# Screening of Potential Hub Genes in Glioma Progression Based on Bioinformatics Analysis

**Chengzhi Cui[1†], Shi Yin[2†], Lei Meng[3*], Jiaming Huang[1*], Dong Chen[1*]**

**1. Department of Neurosurgery, Dalian Municipal Central Hospital, Dalian 116083, China.**

**2. The Central Hospital of Jimusi City, Ordos 017010, China.**

**3. The First Affiliated Hospital of University of South China, Dali 671013, China.**

***Abstract:*** Objectives: Glioma is the most common primary tumor of the central nervous system, and its therapeutic effect is not optimistic. In recent years, related therapeutic technologies have developed rapidly, but unfortunately, the improvement of clinical therapeutic effect is not satisfactory. In addition to conventional therapies, there are some attractive therapies, such as biological therapy (immunotherapy), gene therapy, etc[1]. Therefore, searching for potential target genes of glioma is of great significance for developing new therapeutic directions and designing new biomarkers[2]. Methods: Download gene expression data set, GSE137902 gelatin and GSE13790 matrix through NCBI-G to screen overlapping differential expression genes (DEGs). In order to identify central genes from these genes, we conducted protein protein interaction (PPI) network. To further explore the potential mechanism of central genes in glioma, we performed gene ontology (GO) and Kyoto Gene and Genome Encyclopedia (KEGG) analysis. Then get the intersection of key genes according to five algorithms of Closeness Degree EPC MCC Stress. The intersection is obtained through GSE117423, GSE188256 and GSE90598 in geo database, and finally verified through Receiver Operating Characteristic (ROC) curve. Results: A total of 1274 differentially expressed genes are identified, and then 309 genes are obtained by intersection of the two. 16 Hub genes were obtained, and then the intersection of the two genes with GSE117423, GES188256 and GSE90598 genes was verified to obtain the key gene TIMP1 of glioma. Made the ROC curve of key gene.The intersection with hub gene was determined to identify TIMP1 as the key gene. Conclusion: The DEGs and Hub genes and signal pathways found in this study can confirm that the key gene TIMP1 is closely related to the occurrence and evolution of glioma, and provide candidate targets for the diagnosis and treatment of glioma.

***Keywords:*** Glioma; Degs; Hub Genes; Microarray; Enrichment Analysis

## Introduction

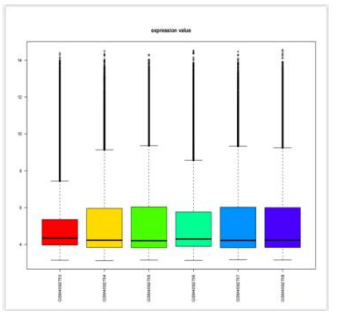
Glioma is the most common malignant brain tumor in central nervous system.Despite advances in the treatment of glioma such as surgery and chemoradiotherapy,most patients are easy to relapse,resulting in adverse clinical outcomes[3]. The current standard treatment methods, including surgical resection and radiotherapy and chemotherapy, do not bring satisfactory therapeutic effect, which is related to the aggressive growth of glioma in the brain, blood-brain barrier limitation and tumor drug resistance. It is one of the most common intracranial malignant tumors, with an annual incidence of about 3-6.4/100,000. The effect of traditional therapy on patients with glioma is poor. Therefore, it is necessary to further study the key genes and pathways related to glioma, and the development of bioinformatics and tumor genomics provides the possibility to discover new tumor biomarkers and therapeutic targets.

As an efficient and large-scale bioinformatics technology, gene array can detect and analyze differentially expressed genes in pathological tissues and normal tissues, so as to understand the changes of the whole genome in the process of tumorigenesis as a whole. By comparing the normal group and glioma patients, the selected differential genes were analyzed by bioinformatics to further understand the molecular mechanism of glioma. Therefore, we can use microarray technology and bioinformatics analysis to screen key genes of glioma

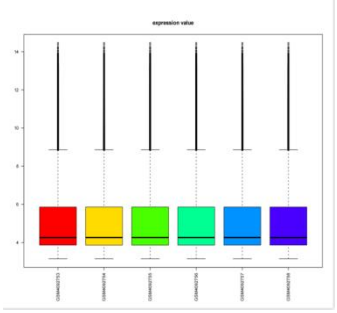
In order to identify the key genes related to glioma, we conducted a series of analysis based on the high-throughput sequencing data obtained from two data sets GSE137902 gelatin and GSE137902 matrix. We first determined the common DEG in the two databases, because the combination of multiple databases can provide more reliable results. Well, we used the online tools of Metascape website and DAVID website to analyze GO and KEGG terms, explore the main ways of DEG enrichment, and explore the research progress of glioma. The protein interaction network between DEGs was constructed by using online tools on STRING website, and was illustrated by using the software Cytoscape. We use the cytoHubba plug-in of Cytoscape to search for hub genes. Here, we use four different models, DEGREE, MCC, EPC and Stress, to screen the most important hub genes. Then, we used the online tool of human protein mapping to explore the genes in the central gene network related to glioma. Finally, we used GSE117423, GSE188256, GSE90598 in GEO database to intersect with hub gene, and concluded that TIMP1 may be the key gene of glioma. After further exploration of ROC curve, we were surprised to find that AUC>0.5 in ROC curve, which verified TIMP1 as a key gene of glioma. In conclusion, our study provides a new potential immunotherapeutic target for glioma biotherapy.

**Materials and methods Dataset selection**

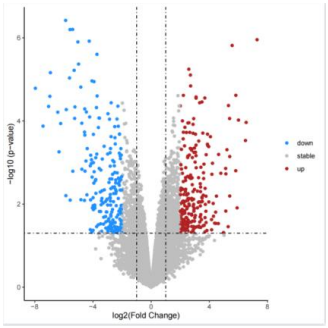
*Microarray Data. GEO.*Microarray Data. GEO (https://www.ncbi.nlm.nih.gov/ geo/) is a database containing high-throughput gene expression data, chips, and microarrays [4]. In this study, we downloaded an original microarray dataset GSE137902 from NCBI Gene Expression Synthesis Database (NCBI-GEO)（GPL13667 [HG-U219] Affymetrix Human Genome U219 Array）Endothelial cells isolated from glioblastoma and normal brain were cultured on gelatin and matrix gel as monolayer, including 3 normal brain endothelial cells as monolayer gelatin, 6 glioblastoma.



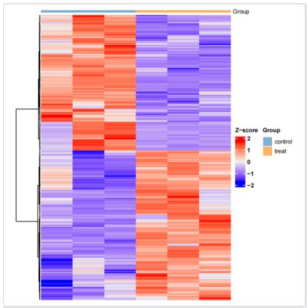
A



B



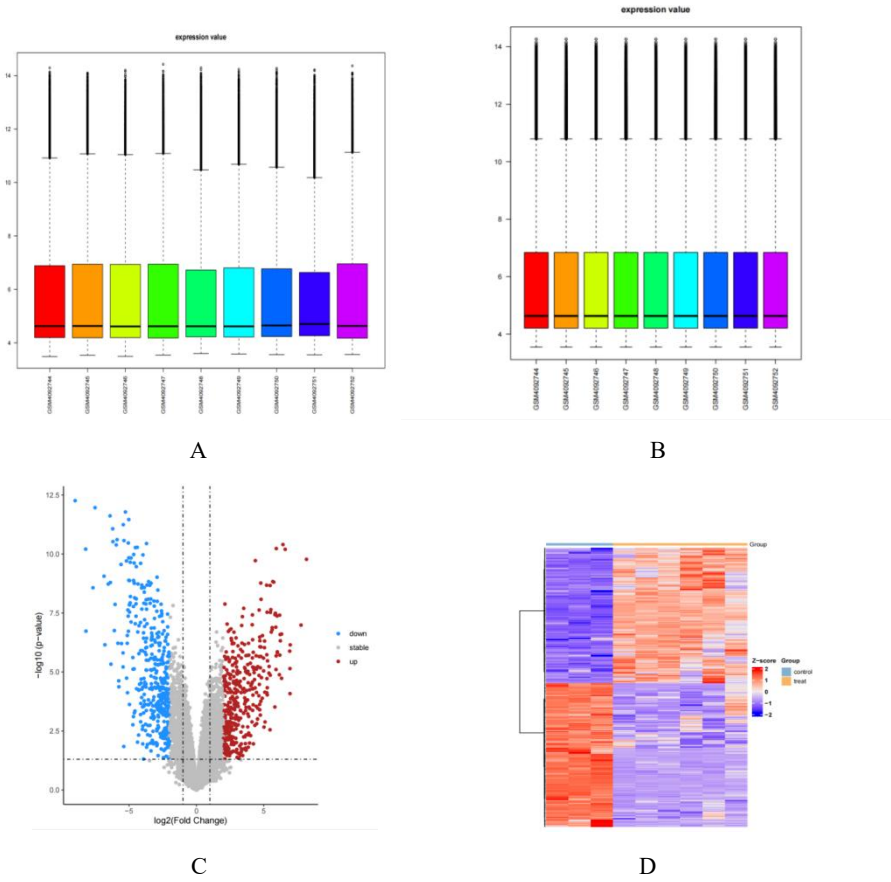

C




D

Figure 1.(A) GSE137902 Substation Culture Box Line Diagram of Substrate Culture Data before Standardization.(B )A GSE137902 Substation Culture Data Standardized Box Line Diagram.(C)Volcano Map, points represent up-regulated genes, and blue points represent down-regulated genes. Genes with no significant difference are shown in black. The differences are set as |logFC|>2 and P<0.05.（D） GSE137902 Substation Culture heatmap.

*Data Processing.*The R languaget includes a data processing and storage facility, operators for array and matrix calculation, a set of data analysis tools graphic functions for analysis and display, and the ability to analyze subsets in geo databases.Data Processing.[5] GEO2R (http://www.ncbi.nlm.nih. gov/geo/geo2r/) is an analysis tool that comes with the GEO database and is used to compare two sets of data; it can be used to analyse any GEO series. P < 0.05 and logFC >1or <–1 were set as the cut-off criteria[6].




A B C D

Figure 2.(A) GSE137902 gelatin culture Box Line Diagram of Substrate Culture Data before Standardization.(B )GSE137902 gelatin culture Data Standardized Box Line Diagram.(C)Volcano Map, points represent up-regulated genes, and blue points represent down-regulated genes. Genes with no significant difference are shown in black. The differences are set as |logFC|>2 and P<0.05.（D） GSE137902 gelatin culture heatmap.

*Identication of Diferentially Expressed Genes (DEGs).*The expression matrix of GSE137902 and GPL13667 platform files is downloaded from GEO website, and the data is converted into expression matrix data and grouping data using manual sorting and Perl language software. The R language is used for standardization of all expressed data. DEGs analysis R package using Limma[7]. Genes which expression multiple changes were greater than 2 and P<0.05 were considered as DEGs. The volcano and heatmap drawings are generated by the R package.

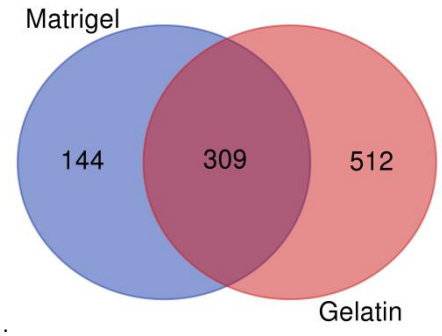
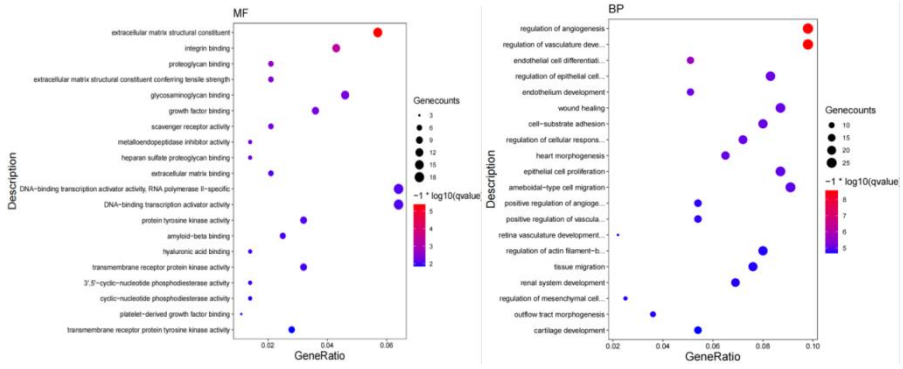
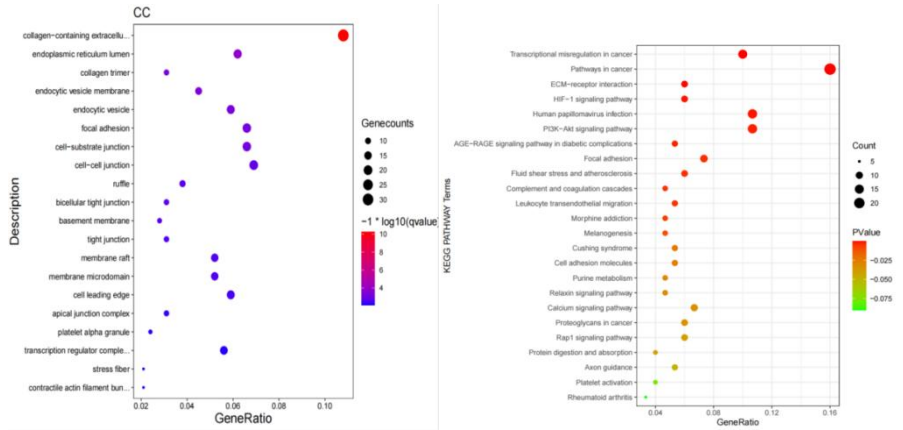



Figure 3.309 genes were obtained from the intersection of gelatin and matrix differential genes

*Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes(KEGG).*Functional and Pathway Enrichment Analyses. First, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis on DEGs by using the Metascape software (http:// metascape.org/gp/index.html#/main/step1) [8]. Metascape is an online analysis tool with integrated discovery and annotation capabilities. To ensure the credibility of the results, we also analysed the data with online tools from the DAVID website and visualized the results via the R language.e DAVID website (https:// david.ncifcrf.gov/home.jsp) is a bioinformatics data resource composed of a comprehensive biological knowledge base and analytical tools [9]. A P value < 0.05 was set as the cut-off criterion.




A B




C D

Figure4.Use R language to enrich, analyze and visualize go and KEGG differential genes after the intersection of colloid and matrix (method: clusterprofiler package）

(A). pathway through which differential genes are enriched in molecular functions(MF).(B) pathway through which differential genes are enriched in biological processes(BP).(C) pathway through which differential genes are enriched in cellular components(CC).(D)Kyoto Encyclopedia of Genes and Genomes

*PPI Network Construction and Module Analysis.*Protein-Protein Interaction (PPI) Network Construction and Module Analysis. STRING (https://string-db.org/) can draw PPI networks after importing common DEGs into search tools to retrieve interacting genes [10]. First, we drew the PPI network diagram of DEGs by using the STRING website. Cytoscape, a free visualization software, was applied to visualize PPI networks and find hub genes [11]. Then, the hub genes were identified by four methods: DEGREE, MCC, DMNC, and MNC in cytoHubba [12].

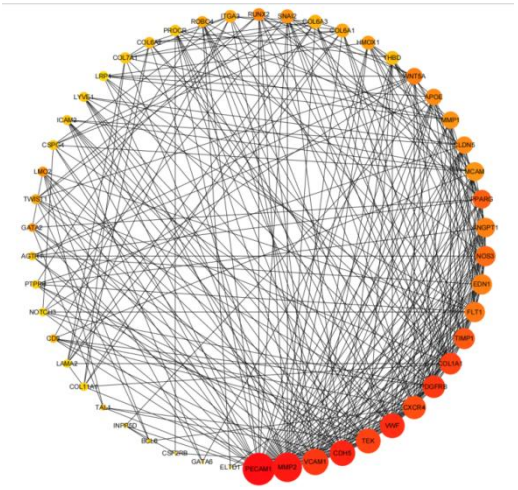


Figure5.Differential genes PPI protein protein interaction network shows the diagram of the top 50 of degree.

*Hub Genes Selection hub.*Gene selection uses a Cytoscape plug-in to identify hub proteins or genes in the PPI network. Using the five algorithms of closeness, degree, EPC, MCC and stress in Cytoscape to obtain the intersection[13], 16 hub genes are obtained; And do enrichment analysis of hub genes.

*Receiver operating characteristic (ROC) curves.*Receiver operating characteristic (ROC) analysis is a tool used to describe the discrimination accuracy of a diagnostic test or prediction model.[14]

# Result

*Identification of DEGs in glioma.*We found the differentially expressed genes on chromosomes of glioma cells through the R studio . R studio was used to investigate the DEGs via mining of the GEO (GSE137902) database (https://www.ncbi.nlm.nih.gov/geo/). We analysed the DEGs in the database and showed them in a heat map and a volcano map. The data were filtered by logFC > 2or<–2 and P< 0.05. .e overlapping DEGs among the 2 datasets were identified, and 309genes were selected and presented using a Venn diagram.

*Gene and pathway enrichment analysis.*GO and KEGG Pathway Analysis.We performed GO analysis through the DAVID website and visualized the data with the R language. Concerning biological processes (BPs), the DEGs were enriched in response to oxygen levels, regulation of vascular permeab, muscle cell proliferation and wound healing .The changes in cellular components (CCs) were significantly enriched in the Collagen-containing extracellues, microvillus, platelet alpha granule and endoplasmic reticulum lumen. The changes in molecular function (MF) were significantly enriched in the platelet-derived growth factor binding, protease binding and growth factor binding.

*Investigation of Glioma hub genes by PPI network analysis.*The identified DEG is submitted to the string database to obtain PPI data. We use Cytoscape 3.6.1 to build a PPI network, and then display the top 50 network diagrams and three

MCODEs in the PPI network. To verify the hub genes related to glioma, we use five algorithms in Cytoscape: proximity, degree, EPC, MCC, and stress to intersect, and obtain hub genes (Cytospace software plug-ins) for determination[15]. All hub genes are MCC, ANGPT1, PECAM1, EDN1, PPARG, VWF, CXCR4, CDH5, TEK, VCAM1, COL1A1, APOE, PDGFRB, HMOX1, TIMP1, CLDN5.

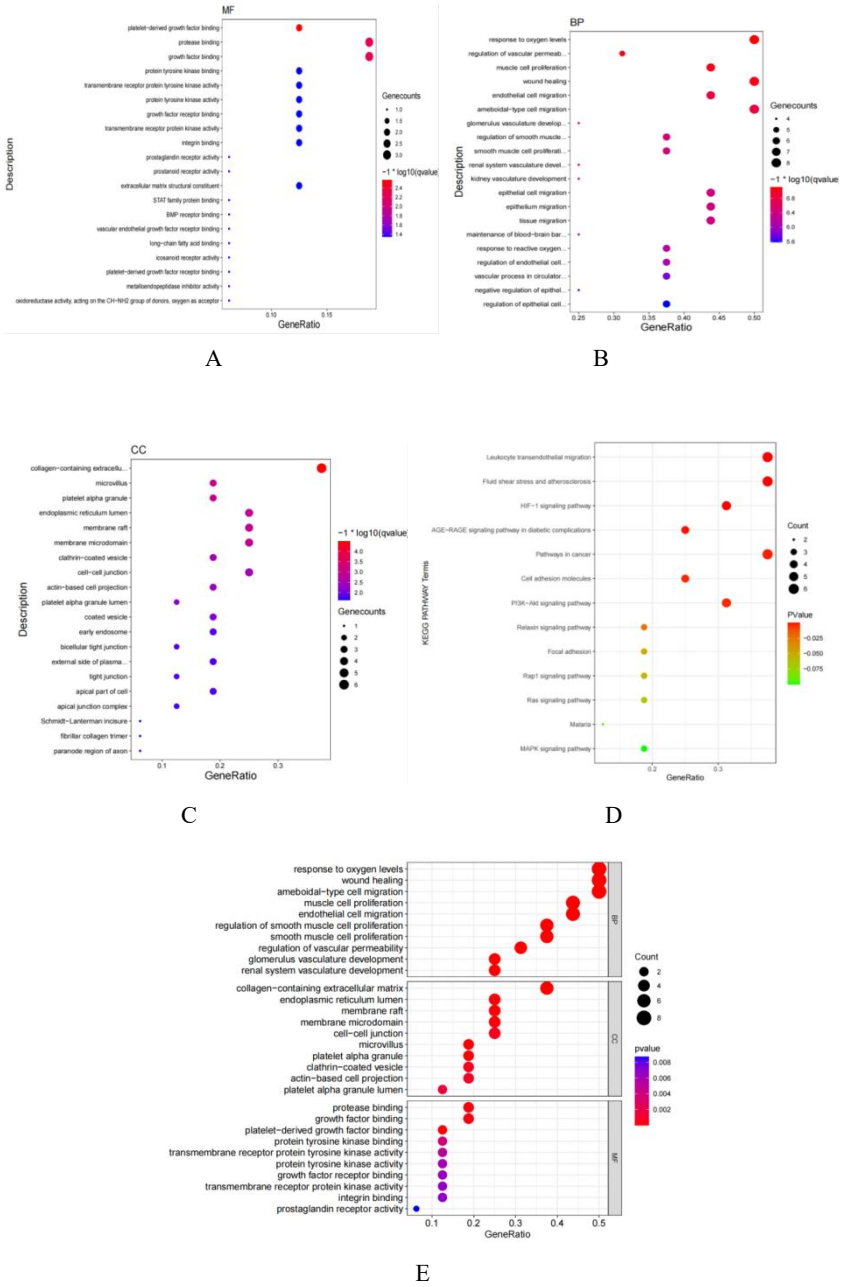



Figure 6. Use R language to enrich, analyze and visualize go and KEGG Hub genes after the intersection of colloid and matrix (method: clusterprofiler package ） (A). pathway through which differential genes are enriched in molecular functions(MF).(B) pathway through which differential genes are enriched in biological processes(BP).(C) pathway through which differential genes are enriched in cellular components(CC).(D)Kyoto Encyclopedia of Genes and Genomes.(E) MF,BP,CC,enrichent of related top ten GO barplot.

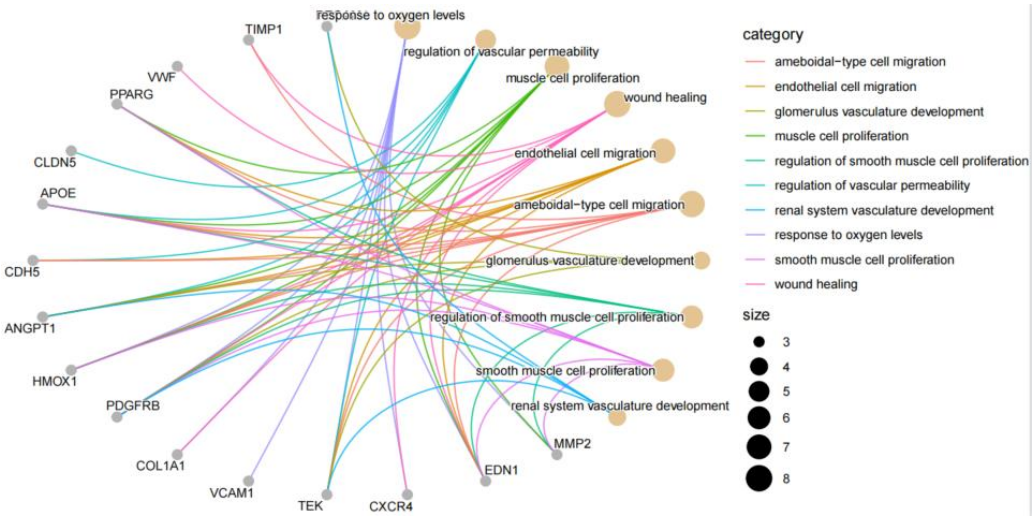
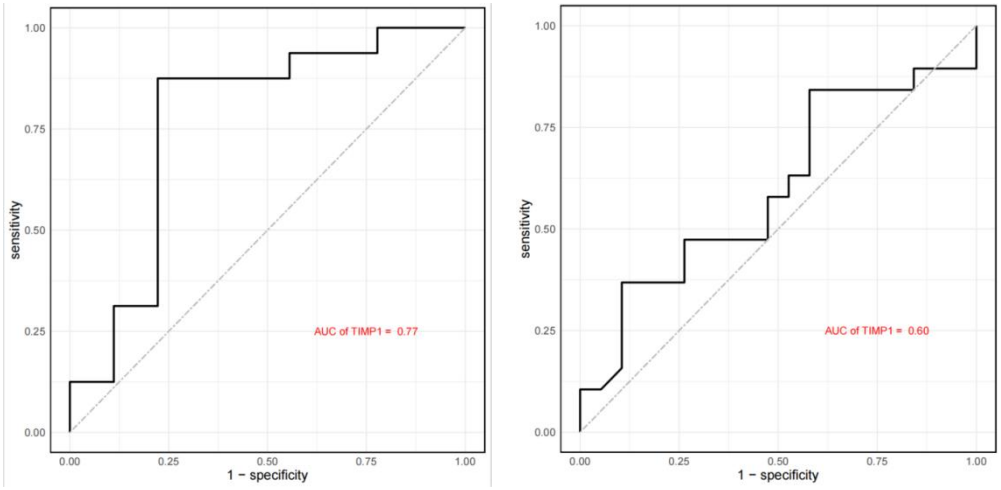



Figure7.Visualization network after GO enrichment analysis

*Verification of hub gene.*In this study, we downloaded the original microarray datasets GSE117423, GSE188256, and GSE90598 from NCBI GEO. Standardize the three datasets, with logFC>0.5 and p.value<0.05 taking differential genes, and obtain 440 DEGs, including 203 down-regulated genes and 237 up-regulated genes. These differential genes intersect with the hub gene, and TIMP1 has been identified as a key gene.

Finally, TIMP1 was verified in the ROC curve, with AUC>0.5 and true positive rate>0.5 for TIMP1. TIMP1 is a key gene associated with glioma that crosses the GSE188256 and GSE90598 data sets.




A B

Figure 8.(A) ROC of TIMP1 in GSE137902 was verified in GSE90598 .(B ) ROC of TIMP1 in GSE137902 was verified in GSE188256.

# Discussion

In this study, 423 down-regulated genes, 398 up-regulated genes, 216 down-regulated genes and 237 up-regulated genes were identified by gse137902 gelatin and matrix respectively. 309 differential genes were obtained from the intersection of the two differential genes. Using R language to analyze go and KEGG enrichment analysis of differential genes, it can be seen that DEGs are mainly enriched in BP, CC, MF. See the following figure for relevant roads. TIMP1[16] gene has been identified as a central gene that may affect the origin and development of glioma, and three pathways have been identified as potential key pathways in glioma.

Like other tumors, glioma is also caused by the interaction between congenital genetic high-risk factors and environmental carcinogens[17]. Some known genetic diseases, such as neurofibromatosis (type I) and tuberculous sclerosis, are genetic predisposing factors of glioma. Patients with these diseases have a much higher chance of glioma than the general population.

Through PPI network analysis, we identified the top 10 Central genes that may affect glioma. According to the interaction score calculated by cytohubba, peacam1, MMP2, vcm1 and cdh5 play an important role[18].

Hub genes are obtained by taking the intersection of five algorithms in Cytoscape: closeness, degree, EPC, MCC and stress. These differential genes are mainly enriched in BP, CC and MF, among which protease binding and growth factor binding are enriched in MF. Collagen containing extractable is enriched in CC. Response to oxygen lever is enriched in BP.[19]

Through the intersection of hub genes and GSE117423 (the overall gene expression profile of human glioma tissue and adjacent normal tissue samples, 6 gliomas and 6 normal tissues), TIMP1 is the key gene. And take the ROC of TIMP1 in GSE137902; It is verified in GSE188256 and GSE90598 respectively, and a conclusion is drawn.There are some limitations in the present study.

# Conclusion

In conclusion, the DEGs, hub genes and signal pathways found in this study can confirm that the key gene TIMP1 is closely related to the occurrence and evolution of glioma[20]. We continue to study the underlying mechanisms by using bioinformatics. Findings of this study may provide candidate targets for the diagnosis and treatment of glioma.

# References

[1]Sun P, Kong Lin. Status and progress of proton and heavy ion therapy in adult gliomas [J]. The Chinese Journal of Radiation Oncology, 2022,31(2):201-207.

[2]Barrett T, Wilhite SE, Ledoux P. et al., “NCBI GEO: archive for functional genomics data sets—update,” Nucleic Acids Research, vol. 41, no. D1, pp. D991–D995, 2012.

[3]Chan BKC. Data Analysis Using R Programming. Adv Exp Med Biol. 2018; 1082: 47-122.

[4]Du Q, Lin YD, Zhang WY, He FM, Xu Y, Chen ZY. Bioinformatics analysis of LMAN1 expression, clinical characteristics, and its effects on cell proliferation and invasion in glioma.[J]. Brain research, 2022, 1789.

[5]Kim WJ, Kim EJ, Kim SK et al., “Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer,” Molecular Cancer, vol. 9, no. 1, 3 pages, 2010.

[6]Zhou Y, Zhou B, Pache L. et al., “Metascape provides a biologist-oriented resource for the analysis of systems-level datasets,” Nature Communications, vol. 10, no. 1, 2019.

[7]Huang DW, Sherman BT, and Lempicki RA, “Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists,” Nucleic Acids Research, vol. 37, no. 1, pp. 1–13, 2009.

[8]Szklarczyk D, Gable AL, Lyon D, et al., “STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” Nucleic Acids Research, vol. 47, no. D1, pp. D607–D613, 2019.

[9]Shannon P, Markiel A, Ozier O, et al., “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” Genome Research, vol. 13, no. 11, pp. 2498–2504, 2003.

[10]Chin CH, Chen SH, Wu HH, et al, “cytoHubba: identifying hub objects and subnetworks from complex interactome,” BMC Systems Biology, vol. 8, no. Suppl 4, S11 pages, 2014.

[11]Singhal A, Cao S, Churas C, et al. Multiscale community detection in Cytoscape.[J]. PLoS computational biology,2020,16(10).

[12]Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves:review of methods with applications in diagnostic medicine. Phys Med Biol. 2018 Mar 29;63(7):07TR01.

[13]Güven E, Afzal M, Kazmi I. Screening the Significant Hub Genes by Comparing Tumor Cells, Normoxic and Hypoxic Glioblastoma Stem-like Cell Lines Using Co-Expression Analysis in Glioblastoma[J]. Genes, 2022, 13(3).

[14]Wang P, Yang X, Zhou N, et al. Identifying a Potential Key Gene, TIMP1, Associated with Liver Metastases of Uveal Melanoma by Weight Gene Co-Expression Network Analysis.[J]. Onco Targets and therapy,2020,13.

[15]Chen WL, Wang YN, Xing H, et al. Progress in the field of glioma in China in 2021 [J / OL]. Union Medical Journal: 1-15 [2022-07-17].

[16]Alevi J. Pulkkanen, Seppo Yla-Herttuala,Gene Therapy for Malignant Glioma: Current Clinical Status,Molecular Therapy,Volume 12, Issue 4,2005,Pages 585-598.

[17]Okura, H., Smith, C.A. & Rutka, J.T. Gene therapy for malignant glioma. Mol and Cell Ther2, 21 (2014).

[18]Perumal E, Devaraj E, Subramanian R. (2022) Quercetin Inhibits the Epithelial to Mesenchymal Transition through Suppressing Akt Mediated Nuclear Translocation of β-Catenin in Lung Cancer Cell Line. Nutrition and Cancer 74:5, pages 1894-1906.

† Chengzhi Cui and Shi Yin have contributed equally to this work and share first authorship.

*Correspondence to:

Dr. Dong Chen, Department of Neurosurgery, Dalian Municipal Central Hospital, Dalian, China

Dr. Jiaming Huang, Department of Neurosurgery, Dalian Municipal Central Hospital, Dalian, China

Dr. Lei Meng, The First Affiliated Hospital of University of South China, Dalian, China