

Development and validation of an immune gene-set based prognostic signature for soft tissue sarcoma

Rui Shen

the Affiliated Hospital of Qingdao University <https://orcid.org/0000-0003-2279-2742>

Xiangying Meng

the people's hospital of dongying district

Jianyi Li

the Affiliated Hospital of Qingdao University

Tengbo Yu (✉ 467199436@qq.com)

Affiliated Hospital of Qingdao University the Affiliated Hospital of Qingdao University

Kuishuai Xu

the affiliated hospital of qingdao university

Research article

Keywords: Soft tissue sarcoma, immune-related gene, nomogram, immune infiltration

Posted Date: August 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-57363/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published on February 8th, 2021. See the published version at <https://doi.org/10.1186/s12885-021-07852-2>.

Abstract

Background

Sarcomas is a group of heterogeneous malignant tumors originated from mesenchymal tissue and different types of sarcomas have disparate outcomes. The present study aims to identify the prognostic value of immune-related genes (IRGs) in sarcoma and establish a prognostic signature based on IRGs.

Methods

We collected the expression profile and clinical information of 255 soft tissue sarcoma samples from The Cancer Genome Atlas (TCGA) database and 2498 IRGs from the ImmPort database. The LASSO algorithm and Cox regression analysis were used to identify the best candidate genes and construct a signature. The prognostic ability of the signature was evaluated by ROC curves and Kaplan-Meier survival curves and validated in an independent cohort. Besides, a nomogram based on the IRGs and independent prognostic clinical variables was developed.

Results

A total of 19 IRGs were incorporated into the signature. In the training cohort, the AUC values of signature at 1-, 2-, and 3-years were 0.938, 0.937 and 0.935, respectively. The Kaplan-Meier survival curve indicated that high-risk patients were significantly worse prognosis ($P < 0.001$). In the validation cohort, the AUC values of signature at 1-, 2-, and 3-years were 0.730, 0.717 and 0.647, respectively. The Kaplan-Meier survival curve also showed significant distinct survival outcome between two risk groups. Furthermore, a nomogram based on the signature and four prognostic variables showed great accuracy in whole sarcoma patients and subgroup analyses. More importantly, the results of the TF regulatory network and immune infiltration analysis revealed the potential molecular mechanism of IRGs.

Conclusions

In general, we identified and validated an IRG-based signature, which can be used as an independent prognostic signature in evaluating the prognosis of sarcoma patients and provide potential novel immunotherapy targets.

Background

Sarcoma is a rare group of heterogeneous malignant tumors originated from mesenchymal tissue, mostly occurring in the soft tissue and ending of the long bone[1]. Sarcoma can be classified into more than 80 histological types according to the histological and molecular features[2]. Osteosarcoma, leiomyoma, lymphosarcoma, and synovial sarcoma are four common histological types[3]. Although

sarcoma only accounts for 1% of malignancies, they account for 10%~12% of malignancies in children and adolescents[2, 4, 5]. The incidence in recent years is 2.49~5.87 per person-year[6-9] and the 5-years survival rate after diagnosis was 56.4%~61.6%[6-8]. However, 40%~50% of sarcoma patients could develop metastasis[3, 10], which makes it difficult to choose an appropriate treatment, such as surgery, chemotherapy, and radiotherapy. Therefore, it's crucial to find effective makers for risk assessment for sarcoma patients.

In recent years, lncRNA, miRNA, and plasmacytoma variant translocation 1 have been identified as prognostic biomarkers associated with the prognosis of sarcoma patients[11-13]. Nevertheless, these markers have been unable to be used in clinical practice due to the poor prognostic ability and lacking validation. Lately, the role of immune-related features in malignant tumors has been widely investigated. Elements of the immune system have proven to be a strong factor in tumorigenesis and progression[14]. For this, immunotherapy is considered as a new and powerful therapy, especially for targeting PD1 and PDL-1 techniques. By acting on the immune checkpoints, it has been used to treat a variety of cancers, including sarcoma[15]. More importantly, previous studies indicated that immune-related genes(IRGs) can serve as an effective prognostic biomarkers in many tumors, such as lung cancer[16-19], ovarian cancer[20, 21], hepatocellular carcinoma[22, 23], head and neck squamous cell carcinoma[24], papillary thyroid cancer[25], bladder urothelial carcinoma[26], and renal cancer[27]. Nonetheless, the prognostic significance of IRGs in sarcoma remains unclear but urgent.

Therefore, the aim of our study was to study the prognostic value of IRGs in sarcoma cohort based on the RAN-sequencing and clinical data from TCGA-SARC (<https://cancergenome.nih.gov/>) and the IRGs from ImmPort database (<https://www.immport.org/shared/>). In addition, the regulatory network between prognostic transcription factors (TFs) and prognostic IRGs was established. Finally, a nomogram based on the IRGs and prognostic clinical variables was developed and evaluated.

Methods

Patient samples and IRGs

RNA-sequencing (RNA-seq) expression profile and corresponding clinical information of soft tissue sarcoma patients were downloaded from TCGA database, and the clinical information of all patients are shown in Table 1. Meanwhile, the IRG set was obtained from the ImmPort database, which covered 2498 genes, including antigen processing and presentation, antimicrobials, BCR signaling pathway, chemokines, chemokine receptors, etc. Then, 170 patients (2/3) were randomly selected to form the training set and the remaining 85 patients (1/3) were incorporated into the testing set.

Table.1 Clinical characteristics of sarcoma patients				
	Training set	Validation set	t/X ²	P
Age	60.74±15.29	60.46±13.50	0.141	0.888
Sex			0.198	0.657
Male	79	37		
Female	91	48		
Race			1.074	0.783
Asian	5	1		
African American	11	7		
WHITE	149	75		
NA	5	2		
Histological type			4.753	0.314
LMS	65	37		
DLP	34	23		
UPS	35	14		
MYX	18	7		
Other	18	4		
Metastasis			1.282	0.527
No	84	36		
Yes	34	21		
NA	52	28		
Surgical margin resection status			0.301	0.860
R0	104	49		
R1-2	49	27		
NA	17	9		
Radiotherapy			0.672	0.715
No	94	43		
Yes	46	27		
NA	30	15		

Multifocal indicator		0.699	0.705
No	131	64	
Yes	24	15	
NA	15	6	

LMS: Leiomyosarcoma; DLP: Dedifferentiated liposarcoma; UPS: Undifferentiated pleomorphic sarcoma; MYX: Myxofibrosarcoma

Identification of OS-related IRGs in sarcoma patients and enrichment analysis.

To identify the prognostic value of IRGs in sarcoma patients, we performed the univariate Cox proportional hazard model to identify the overall survival (OS)-related IRGs in the training set, and genes with a $p < 0.05$ were considered as the OS-related genes. Meanwhile, to further understand the function of OS-related IRGs, Gene Ontology (GO), including molecular function (MF), biological process (BP), and cellular component (CC) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database were used for enrichment analysis. The GO and KEGG analyses were performed by the R package clusterprofiler[28].

Construction of a transcription factor regulatory network

To determine the possible mechanisms behind the regulation of OS-related IRGs in soft tissue sarcoma patients, we analyzed the correlation between the expression profile data of OS-related TFs and OS-related IRGs. TF set was downloaded from the Cistrome Cancer(<http://cistrome.org/>). Then, we used univariate Cox proportional hazard model to detect OS-related TFs. The correlation between OS-related TFs and OS-related IRGs were performed by the Pearson correlation analysis. Correlation with a $p < 0.01$ and $r > 0.5$ was considered to be statistical significance. To better illustrate regulatory relationships between the TFs and IRGs, TF-based regulatory network was generated by Cytoscape[29].

Construction and evaluation of IRGs prognostic signature

Based on the OS-related IRGs identified in the univariate Cox proportional hazard model, we performed the LASSO analysis to avoid overfitting[30]. Then, the significant genes in the LASSO regression were incorporated into the multivariate Cox analysis, and a prognostic signature was established. Meanwhile, the regression coefficients from the multivariate Cox regression model was defined as the weight of the selected genes, which was used to calculate the risk scores for each patient. The formula is follows:

$$Risk\ Score = \sum_{i=0}^n \beta_i * G_i$$

Here, 'G' is the expression of the selected gene, and 'B_i' is the coefficient of the gene from the multivariate Cox proportional hazards analysis.

To identify the prognostic ability of the IRGs-based signature, time-dependent receiver operating characteristic (ROC) curves were generated, and the corresponding time-dependent area under the curve (AUC) was calculated simultaneously. In addition, 170 patients were stratified into the high- and low-risk group based on the median of risk score. Kaplan-Meier(K-M) survival curve and log-rank test was performed to show the difference of survival status between two groups[31].

External validation of IRGs prognostic signature

According to the prognostic signature developed in the training set, risk scores of 85 patients in the testing set were calculated. and all patients were classified into high- and low-risk groups. Similarly, K-M survival curves with log-rank test and ROC analysis were utilized to evaluate the accuracy of the prognostic signature in an independent validation cohort.

Development of a nomogram based on IRGs and clinical variables

In the present study, the data of age, sex, race, histological type of tumor, tumor site and patient metastatic status were obtained from TCGA portal. To investigate whether the risk model is independently associated with the prognosis of soft tissue sarcoma patients, univariate Cox analysis was performed to identify the prognostic variables, which were incorporated into the multivariate Cox analysis to determine the independent prognostic factors. Furthermore, based on the independent prognostic factors, a prognostic nomogram was established [32]. C-index, calibration curve, and decision curve analysis(DCA) were used to evaluate the performance of the nomogram[33].

Identification of immune infiltration cell in low- and high-risk group

Previous studies showed that IRGs can influence the immune cell composition of tumor microenvironment. Therefore, to further understand the difference of immune infiltration cell between low- and high-risk groups, CIBERSORT package was used to calculate the 22 types of immune cells for each sample, and only samples with CIBERSORT $p < 0.05$ were included in further study[34]. Then, Wilconson's rank sum test was used to compare 22 types of immune cells between the high-risk and low-risk groups.

Results

Baseline

According to the criteria, 255 soft tissue sarcoma patients were selected in our research. Then, 170 patients were incorporated into the training set and the remaining 85 patients were used to form the testing set. The baseline information of all patients shown in Table 1. The results indicated that the differences of clinical data between the training and testing sets were not statistically significant.

Identification of prognostic IRGs in sarcoma patients

To identify the prognostic value of IRGs in soft tissue sarcoma patients, univariate Cox proportional hazard model was performed in 170 patients. Totally, 105 IRGs were selected as OS-related IRGs (Supplementary 1). Moreover, GO and KEGG enrichment analyses were performed, and the results are shown in Figure.1, which indicated that the major enriched GO terms of BP were defense response to other organism, positive regulation of cell adhesion, regulation of innate immune response, regulation of leukocyte activation, and positive regulation of leukocyte activation. In CC, the major enriched GO terms were adherens junction, focal adhesion, cell–substrate adherens junction, cell–substrate junction and Schaffer collateral – CA1 synapse. For MF, we can find that the OS-related IRGs were mainly enriched in receptor ligand activity, receptor regulator activity, growth factor activity, cytokine activity, and steroid hormone receptor activity. For KEGG pathway analysis, it showed that many immune- or tumor-related pathways were identified, such as T cell receptor signaling pathway, Natural killer cell mediated cytotoxicity, Kaposi sarcoma–associated herpesvirus infection, PD–L1 expression and PD–1 checkpoint pathway in cancer, Th1 and Th2 cell differentiation, and NF–kappa B signaling pathway.

Construction of a TF regulatory network

To elucidate regulatory mechanisms of OS-related IRGs, we used univariate Cox proportional hazard model to identify the OS-related TFs. Totally, 36 OS-related TFs were confirmed (Supplementary 2). Furthermore, we analyzed the correlation between the expression of OS-related TFs and OS-related IRGs, the correlation results are shown in Supplementary 3. Interestingly, the results indicated that all TFs were positively correlated with IRGs. To better illustrate the regulatory relationship between TFs and IRGs, a TF-based regulatory network was generated (Figure.2).

Construction Development and validation of the IRGs prognostic signature in training cohort

Based on 105 OS-related IRGs, the LASSO regression was used to choose the best appropriate genes as the prognostic predictors of the model. 35 genes were selected in the LASSO regression analysis (Figure.3). Then, the multivariate Cox proportional hazard model was performed based on the significant genes in the LASSO analysis, and a prognostic signature was established based on the 19 prognostic IRGs (Supplementary 4 and Figure.4). The time-dependent ROC of 1-, 2-, and 3-years were shown in Figure. 4C. The AUC values of 1-, 2-, and 3-years were 0.938, 0.937, and 0.935, respectively, which means that the prognostic signature can serve as a valid tool for prognostic prediction in sarcoma patients (Figure.4C). In addition, the risk scores of each patient in the training set were calculated, the median of risk score was used as the cutoff to stratify patients into high-risk(n=85) and low-risk(n=85) groups. The survival curve was generated, and the log-rank test indicated that the patients in the low-risk group had a favorable prognosis (Figure.4D).

External validation of IRGs signature

To further verify the stability and reliability of the risk signature based on the IRGs, an independent set was used. As the formula of risk score in the training set, the risk scores of each patient sample in the testing set were calculated (Figure.5). The time-dependent ROC curves were generated to test the

discrimination of the signature (Figure. 5C). The results showed that the AUC values of 1-,2-, and 3-year were 0.730, 0.717, and 0.647, which also showed good accuracy of nomogram in predicting the OS of sarcoma patients. Furthermore, according to the median of the risk score in the testing set, 85 patients were stratified into the low-risk group (n=43) and high-risk group (n=42). The survival curve of two groups was generated, and the results indicated that patients in the high-risk group have a worse prognosis (Figure.5D).

Development of a nomogram based on the IRG signature and clinical data

To further construct a prognostic nomogram combining IRGs and clinical data, we performed univariate and multivariate Cox regression analysis to assess the independent prognostic variables for soft tissue sarcoma patients. In the univariate analysis, age, disease multifocal indicator, metastatic disease confirmed, surgical margin resection status and risk score were associated with the prognosis of sarcoma patients (Table 2). Then, the significant variables in the univariate Cox analysis were incorporated into the multivariate Cox analysis, and five independent prognostic variables were identified (Table 2). According to the results of the multivariate Cox analysis, we can find that risk score has the greatest impact for OS. In addition, higher age, multifocal sarcoma, tumor metastasis, and surgical resection status(R1-2) were also associated with worse prognosis in soft tissue sarcoma patients (Table 2). To better predict the prognosis of soft tissue sarcoma patients, we constructed a nomogram based on the independent factors determined in the multivariate regression (Figure.6A). The C-index of our nomogram was 0.775(95%CI:0.751-0.799), which showed good accuracy in predicting the prognosis of soft tissue sarcoma patients. The favorable calibration plot of our nomogram indicated that the OS predicted by the nomogram is highly consistent with the actual observation (Figure. 6B-6D). In addition, DCA was also performed, and the results indicated the nomogram can serve as an effective prognostic model for soft tissue sarcoma patients (Figure. 6E-6G).

Table.2 Univariate and multivariate Cox analysis in sarcoma patients								
	Univariate analysis				Multivariate analysis			
	HR	95%CI		P	HR	95%CI		P
Age	1.020	1.004	1.036	0.012	1.025	1.008	1.043	0.004
Sex								
Female								
Male	0.849	0.565	1.274	0.429				
Race								
Asian				0.412				
African American	1.108	0.135	9.098	0.924				
WHITE	0.810	0.111	5.907	0.836				
Unknown	2.004	0.207	19.378	0.548				
Histological type								
DLP				0.903				
LMS	0.855	0.515	1.419	0.543				
MYX	0.730	0.339	1.571	0.421				
UPS	0.960	0.509	1.810	0.899				
Other	0.746	0.321	1.734	0.496				
Multifocal indicator								
No				0.002				0.008
Yes	2.328	1.443	3.754	0.001	2.228	1.324	3.748	0.003
Unknown	1.220	0.606	2.459	0.577	1.135	0.486	2.651	0.770
Metastasis								
No				<0.001				<0.001
Yes	2.962	1.797	4.880	<0.001	2.926	1.721	4.972	<0.001
Unknown	1.795	1.078	2.987	0.024	1.770	1.005	3.115	0.048
Radiotherapy								
No				0.963				
Yes	1.012	0.633	1.620	0.959				

Unknown	1.075	0.638	1.812	0.786				
Surgical margin resection status								
R0				<0.001				0.005
R1-2	2.418	1.572	3.719	<0.001	1.974	1.257	3.099	0.003
Unknown	2.194	1.156	4.165	0.016	2.099	0.960	4.589	0.063
Tumor site								
Extremity				0.745				
Other	1.182	0.699	1.997	0.533				
Retroperitoneum/ Upper abdominal	1.193	0.735	1.934	0.475				
Risk	5.192	3.215	8.382	<0.001	5.362	3.241	8.868	<0.001

LMS: Leiomyosarcoma; DLP: Dedifferentiated liposarcoma; UPS: Undifferentiated pleomorphic sarcoma; MYX: Myxofibrosarcoma

Subgroup analyses of nomogram

To further confirm that the nomogram can perform stably in different histological types of soft tissue sarcoma, patients were divided into different subgroups (Figure. 7). We can find the AUC values of the nomogram were higher than 0.750(range:0.750-0.867) in all subgroups, which means that the nomogram can perform stably in different histological types of soft tissue sarcoma. In addition, in all subgroups, the Kaplan-Meier survival curves and the log-rank test indicated that the patients in the high-risk group have a poorer prognosis than patients in the low-risk group (all $p < 0.05$) (Figure. 7).

Comparison of the immune infiltration cell in low- and high-risk group

After CIBERSORT package was performed, 169 patients with complete data of immune infiltration cells were included in this part of the study (Figure.8A). Among the 169 patients, 87 were in the low-risk group and 82 were in the high-risk group. The results showed that six immune cells were significantly different between the two groups (Figure.8B). The infiltration level of plasma cells and macrophages M0 were significantly higher in the high-risk group, while the infiltration level of NK cells resting, NK cells activated, monocytes, and macrophages M1 were significantly higher in the low-risk group (Figure.8B).

Discussion

Due to the heterogeneity and high metastatic rate of sarcoma, treatment methods were very limited and the outcome was unfavorable. Therefore, it is necessary for us to identify the effective biomarkers for the

prognosis of sarcoma patients. In the present study, 105 IRGs were identified as prognostic IRGs and the regulatory network between prognostic IRGs and TFs was established. More importantly, we constructed a prognostic signature based on 19 IRGs and validated its effectiveness, which also was proved to be associated with immune infiltration cells. Furthermore, a nomogram incorporating IRG-based signature and clinical data was established, which showed excellent performance to predict the outcome of sarcoma patients.

With the wide exploration of immune mechanisms in the pathogenesis and progression of the tumor, many IRGs were identified as prognostic biomarkers in tumor patients[16-20, 22, 23, 26, 27]. More importantly, based on the researches of tumor immunology, many immunotherapies have been developed and become new methods for treating sarcoma patients, and results indicated that the outcome was satisfactory[35-37]. However, as far as we know, it is the first research to establish a signature based on IRGs for sarcoma patients, which may open up a novel perspective for constructing an effective prognostic model and improving sarcoma patient management of immunotherapy.

In our research, a prognostic signature incorporating 19 IRGs was established. The AUC of the model was higher than 0.9 in the training set, which showed great accuracy in predicting the prognosis of sarcoma patients. Moreover, the accuracy of our model was also successfully verified by an independent set. Our results indicated that the model can be used to identify sarcoma patients at high risk and enabling early interventions to improve the prognosis. Among 19 IRGs enrolled in risk signature, VEGFA, CYR61, and RHOA have confirmed to be related to the pathogenesis or prognosis of sarcoma[38-41]. It was reported that VEGFA is a potential neovascular regulatory factor that promotes tumor growth and metastasis through its receptor. Differential expression of VEGFA subtypes regulates sarcoma metastasis and response to anti-VEGFA [38]. Besides, the VEGF family can not only promote tumor-associated immunodeficiency by interfering with the growth of early hematopoietic progenitor T cells—but promote Treg cell proliferation [42, 43]. CYR61 is a member of the CCN family. The expression level of CYR61 is associated with the aggressiveness of osteosarcoma in different pre-clinical models and patient tumor samples, and CYR61 triggers osteosarcoma metastasis and spread through an IGF1R β -dependent EMT-like process[44]. The metastasis of invasive amoebic sarcoma cells depends on the Rho / Rock / MLC signal, and RHOA overexpression is related to tumor cell invasion and migration [41]. The relationship between the remaining 16 genes and sarcoma has few been reported, however, most of them have been confirmed to be associated with other tumors, which suggested that these genes also play essential roles in sarcoma. For example, PSMD10 is a crucial oncoprotein that is up-regulated in a variety of cancers and has potential in the initiation and progression of tumors[45]. Studies have indicated important roles played by PSMD10 in the pathogenesis of liver cancers and colorectal cancers, and PSMD 10 involved in regulating the proliferation and metastasis of thyroid cancers[46].

Combining the results of the above KEGG pathway and immune genes revealed the existence of potential immune mechanisms in sarcoma. Previous studies have suggested that high levels of macrophage infiltration predict a poor prognosis for Ewing's sarcoma, which is consistent with the result that macrophage M0 significantly increased in high-risk groups in our study. [47]. Natural killer (NK) cells are

important immune cells in the innate immune mechanisms. The activation of NK cells can promote the immunotherapy of Ewing's sarcoma [48]. In this study, we found that activated NK cells were higher in the low-risk group and enriched in the KEGG pathway, which can provide a theoretical basis for immunotherapy in sarcoma patients.

In addition, in order to make the clinical application more convenient and accurate, a nomogram was established by combining immune signature and clinical data. The results of the AUC, calibration curve, DCA and subgroup analysis indicated that the nomogram can serve as an effective tool for predicting the prognosis in sarcoma patients. Although a lot of predictive models for sarcoma patients were constructed based on the clinicopathologic data, lncRNA, plasmacytoma variant translocation 1 and other predictors[13, 49, 50]. However, it should point out that the discriminative ability of the previous models was low with AUC or C-statistic less than 0.750[13, 49-51], which means that they are unsatisfactory. Therefore, we think that the IRG-based nomogram can improve the risk stratification for sarcoma patients.

There are some limitations to our research. Firstly, this is a retrospective study, which may lead to bias and a novel sight and large cohort to verify is needed. Second, although we find the prognostic ability of the signature and develop a nomogram with great predictive ability, more clinical features are needed to improve the nomograms, such as AJCC TNM stage. Finally, this study was a bioinformatic study based on a public database, and the experiment should be performed in the future to explore the mechanism of the effects of IRGs in the prognosis of sarcoma.

Conclusion

In summary, we identified and validated an IRG-based signature, which can be used as an independent prognostic signature in evaluating the prognosis of sarcoma patients. However, further experimental exploration is needed to study the potential mechanism of IRGs in sarcoma.

Abbreviations

IRGs: immune-related genes, TCGA:The Cancer Genome Atlas, ROC:Receiver operating characteristic; AUC:Area under the curve, RNA-seq:RNA-sequencing, OS:overall survival, GO:Gene Ontology, MF:molecular function, BP:biological process, CC:cellular component, KEGG:Kyoto Encyclopedia of Genes and Genomes, K-M:Kaplan-Meier, DCA:decision curve analysis, NK:Natural killer.

Declarations

Ethics approval and consent to participate

The research didn't involve animal experiments and human specimens, no ethics related issues.

Consent for publication

Not applicable.

Availability of data and materials

The data of this study are from The Cancer Genome Atlas and ImmPort database.

Competing interests

The authors declare that they have no competing interests.

Funding

We received no external funding for this study.

Author's Contributions

Rui Shen and Tengbo Yu conceived of and designed the study. Rui Shen and Jianyi Li performed literature search. Rui Shen and Xiangying Meng generated the figures and tables. Jianyi Li and Kuishuai Xu analyzed the data. Rui Shen wrote the manuscript and Tengbo Yu critically reviewed the manuscript. Tengbo Yu supervised the research.

Acknowledgements

None

References

1. Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CDM, Devesa SS. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978–2001: An analysis of 26,758 cases. *Int J Cancer*. 2006;119(12):2922–30.
2. Fletcher CDM. **The evolving classification of soft tissue tumours - an update based on the new 2013 WHO classification.** *Histopathology* 2014, 64(1).
3. Cormier JN, Huang X, Xing Y, Thall PF, Wang X, Benjamin RS, Pollock RE, Antonescu CR, Maki RG, Brennan MF, et al. Cohort analysis of patients with localized, high-risk, extremity soft tissue sarcoma treated at two cancer centers: chemotherapy-associated outcomes. *J Clin Oncol*. 2004;22(22):4567–74.
4. Lewis DR, Seibel NL, Smith AW, Stedman MR. Adolescent and young adult cancer survival. *J Natl Cancer Inst Monographs*. 2014;2014(49):228–35.
5. Siegel RL, Miller KD, Jemal A. **Cancer statistics, 2018.** *CA Cancer J Clin* 2018, 68(1).
6. Kim HS, Nam CM, Jang S-Y, Choi SK, Han M, Kim S, Moneta MV, Lee SY, Cho JM, Novick D, et al. Characteristics and Treatment Patterns of Patients with Advanced Soft Tissue Sarcoma in Korea. *Cancer Res Treat*. 2019;51(4):1380–91.

7. Schöffski P, Cornillie J, Wozniak A, Li H, Hompes D. Soft tissue sarcoma: an update on systemic treatment options for patients with advanced disease. *Oncol Res Treat.* 2014;37(6):355–62.
8. Kollár A, Rothermundt C, Klenke F, Bode B, Baumhoer D, Arndt V, Feller A. Incidence, mortality, and survival trends of soft tissue and bone sarcoma in Switzerland between 1996 and 2015. *Cancer Epidemiol.* 2019;63:101596.
9. Bessen T, Caughey GE, Shakib S, Potter JA, Reid J, Farshid G, Roder D, Neuhaus SJ. A population-based study of soft tissue sarcoma incidence and survival in Australia: An analysis of 26,970 cases. *Cancer Epidemiol.* 2019;63:101590.
10. Italiano A, Mathoulin-Pelissier S, Cesne AL, Terrier P, Bonvalot S, Collin F, Michels J-J, Blay J-Y, Coindre J-M, Bui B. Trends in survival for patients with metastatic soft-tissue sarcoma. *Cancer.* 2011;117(5):1049–54.
11. Min L, Garbutt C, Tu C, Hornicek F, Duan Z. **Potentials of Long Noncoding RNAs (LncRNAs) in Sarcoma: From Biomarkers to Therapeutic Targets.** *International journal of molecular sciences* 2017, 18(4).
12. Kohama I, Kosaka N, Chikuda H, Ochiya T. **An Insight into the Roles of MicroRNAs and Exosomes in Sarcoma.** *Cancers* 2019, 11(3).
13. Liu J, Li R, Liao X, Hu B, Yu J. Comprehensive investigation of the clinical significance and molecular mechanisms of plasmacytoma variant translocation 1 in sarcoma using genome-wide RNA sequencing data. *J Cancer.* 2019;10(20):4961.
14. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med.* 2015;21(8):938–45.
15. Zuo W, Zhao L. Recent advances and application of PD-1 blockade in sarcoma. *Onco Targets Ther.* 2019;12:6887–96.
16. Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non–small cell lung cancer. *JAMA oncology.* 2017;3(11):1529–37.
17. Song Q, Shang J, Yang Z, Zhang L, Zhang C, Chen J, Wu X. Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. *Journal of translational medicine.* 2019;17(1):70.
18. Shi X, Li R, Dong X, Chen AM, Liu X, Lu D, Feng S, Wang H, Cai K. IRGS: an immune-related gene classifier for lung adenocarcinoma prognosis. *Journal of Translational Medicine.* 2020;18(1):1–11.
19. Zhang M, Zhu K, Pu H, Wang Z, Zhao H, Zhang J, Wang Y. **An Immune-Related Signature Predicts Survival in Patients With Lung Adenocarcinoma.** *Frontiers in Oncology* 2019, 9(1314).
20. Shen S, Wang G, Zhang R, Zhao Y, Yu H, Wei Y, Chen F. Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. *EBioMedicine.* 2019;40:318–26.
21. Zhang L, Zhu P, Tong Y, Wang Y, Ma H, Xia X, Zhou Y, Zhang X, Gao F, Shu P. An immune-related gene pairs signature predicts overall survival in serous ovarian carcinoma. *OncoTargets therapy.*

- 2019;12:7005.
22. He Y, Dang Q, Li J, Zhang Q, Yu X, Xue M, Guo W. Prediction of hepatocellular carcinoma prognosis based on expression of an immune-related gene set. *Aging*. 2020;12(1):965.
 23. Wang Z, Zhu J, Liu Y, Liu C, Wang W, Chen F, Ma L. Development and validation of a novel immune-related prognostic model in hepatocellular carcinoma. *J Transl Med*. 2020;18(1):67.
 24. She Y, Kong X, Ge Y, Yin P, Liu Z, Chen J, Gao F, Fang S. Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma. *Cancer Cell Int*. 2020;20(1):1–10.
 25. Lin P, Guo Y-n, Shi L, Li X-j, Yang H, He Y, Li Q, Dang Y-w, Wei K-l, Chen G. Development of a prognostic index based on an immunogenomic landscape analysis of papillary thyroid cancer. *Aging*. 2019;11(2):480.
 26. Qiu H, Hu X, He C, Yu B, Li Y, Li J. **Identification and Validation of an Individualized Prognostic Signature of Bladder Cancer Based on Seven Immune Related Genes.** *Frontiers in Genetics* 2020, 11(12).
 27. Shen C, Liu J, Wang J, Zhong X, Dong D, Yang X, Wang Y. Development and validation of a prognostic immune-associated gene signature in clear cell renal cell carcinoma. *Int Immunopharmacol*. 2020;81:106274.
 28. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
 29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003;13(11):2498–504.
 30. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics*. 2019;11(1):123.
 31. Ranstam J, Cook J. **Kaplan-Meier curve.** *British Journal of Surgery* 2017, 104(4).
 32. Harrell FE Jr. **rms: Regression modeling strategies.** *R package version* 2016, 5(2).
 33. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565–74.
 34. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
 35. Gouw LG, Jones KB, Sharma S, Randall RL. Sarcoma immunotherapy. *Cancers*. 2011;3(4):4139–50.
 36. Mackall CL, Rhee EH, Read EJ, Khuu HM, Leitman SF, Bernstein D, Tesso M, Long LM, Grindler D, Merino M, et al. A pilot study of consolidative immunotherapy in patients with high-risk pediatric sarcomas. *Clin Cancer Res*. 2008;14(15):4850–8.
 37. Petitprez F, de Reyniès A, Keung EZ, Chen TW-W, Sun C-M, Calderaro J, Jeng Y-M, Hsiao L-P, Lacroix L, Bougouïn A, et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature*. 2020;577(7791):556–60.
 38. English WR, Lunt SJ, Fisher M, Lefley DV, Dhingra M, Lee Y-C, Bingham K, Hurrell JE, Lyons SK, Kanthou C, et al. Differential Expression of VEGFA Isoforms Regulates Metastasis and Response to

- Anti-VEGFA Therapy in Sarcoma. *Cancer research*. 2017;77(10):2633–46.
39. Croci S, Landuzzi L, Astolfi A, Nicoletti G, Rosolen A, Sartori F, Follo MY, Oliver N, De Giovanni C, Nanni P, et al. Inhibition of connective tissue growth factor (CTGF/CCN2) expression decreases the survival and myogenic differentiation of human rhabdomyosarcoma cells. *Cancer research*. 2004;64(5):1730–6.
40. Manara MC, Perbal B, Benini S, Strammiello R, Cerisano V, Perdichizzi S, Serra M, Astolfi A, Bertoni F, Alami J, et al. The expression of *ccn3*(*nov*) gene in musculoskeletal tumors. *Am J Pathol*. 2002;160(3):849–59.
41. Kosla J, Paňková D, Plachý J, Tolde O, Bicanová K, Dvořák M, Rösel D, Brábek J. Metastasis of aggressive amoeboid sarcoma cells is dependent on Rho/ROCK/MLC signaling. *Cell Commun Signal*. 2013;11:51.
42. JE O, DI G, GD S, KS EK. P, S N, DP C: VEGF inhibits T-cell development and may contribute to tumor-induced immune suppression. *Blood*. 2003;101(12):4878–86.
43. Y T, Y T, T DK, A KES. K, T D, H W, H N, K S: Targeting VEGFR2 with Ramucirumab strongly impacts effector/ activated regulatory T cells and CD8 T cells in the tumor microenvironment. *J immunother Cancer*. 2018;6(1):106.
44. A P-G NHBSDC. F L, O F: CYR61 triggers osteosarcoma metastatic spreading via an IGF1R β -dependent EMT-like process. *BMC Cancer*. 2019;19(1):62.
45. J F, T S: **The Oncoprotein Gankyrin/PSMD10 as a Target of Cancer Therapy**. *Advances in experimental medicine and biology* 2019, 1164:63–71.
46. F L, K L, X Z, J Z, W C, Y Q, Y Z, Y Z, Y Z: **miR-214 regulates papillary thyroid carcinoma cell proliferation and metastasis by targeting PSMD10**. *International journal of molecular medicine* 2018, **42**(6):3027–3036.
47. S TFJF. Y, Y M, N S, Y O, H Y, S O, K W, M O et al: Macrophage infiltration predicts a poor prognosis for human ewing sarcoma. *Am J Pathol*. 2011;179(3):1157–70.
48. dH DHVAS, tD ECMSJSMM, CJ HG, PC M. H, RM E, MJ vT. et al: NK cells recognize and lyse Ewing sarcoma cells through NKG2D and DNAM-1 receptor dependent pathways. *Molecular immunology* 2008, 45(15):3917–25.
49. He R-q, Wei Q-j, Tang R-x, Chen W-j, Yang X, Peng Z-g, Hu X-h, Ma J, Chen G. Prediction of clinical outcome and survival in soft-tissue sarcoma using a ten-lncRNA signature. *Oncotarget*. 2017;8(46):80336.
50. Callegaro D, Miceli R, Bonvalot S, Ferguson P, Strauss DC, Levy A, Griffin A, Hayes AJ, Stacchiotti S, Le Pechoux C. Development and external validation of two nomograms to predict overall survival and occurrence of distant metastases in adults after surgical resection of localised soft-tissue sarcomas of the extremities: a retrospective analysis. *The Lancet Oncology*. 2016;17(5):671–80.
51. Tape TG. Interpretation of diagnostic tests. *Ann Intern Med*. 2001;135(1):72–2.

Figures

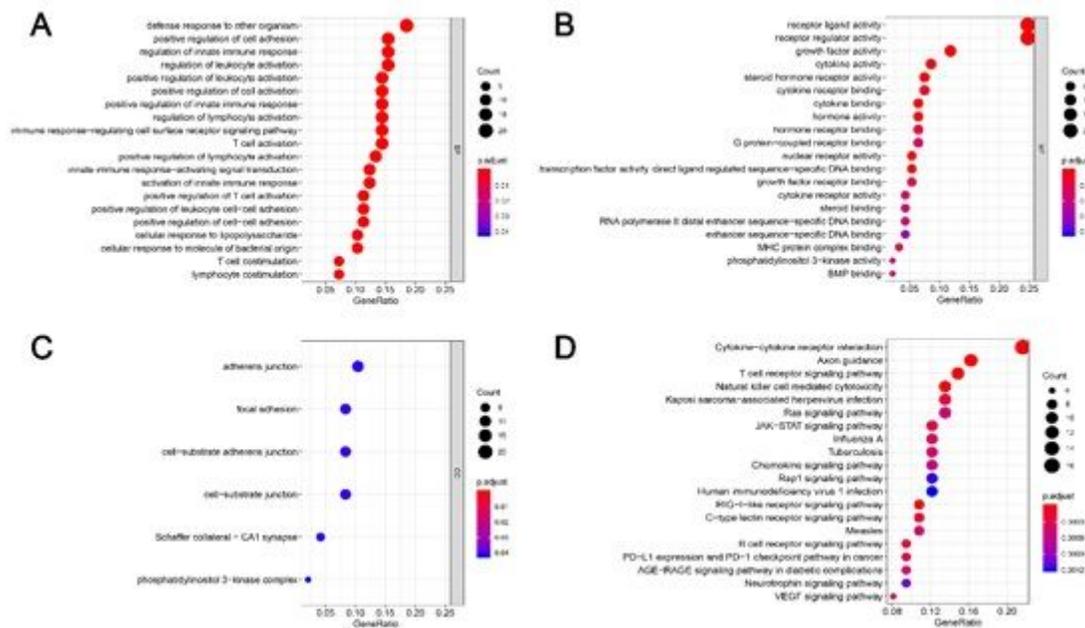


Figure 1

The enrichment analysis of overall survival-related immune-related genes in sarcoma patients. A-C The top 20(if possible) most significant Gene ontology analysis; A. biological process, B. molecular function, C. cellular component. D. The top 20 most significant Kyoto Encyclopedia of Genes and Genomes pathways.

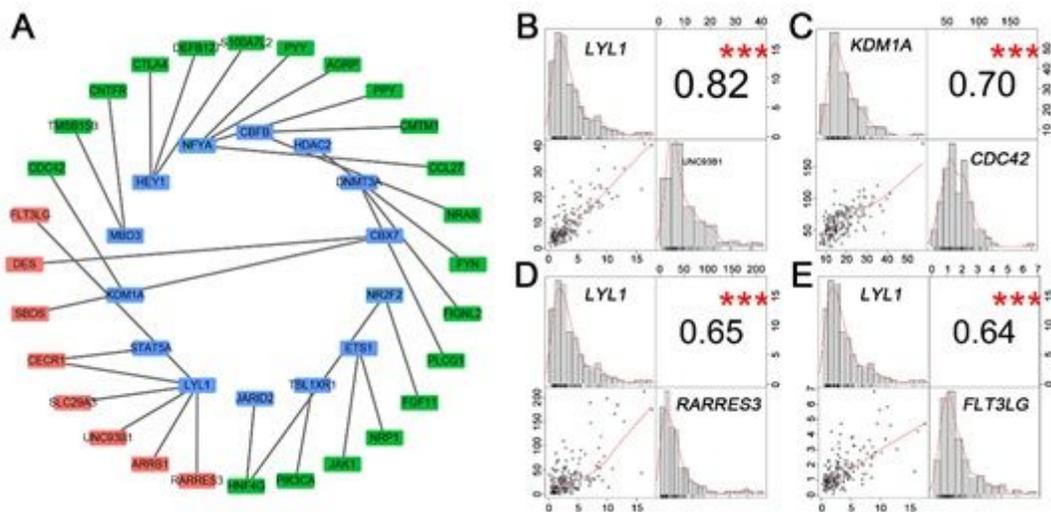


Figure 2

Transcription factors-based regulatory network. A. Regulatory network of prognostic transcription factors and prognostic immune-related genes. The rectangle represent prognostic transcription factors, the green rectangle represent poor prognosis-related immune-related genes, and the red rectangle represent good prognosis-related immune-related genes. B-E. Four examples to show the distribution of prognostic transcription factors and prognostic immune-related genes.

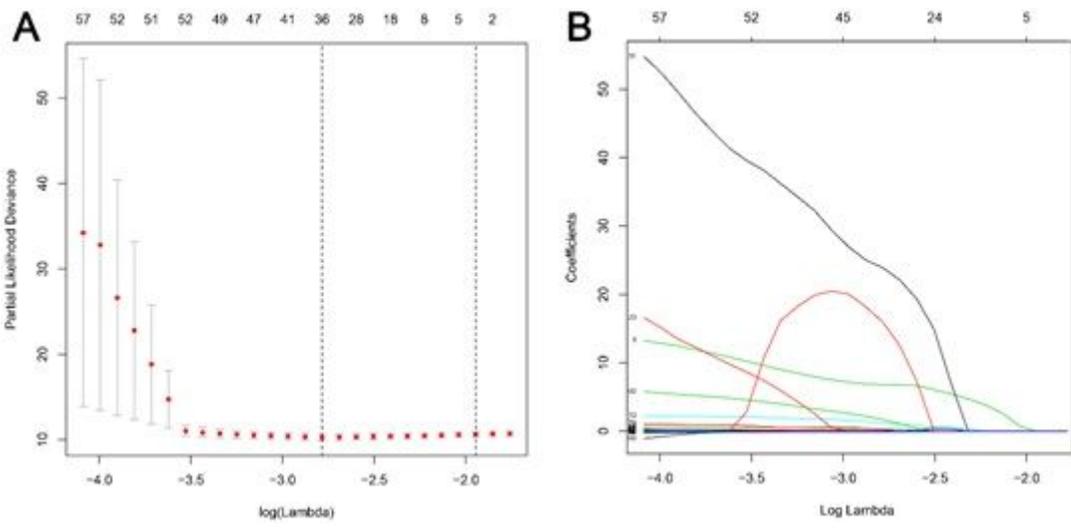


Figure 3

LASSO regression to select the most significant prognosis-related immune-related genes. LASSO: least absolute shrinkage and selection operator.

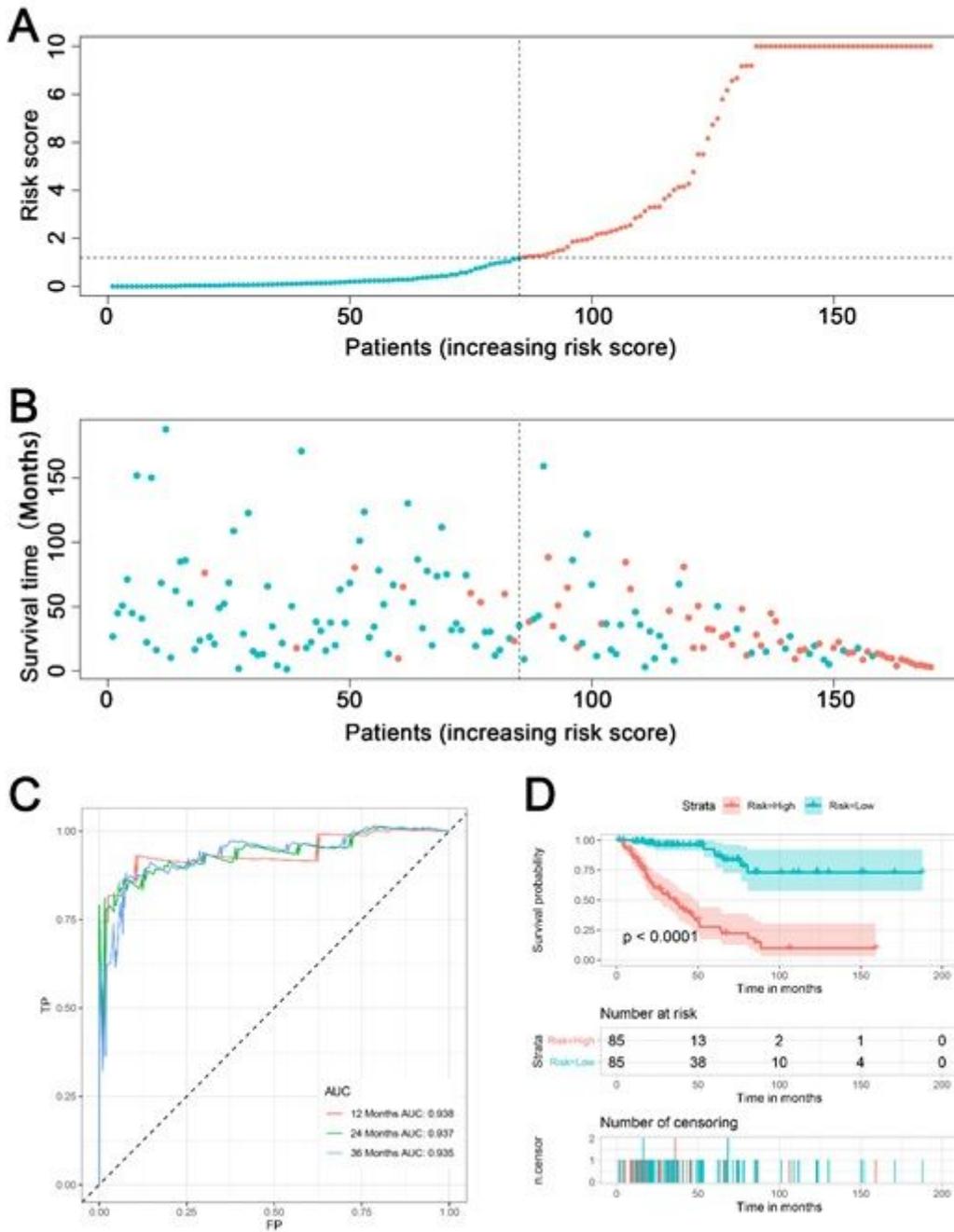


Figure 4

Construction of an immune-related prognostic signature in the training cohort. A. The risk score distribution of patients in the training cohort. B. Survival status scatter plots for patients in the training cohort. C. Time-dependent ROC curve analysis of the immune-related prognostic signature. D. Kaplan-Meier of overall survival in high- and low-risk group of the sarcoma patients in the training cohort.

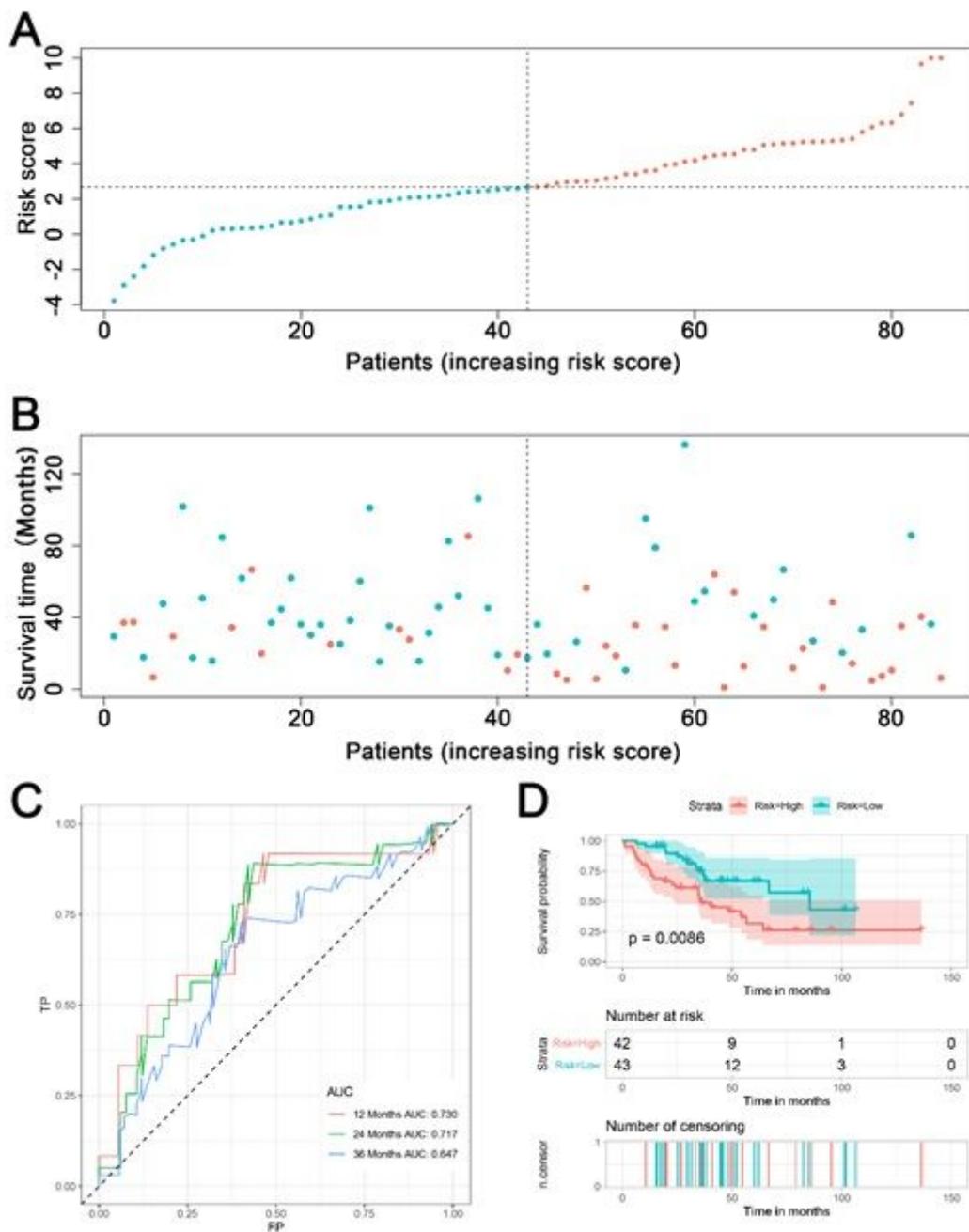


Figure 5

Validation of immune-related prognostic signature in the testing cohort. A. The risk score distribution of patients in the testing cohort. B. Survival status scatter plots for patients in the testing cohort. C. Time-dependent ROC curve analysis of the immune-related prognostic signature. D. Kaplan-Meier of overall survival in high- and low-risk group of the sarcoma patients in testing cohort.

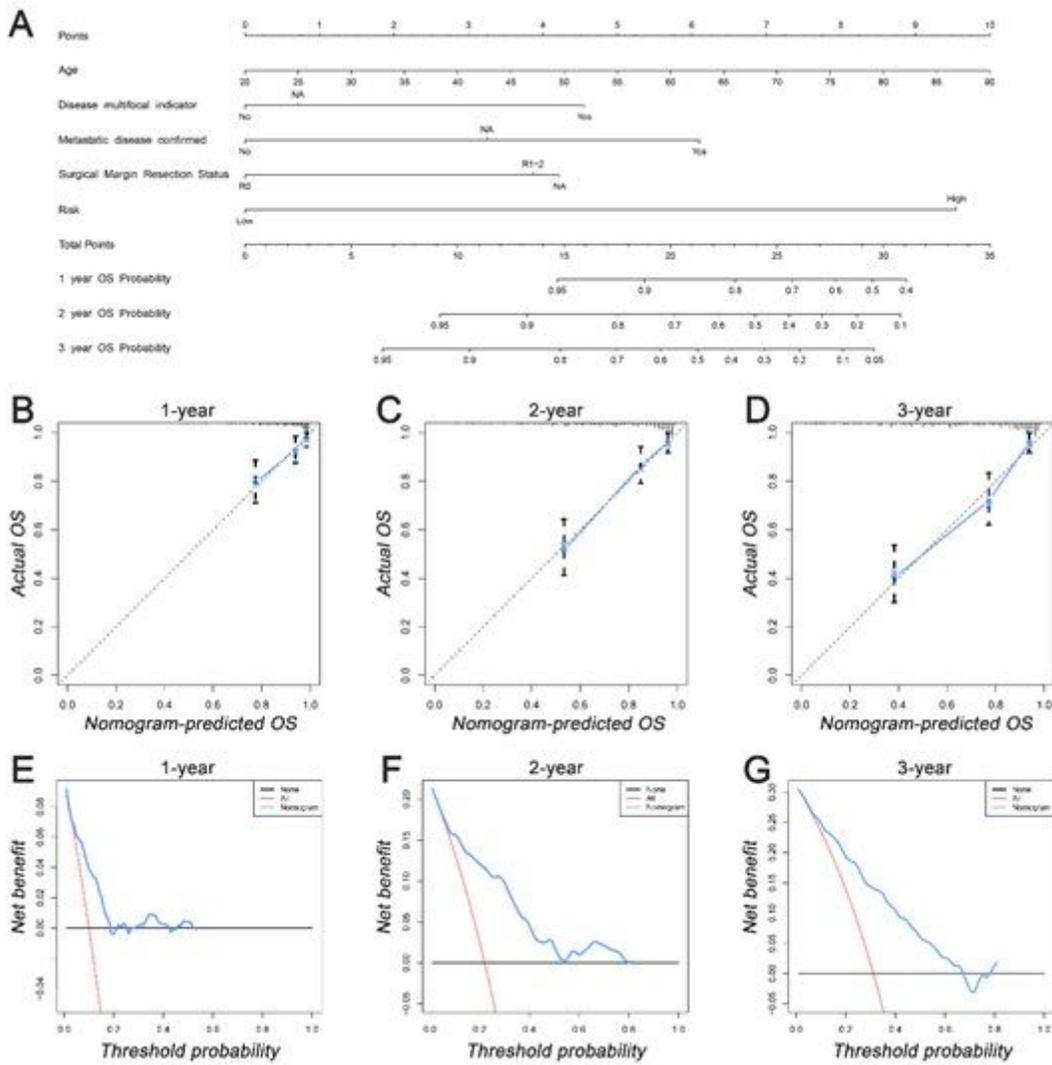


Figure 6

A nomogram based on immune-related prognostic signature and prognostic clinical variables. A. A nomogram for the prediction of the prognosis of sarcoma patients at 1, 2, and 3 years. B-D. Calibration curves of nomogram. E-G. Decision curve analysis of nomogram.

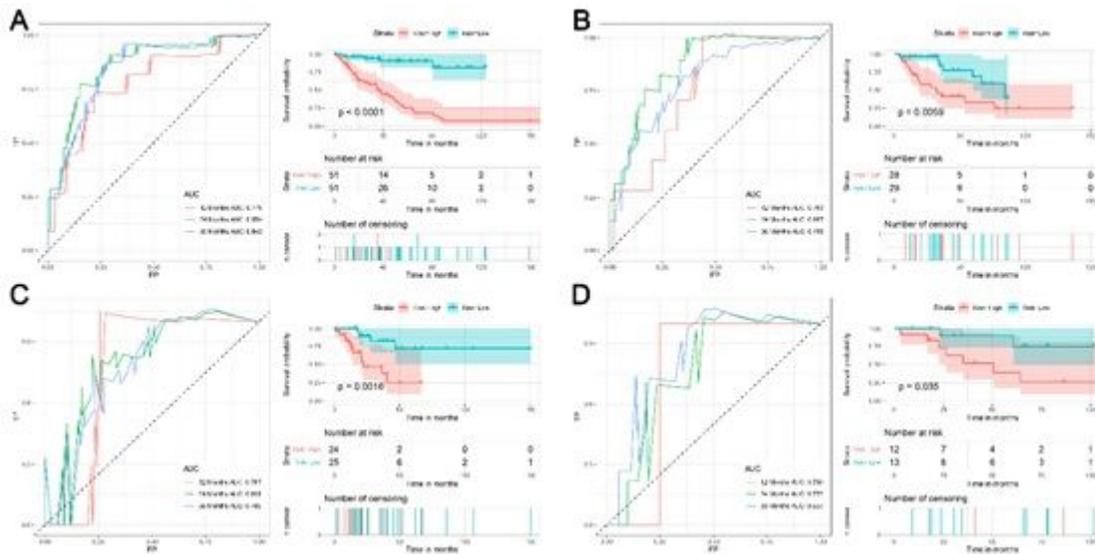


Figure 7

Subgroup analysis of nomogram. A. The ROC curve and Kaplan-Meier curve to show the prognostic value of nomogram in leiomyosarcoma. B. The ROC curve and Kaplan-Meier curve to show the prognostic value of nomogram in dedifferentiated liposarcoma. C. The ROC curve and Kaplan-Meier curve to show the prognostic value of nomogram in undifferentiated pleomorphic sarcoma. D. The ROC curve and Kaplan-Meier curve to show the prognostic value of nomogram in myxofibrosarcoma

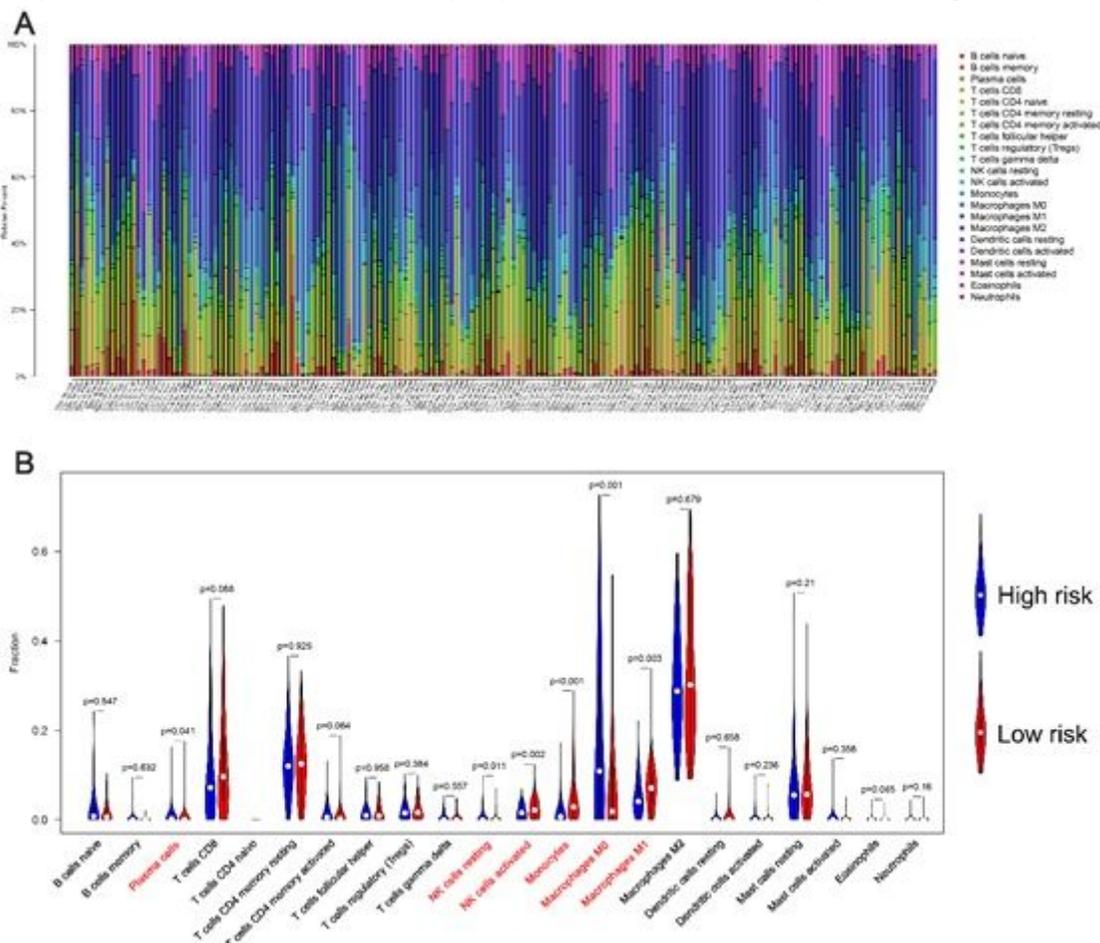


Figure 8

Immune infiltration status of immune-related risk groups. A. The distribution of 22 immune cells in 169 sarcoma patients. B. Comparison of immune infiltration cell between high- and low-risk group.

Supplementary Files

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

- [Supplementary.3Correlationresultes.xlsx](#)
- [Supplementary.2OSrelatedTFs.xlsx](#)
- [Supplementary.1OSrelatedIRGs.xlsx](#)
- [Supplementary.4MultiCox.xlsx](#)