scientific reports

Check for updates

OPEN

# Bioinformatics analysis to identify key invasion related genes and construct a prognostic model for glioblastoma

Jintao Tian[1,5], Jinxi Zhao[1,5], Zhixing Xu[2], Bohu Liu[3], Jun Pu[1], Hongwen Li[4], Qingchun Lei[2], Yu Zhao[1], Weilin Zhou[1], Xuhui Li[1]✉ & Xiaobin Huang[1]✉

**Glioblastoma (GBM) is the most common and lethal brain tumor with limited therapeutic strategies and incomplete studies on its progression and mechanisms. This study aims to reveal potential prognostic marker genes associated with GBM cell invasion, and establish an effective prognostic model for GBM patients. Differentially expressed genes (DEGs) were screened from The Cancer Genome Atlas (TCGA) and the Chinese Glioma Genome Atlas (CGGA), differentially invasive-related genes was obtained, qRT-PCR was used to verify gene expression. The risk scores of individual patients, univariate and multivariate Cox regression were analyzed to investigate the correlation between risk values and glioblastoma, Finally, the risk scores with the prognostic clinical characteristics of the patients, such as PFS, OS were used to build a comprehensive GBM prognostic model. Five DEGs (GZMB, COL22A1, MSTN, CRYGN and OSMR) were significantly associated with GBM prognosis. Pseudotemporal analysis, risk scores (PFS, OS) based on tumor cells revealed that prognostic genes were associated with tumor proliferation and progression. The final prognostic model was developed and validated with good performance with higher accuracy(C-index: 0.675), and it was found that the risk value can serve as an independent prognostic factor for patients with glioblastoma ($p < 0.05$). We constructed a comprehensive prognostic model related to invasion in GBM patients using genetic profiles, survival curves, immune infiltration, and radiotherapy face susceptibility. The model has good predictive ability.**

**Keywords** Glioblastoma, Invasion-related gene, Immunotherapy, Prognostic model

Gliomas are one of the most common malignant tumors of the adult nervous system, accounting for 30–40% of all neurological tumors. On average, 526 out of every 100,000 individuals are diagnosed with gliomas each year, with 17,000 new cases occurring annually. Patients with gliomas are predominantly older, around 60–80 years of age, with a mean age at diagnosis of 64 years[1]. Gliomas are highly aggressive and have a poor prognosis, with a median survival time of 14.6 months after diagnosis, a 2-year survival rate of 26-33%, and a 5-year survival rate of 4-5%[2]. At present, the treatment of glioma patients is still mainly based on surgical treatment, but due to the infiltrative growth of the tumor, there is no obvious boundary between the tumor and the brain tissue, except for the early stage of the tumor is small and is located in the appropriate area, other types of tumors are difficult to do a total resection, generally advocate a comprehensive treatment of postoperative radiotherapy, chemotherapy, etc., which can slow down the recurrence of the disease and prolong the survival period. Although the treatment of glioma has evolved from single surgical treatment to comprehensive treatment combining surgery, radiotherapy, and chemotherapy, its prognosis is still unsatisfactory, and the most common malignant glioma, for which there is no cure due to its special heterogeneity and invasiveness, still has an unsatisfactory prognosis[3]. Therefore, opening up new prognostic assessment treatments as well as therapeutic methods and drugs for GBM has become an urgent goal to be explored. In this study, we aim to develop a new prognostic risk

[1]Department of Neurosurgery, The Second Affiliated Hospital of Kunming Medical University, Kunming 650101, China. [2]Department of Neurosurgery, The Pu'er People's Hospital, Puer 665000, China. [3]Department of Neurosurgery, The Kunming First People's Hospital, Kunming 650011, China. [4]Department of Neurosurgery, The Dali People's Hospital, Dali 671000, China. [5]Jintao Tian and Jinxi Zhao contributed equally to this work. ✉email: 529316932@qq.com; hxbynyx@163.com

model for GBM patients using invasion-related genes, and to explore the response of patients with different risks to chemotherapy and immunotherapy, so as to provide theoretical support for the study of GBM.

Age, gender, mutation status, medications and environmental factors have previously been shown to be associated with clinical GBM outcomes[4]. Among them genetic mapping in GBM turns out to be more and more emphasized. Online databases with a large number of genetically defined GBM samples, such as the Cancer Genome Atlas (TCGA) and the Chinese Glioma Genome Atlas (CGGA), offer great potential to search for other genetic prognostic factors[4].One of the clinical features of GBM is extensive infiltration of parenchymal tissues surrounding the tumor[5]. Gliomas almost never metastasize out of the brain. Infiltration occurs along pre-existing structures such as blood vessels, white matter tracts, and the subarachnoid space and may be orchestrated by specialized cells, leading to collective infiltration[6]. GBM cells can cross tissue barriers by remodeling their own cytoskeleton and extracellular matrix (ECM) and invade either as individual cells or masse[7].The GBM Brain Tumor Initiation Cells (BTIC) as well as differentiated cells can mimic invasion[8]. Glioblastoma (GBM) is composed of highly invasive and destructive tumors, making it difficult to conduct in vitro studies in human derived model systems. The progress of treatment is also limited by the heterogeneity of cells within and between tumors, as well as other factors such as treatment resistance[9].Despite the growing understanding of the mechanisms leading to GBM invasion in vitro and in animal models[8], no prognostic risk model for GBM constructed from invasive genes has been reported to date.

Although researchers have established prognostic models for GBM patients by incorporating prognostic marker genes, no prognostic GBM model has been widely accepted. To date, no invasion-related prognostic models have been studied and/or established. Here, we present a prognostic model for GBM based on TCGA and CGGA data, with special emphasis on correlation with invasion genes.

## Materials and methods

### Data retrieval and processing

The RNA sequencing data and clinical information of 169 GBM patients and 5 normal controls were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) database, and the RNA-seq data of 85 GBM patients were sourced via the Chinese Glioma Genome Atlas (CGGA, http://www.cgga.org.cn) database. The TCGA-GBM cohort and CGGA cohort was treated as the training set the validation set respectively. Additionally, 6664 invasion-related genes (IRGs) of GBM were downloaded from the GBM category in CancerSEA (http://biocc.hrbmu.edu.cn/CancerSEA/).

### Identification of differentially expressed IRGs and enrichment analysis

Initially, the differential analysis was performed in the TCGA-GBM dataset to identify differentially expressed IRGs (DE-IRGs) between GBM samples and control samples, which were related to GBM within the 6664 IRGs with the thresholds of $p < 0.05$ and $|\log(FC)| \geq 1.0$ by edgeR[10,11]. ComplexHeatmap (version 2.6.2) was utilized to plot the volcano plot and heat map of the DE-IRGs. Subsequently, clusterProfiler (version 4.0.2) was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses for DE-IRGs, with a p value < 0.05 was considered statistically significant[12–14]. The enriched result with top 5 p value in each category was selected to visualized.

### Construction and validation of risk model

In order to determine the relationship between DE-IRGs and the survival of GBM patients, the the expression profile of DE-IRGs were extracted from TCGA and combined with the survival data to receive the clinical expression data of 160 GBM samples. Univariate Cox regression analysis was initially employed on the DE-IRGs, ($p < 0.05$, Hazard Ratios (HRs) $\neq 1$), genes related to prognosis were obtained. The genes related to prognosis were further screened using the Least Absolute Shrinkage and Selection Operator (LASSO) regression. LASSO was performed by glmnet (version 3.2-3). The LASSO algorithm, by adjusting the value of λ, was able to selectively compress the coefficients of some unimportant features to zero, thus achieving feature selection. A smaller value of λ meant that the regularization strength was weak, and the model would not compress the coefficients excessively, enabling it to retain more features. Therefore, genes were selected according to the minimum λ value, and the obtained genes were named candidate genes. Subsequently, a multivariate Cox regression analysis was performed on the candidate genes (HRs $\neq 1$). The stepwise regression function step() was used, with the parameter "direction" set to "both" to adjust the multivariate regression model. The genes that passed the multivariate Cox regression analysis were named key genes for subsequent analysis.

The prognostic risk score model was constructed based on the key genes, and the formula was: risk score $= \sum_{n=1}^{n} coef_i * x_i$. Next, a risk score was generated for each of the 160 GBM samples in TCGA. The 160 samples were divided were divided into the high-risk and low-risk groups according to whether its risk score was greater than the median value of the risk scores. Then, a K-M survival curve was constructed for each risk group to compare the survival differences between the them. Beside, a risk curve and a scatter plot of the risk model and a heap map of model genes in each risk group were drawn. The prognostic performance of the risk model was further evaluated by time-dependent receiver operating characteristic (ROC) curve analysis within 1, 3 and 5 years with "pROC" (version 1.17.0.1). Additionally, the same evaluation procedures were employed to the external validation set (85 GBM samples in CGGA) to further verify the applicability of the risk model.

### Identification of independent prognosis factors

Initially, K-M survival analysis was employed between high and low risk groups in stratification of 2 clinicopathologic features which were Age (≥ 60, < 60) and Gender (Male, Female).

Next, to further identify the independent prognosis factors of GBM, the clinicopathologic features with significant survival difference between risk groups of stratification were introduced into univariate and

multivariate Cox proportional hazards regression analyses. The results with $p < 0.05$ in both Cox analyses were considered as the independent prognostic factors of GBM.

### qPCR detection of key gene expression

*Total RNA extraction*

Ten pairs of frozen tissue samples were obtained, the main RNA was extracted using TRIzol reagent, the purified RNA concentration was detected by Fluorescence. Reverse transcription was performed using SureScript-First-strand-cDNA-synthesis-kit from Wuhan Xavier, as follows: remove the components of the reverse transcription kit, melt them at room temperature, centrifuge them briefly, place them on ice, and perform the test according to the requirements of the kit.The qPCR reaction was carried out according to the reaction system of cDNA 3ul, 2xSYBR matser mix 5ul, pre primer (10 μm): 1ul, post primer (10 μm): 1ul.

### Key gene related Immune infiltration

In order to explore the differences in immune cell types between high and low-risk groups, we used the CIBERSORT algorithm (version 1.03) and the LM22 gene set of immune cells to calculate the infiltration ratio of 22 immune cells in the samples. We removed samples with predicted p-values greater than 0.05 and ultimately retained 58 samples. Using Wilcoxon rank sum test to compare the differences of 22 immune cells between high and low-risk groups. In order to effectively identify statistically significant results and rule out the possibility of results occurring by chance, this study selected $p < 0.05$ as the screening threshold[15]. Next, we use the TIDE algorithm to evaluate the sensitivity of all patients to immunotherapy, calculate TIDE scores, and compare TIDE scores between high and low-risk groups using the Wilcoxon rank sum test. Use R package ggplot2 to draw a box plot.

### GSVA analysis

The matrix reflects the relationship between samples and genes, so GSVA transforms a "sample x gene" matrix into a "sample x pathway" matrix, directly reflecting the connection between samples and pathways of interest to readers. Therefore, using the limma package to perform the same analysis on the results of GSVA (still a matrix) can identify pathways with significant differences between samples. These "differentially expressed" pathways, compared to genes, have more biological significance and interpretability. They are a sublimation of GSEA results after the successful combination of statistics and biology, and can be further used for the classification of tumor subtypes and other closely related explorations with biological significance. Since the c5.all.v7.5.1.symbols.gmt encompasses annotation information in three aspects: molecular function, biological process, and cellular component, it integrates a large amount of experimental data and literature reports, ensuring a high level of credibility for the association between the included genes and specific biological functions[16]. Therefore, using c5.all.v7.5.1.symbols.gmt as the reference gene set, we used the R language package "gsva" to score the GO functional items in each sample based on the gene expression profile file. Then, we used the "limma" package to calculate 32 GO functional items with differences between high and low-risk groups (filtering criteria: $|\log2FC| > 0.1$ and p-value $< 0.05$). We used the R package "ComplexHeatmap" to draw a differential pathway map.

### Statistical analysis

All statistical analyzes were performed using R language (version 4.2.2), and $p < 0.05$ was considered statistically significant.

## Results

### Identification and enrichment analysis of DE-IRGs

Through the differential analysis, a total of 2297 DE-IRGs (100 were highly expressed in GBM samples, and 1289 were highly expressed in control samples.) were screened out (Supplementary Table 1, Fig. 1A and B).

Subsequently, it can be observed that the 2297 DE-IRGs were mainly involved in the biological process of regulation of ion transmembrane transport, regulation of membrane potential, T cell activation, modulation of

| Primer | List |
|---|---|
| GZMB F | GTGCGGTGGCTTCCTGATAC |
| GZMB R | CTGGGTCGGCTCCTGTTCTT |
| COL22A1 F | ATGAAAGATGCTGTCAGGGTGG |
| COL22A1 R | CCGTTCCTCGATGGGTAGTG |
| MSTN F | TGAGACCCGTCGAGACTCCTAC |
| MSTN R | CATCAATGCTCTGCCAAATACC |
| CRYGN F | TGAACCGAGTGAACTCCATCC |
| CRYGN R | GTAGTCGCCGTGCTCCAAGA |
| OSMR F | GGTGCTTCTCCTGCTTCTGTAA |
| OSMR R | ACTGTGCTTGTGGTATTGGGAC |
| GAPDH F | CCCATCACCATCTTCCAGG |
| GAPDH R | CATCACGCCACAGTTTCCC |

**Table 1**. Primer list.

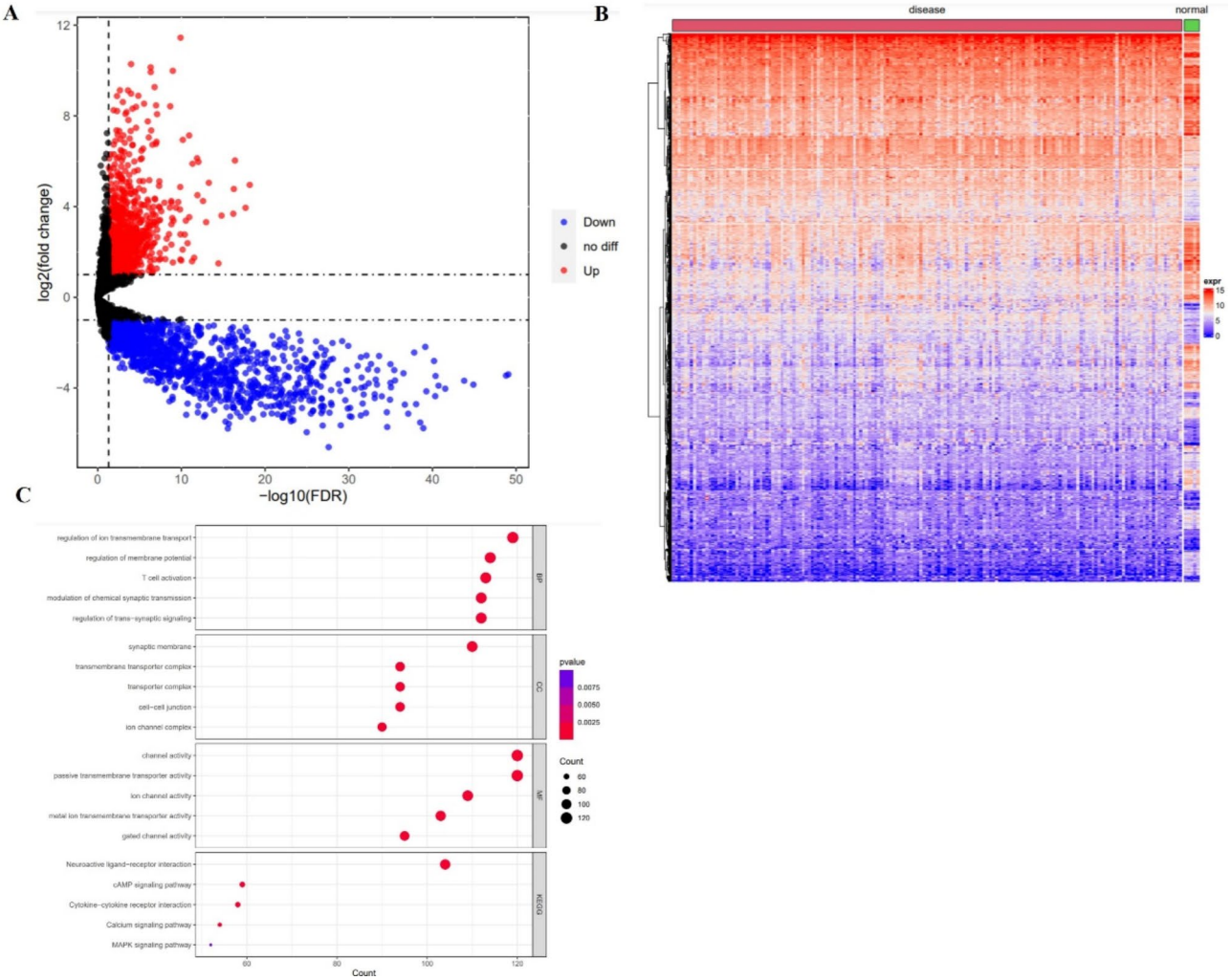


**Fig. 1**. Identification and enrichment analysis of DE-IRGs. (**A**,**B**) A total of 2297 DE-IRGs (100 up-regulated, 1289 down-regulated); (**C**) the KEGG pathways of the 2297 DE-IRGs.

chemical synaptic transmission, and regulation of trans-synaptic signaling; the cellular components of synaptic membrane; and the molecular functions of channel activity and passive transmembrane transporter activity. In terms of the KEGG pathways, the dot chart presented the top 5 enriched pathways which mainly included Neuroactive ligand-receptor interaction, cAMP signaling pathway, Cytokine-cytokine receptor interaction, etc. (Fig. 1C).

## A risk score model of 5 genes was constructed

There were 14 genes related to prognosis were obtained through univariate Cox regression analysis ($p < 0.05$, HRs $\neq 1$) (Fig. 2A). Subsequently, 7 candidate genes were screened out (Fig. 2B). Finally, through multivariate Cox analysis of these 7 candidate genes, 5 key genes were selected, namely GZMB, COL22A1, MSTN, CRYGN and OSMR (Table 2; Fig. 2C). Moreover, MSTN was the protective factor (HRs = 0.9, 95% confidence intervals (CIs): 0.82–0.98), but GZMB (HRs = 1.2, 95% CIs: 1.01–1.35), COL22A1 (HRs = 1.2, 95% CIs: 1.06–1.30), CRYGN (HRs = 1.3, 95% CIs: 1.06–1.49) and OSMR (HRs = 1.2, 95% CIs: 1.00–1.33) were risk factors.

Subsequently, to evaluate the prognostic value of the risk model, the expression levels of the 4 model genes were obtained in the TCGA dataset, and the risk score = GZMB expression × 0.152015597 + COL22A1 expression × 0.158517345118875 + MSTN expression × (-0.108595454) + CRYGN expression × 0.227500042419217 + OSMR expression × 0.141822251623917.

Furthermore, according to the median value of the risk score, the 160 GBM samples were divided into high risk group (80 samples) and low risk group (80 samples). The K-M curves of the risk groups revealed that the survival probability of low risk group was significantly higher ($p = 2.545e\text{-}10$) (Fig. 2D). Moreover, the risk curve of the risk groups illustrated that the risk score of high risk group was much higher than low risk group, and the patients in high risk group were with less survival time, and samples in high risk group tended to be dead (Fig. 2E). Further, the expressions of key genes were shown in the heatmap (Fig. 2F). It can be found from the

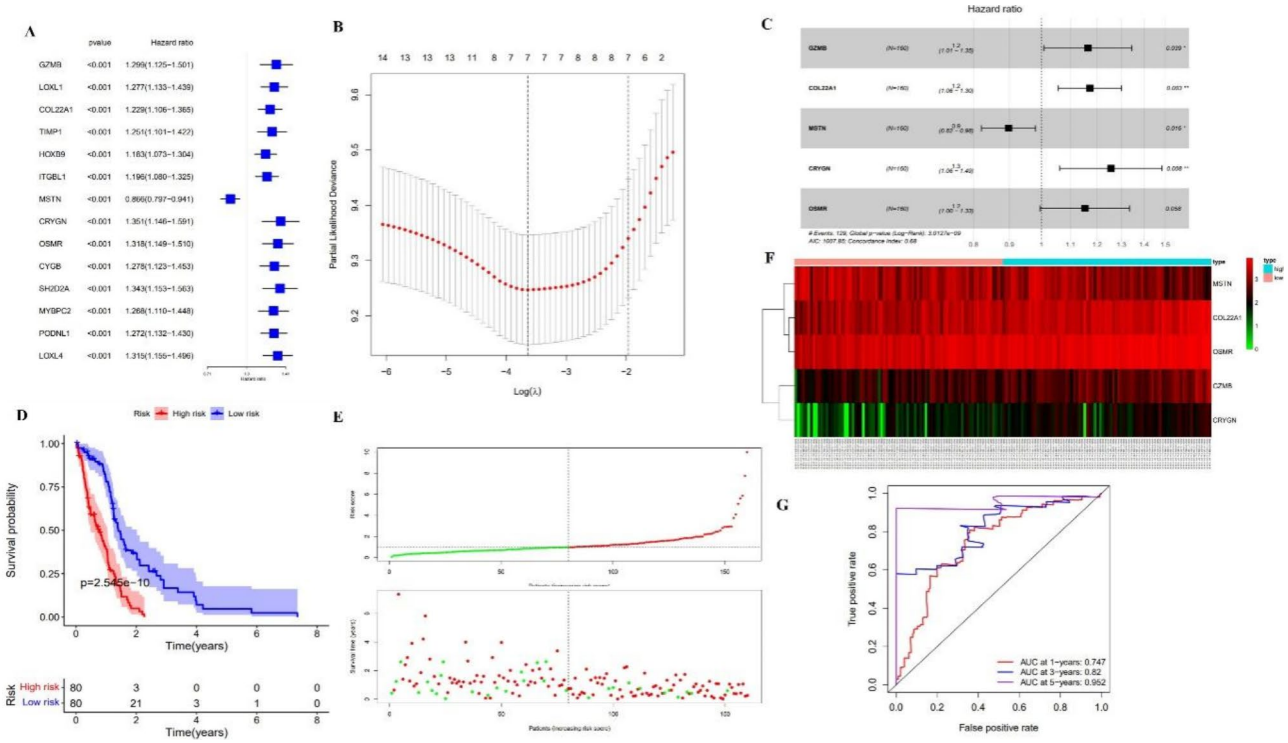


**Fig. 2**. A risk score model of 5 genes was constructed. (**A**) The univariate Cox regression analysis were chosen for Lasso analysis; (**B**) 7 genes were screened out; (**C**) multivariate Cox analysis of the 7 genes; (**D**) The K-M curves of the risk groups; (**E**) the risk curve of the risk groups; (**F**) the heatmap of the model genes were; (**G**) the ROC curves of the model.

| Id | Coef | HR | HR.95 L | HR.95 H | pvalue |
|---|---|---|---|---|---|
| GZMB | 0.1520 | 1.1642 | 1.0077 | 1.3450 | 0.0391 |
| COL22A1 | 0.1585 | 1.1718 | 1.0559 | 1.3003 | 0.0028 |
| MSTN | − 0.1086 | 0.8971 | 0.8211 | 0.9800 | 0.0161 |
| CRYGN | 0.2275 | 1.2556 | 1.0613 | 1.4851 | 0.0079 |
| OSMR | 0.1418 | 1.1524 | 0.9951 | 1.3345 | 0.0582 |

**Table 2**. Multivariate Cox analysis screening yielded five key genes. HR, hazard ratio.

ROC curves of the model that the AUCs of all time periods were greater than 0.7, which indicated the prognostic risk score model was effective (Fig. 2G).

In terms of the validation in the validation set, the 85 GBM samples were separated into 42 in high risk group and 43 in low risk group. The validation results was consistent with them in the training set. The AUCs of the the ROC curves in all 3 year were greater than 0.6, which also suggesting the risk model could effectively predict the prognosis model (Supplementary Fig. 1A-1D).

## qPCR detection of mRNA expression of key genes

The results of qPCR after collecting paracancerous and cancerous tissues and extracting total RNA by TRIzol are shown in Fig. 1, the four invasion-related genes (GZMB, CRYGN, OSMR and MSTN) involved in the construction of GBM prediction model were all significantly increased in GBM tumor tissues ($p < 0.01$), which might be positively correlated with the progression of GBM. In addition the expression of gene COL22A1 gene had a tendency to increase compared with the paraneoplastic group, but $p = 0.6698$, did not reach a significant difference (sees in Fig. 3 and Table 3).

## Immune infiltration

In order to explore the differences in immune cell types between high and low-risk groups, we used the CIBERSORT algorithm (version 1.03) and the LM22 gene set of immune cells to calculate the infiltration ratio of 22 immune cells in the samples. We removed samples with predicted p-values greater than 0.05 and ultimately retained 58 samples. Using Wilcoxon rank sum test to compare the differences of 22 immune cells between

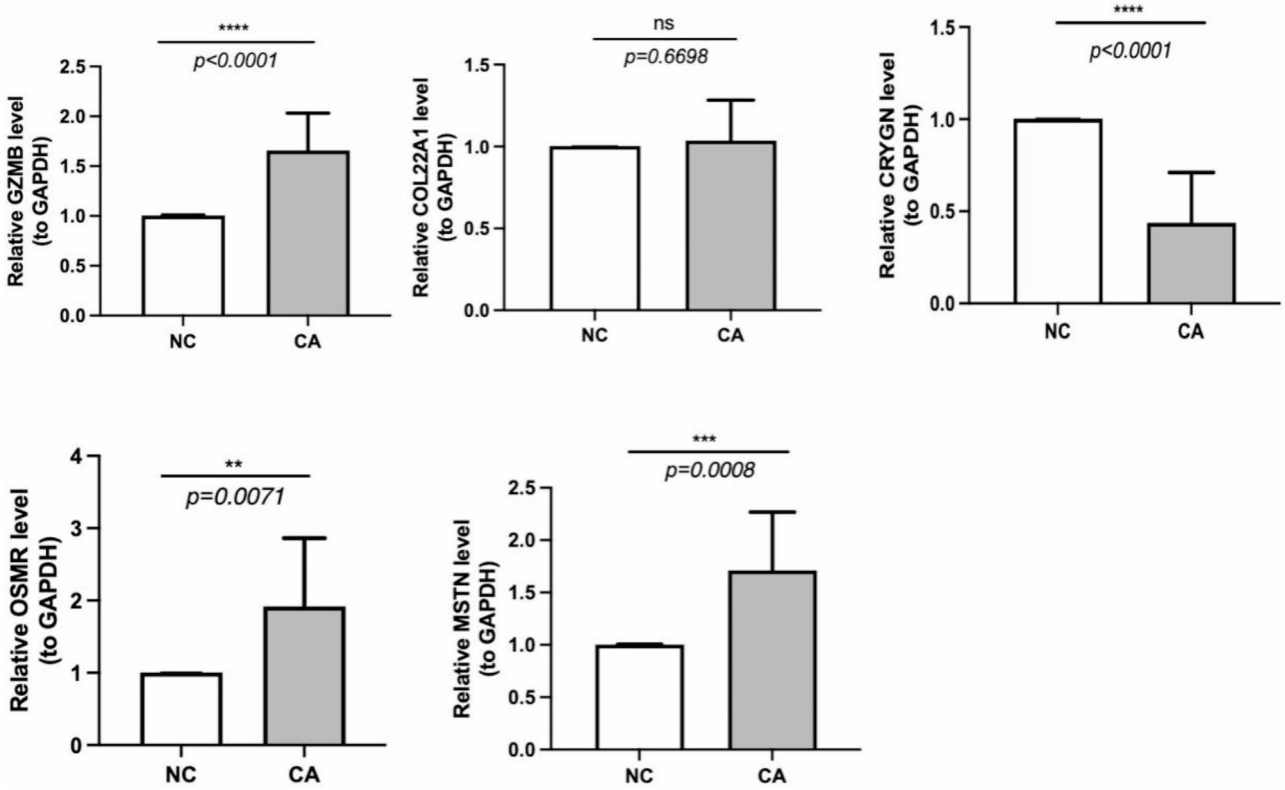


**Fig. 3**. qPCR detection of mRNA expression of key genes. NC stands for Paraneoplastic Peritumoral and CA stands for Carcinoma Disease.

| | NC | Ca | t, df | *p* |
|---|---|---|---|---|
| GZMB | 1.0042 ± 0.0043 | 0.9601 ± 0.5898 | t = 0.2358 df = 9 | 0.8189 |
| COL22A1 | 1.0027 ± 0.0023 | 1.1492 ± 1.0036 | t = 0.4622 df = 9 | 0.6549 |
| MSTN | 1.0022 ± 0.0023 | 1.5106 ± 1.5360 | t = 1.046 df = 9 | 0.3227 |
| CRYGN | 1.0027 ± 0.0024 | 1.2371 ± 0.9301 | t = 0.7968 df = 9 | 0.4461 |
| OSMR | 1.0039 ± 0.0038 | 1.6562 ± 1.7747 | t = 1.162 df = 9 | 0.2751 |

**Table 3**. mRNA expression of key genes detected by qPCR. NC stands for Paraneoplastic Peritumoral and CA stands for Carcinoma Disease.

high and low-risk groups, two immune cells (T. cells. CD4.memory.resting and T. cells. gamma. delta) showed significant differences between high and low-risk groups ($p < 0.05$) (details shown in Fig. 4).

Next, we used the TIDE algorithm to evaluate the sensitivity of all patients to immunotherapy, calculated TIDE scores, and compared the TIDE scores of the high and low-risk groups using the Wilcoxon rank sum test. The results showed a significant difference in TIDE scores between the high and low-risk groups ($p < 0.05$), and a box plot was plotted using the R package ggplot2 (Fig. 5-A). Count the number of people in the high and low-risk groups who respond and do not respond to immunotherapy separately (Fig. 5-B). Finally, a correlation analysis was conducted between the risk value and TIDE score, and the results showed a significant negative correlation between the two ($R$=-0.156, $p = 0.048$) (Fig. 5-C). Using the R package "pRRophetic" to predict patient sensitivity to drugs, it was found that there were significant differences in patient sensitivity to BI.2536, CGP.082996, GW843682X, and KIN001.135 drugs between the high-risk and low-risk groups (Fig. 6).

## GSVA enrichment

GSVA analysis can be understood as follows: the expression matrix reflects the relationship between the sample and the gene, and GSVA transforms a "sample x gene" matrix into a "sample x pathway" matrix, directly reflecting the connection between the sample and the pathway of interest to the reader. Therefore, using the limma package to perform the same analysis on the results of GSVA (still a matrix) can identify pathways with significant differences between samples. These "differentially expressed" pathways, compared to genes, have more biological significance and interpretability. They are a sublimation of GSEA results after the successful combination of statistics and biology, and can be further used for the classification of tumor subtypes and other closely related

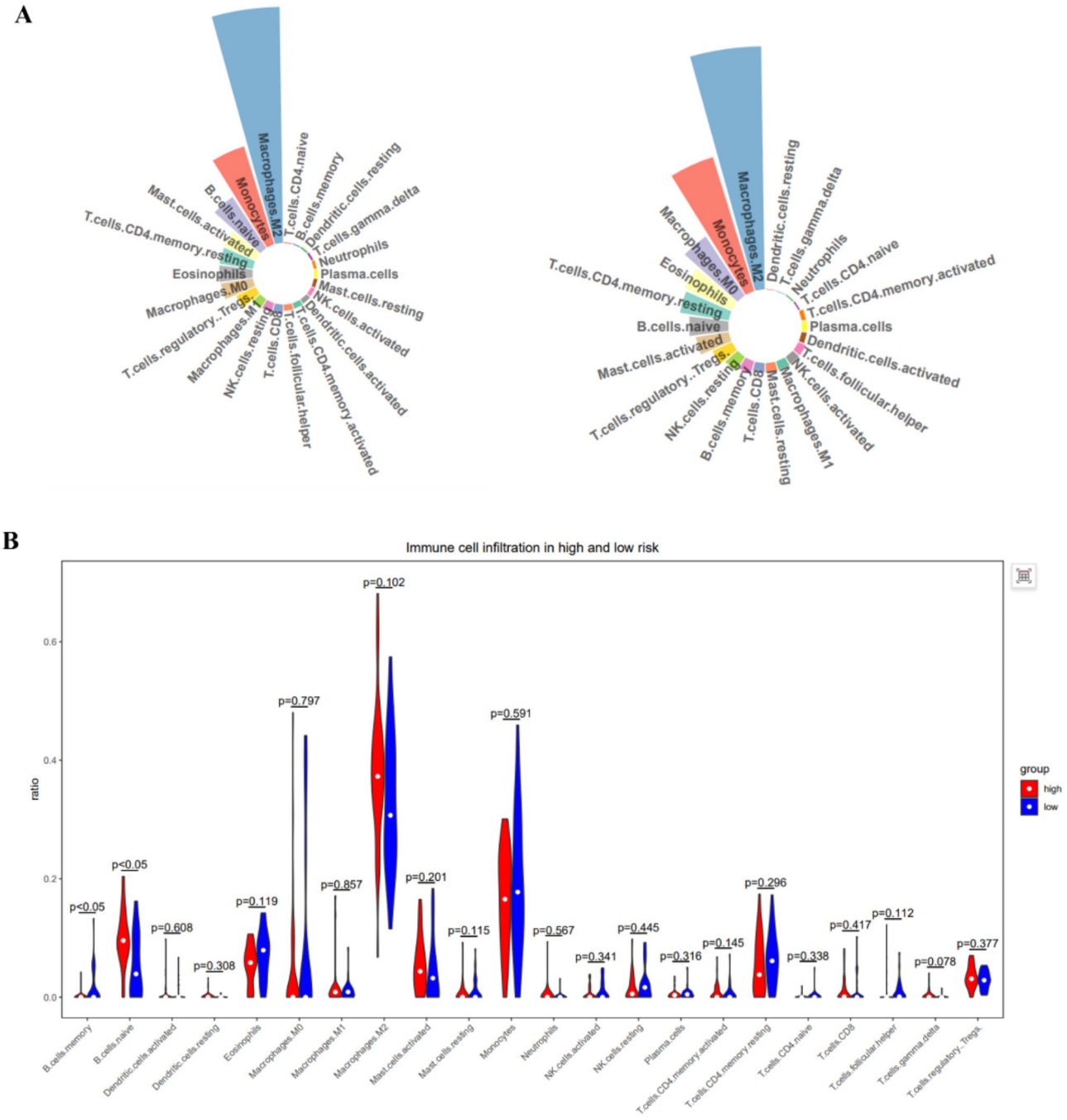


**Fig. 4**. The Immune infiltration related to the risk model. (**A**) The proportion of 22 types of immune cells in the high-risk group, with the left figure representing the proportion of immune cells in the high-risk group and the right figure representing the proportion of immune cells in the low-risk group; (**B**) The differences of 22 immune cells between high and low-risk groups, with the horizontal axis representing the 22 immune cells and the vertical axis representing the proportion of immune cells in the sample. Red represents high-risk groups, and blue represents low-risk groups.

explorations with biological significance. Using c5.all.v7.5.1.symbols.gmt as the reference gene set, we used the R language package "gsva" to score the GO functional items in each sample based on the gene expression profile file. Then, we used the "limma" package to calculate the differences in GO functional items between high and low-risk groups (filtering criteria: | log2FC |>0.1 and p-value < 0.05), totaling 32 items. We used the R package "ComplexHeatmap" to draw the differential pathway map (see in Fig. 7).

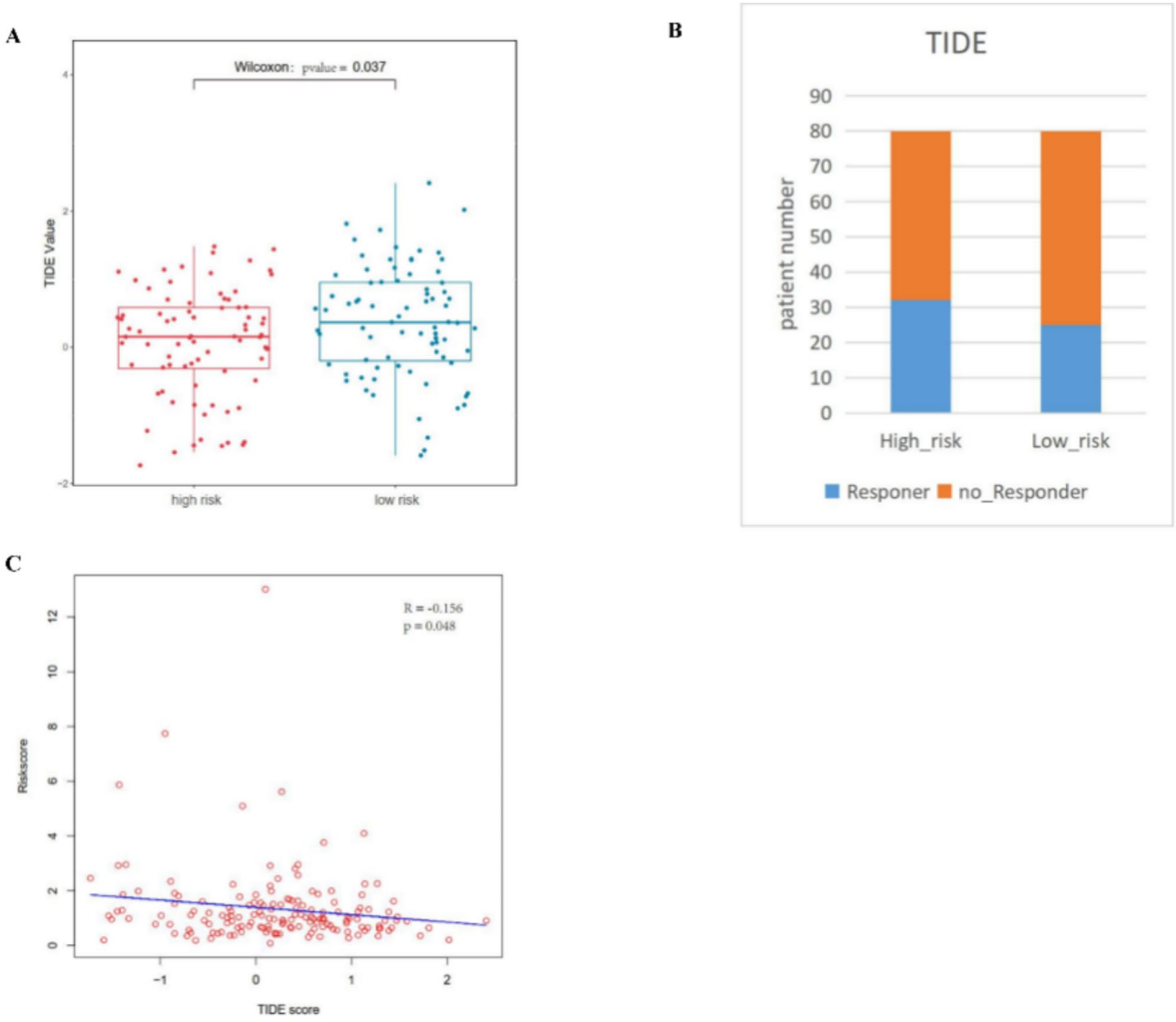


**Fig. 5**. The TIDE and immune therapy response of the risk model. (**A**) The difference in TIDE scores between high and low-risk groups, with red representing the high-risk group, blue representing the low-risk group, and the vertical axis representing the TIDE score. (**B**) Statistics of the number of people responding to immunotherapy in high-risk groups, with blue representing response, orange representing non response, and the vertical axis representing the number of patients. (**C**) The correlation between risk value and TIDE score, with the horizontal axis representing TIDE score and the vertical axis representing risk value.

### Related to age and gender with the risk model

Moreover, we found the survival probability was significantly higher in the low risk group of every subgroup of both Age and Gender. Subsequently, the results of both Cox analyses demonstrated that the p values of Age and risk score were less than 0.05 in both Cox analyses. As a result, risk score and Age were the independent prognostic factors for GBM patients (Supplementary Fig. 2).

## Discussion

In this study, we analyzed the RNA-seq data of glioblastoma (GBM) from the TCGA database bioinformatically and constructed a prediction model by using the five genes screened (GZMB, CRYGN, OSMR, MSTN and COL22A1) as key genes. First, we performed differential analysis, utilizing the edgeR package with the screening condition: $|\log2\text{fold change}| \geq 1.0$; $p < 0.05$), and finally obtained 2297 differential invasion-related genes. Secondly, we performed risk model construction and evaluation based on the 2297 genes. In the TCGA database, the risk model was constructed based on the survival data of 160 samples, and five prognostic genes (GZMB, COL22A1, MSTN, CRYGN, and OSMR) were obtained after unifactorial, LASSO, multifactorial Cox analyses, and qPCR detected the mRNA expression of the above genes, and it was found that the gene expression was significantly increased in the GBM group compared with the control group. The K-M curve ($p < 0.05$), risk curve and ROC curve (AUC > 0.7) showed that the risk model had good predictive ability. MSTN was the

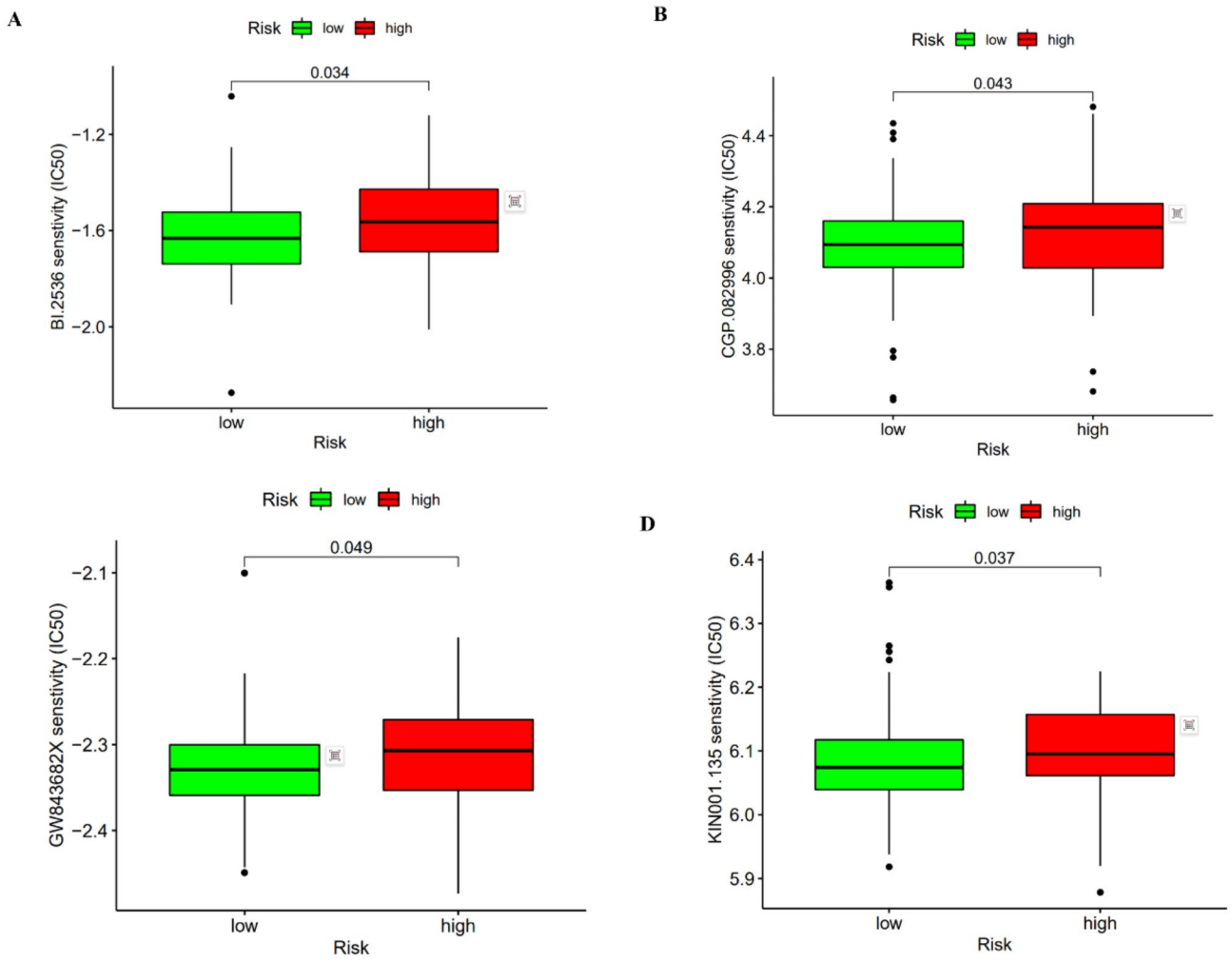


**Fig. 6**. The drug sensitivity of the risk model. Red represents high risk, green represents low risk, and the vertical axis represents IC50, (**A**) BI.2536, (**B**) CGP.082996, (**C**) CGP.082996, (**D**) KIN001.135.

protective factor and the other 4 genes were risk factors. In this study, we confirmed that the re-risk value can be used as an independent prognostic factor for patients with glioblastoma ($p < 0.05$); we constructed a column chart based on the clinical factors and risk values, and plotted a calibration curve to validate the column chart, which showed that the column chart model has a high accuracy for the diagnosis of the disease (C-index: 0.675). The risk model shows that the survival rate of high-risk group patients is lower, which is consistent with the research results of Lin et al.[17].

Meanwhile, we analyzed the differences in immune cell types between high and low-risk groups, and found that two types of immune cells (T. cells. CD4. memory. resting and T. cells. gamma. delta) had significant differences between the high and low-risk groups ($p < 0.05$). Using the TIDE algorithm to evaluate the sensitivity of all patients to immunotherapy, we found that the model risk value was significantly negatively correlated with patients' immune response, and through GSVA analysis, we found that the risk model was closely related to pathways such as glycolysis, immune infiltration, and angiogenesis. Finally, we used the R package "pRRophetic" to predict the sensitivity of patients to drugs, and found that there were significant differences in the sensitivity of patients to BI.2536, CGP.082996, GW843682X, and KIN001.135 drugs between the high-risk and low-risk groups.

Granzyme B (GzmB) is a protease of the chymotrypsin family that is expressed by differentiated cytotoxic T-lymphocytes (CTLs), as well as by certain other cell types such as activated keratinocytes[18]. In cytotoxic cells, GzmB is a component of cytolytic granules[18]. GzmB can escape from the endolysosomal compartment and gain access to many of the key proteins involved in the execution of the apoptotic program[19]. GzmB may also play an important role in antitumor immune responses, effectively inducing tumor cell death, and is also capable of inducing apoptosis in a wide range of drug- and death-receptor-resistant cell lines[20]. In patients with poor response, the proportion of CD8 GZMB + and CD8 proliferation increased after treatment. it is well known that both immunity and angiogenesis play key roles in tumor progression, but no studies have been reported in GBM. In the present study, we demonstrated for the first time that GzmB expression was significantly increased in GBM tumor tissues and positively correlated with GBM progression (OS, PFS), which may be a predictive molecule


**Fig. 7**. The GO analysis of the risk model.

for GBM.The previous research has confirmed that GzmB-mediated autoantigen production can be functionally altered to include cytotoxic particles induced in the redistribution of nuclear antigens to cytoplasmic lysates during cytotoxicity, which may be released during granule-induced death and influence antigen presentation pathways[21]. It is also involved in ECM remodeling and angiogenesis[18]. Therefore, we speculate that GzmB may participate in GBM tumor invasion and immune resistance by regulating tumor angiogenesis or regulating CD8 immune resistance.

Crygn belongs to the γ-crystallins, which are small intracellular proteins of 174 to 182 amino acids with a molecular weight of about 21 kDa[22].γ-crystallin is expressed in the nervous system such as the retina and hippocampus[23]. a recent study reported that Crygn expression was reduced in tissues of patients with cancerous hemangiomas[24]. however, there are no studies that reported its role in the progression of tumors or even GBM. The present study confirmed that Crygn expression tended to increase in the GBM group and was positively correlated with cancer progression.

The tumor suppressor M receptor (OSMR) is a member of the interleukin-6 receptor family and performs a variety of cellular functions including regulation of homeostasis, cell growth and differentiation. OSMR is expressed in many tumor cell types including sarcoma, melanoma, glioma, breast and prostate cancers[25]. It has been reported that oncosuppressor M (OSM), which acts as a ligand for OSMR, also regulates different features of cancer[26]. OSMR expression is up-regulated in mesenchymal and classical glioblastoma subtypes, and OSMR up-regulation is significantly associated with poor patient prognosis[27]. Multivariate analyses of human patient databases housed at TCGA and REMBRANDT indicate that OSMR is a significant predictor of survival, even after controlling for other factors that affect patient survival (including patient age, tumor grade STAT3 expression and IDH1 status). Tumor analysis using human BTSC xenografts in immunodeficient SCID mice showed that knockdown of OSMR impaired their ability to form tumors in vivo[28]. The results of the present study also confirmed that OSMR expression in GBM tumor tissue was upregulated and positively correlated with patient prognosis, which can be used as a GBM predictor. OSM increases tumor growth and metastasis in prostate and breast cancers and may promote epithelial-mesenchymal metastasis[29]. OSMR forms a co-receptor complex with EGFRvIII to amplify receptor tyrosine kinase signaling and glioblastoma development. Deletion of OSMR impairs OXPHOS, promotes reactive oxygen species (ROS) production, and induces cell death. Importantly, deletion of OSMR is sufficient to sensitize glioblastoma tumors to IR therapy and prolong lifespan[30]. Heterocellular OSM-OSMR signaling reprogrammed fibroblasts to promote stem cell carcinoma growth and metastasis[31]. Therefore, it is speculated that OSMR may participate in promoting GBM invasion through multiple aspects such as promoting EMT, inhibiting tumor cell death, and inducing cell stemness.

Muscle growth inhibitor (MSTN), a member of the transforming growth factor-β superfamily, plays a crucial negative regulatory role in skeletal muscle growth and aging[32]. Recently, it has also been found that MSTN is highly enriched in GBM tissues[33,34]. This study found that MSTN was significantly increased in GBM tumor tissues and was positively correlated with tumor progression. By inhibiting the secretion of MSTN in skeletal muscle and its downstream signaling pathway in bone, the Wnt/β-catenin signaling pathway can be activated[35], and Wnt/β-catenin is involved in the progression of GBM[36]. There is limited research on MSTN, and it is speculated that it may promote GBM invasion by activating the Wnt/β-catenin pathway as mentioned above.

Collagen type XXII alpha 1 chain (COL22A1) is present on human chromosome 8q24.2 and exhibits unique localization at tendon junctions, tendons, heart, articular cartilage and skin[37]. Most studies have suggested that the function of COL22A1 is to stabilize tendon junctions and strengthen skeletal muscle attachment during contractile activity[38]. However, a small number of recent studies have shown that COL22A1 transcript levels are elevated in head and neck squamous cell carcinoma (HNSCC) and have been suggested to be a prognostic predictor of HNSCC[39]. suggesting that COL22A1 is involved in tumor progression. However, there is no study on its involvement in GBM correlation. In this study, we innovatively found that COL22A1 expression was significantly increased in GBM, and its expression was correlated with tumor progression and prognosis. COL22A1 maintains vascular stability, and mutations in COL22A1 may be associated with intracranial aneurysms[40]. COL22A1 homologous box 1 transcriptional regulation inhibits nasopharyngeal carcinoma cell senescence through G1/S phase arrest[41].Which showed that COL22A1 may promote GBM invasion and drug resistance progression by regulating tumor angiogenesis and cell cycle.

However, there are currently few reports on the key factors of the above model and the mechanism of GBM progress. More and more evidence suggests that the heterogeneity and functional status of infiltrating T cells in GBM tumors play a critical role in anti-tumor immunity and response to immunotherapy[42]. The glioma microenvironment contains CD4 + and CD8 + T cells, which increase with tumor grade[43]. In this study, we analyzed the differences in immune cell types between high and low-risk groups and found that all patients were sensitive to immunotherapy. Among them, two types of immune cells (T. cells. CD4. memory.resting and T. cells. gamma. delta) showed significant differences between the high and low-risk groups ($p < 0.05$). CD4 + T cells have great plasticity, and the CD4 + T cell subset of interleukin-8 (IL-8) has been shown to be associated with poor prognosis in GBM 16–18. And gamma delta T cells are an important component of non-traditional T cells, which recognize non peptide antigens without the help of classical MHC molecules. The rest include mucosa associated invariant T cells (MAIT) and natural killer T cells (NKT), accounting for only 1–5% of T cells in blood, lymph nodes, and spleen[44]. T cell gamma delta has been used in adoptive immunotherapy, mainly using phosphorylated antigens to amplify human peripheral blood T cell gamma delta. T cell gamma delta mainly includes two subpopulations, namely V δ 1 T cells and V δ 2 T cells. V δ 1 T cells exert immunosuppressive functions to promote tumor development, while V δ 2 T cells mainly participate in the body's immune surveillance of tumors and defense responses against pathogen invasion[45]. Through in vitro activation and expansion, gamma delta T cells can restore their anti-tumor activity and have shown effects on prolonging survival time and slowing tumor progression in animal models[46,47]. Gamma δ T cells are considered a promising immunotherapy method due to their unique tissue localization characteristics and multiple target recognition mechanisms. Especially in targeting tumors that do not respond to existing therapies, gamma delta T cells have shown great potential[48]. Based on the above research results, we speculate that the risk model may exert its effects by regulating gamma delta T and CD4 + T. However, it is worth noting that this study found that T cell CD4 memory restoration is a differential immune cell among GBM patients at different risks.This will help analyze why the immune status of GBM patients varies due to different risks, and open up new ideas for immunotherapy targeting different risk groups.

In addition, we found significant differences in the sensitivity of high and low-risk GBM patients to four drugs, BI.2536, CGP.082996, GW843682X, and KIN001.135, between the high and low-risk groups. Polo like kinase 1 (PLK1) is a serine/threonine kinase that plays a critical role in cell cycle control and is associated with tumor growth and prognosis. Treatment with PLK1 inhibitor BI 2536 for 24 h resulted in mitotic arrest and increased apoptosis of GBM cells[49]. BI 2536 regulates the stemness of EGFRvIII positive glioma stem cells (GSCs) and induces apoptosis through Notch1-SOX2 signaling[50]. The research on the treatment of GBM with the other three drugs is limited. Therefore, it is speculated that the predictive models constructed for the five key genes mentioned above may be mainly related to drug treatments such as BI 2536. The risk model constructed in this study may be involved in the treatment of GBM with PLK1 inhibitor, playing a key regulatory role in the treatment of GBM with sensitive drugs.

## Limitation

There are some limitations in this study, such as a relatively small sample size and reliance on public datasets. In the future, molecular biology experiments will be conducted to screen one or two of the five key genes mentioned above to explore their specific molecular mechanisms in GBM invasion or resistance. And corresponding prospective clinical studies need to be added at the clinical level to validate the risk model and its application in different patient populations.

## Conclusion

In this study, five prognostic genes (GZMB, CRYGN, OSMR, MSTN and COL22A1) were increased in GBM tumor tissues by bioinformatics, and were inversely correlated with the PFS and OS of the tumor patients and positively correlated with the tumor progression, which provided important references for the future diagnosis, mechanism research and treatment of the disease. Meanwhile, the relevant results and findings obtained in this study need to be further confirmed and verified by subsequent related studies, and focused on exploring the molecular mechanisms of the above genes inducing the progression of GBM. In the future, we will delve into

the molecular mechanisms and targeted therapeutic potential of identifying genes alone or in combination to promote GBM progression. And further integrate multiple omics data and prospective studies at the clinical level to improve the prognosis of GBM.

## Data availability

The datasets analyzed during the current study are available from the following public resources: TCGA (The Cancer Genome Atlas, https://portal.gdc.cancer.gov/); CGGA (Chinese Glioma Genome Atlas, http://www.cgga.org.cn/); CancerSEA (Cancer Single-cell State Atlas, http://biocc.hrbmu.edu.cn/CancerSEA).

Received: 31 October 2024; Accepted: 18 March 2025
Published online: 28 March 2025

## References

1. Chen, R. et al. Glioma subclassifications and their clinical significance. *Neurotherapeutics: J. Am. Soc. Experimental Neurother.* **14**(2), 284–297 (2017).
2. Fabian, D. et al. Treatment of glioblastoma (GBM) with the addition of Tumor-Treating fields (TTF): A review. *Cancers.* **11**(2). (2019).
3. Di Delello, L. et al. Improving Temozolomide biopharmaceutical properties in glioblastoma multiforme (GBM) treatment using GBM-targeting nanocarriers. *Eur. J. Pharm. Biopharmaceutics: Official J. Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik E V.* **168**, 76–89 (2021).
4. Aldape, K. et al. Glioblastoma: pathology, molecular mechanisms and markers. *Acta Neuropathol.* **129**(6), 829–848 (2015).
5. Lim, E. et al. Crosstalk between GBM cells and mesenchymal stemlike cells promotes the invasiveness of GBM through the C5a/p38/ZEB1 axis. *Neuro-oncology* **22**(10), 1452–1462 (2020).
6. DeCordova, S. et al. Molecular heterogeneity and immunosuppressive microenvironment in glioblastoma. *Front. Immunol.* **11**, 1402 (2020).
7. Zuo, S. et al. Pan-Cancer analysis of immune cell infiltration identifies a prognostic immune-Cell characteristic score (ICCS) in lung adenocarcinoma. *Front. Immunol.* **11**, 1218 (2020).
8. Xie, X. et al. Quiescent human glioblastoma cancer stem cells drive tumor initiation, expansion, and recurrence following chemotherapy. *Dev. Cell.* **57**(1), 32–46e8 (2022).
9. Arabel, V. Z. et al. Tumor cell invasion in glioblastoma. *Int. J. Mol. Sci.* **21**(6). (2020).
10. Jian-Huang, H. et al. The potential crosstalk genes and molecular mechanisms between glioblastoma and periodontitis. *Sci. Rep.* **14**(1). (2024).
11. B L, W., Intracerebral hematomas. *Neurosurgery.* **17**(3). (1985).
12. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).
13. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Science: Publication Protein Soc.* **28**(11), 1947–1951 (2019).
14. Kanehisa, M. et al. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
15. Goel, H., Raheja, D. & Nadar, S. Evidence-based medicine or statistically manipulated medicine? Are we slaves to the P-value? *Postgrad. Med. J.* **100**(1185), 451–460 (2024).
16. Yingyu, W. et al. The shared biomarkers and pathways of systemic lupus erythematosus and metabolic syndrome analyzed by bioinformatics combining machine learning algorithm and single-cell sequencing analysis. *Front. Immunol.* **13**(0). (2022).
17. Lin, Z. et al. Identification of an immune-related prognostic risk model in glioblastoma. *Front. Genet.* **13**, 926122 (2022).
18. Turner, C., Lim, D. & Granville, D. Granzyme B in skin inflammation and disease. *Matrix Biol. J. Int. Soc. Matrix Biol.* 126–140. (2019).
19. Trapani, J. & Sutton, V. Granzyme B: pro-apoptotic, antiviral and antitumor functions. *Curr. Opin. Immunol.* **15**(5), 533–543 (2003).
20. Gleave, A. & Granville, D. Granzyme B in autoimmune skin disease. *Biomolecules.* **13**(2). (2023).
21. Tang, O. et al. PD1 expression in EGFRvIII-Directed CAR T cell infusion product for glioblastoma is associated with clinical response. *Front. Immunol.* **13**, 872756 (2022).
22. Graw, J. Genetics of crystallins: cataract and beyond. *Exp. Eye Res.* **88**(2), 173–189 (2009).
23. Hartwich, H. et al. Functional role of γ-Crystallin N in the auditory hindbrain. *PLoS ONE.* **11**(8), e0161140 (2016).
24. Zaragoza-Huesca, D. et al. Identification of Thrombosis-Related genes in patients with advanced gastric cancer: data from AGAMENON-SEOM registry. *Biomedicines.* **10**(1). (2022).
25. Caffarel, M. & Coleman, N. Oncostatin M receptor is a novel therapeutic target in cervical squamous cell carcinoma. *J. Pathol.* **232**(4), 386–390 (2014).
26. Guo, L. et al. Stat3-coordinated Lin-28-let-7-HMGA2 and miR-200-ZEB1 circuits initiate and maintain Oncostatin M-driven epithelial-mesenchymal transition. *Oncogene* **32**(45), 5272–5282 (2013).
27. Jahani-Asl, A. et al. Control of glioblastoma tumorigenesis by feed-forward cytokine signaling. *Nat. Neurosci.* **19**(6), 798–806 (2016).
28. Mohan, S., Bonni, A. & Jahani-Asl, A. Targeting OSMR in glioma stem cells. *Oncotarget* **8**(10), 16103–16104 (2017).
29. Lee, M. et al. Oncostatin M promotes mesenchymal stem cell-stimulated tumor growth through a paracrine mechanism involving Periostin and TGFBI. *Int. J. Biochem. Cell Biol.* **45**(8), 1869–1877 (2013).
30. Sharanek, A. et al. OSMR controls glioma stem cell respiration and confers resistance of glioblastoma to ionizing radiation. *Nat. Commun.* **11**(1), 4116 (2020).
31. Lee, B. et al. Heterocellular OSM-OSMR signalling reprograms fibroblasts to promote pancreatic cancer growth and metastasis. *Nat. Commun.* **12**(1), 7336 (2021).
32. Reisz-Porszasz, S. et al. Lower skeletal muscle mass in male Transgenic mice with muscle-specific overexpression of myostatin. *Am. J. Physiol. Endocrinol. Metab.* **285**(4), E876–E888 (2003).
33. Peng, H. et al. Growth and differentiation factor 15 regulates PD-L1 expression in glioblastoma. *Cancer Manage. Res.* **11**, 2653–2661 (2019).
34. Huang, G. et al. Integrative transcriptome analysis identified a BMP signaling pathway-regulated LncRNA AC068643.1 in IDH mutant and wild-type glioblastomas. *Oncol. Lett.* **20**(1), 75–84 (2020).
35. Tang, L. et al. MSTN is an important myokine for weight-bearing training to attenuate bone loss in ovariectomized rats. *J. Physiol. Biochem.* **78**(1), 61–72 (2022).
36. Barzegar Behrooz, A. et al. Wnt and PI3K/Akt/mTOR survival pathways as therapeutic targets in glioblastoma. *Int. J. Mol. Sci.* **23**(3). (2022).
37. Koch, M. et al. A novel marker of tissue junctions, collagen XXII. *J. Biol. Chem.* **279**(21), 22514–22521 (2004).
38. Charvet, B. et al. Knockdown of col22a1 gene in zebrafish induces a muscular dystrophy by disruption of the myotendinous junction. *Dev. (Cambridge England).* **140**(22), 4602–4613 (2013).

39. Misawa, K. et al. Prognostic value of type XXII and XXIV collagen mRNA expression in head and neck cancer patients. *Mol. Clin. Oncol.* **2**(2), 285–291 (2014).
40. Ton, Q. et al. COL22A1 collagen COL22A1 maintains vascular stability and mutations in are potentially associated with intracranial aneurysms. *Dis. Models Mech.* **11**(12) (2018).
41. Huang, M. & Luo, W. Engrailed homeobox 1 transcriptional regulation of COL22A1 inhibits nasopharyngeal carcinoma cell senescence through the G1/S phase arrest. *J. Cell. Mol. Med.* **26**(21), 5473–5485 (2022).
42. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science*. **374**(6574), abe6474 (2021).
43. Mirzaei, R., Sarkar, S. & Yong, V. T cell exhaustion in glioblastoma: intricacies of immune checkpoints. *Trends Immunol.* **38**(2), 104–115 (2017).
44. Sun, Z. et al. Profiles, distribution, and functions of gamma delta T cells in ocular surface homeostasis and diseases. *Front. Biosci. (Landmark edition)* **29**(4), 146. (2024).
45. Liu, J. & Li, G. Identification and validation of a risk signature based on extracellular matrix-related genes in gliomas. *Medicine* **100**(16), e25603 (2021).
46. Choi, H. et al. Human allogenic Γδ T cells kill patient-derived glioblastoma cells expressing high levels of DNAM-1 ligands. *Oncoimmunology* **11**(1), 2138152 (2022).
47. Ahmedna, T. et al. The role of Γδ T-Lymphocytes in glioblastoma: current trends and future directions. *Cancers*. **15**(24). (2023).
48. Kang, I., Kim, Y. & Lee, H. γδ T cells as a potential therapeutic agent for glioblastoma. *Front. Immunol.* **14**, 1273986 (2023).
49. Pezuk, J. et al. Inhibition of polo-like kinase 1 induces cell cycle arrest and sensitizes glioblastoma cells to ionizing radiation. *Cancer Biother. Radiopharm.* **28**(7), 516–522 (2013).
50. Li, X. et al. Dual Inhibition of Src and PLK1 regulate stemness and induce apoptosis through Notch1-SOX2 signaling in EGFRvIII positive glioma stem cells (GSCs). *Exp. Cell Res.* **396**(1), 112261 (2020).


## Author contributions

TJT and ZJX designed the project, XZX and LBH were responsible for data collection and collation, professors PJ gave guidance, LHW and LQC participated in the production of charts, ZWL and ZY collected clinical records, LXH and HXB wrote the manuscript.All authors contributed to revision of the manuscript and approved the final version.

## Funding

This research is supported by National Clinical Medical Research Center for Neurological diseases (GF2020001) and National Natural Science Foundation of China (82160512).

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95067-x.

**Correspondence** and requests for materials should be addressed to X.L. or X.H.