The Signature of Glycometabolism-Related Genes in Predicting the Prognosis of Patients with Osteosarcoma

Fengyan Wang  
The Affiliated Hospital of Guizhou Medical University

Kun Yang  
The Affiliated Hospital of Guizhou Medical University

Ke Li  
Guizhou Provincial People's Hospital

Guoxuan Peng  
Guizhou Medical University

Hua Yang  
The Affiliated Hospital of Guizhou Medical University

Yang Xiang  
The Affiliated Hospital of Guizhou Medical University

Hong Sun  
Guizhou Medical University  
https://orcid.org/0000-0002-4195-501X

Research Article

Keywords: Osteosarcoma, Glycometabolism, Prognosis, Risk model, Biomarkers, Gene signature

Posted Date: February 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1270483/v1

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

Background

Osteosarcoma is a primary malignant bone tumor with high metastatic potential and an inferior prognosis. Glycometabolism also plays a role in the disease. However, the clinical significance of glycometabolism-related genes in patients with osteosarcoma has been unclear.

Methods

We downloaded the expression profile and corresponding clinical data of osteosarcoma samples from database. Glycometabolism-related gene sets were obtained. Regression analyses were performed to construct a glycometabolism-related prognostic gene signature. The independent prognostic value of the signature was further assessed by univariate and multivariate Cox regression analysis, and the correlation between immune cells and the signature was investigated. The regulatory mechanism of the prognostic genes was explored by constructing a ceRNA network.

Results

A glycometabolism-related prognostic gene signature based on PRKACB, SEPHS2, GPX7, and PFKFB3 was constructed, and the survival and receiver operating characteristic curves showed that the glycometabolism-related gene signature had a good performance in predicting the overall survival of patients with osteosarcoma. Univariate and multivariate Cox regression analyses confirmed that the glycometabolism-related gene signature was an independent prognostic factor among patients. Correlation analysis revealed that SEPHS2, PRKACB, and GPX7 were correlated with immune cells. A ceRNA network comprised of four genes, 148 miRNAs, and 91 lncRNAs was constructed.

Conclusions

A novel signature based on four glycometabolism-related genes, was constructed, which may facilitate the prognosis and treatment of osteosarcoma in clinical practice.

Introduction

Osteosarcoma (OS) is a malignancy derived from primitive osteogenic mesenchymal cells. It is the most common type of malignant bone tumor among children and adolescents [1]. OS occurs mostly in the metaphysis of long bones, especially around the knees, and is characterized by high rates of metastasis and progression [2]. Along with the improvement of neoadjuvant therapy and surgical resection, more than two-thirds of patients with localized lesions are likely to achieve long-term survival [3; 4]; however, approximately 30% of patients with non-metastasis at diagnosis suffer from concomitant lung
metastasis after comprehensive treatment. The 5-year survival rate of patients with distant metastasis at diagnosis is still unfavorable, and effective therapeutic interventions for patients with metastatic OS or chemoresistance are lacking [5]. Therefore, it is important to identify new prognostic biomarkers for early diagnosis and to investigate novel targets for OS treatment.

Abnormal glycometabolism has been reported as one of the hallmarks of how malignant tumor cells arise [6]. It is characterized by increased glucose consumption for rapid proliferation and invasion [7]. Aerobic glycolysis, also termed the Warburg effect, is widely and preferentially used in the glycometabolism pathway in tumor cells [8]. Multiple studies suggested that aerobic glycolysis is a critical step in the initiation, proliferation, and migration of OS [9-11]. Considering these findings, targeting the glycolytic pathway may be an attractive therapeutic option for the treatment of OS; however, the function and underlying mechanisms of glycometabolism in OS development remain largely unknown.

In this study, we performed a comprehensive bioinformatics analysis to screen glycometabolism-related genes in OS progression, and investigated the potential molecular mechanisms of abnormal glycometabolism-mediated OS development to explore novel therapeutic strategies for the treatment of OS.

**Materials And Methods**

**Data acquisition**

The expression profile and corresponding clinical data of OS samples were downloaded from the TARGET database (TTD, http://bidd.nus.edu.sg/group/td/ttd.asp) [12]. The dataset includes 85 OS samples with survival information that were divided into training (n=43) and testing (n=42) cohorts with a proportion of 1:1 randomly. Moreover, the GSE39055 dataset, that included 36 OS samples with survival information was downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) as a validation cohort. According the GO Annotation, glycometabolism genes were identified in the molecular signatures database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) [13]. Then eight glycometabolism gene sets were extracted after removing overlapping genes. The entire glycometabolism gene sets contained 291 genes, and 282 genes remained after removing genes whose expression was zero in 50% of the samples.

**Construction and validation of a glycometabolism-related genes signature**

Univariate Cox analysis was performed to assess the prognostic value of glycometabolism-related genes which with a \( P < 0.05 \) were considered potential prognostic genes. Next, the Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm was used to construct an optimal glycometabolism-related gene signature. The risk score for each patient was calculated as follows:

\[
\text{risk score} = \sum_{i=1}^{n} \beta_i \times \text{Exp gene}(i)
\]

where \( n \) is the number of genes in this prognosis model, beta (\( \beta \)) is the
regression coefficient, and \( \text{Expgene} \) is the expression level of each gene. The patients from the training cohort were divided into high- and low-risk groups based on the median value of the risk score. Moreover, Kaplan-Meier curves and receiver operating characteristic (ROC) curves in testing and validation cohort were used to assess the prognostic capacity of the gene signature.

**Evaluation of the relationship between glycometabolism-related gene signature and clinical characteristic**

Clinical data were extracted from the whole TARGET cohort (including samples in the training and testing cohorts), including gender and age. The correlation between the glycometabolism-related gene signature and clinical characteristics calculated by \( t \)-test. A result of \( P < 0.05 \) indicates that the risk score is significantly different between the two groups. Meanwhile, we also analyzed the correlation between the risk score and clinical characteristics in the cohorts, indicating that the risk score was significantly different between the two groups.

**Independent prognostic analysis and construction of a nomogram**

The univariate and multivariate Cox regression analyses were used to assess whether the risk score could be used as an independent prognostic factor in the whole TARGET cohort (including samples in the training and testing cohorts). The results of multivariate regression analysis were used to establish a nomogram to predict the 1-, 3-, and 5-year overall survival. Moreover, the discrimination and accuracy of the nomogram were assessed by the concordance index (C-index), calibration curves, and decision curves.

**Functional enrichment analysis**

The “limma” package was used to analyze the all gene expression profiles between the high- and low-risk groups in the training and testing cohort. Gene set enrichment analysis (GSEA) was performed to analysis the biological functions of the gene. Gene Ontology (GO; http://www.geneontology.org) and Kyoto Encyclopedia of Genes and Genomes (KEGG http://www.kegg.jp/ or http://www.genome.jp/kegg/) pathway enrichment analyses were performed using the ClusterProfiler package [14,15]. We also compared the single-sample GSEA (ssGSEA) scores between the high- and low-risk groups based on gene expression profiles involved in the top 10 GO and KEGG terms [16].

**Immune cell analysis**

To analyze the differences in the proportion of 28 immune cell types between the high- and low-risk groups, ssGSEA was performed in the training and testing cohorts. Then we used “psych” package of R performed a correlation analysis between the proportion of the 28 immune cells and the gene signature.

**Construction of a competitive endogenous RNA (ceRNA) network**

Micro RNAs (miRNAs) that can regulate genes in the signature were predicted based on the miRanda software. Then long non-coding RNAs (IncRNAs) were predicted through the miRanda software by
miRNAs. To improve the accuracy of the competitive endogenous RNA (ceRNA) network, we further screened the results using the following criteria: combined score > 200 was set as the screening criteria of lncRNA-miRNA interactions, and combined score > 200 and MFE score < -200 were set as the screening criteria of miRNA-gene interactions. Finally, a ceRNA network was visualized by Cytoscape.

Statistical analysis

All the above analyses were completed by the R software. A time-dependent ROC analysis was performed by the “pROC” package [17]. In addition, a nomogram was constructed by the “rms” packages [18]. \( P \)-value<0.05 was considered statistically significant.

Results

Identification of the four glycometabolism-related gene signatures

Univariate Cox regression analysis identified 19 glycometabolism-related genes in the training set, which were correlated with the survival of patients with OS (Figure 1A). LASSO regression analysis identified four glycometabolism-related genes, including \textit{PRKACB}, \textit{SEPHS2}, \textit{GPX7}, and \textit{PFKFB3} (Figure 1B and 1C). Subsequently, combining the gene expression and corresponding regression coefficient, a risk-scoring model was established using the following formula: 
\[
\text{risk score} = \text{expression level of } \textit{PRKACB} \times (-0.10894429) + \text{expression level of } \textit{SEPHS2} \times (-0.01563243) + \text{expression level of } \textit{GPX7} \times 0.32559346 + \text{expression level of } \textit{PFKFB3} \times 0.55009921.
\]

According to the median risk score, patients were divided into high- and low-risk groups (Figure 1D). The K-M survival analysis showed that patients in the high-risk group had significantly shorter overall survival than those in the low-risk group. Furthermore, patients with low-risk scores had a survival advantage over those with a high-risk score (Figure 1E), and the area under the ROC curve (AUC) was 0.904, 0.939, 0.900, 0.900, and 0.919 for 1-, 2-, 3-, 4-, and 5-year survival, respectively (Figure 1F).

Validation of the glycometabolism-related gene signature

The Kaplan-Meier curve indicated that patients of the high-risk group had a significantly shorter overall survival than patients of the low-risk group, and the overall survival was significantly decreased with increasing risk score (Figure 2A and 2B), in the testing cohort. The AUC for 1-, 2-, 3-, 4-, and 5-years survival was 0.687, 0.730, 0.729, 0.729, and 0.718, respectively (Figure 2C). The Kaplan-Meier curve showed fewer deaths among patients of the low-risk group than patients of the high-risk group, in the GSE39055 cohort (Figure 2D). The number of patients died rose as the risk score increased (Figure 2E). The AUC for 1-, 2-, 3-, 4-, and 5-years reached 0.907, 0.825, 0.743, 0.743, and 0.743, respectively (Figure 2F). These results indicated that the gene prognostic signature had a robust predictive performance.

Correlation analysis of risk score and clinical features

To assess the relationship between the risk score and clinical features, we profiled the risk score, gender, and age by \( t \)-test. The correlation between the risk score and clinical features in the training cohort is
shown in Figure 3A. Similarly, the correlation between the risk score and clinical features was also analyzed in the testing and validation cohort (Figure 3B and 3C). The results showed that age and gender were not correlated with the risk score in the training cohort (Table S1), and neither age nor gender was correlated with the risk score in the testing and validation cohort (Table S2 and 3).

**Independent prognostic value of glycometabolism-related gene signature**

To assess the independent prognostic value of the glycometabolism-related gene signature, univariate and multivariate Cox regression analyses were used on the risk score and clinical characteristics. Univariate analysis identified that the risk score was an independent factor affecting overall survival (Figure 4A). Then, multivariate analysis showed that the risk score was an independent prognostic factor (Figure 4B). To better predict the prognosis of OS patients in different years after diagnosis, a new nomogram was constructed based on the independent prognostic factors (Figure 4C). When the total score of the patient rose, the prognosis became poor. The C-index of the new risk model was 0.76, and the calibration curve was close to the ideal curve, indicating that the nomogram had a good predictive ability (Figure 4D). Furthermore, the decision curve analysis results indicated that the model can be applied (Figure 4E).

**Functional enrichment analysis**

We conducted GSEA to determine the biological functions of DEGs between the high- and low-risk groups. The results of GO enrichment analysis showed that the DEGs between the high- and low-risk groups were mainly enriched in pathways of the actin filament organization, actin polymerization or depolymerization, activation of innate immune response, adaptive immune response, antigen processing and presentation, and antigen receptor-mediated signaling pathway. The top 10 GO biological processes are shown in Figure 5A. Results from KEGG analysis showed that the DEGs were mainly involved in allograft rejection, antigen processing and presentation, autoimmune thyroid disease, cell adhesion molecule cams, chemokine signaling pathway, complement and coagulation cascades, cytokine-cytokine receptor interaction, graft-versus-host disease, hematopoietic cell lineage, and Leishmania infection pathway (Figure 5B). The pathway score of each sample was calculated by ssGSEA, and the score of the high- and low-risk groups was displayed in the heat map (Figure 5C and 5D).

**Immune cell analysis**

We next evaluated immune infiltration to describe their immune landscape, and an abundance of 28 immune-correlated cell populations were computed in the high- and low-risk groups using the ssGSEA (Figure 6A). Computation showed differential expression in immune cells between the high- and low-risk groups (Figure 6B). SEPHS2 was positively correlated with 14 immune cells, PRKACB was positively correlated with 17 immune cells, and GPX7 was negatively correlated with 11 immune cells (Figure 6C).

**Construction of a ceRNA network**
MiRNAs and lncRNAs correlated with the prognostic genes were screened out using the miRanda software. Then a glycometabolism-related ceRNA network was established by integrating lncRNA-miRNA interactions with miRNA-gene interactions. This network contained four prognostic genes, 148 miRNAs, and 91 lncRNAs (Figure 7).

**Discussion**

In this study, data acquired from the MSigDB, TARGET and GEO database were used for bioinformatics analysis. Four glycometabolism-related prognostic genes (PRKACB, SEPHS2, GPX7, PFKFB3) were identified using the univariate Cox regression analysis followed by LASSO regression analysis. A risk model was established based on these genes. Further analysis showed that the risk model had a good predictive performance and application value. Moreover, immune cell analysis indicated that three prognostic genes including SEPHS2, GPX7, and PFKFB3, were correlated with different proportions of immune cells. To reveal the upstream regulatory mechanisms of these four genes, a ceRNA network was created to show the targeting miRNAs and corresponding lncRNAs related to the prognostic genes. These results may shed new light on the underlying mechanisms of OS from the perspective of glycometabolism.

Glycometabolism in tumor cells is emerging as enhanced glucose uptake and aerobic glycolysis [8]. Aerobic glycolysis allows the conversion of glucose into pyruvate eventually contributing to the production of lactate. This energy metabolic reprogramming promotes energy generation and thus facilitates OS cells proliferation, invasion, and chemoresistance [19]. Several factors including glycolytic enzymes (e.g., GLUT1), oncogenes (e.g., KRT17), transcription factors (e.g., HIF1α), tumor suppressors (e.g., p53), and related signaling pathways have been reported to regulate the glycometabolism in OS cells and play pivotal roles in overall survival of patients [20–22]. These factors may help formulate therapeutic strategies.

This study identified four glycometabolism-related prognostic genes including PRKACB, SEPHS2, GPX7, and PFKFB3. The risk model based on the four prognostic genes showed a good predictive performance in OS patients. Previous studies have suggested that these glycometabolism-related genes are involved in tumorigenesis. Lower PRKACB expression was found in colorectal carcinoma and breast tumor tissues, and was significantly associated with unfavorable overall survival in patients with such tumors [23, 24]. It was shown that SEPHS2 was elevated in breast tumor samples, and SEPHS2 overexpression was correlated with malignancy. This suggests that SEPHS2 may serve as a prognostic marker and therapeutic target for patients with breast cancer [25, 26]. GPX7 is a member of the glutathione peroxidase (GPx) family with weak GPx activity [27]. GPX7 has been confirmed to inhibit tumorigenesis and function as a tumor suppressor [27–30]. This study investigates the prognostic value of these three genes in OS. Accumulated studies have demonstrated that PFKFB3, is a key regulator of glycolysis in tumorigenesis, angiogenesis chemoresistance, and tumor microenvironment [31]. Furthermore, several studies indicated that upregulating PFKFB3 accelerates cell growth and metastasis, making it a potential biomarker for OS [32–34]. Few studies have focused on the molecular mechanisms of these genes in the
pathogenesis of OS. However, their roles in regulating the glycometabolism of OS are warranted in further studies.

To identify more details about the genes mentioned, patients collected from TARGET were divided into high- and low-risk groups based on the median value of the risk score. GSEA was performed to investigate the biological functions of DEGs between groups. These DEGs were significantly enriched in multiple biological processes and signaling pathways. Both GO and KEGG analyses indicated that DEGs were significantly associated with the immune-related processes and pathways. The tumor immune microenvironment has been implicated in the occurrence and progression of OS [35]. Local tumor-infiltrating immune cells function as key regulators in OS cells growth and invasion [36]. In the present study, 26 immune cells were found differentially expressed between groups. Further analysis showed that three genes, including \textit{PRKACB}, \textit{SEPHS2}, and \textit{GPX7}, were significantly associated with different immune cells. The results indicated that the three genes may be involved in the tumorigenesis of OS by regulating the glycometabolism and immune microenvironment.

In order to analysis of the functions of these genes deeply, we constructed the lncRNA-miRNA network. CeRNA network is a prevalent form of post-transcriptional regulation of gene expression in mammals [37]. The mRNA, lncRNA, pseudogene, and circular RNA can affect the stability or translation activity of target RNAs by competitively binding to miRNA. An increasing number of studies have confirmed that ceRNA regulatory network is widely involved in the occurrence and development of OS [38; 39]. To further understand the upstream mechanisms of prognostic genes, a ceRNA was performed to identify the target miRNAs and the corresponding lncRNAs. The results showed that a total of 148 miRNAs and 91 lncRNAs were linked with these four prognostic genes. The ceRNA networks may facilitate a better understanding of the pathophysiologic process of OS.

**Conclusions**

In summary, our research identified a novel signature based on four glycometabolism-related genes and constructed a risk model that can predict the survival of patients with OS. This study provided new insights into the role of glycometabolism-related genes in the molecular mechanisms of OS and may help in the development of novel diagnostic and therapeutic methods.

**Declarations**

**Authors’ contributions**

This study was conceived by H Sun and FY Wang. FY Wang and K Yang performed the bioinformatic and statistical analysis. K Li and GX Peng participated in figures and charts drawing. H Yang supervised the implement of the current study. Data acquisition was accomplished by Y Xiang from online databases. The funding was provided by H Yang and H Sun. K Yang and FY Wang prepared the original draft. H Sun reviewed and edited the manuscript. The final version of the manuscript was approved by all authors.
Funding

The study was funded by the Science and Technology Fund of Guizhou Provincial Health Commission (gzwjky2020-1-120; gzwkj2021-261), Science and Technology Fund of Guizhou Science and Technology Department (QKH-ZK [2021] 391), and the Youth Fund cultivation program of National Natural Science Foundation of Affiliated Hospital of Guizhou Medical University (gyfynfsc-2021-12).

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors read and approved the final manuscript.

Disclosure Statement

The Authors declare that there is no conflict of interest.

References


Figures
Figure 1

Figure 2

Figure 3

Correlation analysis of risk score and clinical features. A, the correlation between risk score and clinical features in the training set. B, the correlation between risk score and clinical features in testing set. C, the correlation between risk score and clinical features in validation set.
Figure 4

**Figure 5**

Figure 6

**Immune cell analysis.** A, Heat map of immune-correlated cell populations between high-risk group and low-risk group. B, the expression of immune cells between high-risk group and low-risk group. C, the correlation between four glycometabolism related genes and immune cells.
Figure 7

**Construction of ceRNA network based on glycometabolism related genes.**

**Supplementary Files**

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

- TableS1.docx
- TableS2.docx
- TableS3.docx