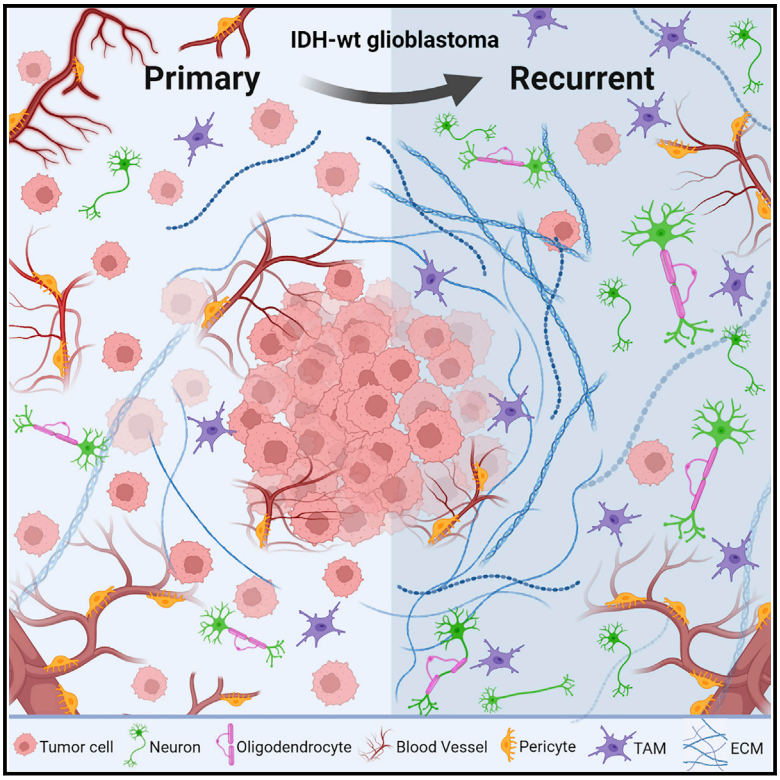
Cancer Cell

Article

Transcriptome analysis reveals tumor microenvironment changes in glioblastoma

Graphical abstract



Highlights

- Glioblastomas evolve by (re)organization of the tumor microenvironment
- Transcriptional glioblastoma subtypes form an interconnecting continuum
- Extracellular matrix genes show prognostic value in recurrent glioblastoma

Authors

Youri Hoogstrate, Kaspar Draaisma, Santoesha A. Ghisai, ..., Martin J. van den Bent, Pierre A. Robe, Pim J. French

Correspondence

y.hoogstrate@erasmusmc.nl (Y.H.), p.french@erasmusmc.nl (P.J.F.)

In brief

Transcriptional changes between primary and recurrent IDH-wt glioblastomas are largely unexplored since resections on recurrent glioblastoma are scarce. Hoogstrate et al. show that longitudinal glioblastoma evolution is shaped predominantly by re-organization of the microenvironment, including a pericyte-expressed extracellular matrix gene set associated with prognosis at recurrence.

Hoogstrate et al., 2023, Cancer Cell *41*, 678–692
April 10, 2023
https://doi.org/10.1016/j.ccell.2023.02.019

CellPress

CellPress
OPEN ACCESS

Cancer Cell

Article

Transcriptome analysis reveals tumor microenvironment changes in glioblastoma

Youri Hoogstrate,[1,25,*] Kaspar Draaisma,[2,23] Santoesha A. Ghisai,[1,23] Levi van Hijfte,[1,3,23] Nastaran Barin,[1,4,23] Iris de Heer,[1] Wouter Coppieters,[5] Thierry P.P. van den Bosch,[6] Anne Bolleboom,[7,8] Zhenyu Gao,[7] Arnaud J.P.E. Vincent,[8] Latifa Karim,[5] Manon Deckers,[5] Martin J.B. Taphoorn,[9,10] Melissa Kerkhof,[9] Astrid Weyerbrock,[11,12] Marc Sanson,[13] Ann Hoeben,[14] Slávka Lukacova,[15] Giuseppe Lombardi,[16] Sieger Leenstra,[8] Monique Hanse,[17] Ruth E.M. Fleischeuer,[18] Colin Watts,[19] Nicos Angelopoulos,[20] Thierry Gorlia,[21] Vassilis Golfinopoulos,[21] Vincent Bours,[22] Martin J. van den Bent,[1] Pierre A. Robe,[2,22,24] and Pim J. French[1,24,*]

[1]Department of Neurology, Erasmus Medical Center, 3015GD Rotterdam, the Netherlands
[2]Department of Neurosurgery, UMC Utrecht, 3584CX Utrecht, the Netherlands
[3]Laboratory of Tumor Immunology, Department of Medical Oncology, Erasmus Medical Center, 3015GD Rotterdam, the Netherlands
[4]Department of Precision and Microsystems Engineering, Delft University of Technology, 2628CD Delft, the Netherlands
[5]Genomics Platform, GIGA Institute, Université de Liège, 4000 Liège, Belgium
[6]Department of Pathology, Erasmus Medical Center, 3015GD Rotterdam, the Netherlands
[7]Deparment of Neuroscience, Erasmus Medical Center, 3015GD Rotterdam, the Netherlands
[8]Department of Neurosurgery, Erasmus Medical Center, 3015GD Rotterdam, the Netherlands
[9]Department of Neurology, Haaglanden Medical Center, 2512VA The Hague, the Netherlands
[10]Department of Neurology, Leiden University Medical Center, 2333ZA Leiden, the Netherlands
[11]Department of Neurosurgery, Medical Center – University of Freiburg, 79106 Freiburg, Germany
[12]Faculty of Medicine, University of Freiburg, 79085 Freiburg, Germany
[13]Sorbonne Université, Inserm, CNRS, UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, AP-HP, Hôpitaux Universitaires La Pitié Salpêtrière - Charles Foix, Service de Neurologie 2-Mazarin, Paris, France
[14]Department of Internal Medicine, Division of Medical Oncology, GROW, Maastricht University Medical Center, 6229ER Maastricht, the Netherlands
[15]Department of Oncology, Aarhus University Hospital, 8200 Aarhus, Denmark
[16]Department of Oncology, Oncology 1, Veneto Institute of Oncology IOV-IRCCS, 35128 Padua, Italy
[17]Department of Neurology, Catharina Hospital, 5623EJ Eindhoven, the Netherlands
[18]Department of Pathology, Elisabeth-TweeSteden Hospital, 5042AD Tilburg, the Netherlands
[19]Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, B15 2SY Birmingham, UK
[20]Systems Immunity Research Institute, Medical School, Cardiff University, CF14 4XN Cardiff, UK
[21]EORTC Headquarters, 1200 Brussels, Belgium
[22]Université de Liège, Department of Human Genetics, 4000 Liège, Belgium
[23]These authors contributed equally
[24]These authors contributed equally
[25]Lead contact
*Correspondence: y.hoogstrate@erasmusmc.nl (Y.H.), p.french@erasmusmc.nl (P.J.F.)
https://doi.org/10.1016/j.ccell.2023.02.019

SUMMARY

A better understanding of transcriptional evolution of IDH-wild-type glioblastoma may be crucial for treatment optimization. Here, we perform RNA sequencing (RNA-seq) (n = 322 test, n = 245 validation) on paired primary-recurrent glioblastoma resections of patients treated with the current standard of care. Transcriptional subtypes form an interconnected continuum in a two-dimensional space. Recurrent tumors show preferential mesenchymal progression. Over time, hallmark glioblastoma genes are not significantly altered. Instead, tumor purity decreases over time and is accompanied by co-increases in neuron and oligodendrocyte marker genes and, independently, tumor-associated macrophages. A decrease is observed in endothelial marker genes. These composition changes are confirmed by single-cell RNA-seq and immunohistochemistry. An extracellular matrix-associated gene set increases at recurrence and bulk, single-cell RNA, and immunohistochemistry indicate it is expressed mainly by pericytes. This signature is associated with significantly worse survival at recurrence. Our data demonstrate that glioblastomas evolve mainly by microenvironment (re-)organization rather than molecular evolution of tumor cells.

INTRODUCTION

Glioblastoma is the most aggressive form of primary brain tumor in adults, with a median overall survival of less than 15 months.[1] The standard of care consists of maximal safe surgical resection in combination with radio- and temozolomide (TMZ) chemotherapy.[1] Tumors invariably recur, and there is no standard of care at disease progression.[2] Although the initial treatment leads to some improvement in survival, the effect is limited and more lasting therapeutic options are needed. Recent efforts and progress in elucidating the genetic makeup of glioblastoma[3,4] have thus far not resulted in the development of more effective treatments.[5] New insights may arise by investigating the temporal evolution of these tumors during therapy, whereby time and treatment resistance-related factors may be uncovered. However, the number of glioblastoma patients undergoing recurrent resection(s) is limited, at least partly because of the short survival time in combination with the unclear clinical benefit of the re-resection.[2]

Studies devoted to glioblastoma tumor evolution showed that primary and recurrent tumors often share large overlap in genomic driver mutations,[6,7] although some show expansion from earlier clones with limited overlap.[8,9] Some studies reported a higher frequency of *MET* and *PDGFRA* amplifications, *CDKN2A/B* deletions, and gains of mutations in *NF1*, *TET2*, *DNMT3A*, *TERT*, and *LTBP4*, although such changes occurred in a minority of samples.[6,9,10] A subset of recurrent tumors with large genetic differences showed a hypermutated phenotype due to mutations in genes linked to DNA mismatch-repair (MMR; e.g., *MSH1*, *MSH2*, *MSH3*, *MSH6*, *POLD3*, and *POLE*). While the resulting mutations are unique per tumor, acquiring this hypermutation phenotype is considered an evolutionary path.[6–12]

Glioblastomas, categorized as IDH-wild type (IDH-wt) or IDH-mutant (IDH-mut), are at the transcriptional level classified into three subtypes: classical (CL), mesenchymal (MES), and proneural (PN).[7] The CL subtype is linked to *EGFR* alterations, the PN subtype to *TP53* mutations and *PDGFRA* alterations, and the MES subtype to *NF1* alterations, and it displays increased proportions of tumor-associated macrophages and/or microglia (TAM) at recurrence.[7] The MES subtype signature includes genes related to the extracellular matrix (ECM). Single-cell RNA sequencing (scRNA-seq) has revealed tumor-intrinsic transcriptional programs reflecting neural progenitor (NPC) and oligodendrocyte progenitor (OPC) cell-like states.[13]

Because expression data can quantify higher-level processes and identify epigenetic changes, investigation of tumor evolution on the transcriptional level may yield novel insights into glioblastoma progression. Here, we provide a large cohort of matching primary-recurrent IDH-wt glioblastoma samples (European Organisation for Research and Treatment of Cancer [EORTC] Study 1542, Glioblastoma, Stability of Actionable Mutations: *G-SAM*). Following RNA sequencing (RNA-seq) with comprehensive differential gene expression (DGE) analysis, we show differences in glioblastoma composition. These are validated on RNAg-seq data from the international Glioma Longitudinal AnalySiS Consortium (GLASS) consortium,[10,14,15] single-nucleus RNA-seq (snRNA-seq), immunofluorescence (IF) and RNA in situ hibridization (RNA-ISH).

RESULTS

Dataset

Of the 172 patients initially included in the G-SAM dataset, all were treated with radiotherapy (RT) and 162 with TMZ.[6] Of the 10 patients who did not receive TMZ, three had *MGMT* promoter methylation status available and were unmethylated. No patients were treated with tumor-treating fields. Eight patients were treated with the anti-angiogenesis agent bevacizumab prior to recurrence. Five samples were taken from biopsies and the remaining from resections. Detailed clinical parameters are provided in Table S1.

After quality control, the final G-SAM RNA-seq dataset was composed of 322 RNA-sequenced IDH-wt tumor samples. Of these, 35 samples had tumor purities <15% and were removed from subtype and differential expression analysis, resulting in 287 samples from 165 patients and 122 complete pairs. Matching targeted exome-sequencing data were sequenced and made available earlier.[6] Our study was extended with 245 IDH-wt grade IV glioblastoma samples from 138 patients of the GLASS cohort (Table S1).[14,15] Of these, 216 samples from 129 patients (87 complete pairs) had a tumor purity >15%.

Transcriptional subtypes

We first assessed transcriptional subtype switching[7] at tumor progression. Classification into discrete classes, however, lacks nuance; for instance, if subtype transitions are borderline or somehow unconfident. We therefore examined the intrinsic subtypes in their relation to tumor progression from a continuous perspective.

The original determination of subtypes was done by performing non-negative matrix factorization (NMF) on 7,425 genes, of which 3 × 50 genes were used to classify external samples using ssGSEA.[7] We used ssGSEA, NMF, and the 3 × 50 gene set as foundation for defining our Glioblastoma Intrinsic Transcriptional Subtype (GITS) space. NMF was performed on the 3 × 50 genes for samples with a purity of 15% or higher. Principal-component analysis showed that the three factors of the NMF's H-matrix were made up of two independent components (Figures S1A–S1D). These two components formed the final grid of our GITS space.

We used ssGSEA-determined subtypes as training labels to define decision boundaries and to determine discrete class labels within the GITS space using Support Vector Machine (SVM). The SVM classifier showed high concordance between the original subtypes, although some shifts near the decision boundaries were observed (Figures S1E and S1F). Importantly, the lower-dimensional GITS space showed that the three subtypes did not form fully isolated clusters but rather formed a continuous spectrum (Figures 1A–1C and S1). Higher tumor purities were observed in CL (Figure S1G). Also, changing between the original 7,425 and 3 × 50 gene sets and using ssGSEA enrichment scores instead of NMF (Figure S1H) did not resolve into isolated clusters. NMF and ssGSEA scores correlated strongly ($R > 0.95$, Figures S1I–S1K). This interconnection underlines that a fully discrete interpretation of subtypes may oversimplify these continuous data.

Distances in the GITS space between resections from the same patient represent its longitudinal subtype evolution. The distances between primary and matching recurrent resections

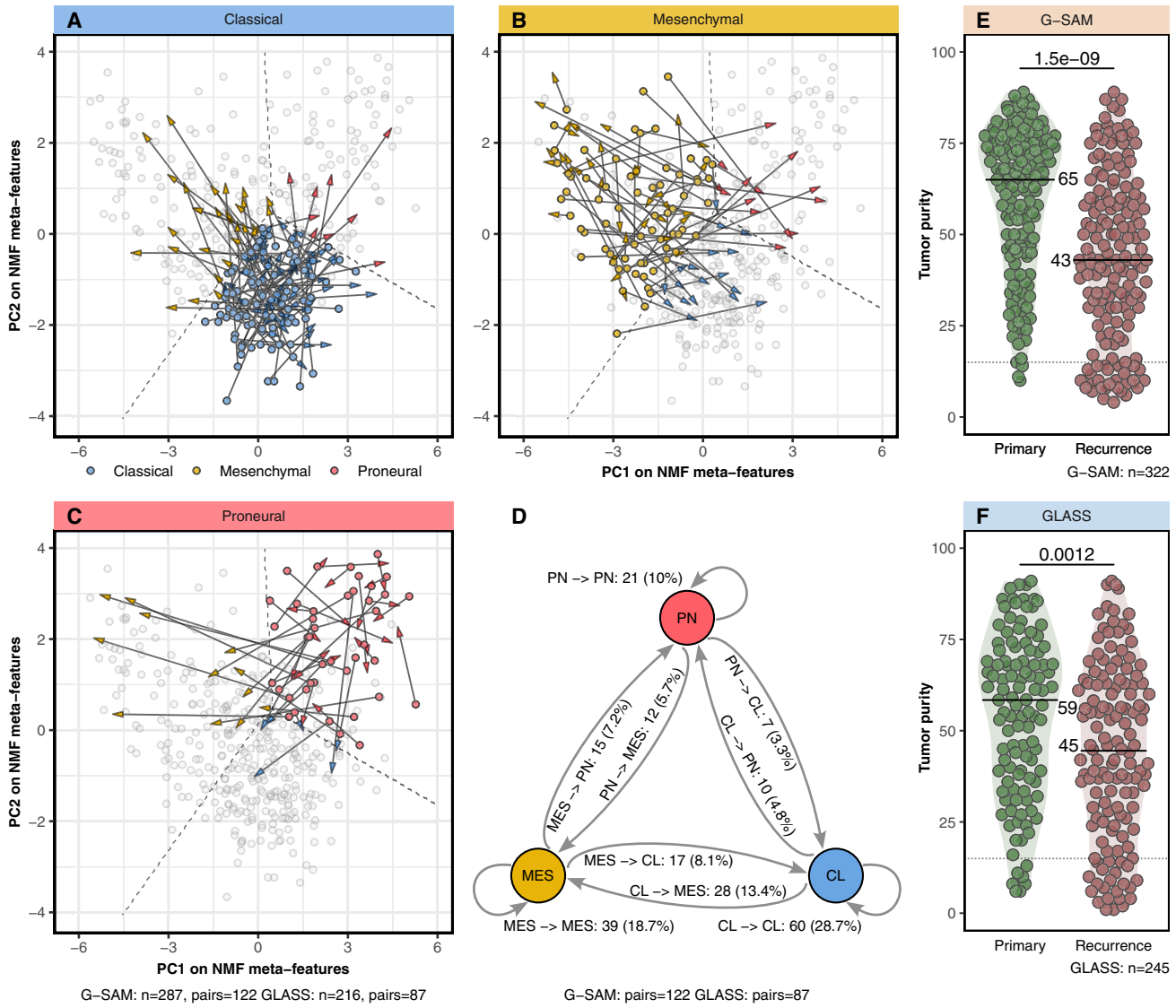


Figure 1. Transcriptional subtypes and longitudinal decrease of tumor purities

(A–C) Directional change of subtype composition only for patients with matching resections, indicated with arrows through the GITS space, separated by the initial resection being classified as classical (A), mesenchymal (B) or proneural (C).

(D) Subtype transitions visualized in a conventional discrete manner. The tumor purities in the G-SAM (E) and GLASS (F) datasets (Wilcoxon rank-sum test). See also Figure S1 and Table S1.

were significantly smaller than between randomly chosen, non-matching, recurrent resections (p = 1.001e−15, Wilcoxon test, Figure S1L). This indicates that, in a continuous sense, tumor pairs resemble each other more closely than random pairs, which is in line with the observation that the majority of glioblastomas retain their subtype.[9,16]

Proportionally, PN most often switched at recurrence (47%), more often to MES (30%, n = 12) compared with CL (17%, n = 7, Figures 1A–1D). Twenty-four percent of the MES tumors switched to CL (n = 17) and 21% to PN (n = 15). CL tumors preferentially switched to MES (29%, n = 28), while 10% switched to PN (n = 7). For all subtypes, Euclidean distances of both transition types are significantly larger than those that remain stable (Figures S1M–S1O). In brief, the non-MES tumors that switched, switched preferentially but not exclusively to MES, while MES tumors switched with a comparable incidence to both CL and PN. In agreement with Wang et al., MES tumors had a significant but not exclusively lower purity (p = 2e−16, Wilcoxon Rank-Sum test, Figure S1P). Hypermutated glioblastomas could arise from all three subtypes and did not show a progression preference toward any subtype (Figure S1Q).

Tumor purity and DGE analysis

We estimated the tumor purity for all samples since differences in tumor purity between primary and recurrent tumors will affect DGE analysis. This was done by finding the best fit between the observed log2 transformed fold change (log2FoldChange) of segments identified by CNV profilers and the expected

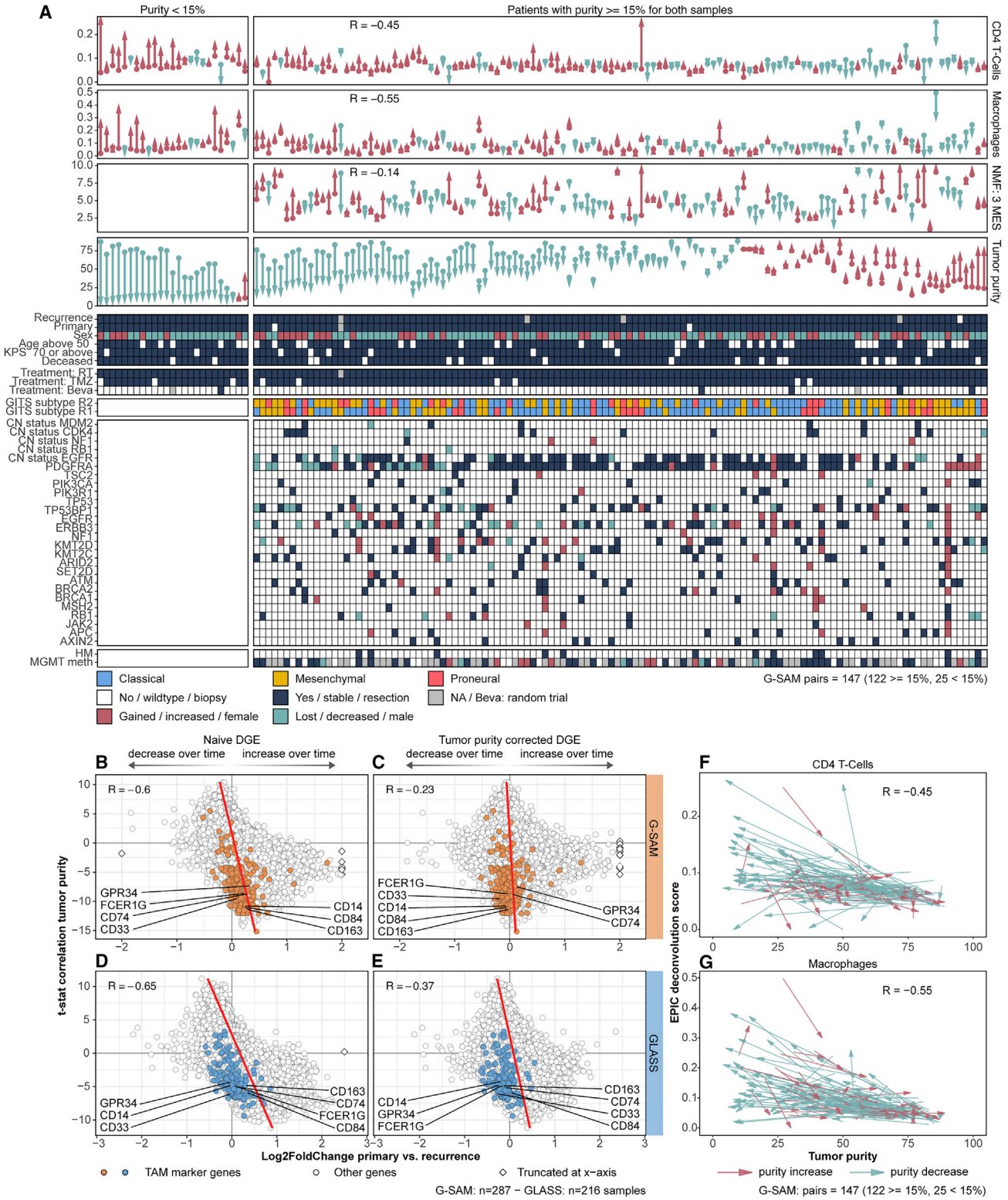


Figure 2. TAM marker genes increase opposed to tumor purity

(A) Changes in immune scores, tumor purity, and the MES representing NMF meta-feature in matching tumor pairs. Pearson correlations compare the change in score with change in purity. Since low-purity samples should yield the highest composition of opposed cell types, these samples are explicitly included to represent data in relation to opposed cell type composition.

(legend continued on next page)

log2FoldChange for all tumor percentages ranging from 1% to 100% for a single loss, single gain, and double gain (STAR Methods). Tumor cell percentages for the 35/245 GLASS samples lacking CNV data were imputed (STAR Methods) using the expression data in a random forest model trained on samples with available DNA sequencing data.

Regions of highest tumor content were marked by the central neuropathologist on H&E-stained slides; DNA/RNA was isolated by macrodissection of these regions on consecutive unstained slides. Despite having selected for these regions, our tumor purity analysis demonstrated that tumor cell percentages differed between primary and recurrent resections, with the purity being significantly lower in recurrent resections ($p = 1.5e{-}09$, Wilcoxon test, Figure 1E). A similar difference was observed in GLASS data ($p = 0.0012$, Wilcoxon test; Figure 1F). This difference in tumor purity across resection types therefore forms a major confounding factor in further analyses. It was not specific for the MES or any other subtype, although patients with the highest stability in tumor cell percentage were mostly CL (Figures 2A and S1P).

To comprehend changes in gene expression by tumor cells specifically, we deconvolved the difference in tumor percentage from the temporal changes in gene expression. This was done by assessing whether the correlation between expression levels (RNA) with tumor cell percentage (DNA) and the difference in per-gene expression between primary and recurrent resections (naive DGE) was independent. We found a clear trend between the outcomes of both tests indicating that the DGE results are dependent on the decreased tumor purity (Figures 2B–2E). Plotting the differences in expression by resection (x axis) compared with the correlation of expression with tumor purity (y axis) highlights the associated dependency. To demonstrate the importance, we specifically analyzed genes located on the commonly gained chromosome 7 (chr7) and commonly lost chr10. Genes located at chr7 are, on average, positively correlated with tumor purity, while genes located at chr10 are on average negatively correlated with tumor purity (Figure S4A). More specifically, the log2FoldChange following naive DGE of the genes located on chr7 are significantly lower than the remaining chromosomes (adjusted $p = 1e{-}14$, t test), while those on ch10 are significantly higher (adjusted $p = 2e{-}16$, t test, Figure S2B). Thus, if not adequately corrected, the difference in tumor purity between primary and recurrent resections will bias DGE.

Tumor cell percentage is anti-correlated with immune cell fraction

Genes specifically expressed in TAM[17] were anti-correlated with tumor purity (*CD14*, $R = -0.27$; *CD74*, $R = -0.31$; *CD163*, $R = -0.3$; *CD33*, $R = -0.36$; *CD84*, $R = -0.35$; *FCER1G*, $R = -0.30$; *GPR34*, $R = -0.29$, Figures 2B–2E). These cell-type marker genes were located near the trend line of the confounding tumor purity effect, which indicates that they are a reflection of lower tumor content and not specifically upregulated by tumor cells at progression. As recurrent tumors generally had a lower tumor purity, we considered this a strong indication that (on average) TAM comprise a larger proportion of the recurrent tumor bulk.

No associations with genomic abnormalities, the hypermutation phenotype or *MGMT* promoter methylation could explain the increase (Figure 2A). Dedicated immune cell deconvolution software[18] confirmed that macrophage and CD4 T cell abundance correlated inversely with change in tumor cell percentage ($R = -0.45$, $R = -0.55$, respectively) and high tumor cell percentages were characterized by low immune cell scores (Figures 2A, 2F, and 2G). These changes in microenvironment-derived immune cell fractions did not associate significantly with time between primary and recurrent tumors (Figure S2C).

Corrected DGE analysis

After linear correction for the total fraction of non-malignant cells (Figures 2C and 2E), we identified 722 genes differentially expressed between primary and recurrent resections (adjusted $p \leq 0.01$ and $|\text{log2FoldChange}| > 0.5$) within the G-SAM dataset. Of these, 484 also had a $|\text{log2FoldChange}| > 0.5$ in the same direction, in the GLASS dataset.

Remarkably, the 484 differentially expressed genes did not include any of the hallmark genes associated with glioblastoma initiation or progression (Figure S2D). *CDKN2A/B* expression decreased after tumor progression (log2FoldChange: −0.48 G-SAM; −0.16 GLASS) but this difference was not significant (adjusted $p = 0.26$ G-SAM, adjusted $p = 0.80$ GLASS). This relatively small decrease in expression likely is a reflection of the *CDKN2A/B* status changes in only in a limited number of patients. In the matching G-SAM DNA data for example, *CDKN2A/B* status was lost at recurrence in 30 out of 146 patients,[6] whereas DNA level analysis of GLASS indicated that *CDKN2A/B* losses in IDH-wt glioblastoma are typically already present at primary disease.[14] Expression of *MGMT*, notorious for inducing TMZ resistance,[19,20] did not significantly change in both datasets (log2FoldChange: 0.149 G-SAM, 0.154 GLASS). This is in concordance with observed changes in its promoter methylation (14 out of 104 gained; 10 out of 104 lost).[6] The only gene linked to glioblastoma[21] with a significant difference in expression was *NODAL* (down).

Because the *EGFRvIII* mutation has features that make it a potentially interesting target for therapy, we analyzed the proportion of *EGFRvIII* relative to total *EGFR* in relation to tumor purity for both resections (Figure 2E). This showed stable fractions but also fractions of *EGFRvIII* that increased or decreased, with no consensus toward a specific direction. The changes did not show a dependency on tumor purity.

An increase in neurons and oligodendrocytes and a decrease in endothelial cells at tumor progression

To elucidate underlying mechanisms regulating the differentially expressed genes, we aimed to perform a combined cluster and

(B–E) The outcomes of naive and purity-corrected DGE analysis and correlation to tumor cell percentage, which for naive analysis are dependent. G-SAM: naive (B) and corrected (C); GLASS: naive (D) and corrected (E). In both datasets, markers for TAM defined by McKenzie (colored dots) are part of the genes most strongly anti-correlated with tumor purity. Genes with an absolute log2FoldChange larger than 2.5 were limited to (−)2.5 and shown as diamond (truncated). Pearson correlation coefficients are provided.

(F and G) Immune deconvolution scores in relation to tumor purity. Pearson correlation coefficients compare the changes over time. Low-purity samples are included for the same reason as specified in (A). See also Figure S2.

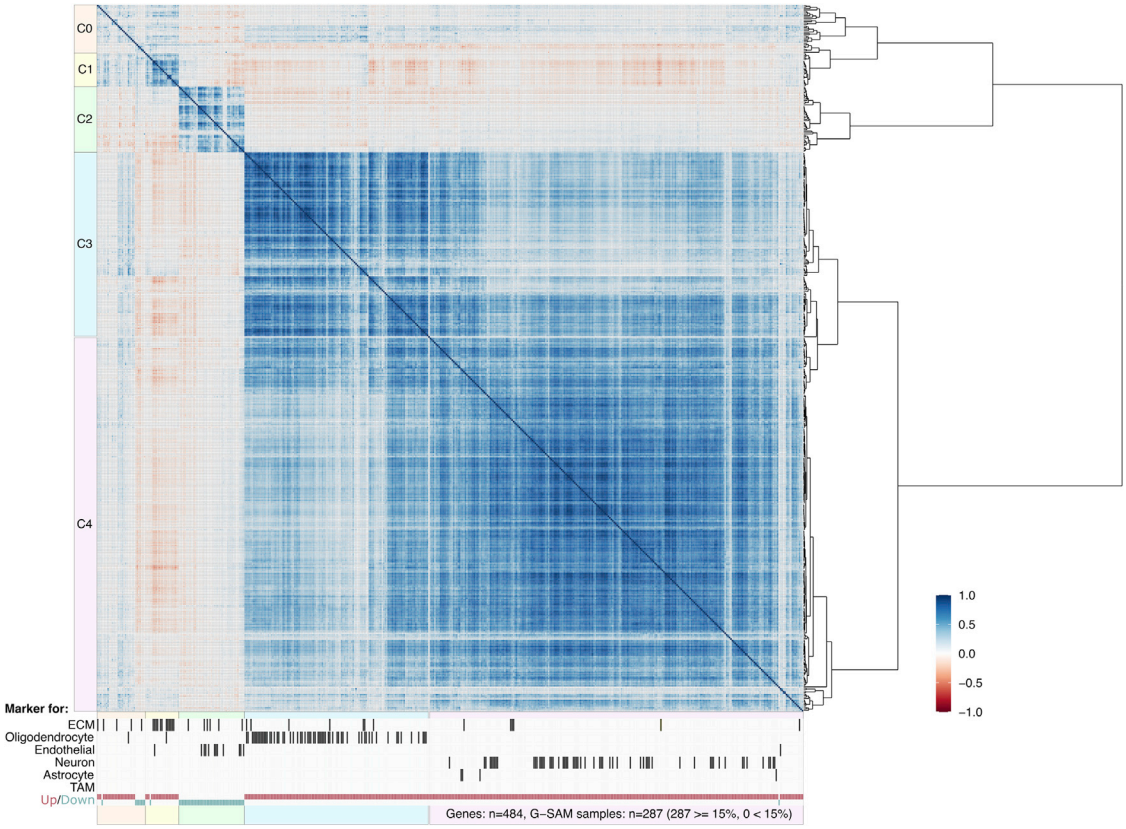


Figure 3. Recursive correlation plot of the 484 genes expressed differentially after correction for fraction of non-malignant cells

VST-transformed expression data from the G-SAM dataset were used. Clusters are represented by colors (red, C0; yellow, C1; green, C2; blue, C3; purple, C4). Markers are defined by McKenzie and from GO: 0031012 (ECM). See also Figures S3 and S4.

correlation analysis. Clustering RNA-seq data is sensitive to overall gene coverage and to outliers, such as hyper-amplified genes. Therefore, we used the correlation of the correlation values on the variance stabilizing transformation (VST) transformed expression values as distance metric; i.e., recursive correlation clustering and observed more natural clustering (STAR Methods). We subsequently performed our recursive correlation-based clustering technique on the DGE results (Figure 3), which divided the 484 genes into five correlated clusters (C0–C4) and one independent gene not assigned to any of these clusters (*SEPT12*). The largest cluster (C4) was enriched with neuron marker genes and the second largest cluster (C3) with oligodendrocyte marker genes. These two clusters were correlated and were both upregulated in recurrent resections. To assess whether upregulation of C4 and C3 originates from the heterogeneously tumor-intrinsic NPC- and OPC-like meta-features, we compared their expression with independent single-cell data, annotated using cell-type-specific marker genes.[17] For this we sequenced and analyzed glioblastoma snRNA-seq samples from the van Hijfte and peri-tumoral Bolleboom-Gao datasets (Figure S3). Virtually all C4 and C3 genes were uniquely expressed in neurons and oligodendrocytes respectively, and not (or to a much lower extent) in tumor cells. This was consistent in s[c/n]RNA-seq samples that contained neurons (n = 10) and/or oligodendrocytes (n = 24) in addition to tumor cells (Figure S3). Varn et al. described increased tumor-intrinsic neuronal signaling in stem-like glioblastoma cells.[14] Therefore, we compared their stem-like signature across C4 genes (Figure S4A). The stem-like signature associates more strongly with NPC-like genes than with C4, implying a tumor-intrinsic neuronal progression by a distinct gene set. We further examined the principal components of NPC-like and OPC-like tumor cell signatures in the G-SAM bulk data (Figure S4B) and found that the neuron marker-enriched C4 and NPC-like genes contributed differently to the principal components, and thus follow distinct expression profiles.

These results indicated that C4 does not represent the NPC-like gene sets or its meta-feature and, similarly (but less prominently), the C3 and OPC-like gene clusters contributed independently to their principal components.

To validate that the C4 signature containing a commonly used neuronal marker (NEUN/*RBFOX3*) is indeed derived from neurons, we performed IF stainings for NEUN, GFAP, and DAPI. This confirmed specificity for neurons, found in both tumor cell-rich and -poor regions (Figures 4A and 4B).

To quantify the finding that the increase in C4 genes is caused by an increase in neuronal cells at recurrence, we performed the multiplex IF staining (GFAP, NEUN, DAPI) on 15 matched primary-recurrent sample pairs and five normal brain controls. To do this in an unbiased manner, image scans of entire sections were separated into equally sized image tiles, processed independently and blinded with respect to sample type (primary/recurrent/control brain; FigureS 4C, S5A, and S5B). This analysis

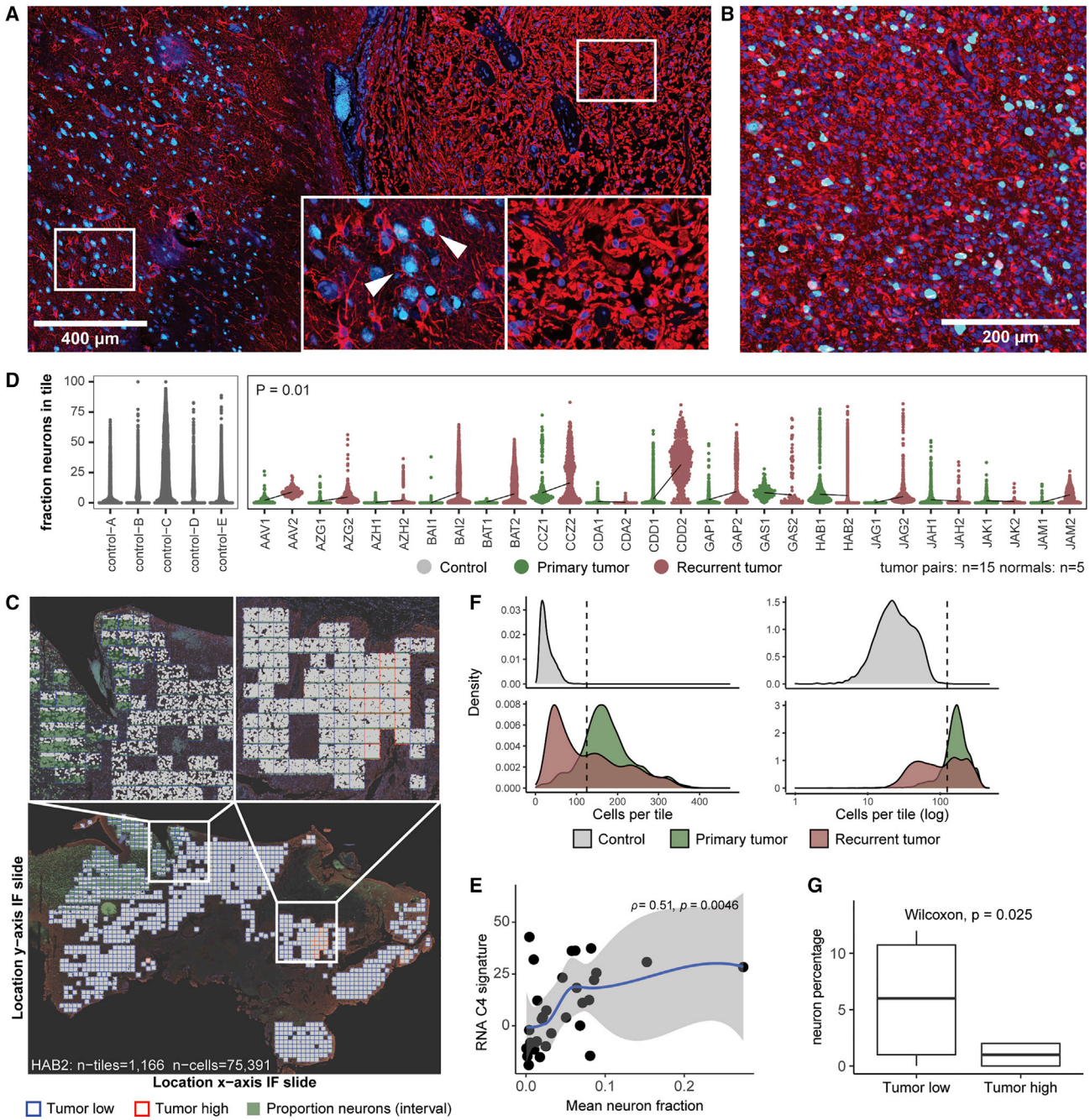


Figure 4. Quantification of whole-slide IF images confirming increased neuron fraction in recurrent resections

(A) IF staining of GFAP (red), NEUN (cyan), and DAPI (blue) for sample CDD2 showing the presence of neurons.

(B) IF staining of GFAP (red), NEUN (cyan), and DAPI (blue) for sample JAM2 showing the presence of neurons between tumor cells (i.e., high cell density and high GFAP staining).

(C) Example of an IF scan in the background overlayed with processed image tiles (sample HAB2). Curated tiles with their computationally defined cells (gray) and neuron fractions (green bars) are indicated in the foreground.

(D) Violin plot showing difference in neuron fraction per tile for paired resections. Significance was tested for neuron ratio per resection (Wilcoxon signed-rank test, paired).

(E) Relation between C1 signature score (RNA) and the IF per-sample neuron ratio (protein). Spearman's rho is indicated.

(F) Density of cell counts from tiles for primary, recurrent, and control samples. Left, raw counts; right, log-transformed.

(G) Difference of neuron ratio between tumor-low and tumor-high regions for primary and recurrent resections combined (Wilcoxon rank-sum test). See also Figure S5.

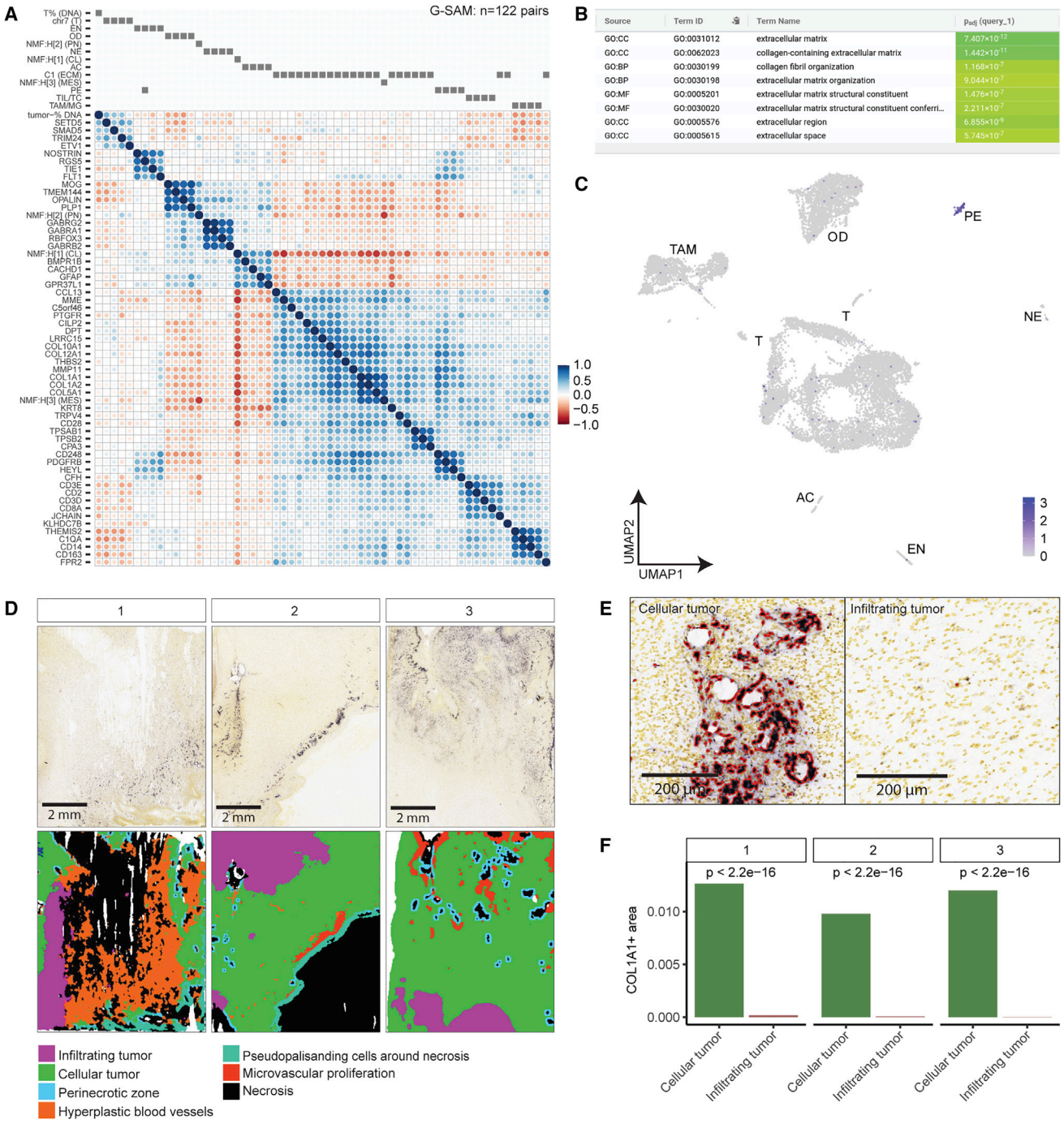


Figure 5. RNA expression of COL1A

(A) Recursive correlation clustering of per-patient log2FoldChange over time, in G-SAM bulk data. This includes the log2FoldChange of VST-transformed gene counts (C1 genes and cell-type marker genes), tumor purity, and the three NMF vectors representing the CL, MES, and PN subtypes. Correlated clusters are indicated in blue, anti-correlated clusters in red. Because log2FoldChanges are used as input, correlations represent correlated changes over time.

(B) Screenshot of g:Profiler output on the cluster C1 genes.

(C) Uniform Manifold Approximation and Projection (UMAP) of the annotated single-nucleus clusters (van Hijfte dataset, primary glioblastoma): tumor (T); oligodendrocyte (OD), pericytes (PE), endothelial cells (EN), TAM, astrocytes (AC), neurons (NE), T cells (TC), and undetermined (?). *COL1A2* is mainly expressed by a cluster of pericytes.

(legend continued on next page)

showed NEUN-positive cells were indeed present at a higher frequency in recurrent tumor samples (p = 0.01, paired Wilcoxon signed-rank test; Figure 4D). Of note, the paired tumor samples show a preserved neuron signal across RNA and protein quantifications (Spearman's ρ = 0.51, p = 0.0046; Figure 4E). Interestingly, a substantial difference in overall tile cell density between primary/recurrent samples was observed (Figures 4F, S5C, and S5D). Regardless of resection type, tumor-low regions showed a significantly higher neuron ratio than tumor-high regions (p = 0.025, Wilcoxon rank-sum test; Figure 4G). Once separated for tumor low and tumor high, the difference in neuron fractions between primary and recurrence was no longer significant, suggesting that a shift in tumor composition at recurrence coincides with the increase of neurons (Figure S5E).

The principal-component analysis, scRNA-seq, and IF indicate that the composition of recurrent resections more often contains higher proportions of neurons and oligodendrocytes. This assumption is strengthened by the observation that *RBFOX3* and the C4 signature are anti-correlated with tumor cell percentage (C4, $R = -0.17$; *RBFOX3*, $R = -0.15$).

Clusters C0 and C2

Cluster C2, which is downregulated at recurrence, contains endothelial cell markers,[17] implying a decrease in endothelial cells. Single-cell/nucleus data confirmed that C2 contains genes that are most strongly expressed by endothelial cells (Figure S6A). In part, the C2 genes were also expressed by pericytes, but this was also the case for the endothelial marker genes that were not present in C2.

Single-cell/nucleus data did not reveal an obvious association with a specific cell type for C0 (Figure S6B). C0 showed correlation with genes in C1, C3, and C4 but not C2. It is therefore likely that C0 genes are not restricted to a single cell type but may reflect a network of interacting genes across various cell types. Enrichment analysis also indicated no clear pathway association.

Pericyte-expressed ECM genes increase at tumor progression

In the G-SAM data, the C1 cluster negatively correlates with tumor cell percentage, astrocyte markers, and oligodendrocyte markers but correlates positively with features associated with low tumor cell percentage (MES signature, T cells, TAM; Figure 5A). Pathway enrichment analysis indicated that the C1 gene cluster is associated with ECM formation, largely due to the expression of genes such as *COL1A1*, *KRT8*, *COL1A2*, *MME*, and *MMP11* (Figure 5B). SnRNA-seq data showed that C1 genes were mainly expressed by a distinct cluster of cells (Figure 5C). These cells are positive for markers expressed by pericytes such as *PDGFRB* and *CD248* (Figure S6C) but negative for endothelial marker CD31/*PECAM1* (Figure S6D).

We evaluated public RNA-ISH data and found high *COL1A1* expression in cellular tumor regions that show microvascular proliferation, supporting the notion that pericytes express the C1 cluster genes (Figures 5D and 5E). Quantification of 3 RNA-ISH samples showed a consistent and significantly lower *COL1A1* expression in infiltrating tumor than in cellular tumor regions (p = 2.2e−16, chi-square test; Figure 5F).

To evaluate the C1 cluster on the protein level, we performed IF on G-SAM tumors, staining for COL1A1, CD31 (endothelial), PDGFRB (pericyte), GFAP (tumor), and DAPI (nuclei) (Figure 6A). Quantification of 1,490 image tiles from six tumors showed that COL1A1 around vessels was more abundant in tumor-high regions (Figures 6B–6D and S7A). Tumor 6 shows outliers that are high in COL1A1 but low in GFAP. Despite the low GFAP signal, these tiles have high vessel density not found in non-neoplastic brain and are surrounded by tumor dense regions (Figure S7B).

Taken together, gene expression analysis (two bulk datasets combined with single-cell/nucleus) and quantification of both public domain immunohistochemistry (IHC) data and in-house IF stainings indicate that the C1 gene set is specific for pericytes.

Additional IF stainings of 26 samples for PDGFRB, CD31, GFAP, and DAPI showed that CD31 and PDGFRB signal co-located near vessels (median Spearman correlation ρ = 0.66 between CD31 and PDGFRB signal intensities of n = 5,032 image tiles; Figure S7C), arguing against recruitment of isolated pericytes by the tumor.

Increased ECM signature associates with survival

To assess potential implications for each of the DGE clusters, we defined a per-sample signature for each cluster (STAR Methods). For each cluster, we split samples into signature high versus low based on either the median recurrent signature score value or, when ranked signature scores indicated presence of two distinct populations (i.e., a "dogleg" in the ranked scores), the split point between the two populations (Figure S7E). For the C1 cluster, we observed distinct populations (Figure 7A) and stratified patients according to this. Interestingly, while the C1 (ECM) signature did not associate with any of the investigated mutations, the hypermutation phenotype, or clinical parameters (Figure S7F), patients with a high ECM signature at recurrence showed a shorter survival from recurrence (adjusted p = 0.00097, log rank test; Figure 7B) and a shorter time to progression (adjusted p = 0.0037; Figure 7C). As a result, the overall survival was significantly shorter for patients with high C1 signatures at recurrence (adjusted p = 0.00097; Figure 7D). These associations remained present in multivariate analyses (survival from recurrence, hazard ratio [HR] = 2.08, 95% confidence interval [CI] [1.27–3.4], p = 0.004; Cox PH, time to progression, HR = 2.53, 95% CI [1.59–4.05], p = 0.0001; Figure 7E). In the 79 patients with *MGMT*-methylation status, the differences also remained significant in multivariate analysis (Figure S7G). A similar difference was observed in the GLASS dataset (adjusted p = 1.2e−7; Figure S7H). However, when stratifying the patients

(D) Glioblastoma RNA-ISH experiments taken from the Ivy GAP database. Panels on top represent spatial gene expression for *COL1A1* using ISH, panels on the bottom represent regional annotations. Annotation legends are in the panels below. *COL1A1* is expressed near blood vessels in regions annotated as microvascular proliferation. Image credit: Allen Institute for Brain Science.[36] Identifiers are provided in the key resources table.
(E) Zoom-in on cellular tumor and invasive tumor region in RNA-ISH *COL1A1* data. Red annotates *COL1A1* positive regions.
(F) Quantification of *COL1A1* in cellular and infiltrating tumor in the RNA-ISH data, showing a significantly higher fraction of *COL1A1* spots per pixel in cellular tumor (false discovery rate [FDR]-adjusted chi-square test p values are provided). See also Figure S6.

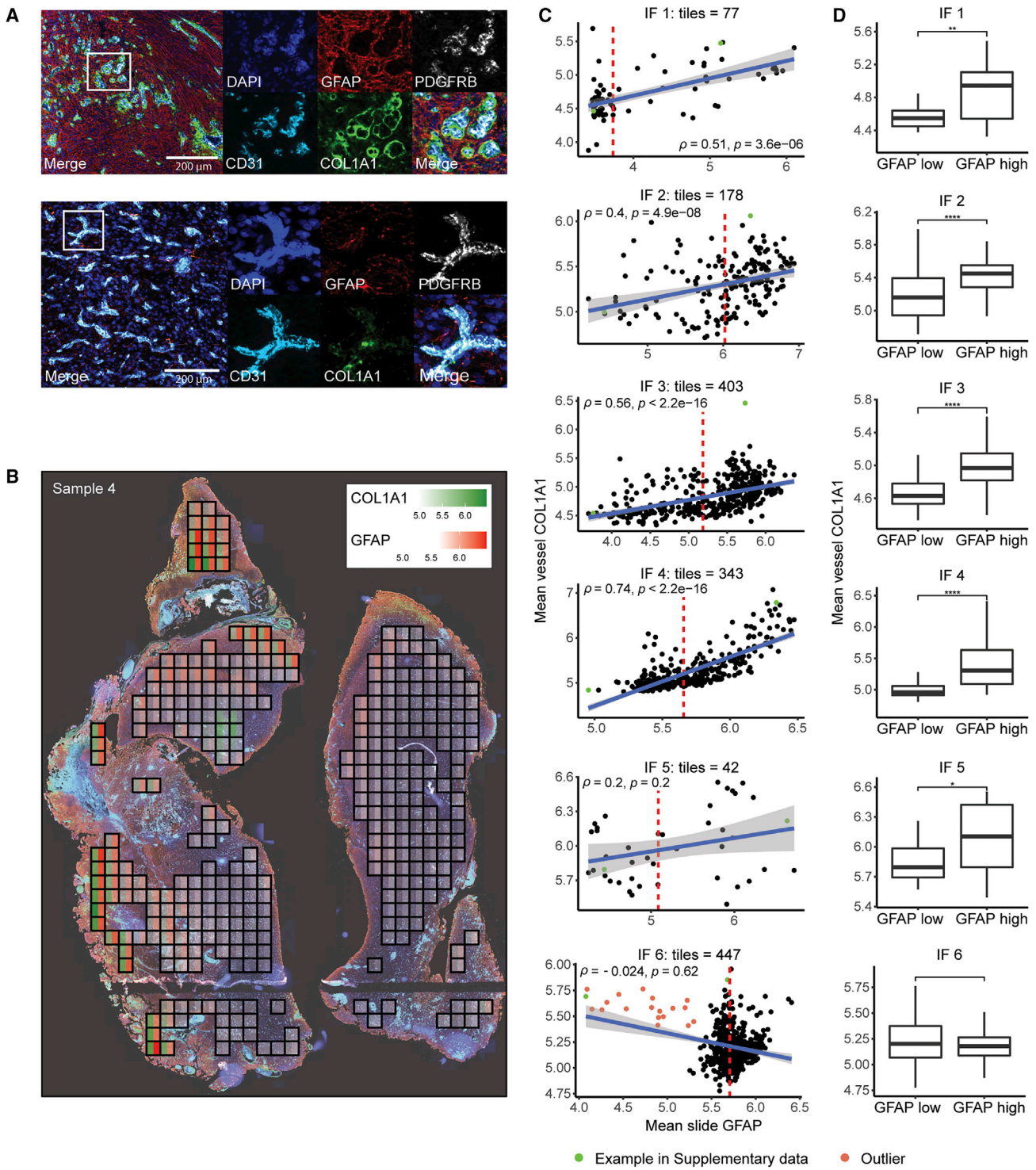
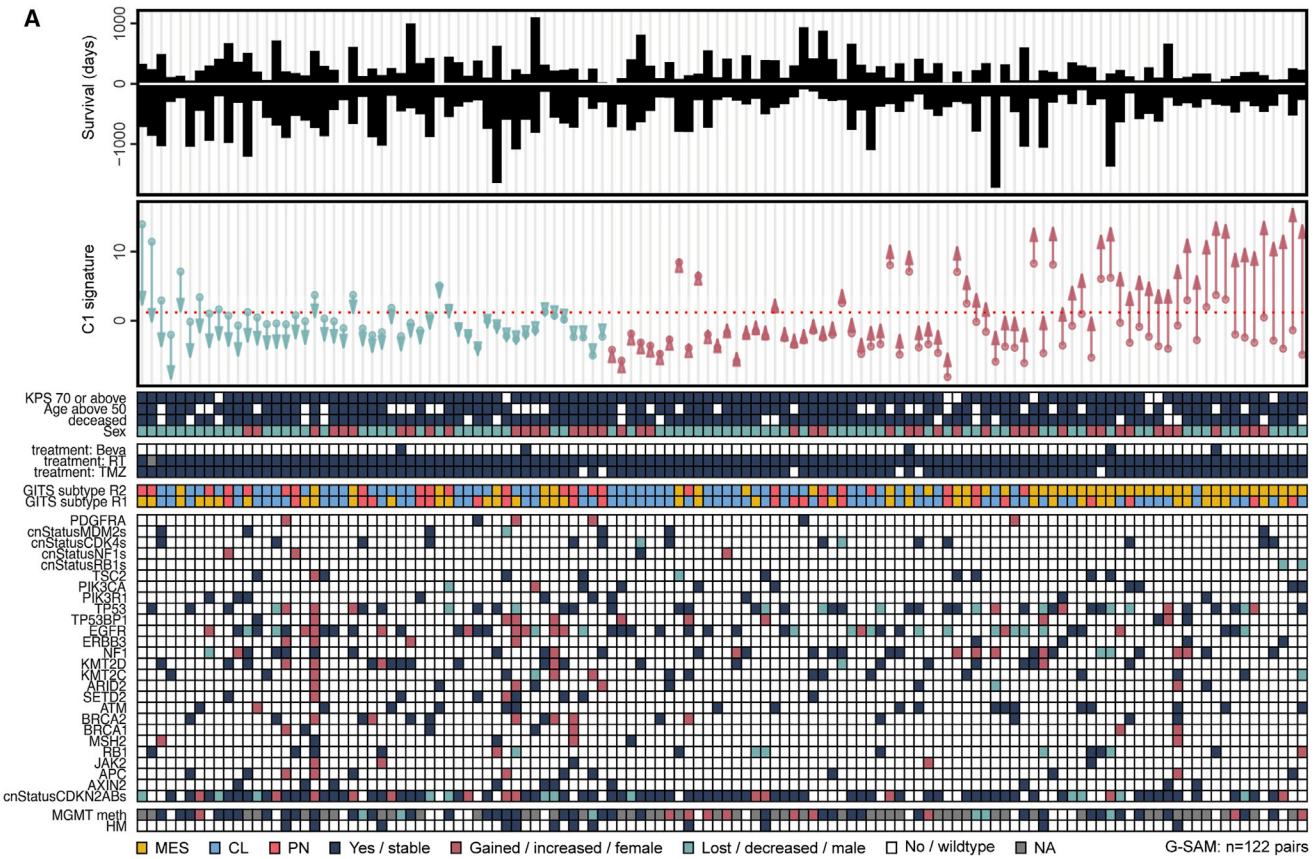


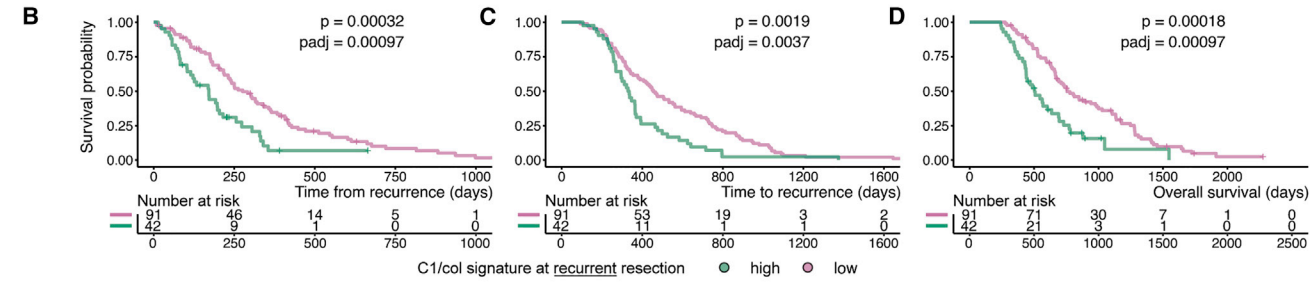
Figure 6. PDGFRB and COL1A stainings

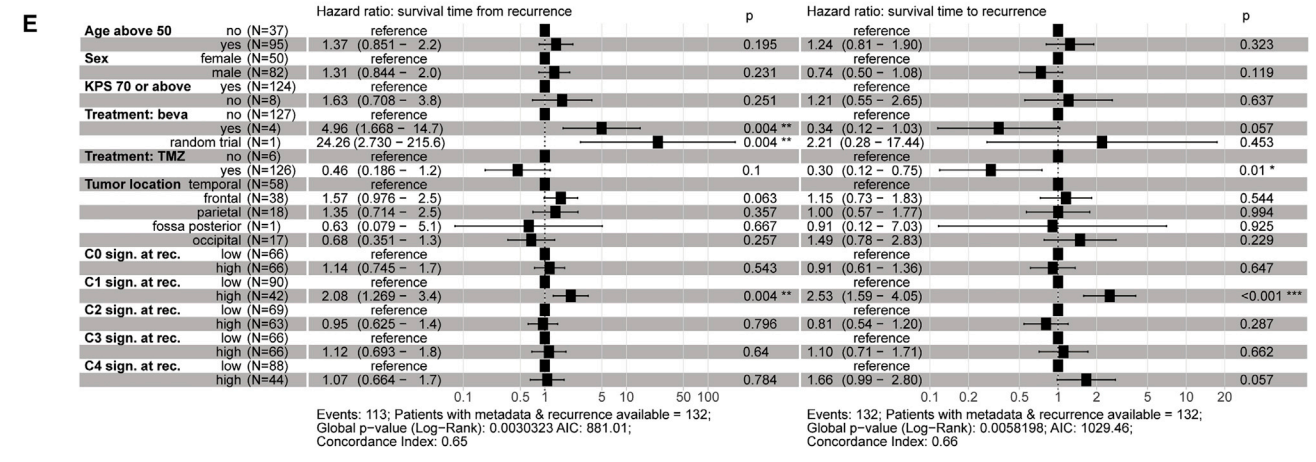
(A) DAPI and IF stainings for endothelial/CD31, COL1A, and tumor/GFAP of different regions within tumor sample GAH1. COL1A proteins show the strongest IF signal near intra-tumoral blood vessels.

(B) Quantification process of IF staining for COL1A1 and GFAP, showing high COL1A1 levels in GFAP-high tiles.

(C and D) Visualization and statistical interpretation indicating in five out of six samples a significant enrichment of COL1A1 in GFAP-high regions (FDR-adjusted Wilcoxon rank-sum test p values are provided). See also Figure S7 (examples: S7A, outliers: S7B).







(legend on next page)

on C1 status at primary disease, no significant differences in survival were found (Figures S7I–S7K). None of the other clusters showed an association with survival (Figure 7E). Longitudinal changes in ECM signatures were associated with MES subtype transitions ($p = 1.373e{-}08$; Wilcoxon test). Although the ECM and MES signature genes had only three collagen genes (*COL1A2*, *COL5A1*, *COL1A1*) in common, the correlation between the ECM signature and the MES-representing NMF factor was high ($R = 0.87$, Pearson correlation).

DISCUSSION

Here, we gathered a large multi-center dataset to study in-depth IDH-wt glioblastoma tumor evolution at the transcriptional level and combined this with publicly available data. We examined the relationship between transcriptional subtypes of primary-recurrent tumors by assigning a position in the two-dimensional GITS space. Contrary to earlier findings,[7,9] we observed CL to be most stable: 61% compared with 55% (MES) and 53% (PN). Importantly, although MES transitions can switch non-preferably to PN and CL, both PN and CL switch preferably to MES, indicating that MES may be considered a transcriptional evolutionary path of preference. Since PN to MES transitions can be diverted by innate immune cells,[22] this evolutionary path may be steered by the tumor's microenvironment.

In glioma, non-tumor cells as part of the microenvironment exert important roles associated with clinical outcome.[16] Prior to DGE analysis, we found a significant decrease in tumor purity over time, in both datasets. Such a decrease has been reported previously,[7] although other studies reported no significant difference.[16,23]

We considered different scenarios that may be responsible, including that tumors evolve to invade surrounding tissue, or generally become more invasive,[24] and that tumors (re-)organize their microenvironment by recruitment of non-malignant cells. Although low purity samples were excluded, it must be taken into consideration that some of the tumor recurrences may be pseudo-progressions that have been operated. The distinction between tumor and pseudo-progression cannot easily be made from an H&E-stained section but is made retrospectively by assessing whether contrast enhancement has decreased at follow-up.[25] Finally, differences in the type of surgery or treatment program may also account for differences in cellular composition. For instance, extent of resection in awake craniotomies and recurrent tumor surgery may be more extensive than expected. Unfortunately, the type of surgery and pseudo-progressions were not noted in our cohort. Since change in expression of TAM is not correlated with the neuron and oligodendrocyte clusters, a combination of factors is plausible.

Using comprehensive analysis techniques, we were able to work around the confounding effect of changes in tumor purity. Using the fraction of non-malignant cells as regression coefficient will correct the effect size of tumor-intrinsic signal specifically. However, non-malignant cell types are treated as one entity and as such does not identify cell-type composition changes. From this analysis followed that change in purity is largely due to a strong proportional increase of TAM combined with the fact that TAM comprise up to 30%–40% of the tumor bulk.[26,27] However, neuron and oligodendrocyte marker expression also increased significantly. This decrease in tumor purity appears to be specific to glioblastoma as no difference was found in IDH-mut astrocytoma (Vallentgoed et al., unpublished data).[14] In a study on the MES subtype specifically, a reduced tumor purity as a result of increased TAM infiltration was noted.[28] In glioblastoma, TAM are known to be actively recruited by glioblastoma stem cells, are associated with the MES subtype, and fulfill critical functions in glioma genesis and therapy resistance.[27,28] The increase in TAM at tumor progression therefore may be a relevant target for therapy.

Interestingly, our tumor-purity-corrected DGE analysis did not reveal hallmark glioblastoma genes being expressed significantly different over time. Earlier reports indicated that these driver genes are generally not further systematically altered on the DNA level at tumor recurrence. It was noted that, except for MMR and hypermutations,[9,29] the greatest evolutionary pressure is attributed to early-stage disease.[10] That the number of reoccurring genomic events at tumor recurrence is limited is intriguing, as most studies report an increased mutational burden, also in non-hypermutators.[10] This indicates progression but no common path for its evolution. Glioblastoma evolution has been described as highly idiosyncratic,[8] stochastic, and not coercing predictable paths,[10] and in some cases convergent,[8,9,12] which can be the underlying reason for our finding at the transcriptional level. Moreover, this may be an indication that primary tumor cells reach a balance between genotype (gene copies and regulatory states) and evolutionary pressure to divide even faster.[30]

Furthermore, our DGE analysis revealed two large gene clusters that are characterized by neuron and oligodendrocyte markers and increase at tumor progression. These results are in concordance with the earlier reported enriched gene sets (LEIN_NEURON_MARKERS and LEIN_OLIGODENDROCYTE_MARKERS).[12] We show that our clusters are independent of the tumor cell-intrinsic NPC-like and OPC-like meta-features, indicating an increase of neurons and oligodendrocytes at tumor recurrence.[13] This neuronal increase was validated at protein level by IF, which showed an increase of NEUN+ cells over time. This increase coincided with lower cell density, but this lower density was higher than observed in non-neoplastic brain. Given the large sample size of the datasets, the agreement with Kim et al., and the strength of the signal, this neuronal increase seems inherent to recurrent resections or disease progression.

Our results seem to contradict earlier conclusions in which stem-like tumor-intrinsic neural signaling increased over

Figure 7. ECM signatures in primary and recurrent glioblastoma (G-SAM dataset)

(A) Panel with overall survival per patient, split at time point of recurrent resection. Panel with changes in ECM signature over time, per patient, ordered on change over time. Patients with a steep increase in ECM signature are characterized by short survival after recurrence. Panel with matching mutation and clinical parameters per patient.

(B–D) Kaplan-Meier curves for survival after recurrence (B), time to recurrence (C), and overall survival (D) when stratifying patients on their ECM status at recurrence. Cutoff low/high is indicated with red dashed lines in (A) and Figure S7E.

(E) Multivariate Cox PH models on the time to and time post recurrence. See also Figure S7.

time.[14] In their manuscript, however, bulk deconvolution was trained without neurons. We show that this stem-like signature associates with the tumor-intrinsic NPC-like programs rather than the C4 genes, and therefore distinct mechanisms are reported. Further matching primary-recurrent s[c/n]RNA-seq data may provide deeper insights into the extent of tumor-intrinsic neural progression and subsequent tumor cell-states.

We found that cluster C1, enriched with ECM genes, had a significant negative prognostic value at recurrence but not when patients were stratified at primary disease, indicating this is an evolutionary path. However, patients included in this study were selected to have undergone a second operation at least 6 months after the primary surgery, and, therefore, our study is biased toward longer time to progression. The ECM gene set strongly overlaps with an earlier described gene set found in The Cancer Genome Atlas (TCGA) micro-array data[31] that also show a negative association with time to recurrence in glioblastoma.[32] ECM has previously been linked to therapy resistance and invasion in glioblastoma.[33,34] Although the ECM signature correlated strongly with the tumor-intrinsic MES subtype, combined bulk RNA, s[c/n]RNA, RNA-ISH, and IF analysis suggests that the signature is mainly expressed by pericytes. Furthermore, according to single-cell/nucleus data, the ECM signature does not appear to be expressed by TAMs while the two are correlated, implying these cell types accompany each other. It seems contradictory that cluster C2 (endothelial cell specific) expression decreases and C1 expression increases over time, while pericytes and endothelial cells are tightly associated in maintenance of the blood-brain barrier. However, as C1 appears specific to pericytes and more abundantly present alongside intra-tumoral vessels, this could point toward a form of vascular proliferation or remodeling[35] that is prognostic.

Our work indicates that the MES subtype, which may be linked to poor survival,[7,9] acts as preferred tumor progression path. Therefore, it is important to better understand the precise relationships between the MES subtype, the ECM signature and, importantly, the cell/tissue types expressing them. Ideally, an equivalent high-quality single-cell/nucleus study might provide deeper insights into these relationships and possibly causality. This could also shed new light on to what extent MES transitions are driven by tumor and by microenvironmental cells. However, in addition to the challenge of collecting sufficient high-quality fresh tissue, this will currently entail exorbitant costs.

In summary, we compared transcriptomes of primary with recurrent IDH-wt glioblastoma. Where the dataset was powerful enough to reveal a ~20% increase in TAM marker gene expression, it did not reveal differences in hallmark glioblastoma-associated genes. Despite transcriptional subtype shifts toward MES, expression differences of genes individually were not due to transcriptional changes of tumor cells. Instead, the globally observed differences point to changes in bulk composition, in which the proportion of tumor and endothelial cells decreased, accompanied by a strong increase in neurons and oligodendrocytes and, independently, TAM. Because the results indicate that temporal glioblastoma evolution involves (re)organization of the tumor microenvironment, including blood-brain barrier composition, a better understanding of the relationship between tumor cells, their environment, and the impact of chemotherapy and RT may be a crucial direction for glioblastoma therapy improvement.

Limitations of the study

Patients were almost exclusively treated with radiotherapy and chemotherapy, preventing determination of whether differences are treatment induced. The study was conducted retrospectively, so clinical implications should be considered with caution.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human samples
- METHOD DETAILS
  - G-SAM RNA-seq cohort
  - GLASS
  - Glioblastoma Intrinsic Transcriptional Subtype (GITS) space
  - Tumor cell percentage in G-SAM
  - Tumor cell percentage GLASS
  - Immune cell composition deconvolution
  - Differential expression analysis
  - Recursive correlation clustering
  - Per cluster (C0-C4) signatures
  - Marker gene list
  - Nuclei isolation and snRNA-seq
  - Single-cell and single nucleus RNA-seq
  - IVY GAP ISH data
  - IF stainings
  - IF image analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ccell.2023.02.019.

ACKNOWLEDGMENTS

This project has been funded by De Westlandse Ride and The Brain Tumour Charity grant number ET_2019/2_10470 and Télévie grant number 7.6514.17 from Belgium. N.A. is supported by the Ser Cymru II program, which is partly funded by Cardiff University and the European Regional Development Fund through the Welsh Government. The graphical abstract was made using BioRender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.H., K.D., S.A.G., L.v.H, P.A.R., and P.J.F.; methodology, Y.H., P.A.R., P.J.F., S.A.G., L.v.H, and N.B.; formal analysis and investigation, Y.H., I.d.H, T.P.P.v.d.B., L.v.H., and N.B.; writing – original draft, Y.H. and P.J.F.; writing – review & editing, all authors; resources, K.D., L.v.H., A.B., Z.G., A.V., W.C., L.K., M.D., M.J.B.T., M.K., A.W., M.S., A.H., S. Lu, G.L., S. Le, M.H., R.E.M.F., C.W., N.A., T.G., V.G., V.B., M.J.v.d.B., P.A.R., and P.J.F.; supervision and funding, P.A.R., P.J.F., and M.J.v.d.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 2, 2022
Revised: July 20, 2022
Accepted: February 14, 2023
Published: March 9, 2023

REFERENCES

1. Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J.B., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N. Engl. J. Med. *352*, 987–996. https://doi.org/10.1056/nejmoa043330.
2. Kraboth, Z., and Kalman, B. (2020). Longitudinal characteristics of glioblastoma in genome-wide studies. Pathol. Oncol. Res. *26*, 2035–2047. https://doi.org/10.1007/s12253-019-00705-1.
3. Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462–477. https://doi.org/10.1016/j.cell.2013.09.034.
4. Roy, A., Sakthikumar, S., Haseeb, L., Pettersson, M.E., Sundström, E., Marinescu, V.D., Lindblad-Toh, K., Forsberg-Nilsson, K., Lindblad-Toh, K., and Forsberg-Nilsson, K. (2020). Whole-genome sequencing of glioblastoma reveals enrichment of non-coding constraint mutations in known and novel genes. Genome Biol. *21*, 127. https://doi.org/10.1186/s13059-020-02035-x.
5. Weller, M., van den Bent, M., Preusser, M., Le Rhun, E., Tonn, J.C., Minniti, G., Bendszus, M., Balana, C., Chinot, O., Dirven, L., et al. (2021). EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. Nat. Rev. Clin. Oncol. *18*, 170–186. https://doi.org/10.1038/s41571-020-00447-z.
6. Draaisma, K., Chatzipli, A., Taphoorn, M., Kerkhof, M., Weyerbrock, A., Sanson, M., Hoeben, A., Lukacova, S., Lombardi, G., Leenstra, S., et al. (2020). Molecular evolution of IDH wild-type glioblastomas treated with standard of care affects survival and design of precision medicine trials: a report from the EORTC 1542 study. J. Clin. Oncol. *38*, 81–99. https://doi.org/10.1200/JCO.19.00367.
7. Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. Cancer Cell *32*, 42–56.e6. https://doi.org/10.1016/j.ccell.2017.06.003.
8. Kim, H., Zheng, S., Amini, S.S., Virk, S.M., Mikkelsen, T., Brat, D.J., Grimsby, J., Sougnez, C., Muller, F., Hu, J., et al. (2015). Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res. *25*, 316–327. https://doi.org/10.1101/gr.180612.114.
9. Wang, J., Cazzato, E., Ladewig, E., Frattini, V., Rosenbloom, D.I.S., Zairis, S., Abate, F., Liu, Z., Elliott, O., Shin, Y.J., et al. (2016). Clonal evolution of glioblastoma under therapy. Nat. Genet. *48*, 768–776. https://doi.org/10.1038/ng.3590.
10. Barthel, F.P., Johnson, K.C., Varn, F.S., Moskalik, A.D., Tanner, G., Kocakavuk, E., Anderson, K.J., Abiola, O., Aldape, K., Alfaro, K.D., et al. (2019). Longitudinal molecular trajectories of diffuse glioma in adults. Nature *576*, 112–120. https://doi.org/10.1038/s41586-019-1775-1.
11. Touat, M., Li, Y.Y., Boynton, A.N., Spurr, L.F., Iorgulescu, J.B., Bohrson, C.L., Cortes-Ciriano, I., Birzu, C., Geduldig, J.E., Pelton, K., et al. (2020). Mechanisms and therapeutic implications of hypermutation in gliomas. Nature *580*, 517–523. https://doi.org/10.1038/s41586-020-2209-9.
12. Kim, J., Lee, I.H., Cho, H.J., Park, C.K., Jung, Y.S., Kim, Y., Nam, S.H., Kim, B.S., Johnson, M.D., Kong, D.S., et al. (2015). Spatiotemporal evolution of the primary glioblastoma genome. Cancer Cell *28*, 318–328. https://doi.org/10.1016/j.ccell.2015.07.013.
13. 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.e21. https://doi.org/10.1016/j.cell.2019.06.024.
14. 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.e16. https://doi.org/10.1016/j.cell.2022.04.038.
15. GLASS Consortium, Amin, S.B., Ashley, D.M., Barnholtz-Sloan, J.S., Bates, A.J., Beroukhim, R., Bock, C., Brat, D.J., Claus, E.B., Costello, J.F., et al. (2018). Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the glioma longitudinal analysis consortium. Neuro Oncol. *20*, 873–884. https://doi.org/10.1093/neuonc/noy020.
16. Zhang, C., Cheng, W., Ren, X., Wang, Z., Liu, X., Li, G., Han, S., Jiang, T., and Wu, A. (2017). Tumor purity as an underlying key factor in glioma. Clin. Cancer Res. *23*, 6279–6291. https://doi.org/10.1158/1078-0432.CCR-16-2598.
17. McKenzie, A.T., Wang, M., Hauberg, M.E., Fullard, J.F., Kozlenkov, A., Keenan, A., Hurd, Y.L., Dracheva, S., Casaccia, P., Roussos, P., and Zhang, B. (2018). Brain cell type specific gene expression and Co-expression network architectures. Sci. Rep. *8*, 8868. https://doi.org/10.1038/s41598-018-27293-5.
18. Racle, J., and Gfeller, D. (2020). EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. In Methods in Molecular Biology, pp. 233–248. https://doi.org/10.1007/978-1-0716-0327-7_17.
19. Alnahhas, I., Alsawas, M., Rayi, A., Palmer, J.D., Raval, R., Ong, S., Giglio, P., Murad, M.H., and Puduvalli, V. (2020). Characterizing benefit from temozolomide in MGMT promoter unmethylated and methylated glioblastoma: a systematic review and meta-analysis. Neurooncol. Adv. *2*, vdaa082. https://doi.org/10.1093/noajnl/vdaa082.
20. Munoz, J.L., Rodriguez-Cruz, V., Ramkissoon, S.H., Ligon, K.L., Greco, S.J., and Rameshwar, P. (2015). Temozolomide resistance in glioblastoma occurs by miRNA-9-targeted PTCH1, independent of sonic hedgehog level. Oncotarget *6*, 1190–1201. https://doi.org/10.18632/oncotarget.2778.
21. De Silva, T., Ye, G., Liang, Y.Y., Fu, G., Xu, G., and Peng, C. (2012). Nodal promotes glioblastoma cell growth. Front. Endocrinol. *3*, 59. https://doi.org/10.3389/fendo.2012.00059.
22. Schmitt, M.J., Company, C., Dramaretska, Y., Barozzi, I., Göhrig, A., Kertalli, S., Großmann, M., Naumann, H., Sanchez-Bailon, M.P., Hulsman, D., et al. (2021). Phenotypic mapping of pathologic cross-talk between glioblastoma and innate immune cells by synthetic genetic tracing. Cancer Discov. *11*, 754–777. https://doi.org/10.1158/2159-8290.CD-20-0219.
23. González-Tablas Pimenta, M., Otero, Á., Arandia Guzman, D.A., Pascual-Argente, D., Ruíz Martín, L., Sousa-Casasnovas, P., García-Martin, A., Roa Montes de Oca, J.C., Villaseñor-Ledezma, J., Torres Carretero, L., et al. (2021). Tumor cell and immune cell profiles in primary human glioblastoma: impact on patient outcome. Brain Pathol. *31*, 365–380. https://doi.org/10.1111/bpa.12927.
24. Drumm, M.R., Dixit, K.S., Grimm, S., Kumthekar, P., Lukas, R.V., Raizer, J.J., Stupp, R., Chheda, M.G., Kam, K.L., McCord, M., et al. (2020). Extensive brainstem infiltration, not mass effect, is a common feature of end-stage cerebral glioblastomas. Neuro Oncol. *22*, 470–479. https://doi.org/10.1093/neuonc/noz216.
25. Leao, D.J., Craig, P.G., Godoy, L.F., Leite, C.C., and Policeni, B. (2020). Response assessment in neuro-oncology criteria for gliomas: practical approach using conventional and advanced techniques. AJNR. Am. J. Neuroradiol. *41*, 10–20. https://doi.org/10.3174/ajnr.A6358.
26. Geribaldi-Doldán, N., Fernández-Ponce, C., Quiroz, R.N., Sánchez-Gomar, I., Escorcia, L.G., Velásquez, E.P., and Quiroz, E.N. (2020). The role of microglia in glioblastoma. Front. Oncol. *10*, 603495. https://doi.org/10.3389/fonc.2020.603495.
27. Buonfiglioli, A., and Hambardzumyan, D. (2021). Macrophages and microglia: the cerberus of glioblastoma. Acta Neuropathol. Commun. *9*, 54. https://doi.org/10.1186/s40478-021-01156-z.

28. Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell *39*, 779–792.e11. https://doi.org/10.1016/j.ccell.2021.05.002.

29. Muscat, A.M., Wong, N.C., Drummond, K.J., Algar, E.M., Khasraw, M., Verhaak, R., Field, K., Rosenthal, M.A., and Ashley, D.M. (2018). The evolutionary pattern of mutations in glioblastoma reveals therapy-mediated selection. Oncotarget *9*, 7844–7858. https://doi.org/10.18632/oncotarget.23541.

30. Hoogstrate, Y., Ghisai, S.A., De Wit, M., De Heer, I., Draaisma, K., Van Riet, J., Van De Werken, H.J.G., Bours, V., Buter, J., Vanden Bempt, I., et al. (2022). The EGFRvIII transcriptome in glioblastoma: a meta-omics analysis. Neuro Oncol. *24*, 429–441. https://doi.org/10.1093/neuonc/noab231.

31. Kim, H., Watkinson, J., Varadan, V., and Anastassiou, D. (2010). Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. BMC Med. Genom. *3*, 51. https://doi.org/10.1186/1755-8794-3-51.

32. Cheng, W.Y., Kandel, J.J., Yamashiro, D.J., Canoll, P., and Anastassiou, D. (2012). A multi-cancer mesenchymal transition gene expression signature is associated with prolonged time to recurrence in glioblastoma. PLoS One *7*, e34705. https://doi.org/10.1371/journal.pone.0034705.

33. Mohiuddin, E., and Wakimoto, H. (2021). Extracellular matrix in glioblastoma: opportunities for emerging therapeutic approaches. Am. J. Cancer Res. *11*, 3742–3754.

34. Virga, J., Szivos, L., Hortobágyi, T., Chalsaraei, M.K., Zahuczky, G., Steiner, L., Tóth, J., Reményi-Puskár, J., Bognár, L., and Klekner, A. (2019). Extracellular matrix differences in glioblastoma patients with different prognoses. Oncol. Lett. *17*, 797–806. https://doi.org/10.3892/ol.2018.9649.

35. Cheng, L., Huang, Z., Zhou, W., Wu, Q., Donnola, S., Liu, J.K., Fang, X., Sloan, A.E., Mao, Y., Lathia, J.D., et al. (2013). Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. Cell *153*, 139–152. https://doi.org/10.1016/j.cell.2013.02.021.

36. Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., Yoon, J.G., Smith, K.A., Lankerovich, M., Bertagnolli, D., Bickley, K., et al. (2018). An anatomic transcriptional atlas of human glioblastoma. Science *360*, 660–663. https://doi.org/10.1126/science.aaf2666.

37. Wang, L., Babikir, H., Müller, S., Yagnik, G., Shamardani, K., Catalan, F., Kohanbash, G., Alvarado, B., Di Lullo, E., Kriegstein, A., et al. (2019). The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. Cancer Discov. *9*, 1708–1719. https://doi.org/10.1158/2159-8290.CD-19-0329.

38. Yuan, J., Levitin, H.M., Frattini, V., Bush, E.C., Boyett, D.M., Samanamud, J., Ceccarelli, M., Dovas, A., Zanazzi, G., Canoll, P., et al. (2018). Single-cell transcriptome analysis of lineage diversity in high-grade glioma. Genome Med. *10*, 57. https://doi.org/10.1186/s13073-018-0567-9.

39. Couturier, C.P., Ayyadhury, S., Le, P.U., Nadaf, J., Monlong, J., Riva, G., Allache, R., Baig, S., Yan, X., Bourgey, M., et al. (2020). Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. Nat. Commun. *11*, 3406. https://doi.org/10.1038/s41467-020-17186-5.

40. Wang, L.B., Karpova, A., Gritsenko, M.A., Kyle, J.E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J.H., Hong, R., et al. (2021). Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell *39*, 509–528.e20. https://doi.org/10.1016/j.ccell.2021.01.006.

41. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

42. Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). Gencode 2021. Nucleic Acids Res. *49*, D916–D923. https://doi.org/10.1093/nar/gkaa1087.

43. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

44. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

45. Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

46. 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. https://doi.org/10.1186/s13059-014-0550-8.

47. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

48. Kursa, M.B., and Rudnicki, W.R. (2010). Feature selection with the boruta package. J. Stat. Software *36*, 1–13. https://doi.org/10.18637/jss.v036.i11.

49. Wei, B., Liu, Z., Fan, Y., Wang, S., Dong, C., Rao, W., Yang, F., Cheng, G., and Zhang, J. (2021). Analysis of cellular heterogeneity in immune microenvironment of primary central nervous system lymphoma by single-cell sequencing. Front. Oncol. *11*, 683007. https://doi.org/10.3389/fonc.2021.683007.

50. Jensen, M.A., Ferretti, V., Grossman, R.L., and Staudt, L.M. (2017). The NCI Genomic Data Commons as an engine for precision medicine. Blood *130*, 453–459. https://doi.org/10.1182/blood-2017-03-735654.

51. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

52. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

53. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: open source software for digital pathology image analysis. Sci. Rep. *7*, 16878. https://doi.org/10.1038/s41598-017-17204-5.

54. Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. *44*, W83–W89. https://doi.org/10.1093/nar/gkw199.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
NEUN, polyclonal	Millipore	#ABN78
MOG, polyclonal	Sigma Aldrich	#HPA021873
GFAP, Clone SP78	Ventana	#760-4345
SOX2, Clone SP76	Cell Marque	#371R-18
CD31, Clone JC70	Cell Marque	#131M-98
PDGFRB, Clone 28E1	Cell Signaling	#3169; RRID: AB_2162497
COL1A1, polyclonal	Novusbio	#NB600-408
Biological samples		
Primary and recurrent tumor samples	EORTC	EORTC Study 1542, Glioblastoma, Stability of Actionable Mutations (G-SAM)
Normal brain control (IF)	ErasmusMC	N/A
Chemicals, peptides, and recombinant proteins		
omnimap anti-mouse HRP	Ventana	#760-4310; RRID: AB_2885182
omnimap anti-rabbit HRP	Ventana	#760-4311; RRID: AB_2811043
CC1 buffer	Ventana	#950-500; RRID: AB_950500
CC2 buffer	Ventana	#950-123; RRID: AB_950123
R6G	Ventana	#760-244; RRID: AB_760244
DCC	Ventana	#760-240
Cy5	Ventana	#760-238
Red610	Ventana	#760-235
FAM	Ventana	#760-243
anti-fading medium	DAKO	#S3023
Nuclei EZ Lysis Buffer	Sigma	#nuc101-1kt
RNaseOUT™ Recombinant Ribonuclease Inhibitor	Invitrogen	#10777019
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Roche	#11873580001
Hoechst	Sigma	#H3570
Critical commercial assays		
AllPrep DNA/RNA FFPE kit	Qiagen	#80234
RNeasy FFPE kit	Qiagen	#73504
Ventana Benchmark ULTRA	Ventana	#05987750001
Optiview universal DAB detection Kit	Roche	#760700
40 μm cell strainer	Corning	#431752
Deposited data		
CNV/tumor purity profiles	This paper	https://doi.org/10.5281/zenodo.7147027
Pipeline and analysis code	This paper	https://github.com/yhoogstrate/gsam
484 DGE genes and clusters table	This paper	https://doi.org/10.5281/zenodo.7555173
G-SAM RNA-seq raw sequencing data	European Genome Phenome Archive	EGA: EGAD00001007860
G-SAM Targeted exome-seq raw sequencing data	European Genome Phenome Archive	EGA: EGAD00001004593
van Hijfte GBM dataset 2022/A (single-nucleus RNA-seq: count data)	This paper	https://doi.org/10.5281/zenodo.6546712

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
van Hijfte GBM dataset 2022/A (single-nucleus RNA-seq: raw reads)	This paper	EGA: EGAD00001009871
Bolleboom-Gao peri-tumoral snRNA-seq GBM dataset (raw reads)	This paper	EGA: EGAD00001009964
Diaz single-cell dataset	Wang et al., 2019[1]	GEO: GSE138794
Couturier single-cell dataset	Couturier et al.,[2] Wei et al.[3]	EGA: EGAS00001004422
Yuan single-cell dataset	Yuan et al.[4]	GEO: GSE103224
CPTAC-3 single-cell dataset	Jensen et al.,[5] Wang et al., 2021[6]	https://portal.gdc.cancer.gov/projects/CPTAC-3
GLASS RNA count data and metadata	Varn et al.[7]	https://www.synapse.org/glass
Stem_like scRNA deconvolution matrix	Varn et al.[7]	https://doi.org/10.1016/j.cell.2022.04.038
COL1A RNA-ISH data	Puchalski et al.[8]	IVY GAP: W8-1-1-C.1.03, W1-1-2-X.1.04 & W11-1-1-G.1.04
Software and algorithms		
fastp v0.21.0	Chen et al.[9]	https://github.com/OpenGene/fastp
STAR 2.7.2a	Dobin et al.[10]	https://github.com/alexdobin/STAR
sambamba 0.7.0	Tarasov et al.[11]	https://lomereiter.github.io/sambamba/
EPIC 1.1	Racle and Gfeller[12]	https://github.com/GfellerLab/EPIC
R 4.1.2	The R Project for Statistical Computing	https://www.r-project.org/
DESeq2	Love et al.[13]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Seurat v4	Hao et al.[14]	https://github.com/satijalab/seurat
msig.library.12.R	Wang et al., 2017[15]	https://doi.org/10.1016/j.ccell.2017.06.003
featureCounts/Subread v2.0.0	Liao et al.[16]	https://subread.sourceforge.net/
recursiveCorPlot v0.1.12	This paper	https://github.com/yhoogstrate/recursiveCorPlot
g:Profiler web portal	Reimand et al.[17]	https://biit.cs.ut.ee/gprofiler/gost
Cell Ranger v3.1.0 pipeline	10x genomics	https://www.10xgenomics.com/
ImageJ	Schneider et al.[18]	https://imagej.nih.gov/ij/index.html
QuPath	Bankhead et al.[19]	https://qupath.github.io/
CNVKit	Talevich et al.[20]	https://github.com/etal/cnvkit/
egfr-v3-determiner v0.7.4	Hoogstrate et al., 2022[21]	https://github.com/yhoogstrate/egfr-v3-determiner
Other		
hg19 + gencode.v31lift37.annotation.gtf	Frankish et al.[22]	https://www.gencodegenes.org/
Single cell/nucleus glioma & TME marker genes: Supplementary File 1 attached to the McKenzie manuscript, sheet: top_human_specificity	McKenzie et al.[23]	https://doi.org/10.1038/s41598-018-27293-5

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Youri Hoogstrate (y.hoogstrate@erasmusmc.nl).

Materials availability

The study did not generate new unique reagents.

Data and code availability

G-SAM bulk RNA-seq data have been deposited in EGA and are publicly available. The accession number is listed in the key resources table. G-SAM targeted exome-seq data are available in EGA and the accession number is listed in the key resources table. Single-nucleus RNA-seq (*van Hijfte* glioblastoma and *Bolleboom-Gao* peri-tumoral glioblastoma) datasets have been deposited at

Zenodo. Accession numbers are listed in the key resources table. Additional s[c/n]RNA-seq data of the Diaz/Wang,[37] Yuan,[38] Couturier,[39] CPTAC-3 dataset[40] were taken from public repositories. Accession numbers are listed in the key resources table. All original code projects (G-SAM and recursiveCorPlot) have been deposited at GitHub and are listed in the key resources table. Code for NMF and single sample Gene Set Enrichment Analysis (ssGSEA) was obtained from the supplementary materials of another study.[7] Fits of copy number variation (CNV) profiles from the tumor purity estimation have been deposited at Zenodo. The accession number is listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human samples

Tissue and metadata from the G-SAM study were accrued through the pan-European Organisation for Research and Treatment of Cancer (EORTC) network. Informed written consent was obtained from all patients. The study design was approved by the institutional review board of Erasmus MC (Rotterdam, The Netherlands), and conducted according to institutional and national regulations. Regarding the normal brain tissue used for IF, this study follows the guidelines of the Dutch Federation of Biomedical Scientific Societies (www.federa.org/codes-conduct, last accessed 02-11-2022), which state that no separate ethical approval is required for the use of anonymized residual tissue procured during regular treatment. Clinical data is provided in Table S1.

SnRNA-seq sample from the van Hijfte dataset was taken from a fresh IDH-wt glioblastoma (male, chr7 gain, chr10 loss, sub-clonal chr3.q gain, sub-clonal chr4 hyperamplification near centromere). SnRNA-seq sample H243-GBM from the Bolleboom-Gao dataset (chr7 gain, chr10 loss, chr9.q loss, chr13.p loss) was taken from the infiltrated peri-tumoral neocortex of a 67-year-old male diagnosed with IDH-wt, grade IV, MGMT unmethylated glioblastoma. Infiltrated peri-tumoral cortex was removed as part of regular tumor resection. This sample is unique because of its peri-tumoral origin and subsequent high neuron fraction.

METHOD DETAILS

G-SAM RNA-seq cohort

Glioblastoma patient samples from primary–recurrent tumor pairs were obtained from adult patients through the EORTC,[6] forming the primary dataset referred to as G-SAM. The minimum interval between resections was six months. FFPE material was collected and RNA was extracted from FFPE using the AllPrep DNA/RNA FFPE kit or the RNeasy FFPE kit (Qiagen, Venlo, The Netherlands) and sent out for sequencing on an Illumina NovaSeq at the GIGA-Genomics Core Facility of the University of Liège. The 2 × 151bp paired-end FASTQ files were obtained from the manufacturer. Reads were cleaned using fastp v0.21.0, using additional arguments: `-y -x -i -l 35`,[41] and aligned to hg19 and gencode.v31lift37.annotation.gtf[42] using STAR 2.7.2a.[43] Duplicate reads were removed using *sambamba markdup* (`-t 3 –overflow-list-size 1000000 –hash-table-size 1000000`).[44] Reads were counted with featureCounts (subread v2.0.0) using the following additional arguments: `-p -s 2 -T 64 –primary -C –ignoreDup`.[45] *EGFRvIII* spliced-in fractions were determined using *egfr-v3-determiner* v0.7.4.[30] The G-SAM cohort consists of 322 RNA-seq samples from 172 individuals (Table S1). This initial full size dataset is used only for statistics on purity and which cell types increase as opposed to tumor cells (Figures 2A, 2F, and 2G). Of these, 35 samples were excluded from further analysis because of low purity as determined from matching targeted exome-seq data (>15%, threshold based on visual inspection of CNV profiles). In total, the dataset used for remaining analysis consisted of 287 RNA-sequenced IDH-wt tumor samples from 165 patients with 122 complete primary–secondary resection pairs.

GLASS

Data from the international *GLASS* consortium[10,14,15] was obtained from the Synapse portal (https://www.synapse.org/glass, most recent data sweep 2022-05-31). Read counts were obtained from the file ‘‘transcript_count_matrix_all_samples.tsv’’. Splice isoform specific read counts were aggregated into per-gene read counts by taking the per-‘gene_id’ sum and rounded to discrete numbers. Survival data was taken from table syn31121181. Clinical information was obtained from table syn31121219. Batches were obtained from table syn31121185. Hypermutation status was obtained from table syn32908224 and HM status was defined as `HM = coverage_adj_mut_freq2 > 10`.[10] Processed copy-number profiles were obtained from table syn31121137. Replicate samples were excluded. Only patients that were consistently classified as grade IV, non-codel, IDH-wt glioblastoma were included. In case a patient had multiple recurrent resections, only the last resection was included. Samples labeled as TP were considered primary, samples labeled as R[1–4] were considered recurrent. Three samples were outliers compared to their batch (PCA analysis) and were excluded: GLSS-CU-R004-TP-01R-RNA-U0DEP1, GLSS-SM-R099-R1-01R-RNA-MNTPMI, GLSS-SM-R111-R1-01R-RNA-WM5ESA.

Glioblastoma Intrinsic Transcriptional Subtype (GITS) space

The intrinsic subtypes of glioblastoma bulk RNA-seq samples can be classified using ssGSEA on a 3 × 50 gene subset of the 7,425 genes originally used for defining the subtypes with non-negative matrix factorization (NMF).[7] We used the same techniques as foundation for defining our GITS space. First, we loaded the raw read-counts of both datasets and took the 150 subtype genes as subset. Gene *DACH1* was represented by different ENSEMBL-IDs in the datasets (G-SAM: `ENSG00000276644`, GLASS: `ENSG00000165659`). We VST transformed the 150 gene raw count matrix (datasets combined) using the DESeq2 package.[46] We applied a batch effect correction with

`limma::removeBatchEffects` for the conditions GSAM and those provided for GLASS (syn31121185). This matrix was the input for ssGSEA and for NMF, set to define 3 meta-features. NMF factorizes the expression matrix in two vectors, W and H, of which W are the per-transcript weights that multiplied with the per-patient H-matrix result in a matrix decompressed by factorization. The H-matrix contains features per patient, most important to refactor the data. In the sense that it reduces dimensions over transcripts, the H-matrix has similarities with principal component analysis, but without the guarantee that the vectors are independently co-variated. We used publicly available NMF code (`msig.library.12.R`),[7] with the same random seed. Principal component analysis showed that the 3 factors of the H-matrix were made up of 2 independent components, which formed the axes of our GITS space. The space was divided into three regions by a linear Support Vector Machine (SVM) classifier using the ssGSEA subtypes as input. For this, the function `e1071::svm` with the following arguments: `scale = F`, `kernel = 'linear'`, `tolerance = 0.0001`, `cost = 3`. Each sample was then re-classified by using the coordinates within the GITS space. For each matching pair the Euclidean distance within the GITS space was calculated after both PC1 and PC2 were scaled to an identical standard deviation.

Tumor cell percentage in G-SAM

Tumor cell percentages were calculated based on targeted exome-seq data. Because we isolated RNA and DNA from combined isolations, the input of both assays comprises the same tumor purity. CNV segments were determined using CNVKit[47] before.[6]

First, we calculated for all tumor cell percentages $\{1\%, \ldots, 100\%\}$ the expected log2FoldChange CNVKit is assumed to provide for a single deletion, a gain and a double gain:

$$f_{p,c} = \log_2\left(\frac{2(1-p)+(p*c)}{2}\right)$$

where p is the theoretical purity, c is the copy number (1, 3 and 4) and $f_{p,c}$ the expected log2FoldChange for the given purity and copy-number for a ploidy of 2. Per targeted exome-seq sample (segment files: *.cnv) we calculated per theoretical purity, the weighted minimal squared error (over c) of the observed log2FoldChange of all segments to the expected log2FoldChange for a single deletion ($f_{c=1}$), no change (0), a single gain or ($f_{c=3}$) a double gain ($f_{c=4}$), weighted to the length of the segment:

$$\varepsilon_p = \sum_s min\begin{cases} ((l_s - f_{p,1}) * w_s)^2 \\ ((l_s - 0) * w_s)^2 \\ ((l_s - f_{p,3}) * w_s)^2 \\ ((l_s - f_{p,4}) * w_s)^2 * \theta \end{cases}$$

where l_s is the observed log2FoldChange for segment s and w_s the CNVKit weight for segment s and ε_p is the over-all-segments squared error towards the expected log2FoldChange for purity p. Here θ is a penalty for double gains, set to 1.1. We then choose the purity that provided the lowest squared error as being the samples' purity. In some samples we noticed the observed log2FoldChange of chr10 (lost) and chr7 (gained) segments within the same tumor represented different purities. As these are indications of sub-clonal differences, for certain samples a manual filtered selection of chromosomes was used (code GitHub). Chromosome X and Y were excluded. Segments corresponding to hyperamplifications (log2 >= 1.1) were excluded. Small segments (<5MB) were excluded.

Tumor cell percentage GLASS

Tumor cell percentages in GLASS were estimated in a similar procedure using CNV segments from the Synapse portal (variants_gatk_seg; syn31121137). For this dataset, log2_copy_ratio's were first weighted median centered. Small segments (num_points <= 60) and segments corresponding to hyperamplifications or homozygous losses (<= –2.0 or >= 1.1) were excluded. Chromosome X and Y were excluded.

For the 35/245 samples for which no CNV segment data was available, purities were taken from imputation using the expression data. For imputation, 10-fold cross-validation was performed. During each iteration, 90% of the data was used for training and 10% was used to predict purity. In each iteration, the top 1,500 genes with the strongest positive and top 1,500 with the negative correlation to CNV purity in the training data were selected as candidate features. With these, the `Boruta::Boruta` function of the Boruta algorithm[48] was used to find features most informative for prediction. This resulted in approximately ~100 genes per cross-validation iteration, marked as either 'Confirmed' or 'Tentative'. The `randomforest::randomForest` function was used with `ntree` set to 5,000, to predict purity on the 10% test. For those that had no CNV purities available, the RF purities were used instead.

Immune cell composition deconvolution

EPIC 1.1[18] was used to deconvolve the composition of individual immune cells. Read counts were converted to TPM. TPM values were used as input for EPIC (`EPIC(bulk = counts.tpm)`).

Differential expression analysis

Differential expression analysis with and without purity correction was performed using DESeq2.[46] The Zenodo accession number for the table with the 484 DGE genes and corresponding clusters is provided in the key resources table.

Recursive correlation clustering

Hierarchical clustering is commonly applied using Euclidean distances on scaled transformed expression values. Scaling is sensitive to outliers, for instance, if there is a small proportion of samples with strong up-regulation of a certain gene, for example by hyper-amplifications, these will affect the spread and will influence Euclidean distance(s). This distance metric is therefore sensitive to outliers. Instead, correlation-based clustering (`d = 1 - correlation(m)`) is less sensitive to outliers, where Spearman's rank can be used to more aggressively suppress outliers. However, this method is relatively sensitive to the overall coverage, and we observed that genes relatively rich in zero counts had a somewhat lower correlation to all other genes. This is likely the result from intrinsically discrete data resulting in ties, combined with a rank-based metric. Regardless, the magnitude and direction of its correlations showed a profile consistent with other genes within the same cluster. Since the magnitude and directions of the correlation are consistent with other genes but the data for that gene did not seem powerful enough to estimate high correlation, we took the correlation of the correlation as the distance metric: `d = 1 - correlation(correlation(m))`. This distance metric was used as input for hierarchical clustering with the "ward.D2" method. For all performed clustering analyses, VST-transformed read counts were used to represent individual genes. The technique is publicly available as open-source R package that provides a 'gg' interface to corrplot-like functionality at: https://github.com/yhoogstrate/recursiveCorPlot.

Per cluster (C0-C4) signatures

We defined a per-sample signature to serve as a substitute for each of the genes within each of the five clusters using PCA. For C1, we performed PCA on these 23 genes in the VST transformed batch corrected bulk RNA data. From this PCA we used the first component (61.1% variance) as the ECM signature. The same technique was performed for the remaining clusters. In case the loadings for a component were in majority negative, the sign of the signature was inverted.

Marker gene list

The top 200 genes used to mark neurons, astrocytes, endothelial, oligodendrocytes, TAM, were obtained from literature.[17]

Nuclei isolation and snRNA-seq

For the *van Hijfte* and the *Bolleboom-Gao* dataset, an in-house sample was used for single nuclei RNA-seq. For van Hijfte, after resection, the sample was immediately processed for snRNA-seq. For Bolleboom-Gao, the sample was snap frozen after surgery and later processed for snRNA-seq. For both, first, nuclei were isolated from the tissue using a two pestle, all glass, Dounce tissue grinder (Kimble, #885300-0002). The tissue was suspended in 1.5 mL of Nuclei EZ Lysis Buffer (Sigma, Darmstadt, Germany, N3408) in the tissue grinder and tissue was broken up with 25 strokes of each pestle. The suspension was transferred to a 15mL tube (Corning, Arizona, USA, #T1943-1000EA), an additional 1.5 mL EZ lysis buffer was added to bring the volume to ~3.0 mL and incubated for 5 min on ice. Next, the suspension was filtered through a 70 μm cell strainer (Corning, #431751) and centrifuged at 500g for 5 min at 4°C in a swinging bucket centrifuge. The supernatant was discarded and the nuclei were resuspended in 3.0 mL EZ lysis buffer. After 5 min incubation on ice, the sample was again centrifuged at 500g for 5 min at 4°C and resuspended in a resuspension buffer (1xPBS, 1.0% BSA, 0.2U/μL RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen, Waltham, MA, USA, #10777019), protease inhibitor cocktail (cOmplete™, EDTA-free Protease Inhibitor Cocktail, Roche, Basel, Germany, #11873580001), 5.0 μg/mL Hoechst (Sigma, # H3570)). After 5 min, the nuclei were filtered through a 40 μm cell strainer (Corning, #431752) and immediately taken to the FANS (Fluorescence Activated Nuclei Sorting) to clear the sample of excess debris. Nuclei quality was assessed and single nucleus RNA sequencing was performed using the 10x V3 technology according to manufacturer protocols. For Bolleboom-Gao, based on the nuclei concentration after FACS sorting the sample was centrifuged for another 5 min at 500g and the supernatant was removed to get an optimized concentration of nuclei and resuspended before using 10x.

Single-cell and single nucleus RNA-seq

We processed the in-house ultra-deep sequenced glioblastoma snRNA-seq sample referred to as the *van Hijfte* dataset (n = 1) and the in-house peri-tumoral neuronal-rich *Bolleboom-Gao* dataset (n = 1). Briefly, for both datasets, after library prep reads were sequenced at a NovaSeq 6000. For generating FASTQ files, Cell Ranger mkfastq (v3.0, 10x Genomics) was used. Transcripts were aligned to a pre-mRNA GRCh38 human reference genome and quantified using Cell Ranger Count (version 3.1.0, 10x Genomics).

Additionally, we obtained and processed single-cell/nucleus RNA-seq data from publicly available studies containing at least a sample which was not specifically chemically enriched for tumor cells; the Couturier (n = 5),[39,49] Diaz/Wang (n = 7),[37] Yuan (n = 5) and CPTAC-3 (n = 9)[40,50] datasets.

Subsequently, all samples from all datasets were pre-processed and normalized using Seurat 4[51] and clusters were extensively manually annotated for cell types, according to marker expression in UMAP projections and DotPlots. For the ECM analysis, we proceeded with 10 samples that harbored cells with pericyte marker expression (Couturier n = 4, Yuan n = 4, Diaz/Wang n = 1, van Hijfte n = 1). For C4, we proceeded with 11 samples yielding neurons (van Hijfte n = 1, Bolleboom-Gao n = 1, CPTAC-3, n = 9).

IVY GAP ISH data

RNA ISH COL1A1 and H&E-stainings were taken from the IVY GAP portal.[36] IVY-GAP samples were selected for presence of regions annotated as cellular tumor and infiltrating tumor. Using QuPath, RNA-ISH images were denoted as cellular or infiltrating tumor

regions according to IVY-GAP annotations. Annotated areas were separated into regions positive or negative for staining by setting a manual threshold for one image. For annotated regions, the number of positive and negative pixels was exported. The fraction of positive pixels was compared between cellular tumor and infiltrating tumor in R using a chi-square test.

IF stainings

Multiplex IF stainings were performed using the Ventana Benchmark Discovery (Ventana Medical Systems Inc). 4μm FFPE tissue sections were deparaffinized and heat-induced antigen retrieval was performed using cell conditioner solution CC1 (#950-500, Ventana) for 64 minutes at 97°C.

For neuron analysis, we implemented two multiplex stainings. First, for illustrative images, a multiplex IF staining with NEUN, MOG, GFAP and SOX2 was performed. Tissue samples were consecutively incubated for 32 min with NEUN (#ABN78, Millipore, 1:100), MOG (#HPA021873, Sigma Aldrich, 1:500), GFAP (#760–4345, Ventana, 1.7 μg/mL) and SOX2 (#371R-18, Cell Marque, 2.4 μg/mL) antibodies, incubated with secondary antibody omnimap anti-rabbit HRP (#760-4311, Ventana, 12 min) and visualized with DCC (#760-240, Ventana, 8 min), R6G (#760-244, Ventana, 4 min), Red610 (#760-235, Ventana, 12 min) and FAM (#760-243, Ventana, 4 min) respectively. Antibody denaturation was performed between antibodies using CC2 (#950-123, Ventana) at 100°C for 20 minutes. An essentially similar staining was performed for neuron quantification where a multiplex IF staining was performed with NEUN and GFAP only. In these experiments, the tissue samples were consecutively incubated with GFAP and NEUN for 32 minutes at 37°C each, followed by secondary antibody detection with omnimap anti-rabbit HRP (12 min) and visualized with Red610 (8 min) and FAM (4 min).

For the analysis of collagen distribution around vessels, multiplex IF staining with CD31, PDGFRB, GFAP and COL1A1 was performed similar to described above. We consecutively incubated slides for 32 min with CD31 (#131M-98, Cell Marque, 0.4 μg/mL), PDGFRB (#3169, Cell Signaling, 1:100), GFAP and COL1A1 (#NB600-408, Novusbio, 1:1000) antibodies, followed by secondary HRP antibody detection and visualized them with DCC (8 min), Cy5 (#760-238, Ventana, 4 min), Red610 (12 min) and FAM (4 min). To analyze the co-localization of pericytes with blood vessels, a multiplex IF staining with CD31, PDGFRB and GFAP antibodies was performed. Tissue samples were incubated with CD31, GFAP and PDGFRB antibodies, followed by incubation with secondary HRP antibodies, and visualized using R6G (8 min), Red610 (12 min) and Cy5 (4 min). All secondary antibodies mentioned above were omnimap anti-mouse HRP (#760-4310, Ventana, 12 min) or omnimap anti-rabbit HRP (12 min).

All slides were incubated in PBS with DAPI for 15 minutes and covered with anti-fading medium (DAKO, S3023).

IF image analysis

To quantify the neuron counts and their relation to cell density in tumor regions, 15 primary–recurrent resection pairs (30 samples) and 5 healthy controls were used. QuPath software[53] was used to analyze whole slide images. Each image was divided to 200 μm × 200 μm tiles. Poor quality tiles were excluded from the analysis (e.g. those with smears or dust), blinded with respect to sample type (primary/recurrent/control brain). In each tile, cells were segmented using the Cell detection function on the DAPI signal. Neurons (NEUN+ cells) were quantified using the mean nuclear NEUN signal intensity, for which a threshold was set manually in all tissue samples. GFAP+ cells were identified using the mean cellular GFAP signal intensity, for which the threshold was determined manually in all tissue samples. Tumor high and tumor low regions were differentiated based on the ranked total number of cells per tile. For paired tests, neuron fractions were compared between primary and recurrent samples. A Wilcoxon signed-rank test was used to test significance. Between tumor low and tumor high regions, the mean absolute neuron counts per sample were compared. Statistical comparisons were performed using Wilcoxon rank-sum tests. For each IF, annotated tiles were exported through R/ggplot and used as overlay for whole slide images. Neuron fractions were split into 10 intervals using the `'cut'` function.

For statistical analysis of collagen distribution around vessels, 8 tumor-resection IF slides were used. 2 slides were excluded from the analysis due to insufficient regions that passed quality control. The remaining 6 tissue section images were divided into 500 μm × 500 μm image tiles using QuPath software. Poor quality tiles were excluded from the analysis. The remaining tiles were exported to the ImageJ software[52] and were analyzed using a custom macro (GitHub folder: `./if`) in which the vessels were first detected using the CD31 signal, and then COL1A1 signal intensity was quantified around these vessels. Areas with signals above a defined CD31 threshold and larger than 150 pixels were selected as vessels. The CD31 threshold was set as a function of the mean, maximum and standard deviation values of the CD31 signal in each tile. After vessel detection, four concentric rings of increasing size were drawn around each vessel and the COL1A1 signal was measured in each ring. Next, the mean COL1A1 signal within the rings per tile was determined. To assess the association between tumor and COL1A1 signal, the correlation between the median GFAP signal intensity and mean COL1A1 signal intensity per tile was assessed using Spearman's rho. Next, tumor high and tumor low regions were differentiated based on the median GFAP signal intensity of tiles across one resection. The difference in mean COL1A1 signal between tumor low and tumor high was tested separately for all samples using a Wilcoxon rank-sum test. Mean and median signal intensities were log-transformed for visualization.

To demonstrate whether PDGFRB (pericytes) co-localized with CD31 (endothelial cells), 26 tumor samples were analyzed. The tissue section images were divided to 500 μm × 500 μm image tiles using QuPath and poor-quality tiles were removed. For each tile, the mean signal intensities for PDGFRB and CD31 were measured using ImageJ (GitHub folder: `./if`) and the Spearman correlation coefficient between the mean intensities per tile was estimated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Pathway enrichment analysis was performed through the web portal of g:Profiler.[54] Plots were made using base R and tidyverse. DESeq2 was used for normalization, VST transformation and differential gene expression.[46] Survival forest plots of Cox PH models were generated using the `survminer:ggforest` function. Kaplan Meier curves were generated using the `survminer::ggsurvplot` function. For the PCA biplots the function `factoextra::fviz_pca_var` was used.