Prediction of Survival Outcome in Lower-Grade Glioma Using a Prognostic Signature with 33 Immune-Related Gene Pairs

Shaohua Chen¹,²,*
Yongchu Sun³,⁴,*
Xiaodong Zhu¹,³–⁵
Zengnan Mo¹,²

¹Department of Urology, The First Affiliated Hospital of Guangxi Medical University, Nanning, People’s Republic of China; ²Guangxi Collaborative Innovation Center for Genomic and Personalized Medicine, Nanning, People’s Republic of China; ³Department of Radiation Oncology, Guangxi Medical University Cancer Hospital, Nanning, People’s Republic of China; ⁴Guangxi Key Laboratory of Early Prevention and Treatment for Regional High Frequency Tumor, Guangxi Medical University, Nanning, People’s Republic of China; ⁵Department of Oncology, Affiliated Wuming Hospital of Guangxi Medical University, Nanning, People’s Republic of China

*These authors contributed equally to this work

Background: Lower-grade glioma (LGG) is one of the prevalent malignancies threatening human health, with considerable intrinsic heterogeneities in their biological behavior. Previous studies have revealed that the immune component is a key factor influencing the formation and development of malignancies. In this study, we aim to use a novel approach to develop a prognostic signature of immune-related gene pairs (IRGPs) to determine the survival outcome of patients with LGG.

Methods: Transcriptomic profiles and clinical data for LGG were obtained from The Cancer Genome Atlas (TCGA) and Chinese Glioma Genome Atlas (CGGA) databases, and used as training and validation data sets, respectively. IRGPs influencing the overall survival (OS) of patients with LGG in the training data set were screened by performing univariate Cox regression analysis. Next, a prognostic IRGPs signature was constructed using least absolute shrinkage and selection operator (LASSO) regression. Finally, we cross-validated the two databases to verify the stability of the prognostic signature.

Results: A total of 33 IRGPs influencing prognosis of LGG in the training data set were included in the prognostic signature. Patients with high risk scores (RSs) in the training and validation data sets had a poorer OS than those with low RSs. Moreover, significant differences were observed in tumor-infiltrating immune cells (TICs) between high- and low-RS groups. Functional enrichment analyses results revealed that genes in the high-RS group were enriched in the immune-related activities and developmental processes.

Conclusion: The prognostic signature containing 33 IRGPs has a significant correlation with OS and relative levels of immune cells associated with LGG. The results of the present study provide new insights into the prediction of survival outcome and therapeutic response of LGG.

Keywords: lower-grade glioma, immune-related gene pairs, tumor-infiltrating immune cells, prognostic signature

Introduction

Gliomas are the most prevalent types of primary malignant tumors in the central nervous system.¹ Based on the histopathological features outlined by the World Health Organization (WHO), grade II and III gliomas are defined as lower-grade gliomas (LGGs),² which include astrocytomas, oligodendrogliomas, and oligoastrocytomas.³ Currently, the standard care for LGG comprises surgical resection combined with postoperative radiotherapy and chemotherapy. LGG exhibits considerable intrinsic heterogeneity in biological behavior, with overall survival (OS) ranging from 1 to 15 years.⁴ Nevertheless, there are some patients who
develop drug-resistant high-grade gliomas, resulting in adverse long-term outcomes. Despite advances in the diagnosis and treatment modalities of LGG in recent years, the prognosis of patients with LGG has not improved substantially over the last few decades. Therefore, identification of novel targets or biomarkers to facilitate early diagnosis and treatment response of LGG is imperative. Risk classification of glioma has been frequently applied in clinical decision-making, and WHO established specific molecular markers for brain tumor classification in 2016, such as chromosome 1p and 19q (chr1p/19q) co-deletions, isocitrate dehydrogenase, and histone H3K27M mutations. This classification is superior to the previous histopathological classification. However, the widespread clinical dissemination of such a classification system remains a huge challenge due to the level of molecular diagnostic technology in each medical center.

Immune components of the tumor microenvironment (TME) that influence the formation and development of malignancies have been elucidated. Consequently, several promising preclinical and clinical immunotherapeutic approaches, including immune checkpoint inhibitors (ICIs), immunotherapy, and gene therapy have been applied in the management of gliomas with satisfactory results. Strikingly, tumor development is thought to be dependent on the infiltration of immune cells, while immune-related genes (IRGs) are thought to be indicative of patient prognosis, which has been reported in various types of cancers, such as esophageal cancer, bladder cancer, lung cancer, and colon cancer. Therefore, the expression profiles of IRGs present great potential in predicting the prognosis and therapeutic response of LGG, which merits further investigation to identify a sophisticated prognostic signature that can predict survival outcome and guide appropriate individualized treatment strategies.

However, traditional methods of data normalization and scaling in the processing of gene expression data have certain limitations, making it difficult to draw reliable conclusions. In the present study, we used a novel method that involved ranking of relative gene expression levels, thereby establishing a prognostic signature for IRGPs. In addition, two separate datasets, including LGG cases from the Cancer Genome Atlas (TCGA) and The Chinese Glioma Genome Atlas (CGGA) databases, were cross-validated to verify its reliability and authenticity as a prognostic marker for LGG.

Materials and Methods

Acquisition and Processing of Transcriptomic and Clinical Data

Transcriptomic data for LGG, which included 529 tumor samples, and corresponding clinical data were obtained from TCGA database (https://portal.gdc.cancer.gov/) and used as the training data set. Other dataset containing 443 LGG samples from CGGA database (http://www.cgga.org.cn/) were used as the validation data set.

Construction of a Prognostic IRGPs Signature for LGG

Next, we obtained IRGs from the ImmPort database (https://www.immport.org/). Subsequently, the transcriptome data of TCGA-LGG cohort were analyzed using limma package of R, intersected with IRGs from the ImmPort database, and retained the IRGs. The relevant IRGs were screened further based on a median absolute deviation (MAD) of >0.5 to attain a comparatively high variation. Afterwards, the transcriptome data for LGG cases obtained from the CGGA database were screened based on MAD value of >0.5. Gene lists obtained from the two databases were intersected to identify a shared list of significant IRGs and their respective expression matrices from the two databases. LGG cases obtained from TCGA and CGGA databases were used as training and validation data sets, respectively. Thereafter, we constructed IRGPs for each patient in two databases: in the pairwise comparison of each sample, the expression level of the first gene was regarded as zero if the expression level of the gene was lower than the second gene; otherwise, it was regarded as one. Samples with ratios of zero and one that were less than 20% and greater than 80% were removed to retain gene pairs that were associated with LGG prognosis. Subsequently, univariate Cox regression analysis of IRGPs from TCGA was performed using the survival package in R based on p-value <0.001 as a filtering condition. Least absolute shrinkage and selection operator (LASSO) regression was performed using the glmnet package in R, and we finally obtained prognostic signatures containing 33 IRGPs. A time-dependent receiver operating characteristic (ROC) curve with three-year survival rate for the training data set was constructed using survROC package. LGG cases in the training and validation data sets were divided into high and low risk score (RS) groups based on the optimal cutoff value of the ROC curve for three-year survival rate.
Cross-Validation of Prognostic IRGPs Signatures

The high- and low-RS groups in the training and validation data sets were compared by performing Kaplan–Meier survival analyses using survival and survminer packages in R. Furthermore, univariate and multivariate Cox regression analyses were used to assess the efficacy of the signatures and the prognostic value of other clinical data.

Tumor-Infiltrating Immune Cell Analysis

To evaluate the relative levels of tumor-infiltrating immune cells (TICs) associated with LGG, we used the CIBERSORT algorithm to calculate the relative levels of 22 TICs in each tumor sample in the training and validation data sets. CIBERSORT is an analytical tool developed by Newman et al, and it requires an input matrix of transcriptomic data, thereby providing an estimation of the relative level of immune cell subtype based on linear support vector regression.17,18 Thereafter, cancer samples with $P > 0.05$ were excluded, and differential analyses of the relative levels of 22 TIC subtypes in the high- and low-RS groups were performed using Wilcoxon test. The results were then visualized using vioplot package in R.

Gene Set Enrichment Analysis of High- and Low-RS Groups

We performed gene set enrichment analysis (GSEA) of differentially expressed genes in high- and low-RS groups using the fgsea package in R. Log2 fold change was made between the two groups, and $P_{adj} < 0.05$ was considered statistically significant. Finally, the results were visualized using the ggplot2 package in R.

Statistical Analysis

Statistical analyses and visualization of data were performed using R software (version 4.0.3). A $p$-value less than 0.05 (two-sided) was considered statistically significant.

Results

Screening of IRGs

A list of 2483 IRGs was obtained from the ImmPort database, which intersected the transcriptome data for LGG obtained from TCGA database, and finally obtained 1692 IRGs. Subsequently, the transcriptome data for glioma in the CGGA database was downloaded, and the LGG cases were retained. Gene expression matrices obtained from the two databases were filtered based on $\text{MAD} > 0.5$, while the IRGs that were shared by the two databases were obtained by intersection analysis for subsequent analysis. A total of 659 IRGs were identified. In the present study, TCGA-LGG and CGGA-LGG cohorts were used as the training and validation data sets, respectively.

Establishment of Prognostic Signature for IRGPs

IRGPs were constructed based on a one-on-one matching of 659 IRGs in the training data set. Univariate Cox regression and Kaplan–Meier survival analyses were used to screen for key IRGPs influencing prognosis of LGG, and 10469 key IRGPs were identified. Finally, LASSO regression was used to further screen the IRGPs, and 33 IRGPs used to establish the prognostic signature were obtained, which included a total of 49 unique IRGs. The names and categories of the IRGs and their coefficients in the prognostic signature are presented in Table 1.

Subsequently, the time-dependent ROC curve analysis results for three-year OS revealed that the optimal cutoff value was 0.113 (Figure 1). The RSs of the cases in the training and validation data sets were calculated and used to classify patients into high- and low-RS groups based on the optimal cutoff value, RSs above the optimal cut-off were assigned to the high-RS group, while RSs below the optimal cut-off were classified into the low-RS group. The heatmap displays the distribution of the 33 IRGPs in the training data set (Figure 2A). Risk curve (Figure 2B) as well as the survival status (Figure 2C) of the LGG cases in the training data set indicated that the proportion of deaths in the high-RS group was significantly higher than in the low-RS group, similar results were obtained in the same analyses of validation data set (Figure 2D–F). Further Kaplan–Meier survival analysis of the training data set demonstrated that the prognosis of patients with LGG in the high-RS group was poorer than that in the low-RS group (Figure 3A). Kaplan–Meier survival analysis of the LGG cases in the validation data set was conducted for cross-validation, confirming that the adverse prognosis of LGG cases in the high-RS group was also presented in the validation data set (Figure 3B).

Furthermore, univariate and multivariate Cox regression analyses were performed on the training data set to analyze prognostic IRGPs signature and clinical features associated with LGG. Overall, age and RS of signature...
were identified as the independent prognostic factors for predicting the survival outcome of patients with LGG (Figure 4A and B). Univariate and multivariate Cox regression analyses of the validation data set also indicated that RS of signature was a detrimental prognostic factor for LGG (Figure 4C and D). In conclusion, the prognostic IRGPs signature exhibited superior stability and could assess the prognosis of patients with LGG accurately.

### Analysis of TICs

To analyze variations in the profiles of TICs, we used the CIBERSORT algorithm to calculate the relative levels of TICs in the TCGA-LGG and CGGA-LGG cohorts, thereby constructing 22 kinds of TICs profiles and their correlation analysis of training data set (Figure 5A and B) and validation data set (Figure S1A and B). Violin plots were used to visualize comparative analyses of TICs between high- and low-RS groups in the training data set (Figure 6). Notably, naïve B cells, activated memory CD4+ T cells, CD8+ T cells, M0 macrophages, M1 macrophages, M2 macrophages, and regulatory T cells (Treg) were highly expressed in the high-RS group. In contrast, eosinophils, activated mast cells, monocytes, and activated NK cells were highly expressed in the low-RS group. Nevertheless, TICs profiles of CGGA-LGG cohort indicated that monocytes were the only TIC subtype that exhibited statistically significant differences between the two groups (Figure 7). The above conclusions shed light

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The Name and Category of Each Immune-Related Gene and Coefficient in the Prognostic Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene (Pair 1)</td>
<td>Category</td>
</tr>
<tr>
<td>MR1</td>
<td>Antigen Processing and Presentation</td>
</tr>
<tr>
<td>HSPA1L</td>
<td>Antigen Processing and Presentation</td>
</tr>
<tr>
<td>KLRC2</td>
<td>Antigen Processing and Presentation</td>
</tr>
<tr>
<td>PSMD5</td>
<td>Antigen Processing and Presentation</td>
</tr>
<tr>
<td>RELB</td>
<td>Antigen Processing and Presentation</td>
</tr>
<tr>
<td>CXCL12</td>
<td>Chemokines and Cytokines</td>
</tr>
<tr>
<td>CXCL12</td>
<td>Chemokines and Cytokines</td>
</tr>
<tr>
<td>CXCL12</td>
<td>Chemokines and Cytokines</td>
</tr>
<tr>
<td>MX1</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>TFRC</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>TFRC</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>NEDD4</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>F2R</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>CYLD</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>CYLD</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>ITGAV</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>ITGAV</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>BST2</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>MSR1</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>APOBEC3C</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>BIRC5</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>BIRC5</td>
<td>Antimicrobials</td>
</tr>
<tr>
<td>RAC3</td>
<td>BCR Signaling Pathway</td>
</tr>
<tr>
<td>PPP3CB</td>
<td>BCR and TCR Signaling Pathway</td>
</tr>
<tr>
<td>SEMA4G</td>
<td>Chemokines and Cytokines</td>
</tr>
<tr>
<td>SEMA4G</td>
<td>Chemokines and Cytokines</td>
</tr>
<tr>
<td>PLAUR</td>
<td>Cytokine Receptors</td>
</tr>
<tr>
<td>ROBO2</td>
<td>Cytokine Receptors</td>
</tr>
<tr>
<td>TGFBI</td>
<td>Cytokines</td>
</tr>
<tr>
<td>IL17RC</td>
<td>Cytokine Receptors</td>
</tr>
<tr>
<td>LGR5</td>
<td>Cytokine Receptors</td>
</tr>
<tr>
<td>NR2E1</td>
<td>Cytokine Receptors</td>
</tr>
<tr>
<td>NR2E1</td>
<td>Cytokine Receptors</td>
</tr>
</tbody>
</table>
on the fact that the relative levels of specific TIC subtypes in the TME of LGG could be associated with the formation and development of LGG.

Functional Enrichment Analysis of Prognostic IRGPs Signature

Next, GSEA was performed to elucidate variations in the gene expression profiles between high- and low-RS groups in the training data set. The results revealed that enriched terms of GSEA analysis in the high-RS group were mostly immune-related activities, including immune effector process, cell activation, and defense response to other organisms. Intriguingly, development-related processes were also involved, including regulation of multicellular organismal development and epithelium development. The enriched terms ranked based on p-value were visualized using bubble plots (Figure 8), and top five enriched terms in GSEA analysis of training data set were immune effector process, cell activation, defense response to other organisms, response to cytokine, and chromatin (Figure 9), which were consistently up-regulated in the high-RS group.

Similar results were acquired in the functional enrichment analysis of validation data set (Figure S2), and top five enriched terms in GSEA analysis of training data set were molecular transducer activity, animal organ morphogenesis, positive regulation of immune system process, regulation of immune system process, and tube development (Figure S3), explaining underlying reasons for the variations in the prognosis of the two groups of patients.

Discussion

Glioma is the most widespread malignancy that influences the nervous system, accounting for approximately 80% of malignant brain tumors. The overall estimated incidence of gliomas is 4.7–5.7 cases per 100,000 people. LGG is one of the common pathological types of gliomas that is characterized by gradual growth. Nevertheless, LGG is associated with an extremely high recurrence rate, which frequently leads to severe disability and death due to its aggressive and invasive nature. ICIs have received increasing attention because of their promising results in the treatment of tumors. Similar results have been obtained in clinical trials of glioma, with a certain therapeutic value and response rate. However, the efficacy of glioma has not been outstanding when compared to other cancers, such as melanoma and lung cancer, while drug efficacy exhibits considerable individual variations and has been associated with varying degrees of treatment-related adverse events. In the present study, we established a prognostic IRGPs signature for predicting the prognosis of LGG with an unbiased method for data analysis based on ranking and pairwise comparisons of relative gene expressions. A similar approach has been used in numerous studies, with remarkable results. Furthermore, the stability and reliability of the prognostic signature were verified by cross-validating the two databases.

The prognostic signature comprised 33 IRGPs, including 49 independent IRGs, which predominantly consist of cytokines and cytokine receptor-related functions. Notably, cytokines have been reported to play an essential role in the communication between different components of TME in glioblastoma (GBM), while cytokine networks constituted by transforming growth factor β (TGF-β), interleukin (IL)-6, and IL-10 can impede T cell proliferation and response, thereby suppressing anti-tumor immunity, and promoting tumor progression and metastasis. The characteristics of low immunogenicity of glioma and the complex cytokine network of immunosuppression in TME could contribute to the poor efficacy of immunotherapy. Similarly, chemokines also have a primary function among the IRGs in the prognostic signature. Chemokines influence the regulation of inflammation, infection, immune response, tissue damage response, and apoptosis by providing optimal host defense against infection with pathogens. Strikingly, chemokines and their receptors also
play a decisive role in tumor progression. Chemokines can indirectly regulate tumor growth by recruiting leukocytes, in turn, influencing the immune status of TME. Previous studies have revealed that chemokines can influence tumor progression in a multifaceted manner in the TME of GBM. CXCL8 is an inflammation-associated chemokine with tumorigenic and pro-angiogenic properties, which has been demonstrated to be upregulated in tumor cells of gliomas, and its expression level is consistent with the pathological grades of gliomas. Studies have revealed that CXCL8 is involved in the cellular process of epithelial–mesenchymal transition (EMT) by activating the JAK/STAT1/HIF-1α/Snail signaling pathway in GBM cells, thus promoting progression of cancer cells. Therefore, targeting cytokines and chemokines could be an alternative solution to the improvement of immunotherapy efficacy in the treatment of glioma.

TICs have been identified as crucial components in tumor biology, and they influence tumor development and growth. The roles of TICs in gliomas have also been reported in previous studies. Several studies have revealed that Tregs are key factors influencing the immunosuppressive state of glioma TME. Tregs inhibit the secretion of pro-inflammatory factors (eg, IL-2 and interferon gamma [IFN-γ]) while promoting the excretion of immunosuppressive factors (eg, TGF-β), thereby suppressing antitumor response in various lymphocytes. In addition, Tregs bind to CD80 or CD86 through CTLA-4, which can substantially inhibit activation of effector T cells.

Figure 2 Establishment of prognostic signature for IRGPs. (A–C) The distribution of 33 IRGPs, risk curve and the survival status of LGG cases in the training data set. (D–F) The distribution of 33 IRGPs, risk curve and the survival status of LGG cases in the validation data sets.
Similarly, macrophages are a canonical subtype of TICs that stably accumulate through the development of gliomas. Macrophages with pro-angiogenic factors have been found to be significantly upregulated in macrophages isolated from tumors, which interact with tumor vessels and are associated with vascular remodeling. In the present study, we observed that the relative levels of Tregs and macrophages in patients with LGG in the high-RS group

Figure 3 Kaplan–Meier survival analyses of LGG cases between high- and low-RS groups. (A) Kaplan–Meier survival analyses of patients in the training data set. (B) Kaplan–Meier survival analyses of patients in the validation data set.

Figure 4 Univariate and multivariate Cox regression analyses results of prognostic IRGPs signature, and clinical features in the training and validation data sets. (A and B) Forest plots showing univariate and multivariate Cox regression analyses results of the training data set, respectively. (C and D) Forest plots showing univariate and multivariate Cox regression analyses results of the validation data set, respectively.
Figure 5 TICs profiles of the training data set. (A) Barplot showing the relative levels of 22 TIC subtypes in tumor tissue of the training data set. (B) Heat map showing the correlation analysis of 22 TIC subtypes. The number and color in the circles indicate the correlation coefficient between two different TIC subtypes.
were relatively higher, suggesting that the two subtypes of TICs are associated with progression and metastasis of LGG. Conversely, the relative levels of activated NK cells and monocytes were relatively higher in the low-RS group and were associated with better prognosis of LGG. The findings of the present study suggest that TICs play a pivotal role in the TME, which provide a novel avenue for the improvement of the immunosuppressive state of TME and guides clinical treatment responses of LGG.

However, the present study had certain limitations. First, we established a prognostic IRGPs signature of LGG based on data obtained from TCGA database and cross-validated the signature with LGG cases in the CGGA database, which achieved satisfactory test efficiency and stability. Nevertheless, future studies involving larger samples are required to validate the reliability of the IRGPs signature. Second, gene expression profiling is still dependent on RNA-seq and microarray data; therefore, extensive promotion remains a huge challenge, which requires further optimization of testing programs to enhance their clinical application.
Conclusion

In general, we established a prognostic IRGPs signature that can accurately evaluate the prognosis of patients with LGG. The risk stratification results based on the signature revealed that the relative levels of TICs and gene expression profiles of patients with low- and high-RSs were significantly different. Therefore, the IRGPs signature is a robust and stable approach for predicting the survival outcome and treatment responses of LGG immunotherapy.

Ethical Statement

This study was approved by the Guangxi Medical University Cancer Hospital (LW2021070).
Acknowledgments
We acknowledge TCGA and CGGA databases for providing their platforms and contributors for uploading their meaningful datasets.

Funding
There is no funding to report.

Disclosure
The authors report no conflicts of interest in this work.

References


