Iran. J. Biotechnol. October 2024;22(4): e3892 DOI: 10.30498/ijb.2024.448826.3892

Research Article

Identification and Verification of Key Genes Associated with Temozolomide Resistance in Glioblastoma Based on Comprehensive Bioinformatics Analysis

Jun Hu[#1], Jingyan Yang[#2], Na Hu1, Zongting Shi[1], Tiemin Hu[3], Baohong Mi[1,4], Hong Wang[5], Weiheng Chen[1,4*]

[1]The Third Affiliated Hospital of Beijing University of Chinese Medicine, Beijing, China
[2]The Third Clinical School of Beijing University of Chinese Medicine, Beijing, China.
[3]Department of Neurosurgery, Affiliated Hospital of Chengde Medical University, Chengde, Hebei, China.
[4]Engineering Research Center of Chinese Orthopaedics and Sports Rehabilitation Artificial Intelligent, Ministry of Education, Beijing, China
[5]Department of Neurosurgery, Affiliated Hospital of Hebei University, Baoding, Hebei, China

[*]*Corresponding author*: Weiheng Chen, The Third Affiliated Hospital of Beijing University of Chinese Medicine, Beijing, China. Tel/Fax: +86-010-84856226, E-mail: drchenweiheng@163.com

[#]Jun Hu and Jingyan Yang contributed equally to this work.

Received: 2024/03/16 ; **Accepted:** 2024/11/26

Background: Glioblastoma (GBM) is the most aggressive form of brain cancer, with poor prognosis despite treatments like temozolomide (TMZ). Resistance to TMZ is a significant clinical challenge, and understanding the genes involved is crucial for developing new therapies and prognostic markers. This study aims to identify key genes associated with TMZ resistance in GBM, which could serve as valuable biomarkers for predicting patient outcomes and potential targets for treatment.

Objectives: This study aimed to identify genes involved in TMZ resistance in GBM and to assess the value of these genes in GBM treatment and prognosis evaluation.

Materials and Methods: Bioinformatics analysis of Gene Expression Omnibus (GEO) datasets (GSE113510 and GSE199689) and The Chinese Glioblastoma Genome Atlas (CGGA) database was performed to identify differentially expressed genes (DEGs) between GBM cell lines with and without TMZ resistance. Subsequently, the key modules associated with GBM patient prognosis were identified by weighted gene coexpression network analysis (WGCNA). Furthermore, hub genes related to TMZ resistance were accurately screened and confirmed using three machine learning algorithms. In addition, immune cell infiltration analysis, TF-miRNA coregulatory network analysis, drug sensitivity prediction, and gene set enrichment analysis (GSEA) were also performed for temozolomide resistance-specific genes. Finally, the expression levels of key genes were validated in our constructed TMZ-resistant cell lines by real-time quantitative polymerase chain reaction (RT–qPCR) and Western blotting (WB).

Results: Integrated analysis of the GEO and CGGA datasets revealed 769 differentially expressed genes (DEGs), comprising 350 downregulated and 419 upregulated genes, between GBM patients and normal controls. Among these DEGs, three key genes, namely, PITX1, TNFRSF11B, and IGFBP2, exhibited significant differences in expression between groups and were prioritized via machine learning algorithms. The expression levels of these genes were found to be closely related to adverse clinical features and immune cell infiltration levels in GBM patients. These genes were also found to participate in several biological pathways and processes. RT–qPCR and WB confirmed the differential expression of these genes in vitro, indicating that they play vital roles in GBM patients with TMZ resistance.

Conclusions: PITX1, TNFRSF11B, and IGFBP2 are key genes associated with the prognosis of GBM patients with TMZ resistance. The differential expression of these genes correlates with adverse outcomes in GBM patients, suggesting that they are valuable biomarkers for predicting patient prognosis and that they could serve as diagnostic biomarkers or treatment targets.

Keywords: Biomarkers, GEO database, Glioblastoma, Machine learning algorithm; Temozolomide resistance

Copyright *© 2024 The Author(s); Published by National Institute of Genetic Engineering and Biotechnology. This is an open access article, distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses /by-nc/4.0/) which permits others to copy and redistribute material just in noncommercial usages, provided the original work is properly cited.*

1. Background

Glioblastoma (GBM) is the most prevalent and lethal malignant primary intracranial tumor and has a high and increasing incidence rate and a dismal prognosis (1, 2). Despite advances in combined therapeutic approaches, including maximal safe tumor resection, concurrent chemoradiotherapy, and adjuvant chemotherapy, the treatment outcome and overall survival for GBM patients are still unsatisfactory (3, 4).

Temozolomide (TMZ) is the commonly used chemotherapeutic regimen after tumor resection in GBM patients. However, the major obstacles to effective treatment are acquired chemotherapy resistance and cancer recurrence after surgery (5-7). Thus, it is important to elucidate the exact molecular processes and pathological mechanisms involved in TMZ resistance. Additionally, reliable biomarkers for evaluating the response to TMZ and new targeted treatments are needed to reduce the incidence of chemotherapeutic resistance.

Weighted gene coexpression network analysis (WGCNA) has been widely used to describe coexpression patterns between genes in disease transcriptomes and to identify gene modules related to clinical characteristics (8). In contrast to the conventional approach of analyzing differentially expressed genes (DEGs), WGCNA is a robust systematic method for identifying higher-order correlations among genes rather than solely focusing on the identification of individual genes associated with diseases. Machine learning methods include least absolute shrinkage and selection operator (LASSO) analysis and analyses based on support vector machines (SVMs) and the random forest (RF) algorithm. These methods have been developed to aid in the diagnosis of cancer and the discovery of sensitizing drug targets (9, 10).

In our study, we first screened DEGs between TMZ-resistant and TMZ-sensitive GBM cell lines from the Gene Expression Omnibus (GEO) database. Next, we downloaded transcriptional data and clinical information on GBM patients from The Chinese Glioblastoma Genome Atlas (CGGA). Subsequently, functional modules related to GBM tumor status and candidate TMZ resistance-related hub genes were screened based on WGCNA and machine learning strategies. Furthermore, we performed immune infiltration analysis, TF-miRNA coregulatory network analysis, and drug sensitivity analysis to better understand the molecular mechanisms involved in the onset of TMZ resistance.

2. Objectives

The objectives of this study are to delineate the genetic landscape of TMZ resistance in GBM by identifying DEGs associated with this resistance. We aim to evaluate the clinical significance of these DEGs in terms of their potential as prognostic indicators and therapeutic targets for GBM patients. Furthermore, the study seeks to validate the expression patterns of key genes identified through bioinformatics and machine learning approaches, and to explore their functional roles and clinical relevance in TMZ-resistant GBM.

3. Materials and Methods

3.1. Raw Data Acquisition and Preprocessing

Two gene expression datasets were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The GSE113510 dataset contained three samples from TMZ-resistant cells and their parental TMZ-sensitive cells and was sequenced using the GPL21047 platform. The GSE199689 dataset included three samples from TMZ-resistant cells and three samples from their parental cell lines and was based on the GPL16956 sequencing platform. The raw data were normalized using the R package "preprocessCore" (version R4.1.3). Transcriptional and clinical information for glioblastoma patients was downloaded from the CGGA (http://www.cgga.org.cn/). Eight clinical characteristics were retained for subsequent analysis: recurrence status, grade, age, OS time, survival status, IDH wild-type status, 1p/19q codeletion status, and MGMT promoter methylation status.

3.2. Identification of DEGs

After normalization and comparison of gene expression levels between the TMZ-sensitive and TMZ-resistant groups, DEGs were screened using the R package "limma" with the cutoff criteria |log2-fold change| >1 and adjusted p value < 0.05. The R packages "ggplot2" and "pheatmap" were subsequently used to present DEG expression levels in volcano plots and heatmaps.

3.3. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses

GO and KEGG (https://www.genome.jp/kegg)

analyses of intersecting genes from the GSE113510 and GSE199689 datasets were performed with the R package "clusterProfiler" (11), and visualization was subsequently performed using the R package "ggplot2". Significantly enriched GO terms and KEGG signaling pathways were screened for analysis (12-14).

3.4. Construction of the WGCNA

Based on the R package "WGCNA", the WGCNA network was constructed to discern gene modules related to TMZ resistance in GBM (8). First, all the samples were clustered, and the clinical features were demonstrated. Second, the soft-thresholding power (β) was calculated by using the "pickSoftThreshold" function to better detect strong correlations between gene modules and convert the correlation matrix into a weighted adjacency matrix. Using a soft threshold of 7 for a correlation coefficient of 0.9, the DEGs were incorporated into a connectivity matrix. Third, the converted topological overlap matrix (TOM) was used to identify coexpressed gene modules (minimum number =50, cutoff height threshold = 0.3). Fourth, key modules were selected based on their correlations with malignant features in GBM, and genes within those modules were considered key genes. Subsequently, the genes with the highest node degree values were subjected to GO and KEGG analyses, as described above.

3.5. Machine Learning-Based Screening of Key Genes Related to TMZ Resistance

The study employed the LASSO logistic regression, SVM, and RF algorithms to ascertain the pivotal genes associated with TMZ resistance and their impact on survival status. The R packages "glmnet", "e1071" and "randomForest" were utilized to identify significant genes using the above three machine learning algorithms. As a dimensionality reduction method, LASSO regression analysis has advantages over traditional regression analysis for evaluating high-dimensional data. For the selection of relevant features and the removal of redundant features, SVM-RFE outperforms linear discriminant analysis and the mean square error. Drug resistance-associated genes that are affected by survival status were ranked using the RF algorithm. A tenfold cross-validation method was used to estimate the prediction performance, identifying twenty genes of relative importance as feature genes. Ultimately, genes that overlapped among the three machine learning algorithms were considered potential markers that can impact survival. Key genes obtained by machine learning were compared between GBM patients with different clinical characteristics. Receiver operating characteristic (ROC) curves were generated with the R package "pROC". The area under the curve (AUC) values were subsequently calculated to estimate the predictive utility of the biomarkers.

3.6. Evaluation of Immune Cell Infiltration

Single-sample gene set enrichment analysis (ssGSEA) was subsequently performed to assess immune cell infiltration. The immune cell enrichment scores and immune function information of the GBM patients were obtained using the R packages "GSVA", "limma", and "GSEABase". To analyze differences in immune function between patients according to survival status, we utilized the R packages "limma", "reshape2", and "ggpubr". Finally, visualization was implemented via the R "ggplot2" package.

3.7. Correlation and Enrichment Analyses

Pearson correlation analysis of the top 50 genes most positively and negatively associated with key genes was performed through the R package "pheatmap". Using the gseGO, gseKEGG, and gsePathway functions of the R package "clusterProfiler", GSEA was conducted with nPerm = 1,000, minGSSize = 10, maxGSSize = 1,000, and *p* value cutoff = 0.05.

3.8. TF-miRNA Coregulatory Network Construction and Drug Sensitivity Analysis

The RegNetwork repository (http://www.regnetworkweb.org/) was used to predict combinatorial regulatory interrelations between microRNAs (miRNAs), transcription factors (TFs), and TMZ resistance-specific target genes. The TF–miRNA–mRNA regulatory network was visualized with Cytoscape. The Gene Set Cancer Analysis (GSCA) database (http://bioinfo.life.hust.edu.cn/GSCA/) was used to analyze the correlation between drug sensitivity and the expression of key genes (15).

3.9. Construction of TMZ-Resistant Cell Lines

Temozolomide was purchased from Sigma–Aldrich (T2577, Sigma–Aldrich, St. Louis, USA). U343 and U251 cells were purchased from Xiamen Immocell

Biotechnology Co., Ltd. (Xiamen, Fujian, China). Based on our previously established methods, we constructed stable TMZ-resistant cells and named them U343-R and U251-R (16). The cells were cultured in a humidified incubator at 37 °C and 5% CO_2 in high-glucose Dulbecco's minimal essential medium (DMEM; Gibco, USA) supplemented with 10% fetal bovine serum (FBS; Meiluncell, Australia) and 1% penicillin–streptomycin (Gibco, USA).

3.10. Cell Counting Kit-8 (CCK-8) Assay

Cell viability was determined after treatment with different concentrations of TMZ via CCK-8 assays (Solarbio, CA1210, China). Briefly, glioma cells were incubated with various concentrations of temozolomide in a 96-well plate for 48 hours. Then, 10 μL of CCK-8 solution was added, and the optical density (OD) at 450 nm was measured. Cell viability (%) was calculated by using the following formula: [OD (TMZ) − OD (blank)]/ [OD (control) − OD (blank)] × 100%. The half-maximal inhibitory concentration (IC50) was calculated by using GraphPad Prism.

3.11. Real-Time Quantitative Polymerase Chain Reaction (RT–qPCR)

Total RNA was extracted with TRIzol reagent (Thermo Scientific, USA). Subsequently, its concentration and purity were assessed using a NanoDrop 2000 (Thermo Scientific, USA). cDNA synthesis was performed using SuperScript III Reverse Transcriptase (Invitrogen, USA), followed by PCR amplification to quantify the expression levels of the hub genes. The relative expression levels of target genes were determined using the $2^{-\Delta\Delta ct}$ method (17, 18). β-actin used as the internal control (or reference gene) (19-21). The primers used are listed in **Table S1**.

3.12. Western Blotting (WB)

The protein expression levels of hub genes were assessed using Western blotting (WB). Cells were lysed in RIPA buffer supplemented with protease and phosphatase inhibitors to extract total proteins. Protein concentrations were quantified using a BCA protein assay kit. Equal amounts of protein samples were resolved by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and subsequently transferred onto polyvinylidene difluoride (PVDF) membranes. The membranes were blocked with 5% non-fat dry milk in Tris-buffered saline with Tween-20 (TBST) for 1 hour at room temperature and then incubated with the following primary antibodies at 4°C overnight: anti-IGFBP2 (1:1000; Cat No. 11065-3-AP, Proteintech), anti-PITX1 (1:1000; Cat No. 10873-1-AP, Proteintech), and anti-TNFRSF11B (1:3000; Cat No. PS-18571, Ab-mart). After incubation with primary antibodies, membranes were washed with TBST and incubated with horseradish peroxidase-conjugated secondary antibodies for 1 hour at room temperature. Immunoreactive bands were visualized using chemiluminescence reagents (Sharebio, China) and detected with a ChemiScope Series imaging system (Clinx Science Instrument Co., China). Relative protein expression levels were quantified using ImageJ software, with β-actin (1:2000; Cat No. 20536-1-AP, Proteintech) serving as an internal loading control.

4. Results

4.1. Identification of TMZ Resistance-Related Degs in GBM

The expression profiles from the 2 GEO datasets (GSE113510 and GSE199689) were normalized using the R package "preprocessCore" (**Fig. S1**). A total of 7911 DEGs were obtained in GSE113510; these DEGs included 3946 upregulated genes and 3965 downregulated genes (**Fig. S2A, S2B**). Analysis of the GSE199689 dataset revealed 2610 upregulated genes and 1938 downregulated genes (**Fig. S2C, S2D**).

4.2. Functional Enrichment Analysis of The Degs

We analyzed the intersection of the 769 DEGs obtained from GSE113510 and GSE199689 (**Fig. 1A, 1B**). GO enrichment and KEGG pathway analyses were subsequently carried out for the up- and downregulated genes to assess their biological functions. The genes were mainly enriched in biological processes (BPs) including extracellular matrix organization, extracellular structure organization, and external encapsulating structure organization. The enriched cellular component (CC) terms included collagen-containing extracellular matrix, adherens junction, basement membrane, and related terms (**Fig. 1C**). KEGG enrichment analyses revealed that the genes were associated with cytokine–cytokine receptor interaction and the lipid and atherosclerosis pathway (**Fig. 1D**).

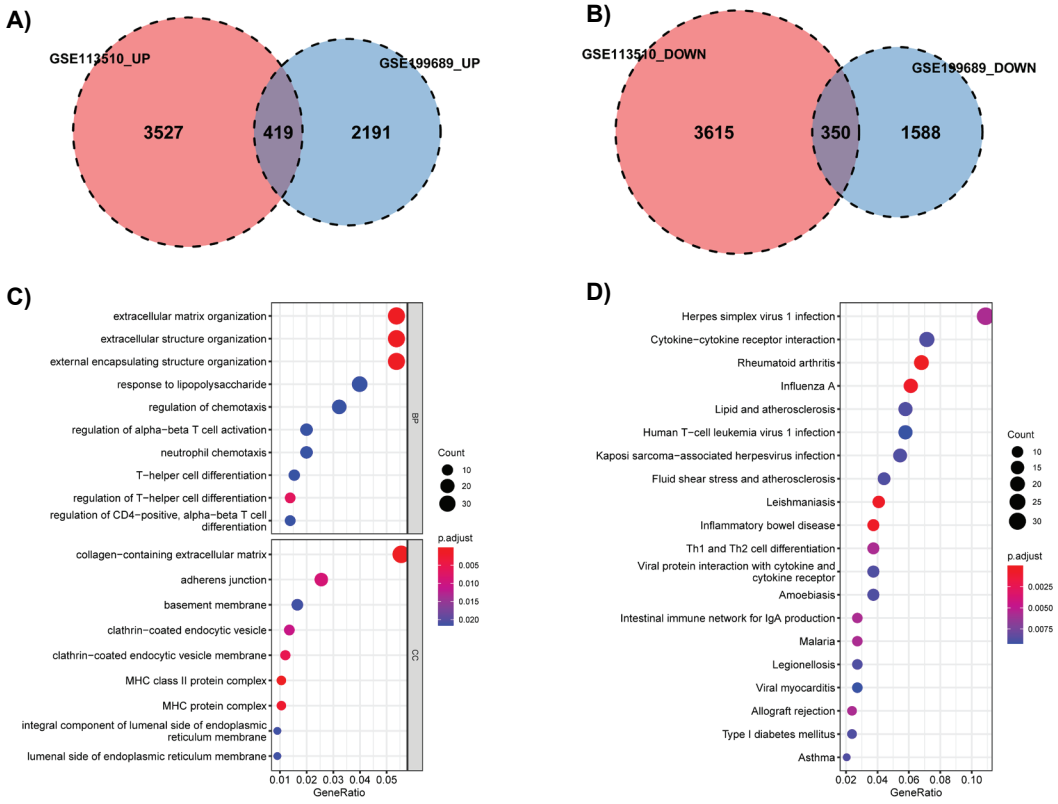

Figure 1. Intersection and functional enrichment analysis of the DEGs. A, B) intersection of the upregulated and downregulated DEGs. C) GO analysis showing enrichment of the DEGs in the BP, CC. **D)** KEGG analysis of the DEGs. BP, biological process; CC, cellular component; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

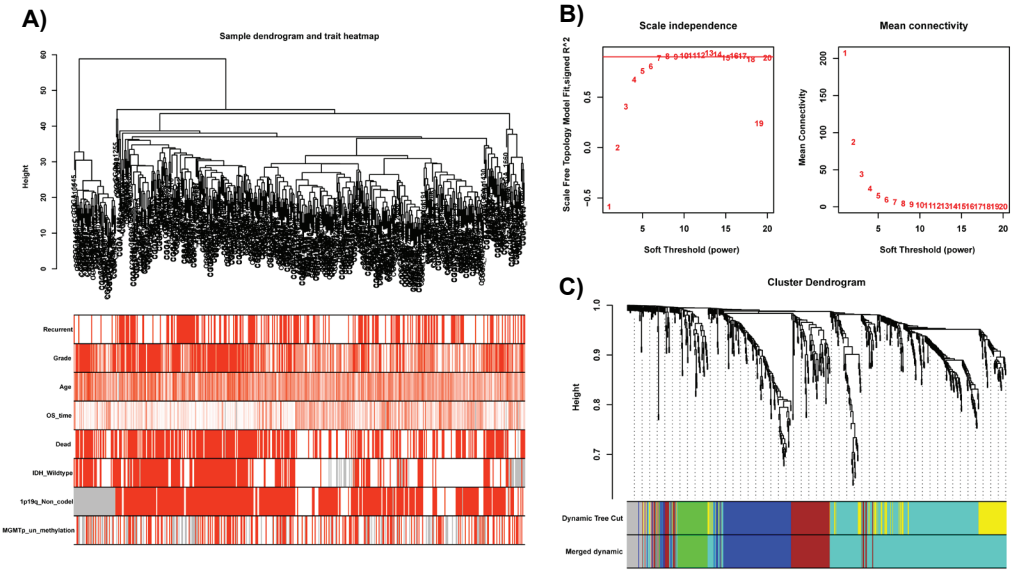

Figure 2. Construction of the weighted gene co-expression network. A). The dendrogram and trait heatmap are accompanied by red or white bar codes beneath the dendrogram, which indicate the presence or absence of a pair of mutually exclusive clinical traits. **B)** An analysis of the scale-free fit index for various soft threshold powers (beta). The merging threshold is shown in the red line. Next, the mean connectivity of various soft threshold powers was analyzed. **C)** Clustering dendrograms showed that genes that were closely related produced 5 gene coexpression modules.

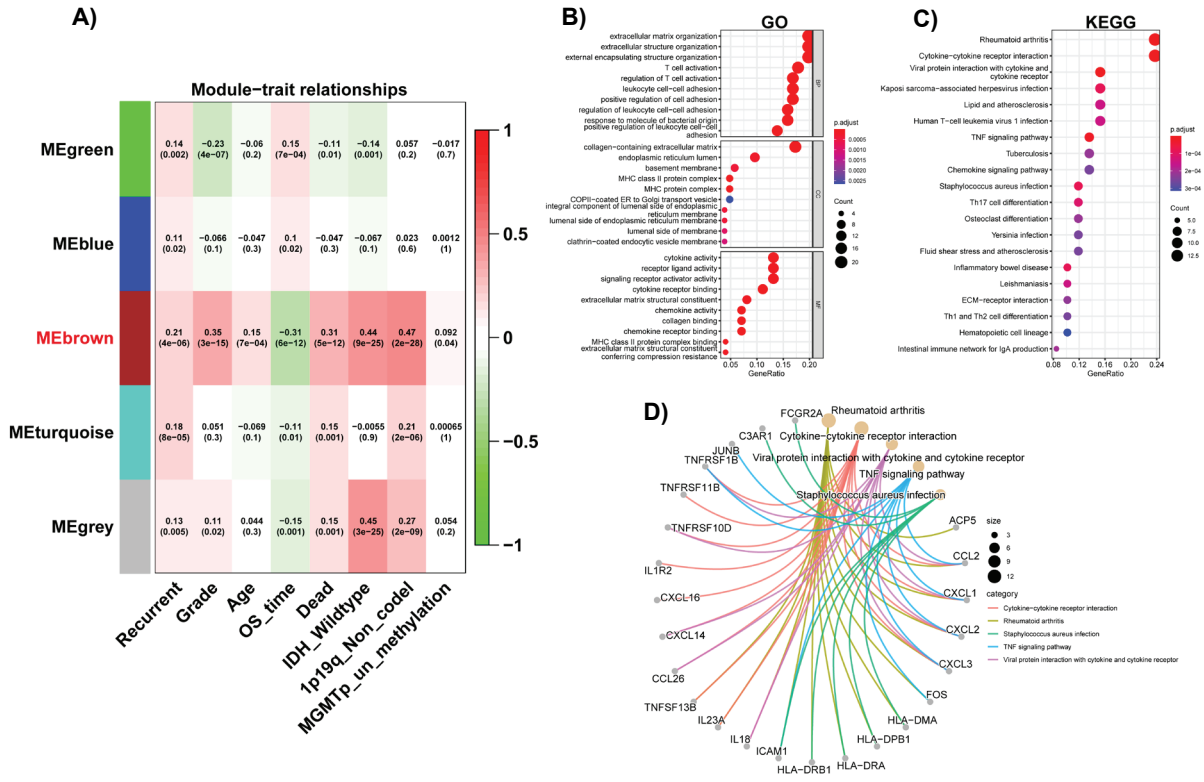

Figure 3. The identification of modules specific to glioblastoma and functional enrichment analyses of module eigengenes. A) Relationships of consensus modules with malignant features of glioblastoma are explored. Each module consists of a set of highly linked genes, with each specified color representing a specific gene module. **B)** GO analyses were conducted to predict the potential functions of genes in the MEbrown module, including CC, MF, and BP. **C)** KEGG potential pathways regarding genes in the MEbrown module were evaluated. **D)** The cord diagram showed the relationship between the top 5 enriched KEGG pathways and eigengenes.

4.3. WGCNA-Based Screening of Crucial Modules Associated with TMZ Resistance In GBM

Data on 23971 transcribed genes in 693 samples were obtained from the CGGA database. A total of 486 patients were enrolled after patients with incompleted clinical feature data were excluded. A total of 769 overlapping DEGs were obtained from the GEO data analysis, and 713 genes were retained after comparison with the DEGs in the CGGA dataset. The above data were subsequently used to screen for TMZ resistance-related hub modules. Both samples showed a better clustering trend (**Fig. 2A**). The gene modules were constructed based on a soft-threshold power of 7 (**Fig. 2B**). A coexpression matrix network was constructed, and five gene modules were visualized as heatmaps via dynamic hybrid shearing (**Fig. 2C**). A strong association was found between the MEbrown module and clinical prognostic signatures in GBM patients (**Fig. 3A**). Eigengenes from the MEbrown module were selected for subsequent functional enrichment analyses. As shown in **Figure 3B**, the functional enrichment of the genes was analyzed in three categories: the main enriched BP terms were extracellular matrix organization, extracellular structure organization, external encapsulating structure organization, and T-cell activation. The main enriched CC terms were collagen-containing extracellular matrix, endoplasmic reticulum lumen, and basement membrane. The main enriched MF terms were cytokine activity, receptor–ligand activity, and signaling receptor activator activity. As shown in **Figure 3C**, KEGG analyses revealed that the module genes were significantly enriched in pathways related to rheumatoid arthritis, cytokine–cytokine receptor interaction, and viral protein interaction with cytokines and cytokine receptors.

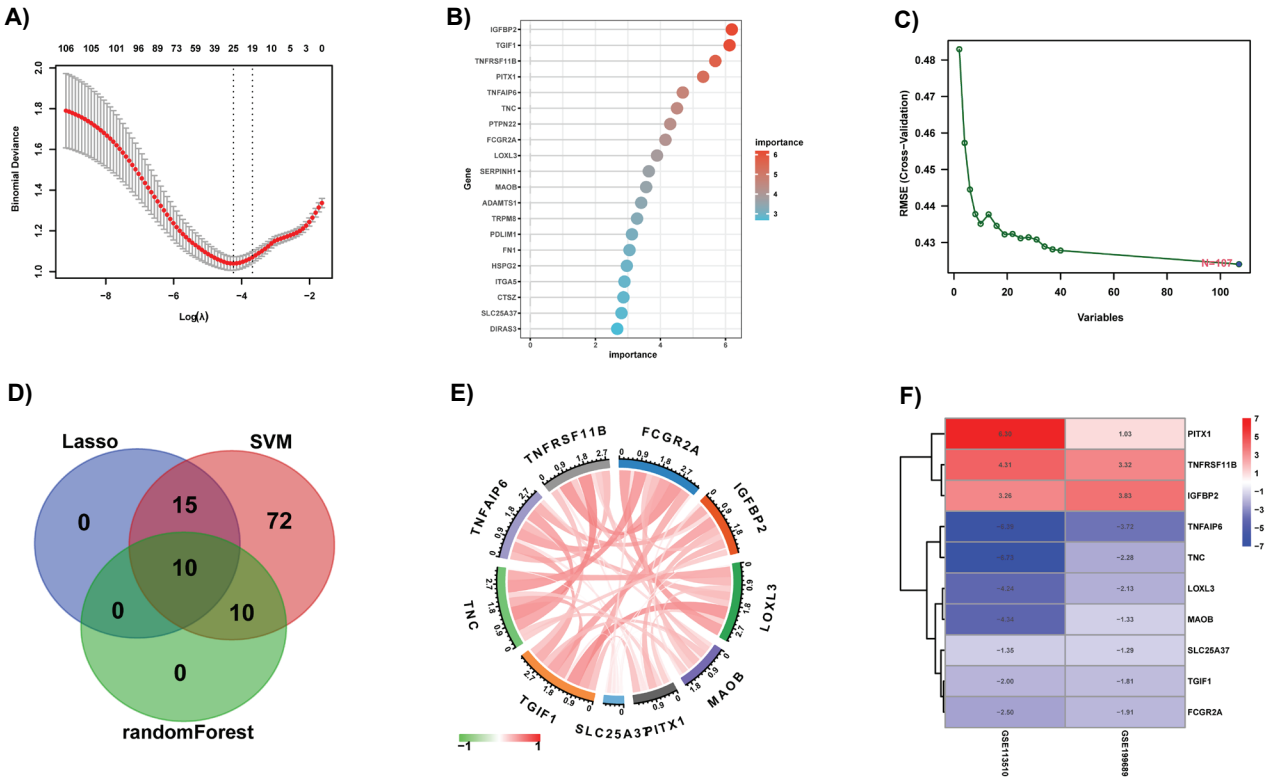

Figure 4. Screening TMZ-resistive related genes via comprehensive strategy. A) The least absolute shrinkage and selection operator (LASSO) logistic regression algorithm, **B)** random forest (RF) algorithm and **C)** support vector machine recursive feature elimination (SVM-RFE) algorithm was used to retain hub genes. **D)** the Venn diagram showed the intersection of genes obtained by about three algorithms. E) The circular barplot was profiled to analyze the correlation between 10 hub genes. **F)** heatmap indicated different expressions of 10 hub genes.

The interrelationships among the five major signaling pathways and the crucial genes are indicated in **Figure 3D**.

4.4. Screening And Confirmation of TMZ Resistance-Related Genes Via a Comprehensive Strategy

Among the DEGs common to both datasets, 25 genes were identified as potential TMZ resistance-related genes via the LASSO logistic regression algorithm (**Fig. 4A**). Twenty candidate resistance-related genes were identified via the RF algorithm (**Fig. 4B**). One hundred and seven candidate resistance-related genes were identified via the SVM-RFE algorithm (**Fig. 4C**). The overlapping DEGs among those identified via the three algorithms were regarded as potential TMZ resistance-related genes; these genes included TNFAIP6, TNFRSF11B, TGIF1, PITX1, FCGR2A, LOXL3, SLC25A37, TNC, IGFBP2, and MAOB (**Fig. 4D**). Further analysis revealed a general positive correlation among all ten genes (**Fig. 4E**). According to the heatmap, three highly expressed genes (PITX1, TNFRSF11B, and IGFBP2) exhibited significant differences in expression levels between GBM patients grouped based on clinical characteristics (**Fig. 5A, 5C, 5E**). The Kaplan–Meier analysis of survival curves revealed that recurrent GBM patients who had high expression of the three genes had shorter overall survival than those with low expression of these genes (**Fig. 5B, 5D, 5F**). The calculated AUCs for TNFRSF11B, PITX1, and IGFBP2 were all greater than 0.7, indicating that these three genes have moderate diagnostic value (**Fig. S3**).

4.5. Analysis of Immune Cell Infiltration

The majority of the correlations between the ratio of infiltrating immune cells and total macrophages were positive, as shown in **Figure S4A**. A violin plot was generated, and the results showed that the level of

most infiltrating immune cells was significantly greater in the decreased patient subgroup than in the live patient subgroup. However, there were no significant differences in the infiltration levels of activated B cells, CD56dim natural killer cells, eosinophils, monocytes, or neutrophils between the groups (**Fig. S4B**). The results of the correlation analysis indicated that there was a strong positive correlation between the expression of TMZ resistance-related genes and the infiltration of multiple immune cell types. Conversely, TNFRSF11B exhibited a negative correlation with the infiltration levels of CD56dim natural killer cells (**Fig. S4C**).

4.6. Correlation And Enrichment Analyses

Utilizing CGGA data, we performed a correlation analysis between these three TMZ resistance-related genes and other genes in GBM samples to predict their functions and related pathways (**Fig. S5A-S7A, Fig. S5B-S7B**). The top 50 genes with the strongest positive and negative associations with the three hub genes were subjected to enrichment analysis. GO analysis via the GSEA website revealed that IGFBP2, PITX1, and TNFRSF11B were primarily associated with neutrophil-mediated immunity, neutrophil activation involved in the immune response, G2/M transition of the mitotic cell cycle, response to hypoxia, regulation of cytokine-mediated signaling pathways, and cytokine production. Additionally, KEGG pathway analysis revealed significant enrichment in protein processing in the endoplasmic reticulum, viral carcinogenesis, the cell cycle, the P53 signaling pathway, cytokine–cytokine receptor interaction, and osteoclast differentiation.

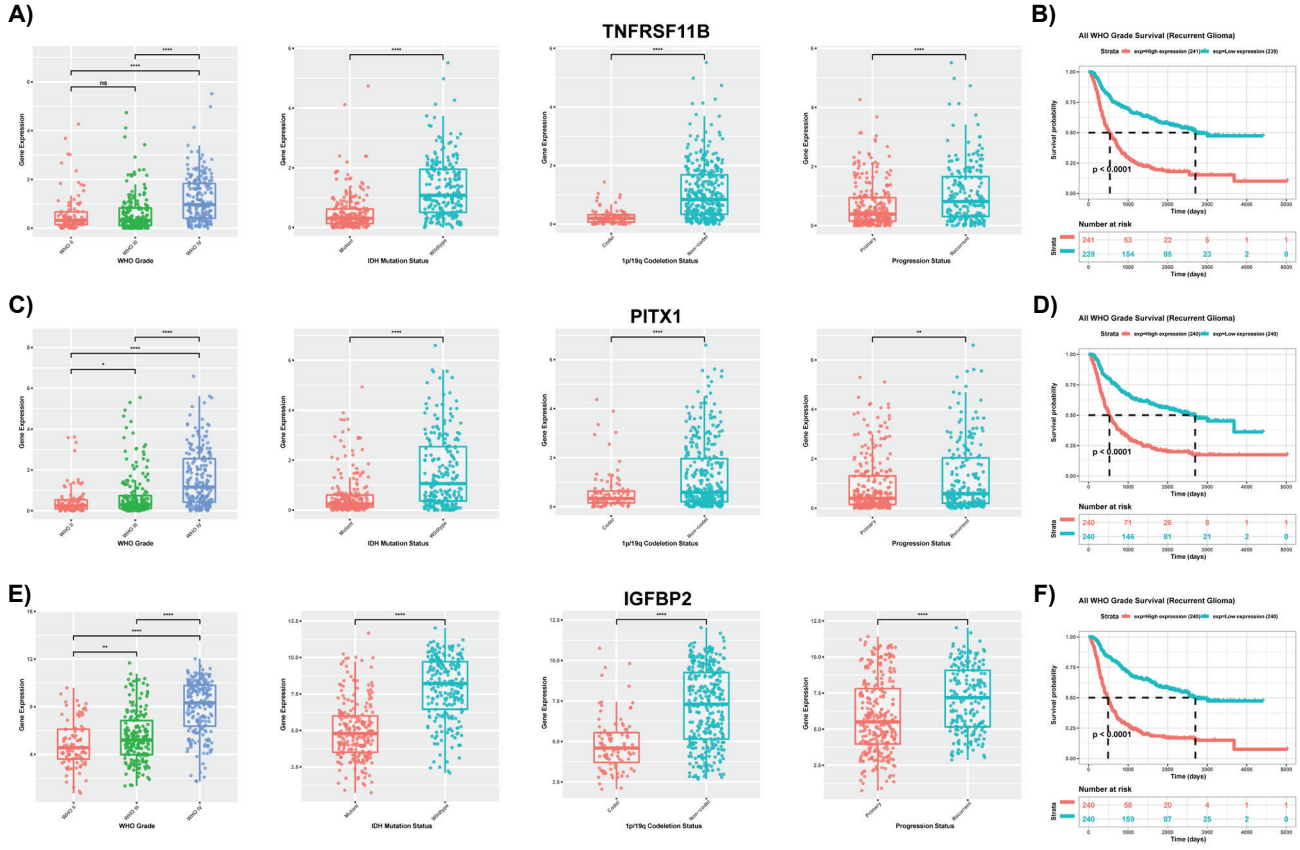

Figure 5. The expression of the three TMZ-resistance genes is related to clinical characteristic subgroups and overall survival for GBM patients. A) The correlation between TNFRSF11B expression levels and glioblastoma clinical characteristic subgroups and **B)** Kaplan–Meier overall survival curves for recurrent GBM patients. **C)** He correlation between PITX1 expression levels and glioblastoma clinical characteristic subgroups and **D)** Kaplan–Meier overall survival curves for recurrent glioblastoma. **E)** The correlation between IGFBP2 expression levels and glioblastoma clinical characteristic subgroups and **F)** Kaplan–Meier overall survival curves for recurrent GBM patients.

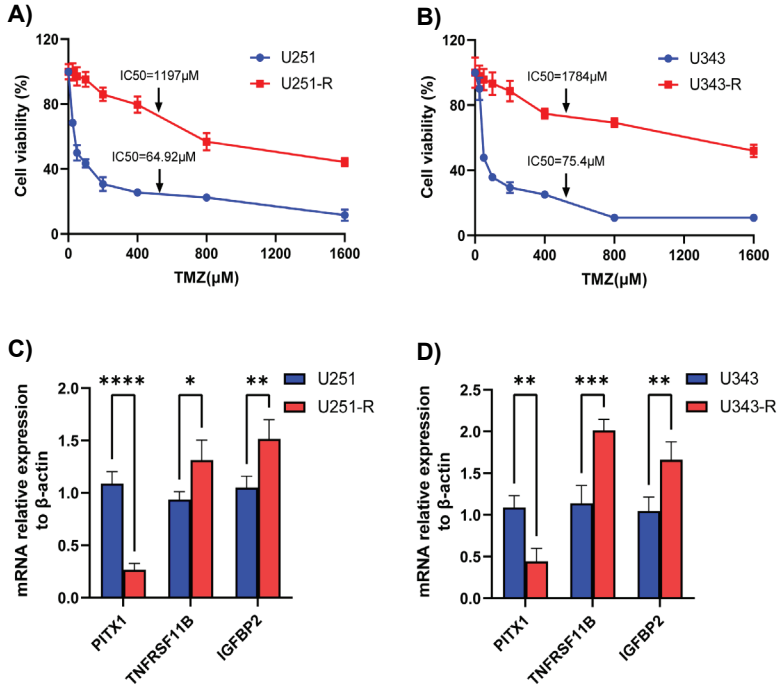

Figure 6. Validation of key genes in established TMZ-resistant cells. A, B) The viability and IC50 index of U343-R cells, U251-R cells, and their parental cells were treated with different concentrations of TMZ. **C, D)** Bar graphs show the expressional differences in PITX1, TNFRSF11B, and IGFBP2 between parental and TMZ-resistant cells, separately. Data indicate means ± SD of three biological replicates. Two-way ANOVA and Tukey's multiple comparison test; **P < 0.01, ***P < 0.001, ****P< 0.0001 (vs. parental cells). TMZ, temozolomide; TR, TMZ-resistant; IC50, 50% inhibiting concentration.

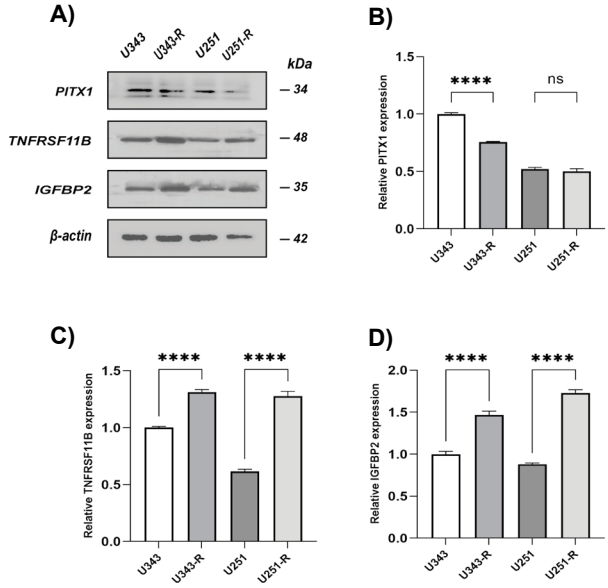

Figure 7. Visualization and statistical analysis of results from WB analysis results. A) Western blot analysis revealed differences in the protein expression levels of three genes, as indicated by varying band intensities, between TMZ-resistant cells and their parental counterparts. **B-D)** Statistical analysis of band intensities from western blot experiments demonstrates the relative protein expression levels of PITX1, TNFRSF11B, and IGFBP2, each normalized to β-actin, across different cell groups. ns indicates not significant. **P < 0.01, ***P < 0.001, ****P< 0.0001 (vs. parental cells). TMZ, temozolomide; TR, TMZ-resistant.

Moreover, the following pathways were strongly enriched according to Reactome pathway analysis: cell cycle mitotic, cell cycle, separation of sister chromatids, mitotic anaphase, neutrophil degranulation, and cytokine signaling in the immune system pathway (**Fig. S5C-S7C**). These results indicate that IGFBP2, PITX1, and TNFRSF11B play important roles in several malignancy-related pathways in GBM, especially those related to immunity, cellular signal transduction, and cell cycle regulation.

4.7. TF-Mirna Coregulatory Network and Analysis

MicroRNAs (miRNAs) and transcription factors (TFs) regulate target gene expression at multiple levels. We performed a comprehensive analysis to investigate the potential dual regulatory role of miRNAs and TFs in GBM. Initially, we screened three genes linked to TMZ resistance and identified their predicted and experimentally verified targets among TFs and miRNAs. The constructed TF-miRNA coregulatory network consisted of 74 nodes, including 21 TFs, 3 core genes, and 50 miRNAs (**Fig. S8**). Three miRNAs, namely, hsa-miR-145, hsa-miR-19a, and hsa-miR-220c, exhibited high connectivity with the core genes. Additionally, the TF gene MZF1 was identified as the most prominent target and exhibited connections with the core genes.

4.8. Drug Sensitivity Analysis

The drug sensitivity of IGFBP2, TNFRSF11B, and PITX1 was represented by bubble plots following GSCA analysis (**Fig. S9**). According to the Clinical Trials Reporting Program (CTRP), high expression of IGFBP2 might be an indicator of sensitivity to austocystin D. High expression of TNFRSF11B might indicate sensitivity to dasatinib. Moreover, the expression levels of IGFBP2, TNFRSF11B, and PITX1 exhibited significant correlations with the IC50 values of some drugs (**Fig. S9A**). Analysis of the GDSC database indicated that patients with high expression of IGFBP2 might exhibit resistance to AZD6482, BEZ235, and 17-AAG, and those with high expression of PITX1 and TNFRSF11B might exhibit resistance to BX-912, GSK1070916, and NPK76-II-72-1. Conversely, the results indicated that patients with high expression of IGFBP2 may exhibit sensitivity to specific drugs, such as SB2334 and navitoclax. Notably, high expression of PITX1 and TNFRSF11B may indicate sensitivity to 17-AAG (**Fig. S9B**).

4.9. Generation of Temozolomide-Resistant Cells and Validation of The Resistance-Related Key Genes

U251 and U343 cells were treated with increasing doses of TMZ to establish cell lines that were resistant TMZ. After a continuous culture period of 6 months, the IC50 values of both the parental and the resistant cells were assessed using the CCK-8 assay. A noticeable association was observed between the concentration of TMZ and the viability of both parental and TMZ-resistant cells, and the correlation became stronger as the dose increased. Notably, compared with that of U251-R cells, the viability rate of U251 cells was significantly lower under treatment with varying concentrations of TMZ, and a similar trend was observed for U343 and U343-R cells (**Fig. 6A, 6B**). Subsequently, a comparative analysis of mRNA expression levels in temozolomide (TMZ)-resistant cells revealed a significant upregulation of TNFRSF11B and IGFBP2 transcripts (all $P < 0.01$) (**Fig. 6C, 6D**). To corroborate these findings at the protein level, we performed western blot validation, which confirmed that the protein expression levels of TNFRSF11B and IGFBP2 were consistent with the observed mRNA expression trends in RT-qPCR ($P < 0.01$) (**Fig. 7**).

5. Discussion

TMZ is a chemotherapeutic agent routinely used to treat high-grade gliomas, especially GBM. However, biological heterogeneity within the GBM patient population significantly influences the treatment response, leading to considerable variations in outcomes and posing a substantial obstacle to the implementation of tailored therapies and the improvement of patient prognosis (22-24). Consequently, identifying TMZ-sensitive patients is of paramount importance in the context of personalized therapeutic interventions. Several studies have investigated the mechanism of chemoresistance in GBM by generating cells resistant to TMZ (25, 26). However, few studies have focused on RNA sequencing (RNA-seq) of TMZ-resistant cells and their parental counterparts to identify key genes that are differentially expressed between cells with and without TMZ resistance. A recent study developed a risk model to predict glioma's response to TMZ and identified DACH1 as a key gene involved

in TMZ resistance, which could serve as a potential therapeutic target for enhancing chemosensitivity (27). However, this single-gene research approach lacks a systematic and comprehensive nature. Therefore, this study systematically investigates gene expression and explores the biological processes involved in the resistance of glioma cells to TMZ chemotherapy.

We identified the DEGs between resistant and nonresistant cells and found that the expression of three key genes (PITX1, TNFRSF11B, and IGFBP2) was notably increased in TMZ-resistant cells and was positively related to adverse clinical features. Furthermore, the expression levels of key genes were validated in our constructed glioma TMZ-resistant cell lines by using RT–qPCR and WB. These findings elucidated the possibility that these three genes play essential roles in TMZ resistance.

The pituitary homeobox 1 gene, commonly referred to as PITX1, has garnered extensive research attention. Recent studies have revealed the dysregulation of PITX1 expression in various malignancies, underscoring its important role in tumor progression and resistance to chemotherapy. According to previous studies, the upregulation of PITX1 in neoplastic cells facilitates drug-induced apoptosis following exposure to mitomycin C and etoposide (28). Sorafenib and regorafenib have been demonstrated to increase the PITX1 protein expression level and promote apoptosis (29, 30). The findings of our study are in concordance with these findings. RT–qPCR showed that PITX1 expression is decreased in TMZ-resistant cell lines, and the inconsistency between these results and that of the bioinformatic analysis may be partly attributed to the use of different GBM cell lines. We speculate that TMZ treatment leads to a noteworthy increase in PITX1 expression in TMZ-sensitive cells, which markedly promotes apoptosis.

Tumor necrosis factor receptor superfamily member 11B (TNFRSF11B), always referred to as osteoprotegerin (OPG), is a constituent of the tumor necrosis factor (TNF) receptor superfamily (31, 32). Patients who exhibit high expression of TNFRSF11B in prostate cancer have a significantly poorer prognosis than those with low expression, and TNFRSF11B serves as a key determinant in cell survival in hormone-resistant prostate cancer cells (33, 34). Moreover, elevated levels of the secreted protein TNFRSF11B are associated with a dismal postoperative prognosis in hepatocellular carcinoma patients (35). Specific to the field of glioma research, A previous study indicates that the methylation-regulated gene TNFRSF11B plays a significant role in GBM progression, with potential as a therapeutic target for drugs like zoledronic acid, which may enhance treatment efficacy (36). Insulin-like growth factor binding protein 2 (IGFBP2) is a prominent constituent of the IGFBP family. Numerous studies have highlighted the pivotal role of IGFBP2, whether by binding to IGFs independently, in influencing tumorigenesis by modulating various cancer-related characteristics (37). In addition, the upregulation of IGFBP2 has been consistently related to enhanced tumor cell proliferation, invasion, migration, and chemotherapy resistance. Multiple investigations have established a positive correlation between heightened IGFBP2 expression and unfavorable outcomes in patients diagnosed with gliomas (38-40). Another study reported that IGFBP2 can facilitate chemotherapy resistance, cell proliferation, and invasion in glioma patients through the integrin β1/ERK signaling pathway (41). Specifically, IGFBP2 was found to hinder the binding of IGF-1 to its receptor. A previous study revealed that IGFBP2 contributes to erlotinib resistance by modulating adipokine leptin-activated IGF-1R signaling (42). Furthermore, Masuo and coworkers demonstrated that the overexpression of the zinc finger protein SNAIL2 in pancreatic cancer promotes gemcitabine resistance through the regulation of transcription and expression of IGFBP2 (43).

Our study revealed that high TNFRSF11B and IGFBP2 expression is significantly correlated with malignant clinicopathological characteristics and a poor prognosis in glioma patients. RT–qPCR and WB analysis further confirmed that TNFRSF11B and IGFBP2 were more highly expressed in TMZ-resistant cell lines than in TMZ-sensitive cell lines. This observed trend is consistent with the findings of previous studies on chemoresistance in glioma and other tumor types. Therefore, we speculated that the TNFRSF11B and IGFBP2 genes possibly participate in similar mechanisms in GBM chemoresistance and proposed a potential therapeutic target for improving the response to GBM TMZ chemotherapy. In addition, our findings align with the theory that aberrant crosstalk among cytokines is linked to chemotherapy resistance and relapse risk. This is supported by the results of the current study, which demonstrated that genes correlated

with TMZ resistance were enriched in the cytokine–cytokine receptor interaction pathway (44, 45).

Compared with previous studies, we utilized more accurate and comprehensive bioinformatics analysis and machine learning methods to screen key TMZ resistance-related genes. In addition to using GEO datasets, we used the CGGA database as validation cohort to increase the sample size of the study and the reliability of the results. However, our study has several limitations. First, it is imperative to acknowledge that the sample size was too small. Studies with larger sample sizes and more comprehensive methods could yield more accurate conclusions regarding the potential value of the identified genes as predictive markers of drug resistance and/or therapeutic targets for glioma. Second, while three pivotal genes exhibiting differential expression between groups were identified, the mechanisms by which these genes are involved in glioma TMZ resistance have not been determined. Consequently, numerous subsequent functional experiments are warranted to determine the biological functions and pathogenesis mechanisms of these key genes.

6. Conclusion

In conclusion, our study identified and validated three potential key genes (PITX1, TNFRSF11B, and IGFBP2) that are specifically associated with resistance to TMZ in GBM patients. Furthermore, we predicted the pathways and biological functions related to these genes. These preliminary findings offer a glimpse into the underlying mechanisms of TMZ resistance and promising therapeutic targets for overcoming chemoresistance in GBM. Ultimately, the application of these results has the potential to improve therapeutic outcomes for patients diagnosed with GBM.

Acknowledgments

The provision of platforms and contributors via the GEO database for uploading meaningful datasets is greatly appreciated.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Author Contribution

JH: Conceptualization and writing of the original manuscript. JYY and NH: Data curation and formal analysis. ZTS and BHM: Methodology, Validation, Visualization. TMH: Funding acquisition. HW and WHC: Revision of the manuscript and supervision the study. All the authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Science and Technology Program of Chengde (No.202006A086) and the Government-funded Clinical Medicine Talent Training Project of Hebei (No.ZF2024235).

Data and Materials Availability

The study's datasets are accessible online via the following GEO accession numbers: GSE113510 and GSE199689, available at NCBI's Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo).

Declarations

Ethics approval and consent to participate

The GEO and CGGA databases are publicly accessible and contain anonymized data; thus, obtaining ethics approval was not necessary.

Consent for publication

This study has not been published before, and this publication has been approved by all authors.

References

1. McKinnon C, Nandhabalan M, Murray SA, Plaha P. Glioblastoma: clinical presentation, diagnosis, and management. *BMJ*. 2021;**374**:n1560. doi: 10.1136/bmj.n1560.
2. Wen PY, Weller M, Lee EQ, Alexander BM, Barnholtz-Sloan JS, Barthel FP, *et al*. Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro Oncol*. 2020;**22**(8):1073-1113. dio: 10.1093/neuonc/noaa287.
3. Bjorland LS, Fluge O, Gilje B, Mahesparan R, Farbu E. Treatment approach and survival from glioblastoma: results from a population-based retrospective cohort study from Western Norway. *BMJ Open*. 2021;**11**(3):e043208. doi: 10.1136/bmjopen-2020-043208.
4. Tan AC, Ashley DM, Lopez GY, Malinzak M, Friedman HS, Khasraw M. Management of glioblastoma: State of the art and future directions. *CA Cancer J Clin*. 2020;**70**(4):299-312. doi: 10.3322/caac.21613.
5. Omuro A, Brandes AA, Carpentier AF, Idbaih A, Reardon DA, Cloughesy T, *et al*. Radiotherapy combined with nivolumab or temozolomide for newly diagnosed glioblastoma with unmethylated MGMT promoter: An international randomized phase III trial. *Neuro Oncol*. 2023;**25**(1):123-134. doi: 10.1093/

neuonc/noac099.
6. van den Bent MJ, Tesileanu CMS, Wick W, Sanson M, Brandes AA, Clement PM, *et al*. Adjuvant and concurrent temozolomide for 1p/19q non-co-deleted anaplastic glioma (CATNON; EORTC study 26053-22054): second interim analysis of a randomised, open-label, phase 3 study. *Lancet Oncol*. 2021;**22**(6):813-823. doi: 10.1016/s1470-2045(21)00090-5.
7. Singh N, Miner A, Hennis L, Mittal S. Mechanisms of temozolomide resistance in glioblastoma - a comprehensive review. *Cancer Drug Resist.* 2021;**4**(1):17-43. doi: 10.20517/cdr.2020.79.
8. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;**9**:559. doi: 10.1186/1471-2105-9-559.
9. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*. 2023;**186**(8):1772-1791. doi: 10.1016/j.cell.2023.01.035.
10. Issa NT, Stathias V, Schurer S, Dakshanamurthy S. Machine and deep learning approaches for cancer drug repurposing. *Semin Cancer Biol*. 2021;**68**:132-142. doi: 10.1016/j.semcancer.2019.12.011.
11. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;**16**(5):284-287. doi: 10.1089/omi.2011.0118.
12. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;**28**(1):27-30. doi: 10.1093/nar/28.1.27.
13. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci*. 2019;**28**(11):1947-1951.
14. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 2023;**51**(D1):D587-D592. doi: 10.1093/nar/gkac963.
15. Liu CJ, Hu FF, Xia MX, Han L, Zhang Q, Guo AY. GSCALite: a web server for gene set cancer analysis. *Bioinformatics.* 2018;**34**(21):3771-3772. doi: 10.1093/bioinformatics/bty411.
16. Geng X, Zhang Y, Lin X, Zeng Z, Hu J, Hao L, *et al*. Exosomal circWDR62 promotes temozolomide resistance and malignant progression through regulation of the miR-370-3p/MGMT axis in glioma. *Cell Death Dis*. 2022;**13**(7):596. doi: 10.1038/s41419-022-05056-5.
17. Song J, Cho J, Park J, Hwang JH. Identification and validation of stable reference genes for quantitative real time PCR in different minipig tissues at developmental stages. *BMC Genomics.* 2022;**23**(1):585. doi: 10.1186/s12864-022-08830-z.
18. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods.* 2001;**25**(4):402-408. doi: 10.1006/meth.2001.1262.
19. Valente V, Teixeira SA, Neder L, Okamoto OK, Oba-Shinjo SM, Marie SK, *et al*. Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. *BMC Mol Biol*. 2009;**10**:1-11. doi: 10.1186/1471-2199-10-17.
20. Kreth S, Heyn J, Grau S, Kretzschmar HA, Egensperger R, Kreth FW. Identification of valid endogenous control genes for determining gene expression in human glioma. *Neuro Oncol*. 2010;**12**(6):570-579. doi: 10.1093/neuonc/nop072.
21. Röhn G, Koch A, Krischek B, Stavrinou P, Goldbrunner R, Timmer M. ACTB and SDHA are suitable endogenous reference genes for gene expression studies in human astrocytomas using quantitative RT-PCR. *Technol Cancer Res Treat.* 2018;**17**:1533033818802318. doi: 10.1177/1533033818802318.
22. Nam Y, Koo H, Yang Y, Shin S, Zhu Z, Kim D, *et al*. Pharmacogenomic profiling reveals molecular features of chemotherapy resistance in IDH wild-type primary glioblastoma. *Genome Med*. 2023;**15**(1):1-15. doi: 10.1186/s13073-023-01165-8.
23. Akgül S, Patch A-M, D'souza RC, Mukhopadhyay P, Nones K, Kempe S, *et al*. Intratumoural heterogeneity underlies distinct therapy responses and treatment resistance in glioblastoma. *Cancers*. 2019;**11**(2):190. doi: 10.3390/cancers11020190.
24. Qazi M, Vora P, Venugopal C, Sidhu S, Moffat J, Swanton C, *et al*. Intratumoral heterogeneity: pathways to treatment resistance and relapse in human glioblastoma. *Ann Oncol*. 2017;**28**(7):1448-1456. doi: 10.1093/annonc/mdx169.
25. Yi GZ, Xiang W, Feng WY, Chen ZY, Li YM, Deng SZ, *et al*. Identification of key candidate proteins and pathways associated with temozolomide resistance in glioblastoma based on subcellular proteomics and bioinformatical analysis. *Biomed Res Int*. 2018;**2018**. doi: 10.1155/2018/5238760.
26. Zeng H, Xu N, Liu Y, Liu B, Yang Z, Fu Z, *et al*. Genomic profiling of long non-coding RNA and mRNA expression associated with acquired temozolomide resistance in glioblastoma cells. *Int J Oncol*. 2017;**51**(2):445-455. doi: 10.3892/ijo.2017.4033.
27. Gu Q, Li L, Yao J, Dong FY, Gan Y, Zhou S, *et al*. Identification and verification of the temozolomide resistance feature gene DACH1 in gliomas. *Front Oncol.* 2023;**13**:1120103. doi: 10.3389/fonc.2023.1120103.
28. Yamaguchi T, Miki Y, Yoshida K. The c-Abl tyrosine kinase stabilizes Pitx1 in the apoptotic response to DNA damage. *Apoptosis.* 2010;**15**:927-935. doi: 10.1007/s10495-010-0488-6.
29. Teng HW, Hung MH, Chen LJ, Chang MJ, Hsieh FS, Tsai MH, *et al*. Protein tyrosine phosphatase 1B targets PITX1/p120RasGAP thus showing therapeutic potential in colorectal carcinoma. *Sci Rep.* 2016;**6**(1):35308. doi: 10.1038/srep35308.
30. Tai WT, Chen YL, Chu PY, Chen LJ, Hung MH, Shiau CW, *et al*. Protein tyrosine phosphatase 1B dephosphorylates PITX1 and regulates p120RasGAP in hepatocellular carcinoma. *Hepatology*. 2016;**63**(5):1528-1543. doi: 10.1002/hep.28478.
31. Lane D, Matte I, Laplante C, Garde-Granger P, Rancourt C, Piché A. Osteoprotegerin (OPG) activates integrin, focal adhesion kinase (FAK), and Akt signaling in ovarian cancer cells to attenuate TRAIL-induced apoptosis. *J Ovarian Res*. 2013;**6**:1-9. doi: 10.1186/1757-2215-6-82.
32. De Toni EN, Thieme SE, Herbst A, Behrens A, Stieber P, Jung A, *et al*. OPG is regulated by β-catenin and mediates resistance to TRAIL-induced apoptosis in colon cancer. *Clin Cancer Res.* 2008;**14**(15):4713-4718. doi: 10.1158/1078-0432.ccr-07-5019.
33. Velletri T, Huang Y, Wang Y, Li Q, Hu M, Xie N, *et al*. Loss of p53 in mesenchymal stem cells promotes alteration of bone remodeling through negative regulation of osteoprotegerin. *Cell Death Differ*. 2021;**28**(1):156-169. doi: 10.1038/s41418-020-0590-4.
34. Holen I, Croucher PI, Hamdy FC, Eaton CL. Osteoprotegerin (OPG) is a survival factor for human prostate cancer cells. *Cancer Res*. 2002;**62**(6):1619-1623.
35. Zhang C, Lin J, Ni X, Li H, Zheng L, Zhao Z, *et al*. Prognostic

value of serum osteoprotegerin level in patients with hepatocellular carcinoma following surgical resection. *Front Oncol*. 2021;**11**:731989. doi: 10.3389/fonc.2021.731989.

36. Xie H, Yuan C, Li JJ, Li ZY, Lu WC. Potential Molecular Mechanism of TNF Superfamily-Related Genes in Glioblastoma Multiforme Based on Transcriptome and Epigenome. *Front Neurol*. 2021;**12**:576382. doi: 10.3389/fneur.2021.576382.
37. Russmueller G, Moser D, Würger T, Wrba F, Christopoulos P, Kostakis G, *et al*. Upregulation of osteoprotegerin expression correlates with bone invasion and predicts poor clinical outcome in oral cancer. *Oral Oncol*. 2015;**51**(3):247-253. doi: 10.1016/j.oraloncology.2014.11.010.
38. Zhang B, Hong CQ, Luo YH, Wei LF, Luo Y, Peng YH, *et al*. Prognostic value of IGFBP2 in various cancers: a systematic review and meta-analysis. *Cancer Med.* 2022;**11**(16):3035-3047. doi: 10.1002/cam4.4680.
39. Yuan Q, Cai HQ, Zhong Y, Zhang MJ, Cheng ZJ, Hao JJ, *et al*. Overexpression of IGFBP2 mRNA predicts poor survival in patients with glioblastoma. *Biosci Rep.* 2019;**39**(6). doi: 10.1042/bsr20190045.
40. McDonald KL, O'Sullivan MG, Parkinson JF, Shaw JM, Payne CA, Brewer JM, *et a*l. IQGAP1 and IGFBP2: valuable biomarkers for determining prognosis in glioma patients. *J Neuropathol Exp Neurol.* 2007;**66**(5):405-417. doi: 10.1097/nen.0b013e31804567d7.
41. Han S, Li Z, Master LM, Master ZW, Wu A. Exogenous IGFBP-2 promotes proliferation, invasion, and chemoresistance to temozolomide in glioma cells via the integrin beta1-ERK pathway. *Br J Cancer.* 2014;**111**(7):1400-1409. doi: 10.1038/bjc.2014.435.
42. Wang F, Zhang L, Sai B, Wang L, Zhang X, Zheng L, *et al*. BMSC-derived leptin and IGFBP2 promote erlotinib resistance in lung adenocarcinoma cells through IGF-1R activation in hypoxic environment. *Cancer Biol Ther.* 2020;**21**(1):61-71. doi: 10.1080/15384047.2019.1665952.
43. Masuo K, Chen R, Yogo A, Sugiyama A, Fukuda A, Masui T, *et al*. SNAIL2 contributes to tumorigenicity and chemotherapy resistance in pancreatic cancer by regulating IGFBP2. *Cancer Sci.* 2021;**112**(12):4987-4999. doi: 10.1111/cas.15162.
44. Liu F, Zhou Q, Jiang HF, Zhang TT, Miao C, Xu XH, *et al*. Piperlongumine conquers temozolomide chemoradiotherapy resistance to achieve immune cure in refractory glioblastoma via boosting oxidative stress-inflamation-$CD8^+$-T cell immunity. *J Exp Clin Cancer Res*. 2023;**42**(1):1-19. doi: 10.1186/s13046-023-02686-1.
45. Hong X, Zhang J, Zou J, Ouyang J, Xiao B, Wang P, *et al*. Role of COL6A2 in malignant progression and temozolomide resistance of glioma. *Cell Signal.* 2023;**102**:110560. doi: 10.1016/j.cellsig.2022.110560.