Apoptosis-related genes-based prognostic signature for osteosarcoma

Fei Yang1, Yi Zhang1

1Department of Orthopaedics, Zibo Central Hospital, Zibo 255036, Shandong, China

Correspondence to: Yi Zhang; email: zhangyi_zbch@outlook.com, https://orcid.org/0000-0002-7956-0937

Keywords: osteosarcoma, risk score, apoptosis-related genes, prognosis

Received: February 10, 2022  Accepted: April 13, 2022  Published: May 3, 2022

Copyright: © 2022 Yang and Zhang. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Osteosarcoma (OS) is a common malignant primary tumor of skeleton, especially in children and adolescents, characterized by high lung metastasis rate. Apoptosis has been studied in various tumors, while the prognostic role of apoptosis-related genes in OS has been seldom studied. Three OS related datasets were downloaded from Gene Expression Omnibus (GEO) database. Univariate Cox and LASSO Cox regression analysis identified optimal genes, which were used for building prognostic Risk score. Subsequent multivariate Cox regression analysis and Kaplan-Meier survival analysis determined the independent prognostic factors for OS. The immune cell infiltration was analyzed in CIBERSORT. Basing on 680 apoptosis-related genes, the OS patients could be divided into 2 clusters with significantly different overall survival. Among which, 6 optimal genes were identified to construct Risk score. In both training set (GSE21257) and validation set (meta-GEO dataset), high risk OS patients had significantly worse overall survival compared with the low risk patients. Besides, high Risk score was an independent poor prognostic factor for OS with various ages or genders. Three immune cells were differentially infiltrated between high and low risk OS patients. In conclusion, a six-gene (TERT, TRAP1, DNM1L, BAG5, PLEKHF1 and PPP3CB) based prognostic Risk score signature is probably conducive to distinguish different prognosis of OS patients.

INTRODUCTION

Osteosarcoma (OS) is a common malignant primary tumor of skeleton, especially in children and adolescents, which is usually characterized by developing sarcoma cells to immature bone or osteoid tissue in metaphysis regions of long bones [1–3]. Along with the development of clinical practices, the overall survival rate of non-metastatic OS patients treated with specialized surgery and chemotherapy has been improved to 50%-70% [4]. Whereas, it has been estimated that there is already micro-metastasis at the diagnosis in more than 50% OS patients [5]. Meanwhile, high aggressiveness and lung metastasis rate of OS are still great challenges limiting survival rate of OS patients [6], leading to a poorer survival rate of approximately 20-30% [7–9]. Immunosuppressive feature of OS is also important in causing undesirable prognosis of patients [10]. Additionally, those OS patients at a similar clinical stage with same treatments might have different outcomes, due to the individual genetic and tumor heterogeneity [11, 12]. Accordingly, deepening mechanisms underlying the progression and prognosis of OS should be further explored, besides, more reliable and clinically meaningful diagnostic or prognostic signatures are urgently demanded for OS.

Increasing genes or signatures have been reported as biomarkers for OS, contributing to improve the outcome of OS patients. For example, targeting BMPR2 and HIF1-α has been indicated to promisingly prevent metastasis and progression in OS patients [7]. A recent study has revealed an immune-related genes based prognostic signature for OS patients, which could also predict the survival of OS patients [13]. Moreover, another risk signature-based on three metastasis-associated genes has been suggested to predict the prognosis of OS patients [14]. However, to the best of our knowledge, the prognostic role of apoptosis-related genes in OS has been seldom studied. Apoptosis, as a
way of programmed cell death, is catalyzed by numerous proteins’ proteolytic cleavage, especially affected by the enzymatic activity of effector caspases 3, 6 and 7, etc. [15, 16]. The apoptotic process comprises in multiple steps, including nuclear membrane breakdown, membrane blebbing, transformation genomic DNA into nucleosomal structures, and so on [17, 18]. Currently, some anti-OS drugs have been indicated to involve in the apoptosis. Apatinib has been evidenced to facilitate the apoptosis in OS patients via regulating VEGFR2/STAT3/BCL-2 signaling, and thereby inhibiting the growth of OS [19]. Moreover, Metformin might have antitumor potential in OS patients, through inducing apoptosis and autophagy via influencing ROS/JNK signaling [20]. Estrogen receptor β (ERβ) is widely involved in the apoptosis and autophagy, which might contribute to suppress the proliferation and metastasis of OS cells [21]. The overexpression of GRIM-19 has been proved to promote the radiation-induced apoptosis of OS cells [22]. Collectively, it will be of great importance to explore the prognostic role of apoptosis-related genes in OS, which would be conducive to those high risk OS patients’ personalized management.

In our study, we herein aimed to explore the prognostic role of apoptosis-related genes in OS patients, basing on the public OS data in Gene Expression Omnibus (GEO) database and the corresponding comprehensive bioinformatics analysis, attempting to construct a reliable prognostic signature for OS. Our findings are expected to give more conducive information for indirect improvement of the prognosis of OS patients.

MATERIALS AND METHODS

Data sources

We totally downloaded 3 datasets from the GEO database (Gene Expression Omnibus) (https://www.ncbi.nlm.nih.gov/geo/). In GSE21257, mRNA chip data of 53 OS patients was obtained, whose complete clinical survival information was shown in Table 1. Moreover, in GSE16091 and GSE39058 datasets, there were 34 and 41 OS samples and corresponding clinical information, respectively. The data in GSE16091 and GSE39058 datasets were analyzed on Affymetrix Human Genome U133A Array and Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip platforms, respectively. Besides, GSE16091 and GSE39058 were merged as meta-GEO dataset.

Cluster analysis

Those OS patients with complete survival information were clustered according to “K-mean” method, using R language (version 4.1.0, the same below).

LASSO Cox regression analysis

Next, the OS samples were subjected to the univariate Cox regression analysis based on the expression data to screen the prognosis related genes (significance threshold P <0.01). The LASSO Cox regression analysis was then conducted to optimize the prognosis related genes, using glmnet of R [23]. The optimal genes and the following formula were used for the Risk score calculation of all OS samples. In below formula, Coefi referred to the risk coefficient of each factor in the LASSO-Cox model, and Xi represented the gene expression value in our study. The best cutoff value of Risk score was determined by using survival, survminer and two-sided log-rank test of R. All OS samples were divided in high and low risk groups according to the cutoff value.

\[
\text{Risk Score} = \sum_{i=1}^{n} \text{Coef}_i \times X_i
\]

Survival analysis

The overall survival rate was estimated basing on Kaplan-Meier method in survival and survminer package of R. The significance of difference among various groups was tested by log-rank test. Multivariate Cox regression analysis was used to find all independent prognostic factors for OS patients.

Nomogram building

Nomogram is usually used to predict the prognosis of cancer. All independent prognostic factors were contained to build Nomogram to predict the 1-year, 2-year and 3-year overall survival of OS patients, utilizing rms package of R. The calibration curves were used to assess the prognostic predictive performance of Nomogram.

The relative proportion of immune cell infiltration

The relative infiltrated proportions of various immune cells in each OS sample were estimated using software CIBERSORT [24]. In CICERSORT, basing on gene expression matrix and preset 547 barcode genes, the immune infiltrating cell composition was characterized according to deconvolution algorithm. The estimated infiltrated proportions of every sample sum up to 1.

The expression of crucial immune checkpoints

The correlation between some key immune checkpoints’ expression (CTLA4, PDL1, LAG3, TDO2) and the Risk score in OS samples was studied. Besides, immune
Table 1. Clinicopathological characteristics of osteosarcoma patients from GSE21257 database.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patients(N=53)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NO.</td>
<td>%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
<td></td>
<td>35.85%</td>
</tr>
<tr>
<td>Male</td>
<td>34</td>
<td></td>
<td>64.15%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤16(Median)</td>
<td>25</td>
<td></td>
<td>47.17%</td>
</tr>
<tr>
<td>&gt;16(Median)</td>
<td>28</td>
<td></td>
<td>52.83%</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>13</td>
<td></td>
<td>24.53%</td>
</tr>
<tr>
<td>II</td>
<td>16</td>
<td></td>
<td>30.19%</td>
</tr>
<tr>
<td>III</td>
<td>13</td>
<td></td>
<td>24.53%</td>
</tr>
<tr>
<td>IV</td>
<td>5</td>
<td></td>
<td>9.43%</td>
</tr>
<tr>
<td>Unknown</td>
<td>6</td>
<td></td>
<td>11.32%</td>
</tr>
<tr>
<td>Survival Time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long(&gt;5 years)</td>
<td>6</td>
<td></td>
<td>11.32%</td>
</tr>
<tr>
<td>Short(&lt;5 years)</td>
<td>47</td>
<td></td>
<td>88.68%</td>
</tr>
<tr>
<td>OS status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>23</td>
<td></td>
<td>43.40%</td>
</tr>
<tr>
<td>Alive</td>
<td>30</td>
<td></td>
<td>56.60%</td>
</tr>
</tbody>
</table>

checkpoints’ expression was also compared between high and low risk OS patients.

Availability of data

All data generated and analyzed in this study are available from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database.

RESULTS

OS patients with different prognosis could be divided based on apoptosis related genes

Firstly, we have downloaded totally 680 apoptosis-related genes from GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) datasets in GSEA database (Supplementary Table 1). Then, basing on these 680 apoptosis-related genes, the OS samples in GSE21257 dataset were clustered according to “K-mean” method using R. Sum of the squared errors (SSE) results indicated that the optimal number of clusters was k=2 (Figure 1A), and all OS samples were clustered into 2 categories (Figure 1B). Kaplan-Meier survival analysis suggested that the OS patients in these two clusters had significantly different overall survival (Figure 1C).

Prognostic Risk score model for OS patients

Taking 680 apoptosis-related genes’ expression as continuous variables, the OS samples in GSE21257 dataset were then subjected to an univariate Cox regression analysis. All apoptosis-related genes’ Hazard ratio (HR) were calculated, among which 7 genes were screened with P value <0.01 (Figure 2A). The subsequent LASSO Cox regression analysis was conducted on these 7 genes, during which lowest lambda value was corresponding to the optimal number of genes (Figure 2B). The 6 optimal genes comprised TERT, TRAP1, DNM1L, BAG5, PLEKHF1 and PPP3CB.

Subsequently, in GSE21257 and meta-GEO (containing GSE16091 and GSE39058) datasets, the mean value and standard deviation (SD) of these 6 genes’ expression value were normalized to 0 and 1, respectively. Then the normalized expression was weighted with the regression coefficient in LASSO Cox regression analysis to establish a prognostic Risk score for OS patients, Risk Score = (0.19843870*TERT)+(0.61052614*TRAP1)+(0.09921504*DNM1L)+(0.19947814*BAG5)+(-0.30350396*PLEKHF1)+(-0.33883222*PPP3CB).

Accordingly, the Risk score of each OS patient was calculated, and the OS samples in GSE21257 and meta-GEO were divided into high and low Risk score groups basing on best cutoff value (-0.627). The survival analysis was then performed on high and low risk OS patients in GSE21257 and meta-GEO, the results of which suggested that in both two datasets, high risk OS patients had poorer overall survival compared with low risk patients (Figure 2C, 2D). Therefore, our data indicated that our Risk score, based on TERT, TRAP1, DNM1L, BAG5, PLEKHF1 and PPP3CB, had a relatively good prognostic predictive performance.

Independent prognostic factor for OS patients -- Risk score

A multivariate Cox regression analysis was performed, containing age, gender, stage and Risk Score, in order to
explore the independent prognostic factors for OS patients (Figure 3A). We found that the Risk score was still significantly correlated with overall survival of patients, besides high Risk score was a poor prognostic indicator for OS patients (HR=4.89, 95% CI: 2.622-9.1, P <0.001).

To further explore the prognostic value of Risk score in OS patients with different ages or genders, we have also conducted Kaplan-Meier survival analysis on different types of patients. The results suggested that in female (Figure 3B), male (Figure 3C), and >16 years old OS patients (Figure 3D), all high risk OS samples had significantly worse overall survival compared with low risk patients. In ≤16 years old patients, high risk OS samples generally tended to have poorer overall survival compared with low risk patients (Figure 3E). Our findings implied that Risk score was a promising independent prognostic factor for OS patients.

**Nomogram for OS patients**

Next, we built a Nomogram based on the independent factor Risk score (Figure 4A). For the OS samples, one line was drawn upwards to get the Points from Risk score, which were matched to the Total Points axis accordingly. Then a line drawn downwards the Total Points axis determined the 1-year, 2-year, and 3-year overall survival of OS patients. The calibration curves of predicted survival probability of OS patients relatively well matched the ideal curve (Figure 4B–4D). Our Nomogram displayed a good predictive performance, which also implied the reliability of the Risk score.

**Immune cell infiltration in high and low risk OS patients**

Basing on CIBERSORT method and LM22 feature matrix, various immune cells’ infiltration was compared

---

**Figure 1. OS patients with different prognosis could be divided basing on apoptosis-related genes.** (A) Elbow diagram indicated that the optimal number of clusters was k=2. (B) The cluster results of OS samples. (C) Kaplan Meier survival curve of OS patients in two clusters. P was calculated according to log-rank test.
between high and low risk OS patients. The immune cell infiltration results of 53 OS patients in GSE21257 dataset (except for 2 abnormal samples) were summarized in Figure 5A, and the differential infiltrating proportions indicated the individual OS patient’s characteristics. Most types of immune cells were differentially infiltrated between high and low risk OS patients (Figure 5B), among which T cells CD8, T cells follicular helper, and Macrophages M0 were significantly differentially infiltrated in high and low risk OS patients (Figure 5C). Whereas, there was a weak correlation among various immune cells’ infiltrating proportions (Figure 5D). The principal component analysis (PCA) based on three significantly differentially infiltrated immune cells suggested that the OS samples could be divided into 2 clusters (Figure 5E).

Correlation between key immune checkpoints and Risk score

Moreover, we analyzed the correlation between several crucial immune checkpoints and Risk score, including

![Image](image-url)

**Figure 2. Construction of prognostic Risk score for OS patients.** (A) Totally 7 genes were significantly related to the prognosis of OS. HR: Hazard ratio; 95% CI: 95% confidence interval. (B) The smallest lambda in LASSO Cox regression analysis was corresponding to the optimal number of genes. (C, D) Kaplan Meier survival curves of OS patients in GSE21257 and meta-GEO datasets, respectively. P was calculated based on log-rank test.
CTLA4, PDL1, LAG3, and TDO2. We found that Risk score was correlated with all these immune checkpoints (Figure 6A). CTLA4 and LAG3 were significantly differentially expressed between high and low risk OS patients (Figure 6B).

**DISCUSSION**

During the past decades, distant metastasis has limited the prognosis of OS patients [25]. Meanwhile, many studies have evidenced that the apoptosis promotion...
would inhibit the progression and attenuates the metastasis of OS [26, 27]. Consequently, we herein investigated the potential role of apoptosis-related genes in OS patients, via analyzing the mRNA data and clinical survival information in GEO database. We have firstly constructed and verified a six-gene based prognostic signature for OS patients.

Many factors have been revealed to regulate the cell apoptosis in OS, including miRNAs, proteins, compounds, and so on [28–31]. Considering the vital influence of apoptosis on various tumors, including OS, we firstly downloaded 680 apoptosis-related genes from GEO database. Based on these genes, the OS patients could be divided into 2 clusters with significantly different overall survival, implying the potential effects of apoptosis-related genes on the prognosis of OS.

Furthermore, using GSE21257, univariate Cox regression analysis determined 7 prognosis apoptosis-related genes, of which 6 optimal genes were identified. Thus, we have built the Risk score basing on TERT, TRAP1, DNM1L, BAG5, PLEKHF1 and PPP3CB. Our data indicated that in both training set (GSE21257) and validation set (meta-GEO dataset), high risk OS patients had significantly worse overall survival compared with the low risk patients. Not only that, subsequent multivariate Cox regression analysis and Kaplan-Meier survival analysis suggested that high Risk score was an independent poor prognostic factor for OS with various ages or genders.

Meanwhile, some clues have been found to support our prognostic signature based on these 6 genes. TERT (telomerase reverse transcriptase) encodes a catalytic subunit of telomerase, which is influenced by multiple

![Figure 4. Nomogram could predict the survival probability of OS patients. (A) Nomogram could predict the 1-year, 2-year, and 3-year overall survival of OS patients. (B–D) The calibration curves of predicted survival probability of OS patients at 1, 2, and 3 years, separately. X-axis: predicted survival probability; Y-axis: actual survival probability.](image-url)
Figure 5. Immune cell infiltration in high and low risk OS patients. (A) The relative proportions of immune infiltrating cells in OS patients. (B) Differentially infiltrated immune cells between high and low risk OS patients. P was calculated with Wilcoxon method. (C) Three types of immune cells significantly differentially infiltrated. (D) Correlation matrix of various immune cells’ infiltrating proportions. Red: positive correlation; blue: negative correlation. (E) The results of PCA.
genetic and epigenetic regulations in many cancers [32]. Moreover, TERT has been evidenced to suppress the cisplatin-induced apoptosis in OS cells after cisplatin treatment, which might be a target for overcoming drug resistance [33]. TERT promoter mutations are associated with poorer overall survival of thyroid cancer [34], while the prognostic role of TERT has been seldom reported in OS. Additionally, TRAP1 (TNF receptor associated protein 1) has been indicated to regulated the apoptosis in lung cancer [35] and thyroid carcinoma cells [36], while TRAP1’s role in the prognosis of OS is firstly revealed in our study as far as we know. Whether TRAP1 could regulate the apoptosis and thereby affect the prognosis in OS in a similar way is still unknown, which deserves further exploration. DNM1L has been recently reported as an important prognostic factor for gastric adenocarcinoma [37], involving invasion and apoptosis, which is partly in line with our findings in OS. Despite few studies of PPP3CB in OS, PPP3CB has been evidenced to correlate with the poor prognosis of neuroblastoma [38], while PPP3CB might have conducive effects on pancreatic cancer [39]. Our data provided more insights into PPP3CB in OS, deserving more exploration. As for BAG5 and PLEKHF1, they have not been investigated in OS up to now, while their role in some other cancers have been documented. For example, BAG5 might involve in the invasion of papillary thyroid cancer cells [40]. Besides, BAG5 is indicated to be a suppressor in pancreatic cancer [41]. More details of BAG5 and PLEKHF1 in OS need to be clarified. Collectively, most genes in our prognostic Risk score could be supported by previous studies directly or indirectly. Besides, the good predictive effects of Nomogram based on Risk score also indicated that our prognostic signature was relatively reliable.

Additionally, immunosuppressive microenvironment in OS tissues has been demonstrated basing on immune and stromal cell type information [42]. In our study, three types of immune cells were found to be differentially infiltrated between high and low risk OS patients, comprising T cells CD8, T cells follicular helper, and Macrophages M0, which probably contributed to the different prognosis of OS patients. Among which, stimulating CD8+ cells might inhibit the development of OS [43]. Macrophages M0, M2 have been demonstrated as the most principal infiltrating immune cells in OS [44], which was compatible with our findings. In short, the complex interaction between OS and these immune cell remains to be unclear, which should be studied in the future.

CONCLUSIONS

To summarize, we have firstly revealed the prognostic role of apoptosis-related genes in OS patients. A six genes-based (TERT, TRAP1, DNM1L, BAG5, PLEKHF1 and PPP3CB) prognostic Risk score signature is probably conducive to distinguish different prognosis of OS patients, which is a promising prognosis prediction alternative for the clinical management.

Figure 6. Correlation between key immune checkpoints and Risk score. (A) Risk score was correlated with immune checkpoints, CTLA4, PD1, LAG3, and TDO2. (B) CTLA4 and LAG3 were significantly differentially expressed between high and low risk OS patients. P was calculated with Wilcoxon method.
AUTHOR CONTRIBUTIONS

All authors contributed to all work of the study including conception and design, material preparation, data collection and analysis and manuscript preparation. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

REFERENCES


SUPPLEMENTARY MATERIALS

Supplementary Table

Please browse Full Text version to see the data of Supplementary Table 1.

Supplementary Table 1. Apoptosis_GeneID.