RNA-Sequencing and immunohistochemistry reveal ZFN7 as a stronger marker of survival than molecular subtypes in G-CIMP-negative glioblastoma

Anna Esteve-Codina¹, Francesc Alameda², Cristina Carrato³, Estela Pineda⁴, Oriol Arpí⁵, Maria Martinez-García⁶, Mar Mallo⁷, Marta Gut⁸, Marc Dabad¹, Avelina Tortosa⁹, Sonia Del Barco¹⁰, Jaume Capellades¹¹, Josep Puig¹², Oscar Gallego¹³, Teresa Pujol¹⁴, Laura Oleaga¹⁴, Miquel Gil-Gil¹⁵, Cristian de Quintana-Schmidt¹⁶, Izaskun Valduvico¹⁷, Anna Martinez-Cardús¹⁸, Beatriz Bellosillo², Ana Maria Muñoz-Marmol³, Anna Esteve¹⁸, Marta Domenech¹⁸, Angels Camins¹⁹, Jordi Craven-Bartle²⁰, Salvador Villa²¹, Jordi Marruecos²², Sira Domenech²³, Nuria de la Iglesia²⁴, Carmen Balana¹⁸

¹CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, 08028 Barcelona, Spain; ²Pathology Department, Neuropathology Unit, Hospital del Mar, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain; ³Pathology Department, Hospital Universitari Germans Trias i Pujol, Badalona, Spain; ⁴Medical Oncology, Hospital Clínic, Translational Genomics and Targeted Therapeutics in Solid Tumors, August Pi i Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain; ⁵Cancer Research Program, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain; ⁶Medical Oncology, Hospital del Mar, Barcelona, Spain; ⁷Institut de Recerca Contra la Leucèmia Josep Carreras, Badalona, Spain; ⁸CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Universitat Pompeu Fabra (UPF), Barcelona, Spain; ⁹Laboratori de Químio-resistència i Càncer, School of Medicine and Health Sciences, University of Barcelona, Department of Fundamental Care and Medical-Surgical Nursing, Bellvitge Biomedical Research Institute (IDIBELL), Bellvitge, Spain; ¹⁰Medical Oncology, Institut Català d'Oncologia (ICO) Girona, Hospital Josep Trueta, Girona, Spain; ¹¹Radiology Department, Hospital del Mar, Barcelona, Spain; ¹²Radiology Department, Institut de Diagnòstic per la Imatge, Hospital Josep Trueta, Girona, Spain; ¹³Medical Oncology, Hospital de Sant Pau, Barcelona, Spain; ¹⁴Radiology Department, Hospital Clínic, Barcelona, Spain; ¹⁵Neuro-Oncology Unit & Medical Oncology Department, Institut Català d'Oncologia (ICO), Institut de Investigació Bellvitge (IDIBELL), L'Hospitalet, Barcelona, Spain; ¹⁶Neurosurgery Department, Hospital de Sant Pau, Barcelona, Spain; ¹⁷Radiation Therapy Department, Hospital Clínic, Barcelona, Spain; ¹⁸ Institut Català d'Oncologia (ICO) Badalona, Badalona Applied Research Group in Oncology (B-ARGO Group), Institut Investigació Germans Trias i Pujol (IGTP), Badalona, Spain; ¹⁹Radiology Department, Institut de Diagnòstic per la Imatge, Hospital de Bellvitge, Bellvitge, Spain; ²⁰Radiation Therapy Department, Hospital de Sant Pau, Barcelona, Spain; ²¹Radiation Therapy Department, Institut Català d'Oncologia (ICO), Badalona, Spain; ²²Radiation Oncology Department, Institut Català d'Oncologia (ICO), Girona, Spain; ²³Radiology Department, Institut de Diagnòstic per la Imatge, Hospital Germans Trias i Pujol, Badalona, Spain; ²⁴Glioma and Neural Stem Cell Group, Translational Genomics and Targeted Therapeutics in Solid Tumors Team, August Pi i Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain

Running Title: RNA-Seq in newly diagnosed GBM

Correspondence to: Carmen Balaña Quintero
Research Coordinator
President of the Internal Scientific Committee of the IGTP
Neuro-oncology and Sarcomas
Institut Català d'Oncologia
Ctra Canyet s/n
08916 Badalona
Spain
Tel: +34 93 497 89 25
Fax: +34 93 497 89 50
Email: cbalana@iconcologia.net
ORCID: 0000-0003-0771-0390

Conflict of interest: The authors declare no conflict of interest for the data they are presenting.

Data availability: The datasets and full information on quality control assessments of sequenced samples generated during the current study are available from the corresponding author on reasonable request.

Molecular data underlying the findings described in the manuscript are fully available without restriction from the Bioproject Sequence Read Archive <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA613395>

Translational relevance: When classifying glioblastoma tumors into TCGA subtypes, there is high variability depending on the algorithm used. Our findings indicate a lack of prognostic impact for both the TCGA and IGS molecular subtypes compared to known prognostic factors, such as type of surgery, age or *MGMT* methylation status. In contrast, we found that the RNA expression levels of several individual genes were strong independent prognostic markers that were validated in external data bases (Gravendeel, TCGA, and/or Rembrandt) . Targeting a few of these genes with immunohistochemistry revealed that both the RNA and protein expression of zinc finger protein 7 had a remarkable independent effect on survival in G-CIMP-negative, a surrogate of wild-typeIDH glioblastomas.

ABSTRACT

Purpose: Glioblastoma is the most aggressive brain tumor in adults and has few therapeutic options. The study of molecular subtype classifications may lead to improved prognostic classification and identification of new therapeutic targets. The TCGA subtype classification has mainly been applied in USA clinical trials, while the IGS has mainly been applied in European trials.

Experimental Design: From paraffin-embedded tumor samples of 432 uniformly treated, newly diagnosed glioblastoma patients, we built tissue microarrays for immunohistochemical analysis and applied RNA-Sequencing to the best samples in order to classify them according to the TCGA and IGS subtypes.

Results: We obtained transcriptomic results from 124 patients. There was a lack of agreement among the three TCGA classificatory algorithms employed, which was not solely attributable to intratumoral heterogeneity. There was overlapping of the TCGA mesenchymal subtype with IGS cluster 23 and of the TCGA classical subtype with IGS cluster 18. Molecular subtypes were not associated with prognosis, but levels of expression of 13 novel genes were identified as independent prognostic markers in G-CIMP-negative patients, independently of clinical factors and *MGMT* methylation. These findings were validated in at least one external database. Three of the 13 genes were selected for immunohistochemical validation. In particular, high *ZNF7* RNA expression and low *ZNF7* protein expression were strongly associated with longer survival, independently of molecular subtypes.

Conclusions: The TCGA and IGS molecular classifications of glioblastoma have no higher prognostic value than individual genes and should be refined before being applied to clinical trials.

Key words: glioblastoma, prognosis, transcriptional subtypes, RNA-Seq, *ZNF7*.

INTRODUCTION

The standard therapy for glioblastoma is surgery followed by radiation therapy with concomitant and adjuvant temozolomide, but median survival remains around 14-16 months, depending on clinical factors, such as extent of surgery, patient age, and Karnofsky performance status (KPS), and on molecular factors, including *IDH* mutation status and *MGMT* promoter methylation. (1,2)

Several systems have been proposed for classifying glioblastoma tumors in molecular subtypes. (3-16) In 2006, Phillips et al (6) used DNA microarray to identify three glioma subtypes with different gene expression patterns, prognoses and biological processes: proneural, which was enriched with markers of neurogenesis; proliferative, with markers of cell proliferation; and mesenchymal, with markers of mesenchymal origin. Verhaak et al (7) used an unsupervised classification approach to data from The Cancer Genome Atlas (TCGA) and reported four subtypes: proneural, characterized by *IDH1* mutations and/or *PDGFRA* amplification; classical, with *EGFR* amplification and/or mutations; mesenchymal, with *NF1* loss and/or mutations; and neural, characterized by the expression of neuron markers. However, they could not confirm the proliferative subtype described by Phillips et al. (6,17) The proneural subtype was redefined by Nushmeh et al (18) with the identification of a glioma-CpG island methylator phenotype (G-CIMP). This phenotype was more prevalent among lower-grade gliomas, associated with *IDH* somatic mutations and younger age, and mostly present in the proneural subtype. A later report by Brennan et al (8) confirmed the survival advantage for the proneural subtype, which was mainly conferred by the G-CIMP-positive phenotype, identified *MGMT* methylation as a predictive factor for response to temozolomide, and described several novel gene mutations, some of which were targetable. (19) Wang et al (9) later explored changes in the microenvironment of different glioblastoma gene expression signatures before and after therapeutic intervention and concluded that the neural subtype described in previous reports (7,8) lacked characteristic gene abnormalities and was, in fact, a contamination of normal neural tissue. Therefore, the classification was refined into only three subtypes (proneural, classical, mesenchymal). In addition, using the single sample Gene Set Enrichment Analysis (ssGSEA) of 50 gene signatures, they found that 8% of cases were enriched in multiple ssGSEA signatures, suggesting that cases can activate more than one transcriptional signature. A survival analysis of the 74 patients (20%) who activated only one signature (highest simplicity scores >0.99), they found that mesenchymal patients had shorter survival than classical and proneural patients ($P = 0.03$). (9) Based on these studies, the TCGA-based subtypes have been used to explore sensitivity to treatment in several clinical trials. (20,21)

Other molecular classifications of glioblastoma subtypes have been developed and used in clinical trials, mainly in Europe. (22-24) Gravendeel et al (12) proposed the Intrinsic Glioma Subtypes (IGS), which segregated gliomas of all histological grades into seven different clusters. Glioblastoma tumors were distributed mainly in clusters 18, 22 and 23 but they could also be seen in other clusters, where prognosis was better, leading the authors to suggest that molecular alterations may drive prognosis better than morphology. In a later study comparing 55 paired fresh-frozen and formalin-fixed paraffin-embedded (FFPE) tumor samples, they confirmed the validity of FFPE samples for molecular profiling. (25)

Nevertheless, when evaluating the prognostic role of molecular alterations, we cannot ignore the impact of clinical variables. In fact, one of the drawbacks in studies of molecular profiling is precisely the partial omission of clinical variables from the analyses, which makes it difficult to separate the effect of such factors as extent of surgery and *MGMT* methylation from that of the molecular subtype.

We have explored the glioblastoma molecular subtypes and other molecular prognostic factors by analyzing FFPE samples from a uniformly treated population of newly diagnosed glioblastoma patients. Our objectives were to evaluate the impact of subtypes on overall survival and to identify other molecular alterations, such as individual genes, related to survival in our patient population.

METHODS

Patients and study design

This study was approved by the Institutional Review Board of the Hospital Germans Trias i Pujol (PI-14-016) and by the Ethics Committees of all the participating institutions and their Biobanks and was conducted in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. All patients or their representatives gave their written informed consent to participate in this study.

From 2004 to 2014, the GLIOCAT project (26) collected clinical data from 432 consecutive glioblastoma patients from six institutions, all of whom had received the standard first-line treatment (surgery followed by radiotherapy with concurrent and adjuvant temozolomide). The pathological diagnosis confirming glioblastoma was performed by pathologists before patients were included in the project. Once selected for inclusion, each case was anonymized and given a number to identify it across all data.

The following data were recorded: age, sex, symptoms, tumor characteristics, type of surgery, post-surgical performance and Mini Mental Status Examination (MMSE) score, details of radiotherapy and temozolomide treatments and treatment at relapse, date of progression, subsequent treatments, date and status at last control, and date and status of death or last control alive. Pre-and post- surgical resonance images were uploaded on a shared platform. Once patients were included in the present study, *MGMT* methylation status was determined if it had not previously been assessed.

DNA extraction and assessment of MGMT methylation

DNA was extracted from two 15- μ m sections of FFPE tissue using the QIAamp DNA Mini Kit (QIAGEN GmbH, Hilden, Germany), following the manufacturer's protocol. In cases with less than 50% of tumor cells, the tumor tissue was macrodissected manually. Then 500 ng of extracted DNA was subjected to bisulfite treatment using the EZ DNA Methylation-Gold Kit (Zymo Research Corporation, Irvine, CA). DNA methylation patterns in the CpG island of the *MGMT* gene were determined by methylation-specific PCR (MSP) using primers specific for either methylated or modified non-methylated DNA, as previously described. (27)

RNA-Sequencing (RNA-Seq)

RNA was extracted and assessed for quality. The highest-quality RNA samples were further anonymized and sent to the Centro Nacional de Análisis Genómico (CNAG-CRG, Barcelona, Spain) for analysis by RNA-Seq. Before performing RNA-Seq in the FFPE samples, we compared results in four cases of paired FFPE and fresh frozen (FF) samples and confirmed the similarity of results in the two types of tissue samples. (26)

RNA extraction and quality assessment

The RNA extraction of FFPE tumor specimens was performed on five 15 μ m-deep tissue cuts using the RNeasy FFPE Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. RNA quantity and purity were measured with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and the Qubit RNA HS Assay Kit (Invitrogen, Eugene, OR) used with the Qubit Fluorometer (Thermo Fisher Scientific). RNA was analyzed with the 2100 Bioanalyzer (Agilent, Waldbronn, Germany) and quality was assessed by Eukaryote Total RNA Nano Bioanalyzer assay

(Agilent) according to the “DV200” standard – the percentage of RNA fragments >200 nucleotides – as recommended by Illumina (San Diego, CA) for FFPE samples.

FFPE samples were sequenced in different batches and we first selected the highest-quality RNAs from patients with the most complete clinical and radiological data. However, as many samples did not have RNA with the minimal quality requirements, we subsequently included cases without available MRIs and, finally, those without informative results from the *MGMT* methylation analyses.

RNA library construction and sequencing

The starting input material for the construction of the RNA library was DNA-free total RNA from FFPE samples, obtained using the KAPA RNA HyperPrep Kit with RiboErase (HMR) (Roche Kapa Biosystems, Wilmington, MA), according to the manufacturer’s protocol. Briefly, the ribosomal RNA (rRNA) was depleted from 0.8-1µg of total RNA using RiboErase. rRNA-depleted RNA samples were fragmented in the presence of divalent ions at 65°C for 1 minute due to the low quality of the initial total RNA. Following RNA fragmentation (first and second strand synthesis), the Illumina-compatible bar-coded adapters were ligated at a concentration of 1.5 µM. Libraries were enriched with 15 cycles of PCR using KAPA HiFi HotStart ReadyMix (Kapa-Roche). The size and quality of the libraries were assessed in a High Sensitivity DNA Bioanalyzer assay (Agilent).

The libraries were sequenced on HiSeq2000 (Illumina) in paired-end mode with a read length of 2x76bp using TruSeq SBS Kit v4. Each sample was sequenced in a fraction of a sequencing v4 flow cell lane, following the manufacturer’s protocol. Sequencing was performed with a minimum of 65 M paired-end reads (2x75 bp) per RNA-seq sample. Image analysis, base calling and quality scoring of the run were processed using Real Time Analysis (RTA 1.18.66.3) software and followed by the generation of FASTQ sequence files by CASAVA (RRID:SCR_001802).

RNA-Seq processing and data analysis

RNA-Seq paired-end reads were mapped against the human reference genome (GRCh38) using STAR (RRID:SCR_015899, version 2.5.2a) (28) with ENCODE parameters for long RNA. Genes were quantified using RSEM (RRID:SCR_013027, version 1.2.28) (29) and GENCODE (RRID:SCR_014966, version 24). Mapping quality metrics were calculated with GEMtools (<http://gemtools.github.io/>) and RSeQC (RRID:SCR_005275). (30) Normalization and transformation of the expression data for plotting

and analysis were done with the variance stabilizing transformation function (vst) of DESeq2 (RRID:SCR_015687). (31) Principal component analysis (PCA) was performed with the R prcomp function and the R ggplot2 library (RRID:SCR_014601).

Classification of glioblastoma molecular subtypes

The TCGA classification of glioblastoma molecular subtypes (7-9) was performed with the GlioVis portal (32), which contains over 6500 tumor samples of approximately 50 expression datasets and facilitates classification with three algorithms: Support Vector Machine (SVM), K-nearest neighbor (KNN) and single sample Gene Set Enrichment Analysis (ssGSEA). G-CIMP status was predicted using the GlioVis SVM algorithm. The GlioVis glioblastoma TCGA cohort was selected as training dataset for both glioblastoma molecular subtype and G-CIMP predictions. The IGS classification of glioblastoma molecular subtypes was done with the R clusterRepro package (9,12), using the centroids for IGS0, IGS9, IGS16, IGS17, IGS18, IGS22 and IGS23 subtypes as described by Gravendeel. (12) (23,24)

Tissue microarray (TMA) construction and immunohistochemical analyses

TMAs were built for immunohistochemical analyses of all samples with sufficient tissue. Between two and four cores of each case were obtained, depending on the amount of tissue available. TMAs were constructed using a Veridiam Tissue Array Instrument (El Cajon, California, USA), model VTA-100, using a 1-mm diameter needle. Consecutive 4- μ m-thick sections were obtained and hematoxylin-eosin staining was done in sections 1, 20, and 40 in order to evaluate the persistence of the tumor at each spot. Samples of normal non-brain tissue were included in the TMAs as a guide for their reading and as positive/negative controls for immunohistochemical staining.

Benchmark XT (Ventana Medical System Inc., Tucson, AZ), Bond Max (Leica Microsystems GmbH, Wetzlar, Germany), and Dako (Dako Cytomation, Glostrup, Denmark) automated immunolabelling systems were used. Cores with less than 50% of tumor were considered non-evaluable, and samples with more than 50% of non-evaluable cores were considered non-informative. Several antibodies were studied. For IDH1-R132H, we used the antibody Dianova Cat# DIA-H09, RRID:AB_2335716. For H3.3 and H3.3 G34R, the antibodies Millipore Cat# ABE419, RRID:AB_2728728 and RevMAb Biosciences Cat# 31-1120-00, RRID:AB_2716433, were used respectively. In all cases results were scored as positive or negative expression.

Gene fusions in glioblastoma molecular subtypes

We also screened for fusions that could characterize each subtype. For fusion detection, RNA-Seq reads were mapped against the reference human genome with STAR-fusion version 0.7.0 with default options using GENCODE version 24. Previously described fusions were selected as final candidates to describe subtypes. (33)

Prognostic value of individual genes and related immunohistochemical protein expression

Finally, we looked for individual genes associated with overall survival. As *IDH* mutation status and G-CIMP status determine populations with different prognosis, we excluded G-CIMP-positive tumors from this analysis in order to assure an *IDH*-wild-type population. (18,34) Genes that were significant in our dataset ($P < 0.05$) and in at least one external glioblastoma gene expression dataset (Gravendeel, TCGA, and/or Rembrandt) were selected for a multivariate survival analysis including clinical variables. The model included *MGMT* methylation status, patient age, extent of surgery, and gene expression (one gene at a time). We identified those genes that were strongly independently related to survival in this G-CIMP-negative population ($HR < 0.7$ or > 1.5 , $P < 0.05$) and selected three for further validation by immunohistochemistry: *YWHAG* (ENSG00000170027.6); *UBXN7* (ENSG00000163960.11); and *ZNF7* (ENSG00000147789.15). A False Discovery Rate (FDR) approach was used to account for multiple testing.

YWHAG, located at chromosome 7q11.23, has only one transcript variant (ENST00000307630.4), which was the isoform detected with RNA-Seq. For immunohistochemical analysis, we selected the antibody Cat# HPA026918, RRID: AB_10610566, which had previously been validated with western blot and protein array (<https://www.proteinatlas.org/ENSG00000170027-YWHAG/antibody>) and matched 100% with the gene transcript.

UBXN7, located at chromosome 3q29, has six transcript variants, four of which are protein-coding. In our data, we observed the expression of only one protein-coding transcript (ENST00000296328.8 [UBXN7-201]). For immunohistochemistry, we used the antibody Anti-UBXN7 (ab185085), which targeted the sequence of the isoform even though it had not previously been validated with western blot, protein array, or orthogonal testing.

ZNF7, located at chromosome 8q24.3, has 18 transcript variants, ten of which are protein-coding. In our data, the most expressed variants were ZNF7 203, 205, 206, 207, 214, and 218, only four of which (203, 206, 207, 218) are protein-coding. For immunohistochemistry, we used the antibody Thermo Fisher Scientific Cat# PA5-38583, RRID:AB_2555182, which recognizes a motif in the N-terminal region of the protein (aminoacids 70-83) that is not present in variant ZNF7 218 (Supplementary Fig. S1).

Details on immunohistochemistry antibodies and scoring systems are summarized in Supplementary data file S1a.

Statistical analyses

Categorical variables were compared with the χ^2 or Fisher's exact test and continuous variables with the Kruskal-Wallis test. Overall survival was calculated from the date of surgery to the date of death or last control alive. Patients who were still alive at the date of last contact were censored. Median overall survival was calculated with the Kaplan-Meier method and compared using the log-rank test. The Cox proportional hazards model was used to calculate HRs with their 95% confidence intervals (CI). Several multivariate analyses of overall survival were performed, including different variables. Clinical data were merged with molecular and immunohistochemical results for statistical analyses. The Pearson's ρ coefficient was used to estimate the correlation between gene and protein expression. The optimal cut-off to predict the maximal difference in overall survival according to immunohistochemical results was calculated with the Maximally Selected Rank Statistics (Maxstat package of R, version 1.1.442). Analyses were performed with R software v3.4.2 and SPSS v24 (IBM; RRID:SCR_002865).

RESULTS

Tumor tissue samples were obtained from 329 of the 432 glioblastoma patients registered in the GLIOCAT project. (26) After multiple RNA extractions from each sample, 357 RNA libraries were prepared and RNA-Seq results were obtained for 151 tumor samples. Twenty-seven of these samples were not included in subsequent analyses: 18 because the RNA was degraded and/or contaminated and nine because they were obtained at second surgeries at recurrence. Quality control assessments and mapping statistics are summarized in Supplementary data file S2. In addition, 17 tissue microarrays (TMAs) were prepared and 280 samples were analyzed by immunohistochemistry.

RNA-Seq characterization of glioblastoma molecular subtypes

Based on the RNA-Seq results, the tumor samples were classified into the three TCGA molecular subtypes (9), using the three facilitated algorithms by GlioVis: SVM, KNN, and ssGSEA (Supplementary data file S1b). (32) There was agreement among all three algorithms on the subtype classification of only 82 samples, which were considered a separate subgroup in further analyses (Table 1).

There was no significant association between clinical characteristics and TCGA subtypes. (Supplementary Table S2). However, tumors classified as proneural by any of the algorithms were more likely to have *IDH1* mutations and to be G-CIMP-positive, while absence of *MGMT* methylation was more frequent in those classified as mesenchymal by SVM (Fig. 1 and Supplementary Table S1). There was no case with a histone H3.3 (K27 or G34 mutant) mutation in our series. There was also no significant association between clinical characteristics and IGS clusters. *IDH1*-mutated and G-CIMP-positive tumors were mostly classified as clusters 9 and 17 (Fig. 1 and Supplementary Table S2). Most tumors classified as IGS cluster 18 were the TCGA classical subtype and those classified as IGS cluster 23 were the TCGA mesenchymal subtype. (Fig. 1)

Prognostic factors

The log-rank test identified age ≤ 65 , greater extent of resection, *MGMT* methylation, G-CIMP-positive, and *IDH1* mutations – but not TCGA subtypes – as associated with longer overall survival (Supplementary Table S3). The multivariate analyses identified age, extent of resection, and *MGMT* methylation status as independent prognostic factors for overall survival. Additionally, the mesenchymal subtype identified by KNN – but not by other algorithms – was associated with shorter survival in the multivariate analysis (HR, 2.53; 95% CI, 1.05-6.12; $P = 0.038$). Patients with tumors classified as IGS 17 had longer survival, but these tumors were enriched with other favorable markers: 83.3% had *MGMT* methylation; 27.3% were *IDH1*-mutated; and 33.3% were G-CIMP-positive.

Gene fusions in TCGA subtypes

Among the 124 samples with informative RNA-Seq results, 25 (20.1%) had actionable gene fusions, which were mainly in the TCGA classical subtype and IGS cluster 18 (Fig 1).

Prognostic value of individual genes in G-CIMP-negative tumors

When we explored the prognostic value of individual genes from our RNA-Seq data, in order to ensure an *IDH*-wild-type subgroup, we eliminated G-CIMP-positive cases and examined gene expression in the

remaining 118 cases (Table 1). Through a univariate analysis and a comparative study with three external molecular data bases, we identified 139 genes associated with overall survival in both our univariate analysis and at least one other data base (Fig. 2). A multivariate model including these genes, along with age, *MGMT* methylation status, and extent of surgery, identified 47 genes as independent markers of overall survival (Fig. 1 and Supplementary Table S4). Five of these 47 genes were strong independent factors of longer survival ($HR < 0.7$; $P < 0.05$) and four of shorter survival ($HR > 1.5$; $P < 0.05$). An additional four genes were related less strongly with survival but were validated in two external data bases in G-CIMP-negative tumors (Table 2). We then narrowed our selection to three genes for further validation by immunohistochemistry and immune phenotyping: *YWHAG*, *UBXN7* and *ZNF7*. Higher median expression levels of *YWHAG* and *UBXN7* were observed in the proneural subtype, while there were no differences in *ZNF7* expression levels across the three subtypes (Fig. 3).

High RNA expression levels of *YWHAG* and *UBXN7* were markers of shorter overall survival (Table 2 and Supplementary Table S5). There was only a trend towards a correlation between *YWHAG* RNA and protein expression ($P = 0.067$), while there was no correlation between *UBXN7* RNA and protein expression ($P = 0.195$). There was no association between protein expression of either *YWHAG* or *UBXN7* and overall survival.

High *ZNF7* RNA expression was a marker of longer survival (Table 2, Fig. 4A). Median expression of *ZNF7* was 15.0 [percentile 25; 75%: 5.00;36.3], Supplementary Fig. S2 and Supplementary data file S1a. There was no correlation between RNA and protein expression ($P = 0.979$). According to a cut-off of 2.3, calculated by the Maximally selected rank statistics method, among the selected G-CIMP-negative population, median overall survival for patients with low *ZNF7* protein expression was 27.2 months (95% CI, 11.7-42.7), compared to 12.9 months (95% CI, 10.4-15.5) for those with high expression (HR , 0.44; 95% CI, 0.21-0.92; $P = 0.029$) (Fig. 4B). Low protein expression was also identified as an independent factor of longer overall survival when adjusted by age, extent of surgery, and *MGMT* methylation status (HR , 0.30; 95% CI, 0.129-0.723; $P = 0.007$) and maintained its prognostic value when we successively included the TCGA subtypes by each of the classification methods in the multivariate model. Among all patients with immunohistochemical results (Table 1), that included unselected G-CIMP or *IDH*-status patients, there was a trend towards association with survival in the univariate analysis: low *ZNF7* protein expression, 20.6 months (95% CI, 15.2-26.1); high *ZNF7* protein expression, 16.0 months (95% CI,

13.6-18.3). The multivariate analysis identified ZNF7 protein expression as a prognostic marker (HR, 0.62; 95% CI, 0.39-1.0; $P = 0.053$) (Supplementary data file S1a).

DISCUSSION

We have examined the prognostic role of the TCGA (9) and IGS (12) molecular subtypes in FFPE tumor samples from a uniformly treated cohort of newly diagnosed glioblastoma patients. We found that samples were classified differently depending on the GlioVis algorithm (32) used and that there was no clear association between molecular subtype and survival. Finally, we identified individual genes that may have a greater impact on survival than the molecular subtypes themselves.

Only 66.1% of cases were classified into the same subtype by all three GlioVis algorithms. (32) Using ssGSEA, which is based on 50 gene signatures strongly associated with the different subtypes, the TCGA network found a 77-93% concordance of unique calls to a specific subtype (7,9) and attributed discordant cases to intratumoral heterogeneity and the activation of more than one transcriptional signature.

However, this explanation cannot by itself account for our divergent findings because our raw data used for classification was the same for all three algorithms.

Nevertheless, we observed a high overlapping between the TCGA mesenchymal subtype and the IGS cluster 23 and between the TCGA classical subtype and the IGS cluster 18, which was especially evident for samples that called only one subtype. Both the TCGA and the IGS classifications have been used in clinical trials (21-23,35,36) to explore treatment susceptibilities. But the variability we have observed in our study would preclude their use, unless the methods used for classification were strengthened in order to find a reproducible and interchangeable classification. (37)

Regardless of the algorithm used for classifying the subtypes, the univariate and multivariate analyses did not identify an association between molecular TCGA subtypes and survival, nor among the glioblastoma IGS subtype independently of the individual status of *IDH1* or *MGMT*. The mesenchymal subtype as identified by KNN, that seemed to be more stringent than the other algorithms by classifying only 11.2% of samples as mesenchymal, was associated with shorted survival but only when the analysis was adjusted for other prognostic factors.

Independently of the molecular subtypes, we identified 13 genes whose expression positively or negatively influenced outcome, even when adjusted for established prognostic factors, in G-CIMP-

negative tumors as surrogates of *IDH*-wild-type glioblastomas. Moreover, the prognostic value of these genes was validated in the TCGA, Gravendeel, and/or Rembrandt in similar population data sets, thus reinforcing our results. Our findings are in line with other reports indicating that individual genes seem to predict outcome more accurately than either morphologic or genomic subtyping. (38-41) We selected three of these genes for validation by immunohistochemistry (YWHAG, UBXN7 and ZNF7). We were able to detect a prognostic role for ZNF7. Interestingly, *ZNF7* gene expression levels were independent of the molecular subtypes. Multivariate analyses showed that low protein expression was an independent marker of longer survival (HR, 0.30; $P = 0.007$) even when adjusted by the other known prognostic factors and the molecular subtypes, while conversely, high gene expression was a marker of longer survival (HR, 0.47; $P = 0.009$). The trend was maintained in the unselected cohort of patients with immunohistochemical but not RNA-Seq assessment..

In an attempt to account for this inverse relationship between gene and protein expression with survival, we first reviewed the lack of correlation between the antibody used for immunohistochemistry and gene expression. ZNF7 is a C2H2 zinc-finger protein that belongs to the large Krüppel-associated box (KRAB) -ZNF subfamily of ZNF transcription factors (<https://www.uniprot.org/uniprot/P17097>). They are characterized by having a KRAB repressor domain in the N-terminus and several C2H2 zinc-finger motifs that recognize specific DNA sequences and are thought to be involved in KRAB-mediated epigenetic silencing activities. Although there is little experimental data on the KRAB C2H2 family, some of the members play an important role in various processes of embryonic development, cell differentiation and proliferation, and regulation of the cell cycle and apoptosis (42). Among the highest expressed coding variants in our samples, only the isoform 207 has both the KRAB domain and the C2H2 zinc-finger motifs, while variants 203 and 206 have only the N-terminal KRAB transcriptional repressor domain (without DNA binding activity) and variant 218 has only the C2H2 zinc-finger motif. It would thus be expected that isoforms 203 and 206, lacking zinc-finger motifs, would not be able to bind DNA, although they could potentially bind co-repressor proteins, and that variant 218 would not recruit transcription repression complexes to ZNF7 DNA binding sites. Therefore, the lack of correlation between gene and protein expression could be in part explained by the inability of the antibody to detect the most expressed transcript variant, ZNF7 218.

Other factors must be considered, however. Discordances between RNA and protein expression are well documented . (43,44) In a comprehensive cross-tissue analysis of gene and protein expression in normal and cancer tissues, (45) there was a significant correlation between mRNA expression and protein abundance in only 6.1% of the studied genes. Intriguingly, higher levels of correlation were observed in normal tissues compared to tumor specimens. Current data demonstrate a substantial role for regulatory processes occurring after mRNA is made in this lack of concordance. For example, translation rates can be influenced by the mRNA sequence, thus causing a difference among ZNF7 transcript variants. Additionally, protein half-life can be modulated by several mechanisms, including post-transcriptional modifications, and in fact, several SUMOylation and ubiquitination sites are predicted in the ZNF7 reference sequence. It is unknown how differences in protein interactions and post-translational modifications can affect the half-life or subcellular localization of each protein variant. Given the complexity of ZNF7 alternative splicing and the great differences among ZNF7 protein isoforms, in-depth *in vivo* and *in vitro* studies and the use of isoform-specific antibodies would be necessary to clarify the contradictory effects on survival of mRNA and protein expression. Functional studies to explore the mechanisms of this lack of concordance are warranted but beyond the scope of the present study.

An important limitation of this study was the dropout of samples when they were processed for RNA-Seq. We obtained FFPE samples of 329 patients and 357 libraries were prepared in different batches and multiple extractions. However, we obtained good-quality results for only 124 cases, which highlights both the feasibility of using FFPE samples for RNA-Seq and also the difficulties involved in chemical modification, cross-linking, and degradation over time due to the fixation and archiving methods. (26) Nevertheless, this is the first study of RNA-Seq analysis of FFPE samples in a sizeable cohort of newly diagnosed, uniformly treated glioblastoma patients. Consequently, we were able to include clinical and molecular variables, such as MGMT methylation, in multivariate models. This allowed us to demonstrate a lack of prognostic impact of the molecular subtypes in the first-line setting and, conversely, the emergence of several genes independently related to survival in non-GCIMP glioblastoma patients. These findings, especially those on the ZNF7 gene and its protein product, warrant validation in experimental models and larger cohorts of patients. .

Acknowledgements:

Funding: This study was funded by the Fundació La Marató TV3 (665/C/2013)

(<http://www.ccma.cat/tv3/marato/projectes-financats/2012/231/>) PIs: Carmen Balana, Francesc Alameda and Nuria de la Iglesia. Anna Esteve-Codina is funded by ISCIII/MINECO (PT17/0009/0019) and co-funded by FEDER. Marta Domenech is funded by ISCIII/AES (CM19/00068). Carmen Balana is funded by ISCIII/AES (INT19/00032).

The authors thank the biobanks of the participating institutions: Fundació Institut Mar d'Investigacions Mèdiques, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP), Fundació Institut de Recerca de l'Hospital de la Santa Creu i St. Pau (IIB Sant Pau), and the Xarxa de Bancs de Tumors sponsored by Pla Director d'Oncologia de Catalunya (XBTC). We also thank Orieta Celiku from the National Cancer Institute for providing useful guidelines for TGCA glioblastoma data retrieval and Natalia Garcia-Balaña for her role as data manager and in coordinating the group. We are grateful for the collaboration of all the investigators of the Glioma Catalonia (GLIOCAT) Project.

Other investigators of the GLIOCAT Project:

Julie Marie Blanc¹, Montserrat Arumí², Raquel Lopez-Martos³, Silvia Menendez⁴, Ivan Archilla⁵, Teresa Ribalta⁵, Carlos Mesia⁶, Silvia Bagué⁷, Rafael Fuentes⁸, Noemí Vidal⁹, Iban Aldecoa^{5,10}

¹CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, 08028 Barcelona, Spain; ²Pathology Department, Neuropathology Unit, Hospital del Mar, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain; ³Pathology Department, Hospital Universitari Germans Trias i Pujol, Badalona, Spain; ⁴Cancer Research Program, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain; ⁵Pathology Department (Neuropathology), Hospital Clínic, Barcelona, Spain; ⁶Neuro-Oncology Unit & Medical Oncology Department, Institut Català d'Oncologia (ICO), Institut de Investigació Bellvitge (IDIBELL), L'Hospitalet, Barcelona, Spain; ⁷Pathology Department, Hospital de Sant Pau, Barcelona, Spain; ⁸Radiation Therapy Department, Institut Català d'Oncologia (ICO), Girona, Spain; ⁹Pathology Department, Hospital de Bellvitge, Bellvitge, Spain; ¹⁰Neurological Tissue Bank, Biobanc-Hospital Clínic-IDIBAPS, Barcelona, Spain

REFERENCES

1. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **2005**;352(10):987-96 doi 10.1056/NEJMoa043330.

2. Bleeker FE, Atai NA, Lamba S, Jonker A, Rijkeboer D, Bosch KS, *et al.* The prognostic IDH1(R132) mutation is associated with reduced NADP+-dependent IDH activity in glioblastoma. *Acta Neuropathol* **2010**;119(4):487-94 doi 10.1007/s00401-010-0645-6.
3. Li A, Walling J, Ahn S, Kotliarov Y, Su Q, Quezado M, *et al.* Unsupervised analysis of transcriptomic profiles reveals six glioma subtypes. *Cancer Res* **2009**;69(5):2091-9 doi 10.1158/0008-5472.CAN-08-2100.
4. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liao LM, *et al.* Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* **2004**;64(18):6503-10 doi 10.1158/0008-5472.CAN-04-0452.
5. Nigro JM, Misra A, Zhang L, Smirnov I, Colman H, Griffin C, *et al.* Integrated array-comparative genomic hybridization and expression array profiles identify clinically relevant molecular subtypes of glioblastoma. *Cancer Res* **2005**;65(5):1678-86 doi 10.1158/0008-5472.CAN-04-2921.
6. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer cell* **2006**;9(3):157-73.
7. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* **2010**;17(1):98-110 doi 10.1016/j.ccr.2009.12.020.
8. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, *et al.* The somatic genomic landscape of glioblastoma. *Cell* **2013**;155(2):462-77 doi 10.1016/j.cell.2013.09.034.
9. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer cell* **2017**;32(1):42-56 e6 doi 10.1016/j.ccell.2017.06.003.
10. Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, *et al.* Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* **2003**;63(7):1602-7.

11. Godard S, Getz G, Delorenzi M, Farmer P, Kobayashi H, Desbaillets I, *et al.* Classification of human astrocytic gliomas on the basis of gene expression: a correlated group of genes with angiogenic activity emerges as a strong predictor of subtypes. *Cancer Res* **2003**;63(20):6613-25.
12. Gravendeel LA, Kouwenhoven MC, Gevaert O, de Rooi JJ, Stubbs AP, Duijm JE, *et al.* Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res* **2009**;69(23):9065-72.
13. Karcher S, Steiner HH, Ahmadi R, Zoubaa S, Vasvari G, Bauer H, *et al.* Different angiogenic phenotypes in primary and secondary glioblastomas. *Int J Cancer* **2006**;118(9):2182-9.
14. Liang Y, Diehn M, Watson N, Bollen AW, Aldape KD, Nicholas MK, *et al.* Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci U S A* **2005**;102(16):5814-9 doi 10.1073/pnas.0402870102.
15. Rickman DS, Bobek MP, Misek DE, Kuick R, Blaivas M, Kurnit DM, *et al.* Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res* **2001**;61(18):6885-91.
16. Sallinen SL, Sallinen PK, Haapasalo HK, Helin HJ, Helen PT, Schraml P, *et al.* Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. *Cancer Res* **2000**;60(23):6617-22.
17. Huse JT, Phillips HS, Brennan CW. Molecular subclassification of diffuse gliomas: seeing order in the chaos. *Glia* **2011**;59(8):1190-9 doi 10.1002/glia.21165.
18. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer cell* **2010**;17(5):510-22 doi 10.1016/j.ccr.2010.03.017.
19. Schreck KC, Grossman SA, Pratilas CA. BRAF Mutations and the Utility of RAF and MEK Inhibitors in Primary Brain Tumors. *Cancers* **2019**;11(9) doi 10.3390/cancers11091262.
20. Cloughesy T, Finocchiaro G, Belda-Iniesta C, Recht L, Brandes AA, Pineda E, *et al.* Randomized, Double-Blind, Placebo-Controlled, Multicenter Phase II Study of Onartuzumab Plus Bevacizumab Versus Placebo Plus Bevacizumab in Patients With Recurrent Glioblastoma: Efficacy, Safety, and Hepatocyte Growth Factor and O-6-Methylguanine-DNA Methyltransferase Biomarker Analyses. *Journal of Clinical Oncology* **2017**;35(3):343-+ doi 10.1200/jco.2015.64.7685.

21. Sandmann T, Bourgon R, Garcia J, Li C, Cloughesy T, Chinot OL, *et al.* Patients With Proneural Glioblastoma May Derive Overall Survival Benefit From the Addition of Bevacizumab to First-Line Radiotherapy and Temozolomide: Retrospective Analysis of the AVAglio Trial. *J Clin Oncol* **2015**;33(25):2735-44 doi 10.1200/JCO.2015.61.5005.
22. Erdem-Eraslan L, Gravendeel LA, de Rooi J, Eilers PH, Idbaih A, Spliet WG, *et al.* Intrinsic molecular subtypes of glioma are prognostic and predict benefit from adjuvant procarbazine, lomustine, and vincristine chemotherapy in combination with other prognostic factors in anaplastic oligodendroglial brain tumors: a report from EORTC study 26951. *J Clin Oncol* **2013**;31(3):328-36 doi 10.1200/JCO.2012.44.1444.
23. Erdem-Eraslan L, van den Bent MJ, Hoogstrate Y, Naz-Khan H, Stubbs A, van der Spek P, *et al.* Identification of Patients with Recurrent Glioblastoma Who May Benefit from Combined Bevacizumab and CCNU Therapy: A Report from the BELOB Trial. *Cancer Res* **2016**;76(3):525-34 doi 10.1158/0008-5472.CAN-15-0776.
24. Gao Y, Weenink B, van den Bent MJ, Erdem-Eraslan L, Kros JM, Sillevs Smitt P, *et al.* Expression-based intrinsic glioma subtypes are prognostic in low-grade gliomas of the EORTC22033-26033 clinical trial. *Eur J Cancer* **2018**;94:168-78 doi 10.1016/j.ejca.2018.02.023.
25. Gravendeel LA, de Rooi JJ, Eilers PH, van den Bent MJ, Sillevs Smitt PA, French PJ. Gene expression profiles of gliomas in formalin-fixed paraffin-embedded material. *Br J Cancer* **2012**;106(3):538-45 doi 10.1038/bjc.2011.547.
26. Esteve-Codina A, Arpi O, Martinez-Garcia M, Pineda E, Mallo M, Gut M, *et al.* A Comparison of RNA-Seq Results from Paired Formalin-Fixed Paraffin-Embedded and Fresh-Frozen Glioblastoma Tissue Samples. *PloS one* **2017**;12(1):e0170632 doi 10.1371/journal.pone.0170632.
27. Esteller M, Hamilton SR, Burger PC, Baylin SB, Herman JG. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res* **1999**;59(4):793-7.
28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**;29(1):15-21 doi 10.1093/bioinformatics/bts635.
29. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**;12:323 doi 10.1186/1471-2105-12-323.

30. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **2012**;28(16):2184-5 doi 10.1093/bioinformatics/bts356.
31. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**;15(12):550 doi 10.1186/s13059-014-0550-8.
32. Bowman RL, Wang Q, Carro A, Verhaak RG, Squatrito M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. *Neuro-oncology* **2017**;19(1):139-41 doi 10.1093/neuonc/now247.
33. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. *Nature communications* **2014**;5:4846 doi 10.1038/ncomms5846.
34. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, *et al.* IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **2012**;483(7390):479-83 doi 10.1038/nature10866.
35. Cloughesy T, Finocchiaro G, Belda-Iniesta C, Recht L, Brandes AA, Pineda E, *et al.* Randomized, Double-Blind, Placebo-Controlled, Multicenter Phase II Study of Onartuzumab Plus Bevacizumab Versus Placebo Plus Bevacizumab in Patients With Recurrent Glioblastoma: Efficacy, Safety, and Hepatocyte Growth Factor and O(6)-Methylguanine-DNA Methyltransferase Biomarker Analyses. *J Clin Oncol* **2017**;35(3):343-51 doi 10.1200/JCO.2015.64.7685.
36. Gao Y, Weenink B, van den Bent MJ, Erdem-Eraslan L, Kros JM, Smitt PAES, *et al.* Expression-based intrinsic glioma subtypes are prognostic in low-grade gliomas of the EORTC22033-26033 clinical trial. *European Journal of Cancer* **2018**;94:168-78 doi 10.1016/j.ejca.2018.02.023.
37. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Statistical applications in genetics and molecular biology* **2007**;6:Article25 doi 10.2202/1544-6115.1309.
38. Louis DN, Wesseling P, Paulus W, Giannini C, Batchelor TT, Cairncross JG, *et al.* cIMPACT-NOW update 1: Not Otherwise Specified (NOS) and Not Elsewhere Classified (NEC). *Acta Neuropathol* **2018**;135(3):481-4 doi 10.1007/s00401-018-1808-0.
39. Louis DN, Giannini C, Capper D, Paulus W, Figarella-Branger D, Lopes MB, *et al.* cIMPACT-NOW update 2: diagnostic clarifications for diffuse midline glioma, H3 K27M-mutant and

- diffuse astrocytoma/anaplastic astrocytoma, IDH-mutant. *Acta Neuropathol* **2018**;135(4):639-42 doi 10.1007/s00401-018-1826-y.
40. Brat DJ, Aldape K, Colman H, Holland EC, Louis DN, Jenkins RB, *et al.* cIMPACT-NOW update 3: recommended diagnostic criteria for "Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV". *Acta Neuropathol* **2018**;136(5):805-10 doi 10.1007/s00401-018-1913-0.
 41. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **2016**;131(6):803-20 doi 10.1007/s00401-016-1545-1.
 42. Fedotova AA, Bonchuk AN, Mogila VA, Georgiev PG. C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. *Acta Naturae* **2017**;9(2):47-58.
 43. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews Genetics* **2012**;13(4):227-32 doi 10.1038/nrg3185.
 44. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **2016**;165(3):535-50 doi 10.1016/j.cell.2016.03.014.
 45. Kosti I, Jain N, Aran D, Butte AJ, Sirota M. Cross-tissue Analysis of Gene and Protein Expression in Normal and Cancer Tissues. *Sci Rep* **2016**;6:24799 doi 10.1038/srep24799.

Fig. 1 Heatmap of all samples showing data on type of surgery, MGMT status, presence of fusions, age group, molecular subtypes according to the three algorithms, and IGS subtypes, as well as the expression of the 47 genes that were independent factors influencing survival when adjusted by these prognostic factors in the non-GCIMP population and that could be validated in at least one external data base.

MET, methylated *MGMT*; UNMET, unmethylated *MGMT*

Fig. 2 Venn diagrams showing the number of genes associated with overall survival in our univariate analysis (GLIOCAT) with (A) $HR < 1$ and (B) $HR > 1$ and coincidences with other data bases

Fig. 3 RNA expression of (A) *YWHAG*, (B) *UBXN7* and (C) *ZNF7* in the classical, mesenchymal and proneural subtypes as classified by each of the three Gliovis algorithms (Support Vector Machine [SVM], K-nearest neighbor [KNN], and single sample Gene Set Enrichment Analysis [ssGSEA] in 118 G-CIMP-negative patients

Fig. 4 Median overall survival according to (A) *ZNF7* gene expression in 118 and (B) *ZNF7* protein expression in 92 G-CIMP-negative patients

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Table 1 Patient characteristics and molecular subtype

	All Patients with RNA-Seq Results <i>n</i> = 124 (%)	Non-GCIMP Patients with RNA- Seq Results <i>n</i> = 118 (%)	All Patients with IHC Results <i>n</i> = 280 (%)
Age, yrs			
median (range)	64.0 (55.0-71.0)	64.5 (56.0-71.0)	62.0 (53.0-69.0)
≤65 yrs	70 (56.5)	64 (54.2)	186 (66.4)
≤50 yrs	15 (12.1)	11 (9.32)	47 (16.8)
Sex			
Male	76 (61.3)	71 (60.2)	164 (58.6)
Female	48 (38.7)	47 (39.8)	116 (41.4)
Glioblastoma type			
Primary	120 (96.8)	116 (98.3)	273 (97.5)
Secondary	4 (3.23)	2 (1.69)	7 (2.50)
Extent of surgery^a			
Total resection	27 (23.9)	26 (24.3)	74 (28.1)
Partial resection	71 (62.8)	66 (61.7)	171 (65.0)
Biopsy	15 (13.3)	15 (14.0)	18 (6.8)
MGMT methylation status^{a,b}			
Methylated	64 (51.6)	60 (50.8)	132 (47.1)
Unmethylated	52 (41.9)	50 (42.4)	138 (49.3)
Unknown	8 (6.5)	8 (6.8)	10 (3.6)
KPS^c			
≤70%	31 (25.0)	30 (25.4)	58 (20.7)
>70%	64 (51.6)	61 (51.7)	169 (60.4)
Unknown	29 (23.4)	27 (22.9)	53 (18.9)
MMSE score			
<27	9 (7.3)	9 (7.8)	40 (14.3)
≥27	31 (25.0)	29 (24.5)	77 (27.5)
Unknown	84 (67.7)	80 (67.7)	163 (58.2)
IDH1 status^d			
Wild-type	107 (86.3)	106 (89.8)	264 (94.3)
Mutated	5 (4.0)	-	11 (3.9)
Unknown	12 (9.7)	12 (10.2)	5 (1.8)
G-CIMP+	6 (4.8)	0 (0.0)	6 (2.1)
Unknown	0 (0.0)	0 (0.0)	167 (59.6)
TCGA subtype by SVM^e			
Classical	46 (37.1)	45 (38.1)	40 (14.3)
Mesenchymal	40 (32.3)	40 (33.9)	36 (12.9)
Proneural	38 (30.6)	33 (28.0)	37 (13.2)
Unknown	0 (0.0)	0 (0.0)	167 (59.6)
TCGA subtype by KNN^e			
Classical	80 (64.5)	80 (67.8)	70 (25.0)

Mesenchymal	14 (11.3)	14 (11.9)	14 (5.0)
Proneural	30 (24.2)	24 (20.3)	29 (10.4)
Unknown	0 (0.0)	0 (0.0)	167 (59.6)
TCGA subtype by ssGSEA^e			
Classical	53 (42.7)	53 (44.9)	45 (16.1)
Mesenchymal	32 (25.8)	32 (27.1)	30 (10.7)
Proneural	39 (31.5)	33 (28.0)	38 (13.6)
Unknown	0 (0.0)	0 (0.0)	167 (59.6)
TCGA subtype by SVM+KNN+ssGSEA (n = 82)^e			
Classical	42 (51.2)	42 (54.5)	36 (12.9)
Mesenchymal	13 (15.9)	13 (16.9)	13 (4.6)
Proneural	27 (32.9)	22 (28.6)	26 (9.3)
Unknown	0 (0.0)	0 (0.0)	167 (59.6)
IGS classification			
0	6 (4.8)	6 (5.0)	5 (1.8)
9	14 (11.3)	12 (10.2)	14 (5.0)
16	2 (1.6)	2 (1.69)	2 (0.7)
17	12 (9.6)	8 (6.78)	11 (3.9)
18	68 (54.8)	68 (57.6)	59 (21.1)
22	4 (3.2)	4 (3.3)	4 (1.4)
23	18 (14.5)	18 (15.3)	18 (6.4)
Unknown	0 (0.0)	0 (0.0)	167 (59.6)

Table 1 Patient characteristics and molecular subtypes

IHC, immunohistochemistry; MMSE, Mini Mental Status Examination

^aPrognostic factor for overall survival in all patient subpopulations

^b*MGMT* methylation status was not available for all patients

^cPrognostic factor for overall survival in patients with IHC results

^dPrognostic factor for overall survival in patients with RNA-Seq or IHC results

^eNo prognostic value for overall survival

Supplementary Table S3 shows the details of the survival analysis.

Table 2 Genes identified as markers of overall survival in the multivariate analysis when adjusted for patient age, *MGMT* methylation status, and extent of surgery

Gene ^a	HR (95% CI)	<i>P</i>	External Data Base for Validation	HUGO Gene Nomenclature Committee Gene Name and ID
High gene expression associated with longer overall survival				
ZNF7 ENSG00000147789	0.47 (0.26-0.83)	0.009	Rembrandt	zinc finger protein 7 (13139)
UBN1 ENSG00000118900	0.53 (0.28-0.98)	0.042	Graveendel	ubinuclein 1 (12506)
LZTS2 ENSG00000123684	0.62 (0.41-0.95)	0.027	Rembrandt	leucine zipper tumor suppressor 2 (29381)
SLC25A25 ENSG00000148339	0.66 (0.46-0.95)	0.024	Graveendel	solute carrier family 25 member 25 (20663)
High gene expression associated with shorter overall survival				
LPGAT1 ENSG00000123684	1.8 (1.2-2.7)	0.007	Graveendel	lysophosphatidylglycerol acyltransferase 1 (28985)
YWHAG ENSG00000170027	1.8 (1.2-2.8)	0.010	Graveendel and TCGA	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma (12852)
DYNLL1 ENSG00000088986	1.8 (1.1-3)	0.021	Graveendel	dynein light chain LC8-type 1 (15476)
DNAJC10 ENSG00000077232	1.8 (1.1-2.9)	0.024	Rembrandt	DnaJ heat shock protein family (Hsp40) member C10 (24637)
UBXN7 ENSG00000163960	2 (1.2-3.5)	0.0092	Rembrandt	UBX domain protein 7 (29119)
High gene expression less strongly associated with shorter overall survival but validated in two external data bases				
SLC17A6 ENSG00000091664	1.1 (1-1.2)	0.012	Graveendel and TCGA	solute carrier family 17 member 6 (16703)
CNDP1 ENSG00000150656	1.1 (1-1.2)	0.031	Graveendel and TCGA	carnosine dipeptidase 1 (20675)
B3GALNT1 ENSG00000169255	1.5 (1.1-2.1)	0.02	Graveendel and Rembrandt	beta-1,3-N-acetylgalactosaminyltransferase 1 (globoside blood group) (918)
KIAA1715 (new name LNPk) ENSG00000144320	1.5 (1-2.1)	0.027	Graveendel and Rembrandt	lunapark, ER junction formation factor (21610)

^aSupplementary Table S5 shows the Gene Ontology terms for these genes.

Supplementary Fig. S1: Figure showing the most expressed variant transcripts. Variants 214 and 215 are non-coding. Variants 203, 206, and 207 - but not 218 – are targeted by the antibody used in the present study.

Supplementary Fig. S2: Representative images of four cases with different percentages of ZNF7 immunopositive tumor cells (a-d). Original magnification x 200

Gene	Antibody	Chromosome	SOURCE/RRID	Clone	Antigen Retrieval	Dilution	Cellular Location	Scoring System	Measurement	ALL (n=113)	G-CIMP negative*	Log-Rank(P)	COX Adj for age and MGMT (HR; 95%CI);P	Cox Adj for age, MGMT and TCGA subtypes**
UBX domain protein 7	UBXN7	3q27	Abcam Cat# 185085, RRID:NA)	polyclonal	Ptlink high	1/200	Nu	Percentage of positive tumoral cells	Median [p25-p75]	61.2 [33.3;76.9]				
									Cut-off for OS ≤76.7	72 [63.7%]	67/90	0.745	1.19; 0.69-2.05;0.523	Only MGMT and age
									Cut-off for OS >76.7	24 [21.2%]	23/90			
Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein eta	YWHAG	7q11.23	(Sigma-Aldrich Cat# HPA026918, RRID:AB_10610566)	polyclonal	PTLink Low	1/500	Nu	Percentage of positive tumoral cells	Median [p25-p75]	0.00 [0.00;0.33]				
									Cut-off for OS ≤0	71 [62.8%]	67/99	0.094	0.71; 0.44-1.15;0.169	Only MGMT and age
									Cut-off for OS >0	34 [30.1%]	32/99			
Zinc finger protein 7	ZNF7	8q24.3	(Thermo Fisher Scientific Cat# PA5-38583, RRID:AB_2555182)	polyclonal	PLink Low	ene-25	Cytoplasm	Percentage of positive tumoral cells	Median [p25-p75]	15.0 [5.00;36.3]				
									Cut-off for OS ≤2.3	17 [15.0%]	14/92	0.025	0.36;0.17-0.75;0.008	MGMT, age and ZNF7.
									Cut-off for OS >2.3	81 [71.7%]	78/92			

*Not all patients had evaluable results of all antibodies

** Subtypes as determined by each algorithm were included different COX analysis

BARCODE	CLASSIFICATION BY KNN				Subtype	CLASSIFICATION BY SVM				Subtype	CLASSIFICATION BY ssGSEA				IGS cluster
	Subtype	probability_CLA	S	probability_PN		probability_CLA	probability_MES	probability_PN	probability_CLA		probability_MES	probability_PN			
AC6230	Classical	0.8	0.2	0	Mesenchymal	0.454	0.523	0.023	Classical	0.0297	0.0594	0.97	18		
AC6223	Classical	0.6	0.27	0.13	Mesenchymal	0.27	0.521	0.209	Mesenchymal	0.208	0.198	0.921	18		
AC0377	Proneural	0.27	0	0.73	Proneural	0	0	1	Proneural	1	1	0.0099	9		
AC0378	Proneural	0.33	0.13	0.53	Proneural	0	0	1	Proneural	1	1	0.0099	9		
AA6392	Classical	0.73	0.07	0.2	Classical	0.998	0	0.002	Classical	0.0099	1	1	18		
AC6259	Classical	0.93	0.07	0	Classical	0.999	0.001	0	Classical	0.0099	1	1	18		
AC6289	Classical	0.47	0.27	0.27	Mesenchymal	0.023	0.554	0.424	Mesenchymal	1	0.0099	0.277	18		
AC0383	Classical	0.6	0.07	0.33	Proneural	0.159	0.02	0.821	Proneural	0.347	0.98	0.0099	0		
AA6390	Classical	0.67	0.07	0.27	Classical	0.994	0	0.006	Classical	0.0099	1	0.802	17		
AC0391	Mesenchymal	0.47	0.53	0	Mesenchymal	0.002	0.998	0	Mesenchymal	0.188	0.0099	1	23		
AC0453	Classical	0.53	0.33	0.13	Mesenchymal	0.006	0.993	0	Mesenchymal	0.96	0.0099	1	18		
AC0454	Proneural	0.33	0.07	0.6	Proneural	0.009	0.106	0.885	Proneural	1	0.97	0.0099	22		
AC0455	Proneural	0.2	0.33	0.47	Mesenchymal	0.024	0.942	0.034	Mesenchymal	0.614	0.0099	0.317	16		
AC0456	Classical	0.73	0.13	0.13	Classical	0.753	0	0.246	Classical	0.0099	0.97	0.436	18		
AC0396	Classical	0.73	0.27	0	Mesenchymal	0.003	0.996	0	Mesenchymal	0.426	0.0099	1	18		
AA6387	Classical	0.47	0.07	0.47	Proneural	0.016	0.007	0.977	Proneural	0.644	1	0.0099	17		
AC0459	Proneural	0.27	0	0.73	Proneural	0	0	0.999	Proneural	1	1	0.0099	9		
AA6393	Classical	0.69	0.19	0.12	Mesenchymal	0.055	0.919	0.026	Classical	0.0297	0.0297	1	18		
AC0461	Classical	0.6	0.07	0.33	Classical	0.897	0.002	0.101	Classical	0.0099	1	0.426	18		
AC6291	Classical	0.6	0	0.4	Classical	1	0	0	Classical	0.0099	1	0.97	17		
AC6287	Classical	0.67	0.2	0.13	Mesenchymal	0.256	0.735	0.009	Classical	0.0099	0.0594	0.98	18		
AC0401	Classical	0.93	0.07	0	Classical	1	0	0	Classical	0.0099	1	1	18		
AA6388	Classical	0.67	0.2	0.13	Mesenchymal	0.4	0.56	0.04	Classical	0.0396	0.426	0.861	18		
AC6286	Classical	0.87	0.07	0.07	Classical	0.996	0.001	0.003	Classical	0.0099	1	0.97	18		
AC0404	Proneural	0.27	0.07	0.67	Proneural	0	0	0.999	Proneural	1	1	0.0099	9		
AC6290	Classical	0.67	0.13	0.2	Classical	0.701	0.264	0.035	Classical	0.0099	0.515	1	18		
AC2091	Classical	0.53	0.07	0.4	Proneural	0.018	0	0.982	Proneural	0.0594	1	0.0099	9		
AC6300	Classical	0.93	0.07	0	Classical	1	0	0	Classical	0.0099	1	1	18		
AC6263	Proneural	0.27	0.2	0.53	Proneural	0.009	0.192	0.799	Proneural	0.941	0.782	0.0099	0		
AC6283	Classical	0.93	0.07	0	Classical	0.988	0.009	0.002	Classical	0.0099	0.941	1	18		
AC6285	Classical	0.8	0.13	0.07	Classical	0.917	0.004	0.079	Classical	0.0099	0.97	0.891	18		
AC6305	Classical	0.6	0.27	0.13	Mesenchymal	0.109	0.878	0.013	Mesenchymal	0.238	0.0396	1	18		
AC6302	Classical	0.81	0.06	0.12	Classical	1	0	0	Classical	0.0099	1	1	18		
AC6208	Mesenchymal	0.4	0.6	0	Mesenchymal	0	1	0	Mesenchymal	0.901	0.0099	1	23		
AC6229	Classical	0.73	0.07	0.2	Classical	0.69	0.073	0.237	Classical	0.0297	0.941	0.446	18		
AC6238	Classical	0.87	0.13	0	Classical	0.999	0.001	0	Classical	0.0099	0.941	0.98	18		
AC0442	Classical	0.6	0.13	0.27	Proneural	0.274	0.011	0.715	Classical	0.0396	1	0.158	18		
AC6222	Mesenchymal	0.27	0.73	0	Mesenchymal	0.006	0.994	0	Mesenchymal	0.822	0.0099	1	23		
AC6269	Classical	0.93	0.07	0	Classical	1	0	0	Classical	0.0099	1	1	18		
AA6367	Classical	0.8	0.2	0	Mesenchymal	0.23	0.768	0.002	Classical	0.0099	0.198	0.98	18		
AC0354	Classical	0.6	0.4	0.153	Mesenchymal	0.001	0.846	0.153	Mesenchymal	0.0198	0.0099	1	23		
AC6220	Mesenchymal	0.2	0.4	0.4	Mesenchymal	0.001	0.684	0.315	Proneural	1	0.0099	0.0099	18		
AC6236	Mesenchymal	0.2	0.67	0.13	Mesenchymal	0	1	0	Mesenchymal	1	0.0099	0.98	23		
AC6255	Classical	0.6	0.13	0.27	Mesenchymal	0.264	0.63	0.105	Mesenchymal	0.703	0.0198	0.861	18		
AC6244	Classical	0.88	0.06	0.06	Classical	1	0	0	Classical	0.0099	1	1	18		
AC6278	Mesenchymal	0.33	0.47	0.2	Mesenchymal	0.002	0.997	0.001	Mesenchymal	0.941	0.0099	0.99	23		
AA6366	Classical	0.6	0.07	0.33	Proneural	0.072	0.216	0.712	Mesenchymal	0.96	0.485	0.564	18		
AC6218	Classical	0.88	0.06	0.06	Classical	0.999	0	0.001	Classical	0.0099	1	1	18		
AC6280	Classical	0.73	0.07	0.2	Classical	0.814	0.008	0.178	Classical	0.158	0.842	0.97	18		
AC6243	Classical	0.6	0.07	0.33	Classical	0.981	0.001	0.018	Classical	0.0099	1	0.455	18		
AC6279	Classical	0.73	0.07	0.2	Classical	0.995	0	0.005	Classical	0.0099	1	0.97	18		
AC0301	Classical	0.67	0.27	0.07	Classical	0.961	0.039	1	Classical	0.0099	0.0594	1	23		
AC6256	Mesenchymal	0.27	0.6	0.13	Mesenchymal	0.008	0.988	0.004	Mesenchymal	0.317	0.0099	0.97	18		
AC6211	Proneural	0.2	0	0	Proneural	0.281	0.002	0.717	Proneural	0.604	1	0.0099	0		
AC6239	Proneural	0.33	0	0.67	Proneural	0.004	0.006	0.99	Proneural	0.98	0.97	0.0099	9		
AC6246	Proneural	0.2	0	0.8	Proneural	0	0	1	Proneural	0.99	1	0.0099	0		
AC6253	Mesenchymal	0.33	0.6	0.07	Mesenchymal	0	1	0	Mesenchymal	0.921	0.0099	1	23		
AC6237	Classical	0.87	0.13	0	Classical	0.811	0.188	0.001	Classical	0.0297	0.614	1	18		
AC6251	Proneural	0.33	0.07	0.6	Proneural	0.014	0.003	0.983	Proneural	0.366	1	0.0099	0		
AC6241	Classical	0.8	0.2	0	Classical	0.86	0.139	0.001	Classical	0.0099	0.0099	1	18		
AC0443	Proneural	0.27	0.13	0.6	Proneural	0.001	0.02	0.978	Proneural	0.98	0.713	0.0099	22		
AC6293	Mesenchymal	0.33	0.53	0.13	Mesenchymal	0.001	0.976	0.023	Mesenchymal	0.95	0.0099	0.158	23		
AC0338	Classical	0.8	0.07	0.13	Classical	0.941	0.046	0.013	Classical	0.0099	1	0.97	18		
AC6296	Classical	0.67	0.07	0.27	Mesenchymal	0.014	0.954	0.032	Mesenchymal	0.683	0.0594	0.317	18		
AA6380	Proneural	0.47	0.07	0.47	Proneural	0.021	0.008	0.971	Proneural	0.703	1	0.0099	22		
AC6250	Mesenchymal	0.33	0.47	0.2	Mesenchymal	0.001	0.995	0.004	Mesenchymal	0.416	0.0099	0.97	23		
AC6258	Classical	0.88	0.12	0	Classical	0.999	0.001	0	Classical	0.0099	0.891	1	18		
AC6270	Classical	0.93	0.07	0	Classical	0.999	0.001	0	Classical	0.0099	0.98	1	18		
AC0340	Proneural	0.33	0	0.67	Proneural	0.002	0	0.998	Proneural	1	1	0.0099	9		
AC6226	Proneural	0.27	0.07	0.67	Classical	0.563	0.004	0.434	Proneural	0.604	1	0.178	17		
AC6265	Classical	0.53	0.33	0.13	Mesenchymal	0.021	0.974	0.005	Classical	0.0297	0.0297	0.98	18		
AC0341	Classical	0.69	0.31	0	Classical	0.641	0.347	0.012	Mesenchymal	0.337	0.0099	1	18		
AC2092	Classical	0.67	0.33	0	Mesenchymal	0.002	0.997	0.001	Mesenchymal	0.644	0.0099	1	18		
AA6376	Classical	0.73	0.27	0	Classical	0.716	0.27	0.014	Mesenchymal	0.396	0.0297	1	18		
AC0360	Classical	0.47	0.2	0.33	Classical	0.181	0.458	0.361	Classical	0.188	0.535	0.446	18		
AA6384	Mesenchymal	0.27	0.6	0.13	Mesenchymal	0	0.998	0.001	Mesenchymal	0.822	0.0099	0.97	23		
AC6276	Classical	0.06	0	0	Classical	0	1	0	Classical	0.0099	1	1	18		
AC6207	Classical	0.53	0.47	0	Mesenchymal	0	1	0	Mesenchymal	0.851	0.0099	1	23		
AC6214	Proneural	0.33	0	0.67	Proneural	0.004	0.005	0.991	Proneural	0.96	1	0.0099	9		
AC6216	Proneural	0.2	0.07	0.73	Proneural	0.007	0.108	0.885	Proneural	1	0.98	0.129	17		
AC6224	Classical	0.53	0.47	0	Mesenchymal	0.002	0.998	0	Mesenchymal	0.465	0.0099	1	23		
AC0344	Classical	0.93	0.07	0	Classical	1	0	0	Classical	0.0099	1	1	18		
AC0359	Proneural	0.2	0.07	0.73	Proneural	0.003	0.002	0.995	Proneural	1	1	0.0099	17		
AC0369	Classical	0.6	0.27	0.13	Mesenchymal	0.02	0.98	0	Mesenchymal	0.317	0.0099	1	23		
AC6272	Classical	0.67	0.13	0.2	Proneural	0.328	0.18	0.492	Proneural	0.891	0.98	0.495	18		
AC0346	Proneural	0.07	0	0.93	Proneural	0	0	1	Proneural	1	1	0.0099	9		
AC6264	Classical	0.69	0.19	0.12	Mesenchymal	0.034	0.913	0.023	Mesenchymal	0.297	0.0594	0.446	18		
AA6369	Classical	0.6	0.07	0.33	Classical	0.765	0.012	0.223	Classical	0.0297	1	0.0495	18		
AC2101	Classical	0.87	0.07	0.07	Classical	0.996	0.001	0.003	Classical	0.0099	1	0.97	18		
AA6372	Classical	0.53	0.07	0.4	Proneural	0.092	0.153	0.756	Proneural	0.812	1	0.525	17		
AC0372	Proneural	0.47	0	0.53	Proneural	0.251	0.036	0.713	Proneural	0.0297	1	0.0099	0		
AA6370	Mesenchymal	0.33	0.67	0	Mesenchymal	0	1	0	Mesenchymal	0.822	0.0099	1	23		
AC0449	Proneural	0.13	0.07	0.8	Proneural	0	0	0.999	Proneural	1	1	0.0099	9		
AC6274	Classical	0.47	0.33	0.2	Mesenchymal	0.099	0.9	0.001	Mesenchymal	0.802	0.0099				

Table S1. Association between clinical characteristics and the three TCGA molecular subtypes (as classified by the GlioVis algorithms KNN, SVM, and ssGSEA).

			KNN				SVM				ssGSEA			
		ALL (N=124)	CLA (N=80)	MES (N=14)	PN (N=30)	<i>P</i>	CLA (N=46)	MES (N=40)	PN (N=38)	<i>P</i>	CLA (N=53)	MES (N=32)	PN (N=39)	<i>P</i>
AGE	Median (range)	64.0 (55.0-71.0)	64.0 (54.8-69.2)	62.5 (56.2-72.0)	65.5 (51.8-72.8)	0.759	65.0 (57.0-71.0)	60.5 (54.0-69.0)	64.5 (55.2-70.8)	0.469	64.0 (55.0-71.0)	63.5 (55.8-68.5)	64.0 (54.0-70.5)	0.912
	≤65 y old	70 (56.5%)	48 (60.0%)	7 (50.0%)	15 (50.0%)	0.561	24 (52.2%)	25 (62.5%)	21 (55.3%)	0.619	29 (54.7 %)	19 (59.4%)	22 (56.4%)	0.916
	≤50 y old	15 (12.1%)	10 (12.5%)	0	5 (16.7%)	0.286	5 (10.9%)	5 (12.5%)	5 (13.2%)	0.943	6 (11.3%)	3 (9.3%)	6 (15.4%)	0.78
SEX	Male	76 (61.3%)	49 (61.3%)	7 (50.0%)	20 (66.7%)	0.572	26 (56.5%)	26 (65.0%)	24 (63.2%)	0.695	29 (54.7%)	21 (65.6%)	26 (66.7%)	0.429
	Female	48 (38.7%)	31 (38.8%)	7 (50.0%)	10 (33.3%)		20 (43.5%)	14 (35.0%)	14 (36.8%)		24 (45.3%)	11 (34.4%)	13 (33.3%)	

TUMOR SIZE	≤50mm	39 (52.7%)	27 (54.0%)	1 (33.3%)	11 (52.4%)	0.918	14 (50.0%)	12 (63.2%)	13 (48.1%)	0.566	18 (54.5%)	7 (53.8%)	14 (50.0%)	0.935
	>50mm	35 (47.3%)	23 (46.0%)	2 (66.7%)	10 (47.6%)		14 (50.0%)	7 (36.8%)	14 (51.9%)		15 (45.5%)	6 (46.2%)	14 (50.0%)	
SURGERY	Complete resection	27 (23.9%)	20 (27.4%)	1 (9.0%)	6 (20.7%)	0.627	7 (17.5%)	11 (29.7%)	9 (25.0%)	0.299	8 (17.0%)	11 (37.9%)	8 (21.6%)	0.124
	Partial resection	71 (62.8%)	42 (57.5%)	9 (81.8%)	20 (69.0%)		24 (60.0%)	23 (62.2%)	24 (66.7%)		29 (61.7%)	16 (55.2%)	26 (70.3%)	
	Biopsy	15 (13.3%)	11 (15.1%)	1 (9.0%)	3 (10.3%)		9 (22.5%)	3 (8.11%)	3 (8.33%)		10 (21.3%)	2 (6.9%)	3 (8.1%)	
MMSE	Unknown	84 (67.7%)	55 (68.8%)	13 (92.9%)	16 (53.3%)	0.12	32 (69.6%)	30 (75.0%)	22 (57.9%)	0.386	35 (66.0%)	26 (81.2%)	23 (59.0%)	0.191
	<27	9 (7.26%)	5 (6.25%)	0	4 (13.3%)		3 (6.52%)	1 (2.50%)	5 (13.2%)		4 (7.5%)	0	5 (12.8%)	
	≥27	31 (25.0%)	20 (25.0%)	1 (7.1%)	10 (33.3%)		11 (23.9%)	9 (22.5%)	11 (28.9%)		14 (26.4%)	6 (18.8%)	11 (28.2%)	
KPS	Unknown	29 (23.4%)	20 (25.0%)	3 (21.4%)	6 (20.0%)	0.516	10 (21.7%)	10 (25.0%)	9 (23.7%)	0.955	12 (22.6%)	8 (25.0%)	9 (23.1%)	0.705

	≤70	31 (25.0%)	19 (23.8%)	6 (42.9%)	6 (20.0%)		13 (28.3%)	10 (25.0%)	8 (21.1%)		14 (26.4%)	10 (31.2%)	7 (18.9%)	
	>70	64 (51.6%)	41 (51.2%)	5 (35.7%)	18 (60.0%)		23 (50.0%)	20 (50.0%)	21 (55.3%)		27 (50.9%)	14 (43.8%)	23 (59.0%)	
MGMT methylation	Met	64 (51.6%)	41 (51.2%)	6 (42.9%)	17 (56.7%)	0.598	30 (65.2%)	14 (35.0%)	20 (52.6%)	0.014	31 (58.5%)	13 (40.6%)	20 (51.3%)	0.329
	UnMet	52 (41.9%)	34 (42.5%)	8 (57.1%)	10 (33.3%)		13 (28.3%)	25 (62.5%)	14 (36.8%)		19 (35.8%)	18 (56.2%)	15 (38.5%)	
IDH1	Mutated	5 (4.59%)	0	0	5 (17.2%)	0.003	1 (2.56%)	0	4 (11.4%)	0.077	0	0	5 (13.9%)	0.004
	Wild-type	104 (95.4%)	67 (100%)	13 (100%)	24 (82.8%)		38 (97.4%)	35 (100%)	31 (88.6%)		44 (100%)	29 (100%)	31 (86.1%)	
G-CIMP	Positive	6 (4.8%)	0	0	6 (20%)	0	1 (2.2%)	0	5 (13.2%)	0.012	0	0	6 (15.4%)	0.001

Table S2. Association between clinical characteristics and the IGS clusters.

		ALL N=124	IGS0 (N=6)	IGS16 (N=2)	IGS17 (N=12)	IGS18 (N=68)	IGS22 (N=4)	IGS23 (N=18)	IGS9 (N=14)	<i>P</i>
AGE	Median (range)	64.0 (55.0-71.0)	67.0 (65.0-69.0)	71.0 (70.0-72.0)	56.5 (50.8-69.5)	64.0 (54.8-70.2)	62.0 (58.5-63.2)	58.0 (55.2-66.0)	66.5 (53.8-74.2)	0.48
	≤65 y old	70 (56.5%)	0	0	3 (25.0%)	9 (13.2%)	0	1 (5.56%)	2 (14.3%)	0.737
	≤50 y old	15 (12.1%)	3 (50.0%)	0	7 (58.3%)	38 (55.9%)	4 (100%)	12 (66.7%)	6 (42.9%)	0.305
SEX	Male	76 (61.3%)	4 (66.7%)	0	7 (58.3%)	41 (60.3%)	3 (75.0%)	10 (55.6%)	11 (78.6%)	0.53
	Female	48 (38.7%)	2 (33.3%)	2 (100%)	5 (41.7%)	27 (39.7%)	1 (25.0%)	8 (44.4%)	3 (21.4%)	
TUMOR SIZE	≤50mm	39 (52.7%)	2 (66.7%)	1 (50.0%)	4 (44.4%)	23 (54.8%)	2 (50.0%)	1 (25.0%)	6 (60.0%)	0.939
	>50mm	35 (47.3%)	1 (33.3%)	1 (50.0%)	5 (55.6%)	19 (45.2%)	2 (50.0%)	3 (75.0%)	4 (40.0%)	

SURGERY	Complete resection	27 (23.9%)	2 (40.0%)	1 (50.0%)	2 (16.7%)	16 (26.2%)	1 (25.0%)	3 (20.0%)	2 (14.3%)	0.956
	Partial resection	71 (62.8%)	3 (60.0%)	1 (50.0%)	8 (66.7%)	36 (59.0%)	2 (50.0%)	11 (73.3%)	10 (71.4%)	
	Biopsy	15 (13.3%)	0	0	2 (16.7%)	9 (14.8%)	1 (25.0%)	1 (6.67%)	2 (14.3%)	
MMSE	Unknown	84 (67.7%)	6 (100%)	1 (50.0%)	4 (33.3%)	47 (69.1%)	1 (25.0%)	17 (94.4%)	8 (57.1%)	0.023
	<27	9 (7.26%)	0	0	3 (25.0%)	4 (5.88%)	0	0	2 (14.3%)	
	≥27	31 (25.0%)	0	1 (50.0%)	5 (41.7%)	17 (25.0%)	3 (75.0%)	1 (5.56%)	4 (28.6%)	
KPS	Unknown	29 (23.4%)	1 (16.7%)	0	3 (25.0%)	17 (25.0%)	0	5 (27.8%)	3 (21.4%)	0.973
	≤70	31 (25.0%)	1 (16.7%)	1 (50.0%)	4 (33.3%)	15 (22.1%)	1 (25.0%)	6 (33.3%)	3 (21.4%)	
	>70	64 (51.6%)	4 (66.7%)	1 (50.0%)	5 (41.7%)	36 (52.9%)	3 (75.0%)	7 (38.9%)	8 (57.1%)	

MGMT methylation	Met	64 (51.6%)	2 (33.3%)	0	10 (83.3%)	35 (51.5%)	1 (25.0%)	7 (38.9%)	9 (64.3%)	0.05
	UnMet	52 (41.9%)	3 (50.0%)	1 (50.0%)	2 (16.7%)	29 (42.6%)	3 (75.0%)	11 (61.1%)	3 (21.4%)	
IDH1	Mutated	5 (4.59%)	0	0	3 (27.3%)	0	0	0	2 (14.3%)	0.006
	Wild-type	104 (95.4%)	5 (100%)	1 (100%)	8 (72.7%)	57 (100%)	4 (100%)	17 (100%)	12 (85.7%)	
G-CIMP	Positive	6 (4.8%)	0	0	4 (33.3%)	0	0	0	2 (14.3%)	

Table S3. Median overall survival (Kaplan-Meier estimate) according to well established prognostic factors and molecular subtypes.

	Median Overall Survival	<i>P</i> ¹
	Months (95% CI)	
MGMT		0.01
Methylated	18.8 (10.7-26.9)	
Unmethylated	14.7 (10.7-28.7)	
Age		<0.001
≤65	19.9 (15.7-24.2)	
>65	10.6 (8.1-13.1)	
Extent of surgery		<0.001
Total resection	19.2 (9.8-28.6)	
Partial resection	16.5 (11.8-21.3)	
Biopsy	6.2 (4.5-7.9)	
G-CIMP		0.003
Positive	30.5 (--)	
Negative	14.5 (11.6-17.3)	
IDH1 mutation status		0.018
Wild-type	30.5 (21.8-39.0)	
Mutated	14.6 (11.7-15.5)	
SVM		0.61
Classical	14.4 (10.1-18.8)	
Mesenchymal	16.4 (13.2-19.7)	
Proneural	13.4 (8.7-18.1)	
KNN		0.088
Classical	15.4 (13.0-17.7)	
Mesenchymal	9.6 (5.8-13.4)	
Proneural	14.7 (1.3-28.2)	
ssGSEA		0.438
Classical	14.4 (11.6-17.3)	
Mesenchymal	16.8 (13.9-19.8)	
Proneural	13.7 (8.4-19.1)	
All three algorithms		0.232
Classical	14.0 (10.3-17.7)	
Mesenchymal	9.6 (5.8-13.5)	
Proneural	13.7 (8.6-18.8)	
IGS clusters		0.276
IGS 0	13.7 (12.1-15.4)	
IGS 9	7.62 (5.1-10.0)	
IGS 16	10.1 (--)	
IGS 17	21.1 (5.2-37.1)	
IGS 18	14.6 (10.6-18.5)	
IGS 22	14.7 (0.4-29.1)	

IGS 23	12.5 (5.0-20.0)	
--------	-----------------	--

¹: Log-rank test.

When mesenchymal tumors were compared with classical and proneural tumors grouped together, there were no significant differences in overall survival: SVM, $P = 0.435$; KNN, $P = 0.129$; ssGSEA, $P = 0.399$; all three algorithms, $P = 0.155$. However, in the multivariate analysis adjusted for established prognostic markers, the mesenchymal subtype identified by KNN – but not by other algorithms – was associated with shorter survival (HR, 2.53; $P = 0.038$).

Table S4 Genes associated with overall survival in at least one external data base

Gene	HRadj ¹	95% CI	P-value	FDR	Validation in other data bases		
<i>ANLN</i>	1.5	(1.1-1.9)	0.0026	0.111081081	Graveendel_noCIMP		
<i>DNAJC6</i>	1.4	(1.1-1.7)	0.003	0.111081081	Graveendel_noCIMP		
<i>FZD3</i>	1.4	(1.1-1.8)	0.0036	0.111081081		Rembrandt_noCIMP	
<i>LPGAT1</i>	1.8	(1.2-2.7)	0.0067	0.111081081	Graveendel_noCIMP		
<i>BCAT1</i>	1.2	(1.1-1.5)	0.0067	0.111081081		Rembrandt_noCIMP	
<i>SNCA</i>	1.2	(1-1.3)	0.0073	0.111081081			TCGA_noCIMP
<i>ZNF7</i>	0.47	(0.26-0.83)	0.0091	0.111081081		Rembrandt_noCIMP	
<i>UBXN7</i>	2	(1.2-3.5)	0.0092	0.111081081		Rembrandt_noCIMP	
<i>FAM49B</i>	1.6	(1.1-2.4)	0.0093	0.111081081	Graveendel_noCIMP		
<i>YWHAG</i>	1.8	(1.2-2.8)	0.0096	0.111081081	Graveendel_noCIMP		TCGA_noCIMP
<i>NAPB</i>	1.4	(1.1-1.9)	0.0099	0.111081081	Graveendel_noCIMP		
<i>SLC17A6</i>	1.1	(1-1.2)	0.012	0.111081081	Graveendel_noCIMP		TCGA_noCIMP
<i>FNIP2</i>	1.6	(1.1-2.4)	0.013	0.111081081	Graveendel_noCIMP		
<i>CLOCK</i>	1.5	(1.1-2.1)	0.014	0.111081081	Graveendel_noCIMP		
<i>HTR1D</i>	0.86	(0.77-0.97)	0.015	0.111081081		Rembrandt_noCIMP	
<i>GLRA3</i>	1.1	(1-1.2)	0.015	0.111081081	Graveendel_noCIMP		
<i>CHI3L2</i>	1.1	(1-1.2)	0.016	0.111081081	Graveendel_noCIMP		
<i>SYN2</i>	1.1	(1-1.2)	0.018	0.111081081	Graveendel_noCIMP		
<i>SLCO1A2</i>	1.1	(1-1.3)	0.019	0.111081081	Graveendel_noCIMP		
<i>B3GALNT1</i>	1.5	(1.1-2.1)	0.02	0.111081081	Graveendel_noCIMP	Rembrandt_noCIMP	
<i>MEF2C</i>	1.4	(1.1-1.9)	0.02	0.111081081	Graveendel_noCIMP		
<i>ROBO1</i>	1.4	(1-1.7)	0.021	0.111081081	Graveendel_noCIMP		

<i>DYNLL1</i>	1.8	(1.1-3)	0.021	0.111081081	Graveendel_noCIMP		
<i>PIGV</i>	0.74	(0.58-0.96)	0.023	0.111081081		Rembrandt_noCIMP	
<i>DNAJC10</i>	1.8	(1.1-2.9)	0.024	0.111081081		Rembrandt_noCIMP	
<i>SLC25A25</i>	0.66	(0.46-0.95)	0.024	0.111081081	Graveendel_noCIMP		
<i>VAMP4</i>	1.4	(1-1.9)	0.025	0.111081081	Graveendel_noCIMP		
<i>MAP3K13</i>	1.7	(1.1-2.7)	0.025	0.111081081	Graveendel_noCIMP		
<i>RNF175</i>	1.2	(1-1.3)	0.025	0.111081081			TCGA_noCIMP
<i>KIAA1715</i>	1.5	(1-2.1)	0.027	0.111081081	Graveendel_noCIMP	Rembrandt_noCIMP	
<i>LZTS2</i>	0.62	(0.41-0.95)	0.027	0.111081081		Rembrandt_noCIMP	
<i>PRKACB</i>	1.4	(1-1.9)	0.028	0.111081081	Graveendel_noCIMP		
<i>NCKAP1</i>	1.7	(1.1-2.7)	0.028	0.111081081		Rembrandt_noCIMP	
<i>PCOLCE2</i>	1.2	(1-1.3)	0.028	0.111081081	Graveendel_noCIMP	Rembrandt_noCIMP	
<i>INPP5F</i>	1.4	(1-1.8)	0.029	0.111081081			TCGA_noCIMP
<i>TSEN15</i>	1.4	(1-1.9)	0.03	0.111081081	Graveendel_noCIMP		
<i>DLG2</i>	1.3	(1-1.6)	0.03	0.111081081	Graveendel_noCIMP		
<i>CNDP1</i>	1.1	(1-1.2)	0.031	0.111763158	Graveendel_noCIMP		TCGA_noCIMP
<i>PER3</i>	0.78	(0.62-0.98)	0.033	0.115923077		Rembrandt_noCIMP	
<i>RNF2</i>	1.3	(1-1.6)	0.035	0.11695122	Graveendel_noCIMP		
<i>THRB</i>	1.3	(1-1.6)	0.035	0.11695122	Graveendel_noCIMP		
<i>SLC7A14</i>	1.1	(1-1.2)	0.037	0.118318182	Graveendel_noCIMP		TCGA_noCIMP
<i>FAM49A</i>	1.4	(1-2)	0.038	0.118318182	Graveendel_noCIMP		
<i>ST18</i>	1.1	(1-1.2)	0.038	0.118318182	Graveendel_noCIMP		
<i>UBN1</i>	0.53	(0.28-0.98)	0.042	0.127866667	Graveendel_noCIMP		
<i>NRAS</i>	1.4	(1-2)	0.043	0.128065217	Graveendel_noCIMP		
<i>DPY30</i>	1.2	(1-1.5)	0.049	0.142829787	Graveendel_noCIMP		

¹: HRadj = adjusted hazard ratio: for each gene multivariate Cox models (adjusted by age, extent of resection, and MGMT methylation status) were estimated).

Table S5 Genes associated with overall survival, showing their Ensembl and HGNC IDs and their Gene Ontology terms

Gene	Ensembl ID	HGNC Gene Name and ID	Gene Ontology Term
ZNF7	ENSG00000147789	zinc finger protein 7 (13139)	regulation of transcription, DNA-templated; regulation of transcription by RNA polymerase II; multicellular organism development
UBN1	ENSG00000118900	ubiquitin 1 (12506)	chromatin organization; viral process; regulation of transcription by RNA polymerase II; DNA replication-independent nucleosome assembly
LZTS2	ENSG00000123684	leucine zipper tumor suppressor 2 (29381)	cell division; cell cycle; Wnt signaling pathway; microtubule severing; negative regulation of Wnt signaling pathway; spindle midzone assembly; mitotic cytokinesis; nuclear export; kidney development; negative regulation of cell proliferation; renal system development; negative regulation of canonical Wnt signaling pathway; negative regulation of protein localization to nucleus
SLC25A25	ENSG00000148339	solute carrier family 25 member 25 (20663)	transmembrane transport; ATP transport; response to dietary excess; response to activity; response to food; multicellular organism growth; camera-type eye development; cellular respiration; ATP metabolic process; adipose tissue development; calcium ion transmembrane transport
LPGAT1	ENSG00000123684	lysophosphatidylglycerol acyltransferase 1 (28985)	lipid metabolic process; phospholipid biosynthetic process; phospholipid metabolic process; phosphatidylglycerol acyl-chain remodeling; positive regulation of fatty acid biosynthetic process

YWHAG	ENSG00000170027	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma (12852)	negative regulation of protein serine/threonine kinase activity; G2/M transition of mitotic cell cycle; regulation of G2/M transition of mitotic cell cycle; ciliary basal body-plasma membrane docking; membrane organization; regulation of signal transduction; regulation of synaptic plasticity; negative regulation of protein kinase activity; positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway; protein targeting; cellular response to insulin stimulus; regulation of neuron differentiation
DYNLL1	ENSG00000088986	dynein light chain LC8-type 1 (15476)	regulation of catalytic activity; apoptotic process; microtubule-based process; intraciliary transport involved in cilium assembly; viral process; G2/M transition of mitotic cell cycle; regulation of G2/M transition of mitotic cell cycle; ciliary basal body-plasma membrane docking; neutrophil degranulation; substantia nigra development; antigen processing and presentation of exogenous peptide antigen via MHC class II; ER to Golgi vesicle-mediated transport; macroautophagy; positive regulation of ATP-dependent microtubule motor activity; plus-end-directed negative regulation of phosphorylation; spermatid development; intraciliary retrograde transport; positive regulation of insulin secretion involved in cellular response to glucose stimulus; motile cilium assembly; positive regulation of non-motile cilium assembly
DNAJC10	ENSG00000077232	DnaJ heat shock protein family (Hsp40) member C10 (24637)	oxidation-reduction process; cell redox homeostasis; ubiquitin-dependent ERAD pathway; negative regulation of protein phosphorylation; response to endoplasmic reticulum stress; intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress; positive regulation of ATPase activity; protein folding in endoplasmic reticulum
UBXN7	ENSG00000163960	UBX domain protein 7 (29119)	post-translational protein modification

SLC17A6	ENSG00000091664	solute carrier family 17 member 6 (16703)	transmembrane transport; ion transport; sodium ion transport; L-glutamate transmembrane transport; synaptic transmission; glutamatergic neurotransmitter transport; regulation of synapse structure or activity; neurotransmitter loading into synaptic vesicle
CNDP1	ENSG00000150656	carnosine dipeptidase 1 (20675)	proteolysis; regulation of cellular protein metabolic process
B3GALNT1	ENSG00000169255	beta-1,3-N-acetylgalactosaminyltransferase 1 (globoside blood group) (918)	protein glycosylation; protein N-linked glycosylation; oligosaccharide biosynthetic process; glycosphingolipid metabolic process
KIAA1715	ENSG00000144320	lunapark, ER junction formation factor (21610)	multicellular organism development; endoplasmic reticulum tubular network organization; endoplasmic reticulum organization; limb development; positive regulation of endoplasmic reticulum tubular network organization; endoplasmic reticulum tubular network maintenance; blood coagulation; regulation of chondrocyte differentiation; embryonic forelimb morphogenesis; embryonic digit morphogenesis