

A novel immune-related gene signature predicting survival in sarcoma patients

Haoyu Ren,¹ Alexandr V. Bazhin,^{1,2} Elise Pretzsch,¹ Sven Jacob,¹ Haochen Yu,¹ Jiang Zhu,³ Markus Albertsmeier,¹ Lars H. Lindner,⁴ Thomas Knösel,⁵ Jens Werner,^{1,2} Martin K. Angele,^{1,2} and Florian Bösch¹

¹Department of General, Visceral, and Transplant Surgery, Ludwig-Maximilians-University Munich, 81377 Munich, Germany; ²German Cancer Consortium (DKTK), Partner Site Munich, Munich, Germany; ³Department of Liver Surgery and Liver Transplantation Centre, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Chengdu, China; ⁴Department of Medicine III, SarKUM, University Hospital, LMU Munich, Munich, Germany; ⁵Institute of Pathology, University Hospital, LMU Munich, Munich, Germany

Sarcomas are a heterogeneous group of rare mesenchymal tumors. The migration of immune cells into these tumors and the prognostic impact of tumor-specific factors determining their interaction with these tumors remain poorly understood. The current risk stratification system is insufficient to provide a precise survival prediction and treatment response. Thus, valid prognostic models are needed to guide treatment. This study analyzed the gene expression and outcome of 980 sarcoma patients from seven public datasets. The abundance of immune cells and the response to immunotherapy was calculated. Immune-related genes (IRGs) were screened through a weighted gene co-expression network analysis (WGCNA). A least absolute shrinkage and selection operator (LASSO) Cox regression was used to establish a powerful IRG signature predicting prognosis. The identified IRG signature incorporated 14 genes and identified high-risk patients in sarcoma cohorts. The 14-IRG signature was identified as an independent risk factor for overall and disease-free survival. Moreover, the IRG signature acted as a potential indicator for immunotherapy. The nomogram based on the risk score was built to provide a more accurate survival prediction. The decision tree with IRG risk score discriminated risk subgroups powerfully. This proposed IRG signature is a robust biomarker to predict outcomes and treatment responses in sarcoma patients.

INTRODUCTION

Sarcomas arise from the skeleton and the soft tissue subdividing in various histologic subtypes and have an increasing incidence of 7.7 cases/100,000 individuals per year.^{1,2} In the European Union, about 27,908 new cases per year are registered.^{3,4} The therapeutic approaches differ between the subgroups, but surgery offers the only chance of cure.^{5,6} However, in large series, recurrence rates are as high as up to 45%, which underlines the importance of a precise diagnostic and therapeutic workup.^{7,8} In this respect, the anatomic heterogeneity complicates the standardization of diagnosis and therapy. Currently the main prognostic criteria for sarcomas are tumor grade, size, histological subtype, and resection margin status,⁹ which makes it challenging to precisely assess the

individual prognosis. Consequently, it is imperative to develop robust biomarkers to predict the prognosis and the therapeutic response of sarcoma patients.

Increasing evidence suggests that tumor-infiltrating immune cells determine the clinical outcome of immunotherapy in different sarcoma subtypes.^{10–13} Immunosurveillance is mediated in part by two major groups of T cells (CD4+, CD8+), which play a pivotal role in tumor formation, tumor progression, and therapy.¹⁴ Moreover, previous studies demonstrated that an abundance of specific T cell subsets exerts favorable effects on the immune status of patients, improving responses to anticancer treatment and, subsequently, prognosis.^{15,16} However, although it is still not clear which genes regulate the microenvironment of sarcomas and the abundance of tumor-infiltrating lymphocytes (TILs), it is reasonable to assume that immune-cell-related genes influence the prognosis of sarcoma patients. In this regard, recently an accurate and unique method called Immune Cell Abundance Identifier (ImmuCellAI) has been implemented. ImmuCellAI allows us to calculate the abundance of 24 tumor-related immune cell types from gene expression data,¹⁷ which enables analyzing the association between the abundance of immune cells, the underlying genes, and the prognosis of cancer patients. Moreover, ImmuCellAI seems to facilitate an accurate prediction of the response to immunotherapy in cancer patients.¹⁷

Thus, the present study aimed to develop a prognostically relevant immune-cell-related gene signature based on sarcoma-related data from The Cancer Genome Atlas (TCGA-SARC) and validate this signature in six independent sarcoma cohorts. In addition, we aimed to analyze the association between immunological features and the gene signature to predict the response to immunotherapy in the sarcoma cohorts. Finally, based on the gene signature and

Received 26 August 2021; accepted 7 December 2021;
<https://doi.org/10.1016/j.omto.2021.12.007>.

Correspondence: Martin K. Angele, Department of General, Visceral, and Transplant Surgery, Ludwig-Maximilians-University, 81377 Munich, Germany.
E-mail: martin.angele@med.uni-muenchen.de

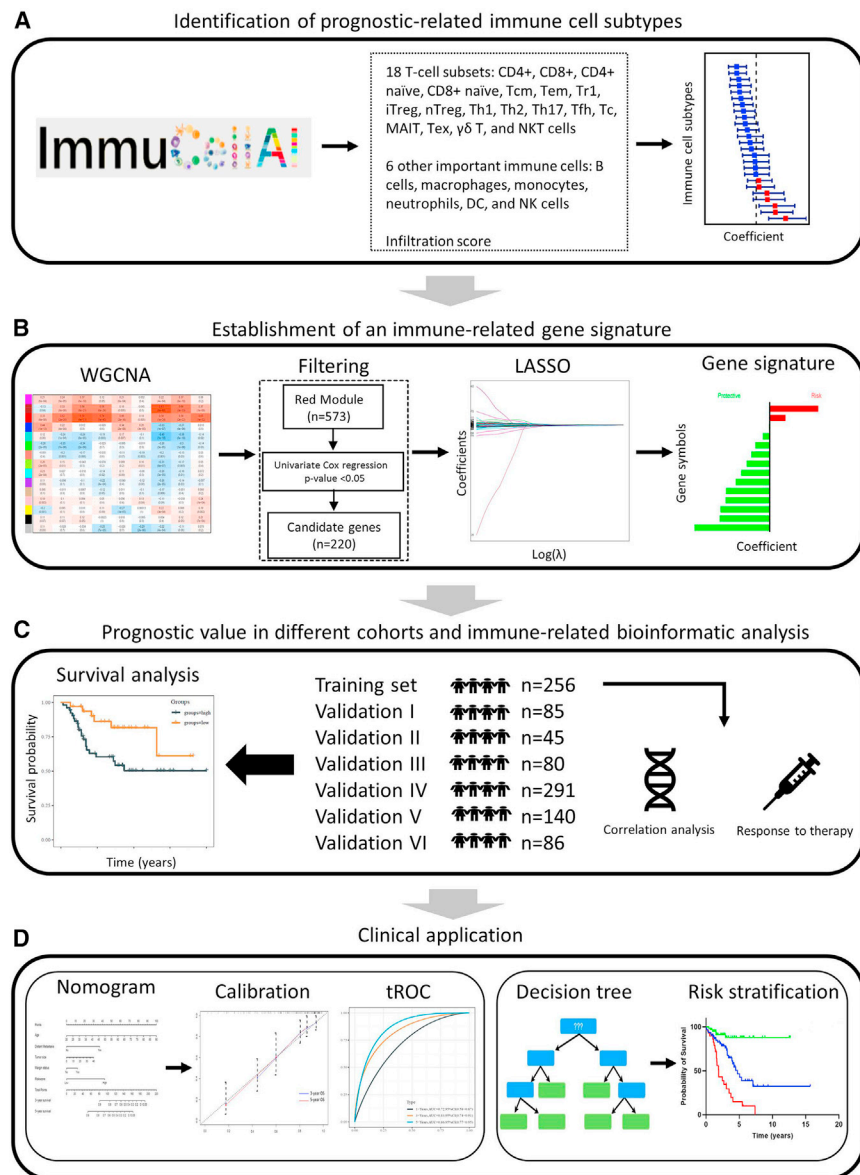


Figure 1. Flow chart of the study design

(A) Identification of the immune cell subtypes with prognostic relevance. (B) The approaches used to establish an immune-related gene signature for prognosis. (C) The prognostic value of the gene signature was investigated in different cohorts. (D) This application of personalized medicine might be used in daily routine. WGCNA, weighted gene co-expression network analysis; tROC, time-dependent receiver operating characteristic.

select the gene module most relevant to these protective immune cell subtypes. In total, 573 genes were extracted from the “MERed module” and further analyzed by univariate Cox regression and LASSO Cox regression analyses. As a result, 14 immune-related prognostic genes were identified and developed into a multiple gene signature with the corresponding coefficients (Figure 1B). Subsequently, the prognostic value of this immune-related gene (IRG) signature was investigated in the TCGA-SARC cohort and six independent validation cohorts. Moreover, the immunologic features and the response to immunotherapy were also explored in high- and low-risk groups based on the signature risk score (Figure 1C). In order to apply the gene signature to clinical practice, the nomogram and decision tree combined clinical information, and IRG risk scores were established to determine the survival outcome and risk levels for sarcoma patients (Figure 1D).

Prognostic relevance of T cell subtypes and NK cells

The univariate Cox regression analysis demonstrated that the abundance scores of eleven immune cells and the infiltration score were significantly associated with overall survival (OS) in the training set. Nine of these

clinicopathological data, we sought to establish a novel nomogram and decision tree to improve risk stratification in daily clinical practice.

RESULTS

Flow chart of the study design

First, the abundance scores of 24 immune cell types and the infiltration scores were downloaded from the ImmuCellAI database and transformed into a Z score using the scale method in R project. The univariate Cox proportional regression analysis was used to identify the protective immune cell types that promote a better prognosis in TCGA-SARC patients (Figure 1A). Then, a weighted gene co-expression network analysis (WGCNA) was performed to

scores, including natural killer (NK), CD4+ T, T follicular helper (Tfh), central memory T (Tcm), T helper type 1 (Th1), natural killer T (NKT), CD8+ T, cytotoxic T cell (Tc), and the infiltration score, correlated positively with survival rates (Figure 2A). The 256 patients of the TCGA-SARC cohort were divided into high-score and low-score groups based on the optimal immune-score-cut-off value, and the high-score group exhibited a better prognosis than the low-score group (Figures 2B–2J).

Construction of the IRG signature for OS

A WGCNA was performed to identify the most significant gene modules related to the nine protective immune scores. When $\beta = 5$ (a soft threshold), the scale-free R2 was 0.85, generating a scale-free co-

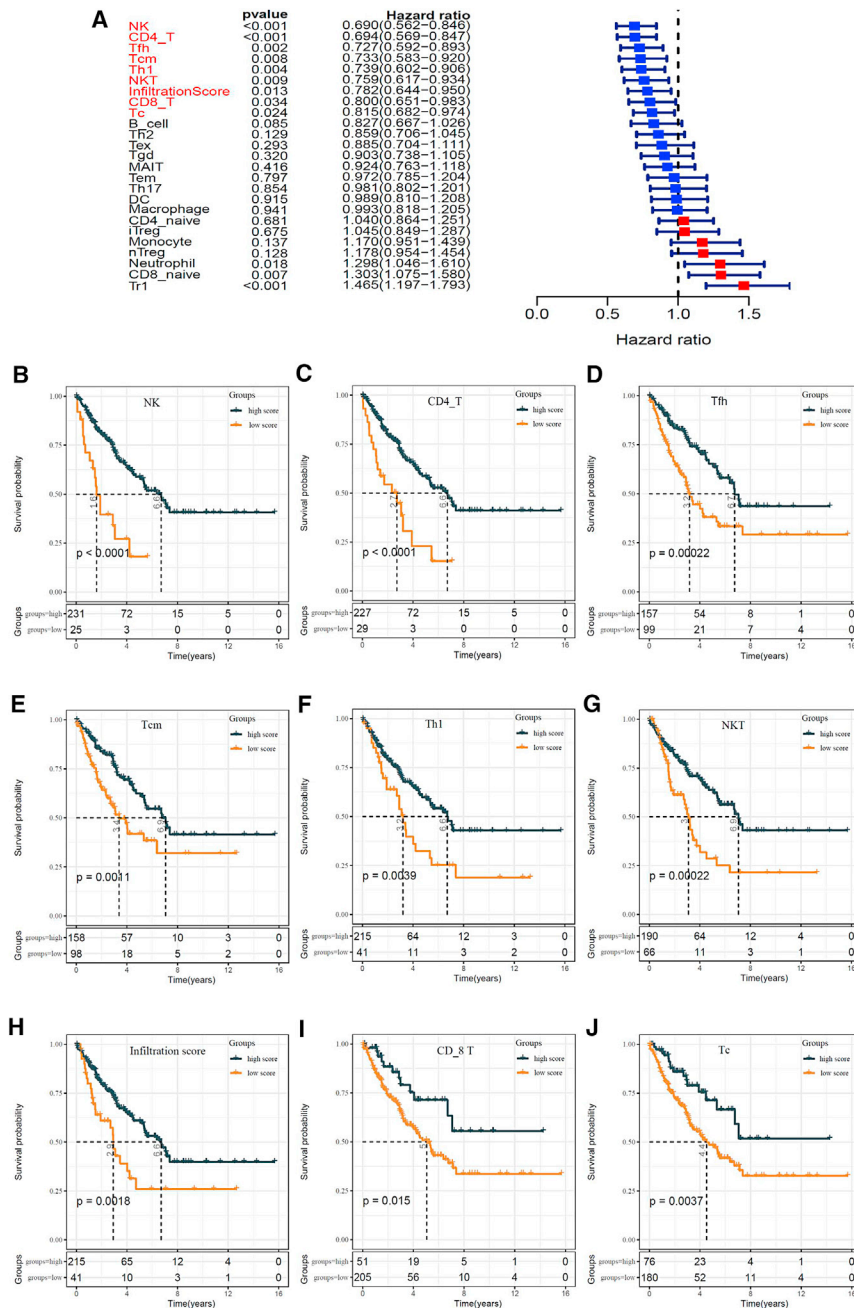


Figure 2. Immune cells were identified as the primary protective factors for survival

(A) Univariate Cox regression analysis showed that 7 T cell types, NK cells, and the infiltration score were protective factors among the scores of various immune cell subsets. (B–H) Kaplan–Meier analysis indicated that patients with higher scores of NK (B), CD4+ T (C), Tfh (D), Tcm (E), Th1 (F), NKT (G), CD8+ T (I), and Tc (J) cells and infiltration score (H) had better OS ($p < 0.05$). NK, natural killer cells; Tfh, follicular helper T cells; Tcm, central memory T cells; Th1, T helper type 1 cells; NKT, natural killer T cells; Tc, cytotoxic T cells.

were identified. Most of these genes exerted a protective effect, and only two genes were negatively correlated (Figure 4A).

A LASSO Cox regression analysis was applied to identify the most robust prognostic genes and the corresponding coefficients (Figures S2A, S2B, and 4B). This analysis revealed 14 prognostically relevant genes (NT5DC2, REC8, GBP2, APOL2, CMA1, TAPBPL, TNFSF15, DHX58, ZNF1, SLC25A20, FCER1A, CRIP1, TNF, TRIM21). Thus, the risk score formula of the IRG signature reads as follows: $NT5DC2 \times 0.11572 + REC8 \times 0.037668 + GBP2 \times (-0.00153) + APOL2 \times (-0.01601) + CMA1 \times (-0.0341) + TAPBPL \times (-0.04386) + TNFSF15 \times (-0.05468) + DHX58 \times (-0.06778) + ZNF1 \times (-0.08096) + SLC25A20 \times (-0.10396) + FCER1A \times (-0.10488) + CRIP1 \times (-0.11699) + TNF \times (-0.11901) + TRIM21 \times (-0.17847)$. The patients of the training set were then divided into a high- and low-risk groups according to the optimal cut-off value generated from the IRG risk score formula. The high-risk group had a highly significant decreased OS compared with the low-risk group ($p < 0.0001$, hazard ratio [HR] = 8.98, 95% CI: 5.06–15.94; Figure 4C). Moreover, the areas under the curve (AUCs) of the IRG signature for 1-, 3-, and 5-year OS were 0.75, 0.8, and 0.75, respectively (Figure 4D). In

expression network (Figures S1A and S1B), and a total of 14 non-gray modules were obtained (black, blue, brown, cyan, green, green-yellow, magenta, pink, purple, red, salmon, tan, turquoise, and yellow) (Figures S1C and 3A). Among these 14 gene co-expression modules, the red module (“MRed module”) had the most significant correlation with all nine protective immune scores (Figure 3B). Subsequently, all genes from the red module were further analyzed by a univariate Cox regression, and thereby 220 prognosis-related genes

in addition, the disease-free survival (DFS) of high-risk patients was also highly significantly shorter than that of low-risk patients ($p < 0.0001$, HR = 2.9, 95% CI: 1.78–4.73; Figure 4E). The AUCs of 1-, 3-, and 5-year DFS were 0.64, 0.72, and 0.68, respectively (Figure 4F).

To confirm whether the IRG signature was an independent prognostic factor, multivariate Cox regression analyses were performed

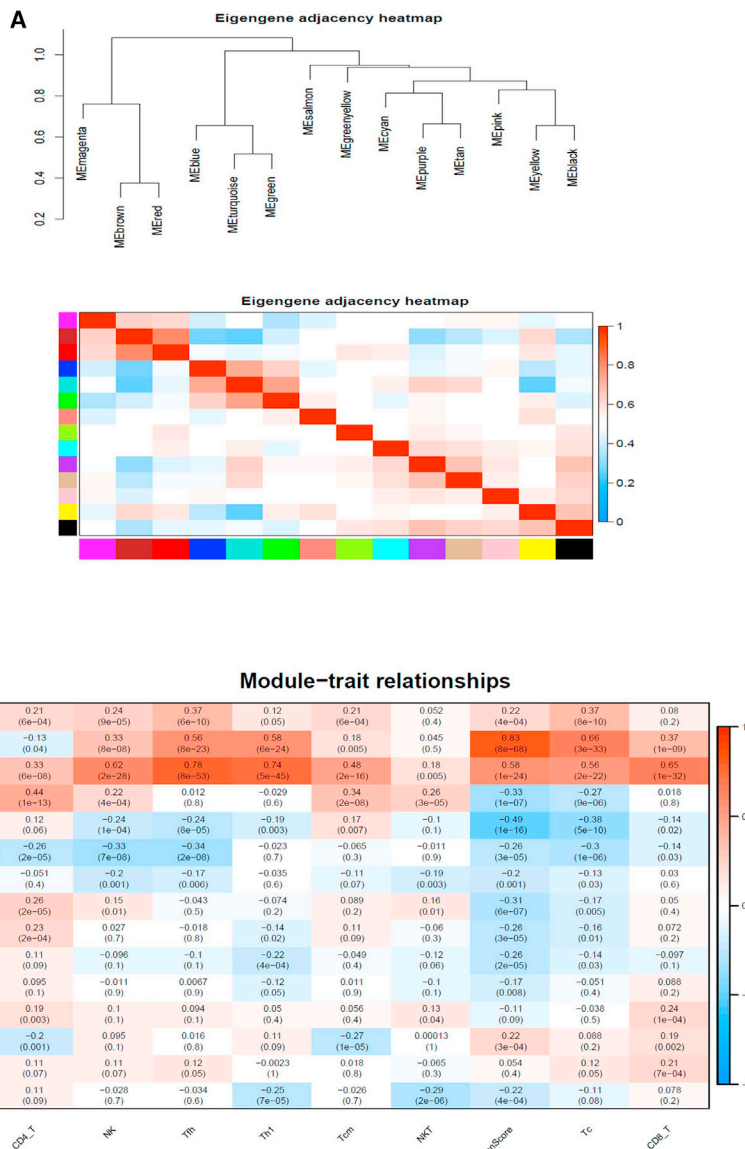


Figure 3. The construction of co-expression modules and module-trait relationships of sarcomas

(A) Visualizing the gene network using a heatmap plot. (B) Correlation of module eigengenes with all the protective immune cell scores. Each unit contains the corresponding correlation coefficient and p value.

high- and low-risk groups according to the generated risk score formula, and thereby the Kaplan–Meier survival analyses proved the prognostic relevance of the IRG signature in all six independent validation groups.

The OS of the high-risk groups was significantly worse compared with the low-risk groups of the validation cohorts I–III (TARGET–Osteosarcoma: HR = 1.86, 95% CI: 0.58–5.96, p = 0.021; GEO: GSE17674: HR = 21.18, 95% CI: 3.71–120.87, p < 0.0001; GEO: GSE119041: HR = 7.92, 95% CI: 1.37–45.87, p = 0.0042; Figures 6A, 6C, and 6E, respectively). The receiver-operating characteristic (ROC) curve analysis of the IRG signature showed a beneficial 1-, 3-, and 5-year AUC in each cohort (TARGET–Osteosarcoma: AUCs of 1-, 3-, and 5-year OS were 0.59, 0.61, and 0.55, respectively; GEO: GSE17674: AUCs of 1-, 3-, and 5-year OS were 0.81, 0.68, and 0.73, respectively; GEO: GSE119041: AUCs of 1-, 3-, and 5-year OS were 0.61, 0.77, and 0.77, respectively; Figures 6B, 6D, and 6F, respectively).

Moreover, the prognostic potential of the IRG signature for DFS was also validated. In the validation cohorts IV–VI, DFS in the high-risk group patients was significantly worse compared with those in the low-risk group (GEO: GSE71118: HR = 1.2, 95% CI: 0.83–1.76, p = 0.0071; GEO: GSE30929: HR = 7.44, 95% CI: 2.49–22.21, p <

in the patients (150 patients for OS and 133 patients for DFS) of the TCGA–SARC cohort with complete clinicopathological information. The multivariate Cox regression analysis showed that age, metastasis, and IRG signature risk score were significantly associated with OS (Figure 5A). Additionally, age, metastasis, resection margin status, and IRG signature risk score were proved to be independent risk factors for DFS (Figure 5B).

Validation of the IRG signature in six independent sarcoma cohorts

To determine the robustness of the identified 14 IRG signatures, validation survival analyses were performed in six independent sarcoma cohorts. The patients of these six validation groups were divided into

0.0001; GEO: GSE40025: HR = 3.32, 95% CI: 1.81–6.12, p < 0.0001; Figures 7A, 7C, and 7E, respectively). Likewise, the predictive accuracy for DFS yielded a valuable 1-, 3-, and 5-year AUC in the three independent validation cohorts (GEO: GSE71118: AUCs of 1-, 3-, and 5-year DFSs were 0.55, 0.54, and 0.57, respectively; GEO: GSE30929: AUCs of 1-, 3-, and 5-year DFS were 0.7, 0.72, and 0.66, respectively; GEO: GSE40025: AUCs of 1-, 3-, and 5-year DFSs were 0.68, 0.7, and 0.75, respectively; Figures 7B, 7D, and 7F, respectively).

The IRG signature serves as a potential indicator for immune characteristics and immunotherapy response

Since the 14-gene signature was closely related to immune cell abundance scores, the difference of immunological features and

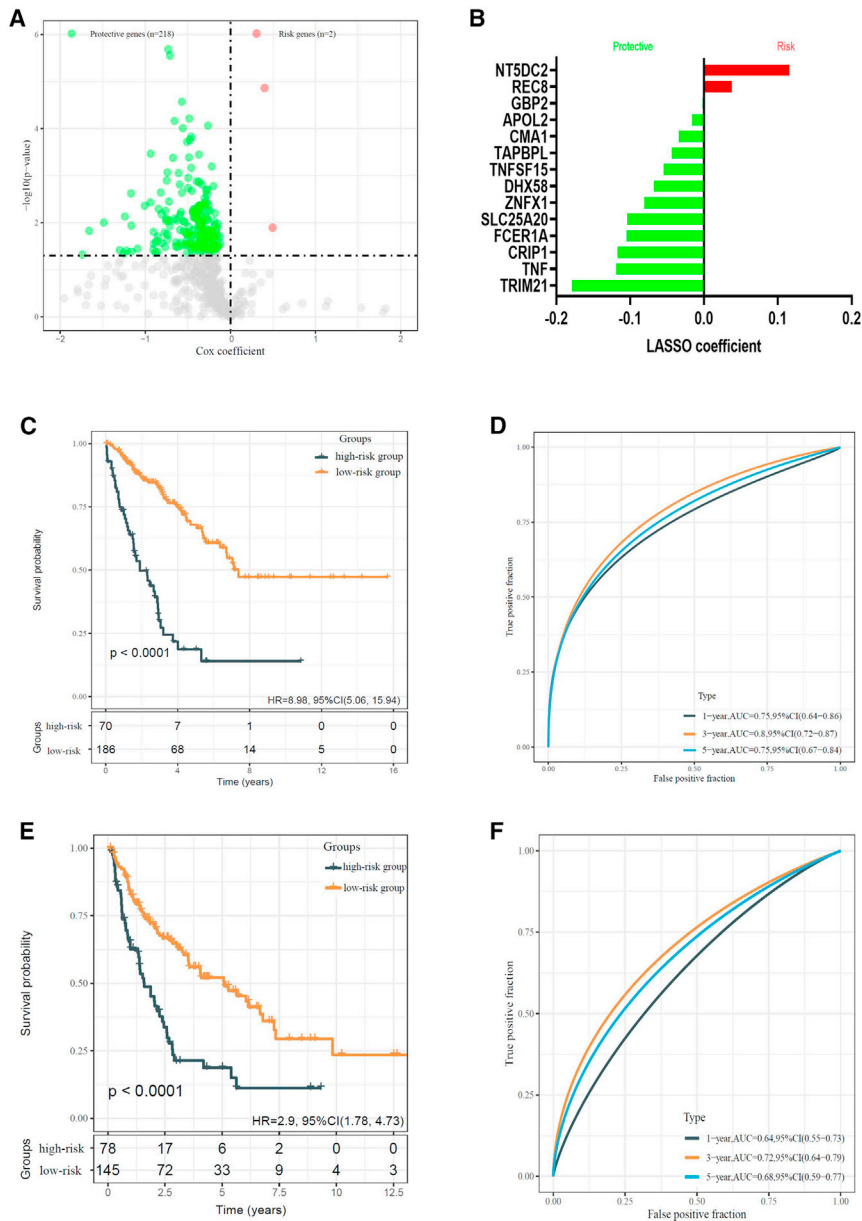


Figure 4. Establishment of the immune-related gene signature

(A) A total of 220 prognosis-related candidates were identified among 573 genes extracted from the red module (“MEred”). (B) Gene symbols and the corresponding LASSO coefficients of the IRG signature. (C) Kaplan–Meier curve revealed that patients with higher IRG risk scores exhibited worse OS. (D) The time-dependent ROC curves of the prognostic signature for OS in the training cohort. IRG, immune-related gene; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic. (E) Kaplan–Meier analysis indicated that patients with higher IRG risk scores showed shorter DFS. (F) The time-dependent ROC curves of the prognostic signature for DFS in the training cohort.

based on the 29 gene sets representing diverse immune cell types, functions, and pathways. The low-risk group showed a higher immune cell infiltration level compared with the high-risk group (Figure 6D). More importantly, there was a significant negative correlation between the IRG risk score and the expression levels of most immune checkpoint genes (ICGs) (Figure 8E). Furthermore, the immunotherapy response ratio between high- and low-risk groups was calculated by the “ImmuCellAI” online tool (<http://bioinfo.life.hust.edu.cn/ImmuCellAI#!/>). As shown in Figure 8F, 63.4% of patients in the low-risk group were sensitive to immunotherapy, which is significantly more patients compared with those in the high-risk group ($p < 0.0001$).

Nomogram and decision tree based on the IRG signature

In order to better apply the newly generated IRG signature to clinical practice, a nomogram for predicting OS was developed (Figure 9A). For this analysis, 150 patients of the TCGA-SARC cohort with complete clinicopathological information were used, and the risk score was incorporated additionally.

The nomogram included various clinicopathological information (age, metastasis, tumor size, resection margin) and the risk score derived from the IRG signature. The integrated nomogram showed a high predictive accuracy for OS with a calculated 1-year AUC of 0.72 (95% CI: 0.58–0.87), a 3-year AUC of 0.83 (95% CI: 0.74–0.91), and a 5-year AUC of 0.86 (95% CI: 0.77–0.95), respectively (Figure 9B). The calibration plots of observed and predicted probabilities of 3- and 5-year OS presented a good consistency (Figure S3A). Moreover, the decision curve analysis (DCA) showed that the net benefit of the nomogram was superior to a model built with traditional clinicopathological information (Figure S3B).

immunotherapy responses in high- and low-risk groups were investigated further. For this purpose, the ESTIMATE (estimation of stromal and immune cells in malignant tumor tissues using expression data) algorithm was applied to assess tumor purity and the ratio of infiltrating stromal/immune cells in tumor tissues based on gene expression profiles. The correlation analysis revealed that the risk score of the IRG signature was significantly negatively correlated with the immune score ($R = -0.48$, $p < 0.0001$), the stromal score ($R = -0.22$, $p = 0.00054$), and the ESTIMATE score ($R = -0.42$, $p < 0.0001$) (Figures 8A–8C).

Next, the single sample gene set enrichment analysis (ssGSEA) scores were used to assess the immunogenic activity in the sarcoma samples

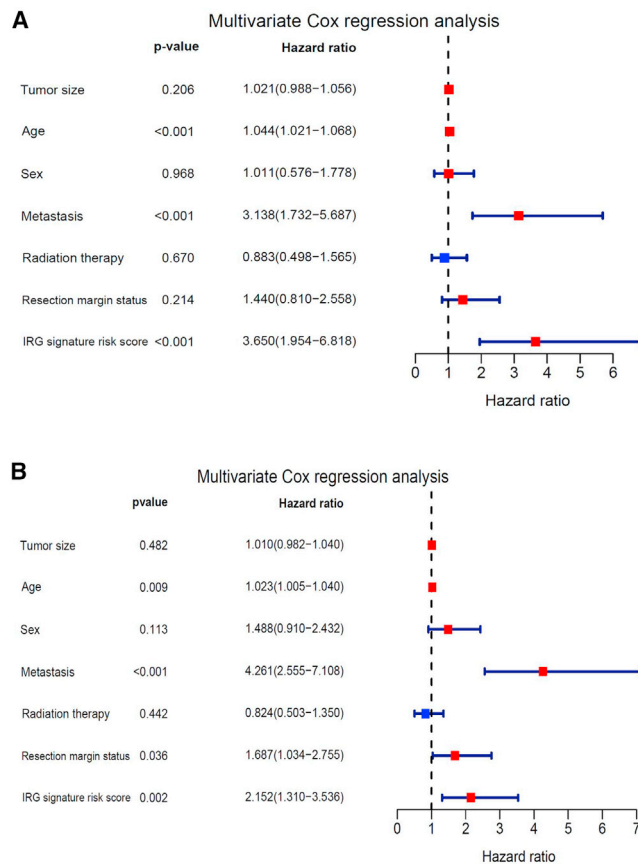


Figure 5. Independent prognostic factors for OS and DFS in the TCGA-SARC cohort

(A and B) Multivariate Cox regression analysis of the relationship between clinico-pathological features (including the risk score) and OS (A)/DFS (B) of patients in the TCGA-SARC dataset.

To further optimize the risk stratification for sarcoma patients, a recursive partitioning analysis was done, and a decision tree for OS was established. Finally, the IRG signature risk score along with age and distant metastasis remained in the decision tree, and three different risk subgroups were identified (Figure 9C). The Kaplan-Meier analysis proved the prognostic relevance of the established decision tree and showed a significant difference in OS among the three subgroups (overall: $p < 0.0001$; high risk versus intermediate risk: $p = 0.0002$, HR = 2.777, 95% CI: 1.44–5.34; intermediate risk versus low risk: $p = 0.0006$, HR = 4.181, 95% CI: 2.07–8.45; high risk versus low risk: $p < 0.0001$, HR = 10.77, 95% CI: 4.71–24.64; Figure 9D). In order to compare with the risk stratification based on the traditional clinico-pathological information alone, the second decision tree was developed, which contains three decision nodes of metastasis, age, and tumor size (Figure S4A). The survival analysis also demonstrated a significant difference in OS among the three risk subgroups generated by the second decision tree (overall: $p < 0.0001$; high risk versus intermediate risk: $p = 0.0002$, HR = 3.346, 95% CI: 1.19–9.40; intermediate risk versus low risk: $p < 0.0001$, HR = 5.191, 95% CI: 2.83–

9.52; high risk versus low risk: $p < 0.0001$, HR = 12.85, 95% CI: 3.05–54.14; Figure S4B).

DISCUSSION

Sarcomas are a heterogeneous malignant disease of mesenchymal origin, which account for approximately 1% of malignant tumors in adults and 15% of all children malignancies.¹⁸ The survival rates of each sarcoma subtype differ due to disease heterogeneity. However, establishing appropriate prognostic biomarkers is critical for precision medicine and risk stratification. In this respect, recent studies demonstrated that distinct immune gene signatures are prognostically relevant in cancer patients.^{19–22} However, the IRGs selected in most studies are derived from the ImmPort database or from ssGSEA. In the present study, the 14-IRG signature for sarcoma patients is based on ImmucellAI and WGCNA, which provides a novel and innovative method to identify prognostic biomarkers. Moreover, the newly presented IRG signature proved to act as a powerful and accurate prognostic parameter in seven different sarcoma cohorts. Furthermore, the risk score derived from the IRG signature was significantly correlated with the immune microenvironment and the response to immune checkpoint therapy. In order to quantify the risk assessment of each patient, a nomogram and a decision tree integrating the IRG signature risk score were established. Both proved to serve as valuable and robust predictive tools for the survival outcomes of sarcoma patients.

In the present study, a significant prognostic correlation between a high IRG signature score and shorter OS and DFS in seven different sarcoma cohorts was demonstrated. This is the most comprehensive study including the biggest collective of sarcoma patients analyzing seven public sarcoma datasets with almost 1,000 patients. In the developed 14-IRG signature, the expression level of twelve genes (TRIM21, TNF, CRIP1, FCER1A, SLC25A20, ZNFX1, DHX58, TNFSF15, TAPBPL, CMA1, APOL2, and GBP2) were negatively correlated with OS. TRIM21 (tripartite motif-containing protein 21) seems to be the most powerful gene since it had the most negative coefficient value. TRIM21 is involved in innate immunity and cancer progression. Previous studies found that the overexpression of TRIM21 inhibits tumor migration and invasion in breast cancer cell lines. Correspondingly, high TRIM21 expression rates were associated with increased survival rates of breast cancer patients.^{23,24} However, Zeng et al. reported that TRIM21 positively regulated osteosarcoma cell proliferation and autophagy.²⁵ Only two genes (NT5DC2 and REC8) act as oncogenes in the IRG signature. NT5DC2 (5'-nucleotidase domain-containing 2) is the most significant risk gene with a high coefficient. It has been shown to be involved in tumor development and progression in various types of carcinomas, such as colorectal cancer, lung cancer, hepatocellular carcinoma, and glioma.^{26–29} Nonetheless, only little is known about the exact function of NT5DC2 in sarcomas. In this respect, the present study firstly reported the potential prognostic value of NT5DC2 in sarcoma patients. Additionally, recent findings further support the present study. Hu et al. demonstrated that the overexpression of NT5DC2 promotes leiomyosarcoma progression and proliferation both *in vitro* and *in vivo*.³⁰ Therefore, NT5DC2

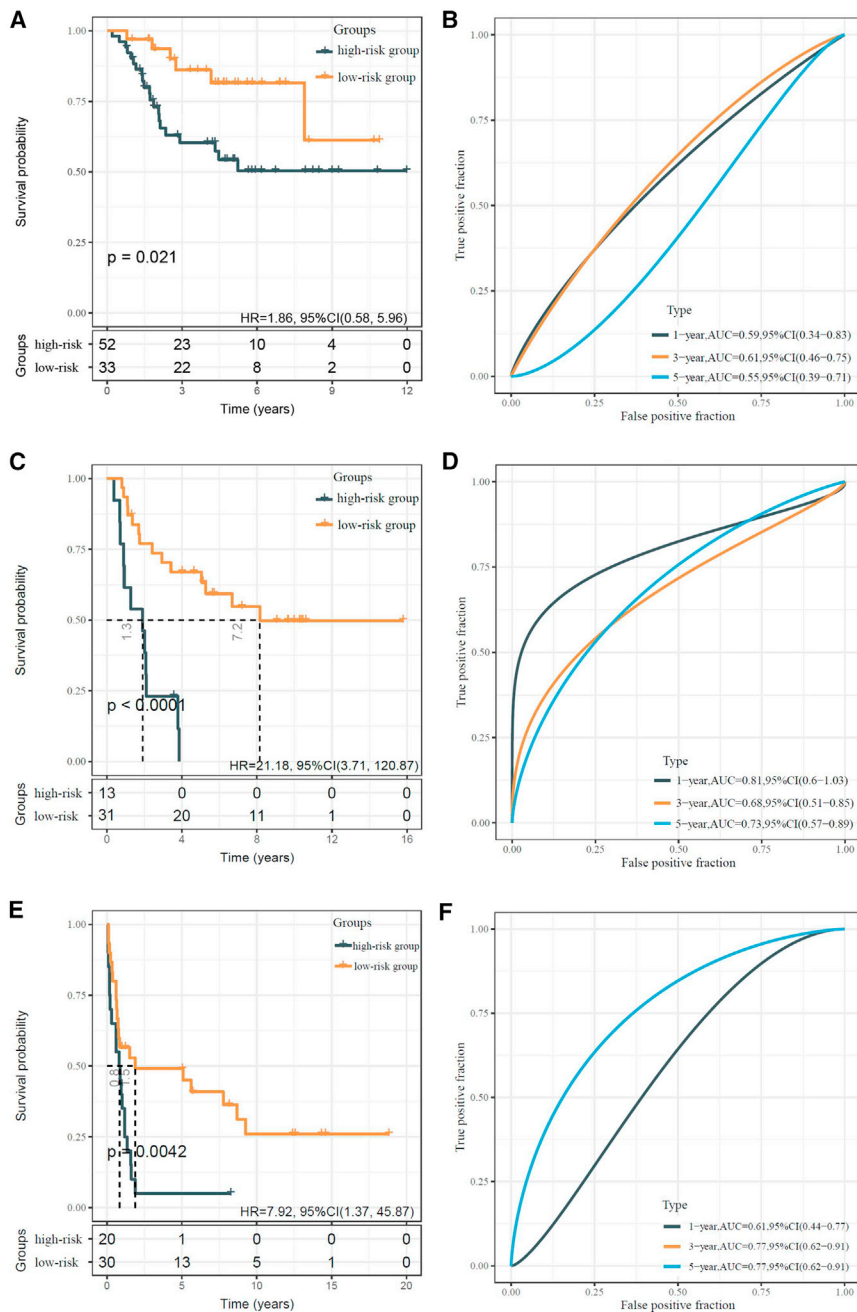


Figure 6. Validation of the OS-related prognostic relevance of the IRG signature with three external datasets

(A) Kaplan-Meier analysis of OS in the TARGET-Osteosarcoma cohort (validation cohort I). (B) The AUC of the risk score predicting the 1-, 3-, and 5-year OS in the TARGET-Osteosarcoma cohort. (C) Kaplan-Meier analysis of OS in the GEO: GSE17674 cohort (validation cohort II). (D) The AUC of the risk score predicting the 1-, 3-, and 5-year OS in the GEO: GSE17674 cohort. (E) Kaplan-Meier analysis of OS in the GEO: GSE119041 cohort (validation cohort III). (F) The AUC of the risk score predicting the 1-, 3-, and 5-year OS in the GEO: GSE119041 cohort.

risk score for 3-year OS was 0.8, which is outperforming previously published genetic risk scores.³¹ These findings indicate that the 14-gene signature may play a critical role in the development and progression of sarcomas and may serve as a prognostic tool. In contrast to previous studies,³²⁻³⁴ the novel IRG signature proved to be suitable for different sarcoma subtypes. The present results suggest that although sarcoma subtypes are partially heterogeneous, these entities share a common genetic background. Therefore, it is necessary to design further experimental research to investigate the biological function of the identified 14 genes in different sarcoma subtypes.

In this respect, adjuvant³⁵⁻³⁷ and neoadjuvant³⁸⁻⁴⁰ systemic therapies reduce the risk of metastatic spread significantly, subsequently leading to prolonged OS. Nonetheless, the findings of the individual trials are sometimes inconclusive, and treatment regimens differ. Recently, it has been demonstrated that a risk-adjusted chemotherapy is associated with significantly increased survival rates.⁴¹ The risk adjustment was done with a nomogram, which was based only on clinicopathological parameters like age, tumor size, and histology, particularly disregarding the tumor biology.^{41,42}

However, a histotype-tailored neoadjuvant chemotherapy had no prognostic benefit in a multicenter trial.³⁸ Thus, the identification of patients at risk is of utmost importance to design further studies investigating multimodal therapy. The present findings suggest that the introduced nomogram and the IRG signature might be appropriate tools to stratify patients.

In an attempt to better apply the IRG signature in a clinical setting, a nomogram and a novel decision tree were established. The results of the AUC, calibration curve, and decision curve analysis (DCA)

may be a novel biomarker and therapeutic target for sarcomas, but further studies are needed to verify its biological function and prognostic value.

Notably, the risk score of the IRG signature demonstrated a significant predictive capacity for survival outcomes in the training set and in all validation cohorts. A multivariate regression analysis showed that the IRG signature risk score was an independent risk factor for both OS and DFS. Moreover, the AUC of the IRG

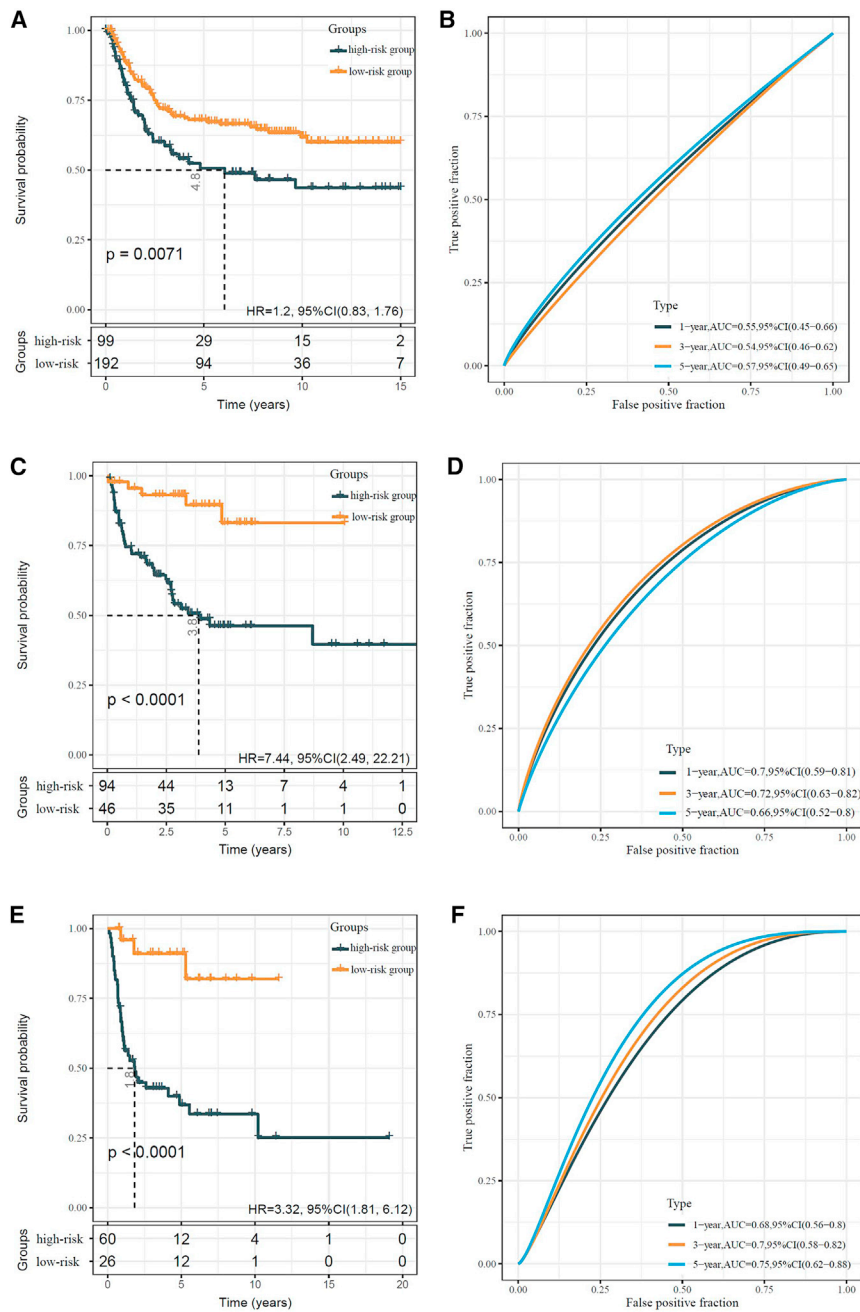


Figure 7. Validation of the DFS-related prognostic relevance of the IRG signature with three external datasets

(A) Kaplan–Meier analysis of DFS in the GEO: GSE71118 cohort (validation cohort IV). (B) The AUC of the risk score predicting the 1-, 3-, and 5-year OS in the GEO: GSE71118 cohort. (C) Kaplan–Meier analysis of OS in the GEO: GSE30929 cohort (validation cohort V). (D) The AUC of the risk score predicting the 1-, 3-, and 5-year DFS in the GEO: GSE30929 cohort (E) Kaplan–Meier analysis of DFS in the GEO: GSE40025 cohort (validation cohort VI). (F) The AUC of the risk score predicting the 1-, 3-, and 5-year DFS in the GEO: GSE40025 cohort.

grade, which has been identified as the most significant independent prognostic factor so far. Nonetheless, adding the IRG signature to the nomogram improved its predictive value significantly. To further improve the performance of predictive models, future studies need to add the IRG signature to their nomograms.

The decision tree was established to optimize risk stratification for OS in sarcoma patients. Finally, the decision tree was built including three components: metastasis acted as the most important determinant; age (cut-off point: 67.5 years) and the IRG risk score were additional factors. During the trimming step of the decision tree, it showed a strong relationship between the IRG signature risk score and high-risk sarcoma patients. The grouping results of the decision tree showed significant differences in the survival curves among the three different risk subgroups, demonstrating its great performance in patient stratification. Thus, using a simple clinical decision tree enables clinicians to identify patients with poor prognoses efficiently and accurately. Interestingly, the risk stratification generated by the IRG-risk-score-based decision tree model has no significant benefit compared with the model based on clinicopathological characteristics alone. However, the decision tree based on the IRG risk score represents a better predictive model because it is not only easier to interpret the threshold of the decision nodes from a medical point of view but also the structure is simpler to avoid model overfitting.

demonstrated that the nomogram including the IRG signature risk score is a powerful and accurate predictive tool for sarcoma patients. Two previously published studies also created similar predictive nomograms based on clinicopathological features or immunoscore. However, these studies evaluated the prognostic value in only two different sarcoma cohorts and reached lower AUC values compared with the present manuscript.^{43,44} Aside from the new IRG signature risk score, the nomogram contains established prognostic factors (age, metastasis, tumor size, and resection margin status).⁴⁵ However, due to missing data, the present nomogram did not include tumor

In general, sarcomas are not highly immunogenic tumors. Nonetheless, in previous studies, immune checkpoint inhibitors (ICIs) exhibited anti-tumor effects in 5%–40% of sarcoma patients.^{46–50} In the present study, the IRG risk score was negatively correlated with the immune score and the expression levels of ICGs. A previously published study revealed

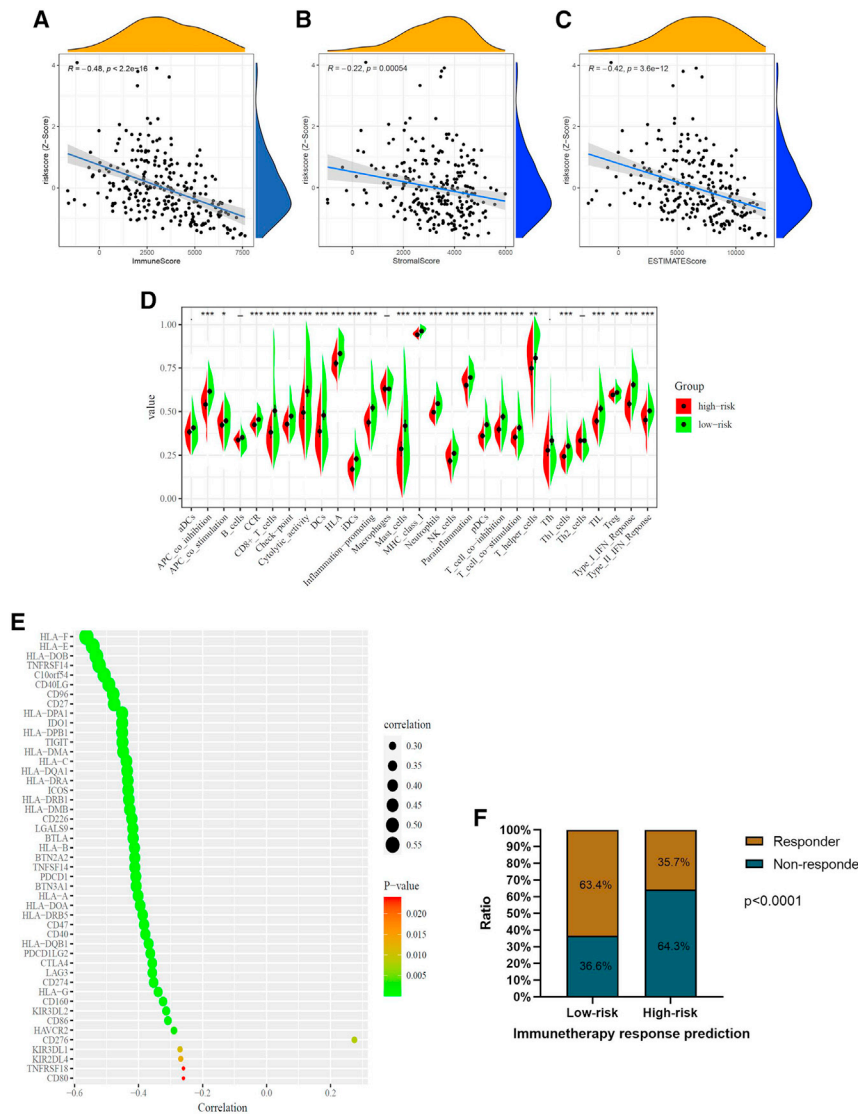


Figure 8. Immune microenvironment analysis and immunotherapy response prediction

(A–C) The correlation analysis indicated that the immune score ($R = -0.48$; $p < 0.001$) (A), the stromal score ($R = -0.22$; $p < 0.001$) (B), and the ESTIMATE score ($R = -0.42$; $p < 0.001$) (C) were significantly negatively correlated with the Z score of the IRG signature risk score. (D) 29 ssGSEA enrichment levels of the immune signatures in the high- and low-risk groups. (E) The correlation analysis bubble diagram showed that 46 immune checkpoint genes were negatively correlated with the risk score. (F) The ratio of immunotherapy response is significantly increased in the low-risk group compared with the high-risk group. ssGSEA, single sample gene set enrichment analysis.

are limitations as well. Firstly, owing to the rarity of sarcomas, a profound investigation of each sarcoma subtype is difficult. Thus, the present study includes all available subtypes of sarcomas introducing heterogeneity. Nonetheless, the IRG risk signature proved to exert a prognostic benefit in every subtype. Thus, further multicenter studies are highly demanded to analyze the risk signature in larger cohorts in a clinical setting. Secondly, the established nomogram and the decision tree were based on one training cohort (TCGA-SARC) and have not been tested in other validation cohorts due to incomplete clinical data. Furthermore, the ratio of immunotherapy responses was estimated *in silico*, which requires more reliable cohort studies in real-world. Nonetheless, this theoretical approach has been successfully validated in several solid tumor cohorts, such as melanoma, lung cancer, bladder cancer, and gastric cancer.¹⁷

In conclusion, using a WGCNA and a Lasso cox regression based on the abundance score of immune cells in sarcomas, a novel 14-IRG signature has been identified to be prognostically relevant. This signature serves as a robust and independent prognostic biomarker in various sarcoma subtypes. In addition, the present study developed a nomogram and a decision tree based on this IRG signature, which potentially act as an accurate and practical predictive tool to identify high-risk patients with low survival rates. Moreover, the IRG signature also had a reliable ability to predict the response to immunotherapy and may help to improve the efficacy of personalized immunotherapy in sarcoma patients.

MATERIALS AND METHODS

Data acquiring and preprocessing

Sarcoma-related clinical data and RNA sequencing (RNA-seq) data from TCGA-SARC were obtained from the UCSC Xena browser (<https://xenabrowser.net/>) and were used in the training set. Patients

that the expression of PD-L1 was associated with more PD-1-positive TILs in soft-tissue sarcomas (STSs).⁵¹ Additionally, the increased TIL counts due to neoadjuvant therapy were beneficial and seemed to prolong DFS of STS patients.⁵² The findings of the present study indicate that low-risk sarcomas with an increased immune cell infiltration have a higher immunogenicity and longer survival time. Furthermore, the prediction model for a response to immunotherapy indicates that low-risk sarcomas are more sensitive to immunotherapy. Taken together, these results underpin that the newly introduced IRG signature may serve as a biomarker to predict responses to ICIs. Nonetheless, this has to be shown in prospective trials. Thus, the IRG signature provides an additional opportunity to identify patients with sarcomas eligible for ICI therapy, which has to be addressed in future studies.

Although this study established a powerful IRG signature, a corresponding nomogram, and a decision tree for sarcoma patients, there

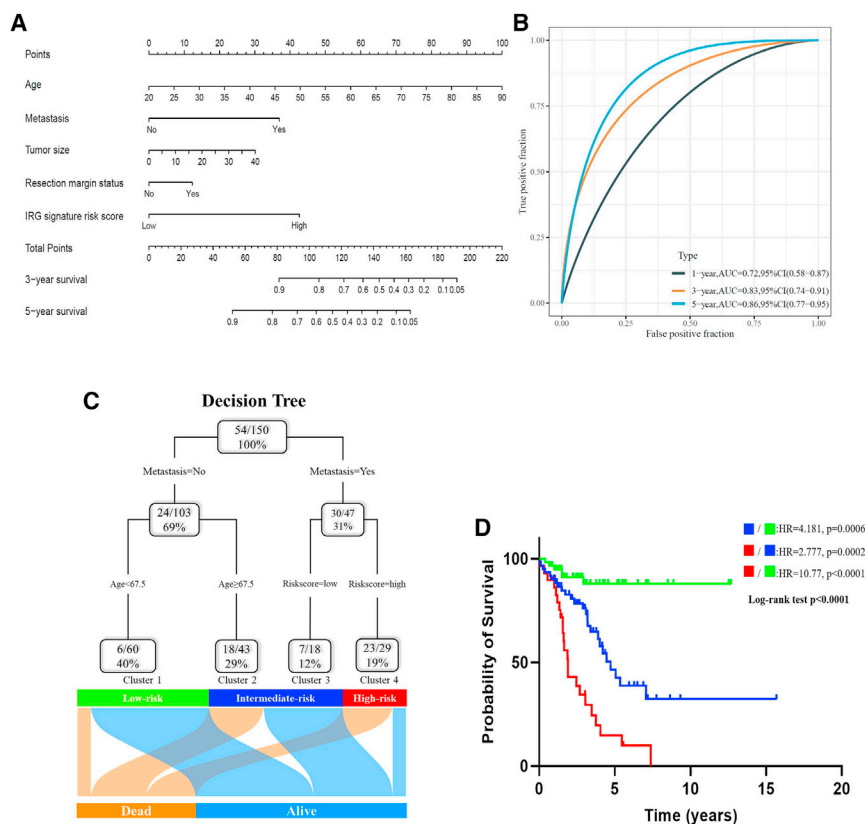


Figure 9. Combination of the IRG signature risk score and clinicopathological features improves prognosis prediction and risk stratification

(A) A nomogram was established to predict 3- and 5-year OS in individual sarcoma patients. (B) The AUC of the nomogram to predict the 3- and 5-year OS in the TCGA-SARC cohort. (C) A decision tree was constructed to categorize patients into three different risk levels. (D) Kaplan-Meier analysis of OS of the three different subgroups.

Gene selection and signature establishment

The infiltration score of each sample and the abundance score of 24 immune cell types including 18 T cell subtypes (CD4+, CD8+, CD4+ naive, CD8+ naive, Tcm, Tem, Tr1, induced regulatory T [iTreg], natural regulatory T [nTreg], Th1, Th2, Th17, Tfh, Tc, mucosal-associated invariant T [MAIT], exhausted T (Tex), gamma delta T ($\gamma\delta$ T), and NKT cells) and six other immune cells (B cell, NK cells, monocytes, macrophages, neutrophils, dendritic cells [DCs]) of TCGA-SARC were downloaded from the ImmuCellAI database. A univariate Cox proportional hazard regression analysis was performed to determine the prognostic value of the infiltration score of the different immune cells in the TCGA-SARC cohort using the R package “survival.”

Then, the gene co-expression network was constructed by the R package “WGCNA” using whole-transcriptome profiling data.⁵³ The threshold for the determination of weighted adjacency matrix was fixed at a soft power of 5 and a scale-free R2 >0.85, respectively.

In a next step, the topological matrix was constructed by using the topological overlap measure (TOM) in the R project. The minimum module size was 80 genes. The featured genes of the modules were calculated, and the similar modules were clustered and merged according to the module dissection threshold. Finally, module trait co-expression similarity and adjacency analyses were performed in the identified gene modules. Thereby, the module with the strongest correlation to the abundance score of the prognosis-related immune cells was identified for further analysis. With a threshold of the p value of univariate Cox regression <0.05, 220 candidate genes with prognostic values were extracted from the “MEred module.” Next, the LASSO Cox regression analysis was applied to select the most powerful prognostic genes of the “MEred module” with R package “glmnet.” The penalty regularization parameter λ was fixed using a 10-fold cross-validation to prevent an overfitting effect. The risk score of the IRG signature was calculated as follows:

$$Risk\ score = \sum_i Coefficient(mRNA_i) \times Expression(mRNA_i),$$

with missing information about survival time, survival status, and clinicopathological characteristics were excluded. In total, 256 sarcoma patients with complete survival data and gene expression profiles were included in the training set. The information about age, sex, resection margin status, tumor size, and radiation therapy was complete for 150 patients. These 150 patients were used for establishing the nomogram and the decision tree analysis.

Six validation groups were established, and the data was downloaded from two databases. The clinical data and gene expression files of 85 osteosarcoma patients derived from TARGET-Osteosarcoma (<https://ocg.cancer.gov/programs/target/projects/osteosarcoma>) were used as the first validation cohort. Moreover, the clinical information and gene expression data of five additional independent validation cohorts (GEO: GSE17674: 45 Ewing sarcoma patients; GEO: GSE119041: 50 uterine sarcoma patients; GEO: GSE71118: 291 various sarcoma patients; GEO: GSE30929: 140 liposarcoma patients; and GEO: GSE40025: 86 synovial sarcoma patients) were downloaded from GEO (<http://www.ncbi.nlm.nih.gov/geo/>). Among them, the TARGET-Osteosarcoma GEO: GSE17674 and GSE119041 cohorts were used to verify the OS, while GEO: GSE71118, GSE30929, and GSE40025 cohorts were used to validate the DFS. All RNA-seq and microarray data included in the present study were normalized and $\log_2(X+1)$ -transformed.

where i is the selected gene in the IRG signature. Patients were classified into high- and low-risk groups according to the optimal cut-off value, which was determined by the `surv_cutpoint` function of the R package “survminer.”

Analysis of immune function and response to immunotherapy

The scores of infiltrating stromal, immune cells, and tumor purity were calculated using the “ESTIMATE” package in R.⁵⁴ The ssGSEA algorithm was applied to comprehensively evaluate the immunological features and the infiltration of immune cells in the TCGA-SARC cohort.⁵⁵ The R packages (“GSVA,” “limma,” and “GSEA-Base”) were employed to analyze the gene expression profiles of sarcoma samples. Previously, 79 ICGs have been demonstrated to reliably be related to the response to immunotherapy. In order to analyze the response to immunotherapy, these 79 ICGs have been used in the present analysis.⁵⁶ The immunotherapy response prediction was based on the transcriptional data via the accurate and reliable online tool “ImmuCellAI.”¹⁷

Construction of the nomogram and the decision tree

A nomogram integrating clinical information and the immune-related risk signature was established using the R package “rms” to predict the survival probability for each patient.⁵⁷ Calibration curves of observed and predicted probabilities of 3- and 5-year OS were plotted to determine the discrimination of the nomogram. The decision tree was developed for risk stratification via a recursive partitioning analysis (RPA) in the R package “rpart.”⁵⁸

Bioinformatic and statistical analyses

R software v.4.0.0 (R Foundation for Statistical Computing, Vienna, Austria) and GraphPad Prism 8.4 (GraphPad Software, San Diego, CA, USA) were used to analyze data and plot figures. Univariate and multivariate Cox regression analyses were performed by the corresponding R packages. The Kaplan–Meier method and log rank tests were used to evaluate the difference of survival outcomes between different risk groups with the R package “survminer” and “survival.” The AUC derived from a time-dependent ROC (tROC) analysis was used to quantify the predictive power with the R package “survival-ROC.”⁵⁹ The HR and 95% confidence interval (CI) were calculated using log rank tests to confirm the risk score associated with survival time. Student’s t tests or one-way ANOVAs were used for the comparison of continuous variables between groups. A p value < 0.05 (two-tailed) was considered to indicate statistical significance, and high significance was indicated by a p value < 0.001 .

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omto.2021.12.007>.

ACKNOWLEDGMENTS

We greatly thank the China Scholarship Council (CSC) for supporting the research and work of H.R. (no. 202008080024) and H.Y. (no. 201908500101).

AUTHOR CONTRIBUTIONS

All authors reviewed the manuscript and agreed to the submission. H.R., M.K.A., and F.B. designed the project. H.R., A.V.B., E.P., S.J., H.Y., and J.Z. provided technical or methodological support. H.R., A.V.B., H.Y., and J.Z. performed statistical analysis. H.R. wrote the manuscript. H.R., M.A., L.H.L., T.K., J.W., M.K.A., and F.B. revised the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Taylor, B.S., Barretina, J., Maki, R.G., Antonescu, C.R., Singer, S., and Ladanyi, M. (2011). Advances in sarcoma genomics and new therapeutic targets. *Nat. Rev. Cancer* *11*, 541–557.
- Gage, M.M., Nagarajan, N., Ruck, J.M., Canner, J.K., Khan, S., Giuliano, K., Gani, F., Wolfgang, C., Johnston, F.M., and Ahuja, N. (2019). Sarcomas in the United States: recent trends and a call for improved staging. *Oncotarget* *10*, 2462–2474.
- Stiller, C.A., Trama, A., Serraino, D., Rossi, S., Navarro, C., Chirlaque, M.D., Casali, P.G., and Group, R.W. (2013). Descriptive epidemiology of sarcomas in Europe: report from the RARECARE project. *Eur. J. Cancer* *49*, 684–695.
- Burningham, Z., Hashibe, M., Spector, L., and Schiffman, J.D. (2012). The epidemiology of sarcoma. *Clin. Sarcoma Res.* *2*, 14.
- Albertsmeier, M., Rauch, A., Roeder, F., Hasenhuttl, S., Pratschke, S., Kirschnick, M., Gronchi, A., Jebsen, N.L., Cassier, P.A., Sargos, P., et al. (2018). External beam radiation therapy for resectable soft tissue sarcoma: a systematic review and meta-analysis. *Ann. Surg. Oncol.* *25*, 754–767.
- Tseng, W.W., Barretta, F., Conti, L., Grignani, G., Tolomeo, F., Albertsmeier, M., Angele, M.K., Rutkowski, P., Skoczylas, J., De Paoli, A., et al. (2021). Defining the role of neoadjuvant systemic therapy in high-risk retroperitoneal sarcoma: a multi-institutional study from the Transatlantic Australasian Retroperitoneal Sarcoma Working Group. *Cancer* *127*, 729–738.
- Bonvalot, S., Rivoire, M., Castaing, M., Stoeckle, E., Le Cesne, A., Blay, J.Y., and Laplanche, A. (2009). Primary retroperitoneal sarcomas: a multivariate analysis of surgical factors associated with local control. *J. Clin. Oncol.* *27*, 31–37.
- Singer, S., Antonescu, C.R., Riedel, E., and Brennan, M.F. (2003). Histologic subtype and margin of resection predict pattern of recurrence and survival for retroperitoneal liposarcoma. *Ann. Surg.* *238*, 358–370, discussion 370–351.
- Tan, M.C., Brennan, M.F., Kuk, D., Agaram, N.P., Antonescu, C.R., Qin, L.X., Moraco, N., Crago, A.M., and Singer, S. (2016). Histology-based classification predicts pattern of recurrence and improves risk stratification in primary retroperitoneal sarcoma. *Ann. Surg.* *263*, 593–600.
- Albertsmeier, M., Altendorf-Hofmann, A., Lindner, L.H., Issels, R.D., Kampmann, E., Durr, H.R., Schubert-Fritschle, G., Angele, M.K., Kirchner, T., Jungbluth, A.A., et al. (2020). Cancer testis antigens and immunotherapy: expression of PRAME is associated with prognosis in soft tissue sarcoma. *Cancers* *12*, 3612.
- Pandya, P.H., Murray, M.E., Pollok, K.E., and Renbarger, J.L. (2016). The immune system in cancer pathogenesis: potential therapeutic approaches. *J. Immunol. Res.* *2016*, 4273943.
- Italiano, A., Bessedè, A., Bompas, E., Piperno-Neumann, S., Chevreau, C., Penel, N., Bertucci, F., Toulmonde, M., Bellera, C.A., Guegan, J.-P., et al. (2021). PD1 inhibition in soft-tissue sarcomas with tertiary lymphoid structures: a multicenter phase II trial. *J. Clin. Oncol.* *39*, 11507.
- Petitprez, F., de Reynies, A., Keung, E.Z., Chen, T.W., Sun, C.M., Calderaro, J., Jeng, Y.M., Hsiao, L.P., Lacroix, L., Bougouin, A., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature* *577*, 556–560.
- Kochenderfer, J.N., Dudley, M.E., Kassim, S.H., Somerville, R.P., Carpenter, R.O., Stetler-Stevenson, M., Yang, J.C., Phan, G.Q., Hughes, M.S., Sherry, R.M., et al. (2015). Chemotherapy-refractory diffuse large B-cell lymphoma and indolent B-cell malignancies can be effectively treated with autologous T cells expressing an anti-CD19 chimeric antigen receptor. *J. Clin. Oncol.* *33*, 540–549.

15. Kamphorst, A.O., Pillai, R.N., Yang, S., Nasti, T.H., Akondy, R.S., Wieland, A., Sica, G.L., Yu, K., Koenig, L., Patel, N.T., et al. (2017). Proliferation of PD-1+ CD8 T cells in peripheral blood after PD-1-targeted therapy in lung cancer patients. *Proc. Natl. Acad. Sci. U S A* *114*, 4993–4998.
16. Lu, H., Dai, W., Guo, J., Wang, D., Wen, S., Yang, L., Lin, D., Xie, W., Wen, L., Fang, J., et al. (2020). High abundance of intratumoral gammadelta T cells favors a better prognosis in head and neck squamous cell carcinoma: a bioinformatic analysis. *Front. Immunol.* *11*, 573920.
17. Miao, Y.R., Zhang, Q., Lei, Q., Luo, M., Xie, G.Y., Wang, H., and Guo, A.Y. (2020). ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci.* *7*, 1902880.
18. Blay, J.Y., and Ray-Coquard, I. (2017). Sarcoma in 2016: evolving biological understanding and treatment of sarcomas. *Nat. Rev. Clin. Oncol.* *14*, 78–80.
19. Jin, K., Qiu, S., Jin, D., Zhou, X., Zheng, X., Li, J., Liao, X., Yang, L., and Wei, Q. (2021). Development of prognostic signature based on immune-related genes in muscle-invasive bladder cancer: bioinformatics analysis of TCGA database. *Aging* *13*, 1859–1871.
20. Xiao, B., Liu, L., Li, A., Xiang, C., Wang, P., Li, H., and Xiao, T. (2020). Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma. *Front. Oncol.* *10*, 607622.
21. Dai, S., Xu, S., Ye, Y., and Ding, K. (2020). Identification of an immune-related gene signature to improve prognosis prediction in colorectal cancer patients. *Front. Genet.* *11*, 607009.
22. Yan, S., Fang, J., Chen, Y., Xie, Y., Zhang, S., Zhu, X., and Fang, F. (2020). Comprehensive analysis of prognostic gene signatures based on immune infiltration of ovarian cancer. *BMC Cancer* *20*, 1205.
23. Zhou, W., Zhang, Y., Zhong, C., Hu, J., Hu, H., Zhou, D., and Cao, M. (2018). Decreased expression of TRIM21 indicates unfavorable outcome and promotes cell growth in breast cancer. *Cancer Manag. Res.* *10*, 3687–3696.
24. Jin, Y., Zhang, Y., Li, B., Zhang, J., Dong, Z., Hu, X., and Wan, Y. (2019). TRIM21 mediates ubiquitination of Snail and modulates epithelial to mesenchymal transition in breast cancer cells. *Int. J. Biol. Macromol.* *124*, 846–853.
25. Zeng, Q.Z., Liu, W.T., Lu, J.L., Liu, X.H., Zhang, Y.F., Liu, L.X., and Gao, X.J. (2018). YWHAZ binds to TRIM21 but is not involved in TRIM21-stimulated osteosarcoma cell proliferation. *Biomed. Environ. Sci.* *31*, 186–196.
26. Zhu, Z., Hou, Q., and Guo, H. (2020). NT5DC2 knockdown inhibits colorectal carcinoma progression by repressing metastasis, angiogenesis and tumor-associated macrophage recruitment: a mechanism involving VEGF signaling. *Exp. Cell Res.* *397*, 112311.
27. Li, K.S., Zhu, X.D., Liu, H.D., Zhang, S.Z., Li, X.L., Xiao, N., Liu, X.F., Xu, B., Lei, M., Zhang, Y.Y., et al. (2020). NT5DC2 promotes tumor cell proliferation by stabilizing EGFR in hepatocellular carcinoma. *Cell Death Dis.* *11*, 335.
28. Guo, S., Ran, H., Xiao, D., Huang, H., Mi, L., Wang, X., Chen, L., Li, D., Zhang, S., Han, Q., et al. (2019). NT5DC2 promotes tumorigenicity of glioma stem-like cells by upregulating fyn. *Cancer Lett.* *454*, 98–107.
29. Jin, X., Liu, X., Zhang, Z., and Xu, L. (2020). NT5DC2 suppression restrains progression towards metastasis of non-small-cell lung cancer through regulation p53 signaling. *Biochem. Biophys. Res. Commun.* *533*, 354–361.
30. Hu, B., Zhou, S., Hu, X., Zhang, H., Lan, X., Li, M., Wang, Y., and Hu, Q. (2021). NT5DC2 promotes leiomyosarcoma tumour cell growth via stabilizing unpalmitoylated TEAD4 and generating a positive feedback loop. *J. Cell Mol. Med.* *25*, 5976–5987.
31. Shen, R., Liu, B., Li, X., Yu, T., Xu, K., and Ma, J. (2021). Development and validation of an immune gene-set based prognostic signature for soft tissue sarcoma. *BMC Cancer* *21*, 144.
32. Ren, E.H., Deng, Y.J., Yuan, W.H., Wu, Z.L., Zhang, G.Z., and Xie, Q.Q. (2021). An immune-related gene signature for determining Ewing sarcoma prognosis based on machine learning. *J. Cancer Res. Clin. Oncol.* *147*, 153–165.
33. Wu, G., and Zhang, M. (2020). A novel risk score model based on eight genes and a nomogram for predicting overall survival of patients with osteosarcoma. *BMC Cancer* *20*, 456.
34. Hou, M., Guo, X., Chen, Y., Cong, L., and Pan, C. (2020). A prognostic molecular signature of N(6)-methyladenosine methylation regulators for soft-tissue sarcoma from the cancer genome Atlas database. *Med. Sci. Monit.* *26*, e928400.
35. Sarcoma Meta-analysis Collaboration (1997). Adjuvant chemotherapy for localised resectable soft-tissue sarcoma of adults: meta-analysis of individual data. *Lancet* *350*, 1647–1654.
36. Sarcoma Meta-Analysis Collaboration (SMAC) (2000). Adjuvant chemotherapy for localised resectable soft tissue sarcoma in adults. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD001419>.
37. Pervaiz, N., Colterjohn, N., Farrokhlyar, F., Tozer, R., Figueredo, A., and Ghert, M. (2008). A systematic meta-analysis of randomized controlled trials of adjuvant chemotherapy for localized resectable soft-tissue sarcoma. *Cancer* *113*, 573–581.
38. Gronchi, A., Ferrari, S., Quagliuolo, V., Broto, J.M., Pousa, A.L., Grignani, G., Basso, U., Blay, J.Y., Tendero, O., Beveridge, R.D., et al. (2017). Histotype-tailored neoadjuvant chemotherapy versus standard chemotherapy in patients with high-risk soft-tissue sarcomas (ISG-STSS 1001): an international, open-label, randomised, controlled, phase 3, multicentre trial. *Lancet Oncol.* *18*, 812–822.
39. Bonvalot, S., Gronchi, A., Le Pechoux, C., Swallow, C.J., Strauss, D., Meeus, P., van Coevorden, F., Stoldt, S., Stoeckle, E., Rutkowski, P., et al. (2020). Preoperative radiotherapy plus surgery versus surgery alone for patients with primary retroperitoneal sarcoma (EORTC-62092: STRASS): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol.* *21*, 1366–1377.
40. Issels, R.D., Lindner, L.H., Verweij, J., Wust, P., Reichardt, P., Schem, B.C., Abdel-Rahman, S., Daugaard, S., Salat, C., Wendtner, C.M., et al. (2010). Neo-adjuvant chemotherapy alone or with regional hyperthermia for localised high-risk soft-tissue sarcoma: a randomised phase 3 multicentre study. *Lancet Oncol.* *11*, 561–570.
41. Pasquali, S., Pizzamiglio, S., Touati, N., Litiere, S., Marreaud, S., Kasper, B., Gelderblom, H., Stacchiotti, S., Judson, I., Dei Tos, A.P., et al. (2019). The impact of chemotherapy on survival of patients with extremity and trunk wall soft tissue sarcoma: revisiting the results of the EORTC-STBSG 62931 randomised trial. *Eur. J. Cancer* *109*, 51–60.
42. Callegaro, D., Miceli, R., Bonvalot, S., Ferguson, P., Strauss, D.C., Levy, A., Griffin, A., Hayes, A.J., Stacchiotti, S., Pechoux, C.L., et al. (2016). 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. *Lancet Oncol.* *17*, 671–680.
43. Fan, J., Qin, X., He, R., Ma, J., and Wei, Q. (2021). Gene expression profiles for an immunoscore model in bone and soft tissue sarcoma. *Aging* *13*, 13708–13725.
44. Lin, Q., Huang, Q., Wang, Q., Yan, W., and Sun, Y. (2021). Novel nomograms-based prediction models for patients with primary undifferentiated pleomorphic sarcomas resections. *Cancers* *13*, 1917.
45. Zhang, L., Akiyama, T., Fukushima, T., Iwata, S., Tsuda, Y., Takeshita, K., Kawai, A., Tanaka, S., and Kobayashi, H. (2021). Prognostic factors and impact of surgery in patients with metastatic soft tissue sarcoma at diagnosis: a population-based cohort study. *Jpn. J. Clin. Oncol.* *51*, 918–926.
46. Maki, R.G., Jungbluth, A.A., Gnjatic, S., Schwartz, G.K., D'Adamo, D.R., Keohan, M.L., Wagner, M.J., Scheu, K., Chiu, R., Ritter, E., et al. (2013). A pilot study of anti-CTLA4 antibody ipilimumab in patients with synovial sarcoma. *Sarcoma* *2013*, 168145.
47. Paoluzzi, L., Cavavio, A., Ghesani, M., Karambelkar, A., Rapkiewicz, A., Weber, J., and Rosen, G. (2016). Response to anti-PD1 therapy with nivolumab in metastatic sarcomas. *Clin. Sarcoma Res.* *6*, 24.
48. D'Angelo, S.P., Mahoney, M.R., Van Tine, B.A., Atkins, J., Milhem, M.M., Jahagirdar, B.N., Antonescu, C.R., Horvath, E., Tap, W.D., Schwartz, G.K., et al. (2018). Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. *Lancet Oncol.* *19*, 416–426.
49. Tawbi, H.A., Burgess, M., Bolejack, V., Van Tine, B.A., Schuetz, S.M., Hu, J., D'Angelo, S., Attia, S., Riedel, R.F., Priebat, D.A., et al. (2017). Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol.* *18*, 1493–1501.
50. Wilky, B.A., Trucco, M.M., Subhawong, T.K., Florou, V., Park, W., Kwon, D., Wieder, E.D., Kolonias, D., Rosenberg, A.E., Kerr, D.A., et al. (2019). Axitinib plus

- pembrolizumab in patients with advanced sarcomas including alveolar soft-part sarcoma: a single-centre, single-arm, phase 2 trial. *Lancet Oncol.* 20, 837–848.
51. Orth, M.F., Buecklein, V.L., Kampmann, E., Subklewe, M., Noessner, E., Cidre-Aranaz, F., Romero-Perez, L., Wehweck, F.S., Lindner, L., Issels, R., et al. (2020). A comparative view on the expression patterns of PD-L1 and PD-1 in soft tissue sarcomas. *Cancer Immunol. Immunother.* 69, 1353–1362.
 52. Issels, R., Büclein, V., Kampmann, E., Knösel, T., Nössner, E., Subklewe, M., and Lindner, L. (2016). 1412P - dissecting the role of tumor-infiltrating lymphocytes (TIL) in patients with high-risk soft-tissue sarcoma (STS) receiving neo-adjuvant chemotherapy (NAC) with regional hyperthermia (RHT). *Ann. Oncol.* 27, vi488.
 53. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.
 54. Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Trevino, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612.
 55. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* 102, 15545–15550.
 56. Hu, F.F., Liu, C.J., Liu, L.L., Zhang, Q., and Guo, A.Y. (2020). Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. *Brief Bioinform.* 22, bbaa176.
 57. Zhang, Z., and Kattan, M.W. (2017). Drawing Nomograms with R: applications to categorical outcome and survival data. *Ann. Transl. Med.* 5, 211.
 58. Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* 14, 323–348.
 59. Heagerty, P.J., Lumley, T., and Pepe, M.S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56, 337–344.

OMTO, Volume 24

Supplemental information

**A novel immune-related gene signature
predicting survival in sarcoma patients**

**Haoyu Ren, Alexandr V. Bazhin, Elise Pretzsch, Sven Jacob, Haochen Yu, Jiang
Zhu, Markus Albertsmeier, Lars H. Lindner, Thomas Knösel, Jens Werner, Martin K.
Angele, and Florian Bösch**

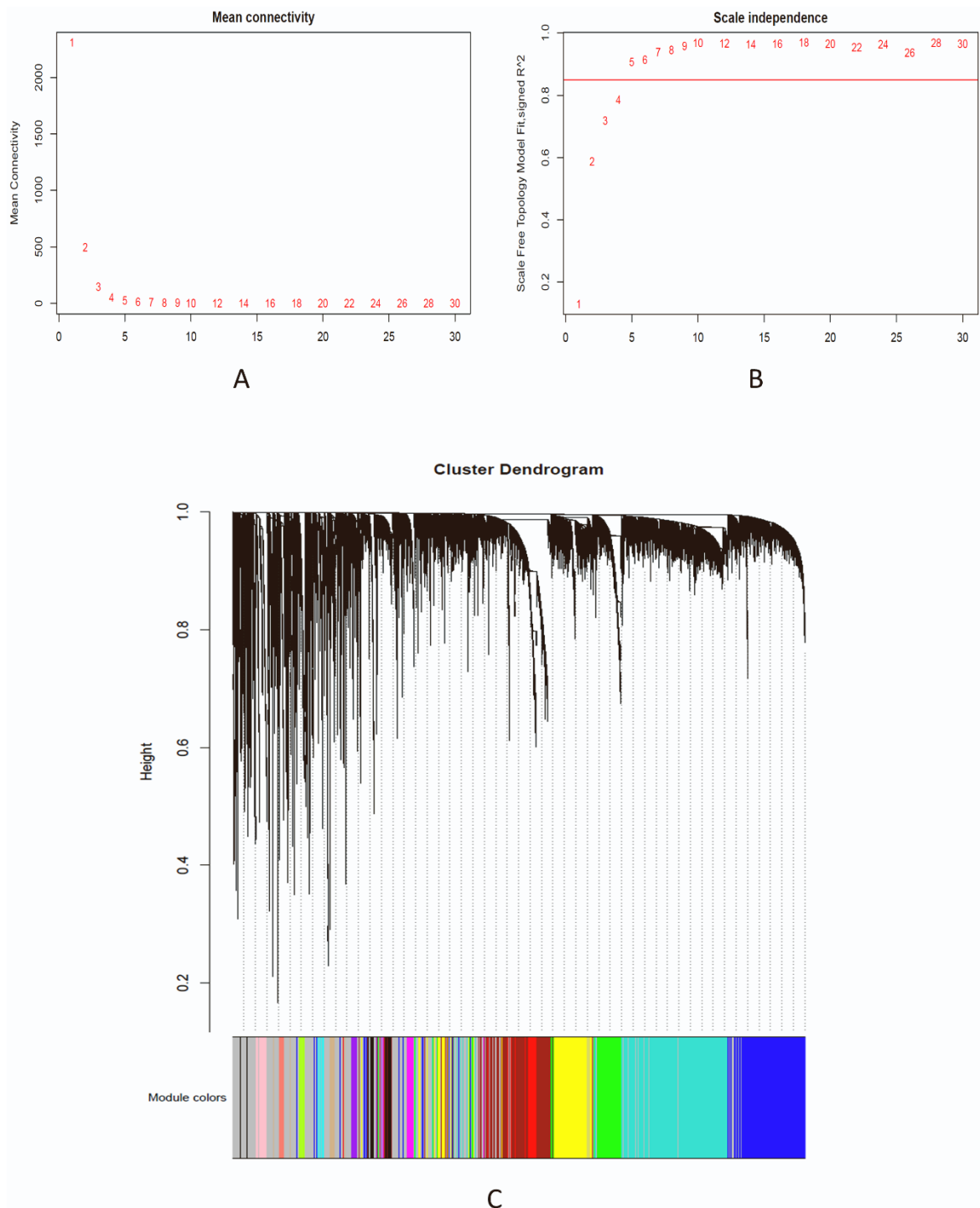


Figure S1. The establishment of a gene co-expression network. (A, B) Soft-thresholding power analysis was used to obtain the scale-free fit index of the network topology. (C) Identification of co-expression modules in TCGA-SARC.

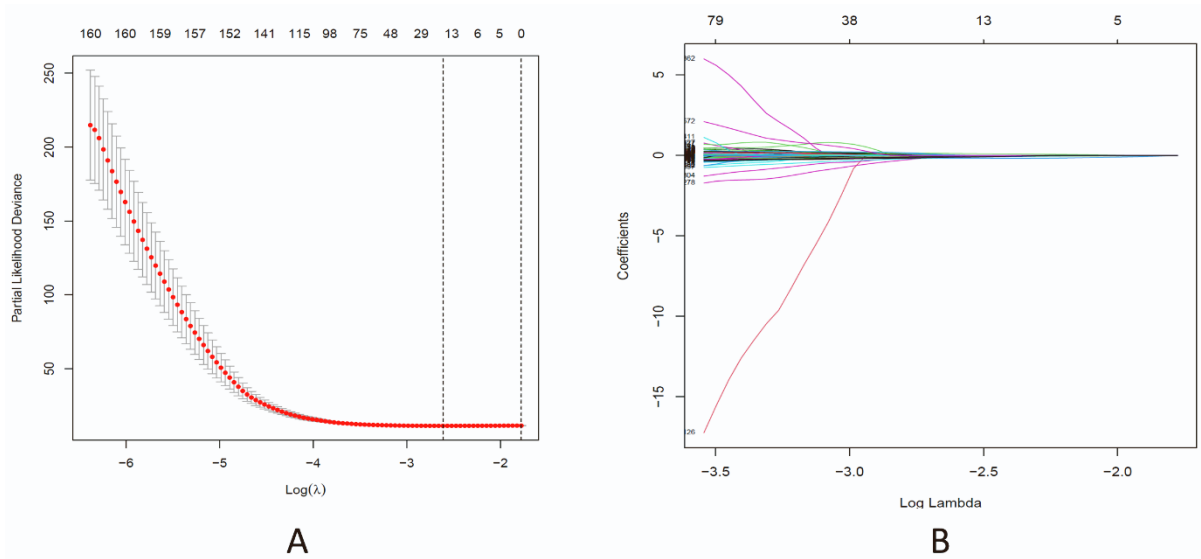


Figure S2. LASSO COX regression analysis for prognostic IRGs. (A) Lambda selection by the partial likelihood deviance. (B) The penalty coefficient of 220 candidate genes was optimized through 10-fold cross-validation in the training set.

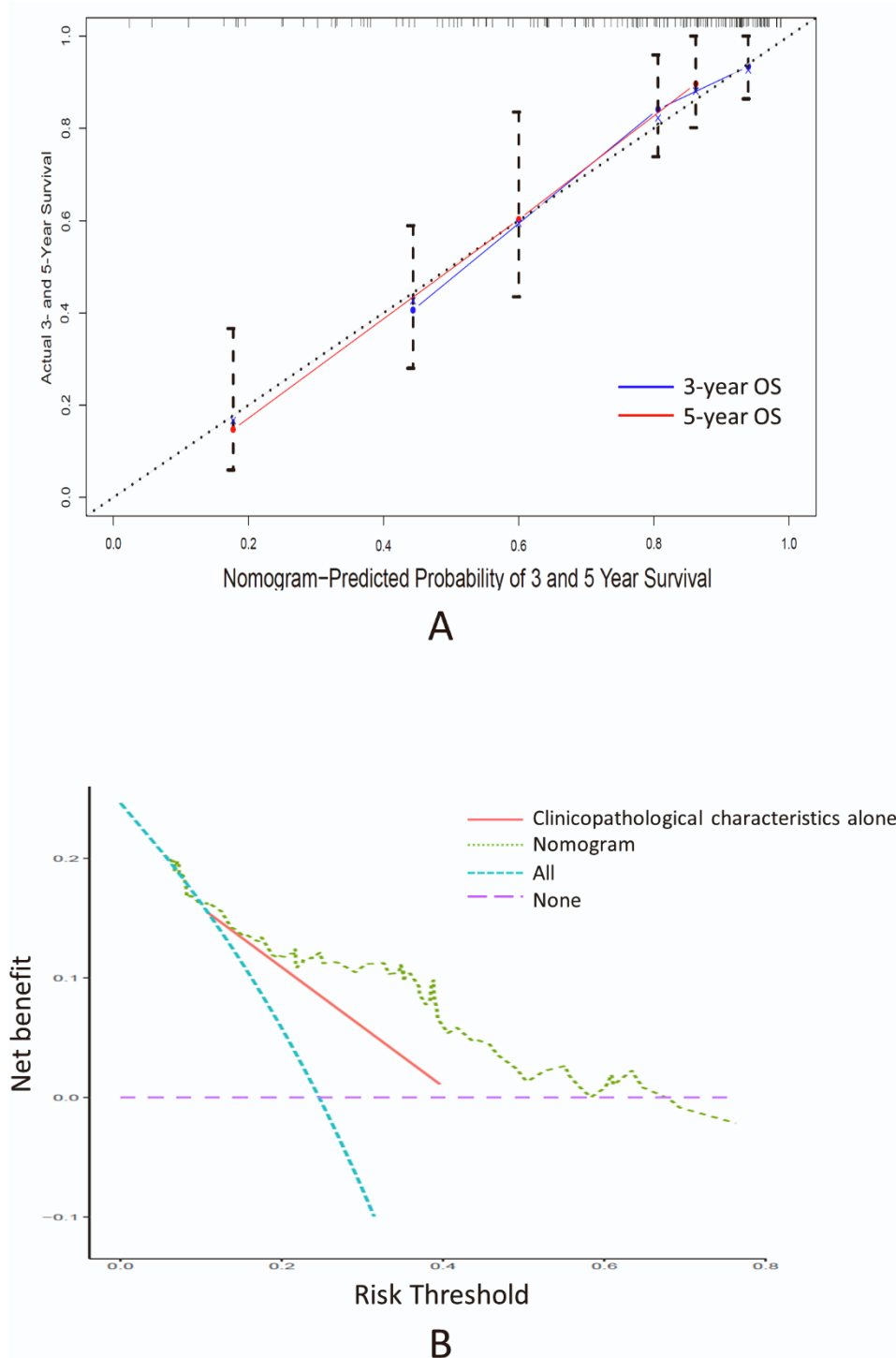
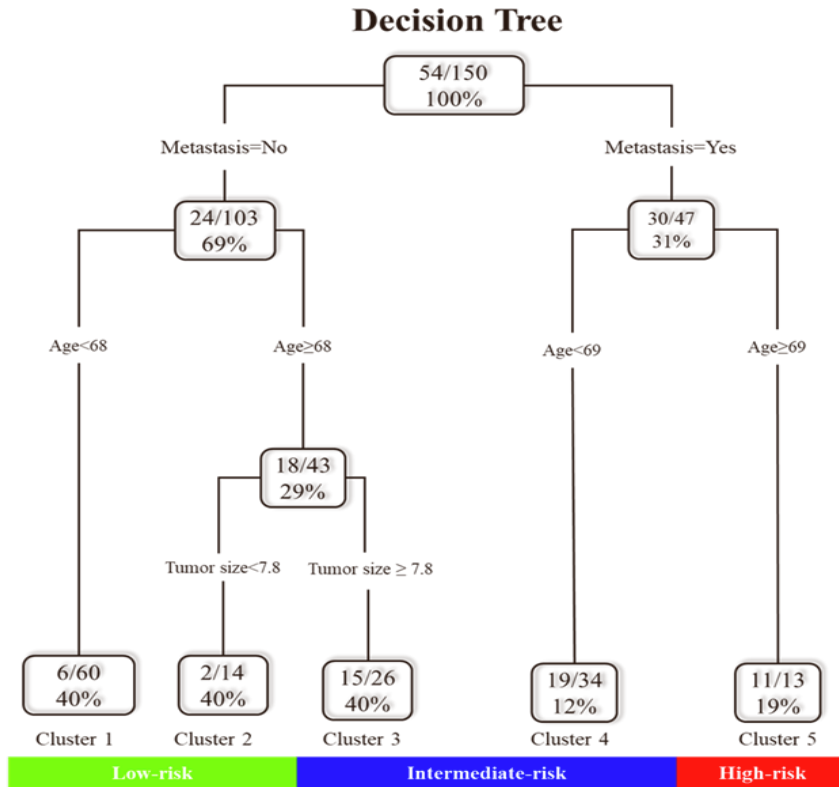


Figure S3. The evaluation of the IRS signature risk score-based nomogram. (A) Calibration analysis indicated a high accuracy of 3- and 5-year survival prediction. (B) Decision curve analysis of nomogram and clinicopathological features alone in terms of overall survival in TCGA-SARC. The x-axis represents the threshold probability, and the y-axis shows the net benefit.



A

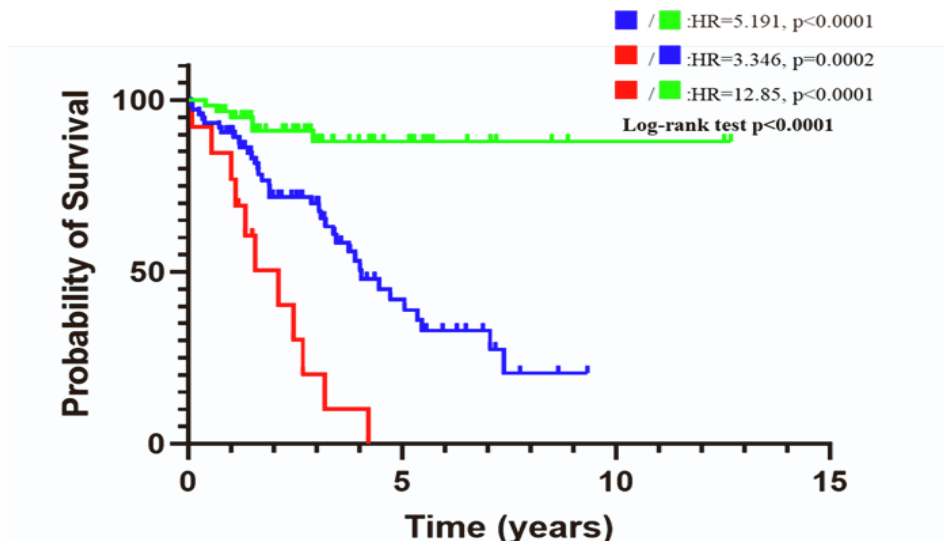


Figure S4. The risk stratification based on clinicopathological information alone. (A) A decision tree based on clinicopathological information alone was established to categorize patients into three different risk levels. (B) Kaplan-Meier analysis of overall survival of the three different subgroups.