Original Article
Whole-exome sequencing, *EGFR* amplification and infiltration patterns in human glioblastoma

Concha López-Ginés[1], Lisandra Muñoz-Hidalgo[2], Teresa San-Miguel[1], Javier Megías[1], Juan Carlos Triviño[3], Silvia Calabuig[1], Pedro Roldán[4], Miguel Cerdá-Nicolás[1], Daniel Monleón[1,2,5]

[1]*Departament of Pathology, University of Valencia, Valencia, Spain;* [2]*Health Research Institute INCLIVA, Valencia, Spain;* [3]*Sistemas Genómicos Inc., Valencia, Spain;* [4]*Department of Neurosurgery, University Clinical Hospital Valencia, Valencia, Spain;* [5]*CIBERFES_ISCIII, Valencia, Spain*

Abstract: Glioblastoma (GBM) is the most common malignant primary brain tumor in adults. This cancer shows rapid, highly infiltrative growth, that invades individually or in small groups the surrounding tissue. The aggressive tumor biology of GBM has devastating consequences with a median survival of 15 months. GBM often has Epidermal Growth Factor Receptor (*EGFR)* abnormalities. Despite recent advances in the study of GBM tumor biology, it is unclear whether mutations in GBM are related to *EGFR* amplification and relevant phenotypes like tumor infiltration. This study aimed to perform whole-exome sequencing analysis in 30 human GBM samples for identifying mutational portraits associated with *EGFR* amplification and infiltrative patterns. Our results show that *EGFR*-amplified tumors have overall higher mutation rates than EGFR-no-amplified. Six genes out of 2029 candidate genes show mutations associated with *EGFR* amplification status. Mutations in these genes for GBM are novel, not previously reported in GBM, and with little presence in the TCGA database. *GPR179*, *USP48,* and *BLK* show mutation only in *EGFR*-amplified cases, and all the affected cases exhibit diffuse infiltrative patterns. On the other hand, mutations in *ADGB, EHHADH,* and *PTPN13,* were present only in the *EGFR*-no-amplified group with a more diverse infiltrative phenotype. Overall, our work identified different mutational portraits of GBM related to well-established features like *EGFR* amplification and tumor infiltration.

Keywords: Glioblastoma (GBM), *EGFR* amplification, whole exome sequencing, infiltration patterns, FISH, somatic mutation

Introduction

Glioblastoma (GBM) is the most common malignant brain tumor in adults. Ninety percent of these tumors are classified as primary GBMs. Primary GBMs initiate and progress rapidly in the absence of clinical or histologic evidence of less malignant precursor lesions [1]. This tumor shows rapid, highly infiltrative growth, and GBM tumor cells tend to invade the surrounding tissue either by individuals cells, which is called diffuse phenotype, or in small groups of cells, which is called a nodular phenotype [2-5]. Despite intensive research and multimodal treatment including surgical resection, radiotherapy, and chemotherapy, patients today still face a dismal prognosis with a median survival of 15 months [1].

According to the 2016 CNS WHO classification, GBM classifies into *IDH*-wildtype GBM or *IDH*-mutant GBM (about 10% of cases). In contrast to *IDH*-wildtype GBM, *IDH*-mutant GBM usually has a history of prior lower grade diffuse glioma and preferentially arises in younger patients [1]. In addition to *IDH* gene status, comprehensive genetic analysis of GBMs demonstrated different molecular alterations that contribute to tumorigenesis. Numerous tumor suppressor genes and oncogenes are inactivated and activated, respectively. As with many other types of cancer and in the context of precision medicine, therapeutic decisions in GBM management rely on biomarker analysis. GBM tumors often contain Epidermal Growth Factor Receptor (*EGFR)* abnormalities including amplification, mutation, protein expression alterations, and changes in

protein function. EGFR protein plays key roles in important cellular functions including proliferation and migration, among others. Alterations such as amplification or mutation of the *EGFR* gene are a hallmark of disease pathogenesis in GBM, observed in approximately 50% of cases [6-8]. Abnormal EGFR activity or function is also associated with changes in the regulation of different signaling pathways including p53, Rb, and receptor tyrosine kinase (RTK)/Ras/phosphoinositide 3-kinase (PI3K) [9, 10]. Recent studies analyzed the frequency of *EGFR* amplification by FISH in a large cohort of GBM patients to evaluate whether *EGFR* amplification differs by geographical origin of patients and if these differences are important to optimize treatments [8]. In addition to *EGFR* alterations, GBMs exhibit other frequent molecular alterations like *TERT* promoter mutations (72%), *TP53* mutations (27%), and *PTEN* mutations (24%) [1, 11]. However, it is still unclear whether these other frequent mutations are related to *EGFR* amplification and how they impact relevant phenotypes like tumor infiltrative patterns.

Next-generation sequencing (NGS) technologies help in unveiling DNA sequences for the characterization of both frequent and rare genomic alterations relevant in many different types of cancer. The amount and quality of data obtained in NGS studies provide a unique and powerful insight into the landscape of genetic alterations and subsequent biological changes in GBM genesis and progression. The study of nucleotide changes through Whole Exome Sequencing (WES) in coding and non-coding exons and RNAs may help for a better understanding of cancer progression and the discovery of new molecular targets for personalized therapies. The analysis of WES data in association with relevant subgroups of patients and response to specific therapies may provide signatures, markers, and algorithms to identify genetic alterations networks relevant to the clinical management of the disease [12, 13]. Recent GBM studies using WES technologies revealed novel genetic alterations related to the age of patients, recurrence, and survival which may play important roles in characterizing and treating GBM [14-17].

In this context, this study aimed to perform WES analysis in 30 tumor samples of human GBM, subdividing them according to the most frequent genetic alteration related to the *EGFR* amplification or not, as well as its possible repercussion in its infiltrative capacity.

Material and methods

Patients and tumor samples

Thirty patients with GBM from the Hospital Clínico Universitario in Valencia were included in the study and biopsies were obtained from their tumors. Patients did not have a previous diagnosis of lower-grade astrocytoma and therefore tumors were clinically diagnosed as primary GBMs. This study was reviewed and approved by the clinical investigation ethics committee at the Hospital Clínico Universitario (CEIC). The samples were diagnosed by two different neuropathologists. The tissue was retrieved from the patients during surgical resection. Immediately after resection, tissue was split into two samples, one for histopathological and immunohistochemical analysis and another one for molecular analysis. The sample for molecular analysis was immediately frozen and stored at -80°C until use. The samples for histopathological and immunohistochemical analysis were fixed in neutral-buffered formalin during 48 h, embedded in paraffin, sectioned, and stained with hematoxylin-eosin (HE).

All tumor samples were analyzed and classified according to the WHO classification criteria [1] and the GBM diagnosis was confirmed Paraffin-embedded sections underwent further immunohistochemistry (IHC) analysis using the avidin-biotin-peroxidase complex method and antibodies directed against the glial fibrillary acidic protein (GFAP), and Ki-67 (MIB-1). Ki-67 proliferation index was evaluated by MIB-1 antibody staining (Dako, Glostrup, Denmark). The final Ki-67 index was calculated as the percentage of immunopositive nuclei in ten fields of two different sections at 20X. Samples were then classified as <5% stained nuclei, 5%-10%, 10%-25%, and >25% stained nuclei.

Molecular status of IDH1/IDH2 and TP53 genes

We used a tissue DNA kit (Qiagen) for extracting genomic DNA from all 30 GBMs samples, according to the manufacturer's instructions. Subsequently, the catalytic domain of IDH1

including codon 132 was amplified using the sense primer IDH1 F: 5'-CGGTCTTCAGAG-AAGCCATT-3' and the antisense primer IDH1 R: 5'-GCAAAATCACATTATTGCCAAC-3'. The catalytic domain of IDH2 including codon 172 was amplified using the sense primer IDH2 F: 5'-AGCCCATCATCTGCAAAAAC-3' and the antisense primer IDH2 R: 5'-CTAGGCGAGGAGC-TCCAGT-3'. PCR was performed in 200 ng of DNA by applying 40 cycles including denaturation at 94°C for 45 s, annealing at 60°C for 45 s, and extension at 72°C for 45 s in a total volume of 25 µl using an AmpliTaq Gold Master Mix (Thermo Fisher Scientific). PCR products were purified using Centricon columns (Amicon, Beverly, MA) and standard manufacturer's instructions. The purified PCR amplification products, both forward and reverse chains, were analyzed on an ABI 310 Sequencer (Foster City, CA). TP53 sequencing was carried out in four different PCR amplification reactions (exons 5-8).

Molecular status of EGFR by Fluorescence in-situ hybridization

To evaluate the EGFR gene amplification status, dual-color fluorescence in-situ hybridization (FISH) was performed in tissue microarrays (TMAs) using the LSI EGFR Spectrum Orange/ CEP 7 Spectrum Green Probe from Vysis (Abbott Laboratories, Downers Grove, IL, USA. Cat. No. 32-191053). For quantification purposes, we counted positive signals in 100-150 non-overlapping tumor cell nuclei. In each case, the mean signal numbers for the *EGFR* gene and the control CEP 7 probe were calculated and used to calculate the EGFR/CEP 7 ratio. The *EGFR* gene was considered to be amplified when the *EGFR/CEP7* signal ratio was greater than 2 in ≥10% recorded cells [7].

Multiplex ligation-dependent probe amplification (MLPA)

DNA for MLPA analysis was obtained by a QIAamp DNA FFPE tissue kit (Qiagen, Inc., Valencia, CA) applied to biopsy punches from selected areas of the paraffin blocks of each sample. When necessary, the quality and quantity of the samples were assessed and improved by standard ethanol precipitation. Multiplex ligation-dependent probe amplification (MLPA) was performed to determine the copy number variations (CN), the mutant EGFRvIII form exhib-iting loss of exons 2 to 7 of genes in a single reaction. SALSA MLPA kits were used following the manufacturer's instructions (MRC-Holland, Amsterdam, Netherlands). Capillary electro-phoresis in an ABI 310 Sequencer (Applied Biosystems Inc., Foster City, CA) separated MLPA fragments The Coffalyser excel-based MLPA analysis software (MRC-Holland) was used for data analysis. Thresholds to detect losses and gains of genetic material were set, respectively, at 0.75 and 1.25. For this multi-genic technique, the two specifically designed sets of probes used contained different genes. Salsa MLPA kit P105-C1 with the genes: PTEN (10q23.31), and CDKN2A/CDKN2B (9p21.3).

DNA methylation status

Sodium bisulfite conversion was performed using an EZ-96 DNA methylation kit according to the manufacturer's protocol (Zymo Research, Freiburg Germany) on 1 µg of genomic DNA. For quantitative methylation measurements, we used Sequenom's MassARRAY platform which utilizes MALDI-TOF mass spectrometry in com-bination with RNA base-specific cleavage (MassCLEAVE). PCR primers for amplification of different regions of the MGMT gene were designed using Epidesigner (Sequenom, San Diego, CA, USA). Whenever feasible, amplicons were designed to cover CGIs in the same region as the 5' UTR. For each reverse primer, an addi-tional T7 promoter tag for in vivo transcription was added, as well as a 10-mer tag on the for-ward primer to adjust for melting-temperature differences. The PCRs were carried out in a 5 µl format with 10 ng bisulfite-treated DNA, 0, 2 units of TaqDNA polymerase (Sequenom), 1× supplied Taq buffer, and 200 mM PCR primers. Mass spectra were collected using a Mass-ARRAY mass spectrometer (Bruker-Sequenom). The resulting spectra were analyzed using pro-prietary peak picking and signal-to-noise calcu-lations after which the spectra' methylation ratios were generated using EpiTYPER software v1.2 (Sequenom, San Diego, CA, USA).

EGFR immunohistochemistry

Paraffin-embedded sections were analyzed by using the avidin-biotin-peroxidase method. We identified the wild-type EGFR and EGFRvIII using the monoclonal mouse anti-human EGFR antibody (clone H11; Dako). Four levels of EGFR expression were established for the analysis

according to the number of stained tumor cells: 0 (no staining), 1 (light or focal, 1%-10% of cells), 2 (moderate, 10%-25% of cells), and 3 (strong, >25% of cells). Samples with scores of 0 or 1 were considered as no overexpressing EGFR whereas scores of 2 and 3 were considered as overexpressing EGFR.

Whole-exome sequencing (WES)

DNA was extracted from snap-frozen tumor samples. For each tumor specimen submitted for WES, sections were reviewed by a neuropathologist to confirm the diagnosis of GBM, and the adequacy for sequencing was assessed. The quality and the quantity of the total DNA have been determined in Nanodrop-1000, by agarose gel electrophoresis and Qubit 2.0. Fragment libraries were obtained and captured using SureSelect Capture Library "SureSelectXT Human All Exon v5", following Agilent protocols and recommendations. The quality and the quantity of the libraries were analyzed in 4200 TapeStation, High Sensitivity assay, and Qubit 2.0. Raw data is accessible in the Zernodo OpenAire data repository (https://doi.org/10.5281/zenodo.4636211).

Computational analysis

The exome was based on an Illumina Hi-Seq2000 sequencing platform using a paired-end 2×100 strategy and an Agilent's SureSelect Target Enrichment System for 51 Mb. Sequencing was done with a 50× of coverage. The reads were aligned against the last version of human genome reference (GRCh38/hg38 assembly) using the Burrows-Wheeler Alignment tool (BWA) [18]. Low-quality reads and sequences flagged as PCR duplicates were removed from the BAM file using the Picard tools. We follow the Best Practices v3 guideline. The variant calling process for SNPs and Indels identification was based on GATK [19] and VarScan [20] methods. Python scripts were developed to combine variants. Variants annotation was based on Ensembl variant effect predictor [21] and NCBI database. For pathological variants selection, we defined different filters: i) novel or low global frequency variants (<0.05) using 1000 Genomes Database [22], ii) variants with a high impact according to Ensembl annotation [23] or, iii) missenses with deleterious prediction according to Sift [24] and polyphen [25] methods. A Fisher exact

test was applied to evaluate differences between different phenotypes. Statistical significance was set at P<0.05 and all statistical tests were two-sided. Multivariate analyses including PCA, DA, and HCA were performed with R (3.3.2 version).

Somatic mutation data analysis and functional enrichments analysis

Based on the statistically significant mutated genes a discriminant analysis was performed using the Adegenet package of R [26, 27]. This method first uses a PCA to reduce the dimension of the data and then applies a discriminant analysis (DA) for detecting the most discriminative variables. In the present analysis, two PC2 dimensions were calculated over the significant mutations. Then, genes with the most categorical discriminant power were selected using the score associated with the DA. The significant genes were also contrasted against the TGCA (Analysis Working Group Data Release Package) databases [28]. Functional enrichment analysis was carried out using the cluster profiler package in R [29].

Results

Patient data and EGFR status

A set of 30 GBM cases, 29 of them *IDH1/2* wild-type, and one case with the mutation c.395G>A (*p.R132H*), was included in this study. None of the patients had a previous astrocytoma diagnosis or previous treatments. Of these, 17 were males and 13 were females. The age of the patients ranged from 31 to 75 years, with a mean of 61,3 years. The location for all tumors was in the supratentorial region. The size of tumors ranged from 1,4 to 8 cm, with a mean of 4,5 cm. The average survival of the patients was 14,8 months (range 2 to 26 months, **Table 1**).

All tumors showed clinical and histological features of GBM including pleomorphic astrocytic tumor cells, prominent microvascular proliferation, and necrosis. All cases presented neoplastic cells with expression of Glial Fibrillary Astrocytic Protein (GFAP). We observed two distinct infiltration patterns independent of the amount of infiltration, as reported in previous works: a diffuse infiltrative pattern and a nodular pattern [2-5]. For infiltration pattern analy-

Table 1. Clinicopathological and genetic features in 30 samples of human GBM

|	Sex/Age	Location	Size (cm)	Survival (months)	Infiltration pattern	Ki-67	EGFR expression	*IDH1*	*EGFR* amp
1	F/72	F	6	8	diffuse	10-25%	3	wt	amp
2	M/69	T	6.5	7	diffuse	10-25%	3	wt	amp
3	F/58	T	6	23	diffuse	>25%	3	wt	amp
4	M/61	T	8	12	nv	>25%	3	wt	amp
5	F/66	F	4	26	nv	10-25%	2	wt	amp
6	M/69	T	5.6	12	diffuse	>25%	2	wt	amp
7	F/61	F	4.5	2	diffuse	>25%	3	wt	amp
8	F/58	P	5.6	20	nv	10-25%	1	wt	amp
9	M/57	F	6	17	diffuse	10-25%	2	wt	amp
10	M/42	P	5	NV	diffuse	10-25%	1	wt	amp
11	M/60	T	5.3	5	nodular	5-10%	2	wt	amp
12	F/31	F	2.6	NV	nodular	10-25%	2	wt	amp
13	F/73	P	4.5	26	nodular	5-10%	1	wt	amp
14	M/35	P	5	26	diffuse	>25%	2	wt	amp
15	M/66	T	4	11	diffuse	5-10%	0	wt	no-amp
16	F/35	F	2.6	NV	nodular	5-10%	0	p.R132H	no-amp
17	M/75	P	3.3	6	nodular	>25%	2	wt	no-amp
18	M/55	F	4.3	26	nodular	5-10%	1	wt	no-amp
19	M/67	P	4	21	diffuse	5-10%	2	wt	no-amp
20	F/75	F	6	2	nodular	10-25%	0	wt	no-amp
21	M/60	F	8	2	diffuse	5-10%	0	wt	no-amp
22	M/50	T	1.9	10	nv	10-25%	0	wt	no-amp
23	F/65	T	2	22	nv	5-10%	2	wt	no-amp
24	M/52	T	6	23	nv	5-10%	0	wt	no-amp
25	M/74	P	1.5	NV	nv	10-25%	0	wt	no-amp
26	F/62	P	5.6	6	diffuse	10-25%	1	wt	no-amp
27	M/55	T	1.5	14	diffuse	5-10%	1	wt	no-amp
28	F/61	P	3.8	10	diffuse	>25%	1	wt	no-amp
29	M/43	P	5	13	nodular	5-10%	0	wt	no-amp
30	F/73	P	1.4	15	nodular	10-25%	0	wt	no-amp

Sex: male (M), female (F). Location: frontal (F), temporal (T), parietal (P), occipital (O). Survival: all cases are exitus at the end of the present study and NV that are cases with unknown evolution. nv: no valuable the infiltration pattern. *EGFR* amplificated (amp).

sis, only samples where, with an objective lens of 25×, at least 75% of the peripheral tumor area could be evaluated. As previously reported, samples with the diffuse infiltrative pattern exhibit a progressive transition from areas with higher counts of neoplastic cells in the center of the tumor towards decreasing counts of neoplastic cells in the periphery of the tumor, infiltrating adjacent nervous tissue (**Figure 1A**). The nodular infiltrative pattern shows a delimitation in the form of a tumor front between high-density neoplastic cells and peripheral nervous tissue. The nodular infiltrative pattern

also shows isolated groups of neoplastic cells in perivascular spaces, clearly separated from the tumor (**Figure 1B**). Our analysis identified 14 cases with a diffuse infiltrative pattern and 9 cases with a nodular infiltrative pattern.

We analyzed the amplification of *EGFR* for all samples with FISH methodology (**Figure 1C** and **1D**). Based on the amplification of *EGFR*, we classified samples into two groups: *EGFR*-amplified (14 cases, for example in **Figure 1C**) and *EGFR*-no-amplified (16 cases, for example in **Figure 1D**). We observed more men in the
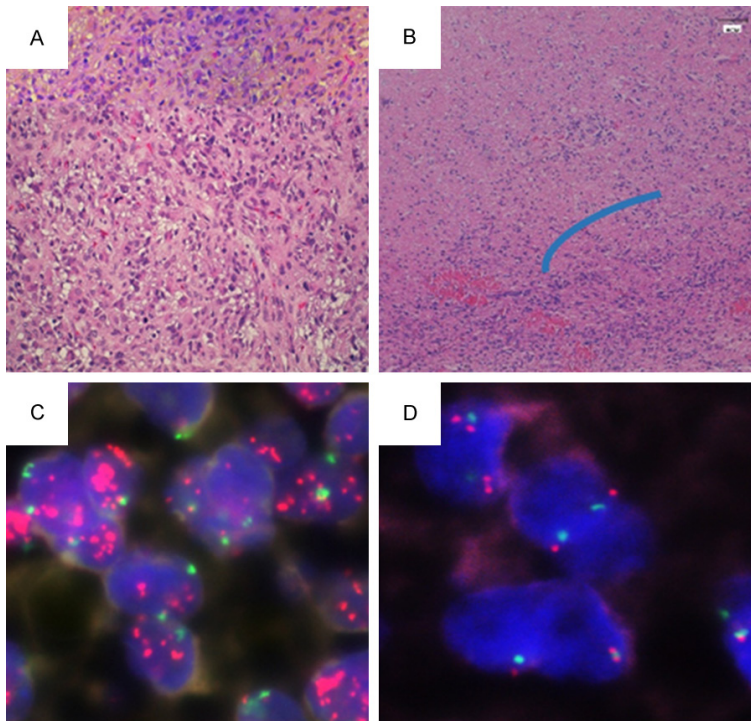
Figure 1. Histopathology, EGFR expression, and EGFR amplification (determined by FISH) in GBM. (A and B) HE stained images of human GBM cells. (A) Individually invading cells giving rise to a diffuse pattern with a progressive transition from the center of the tumor towards the peritumoral region (HE, objective 20×). (B) Infiltration of groups of cells giving rise to a nodular pattern where a front (blue line) partially delimits the center of the tumor and peritumoral region (HE, objective 10×). (C and D) FISH images of human GBM cells showing *EGFR* gene copies (green fluorescence, centromeres; magenta fluorescence, *EGFR*) in GBM *EGFR*-amplified (C) and *EGFR*-no amplified (D) cases.

EGFR-amplified, and all of them with diffuse infiltrative patterns. Proliferation rates were compared by measuring the Ki-67 proliferation index. The expression of Ki67 is strongly associated with tumor cell proliferation and growth and is widely used in the routine pathological investigation as a proliferation marker. Out of 14 *EGFR*-amplified samples, Ki67 immunopositivity was high (>25% of cells) in 5 samples (35,7%), medium (10-25% of cells) in 7 cases (50%); and low (5-10% of cells) in 2 cases (14,3%). On the other hand, out of 16 *EGFR*-no-amplified samples, Ki67 immunopositivity was high (>25% of cells) in 2 samples (12,5%), medium (10-25% of cells) in 5 cases (31,2%); and low (5-10% of cells) in 9 cases (56,3%). In general, the Ki-67 index was higher in the *EGFR*-amplified group. IHC analysis of EGFR protein expression in the 30 GBM samples revealed that EGFR protein was over-expressed in 11 out of 14 *EGFR*-amplified samples. In contrast, EGFR protein was overexpressed only in 3 out of 16 *EGFR*-no-amplified cases. The association between amplification and overexpression of EGFR was statistically significant at the 0.05 *P*-value level (Fisher exact test).

Genetic markers of GBM

We analyzed mutations in exons 5, 6, 7, and 8 of the TP53 gene in all 30 cases of GBM. Two samples (6.6% of cases) showed TP53 mutations. The sample from case 13, from the EGFR-amplified group, showed a heterozygous p.C141R mutation (substitution-missense type) in exon 5 of TP53. The sample from case 16, from the EGFR-*no-amplified* group, showed a heterozygous p.R273C mutation (substitution-missense type) in exon 8 of TP53. Somatic copy number alteration (CNA) analysis showed losses in CDKN2A/CDKN2B for 12 of the 30 GBM samples. Eight of them also showed EGFR

group of *EGFR*-no-amplified, a similar average age in both groups and slightly higher tumor size in the *EGFR* amplified group (5.3 cm vs. 3.8 cm). Mean survival time did not show statistically significant differences neither with respect EGFR amplification status (EGFR-amplified 13.3±7.3 months vs. EGFR-no-amplified 14.2± 7.0 months, *P*-value 0.76) nor with respect infiltration pattern (diffuse 13.6±7 months vs 11.9±7 8.2 months, *P*-value 0.65). In addition, no differences associated with EGFR amplification or tumor infiltration patterns in hazard ratios calculated by a Cox model were identified when adjusted by age and tumor size (see Figure S1). The analysis of infiltrative patterns shows that 73% of cases with *EGFR*-amplified had a diffuse infiltrative pattern, and 50% of cases *EGFR*-no-amplified exhibited a nodular infiltrative phenotype. Mutational analysis of *EGFR* by MLPA showed that mutant EGFRvIII only appeared in 4 tumors, three of them with

amplification. Likewise, we detected losses in the number of copies of PTEN in seven cases, four of them in the EGFR-amplified group (Table S1). The methylation levels of the CpG islands present in the promoter regions of MGMT were evaluated in 23 GBM samples and compared to 6 non-neoplastic control brain tissue samples. The methylation values of these controls ranged from 6,5% to 14,5%. Of the total GBM samples analyzed, 78% showed hypermethylation of the MGMT promoter compared to control levels (Table S1).

Sequence coverage and mutations analysis

We analyzed exome data from 30 GBM cases. Performance and technical quality data for the NGS sequencing process were homogeneous for all samples. The average coverage was focus on 50× with a standard deviation of 2×. We performed a whole-exome analysis and seek for variants with respect to the GRCh38 homo sapiens (human) genome assembly from Genome Reference Consortium. We identified 19867 somatic variations in 2029 genes. These 2029 genes were our candidate genes. The overall number of pathological mutations for each sample ranged from 743 to 1402 with a mean value of 827 and a standard deviation of 127. All samples presented small variations in the mutational rate values. Sample 30 was considered an outlier with 1402 pathological variants. This value was greater than 4 standards deviations above the mean across all samples (Zscore 4,2) (**Table 2**).

The mutation rate in the *EGFR*-amplified group was higher (818.07) than in the *EGFR*-no-amplified group (794.1) (excluding case 30). Moreover, transversions were the most predominant somatic substitution, and the total mutation and transition/transversion rate between both groups did not exhibit statistically significant differences (*P*-values of 0.5 and 0.9 respectively, **Table 2**). Missense mutations were the most frequently observed in both groups. However, frame deletions were more frequent in cases with a diffuse infiltrative pattern whereas frameshift deletion was more frequent in cases with a nodular infiltrative pattern. **Table 3** contains the tumor mutational burden (TMB) for all the samples. Average TMBs were 188.47±3.24 for *EGFR*-amplified group and 186.84±3.44 for the *EGFR*-no-

amplified group. Differences in tumor mutational burden between the *EGFR*-amplified and *EGRF*-no-amplified groups were not statistically significant.

Thirty-six mutated genes exhibited statistical significance in the 30 cases of GBM (**Table 4**). These mutations are distributed throughout all chromosomes except 9, 13, 18, 21, and 22. Chromosomes 6, 3, 4, 8, and 17 contain most of these mutations. A Hierarchical Cluster Analysis (HCA) of all the 30 GBM samples based on these 36 genes also shows a differential distribution of mutations between *EGFR*-amplified and *EGFR*-no-amplified cases. The genes affected by these mutations participate in biological processes such as G-protein cell receptor (GPCR) downstream signaling pathway, apoptosis, cellular proliferation, microtubule motor activity, degradation of receptors, angiogenesis, cell adhesion and migration, transcription factors, and transmembrane transporters, among others (**Table 5**). The analysis of GBM data from The Cancer Genome Atlas (TCGA) public database for mutation on these genes showed that most of them are mutated in GBM TCGA cases in a percentage that oscillates between 0.1% and 6%. Ten of the analyzed genes showed co-occurrence with the amplification status of *EGFR* in the TCGA data (**Table 4**). None of the mutated genes showed an association with patient survival.

Association between mutated genes, EGFR amplification, and infiltration patterns

We compared mutation frequencies in *EGFR*-amplified and *EGFR*-no-amplified groups for identifying differential mutational portraits. We analyzed mutations in 2029 possible candidate genes and evaluate the statistical significance of mutation frequencies with a Fisher exact test. Six mutated genes showed an association at a 0.05 significance level with the status of *EGFR*. The exact impact of these mutations in the function of the expressed proteins (activation or inactivation) remains unclear. From them, *GPR179, USP48,* and *BLK* show mutation only in the *EGFR*-amplified group, and interestingly all the affected cases exhibit a diffuse infiltrative phenotype (**Table 4**; **Figure 2**). Global functional analysis of these three genes reveals roles in different processes such as GPCR downstream signaling pathway, EGFR

Table 2. Rates of mutations in 30 GBM samples and corresponding status for EGFR amplificacion and infiltrative pattern

Case	Pathological variants	*EGFR* amplification*	Infiltrative phenotype	Transition	Transversion	% Transition	% Transversion
1	808	Amp	Diffuse	133	513	16.46	63.49
2	838	Amp	Diffuse	138	545	16.47	65.04
3	838	Amp	Diffuse	138	545	16.47	65.04
4	834	Amp	nv	126	537	15.11	64.39
5	794	Amp	nv	134	513	16.88	64.61
6	917	Amp	Diffuse	150	585	16.36	63.79
7	800	Amp	Diffuse	136	507	17.00	63.38
8	863	Amp	Nv	136	551	15.76	63.85
9	800	Amp	Diffuse	139	478	17.38	59.75
10	811	Amp	Diffuse	131	507	16.15	62.52
11	801	Amp	Nodular	118	505	14.73	63.05
12	802	Amp	Nodular	128	508	15.96	63.34
13	743	Amp	Nodular	122	450	16.42	60.57
14	804	Amp	Diffuse	117	514	14.55	63.93
15	773	No-amp	Diffuse	110	490	14.23	63.39
16	793	No-amp	Nodular	134	489	16.90	61.66
17	797	No-amp	Nodular	128	506	16.06	63.49
18	793	No-amp	Nodular	134	489	16.90	61.66
19	754	No-amp	Diffuse	151	439	20.03	58.22
20	763	No-amp	Nodular	123	482	16.12	63.17
21	818	No-amp	Diffuse	134	527	16.38	64.43
22	797	No-amp	Nv	133	506	16.69	63.49
23	811	No-amp	Nv	127	526	15.66	64.86
24	795	No-amp	Nv	125	505	15.72	63.52
25	775	No-amp	Nv	130	502	16.77	64.77
26	812	No-amp	Diffuse	127	522	15.64	64.29
27	795	No-amp	Diffuse	121	498	15.22	62.64
28	772	No-amp	Diffuse	117	495	15.16	64.12
29	864	No-amp	Nodular	138	553	15.97	64.00
30	1402	No-amp	Nodular	136	1049	9.70	74.82

*Amp, EGFR amplified; No-amp, EGFR no amplified.

degradation regulated by proteasomes, and apoptosis modulation (**Table 5**).

On the other hand, mutations in three other genes, *ADGB, EHHADH,* and *PTPN13,* were present only in the *EGFR*-no-amplified group (**Table 4**; **Figure 2**). These cases exhibit a more diverse infiltrative phenotype including GBMs with nodular and diffuse infiltrative patterns. Global functional analysis on these other three genes shows involvement in the control of energy metabolism and oxidation and apoptosis modulation (**Table 5**; **Figure 2**). NPY4R gene is the most frequent mutation in our data set with 11 out of 30 GBMs distributed in 8 *EGFR*-

amplified cases (57%) and 3 *EGFR*-no-amplified cases (19%). Other genes studied also exhibit interesting patterns although without statistical significance. Overall, we observe a higher number of mutations in the *EGFR*-amplified GBMs than in *EGFR*-no-amplified GBMs (**Table 4**).

For further insights into our data, we explored intrinsic relationships between cases and mutated genes using consecutive Principal Component Analysis (PCA) and Discriminant Analysis (DA) over the 36 significant genes. As expected, the PCA scores plot shows clear separation along with the principal component 1

Table 3. Tumor mutational burden (TMB) in all 30 GBM of the study

Samples	Number of Variants	Exome Size	TMB
1	9624	51 Mbs	188,705882
2	9722	51 Mbs	190,627451
3	9757	51 Mbs	191,313725
4	9713	51 Mbs	190,45098
5	9488	51Mbs	186,039216
6	9785	51 Mbs	191,862745
7	9558	51 Mbs	187,411765
8	9760	51 Mbs	191,372549
9	9582	51 Mbs	187,882353
10	9646	51 Mbs	189,137255
11	9592	51 Mbs	188,078431
12	9607	51 Mbs	188,372549
13	9122	51 Mbs	178,862745
14	9615	51 Mbs	188,529412
15	9411	51 Mbs	184,529412
16	9424	51 Mbs	184,784314
17	9551	51 Mbs	187,27451
18	9460	51 Mbs	185,490196
19	9223	51 Mbs	180,843137
20	9298	51 Mbs	182,313725
21	9698	51 Mbs	190,156863
22	9555	51 Mbs	187,352941
23	9694	51 Mbs	190,078431
24	9503	51 Mbs	186,333333
25	9417	51 Mbs	184,647059
26	9697	51 Mbs	190,137255
27	9548	51 Mbs	187,215686
28	9365	51 Mbs	183,627451
29	9777	51 Mbs	191,705882
30	9840	51 Mbs	192,941176

(PC1) between *EGFR*-amplified and *EGFR*-no-amplified groups (**Figure 3**). PC1 and PC2 explain together 95% of the total data variance. The loadings plot showed that the mutations in *SYNE1, DNHD1, PTPN13,* and *MKI67* are contributing the most to discrimination between *EGFR*-amplified and *EGFR*-no-amplified (**Figure 3**). According to the global functional analysis, functions related to these genes include cell cycle control, microtubule motor activity, apoptosis, and cellular proliferation (**Table 5**).

Discussion

Analysis of mutational status by WES analysis is a novel approach to disentangle relation-ships between frequent mutations and tumor biology. In this study, we identified a set of mutated genes associated with EGFR amplification status. We used FISH for classifying a set of 30 GBMs into *EGFR*-amplified and *EGFR*-no-amplified groups and analyzed their mutational status by WES. The interest in classifying GBMs according to EGFR status lies in the high frequency of alterations in this receptor-tyrosine kinase (RTK). In agreement with classical descriptions, half of our patients displayed EGFR amplification [8]. We observed high mutation rates both in *EGFR*-amplified and EGFR-no-amplified GBM. However, *EGFR*-amplified samples have overall higher mutation rates than EGFR-no-amplified samples. Hypermutagenesis refers to a marked increase in the number of mutations due to the continuous mutagenic process. Hypermutagenesis is a common feature of GBM related to inherited or acquired alterations in DNA repair pathways in several cancer types [30, 31]. Although the impact in tumor biology of EGFR gene amplification has been extensively explored, there is no evidence of overall survival benefit with the use of EGFR-targeted therapies in GBM. However, isolated cases may still benefit from these therapies. In these cases, therapy selection should rely not only on the presence of EGFR amplification or mutations but also on these higher rates of overall mutations [32].

In our study, six genes show mutations with a statistically significant association to *EGFR* amplification status. Mutations in these genes for GBM are novel, not previously reported in GBM, and with little presence in the TCGA database. However, the presence in our tested groups is highly differential and potentially associated with infiltrative phenotypes. Three mutated genes were only present in the *EGFR*-amplified group and the other three were only in the *EGFR*-no-amplified group. *GPR179*, *USP48,* and *BLK* were exclusively mutated in *EGFR*-amplified tumors and all cases with mutations in these genes exhibited a diffuse infiltrative phenotype. *GPR179* gene encodes a member of the glutamate receptor subfamily of GPCR. GPCRs and Receptor Tyrosine Kinases (RTKs), including EGFR, regulate different signaling networks involved in many diseases including cancer. Several mechanisms for the transactivation of the EGFR by GPCR are under investigation [33]. Inactivating mutations in paralogs of

Table 4. Distribution of mutations for the 36 genes significantly associated with EGFR amplification status in 30 GBM and the TCGA database

Gene	*P*-value	EGFR-amp	EGFR-no amp	Diffuse	Nodular	% TCGA	Co-occurrence EGFR *P*-value	Mutation type
GPR179	0.0140	5 (35.7%)	0	4 (28.6%)	0	1.1	0.018	MSV
USP48	0.0140	4 (28.6%)	0	2 (14.3%)	0	0.8	-	FSV
BLK	0.0365	4 (28.6%)	0	3 (21.4%)	0	0.1	0.006	MSV
ADGB	0.0446	0	5 (31.3%)	2 (14.3%)	2 (22.2%)	0	-	MSV, SRV
EHHADH	0.0446	0	5 (31.3%)	2 (14.3%)	2 (22.2%)	0.7	<0.001	MSV
PTPN13	0.0446	0	5 (31.3%)	2 (14.3%)	2 (22.2%)	0.5	-	SGV & MSV
NPY4R	0.0566	8 (57.1%)	3 (18.8%)	4 (28.6%)	4 (44.4%)			MSV
DNHD1	0.0724	5 (35.7%)	1 (6.3%)	3 (21.4%)	2 (22.2%)	0.7	-	SGV & MSV
MKI67	0.0724	5 (35.7%)	1 (6.3%)	3 (21.4%)	0	1.9	0.001	FSV & MSV
ZNF280D	0.0724	5 (35.7%)	1 (6.3%)	3 (21.4%)	1 (11.1%)	0.4	-	MSV
CES5A	0.0859	1 (7.1%)	6 (37.5%)	3 (21.4%)	2 (22.2%)	0	-	MSV & SAV
CHMP6	0.0859	1 (7.1%)	6 (37.5%)	5 (35.7%)	1 (11.1%)	0.1	-	MSV & SDV
MROH6	0.0859	1 (7.1%)	6 (37.5%)	4 (28.6%)	1 (11.1%)	0	-	MSV & FSV
PKHD1	0.0859	1 (7.1%)	6 (37.5%)	1 (7.1%)	4 (44.4%)	5	-	MSV & SGV
RAB44	0.0859	1 (7.1%)	6 (37.5%)	4 (28.6%)	1 (11.1%)	0	-	FSV & MSV
SYNE1	0.0859	1 (7.1%)	6 (37.5%)	1 (7.1%)	5 (55.6%)	6	0.006	SAV & MSV
ARSH	0.0896	3 (21.4%)	0	2 (14.3%)	0	0.3		MSV
BMP2K	0.0896	3 (21.4%)	0	2 (14.3%)	0	0.7	0.015	MSV
CFAP61	0.0896	3 (21.4%)	0	1 (7.1%)	1 (11.1%)	0.3	-	MSV
DEUP1	0.0896	3 (21.4%)	0	1 (7.1%)	0	0.2	-	MSV & SDV & SLV
EPHB6	0.0896	3 (21.4%)	0	2 (14.3%)	1 (11.1%)	1	-	MSV & SGV
ERAP1	0.0896	3 (21.4%)	0	1 (7.1%)	1 (11.1%)	0.2	-	MSV
FLVCR2	0.0896	3 (21.4%)	0	2 (14.3%)	1 (11.1%)	0.1	-	MSV
KALRN	0.0896	3 (21.4%)	0	1 (7.1%)	1 (11.1%)	0.8	-	SGV & MSV
KIF13A	0.0896	3 (21.4%)	0	1 (7.1%)	0	1.5	-	MSV
LPA	0.0896	3 (21.4%)	0	3 (21.4%)	0	1.1	0.004	MSV
PPIG	0.0896	3 (21.4%)	0	1 (7.1%)	0	0.7	-	FSV & MSV
RP1	0.0896	3 (21.4%)	0	2 (14.3%)	0	2.2	-	MSV
RPS3A	0.0896	3 (21.4%)	0	1 (7.1%)	1 (11.1%)	0.1	-	MSV
SCAF1	0.0896	3 (21.4%)	0	2 (14.3%)	1 (11.1%)	0.8	0.006	MSV
RP1	0.0896	3 (21.4%)	0	2 (14.3%)	0	0	-	MSV
SLC26A10	0.0896	3 (21.4%)	0	1 (7.1%)	1 (11.1%)	0.6	0.009	MSV & SGV
STAB1	0.0896	3 (21.4%)	0	2 (14.3%)	1 (11.1%)	1.3	-	MSV
TRPV3	0.0896	3 (21.4%)	0	3 (21.4%)	0	0.1	-	MSV
WWOX	0.0896	3 (21.4%)	0	2 (14.3%)	1 (11.1%)	0	-	MSV
ZBBX	0.0896	3 (21.4%)	0	2 (14.3%)	1 (11.1%)	0.9	0.018	MSV & SRV & SGV

The table includes genes associated with the EGFR amplification group with a *P*-value lower than 0.10 (Fisher exact test). The frequency of the significantly mutated genes was also contrasted using the TGCA database. Key: missense variant: MV; frameshift variant: FSV; stop gained variant: SGV; stop loss variant: SLV; splice region variant: SRV; splice donor variant: SDV; splice acceptor variant: SA.

GPR179, like *GPR158*, relate to prostate cancer growth and progression [34], and associate with lung cancer outcomes [35]. In human gliomas, low expression of GPR158 combined with high levels of miR-449a associates with higher malignancy and poorer survival [36]. However, there are no studies relating GPR179 or its paralogs with GBM tumor biology. *USP48* (Ubiquitin Specific Peptidase 48) is a protein-coding gene related to Ubiquitin-Proteasome Dependent Proteolysis and Deubiquitination. Some studies suggest that the regulation of EGFR degradation is partly mediated by proteosomes, but the underlying mechanism remains unclear [37]. Knockdown of the *USP48* gene in GBM cells inhibits cell proliferation and the expression of *Gli1* downstream targets, which leads to repressed GBM tumorigenesis [38]. Finally,

Table 5. Mutated genes present in 30 cases of GBMs. Gene description and biological function [18]

Gene name	Location	Gene description	Biological function
GPR179	17q12	G Protein-coupledreceptor 179	G-protein
USP48	1p36.12	Ubiquitinspecificpeptidase 48	Degradation
BLK	8p23.1	Proto-oncogene,tyrosine kinase	Apoptosis
ADGB	6q24.3	Androglobin	Iron ion binding and oxygen binding
EHHADH	3q27.2	Enoyl-CoAhydratase and 3-hydroxyacyl CoAdehydrogenase	Peroxisome proliferator-activated receptor
PTPN13	4q21.3	Protein tyrosine phosphatase, non-receptor type 13	Cell growth, Apoptosis modulation and PI metabolism
NPY4R	10q11.2	Neuropeptide Y receptor Y4	G protein-coupled receptor activity
DNHD1	11p15.4	Dyneinheavy chaindomain 1	Microtubule motor activity
MKI67	10q26.2	Markerof proliferation Ki-67	Cellularproliferation
ZNF280D	15q21.2	Zinc fingerprotein 280D	Structural role
CES5A	16q12.2	Carboxylesterase 5A	transcriptionfactork
CHMP6	17q25.3	Chargedmultivesicularbody	Degradation surface receptors. Biosynthesis of endosomes
MROH6	8q24.3	Maestro heat like repeat family	Transcription factor
PKHD1	6p12.3	Fibrocystin/Polyductin	transcription factork
RAB44	6p21.2	RAS oncogenefamily	Metabolism of proteins. GTPase activity
SYNE1	6q25.2	Spectrinrepeat containing nuclear	Cell cycle, centrosome migration
ARSH	Xp22.33	Arylsulfatasefamilymember H	Metabolism of proteins and sphingolipid metabolism
BMP2K	4q21.21	BMP2 inducible kinase	Transcriptional misregulation
CFAP61	20p11.23	Cilia and flagella associated	Oxidoreductaseactivity
DEUP1	11q21	Deuterosomeassemblyprotein 1	Centriolebiogenesis
EPHB6	7q34	EPH receptor B6	Celladhesion and migration
ERAP1	5q15	Endoplasmicreticulumaminopeptidase 1	TNFR1 pathway. Metallopeptidaseactivity
FLVCR2	14q24.3	Feline leukemia virus subgroup C cellular receptor	Transmembrane protein. Calcium transporter
KALRN	3q21.1	KalirinRhoGEFkinase	Rho GTPase family members, neuronal shape, growth, and plasticity
KIF13A	6p22.3	Kinesin familymember 13A	ATPase activity and microtubule motor
LPA	6q25.3	Lipoprotein (A)	Integrin signaling and Lipoprotein metabolism
PPIG	2q31.1	Peptidylprolylisomerase G	Translational control
RP1	8q11.23	Axonemalmicrotubuleassociated	Microtubulebinding.
RPS3A	4q31.3	Ribosomalprotein S3A	Structuralconstituent of ribosome
SCAF1	19q13.3	SR-Related CTD associated F 1	Pre-mRNAsplicing
SLC26A10	12q13.3	Solutecarrier Family 26	Transmembrane transporter activity
STAB1	3p21.1	Stabilin 1	Angiogenesis
TRPV3	17p13.2	Transient receptor potential cation channel subfamily V	CREB Pathway and ion channel transport
WWOX	16q23.1	WW Domaincontainingoxidoreductase	Apoptosis. role in tumor necrosis factor (TNF)-mediated cell death
ZBBX	3q26.1	Zinc finger B-Box domain	Structural role

BLK encodes a nonreceptor tyrosine kinase of the Src kinase family of proto-oncogenes typically involved in cell proliferation and differentiation. According to recent reports, KLF4 could activate BLK via binding at mCpG in enhancer regions, suggesting a role for *BLK* in brain tumor cell migration and invasion [39].

We found mutations in *ADGB, EHHADH,* and *PTPN13* genes only in *EGFR*-no-amplified GBM tumors. This association was no concomitant to a specific infiltration pattern. The first of these genes, *ADGB* (Androglobin), is a protein-coding gene with a suggested oncogenic role in

glioma. *ADGB* knockdown inhibited proliferation and increased apoptosis of glioma cell lines [40]. On the other hand, *EHHADH* (Enoyl-CoA Hydratase And 3-Hydroxyacyl CoA Dehydrogenase) encodes one of the four enzymes of the peroxisomal beta-oxidation pathway [41] and associates with a high risk of metastasis in hepatocellular carcinoma [42]. Finally, the protein encoded by *PTPN13* is a member of the tyrosine phosphatases (PTP) family. PTPs are signaling molecules that regulate a variety of cellular processes including cell growth, cell differentiation, mitotic cycle, and oncogenic transformation. PTPN13 overexpression significantly
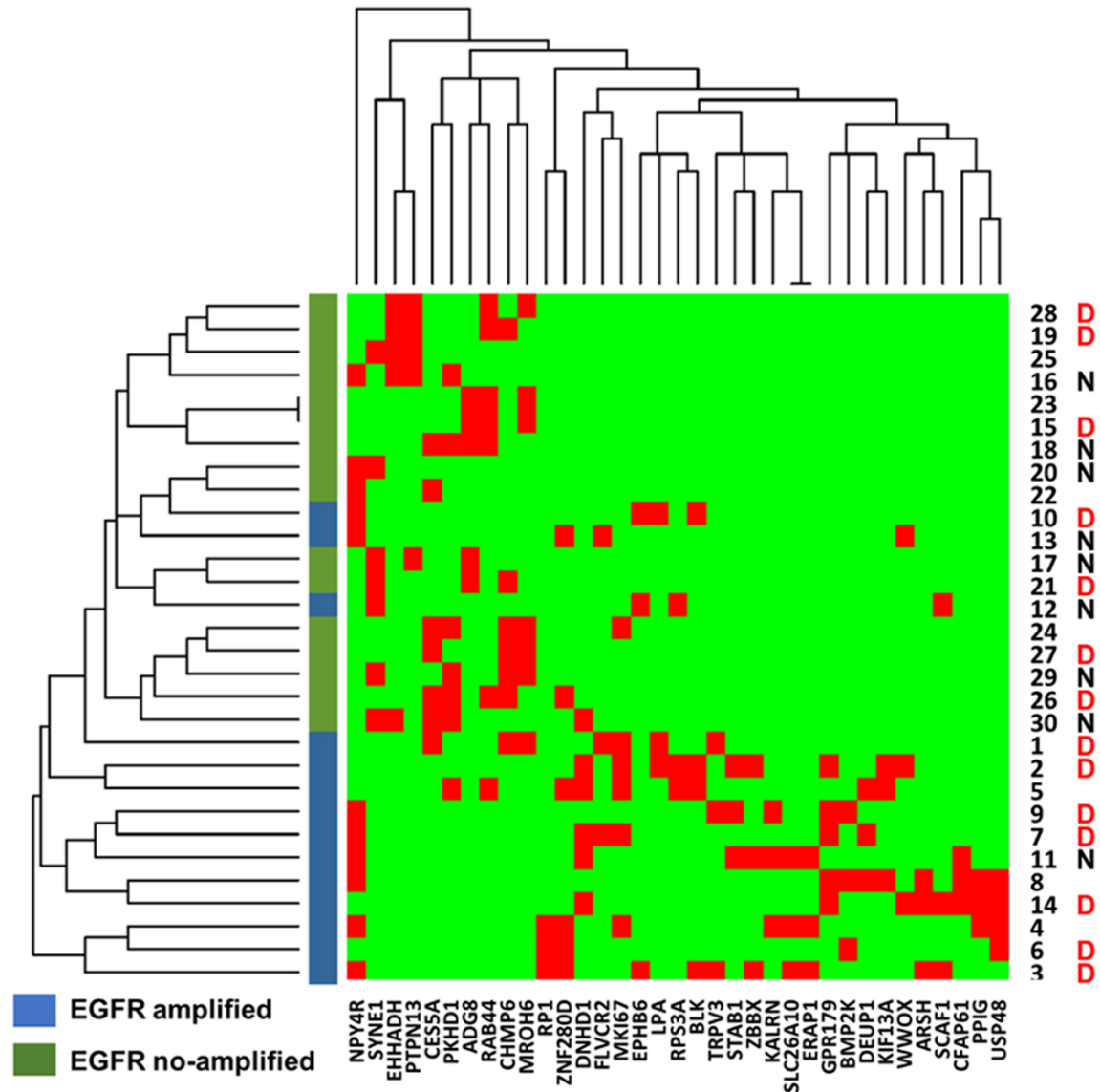
Figure 2. HCA analysis of 30 GBM samples (rows) according to the mutational state of the 36 genes (columns) significantly associated with EGFR amplification. Samples have been colored in green (*EGFR*-no-amplified) or blue (*EGFR*-amplified) for easier visualization of groups. D (in red) and N (in black) after the sample number indicate a diffuse or nodular infiltrative pattern, respectively. Sample 16 (*IDH1* mutant) is enclosed in a black rectangle.

inhibits the progression of hepatocarcinoma cells, possibly by inhibiting epithelial-mesenchymal transition through inactivation of the EGFR/ERK signaling pathway [43]. PTPN13 expression is specifically upregulated in GBM tissue and knockdown of *PTPN13* in GBM cell lines produces increased FAS-mediated apoptosis [44].

PCA shows clear discrimination of the two groups based on the genes detected and revealed that mutations in *SYNE1, DNHD1,*

PTPN13, and *MKI67* genes contribute the most to discrimination of *EGFR* amplification status. *MKI67* and *DNHD1* appear preferentially mutated in *EGFR*-amplified GBM whereas *PTPN13* and *SYNE1* in *EGFR*-no-amplified GBMs. Some of these mutated genes also show association with infiltrative patterns. Mutated *MKI67* is mostly present in diffuse infiltration patterns whereas *SYNE1* mutated samples show a nodular infiltration pattern. *SYNE1* (Spectrin Repeat Containing Nuclear Envelope Protein 1) is a protein-coding gene related to meiosis, mitosis,
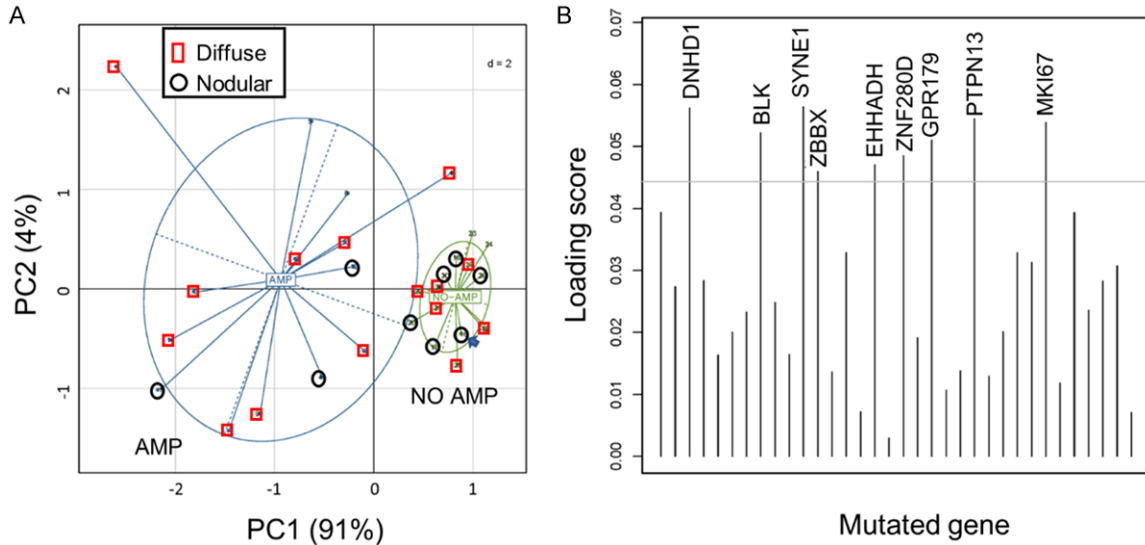
Figure 3. Scores plot (A) and loadings plot (B) for PCA and subsequent discriminant analysis for all 30 GBM samples based on EGFR amplification status. (A) Scores plot for PC1 and PC2. The x-axis is the PC1 (90% of variance explained) and the y-axis is the PC2 (5% additional variance explained). The centroids for each group (*EGFR*-amplified, AMP, blue color, and *EGFR*-no-amplified, NO-AMP, green color) are represented as rectangles. Samples with a diffuse infiltration pattern are surrounded by red squares whereas samples with a nodular infiltration pattern are surrounded by black circles. Sample 16 (*IDH1* mutant) is marked with an arrow. Samples in the *EGFR*-amplified group show larger dispersion because of higher mutational rates. (B) Loadings plot for the orthonormal mutated genes used in the discriminant analysis of principal components to discriminate between *EGFR*-amplified and *EGFR*-no-amplified samples. Higher loadings correspond to more discriminative power.

and cell cycle. *SYNE1* plays some role in inhuman GBM progression and survival [45, 46] and shows mutations in 6% of TCGA GBM cases significantly co-occurring with *EGFR* mutations. *DNHD1* (Dynein Heavy Chain Domain 1) encodes a protein related to microtubule motor activity with mutations observed in pancreatic squamous cell carcinoma [47]. *MKI67* encodes a non-histone nuclear protein (Ki-67) that is associated with and may be necessary for cellular proliferation. Mutations of this gene are present in different types of cancer [28] and Ki-67 expression is broadly used as a diagnostic marker in various cancers including GBM [48, 49]. There is a poor understanding of the correlation between the expression of Ki-67 and overexpression of EGFR in GBMs and the data are controversial [50, 51]. Our study shows that 6 out of 30 GBM have mutations in this gene. Five of them are *EGFR*-amplified tumors, exhibit Ki-67 indexes higher than 15%, and have a diffuse infiltrative pattern suggesting an increase of function mutation.

A limitation of our study is the limited number of cases given the wide heterogeneity of GBM. Although GBM is proposed to display higher morphological than genomic heterogeneity [52], recent NGS analysis demonstrates the wide variety of genetic changes associated with GBM. Moreover, the complexity of the data and the vast array of processes potentially involved hampers interpretation of the functional meaning of our results. Despite the low number of cases included, the use of fresh-tissue tumor samples in our study minimizes the artifactual results that paraffin-embedded specimens could produce. On the other hand, although GBM is the most frequent of the malignant brain tumors, it is rather difficult to collect a large set of cases because of the relatively low incidence. The focus of our study was to study association with genetically (*EGFR* amplification) and phenotypically (infiltration) relevant features for detecting trends and generating new hypotheses.

Concluding remarks

In conclusion, this study provides a global description of the mutational status of GBM in relationship to *EGFR* amplification and relevant phenotypic traits. Our work establishes the basis for different mutational portraits of GBM

related to well-established features like *EGFR* amplification and tumor infiltration. The results also show a trend and potential association between mutations in the *MIKI67* gene, Ki-67 index, and diffuse infiltrative phenotypes. Overall, these findings may help in opening new hypotheses on GBM tumor biology and identifying new potential therapeutic targets.

Disclosure of conflict of interest

Juan Carlos Triviño is an employee for Sistemas Genomicos SA. Daniel Monleón is a member of the editorial board of the American Journal of Cancer Research.

Address correspondence to: Daniel Monleón, Department of Pathology, University of Valencia, Avda Blasco Ibáñez 15, Valencia, Spain. Tel: +34-963864145; E-mail: Daniel.monleon@uv.es

References

[1] Louis DN, Perry A, Reifenberger G, Deimling A, Frigarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P and Ellison DW. The 2016 world organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 2016; 131: 803-820.
[2] Claes A, Idema AJ and Wesseling P. Diffuse glioma growth: a guerilla war. Acta Neuropathol 2007; 114: 443-458.
[3] Monticone M, Daga A, Candiani S, Romeo F, Mirisola V, Viaggi S, Melloni I, Pedemonte S, Zona G, Giaretti W, Pfeffer U and Castagnola P. Identification of a novel set of genes reflecting different in vivo invasive patterns of human GBM cells. BMC Cancer 2012; 12: 358.
[4] Godlewski J, Ferrer-Luna R, Rooj AK, Mineo M, Ricklefs F, Takeda YS, Nowicki MO, Salińska E, Nakano I, Lee H, Weissleder R, Beroukhim R, Chiocca EA and Bronisz A. MicroRNA signatures and molecular subtypes of glioblastoma: the role of extracellular transfer. Stem Cell Reports 2017; 8: 1497-1505.
[5] López-Ginés C, Navarro L, Muñoz-Hidalgo L, Buso E, Morales JM, Gil-Benso R, Gregori-Romero M, Megías J, Roldán P, Segura-Sabater R, Almerich-Silla JM, Monleón D and Cerdá-Nicolás M. Association between epidermal growth factor receptor amplification and ADP-ribosylation factor 1 methylation in human glioblastoma. Cell Oncol (Dordr) 2017; 40: 389-399.
[6] Huncharek M and Kupelnick B. Epidermal growth factor receptor gene amplification as a prognostic marker in glioblastoma: results of a meta-analysis. Oncol Res 2000; 12: 107-112.
[7] López-Ginés C, Gil-Benso R, Ferrer-Luna R, Benito R, Serna E, González-Darder J, Quilis V, Monleón D, Celda B and Cerdá-Nicolas M. New pattern of EGFR amplification in glioblastoma and the relationship of gene copy number with gene expression profile. Mod Pathol 2010; 23: 856-865.
[8] Lassman AB, Aldape KD, Ansell PJ, Bain E, Curran WJ, Eoli M, French PJ, Kinoshita M, Looman J, Mehta M, Muragaki Y, Narita Y, Ocampo C, Roberts-Rapp L, Song M, Vogelbaum MA, Walenkamp AME, Wang TJC, Zhang P and van den Bent MJ. Epidermal growth factor receptor (EGFR) amplification rates observed in screening patients for randomized trials in glioblastoma. J Neurooncol 2019; 144: 205-210.
[9] Rao R and James C. Altered molecular pathways in gliomas: an overview of clinically relevant issues. Semin Oncol 2004; 3: 595-604.
[10] Ohgaki H and Kleihues P. Genetic alterations and signaling pathways in the evolution of gliomas. Cancer Sci 2009; 100: 2235-2241.
[11] Aldape K, Zadeh G, Mansouri S, Reifenberger G and von Deimling A. Glioblastoma: pathology molecular mechanisms and markers. Acta Neuropathol 2015; 129: 829-848.
[12] Jang H and Lee H. Identification of cancer driver genes in focal genomic aberrations from whole-exome sequencing data. Bioinformatics 2018; 34: 519-521.
[13] Coudray A, Battenhouse AM, Bucher P and Iyer VR. Detection and benchmarking of somatic mutations in cancer genomes using RNA-seq data. PeerJ 2018; 31: e5362.
[14] Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, Grimsby J, Sougnez C, Muller F, Hu J, Sloan AE, Cohen ML, Van Meir EG, Scarpace L, Laird PW, Weinstein JN, Lander ES, Gabriel S, Getz G, Meyerson M, Chin L, Barnholtz-Sloan JS and Verhaak RG. Whole-genome and multi-sector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tu-
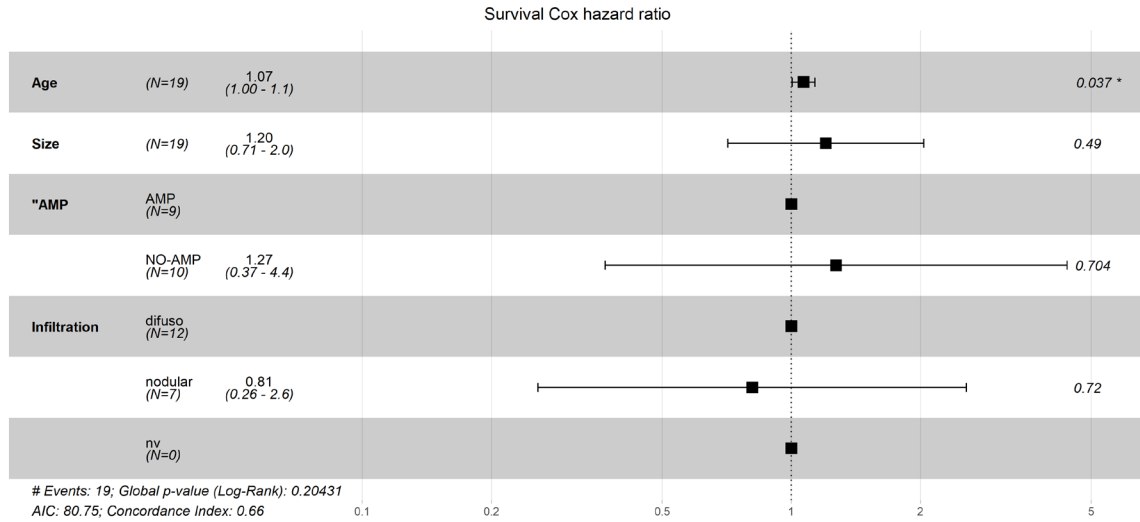
mor evolution. Genome Res 2015; 25: 316-327.

[15] Peng S, Dhruv H, Armstrong B, Salhia B, Legendre C, Kiefer J, Parks J, Virk S, Sloan AE, Ostrom QT, Barnholtz-Sloan JS, Tran NL and Berens ME. Integrated genomic analysis of survival outliers in glioblastoma. Neuro Oncol 2017; 19: 833-844.

[16] Xu G, Zheng H and Li JY. Next-generation whole exome sequencing of glioblastoma with a primitive neuronal component. Brain Tumor Pathol 2019; 36: 129-134.

[17] Ülgen E, Can Ö, Bilguvar K, Oktay Y, Akyerli CB,Danyeli AE, Yakıcıer MC, Sezerman OU, Pamir MN and Özduman K. Whole exome sequencing-based analysis to identify DNA damage repair deficiency as a major contributor to gliomagenesis in adult diffuse gliomas. J Neurosurg 2019; 132: 1435-1446.

[18] Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754-1760.

[19] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-1303.

[20] Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK and Ding L. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 2009; 25: 2283-2285.

[21] Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Garcia Giron C, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Gonzalez Martinez J, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, Ilsley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR and Flicek P. Ensembl 2021. Nucleic Acids Res 2021; 49: D884-D891.

[22] 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and Abecasis GR. A global reference for human genetic variation. Nature 2015; 526: 68-74.

[23] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P and Cunningham F. The ensembl variant effect predictor. Genome Biol 2016; 17: 122.

[24] Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009; 4: 1073-1081.

[25] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS and Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods 2010; 7: 248-249.

[26] Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 2008; 24: 1403-1405.

[27] Jombart T and Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 2011; 27: 3070-3071.

[28] Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008; 455: 1061-1068.

[29] Yu G, Wang L, Han Y and He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-287.

[30] Yuan F, Xu Z, Yang M, Wei Q, Zhang Y, Yu J, Zhi Y, Liu Y, Chen Z and Yang J. Overexpressed DNA polymerase iota regulated by JNK/c-Jun contributes to hypermutagenesis in bladder cancer. PLoS One 2013; 8: e69317.

[31] Sa JK, Choi SW, Zhao J, Lee Y, Zhang J, Kong DS, Choi JW, Seol HJ, Lee JI, Iavarone A, Rabadan R and Nam DH. Hypermutagenesis in untreated adult gliomas due to inherited mismatch mutations. Int J Cancer 2019; 144: 3023-3030.

[32] Lee A, Arasaratnam M, Chan DLH, Khasraw M, Howell VM and Wheeler H. Anti-epidermal growth factor receptor therapy for glioblastoma in adults. Cochrane Database Syst Rev 2020; 5: CD013238.

[33] Wang Z. Transactivation of epidermal growth factor receptor by G protein-coupled receptors: receptor progress challenges and future research. Int J Mol Sci 2016; 17: 95.

[34] Patel N, Itakura T, Jeong S, Liao CP, Roy-Burman P, Zandi E, Groshen S, Pinski J, Coetzee GA, Gross ME and Fini ME. Expression and functional role of orphan receptor GPR158 in prostate cancer growth and progression. PLoS One 2015; 10: e0117758.

[35] Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, Hao D and Sun J. Potential signature of eight

long non-coding RNAs predicts survival in patients with non-small cell lung cancer. J Transl Med 2015; 13: 231.

[36] Li N, Zhang Y, Sidlauskas K, Ellis M, Evans I, Frankel P, Lau J, El-Hassan T, Guglielmi L, Broni J, Richard-Loendt A and Brandner S. Inhibition of GPR158 by microRNA-449a suppresses neural lineage of glioma stem/progenitor cells and correlates with higher glioma grades. Oncogene 2018; 37: 4313-4333.

[37] Alwan HA, van Zoelen EJ and van Leeuwen JE. Ligand-induced lysosomal epidermal growth factor receptor (EGFR) degradation is preceded by proteasome-dependent EGFR de-ubiquitination. J Biol Chem 2003; 278: 35781-35790.

[38] Zhou A, Lin K, Zhang S, Ma L, Xue J, Morris SA, Aldape KD and Huang S. Gli1-induced deubiquitinase USP48 aids glioblastoma tumorigenesis by stabilizing Gli1. EMBO Rep 2017; 18: 1318-1330.

[39] Oyinlade O, Wei S, Kammers K, Liu S, Wang S, Ma D, Huang ZY, Qian J, Zhu H, Wan J and Xia S. Analysis of KLF4 regulated genes in cancer cells reveals a role of DNA methylation in promoter-enhancer interactions. Epigenetics 2018; 13: 751-768.

[40] Huang B, Lu YS, Li X, Zhu ZC, Li K, Liu JW, Zheng J and Hu ZL. Androglobin knockdown inhibits growth of glioma cell lines. Int J Clin Exp Pathol 2014; 7: 2179-2184.

[41] Hoefler G, Forstner M, McGuinness MC, Hulla W, Hiden M, Krisper P, Kenner L, Ried T, Lengauer C, Zechner R, Moser HW and Chen GL. cDNA cloning of the human peroxisomalenoyl-CoA hydratase: 3-hydroxyacylCoA dehydrogenase bifunctional enzyme and localization to chromosome 3q26.3-3q28: a a free left Alu Arm is inserted in the 3' noncoding region. Genomics 1994; 19: 60-67.

[42] Chen P, Wang F, Feng J, Zhou1 R, Chang Y, Liu J and Zhao Q. Co-expression network analysis identified six hub genes in association with metastasis risk and prognosis in hepatocellular carcinoma. Oncotarget 2017; 8: 48948-48958.

[43] Zhan H, Jiang J, Luo C, Sun Q, Ke A, Sun C, Hu J, Hu Z, Hu B, Zhu K, Fan J, Zhou J and Huang X. Tumour-suppressive role of PTPN13 in hepatocellular carcinoma and its clinical significance. Tumour Biol 2016; 37: 9691-9698.

[44] Navis AC, van den Eijnden M, Schepens JT, van Huijsduijnen RH, Wesseling P and Hendriks W. Protein tyrosine phosphatases in glioma biology. Acta Neuropathol 2010; 119: 157-175.

[45] David L and Karchin MR. Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. Cancer Res 2011; 71: 4550-4561.

[46] Serão NV, Delfino KR, Southey B, Beever JE and Rodriguez-Zas SL. Cell cycle and aging morphogenesis and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. BMC Med Genomics 2011; 4: 49.

[47] Xu MD, Liu SL, Feng YZ, Liu Q, Shen M, Zhi Q, Liu Z, Gu DM, Yu J, Shou LM, Gong FR, Zhu Q, Duan W, Chen K, Zhang J, Wu MY, Tao M and Li W. Genomic characteristics of pancreatic squamous cell carcinoma an investigation by using high throughput sequencing after in-solution hybrid capture. Oncotarget 2017; 8: 14620-14635.

[48] Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, Morozova O, Newton Y, Radenbaugh A, Pagnotta SM, Anjum S, Wang J, Manyam G, Zoppoli P, Ling S, Rao AA, Grifford M, Cherniack AD, Zhang H, Poisson L, Carlotti CG Jr, Tirapelli DP, Rao A, Mikkelsen T, Lau CC, Yung WK, Rabadan R, Huse J, Brat DJ, Lehman NL, Barnholtz-Sloan JS, Zheng S, Hess K, Rao G, Meyerson M, Beroukhim R, Cooper L, Akbani R, Wrensch M, Haussler D, Aldape KD, Laird PW and Gutmann DH; TCGA Research Network, Noushmehr H, Iavarone A and Verhaak RG. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. Cell 2016; 164: 550-563.

[49] Henker C, Kriesen T, Schneider B, Glass Ä, Scherer M, Langner S, Erbersdobler A and Piek J. Correlation of Ki-67 index with volumetric segmentation and its value as a prognostic marker in glioblastoma. World Neurosurg 2019; 125: e1093-e1103.

[50] Wong E, Nahar N, Hau E, Varikatt W, Gebski V, Ng T, Jayamohan J and Sundaresan P. Cutpoint for Ki-67 proliferation index as a prognostic marker for glioblastoma. Asia Pac J Clin Oncol 2019; 15: 5-9.

[51] Barresi V, Buttarelli FR, Vitarelli E, Arcella A, Antonelli M and Giangaspero F. Caveolin-1 expression in diffuse gliomas: correlation with the proliferation index epidermal growth factor receptor p53 and 1p/19q status. Hum Pathol 2009; 40: 1738-1746.

[52] Furnari FB, Cloughesy TF, Cavenee WK and Mischel PS. Heterogeneity of epidermal growth factor receptor signalling networks in glioblastoma. Nat Rev Cancer 2015; 15: 302-310.

Survival Cox hazard ratio

Age	(N=19)	1.07 (1.00 - 1.1)			0.037 *
Size	(N=19)	1.20 (0.71 - 2.0)			0.49
"AMP	AMP (N=9)				
	NO-AMP (N=10)	1.27 (0.37 - 4.4)			0.704
Infiltration	difuso (N=12)				
	nodular (N=7)	0.81 (0.26 - 2.6)			0.72
	nv (N=0)				

Events: 19; Global p-value (Log-Rank): 0.20431
AIC: 80.75; Concordance Index: 0.66

0.1   0.2   0.5   1   2   5

Figure S1. Cox hazard ratios models for survival time vs EGFR amplification status and infiltration pattern adjusted by patient age and tumor size.

Table S1. Common genetic markers for GBM (EGFR amplified samples have shadow background)

Case	*EGFR vIII*	*PTEN*	*CDKN2A/CDKN2B*	*TP53*	*MGMT %*
1	mut	LOH	N/N		15,5
2	mut	OH	OH/OH		23,1
3		N	LOH/LOH		31,8
4		N	N/N		22,3
5		N	LOH/LOH		20,5
6	mut	N	LOH/LOH		13,8
7		N	LOH/N		19,9
8		N	N/N		30,7
9		LOH	LOH/LOH		nv
10		N	N/N		18,5
11		LOH	LOH/LOH		12,1
12		N	N/N		nv
13		N	N/N	p,C141R	47,7
14		N	LOH/LOH		30,1
15		LOH	LOH/LOH		15,6
16		N	N/N	p,R273C	nv
17		N	N/N		11,8
18		N	N/N		12,3
19		N	N/N		27,5
20		N	LOH/LOH		14,1
21	mut	N	LOH/LOH		36,3
22		LOH	N/N		25,9
23		N	N/N		nv
24		N	N/N		16,5
25		N	N/N		18,9
26		N	N/N		27,9
27		N	N/N		nv
28		LOH	N/N		nv
29		N	LOH/LOH		16,1
30		N	N/N		nv

mut: mutated; LOH: Loss of number of copies in heterozygosity; OH: Loss of number of copies in homocigosity; N: number of copies normal. Nv: not available.