Development and Validation of a Robust Apoptosis-Related Prognostic Classifier in Patients with Osteosarcoma

Zhifeng Zhang, Yi Wang, Fengmei Chen, Zhengmao Guan*, Yinquan Zhang*

Department of orthopedics, Third Affiliated Hospital of Naval Medical University, Shanghai 200011, PR China

Corresponding author
Y Zhang, Z Guan, Department of orthopedics, Third Affiliated Hospital of Naval Medical University, Shanghai 200011, PR China

Submitted: 03 Jun 2022; Accepted: 14 Jun 2022; Published: 23 Jul 2022


Abstract

**Background:** Apoptosis plays an important role in the tumorigenesis and the development of osteosarcoma, but the reliable biomarkers for individual treatment and prognosis of osteosarcoma based on apoptosis is lacking.

**Methods:** A total of 1476 apoptosis-related genes were extracted from pathways and biological processes associated with apoptosis downloaded from MSigDB. All of those genes were used to identify the prognosis-related genes by univariate cox regression in the TARGET dataset and the ARS was constructed using the LASSO regression. The performance of the classifier was verified in the training and validation groups. The infiltration of immune cells and the expression levels of the immune checkpoint in different groups were also analyzed. Finally, a nomogram based on ARS and other Clinicopathological factors was constructed to facilitate clinical application.

**Results:** ARS containing 22 apoptosis-related genes were identified, and its predictive ability performed well in both the training and validation groups. Macrophages M1 were highly expressed in the low-score group, and NK cells resting was highly expressed in the high-score group. The samples with low-score had higher expression of CTLA4 and PDL1. A nomogram with excellent predictive effectiveness (AUC= 0.932, 0.984, 0.939, 0.939, 0.948) was constructed to facilitate clinical decision-making.

**Conclusion:** A prognostic classifier based on 22 apoptosis-related genes and a nomogram were constructed to predict the overall survival of patients with osteosarcoma. The classifier also provides a reference for selecting suitable patients for immunotherapy and targeted therapy.

Keywords: Osteosarcoma; Apoptosis; Prognosis; Immune; TARGET

Introduction

Osteosarcoma is the prevailing primary bone tumor with highly malignant, which is tend to occur in teenagers. The 5-year survival rate is between 50 percent and 60 percent due to local progression and early metastasis [1]. The therapeutic effect of osteosarcoma patients depends mainly upon the time of diagnosis. However, the progression of osteosarcoma is often severe at the time of diagnosed due to the lack of specific symptoms in early stages[2,3]. New biomarkers for osteosarcoma can not only conduce to early screening, diagnosis and predicting prognosis, but also provide a new perspective for the treatment of osteosarcoma.

Apoptosis is one of the main regulative ways of cell death and it can be influenced by genes or directly affected by cytokines. Inhibition of apoptosis is associated with tumorgenesis while activation of apoptosis can be used as a way of oncotherapy [4, 5]. For example, Chen et al has proved that apoptosis is related to the occurrence of gastric cancer, and Wahba et al has proved that apoptosis is related to the treatment of epithelial ovarian cancer [6,7]. There are many studies focus on the relationship between osteosarcoma and apoptosis by single gene or gene family[8-10]. Comprehensive studies of multiple pathways and biological processes associated with apoptosis are still lacking.

In this study, all genes were extracted form 153 apoptosis-related pathways and biological processes, and the datasets of TARGET and GEO (https://www.ncbi.nlm.nih.gov/gds/) were used to construct and validate ARS. ESTIMATE algorithm and CIBERSROT were performed to further investigated the relationship among ARS, immunity and tumor composition. In addition, the expression levels of the immune checkpoint in different ARS groups were also be analyzed. Finally, a nomogram was constructed and tested to facilitate clinical application. Our find-
ings put forward a new perspective for treatment and exploring the potential genes of apoptosis in osteosarcoma.

**Materials and Methods**

**Data Acquisition and Processing**

FPKM and clinicopathological information of 89 samples with osteosarcoma were downloaded from TARGET datasets of UCSC Xena (http://xena.ucsc.edu/; Accessed 6 November 2021) [12]. The samples without survival information were removed. Ultimately, the training cohort with 84 osteosarcoma samples and corresponding clinicopathological and prognostic information were brought into the analysis.

The validation cohort with 53 osteosarcoma samples was derived from microarray dataset GSE21257 of GEO DataSets. This dataset was produced by Illumina human-6 v2.0 expression beadchip (using nulDs as identifier) and contained survival information of each sample.

The data processing is shown in Figure 1. All microarray and RNA-seq data included in our study were normalized and log2 transformed.

**Apoptosis-Related Pathways and Biological Processes**

153 apoptosis-related pathways or biological processes were extracted from hallmark gene sets (H collection), curated gene sets (C2 collection), and ontology gene sets (C5 collection) in MSigDB (http://software.broadinstitute.org/gsea/index.jsp) [13, 14]. All these pathways, biological processes, and their corresponding genes are fused into Table S1.

**Signature Establishment in Training Cohort**

All genes in apoptosis-related pathways or biological processes were extracted and prognosis-related genes were filtered by the univariate Cox regression with the same criterion of P-value < 0.05. Then, LASSO regression model was performed to find the most robust prognosis-related markers. ARS of each sample was established by the formula:

\[ ARS = \sum_i \text{Coefficient(mRNA}_i\text{)} \times \text{Expression(mRNA}_i\text{)} \]

**The Power of ARS in Different Cohorts**

The training cohort was split into a high-score group and a low-score group with the median value of ARS. The survival state and ARS in different group are depicted by risk factor curve,
survival status scatter plot, and Kaplan-Meier curve. To verify the predictive capacity of ARS in the training cohort, AUC of 3- and 5-years were calculated and plotted separately using the survivalROC package. And then, the reliability and applicability of ARS were further validated in the validation cohort.

Relationship between ARS and Clinicopathological Features
To explore the relationship between ARS and different clinicopathological features containing age, sex, location and metastasis, the Kaplan-Meier curve was plotted to evaluate the predictive capacity of ARS in each subgroup based on the median value.

ESTIMATE, Immune Profile and Immune Checkpoint Molecules
To estimate the proportion of stromal and immune cells in each sample, ESTIMATE algorithm (https://bioinformatics.mdander son.org/estimate/) was used to calculate the stromal score and immune score respectively [15]. To determine the depth of immune cell infiltration, the CIBERSORT package was performed to calculate the expression of 22 immune cells in different ARS groups [16]. Besides, the expression levels of the immune checkpoint (PD-1, PD-L1, and CTLA4) in different ARS groups were compared.

Construction of a Prognostic Nomogram
All TARGET osteosarcoma samples were used to establish the nomogram. The univariate and multivariate Cox regression analyses were performed to select independent risk factors from ARS and clinicopathological features by the criterion of P-value < 0.05. The nomogram was constructed by the “RMS” package and the stability of the nomogram performance was evaluated by the calibration curve and iROC.

Statistical Analysis
All statistical analyses were performed using the R software (www.r-project.org). Clinicopathological characteristics were compared within the training group and validation group using the Chi-square test, Fisher exact probability test, and Student’s t test.

Results

Data Processing
The expression data and clinicopathological information of samples in training cohort and validation cohort were downloaded respectively. The clinicopathological characteristics of the two cohorts are detailed in Table 1. All 153 apoptosis-related pathways or biological processes were downloaded and corresponding 1807 nonredundant genes were used to take intersection with training and validation cohorts. Ultimately, 1476 mutual genes were identified.

Table 1: characteristics of training and validation cohorts

<table>
<thead>
<tr>
<th>characteristics</th>
<th>TARGET (n=84)</th>
<th>GSE21257 (n=53)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>age(mean±sd)</td>
<td>14.99±4.82</td>
<td>18.71±12.20</td>
<td>0.013</td>
</tr>
<tr>
<td>sex(%)</td>
<td></td>
<td></td>
<td>0.440</td>
</tr>
<tr>
<td>Female</td>
<td>37(44.0)</td>
<td>19(35.8)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>47(56.0)</td>
<td>34(64.2)</td>
<td></td>
</tr>
<tr>
<td>metastasis state(%)</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>metastasis</td>
<td>63(75.0)</td>
<td>0(0.0)</td>
<td></td>
</tr>
<tr>
<td>non-metastasis</td>
<td>21(25.0)</td>
<td>0(0.0)</td>
<td></td>
</tr>
<tr>
<td>location(%)</td>
<td></td>
<td></td>
<td>0.145</td>
</tr>
<tr>
<td>foot</td>
<td>76(90.5)</td>
<td>44(83.0)</td>
<td></td>
</tr>
<tr>
<td>hand</td>
<td>6(7.1)</td>
<td>8(15.1)</td>
<td></td>
</tr>
<tr>
<td>pelvis</td>
<td>2(2.4)</td>
<td>0(0.0)</td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>0(0.0)</td>
<td>1(1.9)</td>
<td></td>
</tr>
</tbody>
</table>

The Construction of ARS
Univariate Cox regression analysis was performed for 1476 mutual genes, and 180 genes with P-value < 0.05 in univariate analysis were included in the LASSO regression analysis (Table S2) to construct the ARS. The ten-fold cross-validation was performed to determine the penalty parameter (λ) of the model. A total of 22 genes (RPS6, IFITM3, GRN, ATF4, MYC, DYNLL2, G6PD, BNIP3, PTGIS, BCL10, TRIM32, MAGEA3, PDK2, NMNAT1, EN1, TNFRSF11B, PPARG, SYNGAP1, GAL, GRIK2, MCF2, TERT) were included in the LASSO model (Figure 2). The coefficient of each gene was shown in Table S3.
The Power of ARS in Different Cohorts

In training cohort, the high-score group exhibited worse overall survival data as compared to the low-score group, with P-value < 0.01 (Figure 3E). Meanwhile, the distribution patterns of risk scores and survival status were plotted (Figure 3A and 3C). With the increasing ARS score, the overall survival time decreased and mortality increased. The 3- and 5-years AUC (AUC = 0.937, 0.947) of ARS was shown in Figure 3G, calculated by the “survivalROC” package.

These results indicate that the higher ARS represent the worse prognosis, and ARS has a good ability to predict the prognosis of osteosarcoma patients. The same result was also shown in the validation cohort. We calculate the ARS in validation cohort according to the LASSO formula. The high-score group exhibited worse overall survival as compared with the low-score group (P-value < 0.05; Figure 3F). The trend of survival time and mortality was the same as in the training cohort (Figure 3B and 3D). We found that the 3- and 5-years AUC (AUC = 0.771, 0.737) of ARS could still accurately predict survival state of osteosarcoma patients in validation cohort (Figure 3H).
**Figure 3:** The performance of ARS in different cohorts. (A, B) The distribution of ARS in training and validation cohorts. (C, D) The distributions of overall survival status, overall survival and risk score in training and validation cohorts. (E, F) Kaplan-Meier curves for the overall survival of the high- and low-score groups in training and validation cohorts. (G, H) AUC plots showed that ARS was an accurate variable for survival prediction in training and validation cohorts.

**Relationship Between ARS and Clinicopathological Features**

To explore the relationship between ARS and clinicopathological features including age, sex, location and metastasis, we plotted the Kaplan-Meier curve in a different subgroup of clinicopathological features. As shown in Figure 4, the high-score group exhibited worse overall survival compared with the low-score group in each subgroup except in location of hand.
**Figure 4:** The performance of ARS in different clinicopathological including age, gender, location and metastasis state subgroups.

**ESTIMATE, Immune Profile and Immune Checkpoint Molecules**

The ESTIMATE algorithm revealed the immune (-1439.104 to 2560.632) and stromal scores (-689.9769 to 1927.611) of the training cohort (Figure 5A). It can be seen from the heatmap that there are significant differences in tumor purity, immune score and stromal score between the two groups. To further elucidate the relationship between ARS and immune/stromal score, both of the scores in ARS subgroups were compared. The Figure 5B-C shows that both immune scores and stromal scores in the low-score group were higher than that in the high-score group. This verified that immune score and ratio of tumor to stroma were significantly associated with ARS.
Figure 5: (A) A heat map of the estimate results. (B, C) The expression levels of immune score and stromal score in high- and low-score groups of ARS.

CIBERSORT package was performed to calculate the expression of 22 immune cells in different ARS groups. The result showed that NK cells resting is significantly more prevalent in the high-score group, while Macrophages M1 is significantly more prevalent in the low-score group (Figure 6).

To analyze the expression levels of immune checkpoint proteins in ARS subgroup, the expression of PD1, PDL1, and CTLA4 was calculated. The samples with low-score had higher expression of CTLA4 and PDL1 (Figure 6B-C), suggesting that patients with low-score ARS may respond better to immune checkpoint inhibitors targeting CTLA4 and PDL1.
Figure 6: (A) Violin plot showed the different proportions of tumor-infiltrating cells between high- and low-score groups. (B-D) The expression levels of immune checkpoint molecules in high- and low-score groups.

Construction of a Prognostic Nomogram

The clinicopathological features and ARS of training cohort were used to construct a prognostic nomogram. Ultimately, the metastasis state and ARS were identified as the independent factors for prognosis of osteosarcoma according to univariate and multivariate cox regression analysis by the criterion of P-value < 0.05. A prognosis-related nomogram was constructed by the “rms” package with independent factors (Figure 7C). The tAUC of clinicopathological features, ARS and nomogram were shown in Figure 7B. The figure shown that the prediction ability of ARS (AUC = 0.887, 0.976, 0.937, 0.937, 0.947) and nomogram (AUC = 0.932, 0.984, 0.939, 0.939, 0.948) was higher than all other clinicopathological features. Then, the calibration curves of 1-, 3- and 5-year also shown that the prediction ability of the nomogram was very stable in different time points (Figure 7D-F).
Figure 7: (A) tROC analysis showed that ARS was an accurate variable for survival prediction in training cohort. (B) Univariate and multivariate cox regression analysis indicated that ARS was the independent risk factor among various features. (C) A nomogram was constructed to quantify risk assessment for individual patients. (D-F) Calibration analysis indicated a high accuracy of survival prediction in 1-, 3- and 5-year.
Discussion

Although the therapeutic effect of surgery combined with chemotherapy is satisfactory in most patients with osteosarcoma, the treatment regimen did not significantly improve overall survival in those patients with metastasis or recurrence and the overall 5-year survival rate of those patients is only 20 percent [17-19]. It has been proved that preoperative response to chemotherapy can predict the overall survival of patients, but its accuracy can be further improved [20]. It is still of great significance to find new genes as therapeutic targets and prognostic indicators. Currently, biotherapy is a main direction of tumor therapy, and inducing tumor cell apoptosis is the foundation for biotherapy to achieve therapeutic effect. There are also many potential mechanisms of apoptosis, such as death receptor-dependent pathway, mitochondrial-dependent pathway and caspase-independent apoptosis [21].

In this study, 153 apoptosis-related pathways and biological processes and corresponding genes were used to explore the role of apoptosis in osteosarcoma.

In order to meet the requirements of subsequent analysis, the TARGET cohort was used to construct ARS consisting of 22 apoptosis-related genes by the univariate cox regression and LASSO regression. Next, we found that ARS had a very good predictive accuracy for overall survival of osteosarcoma in both training and validation cohorts by using the Kaplan-Meier curve and AUC.

For the sake of illustrating the applicability of ARS in different situations, subgroup analysis was conducted. The performance of ARS was very stable in each subgroup of age, sex, metastasis and location, except for the hand subgroup. The small sample size of the subgroup may seriously affect the results of Kaplan-Meier curve. Nevertheless, the above results can be used to demonstrate that ARS has the ability to distinguish the prognosis of patients with osteosarcoma in different subgroups.

Recently, great progress has been made in immunotherapy. Sipuleucel-t has been proved to treat castration-resistant prostate cancer, and the HPV vaccine can prevent and treat infections caused by the HPV virus as well as cancers induced by it [22,23]. ESTIMATE algorithm was performed to explore the relationship among ARS, immune infiltration and proportion of stroma in osteosarcoma. We found a correlation between ARS and immune infiltration as well as proportion of stroma in osteosarcoma. CIBERSORT was used to further analyze which immune cell infiltrates were different in different ARS group. The result showed that NK cells resting is significantly more prevalent in the high-score group, while Macrophages M1 is significantly more prevalent in the low-score group. Natural killer (NK) cells have significant capability in tumor immune-surveillance but the penetration of NK cells in tumor is a huge obstacle for cancer immunotherapy. Verhoeven et al had shown that NK cells can recognize and lyse Ewing sarcoma cells through NKG2D and DNAM-1 receptor dependent pathways. In addition, cetuximab can enhance the the cytolytic activity of resting NK cells in osteosarcoma[24-26]. These results suggest that NK cells may contribute to anticancer activity in osteosarcoma with high-score ARS. Macrophages play an important role in cancer development and metastasis and Macrophages M1 also can phagocytose tumor cells. It can be used as drug carriers for tumor therapy because it can anchor tumor cells. Cerussimo et al had shown that increased infiltration of macrophages M2 is associated with the metastasis and prognosis of osteosarcoma[27,28]. According to the results of our study, Macrophages M1 may contribute to anticancer activity in osteosarcoma with low-score ARS.

We also analyzed the relationship between immune checkpoint inhibitors and ARS, and the results showed that the expression of CTLA4 and PD-L1 was higher in the low-score group, indicating that patients with low-score ARS may benefit from the treatment of immune checkpoint inhibitors.

Among the 22 genes of ARS, some genes, such as IFITM3, MYC, DYNLL2 and G6PD, had been proved associated with the occurrence and development of osteosarcoma while other genes, such as RPS6, GRN, GRN and PTGS2, has not been studied in osteosarcoma and further research is needed [29-33].

Several limitations exist in this study. On the one hand, the number of osteosarcoma samples is not enough in both the training and validation groups, and larger-scare samples are needed to verify the stability of the model. On the other hand, the relationship between ARS and immune cells or immune checkpoints requires further experimental verification.

Conclusion

In conclusion, a prognostic classifier based on 22 apoptosis-related genes and a nomogram was constructed to predict the overall survival of patients with osteosarcoma. The 22 genes of ARS provide a new direction for exploring the mechanisms of apoptosis in osteosarcoma. ARS also provides a reference for selecting suitable patients for immunotherapy and targeted therapy.

Declarations

Ethics approval and consent to participate

The study is the bioinformatics analysis article. The data in our study were obtained from the free online databases and the ethical approval was not necessary. There are no animal and human experiments involved in the study. There are no human subjects in the article and informed consent was not necessary.

Consent for publication

Not applicable.

Availability of data and materials


Competing interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
Not applicable.

Authors’ contributions
Yinquan Zhang and Zhifeng Zhang performed the study conceptualization. Zhifeng Zhang and Yi Wang performed the data collection and analysis. Fengmei Chen performed writing the original manuscript draft. Zhifeng Zhang contributed to the manuscript review and editing. Yi Wang contributed to the drawing pictures. Zhifeng Zhang, Yi Wang and Fengmei Chen contributed equally to this work.

Acknowledgements
Not applicable.

Abbreviations
FPKM: The fragments per kilobase per million mapped reads; TARGET: Therapeutically Applicable Research to Generate Effective Treatments; GEO: Gene Expression Omnibus; MSigDB: Molecular Signatures Database; ARS: apoptosis-related risk score; LASSO: Least Absolute Shrinkage and Selection Operator; AUC: area under the curve; tROC: time-dependent receiver operating characteristic; ICI: immune checkpoint inhibitors; PD1: programmed cell death protein 1; PD1L: programmed death-ligand 1; CTLA4: cytotoxic T-lymphocyte-associated protein 4.

References
22. Markowitz, L. E., Liu, G., Hariri, S., Steinau, M., Dunne,


