

1 Title: Tumor cell phenotype and heterogeneity differences in IDH1 mutant vs wild-type gliomas

2 Short title: Tumor cell phenotype and heterogeneity differences in IDH1 mutant vs wild-type gliomas

3 Michael E. Berens^{1¶}, Anup Sood^{3,¶}, Jill S. Barnholtz-Sloan^{2,¶}, John F. Graf³, Sanghee Cho³, Seungchan
4 Kim⁴, Jeffrey Kiefer^{1, #a}, Sara A. Byron¹, Rebecca F. Halperin¹, Sara Nasser¹, Jonathan Adkins¹, Lori
5 Cuyugan¹, Karen Devine², Quinn Ostrom², Marta Couce², Leo Wolansky^{2, #b}, Elizabeth McDonough³,
6 Shannon Schyberg³, Sean Dinn³, Andrew E. Sloan⁷, Michael Prados⁵, Joanna J. Phillips⁵, Sarah J. Nelson⁶,
7 Winnie S. Liang¹, Mirabela Rusu^{3, #c}, Maria I. Zavodszky^{3, #d}, Fiona Ginty³

8

9 ¹Brain Tumor Unit, Cancer and Cell Biology Division, Translational Genomics Research Institute, Phoenix,
10 Arizona, United States of America

11 ²Department of Population and Quantitative Health Sciences and Case Comprehensive Cancer Center,
12 Case Western Reserve University School of Medicine, Cleveland, Ohio, United States of America

13 ³GE Research Center, Niskayuna, New York, United States of America

14 ⁴Department of Electrical and Computer Engineering, Roy G. Perry College of Engineering, Prairie View
15 A&M University, Prairie View, Texas, United States of America

16 ⁵Department of Neurological Surgery, Helen Diller Cancer Center, University of California San Francisco,
17 San Francisco, California, United States of America

18 ⁶Department of Radiology and Biomedical Imaging, University of California, San Francisco, California,
19 United States of America

20 ⁷Department of Neurosurgery, University Hospitals-Seidman Cancer Center, Cleveland, Ohio, United
21 States of America

22 ^{#a}Current Address: Systems Oncology, Scottsdale, Arizona, United States of America

^{#b}Current Address: Department of Diagnostic Imaging and Therapeutics, UConn Health, Connecticut, United States of America

^{#c}Current Address: Stanford University, Department of Radiology, Stanford, California, United States of America

^{#d}Current Address: Biogen, Cambridge, Massachusetts, United States of America

*Corresponding author

Email: mberens@tgen.org

[¶] Joint first authors

Abstract

Glioma is recognized to be a highly heterogeneous CNS malignancy, whose diverse cellular composition and cellular interactions have not been well characterized. To gain new clinical- and biological-insights into the genetically-bifurcated IDH1 mutant (mt) vs wildtype (wt) forms of glioma, we integrated multiplexed immunofluorescence single cell data for 43 protein markers across cancer hallmarks, in addition to cell spatial metrics, genomic sequencing and magnetic resonance imaging (MRI) quantitative features. Molecular and spatial heterogeneity scores for angiogenesis and cell invasion differ between IDHmt and wt gliomas irrespective of prior treatment and tumor grade; these differences also persisted in the MR imaging features of peritumoral edema and contrast enhancement volumes. Longer overall

survival for IDH1mt glioma patients may reflect generalized altered cellular, molecular, spatial heterogeneity which manifest in discernable radiological manifestations.

Introduction

Gliomas represent the most common type of malignant brain tumor, comprising 81% of malignant brain and central nervous system (CNS) tumors and 27% of all brain and CNS tumors in the United States(1). While gliomas are relatively rare in the general population with an average annual age-adjusted incidence of 6.2 per 100,000, these primary brain tumors contribute significant morbidity and mortality, with glioblastoma carrying a 5-year survival rate of less than 6%(1).

The landscape of our knowledge about molecular features required for accurate diagnosis and prognosis for glioma patients has advanced greatly in the last decade (2-5). Molecular subclassification highlights different genetic underpinnings of glioblastoma (6), which offer some prognostic insight (7), likely attributable, in part, to gene expression patterns influencing vulnerability to radiation (8). The World Health Organization (WHO) classifies gliomas into defined categories based upon histologic and molecular features and are assigned into four grades of increasing aggressiveness. Additionally, the methylation status of O6-methylguanine-DNA methyltransferase (MGMT) has been implicated as a useful biomarker for conferring tumor resistance to alkylating chemotherapies; methylation of the MGMT promoter leads to transcriptional silencing of MGMT, which is associated with loss of MGMT expression and increased response to alkylating chemotherapies such as temozolomide (TMZ) (9). Analysis of DNA methylation from gliomas identified a DNA methylation-based phenotype, G-CIMP, which is characterized by global hypermethylation of CpG islands and is predictive of increased survival; this G-CIMP phenotype is associated with isocitrate dehydrogenase (IDH) mutation status (3, 4, 10).

IDH wild type (wt) in histologically defined low-grade gliomas is associated with poor clinical prognosis that more resembles glioblastoma multiforme (GBM), which generally lack *IDH* mutation (*IDH*mt) (3, 11). Conversely, *IDH* mutations are observed in the majority of lower-grade gliomas and are associated with better clinical outcomes. In low-grade gliomas with *IDH* mutations, 1p/19q codeletion is further associated with oligodendrogliomas and better chemotherapeutic response (12). The validation of some of these molecular biomarkers for diagnosis and prognosis has prompted WHO to include molecular subclasses into their latest classification schema for CNS tumors, including addition of MGMT methylation and *IDH*-mutant/*IDH*-wildtype classifications for glioblastoma, as well as *IDH*-mutant and 1p/19q-codeleted classifications for oligodendrogliomas and anaplastic oligodendrogliomas (13).

Intratumoral heterogeneity, even across molecular subtypes, is now also appreciated as a characteristic of glioma and glioblastoma (14) and has been shown to occur temporally (15), spatially (16) (17), for oncogenic drivers (18), and through the stem cell lineage (19). Heterogeneity features have been identified by radiologic imaging with quantitative features, including distinguishing between *IDH*1mt vs wt gliomas(20) . While these and other studies have interrogated glioma heterogeneity using bulk transcriptomics and single cell sequencing, medical imaging has also provided valuable heterogeneity insights (albeit limited by resolution e.g. 1 voxel, the volumetric unit, in a 1.5 T MRI image contains approx. 1-2 million cells). There have been no investigations to date of cell-level spatial heterogeneity in protein expression or cell types and how they relate to the radiological appearance of these tumors on MRI. Understanding malignant progression in *IDH*1 mt and wt patients at multiple scales and in a spatial context is pivotal to delineating biological events underlying glial tumors and may facilitate tailored treatment approaches as well as reveal new therapeutic targets. Moreover, this multi-scale characterization may facilitate the identification of quantitative metrics derived from non-invasive imaging, i.e. MRI, which correlate with or predict molecular and cellular phenotypes. Such metrics may

be evaluated for new patients prior to biopsy or surgery and might inform about the presence of certain cellular characteristics that may affect treatment response or outcome.

To discern multimodal differences in relation to IDHmt status, we conducted a multiscale interrogative workflow which combines multiplexed immunofluorescence and single cell spatial analysis of fixed glioma tissue, bulk genomic tumor sequencing, MR imaging quantitative features of the whole tumor and subregions, and patient outcomes. Multiscale datasets were assembled from treatment-naïve cases of grade 2, 3, and 4 astrocytoma/oligodendroglioma (n=20, referred as treatment-naïve glioma) as well as from recurrent (previously-treated) grade 4 astrocytoma (glioblastoma) (n=16, referred as recurrent GBM). Tumor tissue punches from diagnostic paraffin blocks were assembled in duplicate (glioma) or triplicate (recurrent GBM) into tissue microarrays for multiplex immunofluorescence staining (21) using 43 markers to identify cell types and functional states corresponding to cancer hallmarks (22). Exome sequencing data was processed for mutations, copy number aberrations, as well as insertions and deletions. Deconvolution of gene expression data from bulk tumor specimens afforded comparisons of protein levels and transcript levels across cognate specimens. An expert neuroradiologist (LW) outlined on MRI of the treatment naïve glioma, and of recurrent GBM (SJM), while advanced deep learning methods were utilized to delineate necrotic and enhancing cores, as well as peri-tumoral edema. Morphologic features assessed the volumes of the different regions and their ratios, while simple features, T1 weighted post contrast (T1 Post), Apparent Diffusion Coefficient (ADC), and Fluid Attenuated Inversion Recovery (FLAIR), were extracted from different MRI protocols.

Various MRI-focused studies (23-26) have investigated the ability of imaging features to predict IDH1 mutational status. Studies focused on assessing the tumor volume, contrast enhancement status (27), Visually Accessible Rembrandt Images (Vasari) feature set (28, 29), radiomics features (30) or features that were derived via convolutional neural networks (31), among others and used these to train

predictive models of IDH1 mutational status. These studies showed great ability to predict IDH1 mutational status with accuracies as high as 89.1% and area under the receiver operator curves (AUC) of 0.95. Other radiogenomic studies have revealed the correlation of IDH1 mutational status with hypoxia induced angiogenesis and identified that the relative cerebral blood volume (rCBV) MRI was able to predict IDH1 mutations status with an 88% accuracy (32). Unlike the latter studies that predict IDH1 mutational status, we seek to reveal correlations between MRI derived quantitative features, cellular composition and spatial cellular heterogeneity to understand the mechanism of disease progression in relation to IDH1 mutational status. Such knowledge could enable creation of predictive models on MRI of disease progression or treatment response without the need for an invasive biopsy.

We show lower cell-level protein expression in IDH1mt vs wt cases. Further, IDH1mt gliomas, irrespective of grade, showed greater spatial heterogeneity but lower molecular heterogeneity of biomarkers associated with angiogenesis (VEGR2, CD31, SMA, S100A4) and invasion (n-cadherin, cofilin, collagen IV, GFAP and vimentin). Similarly, cell classes derived from deconvolution of bulk gene expression data showed the cell class with high expression of most hallmark genes, particularly those belonging to enabling replicative immortality, evading growth suppressors and inducing angiogenesis, were significantly under represented (<10%) in the IDHmt tumors. IDH mutation was co-expressed with ATRX mutations and was mutually exclusive of EGFR and PTEN mutations consistent with known tumor biology. Longer overall survival following diagnosis for IDH1mt glioma patients may reflect generalized altered cellular, molecular and spatial heterogeneity, which is also reflected in the MR images as lower enhancement and higher edema.

Materials and Methods

Patient cohorts

Cohorts of 20 treatment-naïve gliomas (grades 2, 3, and 4 from the Ohio Brain Tumor Study) and 16 post-treatment recurrent glioblastoma (grade 4 from University of California San Francisco(33)) were retrieved based on appropriate patient consent, suitable MR images, FFPE tissue availability, and specimens suitable for next-generation sequencing (**Table 1 for patient summary and S-Tables S-1 and S-2 for additional details**).

Table 1: Summary of patient characteristics of glioma and recurrent GBM cohorts

Cohort	Treatment naïve primary glioma patients	Recurrent/Refractory GBM patients
Patient number	20	16
Median (range) age at diagnosis (years)	57 (26-77)	51 (29-66)
Gender		
Male	12	12
Female	8	4
Ethnicity		
Caucasian	18	15
Hispanic, Asian, African American	0,1,1	1,0,0
Histologic grade		
II	5	-
III	7	-
IV	8	16
IDH1/2 mutation status		
Mutant (IDH1 R132H)	8	3
Wildtype	12	13
1p19q codeletion		
Codeletion	4	-
Non-codeletion	11	-
Not available	5	-
Median (range) survival (days)		
Grade II	1120 (420-2326)	-
Grade III	487 (370-2964)	-
Grade IV (GBM)	438 (222-541)	1031 (396-3771)

Workflow for multi-modal data generation and integration

Using the methods provided below, three parallel analytical interrogations of the treatment-naïve glioma and recurrent cases were pursued: multiparametric MRI; multiplexed immunofluorescence tissue imaging; and RNA and DNA sequencing. **Fig. 1** depicts the overall workflow for this multi-modal data generation, including multiple analytical approaches to cluster and differentiate clinically variable phenotypes. Given the two cohorts of different clinical characteristics and the multi-modal nature of the data, our analysis was performed stratified by cohort, yet we aimed at identifying associations that are consistent across the two cohorts.

Fig. 1. Overall workflow for generating multi-scale, multiparametric data, extraction of various features and/or conversion to higher scales and multiple analysis approaches to differentiate clinically variable phenotypes. Multi-parametric MRIs (Panels A & B) were segmented for ROIs and various image features to characterize tumor and subregions (necrosis, enhancing and edema) within the tumor. Multiplexed immunofluorescence tissue analysis (Cell DIVE) (Panel C) provides (left-to-right) a virtual H&E (vH&E), which is a pseudo-colored DAPI and AF image, and corresponding overlays of 46 markers (examples shown are for proliferation and angiogenesis markers). Single cell data were generated for every multiplexed marker and intensity binned into 3 tiers (low, medium or high) for each cell. Cell level biomarker data was integrated with known biological pathways knowledge base and used to compute molecular states of individual cells. For visualization purposes, the molecular state of a cell is overlaid on the vH&E image (Panel D). Genomics data (Panel E & F), including IDH1 mutation status, were summarized into pathways, cancer hallmarks, and enrichments for each tumor. Cell-level and MRI feature data were clustered across all glioma patients and by IDH 1 status (Panel G); finally, molecular and spatial heterogeneity were analyzed relative to IDH1 mutation status or tumor grade (Panel H).

Multiplexed immunofluorescence imaging of disease and cellular biomarkers

Using the original diagnostic FFPE tissue blocks of each case studied, dual (treatment-naïve glioma) or triplicate punches (recurrent GBM) were selected for tissue microarray (TMA) construction and subsequent multiplex immunofluorescence staining and imaging (MxIF). Two replicate slides were used

for the treatment-naïve glioma TMAs and 3 replicate slides were used for the recurrent GBM TMAs. Control cores (2 per cancer type) were included on all slides for glioma, prostate, melanoma, lung, breast cancer to verify antibody performance. Briefly, the Cell DIVE™ platform (GEHC), which allows in situ probing of up to 60 biomarkers in a single 5µm FFPE TMA tissue section, followed by image processing, registration and single cell analysis, was used (21) (**S1 A-C Fig.**). After a two-step antigen retrieval step, the sample underwent repeated cycles of staining, imaging and dye signal quenching (**S1-A Fig.**) for a total of 43 biomarkers (**S3 Table**), representing members of different cancer hallmarks, cell lineage and cell segmentation(22). Markers of iron metabolism were also included as ferroptosis is an emerging field of study with mechanistic ties to glioma cell resistance to therapy (34-36). Antibody clones, staining concentrations and staining sequence are provided in **S3 Table**. The detailed process for antibody validation (testing, conjugation and verification) is described in **S2 Fig.** and described in S information of Gerdes et al (21). Prior to storage, images are automatically processed for illumination correction, registration from multiple rounds using the DAPI image acquired in each round and background (tissue autofluorescence, AF) removal by subtracting the image of tissue acquired prior to staining from the image after staining (**S1-B Fig.**).

Image quality checks, processing and cell segmentation

Staining quality of all multiplexed images was assessed by visual assessment of staining patterns of individual markers in all samples and compared to controls and/or expected patterns. Since replicate slides were also available, staining intensities were compared across slides. Staining profiles between the treatment-naïve glioma and recurrent GBM cohorts were also compared. Markers that failed or had non-specific staining or very low or negative expression across the whole cohort were excluded from analysis (**S3 A-C Fig.**). The single cell analysis workflow consists of segmentation and quantification steps

(S1-C Fig.) First, image background was suppressed using top-hat filtering followed by multi-level image thresholding. Second, nuclei were segmented using a wavelet-based algorithm that uses both nuclei intensity and shape (blobness) information (37). Nuclear segmentation was followed by whole-cell segmentation, where synthetic cell boundary was extracted by applying Voronoi tessellation using the nuclei as seeds. To avoid producing very large cells from isolated nuclei, a constraint on the maximum distance between the nucleus and the corresponding cell boundary was applied. Segmented images were visually assessed for segmentation quality and compared with images of DAPI staining and virtual H&E (generated from pseudo-color overlays of DAPI and tissue AF). A single image (1 of 40 (treatment-naïve glioma) failed segmentation due to poor tissue quality. Five images (of 46 total images) from recurrent GBM patients were removed from analysis as these cores contained few (<10%) tumor cells or were cauterized.

The cell segmentation steps were followed by quantification of biomarker intensities in each cell, as well as cell features and morphological properties. The entire set of cell-based measurements, including their IDs and spatial coordinates, are saved as .csv files for statistical analysis in R. For each tissue core, image registration quality was determined based on DAPI correlation with baseline round. Further, using correlation of DAPI signal at cell level from each staining/imaging round, a quality score was generated for every cell in each image, which ranges from 0-1 (0 being no registration, up to 1 for perfect registration). Only cells with quality score above 0.85 were included in the analysis. Scores below 0.5 are generally due to tissue shifting/movement and loss. Excellent correlations (**S4 Fig.**) in number of cells per replicate slide were found for the replicate treatment-naïve glioma TMAs and 2 of the recurrent GBM slides (>0.98). Slightly greater cell heterogeneity was found for one of the recurrent GBM slides but slide to slide correlation was still high (0.74).

Identification of cell clusters and biomarker co-expression

After exclusion of segmentation markers and single-cell MxIF markers which failed QC or staining criteria as described above and shown in **S3 Fig.**, unsupervised cell clustering was performed with all the remaining markers (n=24) using data from reported subcellular staining location (compartments used, shown in **S3-A Fig.**). In total, 85,767 cells (from 20 treatment-naïve glioma cases) and 56,304 cells (from recurrent GBM cases) remained for analysis. Separately, clustering was also conducted with smaller subsets of markers representing individual hallmarks (angiogenesis, proliferation, invasion and motility). Log2-transformed median cell intensity for each marker was used for K-means clustering. After trimming to reduce the impact of extreme outliers at both 2.5% tails, and since the distribution of marker intensity/expression values varies significantly within and between marker type, median cell biomarker values were standardized by the overall marker mean and standard deviation.

Cells were clustered into K groups based on the multi-dimensional marker space (equivalent to number of markers used for clustering). The *kmeans* function provided by *stat* package of *R* (v. 3.4.1) was used with K (= 2 to 15). We used 10 random starts (*nstart* = 10) to address K-means clustering algorithm's sensitivity to initial seeds. We also used multiple metrics to determine the best number of clusters for the data such as Silhouette width, Calinsky criterion, Sum of squares of errors, and consensus clustering metrics. For consensus clustering (*R* ConsensusClusterPlus package), a subset of 5,000 randomly selected cells (due to computational constraints) were used. Consensus clustering iterates the clustering algorithm and examines if each pair of samples consistently clusters together or not. K-means clustering with Euclidean distance as metric was used for 1,000 iterations with 80% resampling. The cumulative distribution function (CDF) plot and the heatmap from consensus clustering were evaluated to guide us to determine the best number of clusters, aided by other metrics mentioned above. For a given K, each cell was assigned to one of the K clusters, and each tumor sample represented according to the

proportion of cells belonging to one of the K clusters. For the purposes of data visualization and interpretation, data was aligned by cluster, IDH1 mutation and patient ID. Biomarker intensities were grouped by cancer hallmarks (invasion; energy metabolism; angiogenesis; stem cells; immune response; proliferation; resisting cell death; DNA damage) and iron metabolism.

Exome and RNA Sequencing

Tumor and normal whole-exome sequencing and tumor RNA-sequencing data from the 20 treatment-naïve gliomas was studied; data was either produced from fresh-frozen tissue (n=16, 8 of which had been sequenced in The Cancer Genome Atlas) or from FFPE tissue (n=4) (**S1 Table**). Twelve of these were newly accessed for de novo analysis, and the remaining data was already available. Pathology estimates suggested those 12 samples all had greater than 70% tumor cell density and less than 50% necrosis. Data from sixteen post-treatment recurrent fresh-frozen glioblastoma tumors previously sequenced as part of a clinical trial (33) (data available in the database of Genotypes and Phenotypes (dbGaP) under accession number phs001460.v1.p1) was also included (**S2 Table**). All 16 of these tumors had whole-exome sequencing data, and fourteen had cognate RNA-sequencing data available.

Constitutional DNA from PBMCs was available for all 36 samples. For the eight fresh frozen glioma samples, Qiagen AllPrep DNA/RNA Mini Kit (cat#80204) was used to isolate DNA and RNA; for the four FFPE treatment-naïve glioma samples, Qiagen AllPrep DNA/RNA FFPE Kit (cat# 80234) was used. Exome libraries were constructed from 200ng of DNA (DIN=3-5 for FFPE samples, DIN >8 for blood and fresh frozen samples) using KAPA Biosystems' Hyper Prep Kit (cat#KK8504) and Agilent's SureSelectXT V5 baits, containing custom content, following the manufacturer's protocols. Custom bait content included copy number probes distributed across the entire genome, along with additional probes targeting tumor suppressor genes and genes involved in common cancer translocations to enable structural analysis. For

high quality RNA (RIN>6.0, DV200>90%), RNA libraries were constructed using Illumina's TruSeq RNA Library Preparation Kit V2 (cat#RS-122-2001) with 500ng inputs. For remaining RNAs (RIN<6, DV200>30%), libraries were prepared using Illumina's TruSeq RNA Access Library Prep Kit (cat#RS-301-2001) with either 40ng or 100ng inputs following the manufacturer's protocol and sample quality/input recommendations. Libraries were equimolarly pooled, quantitated, and sequenced by synthesis on the Illumina HiSeq 4000 for paired 82bp reads. FASTQ were aligned using bwa-mem (version 0.7.8) to the reference genome from 1000 Genomes project build hs37d5 with decoy contigs [b37d5] and Ensembl v74 for annotations. Somatic variants were called using lumosVar2 (38). For this study, a tumor-normal mode was used which the sample fraction of clonal variant groups is set to zero in the constitutional sample.

Deconvolution of samples into cell classes from RNAseq data of bulk samples

Multiple cell classes, characterized by different dominant biological processes, can be discerned by computational deconvolution of bulk gene expression data obtained from complex samples (39, 40). This approach is a practical alternative when available samples are not suitable or available for single-cell sequencing (scRNAseq). Deconvolution assumes that the analyzed sample is composed of a certain number of cell types or different cell states, called *classes*. These classes do not necessarily fall into mutually-exclusive cell types. Instead, they represent quantifiable components of the analyzed samples that exhibit distinct gene- or pathway-attributable behaviors. We employed the previously published CellDistinguisher algorithm to identify sets of genes that are expressed predominantly in one class relative to the others (41). As demonstrated in the Results, gene sets of ~50 genes led to robust assignments of cells into three classes. These distinguisher gene sets were then used to derive class signatures and compute sample compositions (fractions of cell types or classes in each sample) using the

SSKL algorithm from the CellMix package (42). To validate and support our findings with the multiplexed single cell data, we also explored how well cell type assignments based on gene expression data compared to those based on protein expression measured by MxIF.

Calculation of molecular and spatial cell heterogeneity metrics

Molecular and spatial heterogeneity metrics were computed for the MxIF spatially resolved cell data using a previously published heterogeneity analysis algorithm (MOHA) (43). As described in more detail below, this technique computes the molecular “state” of each cell in a tissue section based on the fluorescence intensity of proteins within a given pathway, gene set or cancer hallmark (22). Spatial “states” is a summated score which depicts the degree to which adjacent cells are of the same molecular state. The MOHA algorithm computes heterogeneity (or similarity or divergent states) metrics based on the distributions of these molecular and spatially defined states.

The molecular state of a given cell was defined as an ordered set of the values for each individual marker. A complete list of the cancer hallmark gene sets and the markers that were assigned to them is shown in **S3 Table**. The state of each marker was quantized into an ordinal value representing either a high, medium or low state, using the 33rd and 67th quantiles as the thresholds. The specific ordering of the markers in a given gene set (i.e. concatenation sequence) is arbitrary but was maintained consistently throughout the analysis. This process of computing the molecular state was repeated for each cancer hallmark marker set and for each cell. Next molecular heterogeneity metrics were computed as a normalized Shannon’s entropy of molecular states:

$$\text{Molecular Heterogeneity} = \frac{-\sum_{i=1}^{Nm} Pm_i \ln(Pm_i)}{\ln(Nm)}$$

The Pm_i is the fraction of cells in molecular state i , and Nm is the number of possible molecular states in the system. The number of possible states for a gene set was defined as three raised to the power of the number of markers assigned to the gene set (e.g. $3^{\text{number of markers}}$). The molecular heterogeneity metric value can range from zero to unity (i.e. maximum heterogeneity). For each patient tissue sample, a molecular heterogeneity metric was computed for each cancer hallmark.

Cell Spatial Heterogeneity is a summated score which depicts the degree to which adjacent cells are of the same molecular state as that of an index cell, with each cell in the tissue section serving as an index cell (Example shown in **S5 Fig.**). Identifying neighboring cells is necessary for computing the spatial heterogeneity metrics. Two cells were classified as neighbors if the Euclidean distance between the centers of the two cells was less than 1.3 times the sum of their radii. The cell radii were computed from the segmented cell area after approximating the cell as a circle. The spatial state metric was computed by surveying the neighbors of each cell and counting only the number of neighbors in the same molecular state. This number of neighbors represents the cell spatial state for each pathway or gene set. Having no neighbors in the same molecular state is a valid cell spatial state. Therefore, the cell spatial state can range from zero to the maximum number of neighbors a cell has. After going through every cell and their neighbors, a frequency distribution was established for these cell spatial states. The cell spatial heterogeneity was then computed as a normalized Shannon's entropy of spatial states:

$$\text{Cell Spatial Heterogeneity} = \frac{-\sum_{k=0}^{Z_{max}} Ps_k \ln(Ps_k)}{\ln(Z_{max} + 1)}$$

where, P_{s_k} is the probability of state k , and $Z_{\max}$ is the maximum number of neighbors a cell can have as measured in the tissue sample. For each patient tissue sample, a spatial heterogeneity metric was computed for each cancer hallmark.

MRI imaging protocols and image feature extraction

The multi-parametric MRI (mpMRI) exams of the brain consisted of T2-weighted (T2), T1 weighted pre-contrast (T1 Pre), T1 weighted post contrast (T1 Post), Apparent Diffusion Coefficient (ADC) derived from diffusion-weighted imaging (DWI), and Fluid Attenuated Inversion Recovery (FLAIR) images. The subjects with recurrent GBM were imaged using 3 Tesla GE scanners, while the treatment naïve subjects were imaged at a different institution using 3 Tesla Siemens scanners. Although the acquisitions were consistent in sequence types across institutions, parameters such as relaxation and echo times were different, thus prompting separate image analysis for the two cohorts.

Tumor annotations on the MR images were manually outlined by an expert neuroradiologist to depict the extent of the whole tumor, including peritumoral regions, relative to the FLAIR sequence. To the extent possible, an equivalent normal region on the contra-lateral side of the brain was demarcated. A deep learning approach was trained on the Brain Tumor Segmentation (BraTS) challenge data (44) and was utilized to divide the whole tumor segmentation into enhancing core and necrotic core based on T1-post contrast MRI. A U-net network was trained using the T1 Post contrast MRI to identify the extent of the enhancing and necrotic cores on the BraTS data. The training code and trained model are available (https://github.com/mirabelarusu/deep_learning_inference_browser). The trained model was subsequently applied on the T1 post contrast MR images for the patients in our cohort to segment the enhancing and necrotic cores. The peri-tumoral (edema) regions were obtained by subtracting the enhancing and necrotic core from the whole tumor segmentation. Manual corrections and automatic

postprocessing were utilized when appropriate to improve the precision of the annotations or remove minor disconnected regions. At the completion of these processing steps, an annotation of the whole tumor, the peritumoral (edema) region, enhancing core, and necrosis were obtained for each subject relative to the FLAIR protocol.

Pre-processing steps were applied on the mpMRI prior to feature extraction, including spatial registration to align the FLAIR protocol relative to the others, in order to project the region annotations on the rest of the protocols. Intensity normalization was applied in the entire organ by using the normal regions as reference. Specifically, the intensities were normalized such that the average intensity in the normal region had a value of 1. To perform this normalization, we divided the intensity of each voxel by the average of intensities within the normal region.

Image derived quantitative features were evaluated for each subject. Due to the limited number of subjects in our study, the large number of protocols (n=5) available for each subject and the multiple subregions available for each tumor (n=4), we chose to consider only three protocols (T1-post, FLAIR and ADC) and three tumor subregions (the whole tumor, the peritumoral edema and enhancing core). We represented the tumor subregions by two image-derived quantitative features (mean and standard deviation), resulting in 18 image-derived features per subject. Also, for each subject, we included three morphologic features (the volume of the enhancing core, the volume of the entire tumor and their ratio – which we refer to as the normalized enhancing core volume).

Multimodality data integration and clustering

Finally, we investigated the associations between imaging quantitative features and other variables including cell cluster data, clinical parameters and cancer hallmarks based on cell protein expression, RNA and DNA. Due to the different source and scales of the multimodal data (clinical, MxIF, genomic, MRI), we discretized the most relevant features into “low”, “medium” and “high” groups, based on the

data ranges across the individual cohorts. Features were considered to be relevant for the multimodal association analysis either because there were clinically utilized for decision making, e.g. IDH1 mutation status, age, and grade, or because they showed consistent trends across both treatment naïve subjects as well as recurrent GBM subjects. Based on the discretized variables, subjects were then clustered using hierarchical clustering with the Euclidean distance metrics.

Results

Marker expression differences between IDH1 mt and wt tumors

Univariate and multivariate analysis of biomarker expression in the treatment-naïve glioma cohort showed significant differences in the mean expression of vimentin ($p=0.0002$), VEGFR2 ($p=0.0002$), Nestin ($p=0.003$), Ki67 ($p=0.006$) and HLA1 ($p=0.008$) proteins between the IDHmt and IDHwt tumors (**S6-A Fig.**). Three of these, VEGFR2, Vimentin and HLA1 were also included in the multi-variate model using Random Forest which provided an AUC of 0.87 (error rate 5%) in predicting IDH mutation status (**S6-B Fig.**). Since a majority of IDHmt tumors are derived from oligodendrogliomas which minimally express vimentin and IDHmt tumors are known to have suppressed angiogenic pathways, differential expression of VEGFR2 and vimentin between IDHmt and IDHwt is not surprising.

Cellular and genomic analysis shows cancer hallmark differences in

IDH1 mt vs wt tumors

Cellular differences in IDHmt vs wt tumors

In total, 24 markers across 85,000 cells from the 20 treatment-naïve glioma cases underwent k-means clustering. **Fig. 2** shows unsupervised clustering and segregation of the cells into 7 clusters; marker

intensity organized by cluster, IDH1 mutation and cancer hallmarks (invasion; energy metabolism; angiogenesis; stem cells; immune response; proliferation; resisting cell death; DNA damage) and iron metabolism. Relative biomarker intensities (compared to population mean) for each cluster are shown in **S7 Fig**. Clusters 1 and 4 with above average expression of most hallmarks were composed of cells from just two IDH1wt patients (**Fig. 2**). Clusters 2 and 6 contained the largest numbers of cells (21.0% and 21.9%, respectively, **S7 Fig**.) from the greatest number of cases (12 and 11 cases, respectively, **Fig. 2**), cluster 2 being dominantly composed of cells from IDH1wt tumors while cluster 6 contained cells exclusively from IDHmt cases. Cluster 2 shows lower expression of γ H2AX, Sox2, SMA, and Ncad and higher expression of FTL and FTH1, while most other protein expression was near average of the all clusters. Cluster 6 had lower expression of most of cancer hallmarks (**S7 Fig**.) and only pERK, CD31 and Ncad had slightly above average expression. Clusters 5 with lower than average expression of all hallmarks and cluster 7 with above average expression of most hallmarks were evenly occupied by cells from both IDHwt and IDHmt cases. Notably, both angiogenesis and metabolism-related markers were lower in IDH1mt cases, as was expression of antigen presenting machinery, i.e. HLA1, and invasion markers collagen IV and vimentin. Lower expression of vimentin is consistent with IDH1 mutant tumors originating from oligodendrocyte progenitor cells(45, 46), which minimally express vimentin. Interestingly, IDHwt cells had higher expression of ferritin light and heavy chains, indicating increased iron storage in these cells. Removal of free iron by enhanced iron storage has been implicating in evading ferroptosis by cancer cells. A more in-depth analysis to this pathway in this cell cluster is necessary to determine if evasion of ferroptosis is indeed driving the tumor growth in these patients. **Fig. 3** shows two representative examples of IDHwt and mt tumor samples, with biomarker staining and relative biomarker expression for clusters 2 and 6. Clustering of cells by expression of individual hallmarks (angiogenesis, invasion and reprogramming cellular energetics) also showed significant differences in cluster profiles and distribution of clusters among IDH mt and wt tumors (e.g. dominantly

higher representation of proangiogenic clusters in IDHwt tumors, **S8 Fig.**). Overall, similar staining profiles and biomarker patterns in IDHmt vs wt cases were found in the recurrent GBM cohort.

Fig. 2. Distribution and clustering of cells based on protein expression from all treatment-naïve patients. Unsupervised clustering of MxIF data revealed 7 distinct subsets (clusters) of cells derived from all patients. Cluster 2 is dominated by IDH1wt and Cluster 6 is dominated by IDH1mt cases. Clusters 1, 4 and 7 (which were less diverse patient groups) show higher staining intensities of most MxIF markers (cancer hallmarks) compared to Clusters 2, 5 and 6. Iron Metabolism hallmark was generally high in Cluster 2, but low in Cluster 6.

Fig. 3. IF images & Lolipop plots for Cell-DIVE cluster 2 and cluster 6 hallmarks. MxIF images for representative cases in Cluster 2 (A) and Cluster 6 (D), including a vH&E image (top left), segmented image (top middle) showing individual cells, an image with cluster assignment to individual cells (top right) and a number of single marker or multi-marker overlays representing expression of different hallmark proteins (a: DNA breaks, gH2AX. b: Iron metabolism; FTL, FTH1; c: Cell Death, Cleaved Caspase3; d: Proliferation, EGFR, pERK, Ki67; e: Immune MHC1, PDL1; f: Stemness, Nestin, SOX2; g: Angiogenesis, VEGFR2, SMA, S100A4, CD31; h & i: Metabolism, FASN overlaid on DAPI (h) & GSK3b, PKM2, CA9 (i); j & k: Invasion, GFAP, Collagen IV (j) and Vimentin, Cofilin & NCad (k). Panel B and Panel E show the protein expression profiles of individual clusters (2 & 6, respectively); “lolipop” lines originate at the average expression of proteins in all cells measured from all cases and dots reside at the expression of the proteins in the cluster. Lines moving to the left show lower than average expression, while to the right show higher than average expression. Cluster 2 (Panel C) and cluster 6 (Panel F) trend towards

separating cases by IDH1 mutation status. Specifically, Cluster 6, which shows a lower than average expression of most hallmark proteins, is significantly positively correlated to IDH1 mutation (Panel F); Cluster 2 cells with higher iron metabolism (FTL, FTH1) show a trend towards lower representation in IDH1 mutant samples (Panel C). These positive and negative correlations are even stronger when cluster representation in individual cores is correlated to the overall patient status indicating that there is heterogeneity within these tumors.

Cell cluster alignment with IDH and other glioma related mutations

Fig. 4 shows cluster distribution aligned with IDH mutation status and the other most common mutations in treatment-naïve glioma. In concordance with known biology, IDH1 mutations were found to be mutually exclusive of EGFR and PTEN mutations (**Fig. 4, panel B**). IDH1mt samples appeared to be more homogenous, particularly those with concurrent ATRX mutation, and were mostly dominated by the cluster 6 cell phenotype (lower than average expression of most markers (panel A). Approx. 50% of IDH1wt cases with EGFR amplification had a high proportion of cluster 2 cells (overall, average biomarker expression, and lower DNA damage and stem cell markers, higher iron metabolism markers).

Fig. 4. Cell cluster composition and Oncoprint of treatment naïve gliomas. For each glioma case, Panel A portrays the fractional distribution of its cells within each of the 7 clusters. Panel B depicts the genomic profile of each glioma case.

Cell cluster alignment with RNA expression and IDH status

The degree to which single cell clusters agreed with deconvoluted, transcript-based cell class assignments across treatment-naïve gliomas with IDHmt or wt was also evaluated. Based on the gene

expression data of all measured genes, we identified three cell classes using CellDistinguisher, each class having 50 or more distinguisher genes (**S9 Fig.**). Exceeding three classes resulted in a very short list of distinguisher genes for some classes, which diminishes the utility of comparing behavior or functions across the classes. Classes 2 & 3 were qualitatively similar to protein derived cell clusters 6 and 2 respectively. Ratios of the average staining intensities for 21 markers in clusters 6 and 2 were calculated (**Fig. 5A**). The ratios of the expression values for the same 21 genes were compared between RNA classes 2 and 3 (**Fig. 5B**). Fractional composition of IDHmt and wt cases within cell cluster 2 or 6 (**Fig. 5C**) or within RNA class 2 or 3 (**Fig. 5D**) was determined. Consistent with earlier results, tumors dominated by cluster 2 cells were more likely to be IDHwt, while cases with dominance of cluster 6 were mostly IDH1mt. Similarly, the IDHwt tumors were mainly comprised of RNA class 3 markers while class 2 was more abundant in the IDH1mt (**Fig. 5D**). IDH1wt tumors were enriched in class 3 cells (enriched in genes related to the cancer hallmarks of inducing angiogenesis, enabling replicative immortality and evading growth suppression), while the IDH1mt samples had a lower abundance of genes related to these cancer hallmarks.

Fig. 5. IDH1 mutation status drives cell phenotype at both the gene and the protein

level. Ratios of the average staining intensities for 21 MxIF markers in clusters 6 and 2 were calculated (Panel A). Following deconvolution of the transcriptomes using CellDistinguisher, RNA expression counts (FPKM) for the mRNAs were used to distinguish “class types” (n=3) across the bulk sequenced specimens, then ratios of the expression values for the same 21 genes compared between Class 2 and Class 3 (Panel B. Fractional composition of each patient case within Cluster 2 or 6 (Panel C) or within Class 2 or 3 (Panel D) was determined. Cases dominated by cells belonging to protein cluster 2 were more likely to be found in IDH1 wild-type tumors, while cases for which cells from cluster 6 dominated were mostly IDH1 mutated tumors (Panel C). Similarly,

the fractional composition of glioma cases comprised of gene expression class 3 were present in higher proportions in IDH1 wild type samples, while class 2 cell types were more abundant in the IDH1 mutant ones (Panel D). The distinguisher genes of class 3 were enriched in genes related to cancer hallmarks of “inducing angiogenesis”, “enabling replicative immortality” and “evading growth suppression” (see S4 Fig. and S2 Table).

We have found noteworthy similarity between the cell types and patient compositions identified from the MxIF biomarker intensities and the gene expression data. Except for FASN, GSK3b and NCad, good directional correlation was observed in differential protein and gene expression between cell clusters and RNA classes in the IDH1mt and IDHwt populations (**Fig. 5**). Lack of concordance between H2AX protein and transcript likely is due to staining intensity by anti- γ H2AX antibody reporting only the post-translationally phosphorylated form of the protein (instead of total protein, which the transcript count would more reasonable reflect). The high concordant directionality of 17 of the 21 markers argues for robustness of the biological inference that molecular features in cells from treatment-naïve gliomas are related to IDH1 mutation status. We conclude that biomarker-based clusters 6 and 2 refer to the same cells and/or processes as gene-expression-based classes 2 and 3. Although at individual gene levels, mRNA and protein expression values don’t evidence quantitative direct, strong correlation, our findings indicate that looking at the behavior of cells at the gene set or pathway level can lead to consistent patterns starting from different data types (47, 48).

Intratumor and spatial heterogeneity

In addition to the cell level protein expression and cell composition within the IDHmt and wt tumors, we further investigated molecular and spatial heterogeneity of the biomarkers in each of the hallmark

categories. Examples of the heterogeneity metrics for the cell proliferation hallmark (comprising Ki67, nestin and EGFR) in gliomas and recurrent GBMs are shown in **Fig. 6A**, which shows the discretized (high (2), medium (1), low (0)) expression values for each marker, and corresponding color-coding for each cell. Heterogeneity calculated from the distribution of these states in different tumors shows an inverse correlation between molecular and spatial heterogeneity in both treatment-naïve glioma and recurrent GBM cohorts. IDHwt tumors had higher molecular heterogeneity while IDHmt tumors were more spatially heterogeneous (S10 Fig.). Similar trends were present in both cohorts. **Fig. 6B** shows a scatter plot of heterogeneity in the inducing angiogenesis hallmark with the range of spatial and molecular heterogeneity metrics for gliomas and recurrent GBM samples, also encoded by IDHmt (red) and wt (blue) status. Trends in heterogeneity of this hallmark were similar to those observed for the proliferation hallmarks as well as activating invasion motility hallmark (S10 Fig.). No other significant differences in heterogeneity were found.

Fig. 6. Computed molecular and spatial heterogeneity metrics using the multi-omics heterogeneity analysis (MOHA) tool. The method first converts the continuous marker intensity measures of each segmented cell into an ordinal value representing either a high, medium, or low state. Panel (a) presents an example for the Sustaining Proliferative Signaling cancer hallmark. This gene set is composed of three markers: EGFR, Ki67, Nestin. The state of each of these markers can either be high (2), medium (1), or low (0). Therefore, the three-marker gene set has 27 possible molecular states presented in the color-coded legend (far left). The scatter plot (center) presents the spatial and molecular heterogeneity of treatment naïve gliomas and recurrent GBM samples. Images of tissues from four treatment naïve gliomas (A-D) and four recurrent GBM (E-H) are presented with each segmented cell colored by their expressed molecular state. The spatial state distributions of these eight samples are presented

above the scatter plot. For the 4-gene set “inducing angiogenesis” (SMA [ACTA2], VEGFR2 [KDR], CD31 [PECAM1], and S100A4) hallmark, IDH1 mutation status discriminates those cases with relatively lower molecular heterogeneity and relatively higher spatial heterogeneity in grade III treatment-naïve glioma or recurrent glioblastoma (panel b).

MR feature differences between IDH1 mutant and wildtype patients

Simple features derived from the MR images uncovered differences in discernable elements of brain tumor dispersion from IDH1wt and IDH1mt patients. IDH1wt patients had larger enhancing cores (feature “Normalized enhancing core volume”), but less contrast uptake in the peri-tumoral edema regions (feature “Edema T1 post”). On the other hand, the IDH1mt patients lack a clearly defined enhancing core, but have increased contrast uptake on the T1 post contrast MRI protocol in the peri-tumoral edema region (**Fig. 7**). These trends were observed both in the treatment-naïve glioma as well as the recurrent GBM, and are not surprising since the IDH1mt are known to have less contrast enhancement than the IDH1wt (49).

Fig. 7. MRI-derived features appear to differentiate patients that carry an IDH1 mutation (IDH1mt) and those that are wild type (IDH1wt), regardless if subjects are treatment naïve or have recurring GBM. T1w post contrast MRI for IDH1wt subjects (a,e), and IDH1mt (b,f). The white outlines show the extent of the tumor as delineated by the expert neuroradiologist (LW). (c,g) Across the two cohorts, a similar trend may be notice when comparing the mean T1 post-contrast intensity signal in the peri-tumoral edema region, suggesting an increase in enhancement in the IDH1mt in the peri-tumoral

edema region when compared to the IDH1wt (c and g). An opposite trend is observed when comparing the normalized enhancing core volume across IDH1wt and IDH1mt (d and h), indicating that subjects with IDH1 mutants have limited to no enhancement. None of these comparisons reach statistical significance after multiple comparison correction using false discovery rate.

Other intensity and volumetric features were evaluated on clinically important MRI protocols, e.g. ADC or FLAIR, but they failed to show separation between IDH1 mutational status or a consistent trend across the two cohorts. Thus, our analysis focuses on the normalized enhancing core volume – measuring the enhancing core volume normalized to the entire tumor volume, and the T1w MRI post contrast uptake in the peritumoral edema region. Statistical significance was not achieved for any features after multiple comparison corrections likely due to the small number of patients in each cohort.

Multimodal data association

Unlike previous studies (27),(28-31) that focused on predicting IDH1 mutational status using MRI features, we assessed the correlations of MRI features with genomic and proteomic markers within the angiogenesis hallmark to characterize the differences between IDH1 mutational status. **S11 Fig.** shows that larger enhancing cores are associated with higher RNA expression levels in the Inducing Angiogenesis hallmark. A similar association is observed with the expression levels of protein markers, i.e. S100A4 that is known to promote angiogenesis and metastasis development (50) , and VGFR2 that plays a fundamental role in neovascularization (51). These found associations were consistent regardless of the type of tumor, treatment naïve glioma or recurrent GBM.

When investigating multimodal associations (**Fig. 8**), we can also observe a consistent trend across the two cohorts of patients. Not surprisingly, IDH1 mutations are found in lower grade tumors, younger patients and have better overall survival. As also shown in **Fig. 7** and **S11 Fig.**, IDH1mt tumors have smaller enhancing cores but more contrast uptake in the edema regions and show reduced expression levels of RNA and protein from the Inducing Angiogenesis hallmark (**Fig. 8**, highlighted box). Of the five angiogenesis hallmark cell clusters, cluster 4 (above average expression of VEGFR2, SMA and CD31) and cluster 5 (above average expression of VEGFR2 and S100A4), which are characterized by higher expression of angiogenesis markers, show low cell percentages in the subjects with IDH1 mutations. On the other, the IDH1wt tumors are molecularly more diverse and show more heterogeneous multi-modal variables, yet still a general trend of higher expression levels of RNA and protein markers involved in inducing angiogenesis and reduced overall survival. Clusters with average (cluster 3) and lower than average expression (clusters 1 & 2) were distributed among all patients, however, relative proportion of these compared to the other two clusters was much higher in the IDHmt patients. Age, grade and histology are confounding factors in the recurrent GBM progression cohort as IDH1mt tumors tend to occur at younger age and are generally low grade oligodendrogliomas, however, as the similar trends were apparent in the recurrent cohort, which are all grade IV GBMs, these observations probably reflect differences in biology between the IDH1mt and IDH1wt tumors.

Fig. 8. Comprehensive rendering of multi-scale measurements in gliomas. Multiscale modalities depicted include: 1) clinical information (red), 2) IDH1 Mutational status (blue), 3) MRI derived variables (green), 4) RNA expression level of genes involved in the Inducing Angiogenesis Hallmark (black), and 5) Multiplex Immunofluorescence Angiogenesis markers or Cell clusters (magenta). The data is binned in low, medium and high categories. Across the treatment-naïve gliomas (a) and the recurrent (post-treatment) glioblastoma* (b) cohorts, it can be observed that subjects that carry the

IDH1 mutation have low angiogenesis according to RNA expression levels and expression of S100A4 and VEGFR. The subjects also have high fraction of cells in clusters 1 and 2, and low fraction of cells in cluster 4 and 5 (**S8 Fig.**), cluster profiles of angiogenesis clusters). Moreover, MR Images for the same subjects have lower normalized enhancing cores volumes and measure higher intensities on T1 post contrast. *Recurrent GBM (5 subjects are not shown since they were missing MxIF.)

Discussion

We deployed a multiscale workflow that accommodates biomedical imaging (multi-parameter MR imaging) of glial tumors, *in situ* multiplex immunodetection of discrete biochemical functional states in tissue sections from tumors, and next generation sequencing of DNA and RNA from those same tumors. The data produced by each technology was post-processed to regions-of-interest and features (MRI), molecular state assignments of individual cells in tissue (based on gene sets and signaling pathways interrogated by specific antibodies), and molecular subtyping, pathway and hallmark mapping (determined by mutations and cellular deconvolution from bulk RNA sequencing). A coherent picture of enhanced angiogenesis in IDHwt tumors evident in non-invasive *in vivo* imaging features emerges from the data derived from multiple platforms (genomic, proteomic and imaging) and scales from individual proteins to cell clusters/states as well as bulk tumor. Results are consistent with known observations at the molecular (suppression of proangiogenic markers in IDHmt tumors) and imaging scales (no or low enhancement in IDHmt tumor), but now fill in the gaps on how the two are linked through the intermediate scales of cellular states and their spatial organization. Multiplexed immunofluorescence (MxIF) staining using 43 antibodies on individual tissue sections (duplicate punches in a tissue microarray) afforded insight into the clustering of single cell functional states from 20 treatment-naïve

gliomas (grades 2 – 4) into 7 clusters. Discreet patterns of protein abundance across 7 hallmark phenotypes and 2 biochemical signature events (iron metabolism and DNA damage) suggest that broad segregation of such functional states may be associated with IDH1 mutation status. Among the more robustly discriminating hallmarks between IDH1 wildtype from IDH1mutant gliomas is that of angiogenesis. The enhancement patterns, specifically how much of the tumor enhances (assessed by the “normalized enhancing core volume” feature) and the contrast uptake in the peri-tumoral edema region (Edema T1 post intensity), appear to be consistently correlated with the IDH1 mutational status, a trend that is conserved across the two independent cohorts we investigated. Our findings suggest that the IDH1wt tumors have a more consistent enhancing pattern with a clearly defined enhancing rim and little uptake elsewhere. On the other hand, the IDH1mt tumors have a diffuse appearance on MRI without a well-defined enhancing rim and with higher uptake in the edema region, on account of infiltrating cells. Previous studies have linked poor survival with the peritumoral edema volume (52) and tumor volume (27). Moreover, IDH1mt tumors are known to have less edema (49). From the richness of the molecular heterogeneity portrayed from MxIF scoring, comparing the functional states of adjacent cells (whether they are similar or dissimilar) affords a calculation of spatial heterogeneity across the different hallmark phenotypes. Here we find the unanticipated segregation of both treatment-naïve gliomas as well as recurrent glioblastoma based on IDH1 mutation status within hallmarks of “invasion motility”, “proliferative signaling”, and “angiogenesis”. The genomic profiling depicted what is already known about glial tumors, (the mutual exclusivity of IDH1 mutations with EGFR and PTEN mutations, the co-existence of ATRX mutations only within a subset of IDH1 low grade tumors, etc), but also revealed the heretofore unknown frequent, diminished molecular heterogeneity of IDH1mt low grade tumors. Removal of free iron by enhanced iron storage has been implicated in evading ferroptosis by cancer cells (34, 35). Cluster 2, which was highly represented in IDH1wt tumors showed an increased expression of iron storage markers (FTL and FTH1) and decreased expression of γ H2AX, a marker of DNA breaks (**S7**

Fig.). This is consistent with increased sequestration of iron, making it unavailable for oxidative DNA damage leading to evasion of ferroptosis. A more in-depth analysis of this pathway that includes iron transport, storage and utilization is necessary to determine if evasion of ferroptosis is indeed driving the tumor growth in these patients (53, 54). Inter- and intra- tumoral molecular heterogeneity is a well-recognized feature of GBM (6, 55, 56) and is believed to be the main reason behind treatment failure. Emergence of several single cell analysis platforms has fueled the investigations of intra-tumoral heterogeneity of glioma (17, 57-59) , including tumor-stromal cell interactions (60, 61) as well as interactions between the diverse tumor cell populations (62, 63). Importance of the intercellular interactions among heterogenous tumor cell population is highlighted by the observations of Inda et. al. (62) that EGFR^{mt} cells that are far outnumbered by the EGFR^{wt} population drive enhanced proliferation of these cells by paracrine signaling thereby driving tumor growth. Thus, tools to evaluate molecular and spatial heterogeneity and cell-cell interactions are likely to unravel heretofore unknown mechanisms that drive tumor growth and/or treatment failure. IDH mutation induced suppression of immune response has also been noted previously, however, it has been linked to decreased expression of effector T cell response related genes (64) . Whether this in turn affects the expression of HLA1 in IDH^{mt} tumors is not known.

Study limitations

The key limitations of this study include small sample size, lack of registration of sample derived for molecular analysis to MR images and a limited number of markers representing different hallmarks. The intent of this study was not to generate a diagnostic signature but to evaluate correlation between imaging and molecular features at the hallmark level and to generate a work flow for integrating multiscale multiparametric data to study disease biology. While the sample size (n=20) in the

treatment-naïve glioma cohort was limited, the fact that similar cell clusters existed in another cohort (recurrent GBM) and the correlations between MR and molecular features of angiogenesis hallmark hold for both cohorts is encouraging. Having developed methods to integrate and evaluate such a complex data set, we are in the process of designing a more focused study to interrogate the biology of a specific molecular subtype of GBM that will consider and address the aforementioned shortcomings.

Acknowledgements

We acknowledge the significant support from Alex Corwin, Sireesha Kaanumalle and Christina Lowes for antibody validation testing.

References

1. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. *Neuro Oncol.* 2018;20(suppl_4):iv1-iv86.
2. Brennan CW, Verhaak RG, McKenna A, Campos B, Nounshmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell.* 2013;155(2):462-77.
3. Cancer Genome Atlas Research N, Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med.* 2015;372(26):2481-98.
4. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell.* 2016;164(3):550-63.
5. Consortium G. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro Oncol.* 2018;20(7):873-84.

721 6. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human
722 glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061-8.

723 7. Phillips HS, Kharbanda S, Chen R, Forrester WF, Soriano RH, Wu TD, et al. Molecular subclasses of
724 high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in
725 neurogenesis. *Cancer Cell*. 2006;9(3):157-73.

726 8. Bhat KPL, Balasubramanian V, Vaillant B, Ezhilarasan R, Hummelink K, Hollingsworth F, et al.
727 Mesenchymal differentiation mediated by NF-kappaB promotes radiation resistance in glioblastoma.
728 *Cancer Cell*. 2013;24(3):331-46.

729 9. Hegi ME, Diserens AC, Godard S, Dietrich PY, Regli L, Ostermann S, et al. Clinical trial
730 substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation
731 in glioblastoma patients treated with temozolomide. *Clin Cancer Res*. 2004;10(6):1871-4.

732 10. Noshmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification
733 of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*.
734 2010;17(5):510-22.

735 11. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in
736 gliomas. *N Engl J Med*. 2009;360(8):765-73.

737 12. Cimino PJ, Zager M, McFerrin L, Wirsching HG, Bolouri H, Hentschel B, et al. Multidimensional
738 scaling of diffuse gliomas: application to the 2016 World Health Organization classification system with
739 prognostically relevant molecular subtype discovery. *Acta Neuropathol Commun*. 2017;5(1):39.

740 13. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The
741 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary.
742 *Acta Neuropathol*. 2016;131(6):803-20.

743 14. Aum DJ, Kim DH, Beaumont TL, Leuthardt EC, Dunn GP, Kim AH. Molecular and cellular
744 heterogeneity: the hallmark of glioblastoma. *Neurosurg Focus*. 2014;37(6):E11.

745 15. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, et al. Intratumor
746 heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*.
747 2013;110(10):4009-14.

748 16. Kumar A, Boyle EA, Tokita M, Mikheev AM, Sanger MC, Girard E, et al. Deep sequencing of
749 multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes.
750 *Genome Biol*. 2014;15(12):530.

751 17. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq
752 highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014;344(6190):1396-401.

753 18. Furnari FB, Cloughesy TF, Cavenee WK, Mischel PS. Heterogeneity of epidermal growth factor
754 receptor signalling networks in glioblastoma. *Nat Rev Cancer*. 2015;15(5):302-10.

755 19. Morokoff A, Ng W, Gogos A, Kaye AH. Molecular subtypes, stem cells and heterogeneity:
756 Implications for personalised therapy in glioma. *J Clin Neurosci*. 2015;22(8):1219-26.

757 20. Ryu YJ, Choi SH, Park SJ, Yun TJ, Kim JH, Sohn CH. Glioma: application of whole-tumor texture
758 analysis of diffusion-weighted imaging for the evaluation of tumor heterogeneity. *PLoS One*.
759 2014;9(9):e108335.

760 21. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, et al. Highly multiplexed single-
761 cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc Natl Acad Sci U S A*.
762 2013;110(29):11982-7.

763 22. Kiefer J, Sara N, Graf JL, Kodira C, Ginty F, Newberg L, et al. Hallmarks of Cancer Gene Set
764 Annotation2017.

765 23. Bai HX, Lee AM, Yang L, Zhang P, Davatzikos C, Maris JM, et al. Imaging genomics in cancer
766 research: limitations and promises. *Br J Radiol*. 2016;89(1061):20151030.

767 24. Ellingson BM. Radiogenomics and imaging phenotypes in glioblastoma: novel observations and
768 correlation with molecular characteristics. *Curr Neurol Neurosci Rep*. 2015;15(1):506.

769 25. Pinker K, Shitano F, Sala E, Do RK, Young RJ, Wibmer AG, et al. Background, current role, and
770 potential applications of radiogenomics. *J Magn Reson Imaging*. 2018;47(3):604-20.

771 26. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour
772 heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin*
773 *Radiol*. 2017;72(1):3-10.

774 27. Carrillo JA, Lai A, Nghiemphu PL, Kim HJ, Phillips HS, Kharbanda S, et al. Relationship between
775 tumor enhancement, edema, IDH1 mutational status, MGMT promoter methylation, and survival in
776 glioblastoma. *AJNR Am J Neuroradiol*. 2012;33(7):1349-55.

777 28. Park YW, Han K, Ahn SS, Bae S, Choi YS, Chang JH, et al. Prediction of IDH1-Mutation and
778 1p/19q-Codeletion Status Using Preoperative MR Imaging Phenotypes in Lower Grade Gliomas. *AJNR*
779 *Am J Neuroradiol*. 2018;39(1):37-42.

780 29. Su CQ, Lu SS, Zhou MD, Shen H, Shi HB, Hong XN. Combined texture analysis of diffusion-
781 weighted imaging with conventional MRI for non-invasive assessment of IDH1 mutation in anaplastic
782 gliomas. *Clin Radiol*. 2018.

783 30. Li ZC, Bai H, Sun Q, Zhao Y, Lv Y, Zhou J, et al. Multiregional radiomics profiling from
784 multiparametric MRI: Identifying an imaging predictor of IDH1 mutation status in glioblastoma. *Cancer*
785 *Med*. 2018;7(12):5999-6009.

786 31. Chang K, Bai HX, Zhou H, Su C, Bi WL, Agbodza E, et al. Residual Convolutional Neural Network
787 for the Determination of IDH Status in Low- and High-Grade Gliomas from MR Imaging. *Clin Cancer Res*.
788 2018;24(5):1073-81.

789 32. Kickingeder P, Sahm F, Radbruch A, Wick W, Heiland S, Deimling A, et al. IDH mutation status
790 is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively
791 predictable with rCBV imaging in human glioma. *Sci Rep*. 2015;5:16238.

792 33. Byron SA, Tran NL, Halperin RF, Phillips JJ, Kuhn JG, de Groot JF, et al. Prospective Feasibility Trial
793 for Genomics-Informed Treatment in Recurrent and Progressive Glioblastoma. Clin Cancer Res.
794 2018;24(2):295-305.

795 34. Hangauer MJ, Viswanathan VS, Ryan MJ, Bole D, Eaton JK, Matov A, et al. Drug-tolerant persister
796 cancer cells are vulnerable to GPX4 inhibition. Nature. 2017;551(7679):247-50.

797 35. Viswanathan VS, Ryan MJ, Dhruv HD, Gill S, Eichhoff OM, Seashore-Ludlow B, et al. Dependency
798 of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. Nature. 2017;547(7664):453-7.

799 36. Xie Y, Hou W, Song X, Yu Y, Huang J, Sun X, et al. Ferroptosis: process and function. Cell Death
800 Differ. 2016;23(3):369-79.

801 37. Padfield D, Rittscher J, Roysam B. Coupled minimum-cost flow cell tracking for high-throughput
802 quantitative analysis. Med Image Anal. 2011;15(4):650-68.

803 38. Halperin RF, Carpten JD, Manojlovic Z, Aldrich J, Keats J, Byron S, et al. A method to reduce
804 ancestry related germline false positives in tumor only somatic variant calling. BMC Med Genomics.
805 2017;10(1):61.

806 39. Aran D, Butte AJ. Digitally deconvolving the tumor microenvironment. Genome Biol.
807 2016;17(1):175.

808 40. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for
809 correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinformatics.
810 2017;18(1):105.

811 41. Newberg LA, Chen X, Kodira CD, Zavodszky MI. Computational de novo discovery of
812 distinguishing genes for biological processes and cell types in complex tissues. PLoS One.
813 2018;13(3):e0193067.

814 42. Gaujoux R, Seoighe C. CellMix: a comprehensive toolbox for gene expression deconvolution.
815 Bioinformatics. 2013;29(17):2211-2.

816 43. Graf JF, Zavodszky MI. Characterizing the heterogeneity of tumor tissues from spatially resolved
817 molecular measures. PLoS One. 2017;12(11):e0188878.

818 44. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain
819 Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging. 2015;34(10):1993-2024.

820 45. Alcantara Llaguno SR, Wang Z, Sun D, Chen J, Xu J, Kim E, et al. Adult Lineage-Restricted CNS
821 Progenitors Specify Distinct Glioblastoma Subtypes. Cancer Cell. 2015;28(4):429-40.

822 46. Alcantara Llaguno SR, Xie X, Parada LF. Cell of Origin and Cancer Stem Cells in Tumor Suppressor
823 Mouse Models of Glioblastoma. Cold Spring Harb Symp Quant Biol. 2016;81:31-6.

824 47. Akan P, Alexeyenko A, Costea PI, Hedberg L, Solnestam BW, Lundin S, et al. Comprehensive
825 analysis of the genome transcriptome and proteome landscapes of three tumor cell lines. Genome Med.
826 2012;4(11):86.

827 48. Maier T, Guell M, Serrano L. Correlation of mRNA and protein in complex biological samples.
828 FEBS Lett. 2009;583(24):3966-73.

829 49. Wasserman JK, Nicholas G, Yaworski R, Wasserman AM, Woulfe JM, Jansen GH, et al.
830 Radiological and pathological features associated with IDH1-R132H mutation status and early mortality
831 in newly diagnosed anaplastic astrocytic tumours. PLoS One. 2015;10(4):e0123890.

832 50. Semov A, Moreno MJ, Onichtchenko A, Abulrob A, Ball M, Ekiel I, et al. Metastasis-associated
833 protein S100A4 induces angiogenesis through interaction with Annexin II and accelerated plasmin
834 formation. J Biol Chem. 2005;280(21):20833-41.

835 51. Basagiannis D, Zografou S, Murphy C, Fotsis T, Morbidelli L, Ziche M, et al. VEGF induces
836 signalling and angiogenesis by directing VEGFR2 internalisation through macropinocytosis. J Cell Sci.
837 2016;129(21):4091-104.

838 52. Wu CX, Lin GS, Lin ZX, Zhang JD, Liu SY, Zhou CF. Peritumoral edema shown by MRI predicts poor
839 clinical outcome in glioblastoma. World J Surg Oncol. 2015;13:97.

840 53. Tong L, Yi L, Liu P, Abeysekera IR, Hai L, Li T, et al. Tumour cell dormancy as a contributor to the
841 reduced survival of GBM patients who received standard therapy. *Oncol Rep.* 2018;40(1):463-71.

842 54. Wu T, Li Y, Liu B, Zhang S, Wu L, Zhu X, et al. Expression of Ferritin Light Chain (FTL) Is Elevated in
843 Glioblastoma, and FTL Silencing Inhibits Glioblastoma Cell Proliferation via the GADD45/JNK Pathway.
844 *PLoS One.* 2016;11(2):e0149361.

845 55. Friedmann-Morvinski D. Glioblastoma heterogeneity and cancer cell plasticity. *Crit Rev Oncog.*
846 2014;19(5):327-36.

847 56. Meyer M, Reimand J, Lan X, Head R, Zhu X, Kushida M, et al. Single cell-derived clonal analysis of
848 human glioblastoma links functional and genomic heterogeneity. *Proc Natl Acad Sci U S A.*
849 2015;112(3):851-6.

850 57. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly
851 multiplexed RNA profiling in single cells. *Science.* 2015;348(6233):aaa6090.

852 58. Shah S, Lubeck E, Zhou W, Cai L. In Situ Transcription Profiling of Single Cells Reveals Spatial
853 Organization of Cells in the Mouse Hippocampus. *Neuron.* 2016;92(2):342-57.

854 59. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and
855 analysis of gene expression in tissue sections by spatial transcriptomics. *Science.* 2016;353(6294):78-82.

856 60. Chen Z, Hambardzumyan D. Immune Microenvironment in Glioblastoma Subtypes. *Front*
857 *Immunol.* 2018;9:1004.

858 61. Dey MF, H.; Heimberger, A.B. The Role of Glioma Microenvironment in Immune Modulation:
859 Potential Targets for Intervention. *Letters in Drug Design & Discovery.* 2006;3(7):11.

860 62. Inda MM, Bonavia R, Mukasa A, Narita Y, Sah DW, Vandenberg S, et al. Tumor heterogeneity is
861 an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. *Genes Dev.*
862 2010;24(16):1731-45.

63. Inda MM, Bonavia R, Seoane J. Glioblastoma multiforme: a look inside its heterogeneous nature. Cancers (Basel). 2014;6(1):226-39.
64. Kohanbash G, Carrera DA, Shrivastav S, Ahn BJ, Jahan N, Mazor T, et al. Isocitrate dehydrogenase mutations suppress STAT1 and CD8+ T cell accumulation in gliomas. J Clin Invest. 2017;127(4):1425-37.

Supporting Information

S1 Table. Detailed patient characteristics and datasets for treatment naïve glioma cohort

S2 Table. Detailed patient characteristics and datasets for recurrent GBM cohort

S3 Table: Antibody information and staining sequence

S4 Table: Min and max number of cells per core in glioma and recurrent GBM TMAs

S1 Fig. Multiplexed immunofluorescence (MxIF) workflow

The Cell DIVE MxIF workflow involves repeated cycles of staining, imaging and signal inactivation (panel A), following slide clearing and antigen retrieval. Prior to antibody staining, tissue is stained with DAPI and imaged in all channels of interest to record background autofluorescence (AF) of the tissue. Following background imaging, tissue is stained with 2-3 antibodies and reimaged to capture antigen-specific signal and then undergoes a dye inactivation step to remove the signal. The slide is re-imaged to measure background fluorescence intensity. These cycles are repeated multiple times until all targets of interest have been imaged. Panel 2 shows various image processing steps prior to generating single cell

data. Some of these are performed during imaging itself while others are performed post image acquisition. The steps include, illumination correction, to correct for uneven illumination across the FOV, registration of images from all rounds (using DAPI signal from each round) and tissue AF removal. Panel C: Staining intensity of various cellular and subcellular markers is used to generate cellular segmentation masks. Segmented images are compared with real or virtual H&Es (generated from DAPI stained background images at the beginning of multiplexing) by a trained biologist or pathologist, and images with poor segmentation are removed from analysis. In parallel, marker staining is evaluated by reviewing AF removed images and markers that failed to stain or images with large artefacts are removed from analysis. Marker expression is quantified at cellular and subcellular compartments and data is generated in an easy to use .csv or Excel format which is then analyzed by a variety of different tools/approaches including simple statistical correlations, cluster analysis as well as heterogeneity analysis.

S2 Fig. Antibody validation workflow

A typical antibody validation workflow: Starting with literature reports to identify antibody clones previously used for IHC on FFPE tissue, 3 or more clones per target are identified and evaluated for sensitivity and specificity of the signal on a multi-tissue array (TMA) comprising all major tumor types and corresponding normal tissues. The down-selected antibody is conjugated with CY3, Cy5 or Cy7 at 2 different dye/protein ratio and conjugates validated by staining comparison with unconjugated primary on serial sections of the same TMA. The down-selected conjugate is tested at different concentrations on a TMA with tumor tissue of interest to determine the optimal concentration for staining. In parallel, a set of TMA serial sections are pre-treated with different rounds of bleaching and evaluated for bleaching solution's effect on antigen of interest by comparing the staining among this set. Antigens

with discernible effects are prioritized for staining early in the sequence, immediately after primary secondary staining of targets which failed to conjugate.

S3 Fig. Marker Staining quality assessment

A: Marker staining performance in each cohort (True-positive, False-negative), staining round, subcellular location used for analysis and gene symbol, B: examples of quantitative FOV level correlation of marker intensities on replicate slides, C: Examples of fluorescence image overlays of various hallmark markers showing heterogeneity of expression in astrocytoma.

S4 Fig. Number of segmented cells in serial sections

High correlation in number of segmented cells was observed between serial sections, particularly for the treatment naïve glioma cohort and two out of three sections of the recurrent GBM cohort.

S5 Fig. Example workflow for calculating cell molecular state and cell spatial heterogeneity

Example of how molecular state and cell spatial heterogeneity metrics are calculated, using EGFR as an example. A. Segmentation of cells using DAPI staining and generation of nuclear and extra-nuclear masks; B. EGFR fluorescence intensity is quantified for each cell and discretized as low, moderate, and high. The different levels of cell expression are shown as red (high), green (moderate) or blue (low). C. For each cell (I through v in this cartoon), adjacent neighboring (touching) cells are counted, and their Spatial State is used to sum the Spatial Heterogeneity.

S6 Fig. Uni- (A) and multi-variate (B) analysis of biomarker expression and overall survival as a function of IDH mutation status

A. Differences in individual biomarker expression and survival of IDHmt and IDHwt patients. B. A predictive multivariate model of IDH mutation status.

S7 Fig. Lollipop plots for biomarker expression in each cluster, relative to population median

Protein expression profiles of individual clusters plotted relative to median expression in the whole population. Solid circles represent the average expression in the cluster while direction and length of the lollipop shows difference in expression relative to population median (left-lower, right-higher).

S8 Fig. Cell clusters based on angiogenesis hallmark proteins

Unsupervised clustering of cells using angiogenesis hallmark proteins identified a 5 cluster set. Clusters with lower than average hallmark protein expression (1 & 2) are highly represented in samples with IDH1 mutation. Cluster 4 & 5 with higher expression are proportionally more abundant in IDH1wt samples.

S9 Fig. Abundance of distinguisher genes (mRNA)/class per patient

A: Relative proportion of cells belonging to different CellDistinguisher classes in each sample. Class 3 is highly represented in IDHwt samples. B: shows relative abundance of distinguisher genes grouped by hallmarks in individual classes.

S10 Fig. Molecular and spatial heterogeneity in grade III gliomas and recurrent GBM IDHwt and IDHmt tumors

Molecular and spatial heterogeneity in grade III gliomas and recurrent GBM IDHwt and IDHmt tumors according to the following hallmarks: Invasion and Motility, Cell Proliferative Signaling and Inducing Angiogenesis.

956

957 **S11 Fig. Differences in MR features across the population range of RNA and protein marker**

958 **expression for angiogenesis**

959 Correlation between Normalized enhancing core volume (derived from MRI) and Angiogenesis
 960 estimated from (a,d) RNA expression levels, and, based on multiplex immunofluorescence (MxIF)
 961 angiogenesis markers (b, e) S100A4 and (c,f) VEGFR2; (a-c) shows the plots on Cohort 1 (CW Glioma,
 962 treatment naive) while (d-f) show cohort 2 (UCSF, recurrent GBM). A progressive increasing trend may
 963 be observed in both cohorts when examining the normalized enhancing core volume for low, medium
 964 and high angiogenesis. The trends across the enhancement ratio are also conserved when comparing
 965 RNA with MxIF Angiogenesis. None of these comparisons reach statistical significance after multiple
 966 comparison correction using false discovery rate.

967

Figure 7: MRI-derived features T1 edema and enhancement in IDH1wt and IDH1mt tumors

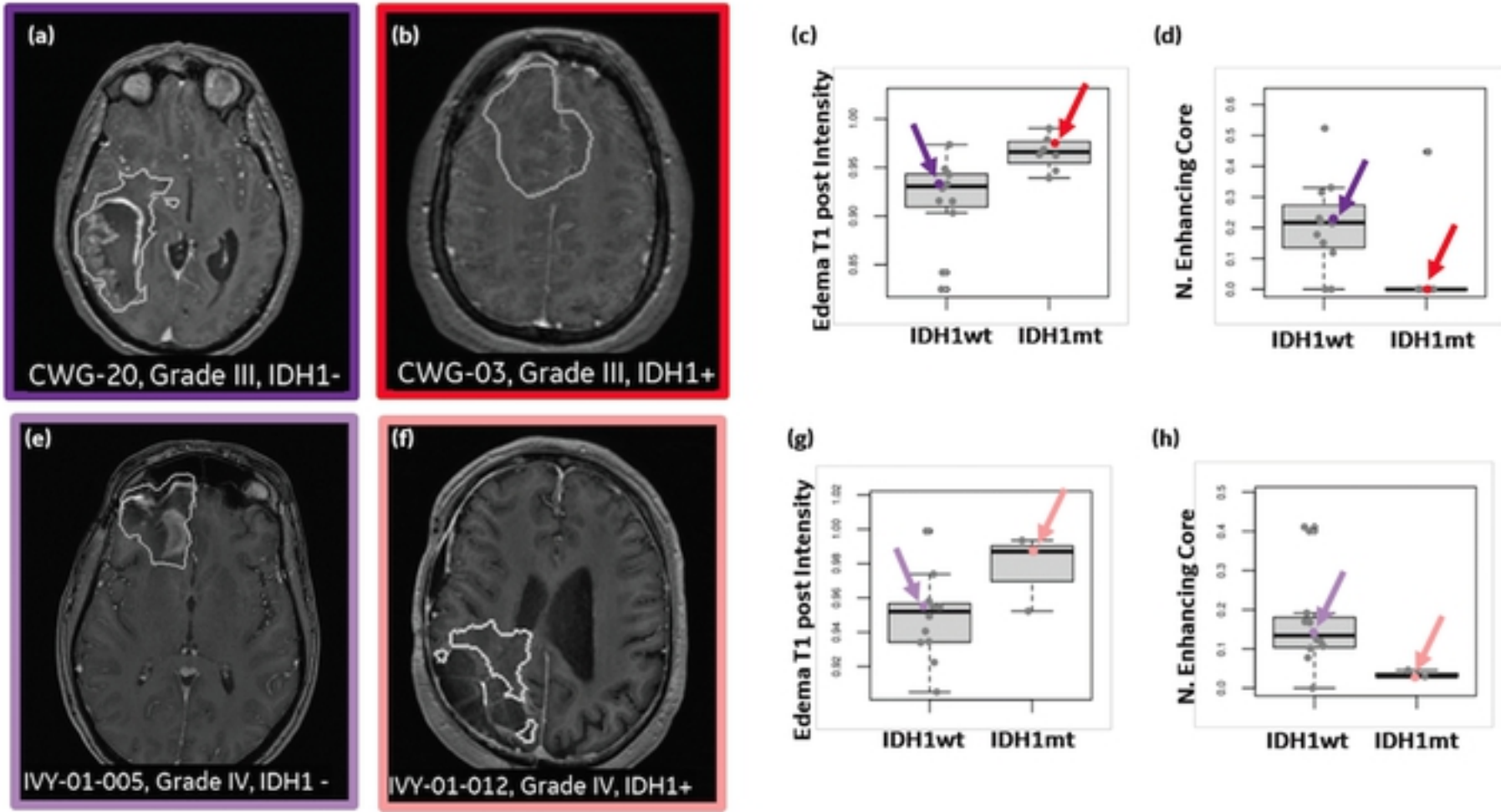


Figure 7

Paper Figures:

Table 1: Patient characteristics summary

Figure 1: Multimodal data workflow

Figure 2: Distribution and clustering of cells based on protein expression from all treatment-naïve patients.

Figure 3: Cluster 2 and 6 profiles and example images for selected biomarkers in IDHwt and IDHmt tumors

Figure 4: Cell cluster composition and oncoprint of treatment naïve gliomas

Figure 5: Correlation between IDH1 mutation status and cell phenotype at both gene and protein level.

Figure 6: Computed molecular and spatial heterogeneity metrics using the multi-omics heterogeneity analysis (MOHA) tool

Figure 7: MRI-derived features T1 edema and enhancement in IDH1wt and IDHmt tumors

Figure 8: Multiscale modality integration including IDH1 mutation, clinical data, survival, RNA expression (angiogenesis markers), protein expression and cell clusters

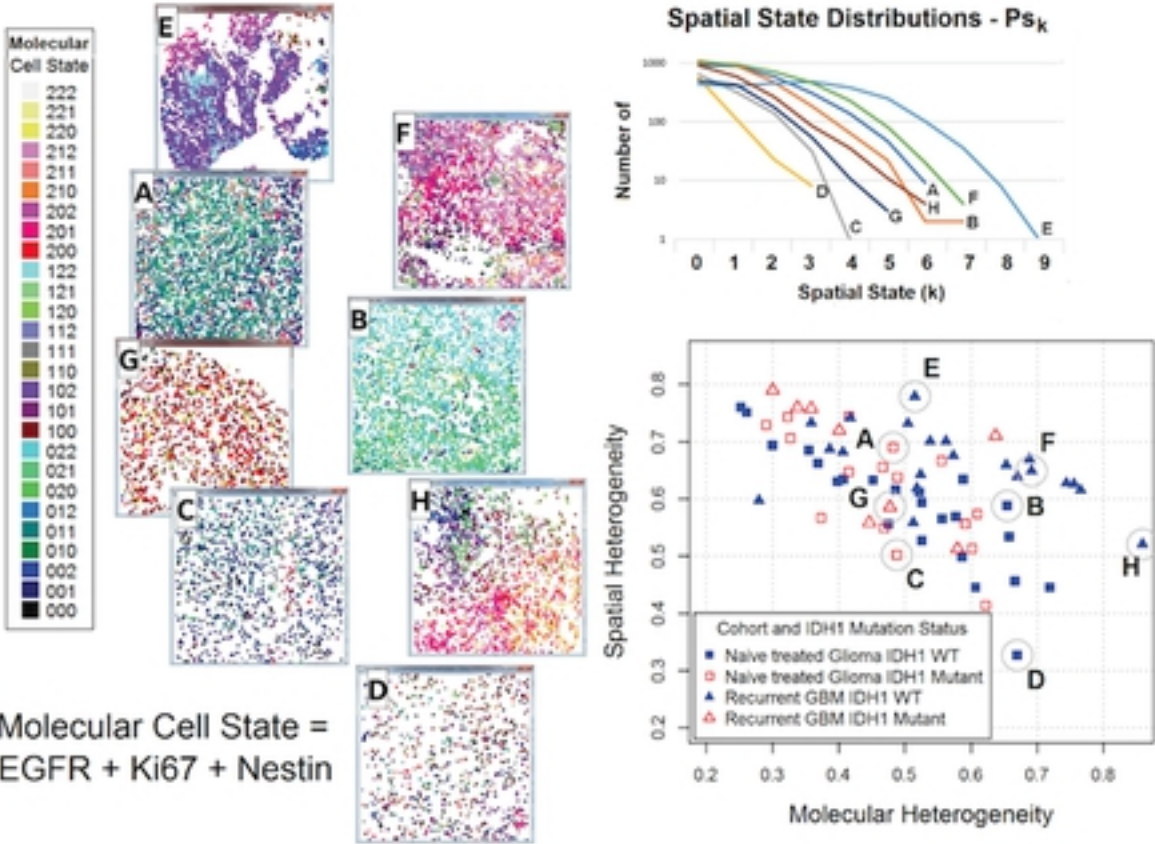
Table 1: Patient Characteristics

Cohort	Treatment naïve primary glioma patients	Recurrent/Refractory GBM patients
Patient number	20	16
Median (range) age at diagnosis (years)	57 (26-77)	51 (29-66)
Gender		
Male	12	12
Female	8	4
Ethnicity		
Caucasian	18	15
Hispanic, Asian, African American	0, 1, 1	1, 0, 0
Histologic Grade		
II	5	-
III	7	-
IV	8	16
IDH1/2 Mutation status		
Mutant (IDH1 R132H)	8	3
Wildtype	12	13
1p19q codeletion		
Codeletion	4	-
Non-codeletion	16	-
Median (range) Survival (days)		
Grade II	1120 (420-2326)	
Grade III	487 (370-2964)	
Grade IV (GBM)	438 (222-541)	1031 (396-3771)



Figure 6: Computed molecular and spatial heterogeneity metrics using the multi-omics heterogeneity analysis (MOHA) tool

(a) Heterogeneity of cell proliferative markers



(b) Heterogeneity of inducing angiogenesis

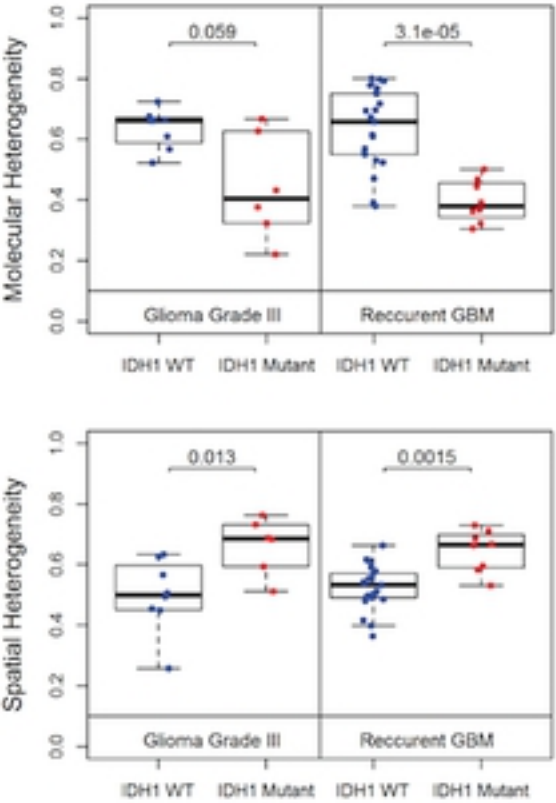


Figure 6

Figure 1: Overall workflow for generating multi-scale, multiparametric data, extraction of features and/or conversion to higher scales and analysis approaches

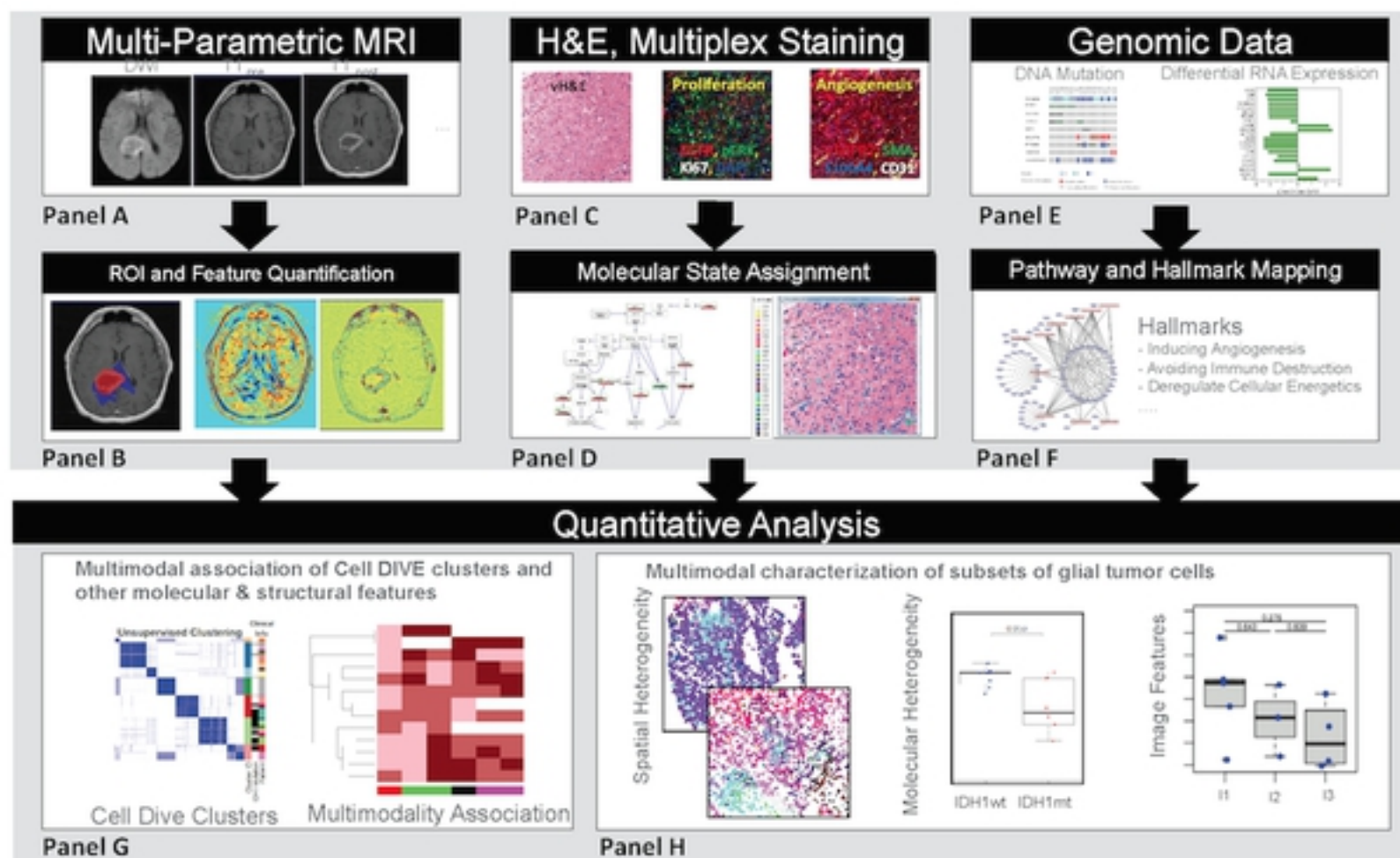


Figure 3: Cluster 2 and 6 profiles and example images for selected biomarkers in IDHwt and IDHmt tumors

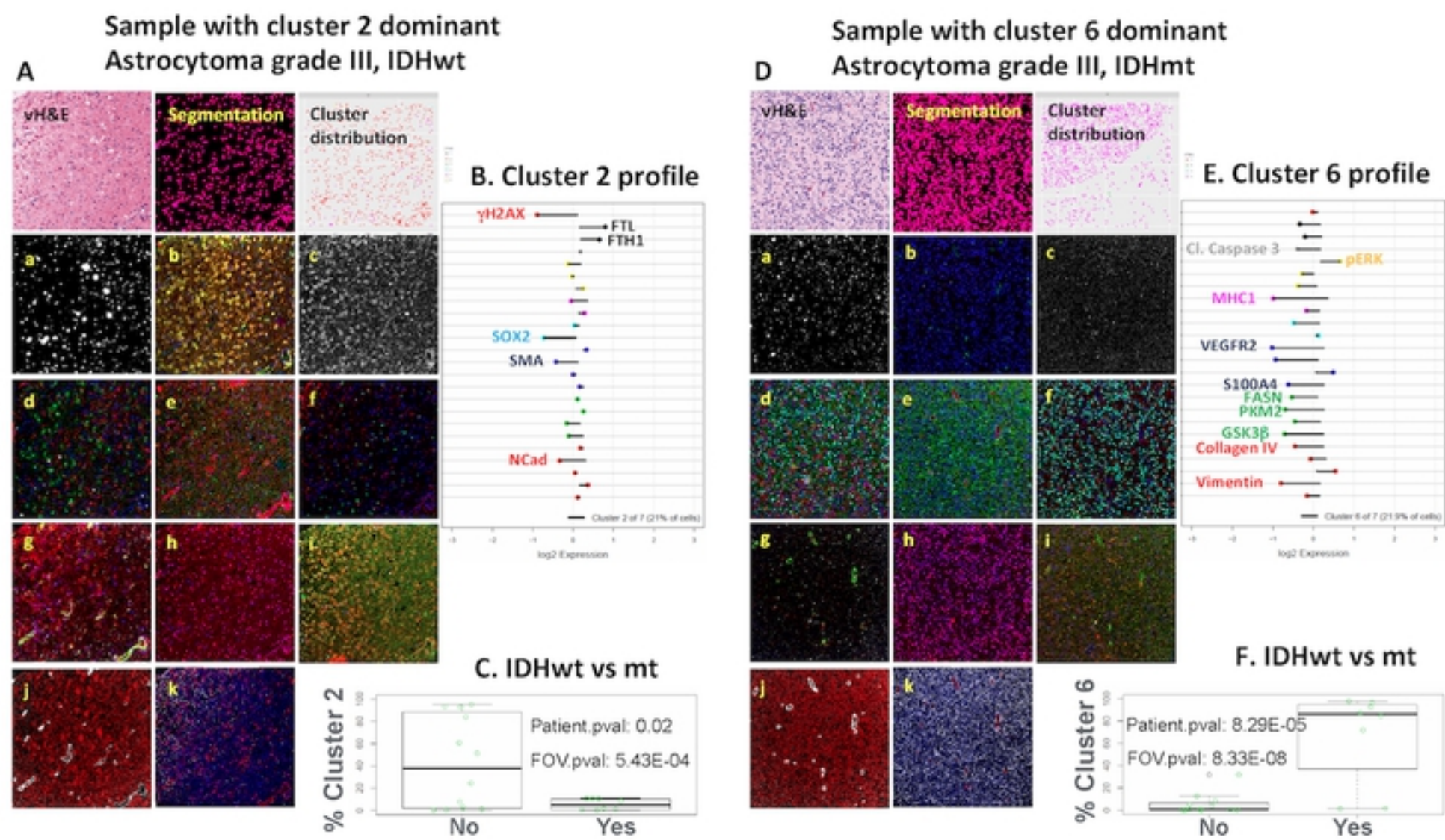


Figure 3

Figure 2: Distribution and clustering of cells based on protein expression in treatment naïve glioma patients

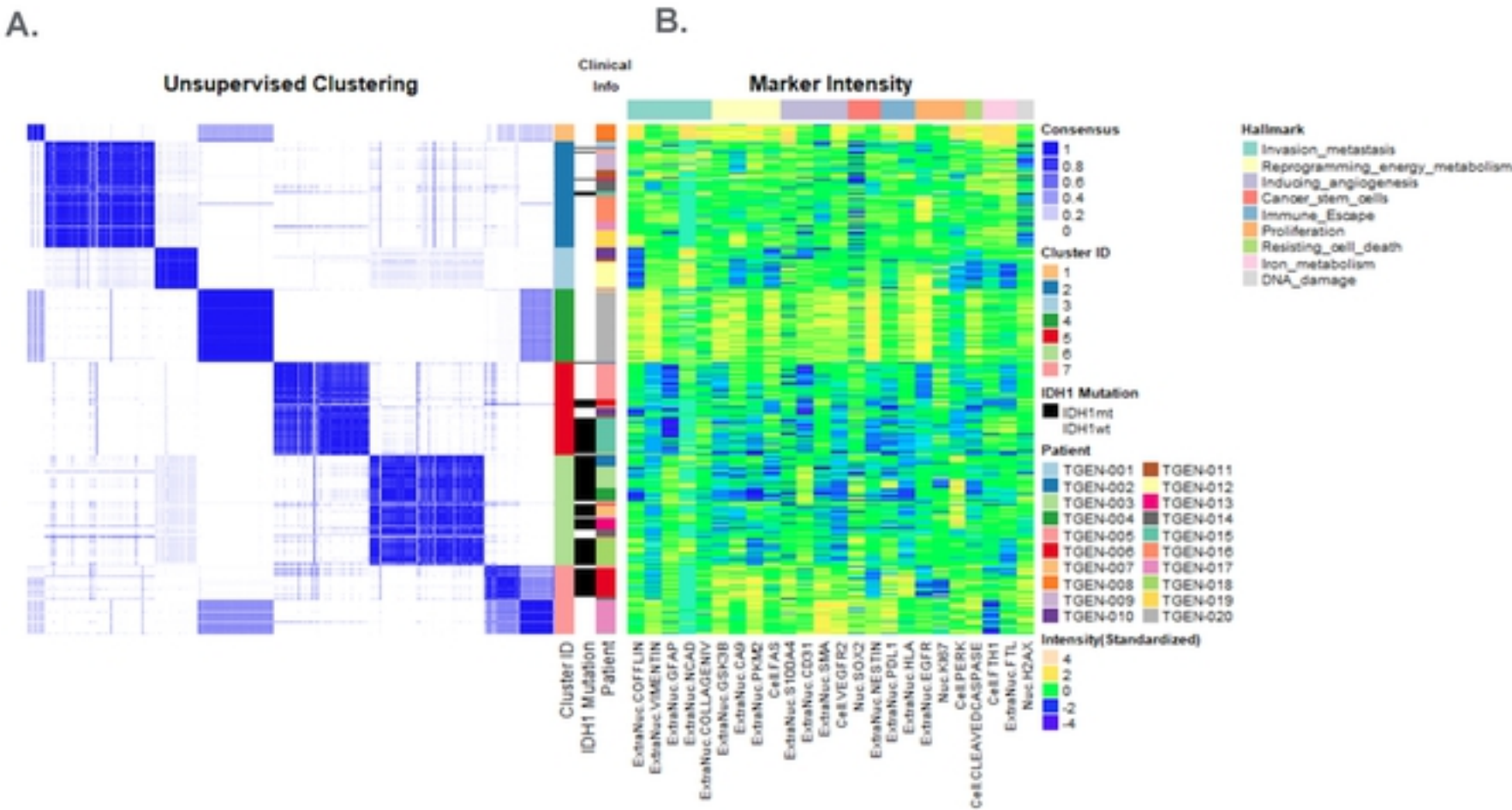


Figure 2

Figure 5: Correlation between IDH1 mutation status and cell phenotype at both gene and protein level.

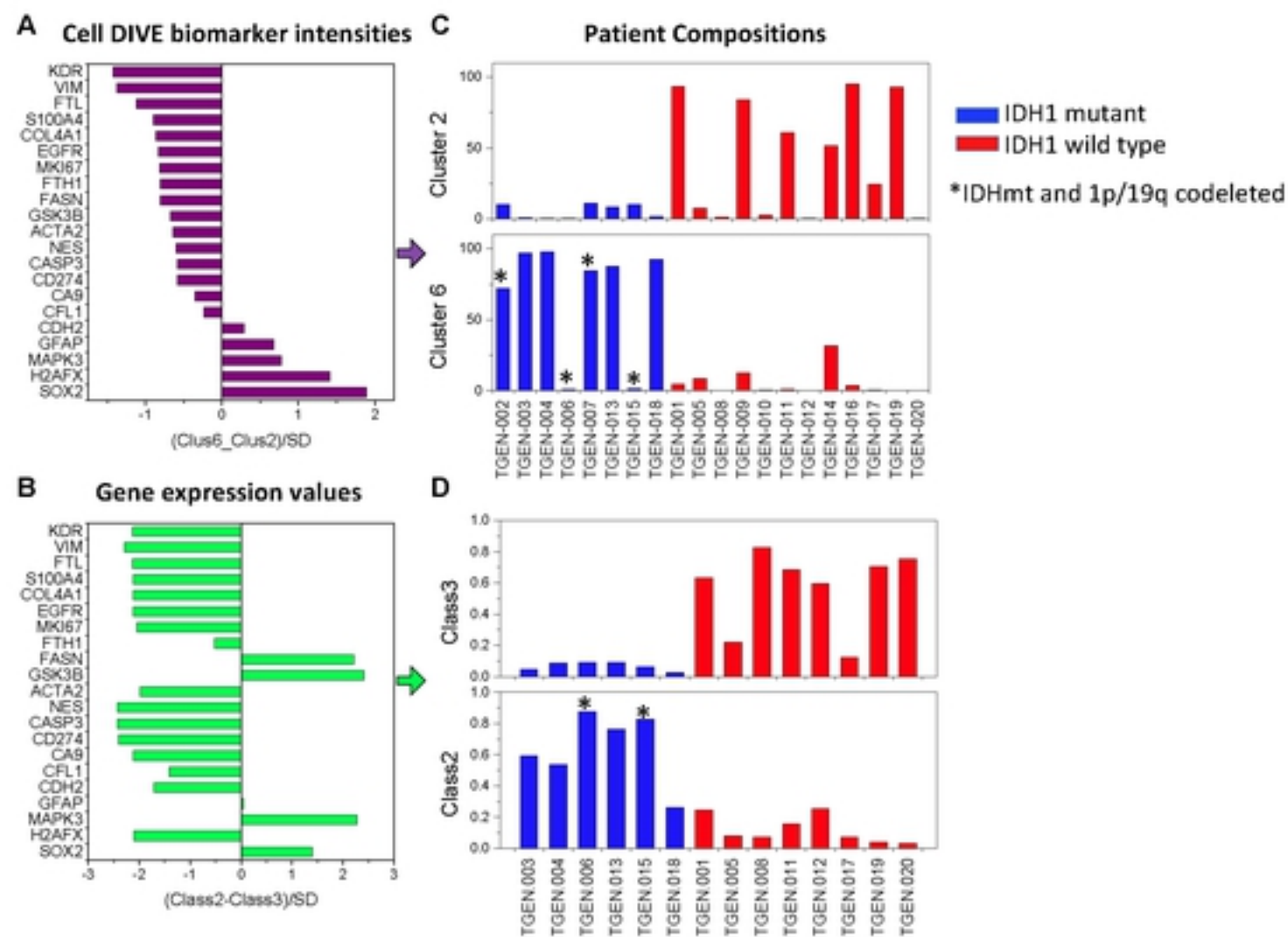


Figure 5

Figure 8: Multiscale modality integration including IDH1 mutation, clinical data, MRI, RNA and protein expression for the angiogenesis markers and corresponding cell clusters.

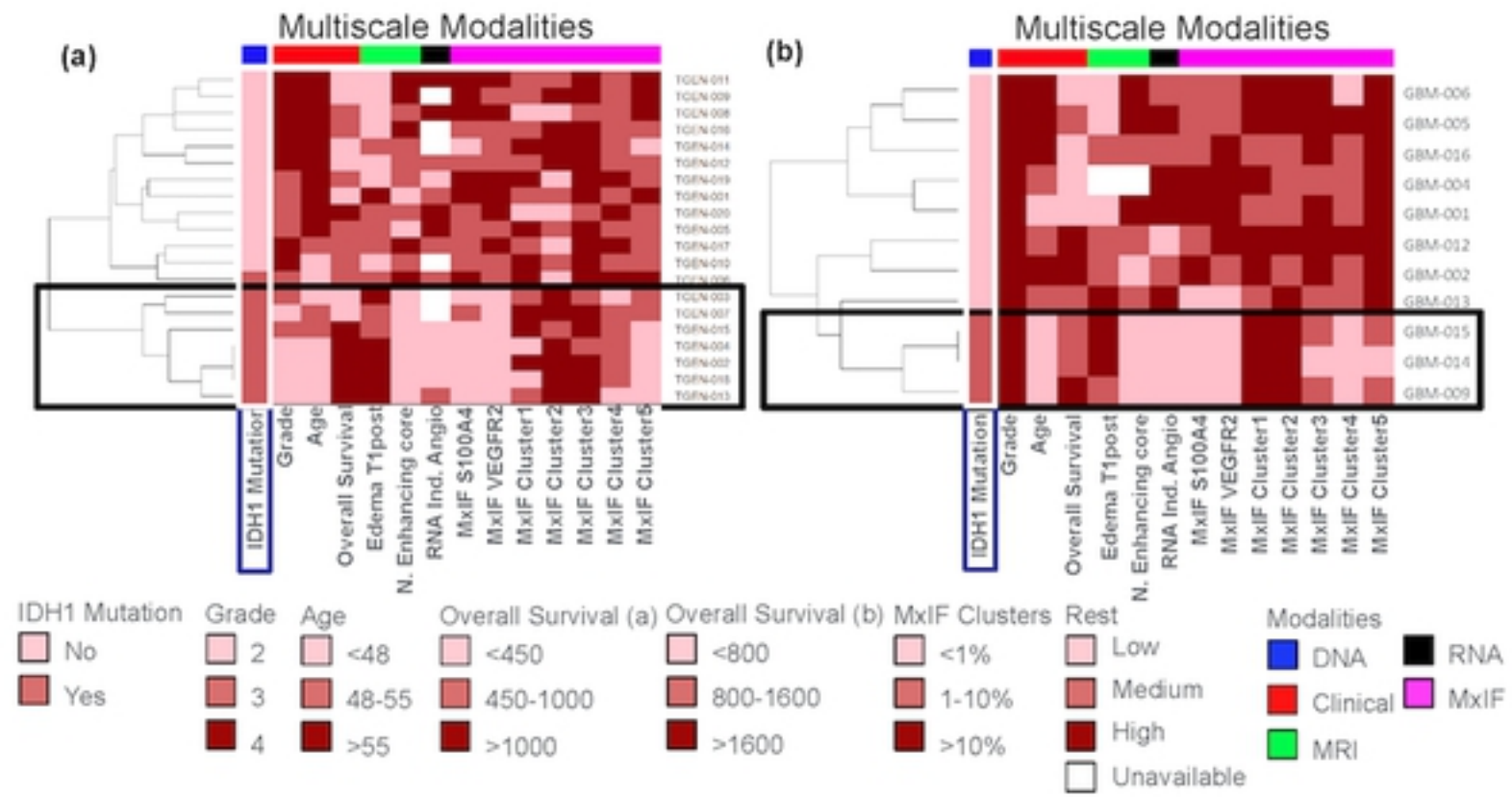


Figure 8

Figure 4: Cell cluster composition and oncoprint of treatment naive gliomas

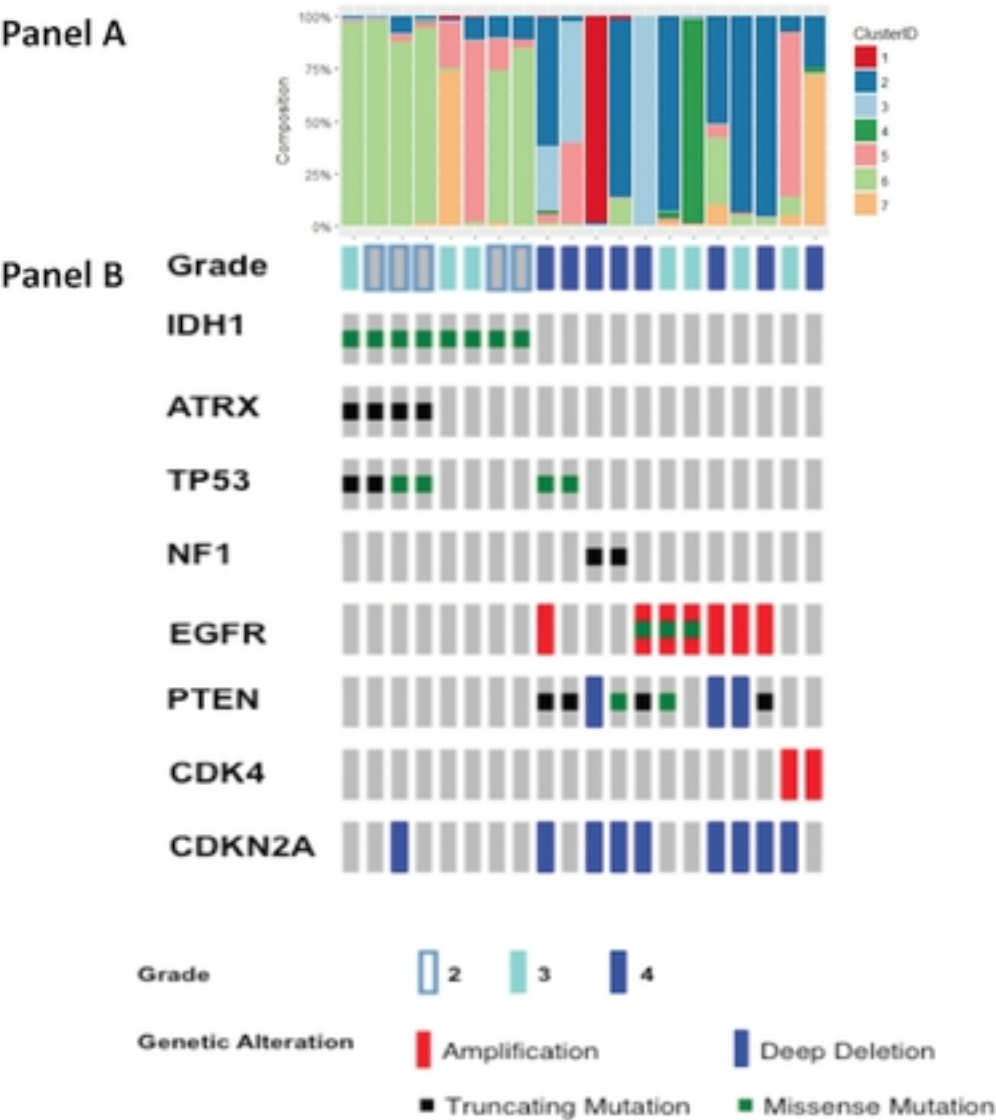


Figure 4