



Identification and Development of a Novel 4-Gene Immune-Related Signature to Predict Osteosarcoma Prognosis

Mingde Cao^{1,2}, Junhui Zhang¹, Hualiang Xu¹, Zhujian Lin¹, Hong Chang¹, Yuchen Wang¹, Xusheng Huang¹, Xiang Chen¹, Hua Wang¹ and Yancheng Song^{1*}

¹ Department of Orthopedics, The First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou, China,

² Department of Orthopaedics and Traumatology, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong

OPEN ACCESS

Edited by:

Pier Paolo Piccaluga,
University of Bologna, Italy

Reviewed by:

Ramya Sivakumar,
University of Washington,
United States
Jacopo Junio Valerio Branca,
University of Florence, Italy

*Correspondence:

Yancheng Song
songyancheng@21cn.com

Specialty section:

This article was submitted to
Molecular Diagnostics and
Therapeutics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 20 September 2020

Accepted: 24 November 2020

Published: 23 December 2020

Citation:

Cao M, Zhang J, Xu H, Lin Z,
Chang H, Wang Y, Huang X, Chen X,
Wang H and Song Y (2020)
Identification and Development of a
Novel 4-Gene Immune-Related
Signature to Predict Osteosarcoma
Prognosis.
Front. Mol. Biosci. 7:608368.
doi: 10.3389/fmolb.2020.608368

Osteosarcoma (OS) is a malignant disease that develops rapidly and is associated with poor prognosis. Immunotherapy may provide new insights into clinical treatment strategies for OS. The purpose of this study was to identify immune-related genes that could predict OS prognosis. The gene expression profiles and clinical data of 84 OS patients were obtained from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database. According to non-negative matrix factorization, two molecular subtypes of immune-related genes, C1 and C2, were acquired, and 597 differentially expressed genes between C1 and C2 were identified. Univariate Cox analysis was performed to get 14 genes associated with survival, and 4 genes (*GJA5*, *APBB1IP*, *NPC2*, and *FKBP11*) obtained through least absolute shrinkage and selection operator (LASSO)-Cox regression were used to construct a 4-gene signature as a prognostic risk model. The results showed that high *FKBP11* expression was correlated with high risk (a risk factor), and that high *GJA5*, *APBB1IP*, or *NPC2* expression was associated with low risk (protective factors). The testing cohort and entire TARGET cohort were used for internal verification, and the independent GSE21257 cohort was used for external validation. The study suggested that the model we constructed was reliable and performed well in predicting OS risk. The functional enrichment of the signature was studied through gene set enrichment analysis, and it was found that the risk score was related to the immune pathway. In summary, our comprehensive study found that the 4-gene signature could be used to predict OS prognosis, and new biomarkers of great significance for understanding the therapeutic targets of OS were identified.

Keywords: gene signature, prognosis, osteosarcoma, immunotherapy, bioinformatics analysis

INTRODUCTION

Osteosarcoma (OS) is the prevailing primary malignant tumor of the bone, and it is most likely to occur in children and adolescents, with 80–90% of OS tumors occurring in the metaphysis of the long tubular bones, distal femur, proximal tibia, and proximal humerus (Jackson et al., 2016). OS is highly malignant and develops rapidly, and it is likely to metastasize to the lung

in the early stage (Pan et al., 2019). The average 5-year survival rate of those with local OS is about 80%, but the outcome for those with metastatic disease is much worse (Mirabello et al., 2009). Surgery, chemotherapy, and radiotherapy developments play an extremely important role in reducing the rate of lung metastasis and improving the long-term survival rate of patients with OS, making the 5-year survival rate reach 60–80% (Jawad et al., 2011). Targeted therapy and immunotherapy may provide new opportunities for comprehensive OS treatment. In recent years, the discovery of immune checkpoints has pushed cancer immunotherapy to a new level, with specific blocking of immunosuppression effects and enhanced anti-tumor immune responses. As an emerging treatment, immunotherapy has shown promising results for OS (Miwa et al., 2019). However, there are still many problems to be solved in immunotherapy for OS, especially in terms of predicting immunotherapy biomarkers and identifying new effective therapeutic targets.

Immunotherapies, particularly those based on immune checkpoint inhibitors, have shown promising activity in a variety of tumors, with clinically significant improvements in response rates, progression-free survival, and overall survival in lung cancer, head and neck cancer, and bladder cancer (Chen et al., 2017). In many malignancies, innate and adaptive immune cells play a role in the tumor microenvironment, communicating with antigen-presenting cells, such as natural killer cells, macrophages, and dendritic cells, as well as lymphocytes, thereby allowing for effective tumor control (Gajewski et al., 2013). Recently, immunotherapy has shown promise in a variety of cancers, but its application in OS remains unexplored. It has been suggested that OS may be sensitive to immunotherapies. The percentage of CD8+ infiltrating lymphocytes in OS is higher than in other sarcoma subtypes (van Erp et al., 2017), and the degree of infiltration is positively correlated with survival (Gomez-Brouchet et al., 2017). Some studies have shown that the use of immune checkpoints can be a promising treatment for OS. High PD-L1 levels have been observed in patients with OS, and PD-L1 levels have been associated with tumor-infiltrating lymphocyte levels (Shen et al., 2014). In addition, the median overall survival of patients with low PD-L1 levels was shown to be longer than that of patients with high PD-L1 levels. A recent study reported that PD-1 levels in peripheral blood CD4+ and CD8+ T cells in patients with OS were high and that PD-1 levels in CD4+ T cells in patients with metastasis were significantly higher than in patients without metastasis, and PD-1 and PD-L1 levels were found to be negatively correlated with prognosis (Zheng et al., 2015). OS has high levels of genomic instability, and some tumors express PD-L1, suggesting their potential sensitivity to inhibitors of the PD-1/PD-L1 axis (Hacohen et al., 2013; Kansara et al., 2014; Mouw et al., 2017). This all suggests that PD-1 is involved in the pathogenesis and progression of OS. Further, a prospective, randomized phase III trial showed that muramyl tripeptide phosphatidyl-ethanolamine, which is an investigational agent only available in clinical trials, could activate monocytes, and macrophages to improve tumor control. The addition of muramyl tripeptide to chemotherapy was shown to increase the 6-year overall survival rate of patients with OS from 70 to 78%, and EFS tended to improve (Meyers et al., 2008).

In summary, several biological characteristics of OS suggest that regulation of the immune response may bring benefits, and a better understanding of OS may provide new insights for immunotherapy in terms of antibody targeting of cell surface proteins, tumor vaccines using dendritic cells, oncolytic viruses, adoptive cell therapy, and checkpoint inhibitors (Wedekind et al., 2018). It is hoped that immunotherapy will lead to a breakthrough in OS therapy.

To increase immunotherapy efficacy, it is especially pivotal to determine immune-associated prognostic biomarkers. OS is characterized by high heterogeneity of somatic copy number changes and structural rearrangement, suggesting that genes included in somatic copy number changes are the key drivers of cancer (Sayles et al., 2019). By focusing on genes that have an underlying role, we were able to reduce the complexity of the OS genomic pattern to identify the changes most likely to be relevant to clinical practice. This suggested the need for genomic information to perform targeted therapy. Here, we combined the expression data derived from multiple OS genes from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and Gene Expression Omnibus (GEO) databases, and we used the non-negative matrix factorization (NMF) method to perform molecular subtyping based on immune-related genes. Further, we used a robust likelihood-based survival model to develop a 4-gene signature to predict OS prognosis. A prognostic risk score model was established, and internal and external verifications were performed. Our study may provide new biomarkers for predicting the therapeutic efficacy of immunotherapy and OS prognosis.

MATERIALS AND METHODS

Data Sources and Preprocessing

A total of 84 OS samples with RNA sequencing (RNA-Seq) data and corresponding clinical follow-up information were downloaded from the TARGET database (<https://ocg.cancer.gov/programs/target>), and the GSE21257 cohort containing 53 OS samples with prognostic information was downloaded from the GEO database (www.ncbi.nlm.nih.gov/geo).

The patients collected from the TARGET database were used for the expression profiles sequenced by RNA-Seq, while the GSE21257 cohort was used for chip data (the platform was the Illumina Human-6 v2.0 Expression BeadChip). In the probe design process, some genes were not designed to be detected as probes. For the RNA-Seq data, different genomic backgrounds could be different, so the genes detected by the different data platforms could also be somewhat different.

In addition, 864 genes were collected from the immune-related literature (Nirmal et al., 2018). We preprocessed the TARGET dataset to filter for immune-related genes with low expression. The filter criterion was to remove genes with expression <1, which accounted for more than 50% of all samples. A total of 60 genes were removed from the TARGET data due to low expression. The details of those genes are given in **Supplementary Table 1**. The clinical features of the samples are provided in **Table 1**.

TABLE 1 | Sample clinical features.

Clinical features	TARGET-OS	GSE21257
Status		
Censored	55	30
Dead	29	23
Gender		
Male	47	34
Female	37	19
Metastatic		
Yes	21	34
No	63	19
Age		
≤15	46	21
>15	38	32

Identification of Molecular Subtypes Based on Immune-Related Genes

We extracted the expression profiles of 804 immune-related genes from the TARGET database, then univariate Cox analysis was performed using the `coxph` function to acquire prognosis-related genes. NMF was applied to cluster the OS samples, with the standard “brunet” and 50 iterations. We set the number of clusters, k , from 2 to 10, and we determined the average silhouette width of the membership matrix by using the NMF package in R, with the minimum member of each subclass set to 10. Kaplan–Meier (KM) survival curves were drawn to compare prognoses between the molecular subtypes. The `limma` package in R was used to identify the differentially expressed genes (DEGs) between the different molecular subtypes, and the `ClusterProfiler` package was used to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the DEGs.

Construction of the Prognostic Risk Model Random Grouping of Samples

The expression profiles of 597 DEGs between the molecular subtypes were retained. A total of 84 OS samples from the TARGET database were divided into a training cohort and a testing cohort. To avoid random allocation bias from affecting the stability of subsequent modeling, all samples were randomly grouped with replacement 100 times in advance. Group sampling was based on a training cohort-to-verification cohort ratio of 0.5:0.5.

Univariate Cox Analysis of the Training Cohort

Univariate Cox proportional hazards regression was performed for each immune-related gene in the training cohort, and the `coxph` function of the survival package in R was performed with $P < 0.01$ as the threshold for filtering.

LASSO-Cox Analysis

To reduce the number of genes in the risk model, we performed least absolute shrinkage and selection operator

(LASSO) regression. LASSO is a penalized regression method that reduces overfitting by performing shrinkage and model selection simultaneously. A more refined model is obtained by constructing a penalty function, which can compress some coefficients and set others equal to 0. Therefore, it retains the advantage of subset contraction. Further, it is a biased estimate for processing data with multicollinearity, which can realize the selection of variables while estimating parameters and better solve the multicollinearity problem in regression analysis.

Relationships Between Risk Scores and Pathways

To observe the relationships between the risk scores of the different samples and biological functions, we selected the gene expression profiles corresponding to these samples. The GSVA package in R was applied for single-sample gene set enrichment analysis (ssGSEA). We calculated the score of each sample for different functions, thus obtaining the ssGSEA score of each sample corresponding to each function. Further, we observed the correlations between these functions and risk scores.

RESULTS

Flow Chart of Analysis

We designed a protocol for the analysis of OS subtypes and construction of the prognostic model. The analysis process was carried out in strict accordance with the protocol (Figure 1).

Identification of Molecular Subtypes Based on Immune-Related Genes

Univariate Cox analysis was used to obtain 142 genes related to prognosis ($P < 0.05$; Supplementary Table 2). According to indicators such as cophenetic, dispersion, and silhouette in the NMF method, $k=2$ was chosen as the optimal number of clusters (Figures 2A,B). At the same time, cluster analysis was conducted on the expression profiles of these 142 genes (Figure 2C). Compared with the Cluster 2 (C2) subtype, most genes were highly expressed in the Cluster 1 (C1) subtype.

Prognosis, Immune Score, and Biological Functional Analysis Between Subtypes

KM survival curves were drawn to compare the survival of the two molecular subtypes. The results showed that C2 had worse prognosis than C1, for both overall survival and progression-free survival ($P < 0.05$; Figures 3A,B).

To determine the relationships of the immune and stromal scores between the two molecular subtypes, we estimated the immune and stromal scores of each sample by using the R software package and then made comparisons. The results showed that there were significant differences in stromal scores, immune scores, and ESTIMATE scores between C1 and C2, and the immune score of C1 was significantly higher than that of C2 ($P < 1e-5$; Figures 3C–E).

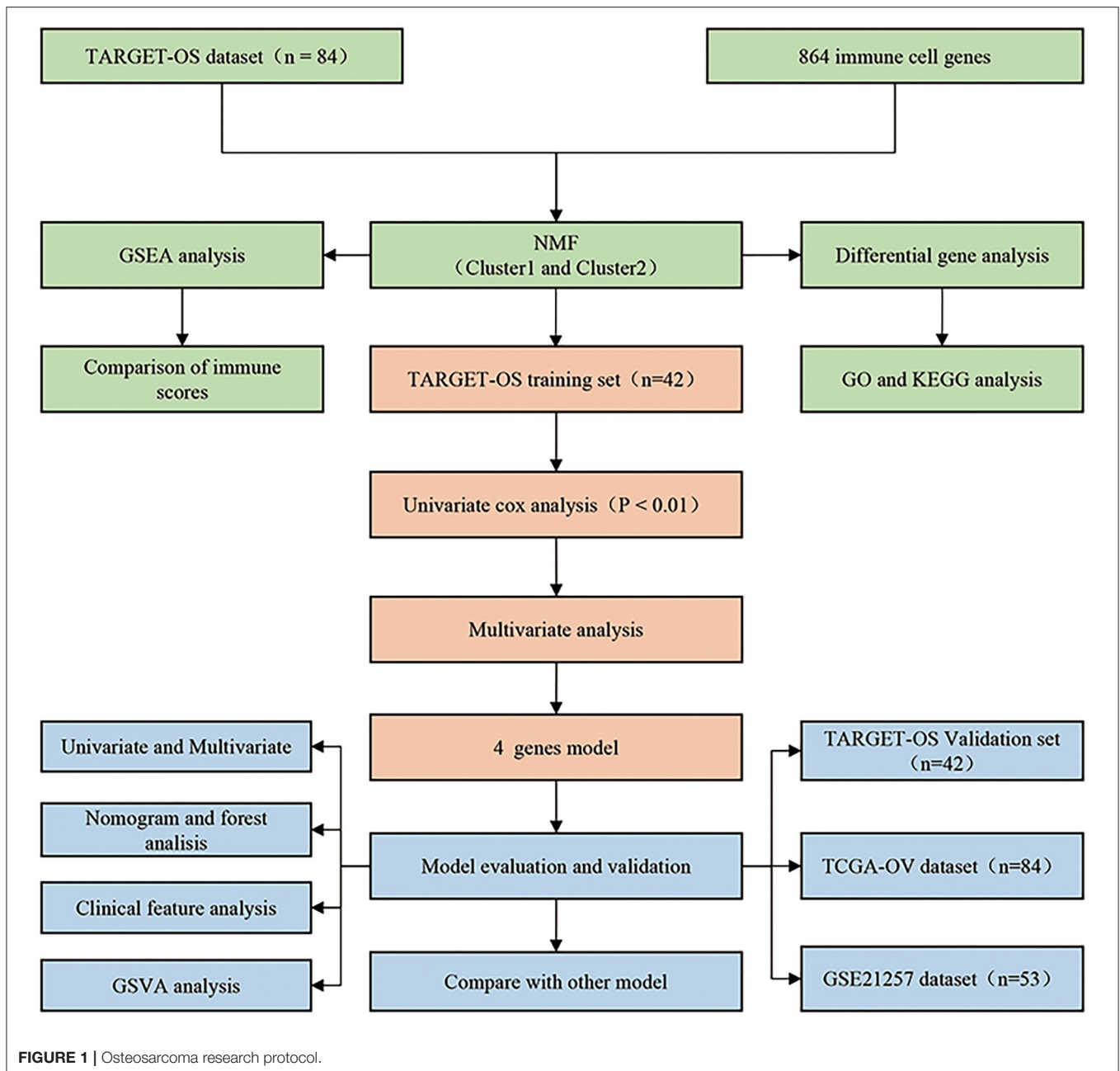


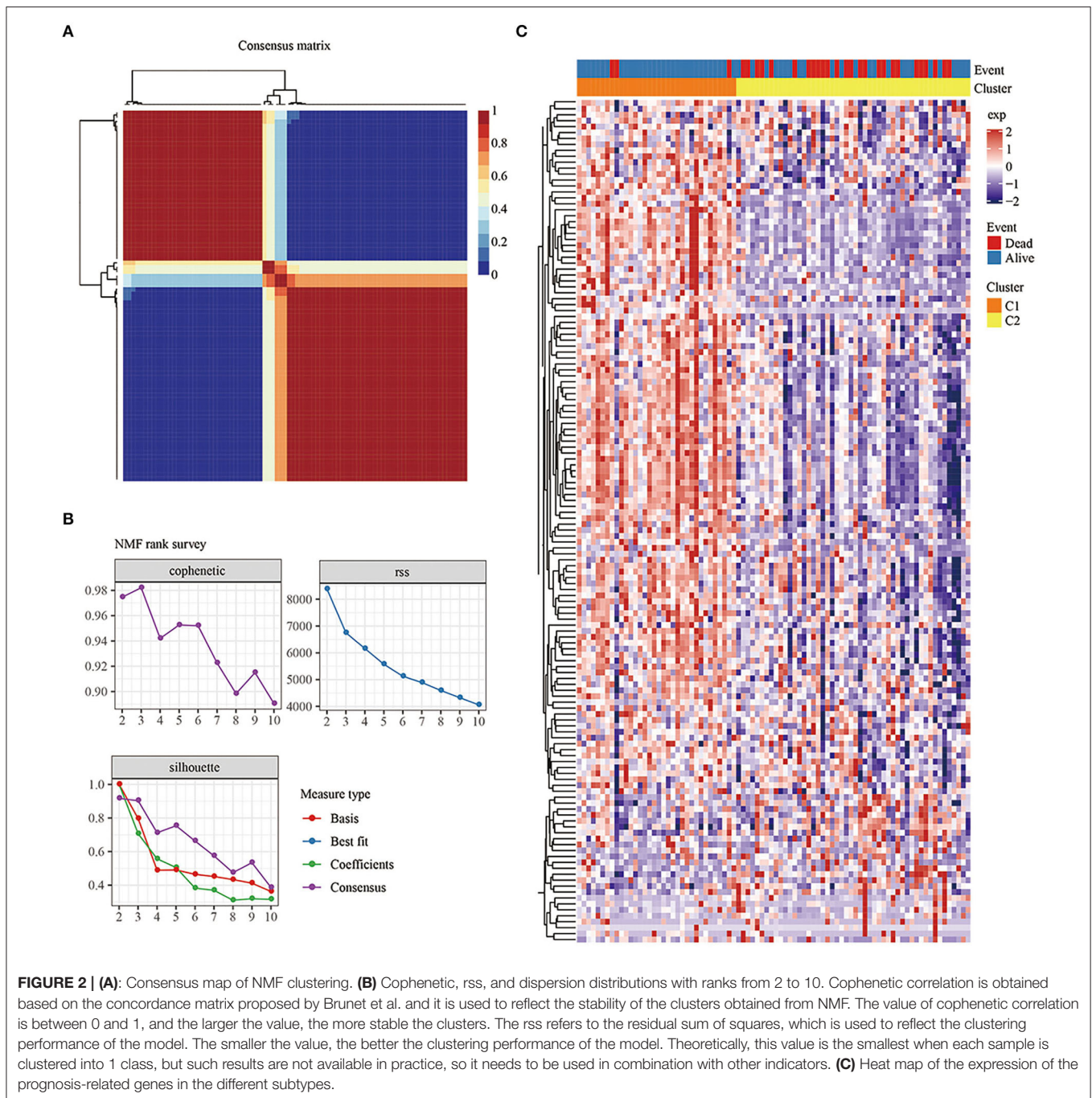
FIGURE 1 | Osteosarcoma research protocol.

GSEA was performed to analyze the significantly enriched pathways in C1 and C2. More immune-related pathways were enriched in C1, including KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY, KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY, KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY, and KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY, suggesting that C1 had a close relationship with immunity (**Supplementary Figure 1**).

Compared with C2, there were 597 DEGs, including 552 upregulated genes and 45 downregulated genes, in C1. The results suggested that the C1 and C2

subtypes were dominated by upregulated DEGs (**Supplementary Table 3**).

Further, KEGG pathway analysis of the DEGs between the subtypes of the TARGET data was performed by using ClusterProfiler (v3.14.0) in R. For the upregulated DEGs, a total of 62 pathways were annotated, among which 46 had a significant difference (false discovery rate <0.05). The top 10 results are shown in **Figure 3F**. Natural killer cell-mediated cytotoxicity, the toll-like receptor signaling pathway, the T cell receptor signaling pathway, the B cell receptor signaling pathway, and other immune-related pathways were significantly enriched (**Supplementary Table 4**).



Similarly, seven pathways were annotated among the downregulated DEGs shown in **Figure 3G**. The HIF-1 signaling pathway, pathways in cancer, and central carbon metabolism in cancer were enriched (**Supplementary Table 5**).

Construction of The Prognostic Risk Model Based on Immune-Related Genes

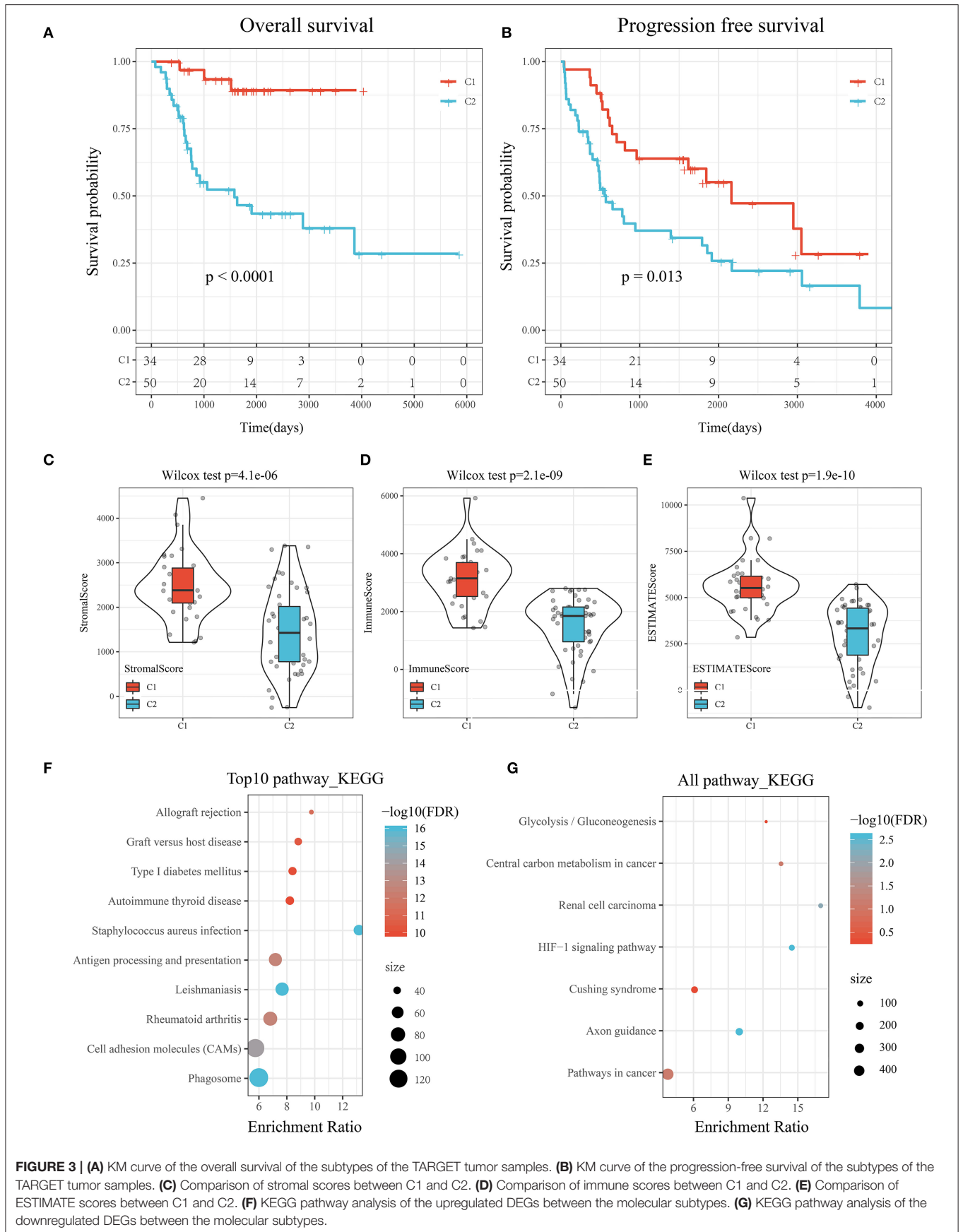
Random Grouping of Samples and Univariate Cox Analysis

Group sampling was based on a training cohort-to-verification cohort ratio of 0.5:0.5. There were 42 training cohort samples

and 42 testing cohort samples (**Supplementary Table 6**). Univariate Cox proportional hazards regression was performed for each immune-related gene. The results showed that there were 14 prognostic genes in the training cohort (**Supplementary Table 7**).

Construction of the Gene Signature by LASSO-Cox Analysis

Fourteen was considered a large number of prognostic genes and inappropriate for the clinical context, so LASSO-Cox regression analysis was performed to further narrow the range of immune-related genes while maintaining a high accuracy rate. First, the



change trajectory of each independent variable was analyzed (**Figure 4A**). The results showed that the number of independent variable coefficients approaching 0 increased with the gradual increase of lambda. A 10-fold cross-validation was used to construct the model. The confidence interval of each lambda is shown in **Figure 4B**. A total of 8 genes were selected as the target genes for subsequent analysis when the model reached the optimal value, with lambda equal to 0.1030521.

To further reduce the number of genes, the Akaike information criterion (AIC) was used for stepwise regression. AIC takes into account a model's goodness of fit and its simplicity in terms of the number of parameters needed to achieve this fit. The stepAIC function of the MASS package in R started with the most complex model and successively deleted variables in turn to reduce the AIC. The smaller the value, the better the model. It gave the model sufficient fit with fewer parameters. Finally, a total of 4 genes, *GJA5*, *APBB1IP*, *NPC2*, and *FKBP11*, was obtained.

The prognostic KM curves of the 4 genes are shown in **Figures 5A–D**. *APBB1IP* and *NPC2* could divide the TARGET training cohort samples into high- and low-risk groups with prognostic significance ($P < 0.05$). *FKBP11* had marginal prognostic significance ($P = 0.05152$), and *GJA5* had no prognostic significance ($P > 0.05$).

Evaluation of the Risk Model

The 4 genes were used to construct the risk model, as follows: $\text{RiskScore} = -0.873 * \text{GJA5} - 1.016 * \text{APBB1IP} - 1.192 * \text{NPC2} + 0.915 * \text{FKBP11}$. The risk score of each sample according to expression was calculated and shown as the risk score distribution (**Figure 6A**). The overall survival of the samples with high risk scores was significantly lower than that of the samples with low risk scores, suggesting that samples with high risk scores had worse prognoses. It was determined that the high expression of *FKBP11* was correlated with high risk (a risk factor), while the high expression of *GJA5*, *APBB1IP*, or *NPC2* was associated with low risk (protective factors). Further, receiver operating characteristic (ROC) analysis was conducted by using the timeROC package in R for the prognostic classification of risk scores. We analyzed the classification efficiency of prognosis prediction at 2, 3, and 5 years. The average area under the curve (AUC) reached 0.93 (**Figure 6B**). Finally, the Z-score method was applied to the preprocessing of risk scores. Samples with a risk score > 0 were put into the high-risk group, while those with a risk score < 0 were put into the low-risk group. A KM curve between the 2 risk groups was plotted (**Figure 6C**). There was a very significant difference ($P < 0.0001$), with 18 samples classified as high risk and 24 as low risk.

Robustness Verification of The Risk Model

Risk Model Validation in the Testing Cohort and Entire TARGET Cohort

To determine the robustness of the model, we used the same model and the same coefficient in the training cohort for the testing cohort and entire TARGET cohort. We calculated the risk score of each sample according to expression and then plotted the risk score distribution.

The risk score distribution of the testing cohort is shown in **Figure 7A**. The overall survival of the samples with high risk scores was significantly lower than that of the samples with low risk scores, suggesting that the samples with high risk scores had worse prognoses. Regarding changes in the expression of the 4 signature genes with increased risk scores, it was determined that the high expression of *FKBP11* was correlated with high risk (a risk factor). The high expression of *GJA5*, *APBB1IP*, or *NPC2* was associated with low risk (protective factors). This was consistent with the TARGET training cohort. The average AUC reached 0.78 (**Figure 7B**). The KM curve showed a very significant difference ($P = 0.019$), with 24 samples classified as high risk and 18 as low risk (**Figure 7C**).

The risk score distribution of the entire TARGET cohort is shown in **Figure 8A**. It was consistent with the performance of the TARGET training cohort. The average AUC reached 0.88. Prognoses in the high-risk group were significantly worse than those in the low-risk group ($P < 0.001$; **Figure 8B**). The KM curve showed a very significant difference ($P < 0.0001$), with 43 samples classified as high risk and 41 as low risk (**Figure 8C**).

Risk Model Validation in the External Cohort

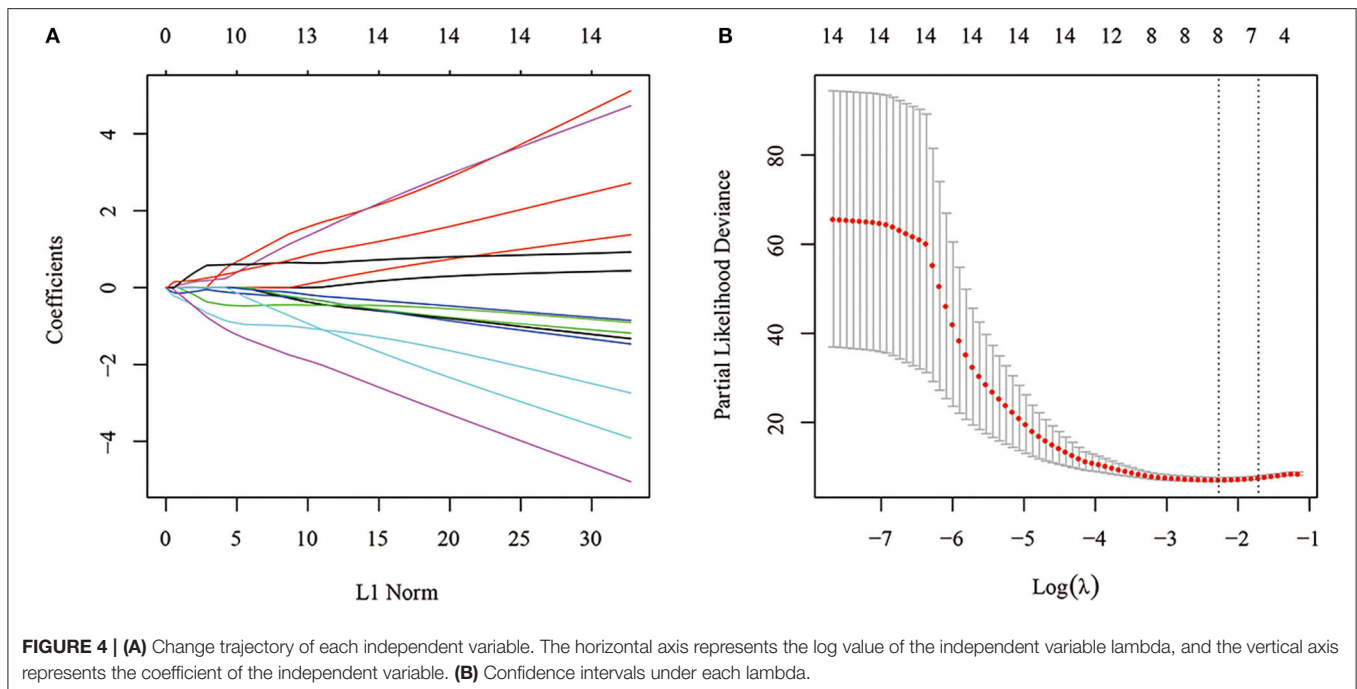
To determine the robustness of the model, we used the same model and the same coefficient in the training cohort for the external testing cohort. The risk score distribution of the GSE21257 testing cohort is shown in **Figure 9A**. The results were consistent with those of the TARGET training cohort. The average AUC reached 0.7 (**Figure 9B**). Prognoses in the high-risk group were worse than those in the low-risk group. The KM curve showed a very significant difference ($P = 0.026$), with 27 samples classified as high risk and 24 as low risk (**Figure 9C**).

Prognostic Analysis of the Risk Score Model and Clinical Features

We performed subgroup survival analysis of the clinical variables based on risk scores and found that the 4-gene signature divided the samples into the high- and low-risk groups appropriately in terms of age, sex, and metastasis ($P < 0.01$; **Figure 10**). The results suggested that the risk score model had good predictive ability in terms of different clinical features.

Relationships Between Risk Scores and Pathways

To observe the relationships between the risk scores of the different samples and biological functions, ssGSEA was performed by using the GSVA package in R. The functions with correlation coefficients > 0.35 were selected (**Figure 11A**). A total of 26 pathways were negatively correlated with risk score. The KEGG pathways used for cluster analysis based on enrichment scores are shown in **Figure 11B**. Some of the 26 pathways negatively correlated with increased risk score, including KEGG_JAK_STAT_SIGNALING_PATHWAY, KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY, KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY, and KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY.



Univariate and Multivariate Cox Analyses of the 4-Gene Signature

Univariate and multivariate Cox regression analyses were used to determine the independence of the 4-gene signature model in clinical applications (Table 2). Univariate Cox regression analysis results showed that risk score was significantly correlated with prognosis; at the same time, multivariate Cox analysis results showed that it was an independent prognostic risk factor (hazard ratio = 2.9103, 95% confidence interval = 1.8364–4.6120, $P = 5.4e-06$). Forest maps can simply and intuitively display the statistical summary results of different factors. The multivariate Cox results in a forest diagram are shown in Supplementary Figure 2. The results suggested that the 4-gene signature model had good predictive performance in clinical applications.

Comparison of the Risk Score Model With Other Models

Three OS prognosis-related risk models were identified by reviewing the literature, including a 19-gene signature (Goh et al., 2019), an 8-gene signature (Zhang et al., 2019), and a 3-gene signature (Shi et al., 2020). To make the models comparable, the same method was conducted to calculate the risk score of each OS sample in the TARGET cohort based on the corresponding genes in the 4 models. The Z-score method was applied to the preprocessing of risk scores, and samples with a risk score >0 were put into the high-risk group, while those with a risk score <0 were put into the low-risk group. The ROC and KM curves of the 3 comparison models are shown in Figure 12. The AUCs of these 3 models were lower than that of our 4-gene risk score model at 2, 3, and 5 years. Our 4-gene risk score model included a reasonable number of genes and had better performance. In

addition, the 3 comparison models could significantly divide samples into high- and low-risk groups ($P < 0.05$).

Epigenetic Modification of the 4 Genes

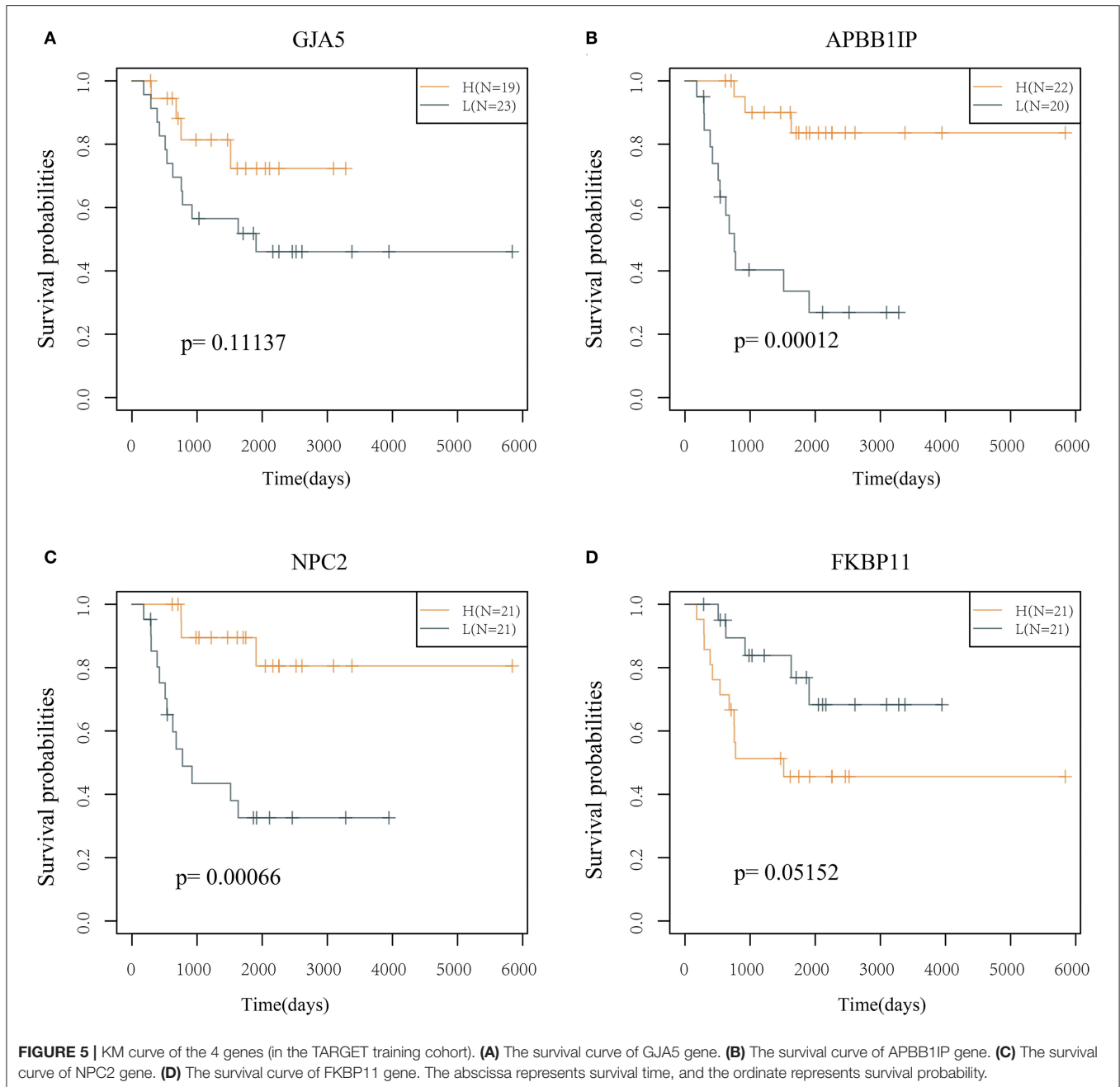
We analyzed the methylation of the 4 genes in the high- and low-risk groups, and the results showed that the level of methylation of *APBB1IP* in the high-risk group was higher than in the low-risk group, and the risk score heat map showed that in the high-risk group, *APBB1IP* was lowly expressed. The level of methylation of *FKBP11* in the low-risk group was higher than in the high-risk group, and the risk score heat map showed that *FKBP11* was lowly expressed in the low-risk group (Supplementary Figure 3).

Thus, the degree of methylation had a negative correlation with gene expression. The expression of *APBB1IP* and *FKBP11* was found to be regulated by methylation to a certain extent. There was no significant relationship between the expression of *GJA5* or *NPC2* and the level of methylation.

DISCUSSION

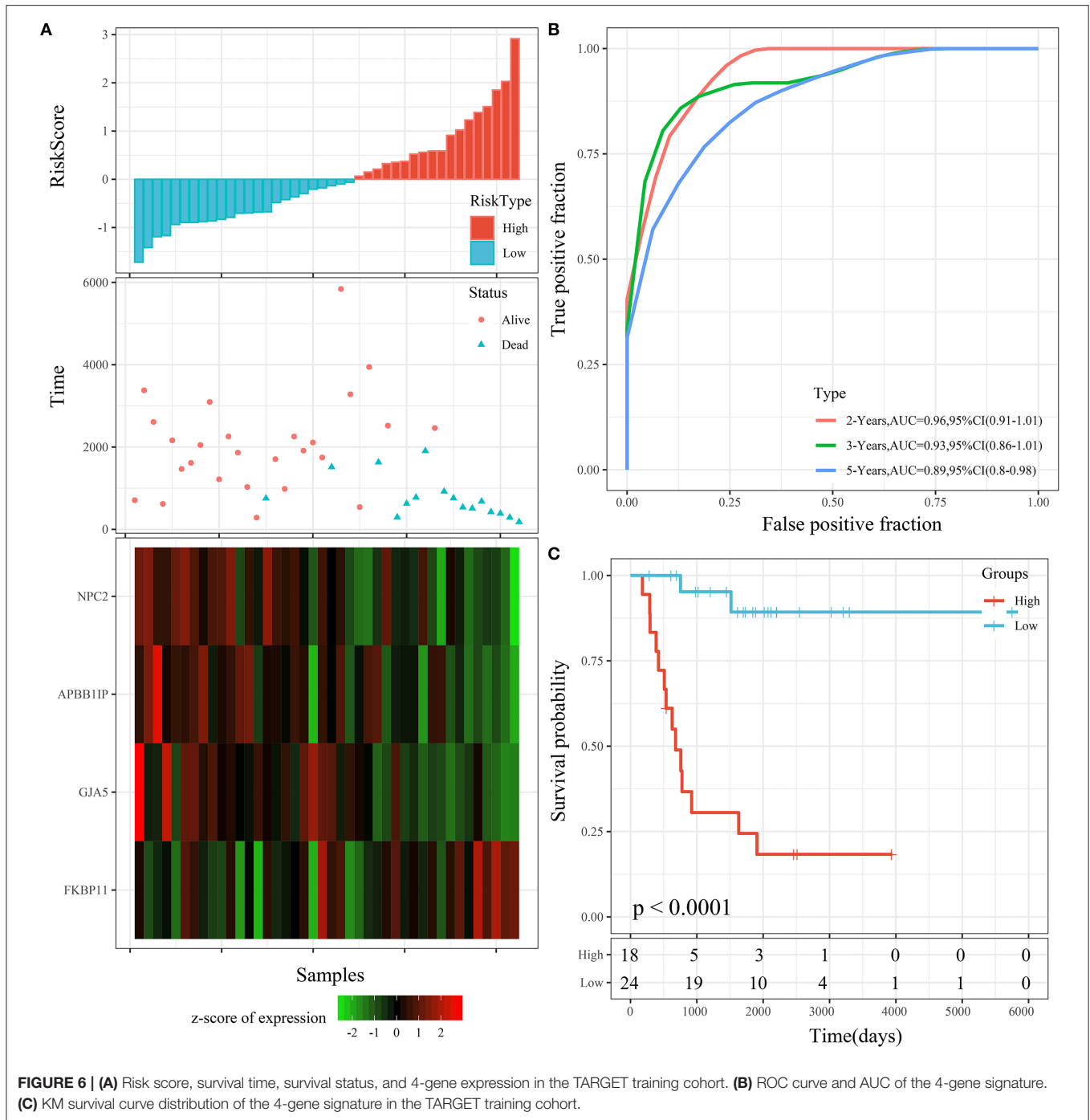
The treatment and prognosis of OS have not changed much in the past 30 years. Due to the poor prognosis of OS, it is necessary to determine prognostic biomarkers for OS patients in order to then use appropriate treatment strategies, including immunotherapy. We conducted a comprehensive study to establish a 4-gene signature, which was verified through internal and external testing cohorts. An immune-related risk model was established to explore the potential association between OS risk score and survival and to provide new biomarkers for OS prognosis.

In our study, the gene expression profiles and clinical data of 84 OS samples were obtained from the TARGET database, and 42 samples were randomly selected as the training cohort. Using



the NMF algorithm to cluster molecular subtypes, two molecular subtypes (C1 and C2) were obtained, and 597 DEGs between C1 and C2 (including 552 upregulated genes and 45 downregulated genes) were identified. Functional and pathway enrichment analysis results showed that immune-related pathways, such as natural killer cell-mediated cytotoxicity, the toll-like receptor signaling pathway, the T cell receptor signaling pathway, the B cell receptor signaling pathway, and other immune-related pathways, were significantly enriched in the C1 subtype, suggesting that C1 was closely related to immunity. Moreover, the immune score of C1 was significantly higher than that of C2, and the prognosis of C1 was better than that of C2. Further,

the results of the KEGG pathway analysis of the DEGs between the subtypes of the TARGET cohort showed that some immune-related pathways had a significant difference. Specifically, the HIF-1 signaling pathway was enriched. HIF-1 α promotes tumor cell growth, migration, and invasion in OS through activation of the AKT/cyclin D1 signal cascade (Zhang et al., 2018). In recent years, immunotherapy has become a novel and effective treatment method, and it has been used in the treatment of various tumors, including OS (Lettieri et al., 2016; Wang et al., 2016; Thanindratarn et al., 2019). However, only a few patients receiving immunotherapy have responded to this treatment due to the immunosuppressive microenvironment (de Visser et al.,

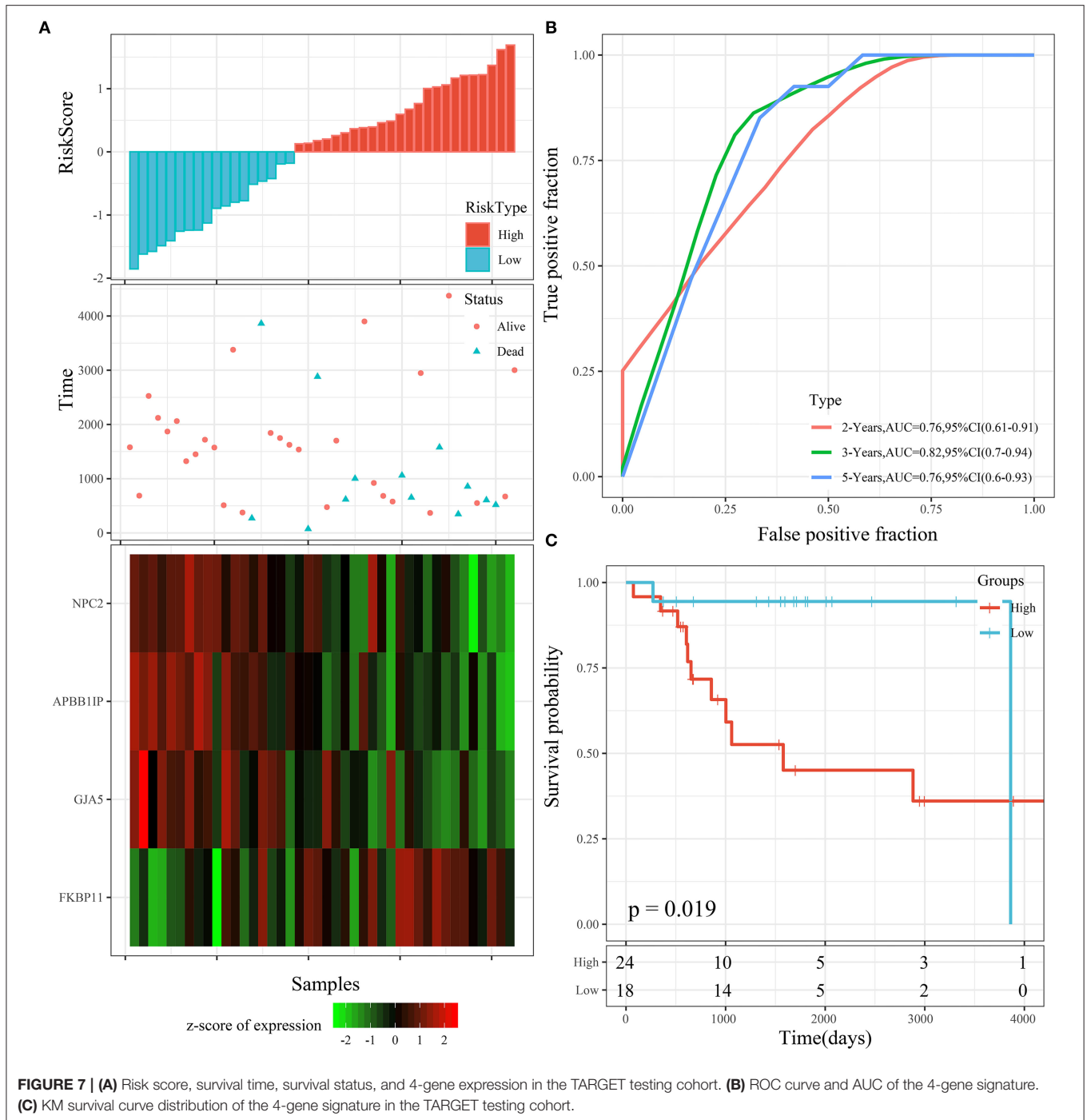


2006). Therefore, it is necessary to study the biomarkers that enable the prediction of the benefits of immunotherapy, which could aid clinical decision-making for individualized treatment.

A total of 14 genes associated with survival were obtained through univariate Cox analysis, and 8 genes were selected as target genes by LASSO regression. Further, the AIC was used to obtain 4 genes (*GJA5*, *APBB1IP*, *NPC2*, and *FKBP11*) to include in a 4-gene signature as a prognostic risk model. The results showed that high *FKBP11* expression was correlated with

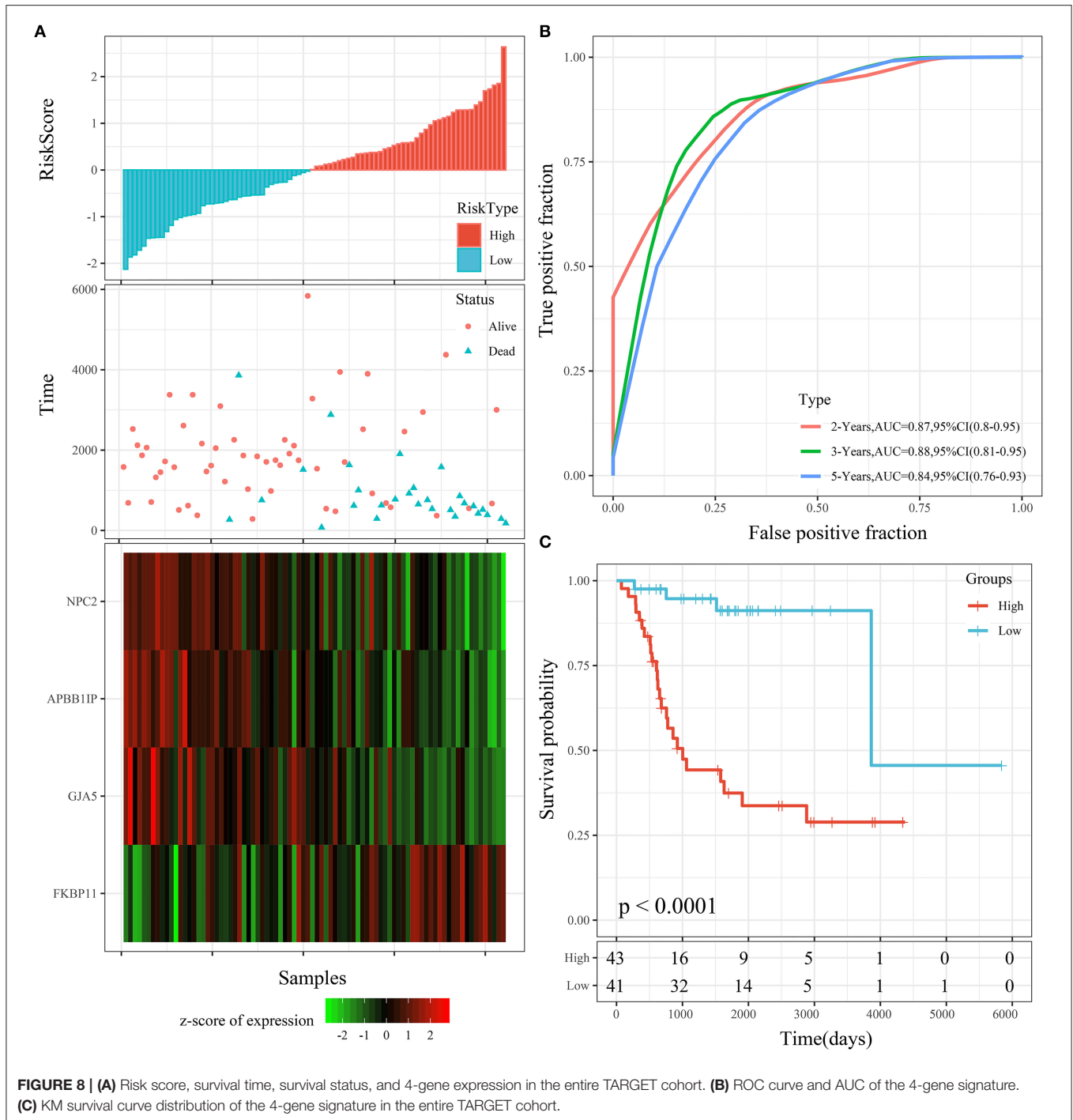
high risk (a risk factor), while high *GJA5*, *APBB1IP*, or *NPC2* expression was associated with low risk (protective factors).

FK506 binding protein 11 (*FKBP11*) belongs to the FK506 binding protein family (Rulten et al., 2006), and *FKBP11* mRNA is present at high levels in many secreting tissues, including pancreas, stomach, and salivary gland tissues (Bonner and Boulianne, 2017). *FKBP11* is involved in the regulation of mTOR (Hausch et al., 2013; Wang et al., 2019) and the pathogenesis of stress-related inflammatory diseases, including type 2 diabetes



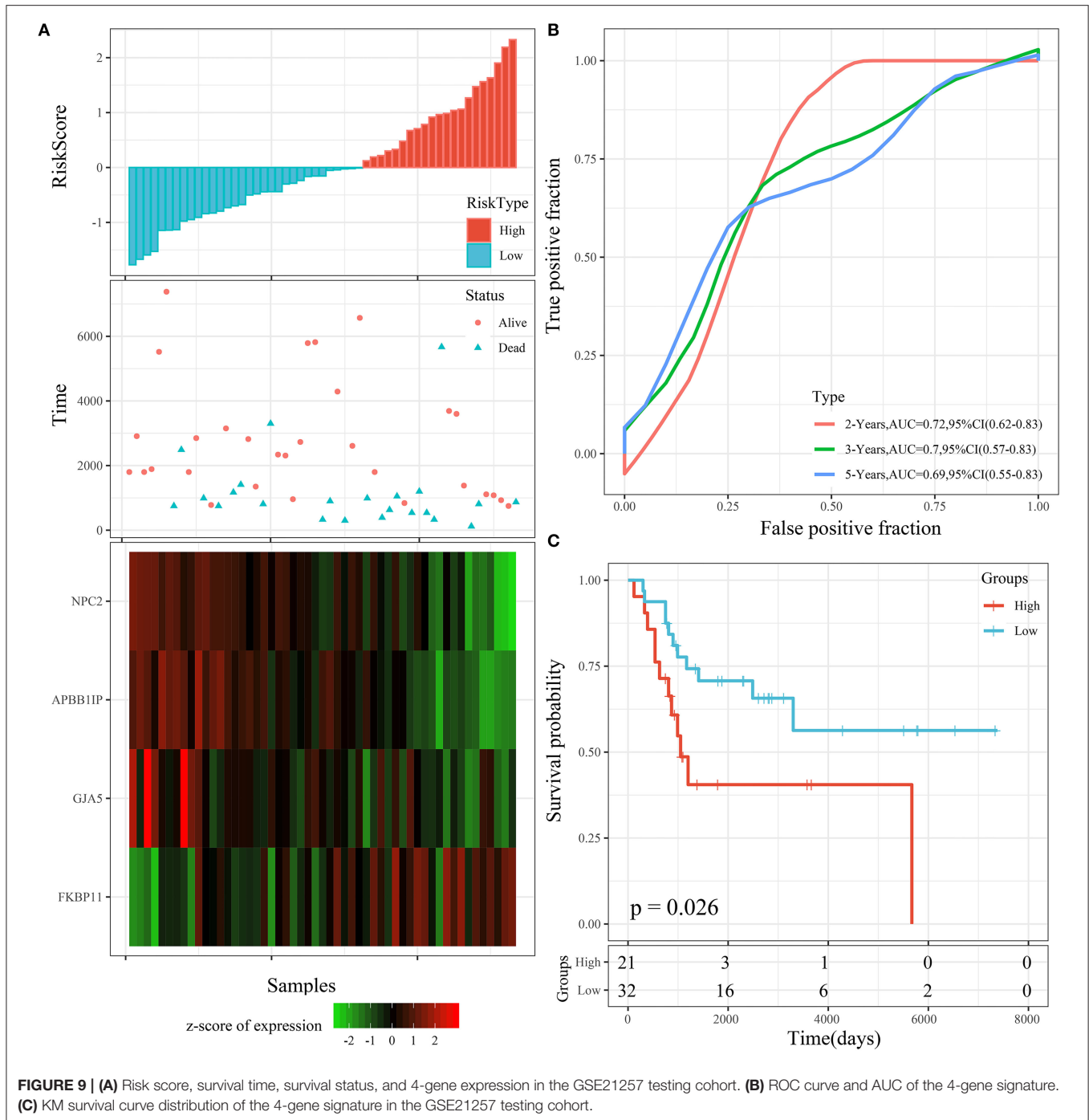
mellitus, systemic lupus erythematosus, and hepatitis (Ruer-Laventie et al., 2015; Wang et al., 2017, 2018). In addition, a gradual increase in *FKBP11* expression has been detected in the development of hepatocellular carcinoma (HCC), suggesting that *FKBP11* may be a potential early biomarker for HCC (Lin et al., 2013). Gap junction alpha-5 protein (*GJA5*, also known as connexin 40) is a constitutive vascular gap junction protein that plays an important role in the coupling between vascular wall cells (Bai, 2014; Lu and Wang, 2017). New ultraviolet

target genes, including *GJA5*, have been identified, and they are often dysregulated in human squamous cell carcinoma (Shen et al., 2017). However, the significance of *GJA5* in tumorigenesis remains unclear. Amyloid beta precursor protein binding family B member 1 interacting protein [*APBB1IP*, also known as Rap1-GTP-interacting adaptor molecule (RIAM)], appears to play a role in signal transduction from Ras activation to actin cytoskeleton remodeling, and it seems to mediate Ras-related protein 1-induced adhesion (Bromberger et al.,



2019). APBB1IP integrates signaling events that are critical for integrin-mediated immune function control and cancer progression (Patsoukis et al., 2019). Some studies have shown that APBB1IP is a new biological marker associated with gastric cancer and head and neck squamous cell carcinoma (Sanati et al., 2018; Shen et al., 2019). The Niemann–Pick type C2 (NPC2) protein is involved in the regulation of intracellular

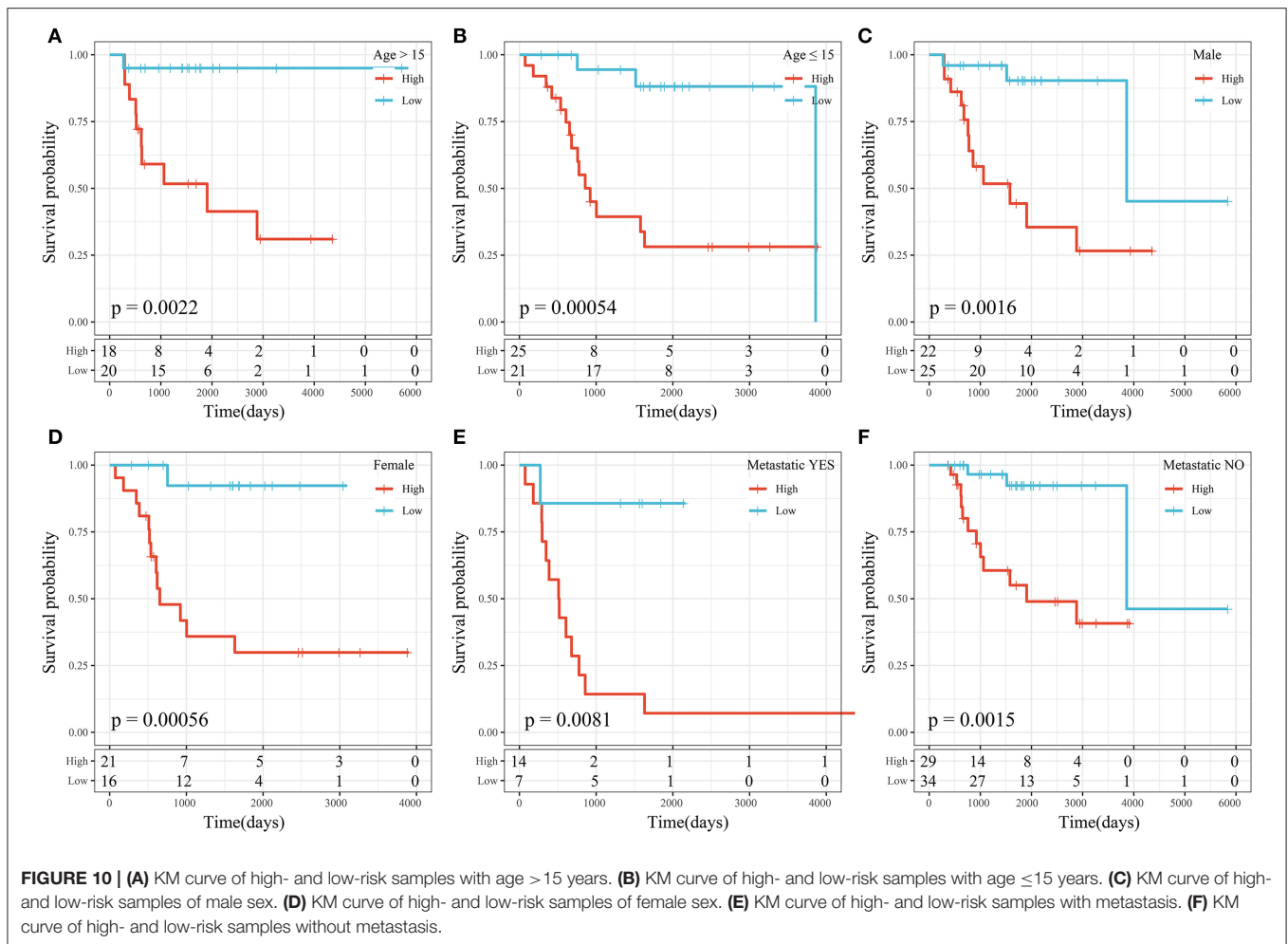
cholesterol homeostasis via direct binding with free cholesterol (Kamata et al., 2015). NPC2 is abundant in normal liver tissue, but it is downregulated in HCC. Further, low NPC2 levels may predict poor prognosis and regulate the progression of HCC by regulating the ERK1/2 pathway (Liao et al., 2015; Chen et al., 2018). To further understand the biological correlations of these genes, we conducted functional enrichment analysis and found



that these genes were mainly concentrated in immune-related pathways. The 4-gene signature was reliable and performed well in predicting OS prognosis and the effect of immunotherapy.

Furthermore, the 4-gene signature was able to divide samples into high- and low-risk groups appropriately in terms of age, sex, and metastasis. The results suggested that the risk model had good predictive ability in terms of different clinical features and that the risk score could be used as an independent prognostic risk factor.

To observe the relationships between the risk scores of the different samples and biological functions, KEGG_JAK_STAT_SIGNALING_PATHWAY, KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY, KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY, and KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY were selected. JAK/STAT signal transduction is an essential part of growth factor and cytokine signaling, which is involved in virtually all immune regulatory processes and has become



an attractive target for many immune, inflammatory, and hematopoietic diseases (Waldmann and Chen, 2017; Gao et al., 2018; Owen et al., 2019). Natural killer cells are critical effector lymphocytes that mediate tumor immune surveillance and clearance and play an important role in innate and adaptive immune responses to tumors (Malmberg et al., 2017). Toll-like receptor signaling is involved in activating innate and adaptive immune responses, and there is substantial evidence of the benefit of targeting this pathway in cancer treatment (Li et al., 2014; Moradi-Marjaneh et al., 2018). B cell receptor signaling is critical for normal B cell development and adaptive immunity, and B cell receptor signal transduction supports the survival and growth of malignant B cells in patients with B cell leukemia or lymphoma (Burger and Wiestner, 2018). These pathways were all negatively correlated with risk score.

More studies of OS prognosis models are being reported as time goes on. We compared three published OG prognosis signatures to prove the superiority of our model. Goh et al. (2019) developed a novel OS prognostic score with grouped-variable selection using network-regularized high-dimensional Cox regression analysis. Zhang et al. (2019) built a predictive tool for OS lung metastasis and progression by using co-expression

network analysis. Shi et al. (2020) constructed a 3-gene risk signature based on metastasis-associated genes for the prediction of OS prognosis and therapeutic targets. To make the models comparable, the same method was conducted to calculate the risk score of each OS sample in the TARGET cohort based on the corresponding genes in these 4 models. The AUCs of the 3 comparison models were lower than that of our 4-gene risk score model at 2, 3, and 5 years. Our 4-gene risk score model used a reasonable number of genes and had significantly high discriminatory power in predicting overall survival.

Our study has some limitations. Because the incidence of human OS is low, the number of OS samples in the TARGET database is relatively small. Only 84 OS samples were obtained, which may lead to selection bias. In addition, some important and meaningful genes may have been missed in our multiple screening processes. For better clinical applications, more samples are needed in future research to verify our findings. Therefore, it is necessary to do more functional research studies of the 4 genes mentioned in this paper. Furthermore, numerous clinical studies as well as animal and cell experiments should be carried out to verify the prediction accuracy of our risk model and discover potential immune-related mechanisms.

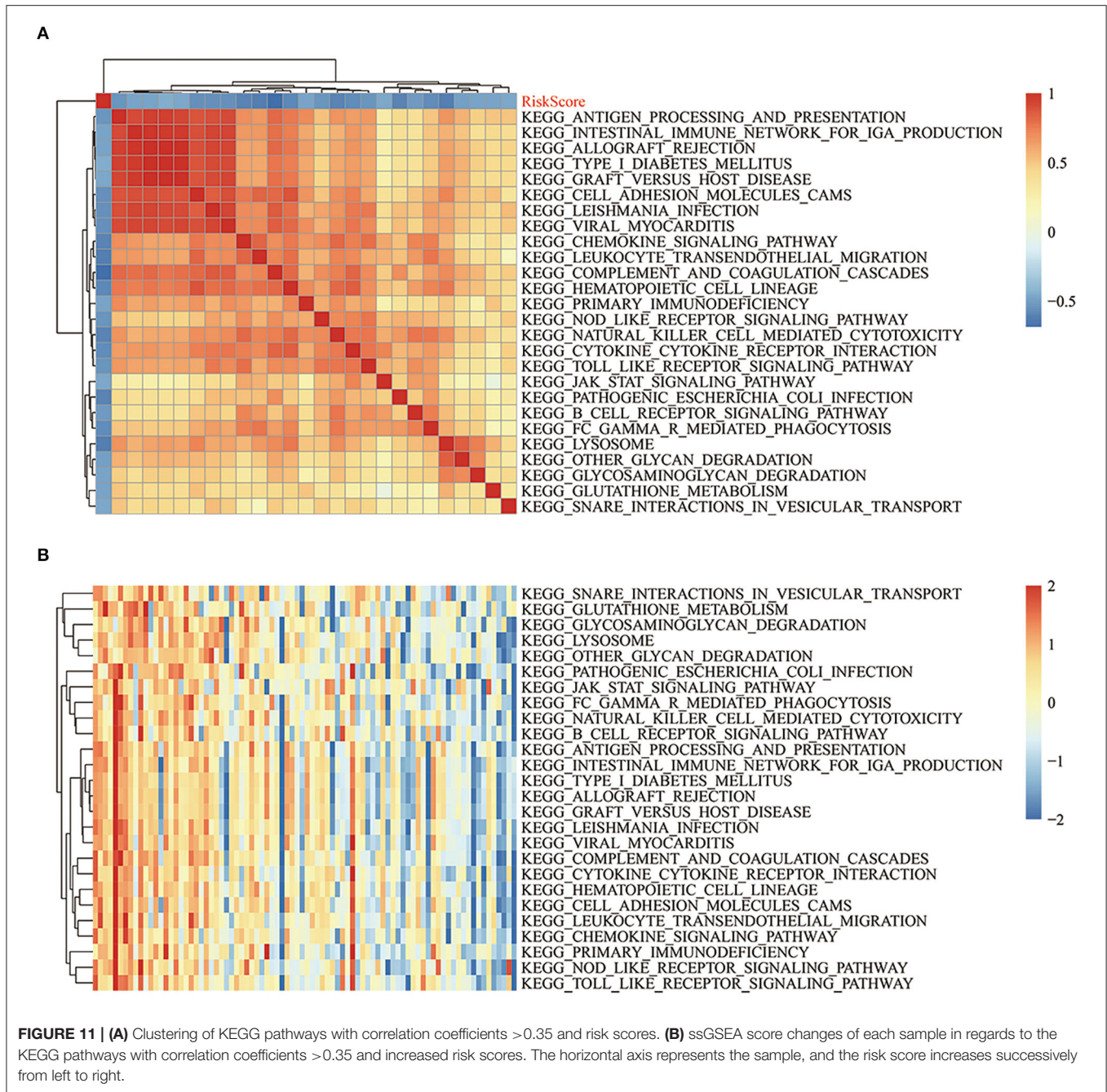


TABLE 2 | Univariable analysis and multivariable analysis of the entire TARGET.

Variables	Univariable analysis				Multivariable analysis			
	HR	95% CI of HR		P	HR	95% CI of HR		P
		Lower	Upper			Lower	Upper	
Age	0.9901	0.9118	1.0750	0.8130	1.0191	0.9298	1.1170	0.6858
Gender	0.6870	0.3304	1.4290	0.3150	0.6521	0.3005	1.4150	0.2793
Metastatic	4.7400	2.2710	9.8950	3.4E-05	3.1864	1.4696	6.9090	0.0033
RiskScore	3.4030	2.1600	5.3610	1.3E-07	2.9103	1.8364	4.6120	5.4E-06

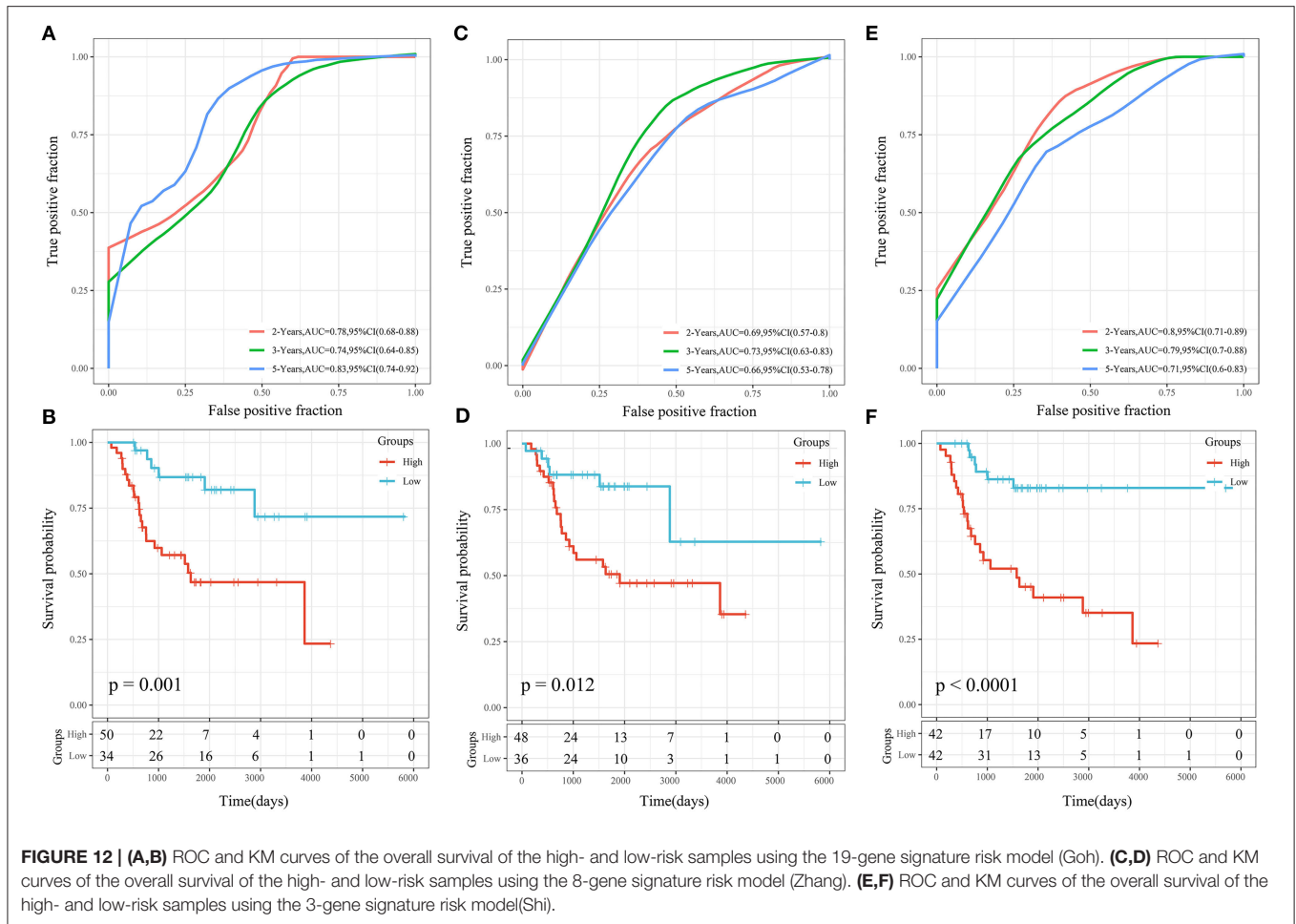


FIGURE 12 | (A,B) ROC and KM curves of the overall survival of the high- and low-risk samples using the 19-gene signature risk model (Goh). **(C,D)** ROC and KM curves of the overall survival of the high- and low-risk samples using the 8-gene signature risk model (Zhang). **(E,F)** ROC and KM curves of the overall survival of the high- and low-risk samples using the 3-gene signature risk model (Shi).

In conclusion, we constructed an effective 4-immune gene signature (including *GJA5*, *APBB1IP*, *NPC2*, and *FKBP11*) based on TARGET and GEO data for predicting OS prognosis, and the stability and accuracy of the model were assessed. Our work will help clinicians evaluate OS prognosis and select appropriate immunotherapy targets. In the future, this risk prognosis model should be further validated in more OS patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MC: conceptualization, methodology, software, and writing—reviewing and editing. JZ: data curation and writing—original

draft preparation. HX: visualization and investigation. ZL, XC, and HW: software and validation. YS: supervision. All authors contributed to the article and approved the submitted version.

FUNDING

The work was funded by Medical research fund project of Guangdong Province (Grant No. B2019184) and Project of Administration of Traditional Chinese Medicine of Guangdong Province of China (Grant No. 20191204).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.608368/full#supplementary-material>

Supplementary Figure 1 | Immune-related pathways between C1 and C2.

REFERENCES

- Bai, D. (2014). Atrial fibrillation-linked GJA5/connexin40 mutants impaired gap junctions via different mechanisms. *FEBS Lett.* 588, 1238–1243. doi: 10.1016/j.febslet.2014.02.064
- Bonner, J. M., and Boulianne, G. L. (2017). Diverse structures, functions and uses of FK506 binding proteins. *Cell. Signal* 38, 97–105. doi: 10.1016/j.cellsig.2017.06.013
- Bromberger, T., Zhu, L., Klapproth, S., Qin, J., and Moser, M. (2019). Rap1 and membrane lipids cooperatively recruit talin to trigger integrin activation. *J. Cell Sci.* 132:jcs235531. doi: 10.1242/jcs.235531
- Burger, J. A., and Wiestner, A. (2018). Targeting B cell receptor signalling in cancer: preclinical and clinical advances. *Nat. Rev. Cancer* 18, 148–167. doi: 10.1038/nrc.2017.121
- Chen, K. J., Jin, R. M., Shi, C. C., Ge, R. L., Hu, L., Zou, Q. F., et al. (2018). The prognostic value of Niemann-Pick C1-like protein 1 and Niemann-Pick disease type C2 in hepatocellular carcinoma. *J. Cancer* 9, 556–563. doi: 10.7150/jca.19996
- Chen, Y. P., Zhang, Y., Lv, J. W., Li, Y. Q., Wang, Y. Q., He, Q. M., et al. (2017). Genomic analysis of tumor microenvironment immune types across 14 solid cancer types: immunotherapeutic implications. *Theranostics* 7, 3585–3594. doi: 10.7150/thno.21471
- de Visser, K. E., Eichten, A., and Coussens, L. M. (2006). Paradoxical roles of the immune system during cancer development. *Nat. Rev. Cancer* 6, 24–37. doi: 10.1038/nrc1782
- Gajewski, T. F., Schreiber, H., and Fu, Y. X. (2013). Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* 14, 1014–1022. doi: 10.1038/ni.2703
- Gao, Q., Liang, X., Shaikh, A. S., Zang, J., Xu, W., and Zhang, Y. (2018). JAK/STAT signal transduction: promising attractive targets for immune, inflammatory and hematopoietic diseases. *Curr. Drug Targets* 19, 487–500. doi: 10.2174/1389450117666161207163054
- Goh, T. S., Lee, J. S. Il, Kim, J., Park, Y. G., Pak, K., Jeong, D. C., et al. (2019). Prognostic scoring system for osteosarcoma using network-regularized high-dimensional Cox-regression analysis and potential therapeutic targets. *J. Cell. Physiol.* 234, 13851–13857. doi: 10.1002/jcp.28065
- Gomez-Brouchet, A., Illac, C., Gilhodes, J., Bouvier, C., Aubert, S., Guinebretiere, J. M., et al. (2017). CD163-positive tumor-associated macrophages and CD8-positive cytotoxic lymphocytes are powerful diagnostic markers for the therapeutic stratification of osteosarcoma patients: an immunohistochemical analysis of the biopsies from the French OS2006 phase 3 trial. *Oncoimmunology* 6:e1331193. doi: 10.1080/2162402X.2017.1331193
- Hachen, N., Fritsch, E. F., Carter, T. A., Lander, E. S., and Wu, C. J. (2013). Getting personal with neoantigen-based therapeutic cancer vaccines. *Cancer Immunol. Res.* 1, 11–15. doi: 10.1158/2326-6066.CIR-13-0022
- Hausch, F., Kozany, C., Theodoropoulou, M., and Fabian, A. K. (2013). FKBP and the Akt/mTOR pathway. *Cell Cycle* 12, 2366–2370. doi: 10.4161/cc.25508
- Jackson, T. M., Bittman, M., and Granowetter, L. (2016). Pediatric malignant bone tumors: a review and update on current challenges, and emerging drug targets. *Curr. Probl. Pediatr. Adolesc. Health Care* 46, 213–228. doi: 10.1016/j.cppeds.2016.04.002
- Jawad, M. U., Cheung, M. C., Clarke, J., Koniaris, L. G., and Scully, S. P. (2011). Osteosarcoma: improvement in survival limited to high-grade patients only. *J. Cancer Res. Clin. Oncol.* 137, 597–607. doi: 10.1007/s00432-010-0923-7
- Kamata, T., Jin, H., Giblett, S., Patel, B., Patel, F., Foster, C., et al. (2015). The cholesterol-binding protein NPC2 restrains recruitment of stromal macrophage-lineage cells to early-stage lung tumours. *EMBO Mol. Med.* 7, 1119–1137. doi: 10.15252/emmm.201404838
- Kansara, M., Teng, M. W., Smyth, M. J., and Thomas, D. M. (2014). Translational biology of osteosarcoma. *Nat. Rev. Cancer* 14, 722–735. doi: 10.1038/nrc3838
- Lettieri, C. K., Appel, N., Labban, N., Lussier, D. M., Blattman, J. N., and Hingorani, P. (2016). Progress and opportunities for immune therapeutics in osteosarcoma. *Immunotherapy* 8, 1233–1244. doi: 10.2217/imt-2016-0048
- Li, T. T., Ogino, S., and Qian, Z. R. (2014). Toll-like receptor signaling in colorectal cancer: carcinogenesis to cancer therapy. *World J. Gastroenterol.* 20, 17699–17708. doi: 10.3748/wjg.v20.i41.17699
- Liao, Y. J., Fang, C. C., Yen, C. H., Hsu, S. M., Wang, C. K., Huang, S. F., et al. (2015). Niemann-Pick type C2 protein regulates liver cancer progression via modulating ERK1/2 pathway: clinicopathological correlations and therapeutic implications. *Int. J. Cancer* 137, 1341–1351. doi: 10.1002/ijc.29507
- Lin, I. Y., Yen, C. H., Liao, Y. J., Lin, S. E., Ma, H. P., Chan, Y. J., et al. (2013). Identification of FKBP11 as a biomarker for hepatocellular carcinoma. *Anticancer Res.* 33, 2763–2769.
- Lu, X. J., and Wang, H. T. (2017). Reduced Gja5 expression in arterial endothelial cells impairs arteriogenesis during acute ischemic cardiovascular disease. *Exp. Ther. Med.* 14, 4339–4343. doi: 10.3892/etm.2017.5068
- Malmberg, K. J., Carlsten, M., Björklund, A., Sohlberg, E., Bryceson, Y. T., and Ljunggren, H. G. (2017). Natural killer cell-mediated immunosurveillance of human cancer. *Semin. Immunol.* 31, 20–29. doi: 10.1016/j.smm.2017.08.002
- Meyers, P. A., Schwartz, C. L., Krailo, M. D., Healey, J. H., Bernstein, M. L., Betcher, D., et al. (2008). Osteosarcoma: the addition of muramyl tripeptide to chemotherapy improves overall survival—a report from the Children’s oncology group. *J. Clin. Oncol.* 26, 633–638. doi: 10.1200/JCO.2008.14.0095
- Mirabello, L., Troisi, R. J., and Savage, S. A. (2009). International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. *Int. J. Cancer* 125, 229–234. doi: 10.1002/ijc.24320
- Miwa, S., Shirai, T., Yamamoto, N., Hayashi, K., Takeuchi, A., Igarashi, K., et al. (2019). Current and emerging targets in immunotherapy for osteosarcoma. *J. Oncol* 2019:7035045. doi: 10.1155/2019/7035045
- Moradi-Marjaneh, R., Hassanian, S. M., Fiuji, H., Soleimanpour, S., Ferns, G. A., Avan, A., et al. (2018). Toll like receptor signaling pathway as a potential therapeutic target in colorectal cancer. *J. Cell. Physiol.* 233, 5613–5622. doi: 10.1002/jcp.26273
- Mou, K. W., Goldberg, M. S., Konstantinopoulos, P. A., and D’Andrea, A. D. (2017). DNA damage and repair biomarkers of immunotherapy response. *Cancer Discov.* 7, 675–693. doi: 10.1158/2159-8290.CD-17-0226
- Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., and Freeman, T. C. (2018). Immune cell gene signatures for profiling the microenvironment of solid tumors. *Cancer Immunol. Res.* 6, 1388–1400. doi: 10.1158/2326-6066.CIR-18-0342
- Owen, K. L., Brockwell, N. K., and Parker, B. S. (2019). JAK-STAT signaling: a double-edged sword of immune regulation and cancer progression. *Cancers* 11:2002. doi: 10.3390/cancers11122002
- Pan, Y., Chen, D., Hu, T., Lv, G., and Dai, Z. (2019). Characteristics and prognostic factors of patients with osteosarcoma older than 60 years from the SEER database. *Cancer Control* 26:1073274819888893. doi: 10.1177/1073274819888893
- Patsoukis, N., Bardhan, K., Weaver, J. D., Sari, D., Torres-Gomez, A., Li, L., et al. (2019). The adaptor molecule RIAM integrates signaling events critical for integrin-mediated control of immune function and cancer progression. *Sci. Signal* 10:eaam8298. doi: 10.1126/scisignal.aam8298
- Ruer-Laventie, J., Simoni, L., Schickel, J. N., Soley, A., Duval, M., Knapp, A. M., et al. (2015). Overexpression of Fkbp11, a feature of lupus B cells, leads to B cell tolerance breakdown and initiates plasma cell differentiation. *Immun. Inflamm. Dis.* 3, 265–279. doi: 10.1002/iid3.65
- Rulten, S. L., Kinloch, R. A., Tateossian, H., Robinson, C., Gettins, L., and Kay, J. E. (2006). The human FK506-binding proteins: characterization of human FKBP19. *Mamm. Genome* 17, 322–331. doi: 10.1007/s00335-005-0127-7
- Sanati, N., Iancu, O. D., Wu, G., Jacobs, J. E., and McWeeney, S. K. (2018). Network-based predictors of progression in head and neck squamous cell carcinoma. *Front. Genet.* 9:183. doi: 10.3389/fgene.2018.00183
- Sayles, L. C., Breese, M. R., Koehne, A. L., Leung, S. G., Lee, A. G., Liu, H. Y., et al. (2019). Genome-informed targeted therapy for osteosarcoma. *Cancer Discov.* 9, 46–63.
- Shen, J. K., Cote, G. M., Choy, E., Yang, P., Harmon, D., Schwab, J., et al. (2014). Programmed cell death ligand 1 expression in osteosarcoma. *Cancer Immunol. Res.* 2, 690–698. doi: 10.1158/2326-6066.CIR-13-0224
- Shen, Q., Polom, K., Williams, C. F., de Oliveira, M. S., Guergova-Kuras, M., et al. (2019). A targeted proteomics approach reveals a serum protein signature as diagnostic biomarker for resectable gastric cancer. *EBioMedicine* 44, 322–333. doi: 10.1016/j.ebiom.2019.05.044
- Shen, Y., Stanislaukas, M., Li, G., Zheng, D., and Liu, L. (2017). Epigenetic and genetic dissections of UV-induced global gene dysregulation in skin cells through multi-omics analyses. *Sci. Rep.* 7:42646. doi: 10.1038/srep42646

- Shi, Y., He, R., Zhuang, Z., Ren, J., Wang, Z., Liu, Y., et al. (2020). A risk signature-based on metastasis-associated genes to predict survival of patients with osteosarcoma. *J. Cell Biochem.* 121, 3479–3490. doi: 10.1002/jcb.29622
- Thanindrarn, P., Dean, D. C., Nelson, S. D., Hornicek, F. J., and Duan, Z. (2019). Advances in immune checkpoint inhibitors for bone sarcoma therapy. *J. bone Oncol.* 15:100221. doi: 10.1016/j.jbo.2019.100221
- van Erp, A. E. M., Versleijen-Jonkers, Y. M. H., Hillebrandt-Roeffen, M. H. S., van Houdt, L., Gorris, M. A. J., van Dam, L. S., et al. (2017). Expression and clinical association of programmed cell death-1, programmed death-ligand-1 and CD8(+) lymphocytes in primary sarcomas is subtype dependent. *Oncotarget* 8, 71371–71384. doi: 10.18632/oncotarget.19071
- Waldmann, T. A., and Chen, J. (2017). Disorders of the JAK/STAT pathway in T cell lymphoma pathogenesis: implications for immunotherapy. *Annu. Rev. Immunol.* 35, 533–550. doi: 10.1146/annurev-immunol-110416-120628
- Wang, T., He, X., Liu, X., Liu, Y., Zhang, W., Huang, Q., et al. (2017). Weighted gene co-expression network analysis identifies FKBP11 as a key regulator in acute aortic dissection through a NF-kB dependent pathway. *Front. Physiol.* 8:1010. doi: 10.3389/fphys.2017.01010
- Wang, X., Cui, X., Zhu, C., Li, M., Zhao, J., Shen, Z., et al. (2018). FKBP11 protects intestinal epithelial cells against inflammation-induced apoptosis via the JNK-caspase pathway in Crohn's disease. *Mol. Med. Rep.* 18, 4428–4438. doi: 10.3892/mmr.2018.9485
- Wang, X., Deng, Y., Zhang, G., Li, C., Ding, G., May, H. I., et al. (2019). Spliced X-box binding protein 1 stimulates adaptive growth through activation of mTOR. *Circulation* 140, 566–579. doi: 10.1161/CIRCULATIONAHA.118.038924
- Wang, Z., Li, B., Ren, Y., and Ye, Z. (2016). T-cell-based immunotherapy for osteosarcoma: challenges and opportunities. *Front. Immunol.* 7:353. doi: 10.3389/fimmu.2016.00353
- Wedekind, M. F., Wagner, L. M., and Cripe, T. P. (2018). Immunotherapy for osteosarcoma: where do we go from here? *Pediatr. Blood Cancer* 65:e27227. doi: 10.1002/pbc.27227
- Zhang, B., Li, Y. L., Zhao, J. L., Zhen, O., Yu, C., Yang, B. H., et al. (2018). Hypoxia-inducible factor-1 promotes cancer progression through activating AKT/Cyclin D1 signaling pathway in osteosarcoma. *Biomed. Pharmacother.* 105, 1–9. doi: 10.1016/j.biopha.2018.03.165
- Zhang, H., Guo, L., Zhang, Z., Sun, Y., Kang, H., Song, C., et al. (2019). Co-expression network analysis identified gene signatures in osteosarcoma as a predictive tool for lung metastasis and survival. *J. Cancer* 10, 3706–3716. doi: 10.7150/jca.32092
- Zheng, W., Xiao, H., Liu, H., and Zhou, Y. (2015). Expression of programmed death 1 is correlated with progression of osteosarcoma. *APMIS* 123, 102–107. doi: 10.1111/apm.12311

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cao, Zhang, Xu, Lin, Chang, Wang, Huang, Chen, Wang and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.