Exploring key genes and potential therapeutic and predictive targets between pyroptosis genes and adult extent of malignant glioma based on bioinformatics analysis

Zhengbo Yuan
Binzhou Medical University Affiliated Hospital

Chenglong Li
Binzhou Medical University Affiliated Hospital

Qingbo Wang
Binzhou Medical University Affiliated Hospital

Qi Wei
Binzhou Medical University Affiliated Hospital

Fengzhen Xiong
Binzhou Medical University Affiliated Hospital

Zefu Li (lizefu163@163.com)
Binzhou Medical University Affiliated Hospital

Research Article

Keywords: Low-grade glioma, Pyroptosis, Bioinformatics, Prognostic biomarkers, Immune infiltration, Tumor mutational burden, prognostic model

Posted Date: February 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1298963/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

**Background:** Adult glioma is the most common primary nervous system tumor, and there is a direct link between the grade of glioma and the prognosis of the patient. Furthermore, its treatment is still very challenging. Pyroptosis is another research direction that has made breakthroughs in recent years, and it has been confirmed that it plays an important role in a variety of tumors. We are currently focusing on analyzing the data and clinical data of a large number of pyroptosis genes, exploring the regulatory mechanism of tumor cell pyroptosis and the potential role of pyroptosis in cancer, in order to explore potential cancer diagnosis and treatment markers, for The pursuit of personalized or precise treatment provides new ideas and directions.

**Design:** Bioinformatics analysis of GEO database, TCGA database and GSEA data.

**Methodology:** Differential gene expression, Prognostic Network Construction, Construction of a prognostic model, Differential analysis of tumor microenvironment, immune cell expression, Tumor mutation burden correlation analysis, stem cell correlation analysis, drug sensitivity analysis

**Conclusions:** We have identified differential genes with high mutation frequencies in adult gliomas, as well as key genes associated with cell scorch and differentially expressed in different extent of malignant glioma, followed by theoretical validation and construction of prognostic models and assessment of high- and low-risk, and analysis of immune infiltration, tumor mutational load, tumor microenvironment, stem cell relevance, and drug sensitivity, revealing potential prognostic and predictive targets.

Introduction

Adult low-grade gliomas (LGGs) are a heterogeneous group of neuroepithelial tumors arising from supporting glial cells of the central nervous system (CNS). The World Health Organization (WHO) classifies gliomas into 4 grades based on their histopathological features, and only WHO grades I and II gliomas are considered low-grade[1,2]. According to the US CBTRUS statistical report[3]: Low-grade gliomas account for 6.4% of adult CNS primary tumors. There are differences in the incidence characteristics of race, gender and age[4]. Risk factors for low-grade gliomas and high-grade gliomas have not yet been definitively established, although certain inherited tumor susceptibility syndromes (neurofibromatosis type 1, Li-Fraumeni syndrome, Lynch syndrome) have been identified. sign) only a small part[5]

Low-grade gliomas currently remain a group of incurable diseases, and despite efforts to research and apply different treatments, these tumors still cause premature death. However, some interventions have been shown to improve median survival time[1]. The majority of untreated, symptomatic and incidentally discovered low-grade gliomas eventually undergo malignant transformation, leading to a more complex disease process, worse quality of life, and ultimately shortened patient survival[6,7].
Under currently accepted medical treatment guidelines, including surgical resection with the maximum safety margin, radiotherapy, temozolomide (TMZ) combined with maintenance chemotherapy, the median overall survival (OS) for patients with first-time diagnosis of glioma is only 12–18 months\(^{(8,9)}\). Overall, there is no absolute advantage of various monotherapy regimens, and the combination of multiple regimens will also bring new considerations\(^{(10)}\). In the future, further research at the gene level is expected to provide a deeper understanding of the occurrence and development of gliomas, and may provide new and more effective targets for the intervention and treatment of gliomas.

Douglas Hanahan and Robert A. Weinberg proposed six characteristics of cancer that are unique abilities that lead to cancer growth and metastasis. Resistance to cell death is one of its characteristics\(^{(11)}\). Cell death is not only a physiological regulation mechanism of cell proliferation, stress response and homeostasis, but also a specific research direction that we have long hoped to suppress tumors. In normal cells, cell death is a protective mechanism, and tumors can employ a variety of strategies to circumvent or limit cell death pathways\(^{(12)}\). Cell death is undoubtedly beneficial to our health in some special cases, such as cancer treatment\(^{(13)}\).

The three classical types of cell death are mainly: apoptosis\(^{(14)}\), autophagy\(^{(15)}\), necrosis\(^{(16)}\). The internationally accepted classification is mainly based on morphological features, and apoptosis is characterized by cell shrinkage and chromatin condensation\(^{(17)}\). In contrast, autophagic death is characterized by cytoplasmic vacuolation, phagocytic uptake, and lysosomal degradation\(^{(18)}\). Necrosis - manifested by cell swelling, loss of membrane integrity, DNA degradation, and release of cytoplasmic content\(^{(19)}\). Recently, researchers elucidated the detailed mechanism of programmed necrotic cell death, characterized by mitochondrial permeability transition (MPT)-dependent necrosis, ferroptosis, and pyroptosis\(^{(20)}\). Among these novel mechanisms, there is growing evidence that pyroptosis, a process of programmed inflammation-related cell death, plays a key role in various diseases of the central nervous system.

A recent literature based on experimental studies on the role of pyroptosis in glioma pointed out that miR–214 can regulate caspase–1 (caspase–1)-mediated glioma U87 and T98G cells. Pyroptosis inhibits cell proliferation and migration and may suggest a new therapy for intervention in glioma\(^{(21)}\).

In this paper, through the joint arrangement and analysis of transcriptome (expression), clinical, mutation, and copy number related data in the TCGA database and the transcriptome (expression) and clinical data in the GEO database, the aim of this paper is to analyze the effect of pyroptosis in low-grade glial cells. The progress in glioma (WHO grade II)\(^{(22)}\) is analyzed to provide new ideas and methods for clinical treatment and medical research, and provide more powerful data support for optimizing glioma treatment plan.

**Materials And Methods**
Acquisition of target data

The required glioma and pyroptosis-related datasets were retrieved on the NCBI platform and UCSC Xena, respectively. The datasets were obtained from the Cancer Genome Atlas Project (TCGA, https://portal.gdc.cancer.gov/repository), gene expression Comprehensive database (GEO, https://www.ncbi.nlm.nih.gov/geo/), UCSC Xena (https://xena.ucsc.edu/), and GSEA (https://www.gsea-msigdb.org/gsea/). The data selected for the download included low-grade glioma (WHO grade II) (TCGA-LGG) transcriptome analysis data in the TCGA database, clinical sample information, copy number variation data, and human glioblastoma with a sample size of 180 cases in the GEO database.

Further, the modeling data for malignant progression of gliomas (GSE184941), gene-level copy numbers of pyroptosis-related genes in glioma cells, and pan-cancer gene stem cell data. Non-human samples and expression profile combinations were excluded from the present analysis. Among them, the data in the TCGA database comes from the National Cancer Institute Cancer Genomics Center and the National Human Genome Research Institute funded by the U.S. government, collated and published shared data, including mRNA profiles of low-grade and high-grade human glioma samples, and clinical sample information among others.

The Cancer Genome Atlas Project database and Gene Expression Comprehensive Database were separately searched using the key search term "low-grade glioma", and the target datasets were obtained through category limitation and manual screening of the search results. A total of 529 target data that met the required conditions in the TCGA and GSE43378 in the GEO were separately downloaded, including a total of 180 high- and low-grade gliomas (90 cases each) clinical samples and mRNA profile data.

All raw data applied in this paper are from public databases, and all raw data are shared. The data collection and study is done in compliance with the Chinese or international norms and regulation. Any discrepancies in the obtained data were resolved through the participation or discussion with a third researcher. The present study was approved by the review board of Binzhou Medical University Hospital.

Differential analysis of pyroptosis-related genes

For screening of copy number variation data downloaded from The Cancer Genome Atlas Project database, mutation data analyzed using the varscan program was selected. Raw data of the dataset were processed using the "maftools" package in the R software (version 4.1.1.) to draw waterfall plots and visualize the expression frequencies of differential gene mutation.

The number of copy variations was calculated, and the file of the gene mutation was obtained using Strawberry Perl (version 5.34.0). Information about tumor mutational burden (TMB), the total number of somatic gene coding errors, base substitutions, and gene insertion or deletion errors detected per megabase, reflects the total number of mutations carried by tumor cells.
The raw data of the dataset can then be visualized using R software (version 4.1.1.) for the downloaded copy number of pyroptosis-related genes in glioma cells, resulting in a line graph of copy number variation frequency. To clarify the relationship between mutated genes and chromosomal locations as well as to visualize whether copy number is up- or down-regulated, the "RCircos" package in R software (version 4.1.1.) was used to visualize the chromosome-related copy number and obtain a circle graph.

**Establishment of Survival Analysis and Prognosis Network**

For the data files downloaded in the TCGA database and those downloaded in the GEO database the "limma" and "sva" packages\(^\text{(25)}\) in the R software (version 4.1.1.) were used to correct, merge, and then correct, and then merge the data files. The R software (version 4.1.1.) "limma" package was used to extract the expression of pyroptotic genes and after obtaining the expression of pyroptotic genes, perform univariate Cox analysis, and perform km analysis according to the findings of the univariate Cox analysis to compare the high and low expression.

Survival difference, screened out the related genes that met the condition of P value < 0.05 and divided them into two groups with HR = 1 as high and low expression. The high and low expression can be analyzed in combination with the clinical sample information downloaded from TCGA and GEO databases. The survival curve was drawn between the two groups to compare the difference in survival time between the two groups and hence to judge the relationship between the pyroptosis genes and prognosis of the patients. Similarly, for the outcome data of the univariate Cox analysis, the "igraph", "psych", "reshape2", "RCColorBrewer" packages\(^\text{(26)}\) in the R software (version 4.1.1.) were used to process its outcome data to obtain a prognostic network picture. Statistical significant difference was set at P value < 0.05.

**Pyroptosis genotyping and inter-type survival analysis**

According to the level of pyroptosis gene expression, use the "ConsensusClusterPlus" package\(^\text{(27)}\)in the R software (version 4.1.1.) was used to cluster the data (Kmax = 9), and then the internal correlation of the typing was manually selected after obtaining the typing result.

High, K value (K = 2) with significant difference between the types and then it was judged that it was most meaningful to divide the downloaded sample data into two groups according to the level of expression. For the two well-typed groups, a survival curve can be drawn to compare whether there is a difference in survival time between the types. Similarly, for the grouping of pyroptosis genes, the R software (version 4.1.1.) can also be used. The "pheatmap" visualizes the typing data combined with the clinical traits, so as to more intuitively reflect the differences in the level of pyroptosis-related genes between the two groups. Statistically significant difference was set at P value < 0.05.

**PCA analysis**

For the two groups of pyroptotic gene expression types, a linear dimensionality reduction algorithm variance was used to measure the difference of the data and projected the high-dimensional data with
large differences into a low-dimensional space for visualization of expression. This was to judge whether the two groups formed by this classification method had real differences. Specifically, data in this study were processed using the "ggplot2" package\(^{(28)}\) in R software (version 4.1.1.)

**Gene Set Variation Analysis**

For the data grouped based on the expression levels of pyroptotic genes, the "limma", "GSEABase", "GSVA", and "pheatmap" packages\(^{(29)}\) in the R software (version 4.1.1.) were used to analyze the expression levels of the two groups of data separately. The enrichment score values were then normalized and compared between the different samples to evaluate the enrichment of different metabolic pathways between the two groups of samples, and the results obtained were finally visualized.

**Differential analysis of immune cells**

There may be differences in the degree of immune cell infiltration between the two groups with different types. Furthermore, the infiltration of immune cells in tumor tissue is closely related to clinical outcomes. The immune cells infiltrated in tumors are most likely to be used as drug targets to improve patients and hence, the "ggpubr", "reshape2", "limma", "GSEABase", and "GSVA" packages\(^{(27,29)}\) in the R software (version 4.1.1.) were used to perform immune cell differential analysis between the two different samples as well as visualize the results obtained. Each immune cell was compared in two groups with different types and a P value < 0.05 was considered statistically significant.

**Difference analysis among pyroptosis genotypes**

Through differential analysis of genes between the types it was intended that the different genes between these two groups are known. The "limma" and "VennDiagram" packages\(^{(30)}\) in the R software (version 4.1.1.) were used to process the typed data and obtain the t value after the test. The corrected P.Val value can be obtained according to the t value, all genes can be filtered and screened according to |logFC| = 1 as well as P.Val < 0.05, find the significantly different genes that meet the screening conditions and draw a Venn diagram..

**GO enrichment analysis and KEGG analysis of differential genes between phenotypes**

The functions of the differential genes enriched by GO analysis can be observed for the previously screened differential genes. Clusterprofiler is used to analyze and visualize the critical data on GO along with KEGG analyses\(^{(31)}\). First, the "clusterProfiler", "org.Hs.eg.db", "enrichplot", and "ggplot2" packages\(^{(27,32)}\) in the R software (version 4.1.1.) were used to process and screen the previously selected differential genes. Finally, the screening results were visualized whereby the condition were P.Val < 0.05 and adj. P.Val < 0.05.

Similarly, the specific roles of the differential genes and their products in the cellular metabolic pathways can also be observed by performing KEGG analysis on the differential gene data. The "clusterProfiler", "org.Hs.eg.db", "enrichplot", and "ggplot2" packages\(^{(27)}\) in R software (version 4.1.1.) were used to process
and filter the data whereby the filter condition were P.Val < 0.05 and adj. P.Val < 0.05 and the screening results were hence finally visualized.

**Survival analysis and retyping of differential genes between types**

By performing survival analysis on differential genes between subtypes, it was hoped that differential genes associated with prognosis would be found. A univariate Cox analysis on the differential gene expression files and survival data (both from the results of the previous analysis) was performed using the "limma", "survival" packages\(^{33}\) in the R software (version 4.1.1.), for the genes obtained. Further, if the P.Val was < 0.05, it meant that there was statistical significant difference and the HR value was used to measure the prognostic risk, with HR = 1 as the threshold, and HR > 1 means high prognostic risk.

For the results of univariate Cox analysis, the "clusterProfiler" and "limma" packages\(^{27}\) in the R software (version 4.1.1.) was used to cluster the resulting data (Kmax = 9) and combine CDF curve after obtaining the typing results, then artificially select the K value (K = 3) with high internal correlation between the types, significant difference between types and large cumulative increase as the grouping basis and then determine that the downloaded sample data would be divided into three according to the difference of genes. Groups make the most sense.

**Survival analysis between retypes**

For data that had been retyped, the "survival" and "survminer" packages in the R software (version 4.1.1.) were used to perform survival analysis on the typed data. The P.Val < 0.05 between the groups indicated that the type is statistically significant meaning the processed data can be visually expressed.

**Clustering heatmap between retypes**

For the data file after re-typing, it was clustered with the clinical data and the typed data based on the expression of pyroptotic genes and then analyzed together, using the "pheatmap" package in the R software (version 4.1.1.), the data was processed, the results were visualized, and a cluster heatmap drawn.

**Screening of differential genes between retypes**

The "limma", "reshape2", and "ggpubr" packages in the R software (version 4.1.1.) were used to process the genotyping result data and the pyroptosis gene expression level file in each sample to screen for different Differential genes between groups as well as to visualize the expression of the screening results.

**Construction of a prognostic model**

We used the results of the previous one-way Cox analysis to randomize the data using the "survival", "caret", "glmnet", "survminer", "timeROC" packages\(^ {34}\) in the R software (version 4.1.1.) (randomly grouped once, divided into train group and test group), lasso regression model construction and cross-validation, the gene corresponding to the point with the smallest error was hence obtained whereas the expression
amount was extracted as the key gene for building the model to prepare for subsequent analysis. In addition, the Cox model formula was constructed, risk calculation was performed, then the area under the ROC curve was calculated, and the risk data of the train group, the test group and all samples were finally output. The "ggalluvial", "ggplot2", and "dplyr" packages in R software (version 4.1.1.) were used for the entire prognostic model building process to draw Sankey diagrams for a more intuitive display.

**Difference Analysis of Risk Score and Difference Analysis of Risk Gene**

The risk scores of patients were evaluated on whether they are different between different groups and then judged whether the classification is meaningful. Through the description of the previous steps, a total of two classifications were carried out to classify the two classifications. The results were separately analyzed for differences in risk scores and "limma" and "ggpubr" in R software (version 4.1.1.) were used to analyze the differences in risk scores. The P.Val < 0.05 in each group between the two classifications indicated that the classification was statistically significant. The analysis results were visualized and boxplots finally obtained.

Similarly; for different risk groups, meaningful differential genes, was to be obtained and hence the "limma", "reshape2", "ggplot2", and "ggpubr" packages in R software (version 4.1.1.) were used to focus. The list, expression, and risk data of the dead genes were processed and the obtained results were visualized and boxplots were also drawn.

**Prognostic Model Accuracy Verification**

In the process of constructing the prognostic model, the result data of the univariate Cox analysis were randomly divided into the train group and the test group. To verify the accuracy of the model built, whether it can distinguish patients in high and low risk groups, the data in the random group were divided into high and low risk groups in the train group according to the median value of their risk scores and in the test group.

The group was also divided into high and low risk groups as well as the difference analysis of survival was performed in the high and low risk groups whereas the survival difference between the high and low risk groups was compared. Survminer" package was used for processing and to obtain the P value of the difference in survival. If the P.Val < 0.05 between the high and low risk groups was significant, it was the real difference whereby the survival curve was drawn. To judge the accuracy of the prognostic model in predicting the survival time of patients, the risk data was used to draw the ROC curve. The larger the area under the ROC curve, the more accurate the prediction of the survival time of the prognosis of patient. The "survival", "survminer", and "timeROC" packages in R software (version 4.1.1.) were used to process the risk data and plot the curves.

Taking the 16th patient in the clinical data file as an example, the "survival", "regplot" and "rms" packages in the R software (version 4.1.1.) was use to process the patient's data and the patient's survival time was finally predicted and visualized. To display the risk score more intuitively the train group, the test group,
and all the data was sort according to their risk scores. They were then divided into two groups, high and low risk according to the median value of the risk score, and then R software (version 4.1.1.) in the "pheatmap" package used to draw risk curves, survival status maps, and risk heat maps for the sorted patients.

**Determination of immune infiltration with CIBERSORT**

To obtain the content of immune cells in each sample the expression data of all pyroptotic genes in the samples and CIBERSORT was analyzed through the immune cell infiltration algorithm\(^{(36)}\) in the R software (version 4.1.1). The experimental data of the immune cell chip was calculated, processed, and the final result was the relative content of immune cell infiltration in each sample. If the calculated P.Val of the sample is < 0.05, it is defined as the calculated immune cell infiltration. The relative content of the samples was highly accurate and met analytical requirements of the present study.

To determine the correlation between modeled key genes and immune cells, as well as the risk scores of patient with immune cells, "limma", "reshape2", "tidyverse", The "ggplot2", "ggpubr", and "ggExtra" packages\(^{(37)}\) was used to analyze and process the risk data of all samples as well as the relative content data of immune cell infiltration in each sample, and visualize the expression of the samples whose P.Val < 0.05 is calculated.

**Differential analysis of tumor microenvironment**

This study expected to indirectly judge the tumor purity through the stromal cell score, immune cell score, and comprehensive score, as well as compare the high and low risk groups to judge the differences between different groups. Specifically, the gene expression data files was processed and calculated through the "limma" and "estimate" packages in the R software (version 4.1.1.). Further, after obtaining the score data, combined with the risk data files of all samples, the "reshape2" and "ggpubr" packages\(^{(38)}\) in the R software (version 4.1.1.) were passed to analyze the data and draw violin plots to visualize the results.

**The relationship between gene mutation status and risk division**

The downloaded gene copy number variation data were grouped into high and low risk groups, and used the "maftools" package in R software (version 4.1.1.) to count the mutation frequencies of mutated genes in high and low risk groups. The top 20 genes were set with the highest mutation frequency, visualized, and expressed, whereas a waterfall chart of high and low risk groups was also drawn.

**Tumor mutation burden correlation analysis**

To determine whether tumor mutational burden differed in between high and low risk groups, the "ggpubr" in R software (version 4.1.1.) were used for data on patient risk scores, high and low risk groups, genotyping results and tumor mutational burden, "reshape2" package for processing, P.Val < 0.05
obtained after processing indicates that the difference is statistically significant and the results were visually expressed, as well as the boxplots and correlation scatter plots are drawn.

**Stem Cell Correlation Analysis**

To clarify the correlation between the stem cell content in the samples and different risk scores, the "limma", "ggplot2", "ggpubr", and "ggExtra" packages in R software (version 4.1.1.) were used to process the data. In the obtained results, it was found that the larger the stem cell index, the higher the content of stem cells in the sample. In addition, the P.Val < 0.05 indicated that the correlation between the two was statistically significant and the results were hence finally visualized.

**Drug Sensitivity Analysis**

Finally, the high- and low-risk group data and drug sensitivity data was processed and analyzed to identify which drugs have different sensitivities in the high- and low-risk groups and the "limma" "ggpubr" "pRRophetic", and "ggplot2" packages(39) were used. All data were processed, P.Val < 0.001 was used as the screening condition whereas the drugs with high sensitivity in high and low risk groups were screened, the results separately visualized as well as expressed, and boxplots were drawn.

**Results**

**Identification of differentially expressed genes linked to LGG**

We analyzed gene expression in low-grade glioma and pyroptosis-related datasets, then generated waterfall diagram of pyroptosis-related genes and their copy number variation. Frequency line chart, copy number and chromosome correlation circle chart, survival analysis of pyroptotic genes, and prognosis network chart (adjusted p < 0.05, |logFC| > 1 Fig. 1A-1E). Results revealed that TP53 had a very high mutation frequency (46%) across all samples. Notably, these mutations were highly diverse, and included missense, nonsense, frameshift and whole-frame types, of which missense mutations were the dominant type. At the same time, we found a significantly high proportion of cytosine, after mutation, (Fig. 1A). Copy number variation is an important component of genomic structural variation during tumor mutational burden (Fig. 1B). This was mainly manifested as submicroscopic level deletions and duplications, as well as an increase in GSDMC, and GSDMD copy numbers, but notably NLRP7. There was a higher loss of copy number for NLRP2, as was the case for CHMP2A, BAX and NLRP6. Among pyroptosis-related genes (Fig. 1C), we found a high copy number for ELANE on chromosome 19, GSDMD and GSDMC on chromosome 8 (also low-grade glioma-associated genes). Moreover, genes with deletions were distributed differently across autosomes 1 to 20, although the precise role of the deletion requires further exploration. Next, we generated survival curves targeting the 45 prognosis-related genes screened herein, in combination with previous findings. The results are presented in Fig. 1D. Summarily, GSDMD expression was significantly correlated with survival times of patients (p < 0.001), while its overexpression was significantly associated with low survival rates. NLRP7 (p = 0.001) and NLRP2 (p = 0.021) were also significantly associated with patient survival, with their upregulation having a negative effect on the survival times of patients. Moreover, CHMP2A (p < 0.001), BAX (p < 0.001) and upregulation
of NLRP6 (p < 0.001) all had a significant negative correlation with patient prognosis. A prognostic network revealed that most of the pyroptosis-related genes were indicators of high risk in patients, and there was a complex co-expression relationship among genes, especially GSDMD, BAX, CHMP2A (Fig. 1E). This strong prognostic value suggested that these genes could be potential targets, although further explorations are required to validate this hypothesis. NLRP2 has inhibitory effects on 4 genes that are related to pyroptosis and are high risk, and it is particularly noteworthy that its effect on TP53 the presence of inhibition.

**Differential analysis of pyroptosis genotyping**

According to the level of the expression of pyroptosis genes, we first typed the sample data and performed differential analysis of the typing results. For the typing process and the data analysis results after typing, we used visual expression, and the drawn cells Pyroptosis genotyping map, PCA analysis map, survival analysis curve between types, cluster heat map of pyroptosis type, boxplot of immune cell difference analysis between types, Venn diagram of differential genes, GSVA analysis result graph, GO enrichment analysis result box Box and Bubble Plots, Box Plots and Bubble Plots of KEGG Analysis Results. (Fig. 2A-2H)

First of all, for the results of typing, when K = 2, the intra-type correlation is high and the differences between types are obvious, so we divided all samples into two groups, in order to verify the feasibility of typing, we performed PCA Analysis (Fig. 2A), in order to prove that the classification is feasible; and then we draw the survival curve between the classifications (Fig. 2B), which can intuitively reflect that there is a significant difference in survival time between groups A and B (p < 0.001), The survival of type A samples was significantly better than that of type B samples. Similarly, after we combined the clinical information of the samples and drew a cluster heat map (Fig. 2C), we found that most of the pyroptotic genes were highly expressed in type B, especially GSDMD, BAX, CHMP2A and TP53 that we analyzed earlier The target genes are also actively expressed and may be associated with high-grade gliomas.

Next, from the visualization results of immune cell differential analysis (Fig. 2D), we can find that, except for Monocytena, the vast majority of immune cells are significantly different between types (p < 0.001) and except for CD56dim.natural.killer. In addition to cellna, the remaining immune cells with significant differences are actively expressed in the B family with poor prognosis, and we infer that the massive infiltration of immune cells has a negative effect on the prognosis of low-grade glioma patients with active pyroptotic gene expression. A total of 818 differential genes were screened this time in Veen diagram (Fig. 2E), the results of GSVA analysis (Fig. 2F) indicated that in type B, most low-grade glioma-related metabolic pathways were actively expressed, which more reflected the accuracy of the classification and the reliability of the analysis results after classification. After I got the typed differential genes, I performed GO analysis and KEGG analysis to observe the results (Fig. 2G, Fig. 2H),

It is not difficult to find that: through GO analysis, differential genes are involved in neutrophil activation, neutrophil activation involved in immune response, neutrophil mediated immunity, neutrophil degranulation, these functions account for a large proportion, and the enrichment is significantly high; while in calcium-dependent protein binding, peptide Antigen binding, Toll-like receptor binding, and
syntaxin-1 binding account for a very small functional proportion, the enrichment is not significant, and there is no potential value for further analysis. Moreover, the number of differential genes enriched in the disease-related pathways of Tuberculosis, Phagosome, Epstein-Barr virus infection, Human T-cell leukemia virus 1 infection, Coronavirus disease - COVID-19 by KEGG analysis; Tuberculosis, Phagosome, Epstein-Barr virus infection, Human T-cell leukemia virus 1 infection, Coronavirus disease-COVID-19 disease-related pathways have a high proportion, a large number of enrichment, and a significant degree of enrichment; but in Cytokine-cytokine The enrichment degree of receptor interaction pathway is very insignificant and has no research value.

**Construction of a pyroptosis-related-gene prognostic model**

Next, we built a model based on pyroptosis-related genes for prognosis of glioma patients. We generated a Sankey diagram (Fig. 3I), then generated survival curves an a heat map for the clinical data. We used the median score to stratify the samples into high- and low-risk groups, then validated accuracy of the prognostic model using area under curve (AUC) of the receiver operating characteristic (ROC) (Fig. 3A-H). The selection method for re-typing was the same as before, except that 3 groups (K = 3) were used this time. Differential gene expression analysis for the 3 groups revealed 44 targets, two of which were significantly different (p < 0.001) (Fig. 3A). Among them, BAX, GSDMD, CHMP2A, and TP53 were upregulated and downregulated in group B and A, respectively, based on our earlier results. This trend further confirmed that GSDMD, BAX, CHMP2A and TP53 are indeed associated with poor patient prognosis. In addition, NLRP2 was upregulated in group A than in group C and similar group B. Positive effect of this gene on prognosis. Survival analysis results revealed significant differences among the 3 groups (p < 0.001), with group A exhibiting the longest survival times, group B showing poor survival times, and group C in between them (Fig. 3B). Further analysis indicated that B group, which had the shortest survival times, had significant differential gene expression across samples (Fig. 3C). This may explain the differences in survival times between groups.

Next, we performed risk score analysis on the constructed prognostic model and found statistically significant differences in the first classification (p < 0.001) (Fig. 3D). Notably, group B had higher risk scores; the second classification had a higher risk score; Risk scores were also different (p < 0.001), with group B having the highest risk score and group A having the lowest risk score. Differential gene expression analysis revealed 42 targets, of which BAX, CHMP2A, GSDMD, and TP53 were significantly upregulated in the high-risk group (Fig. 3E). There were significant differences between the high- and low-risk groups (p < 0.001), affirming the feasibility of our model. Next, we randomly divided the samples into train and test sets and generated survival curves (Fig. 3F). Analysis of risk scores revealed significant differences in survival times between groups (p < 0.001), with results of the survival time response found to be consistent with those of survival times in all samples, further affirming feasibility of the model. Furthermore, we obtained an AUC greater than 0.65 for all samples (Fig. 3G), indicating that our model had high accuracy. Overall, overexpression of TGIF1, GDF15, PDLIM4, and TOP2A genes, as well as downregulation of CRTAC1, were associated with high disease risk (Fig. 3H). This suggests that TGIF1,
GDF15, PDLIM4, and TOP2A genes are predictors of high-risk prognosis while CRTAC1 has potential to improve prognosis and survival of patients.

**Correlation gene expression and immune cell infiltration**

Correlating gene expression with immune cell infiltration revealed no significant association on prognostic risk genes TGIF1 and TOP2A (Fig. 4A). GDF15 expression had a strong positive correlation with Macrophages M0, Macrophages M1, T cells CD8, T cells follicular helper and T cells gamma delta (p < 0.001) but a significant association with Monocytes. Moreover, PDLIM4 was strongly correlated with Macrophages M0, Macrophages M1, T cells CD8, and T cells gamma delta (p < 0.001) but negatively with Monocytes and T cells CD4 memory resting. Finally, CRTAC1, a key gene associated with patient prognosis, had a positive and negative correlation with Mast cells activated and T cells CD8, respectively.

**Differential analysis of tumor microenvironment**

According to the difference analysis results of tumor microenvironment scores, Differentially expressed genes were significantly associated with stromal cell score, immune cell score and comprehensive score between high and low risk groups (p < 0.001) (Fig. 5A). Proportions of total immune cells were all higher in the high-, relative to the low-risk group.

**Differential Analysis of Gene Mutations**

The low-risk group exhibited higher gene mutation frequencies than the high-risk group, with IDH1, TP53 and ATRX having the highest frequencies (Fig. 6A).

**Tumor mutation burden**

Analysis of tumor mutation burden revealed significant differences between patients in the high- and low-risk groups (p < 0.001) (Fig. 7A). Notably, risk scores were positively correlated with tumor mutational burden (R = 0.42, p < 2.2e-16).

**Stem Cell Correlation Analysis**

Patient's risk scores were negatively correlated with stem cells (R = -0.36, p < 2.2e-16) (Fig. 8A).

**Drug Sensitivity Analysis**

Analysis of drug sensitivity, targeting 103 drugs that met our screening criteria, revealed that 86 drugs had different degrees of sensitivity to the high-risk group (Table 1).
Table 1

<table>
<thead>
<tr>
<th>Sensitive drugs in the high-risk group</th>
</tr>
</thead>
</table>

Discussion

In 2016, the World Health Organization (WHO) updated genotypic and expression data of gliomas with the aim of improving accuracy of clinical, experimental and epidemiological studies. This new classification now includes information on the genetic basis of tumorigenesis, adding predictive data for more effective targeted therapy\(^{(40)}\). Prognostic and predictive markers play an important role in clinical practice, particularly during determination of patient prognosis and selection of appropriate therapies. This is imperative especially for disease such as glioma. In the present study, we integrated multiple gene expression datasets from 2 different databases and analyzes them using bioinformatics and R software. Our results revealed that TP53 has a very high and diverse mutation frequency in all low-grade glioma samples, including missense, nonsense, frameshift and whole-frame mutation types. Moreover, we found that cytosine was significantly increased, and the increase is statistically significant; the somatic mutation of the TP53 gene is one of the most common changes in human cancer, and most mutations are single mutations distributed throughout the coding sequence. base substitution. Their different types and locations may inform the nature of the mutagenic mechanisms underlying cancer etiology, since TP53 mutations have been shown to be potential prognostic and predictive markers, as well as target genes for drug intervention\(^{(41)}\). Based on our results, it is evident that mutations in this gene plays an important role in low-grade glioma. Our results also revealed an increase in the copy number of the pyroptosis-related gene GSDMD, as well as deletion of CHMP2A, BAX, NLRP6 and NLRP2 in low-grade glioma cells. Moreover, chromosome number and position corresponded to the increase or deletion. Correspondence, and in the subsequent series of sample typing, differential analysis and prognostic model construction results confirmed the relationship between GSDMD, CHMP2A, BAX on the adverse effects of low-grade glioma patients and may be related to high-grade glioma There is a relationship between them. Recent studies have shown that GSDMD is an effector of pyroptosis\(^{(42)}\). Moreover, some MicroRNAs (miRNAs) have been implicated in pyroptosis of glioma cells. For example, miRNA-214 was found to inhibit proliferation and migration of glioma cells by directly targeting caspase-1. Moreover, Caspase-1 induced cells by cleaving GSDMD Pyroptosis occurs, leading to the formation of plasma membrane pores\(^{(43)}\).
GO terms indicated that the differentially expressed genes were involved in neutrophil activation, and immune response, as well as neutrophil mediated immunity, and neutrophil degranulation. Furthermore, KEGG pathway analysis indicated that these genes were enriched in Tuberculosis, Phagosome, Epstein-Barr virus infection, Human T-cell leukemia virus 1 infection, Coronavirus disease-COVID-19 disease-related pathways. The specific functional mechanisms of target genes that play an important role in the development and progression of low-grade gliomas provide potential directions. Next, we constructed a prognostic model comprising BAX, CHMP2A, GSDMD, and TP53, genes as well as high-risk factors, namely TGIF1, GDF15, PDLIM4, TOP2A, and CRTAC1. Conclusions on genes critical for survival

Correlation between gene expression and immune cell infiltration revealed a positive relationship between GDF15 and Macrophages M0, Macrophages M1, T cells CD8, T cells follicular helper and T cells gamma delta, but a negative correlation with Monocytes. In addition, PDLIM4 expression was also positively correlated with Macrophages M0, Macrophages M1, T cells CD8, T cells gamma delta, and Monocytes, but negatively with T cells CD4 memory resting. Moreover, CRTAC1, a key gene related to patient prognosis, had a strong positive and negative correlation with Mast cells activated and T cells CD8, respectively. Previous studies have revealed the relationship between the tumor immune microenvironment and efficacy of immunotherapy. Analysis of the tumor immune microenvironment revealed that the total amount of stromal cells and immune cells were higher in the high-risk group. The tumor mutational burden (TMB) is an emerging biomarker that has received numerous attention due to its role in predicting efficacy of tumor immunotherapy. Previous studies have shown that to a certain extent, TMB levels reflect DNA repair damage in tumor cells and are closely related to the ability to generate tumor neoantigens. In the present study, we analyzed tumor mutational burden between high and low-risk groups and found that tumor mutational burden also increased as risk scores increased.

Previous studies have shown that in many solid tumors, only a handful of cancer cells have the ability to maintain tumor formation, with most of them shown to possess limited ability to proliferate. Therefore, in-depth study of tumor stem cells and the search for corresponding targeted drugs are of great significance to the clinical treatment of cancer. Results of our stem cell correlation analysis revealed a significant negative correlation between patient risk scores with stem cell index indicating that cancer stem cells have potential significance in predicting prognosis of patients with low-grade glioma. Drug sensitivity analysis plays a direct role in combination with clinical practice. Results of the present study revealed that a total of 86 drugs exhibited different degrees of sensitivity to high-risk groups, which is crucial for the development of clinical treatment. Prompting.

Although our study provides new insights to guide future research seeking to combine pyroptosis-related genes with low-grade glioma treatment and prognosis, it still had some limitations. Firstly, the reliability of our molecular mechanism results is somewhat limited by lack of in vitro or in vivo experimental data. Secondly, we did not have extra strict controls on data collection, due to the fact that we entirely used data from public databases, whose quality is uncertain.

Conclusions
In summary, we identified differentially expressed pyroptosis-related genes with high mutation frequencies in adult low-grade gliomas. We theoretically verified their functions by constructing a prognostic model. Analysis of the relationship between gene expression with immune infiltration, tumor mutation burden, tumor microenvironment, stem cells, and drug sensitivity, revealed the potential prognostic value of these targets, although this needs to be experimentally validated. Future explorations of the molecular signatures, based on genomics and transcriptomics, are expected to guide clinicians on how to rationally select drugs and provide actionable targets for the development of treatment plans for each patient.

**Declarations**

**Acknowledgments:**

The authors acknowledge assistance from the Neurosurgery Research Group members (Department of Neurosurgery, Binzhou Medical University Hospital, Binzhou, Shandong, China).

**Data deposition information item**

TCGA: https://portal.gdc.cancer.gov/repository


UCSC Xena: https://xena.ucsc.edu/

GSEA: https://www.gsea-msigdb.org/gsea/

**Data availability statement**


All raw data applied in this paper are from public databases, and all raw data are shared. The data collection and study is done in compliance with the Chinese or international norms and regulation. Any discrepancies in the obtained data were resolved through the participation or discussion with a third researcher. The present study was approved by the review board of Binzhou Medical University Hospital.

**References**


chemistry, 90, 107358. https://doi.org/10.1016/j.compbiolchem.2020.107358


**Figures**

**Figure 1**

A: Differential gene waterfall chart

The abscissa represents the sample, and the ordinate represents the genes related to pyroptosis; the different color modules in the waterfall diagram represent different mutation types, the color modules below the waterfall diagram represent the types of base changes in the sample, and the legend below is the specific corresponding relationship.

B: Copy number variation frequency line graph

The abscissa represents pyroptosis-related genes, and the ordinate represents the frequency of copy number variation. The red dots in the figure represent the frequency of copy number gain, and the green dots represent the frequency of copy number deletion.

C: Circle diagram of the relationship between mutant genes and chromosomal locations

The outermost circle of the circle represents the chromosome, the genes marked in the inner circle are pyroptosis-related genes, and the red color represents the higher frequency of the gene copy number increase; the blue represents the higher frequency of the gene copy number deletion.

D: differential gene survival curve

This is the survival curve of the 6 genes related to prognosis (P-value<0.05) that we screened, the abscissa represents the survival time, the unit is years; the ordinate represents the survival rate, and the survival rate of patients decreases with the passage of time, the P value in the figure represents whether there is a difference between the two groups with high and low expression (P-value<0.05).
E: The establishment of the prognosis network

A node in the figure represents a gene, the left semicircle of the node represents the gene attribute; the right semicircle represents the risk of the gene, purple represents high risk, green represents low risk, the size of the node represents the correlation between the gene and prognosis, and the area of the node The larger the value, the higher the correlation between the gene and the patient's prognosis; the connection between the dots represents the co-expression relationship between the two genes, the red line represents a positive correlation, and the blue line represents a negative correlation.

Figure 2

A: The first typing and verification

The top of the figure is the result of PCA analysis, the blue dots represent the samples of type A, and the orange dots represent the samples of type B. The middle part is the cluster diagram of typing, the blue squares in the lower left represent a group of typing, the typing in the upper right represents a group, and the bottom is the CDF curve of typing.

B: Survival curves between types

The abscissa of the survival curve represents the survival time, the unit is years; the ordinate represents the survival rate, and the survival rate of patients decreases with the passage of time, and the P value in the figure reflects the difference between the two groups.

C: Heatmap between types

The samples represented by the abscissa, the Horizontal axis shows clustering information of samples which were divided into two major clusters. The ordinates represent the pyroptotic genes, blue in the figure represents low expression, and red represents high expression.

D: Differential analysis of immune cells

The horizontal axis represents the name of immune cells, the vertical axis represents the score of immune cells, the asterisk above represents the degree of difference, three stars represent P-value<0.001, two stars represent P-value<0.01, and one star represents P-value value<0.05.

E: Venn diagram of differential genes 818 differential genes.

F: GSVA analysis

The abscissa and ordinate represent samples, and the ordinate represents the name of low-grade glioma-related metabolic pathways; blue color modules represent low expression, and red color modules
represent high expression.

G: GO analysis

In the top bar chart, the abscissa represents the number of genes enriched in the GO pathway, and the ordinate represents the name of the GO pathway; in the lower bubble chart, the abscissa represents the proportion of genes in the pathway, and the size of the bubble represents the gene. The number of genes in the pathway. In the whole figure, red represents a significant degree of enrichment, and blue represents a low degree of enrichment.

H: KEGG analysis

In the top bar chart, the abscissa represents the number of genes enriched in the KEGG pathway, and the ordinate represents the name of the KEGG pathway; in the lower bubble chart, the abscissa represents the proportion of genes in the pathway, and the size of the bubble represents the gene. The number of genes in the pathway. In the whole figure, red represents a significant degree of enrichment, and blue represents a low degree of enrichment.

**Figure 3**

A: Boxplots of expression between retypes and differential gene retypes

The top of the figure is the re-typed CDF curve, the middle part is the re-typed cluster diagram, and the three blue squares represent different groups.

B: Survival curves between retypes

The abscissa of the survival curve represents the survival time, the unit is years; the ordinate represents the survival rate, and the survival rate of patients decreases with the passage of time, and the P value in the figure reflects the difference between the three groups.

C: Heatmap between types 2

The samples represented by the abscissa, the Horizontal axis shows clustering information of samples which were divided into three major clusters. The ordinates represent the pyroptotic genes, in the figure blue represents low expression, and red represents high expression.

D: Box Plot for Risk Score Difference Analysis

The upper graph is the risk score difference of the first classification, the blue is the A group, the orange is the B group, the P value response difference between the two; the lower graph is the risk difference score of the second classification, blue is group A, orange is group B, and red is group C.
E: Box Plot for Risk Difference Analysis

The abscissa represents the genes related to pyroptosis, the ordinate represents the expression level of the genes, the red represents the high-risk group, the blue represents the low-risk group, the asterisk above represents the degree of difference, three stars represent P-value<0.001, two A star represents P-value < 0.01, and one star represents P-value < 0.05.

F: Survival curves between all samples and random groups

The figure represents the survival curves of all samples and random groups (train group and test group) respectively. The abscissa represents the survival time, in years; the ordinate represents the survival rate. The survival rate of patients decreases with the passage of time. The P value in represents whether there is a difference between the high and low expression groups (P-value<0.05).

G: ROC curve

The figure shows the ROC curve of all samples and random groups from top to bottom. The horizontal axis represents the false positive rate (represented by 1-specificity), the vertical axis represents the true positive rate (represented by sensitivity), and the green represents the ROC of 1 year. Curve, blue represents the 3-year ROC curve, red represents the 5-year ROC curve; the area under the ROC curve is between 0.5 and 1.

H: risk curve

From left to right are the risk curves of all samples, the train group and the test group; the top of the risk curve of each group is the curve graph of the sample risk score, and the patients are arranged according to the risk score from low to high, with the median of the risk score. The number is the boundary, the low-risk group is represented in blue, and the high-risk group is represented in red; the ordinate of the middle scatterplot represents the survival status, the blue represents the surviving patients, and the red represents the dead patients; the bottom risk heat map In the middle, the ordinate is the key gene for modeling, blue represents low expression, and red represents high expression.

I: Sankey diagram of the typing process

It reflects the whole process of two classifications, the process of risk grouping and the relationship between survival status.

Figure 4

A: Immune cell correlation of key genes
The abscissa represents the genes involved in model construction, the ordinate represents the name of immune cells, red represents positive correlation, blue represents negative correlation, and the number of asterisks represents the strength of the correlation.

**Figure 5**

A: TME Difference Analysis

The horizontal axis represents the type of tumor microenvironment score, the vertical axis represents the tumor microenvironment score, blue represents the low-risk group, red represents the high-risk group, and the asterisk above represents the degree of difference.

**Figure 6**

From top to bottom are the waterfall charts of pyroptosis-related genes in the low-risk group and high-risk group, respectively. The abscissa represents the sample, and the ordinate represents the pyroptosis-related genes. Different color modules in the waterfall chart represent different mutation types.

**Figure 7**

Tumor mutation burden correlation analysis

The top is the boxplot of the difference in tumor mutation between the high and low risk groups, the abscissa represents the high and low risk groups, and the ordinate represents the tumor mutation burden; the bottom is the scatter plot of the difference between the risk score and the tumor mutation burden, the abscissa Represents the patient's risk score, the ordinate represents the patient's tumor mutation burden, and the color of the dots in the figure represents the results of the typing.

**Figure 8**

Stem Cell Correlation Analysis

The abscissa represents the patient risk score, and the ordinate represents the stem cell index (RNAss). The larger the stem cell index, the higher the content of stem cells in the sample.

**Supplementary Files**
This is a list of supplementary files associated with this preprint. Click to download.

- supplementfilm.docx