferation markers were enriched in tumor subpopulation 1 (Fig. 1E), whereas the top 10 DEGs in the various matrisome groupings showed enrichment in tumor cell subclusters 1 and 3 (Fig. 1F). Analysis of spatial positions for the four different cancer cell clusters (Fig. 1G) revealed distinct and largely non-overlapping regions within the GBM sample (Supp. Fig. 1). Cancer cells in clusters 1 and 3 were more diffusely located throughout the tumor including in regions containing microvascular proliferation, which is one defining pathological feature of GBM (Fig. 1G). In contrast, cancer cells in clusters 2 and 4 were localized to hypoxic and necrotic regions, with cluster 4 showing association with areas of pseudopalisading necrosis, another unique pathology found in GBM (Fig. 1G).

Principle component analyses revealed distinguishable gene expression profiles for clusters 1 and 3 (Fig. 2A). Some notable genes that showed comparatively enhanced spatial transcript expression in tumor cells residing in cluster 1 include COL20A1, TNR, and FREM2, whereas transcripts for FBLN5, ITGA3, and MUC1 were higher in cluster 3 tumor cells (Fig. 2B, C). Tumor cells in clusters 2 and 4 were distinguishable by principal component analysis (Fig.

2D). In comparison to cluster 4 cells, tumor cells belonging to cluster 2 expressed higher LGALS3 transcripts, indicating prevalence of more pseudopalisading cell features. Tumor cluster 2 cells showed higher abundance of MGP, ITGA5, and SDC4 whereas cluster 4 cells displayed enriched expression of MBP, FREM2, and NCAN transcripts (Fig. 2E, F).

We next performed high-resolution spatial transcript mapping of matrisome gene expression using a human grade 3 astrocytoma sample (~1.5 cm^2 area). Unlike GBM, these lower grade 3 tumors are slower growing and often harbor R132H mutations in the IDH1 gene [1]. Spatial transcriptomic profiling and gene enrichment based single cell clustering by UMAP analyses identified diverse cancer and stromal cell types in the grade 3 tumor (Fig. 3A-C and Supp. Fig. 2). Next, we compared matrisome gene expression across different cell clusters in the grade 3 sample. Like the GBM data, additional in-depth UMAP analyses of cancer cells from the grade 3 astrocytoma revealed the presence of four subclusters based on differential gene expression (Fig. 3D). Several DEGs linked to cell proliferation, e.g., MKI67 and PCNA, were enriched in subclusters 1and 4 in comparison to tumor cell subclusters 2 and 3 (Fig. 3E). Similarly, several DEGs belonging to different ECM categories of the matrisome were more enriched in tumor subcluster 4 (Fig. 3F). Analysis of spatial positioning of cells in the four different cancer clusters revealed largely uniform and overlapping localization across the tumor tissue sample (Fig. 3G and Supp. Fig. 2), which is unlike the distinct regions of cancer cell cluster localization found in the GBM sample (Fig. 2).

We next analyzed matrisome gene expression profiles in the various non-cancerous stromal cell types in the grade 3 astrocytoma and GBM samples. Heat maps identified several stromal cell types based on differentially expressed matrisome genes, including microglia, vascular endothelial cells, neural cells, oligodendrocyte progenitor cells (OPCs) and related oligodendrocytes (OLs) as well as astrocytes (Fig. 4A, B). Given that astrocytomas are derived from transformed astroglial/neural stem cells of origin, we will refer to the non-cancerous

astrocyte population as ‘astrocyte-like’, since it may contain tumor cells that retain expression of astrocyte markers. Furthermore, the microglial population likely contains infiltrating macrophages that express common markers. Among the non-cancerous stromal cells in the grade 3 sample, the ten most differentially expressed genes (DEGs) for ECM glycoproteins were enriched in vascular endothelial cells and in the microglial cell population, whereas overall expression levels of collagen genes were higher in neural cell types and in OPCs/OLs (Fig. 4A). While the top ten differentially expressed integrin and proteoglycan genes were enriched in multiple cell types, ECM-affiliated genes showed highest enrichment in endothelial cells followed by microglia and neural cells (Fig. 4A). In the GBM sample, the top ten DEGs encoding ECM glycoproteins and integrins were mostly enriched in vascular endothelial and microglial cells (Fig. 4B), whereas the top differentially expressed collagen and proteoglycan genes were expressed more broadly across multiple cell types. Selected ECM-affiliated genes showed higher expression in vascular endothelial cells, whereas others were mostly elevated in neural cell types (Fig. 4B). Differential levels of matrisome gene expression between grade 3 and GBM samples were next compared by integrative comparisons of transcript counts in the various stromal cell clusters (Fig. 4C). As shown in Fig. 4D, we detect multiple matrisome transcripts showing elevated expression in GBM versus the grade 3 tumor. Genes including IGFBP2 (ECM glycoprotein), ANXA1 (ECM-affiliated), ITGA7 (integrin), and SRGN (proteoglycan) displayed increased transcript levels in multiple GBM stromal cell types (Fig. 4D). Interestingly, we detect upregulation of specific matrisome factors such as MMRN1, SEMA7A, ITGA2, and ESM1 uniquely in GBM vascular endothelial cells that are not differentially expressed in other stromal cell types (Fig. 4D). Several genes showing increased expression levels in distinct stromal cell types in the grade 3 tumor versus GBM were also identified (Supp. Fig. 3).

ECM glycoproteins of the matrisome play diverse roles in mediating tumor cell adhesion and migration and influencing stromal cell processes like angiogenesis [18]. Therefore, we next

analyzed the cellular transcript levels and spatial expression patterns of 104 genes encoding ECM glycoproteins. UMAP plots confirmed upregulation of select genes, including IGFBP2 and MGP, in diverse cell types in the GBM sample versus grade 3 astrocytoma (Fig. 5A-C). Spatial analysis of IGFBP2 and MGP transcript expression in grade 3 and GBM revealed upregulated expression in tumor cells, endothelial cells, and astrocyte-like populations surrounding blood vessels in GBM (Fig. 5B, C). The same GBM regions showed increased MGP transcripts in endothelial cells and surrounding perivascular cell populations, as compared to the grade 3 sample (Fig. 5B, C). In comparison to vascular endothelial cell-specific expression in the grade 3 tumor, there was markedly elevated FN1 transcript abundance in endothelial cells along with perivascular cells in the GBM sample (Supp. Fig. 4). The matrisome probe panel was also used to profile the expression of 23 different genes encoding integrins, which are the primary cell surface receptors for ECM glycoproteins [19]. The integrin genes ITGA5 and ITGA7 showed significantly enhanced transcript counts in distinct cell types in GBM compared to the grade 3 tumor (Fig. 5D-F). For example, elevated spatial expression of ITGA5 transcripts was detected primarily in GBM endothelial cells whereas ITGA7 transcripts were expressed mainly in GBM cells and astrocyte-like cells (Fig. 5E, F).

Collagens are heterotrimeric proteins that promote tissue integrity and serve as ligands for integrins and other cell receptors. Among the 19 collagen gene transcripts we profiled, COL16A1, COL8A1 and COL2A1 exhibited higher transcript levels in the GBM sample versus grade 3 astrocytoma (Supp. Fig. 5). There was increased spatial distribution of COL16A1 transcripts primarily in vascular endothelial cells and perivascular microglia, and some expression in surrounding tumor and astrocyte-like cells in the GBM sample (Supp. Fig. 5). Similarly, unlike in grade 3 astrocytoma, COL8A1 transcripts showed enhanced expression primarily in perivascular microglial cells in GBM. We also detected increases in COL8A1

transcript expression in GBM versus grade 3 astrocytoma based on UMAP analysis (Supp. Fig. 5).

Profiling of 19 genes encoding proteoglycans, which are a major component of the interstitial ECM in the brain, revealed that various transcripts, including LUM and SRGN, exhibit significantly higher counts in the GBM sample as determined by UMAP analyses (Fig. 6A). We detected enhanced spatial distribution of LUM primarily in GBM endothelial cells and SRGN transcripts mainly in GBM microglial cells and in other stromal cell types (Fig. 6B, C). ECM-affiliated genes encode the largest category of factors in the human matrisome. ECM-affiliated proteins such as proteases regulate ECM remodeling, control the availability of growth factors and cytokines, and influence tissue rigidity via interactions with ECM glycoproteins and proteoglycans [13]. The ECM-affiliated genes ANXA1 and ANXA2 showed robust spatial enrichment in GBM compared to grade 3 tumor sample (Fig. 6D). There was markedly enhanced expression of ANXA1 and ANXA2 transcripts primarily in GBM cells residing in close juxtaposition with vascular endothelial cells (Fig. 6E, F). It is important to note that not all DEGs encoding matrisome components showed elevated expression in GBM versus grade 3 astrocytoma. For example, we detect lower levels of genes encoding specific ECM glycoproteins, collagens, proteoglycans, , and ECM-affiliated components in GBM versus grade 3 astrocytoma (Supp. Fig. 6).

We also analyzed regions of perivascular microgliosis and pseudopalisading necrosis (Supp. Fig. 7), which are unique pathological features of GBM and are absent in lower grade gliomas. Three ECM-affiliated genes (COLEC12, SDC2 and SEMA5A) showed obvious spatial differences in expression (Supp. Fig. 7). In comparison to surrounding astrocyte-like cell populations, LGALS3 showed significantly elevated transcript localization in pseudopalisading tumor cells near necrotic regions. Unlike GBM, the grade 3 astrocytoma sample did not show

such pseudopalisading pathologies and all four select ECM-affiliated genes (COLEC12, SDC2, SEMA5A and LGALS3) did not reveal obvious increases or changes in the spatial distribution of cellular transcripts (Supp. Fig. 7).

We next analyzed matrisome gene expression in a tumor microarray containing punch biopsy samples taken from grade 2 diffuse gliomas (n=9), grade 3 astrocytomas (n=10), and grade 4/GBM (n=7) along with some matching non-cancerous brain tissue samples (n=4). All grade 2 and grade 3 samples harbored R132H mutations in the IDH1 gene, whereas the GBM samples were wild type for IDH1. The top three DEGs in tumor versus normal tissue were identified for each matrisome gene group. In comparison to the matched normal samples, the top three DEGs encoding ECM glycoproteins (SMOC1, ABI3BP and SPARC) showed the highest transcript abundance followed by ECM-affiliated DEGs (PLXNB3, SEMA6A and PLXNB1) in grade 2 diffuse glioma samples (Supp. Fig. 8). All three DEGs encoding proteoglycans (HAPLN2, ACAN and HAPLN1) showed higher transcript enrichment in the grade 2 gliomas in comparison to the matched normal samples. Interestingly, the only differentially expressed collagen gene (COL24A1) showed decreased transcript abundance in perivascular cells of the grade 2 glioma sample (Supp. Fig. 8). In comparison to normal tissue, spatial enrichment for SMOC1, ACAN, and SEMA6A transcripts were evident in cells of grade 2 glioma samples. On the other hand, comparison of the grade 3 astrocytoma to the matched normal sample, revealed the top three DEGs encoding ECM glycoproteins (NTN1, AEBP1 and SPARC) and ECM-affiliated factors (C1QA, LGALS9 and C1QL1) (Supp. Fig. 8). The top three proteoglycan DEGs (SPOCK1, SPOCK3 and HAPLN2) displayed reduced transcript enrichment in single cells of the grade 3 sample when compared to the matched normal sample. Similarly, two differentially expressed collagen genes (COL9A2 and COL16A1) exhibited negative fold-changes in transcript counts for the grade 3 tumor sample (Supp. Fig. 8). Representative images revealing comparative spatial cellular enrichment of transcripts for SPARC (ECM

glycoprotein), COL20A1 (collagen), ITGB4 (integrin), HAPN2 (proteoglycan) and LGALS9 (ECM-affiliated) in grade 3 astrocytoma and matched-normal samples are shown in Supp. Fig. 8. Plots showing variance-based distribution of single cells from grade 2 glioma, grade 3 astrocytoma and corresponding normal samples are shown in Supp. Fig. 9. Also shown are DEG analysis and variance-based single-cell principal component analysis plots from additional punch biopsy samples for matched grade 3 astrocytoma versus normal, and grade4/GBM versus normal from the tumor microarray.

Single cell transcript count-based expression analyses for core matrisome and affiliated genes were next performed with human grade 2 (n=9), grade 3 (n=10), and GBM (n=7) samples. Some core matrisome and affiliated ECM genes revealed reduced transcript levels in GBM in comparison to grade 2 and grade 3 samples (Supp. Fig. 10). Plots showing variance-based comparative distribution of the different patient-derived progressive glioma samples analyzed are shown in Supp. Fig. 10. The majority of the matrisome genes showing significantly elevated transcript abundance in GBM samples belonged to the ECM glycoprotein and ECM-affiliated gene categories (Fig. 7A-D). In comparison to grade 2 diffuse glioma and grade 3 astrocytoma samples, several unique core matrisome and affiliated genes, namely IGFBP2, MGP, IBSP, POSTN and TGFBI (ECM glycoproteins); ANXA1, ANXA2, LGALS1, LGALS3 and SDC1 (ECM-affiliated); ITGA5 and ITGA7 (integrins); COL8A1 (collagens); and LUM and SRGN (proteoglycans), identified to be upregulated in high grade GBM samples (Fig. 7A-D). In comparison to grade 2 and grade 3 samples, elevated spatial enrichment of transcripts for IBSP (ECM glycoprotein), COL27A1 (collagen), ITGB3 (integrin), FMOD (proteoglycan) and LGALS1 (ECM-affiliated) in GBM samples are shown in representative high-resolution images (Fig. 7E). Lastly, by immunofluorescence labeling we detected robust upregulation of annexin-a2 protein expression in GBM versus matched normal brain as well as lower grade glioma samples (Fig. 7F), which further supports the *in situ* spatial transcript profiling results.

Discussion

Resolving the expression patterns and functions for matrisome components of the brain tumor microenvironment requires in-depth understanding of what cell types express which ECM components as well as how these different components mediate adhesion and signaling between stromal cells and cancer cells. Here, we have performed the first *in situ* single-cell spatial transcriptomic mapping of nearly 350 genes using a targeted probe panel to detect expression of core matrisome, ECM-affiliated, and other selected markers in the normal brain and in human glioma samples representing various grades of malignancy. This high throughput strategy has enabled us to dissect the cellular architecture of the tumor and to quantify expression profiles of matrisome genes at single-cell resolution. Interestingly, in GBM we have identified at least four different tumor cell subclusters based on matrisome gene expression. In this study we have done novel spatial analyses, revealing largely non-overlapping locations for these cell types in GBM. In comparison to grade 3 astrocytoma where the cancer cell clusters show overlapping spatial patterns, cell clusters 2 and 4 in GBM are enriched and localize to regions of necrosis and hypoxia that are absent in the grade 3 sample. In contrast, GBM clusters 1 and 3 localize to regions o microvascular proliferation and neuroinflammation, which are also absent in grade 3 tumors. Hence, these defining pathological features of GBM are linked to unique matrisome factors expressed by distinct cancer and stromal cell types. It will be interesting to determine if the multiple cell clusters we have identified are unique or overlapping with sub-populations of GBM cells identified using single cell whole transcriptome sequencing efforts [4,20,21].

Two ECM-affiliated matrisome components that are highly upregulated in GBM versus normal brain and low-grade astrocytoma samples are ANXA1 and ANXA2 [22]. Members of the Annexin A (ANXA) family belong to a larger superfamily of calcium and phospholipid-binding

ECM-affiliated annexins. A prior TCGA analysis revealed that ANXA1 and ANXA2 are upregulated in GBM [23]. Our *in situ* spatial mapping of ANXA1 and ANXA2 revealed highly enhanced transcript localization in different cell types around GBM tumor endothelial cells. Further transcript enrichment and UMAP analyses displayed over-expression of both ANXA1 and ANXA2 genes in nearly all cells of the GBM tissue sample, and similar protein expression patterns were observed for annexin-a2 in GBM core tissue sections from multiple patients. We also found significantly elevated transcript levels of ECM-affiliated genes LGALS1 and LGALS3 in GBM samples in comparison to low-grade gliomas and matched normal samples. LGALS genes encode galectins, a family of ECM-affiliated proteins that form ordered cell surface Gal-glycan lattices critical for engagement of specific glycoconjugates through cell receptor-ligand interactions, clinically relevant in immune responses during inflammation, and in pancreatic cancer [24,25]. LGALS genes have prognostic values and are upregulated in IDH1 wild type tumors, promoting GBM stem cell maintenance [26]. Spatially, LGALS3 was enriched in tumor cells near regions of pseudopalisading necrosis whereas ANXA2 and SEMA5A were enriched in perivascular tumor cells. The ECM-affiliated protein SEMA5A orchestrates RhoGDI/Rac1-mediated invasiveness of motile human glioma cells [27,28]. The transmembrane C-type lectin COLEC12 is aberrantly expressed in several cancers and promotes proliferation, migration, and invasion of gastric cancer cells in vitro, with additional bioinformatic analysis predicting immunosuppressive roles for COLEC12 [29]. Another ECM-affiliated type I transmembrane heparan sulfate proteoglycan SDC2 plays diverse roles during the pathogenesis and progression of various cancers [30]. We detected highly enhanced spatial enrichment of COLEC12, SDC2 and SEMA5A in the GBM sample. Aberrantly high transcript levels of all these genes were observed in tumor cells and tumor-associated microglial, astrocyte-like cells, and neural cell populations proximal to blood vessels in high grade GBM. However, specific

functions for these ECM-affiliated genes in low grade glioma or high grade GBM remain largely unknown and will be topics of future studies.

We have identified several integrins showing differential spatial expression patterns *in situ* across diverse cell types of the tumor microenvironment, with robust enrichment in endothelial cells in both grade 3 astrocytoma and GBM samples. For example, enhanced spatial enrichment of ITGA5 and ITGB1 in tumor endothelial cells, and ITGA7 in perivascular tumor cells and astrocyte-like cells were detected in GBM. Moreover, cellular transcript expression of ITGA4, ITGB3 and ITGA7 significantly increases in low-grade glioma tumors versus the GBM samples. In a recent study, the runt-related transcription factor 1 (RUNX1) which promotes cell proliferation and invasion in GBM was shown to regulate expression of the ECM glycoprotein FN1, collagen COL4A1, and the proteoglycan LUM in the U87 GBM cells [31]. We detected similarly elevated spatial expression of LUM and another novel ECM proteoglycan SRGN in endothelial cells, and different stromal cells including perivascular microglia. While we refer to these cells as microglia, it is likely that this population also includes infiltrating macrophages and/or other immune cells [3]. In comparison to low-grade gliomas, LUM and SRGN transcripts were also found to be upregulated in GBM. Earlier work involving tumor cell lines and tissues samples has reported increased expression and angiogenic roles of specific collagens in GBM growth and progression [32-34]. For example, some collagen genes were found to be upregulated in a syngeneic GBM mouse model [35] and COL8A1 expressed by macrophages/microglia has been identified as a prognostic biomarker crucial for GBM proliferation and migration [36]. We spatially mapped COL8A1 transcripts to microglial cells/macrophages, tumor cells, and neural subpopulations in GBM samples. We also identified expression of the COL16A1 gene in cancer cells and endothelial cells, as well as in microglial and astrocyte-like cell populations of the GBM microenvironment. COL16A1 exhibited lower enrichment in OPC/OLs and was rarely detected in stromal cells of the grade 3 astrocytoma sample. It will be important to determine the

differential roles for these various matrisome components in GBM initiation and growth as well as possible contributions to therapy resistance.

In addition to our data showing that previously characterized ECM gene products, such as IGFBP2 [37], are linked to GBM malignancy, we have discovered several other novel factors with putative functions in GBM. For example, we have identified the MGP gene as upregulated in tumor cells and astrocyte-like cells in GBM. Matrix Gla protein (Mgp) is a member of the family of vitamin k_2-dependent Gla-containing proteins that display high affinity binding to extracellular Ca^{2+} ions [38]. Mutations in the MGP gene are linked to Keutel Syndrome, a rare condition associated with elevated calcium deposition in multiple tissues [39] as well as some coincident vascular pathologies in the brain and retina [40]. Mgp-/- mice develop Keutel Syndrome-like pathologies but survive for several months after birth [41]. It will be interesting to determine roles for MGP in brain tumor growth, progression, and recurrence after therapy.

In summary, we have spatially profiled transcripts of a large array of ECM and related gene signatures belonging to diverse categories of the brain matrisome at the single cell level within the brain tumor tissue. Our study paves the way for additional mechanistic and functional studies of components of the core matrisome, ECM-affiliated components, and related factors. An in-depth understanding of the pathophysiological relevance of the brain tumor matrisome is crucial for determining potential therapeutic targets to benefit patients with GBM and other brain tumors.

Materials and Methods

Human tissue samples

All human tissue samples were collected in accordance with the Institutional Review Board of The University of Texas MD Anderson Cancer Center, Houston, and subsequently de-identified for all experimental studies. Patient-derived samples included grade 2 glioma (n=9), grade 3 astrocytoma (n=11), GBM (n=8), normal (n=4), and tissue matched GBM core and normal (n=3) tissue samples.

Xenium gene panel design and analysis

The panel was designed to detect mRNA expression of genes categorized as core matrisome and matrisome-affiliated (https://matrisomedb.org), in addition to cellular markers relevant to normal brain functions and brain tumor growth/progression. A total of 478 genes were selected and curated (Supplemental File #1) primarily based on single-cell atlas data [42] and the MatrisomeDB (https://matrisomedb.org). Fluorophore-tagged probes were designed by 10x Genomics (Pleasanton, CA). A total of two independent Xenium runs were performed in the manufacturer optimized Xenium analyzer instrument following the protocol provided by 10x Genomics (Pleasanton, CA). All parameters for Xenium analysis are manufacturer-designed and optimized [16]. Summary of metrics obtained from the Xenium *In Situ* pipelines is included in the Supplemental Table 2.

Formalin-fixed and paraffin-embedded tissue blocks were used for microdissection to obtain 5 µm tissue sections which were placed on specialized Xenium slides (Research Histology Core at the Brain Tumor Center, MD Anderson Cancer Center, Houston). Slides were air-dried, incubated at 42°C for 3 hours and kept in the desiccator till ready for further processing. Tissue samples were processed according to the protocol provided by 10x Genomics (Pleasanton, CA), and all steps that included tissue deparaffinization,

permeabilization, probe hybridization, probe ligation, gene amplification, and autofluorescence quenching were performed by stringently following the manufacturer's instructions.

Cell segmentation, image acquisition, and downstream analysis

Grade 3 astrocytoma (n=1), GBM (n=1), and tissue matched normal and GBM core samples (n=3) were used for cell segmentation on the Xenium analyzer with Xenium ranger version 1.7.1.0 (10x Genomics, Pleasanton, CA) which utilizes nuclear expansion distance of 15 µm to establish cell boundaries. Grade 2 glioma (n=9), grade 3 astrocytoma (n=10), GBM (n=7) and non-cancerous brain tissues (n=4) were used for cell segmentation on the Xenium analyzer with Xenium ranger version 2.0.0.10 (10x Genomics, Pleasanton, CA) which utilizes nuclear expansion distance of 5 µm to establish cell boundaries. High-resolution fluorescence images displaying spatial distribution of cellular transcripts were acquired from Xenium Explorer version 3.0 (10x Genomics, Pleasanton, CA). Image acquisition parameters were fixed for all samples. The cell-feature matrix files (h5 file) produced by the Xenium analyzer instrument for each group of samples were used for downstream analysis. Single cell IDs for regions of interest were exported from the Xenium Explorer software (10x Genomics, Pleasanton, CA) as CSV files and used for downstream analysis. Off-instrument downstream analyses were performed on RStudio 4.4.0. H5 matrix files from each xenium experiment were imported using the scCustomize package and CSV files containing cell IDs for regions of interest were imported into R using the csv package (Bergsma 2022). Cluster analyses and differential gene expression were performed using the Seurat package [43]. Pseudo-bulk differential gene expression was performed using the DESeq2 package [44]. Data integration was performed using the Harmony package [45] . Data manipulation and visualizations were performed using the tidyverse, Seurat, and ggplot2 packages.

Cluster analysis

Data normalization, quality control, dimensionality reduction (PCA and UMAPs), and unsupervised cluster analyses for grade 3 and GBM samples were performed in RStudio following the standard Seurat 5 workflow (https://satijalab.org/seurat). Only transcripts within cells were included in downstream analyses. Cluster identity was assigned based on the differential expression of cell type marker genes included in the Xenium targeted gene panel. For our grade 3 and GBM samples, we identified a total of 18 clusters and 15 clusters respectively which were annotated as astrocyte-like, endothelial, microglia, oligodendrocyte (OPC/OL), neural, or tumor subclusters. The FeaturePlot function was utilized to visualize the expression of individual genes in the UMAP.

Data Integration

Grade 3 astrocytoma and GBM data were integrated in R studio by merging the two data sets into one Seurat object and then performing the Harmony integration method which projects all cells into a shared embedding. This allows cells to be grouped by cell type rather than sample identity. Cluster analysis for the integrated data was performed by assessing the differential expression of conserved marker genes following the Seurat workflow for integrated single cell data.

Differential gene expression

Differential gene expression analyses in grade 3 astrocytoma, GBM, integrated grade 3 and GBM data, GBM tumor subclusters (1 vs. 3 and 4 vs. 2), and tissue matched grade 3 vs. normal (n=1), grade 3 vs. normal (n=2) and GBM vs. normal (n=1) were performed using the FindAllMarkers and FindMarkers functions in the Seurat workflow which identifies differentially expressed genes between groups of cells using a Wilcoxon Rank Sum test on normalized data

counts. A log2fold-change (log2FC) greater than 1 or less than -1 and an adjusted p-value (Bonferroni correction for multiple comparisons) of less than 0.05 were used as cutoffs for statistical significance. Differential gene expression analysis for patient brain tumor grade comparisons (grade 2 vs. GBM and grade 3 vs. GBM) was performed following the DESeq2 protocol for pseudo-bulk data (Harvard Chan Bioinformatics Core website; https://bioinformatics.sph.harvard.edu/training/) which runs a negative binomial model on normalized count data and performs hypothesis testing using the Wald test. Log2FC greater than 1 or less than -1 and adjusted p-value (Benjamini-Hochberg correction for multiple comparisons) of less than 0.05 were used as cutoffs for statistical significance.

Z-score calculations

For grade 3 astrocytoma and GBM samples, the top 10 differentially expressed genes for glycoproteins, collagens, integrins, proteoglycans, and ECM-affiliated proteins marker genes were filtered for astrocyte-like, endothelial, microglia, OPCs, and neural cells. The same genes were filtered for all tumor subclusters. All tumor and proliferation markers were filtered for all tumor subclusters. Z-scores were calculated by taking the average transcript count of a gene for each cell type or tumor subcluster minus the average transcript count for all cell types or subclusters and dividing the difference over the standard deviation of all cell types or subclusters.

H&E staining and immunohistochemistry

H&E staining was performed on the slides using standard protocols. Bright-field digital scanning of H&E slides was performed with Aperio AT2 scanner (Leica Biosystems). Immunohistochemical staining was performed on formalin-fixed, paraffin-embedded brain tumor and normal tissue sequential sections from the same tissue blocks used for sectioning during

Xenium sample preparation. The tissue sections were deparaffinized at 65°C and rehydrated in decreasing concentrations of alcohol. Heat-induced antigen retrieval was performed using a Tris-EDTA buffer at pH 9.0. The samples were blocked in 10% serum of the same species as the secondary antibody for 1 hour at room temperature and incubated overnight at 4°C in primary unconjugated antibodies diluted in blocking buffer. Immunohistochemical analyses were performed with the following primary antibodies: rabbit anti-annexin A2 (1:200; catalog #11256–1-AP, Proteintech) and mouse anti-fibronectin antibody (1:300, catalog #66042-1-Ig, Proteintech). Slides were then washed in 1× PBS and incubated with biotinylated secondary antibodies (1:500–1:1000) in the blocking buffer for one hour at room temperature. Following another series of washes in PBS, the slides were incubated in ABC Reagent (catalog #PK-4000, Vector Laboratories) for 30 min at RT. Following more washes with PBS, ImmPACT DAB Substrate (catalog #SK-4105, Vector Laboratories) was added for 5–10 min before rinsing the slides in double-distilled H_2O and counterstaining with hematoxylin for 10–15 s. After rinsing in cold tap water, the slides were then dehydrated in increasing concentrations of alcohol and mounted using mounting media and stored until images were collected. Light microscope images were acquired using a microscope (model BX43, Olympus) with a 20× objective.

References

1. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol *23*, 1231-1251. 10.1093/neuonc/noab106.
2. Broekman, M.L., Maas, S.L.N., Abels, E.R., Mempel, T.R., Krichevsky, A.M., and Breakefield, X.O. (2018). Multidimensional communication in the microenvirons of glioblastoma. Nat Rev Neurol *14*, 482-495. 10.1038/s41582-018-0025-8.
3. Khan, F., Pang, L., Dunterman, M., Lesniak, M.S., Heimberger, A.B., and Chen, P. (2023). Macrophages and microglia in glioblastoma: heterogeneity, plasticity, and therapy. J Clin Invest *133*. 10.1172/JCI163446.
4. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835-849 e821. 10.1016/j.cell.2019.06.024.
5. Zarodniuk, M., Steele, A., Lu, X., Li, J., and Datta, M. (2023). CNS tumor stroma transcriptomics identify perivascular fibroblasts as predictors of immunotherapy resistance in glioblastoma patients. NPJ Genom Med *8*, 35. 10.1038/s41525-023-00381-w.
6. Seano, G., and Jain, R.K. (2020). Vessel co-option in glioblastoma: emerging insights and opportunities. Angiogenesis *23*, 9-16. 10.1007/s10456-019-09691-z.
7. Arvanitis, C.D., Ferraro, G.B., and Jain, R.K. (2020). The blood-brain barrier and blood-tumour barrier in brain tumours and metastases. Nat Rev Cancer *20*, 26-41. 10.1038/s41568-019-0205-x.

8. Venkataramani, V., Yang, Y., Schubert, M.C., Reyhan, E., Tetzlaff, S.K., Wissmann, N., Botz, M., Soyka, S.J., Beretta, C.A., Pramatarov, R.L., et al. (2022). Glioblastoma hijacks neuronal mechanisms for brain invasion. Cell *185*, 2899-2917 e2831. 10.1016/j.cell.2022.06.054.

9. McCarty, J.H. (2013). Glioblastoma resistance to anti-VEGF therapy: has the challenge been MET? Clin Cancer Res *19*, 1631-1633. 10.1158/1078-0432.CCR-13-0051.

10. Jahangiri, A., Nguyen, A., Chandra, A., Sidorov, M.K., Yagnik, G., Rick, J., Han, S.W., Chen, W., Flanigan, P.M., Schneidman-Duhovny, D., et al. (2017). Cross-activating c-Met/beta1 integrin complex drives metastasis and invasive resistance in cancer. Proc Natl Acad Sci U S A *114*, E8685-E8694. 10.1073/pnas.1701821114.

11. Peyton, S.R., Platt, M.O., and Cukierman, E. (2023). Challenges and Opportunities Modeling the Dynamic Tumor Matrisome. BME Front *4*, 0006. 10.34133/bmef.0006.

12. Naba, A. (2024). Mechanisms of assembly and remodelling of the extracellular matrix. Nat Rev Mol Cell Biol. 10.1038/s41580-024-00767-3.

13. Apte, S.S., and Naba, A. (2023). Beyond the matrisome: New frontiers in ECM research. Matrix Biol *115*, 133-138. 10.1016/j.matbio.2022.12.003.

14. Piwecka, M., Rajewsky, N., and Rybak-Wolf, A. (2023). Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease. Nat Rev Neurol *19*, 346-362. 10.1038/s41582-023-00809-y.

15. Gulati, G.S., D'Silva, J.P., Liu, Y., Wang, L., and Newman, A.M. (2024). Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. Nat Rev Mol Cell Biol. 10.1038/s41580-024-00768-2.

16. Janesick, A., Shelansky, R., Gottscho, A.D., Wagner, F., Williams, S.R., Rouault, M., Beliakoff, G., Morrison, C.A., Oliveira, M.F., Sicherman, J.T., et al. (2023). High

resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. Nat Commun *14*, 8353. 10.1038/s41467-023-43458-x.

17. Zhang, Q., Jiang, S., Schroeder, A., Hu, J., Li, K., Zhang, B., Dai, D., Lee, E.B., Xiao, R., and Li, M. (2023). Leveraging spatial transcriptomics data to recover cell locations in single-cell RNA-seq with CeLEry. Nat Commun *14*, 4050. 10.1038/s41467-023-39895-3.

18. Naba, A., Clauser, K.R., Hoersch, S., Liu, H., Carr, S.A., and Hynes, R.O. (2012). The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics *11*, M111 014647. 10.1074/mcp.M111.014647.

19. Hynes, R.O. (2014). Stretching the boundaries of extracellular matrix research. Nat Rev Mol Cell Biol *15*, 761-763. 10.1038/nrm3908.

20. Mathur, R., Wang, Q., Schupp, P.G., Nikolic, A., Hilz, S., Hong, C., Grishanina, N.R., Kwok, D., Stevers, N.O., Jin, Q., et al. (2024). Glioblastoma evolution and heterogeneity from a 3D whole-tumor perspective. Cell *187*, 446-463 e416. 10.1016/j.cell.2023.12.013.

21. Varn, F.S., Johnson, K.C., Martinek, J., Huse, J.T., Nasrallah, M.P., Wesseling, P., Cooper, L.A.D., Malta, T.M., Wade, T.E., Sabedot, T.S., et al. (2022). Glioma progression is shaped by genetic evolution and microenvironment interactions. Cell *185*, 2184-2199 e2116. 10.1016/j.cell.2022.04.038.

22. Maule, F., Bresolin, S., Rampazzo, E., Boso, D., Della Puppa, A., Esposito, G., Porcu, E., Mitola, S., Lombardi, G., Accordi, B., et al. (2016). Annexin 2A sustains glioblastoma cell dissemination and proliferation. Oncotarget *7*, 54632-54649. 10.18632/oncotarget.10565.

23. Xu, H., Wu, X., Dou, Y., and Zheng, W. (2022). The prognostic significance of annexin A family in glioblastoma. Ir J Med Sci *191*, 1539-1547. 10.1007/s11845-021-02737-6.

24. Manero-Ruperez, N., Martinez-Bosch, N., Barranco, L.E., Visa, L., and Navarro, P. (2020). The Galectin Family as Molecular Targets: Hopes for Defeating Pancreatic Cancer. Cells *9*. 10.3390/cells9030689.

25. Tribulatti, M.V., Carabelli, J., Prato, C.A., and Campetella, O. (2020). Galectin-8 in the onset of the immune response and inflammation. Glycobiology *30*, 134-142. 10.1093/glycob/cwz077.

26. Zhu, H., Liu, D., Cheng, L., Liu, J., Wang, G., Li, H., Zhang, Y., Mi, H., Zhang, S., Shu, K., and Yu, X. (2022). Prognostic Value and Biological Function of Galectins in Malignant Glioma. Front Oncol *12*, 834307. 10.3389/fonc.2022.834307.

27. Li, X., and Lee, A.Y. (2010). Semaphorin 5A and plexin-B3 inhibit human glioma cell motility through RhoGDIalpha-mediated inactivation of Rac1 GTPase. J Biol Chem *285*, 32436-32445. 10.1074/jbc.M110.120451.

28. Li, X., Law, J.W., and Lee, A.Y. (2012). Semaphorin 5A and plexin-B3 regulate human glioma cell motility and morphology through Rac1 and the actin cytoskeleton. Oncogene *31*, 595-610. 10.1038/onc.2011.256.

29. Sun, X., Zhang, Q., Shu, P., Lin, X., Gao, X., and Shen, K. (2023). COLEC12 Promotes Tumor Progression and Is Correlated With Poor Prognosis in Gastric Cancer. Technol Cancer Res Treat *22*, 15330338231218163. 10.1177/15330338231218163.

30. Mytilinaiou, M., Nikitovic, D., Berdiaki, A., Kostouras, A., Papoutsidakis, A., Tsatsakis, A.M., and Tzanakakis, G.N. (2017). Emerging roles of syndecan 2 in epithelial and mesenchymal cancer progression. IUBMB Life *69*, 824-833. 10.1002/iub.1678.

31. Cui, X., Huo, D., Wang, Q., Wang, Y., Liu, X., Zhao, K., You, Y., Zhang, J., and Kang, C. (2024). RUNX1/NPM1/H3K4me3 complex contributes to extracellular matrix remodeling via enhancing FOSL2 transcriptional activation in glioblastoma. Cell Death Dis *15*, 98. 10.1038/s41419-024-06481-4.

32. Huijbers, I.J., Iravani, M., Popov, S., Robertson, D., Al-Sarraj, S., Jones, C., and Isacke, C.M. (2010). A role for fibrillar collagen deposition and the collagen internalization receptor endo180 in glioma invasion. PloS one *5*, e9808. 10.1371/journal.pone.0009808.

33. Tsai, H.F., Chang, Y.C., Li, C.H., Chan, M.H., Chen, C.L., Tsai, W.C., and Hsiao, M. (2021). Type V collagen alpha 1 chain promotes the malignancy of glioblastoma through PPRC1-ESM1 axis activation and extracellular matrix remodeling. Cell Death Discov *7*, 313. 10.1038/s41420-021-00661-3.

34. Mammoto, T., Jiang, A., Jiang, E., Panigrahy, D., Kieran, M.W., and Mammoto, A. (2013). Role of collagen matrix in tumor angiogenesis and glioblastoma multiforme progression. Am J Pathol *183*, 1293-1305. 10.1016/j.ajpath.2013.06.026.

35. Mirzaei, R., D'Mello, C., Liu, M., Nikolic, A., Kumar, M., Visser, F., Bose, P., Gallo, M., and Yong, V.W. (2023). Single-Cell Spatial Analysis Identifies Regulators of Brain Tumor-Initiating Cells. Cancer Res *83*, 1725-1741. 10.1158/0008-5472.CAN-22-3004.

36. Liu, M., Liang, W., Su, Y., Wen, Y., Qi, J., Wang, L., Su, S., Zhao, J., Shan, J., and Wang, J. (2024). COL8A1 is a potential prognostic biomarker associated with migration, proliferation, and tumor microenvironment in glioma. Experimental cell research *439*, 114076. 10.1016/j.yexcr.2024.114076.

37. Li, T., Forbes, M.E., Fuller, G.N., Li, J., Yang, X., and Zhang, W. (2020). IGFBP2: integrative hub of developmental and oncogenic signaling network. Oncogene *39*, 2243-2257. 10.1038/s41388-020-1154-2.

38. Schurgers, L.J., Uitto, J., and Reutelingsperger, C.P. (2013). Vitamin K-dependent carboxylation of matrix Gla-protein: a crucial switch to control ectopic mineralization. Trends Mol Med *19*, 217-226. 10.1016/j.molmed.2012.12.008.

39. Cancela, M.L., Laize, V., Conceicao, N., Kempf, H., and Murshed, M. (2021). Keutel Syndrome, a Review of 50 Years of Literature. Front Cell Dev Biol *9*, 642136. 10.3389/fcell.2021.642136.

40. Hur, D.J., Raymond, G.V., Kahler, S.G., Riegert-Johnson, D.L., Cohen, B.A., and Boyadjiev, S.A. (2005). A novel MGP mutation in a consanguineous family: review of the clinical and molecular characteristics of Keutel syndrome. Am J Med Genet A *135*, 36-40. 10.1002/ajmg.a.30680.

41. Marulanda, J., Gao, C., Roman, H., Henderson, J.E., and Murshed, M. (2013). Prevention of arterial calcification corrects the low bone mass phenotype in MGP-deficient mice. Bone *57*, 499-508. 10.1016/j.bone.2013.08.021.

42. Karlsson, M., Zhang, C., Mear, L., Zhong, W., Digre, A., Katona, B., Sjostedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single-cell type transcriptomics map of human tissues. Sci Adv *7*. 10.1126/sciadv.abh2169.

43. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol *42*, 293-304. 10.1038/s41587-023-01767-y.

44. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.

45. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods *16*, 1289-1296. 10.1038/s41592-019-0619-0.

Figure Legends

Figure 1. *In situ* spatial single-cell analysis of matrisome gene expression in GBM. (A); UMAP plot from a human GBM sample showing 6 different clusters of stromal and cancer cell types. Cell clusters are defined and pseudocolored based on differential matrisome and cell-type specific marker expression profiles. **(B);** Graph showing the percentages for the six different stromal and cancer cell types within the GBM sample identified from the UMAP plot. **(C);** A representative region from the GBM sample (upper panel) showing single cells assigned pseudo-colors based on UMAP clustering (A). A magnified image of the boxed area is shown in the lower panel. Scale bars: 200 μm (upper) and 100 μm (lower)**. (D);** UMAP plot showing four tumor sub-clusters identified by in-depth single cell transcript expression profiling (A). **(E);** Heatmap showing expression profiles of tumor cell and proliferation markers in the four GBM sub-clusters. **(F);** Heatmap showing transcript enrichment of select matrisome genes in different tumor cell subclusters. Shown on the horizontal axis are top DEGs from each group: ECM glycoproteins, Collagens, Integrins, Proteoglycans and ECM-affiliated factors. **(G);** Representative H&E (top) and pseudocolored cells (bottom) for four different GBM tissue areas reveals distinct spatial positioning of tumor cells belonging to different subclusters, with clusters 1 and 3 co-localizing with blood vessels and microglia and clusters 2 and 4 showing enrichment in hypoxic and necrotic areas. Scale bars are 500 μm.

Figure 2. GBM cell clusters show distinct spatial positioning and heterogeneous expression of matrisome genes. (A); Variance-based comparative distribution of GBM tumor subclusters 1 and 3 revealing unique gene expression profiles as represented by PCA plot. **(B);** Single-cell transcript analysis of GBM cell subclusters 1 and 3 shows 20 differentially expressed

genes with enhanced transcript levels in tumor subcluster 1 (left) and 20 genes displaying upregulation in cells of GBM subcluster 3 (right). Multiple ROIs (n=4) were analyzed, and all differences are statistically significant ($p < 0.0001$). **(C);** Representative high-resolution single-cell images showing enhanced transcript enrichment of select genes (COL20A1, TNR, and FBLN5) in region 3 (bottom) compared to region 1 (top). Transcripts are shown in bark blue (region 1, top) and yellow (region 3, bottom) with pseudo-colored cell types (indicated alongside). Scale bars are 200 μm. **(D);** Variance-based distribution of GBM tumor subclusters 2 and 4 revealing unique gene expression profiles represented by the PCA plot. **(E);** Single-cell transcript analysis of GBM tumor subclusters 2 and 4 shows differential gene expression profiles with 20 genes showing enhanced transcript levels in tumor subcluster 2 (left) compared to 20 genes displaying upregulation in cells of GBM subcluster 4 (right). Multiple ROIs (n=4) were analyzed, and all differences are statistically significant ($p < 0.0001$). **(F);** Representative H&E (left) and corresponding high-resolution single-cell images (right) showing enhanced transcript enrichment of select genes; MGP and ITGA5 in region 2 (top) compared to region 4 (bottom), and increased levels of FREM2 in region 4 (bottom) compared to region 2 (top). Transcripts are shown in yellow (region 2, top) and dark blue (region 4, bottom) with pseudo-colored cell types (shown alongside). Scale bars are 200 μm.

Figure 3. *In situ* spatial single cell analysis of matrisome gene expression in grade 3 astrocytoma. (A); UMAP plot of a human grade 3 astrocytoma sample showing pseudo-colored cell types based on matrisome and cell-type specific marker gene expression profiles from single cell cluster analysis. **(B);** Graph showing the percentages for the six different stromal and cancer cell types within the grade 3 sample identified from the UMAP plot. **(C);** Representative images of the grade 3 astrocytoma sample showing different cell types assigned

pseudo-colors based on UMAP clustering (A). Magnified image of the boxed area is shown in the bottom panel. Scale bars: 200 μm (upper) and 100 μm (lower)**. (D);** UMAP plot showing four tumor sub-clusters identified by in-depth single cell transcript expression profiling (A). **(E);** Heatmap showing expression profiles of tumor cell and proliferation markers in the four cancer cell sub-clusters. **(F);** Heatmap showing transcript enrichment of select matrisome genes in different tumor cell subtypes. Shown on the horizontal axis are top DEG from each group: ECM glycoproteins, Collagens, Integrins, Proteoglycans and ECM-affiliated factors. (**G);** Representative H&E (top) and pseudocolored cells (bottom) for four different grade 3 astrocytoma areas reveals non-distinct and largely overlapping spatial positioning of tumor cells belonging to the four different subclusters. Scale bars are 500 µm.

Figure 4. Spatial mapping and gene expression analysis of stromal cell types in human grade 3 astrocytoma and GBM. (A); Heatmap showing transcript enrichment of selected matrisome genes in different single cell types in the grade 3 astrocytoma sample. Shown on the horizontal axis are the top DEGs from each group: ECM glycoproteins, collagens, integrins, proteoglycans, and ECM-affiliated factors. **(B);** Heatmap showing transcript enrichment of selected matrisome genes in different single cell types in the GBM sample. Shown on the horizontal axis are top DEGs from each group: ECM glycoproteins, collagens, integrins, proteoglycans, and ECM-affiliated factors. **(C);** Integrated UMAP plots from the human grade 3 astrocytoma sample (left) and GBM sample (right) showing pseudo-colored cell clusters based on matrisome gene expression profiles from single cell analysis. **(D);** Single-cell transcript analysis, presented as lollipop plots, of integrated UMAP subclusters (C) showing differential gene expression profiles for different stromal cells and cancer cell clusters, with all differential gene expression being statistically significant ($p < 0.0001$).

Figure 5. High resolution *in situ* single cell spatial mapping of ECM glycoprotein and integrin gene expression. (A); Integrated UMAP plots showing elevated IGFBP2 (left) and MGP (right) gene expression in multiple cell types in the GBM sample versus the grade 3 astrocytoma sample. **(B, C);** Representative images showing spatial distribution of ECM glycoprotein transcripts IGFBP2 and MGP at single-cell resolution. Cells are pseudo-colored based on UMAP analyses of grade 3 and GBM samples. The right panels show magnified regions from the boxed ROI (H&E) of the grade 3 (B) and GBM (C) samples. In comparison to grade 3 astrocytoma, both glycoprotein gene transcripts show significantly increased expression levels in different cell types proximal to endothelial cells in the GBM sample. PECAM1; cyan, KDR; yellow and ECM glycoprotein gene transcripts; dark blue. Scale bars: 200 μm for H&E images and 50 μm for pseudocolored images. **(D);** Integrated UMAP plots showing elevated ITGA5 (left) and ITGA7 (right) gene expression in multiple cell types in the GBM sample versus the grade 3 astrocytoma sample. **(E, F);** Representative images showing spatial localization of select integrin transcripts ITGA5 and ITGA7 at single-cell resolution (cells pseudo-colored). Magnified regions from boxed ROIs (H&E) of the grade 3 astrocytoma (E) and GBM (F) are shown. In comparison to grade 3 sample, ITGA5 transcripts show higher expression in endothelial cells whereas ITGA7 transcripts show elevated expression in perivascular astrocyte-like cells and cancer cells within the GBM sample. PECAM1; cyan, KDR; yellow and integrin transcripts; dark blue. Scale bars: 500 μm for H&E images and 200 μm for single cell images.

Figure 6. High resolution *in situ* single cell spatial mapping of proteoglycan and ECM-affiliated gene expression. (A); Integrated UMAP plots showing elevated LUM (left) and SRGN (right) gene expression in different cell types in the GBM sample versus the grade 3

astrocytoma sample. Note that LUM expression increases in endothelial cells whereas increased SRGN expression occurs in multiple cell types. **(B, C);** Representative images showing spatial distribution of proteoglycan transcripts LUM and SRGN at single-cell resolution in the magnified region from boxed ROI (H&E) of the grade 3 (B) and GBM sample (C). In comparison to grade 3 astrocytoma sample, LUM shows enhanced transcript levels primarily in endothelial cells whereas SRGN shows transcript enrichment in tumor and different stromal cell types in the GBM sample. PECAM1; cyan, KDR; yellow and proteoglycan gene transcripts; dark blue. Scale bars: 500 μm for H&E images and 200 μm for pseudocolored images. **(D);** Integrated UMAPs showing elevated ANXA1 (left) and ANXA2 (right) gene expression in different cell types in the GBM sample versus the grade 3 astrocytoma sample. **(E, F);** Representative images showing *in situ* spatial distribution of ECM-affiliated transcripts ANXA1 and ANXA2 at single cell resolution (cells pseudo-colored) in the magnified region from boxed ROI (H&E) of the grade 3 tissue (E) and GBM (F). Both ANXA1 and ANXA2 transcripts show significantly increased localization predominantly in tumor cells (dark blue) and other cell types: astrocyte-like (green) and microglia (light blue), as well as in endothelial cells (red) in the GBM sample (bottom). PECAM1; cyan, KDR; yellow and ECM-affiliated gene transcripts; dark blue. Scale bars: 200 μm for H&E images and 50 μm for pseudo-colored images.

Figure 7. Spatial analysis of matrisome gene expression in lower grade gliomas versus GBM. (A); Volcano plot showing DEGs representing different cell markers and matrisome groups compared between grade 2 glioma (n=9) and GBM (n=7) pseudo-bulk samples resected from different patients. Selected genes are highlighted that show statistically significant (above horizontal red dotted line) $\log_2$ fold change ≥ +1.0 or ≤ -1.0 (right or left of vertical blue dotted lines respectively) in cellular transcript counts for GBM in comparison to grade 2 glioma

samples. All analyses were based on pseudo-bulk transcript counts from single cells. **(B);** Pie-chart showing relative representation of matrisome DEGs when grade 2 diffuse glioma and GBM samples are compared based on transcript counts from single cells. Note that only selected DEGs that show statistically significant $\log_2$ fold change ≥ +1.5 are being highlighted with larger and bolder fonts correlating with higher cellular transcript counts in GBM compared to grade 2 samples. **(C);** Volcano plot showing DEGs representing different cell markers and matrisome groups to compare between grade 3 glioma (n=10) and GBM (n=7) pseudo-bulk samples from different brain tumor patients. Selected genes are being highlighted that show statistically significant (above horizontal red dotted line) $\log_2$ fold change ≥ +1.0 or ≤ -1.0 (right or left of vertical blue dotted lines respectively) in cellular transcript counts for GBM in comparison to grade 3 glioma samples. All analyses were based on pseudo-bulk transcript counts from single cells. **(D);** Pie-chart showing relative representation of matrisome DEGs when grade 3 astrocytoma and GBM samples were compared based on transcript counts from single cells. Note that only selected DEGs that show statistically significant $\log_2$ fold change ≥ +1.5 are being highlighted with larger and bolder fonts correlating with higher fold changes in cellular transcript counts in GBM compared to grade 3 glioma. **(E);** Representative images showing *in situ* spatial distribution of transcripts in single cells identified in grade 2 and grade 3 glioma, and GBM samples from different brain tumor patients. Note the overall enhanced enrichment of cellular transcripts of all selected genes; IBSP (ECM glycoprotein), COL27A1 (collagen), ITGB3 (integrin), FMOD (proteoglycan) and LGALS1 (ECM-affiliated protein) in GBM compared to both grade 2 and grade 3 glioma. All transcripts for genes of interest are shown in green, PECAM1 (red), KDR (orange) and DAPI (blue). Cell outlines are shown in white. Scale bars: 50 μm. **(F);** Expression of annexin-a2 protein was analyzed in fixed sections from non-cancerous brain, grade 2 diffuse glioma, grade 3 astrocytoma, and GBM by immunofluorescence. Note that in support of the spatial transcriptomic data for ANXA2

expression (A-D), there is an increase in annexin-a2 protein levels in GBM versus lower grade brain tumor samples.

Figure 1

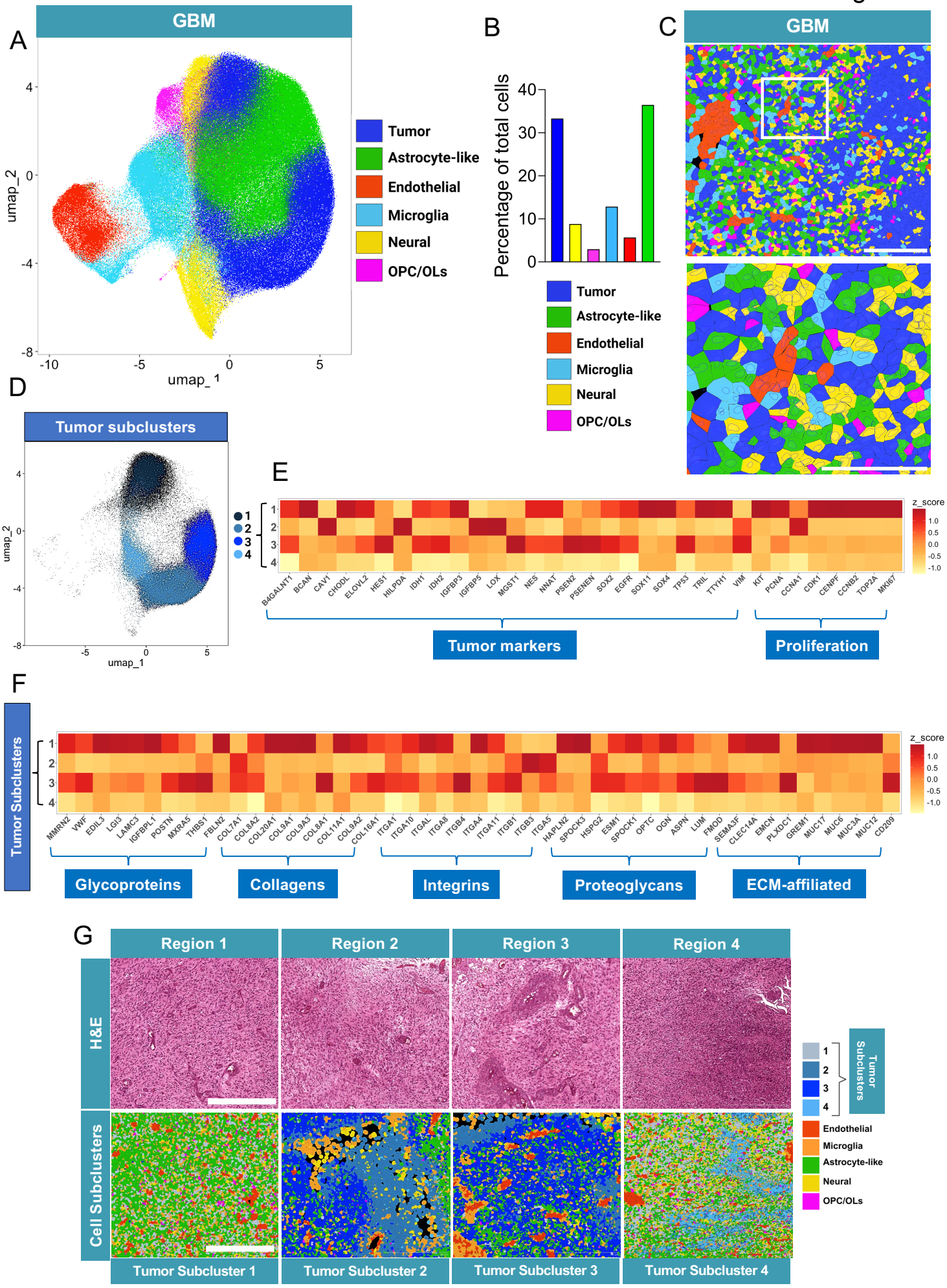

A
GBM
umap_2
umap_1
Tumor
Astrocyte-like
Endothelial
Microglia
Neural
OPC/OLs
B
Percentage of total cells
Tumor
Astrocyte-like
Endothelial
Microglia
Neural
OPC/OLs
C
GBM
D
Tumor subclusters
umap_2
umap_1
E
z_score
Tumor markers
Proliferation
F
Tumor Subclusters
z_score
Glycoproteins
Collagens
Integrins
Proteoglycans
ECM-affiliated
G
Region 1
Region 2
Region 3
Region 4
H&E
Cell Subclusters
Tumor Subclusters
Endothelial
Microglia
Astrocyte-like
Neural
OPC/OLs
Tumor Subcluster 1
Tumor Subcluster 2
Tumor Subcluster 3
Tumor Subcluster 4

Figure 2

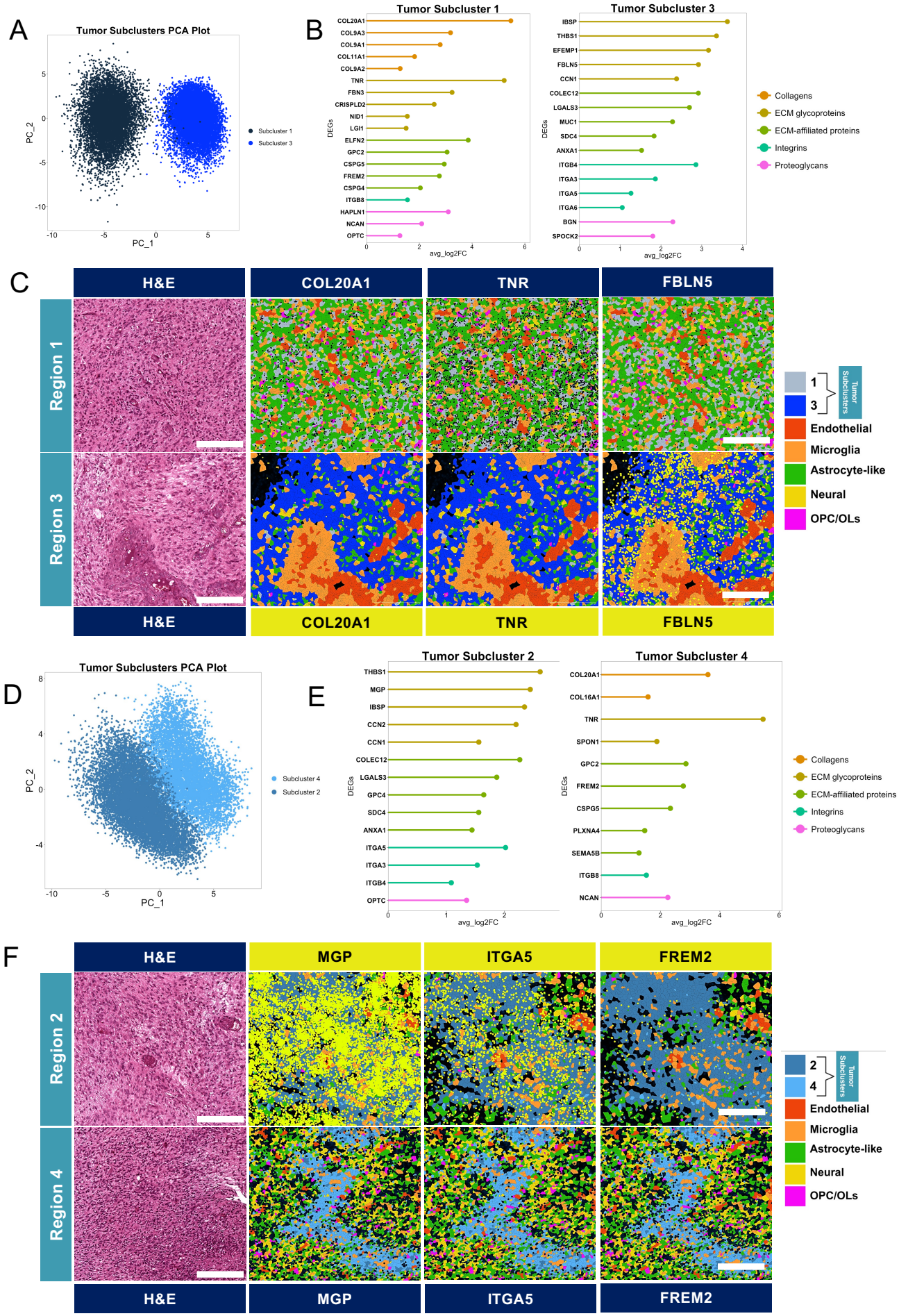

A
Tumor Subclusters PCA Plot
PC_2
PC_1
Subcluster 1
Subcluster 3
B
Tumor Subcluster 1
COL20A1
COL9A3
COL9A1
COL11A1
COL9A2
TNR
FBN3
CRISPLD2
NID1
LGI1
ELFN2
GPC2
CSPG5
FREM2
CSPG4
ITGB8
HAPLN1
NCAN
OPTC
DEGs
avg_log2FC
Tumor Subcluster 3
IBSP
THBS1
EFEMP1
FBLN5
CCN1
COLEC12
LGALS3
MUC1
SDC4
ANXA1
ITGB4
ITGA3
ITGA5
ITGA6
BGN
SPOCK2
Collagens
ECM glycoproteins
ECM-affiliated proteins
Integrins
Proteoglycans
C
H&E
COL20A1
TNR
FBLN5
Region 1
Region 3
Tumor Subclusters
1
3
Endothelial
Microglia
Astrocyte-like
Neural
OPC/OLs
D
Tumor Subclusters PCA Plot
PC_2
PC_1
Subcluster 4
Subcluster 2
E
Tumor Subcluster 2
THBS1
MGP
IBSP
CCN2
CCN1
COLEC12
LGALS3
GPC4
SDC4
ANXA1
ITGA5
ITGA3
ITGB4
OPTC
DEGs
avg_log2FC
Tumor Subcluster 4
COL20A1
COL16A1
TNR
SPON1
GPC2
FREM2
CSPG5
PLXNA4
SEMA5B
ITGB8
NCAN
Collagens
ECM glycoproteins
ECM-affiliated proteins
Integrins
Proteoglycans
F
H&E
MGP
ITGA5
FREM2
Region 2
Region 4
Tumor Subclusters
2
4
Endothelial
Microglia
Astrocyte-like
Neural
OPC/OLs

Figure 3

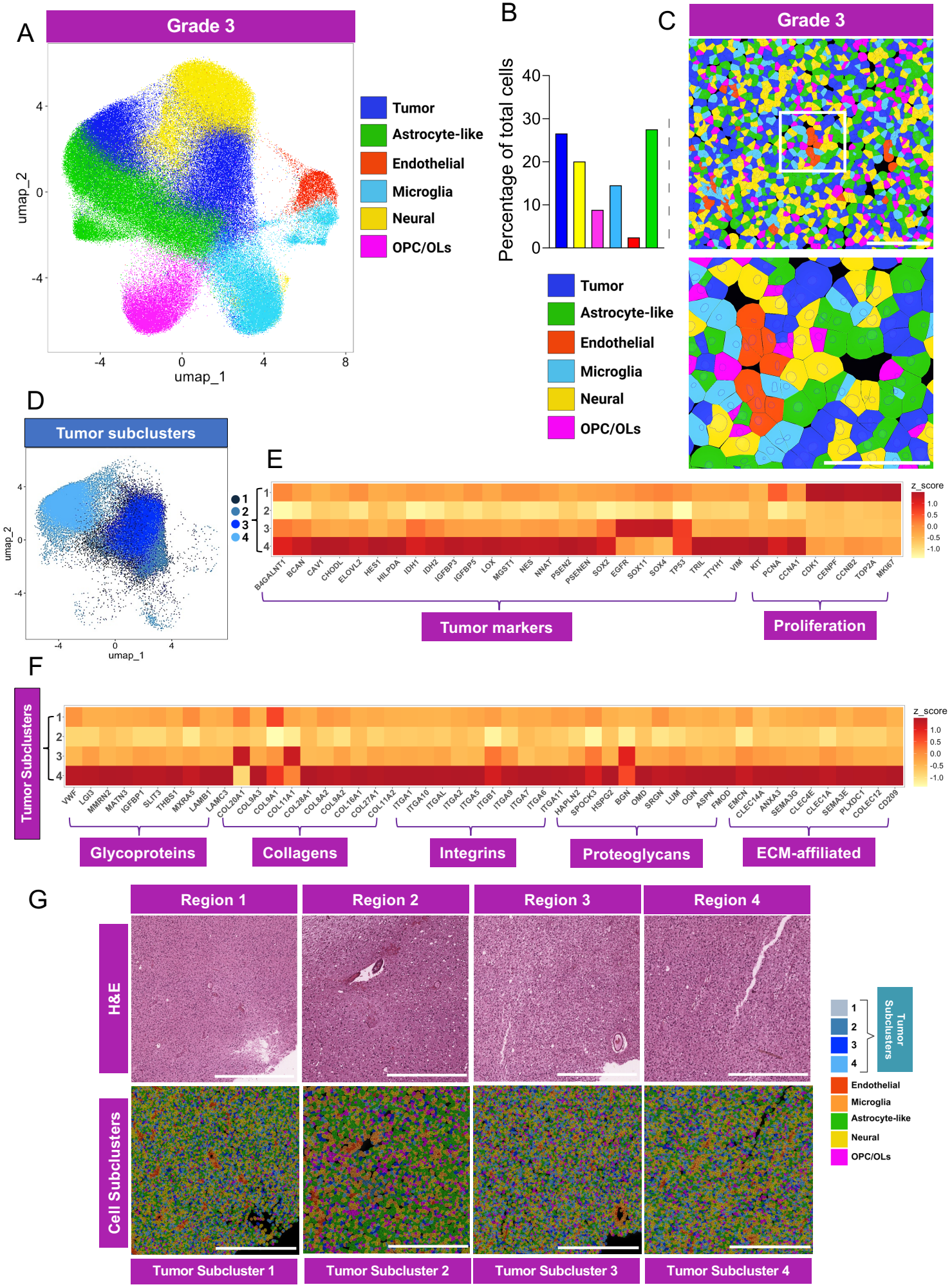

A
Grade 3
umap_2
umap_1
Tumor
Astrocyte-like
Endothelial
Microglia
Neural
OPC/OLs
B
Percentage of total cells
C
Grade 3
D
Tumor subclusters
E
z_score
Tumor markers
Proliferation
F
Tumor Subclusters
Glycoproteins
Collagens
Integrins
Proteoglycans
ECM-affiliated
G
Region 1
Region 2
Region 3
Region 4
H&E
Cell Subclusters
Tumor Subcluster 1
Tumor Subcluster 2
Tumor Subcluster 3
Tumor Subcluster 4

Figure 4

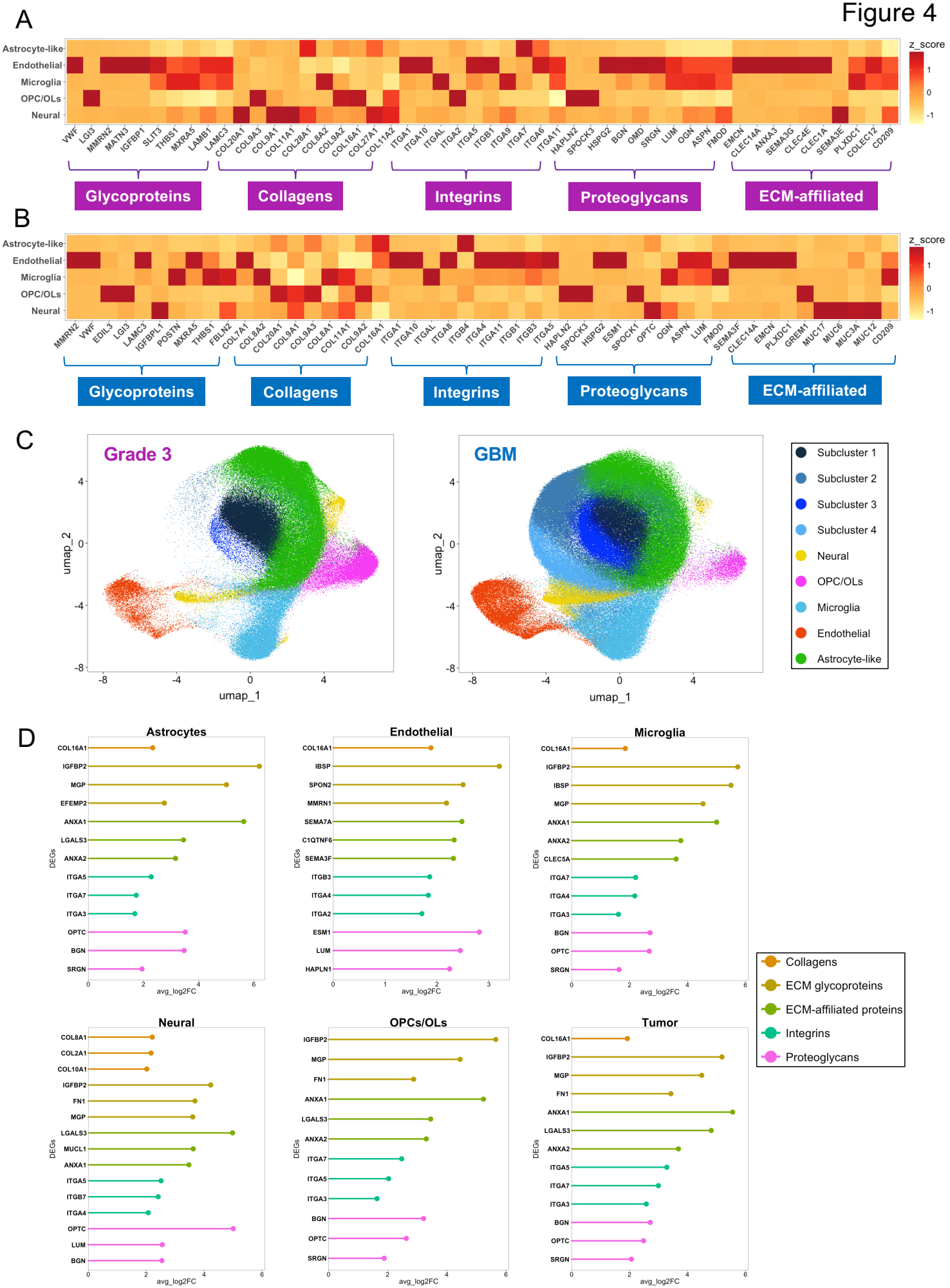

A
Astrocyte-like
Endothelial
Microglia
OPC/OLs
Neural
z_score
1
0
-1
VWF LGI3 MMRN2 MATN3 IGFBP1 SLIT3 THBS1 MXRA5 LAMB1 LAMC3 COL20A1 COL9A3 COL9A1 COL11A1 COL28A1 COL8A2 COL9A2 COL16A1 COL27A1 COL11A2 ITGA1 ITGA10 ITGAL ITGA2 ITGA5 ITGB1 ITGA9 ITGA7 ITGA6 ITGA11 HAPLN2 SPOCK3 HSPG2 BGN OMD SRGN LUM OGN ASPN FMOD EMCN CLEC14A ANXA3 SEMA3G CLEC4E CLEC1A SEMA3E PLXDC1 COLEC12 CD209
Glycoproteins
Collagens
Integrins
Proteoglycans
ECM-affiliated
B
Astrocyte-like
Endothelial
Microglia
OPC/OLs
Neural
z_score
1
0
-1
MMRN2 VWF EDIL3 LGI3 LAMC3 IGFBPL1 POSTN MXRA5 THBS1 FBLN2 COL7A1 COL8A2 COL20A1 COL9A1 COL9A3 COL8A1 COL11A1 COL9A2 COL16A1 ITGA1 ITGA10 ITGAL ITGA8 ITGB4 ITGA4 ITGA11 ITGB1 ITGB3 ITGA5 HAPLN2 SPOCK3 HSPG2 ESM1 SPOCK1 OPTC OGN ASPN LUM FMOD SEMA3F CLEC14A EMCN PLXDC1 GREM1 MUC17 MUC6 MUC3A MUC12 CD209
Glycoproteins
Collagens
Integrins
Proteoglycans
ECM-affiliated
C
Grade 3
GBM
umap_2
umap_1
4
0
-4
-8
Subcluster 1
Subcluster 2
Subcluster 3
Subcluster 4
Neural
OPC/OLs
Microglia
Endothelial
Astrocyte-like
D
Astrocytes
COL16A1 IGFBP2 MGP EFEMP2 ANXA1 LGALS3 ANXA2 ITGA5 ITGA7 ITGA3 OPTC BGN SRGN
Endothelial
COL16A1 IBSP SPON2 MMRN1 SEMA7A C1QTNF6 SEMA3F ITGB3 ITGA4 ITGA2 ESM1 LUM HAPLN1
Microglia
COL16A1 IGFBP2 IBSP MGP ANXA1 ANXA2 CLEC5A ITGA7 ITGA4 ITGA3 BGN OPTC SRGN
Neural
COL8A1 COL2A1 COL10A1 IGFBP2 FN1 MGP LGALS3 MUCL1 ANXA1 ITGA5 ITGB7 ITGA4 OPTC LUM BGN
OPCs/OLs
IGFBP2 MGP FN1 ANXA1 LGALS3 ANXA2 ITGA7 ITGA5 ITGA3 BGN OPTC SRGN
Tumor
COL16A1 IGFBP2 MGP FN1 ANXA1 LGALS3 ANXA2 ITGA5 ITGA7 ITGA3 BGN OPTC SRGN
DEGs
avg_log2FC
Collagens
ECM glycoproteins
ECM-affiliated proteins
Integrins
Proteoglycans

Figure 5

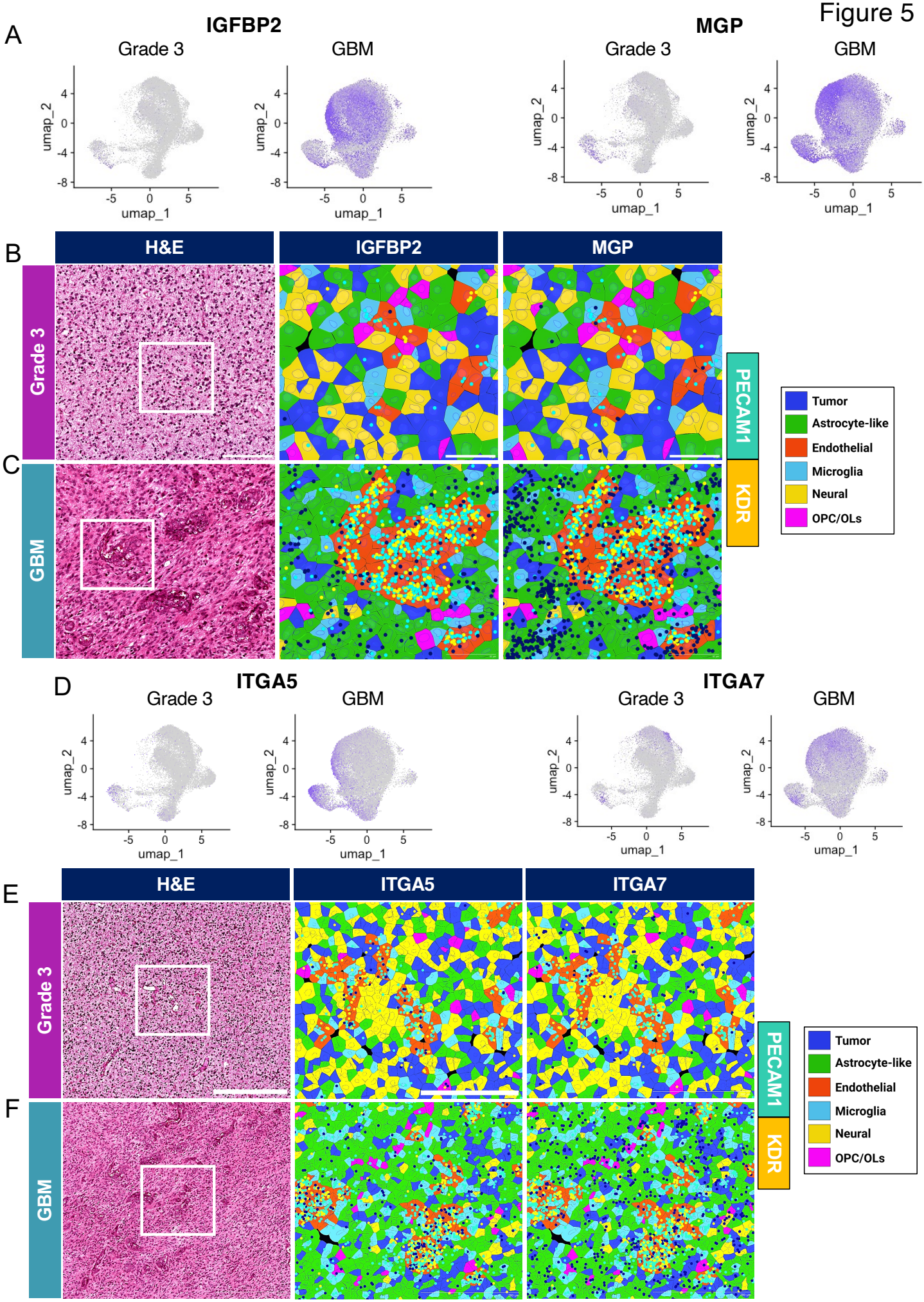

A
IGFBP2
Grade 3
GBM
MGP
Grade 3
GBM
umap_2
umap_1
B
C
H&E
IGFBP2
MGP
Grade 3
GBM
PECAM1
KDR
Tumor
Astrocyte-like
Endothelial
Microglia
Neural
OPC/OLs
D
ITGA5
Grade 3
GBM
ITGA7
Grade 3
GBM
E
F
H&E
ITGA5
ITGA7
Grade 3
GBM
PECAM1
KDR
Tumor
Astrocyte-like
Endothelial
Microglia
Neural
OPC/OLs

Figure 6

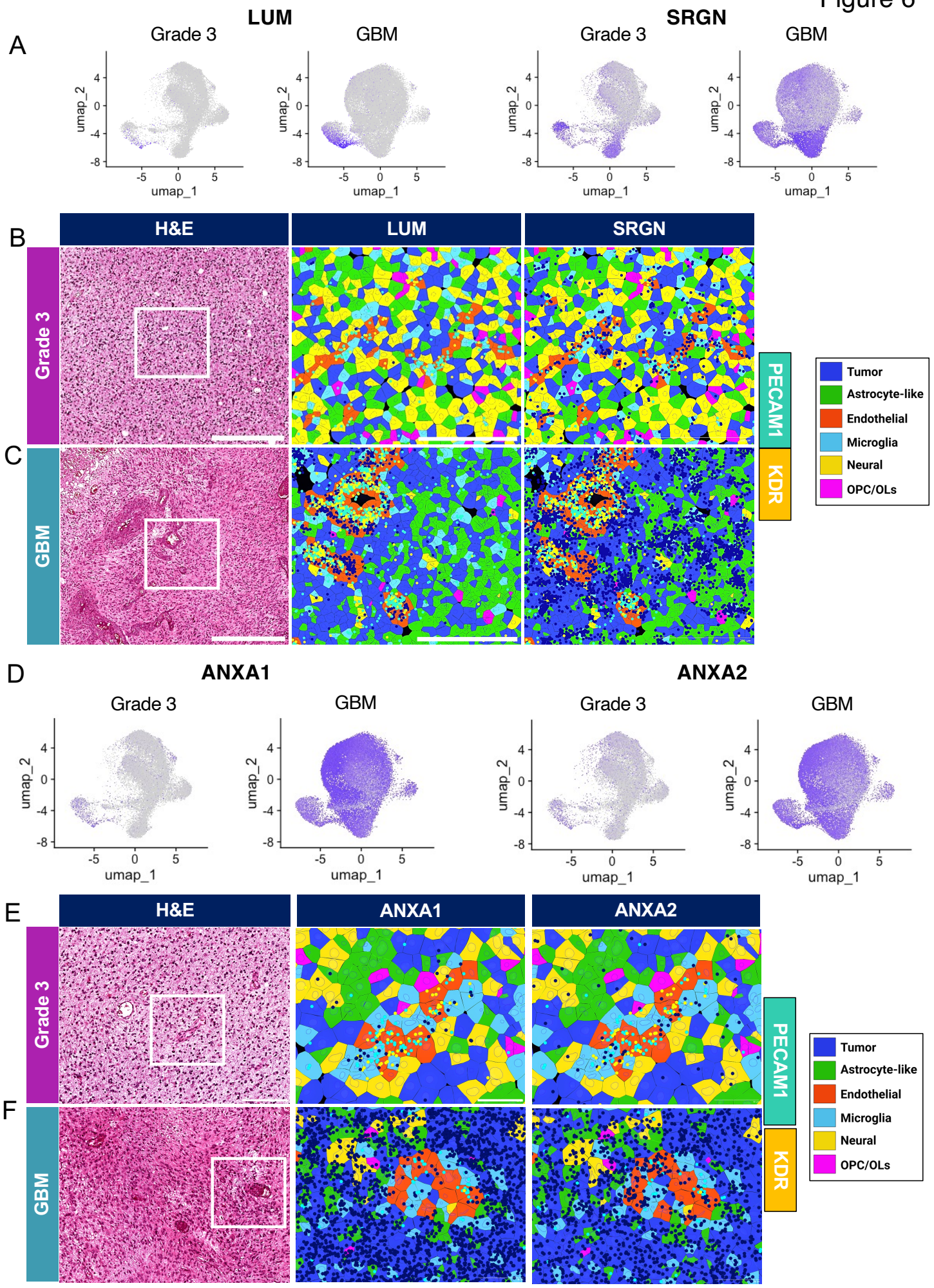
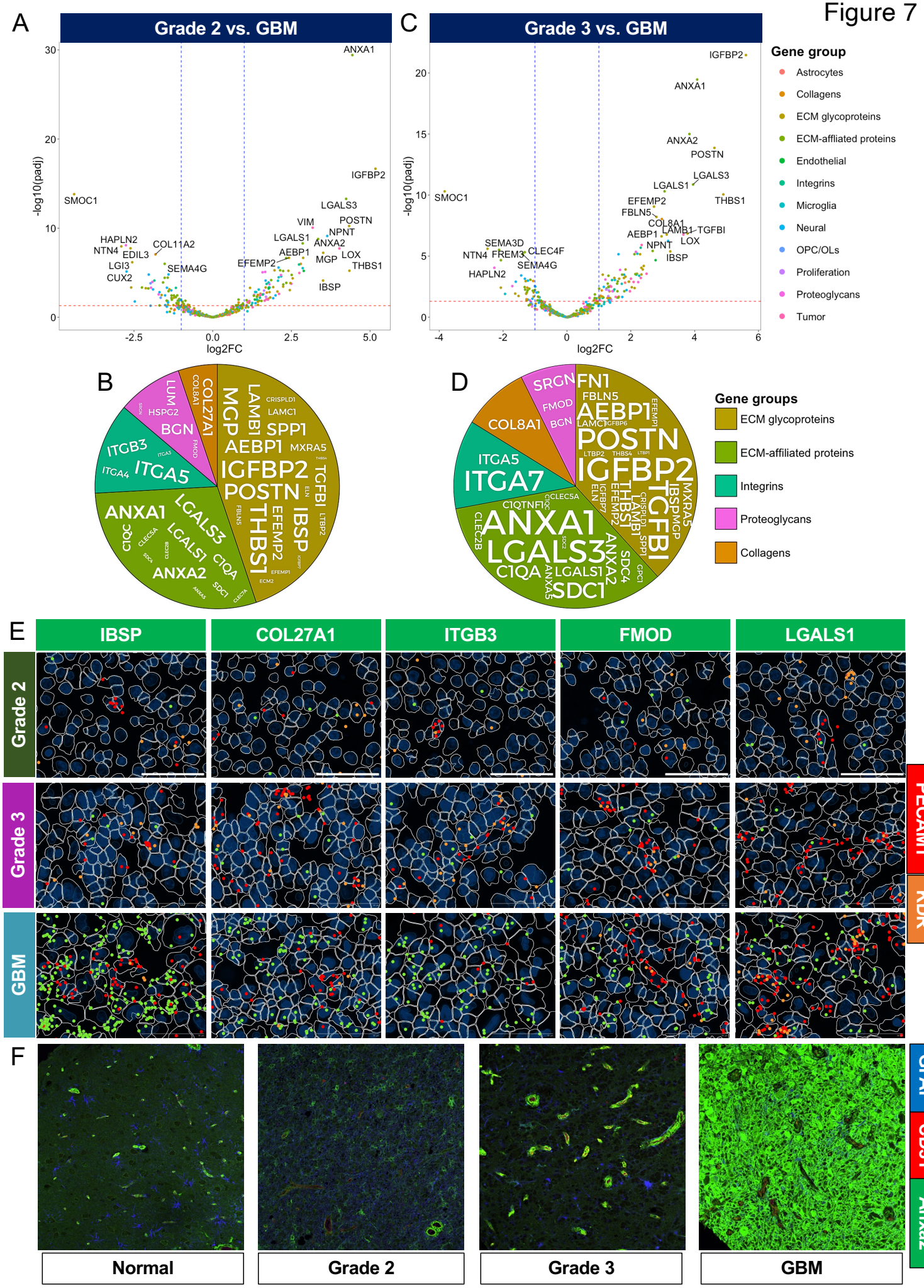

A
LUM
Grade 3
GBM
SRGN
Grade 3
GBM
umap_2
umap_1
B
H&E
LUM
SRGN
Grade 3
C
GBM
PECAM1
KDR
Tumor
Astrocyte-like
Endothelial
Microglia
Neural
OPC/OLs
D
ANXA1
ANXA2
Grade 3
GBM
Grade 3
GBM
E
H&E
ANXA1
ANXA2
Grade 3
F
GBM
PECAM1
KDR
Tumor
Astrocyte-like
Endothelial
Microglia
Neural
OPC/OLs

Figure 7


A
Grade 2 vs. GBM
-log10(padj)
log2FC
ANXA1
IGFBP2
SMOC1
LGALS3
VIM
POSTN
NPNT
LGALS1
ANXA2
AEBP1
LOX
MGP
THBS1
EFEMP2
IBSP
HAPLN2
NTN4
EDIL3
COL11A2
LGI3
CUX2
SEMA4G
C
Grade 3 vs. GBM
IGFBP2
ANXA1
ANXA2
POSTN
LGALS3
LGALS1
SMOC1
EFEMP2
THBS1
FBLN5
COL8A1
LAMB1
TGFBI
AEBP1
LOX
NPNT
IBSP
SEMA3D
NTN4
FREM3
CLEC4F
SEMA4G
HAPLN2
Gene group
Astrocytes
Collagens
ECM glycoproteins
ECM-affliated proteins
Endothelial
Integrins
Microglia
Neural
OPC/OLs
Proliferation
Proteoglycans
Tumor
B
LUM
COL8A1
COL27A1
MGP
LAMB1
CRISPLD1
LAMC1
SPP1
HSPG2
BGN
AEBP1
MXRA5
ITGB3
ITGA4
ITGA5
IGFBP2
POSTN
TGFBI
ANXA1
LGALS3
LGALS1
C1QC
C1QA
ANXA2
THBS1
EFEMP2
IBSP
D
SRGN
FN1
FMOD
FBLN5
AEBP1
COL8A1
BGN
LAMC1
POSTN
ITGA5
ITGA7
IGFBP2
TGFBI
ANXA1
LGALS3
C1QA
LGALS1
SDC1
ANXA2
SDC4
Gene groups
ECM glycoproteins
ECM-affiliated proteins
Integrins
Proteoglycans
Collagens
E
IBSP
COL27A1
ITGB3
FMOD
LGALS1
Grade 2
Grade 3
GBM
PECAM1
KDR
F
GFAP
CD31
Anxa2
Normal
Grade 2
Grade 3
GBM