Bioinformatics analysis of high-throughput data to validate potential novel biomarkers and small molecule drugs for glioblastoma multiforme

Fuwei Qi*, Qing Li*, Xiaojun Lu and Zhihua Chen ⓘD

Abstract

Objective: There have been no recent improvements in the glioblastoma multiforme (GBM) outcome, with median survival remaining 15 months. Consequently, the need to identify novel biomarkers for GBM diagnosis and prognosis, and to develop targeted therapies is high. This study aimed to establish biomarkers for GBM pathogenesis and prognosis.

Methods: In total, 220 overlapping differentially expressed genes (DEGs) were obtained by integrating four microarray datasets from the Gene Expression Omnibus database (GSE4290, GSE12657, GSE15824, and GSE68848). Then a 140-node protein–protein interaction network with 343 interactions was constructed.

Results: The immune response and cell adhesion molecules were the most significantly enriched functions and pathways, respectively, among DEGs. The designated hub genes *ITGB5* and *RGS4*, which have a high degree of connectivity, were closely correlated with patient prognosis, and GEPIA database mining further confirmed their differential expression in GBM versus normal tissue. We also determined the 20 most appropriate small molecules that could potentially reverse GBM gene expression, Prestwick-1080 was the most promising and had the highest negative scores.

Conclusions: This study identified *ITGB5* and *RGS4* as potential biomarkers for GBM diagnosis and prognosis. Insights into molecular mechanisms governing GBM occurrence and progression will help identify alternative biomarkers for clinical practice.

*These authors contributed equally to this study.

Corresponding author:
Zhihua Chen, The First People's Hospital of Taicang City, Taicang Affiliated Hospital of Soochow University, Suzhou 215400, P. R. China.
Email: czh19821127@163.com

The First People's Hospital of Taicang City, Taicang Affiliated Hospital of Soochow University, Suzhou, P. R. China

Introduction

Glioblastoma multiforme (GBM) is the most common malignant central nervous system tumor and is highly fatal. Although multimodal treatments for GBM, including surgical resection, chemotherapy, radiation, and targeted therapy, have improved, the overall survival of patients remains far from ideal with a median time of approximately 15 months. Additionally, after recurrence, survival time drops to below 6 months.[1–4]

Recent studies have focused on targeting tumor growth factor receptors, epigenetic regulation, angiogenesis, and immune responses to improve overall survival and decrease the recurrence rate.[5] To date, individualized therapies against molecular targets and signaling pathways that drive GBM malignancy have been unsuccessful in randomized clinical trials.[6] Thus, identifying promising novel biomarkers and understanding their molecular mechanisms will enhance our understanding of GBM initiation and progression.

Approximately 20% of GBM patients have deletions of the epidermal growth factor receptor gene (EGFR) in exons 2 to 7 that induce a constitutively-active carcinogenic variant, which is also commonly related to EGFR amplification (EGFRvIII). Results from phase III randomized trials with newly diagnosed EGFRvIII-positive GBM patients showed that rindopepimut, an EGFRvIII peptide vaccine, did not significantly prolong overall survival compared with the controls.[7]

Amplification of the platelet-derived growth factor receptor A gene (PDGFRA) is present in 15% of GBM patients. This receptor is highly active in all types of gliomas and could be the most promising target for treating GBM. However, multiple multi-kinase inhibitors that target PDGFR showed no clinical activity or response in GBM patients.[8,9] Finding new biomarkers that correlate with GBM pathogenesis and prognosis is therefore a crucial necessity.

This study integrated gene expression profiles from four Gene Expression Omnibus (GEO) datasets (GSE4290, GSE12657, GSE15824, and GSE68848) to identify differentially expressed genes (DEGs) in GBM versus adjacent normal tissue. We also mined the Connectivity Map (CMap) database for specific candidate small molecules that could conceivably reverse the DEGs in GBM, which yielded two novel biomarkers that were significantly correlated with patient diagnoses and prognoses. These new biomarkers may also help reveal the molecular mechanisms involved in GBM initiation and progression. In summary, this study was designed to deliver new insights into GBM, which is a multi-gene hereditary disease, while also exploring promising novel biomarkers for diagnosis, prognosis, and targeted therapies. Figure 1 shows the study workflow.

Materials and methods

Data resources

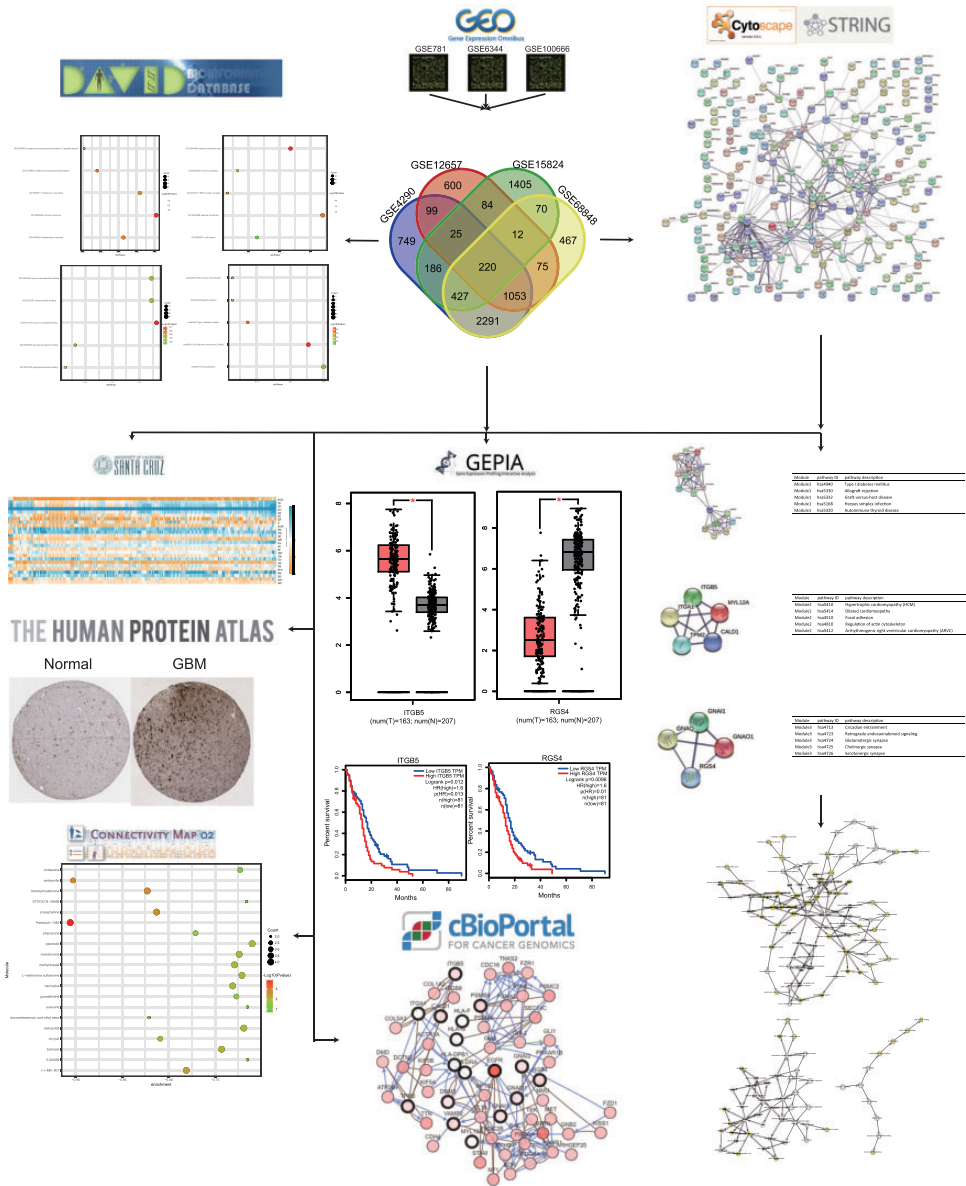This study investigated DEGs in GBM versus normal samples by analyzing

Figure 1. Study workflow to identify key genes and pathways in GBM.
GBM, glioblastoma multiforme.

GSE4290, GSE12657, GSE15824 and GSE68848 GEO gene expression profiles downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). GEO serves as a public repository for experimental high-throughput microarray data. These RNA profiles were generated with the GPL570 (Affymetrix Human Genome U113 Plus 2.0) and GPL8300 (Affymetrix Human Genome U95 Version 2) array platforms. We aggregated 345 GBM and 61 control samples for our study, with

individual profile counts being: GSE4290, 77 tumor and 23 normal; GSE12657, seven tumor and five normal; GSE15824, 33 tumor and five normal; and GSE68848, 228 tumor and 28 normal.

Identification of DEGs

Original CEL files downloaded were classified into GBM and normal groups. The Bioconductor package 'affy' (http://www.bioconductor.org/) facilitated data standardization and converted raw data into expression values. DEGs between GBM samples and normal tissues were identified *via* a significance analysis of the empirical Bayes methods, which was applied within the Limma package.[10] The cut-off criteria set to designate significant DEGs were P value <0.05 and $|logFC| >1$.

Functional enrichment analysis

Gene ontology (GO) enrichment analysis using the Database for Annotation Visualization and Integrated Discovery (DAVID) was conducted to study potential biological processes, molecular functions, and cellular components related to the overlapping DEGs. DAVID is a universal online biological information database with comprehensive annotations on gene and protein functions (version 6.7; http://david.ncifcrf.gov).[11–14] The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis further elucidated the potential signaling pathways associated with the overlapping DEGs. KEGG is a compilation of several databases with large amounts of information on genomes, diseases, biological pathways, chemical substances, and drugs that is extensively used to ascertain functional and metabolic pathways.[15] A P value <0.05 was deemed statistically significant.

Protein–protein interaction (PPI) network construction and module analysis

The Search Tool for the Retrieval of Interacting Genes (STRING; https://string-db.org/) database is designed to evaluate PPI information.[16] The previously identified DEGs were all submitted to this database to explore potential interactions; those with a combined score >0.4 were considered significant and extracted for PPI network construction in Cytoscape.[17] Molecular Complex Detection (MCODE) was subsequently used to screen significant modules from the PPI network, with the degree cutoff set as 2, node score cutoff as 0.2, k-core as 2, and maximum depth as 100.[18] Functional and pathway enrichment analyses of these modules were also conducted. The Biological Process analysis of the hub genes was visualized *via* the Cytoscape plugin Biological Networks Gene Oncology tool (BiNGO).[19]

Analysis and validation of hub genes

The UCSC Cancer Genomics Browser (http://genome-cancer.ucsc.edu) was used to construct a hierarchical clustering of module genes, and the cBioPortal online platform (http://www.cbioportal.org) established a network of co-expressed genes. The reliability of hub gene expression levels in GBM relative to normal samples was further confirmed *via* Gene Expression Profiling Interactive Analysis (GEPIA), which is an interactive web application that comprises 8587 normal and 9736 tumor samples from the Genotype–Tissue Expression and The Cancer Genome Atlas databases.[20–22] The prognostic significance of hub genes was additionally explored on the GEPIA platform. A hazard ratio for overall survival (OS) was calculated with a 95% confidence interval (CI), and Kaplan–Meier curves and boxplots used to show correlations with DEGs and patient

prognosis. The protein expression of hub genes in GBM relative to normal tissues was investigated *via* the Human Protein Atlas (HPA; www.proteinatlas.org) database, which is an online tool allowing users to analyze protein levels in clinical samples.

Identification of small molecules

CMap (http://www.broadinstitute.org/cmap/) was used to predict potential small molecules that may dampen or even reverse the existing biological status of GBM regarding its particular gene expression signature.[23] We compared the DEGs in our study with those that contribute to small active molecular interference in the CMap database to discover potentially correlated small molecules. The overlapping DEGs were initially allocated as downregulated and upregulated groups, and probe sets from each with significant differential expression were then designated for Gene Set Enrichment Analysis. The scores signified a high to low similarity amongst the compared genes, ranging correspondingly from −1 to +1, such that a positive value approaching +1 denoted the small molecules that were predicted to induce a GBM-like cell state, while a negative score nearing −1 implied that the small molecule reversed the GBM-like state of gene expression.

Results

Identification of DEGs in GBM

A Limma package analysis mined 220 overlapping DEGs in GBM samples from the GSE4290, GSE12657, GSE15824, and GSE68848 GEO datasets. Figure 2a presents the volcano plot, and Figure 2b the Venn diagrams for these DEGs.

Enrichment analyses

The biological functions of the identified DEGs were explored *via* functional and pathway enrichment analyses, using DAVID. GO analysis showed significant enrichment of up- and down-regulated DEGs in immune, inflammatory, and wound responses, as well as in antigen processing and the expression of peptide antigen. Cell component analysis revealed that these DEGs were specifically active in the plasma membrane compartment, major histocompatibility complex (MHC) protein complex, neuron projections, and cell fraction. Likewise, the molecular functions of the DEGs were significantly enriched in purine nucleotide, ribonucleotide, purine ribonucleotide, lipopolysaccharide, and glycosaminoglycan binding. Furthermore, KEGG pathway analysis suggested that these DEGs were also significantly enriched for cell adhesion molecules, type I diabetes mellitus, focal adhesion, allograft rejection, and graft-versus-host disease (Figure 3, Table 1).

PPI network construction and module analysis

Using Cytoscape, a STRING database was used to establish a PPI network of these DEGs, with 140 nodes and 343 interactions (Figure 4). The three most significant modules were extracted by MCODE (Figure 5). The signaling pathway enrichment analysis suggested the existence of significantly enriched genes of module 1 in type I diabetes mellitus, allograft rejection, and graft-versus-host disease (Figure 5a), of module 2 in hypertrophic cardiomyopathy, dilated cardiomyopathy, and focal adhesion (Figure 5b), and of module 3 in circadian entrainment, retrograde endocannabinoid signaling, and glutamatergic synapse (Figure 5c, Table 2). Hierarchical clustering suggested these module genes could essentially discern GBM and noncancerous
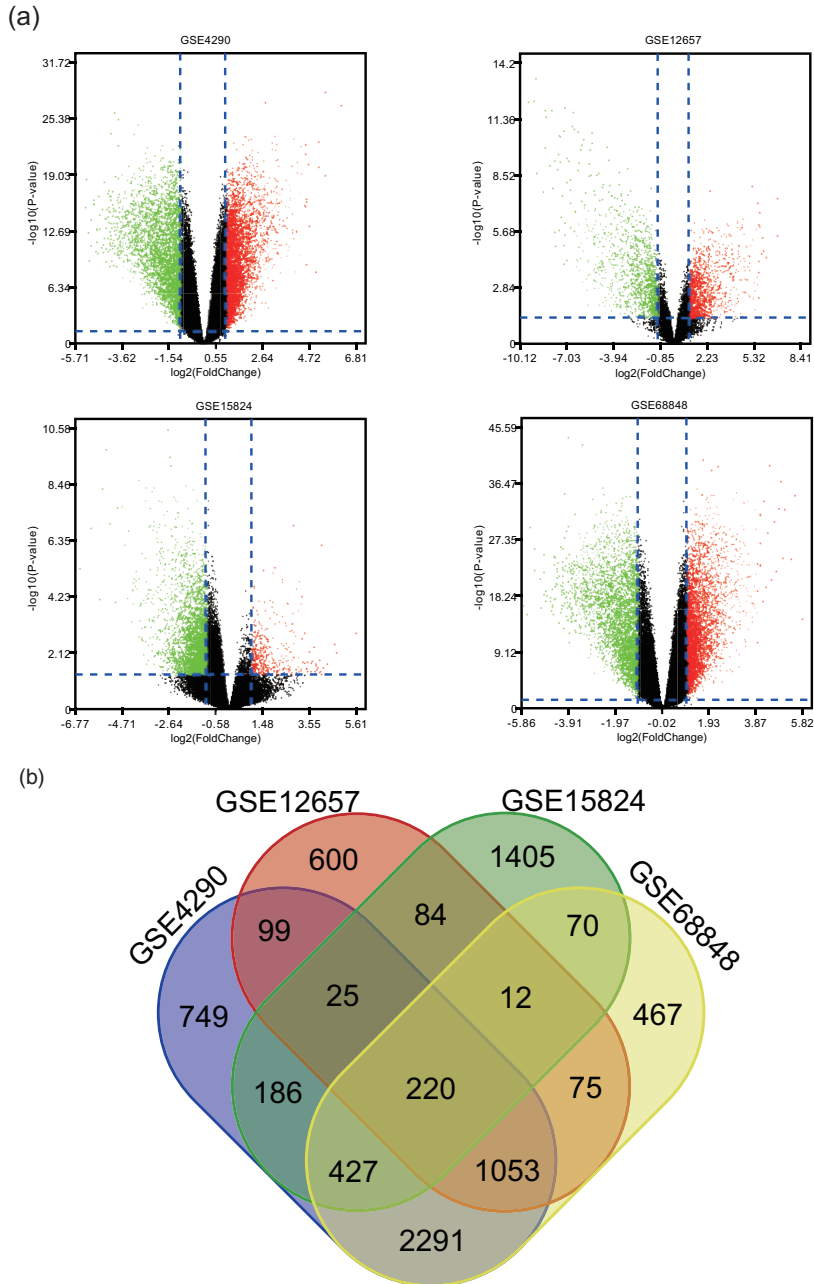
(a)

(b)

Figure 2. (a) Volcano plot of gene expression profile data between GBM and normal tissues in each dataset. Red dots: significantly upregulated genes in GBM; green dots: significantly downregulated genes in GBM; black dots: non-DEGs. $P<0.05$ and $|\log_2 FC|>1$ were considered significant. (b) Venn diagram of the 220 overlapping DEGs from the GSE4290, GSE12657, GSE15824, and GSE68848 datasets.
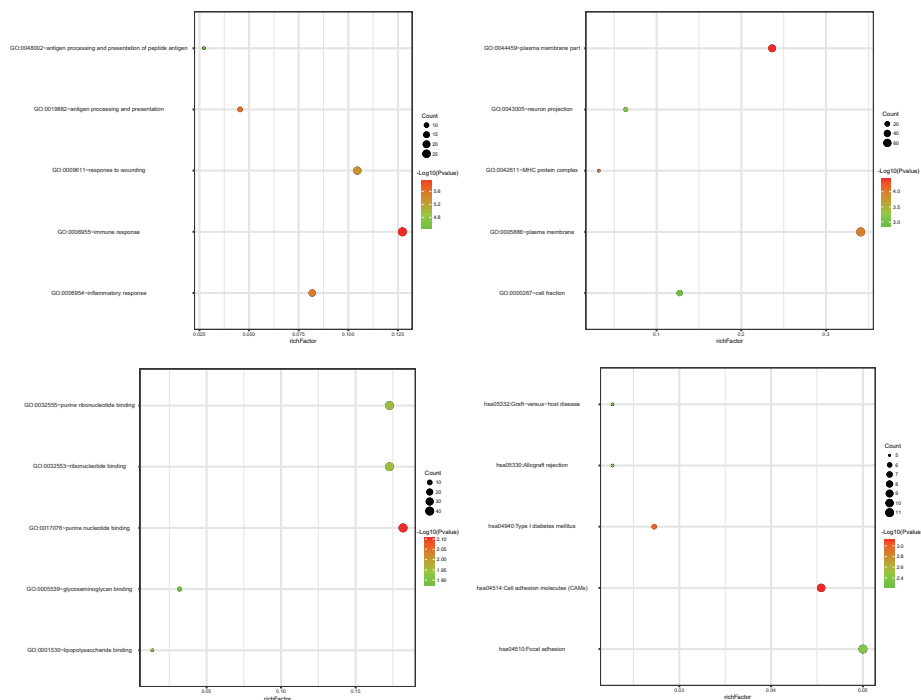GBM, glioblastoma multiforme; DEG, differentially expressed gene.

Figure 3. Functional and signaling pathway analysis of the overlapping DEGs in GBM. (a) Biological processes. (b) Cellular components. (c) Molecular function. (d) KEGG pathways. DEG, differentially expressed gene; GBM, glioblastoma multiforme; KEGG, Kyoto Encyclopedia of Genes and Genomes.

tissues (Figure 6a). Biological process analysis revealed that module genes were significantly correlated with cell division site, cell division site part, and actomyosin contractile ring (Figure 6b). Table 3 lists the full names and functions of all hub genes. *ITGB5* and *RGS4* were selected for further analysis because of their high degree of connectivity.

Analysis and validation of hub genes

The HPA and cBioPortal for Cancer Genomics databases facilitated further analysis, which validated the correlation of hub gene expression with the clinical aspects of GBM. Figure 7a shows the median expression of *ITGB5* and *RGS4* in normal and tumor samples from BodyMap. GEPIA database mining also established the significant differential expression of

ITGB5 and *RGS4* in GBM versus normal tissues. These findings corroborated that expression levels of these hub genes were closely related to GBM onset (Figure 7b).

The GEPIA database had 162 GBM patients for OS analysis, and these were divided into high- and low-expression groups. This demonstrated that *ITGB5* upregulation and *RGS4* downregulation were significantly correlated with poorer OS in GBM patients (Figure 7c), and that the expression levels of *ITGB5* and *RGS4* could collectively serve as prognostic biomarkers to predict survival in GBM patients. However, because gene expression is not always consistent with protein levels,[24] we further analyzed the protein levels of hub genes in clinical GBM tissues. Immunohistochemical staining results from the HPA database suggested considerably

Table 1. Functional and pathway enrichment analysis of overlapping DEGs.

Category	Term	Count	P-value
GOTERM_BP_FAT	GO: 0006955~immune response	28	1.04E-06
GOTERM_BP_FAT	GO: 0019882~antigen processing and presentation	10	2.16E-06
GOTERM_BP_FAT	GO: 0006954~inflammatory response	18	2.89E-06
GOTERM_BP_FAT	GO: 0009611~response to wounding	23	4.46E-06
GOTERM_BP_FAT	GO: 0048002~antigen processing and presentation of peptide antigen	6	3.73E-05
GOTERM_CC_FAT	GO: 0044459~plasma membrane part	52	3.72E-05
GOTERM_CC_FAT	GO: 0005886~plasma membrane	75	1.09E-04
GOTERM_CC_FAT	GO: 0042611~MHC protein complex	7	1.12E-04
GOTERM_CC_FAT	GO: 0043005~neuron projection	14	7.29E-04
GOTERM_CC_FAT	GO: 0000267~cell fraction	28	0.001274
GOTERM_MF_FAT	GO: 0017076~purine nucleotide binding	40	0.007919
GOTERM_MF_FAT	GO: 0032553~ribonucleotide binding	38	0.011224
GOTERM_MF_FAT	GO: 0032555~purine ribonucleotide binding	38	0.011224
GOTERM_MF_FAT	GO: 0001530~lipopolysaccharide binding	3	0.011515
GOTERM_MF_FAT	GO: 0005539~glycosaminoglycan binding	7	0.013306
KEGG_PATHWAY	hsa04514: cell adhesion molecules	10	7.24E-04
KEGG_PATHWAY	hsa04940: type I diabetes mellitus	6	0.001008
KEGG_PATHWAY	hsa04510: focal adhesion	11	0.004007
KEGG_PATHWAY	hsa05330: allograft rejection	5	0.004261
KEGG_PATHWAY	hsa05332: graft-versus-host disease	5	0.005702

DEG, differentially expressed gene; MHC, major histocompatibility complex.

higher ITGB5 positivity in cancer tissues relative to adjacent normal tissues (Figure 7d). The cBioPortal online platform facilitated construction of a network for this module and gene co-expression (Figure 7e).

We further explored molecular mechanisms of these hub genes in GBM by predicting transcription factors that could regulate their expression (Figure 8a), thereby also constructing a regulatory long non-coding (lnc)RNA, micro (mi)RNA, and mRNA network *via* Gene-Cloud Biotechnology Information analysis (Figure 8b).

Identification of potentially active small molecules

Gene Set Enrichment Analysis of up- and downregulated DEGs was uploaded to the CMap database to screen for candidate small molecules that could potentially serve as therapeutic drugs for GBM. This primarily aimed to discover small molecules that could reverse the altered gene expression in GBM. Table 4 and Figure 8c list the 20 most significant small molecules along with their enrichment scores and p values; amongst these, Prestwick-1080 (enrichment score: -0.906) and verteporfin (enrichment score: -0.903) had highly significant negative scores ($P = 0.00014$ and 0.00176, respectively), suggesting they could restore the normal cell state.

Ethics

This research was based on GEO database and bioinformatics analysis. No human and animal studies were conducted.

Discussion

The recently developed high-throughput sequencing technologies instill hope, as
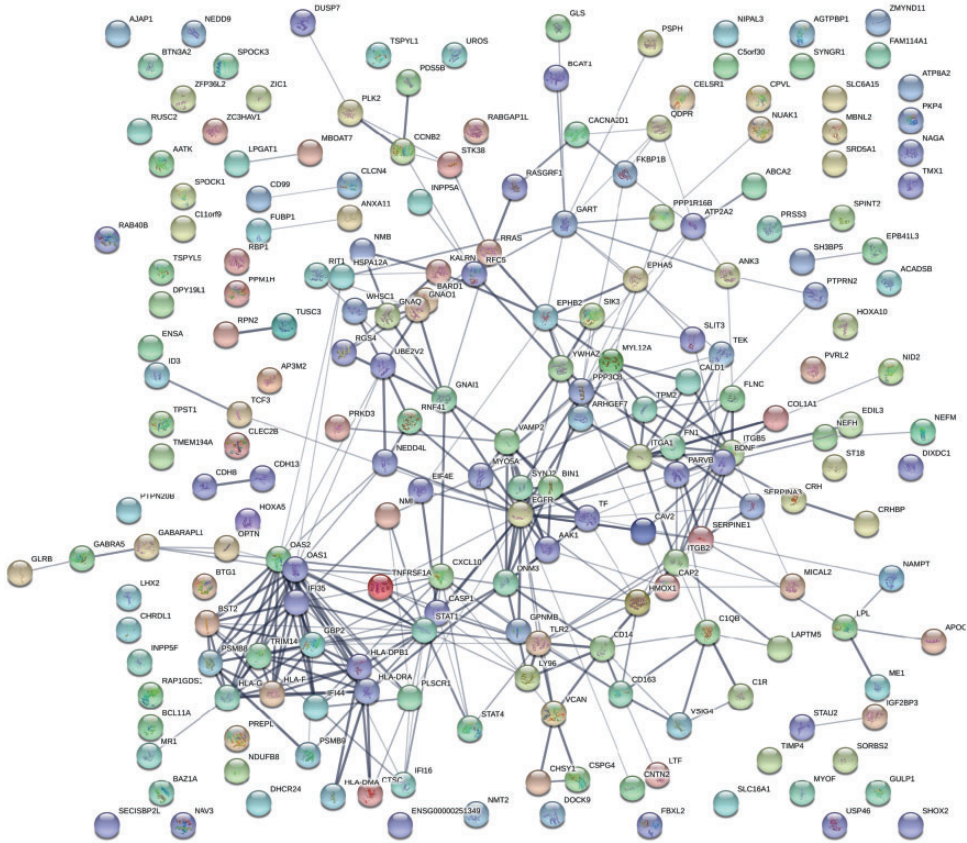
Figure 4. The constructed protein–protein interaction (PPI) network.

they can uncover vital genetic and epigenetic changes that accompany tumor initiation and progression. Integrated bioinformatics analysis plays a vital role in screening DEGs, hub node discovery from PPI networks, and prognostic analysis. This technology has been widely used to identify potential novel biomarkers related to tumor diagnosis, prognosis, and treatment.[25–27] The present study aimed to uncover novel diagnostic and prognostic biomarkers, as well as targeted therapies for GBM. We integrated four microarray datasets from the GEO database to detect DEGs between GBM and adjacent normal tissues. Additionally, we identified candidate small molecules that could reverse the tumoral status of GBM, providing new

directions and molecular mechanisms for the development of new anti-GBM drugs.

A total of 202 overlapping DEGs were identified that could be significantly associated with the occurrence and progression of GBM, suggesting they could be used as potential biomarkers for GBM diagnosis, prognosis, and treatment. We performed a GO enrichment analysis of these overlapping DEGs with regard to biological process, molecular function, and cellular component to reveal the potential molecular mechanisms underlying GBM pathogenesis. Immune response, antigen processing and presentation, and inflammatory response were the top three major functions among the biological processes. Molecular function enrichment for the DEGs was
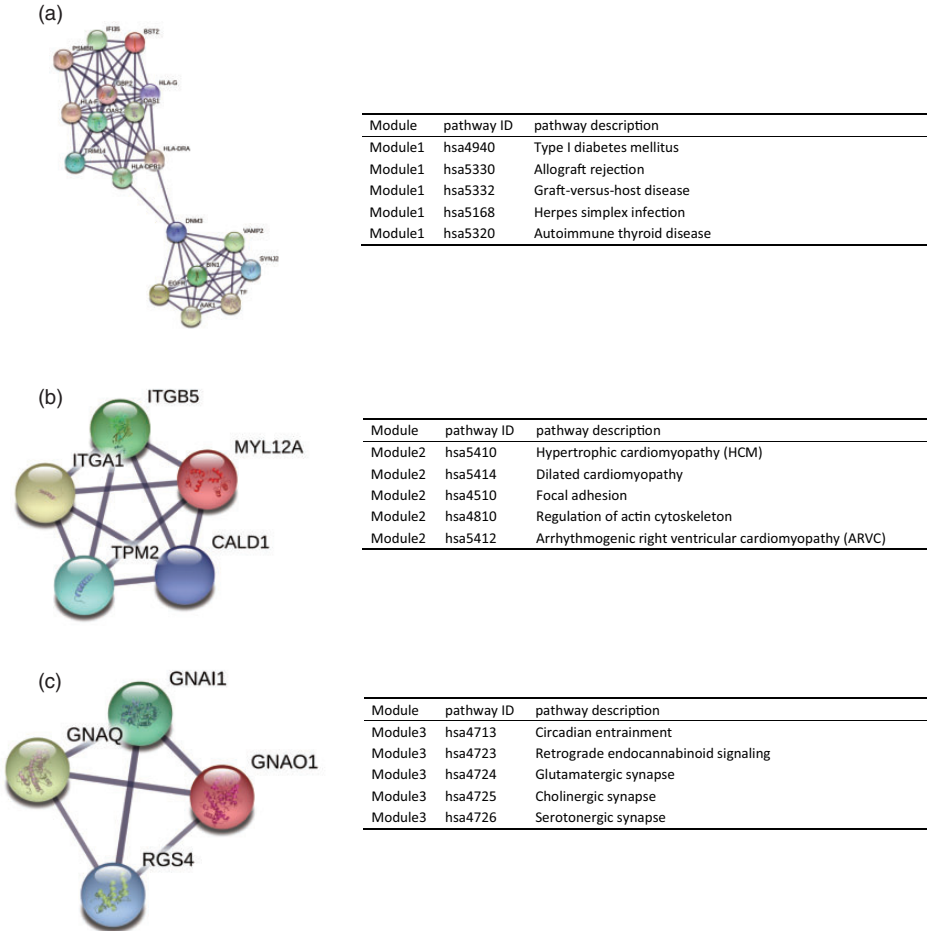
Figure 5. The top three most significant modules extracted from the PPI network and the KEGG pathway analysis of the module genes.
PPI, protein–protein interaction; KEGG, Kyoto Encyclopedia of Genes and Genomes.

**Table 2.** Functional and pathway enrichment analyses of genes in the most significant modules.

Module	Pathway ID	Pathway description	Observed gene count	False discovery rate
Module 1	hsa4940	Type I diabetes mellitus	3	0.00053
Module 1	hsa5330	Allograft rejection	3	0.00053
Module 1	hsa5332	Graft-versus-host disease	3	0.00053
Module 1	hsa5168	Herpes simplex infection	4	0.000698
Module 1	hsa5320	Autoimmune thyroid disease	3	0.000698
Module 2	hsa5410	Hypertrophic cardiomyopathy (HCM)	3	0.00011
Module 2	hsa5414	Dilated cardiomyopathy	3	0.00011
Module 2	hsa4510	Focal adhesion	3	0.000732

(continued)

Table 2. Continued

| Module	Pathway ID	Pathway description	Observed gene count	False discovery rate
Module 2	hsa4810	Regulation of actin cytoskeleton	3	0.000732
Module 2	hsa5412	Arrhythmogenic right ventricular cardiomyopathy	2	0.00716
Module 3	hsa4713	Circadian entrainment	2	0.00534
Module 3	hsa4723	Retrograde endocannabinoid signaling	2	0.00534
Module 3	hsa4724	Glutamatergic synapse	2	0.00534
Module 3	hsa4725	Cholinergic synapse	2	0.00534
Module 3	hsa4726	Serotonergic synapse	2	0.00534

(a)
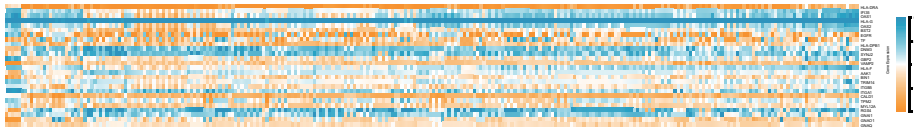
(b)
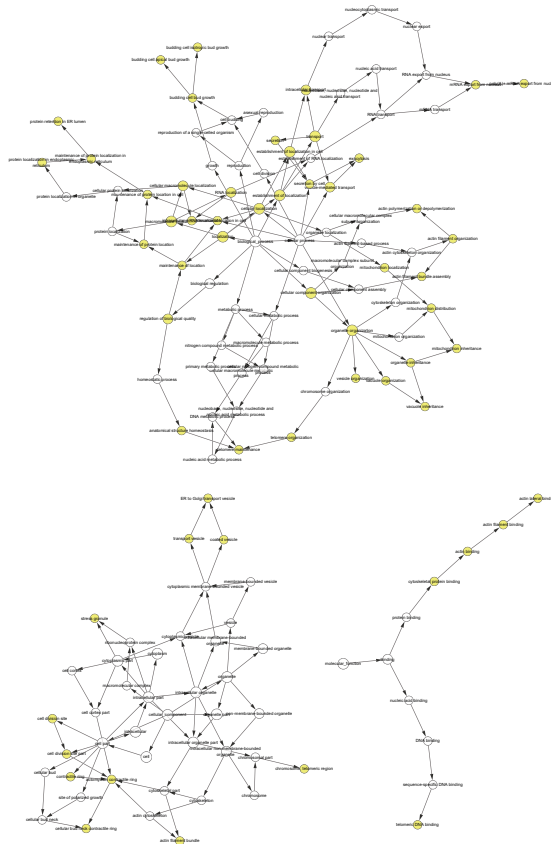
Figure 6. (a) Heatmap of module genes between GBM and normal samples. (b) The biological processes of module genes as analyzed by BiNGO. The color depth of nodes represents the corrected P-value; node size represents the number of genes involved.

GBM, glioblastoma multiforme; BiNGO, Biological Networks Gene Oncology tool.

Table 3. The full name and functional roles of eight hub genes.

Gene symbol	Full name	Function
ITGB5	Integrin Subunit Beta 5	This gene encodes a beta subunit of integrin, which can combine with different alpha chains to form a variety of integrin heterodimers. Integrins are integral cell-surface receptors that participate in cell adhesion as well as cell-surface mediated signaling. The alpha v beta 5 integrin is involved in adhesion to vitronectin.
RGS4	Regulator Of G Protein Signaling 4	Regulator of G protein signaling (RGS) family members are regulatory molecules that act as GTPase activating proteins (GAPs) for G alpha subunits of heterotrimeric G proteins. RGS proteins are able to deactivate G protein subunits of the Gi alpha, Go alpha, and Gq alpha subtypes. They drive G proteins into their inactive GDP-bound forms. Regulator of G protein signaling 4 belongs to this family. All RGS proteins share a conserved 120-amino acid sequence termed the RGS domain. Regulator of G protein signaling 4 protein is 37% identical to RGS1 and 97% identical to rat Rgs4. This protein negatively regulates signaling upstream or at the level of the heterotrimeric G protein and is localized in the cytoplasm. Alternatively spliced transcript variants have been found for this gene.

primarily within purine nucleotide binding, ribonucleotide binding, and purine ribonucleotide binding. Changes in cell components were mainly associated with the plasma membrane and MHC protein complex. Additionally, the overlapping DEGs were enriched in five KEGG pathways, including cell adhesion molecules, type I diabetes mellitus, focal adhesion, and allograft rejection.

A wide variety of cell adhesion molecules are involved in the intercellular and cell–extracellular matrix interactions of tumors. Some adhesion molecules play a vital role in tumor recurrence, metastasis, and invasion. For example, the high expression of focal adhesion kinase (FAK) has been reported to be associated with increased fibrosis and poor infiltration of CD8+ T cells, while inhibiting FAK can significantly limit tumor progression and prolong the OS of patients.[28,29] Our bioinformatics findings were consistent with this. The PPI network

among the overlapping DEGs revealed the top three most significant network modules. *ITGB5* and *RGS4* had a high degree of connectivity, and thus were selected as hub genes. To validate these results, we used the GEPIA database to determine their expression levels and prognostic value in GBM patients. Mining the GEPIA database confirmed the same trend found in the GEO database, with *ITGB5* and *RGS4* expression having a significant impact on the prognosis of GBM patients. High *ITGB5* expression was significantly correlated with a worse OS in GBM patients, while low *RGS4* expression indicated a better OS in GBM patients.

Immunohistochemical staining results from the HPA database were consistent with sequencing data, further supporting our findings. No studies have yet reported a role for these two key genes in the initiation or progression of GBM, and ours is the first to reveal their diagnostic, prognostic,
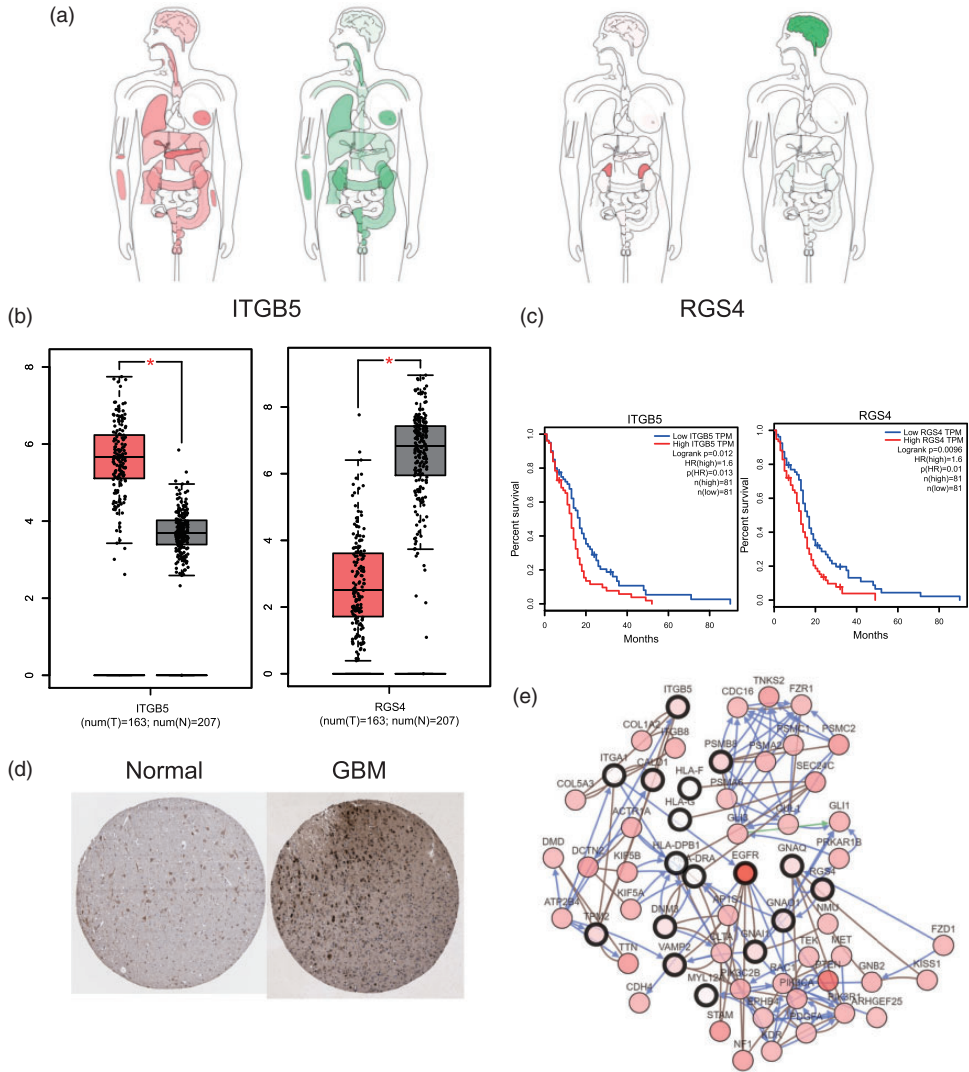
Figure 7. (a) Median expression of *ITGB5* and *RGS4* in tumor and normal samples in BodyMap. (b&c) The expression levels of *ITGB5* and *RGS4* and their prognostic value according to the GEPIA database. (d) Representative immunohistochemistry results revealing the protein levels of ITGB5 in GBM and normal tissues. (e) The network of module genes and their co-expressed genes constructed by cBioPortal. Nodes with thick outline: hub genes; nodes with thin outline: co-expressed genes.
GEPIA, Gene Expression Profiling Interactive Analysis; GBM, glioblastoma multiforme.

and therapeutic value. To further explore the potential mechanism of *ITGB5* and *RGS4* in the pathogenesis of GBM and enhance our understanding of this multigene hereditary disease, we predicted potential transcription factors that could be involved in the regulation of their expression and constructed a molecular regulatory network of lncRNAs, miRNAs, and mRNAs. *ITGB5* overexpression was previously shown to abolish the inhibition of glycolysis, growth, and proliferation induced
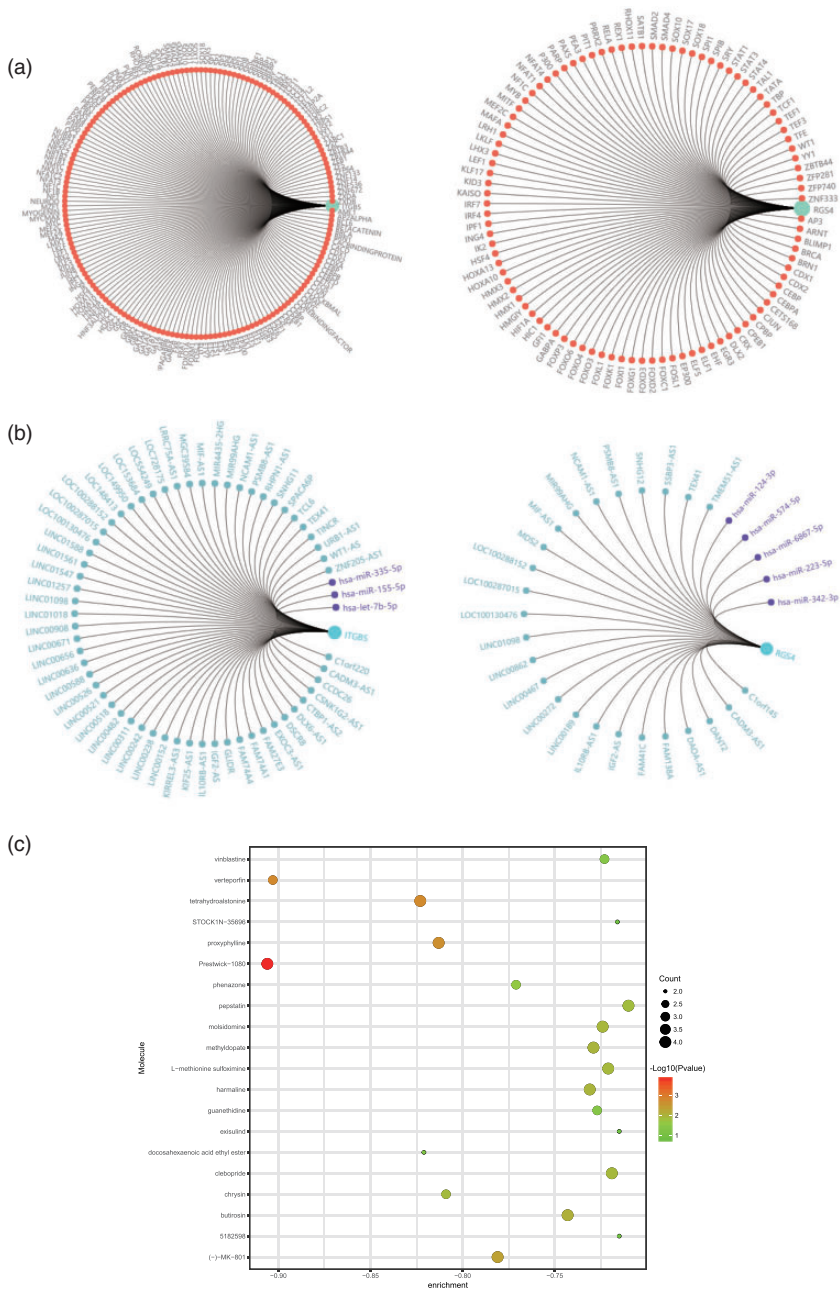
Figure 8. (a) The transcription factors that could be involved in regulating the expression of *ITGB5* and *RGS4*. (b) A regulatory network of lncRNAs, miRNAs, and mRNAs constructed by GCBI. Purple nodes: related lncRNAs; blue nodes: targeted miRNAs. (c) Pop plot of the top 20 identified small molecules that could reverse the gene expression of GBM.

GBM, glioblastoma multiforme; GCBI, Gene-Cloud of Biotechnology Information.

Table 4. Top 20 significant small molecule drugs that can reverse the tumoral status of GBM.

CMap name	Enrichment	P-value
Prestwick-1080	−0.906	0.00014
verteporfin	−0.903	0.00176
tetrahydroalstonine	−0.823	0.00185
docosahexaenoic acid ethyl ester	−0.821	0.06304
proxyphylline	−0.813	0.00225
chrysin	−0.809	0.0138
(-)-MK-801	−0.781	0.00465
phenazone	−0.771	0.02462
butirosin	−0.743	0.00871
harmaline	−0.731	0.01056
methyldopate	−0.729	0.01074
guanethidine	−0.727	0.0418
molsidomine	−0.724	0.0118
vinblastine	−0.723	0.04383
L-methionine sulfoximine	−0.721	0.01239
clebopride	−0.719	0.01265
STOCK1N-35696	−0.716	0.16108
5182598	−0.715	0.16142
exisulind	−0.715	0.16213
pepstatin	−0.71	0.01446

GBM, glioblastoma multiforme.

by cisplatin in breast and cervical cancer cells,[30] while the absence of RGS4 in pediatric nephroblastoma tissue correlates with poor prognosis. Moreover, pristimerin upregulated RGS4 expression, thereby suppressing lamellipodia formation and invasion in breast cancer cells.[31] The exact mechanisms of ITGB5 and RGS4 in GBM initiation and progression should be investigated in future studies.

Based on the overlapping DEGs and the CMap database, we identified a set of potential small molecules that could reverse the cellular status of GBM. Candidate small molecules with highly significant negative enrichment values may reverse the abnormal gene expression induced by GBM, thereby offering a fresh perspective to advance future targeted therapies. Prestwick-1080 (enrichment score: −0.906) was the most significant small molecule related to the status of GBM cells, but it

has not been investigated for its efficacy or safety in any cancer type. However, a recent study reported that verteporfin (enrichment score: −0.903) selectively kills hypoxic glioma cells by iron binding.[32] Nevertheless, studies to validate their application in GBM pathogenesis are needed to further our understanding of therapeutic mechanisms from the perspective of DEGs induced by GBM.

By mining the gene expression profiles of GBM and performing a comprehensive bioinformatics analysis, we uncovered two key genes that could contribute to understanding the molecular mechanism of GBM initiation and progression. ITGB5 and RGS4 may serve as novel biomarkers for the diagnosis, prognosis, and treatment of GBM, and a series of detailed analyses further confirmed their vital roles in GBM pathogenesis. We additionally identified a group of candidate small molecules that could

offer new insights into the molecular mechanisms of GBM and help develop new targeted therapies. In summary, we uncovered several promising novel biomarkers for GBM that provide new insights into the altered disease biology.

Author contributions

QFW, LQ, and CZH analyzed the data. LXJ and QFW wrote the article. All authors read and approved the final manuscript.

Acknowledgments

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

ORCID iD

Zhihua Chen   https://orcid.org/0000-0003-2457-4802

References

1. Stupp R, Hegi ME, Mason WP, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 2009; 10: 459–466.
2. Weller M, van den Bent M, Hopkins K, et al. EANO guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma. *Lancet Oncol* 2014; 15: e395–e403.
3. Van Meir EG, Hadjipanayis CG, Norden AD, et al. Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma. *CA Cancer J Clin* 2010; 60: 166–193.
4. Ryken TC, Kalkanis SN, Buatti JM, et al. The role of cytoreductive surgery in the management of progressive glioblastoma: a systematic review and evidence-based clinical practice guideline. *J Neurooncol* 2014; 118: 479–488.
5. Touat M, Idbaih A, Sanson M, et al. Glioblastoma targeted therapy: updated approaches from recent biological insights. *Ann Oncol* 2017; 28: 1457–1472.
6. Osuka S and Van Meir EG. Overcoming therapeutic resistance in glioblastoma: the way forward. *J Clin Invest* 2017; 127: 415–426.
7. Schuster J, Lai RK, Recht LD, et al. A phase II, multicenter trial of rindopepimut (CDX-110) in newly diagnosed glioblastoma: the ACT III study. *Neuro Oncol* 2015; 17: 854–861.
8. Lassman AB, Pugh SL, Gilbert MR, et al. Phase 2 trial of dasatinib in target-selected patients with recurrent glioblastoma (RTOG 0627). *Neuro Oncol* 2015; 17: 992–998.
9. Reardon DA, Dresemann G, Taillibert S, et al. Multicentre phase II studies evaluating imatinib plus hydroxyurea in patients with progressive glioblastoma. *Br J Cancer* 2009; 101: 1995–2004.
10. Gautier L, Cope L, Bolstad BM, et al. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004; 20: 307–315.
11. Ritchie ME, Belinda P, Di W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
12. Dennis G, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; 4: P3.
13. Consortium GO. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res* 2006; 34: 322–326.
14. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of

biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25–29.

15. Huang DW, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4: 44.

16. Kanehisa M and Goto S. KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucleic Acids Res* 2000; 28: 27–30.

17. Damian S, Andrea F, Stefan W, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015; 43: D447.

18. Smoot ME, Ono K, Ruscheinski J, et al. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011; 27: 431–432.

19. Bandettini WP, Kellman P, Mancini C, et al. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *J Cardiovasc Magn Reson* 2012; 14: 83.

20. Maere S, Heymans K and Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005; 21: 3448–3449.

21. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; 2: 401–404.

22. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; 6: pl1.

23. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017; 45: W98–W102.

24. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006; 313: 1929–1935.

25. Kulasingam V and Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol* 2008; 5: 588–599.

26. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014; 513: 202–209.

27. Liu X, Wu J, Zhang D, et al. Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis. *Front Genet* 2018; 9: 265.

28. Zhang X, Pan Y, Fu H, et al. Nucleolar and Spindle Associated Protein 1 (NUSAP1) inhibits cell proliferation and enhances susceptibility to epirubicin in invasive breast cancer cells by regulating Cyclin D Kinase (CDK1) and DLGAP5 expression. *Med Sci Monit* 2018; 24: 8553–8564.

29. Jiang H, Hegde S, Knolhoff BL, et al. Targeting focal adhesion kinase renders pancreatic cancers responsive to checkpoint immunotherapy. *Nat Med* 2016; 22: 851–860.

30. Wang S, Xie J, Li J, et al. Cisplatin suppresses the growth and proliferation of breast and cervical cancer cell lines by inhibiting integrin beta5-mediated glycolysis. *Am J Cancer Res* 2016; 6: 1108–1117.

31. Mu XM, Shi W, Sun LX, et al. Pristimerin inhibits breast cancer cell migration by upregulating regulator of G protein signaling 4 expression. *Asian Pac J Cancer Prev* 2012; 13: 1097–1104.

32. Eales KL, Wilkinson EA, Cruickshank G, et al. Verteporfin selectively kills hypoxic glioma cells through iron-binding and increased production of reactive oxygen species. *Sci Rep* 2018; 8: 14358.