



Article

# Recognition of a Novel Gene Signature for Human Glioblastoma

Chih-Hao Lu <sup>1,2,3,†</sup> , Sung-Tai Wei <sup>4,†</sup>, Jia-Jun Liu <sup>1</sup>, Yu-Jen Chang <sup>1</sup>, Yu-Feng Lin <sup>5</sup> , Chin-Sheng Yu <sup>6</sup> and Sunny Li-Yun Chang <sup>3,\*</sup>

<sup>1</sup> The Ph.D. Program of Biotechnology and Biomedical Industry, China Medical University, Taichung 404333, Taiwan; chlu@mail.cmu.edu.tw (C.-H.L.); u107305606@cmu.edu.tw (J.-J.L.); u107311201@cmu.edu.tw (Y.-J.C.)

<sup>2</sup> Department of Medical Laboratory Science and Biotechnology, China Medical University, Taichung 404333, Taiwan

<sup>3</sup> Graduate Institute of Biomedical Sciences, China Medical University, Taichung 404333, Taiwan

<sup>4</sup> Department of Neurosurgery, China Medical University Hospital, Taichung 404332, Taiwan; zaireeriaz@gmail.com

<sup>5</sup> Department of Medical Laboratory Science and Biotechnology, Asia University, Taichung 413305, Taiwan; yflin@asia.edu.tw

<sup>6</sup> Department of Information Engineering and Computer Science, Feng Chia University, Taichung 407102, Taiwan; yucs@fcu.edu.tw

\* Correspondence: sunnylyc@gmail.com

† These authors contributed equally to this work and share first authorship.

**Abstract:** Glioblastoma (GBM) is one of the most common malignant and incurable brain tumors. The identification of a gene signature for GBM may be helpful for its diagnosis, treatment, prediction of prognosis and even the development of treatments. In this study, we used the GSE108474 database to perform GSEA and machine learning analysis, and identified a 33-gene signature of GBM by examining astrocytoma or non-GBM glioma differential gene expression. The 33 identified signature genes included the overexpressed genes *COL6A2*, *ABCC3*, *COL8A1*, *FAM20A*, *ADM*, *CTHRC1*, *PDPN*, *IBSP*, *MIR210HG*, *GPX8*, *MYL9* and *PDLIM4*, as well as the underexpressed genes *CHST9*, *CSDC2*, *ENHO*, *FERMT1*, *IGFN1*, *LINC00836*, *MGAT4C*, *SHANK2* and *VIPR2*. Protein functional analysis by CELLO2GO implied that these signature genes might be involved in regulating various aspects of biological function, including anatomical structure development, cell proliferation and adhesion, signaling transduction and many of the genes were annotated in response to stress. Of these 33 signature genes, 23 have previously been reported to be functionally correlated with GBM; the roles of the remaining 10 genes in glioma development remain unknown. Our results were the first to reveal that GBM exhibited the overexpressed *GPX8* gene and underexpressed signature genes including *CHST9*, *CSDC2*, *ENHO*, *FERMT1*, *IGFN1*, *LINC00836*, *MGAT4C* and *SHANK2*, which might play crucial roles in the tumorigenesis of different gliomas.

**Keywords:** glioblastoma; astrocytoma; oligodendrocytoma; glioma; biomarker signature; gene signature; survival; machine learning



**Citation:** Lu, C.-H.; Wei, S.-T.; Liu, J.-J.; Chang, Y.-J.; Lin, Y.-F.; Yu, C.-S.; Chang, S.L.-Y. Recognition of a Novel Gene Signature for Human Glioblastoma. *Int. J. Mol. Sci.* **2022**, *23*, 4157. <https://doi.org/10.3390/ijms23084157>

Academic Editor: Martin Berghoff

Received: 21 March 2022

Accepted: 7 April 2022

Published: 9 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Brain tumors are among the most feared and deadliest of all forms of cancers. Primary brain tumors are categorized as glial (arising from glial cells) or nonglial (derived from diverse brain structures including nerves, blood vessels and glands), benign or malignant. Gliomas produced by glial cells are the most common and prevalent type of central nervous system (CNS) tumors. According to their histological features, gliomas are classified as astrocytomas (derived from astrocytes), oligodendroglial tumors, or ependymomas and are assigned World Health Organization (WHO) grades I–IV, according to the presence of anaplastic features indicating varying degrees of malignancy [1]. The most common

astrocytomas are pilocytic astrocytomas (WHO grade I), diffuse astrocytomas (grade II), anaplastic astrocytomas (grade III) and glioblastoma multiforme (GBM, grade IV).

In the most recent US population-based cancer registry data on primary brain and other CNS tumors diagnosed between 2014 and 2018, GBM was the most common (14.3% of all tumors and 49.1% of malignant tumors) [2]. GBM is a very aggressive tumor, with a rapid growth rate, as well as high biological and genetic heterogeneity. The prognosis is very poor, with a median survival of only 8 months [2]. Overall 1-, 2- and 5-year relative survival rates for GBM were 40.9%, 6.6% and 4.3%, respectively, amongst 86,355 cases of primary malignant and nonmalignant CNS tumors diagnosed in the USA between 2014 and 2018 [2]. Current standard treatment for GBM consists of maximal surgical resection followed by aggressive chemoradiotherapy using temozolomide, but this fails to prevent the local recurrence of almost all GBM tumors [3]. It is hoped that ongoing explorations into future therapeutic strategies, such as targeted inhibitors of specific molecular processes, nanotherapy and immunotherapy, will improve GBM treatment [3].

The heterogeneity of GBM tumors complicates their diagnoses, predictions of outcomes and decisions on treatment strategies. Recent investigations using biomedical imaging technologies to explore GBM clinical and histological features reflect the aggressiveness of such tumors and provide valuable information for improving the accuracy of prognosis [4–6]. In addition to the histological features, genetic features, such as DNA methylation data and molecular biomarkers, provide insights into the mechanisms underlying GBM and its diagnosis and have been included in the WHO classification of brain tumors [7,8]. Gene- and protein-based signatures that have been identified in bioinformatics studies have revealed gene and protein biomarkers of GBM survival [9–11]. Identified molecular biomarkers of GBM include immune-related molecules [10], cytoskeletal proteins [12,13], nonprotein-coding RNAs [14] and signal-related molecules revealing signaling pathways that play crucial roles in GBM, i.e., EGFR [15,16], VEGF [17], sonic hedgehog (SHH) [18,19], Notch [20] and Wnt/ $\beta$ -catenin signaling [21,22]. These molecular biomarkers not only greatly assist with the diagnosis and classification of GBM, but also have enormous potential as drug targets in therapeutic drug development [20,23–27]. However, it is important to realize that while changes in expression levels of many GBM biomarkers might be associated with the suppression of tumor proliferation in GBM, none of these biomarkers are capable of destroying all of the tumor cells [20,23–27]. This suggests that identifying therapeutic strategies involving biomarkers with potential to destroy tumor cells in GBM remains an urgent task.

In recent years, large-scale human genomic data emerging from several projects have been stored in public databases, including the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) repository [28,29], The Cancer Genome Atlas (TCGA) [30] and the Catalogue Of Somatic Mutations In Cancer (COSMIC) database [31]. As some researchers have proposed, the efficient storage and processing of big data may help us to extract essential information and to understand the complicated mechanisms of life [32]. Machine learning methods offer us the capacity to leverage big data and to analyze complex problems, so that we may clarify information conveyed by these data. This study has consulted records from the Rembrandt (REpository for Molecular BRAin Neoplasia DaTa) brain cancer patient-derived dataset, a joint project established by the US National Institutes of Health (NIH) National Cancer Institute (NCI) and National Institute of Neurological Disorders and Stroke (NINDS) [33]. The genomic data from this project are available on the open-access Georgetown Database of Cancer (G-DOC) platform and in the NCBI GEO repository as a super series GSE108474 [33]. Using gene set enrichment analysis (GSEA) and support vector machine (SVM) learning analysis, our analysis of the biological signatures identified a novel 19-gene GBM signature in the glioblastoma and astrocytoma gene expression data (GBM1 dataset) and a novel 17-gene GBM signature in the glioblastoma and nonglioblastoma patient samples (GBM2 dataset).

Our findings not only provide a new point of reference for the prognostic prediction of GBM, but also contributed to the understanding of molecular mechanisms in GBM

development. Furthermore, these novel signature genes might potentially be exploited as therapeutic targets for GBM.

## 2. Results

### 2.1. Recognition of Novel Gene Signatures in GBM

This study employed the machine learning method to construct prediction models and select critical genes from two datasets, GBM1 and GBM2. Each prediction model was subjected to the five-fold cross-validation technique to evaluate predictive performances. Table 1 presents predictive performances from the GBM1 and GBM2 datasets, optimized with the Matthews correlation coefficient (MCC). The performance of the GBM1 dataset was slightly superior to that of the GBM2 dataset, with an accuracy of 92%, an MCC of 0.83 and an F1 score of 0.93, which might be because the GBM1 dataset consisted of only two tumor types (glioblastoma multiforme and astrocytoma), whereas the GBM2 dataset had five (glioblastoma multiforme, astrocytoma, oligodendroglioma, mixed and unclassified).

**Table 1.** Predictive performances of the GBM1 and GBM2 datasets. All predictions were optimized using the MCC as the fitness function.

Datasets	Accuracy	Sensitivity	Specificity	MCC	Precision	F1 Score
GBM1	0.9176	0.9517	0.8648	0.8265	0.9156	0.9333
GBM2	0.8835	0.8947	0.8722	0.7672	0.8755	0.8850

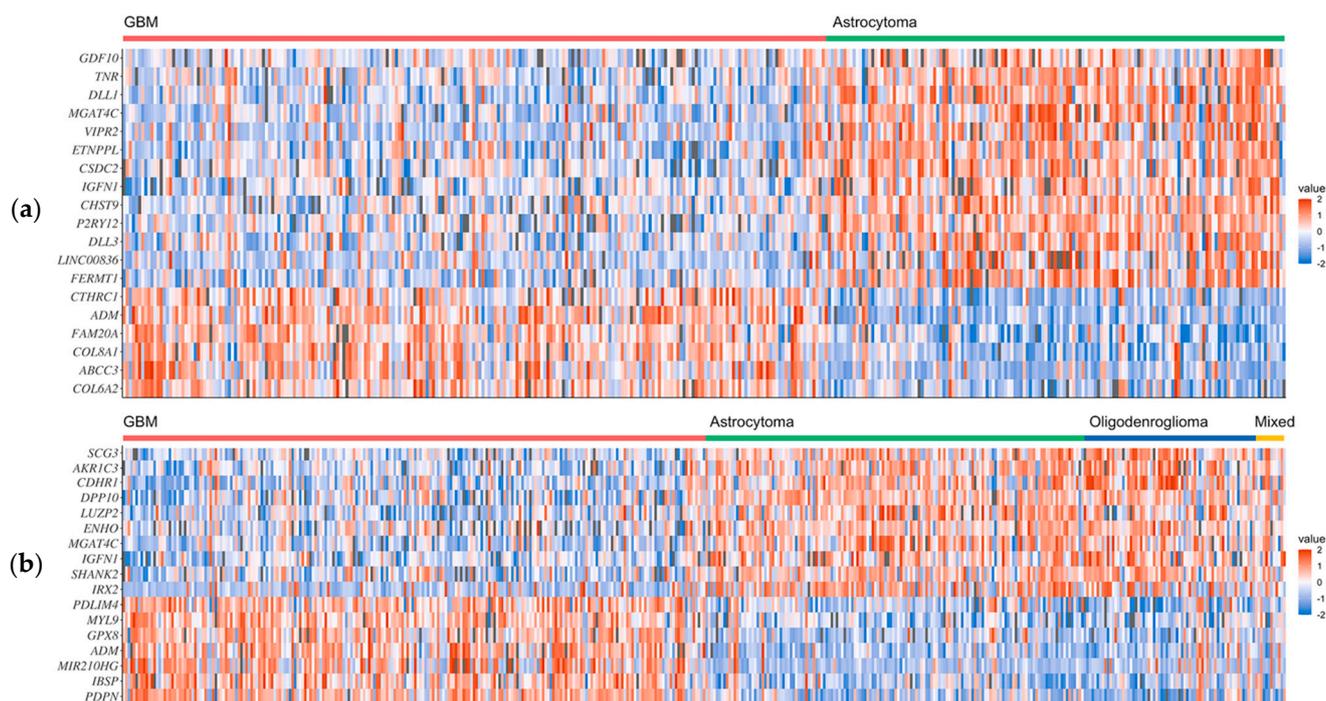
Gene scores ranged from 0 to 3, where the highest value had the most impact, representing the most frequently selected gene. A total of 19 and 17 genes scored  $\geq 2$  with their gene signatures in the GBM1 and GBM2 datasets, respectively (Table 2). Three genes, including *IGFN1*, *MGAT4C* and *ADM*, were selected in both datasets. Thus, a total of 33 GBM gene signatures were identified. Twelve of these were overexpressed in GBM, including the *MIR210HG* nonprotein-coding gene and 11 protein-coding genes (*COL6A2*, *ABCC3*, *COL8A1*, *FAM20A*, *ADM*, *CTHRC1*, *PDPN*, *IBSP*, *GPX8*, *MYL9* and *PDLIM4*) (Table 2). The underexpressed GBM genes included the *LINC00836* nonprotein-coding gene and 20 protein-coding genes (*FERMT1*, *DLL3*, *P2RY12*, *CHST9*, *IFGN1*, *CSDC2*, *ETNPPL*, *VIPR2*, *MGAT4C*, *DLL1*, *TNR*, *GDF10*, *IRX2*, *SHANK2*, *ENHO*, *LUZP2*, *DPP10*, *CDHR1*, *AKR1C3* and *SCG3*) (Table 2).

**Table 2.** Expression signatures of the 33 genes exhibiting high selective scores ( $\geq 2$ ) in the GBM1 and GBM2 datasets.

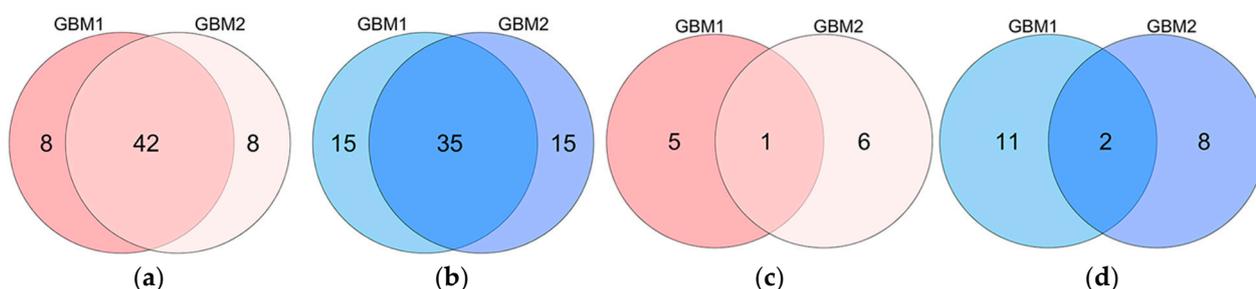
Datasets	Overexpressed Genes	Underexpressed Genes
GBM1	<i>COL6A2</i> , <i>ABCC3</i> , <i>COL8A1</i> , <i>FAM20A</i> , <i>ADM</i> <sup>1</sup> , <i>CTHRC1</i>	<i>FERMT1</i> , <i>LINC00836</i> <sup>2</sup> , <i>DLL3</i> , <i>P2RY12</i> , <i>CHST9</i> , <i>IGFN1</i> <sup>1</sup> , <i>CSDC2</i> , <i>ETNPPL</i> , <i>VIPR2</i> , <i>MGAT4C</i> <sup>1</sup> , <i>DLL1</i> , <i>TNR</i> , <i>GDF10</i>
GBM2	<i>PDPN</i> , <i>IBSP</i> , <i>MIR210HG</i> <sup>2</sup> , <i>ADM</i> <sup>1</sup> , <i>GPX8</i> , <i>MYL9</i> , <i>PDLIM4</i>	<i>IRX2</i> , <i>SHANK2</i> , <i>IGFN1</i> <sup>1</sup> , <i>MGAT4C</i> <sup>1</sup> , <i>ENHO</i> , <i>LUZP2</i> , <i>DPP10</i> , <i>CDHR1</i> , <i>AKR1C3</i> , <i>SCG3</i>

<sup>1</sup> Selected by both the GBM1 and GBM2 datasets. <sup>2</sup> Nonprotein-coding RNA genes.

The microarray expression data of the 19 GBM1 and 17 GBM2 selected genes are shown as heatmaps in Figure 1a, b, respectively. The GSEA analysis identified intersections of 42 overexpressed genes and 35 underexpressed genes from the combined GBM1 and GBM2 datasets (Figure 2). In the 33-gene signature, our methods selected only one overexpressed (*ADM*) and two underexpressed (*IGFN1* and *MGAT4C*) intersection genes. Twelve overexpressed and 21 underexpressed signature genes were recognized from 58 overexpressed and 65 underexpressed genes in both GBM1 and GBM2 datasets. Of particular interest was that only three intersection genes were recognized as signature genes from 77 intersection genes, while 30 genes were recognized by 46 genes that were either in the GBM1 or GBM2 dataset.



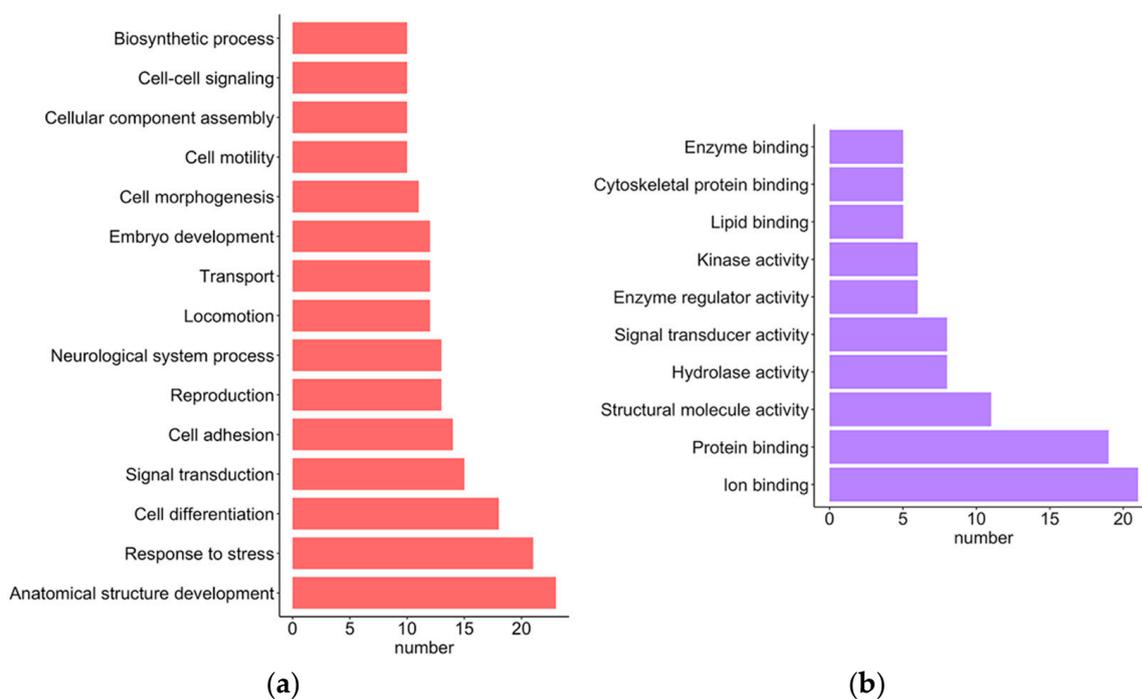
**Figure 1.** Heatmaps of the selected signature genes of GBM from the (a) GBM1 and (b) GBM2 datasets.



**Figure 2.** Venn diagrams of the numbers of original (a,b) and selected (c,d) genes in the GBM1 and GBM2 datasets, respectively. (a,c) represents the overexpressed genes in GBM samples; (b,d) represents the underexpressed genes in GBM samples.

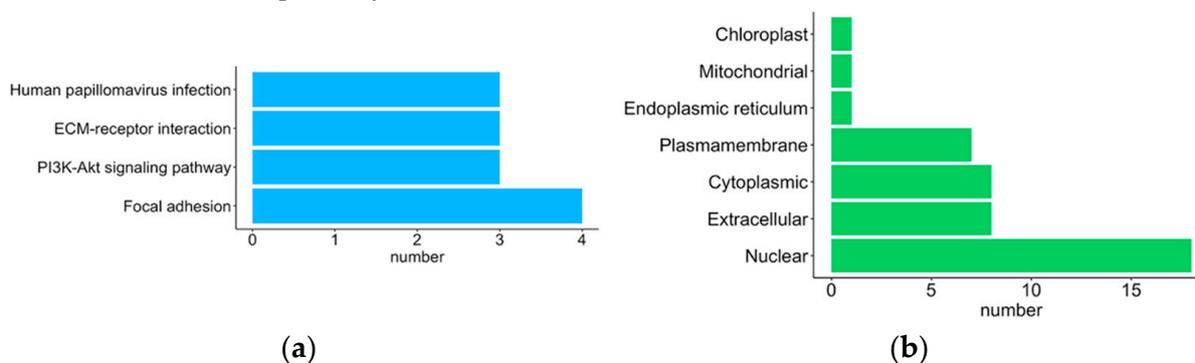
### 2.2. Functional Analysis

We performed the protein functional analysis with CELLO2GO [34], a web server that uses BLAST homology searching approaches to discern functional gene-ontology (GO) annotation. Two GO functional annotation categories, biological processes (BPs) and molecular functions (MFs) were retrieved for each selected protein-coding gene signature. An analysis of the selected 31 protein-coding gene signatures revealed that they were annotated to 50 and 24 GO-enriched groups in the BP and MF categories, respectively (Supplementary Materials). Statistically, 15 GO terms for the BP category retrieved over 10 annotated proteins. The top five GO terms retrieved were “anatomical structure development”, “response to stress”, “cell differentiation”, “signaling transduction” and “cell adhesion” (Figure 3a; Supplementary Materials). Ten of the retrieved GO terms in the MF category had more than five annotated proteins; the top three were “ion binding”, “protein binding” and “structural molecule activity” (Figure 3b; Supplementary Materials). Notably, more than half of the proteins encoded by signature genes were annotated to the “ion binding” and “protein binding” GO terms associated with the MF category.



**Figure 3.** The GO annotations of the proteins encoded by selected signature genes of GBM. The lengths of the bars represent the number of proteins involved. (a) The GO terms of “biological process”, which involve more than 10 proteins. (b) The GO terms of “molecular function”, which involve more than five proteins.

The KEGG database [35] was used to identify potential signaling pathways involving the selected protein-coding gene signatures. Four of the retrieved signaling pathways (“focal adhesion”, “PI3K-Akt signaling pathway”, “ECM-receptor interaction” and “human papillomavirus infection”) each had more than two annotated proteins (Figure 4a; Supplementary Materials). CELLO2GO also predicted protein subcellular localization. Proteins encoded by the 31 gene signatures were mostly located in the nuclear (18 proteins), extracellular (8 proteins) or cytoplasmic regions (8 proteins), or were sited within the plasma membrane (7 proteins) (Figure 4b; Supplementary Materials). CELLO2GO predicted the locations of proteins SCG3 and CHST9 in the endoplasmic reticulum and mitochondria, respectively.



**Figure 4.** The KEGG pathways and predicted subcellular localization of the proteins encoded by selected signature genes of GBM. The lengths of the bars represent the number of proteins involved. (a) The KEGG pathways, which involve more than three proteins. (b) Predicted subcellular localization by CELLO2GO.

### 2.3. Roles of the Overexpressed GBM Gene Signatures

A literature search revealed that all of the selected overexpressed gene signatures were differentially expressed in gliomas and associated with either poor prognosis or low survival rates, except for the *GPX8* gene (Table 3). Of the 12 overexpressed genes, *ABCC3*, *COL8A1*, *IBSP* and *PDPN* are reportedly associated with GBM [10], while *CTHRC1*, *PDLIM4* and *MYL9* are associated with high-grade and malignant gliomas, respectively [36,37]. These findings were consistent with the evidence of their overexpression in GBM revealed by our study, except for *ABCC3*, which in one study was apparently expressed at lower levels in GBM tissue compared with normal brain tissue samples [38]. Correlations between the *ABCC3*, *COL6A2* and *FAM20A* genes and low-grade glioma suggested that these genes might play more complex roles in gliomas [38–40].

**Table 3.** Roles/functions of overexpressed GBM gene signatures in glioma.

Gene	Encoded Protein	Reported Roles/Functions in Gliomas	References
<i>ABCC3</i>	ATP-binding cassette subfamily C member 3	Lower expression in GBM tissue versus normal brain tissue; low expression is associated with low survival rates. Upregulated expression correlates with poor overall survival in GBM.	Su et al., 2020 [38] Jiang et al., 2021 [10]
<i>ADM</i>	Proadrenomedullin	Positively regulated by STAT-3 signaling; enhances the migration of astroglia cells.	Lim et al., 2014 [41]
<i>COL6A2</i>	Collagen alpha-2(VI) chain	High expression associated with worse prognosis; induces tumor cell proliferation in recurrent and high-risk low-grade glioma.	Chen et al., 2020 [39]
<i>COL8A1</i>	Collagen alpha-1(VIII) chain	High expression correlated with poor overall survival in GBM.	Jiang et al., 2021 [10]
<i>CTHRC1</i>	Collagen triple helix repeat containing protein-1	Increased expression in glioma tissue is associated with WHO disease stage; regulates tumor cell invasion, migration and adhesion.	Mei et al., 2017 [42]
<i>FAM20A</i>	Pseudokinase FAM20A	Biomarker for low-grade glioma; overexpression predicts poor outcomes. Associated with disrupted immune responses in the GBM microenvironment.	Feng et al., 2021 [40] Du et al., 2020 [43]
<i>GPX8</i>	Glutathione peroxidase 8	N/A.	
<i>IBSP</i>	Integrin-binding bone sialoprotein 2	High expression correlated with poor overall survival in GBM.	Jiang et al., 2021 [10]
<i>MIR210HG</i>	Nonprotein-coding gene	Identified as an EMT-related lncRNA in gliomas. Serves as a biomarker for glioma diagnosis.	Tao et al., 2021 [44] Min et al., 2016 [45]
<i>MYL9</i>	Myosin regulatory light polypeptide 9	High expression is associated with a poor prognosis and is increased in patients with recurrent disease. The DAPK1-ITPRIP-MYL9 complex promotes the progression of malignant glioma.	Kruthika et al., 2019 [13] Cao et al., 2021 [37]
<i>PDLIM4</i>	PDZ and LIM domain protein 4	Biomarker for high-grade glioma.	de Tayrac et al., 2013 [36]
<i>PDPN</i>	Podoplanin	Correlated with poor overall survival in GBM. Increases tumor cell migration and angiogenesis in malignant glioma.	Jiang et al., 2021 [10] Grau et al., 2015 [46]

### 2.4. Roles of the Underexpressed GBM Gene Signatures

Of the 21 underexpressed genes, 13 (*AKR1C3*, *CDHR1*, *DLL1*, *DLL3*, *DPP10*, *ETNPPL*, *GDF10*, *IRX2*, *LUZP2*, *P2RY12*, *SCG3*, *TNR* and *VIPR2*) are reportedly associated with gliomas (Table 4). These genes are underexpressed in GBM and overexpressed in low-grade gliomas, including diffuse astrocytomas, anaplastic astrocytoma and oligodendrocytoma. Observed correlations between decreasing levels of expression of these genes and increasingly malignant gliomas suggested that these genes might be worth targeting in glioma treatment [20,24–27,47–49]. No published evidence was available as to the roles of the remaining eight underexpressed gene signatures (*CHST9*, *CSDC2*, *ENHO*, *FERMT1*, *IGFN1*, *LINC00836*, *MGAT4C* and *SHANK2*) regarding glioma pathogenesis or prognosis.

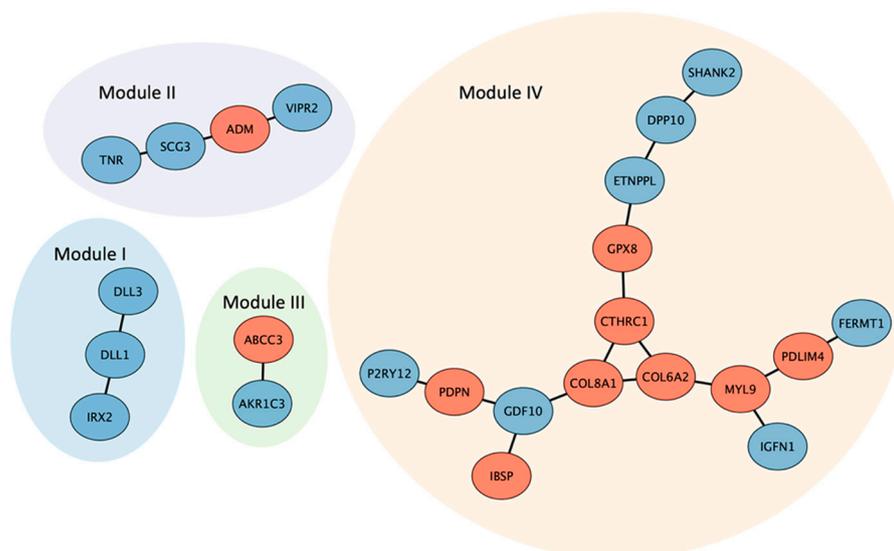
**Table 4.** Roles/functions of the underexpressed GBM gene signatures in glioma.

Gene	Encoded Protein	Reported Roles/Functions in Gliomas	References
<i>AKR1C3</i>	Aldo-keto reductase family 1 member C3	A hormone activity regulator and prostaglandin F synthase that is expressed in GBM and oligodendrogliomas; associated with the duration of overall survival in patients with gliomas.	Park et al., 2010 [50]
<i>CDHR1</i>	Cadherin-related family member 1	Downregulated in GBM and other gliomas (compared with normal brain tissue); lower expression of <i>CDHR1</i> is associated with worse clinical prognosis in GBM.	Wang et al., 2021 [24]
<i>CHST9</i>	Carbohydrate sulfotransferase 9	N/A.	
<i>CSDC2</i>	Cold shock domain-containing protein C2	N/A.	
<i>DLL1</i>	Delta-like ligand 1	Contributes to Notch signaling, which suppresses glioma stem cell differentiation and maintains their stem cell properties that contribute to GBM tumorigenesis.	Bazzoni et al., 2019 [20], Talukdar et al., 2016 [25]
<i>DLL3</i>	Delta-like ligand 3	An inhibitory ligand-driven activation of the Notch pathway and is a potent prognostic factor for malignant glioma; low <i>DLL3</i> expression is linked to shorter overall survival.	Maimaiti et al., 2021 [26]
<i>DPP10</i>	Inactive dipeptidyl peptidase 10	Underexpressed in GBM but overexpressed in diffuse astrocytomas and anaplastic astrocytomas.	Gonzalez-Garcia et al., 2020 [51]
<i>ENHO</i>	Adropin (energy homeostasis-associated protein)	N/A.	
<i>ETNPPL</i>	Ethanolamine phosphate phospholyase	Underexpressed in GBM but overexpressed in diffuse astrocytomas and anaplastic astrocytomas.	Gonzalez-Garcia et al., 2020 [51]
<i>FERMT1</i>	Fermitin family member 1	N/A.	
<i>GDF10 (BMP3b)</i>	Growth differentiation factor 10	Associated with progression-free survival in GBM in a gender-dependent manner (PFS probability falls faster in males with high <i>GDF10</i> expression than in females).	Serao et al., 2011 [47]
<i>IFGN1</i>	Immunoglobulin-like and fibronectin type III domain-containing protein 1	N/A.	
<i>IRX2</i>	Iroquois-class homeodomain protein <i>IRX-2</i>	Biomarker of pilocytic astrocytoma localization.	Antonelli et al., 2018 [52]
<i>LINC00836</i>	Long intergenic nonprotein-coding RNA 836	N/A.	
<i>LUZP2</i>	Leucine zipper protein 2	Crucial for nervous system extracellular matrix development; downregulated expression corresponds with increasing tumor stage in low-grade gliomas.	Chen et al., 2020 [48]
<i>MGAT4C</i>	$\alpha$ -1,3-mannosyl-glycoprotein 4- $\beta$ -N-acetylglucosaminyltransferase C	N/A.	
<i>P2RY12</i>	P2Y purinoceptor 12	A specific marker for resident microglia in gliomas; its expression and localization correspond with tumor stage and M1/M2 immune responses.	Zhu et al., 2017 [53]
<i>SCG3</i>	Secretogranin III	Expression is inversely correlated with malignancy grade; high in oligodendrogliomas and low in GBM.	Wang et al., 2021 [27]
<i>SHANK2</i>	SH3 and multiple ankyrin repeat domains 2	N/A.	
<i>TNR</i>	Tenascin-R	Low expression in GBM; <i>TNR</i> dysregulation in GBM is associated with glioma malignancy.	Bi et al., 2017 [49]
<i>VIPR2</i>	Vasoactive intestinal polypeptide receptor 2	Overexpressed in gliomas, particularly in oligodendrogliomas.	Jaworski et al., 2000 [54]

### 2.5. Protein–Protein Interaction Network Analysis Using STRING

We used STRING (Search Tool for the Retrieval of Interacting Genes; <https://string-db.org>, accessed on 1 March 2022) to assess potential interactions of the proteins encoded by the 31 GBM gene signatures. The predicted protein–protein interaction (PPI) networks

contained four linkage groups, modules I, II, III and IV, containing 3, 4, 2, and 15 nodes, respectively (Figure 5). All three members of module I (IRX2, DLL1 and DLL3) were encoded by underexpressed genes in GBM and all were annotated with the GO terms of “embryo development”, “anatomical structure formation involved in morphogenesis”, “anatomical structure development”, “cell differentiation” and “reproduction” in the BP category (Supplementary Materials).



**Figure 5.** Four linkage modules (I, II, III and IV) of predicted protein–protein interaction networks by STRING. The nodes in red and blue represent the proteins encoded by the over- and underexpressed GBM signature genes, respectively.

Members in module II (TNR, SCG3, ADM and VIPR2) were all annotated to the GO terms of “transport”, “vesicle-mediated transport” and “response to stress” in the BP category. Module III, containing a two-node linkage of AKR1C3 and ABCC3, was also annotated to the GO term “response to stress”, in addition to the “lipid metabolic”, “biosynthetic process”, “catabolic process” and “metabolic process” GO terms in the BP category (Supplementary Materials).

Of particular interest was that the largest module (IV) of 15 proteins was centrally linked by COL6A2, COL8A1 and CTHRC1, a triangular interactive network of collagen-related proteins that were all overexpressed in GBM and have previously been reported as correlating with increasing tumor cell invasion, migration and adhesion in GBM (Table 3, Figure 4) [10,39,42]. The other five nodes that reportedly correlate with GBM malignancy (PDLIM4, MYL9, PDPN, IBSP and GPX8) [10,13,36,37,46] were all (with the exception of GPX8) overexpressed in the GBM gene signatures (Figure 4). Among the 15 node proteins, only GPX8 and DPP10 were not annotated to the GO term “anatomical structure development”. The majority of nodes (7 of 10) in the COL6A2 and COL8A1 side chains were annotated to the GO term “cell adhesion” in the BP category (MYL9, GDP10 and P2RY12 were not).

### 3. Discussion

This study found a novel 33-gene GBM signature by using the GSE108474 database to compare differentially expressed genes amongst GBM and astrocytomas or non-GBM gliomas. Notably, there were more overexpressed than underexpressed intersection genes in the selected genes of both the GBM1 and GBM2 datasets. This was not unexpected because these datasets shared the same GBM patient samples and there might be a greater association between the overexpressed genes and GBM, while the underexpressed genes could be more associated with the partially overlapping gliomas between the datasets. Our

results implied that our methodological approaches were independent and were not biased by the intersection genes.

Of the 11 protein-coding, overexpressed signature genes, only *GPX8* was not associated with either GBM or glioma (Table 3). *GPX8* is an endoplasmic reticulum-resident protein that plays an important role in various cancers, such as gastric, breast and non-small cell lung cancer [55–57]. FoxC1-induced transcriptional activation of *GPX8* activates Wnt signaling and subsequent gastric cancer cell proliferation [55]. The Wnt/ $\beta$ -catenin signaling pathway has pleiotropic functions in neurogenesis and is one of the main signaling pathways in glioma tumorigenesis [58,59]. High levels of FoxC1 expression have been found in gliomas and FoxC1 may regulate epithelial-to-mesenchymal transition (EMT) via Wnt/ $\beta$ -catenin signaling [60]. Furthermore, evidence has shown that *GPX8* maintains an aggressive breast cancer phenotype by regulating the interleukin 6 (IL-6)/JAK/STAT3 signaling pathway [57], which is hyperactivated in many different malignancies and is generally linked to a poor prognosis [61]. Indeed, aberrantly activated STAT3 signaling is positively correlated with tumor grade and survival rates of patients with GBM [62]. Future research could usefully investigate whether the overexpression of *GPX8* in GBM is activated by FoxC1 and consequently regulates GBM tumorigenesis via the Wnt/ $\beta$ -catenin pathway or via *GPX8*/IL-6/STAT3 signaling.

The identification of the underexpressing genes in GBM might contribute to the development of a therapeutic strategy. Twelve of the 13 protein-coding, underexpressed signature genes were associated with GBM or other gliomas (Table 4). Many of them, such as *CDHR1*, *DLL1*, *DLL3* and *SCG3*, have been reported to be potential therapeutic targets for glioma treatment [20,23–27]. We failed to find any reports in the literature revealing the role of the other seven selected protein-coding, underexpressed signature genes (*CHST9*, *CSDC2*, *ENHO*, *FERMT1*, *IGFN1*, *MGAT4C* and *SHANK2*). However, all reportedly play crucial roles in various tumors and thus may also play important roles in gliomas.

Genetic variants of *CHST9* contribute to the prognosis of triple-negative breast cancer [63] and the copy number variants of *CHST9* are associated with hematologic malignancies [64]. *CSDC2* has been shown to be a potential diagnostic biomarker for early-onset colorectal cancer [65] and prostate cancer [66]. The protein encoded by the *ENHO* gene, adropin, is a secreted peptide hormone related to energy homeostasis [67] and is reportedly associated with the pathogenesis of endometrium cancer [68]. Evidence has suggested that *FERMT1* silencing inhibits oral squamous cell cancer EMT and invasion by inactivating the PI3K/AKT signaling pathway, one of the most important pathways in GBM [69]. As for *IGFN1*, although the biological function of this gene remains unclear, *IGFN1* expression has been associated with susceptibility to primary retroperitoneal liposarcoma and renal cell carcinoma, and the radiotherapy response in non-small cell lung cancer [70–72]. *MGAT4C* overexpression in benign and cancer prostate cell lines significantly increases their proliferation and migration, and increasingly higher *MGAT4C* transcript levels are associated with prostate cancer progression [73].

*SHANK2* is the most frequently amplified gene on 11q13, a major tumor amplicon in human cancer [74]. *SHANK2* plays an evolutionarily conserved role in the regulation of Hippo signaling [74], which promotes tumorigenesis and metastasis of several cancers including GBM, although scant data exist on the role of Hippo signaling in brain tumors [75,76]. *SHANK2* may be important for the development of GBM [74]. *VIPR2*, also known as VIP and PACAP receptor 2, is a receptor subtype for pituitary adenylate cyclase-activating polypeptide (PACAP) [77]. VIP and PACAP are neurotransmitters and neuromodulators that regulate neurons in various aspects, such as neuronal division, differentiation and survival [78]. Under normal culture conditions, VIP and PACAP induce the proliferation of rat GBM-derived C6 glioma cells, while under serum-starved conditions, VIP and PACAP possess antiproliferative properties [78]. As the receptor of VIP and PACAP, *VIPR2* may regulate GBM development [77]. Importantly, the associations of these eight underexpressed GBM genes with various tumors imply that they may also play crucial roles in glioma development.

This study identified two nonprotein-coding RNA genes, *MIR210HG* and *LINC00836*, as GBM signature genes. *MIR210HG* was overexpressed in GBM compared with non-GBM gliomas. Previous research has found *MIR210HG* within a nine-EMT-related lncRNA signature in patients with glioma, while other evidence suggests that *MIR210HG* is an important diagnostic biomarker for glioma [44,45]. No reports have documented the underexpression of *LINC00836* in GBM compared with astrocytomas. Nevertheless, although no evidence exists as to the differential expression of *LINC00836* in gliomas, this gene is known to be a key Alzheimer's disease-related immune hub gene that potentially contributes to immune-related phenomena in Alzheimer's disease by regulating other immune-related hub genes in the Alzheimer's brain [79]. Further analysis should explore the role of *LINC00836* in glioma pathogenesis.

#### 4. Materials and Methods

##### 4.1. Brain Tumor Dataset

GEO Series Experiment 108474 (GSE108474), which was part of the Rembrandt brain cancer project from 2004 to 2006, contains clinical and biospecimen data from 671 patients collected from 14 institutions. Records for 550 patients that had both gene expression and clinical metadata are held in the NCBI GEO database [33]. Table 5 lists these 550 samples by tumor types obtained from the NCBI GEO repository. Two datasets were collected based on the types of brain tumors documented in the metadata. The first dataset (GBM1) contained 148 astrocytoma and 228 GBM samples, while the second dataset (GBM2) contained 228 GBM samples and 227 other brain carcinoma samples, including 148 astrocytomas, 11 mixed, 67 oligodendrogliomas and 1 unclassified tumor.

**Table 5.** Clinical attributes of the GSE108474 metadata from the NCBI GEO repository.

Tumor Type	Number of Patients
Astrocytoma	148
Glioblastoma multiforme	228
Mixed	11
Oligodendroglioma	67
Unclassified <sup>1</sup>	1
Unknown <sup>2</sup>	67
Control <sup>3</sup>	28

<sup>1</sup> The type of glioma was unclassified for this patient. <sup>2</sup> The type of brain tumor was unknown in these patient samples. <sup>3</sup> Normal, healthy brain samples.

##### 4.2. Gene Feature Generation

The gene expression data of 550 patient samples in GSE108474 were sequenced on the GEO platform 570 (GPL570, Affymetrix Human Genome U133 Plus 2.0 Array [HG-U133\_Plus\_2]) containing 54,675 human probsets. The gene expression values from the microarray data were first normalized using MAS5 [80] (linear scaling) before adopting the log base 2 scale. The GSEA method [81] was then used for the preliminary screening of genes, to determine whether any statistically significant differences were apparent between GBM and other types of brain tumors. The normalized microarray data from each sample were first labeled by categorical classes, before comparing the genes inside the microarray with the *a priori* defined gene set from the REACTOME [82] subset of canonical pathways curated by the Molecular Signatures Database (MSigDB) [83]. We used the default algorithm signal-to-noise ratio to compare differences in gene expression amongst different phenotypes and ranked the genes based on the value of the formula. To evaluate the distribution of genes in the REACTOME gene set across the entire ranked list, the GSEA method was used to perform a running sum statistic progressing down the ranked list: if a gene in the ranked list belonged to the gene in the REACTOME set, the GSEA increased the accumulative score and reduced it when genes did not belong. In this process, the maximum value was denoted as the enrichment (EC) score. The EC score reflected which

genes in the REACTOME set were overexpressed at the top or bottom of the ranked list. Finally, we obtained the top 50 over- or underexpressed genes ranked by the EC score as our classification features for input into the machine learning method.

#### 4.3. Selection of Critical Genes

The genetic algorithm (GA) [84–86] was used to select critical genes and optimize classification performance. The GA procedures in this work were as follows: in the initial population, we randomly generated 80 solutions ( $S_i$ ,  $i = 1, \dots, 80$ ), where each solution  $S_i$  was represented as a set of 100-dimensional feature vectors ( $f_j^i$ ,  $j = 1, \dots, 100$ ), indicating the binary representations of 100 genes selected from GSEA. If  $f_j^i = 1$ , the  $j^{\text{th}}$  gene was kept; if  $f_j^i = 0$ , the gene  $j^{\text{th}}$  was eliminated for feeding into the SVM, a supervised learning method that used the principle of statistical risk minimization to estimate the hyperplane of a classification. All SVM calculations were performed using LIBSVM (version 3.24) [87,88], with the radial basis function (RBF) kernel. The parameters (penalty and gamma values of the RBF kernel) were both trained by exponentially increasing the grid search from  $2^{-15}$  to  $2^{15}$ , incorporating the best values of informative measures with a five-fold cross-validation during model training.

In this work, 200 generations were iterated. For each generation,  $\tau$ , the three basic mechanisms driving the evolutionary processes were performed consisting of the selection, mutation and crossover processes. The selection operators were defined as  $\alpha^\tau = \max\{S_1^\tau, \dots, S_{40}^\tau, \alpha^{\tau-1}\}$  and  $\beta^\tau = \max\{S_{41}^\tau, \dots, S_{80}^\tau, \beta^{\tau-1}\}$ . The solutions  $\alpha^\tau$  and  $\beta^\tau$  had the best fitness values in each half of 80 solutions and  $\alpha^{\tau-1}$  and  $\beta^{\tau-1}$  in the previous generation, respectively. The best solutions  $\omega^\tau$  had the best fitness values between  $\alpha^\tau$  and  $\beta^\tau$ . Note that for the special case of  $\tau = 0$ , the fitness values of  $\alpha^0$  and  $\beta^0$  were defined as 0. A new solution in the next generation  $\tau + 1$ ,  $S_i^{\tau+1}$ , was equal to  $\alpha^\tau$  if  $i$  was odd, while  $S_i^{\tau+1}$  was equal to  $\beta^\tau$  if  $i$  was even.

Four informative measures (Equations (1)–(4)) calculated using five-fold cross-validation during model training were used as the fitness functions in the selection process. They consisted of accuracy (*Acc*), the *MCC*, the F1 score (*F1*), summation of sensitivity and weighted specificity (hybrid), calculated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2)$$

$$F1 = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (3)$$

$$Hybrid = Sensitivity + \delta \times Specificity \quad (4)$$

where  $Precision = \frac{TP}{TP+FP}$ ,  $Sensitivity = \frac{TP}{TP+FN}$ ,  $Specificity = \frac{TN}{TN+FP}$ , *TP* represents true-positives, *TN* represents true-negatives, *FP* represents false-positives, *FN* represents false-negatives and  $\delta$  is the ratio of the number of positives to negatives.

After adopting the selection operators, two types of mutation were applied to all solutions ( $S_i$ s). In the first half of the solutions, every  $b$  bit of the vectors was subject to mutation:  $b = \sim b$ , if the mutation rate was less than a mutation threshold  $\mu_{mu} = 0.1$ . In the second half of the solutions, we randomly chose a bit from the 100 vectors subject to mutation:  $b = \sim b$  without any mutation thresholds. The one-point crossover operations were carried out between  $S_{2k-1}$  and  $S_{2k}$ , where  $k = 1, \dots, 40$  and proceeded as follows: the feature vectors from  $\lambda$  to 100 of  $S_{2k-1}$  and  $S_{2k}$  were swapped if the crossover rate was less than the crossover threshold  $\mu_{cr} = 0.5$ , where  $\lambda$  was randomly selected from 1 to 100.

The selection procedure of critical genes, performed by the genetic algorithm, was repeated 10 times with each informative measure. Each repeat produced 200 best solutions ( $\omega^\tau$ ,  $\tau = 1, \dots, 200$ ) for every 200 generations. Thus, we generated a total of 2000 solutions.

After deleting the redundant solutions, the top 10 were obtained from the remainder and ranked by fitness value. The selective scores ( $r_j$ ,  $j = 1, \dots, 100$ ) of 100 genes were calculated from the top 10 solutions, as follows:

$$r_j = \frac{1}{4} \sum_{p=1}^4 \left( \frac{1}{10} \sum_{q=1}^{10} f_j^q + \frac{1}{5} \sum_{q=1}^5 f_j^q + \sum_{q=1}^1 f_j^q \right) \quad (5)$$

where  $p$  is the specific informative measure (Equation  $p$ ) used as the fitness function, and  $q$  represents the top  $q$ -th solution.

## 5. Conclusions

GBM is one of the most common malignant and incurable brain tumors. The molecular mechanisms underlying the tumorigenesis of GBM, a biologically heterogeneous tumor, remain unclear. The identification of a gene signature for GBM may be helpful for the diagnosis, treatment, prediction of prognosis, and even new treatments for GBM. The identification of a 33-gene signature of GBM by GSEA and machine learning analysis revealed reported associations with GBM for 24 of these genes; the roles of the remaining nine in glioma development remain unknown. Our results were the first ever to report that the overexpressed *GPX8* gene and the underexpressed GBM *CHST9*, *CSDC2*, *ENHO*, *FERMT1*, *IGFN1*, *LINC00836*, *MGAT4C* and *SHANK2* genes might play crucial roles in the tumorigenesis of different gliomas.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23084157/s1>.

**Author Contributions:** Conceptualization, C.-H.L.; methodology, C.-H.L. and S.-T.W.; validation, S.-T.W. and C.-S.Y.; investigation, J.-J.L.; data curation, J.-J.L. and Y.-J.C.; writing—original draft preparation, Y.-F.L. and C.-H.L.; writing—review and editing, S.L.-Y.C.; visualization, J.-J.L.; supervision, S.L.-Y.C.; project administration, C.-H.L.; funding acquisition, C.-H.L. and S.L.-Y.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by C.-H.L. of China Medical University, Taiwan, grant number CMU110-ASIA-07.

**Informed Consent Statement:** Not applicable. The data used in this study was downloaded from public database (NCBI GEO database).

**Data Availability Statement:** The authors confirm that the data supporting the findings of this study are available within the article and its Supplementary Materials.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Louis, D.N.; Ohgaki, H.; Wiestler, O.D.; Cavenee, W.K.; Burger, P.C.; Jouvet, A.; Scheithauer, B.W.; Kleihues, P. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **2007**, *114*, 97–109. [[CrossRef](#)] [[PubMed](#)]
- Ostrom, Q.T.; Gittleman, H.; Farah, P.; Ondracek, A.; Chen, Y.; Wolinsky, Y.; Stroup, N.E.; Kruchko, C.; Barnholtz-Sloan, J.S. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. *Neuro Oncol.* **2013**, *15* (Suppl. 2), ii1–ii56. [[CrossRef](#)] [[PubMed](#)]
- Wu, W.; Klockow, J.L.; Zhang, M.; Lafortune, F.; Chang, E.; Jin, L.; Wu, Y.; Daldrup-Link, H.E. Glioblastoma multiforme (GBM): An overview of current therapies and mechanisms of resistance. *Pharm. Res.* **2021**, *171*, 105780. [[CrossRef](#)] [[PubMed](#)]
- Mikkelsen, V.E.; Solheim, O.; Salvesen, O.; Torp, S.H. The histological representativeness of glioblastoma tissue samples. *Acta Neurochir.* **2021**, *163*, 1911–1920. [[CrossRef](#)]
- Mobadersany, P.; Yousefi, S.; Amgad, M.; Gutman, D.A.; Barnholtz-Sloan, J.S.; Velazquez Vega, J.E.; Brat, D.J.; Cooper, L.A.D. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E2970–E2979. [[CrossRef](#)]
- Hara, A.; Kanayama, T.; Noguchi, K.; Niwa, A.; Miyai, M.; Kawaguchi, M.; Ishida, K.; Hatano, Y.; Niwa, M.; Tomita, H. Treatment Strategies Based on Histological Targets against Invasive and Resistant Glioblastoma. *J. Oncol.* **2019**, *2019*, 2964783. [[CrossRef](#)]

7. Louis, D.N.; Perry, A.; Reifenberger, G.; von Deimling, A.; Figarella-Branger, D.; Cavenee, W.K.; Ohgaki, H.; Wiestler, O.D.; Kleihues, P.; Ellison, D.W. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol.* **2016**, *131*, 803–820. [[CrossRef](#)]
8. Louis, D.N.; Perry, A.; Wesseling, P.; Brat, D.J.; Cree, I.A.; Figarella-Branger, D.; Hawkins, C.; Ng, H.K.; Pfister, S.M.; Reifenberger, G.; et al. The 2021 WHO Classification of Tumors of the Central Nervous System: A summary. *Neuro Oncol.* **2021**, *23*, 1231–1251. [[CrossRef](#)]
9. Li, L.; Liu, X.; Ma, X.; Deng, X.; Ji, T.; Hu, P.; Wan, R.; Qiu, H.; Cui, D.; Gao, L. Identification of key candidate genes and pathways in glioblastoma by integrated bioinformatical analysis. *Exp. Med.* **2019**, *18*, 3439–3449. [[CrossRef](#)]
10. Jiang, Z.; Shi, Y.; Zhao, W.; Zhang, Y.; Xie, Y.; Zhang, B.; Tan, G.; Wang, Z. Development of an Immune-Related Prognostic Index Associated with Glioblastoma. *Front. Neurol.* **2021**, *12*, 610797. [[CrossRef](#)]
11. Yin, W.; Tang, G.; Zhou, Q.; Cao, Y.; Li, H.; Fu, X.; Wu, Z.; Jiang, X. Expression Profile Analysis Identifies a Novel Five-Gene Signature to Improve Prognosis Prediction of Glioblastoma. *Front. Genet.* **2019**, *10*, 419. [[CrossRef](#)] [[PubMed](#)]
12. Zottel, A.; Jovcevska, I.; Samec, N.; Komel, R. Cytoskeletal proteins as glioblastoma biomarkers and targets for therapy: A systematic review. *Crit. Rev. Oncol. Hematol.* **2021**, *160*, 103283. [[CrossRef](#)] [[PubMed](#)]
13. Kruthika, B.S.; Sugur, H.; Nandaki, K.; Arimappamagan, A.; Paturu, K.; Santosh, V. Expression pattern and prognostic significance of myosin light chain 9 (MYL9): A novel biomarker in glioblastoma. *J. Clin. Pathol.* **2019**, *72*, 677–681. [[CrossRef](#)] [[PubMed](#)]
14. Cao, Y.; Wang, P.; Ning, S.; Xiao, W.; Xiao, B.; Li, X. Identification of prognostic biomarkers in glioblastoma using a long non-coding RNA-mediated, competitive endogenous RNA network. *Oncotarget* **2016**, *7*, 41737–41747. [[CrossRef](#)]
15. An, Z.; Aksoy, O.; Zheng, T.; Fan, Q.W.; Weiss, W.A. Epidermal growth factor receptor and EGFRvIII in glioblastoma: Signaling pathways and targeted therapies. *Oncogene* **2018**, *37*, 1561–1575. [[CrossRef](#)]
16. Saadeh, F.S.; Mahfouz, R.; Assi, H.I. EGFR as a clinical marker in glioblastomas and other gliomas. *Int. J. Biol. Mark.* **2018**, *33*, 22–32. [[CrossRef](#)]
17. Loureiro, L.V.M.; Neder, L.; Callegaro-Filho, D.; de Oliveira Koch, L.; Stavale, J.N.; Malheiros, S.M.F. The immunohistochemical landscape of the VEGF family and its receptors in glioblastomas. *Surg. Exp. Pathol.* **2020**, *3*, 9. [[CrossRef](#)]
18. Wu, X.; Xiao, S.; Zhang, M.; Yang, L.; Zhong, J.; Li, B.; Li, F.; Xia, X.; Li, X.; Zhou, H.; et al. A novel protein encoded by circular SMO RNA is essential for Hedgehog signaling activation and glioblastoma tumorigenicity. *Genome Biol.* **2021**, *22*, 33. [[CrossRef](#)]
19. Gu, W.; Shou, J.; Gu, S.; Sun, B.; Che, X. Identifying hedgehog signaling specific microRNAs in glioblastomas. *Int. J. Med. Sci.* **2014**, *11*, 488–493. [[CrossRef](#)]
20. Bazzoni, R.; Bentivegna, A. Role of Notch Signaling Pathway in Glioblastoma Pathogenesis. *Cancers* **2019**, *11*, 292. [[CrossRef](#)]
21. Latour, M.; Her, N.G.; Kesari, S.; Nurmammedov, E. WNT Signaling as a Therapeutic Target for Glioblastoma. *Int. J. Mol. Sci.* **2021**, *22*, 8428. [[CrossRef](#)] [[PubMed](#)]
22. Chen, Y.; Fang, R.; Yue, C.; Chang, G.; Li, P.; Guo, Q.; Wang, J.; Zhou, A.; Zhang, S.; Fuller, G.N.; et al. Wnt-Induced Stabilization of KDM4C Is Required for Wnt/beta-Catenin Target Gene Expression and Glioblastoma Tumorigenesis. *Cancer Res.* **2020**, *80*, 1049–1063. [[CrossRef](#)] [[PubMed](#)]
23. Sasmita, A.O.; Wong, Y.P.; Ling, A.P.K. Biomarkers and therapeutic advances in glioblastoma multiforme. *Asia Pac. J. Clin. Oncol.* **2018**, *14*, 40–51. [[CrossRef](#)] [[PubMed](#)]
24. Wang, H.; Wang, X.; Xu, L.; Lin, Y.; Zhang, J.; Cao, H. Low expression of CDHR1 is an independent unfavorable prognostic factor in glioma. *J. Cancer* **2021**, *12*, 5193–5205. [[CrossRef](#)]
25. Talukdar, S.; Das, S.K.; Pradhan, A.K.; Emdad, L.; Shen, X.N.; Windle, J.J.; Sarkar, D.; Fisher, P.B. Novel function of MDA-9/Syntenin (SDCBP) as a regulator of survival and stemness in glioma stem cells. *Oncotarget* **2016**, *7*, 54102–54119. [[CrossRef](#)]
26. Maimaiti, A.; Wang, X.; Hao, Y.; Jiang, L.; Shi, X.; Pei, Y.; Feng, Z.; Kasimu, M. Integrated Gene Expression and Methylation Analyses Identify DLL3 as a Biomarker for Prognosis of Malignant Glioma. *J. Mol. Neurosci.* **2021**, *71*, 1622–1635. [[CrossRef](#)]
27. Wang, Y.; Ji, N.; Wang, J.; Cao, J.; Li, D.; Zhang, Y.; Zhang, L. SCG3 Protein Expression in Glioma Associates With less Malignancy and Favorable Clinical Outcomes. *Pathol. Oncol. Res.* **2021**, *27*, 594931. [[CrossRef](#)]
28. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)]
29. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets-update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [[CrossRef](#)]
30. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **2008**, *455*, 1061–1068. [[CrossRef](#)]
31. Bamford, S.; Dawson, E.; Forbes, S.; Clements, J.; Pettett, R.; Dogan, A.; Flanagan, A.; Teague, J.; Futreal, P.A.; Stratton, M.R.; et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **2004**, *91*, 355–358. [[CrossRef](#)]
32. Thai, M.T.; Wu, W.; Xiong, H. *Big Data in Complex and Social Networks*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2016. [[CrossRef](#)]
33. Gusev, Y.; Bhuvaneshwar, K.; Song, L.; Zenklusen, J.C.; Fine, H.; Madhavan, S. The REMBRANDT study, a large collection of genomic data from brain cancer patients. *Sci. Data* **2018**, *5*, 180158. [[CrossRef](#)] [[PubMed](#)]
34. Yu, C.S.; Cheng, C.W.; Su, W.C.; Chang, K.C.; Huang, S.W.; Hwang, J.K.; Lu, C.H. CELLO2GO: A web server for protein subCELLular LOcalization prediction with functional gene ontology annotation. *PLoS ONE* **2014**, *9*, e99368. [[CrossRef](#)] [[PubMed](#)]

35. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **1999**, *27*, 29–34. [[CrossRef](#)] [[PubMed](#)]
36. De Tayrac, M.; Saikali, S.; Aubry, M.; Bellaud, P.; Boniface, R.; Quillien, V.; Mosser, J. Prognostic significance of EDN/RB, HJURP, p60/CAF-1 and PDL14, four new markers in high-grade gliomas. *PLoS ONE* **2013**, *8*, e73332. [[CrossRef](#)] [[PubMed](#)]
37. Cao, C.; He, K.; Li, S.; Ge, Q.; Liu, L.; Zhang, Z.; Zhang, H.; Wang, X.; Sun, X.; Ding, L. ITPRIP promotes glioma progression by linking MYL9 to DAPK1 inhibition. *Cell Signal.* **2021**, *85*, 110062. [[CrossRef](#)]
38. Sun, Z.; Qi, X.; Zhang, Y. Bioinformatics Analysis of the Expression of ATP Binding Cassette Subfamily C Member 3 (ABCC3) in Human Glioma. *Open Med.* **2020**, *15*, 107–113. [[CrossRef](#)]
39. Chen, P.Y.; Li, X.D.; Ma, W.N.; Li, H.; Li, M.M.; Yang, X.Y.; Li, S.Y. Comprehensive Transcriptomic Analysis and Experimental Validation Identify lncRNA HOXA-AS2/miR-184/COL6A2 as the Critical ceRNA Regulation Involved in Low-Grade Glioma Recurrence. *Onco Targets* **2020**, *13*, 4999–5016. [[CrossRef](#)]
40. Feng, J.; Zhou, J.; Zhao, L.; Wang, X.; Ma, D.; Xu, B.; Xie, F.; Qi, X.; Chen, G.; Zhao, H.; et al. Fam20C Overexpression Predicts Poor Outcomes and is a Diagnostic Biomarker in Lower-Grade Glioma. *Front. Genet.* **2021**, *12*, 757014. [[CrossRef](#)]
41. Lim, S.Y.; Ahn, S.H.; Park, H.; Lee, J.; Choi, K.; Choi, C.; Choi, J.H.; Park, E.M.; Choi, Y.H. Transcriptional regulation of adrenomedullin by oncostatin M in human astrogloma cells: Implications for tumor invasion and migration. *Sci. Rep.* **2014**, *4*, 6444. [[CrossRef](#)]
42. Mei, P.J.; Bai, J.; Miao, F.A.; Chen, C.; Zhu, Y.S.; Li, Z.L.; Zheng, J.N.; Fan, Y.C. CTHRC1 mediates multiple pathways regulating cell invasion, migration and adhesion in glioma. *Int. J. Clin. Exp. Pathol.* **2017**, *10*, 9318–9329. [[PubMed](#)]
43. Du, S.; Guan, S.; Zhu, C.; Guo, Q.; Cao, J.; Guan, G.; Cheng, W.; Cheng, P.; Wu, A. Secretory Pathway Kinase FAM20C, a Marker for Glioma Invasion and Malignancy, Predicts Poor Prognosis of Glioma. *Onco Targets* **2020**, *13*, 11755–11768. [[CrossRef](#)] [[PubMed](#)]
44. Tao, C.; Luo, H.; Chen, L.; Li, J.; Zhu, X.; Huang, K. Identification of an epithelial-mesenchymal transition related long non-coding RNA (lncRNA) signature in Glioma. *Bioengineered* **2021**, *12*, 4016–4031. [[CrossRef](#)] [[PubMed](#)]
45. Min, W.; Dai, D.; Wang, J.; Zhang, D.; Zhang, Y.; Han, G.; Zhang, L.; Chen, C.; Li, X.; Li, Y.; et al. Long Noncoding RNA miR210HG as a Potential Biomarker for the Diagnosis of Glioma. *PLoS ONE* **2016**, *11*, e0160451. [[CrossRef](#)]
46. Grau, S.J.; Trillsch, F.; Tonn, J.C.; Goldbrunner, R.H.; Noessner, E.; Nelson, P.J.; von Luettichau, I. Podoplanin increases migration and angiogenesis in malignant glioma. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 8663–8670.
47. Seroo, N.V.; Delfino, K.R.; Southey, B.R.; Beever, J.E.; Rodriguez-Zas, S.L. Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. *BMC Med. Genom.* **2011**, *4*, 49. [[CrossRef](#)]
48. Chen, J.; Zhao, X.; Cui, L.; He, G.; Wang, X.; Wang, F.; Duan, S.; He, L.; Li, Q.; Yu, X.; et al. Genetic regulatory subnetworks and key regulating genes in rat hippocampus perturbed by prenatal malnutrition: Implications for major brain disorders. *Aging* **2020**, *12*, 8434–8458. [[CrossRef](#)]
49. Bi, B.; Li, F.; Guo, J.; Li, C.; Jing, R.; Lv, X.; Chen, X.; Wang, F.; Azadzi, K.M.; Wang, L.; et al. Label-free quantitative proteomics unravels the importance of RNA processing in glioma malignancy. *Neuroscience* **2017**, *351*, 84–95. [[CrossRef](#)]
50. Park, A.L.; Lin, H.K.; Yang, Q.; Sing, C.W.; Fan, M.; Mapstone, T.B.; Gross, N.L.; Gumerlock, M.K.; Martin, M.D.; Rabb, C.H.; et al. Differential expression of type 2 3alpha/type 5 17beta-hydroxysteroid dehydrogenase (AKR1C3) in tumors of the central nervous system. *Int. J. Clin. Exp. Pathol.* **2010**, *3*, 743–754.
51. Gonzalez-Garcia, N.; Nieto-Librero, A.B.; Vital, A.L.; Tao, H.J.; Gonzalez-Tablas, M.; Otero, A.; Galindo-Villardón, P.; Orfao, A.; Taberner, M.D. Multivariate analysis reveals differentially expressed genes among distinct subtypes of diffuse astrocytic gliomas: Diagnostic implications. *Sci. Rep.* **2020**, *10*, 11270. [[CrossRef](#)]
52. Antonelli, M.; Fadda, A.; Loi, E.; Moi, L.; Zavattari, C.; Sulas, P.; Gentilini, D.; Cameli, C.; Bacchelli, E.; Badiali, M.; et al. Integrated DNA methylation analysis identifies topographical and tumoral biomarkers in pilocytic astrocytomas. *Oncotarget* **2018**, *9*, 13807–13821. [[CrossRef](#)] [[PubMed](#)]
53. Zhu, C.; Kros, J.M.; van der Weiden, M.; Zheng, P.; Cheng, C.; Mustafa, D.A. Expression site of P2RY12 in residential microglial cells in astrocytomas correlates with M1 and M2 marker expression and tumor grade. *Acta Neuropathol. Commun.* **2017**, *5*, 4. [[CrossRef](#)] [[PubMed](#)]
54. Jaworski, D.M. Expression of pituitary adenylate cyclase-activating polypeptide (PACAP) and the PACAP-selective receptor in cultured rat astrocytes, human brain tumors, and in response to acute intracranial injury. *Cell Tissue Res.* **2000**, *300*, 219–230. [[CrossRef](#)] [[PubMed](#)]
55. Chen, H.; Xu, L.; Shan, Z.L.; Chen, S.; Hu, H. GPX8 is transcriptionally regulated by FOXC1 and promotes the growth of gastric cancer cells through activating the Wnt signaling pathway. *Cancer Cell Int.* **2020**, *20*, 596. [[CrossRef](#)]
56. Zhang, J.; Liu, Y.; Guo, Y.; Zhao, Q. GPX8 promotes migration and invasion by regulating epithelial characteristics in non-small cell lung cancer. *Thorac. Cancer* **2020**, *11*, 3299–3308. [[CrossRef](#)]
57. Khatib, A.; Solaimuthu, B.; Ben Yosef, M.; Abu Rmaileh, A.; Tanna, M.; Oren, G.; Schlesinger Frisch, M.; Axelrod, J.H.; Lichtenstein, M.; Shaul, Y.D. The glutathione peroxidase 8 (GPX8)/IL-6/STAT3 axis is essential in maintaining an aggressive breast cancer phenotype. *Proc. Natl. Acad. Sci USA* **2020**, *117*, 21420–21431. [[CrossRef](#)]
58. Shahcheraghi, S.H.; Tchokonte-Nana, V.; Lotfi, M.; Lotfi, M.; Ghorbani, A.; Sadeghnia, H.R. Wnt/beta-catenin and PI3K/Akt/mTOR Signaling Pathways in Glioblastoma: Two Main Targets for Drug Design: A Review. *Curr. Pharm. Des.* **2020**, *26*, 1729–1741. [[CrossRef](#)]

59. Guan, R.; Zhang, X.; Guo, M. Glioblastoma stem cells and Wnt signaling pathway: Molecular mechanisms and therapeutic targets. *Chin. Neurosurg. J.* **2020**, *6*, 25. [[CrossRef](#)]
60. Cao, Q.; Wang, X.; Shi, Y.; Zhang, M.; Yang, J.; Dong, M.; Mi, Y.; Zhang, Z.; Liu, K.; Jiang, L.; et al. FOXC1 silencing inhibits the epithelial-to-mesenchymal transition of glioma cells: Involvement of betacatenin signaling. *Mol. Med. Rep.* **2019**, *19*, 251–261. [[CrossRef](#)]
61. Johnson, D.E.; O’Keefe, R.A.; Grandis, J.R. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 234–248. [[CrossRef](#)]
62. Kim, B.H.; Lee, H.; Park, C.G.; Jeong, A.J.; Lee, S.H.; Noh, K.H.; Park, J.B.; Lee, C.G.; Paek, S.H.; Kim, H.; et al. STAT3 Inhibitor ODZ10117 Suppresses Glioblastoma Malignancy and Prolongs Survival in a Glioblastoma Xenograft Model. *Cells* **2020**, *9*, 722. [[CrossRef](#)] [[PubMed](#)]
63. Yuan, J.; Zhang, N.; Zhu, H.; Liu, J.; Xing, H.; Ma, F.; Yang, M. CHST9 rs1436904 genetic variant contributes to prognosis of triple-negative breast cancer. *Sci. Rep.* **2017**, *7*, 11802. [[CrossRef](#)] [[PubMed](#)]
64. Zhao, X.; Wu, Q.; Fu, X.; Yu, B.; Shao, Y.; Yang, H.; Guan, M.; Huang, X.; Zhang, W.; Wan, J. Examination of copy number variations of CHST9 in multiple types of hematologic malignancies. *Cancer Genet. Cytogenet.* **2010**, *203*, 176–179. [[CrossRef](#)] [[PubMed](#)]
65. Mo, X.; Su, Z.; Yang, B.; Zeng, Z.; Lei, S.; Qiao, H. Identification of key genes involved in the development and progression of early-onset colorectal cancer by co-expression network analysis. *Oncol. Lett.* **2020**, *19*, 177–186. [[CrossRef](#)]
66. Xing, Q.; Liu, S.; Luan, J.; Wang, Y.; Ma, L. A novel 13 RNA binding proteins (RBPs) signature could predict prostate cancer biochemical recurrence. *Pathol. Res. Pr.* **2021**, *225*, 153587. [[CrossRef](#)] [[PubMed](#)]
67. Jasaszwilli, M.; Billert, M.; Strowski, M.Z.; Nowak, K.W.; Skrzypski, M. Adropin as A Fat-Burning Hormone with Multiple Functions-Review of a Decade of Research. *Molecules* **2020**, *25*, 549. [[CrossRef](#)]
68. Nergiz, S.; Altinkaya, S.O.; Kurt Omurlu, I.; Yuksel, H.; Kucuk, M.; Demircan Sezer, S. Circulating adropin levels in patients with endometrium cancer. *Gynecol. Endocrinol.* **2015**, *31*, 730–735. [[CrossRef](#)]
69. Wang, X.; Chen, Q. FERMT1 knockdown inhibits oral squamous cell carcinoma cell epithelial-mesenchymal transition by inactivating the PI3K/AKT signaling pathway. *BMC Oral Health* **2021**, *21*, 598. [[CrossRef](#)]
70. Miao, C.; Liu, D.; Zhang, F.; Wang, Y.; Zhang, Y.; Yu, J.; Zhang, Z.; Liu, G.; Li, B.; Liu, X.; et al. Association of FPGS genetic polymorphisms with primary retroperitoneal liposarcoma. *Sci. Rep.* **2015**, *5*, 9079. [[CrossRef](#)]
71. Verma, S.P.; Das, P. Novel splicing in IGFN1 intron 15 and role of stable G-quadruplex in the regulation of splicing in renal cell carcinoma. *PLoS ONE* **2018**, *13*, e0205660. [[CrossRef](#)]
72. Ma, Q.; Geng, K.; Xiao, P.; Zeng, L. Identification and Prognostic Value Exploration of Radiotherapy Sensitivity-Associated Genes in Non-Small-Cell Lung Cancer. *Biomed. Res. Int.* **2021**, *2021*, 5963868. [[CrossRef](#)] [[PubMed](#)]
73. Demichelis, F.; Setlur, S.R.; Banerjee, S.; Chakravarty, D.; Chen, J.Y.; Chen, C.X.; Huang, J.; Beltran, H.; Oldridge, D.A.; Kitabayashi, N.; et al. Identification of functionally active, low frequency copy number variants at 15q21.3 and 12q21.31 associated with prostate cancer risk. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6686–6691. [[CrossRef](#)] [[PubMed](#)]
74. Xu, L.; Li, P.; Hao, X.; Lu, Y.; Liu, M.; Song, W.; Shan, L.; Yu, J.; Ding, H.; Chen, S.; et al. SHANK2 is a frequently amplified oncogene with evolutionarily conserved roles in regulating Hippo signaling. *Protein Cell* **2021**, *12*, 174–193. [[CrossRef](#)]
75. Masliantsev, K.; Karayan-Tapon, L.; Guichet, P.O. Hippo Signaling Pathway in Gliomas. *Cells* **2021**, *10*, 184. [[CrossRef](#)]
76. Xue, J.; Sang, W.; Su, L.P.; Gao, H.X.; Cui, W.L.; Abulajiang, G.; Wang, Q.; Zhang, J.; Zhang, W. Proteomics reveals protein phosphatase 1gamma as a biomarker associated with Hippo signal pathway in glioma. *Pathol. Res. Pr.* **2020**, *216*, 153187. [[CrossRef](#)] [[PubMed](#)]
77. Lundberg, P.; Lundgren, I.; Mukohyama, H.; Lehenkari, P.P.; Horton, M.A.; Lerner, U.H. Vasoactive intestinal peptide (VIP)/pituitary adenylate cyclase-activating peptide receptor subtypes in mouse calvarial osteoblasts: Presence of VIP-2 receptors and differentiation-induced expression of VIP-1 receptors. *Endocrinology* **2001**, *142*, 339–347. [[CrossRef](#)]
78. D’Amico, A.G.; Scuderi, S.; Saccone, S.; Castorina, A.; Drago, F.; D’Agata, V. Antiproliferative effects of PACAP and VIP in serum-starved glioma cells. *J. Mol. Neurosci.* **2013**, *51*, 503–513. [[CrossRef](#)]
79. Xu, H.; Jia, J. Immune-Related Hub Genes and the Competitive Endogenous RNA Network in Alzheimer’s Disease. *J. Alzheimers Dis.* **2020**, *77*, 1255–1265. [[CrossRef](#)]
80. *Affymetrix Microarray Suite User Guide*, 5th ed.; Affymetrix: Santa Clara, CA, USA, 2001.
81. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
82. Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; Sidiropoulos, K.; Cook, J.; Gillespie, M.; Haw, R.; et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **2020**, *48*, D498–D503. [[CrossRef](#)]
83. Liberzon, A.; Birger, C.; Thorvaldsdottir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [[CrossRef](#)] [[PubMed](#)]
84. Lu, C.H.; Chen, Y.C.; Yu, C.S.; Hwang, J.K. Predicting disulfide connectivity patterns. *Proteins* **2007**, *67*, 262–270. [[CrossRef](#)]
85. Yu, C.S.; Lu, C.H. Identification of antifreeze proteins and their functional residues by support vector machine and genetic algorithms based on n-peptide compositions. *PLoS ONE* **2011**, *6*, e20445. [[CrossRef](#)] [[PubMed](#)]

- 
86. Liu, J.J.; Yu, C.S.; Wu, H.W.; Chang, Y.J.; Lin, C.P.; Lu, C.H. The structure-based cancer-related single amino acid variation prediction. *Sci. Rep.* **2021**, *11*, 13599. [[CrossRef](#)] [[PubMed](#)]
  87. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27. [[CrossRef](#)]
  88. Lin, C.-J. Formulations of support vector machines: A note from an optimization point of view. *Neural Comput.* **2001**, *13*, 307–317. [[CrossRef](#)]