Identification of IFN-β Associated Genes Signature Predicting Overall Survival for Glioblastoma

Cheng Lijing
The First Affiliated Hospital of Dali university

Yuan Meiling
The First Affiliated Hospital of Dali University

Li Shu
The First Affiliated Hospital of Dalian Medical University

Chen Junjing
Jiangxi Cancer Hospital

Zhong Shupeng
Zhongshan City People's Hospital

zhou jian (✉ zjsmu362324@126.com)
Zhujiang Hospital  https://orcid.org/0000-0001-5757-0010

Primary research

Keywords: TCGA, GEO, CGGA, IFN-β, glioblastoma, a prognostic model

Posted Date: March 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-212352/v1

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

**Background:** Brain glioblastoma (GBM) is the most common primary malignant tumor of intracranial tumors. The prognosis of this disease is extremely poor. While the introduction of IFN-β regimen in the treatment of gliomas has significantly improved the outcome of patients, the underlying mechanism remains to be elucidated.

**Materials and methods:** mRNA expression profiles and clinicopathological data were downloaded from TCGA-GBM and GSE83300 data set from the GEO. Univariate Cox regression analysis and lasso Cox regression model established a novel four-gene IFN-β signature (including PRDX1, SEC61B, XRCC5, and BCL2L2) for GBM prognosis prediction. Further, GBM samples (n=50) and normal brain tissues (n=50) were then used for real-time polymerase chain reaction (PCR) experiments. Gene Set Enrichment Analyses (GSEA) was performed to further understand the underlying molecular mechanisms. Pearson correlation was applied to calculate the correlation between the lncRNAs and IFN-β associated genes. A lncRNA with a correlation coefficient $|R^2| > 0.3$ and $P < 0.05$ was considered to be an IFN-β associated lncRNA.

**Results:** Patients in the high-risk group shown significantly poorer survival than patients in the low-risk group. The signature was found to be an independent prognostic factor for GBM survival. Furthermore, GSEA revealed several significantly enriched pathways, which might help explain the underlying mechanisms. Our study identified a novel robust four-gene IFN-β signature for GBM prognosis prediction. The signature might contain potential biomarkers for metabolic therapy and treatment response prediction in GBM.

**Conclusions:** Our study established a novel IFN-β associated genes signature to predict overall survival of GBM, which may help in clinical decision making for individual treatment.

Introduction

Brain glioblastoma (GBM) is the most common primary malignant tumor of intracranial tumors, accounting for ~50% of intracranial tumors with high morbidity and mortality in adults and children. Currently, surgery, radiation, and chemotherapy are the mainstays of GBM management. Chemotherapy is a critical process in the postsurgical treatment of GBM$^{1-3}$. Alkylating agents, such as temozolomide, remain the standard-of-care in GBM chemotherapy, but responses remain very poor$^{4}$.

The DNA repair protein O6-methylguanine-DNA methyltransferase (MGMT) plays an essential role in cellular resistance to alkylating agents$^5$. IFN-β can act as a drug sensitizer, enhancing toxicity against various neoplasias, and is widely used in combination with other antitumor agents such as nitrosoureas$^{6-8}$. IFN-β sensitizes glioma cells that harbor the unmethylated MGMT promoter and are resistant to temozolomide$^{6,9,10}$. Nevertheless, the specific mechanisms and molecules associated with this phenomenon have not yet been completely elucidated, and further research is needed.
In this study, we explored and analyzed differentially expressed IFN-β associated genes (GSEA-M2567) by systematic bioinformatics analysis. Five hundred and ninety-six were included in TCGA-GBM to train the prognostic model. Univariate Cox regression model found 5 survival-related genes, then Lasso-penalized Cox analysis identified 5 genes to construct the prognostic model. Using the methodology previously described, the proposed approach is validated on the GEO datasets (GSE83300). We found that 4 IFN-β associated genes (PRDX1, SEC61B, XRCC5, and BCL2L2) signature was a robust marker of seizure prognosis in patients with GBM. With CGGA data, we validated that 4 IFN-β associated genes were an independent biomarker of prognosis. Pathway enrichment analysis results show that several modules are enriched in pathways related to GBM.

LncRNA has been demonstrated that played an important role in human disease\textsuperscript{11}, especially in the GBM. IncRNA AC003092.1 regulates TFPI2 expression through ceRNA mechanism, and LncRNA SOX2OT interacts with RNA binding proteins to promote the expression level of SOX2 is involved in the resistance of glioma chemotherapy\textsuperscript{12,13}. However, the relevance between IFN-β and IncRNA has not been fully elucidated in GBM. Pearson correlation was applied to calculate the correlation between the IncRNAs and IFN-β associated genes. A IncRNA with a correlation coefficient $|R^2| > 0.3$ and $P < 0.05$ was considered to be an IFN-β associated IncRNA. Univariate and multivariate Cox regression were used for survival analysis indicated that AC093278.2, AC004067.1, LINC01116, and AC017104.1 was an independent prognostic factor for GBM patients.

In a word, systems-based biomarker discovery approaches may more accurately reflect the underlying biology, and this laid the foundation for our research on GBM. The present study may lay the foundations for future studies aimed to investigate GBM.

**Methods And Materials**

**Clinical specimens and Data collection**

All the 50 human glioma tissue samples and 50 normal brain tissues were obtained from surgery at Jiangxi Cancer Hospital of Nanchang University, China. Samples were then frozen in liquid nitrogen after surgical resection. The study was approved by the Ethics Committee of the Jiangxi Cancer Hospital of Nanchang University with informed consent from the patients or guardians.


**Identification of differentially expressed gene in TCGA-GBM**
Limma package was used to screen the interestingly differentially expressed genes with R 4.0.2. The expression pattern of the 120 IFN-β associated genes mentioned above was then investigated in TCGA. Genes were selected as consistently altered IFN-β associated genes for subsequent prognostic analysis if: (a) they showed consistent expression pattern in TCGA cohort; and (b) they were listed in the GSE83300 data set.

**Construction of the prognostic IFN-β associated gene signature**

Univariate Cox regression analysis and lasso-penalized Cox regression analysis were performed to identify the prognosis-related IFN-β associated genes and construct the prognostic gene signature. $P < 0.05$ in univariate Cox regression analysis were considered statistically significant. The prognostic gene signature was shown as risk score = $(\text{CoefficientmRNA}_1 \times \text{expression of mRNA}_1) + (\text{CoefficientmRNA}_2 \times \text{expression of mRNA}_2) + \cdots + (\text{CoefficientmRNA}_n \times \text{expression of mRNA}_n)$. The R package “survival” and “survminer” was used to explore the optimal cut-off of risk score and drawn the Kaplan–Meier survival curve. In particular, the “surv_cutpoint” function of the “survminer” R package was used to determine the optimal cut-off value to divide patients into high- and low-risk group. The R package “survivalROC” was used to investigate the time-dependent prognostic value of the gene signature. A two-sided log-rank $p < 0.05$ were considered significant for survival analysis.

**IFN-β associated IncRNA screening**

The profiles of IncRNAs and IFN-β associated genes were obtained from the TCGA RNAseq dataset. Pearson correlation was applied to calculate the correlation between the IncRNAs and differential genes. A IncRNA with a correlation coefficient $|R^2| > 0.3$ and $P < 0.05$ was considered to be IFN-β associated IncRNA.

**Construction of the prognostic IFN-β associated IncRNA signature**

Construction of the prognostic IFN-β associated IncRNA signature was performed as previously described.

**Gene set enrichment analysis (GSEA)**

GSEA was applied to investigate potential mechanisms underlying the influence of differential gene expression on GBM prognosis. GSEA was also applied to detect whether a priori defined set of genes showed statistically significant differential expression between the high and low-risk groups. Gene sets with a standard $P < 0.05$ were considered to be significantly enriched.

**Immunohistochemistry staining**

The paraffin-embedded glioma tissues were cut into thin slices and then installed on glass slides for immunohistochemistry experiments. The specimens were bred with rabbit anti-BCL2L2, anti-PRDX1, anti-XRCC5, and SEC61B antibody (1:200 dilutions, Abcam, USA) at 4 °C overnight, followed by 1 h incubation
of biocatalyst secondary antibody (1:200 dilutions, Santa Cruz Biotechnology, USA) under room condition. Then, the avidin biocatalyst peroxidase complex methods were adopted to determine the target protein's location and relative expression to visualize the bound antibodies.

**Result**

**Construction and validation of the prognostic IFN-β associated genes signature**

Five hundred and ninety-six and 120 IFN-β associated genes (70 upregulated and 51 downregulated) were included in TCGA-GBM to train the prognostic model. Differential gene expression analysis identified 14 downregulated and 53 upregulated IFN-β associated genes, respectively (Supplementary Table 1). In this part, Firstly, we displayed the heatmap by differential genes and analyzed these significant genes further (Figure.1A). Univariate Cox regression model found 5 survival-related genes (Supplementary Table 2), then Lasso-penalized Cox analysis identified 5 genes to construct the prognostic model (Supplementary Table 3). Using the methodology previously described, the proposed approach is validated on the GEO datasets (GSE83300) (Supplementary Table 4). Patients were divided into a high- and low-risk group for risk score. GBM patients with high-risk scores indicated poor prognosis (Figure.1B-C). Moreover, increased expression of the 4 different signature genes (PRDX1, SEC61B, XRCC5, and TXN) and reduced expression of the 1 signature genes (BCL2L2) was observed as the risk value increased (Figure.1D-E). Taking all of our results together, four genes were found to be correlated with unfavorable clinical outcomes.

**Prognostic significance of the four signature genes expression in GBM**

To further validate the expression of the prognostic genes constructing the gene signature, Kaplan-Meier survival analysis was used. Results are presented that the high expression of the SEC61B and XRCC5 is associated with poor prognosis in the GEO dataset (Figure.2A-B). Moreover, the high expression of the SEC61B, XRCC5, and PRDX1 is associated with poor prognosis, and the low expression of the BCL2L2 is associated with poor prognosis in the TCGA dataset (Figure.2C-F). To further verify whether the expression of SEC61B, XRCC5, BCL2L2 and PRDX1 was associated with prognosis in GBM, the GEPIA database (https://gepia.cancer-pku.cn/) was used. The expression of SEC61B, XRCC5, and PRDX1 have significantly high expression in tumor samples compared to normal samples, and the expression of BCL2L2 has significantly low expression in tumor samples compared to normal samples (Figure.2G). Further, GBM samples (n=50) and normal brain tissues (n=50) were then used for real-time polymerase chain reaction (PCR) experiments. The results were united with the GEPIA database (Figure.2H). Taken together, four signature-genes expressions are considered to be of clinical significance in GBM.

**Validation of the four signature genes in the CGGA database**

To further validate these results, we used the Chinese Glioma Genome Atlas (CGGA) database (http://www.cgga.org.cn/). Kaplan-Meier survival analysis of CGGA data set shows that high expression of SEC61B, XRCC5 and PRDX1, and low expression of BCL2L2 informs poor patient prognosis (Figure.3A-
D). The expression level of BCL2L2 was significantly decreased with higher grade glioma (Figure.3E). Moreover, the expression of SEC61B, XRCC5 and PRDX1 was significantly increased with higher grade glioma (Figure.3F-H). To further validate these results, we performed immunohistochemical experiments. The immunohistochemistry results obtained in the present study were consistent with the results of the CGGA database (Figure.3I). In all, four signature-genes expressions are further considered to be of clinical significance in GBM.

**GSEA analysis of the four signature genes**

To further clarify the impact of the four signature genes on GBM, gene ontology and pathway enrichment analyses were performed using GSEA. The results revealed that these genes were mainly enriched in 14 pathways based on the TCGA-GBM database, involved in the calcium signaling pathway, cell cycle, ERBB signaling pathway, GAP junction, Glioma, inositol phosphate metabolism, MAPK signaling pathway, oxidative phosphorylation, phosphatidylinositol signaling system, purine metabolism, ribosome, RNA degradation, spliceosome, and VEGF signaling pathway (Figure.4A). Moreover, In the GEO dataset, these genes were mainly enriched in 9 pathways, involved in the calcium signaling pathway, cell cycle, ECM receptor interaction, ERBB signaling pathway, inositol phosphate metabolism, oxidative phosphorylation, P53 signaling pathway, phosphatidylinositol signaling system, and pyrimidine metabolism (Figure.4B). These genes may be involved in the proliferation of GBM.

**The prognostic impact of the IFN-β associated IncRNAs signature for GBM**

Considering the critical role of IncRNA in GBM, the identification of important IncRNAs in cancer and developing IncRNA-based therapeutic strategies would be meaningful in the future. Pearson correlation was applied to calculate the correlation between the IncRNAs and IFN-β associated genes. A IncRNA with a correlation coefficient $|R^2| > 0.3$ and $P < 0.05$ was considered to be an IFN-β associated IncRNA. Univariate and multivariate Cox regression were used for survival analysis indicated that AC093278.2, AC004067.1, LINC01116, and AC017104.1 was an independent prognostic factor for GBM patients (Figure.5A-C). Moreover, the high expression of the AC004067.1, AC017104.1, and LINC01116 is associated with poor prognosis, and the low expression of the AC093278.2 is associated with poor prognosis in the TCGA dataset (Figure.5D-G). Further, GBM samples (n=50) and normal brain tissues (n=50) were then used for real-time polymerase chain reaction (PCR) experiments to validate the expression of the IFN-β associated IncRNAs in GBM. qRT-PCR showed that compared with the normal brain tissues, AC004067.1, AC017104.1, and LINC01116 was highly expressed, AC093278.2 was low expressed in GBM (Figure.5H). Taken together, IFN-β associated IncRNAs have a high diagnostic value for GBM.

**Discussion**

The understanding of cancer treatment has gradually changed powerful with high-throughput sequencing technologies became more pervasive and user-friendly. As the scope of analyzed genes and diseases
expands, bioinformatic analysis is increasingly critical. In the present study, we analyzed the biological functions of the prognostic IFN-β associated genes signature using bioinformatics analysis. This approach may offer clues for therapeutic targets.

Univariate Cox regression model found 5 survival-related genes, then Lasso-penalized Cox analysis identified 5 genes to construct the prognostic model. Using the methodology previously described, the proposed approach is validated on the GEO datasets (GSE83300). We found that 4 IFN-β associated genes (PRDX1, SEC61B, XRCC5, and BCL2L2) signature was a robust marker of seizure prognosis in patients with GBM. With CGGA data, we validated that 4 IFN-β associated genes was an independent biomarker of prognosis and played important roles in many biological processes. For example, the PRDX1 (Peroxiredoxin 1) this gene encodes a member of the peroxiredoxin family of antioxidant enzymes, which reduce hydrogen peroxide and alkyl hydroperoxides. PRDX1 forms a heterodimer with p38α mitogen-activated protein kinase 14 (MAPK14), stabilizing phosphate-p38α in glioma cells and epigenetic silencing of PRDX1 is frequent in 1p/19q-deleted oligodendrogial tumors and likely contributes to radio- and chemosensitivity of these tumours; XRCC5 (X-Ray Repair Complementing 5) is the 80-kilodalton subunit of the Ku heterodimer protein, which is also known as ATP-dependant DNA helicase II or DNA repair protein XRCC5. The Polymorphisms of XRCC5 play an important role in astrocytoma prognosis in the Chinese Han population, which could be used in the determination of astrocytoma prognosis in clinical researches; BCL2L2 (BCL2 Like 2) is a member of the BCL-2 protein family. The proteins of this family form hetero- or homodimers and act as anti- and pro-apoptotic regulators. The expression of BCL2L2 mRNA or protein in various cancer cell types. Interestingly, BCL2L2 mRNA is highly expressed in the mesenchymal type of GBM. SEC61B (SEC61 Translocon Subunit Beta) is the central component of the protein translocation apparatus of the endoplasmic reticulum (ER) membrane. However, to the best of our knowledge, the expression pattern and function of SEC61B in GBM have not been previously reported. Further study is warranted.

Considering the critical role of lncRNA in GBM, the identification of important lncRNAs in cancer and developing lncRNA-based therapeutic strategies would be meaningful in the future. Pearson correlation was applied to calculate the correlation between the lncRNAs and IFN-β associated genes. Univariate and multivariate Cox regression were used for survival analysis indicated that AC093278.2, AC004067.1, LINC01116, and AC017104.1 was an independent prognostic factor for GBM patients. LncRNA genes play important roles in many biological processes. For example, LINC01116 promotes tumor proliferation migration and invasion in glioma cell. However, the role of AC093278.2, AC004067.1, and AC017104.1 in GBM has not been reported, is urgent to be elucidated.

In conclusion, we explored and analyzed differentially expressed IFN-β associated genes by systematic bioinformatics analysis.

Conclusions
Our study established a novel IFN-β associated genes signature to predict overall survival of GBM, which may help in clinical decision making for individual treatment.

**Abbreviations**

BCL2 Like 2, BCL2L2; Chinese Glioma Genome Atlas, CGGA; Gene Expression Omnibus, GEO; glioblastoma, GBM; O6-methylguanine-DNA methyltransferase, MGMT; Interferon Beta 1, IFN-β; SEC61 Translocon Subunit Beta, SEC61B; The Cancer Genome Atlas, TCGA; Peroxiredoxin 1, PRDX1; X-Ray Repair Cross Complementing 5, XRCC5;

**Declarations**

**Ethics approval and consent to participate**

The study was approved by the Ethics Committee of the Jiangxi Cancer Hospital of Nanchang University with informed consent from the patients or guardians.

**Consent for publication**

All contributing authors agree to the publication of this article.

**Availability of data and materials**

All raw data used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Competing interests**

The authors declare that there are no conflict of interests.

**Funding**

No funding

**Author contributions**

LC and MY conceived of the study and participated in experiment design and coordination, drafted and revised the manuscript. JZ performed gene differential analysis and survival analysis using TCGA, GEO.
and CGGA data. JC collected tissue samples and performed experiment. SL and SZ performed statistical analysis and revised manuscript.

**Acknowledgements**

We are very thankful to all the participants in the present research and the highly valuable comments of Profs. We also thank the reviewers for their valuable advice.

**References**


**Figures**
Figure 1

Construction and validation of the prognostic IFN-β associated genes signature. (A) Heatmap showing differential expression of 67 genes between normal brain tissues and GBM tissues that differed significantly (p<0.05). (B) Survival curve of GBM patients based on risk score model in TCGA. (C) Survival curve of GBM patients based on risk score model in TCGA. (D-E) The survival duration and status of GBM cases. Interferon-beta associated gene risk score analysis of glioma patients in TCGA and GEO. Heatmap of the 5 key genes expressed in GBM.
Figure 2

Prognostic significance of four signature-genes expression in GBM. (A-B) Survival analysis of the 2 prognostic interferon-beta associated genes based on the GEO database. (C-F) Survival analysis of the 4 prognostic interferon-beta associated genes based on the TCGA database. (G) The expression analysis of 4 prognostic interferon-beta associated genes according to GEPIA database. (H) Real-time PCR expression analysis of 4 prognostic interferon-beta associated genes in normal brain tissues and GBM tissues (N=50), ***p<0.001.
Figure 3

Figure 4

GSEA analysis of the four signature genes. (A-B) GSEA analysis of the 4 prognostic interferon-beta associated genes based on TCGA and GEO database.
Figure 5

The prognostic impact of the IFN-β associated lncRNAs signature for GBM. (A) Network of prognostic lncRNAs with co-expressed interferon-beta associated lncRNAs in GBM. (B) Survival curve of GBM patients based on risk score model in TCGA. (C) The survival duration and status of GBM cases. Interferon-beta associated lncRNAs risk score analysis of GBM patients in TCGA. Heatmap of the 4 key lncRNAs expressed in GBM. (D-G) Survival analysis of the 4 prognostic interferon-beta associated lncRNAs based on the TCGA database. (H) Real-time PCR expression analysis of 4 prognostic interferon-beta associated lncRNAs in normal brain tissues and GBM tissues (N=50), ***p<0.001.