Glioblastoma multiforme: a multi-omics analysis of driver genes and tumour heterogeneity

Gabriel Emilio Herrera-Oropeza1,2,†, Carla Angulo-Rojo2,†, Santos Alberto Gástelum-López4, Alfredo Varela-Echavarría1, Maribel Hernández-Rosales5 and Katia Aviña-Padilla1,5

1Instituto de Neurobiología, Universidad Nacional Autónoma de México, Querétaro, Mexico
2Centre for Developmental Neurobiology, Institute of Psychiatry, Psychology, and Neuroscience, King’s College London, London, UK
3Centro de Investigación Aplicada a la Salud, Facultad de Medicina, Universidad Autónoma de Sinaloa, Culiacán, Sinaloa, Mexico
4Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional, Instituto Politécnico Nacional, Guasave, Sinaloa, Mexico
5Centro de Investigación y de Estudios Avanzados del IPN, Unidad Irapuato, Guanajuato, Mexico

KA-P, 0000-0002-6338-7930

Glioblastoma (GBM) is the most aggressive and common brain cancer in adults with the lowest life expectancy. The current neuro-oncology practice has incorporated genes involved in key molecular events that drive GBM tumorigenesis as biomarkers to guide diagnosis and design treatment. This study summarizes findings describing the significant heterogeneity of GBM at the transcriptional and genomic levels, emphasizing 18 driver genes with clinical relevance. A pattern was identified fitting the stem cell model for GBM ontogenesis, with an upregulation profile for MGMT and downregulation for ATRX, H3F3A, TP53 and EGFR in the mesenchymal subtype. We also detected overexpression of EGFR, NES, VIM and TP53 in the classical subtype and of MKi67 and OLIG2 genes in the proneural subtype. Furthermore, we found a combination of the four biomarkers EGFR, NES, OLIG2 and VIM with a remarkable differential expression pattern which confers them a strong potential to determine the GBM molecular subtype. A unique distribution of somatic mutations was found for the young and adult population, particularly for genes related to DNA repair and chromatin remodelling, highlighting ATRX, MGMT and IDH1. Our results also revealed that highly lesioned genes undergo differential regulation with particular biological pathways for young patients. This multi-omic analysis will help delineate future strategies related to the use of these molecular markers for clinical decision-making in the medical routine.

1. Introduction

Glioblastoma multiforme (GBM) is the most frequent and aggressive deadly primary brain tumour in adults, accounting for approximately 82% of all malignant gliomas [1]. Although it can affect children, its incidence rises with age. GBM tumours are characterized by increased cell proliferation, aggressive invasion, active angiogenesis and a remarkable genetic heterogeneity [2]. Histologically, tumours display a high morphological variability as they contain pleomorphic and multinucleated cells with high mitotic activity, show microvascular proliferation, undergo severe and characteristic endothelial hyperplasia, contain intravascular microthrombi, and extensive necrosis of an ischaemic or pseudo-emphalized nature. The multiforme denomination of GBM tumours is due to the
diverse and heterogeneous microenvironments that parallel their multiple histological patterns and cytological features.

According to their ontogeny, most GBMs are primary tumours that develop de novo in the absence of previous neoplasia. Primary GBMs are highly aggressive and invasive, tend to extend to both cerebral hemispheres, or are bilateral, and they are most commonly manifested in elderly patients. Secondary GBMs, in contrast, are located in the frontal lobe and develop mainly in younger patients suffering from anaplastic astrocytoma or low-grade astrocytoma, presenting a much better prognosis [3]. Recent reports have determined that primary and secondary glioblastomas have distinct genetic alterations related to particular biological pathways [1,3,4], suggesting they require different therapeutic approaches. Hence, from the clinical perspective, discerning between primary and secondary GBM is highly relevant [2]. Usually, primary GBMs present overexpression and gene amplification of epidermal growth factor receptor (EGFR) and mutations in cyclin-dependent kinase inhibitor 2A (CDKN2A/p16INK4A) and phosphatase tensin homologue (PTEN) genes. Molecular biomarkers of secondary GBM include mutations in tumour protein 53 (TP53) and isocitrate dehydrogenase-1 (IDH1) genes, which correlate strongly with O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation [3,5].

Initiation and progression of GBM tumorigenesis are related to genetic and epigenetic alterations and molecular subtypes of GBM have unique transcriptional profiles. Based on expression features, GBM tumours were originally classified into four subtypes: neural, proneural, classical (proliferative) and mesenchymal [6], a scheme that has been recently revised using transcriptomic information. The improved classification eliminates the neural subtype and considers tumours of this molecular type as containing normal brain tissue contamination [7,8]. GBM molecular subtypes are also associated with different spatial zones, heterogeneity and aggressiveness of the tumour [9].

GBMs belonging to the proneural subtype have alterations in TP53, PDGFR, PIK3CA and IDH1 genes [10,11]. The classical subtype, also known as a proliferative subtype, has been associated with high levels of cell proliferation and upregulation of EGRF [12]. Mesenchymal GBMs show overexpression of mesenchymal and astrocytic markers (CD44, and MERTK) and downregulation of neurofibromatosis type-1 (NF1) and upregulation of chitinase 3 like 1 (CHI3L1/YKL-40) and MET are frequently observed [10]. While the proneural subtype has been mostly reported in younger patients and is associated with a favourable prognosis, the mesenchymal and the classical subtypes are usually linked to more aggressive high-grade gliomas that appear in adult or elderly life.

Recent advances employing next-generation sequencing have led to a better insight into the molecular biology of gliomas contributing potential markers for better diagnosis and new approaches to finding specific treatment strategies [13]. GBM remains an incurable deadly disease with an abysmal prognosis that has not significantly shown improvement, causing an enormous individual and societal burden. Thus, there is a need for tumour-specific drug targets and pharmacological agents to inhibit cell migration, dispersal and angiogenesis [7]. For a current detailed review, see [14].

In the last years, the clinical relevance of GBM heterogeneity has been highlighted [15]. This particular feature makes this type of cancer one of the most challenging to treat and consists of inter-tumour and intra-tumour feature variations. Inter-tumour heterogeneity refers to GBMs from different patients with altered and differing genotypes and phenotypes related to diverse etiological and environmental factors. On the other hand, intra-tumour heterogeneity refers to the presence of multiple and different cell subpopulations within the same tumour, defining its topology and architecture [16]. The comprehensive genomic classification of GBM paves the way for an improved understanding of tumour progression, which in the future may result in personalized therapy. Hence, there is an urgent need to further our knowledge of tumour heterogeneity as it will help design better therapies against GBM and tumour recurrence. Based on a multi-omic analysis, in this study we describe the heterogeneity of GBM at the transcriptional and genomic levels, with emphasis on driver genes currently used as biomarkers. For that purpose, from 60 clinical reports, we selected and analysed 18 driver genes that have shown deregulated behaviour in patient samples. Using bioinformatics pipelines and the TCGA database, we examined their mRNA expression in the different GBM molecular subtypes and the presence of somatic mutations linked to possible disruption of protein function. We hope that the new knowledge generated in this study leads to novel therapeutic intervention strategies.

2. Material and methods

2.1. Data mining for selection of GBM driver genes currently used as genomic markers in the clinic

The literature research was performed using a systematic approach to identify GBM biomarkers in the routine clinical diagnosis that yielded differential transcriptomics or genomic profiles on tumour samples. Using a combination of three terms: (1) ‘Glioblastoma’, (2) ‘Clinical’ and (3) ‘Case’, a total of 3238 clinical reports were found using the BVS (1548), Cochrane (0), Karger (271) and PubMed (1419) databases. Clinical reports were identified and selected by title and summary. All articles were evaluated using the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (http://prisma-statement.org/) to determine their eligibility, resulting in 60 reports, as described in electronic supplementary material, file 1. The search was conducted in July 2020 and focused on studies published in June 2005–June 2020.

2.2. Data source for the gene expression analysis

Eighteen genes were found to be involved in GBM diagnosis during the neuro-oncology clinical routine and evaluated for their mRNA expression analysis using data from the Glioblastoma BioDiscovery Portal (GBM-BioDP) https://gbm-biodp.nih.gov [17]. The gene expression data include normalized (level 3) data from Verhaak 840 Core, a filtered dataset composed of three microarray platforms: HT_HG-U133A (488 patient samples/612 042 features), HuEx-1_0-st-v2 (437 patient samples/618 631 features), and AgilentG4502A_07_1/2 (101 + 396 patient samples/617 813 features). GBM molecular subtypes were assigned according to the Verhaak classification [11].

2.3. Determination of gene expression of GBM driver genes

We classified the mRNA expression analysis of the driver genes according to their biological ontology into three groups: (1) DNA repair and chromatin remodelling, (2) cytoskeleton and cellular
proliferation, and (3) tumour suppressors genes. Using Python scripts (https://github.com/kap8416/GBM-META-ANALYSIS-OF-DRIVER-GENES), we determined the average and standard deviation of the z-score expression values for all patient results and classified them into the molecular GBM subtypes (Classic, Proneural, Mesenchymal), and grouped them by their corresponding biological gene ontology group.

We examined the mRNA expression patterns of the driver genes clustered by patient subgroups taking their age into account. From a total of 166 patients, three subgroups were created: 10–29-year-old patients (young subgroup), 30–59 years old (adult subgroup) and 60–89 years old (elderly subgroup). The average of the z-score values among the patient subgroups was clustered into the molecular GBM subtypes.

Finally, the Mann–Whitney test was used to examine the statistical difference in the mRNA expression z-scores between GBM molecular subtypes and patient subgroups and between each gene and GBM subtypes. Multiplicity adjustments were performed on the obtained p-values by using the Benjamini–Hochberg method. Statistical significance for the test was set to \( p < 0.05 \).

2.4. Data source for somatic mutations of GBM
Genomic data from 588 patients for the 18 genes previously identified as molecular markers was downloaded from the NIH website https://portal.gdc.cancer.gov/ using the following restriction criteria: Primary site: brain; Program: TCGA; Project: TCGA-GB; Disease Type: glioblastoma; Sample type: primary tumour; Clinical age of diagnosis: 10–29 years, 30–59 years and 60–89 years.

2.5. Determination of mutations in GBM and driver genes
Using Python scripts (https://github.com/kap8416/GBM-META-ANALYSIS-OF-DRIVER-GENES), the number of mutations per gene in the TCGA-GBM project was determined by calculating the amount of different genomic DNA changes reported in each gene. Subsequently, the relative percentage of mutations per chromosome was calculated by taking into account the total length (base pairs) of the respective chromosome. Substitutions, deletions and insertions were identified, then the number of nucleotide changes occurring in all genes was determined, and their distribution was compared to the distribution of those present in the driver genes. Moreover, the total number of mutations per gene and the genome location of the somatic mutations were compared among patient subgroups according to their age. Finally, the protein phenotype impact values (polyphen) of all the canonical missense variant consequences of the driver genes in the TCGA-GBM project were determined, analysed and compared between patient subgroups clustered by age.

2.6. Functional enrichment for driver genes, unique or shared pathways
GO enrichment analysis was performed using the Metascape tool (http://metascape.org/). We then used the meta-analysis workflow to compare the driver gene pathways with those of the highly mutated genes to identify unique or shared biological pathways in which they are involved. Using Python scripts, the top 50 mutated genes observed in the TCGA-GBM project were clustered by age group. Those genes were selected and analysed by their GO enriched terms. Finally, affected genes in their protein polymorphism phenotype with more than three probably damaging consequences (PR) were clustered by the patient subgroups for the GO terms and TRRUST enrichment analysis.

3. Results
3.1. Selection of the GBM driver genes as genomic markers in the clinic
First, we aimed to identify and select the top used biomarkers in the clinic. Sixty clinical reports were found from 2005 to 2020 that were fully text reviewed (described in electronic supplementary material, E1). A total of 73 patients with GBM were characterized and described in table 1. Patient demographics consisted of 43 men and 30 women with ages ranging from four to 78 years and a mean age of 43.31. Twenty-two patients were classified as young (4–29 years), 30 adults (30–59 years) and 21 as elderly (60–78 years).

Patients underwent a biopsy procedure to evaluate the expression and mutations of biomarkers, which were the most representative genes used in clinical cases over the last 15 years. More than 80% of the clinical cases highlighted the use of a combination of 2–11 of the 18 markers. The most-reported were IDH1, GFAP, Mki67 and MGMT, followed by TP53, ATRX and EGFR.

In this systematic review, only the biomarkers with differential positive results for patient diagnosis in the clinical reports were selected for further analysis (table 1). According to their Biological Process Gene Ontology, driver genes were clustered using k-means into three groups to determine their possible role in common pathways. The first group includes ATRX, H3F3A, IDH1, MGMT and TERT driver genes related to DNA repair and involved in chromatin remodelling pathways. The second group includes the cytoskeleton and cellular proliferation-related genes EGER, FLT1 (VEGFR), BRAF, GFAP, Mki67, NES, OLIG2, PIK3CA, SMAD3, S1001A and VIM. In particular, EGER has an essential role in activating the receptor tyrosine kinase/Ras/ phosphoinositide 3-kinase RTK/RAS/PK B pathway. Alterations in this pathway disrupt the G1-S transition in the cell cycle, which is highly relevant in the progression and excessive proliferation of GBM tumour cells. The third group included tumour suppressor genes SMARCBI/INI1 and TP53 which are negative regulators of cell growth control, normally acting to inhibit tumour development.

3.2. Transcriptomics analysis of driver genes of GBM tumorigenesis
Due to the high inter- and intra-tumour heterogeneity in GBM and to gain insight into this complex process, the expression profiling pattern of the top 18 genes used as biomarkers in the clinical report systematic review was analysed using gene expression data from the Glioblastoma BioDiscovery Portal. We focused on this analysis according to the Verhaak molecular classification of GBM, which groups tumours as proneural, classical and mesenchymal [11,16]. The gene expression analysis included all data available from the GBM-BioDP, including a total of 166 patients, from which 56 were proneural, 53 classical and 57 mesenchymal subtypes. Gene expression data from each patient were available for the 18 driver genes analysed (table 2).

First, we analysed the overall profile expression pattern of each gene among GBM subtypes (table 2). For the DNA repair and chromatin remodelling genes, such as ATRX and H3F3A, we observed a tendency to a lower expression level in mesenchymal and an increased expression in proneural compared to the classic subtype. An inverse pattern was
| Reference | Database | Gender | Age | Symptoms | Type of GBM Tumor | Surgical Resection | Therapy | Tumor Recurrence | Molecular Markers (Expression) | Molecular Markers (Somatic Mutations) | Survival Time |
|-----------|----------|--------|-----|----------|-----------------|-------------------|---------|-----------------|-------------------------------|--------------------------------------|--------------|
| Dormegny et al. [35] | BVS | male | 21 | hemiparesis, seizures | primary | n.a. | n.a. | yes | n.a. | H3K27M(+), IDH R132H(−) | 3 months |
| Kajitani et al. [36] | BVS | male | 13 | headache | secondary | partial | QT, RT, TMZ | yes | GFAP(+), VIMENTIN(+), Ki67(−), MGMT(−) | ATRX(−), IDH−1 wt(−), H3K27M(−), BRAFV600E(−), TP53(−) | 4 months |
| Kajitani et al. [36] | BVS | female | 16 | seizures | secondary | partial | QT, RT, TMZ | yes | GFAP(+), ATRX(−), VIMENTIN(+), Ki67(−), MGMT(−) | IDH-1 wt(−), TP53(−), BRAFV600E(−), H3K27M(−) | 6 months |
| Kajitani et al. [36] | BVS | female | 16 | facial nerve paralysis | secondary | total | QT, RT, TMZ | no | GFAP(+), ATRX(−), VIMENTIN(+), Ki67(−) | IDH-1 wt(−), BRAFV600E(−), H3K27M(−), TP53(−) | alive |
| Kajitani et al. [36] | BVS | male | 65 | headache, personality disorder, dizziness | primary | total | n.a. | yes | GFAP(+) | IDH-1 wt(+) | 17 months |
| McGree et al. [38] | BVS | male | 57 | headache, hemiparesis | primary | partial | QT, RT, TTF | yes | GFAP(+) | VIMENTIN(+) | H3K77(+) | TP53(+) | IDH−1(−) | 25 months |
| Petzold et al. [39] | BVS | female | 28 | headache, aphasia, dizziness, nausea | primary | total | QT, RT | no | n.a. | | | | | alive |
| Prelaj et al. [40] | BVS | male | 60 | aphasia, hemiparesis | primary | total | QT, RT, TMZ | yes | GFAP(+), EGFR(+) | MKi67(+) | TP53(+) | IDH−1 wt(+) | 6 months |
| Ranjan et al. [41] | BVS | female | 51 | headache, hemiparesis | primary | total | QT, RT, TMZ | no | Ki67(+) | | | | alive |
| Ranjan et al. [41] | BVS | male | 63 | aphasia | primary | total | QT, RT, TMZ | no | Ki67(+) | | | | alive |
| Ranjan et al. [41] | BVS | male | 47 | headache, nausea, vomiting | primary | total | QT, RT, TMZ | no | Ki67(+) | | | | alive |
| Ranjan et al. [41] | BVS | male | 47 | headache, aphasia | primary | total | QT, RT, TMZ | no | Ki67(+) | | | | alive |
| Richard et al. [42] | BVS | male | 28 | seizures | primary | total | QT, RT, TMZ | no | GFAP(+), OLIG2(+), ATRX(−), MND-7(+) | MKMT(−), TP53(−), TERT(+) | alive |
| Rosen et al. [43] | BVS | female | 48 | aphasia, hemiparesis | primary | partial | QT, RT, TMZ, bevacizumab | yes | MGMT(−) | | IDH-1 wt(+) | 13 months |
| Wang et al. [44] | BVS | male | 4 | headache, hemiparesis, vomiting | primary | total | QT, RT, TMZ, nimotuzumab, ipilimumab | yes | GFAP(+), VIMENTIN(+), OLIG2(+), S−X0(+) | ATRX(−), MGMT(+), SMAD3(−), SMARCB1(+) | TP53, IDH−1(−), SMAX1(−), SMAC(−) | 8 years |
| Bartsch et al. [45] | BVS | male | 44 | hemiparesis | primary | total | QT, RT, nimotuzumab | yes | S100A(+) | | | | alive |

(Continued)
| Reference | Database | Gender | Age | Symptoms | Type of Gbm Tumor | Surgical Resection | Therapy | Tumor Recurrence | Molecular Markers (Expression) | Molecular Markers (Somatic Mutations) | Survival Time |
|-----------|----------|--------|-----|----------|-----------------|------------------|---------|-----------------|---------------------------------|-------------------------------------|--------------|
| Porto et al. [46] | BVS | Male | 72 | Headache | Primary | Partial | QT/RT | Yes | ATRX (+) | IDH-1 wt (+) | 5 months |
| Awadulla et al. [47] | BVS | Male | 60 | Aphasia, Hemiplegia | Primary | Partial | QT/RT | No | GFAP (+), VIMENTIN (+) | n.a. | 8 months |
| Gestrich et al. [48] | BVS | Male | 64 | Altered Mental Status | Primary | Total | QT/RT | Yes | GFAP (+), S100 (+) | IDH-1 wt (+) | 10 months |
| Machi et al. [49] | BVS | Female | 43 | Seizures, Memory Loss | Primary | N.a. | QT/RT TMZ | No | n.a. | IDH-1 wt (+) | 9 months |
| Watanabe et al. [50] | BVS | Female | 19 | Headache | Secondary | Total | QT/RT TMZ | Yes | OLIG2 (+), IDH-1 wt (+) | BRAF V600E (+), IDH1 R132H (−), SMARCB1 (−), SMAD2/3 (−), TP53 (−) | Alive |
| Porto et al. [46] | BVS | Male | 72 | Headache, Hemiplegia | Primary | Partial | S-ALA Fluorescence | QT/RT | Yes | ATRX (+) | IDH-1 wt (+) | 5 months |
| Awadulla et al. [47] | BVS | Male | 60 | Aphasia, Hemiplegia | Primary | Partial | QT/RT | No | GFAP (+), VIMENTIN (+) | n.a. | 8 months |
| Gestrich et al. [48] | BVS | Male | 64 | Altered Mental Status | Primary | Total | QT/RT | Yes | GFAP (+), S100 (+) | IDH-1 wt (+) | 10 months |
| Machi et al. [49] | BVS | Female | 43 | Seizures, Memory Loss | Primary | N.a. | QT/RT TMZ | No | n.a. | IDH-1 wt (+) | 9 months |
| Watanabe et al. [50] | BVS | Female | 19 | Headache | Secondary | Total | QT/RT TMZ | Yes | OLIG2 (+), IDH-1 wt (+) | BRAF V600E (+), IDH1 R132H (−), SMARCB1 (−), SMAD2/3 (−), TP53 (−) | Alive |
| Porto et al. [46] | BVS | Male | 72 | Headache, Hemiplegia | Primary | Partial | S-ALA Fluorescence | QT/RT | Yes | ATRX (+) | IDH-1 wt (+) | 5 months |
| Awadulla et al. [47] | BVS | Male | 60 | Aphasia, Hemiplegia | Primary | Partial | QT/RT | No | GFAP (+), VIMENTIN (+) | n.a. | 8 months |
| Gestrich et al. [48] | BVS | Male | 64 | Altered Mental Status | Primary | Total | QT/RT | Yes | GFAP (+), S100 (+) | IDH-1 wt (+) | 10 months |
| Machi et al. [49] | BVS | Female | 43 | Seizures, Memory Loss | Primary | N.a. | QT/RT TMZ | No | n.a. | IDH-1 wt (+) | 9 months |
| Watanabe et al. [50] | BVS | Female | 19 | Headache | Secondary | Total | QT/RT TMZ | Yes | OLIG2 (+), IDH-1 wt (+) | BRAF V600E (+), IDH1 R132H (−), SMARCB1 (−), SMAD2/3 (−), TP53 (−) | Alive |

(Continued)
| Reference          | Database | Gender | Age  | Symptoms                      | Type of Gbm Tumor | Surgical Resection | Therapy | Tumor Recurrence | Molecular Markers (Expression) | Molecular Markers (Somatic Mutations) | Survival Time |
|--------------------|----------|--------|------|-------------------------------|-------------------|--------------------|---------|-----------------|-------------------------------|--------------------------------------|---------------|
| Zhang et al. [63]  | PubMed   | Male   | 17   | Dysphagia, hypokinesia       | Primary           | Partial            | QT TMZ  | Yes             | GFAP(+)                        | S100A1(+), VIMENTIN(+), MGMT(−), M667(−) | 37 months     |
| Zucali et al. [64] | PubMed   | Female | 65   | Headache, nausea, memory loss | Primary           | Partial            | QT RT TMZ    | Yes             | MGMT(+)                         | n.a.                                                | Alive         |
| Miao-Xia He et al [65] | PubMed   | Male   | 31   | Headache                      | Primary           | Total              | QT RT       | Yes             | GFAP(+), S100A1(+), Vimentin(+), M667(+) | SMARCB1(+)  | 4 months      |
| Paraskevopoulos et al [66] | PubMed   | Female | 12   | Hemiparesis, dysesthesia     | Primary           | Total              | QT RT vincerine, etoposide, carboplatin | Yes             | GFAP(+), S100A1(+), M667(+) | n.a.                                                | 12 months     |
| Jeong et al. [67]   | PubMed   | Male   | 32   | Headache                      | Primary           | Total              | QT RT TMZ     | No              | GFAP(+)                         | M667(+), SMARCB1(+) | IDH wt R132H(−)  | Alive         |
| Lakićević et al. [68] | PubMed   | Male   | 53   | Headaches, nausea, vomiting  | Primary           | Total              | QT RT TMZ     | No              | GFAP(+)                         | n.a.                                                | Alive         |
| Matsuda et al. [69] | PubMed   | Male   | 69   | Facial pain                   | Primary           | Partial            | QT RT TMZ     | No              | GFAP(+)                         | M667(+), SMARCB1(+) | TP53(−), IDH wt R132H(−) | Alive         |
| Theeler et al. [70] | PubMed   | Female | 36   | Progressive neurologic deficits | Secondary         | N. A.              | QT RT TMZ     | Yes             | N. A.                           | IDH1 wt R132H(+)  | BRAF V600E(+)  | Alive         |
| Theeler et al. [70] | PubMed   | Male   | 32   | Progressive neurologic deficits | Primary           | Partial            | QT RT TMZ, erlotinb | Yes             | P16(−)                          | IDH wt R132H(+)  | Alive         |
| Johnson et al. [71] | PubMed   | Male   | 73   | Hemiparesis, seizures         | Primary           | Total              | QT RT TMZ     | Yes             | M667(−)                         | n.a.                                                | 24 months     |
| Johansen et al. [72] | PubMed   | Female | 59   | Headache, blurred vision      | Primary           | Total              | QT RT TMZ     | No              | GFAP(+), O68G(+), M667(+), IDH1 wt (−), TP53(−) | n.a.                                                | 8 months       |
| Johansen et al. [72] | PubMed   | Male   | 60   | Seizures, cerebral haemorrhage | Primary           | Total              | N. A.          | No              | GFAP(+), O68G(+), M667(+), IDH1 wt (−), TP53(−) | n.a.                                                | 10 months      |
| Anghierez et al. [73] | PubMed   | Male   | 43   | Headache                      | Primary           | Total              | QT/QT, TMZ, bevacizumab | Yes             | GFAP(+), VIMENTIN(+), M667(−), IDH1 wt (−), TP53(−) | n.a.                                                | 25 months     |
| Bhagat et al. [74]  | PubMed   | Male   | 30   | Seizures                      | Primary           | Total              | QT RT TMZ, bevacizumab | Yes             | GFAP(+), VIMENTIN(+), M667(−), IDH1 wt (−), TP53(−) | n.a.                                                | 6 years        |
| Chen et al. [75]    | PubMed   | Male   | 5    | Fever, vomiting               | Primary           | Total              | QT RT         | No              | GFAP(+)                         | M667(−), TP53(−)  | 2 months       |
| Gandhi et al. [5]   | PubMed   | Female | 45   | Aphasia                       | Primary           | Partial            | QT RT         | Yes             | M667(+)                         | TP53(−), EGFR(+) | IDH1 wt (−)   | 26 months     |

(Continued)
| reference | database | gender | age | symptoms | type of glioblastoma tumour | surgical resection | therapy | tumour recurrence | molecular markers (expression) | molecular markers (somatic mutations) | survival time |
|-----------|----------|--------|-----|----------|-----------------------------|-------------------|---------|------------------|--------------------------------|----------------------------------|--------------|
| Efferth et al. [76] | PubMed | male | 65 | headache, seizures | primary | partial | QT/RT, TMZ | no | MGMT(+) | n.a. | alive |
| Shen et al. [77] | PubMed | female | 15 | hemiparesis | primary | partial | QT/RT, TMZ | no | GFAP(+), H3K27M(+) | n.a. | 13 months |
| Tokuda et al. [78] | PubMed | male | 27 | seizures, headache | secondary | total | QT/RT, TMZ, bevacizumab | yes | M667Y(+), VEGFR/FLT1(+) | IDH-1 Mutant(+) | alive |
| Wang et al. [79] | PubMed | female | 50 | headache, hemiparesis, nausea, vomiting | primary | total | RT | no | VIMENTIN(+), GFAP(+), OLIG2(+), NESTIN(+) | IDH1-R132H(−), TP53(−), BRAFV600E(+) | alive |
| Wang et al. [79] | PubMed | male | 36 | headache, nausea, vomiting | primary | total | QT/RT, TMZ | yes | VIMENTIN(+), GFAP(+), OLIG2(+), NESTIN(+) | IDH1-R132H(−), TP53(−), BRAFV600E(+) | 8 months |
| Zhang et al. [80] | PubMed | male | 40 | headache, hemiparesis, vomiting | primary | total | RT, TMZ | yes | Ki-67(−), MGMT(+) | TP53(+) | alive |
| Zhou et al. [81] | PubMed | male | 31 | headache, vomiting | primary | total | QT/RT, TMZ | yes | GFAP(+), VIMENTIN(+), NESTIN(+), OLIG2(+), M667Y(+) | EGFR(+), S100A1(+) | 15 months |
| Comito et al. [82] | PubMed | female | 57 | headache, nausea, photopsia | primary | total | QT/RT, TMZ, lomustine n.a., nivolumab | yes | M667Y(+), GFAP(+), MGMT(+), IDH1 wt(+) | alive |
| Fimmers et al. [83] | PubMed | female | 29 | aphasia, headache, confusion | secondary | total | RT | no | GFAP(+), MGMT(-), S-100A(-) | EGFR(-), SMARCB(−), TP53(−), IDH1 wt(-) | alive |
| Homma et al. [84] | PubMed | female | 78 | speech difficulty and forgetfulness | primary | partial | QT/RT, TMZ | no | S-100A(+), GFAP(+), OLIG2(+), ATRX(−), M667Y(+) | SMARCB(1)(−), BRAFV600E(−), IDH1-R132H(−) | alive |
| Jarvik et al. [85] | PubMed | male | 51 | headache, memory loss | primary | total | QT/RT, TMZ | yes | GFAP(+), M667Y(+) | TP53(+) | BRAFV600E(+) | 23 months |
| Narasimhaiah et al. [86] | PubMed | male | 16 | headache, vomiting, diplopia | primary | partial | QT/RT | yes | S-100A(+), GFAP(+), M667Y(−), ATRX(−) | TP53(+) | IDH1 wt(−) | alive |
| Narasimhaiah et al. [86] | PubMed | female | 21 | headache, seizures | primary | total | QT/RT | no | GFAP(+), M667Y(+), S100(+), ATRX(−) | TP53(−), IDH1-R132H-mutant(−) | alive |
| Nance et al. [87] | PubMed | male | 62 | confusion, aphasia | primary | partial | QT/RT, bevacizumab, irinotecan | yes | ATRX(−), MGMT(+) | IDH1 wt(+) | 15 months |
| Nance et al. [87] | PubMed | male | 30 | headache, seizures, confusion | secondary | partial | QT/RT, TMZ | yes | GFAP(+), ATRX(+), MGMT(+) | IDH1(+) | alive |
| Okamoto et al. [88] | PubMed | female | 27 | hemiparesis, dysuria | primary | partial | QT/RT, TMZ | no | GFAP(+), H3K27M(+), OLIG2(+) | ATRX(−), Nestin(−), H3K27M(+) | alive |

(Continued)
| reference   | database | gender | age  | symptoms                                      | type of gbm tumour | surgical resection | therapy | tumour recurrence | molecular markers (expression) | molecular markers (somatic mutations) | survival time |
|------------|----------|--------|------|-----------------------------------------------|-------------------|-------------------|---------|------------------|-------------------------------|--------------------------------------|--------------|
| Cuoco et al. [89] | PubMed   | male   | 76   | hemiparesis, clumsiness                       | primary           | partial           | QT/RT   | no               | MGMT(+), EGFR(−)              | IDH-1 wt (+), TP53(−)                  | 1 months     |
| Romo et al. [90]  | PubMed   | male   | 28   | headache, nausea, personality changes, aphasia | primary           | total             | QT/RT, TMZ, VPC | yes              | GFAP(+), OLIG2(+), ATRX(+), MGMT(+), S100B(+) | IDH-1 mutant(+), TP53 mutant(+), SMARCB1(+), H3K27M(−) | 3 months     |
| Uppar et al. [91]  | PubMed   | female | 28   | hemiparesis                                  | primary           | total             | n.a.    | yes              | GFAP(+), MGMT(+), H3K27M(+)     | IDH-1 wt(+), BRAFV600E(+), IDH-1 wt(+) | 1 month      |
| Woo et al. [92]   | PubMed   | female | 22   | headache                                     | primary           | total             | RT, dabrafenib, trametinib | yes              | MGMT(+)                       | IDH-1 wt(+), BRAFV600E(+), TERT(+), EGFR(−) | 7 months     |
| Sajan et al. [93] | PubMed   | female | 39   | headache                                     | primary           | n.a.              | QT/RT TMZ | no               | GFAP(+), MGMT(+)               | EGFR(+), IDH-1 wt(+), H3K27M(+)        | alive        |
| Gupta et al. [94] | PubMed   | male   | 58   | seizures                                     | primary           | total by 5-ALA fluorescence | QT/RT   | yes              | n.a.                          | IDH-1 wt R132H(+)                     | alive        |

*Reference of the clinical report.

Database where the clinical report was found.

Sex of the patient of the clinical case reported.

Age of the patient of the clinical case reported.

Symptoms described during clinical routine.

GBM tumour according to ontogenesis subtypes.

Surgical procedures during patient treatment.

Medical and drugs administered for treatment.

Recurrence of tumour after surgical procedures.

Gene expression measured for diagnosis.

Molecular markers to identify somatic mutations.

Survival time of patients after surgery and therapy treatment.
observed for MGMT with a tendency to be upregulated in mesenchymal and downregulated in proneural subtypes. Related to TERT, no expression differences are observed between the subtypes.

Among the cytoskeleton and cellular proliferation genes, the most substantial differences among subtypes are for EGFR, with a general tendency to be upregulated in the classical proliferative subtype and downregulated in the proneural and mesenchymal subtypes (table 2). Another tyrosine kinase growth factor, FL1, did not show big differences in expression among GBM subtypes; meanwhile, IDH1 has an upregulation tendency only for the classical subtype. The downstream effectors for growth factors, PIK3CA and SMAD3, showed upregulation and downregulation, respectively, for the proneural subtype; meanwhile, the pattern is inverse, downregulation for PIK3CA and upregulation for SMAD3 for both the classical and mesenchymal subtypes (table 2). Another proliferation biomarker, MKi67, showed a marked overexpression in the proneural subtype and a tendency to downregulation in the classical and mesenchymal subtypes. NES and VIM appeared to be expressed more in the classic subtype than in other subtypes. Moreover, no relevant changes were observed for GFAP, another intermediate filament expressed in neural stem cells. Nevertheless, OLIG2, an oligodendrocyte marker, is upregulated in the proneural and downregulated in the mesenchymal subtype. The same behaviour was observed for the differentiation marker S100A1 (table 2).

For the tumour suppressor genes, TP53 is clearly upregulated in the proliferative classical subtype. The other gene, SMARCB1, is also overexpressed in the proneural and classical, but downregulated in the mesenchymal subtype (table 2).

We then analysed the expression patterns of the driver genes clustered into three subgroups of patients according to their age (tables 3–5). An important observation is that among tumours showing expression of these genes in patients under 30 years, the mesenchymal subtype was not observed (table 3). On the other hand, the driver gene expression in the mesenchymal subtype is only present in patients older than 80 years (data not shown).

For the young subgroup, the samples were determined to belong only to proneural and classical subtypes, and from the 18 genes analysed, only OLIG2 and VIM showed a differential pattern in gene expression. OLIG2 is upregulated in the proneural tumours, according to its role as a differentiation biomarker. Meanwhile, our analysis revealed a downregulation tendency in the classical subtype. An inverse pattern was observed for VIM, which is downregulated in proneural and upregulated in the classical subtype. Interestingly, EGFR is downregulated in both subtypes (table 3).

Among the subgroup of adult patients, the behaviour of the EGFR stands out as it is upregulated in the classic subtype and downregulated in proneural and mesenchymal subtypes (table 4). The same pattern was observed in the elderly subgroup, but with a larger gap between subtypes (table 5). Analyses of genes ATRX, H3F3A, MGMT, MKi67, NES,
young subgroup (10–29 years) | proneural | classical
--- | --- | ---
DNA repair and chromatin remodelling genes | | |
ATRX | 0.105 ± 1.316 | -0.105 ± 0.711
BRAF | 0.257 ± 0.466 | 0.034 ± 0.777
H3F3A | 0.375 ± 0.626 | -0.202 ± 0.619
MGMT | 0.297 ± 0.494 | 0.176 ± 1.782
TERT | 0.066 ± 0.335 | 0.196 ± 0.464

cytoskeleton and cellular proliferation genes | | |
EGFR | -3.133 ± 1.228 | -4.563 ± 2.382
FLT1 | -0.946 ± 0.61 | -0.955 ± 0.539
GFAP | -0.106 ± 0.914 | 0.262 ± 0.113
IDH1 | -0.679 ± 0.679 | -0.910 ± 0.204
MKI67 | 0.820 ± 2.157 | 0.572 ± 1.801
NES | -0.207 ± 1.018 | 0.822 ± 1.167
OLIG2 | 0.998 ± 1.427 | -1.341* ± 0.859
PIK3CA | 0.058 ± 0.526 | 0.272 ± 0.844
S100A1 | 0.452 ± 1.032 | 0.101 ± 0.807
SMAD3 | 0.104 ± 0.732 | 0.664 ± 0.438
VIM | -1.127 ± 1.343 | 1.455* ± 0.353

tumour suppressor genes | | |
SMARCB1 | 0.830 ± 0.653 | 0.653 ± 0.874
TP53 | 0.286 ± 0.993 | -0.336 ± 1.242

*p < 0.05, **p < 0.01 and ***p < 0.001.

Table 3. Summary of driver gene expression in GBM molecular subtypes in the 10–29 year subgroup with significant p-value. Data represent mean ± standard deviation for z-score in each gene. Statistical significance is represented by asterisks.

3.3. Somatic mutation analysis on driver genes

Gene mutation profiling has also served as a biomarker for the diagnosis and treatment of GBM. We used high-throughput data from the TCGA-GBM project and obtained the genomic profiles of a total of 588 clinical GBM cases.

Among the driver genes, cases showed that the most frequently affected genes in patients were TP53 (26%), EGFR (22%), ATRX/PIK3CA (approx. 10%) and IDH1/MKi67 (approx. 5%) (electronic supplementary material, figure S1). For an overall view of GBM aberrations, the distribution of the total mutated genes and their DNA changes was determined using the relative percentage of gene mutations according to the total length (base pairs) of each chromosome. The highest rate was found in chromosomes 19, 17 and 11, and the lowest levels were found in chromosomes 18, 13 and Y (figure 2a). Chromosome 1, which contains the highest number of coding genes (2076), showed a lower percentage of mutations than chromosome 17, which contains less than 60% the number of genes (1209). TP53 (17p13.1), which suffers from a broad amount of mutations, and GFAP (17q21.31), two of the most commonly used genomic markers for GBM, are found in this chromosome (figure 2b). Among all mutations, 95% substitutions, 3% deletions and 2% insertions were identified (figure 2a).

A comparison was done to determine whether the relative abundance of the types of DNA changes present in driver genes was similar to that of the whole genome. This revealed that the base substitutions were the highest both in driver genes and in the whole genome and that the nucleotide G-C change the most common (figure 2c and data not shown). However, mutations in the driver genes displayed a higher number of deletions and insertions than the whole genome.

The genomic location and frequency of mutations were determined according to the patient age subgroup. Chromosomes 19, 17, 11 and 16 had the highest percentage of mutations in all subgroups. However, some chromosomes, such as 6 and 18, showed different patterns according to patient age. Regarding mutation types, substitutions were the highest in all patients, but an increase of deletions and insertions was found according to patient age (figure 3a). We also observed that mutations in the driver genes reflect the parallel distribution of the genome-wide mutations (figure 3b), as is the case in other cancers [18]. However, the frequency of mutations varies according to age group, highlighting the different mutational behaviour of driver genes in the young subgroup. In particular, TP53 and EGFR, which are shown to be the most mutated genes in the adult and elderly subgroups, are not so in the young subgroup, where ATRX is the most affected driver gene. Among other DNA repair and chromatin remodelling genes, the mutation frequency behaviour of IDH1, and MGMT increases at 30 years of age and decreases at 60 years (figure 3b). When analysing these mutations in more detail, we observed that most of the mutations in all subgroups are substitutions: 91% in young, 80% in adults and 87% in the elderly.
As we found that polyphen impact was mainly reported for the function of the proteins encoded by the driver genes. We also studied the phenotypic consequences of each mutation, which can often cause several of them. In the case of TP53, for example, a single mutation affects its 27 transcripts, causing consequences of different types. The missense variant consequence appears to be by far the most abundant, representing 47% of all consequences elicited by somatic mutations. Downstream and upstream gene variants, frameshift and intron variants, and stop gain, represent 35% of the consequences caused by mutations, and the remaining percentage is distributed among all other consequences.

Then, we focused on analysing the biological relevance of mutations on the driver genes. Polymorphism Phenotyping (polyphen) helps to predict the functional significance of an allele replacement from its features by a Naive Bayes classifier [19]. The polyphen impact reported in TCGA is a prediction of a mutation consequence being probably damaging, possibly damaging, or benign. Therefore, we used this data to indicate the possible impact of the consequence types on the function of the proteins encoded by the driver genes. As we found that polyphen impact was mainly reported for the missense variant consequence, we focused on the possible impact of amino acid substitutions.

3.4. Phenotypic consequences of mutations on driver genes

We also studied the phenotypic consequences of each mutation, which can often cause several of them. In the case of TP53, for example, a single mutation affects its 27 transcripts, causing consequences of different types. The missense variant consequence appears to be by far the most abundant, representing 47% of all consequences elicited by somatic mutations. Downstream and upstream gene variants, frameshift and intron variants, and stop gain, represent 35% of the consequences caused by mutations, and the remaining percentage is distributed among all other consequences.

Then, we focused on analysing the biological relevance of mutations on the driver genes. Polymorphism Phenotyping (polyphen) helps to predict the functional significance of an allele replacement from its features by a Naive Bayes classifier [19]. The polyphen impact reported in TCGA is a prediction of a mutation consequence being probably damaging, possibly damaging, or benign. Therefore, we used this data to indicate the possible impact of the consequence types on the function of the proteins encoded by the driver genes. As we found that polyphen impact was mainly reported for

### Table 4. Summary of driver gene expression in GBM molecular subtypes in the 30–59 year subgroup with significant p-value. Data represent mean ± standard deviation for z-score in each gene. Statistical significance is represented by asterisks.

| DNA repair and chromatin remodelling genes | proneural | classical | mesenchymal |
|-----------------------------------------|-----------|-----------|-------------|
| ATRX                                    | 0.279 ± 0.920 | 0.113 ± 0.608 | −0.180** ± 0.583 |
| BRAF                                    | −0.125 ± 0.392 | −0.122 ± 0.561 | 0.064 ± 0.505 |
| H3F3A                                   | 0.375 ± 0.446 | −0.121**** ± 0.609 | −0.451**** ± 0.602 |
| MGMT                                    | −0.354 ± 1.236 | 0.200 ± 1.339 | 0.786* ± 1.272 |
| TERT                                    | 0.099 ± 0.402 | 0.357 ± 0.472 | 0.138 ± 0.451 |
| cytoskeleton and cellular proliferation genes |         |           |             |
| EGFR                                    | −2.693 ± 3.863 | 3.134**** ± 4.265 | −1.970 ± 3.678 |
| FL1                                     | −0.637 ± 0.735 | −0.259 ± 1.091 | 0.194* ± 1.220 |
| GFAP                                    | 0.055 ± 1.018 | 0.249 ± 0.541 | −0.399 ± 1.054 |
| IDH1                                    | −0.054 ± 0.833 | 0.781** ± 1.065 | −0.115 ± 0.889 |
| MKI67                                    | 1.065 ± 1.551 | −0.052** ± 1.278 | −0.061** ± 0.938 |
| NES                                     | −0.198 ± 0.700 | 1.380**** ± 0.884 | −0.151 ± 0.941 |
| OLIG2                                   | 1.396 ± 1.021 | −0.029**** ± 1.083 | −1.581**** ± 1.063 |
| PIK3CA                                   | 0.415 ± 1.229 | −0.370* ± 0.785 | −0.328* ± 0.625 |
| S100A1                                   | 0.467 ± 1.287 | −0.674** ± 0.930 | 0.127 ± 1.516 |
| SMAD3                                   | −0.207 ± 0.838 | 0.271 ± 0.431 | 0.198 ± 0.542 |
| VIM                                      | −0.369 ± 1.083 | 0.564* ± 0.850 | 0.778** ± 0.944 |
| tumour suppressor genes                 |         |           |             |
| SMARCB1                                  | 1.008 ± 0.951 | 0.311* ± 0.998 | −0.297**** ± 0.862 |
| TP53                                     | 0.194 ± 1.162 | 0.829* ± 0.654 | −0.124 ± 0.809 |

*p < 0.05, **p < 0.01, ***p < 0.001 and ****p < 0.00001.

Summarizing, the TP53 tumour suppressor gene was found to have the highest frequency of mutations among all patient groups. For SMARCB1, another tumour suppressor gene, we found few mutations in adult and elderly subgroups, and none for the young subgroup (figure 3b).

3.5. Driver gene biological pathways compared to the highest affected genes in GBM

Functional enrichment analysis was carried out for driver genes and for other genes identified with the worst protein polyphen impact. Driver genes are significantly enriched in
Table 5. Summary of driver gene expression in GBM molecular subtypes in the 60–89 year subgroup with significant p-value. Data represent mean ± standard deviation for z-score in each gene. Statistical significance is represented by asterisks.

| DNA repair and chromatin remodelling genes | proneural | classical | mesenchymal |
|-------------------------------------------|-----------|----------|-------------|
| ATRX                                      | 0.557 ± 0.718 | 0.077*** ± 0.557 | −0.255**** ± 0.597 |
| BRAF                                      | −0.331 ± 0.498 | −0.474 ± 0.408 | 0.047* ± 0.416 |
| H3F3A                                     | 0.293 ± 0.649 | 0.014 ± 0.606 | −0.551**** ± 0.606 |
| MGMT                                      | −0.07 ± 1.377 | −0.57 ± 1.449 | 0.394 ± 1.230 |
| TERT                                      | 0.226 ± 0.372 | 0.121 ± 0.463 | 0.178 ± 0.535 |

| Cytoskeleton and cellular proliferation genes | TERT | GFAP | IDH1 | MVI67 |
|----------------------------------------------|------|------|------|-------|
| EGFR                                         | −4.398 ± 1.122 | 5.495**** ± 2.32 | −2.045** ± 3.921 |
| FLI1                                          | −0.368 ± 0.887 | −0.229 ± 0.939 | −0.062 ± 0.884 |
| OLIG2                                        | 1.36 ± 1.204 | 0.524* ± 1.091 | −1.293**** ± 0.792 |
| PIK3CA                                       | 0.143 ± 0.968 | 0.03 ± 1.056 | 0.086 ± 0.855 |
| S100A1                                       | 0.596 ± 1.211 | −0.973*** ± 1.2 | −0.192* ± 1.373 |
| SMAD3                                        | −0.387 ± 0.489 | 0.269*** ± 0.376 | 0.341*** ± 0.614 |
| VIM                                          | −0.629 ± 1.021 | 1.048**** ± 1.111 | 0.535*** ± 0.764 |

| Tumour suppressor genes | SMARCB1 | TP53 |
|-------------------------|---------|------|
| SMARCB1                 | 0.901 ± 0.886 | 0.602 ± 1.016 | −0.515*** ± 0.916 |
| TP53                    | −0.058 ± 0.864 | 0.724** ± 0.767 | 0.327 ± 0.645 |

* p < 0.05, **p < 0.01, ***p < 0.001 and ****p < 0.0001.

hsa0513 and hsa0512 for pancreatic and endometrial cancer from the KEGG pathway (−log 10, 9.05 > −7.3), and the top GO terms include dsRNA processing, multicellular organism growth, negative regulation of cell differentiation, regulation of DNA metabolic process and regulation of neuron apoptotic process (−log10−7.3 > −4.80) (figure 5a). We also observed that the most affected protein phenotypes are functionally enriched in biological processes such as blood circulation, purine containing compound biosynthetic process, cellular response to nitrogen compound and vascular process in the circulatory system (−log10−30.02 < −22.68) (figure 5b). The biological pathways enriched were Reactome has R-HSA-382551: Transport of small molecules, (−log10−46.69), KEGG has:04022 cGMP PKG signalling pathway, has:0513 and 00071 Fatty acid degradation and has:00010 Glycolysis/gluconeogenesis pathways (−log10−39.80 > −16.74) (figure 5b). Those lesioned genes were linked to seizures, epilepsy, weight loss, paediatric failure to thrive, mental depression, irritation and vomiting symptoms (−log10−18 < −8.3) (figure 5b).

4. Discussion

Current clinical standard methods in neuro-oncology for GBM diagnosis consist of tumour surgery resection and biopsy followed by pathology analysis. We searched the literature over the last 15 years and found 60 clinical reports of 73 clinical cases in which patient tumour biopsy or fluid sample underwent the analysis of a combination of biomarkers which mainly consisted in IDH1, GFAP, MVI67 and MGMT coupled in sets with more than two and up to 11 additional markers per sample for diagnosis. Molecular markers were reported for their relevance as measurable indicators of the presence and severity of GBM. Among those genes, the measures on the expression of ATRX, MGMT, FLI1, GFAP, MVI67, NES, OLIG2, S100A1, VIM, PIK3CA, as well as the genetic analysis of driver mutation events in BRAF, H3F3A, TERT, EGFR, IDH1, SMAD3, TP53 and SMARCB1 were highlighted from our literature search strategy. We searched among clinical results for a pattern of biomarker behaviour in the analysed samples with unsuccessful results. Aware of inter-tumour molecular heterogeneity as a significant challenge, and due to the remarkable importance of driver genes for the routine clinical role, we delved into their biological behaviour. A compendium of summarized findings of driver genes is shown in electronic supplementary material, file 2.

GBM inter-tumour heterogeneity allows molecular subclassification based on genomic profiling. This is also affected by intra-tumour heterogeneity, originating from two proposed mechanisms, clonal evolution and cancer stem cells. Clonal evolution is the process by which a single cell undergoes iterative genetic changes which allows it to evolve and disseminate, forming a tumour [20]. By contrast, cancer cells in GBM could possess different stemness according to their cellular ontology, being a direct transformation from a normal stem cell or a reprogramming process from a cancer stem cell with less proliferative or differentiation capacity [17]. The GBM tumour consists of a core region of
high cell proliferation and inflammation, delimited by a margin between the tumour tissue and the normal brain cells, and then the peritumoral brain zone mainly composed of normal tissue with some infiltrative and isolated tumour cells [16].

Based on a multi-omic analysis, we herein describe the heterogeneity of GBM at the transcriptional and the genomic levels, with an emphasis on tumorigenesis driver genes currently used in the clinic as molecular markers. Altogether, our results suggest that a combination of these biomarkers would provide a multidimensional approach for a better diagnosis and GBM subtype molecular classification for patient prognosis. Besides, our studies for gene expression and somatic mutations will provide information on the heterogeneity of primary GBM types due to their clinical relevance. Our transcriptomics analysis from mRNA expression data agrees with previous reports with respect to the mesenchymal subtype. This subtype is characterized by its poor prognosis, stem cell biomarkers, angiogenesis and prominent radio- and chemoresistance. From the 18 genes analysed, we found upregulation of \textit{MGMT}, which may be related to its own promoter’s unmethylated status frequently observed in this GBM subtype and related to temozolomide treatment resistance and short patient survival [21]. In our analysis, this expression profile was conserved during adult and elderly life stages.

Furthermore, the downregulation of \textit{ATRX}, \textit{H3F3A} and \textit{EGFR} was observed. \textit{ATRX} encodes an adaptor protein that contributes to the Methyl-CpG binding protein 2 (M\textsubscript{c}CP2)-mediated pericentric heterochromatin organization, which is very important for neural differentiation [22]; thus, downregulation of this gene might be expected in cells of a less differentiated subtype with more stemness such as the mesenchymal GBM subtype. The opposite,
upregulated behaviour, was observed in the proneural subtype, which has less stemness and more characteristics of differentiated cells. Another chromatin remodeler, H3F3A, whose driver mutations HK27M and G34R induce dysfunction of Polycomb repressive complex 2 (PRC2) and dramatic alterations of gene expression [23,24], may contribute to high alterations in profile expression for mesenchymal GBM subtype. EGFR, which is perhaps one of the best-characterized molecules in primary GBM [25], showed a downregulation in mesenchymal and proneural subtypes, but a clear upregulation in the classical GBM subtype. This behaviour is conserved across all age groups and strikingly marked for the elderly population. This expression profile could be dependent on mesenchymal GBM increased mutation rates, which may play a feedback role in downregulating EGFR gene expression. The coexistence of mutations in

![Figure 2. Distribution of the percentage of mutations in genes per chromosome observed in the TCGA-GBM project and the location of their mutations.](image-url)

(a) Relative percentage of gene mutations per chromosome shown by mutation type: substitutions (blue), deletions (orange) and insertions (green). (b) Cytogenetic representation of human chromosomes, rendered with standard banding patterns, showing the chromosomal location of the driver genes (one coloured arrow per gene). (c) Number of mutations per nucleotide found in the entire genome (left) driver gene mutations (right).
critical molecules from downstream EGFR signalling such as Ras or PTEN, which maintain active signalling without a ligand to the receptor, could play a role as an alternative mechanism.

We observed other genes with striking profile expression, including NES, VIM and TP53, with upregulation behaviour. NES and VIM encode the intermediate filament proteins Nestin and Vimentin. Vimentin is expressed mainly in mesenchymal cell types, while Nestin mainly in neural stem and progenitor cells in the central nervous system [26]. These proteins function not only as part of the cytoskeleton, but also impact several key cellular processes such as proliferation, death, migration and invasiveness [26]. Our analysis showed that VIM is upregulated in both mesenchymal and classical GBM subtypes and NES only in the classical subtype. This pattern may be related to the ontogenesis of these tumours and suggest the transition state for classical GBM to a possible mesenchymal GBM, with a neural stem cell marker remaining.

The proneural GBM subtype showed upregulation of MKi67 and OLG2. MKi67 encodes the DNA binding protein Ki-67 and is widely used as a proliferation marker as it participates in chromosome motility and chromatin organization during the cell cycle [27]. OLG2 encodes a central nervous system transcription factor that plays an essential role in the proliferation of oligodendrocyte precursors and their differentiation [28]. OLG2 also showed down-regulation in classical and mesenchymal GBM subtypes. Therefore, these expression patterns support the idea that the proneural GBM subtype arises from central nervous system progenitors with fewer stemness properties but with proliferative capacity.

Our analysis in the expression profile for the 18 driver genes supports the GBM ontogenesis hypothesis from Celiku et al. [17], which proposes that proneural subtypes can be generated from neural progenitors, and these cells may gain somatic mutations to become classical and consecutively mesenchymal subtypes. It is also possible that classical

Figure 3. Genome location and percentage of gene mutations according to patient age subgroup. (a) Relative percentage of gene mutations per chromosome according to patient age: 10–29 (top), 30–59 (centre) and 60–89 years (bottom). Mutations are classified according to their type: substitutions (blue), deletions (orange) and insertions (green). (b) X-axis shows the number of genes sharing the same number of mutations, shown in Y-axis, grouped by patient age (same as in (a)); groups of genes that share the same number of mutations with driver genes represented as red dots. Data obtained from the TCGA-GBM project.
or mesenchymal subtypes originate from central nervous system progenitors with high stemness.

In this study, we found that all driver genes have reported mutations in GBM patients. However, genes that are significantly mutated and that display multiple biological consequences include TP53, ATRX, PIK3CA and EGFR. Abnormalities of TP53 have been the most extensively investigated genetic variations found in more than 50% of human tumours [29]. Contrary to other reports where TP53 mutations are more related to paediatric tumours [30], we found an increasing behaviour from the young to elderly subgroups. The same behaviour is observed for genes ATRX, PIK3CA and EGFR. However, TP53 and EGFR were found to be the most mutated genes in adult and elderly subgroups, while these mutational behaviour changes in the young subgroup, in which ATRX is the most affected gene (figure 4b).

Impairment of DNA repair is expected to increase the overall frequency of mutations and, hence, the likelihood of cancer-causing mutations. In comparison to other studies in which ATRX was found to be mutated only rarely in adult primary GBM, but frequently found in younger adults with lower-grade glioma (WHO grade II/III) [31], we found a high frequency at 30 years that decreases in elderly patients. A similar behaviour was observed for the DNA repair and chromatin remodelling genes IDH1 and MGMT.

Additionally, NES and VIM mutations were absent in the elderly subgroup and are present only in patients below 60 years of age with an unfavourable consequence in protein structure and function. By contrast, OLIG2 mutations with negative impact consequences were found only in the elderly patient subgroup.

Some driver mutations on key genes have been pivotal for the diagnosis and prognosis of GBM patients. We focused particularly on the effects of mutations with non-synonymous changes, also called missense mutations, which alter the codons so that they specify different amino acids during protein synthesis (electronic supplementary material, figure S2), and carried out a comparison of GO enriched terms of the selected driver genes with those identified with a higher probability of damaging consequences. Similar in lethality and aggressiveness to GBM, pancreatic cancer is a solid tumour difficult to treat and often fatal, characterized by the absence of early symptoms. Driver genes of tumorigenesis shared between GBM and the hsa:0513 pancreatic cancer pathway are BRAF, EGFR, SMAD3, PIK3CA, TP53, IDH1, TERT, VIM, ATRX and GFAP. Owing to their high proliferative condition, cancer cells have an increased demand for nutrients. As a mechanism, tumours alter their metabolism to feed their extensive anabolic requirements having a uniquely high demand for amino acids. Accordingly, upregulation of selective amino acid transporters has been reported [32]. R-HSA-382551 transport of small molecules pathway is involved in the regulation and movement of small molecules across plasma membranes and between cellular organelle compartments within cells. Our functional enrichment analysis on highly affected proteins shows a significant abundance of the solute carrier (SLC) superfamily related to this pathway. Examples of highly lesioned solute carrier proteins found are SLC1A6 (with 10 probably damaging (PR) consequences), SLC5A7 (6: PR), SLC9A2 (6: PR) and SLC6A19 (4: PR). In pancreatic cancer, the clinical potential of an amino acid transporter SLC6A14 as a drug target has been recently reported [32]. We also found that highly affected pathways such as blood circulation and vascular processes in the circulatory system are consistent with alterations in angiogenesis in GBM. We also identified a link of the lesioned proteins to seizures, epilepsy, weight loss, paediatric failure to thrive, mental depression, irritation and vomiting, among other symptoms that are in agreement with those reported in the clinical cases reviewed.

Efforts have been made for the identification of relevant biomarkers to assess GBM progression by targeting genes with the highest density of missense mutations. For example, tumours with the BRAF V600E mutation tend to be more

---

**Figure 4.** Distribution of protein phenotype impact of mutation consequences of missense variants in driver genes grouped by patient age X-axis depicting selected driver genes, while Y-axis represents the number of consequences per missense variant mutation in their corresponding canonical transcript. Consequences were classified by their protein phenotype impact regarding aggressiveness affections in probably damaging (PR, red); possibly damaging (PO, golden) and benign (BE, green). Data are shown according to the age of patients as shown on top of each panel. Data obtained from the TCGA-GBM project.
Figure 5. Comparison of driver gene ontology enrichment analysis with the most lesioned genes in the TCGA-GBM project. (a) Gene ontology enrichment and pathways for driver genes. (b) Heatmaps showing enrichment for most affected genes on their gene ontology and pathways (top), and DisGeNET terms (bottom). The colour key from yellow to brown indicates high to low p-values, respectively. Data obtained from the TCGA-GBM project.
severe. This somatic mutation prevents the Bmi1 protein from controlling cell proliferation (electronic supplementary material, figure S3), which has been reported in the TCGA database, appearing at all ages but more frequently in elderly patients.

$TP53$ mutations were predominantly point mutations, which lead to amino acid substitutions in the DNA binding domain (DBD). The substitution of arginine residues within the DBD, such as R175, R248 and R273, was reported in other studies and was also found in GBM patients [33]. However, this was not the most abundant amino acid substitution, being G105R, S127Y, P152S and V157G, examples of some amino acid changes abundantly reported in the TCGA cohort.

The most cited biomarker for diagnosis IDH1 R132H has also been reported in the TCGA database as a mutation in all age subgroups with a negative polyphen impact [3]. On the other hand, the H3K27M mutation that has been highly linked to paediatric thalamic gliomas and is associated with a worse prognosis than low-grade tumours were not found in the TCGA cohort, which is the case of other biomarkers used in clinical studies, such as H3G34R, H3G34N, $EGFR$ R776C, and the $TERT$ promoter mutations C228T and C250T [23,24,31].

To understand better the behaviour of mutations in young patients, we analysed genes that are involved in GBM with the worst polyphen impact consequences and analysed the transcription factors that regulate them. Our results showed that the young subgroup behaves differently, as genes that are mutated are regulated by different transcription factors (TFs). Moreover, the TFs that regulate genes with mutations in the young subgroup share almost no TFs with adult and elderly subgroups. This might explain why the young subgroup has a divergent behaviour in comparison with the other subgroups. On the other hand, the adult and elderly subgroups share most of the biological pathways, while microtubule cytoskeleton organization, regulation of microtubule-based process, adenylate cyclase-inhibiting G protein-coupled glutamate receptor signalling among others are GO terms unique for the adult subgroup, while protein–protein interactions at synapses, regulation of cyclase activity and carbohydrate digestion and absorption are unique functional terms for the elderly subgroup. In particular, genes with mutations with a negative polyphen impact in the 10–29-year-old subgroup share fewer identities with the 60–89-year-old subgroup (electronic supplementary material, figure S3).

Figure 6. Proposed biomarker panel to determine the GBM molecular subtype. Summarized gene expression analysis showed that the altered expression pattern for the 18 driver genes supports the GBM glioma progression model, which proposes that proneural subtypes can be generated from neural progenitors, and these cells may gain somatic mutations to become classical and consecutively mesenchymal subtypes. Combination of the gene biomarkers $EGFR$, $H3F3A$, $FLT1$, $MGMT$, $Mki67$, $NES$, $S100A$, $TP53$, $OLIG2$ and $VIM$ could help to determine the GBM molecular subtype for patient prognosis. $EGFR$, $NES$, $OLIG2$ and $VIM$ (highlighted in dark blue) show a strong differential expression pattern.
tumour heterogeneity for a more accurate diagnosis. To achieve this objective, it is crucial to analyse the raw data for other key molecules involved in the mechanisms that drive the balance between proliferation and differentiation in the stem and cell precursors for the central nervous system.

Currently, expression levels of ATRX, MGMT, FLI1, GFAP, Mki67, NES, OLIG2, S100A1, VIM and PIK3CA are used in the clinic for patient GBM diagnosis and prognosis. Our results indicate that the biomarker set integrated by EGFR, H3F3A, FLI1, MGMT, Mki67, NES, S100A1, TP53, OLIG2 and VIM genes could be a strong combination to determine the GBM molecular subtype (figure 6). For example, the mesenchymal subtype, known as the most aggressive GBM, showed overexpression of MGMT and VIM, and the repression of EGFR, H3F3A, OLIG2, S100A1 and TP53. On the other hand, while overexpression of EGFR, NES, VIM and TP53 was characteristic of the proliferative or classical subtype, concomitant overexpression of Mki67 and OLIG2 could be more favourable prognosis owing to their association with the less aggressive proneural subtype. Recently, Teo et al. based on TCGA data used a set of 1500 genes from GBM subtypes across Caucasian, Korean and Chinese populations [34]. In comparison, from our selected driver genes EGFR, IDH1, Mki67, NES, S100A1 and VIM were reported to be differentially expressed and overlapped with the three datasets of TCGA-GBM populations samples. Moreover, EGFR, NES and S100A1 are among their selected 500 genes used for the classification of the three GBM subtypes. Furthermore, in accordance with their study we also identified that EGFR presents subtype specificity. In addition, our study suggests that NES, OLIG2 and VIM are also subtype specific genes. Altogether, our findings indicate that EGFR, NES, OLIG2 and VIM genes represent an outstanding selection of biomarkers for patient prognosis, since a remarkable differential pattern from the combination of them was revealed by our transcriptomic analysis (figure 6). Further clinical trials with patient samples for expression analysis, together with the development and application of gene expression-based classifier algorithms for molecular subtypes testing the above-mentioned biomarkers could provide confirmatory evidence for their clinical potential.

5. Conclusion

GBM is a highly heterogeneous cancer that consists of multiple molecular alterations. Despite the vertiginous advances in the clinical medical area, the prognosis of patients continues to be unfavourable, with an average survival of less than 1 year. The differential molecular characteristics of histologically similar tumours make it difficult to predict clinical outcomes and select optimal treatment strategies. Given the heterogeneity of GBM and the multitude of factors that influence disease progression, general clinical characteristics are insufficient to predict individual prognosis and survival accurately. In the clinical routine, a combination of biomarkers is necessary for differential diagnosis and prognosis being IDH1, GFAP, Mki67 and MGMT the most reported ones. The inter-tumour molecular heterogeneity remains the hardest challenge in neuro-oncology practice. In our study, the expression profiles of those markers revealed a consistent link with the progression model for GBM tumour ontogenesis, supporting that tumours display a unique behaviour and that ‘personalized’ treatment must be required for each molecular subtype. Our results indicate that a combination of the biomarker genes EGFR, NES, OLIG2 and VIM could be a strong set to determine the GBM molecular subtype for patient prognosis. Notably, the frequency of mutations varies according to age group, highlighting the different mutational behaviour of driver genes in the young subgroup. In particular, TP53 and EGFR, which are the most mutated genes in the adult and elderly subgroups, are not mutated in the young subgroup, in which ATRX is the most affected driver gene. Besides, a unique distribution of somatic mutations was found for the young and adult populations, particularly for the genes related to DNA repair and chromatin remodelling ATRX, MGMT and IDH1. We also identified regulatory and biological pathway behaviours that varied with age which could serve as a basis for further analysis in the journey of the development of improved therapy for patients suffering from this disease.

Data accessibility. Data source for the gene expression analysis Glioblastoma BioDiscovery Portal (GBM-BioDP) https://gbm-biodp.nci.nih.gov. Data source for somatic mutations of GBM https://portal.gdc.cancer.gov/. Github Repository for codes Python scripts (https://github.com/kap8416/GLIOBLASTOMA-MULTIFORME-A-META-ANALYSIS-OFDRIVER-GENES-CURRENT-DIAGNOSIS-AND-TUM-OR-HETEROGENEITY).

Authors’ contributions. G.E.H.O.: provided ideas for project design, manuscript writing and carried out data analysis, prepared the figures and their interpretation. C.E.A.R.: neurosciences’ expertise provided ideas for the study designing, data analysis, prepared and the interpretation of data analysis, writing and proofreading. S.A.G.L. performed data collection and literature search, data analysis and figures. A.V.E.: neuroscience’s expertise contributed to study development, writing, and proofreading. M.H.R: provided advice on the research strategy, performed bioinformatic analysis, data analysis and interpretation, manuscript writing and proofreading. K.A.P.: conceived the project, supervised the study development, performed data analysis, figures and interpretation, writing and proofreading.

Competing interests. Authors declare no conflicts of interest.

Funding. This project was supported by CONACYT-SEP Investigación en Ciencia Básica grant 254206 and CONACYT grant 88344. K.A.P. received a postdoctoral fellowship from DGAPA-UNAM, and is a current holder of support from CONACYT (CVU/227919).

Acknowledgements. K.A.P. thanks to the CABANA program for bioinformatics training. For technical support, we thank Luis Alberto Aguilar Bautista, Alejandro de León Cuevas, Carlos Sair Flores Bautista and Jair García of the Laboratorio Nacional de Visualización Científica Avanzada (LVAS).
20

doi:10.1055/s-0043-170241

20

doi:10.1055/s-0043-170241

20

doi:10.1055/s-0043-170241

20
Papillary glioblastoma exhibiting a neuroradiological cyst with a mural nodule: a case report. Medicine 98, e14102. (doi:10.1097/MD.0000000000014102)

Janik K et al. 2019 Glioblastoma with BRAFV600E mutation and numerous metastatic foci: a case report. Folia Neuropathol 57, 72–79. (doi:10.5114/fn.2019.83833)

Narasimhaiah D, Sridutt BS, Thomas B, Vilanilam GC. 2019 Glioblastoma in adults with neurofibromatosis type I: a report of two cases. Neuropathology 39, 368–373. (doi:10.1111/neup.12579)

Nørøxe DS, Michaelsen SR, Broholm H, Møller S, Poulsen HS, Lassen U. 2019 Extracranial metastases in glioblastoma—two case stories. Clin. Case Rep. 7, 289. (doi:10.1002/ccr3.1980)

Chanchotisatien A, Xiong J, Yu J, Chu S. 2019 Exophytic primary intramedullary spinal cord glioblastoma: case report and critical review of literature. World Neurosurg. 122, 573–576. (doi:10.1016/j.wneu.2018.11.113)

Cuoco JA, Klein BJ, Busch CM, Guiliams EL, Olausunrammi AL, Entwistle JJ. 2019 Corticosteroid-induced regression of glioblastoma: a radiographic conundrum. Front. Oncol. 9, 1288. (doi:10.3389/fonc.2019.01288)

Romo CG et al. 2019 Widely metastatic IDH1-mutant glioblastoma with oligodendroglial features and atypical molecular findings: a case report and review of current challenges in molecular diagnostics. Diagn. Pathol. 14, 16. (doi:10.1186/s13000-019-0793-5)

Uppar A, Konar SK, Nandeesh BN, Shukla D. 2019 H3K27M-positive primary spinal glioblastoma presenting with hemorrhage—a rare clinical entity. World Neurosurg. 126, 223–227. (doi:10.1016/j.wneu.2019.03.025)

Woo PY et al. 2019 Regression of BRAFV600E mutant adult glioblastoma after primary combined BRAF-MEK inhibitor targeted therapy: a report of two cases. Oncotarget 10, 3818. (doi:10.18632/oncotarget.26932)

Sajan A, Hiewitt K, Soleiman A, Velayudhan V. 2020 Pineal glioblastoma: case report and literature review of an exceedingly rare etiology for pineal region mass. Clin. Imaging. 60, 95–99. (doi:10.1016/j.clinimag.2019.12.002)

Gupta S et al. 2020 Multi-modality imaging assisted fluorescence-guided resection of glioblastoma: case report. Interdiscip. Neurosurg. 19, 100593. (doi:10.1016/j.inat.2019.100593)