Comprehensive analysis of immune infiltration and gene expression for predicting survival in patients with sarcomas

Hongmin Chen[1,2], Yijiang Song[1,2], Chuangzhong Deng[1,2], Yanyang Xu[1,2], Huaiyuan Xu[1,2], Xiaojun Zhu[1,2], Guohui Song[1,2], Qinglian Tang[1,2], Jinchang Lu[1,2], Jin Wang[1,2]

[1]Department of Musculoskeletal Oncology, Sun Yat-Sen University Cancer Center, Guangzhou 510060, Guangdong, P. R. China
[2]State Key Laboratory of Oncology in Southern China, Collaborative Innovation Center of Cancer Medicine, Guangzhou 510060, P. R. China

ABSTRACT

Tumor microenvironments are strongly related to tumor development, and immune-infiltrating cells and immune-related molecules are potential prognostic markers. However, the shortcomings of traditional measurement methods limit the accurate evaluation of various components in tumor microenvironments. With the rapid advancement of Next-Generation RNA Sequencing technology, dedicated and in-depth analyses of immune filtration within the tumor microenvironment has been achieved. In this study, we combined the bioinformatics analysis methods ESTIMATE, CIBERSORT, and ssGSEA to characterize the immune infiltration of sarcomas and to identify specific immunomodulators of different pathological subtypes. We further extracted a functional enrichment of significant immune-related genes related to improved prognosis, including NR1H3, VAMP5, GIMAP2, GBP2, HLA-E and CRIP1. Overall, the immune microenvironment is an important prognostic determinant of sarcomas and may be a potential resource for developing effective immunotherapy.

INTRODUCTION

Sarcomas, a broad family of mesenchymal malignancies, exhibit remarkable histologic diversity [1]. Soft tissue sarcomas are malignancies caused by extra-skeletal connective tissue (including the peripheral nervous system) and consist of more than 80 pathological types [2]. Sarcomas are rare; an estimated 13,130 sarcomas are diagnosed each year in the United States, accounting for approximately 1% of 1,806,590 new malignancies [3]. Many are highly aggressive and account for a high proportion of cancer mortality in young adults (in SEER data, seer.cancer.gov). Soft tissue sarcomas occur with great frequency in patients with specific mutations, such as the APC mutation and TP53 mutation [1, 4, 5].

The tumor microenvironment (TME) is composed of various cell types (e.g., endothelial cells, fibroblasts and immune cells) and extracellular components (e.g., extracellular matrix, cytokines and growth factors) [6]. In the TME, immune cells and stromal cells are the two main types of non-tumor components and are indicated to be of great significance in the prognosis evaluation [7, 8]. A previous study reported that immune surveillance may play an important role in the progression of sarcomas [9]. The response rates were less than 20% with either pembrolizumab monotherapy or with combination ipilimumab/nivolumab therapy [10, 11]. However, effective responses to immune checkpoint blockade therapy have been observed in specific subtypes, including high-grade undifferentiated pleomorphic sarcomas, clear cell sarcomas and dedifferentiated

liposarcomas [10–13]. Different immune characteristics may have a significant impact on efficacy. Therefore, it is urgent to explore the features of sarcoma immune microenvironment cells and immune checkpoints, which may provide potential prognostic factors and treatment targets for clinical therapy.

Immune microenvironment-related bioinformatic algorithms have been applied to cervical cancer [14], breast cancer [15], and gliomas [16], showing the effectiveness of the algorithms [17]. However, the practicality of immune scores for addressing sarcomas has not been studied. ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data), an algorithm to calculate immune and stromal scores, predicts the infiltration of non-cancer components [18]. The Cancer Genome Atlas sarcoma (TCGA SARC) database is available to understand potential correlations between gene set profiles and overall survival from malignancies [19]. To better understand the proportions of immune cells in the TME, we used CIBERSORT (Cell type Identification By Estimating Relative Subsets Of RNA Transcripts) deconvolution software and ssGSEA (single sample Gene Set Enrichment Analysis) to determine the relative proportions of several distinct leukocyte cell types in sarcomas from microarray gene expression data of sarcoma patients [20, 21].

Thus, we first obtained a list of genes that predict good outcomes in sarcoma patients by making use of the ESTIMATE, CIBERSORT and ssGSEA algorithms. Finally, we validated these genes in an independent sarcoma cohort from the Gene Expression Omnibus (GEO) dataset GSE17679.

RESULTS

Immune scores and stromal scores of sarcoma subtypes

The gene expression profiles and clinical information of all 254 sarcoma patients with an initial pathologic diagnosis and their overall survival were downloaded from the database of TCGA. All sarcoma cases with complete gene expression data and clinical information in TCGA were included in our analysis. Among all cases, the pathological diagnoses included 58 (22.8%) cases of dedifferentiated liposarcoma (DDLPS), 103 (40.6%) cases of leiomyosarcoma (LMS), 25 (9.8%) cases of myxofibrosarcoma (MFS), 9 (3.5%) cases of malignant peripheral nerve sheath tumor (MPNST), 10 (3.9%) cases of synovial sarcoma (SS) and 49 (19.3%) cases of undifferentiated pleomorphic sarcoma (UPS). According to the ESTIMATE algorithm, the average immune scores of the UPS cases were ranked highest

among all six pathological subtypes, followed by those of MFS, DDLPS, MPNST and LMS. The SS cases had the lowest immune scores (Figure 1A, $P<0.001$, one-way ANOVA). Meanwhile, the stromal scores of DDLPS, MFS and UPS were ranked the highest among all subtypes, followed by MPNST and LMS. The stromal scores of SS cases were also the lowest (Figure 1B, $P<0.001$, one-way ANOVA). These results showed the patterns of microenvironmental variance across these pathological subtypes.

Based on the TCGA somatic mutation database, TP53, TTN, ATRX, MUC16, RB1 and MUC4 make up the highest frequency of somatic mutations in sarcomas (Figure 1C). We explored the relationship between the mutations of these genes and the immune scores and found that the mutation status was not correlated with immune scores (Figure 1D–1I, Student's t test), indicating that the main type of mutation in sarcoma might not cause specific changes in immune infiltration.

To examine the correlations between overall survival and immune scores or stromal scores, we divided the cases into two halves (high scores and low scores). A Kaplan-Meier survival curve showed that the survival time of the high-immune-score group was longer than that of the low-score group (Figure 1J, $P=0.001$). Consistently, patients in the low-stromal-score group were significantly worse off in their overall survival times compared to the high-score group (Figure 1K, $P=0.035$).

Gene expression profiles of immune/stromal scores in sarcomas

To reveal the differential gene expressions in the immune score/stromal scores, we compared the RNA-seq data of 254 sarcoma patients obtained from the database of TCGA. The differentially expressed genes (DEGs) in the immune/stromal groups with high or low scores have different characteristics. In terms of the immune-score groups, 1,396 genes were upregulated and 949 genes were downregulated in the high-score group compared with the low-score group (fold change>2, $P<0.05$). Similarly, for the stromal-score groups, the high-score group had 1,533 upregulated genes and 969 downregulated genes (fold change>2, $P<0.05$) compared with the low-score group. Figure 2A, 2B plots the heatmap of the top 100 DEGs with high or low scores. In addition, the CIBERSORT deconvolution algorithm was used to calculate the proportions of 22 types of immune cells in sarcoma samples (Figure 2C and Supplementary Figure 1A). The type 2 macrophages in sarcomas made up the largest composition of all immune cells in the TME of sarcomas (except synovial sarcoma). The results indicate a suppressive TME in sarcomas.

To further reveal the biological processes, cellular components and molecular functions of the DEGs, we performed a functional enrichment analysis of the genes that were upregulated in the high-immune-score group that fold change>3. The top gene ontology (GO) terms identified included immune (including innate and adaptive) response, plasma membrane and chemokine activities. (Figure 2D–2F).

Genes for survival prediction in sarcomas

To explore the potential roles of individual immune-related DEGs on the overall survival of sarcoma patients, we used univariate Cox regression analyses and Kaplan-Meier survival curves. Among the DEGs upregulated in the high-immune-score group, a total of 528 DEGs were significantly related to good overall
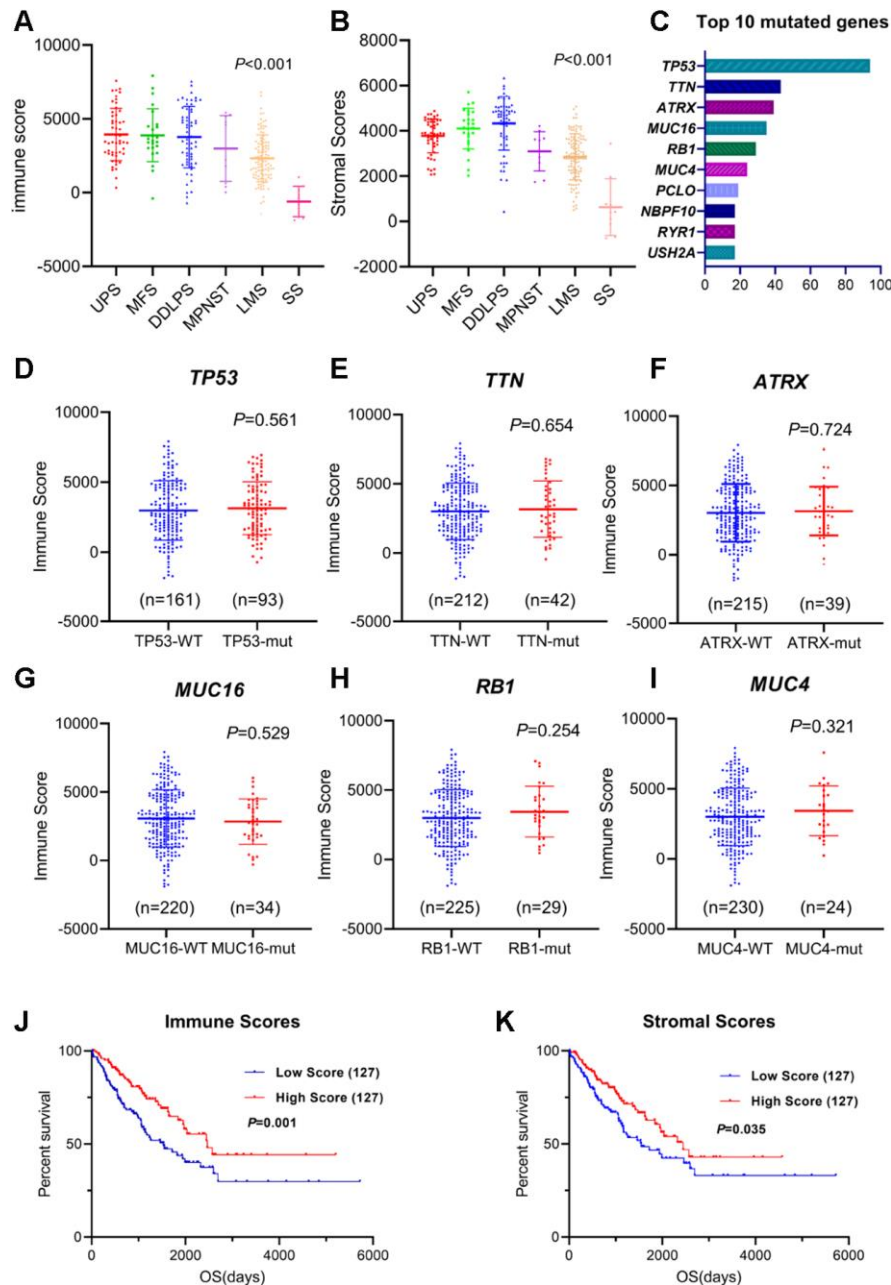
Figure 1. Immune/stromal scores of different sarcoma subtypes and survival. (**A**, **B**) Immune scores and stromal scores for different sarcoma subtypes. One-way ANOVA was applied. (**C**) Top 10 somatic mutation of TCGA sarcoma cohort. (**D**–**I**) Distribution of immune scores for *TP53*, *TTN*, *ATRX*, *MUC16*, *RB1* and *MUC4* mutant/wildtype sarcoma cases. Student's t test was applied. (**J**, **K**) The Kaplan-Meier survival curves of immune scores and stromal scores. Sarcoma cases were divided into high- and low-score groups based on the median. The sample number for each group was listed in brackets. Statistical significance was determined using the log-rank test.

survival, as determined by univariate Cox regression analysis. Functional enrichment analysis showed a strong association between these genes and the immune response as well. The top GO terms mainly included immune responses, plasma membrane, T cell receptor complex, chemokine activity and receptor binding (Figure 3A–3C, Supplementary Table 1).

After performing log-rank tests in the TCGA SARC cohort and validating these good-prognosis-related genes with the GEO dataset GSE17679, we extracted

information indicating that NR1H3, VAMP5, GIMAP2, GBP2, HLA-E and CRIP1 were highly expressed in the immune microenvironment and were most significantly associated with predicting good outcomes in sarcoma patients (Figure 3D–3I, Supplementary Figure 2).

Protein-protein interactions among prognosis-related genes

To better understand the interplay among the identified DEGs, we obtained protein-protein interaction (PPI)
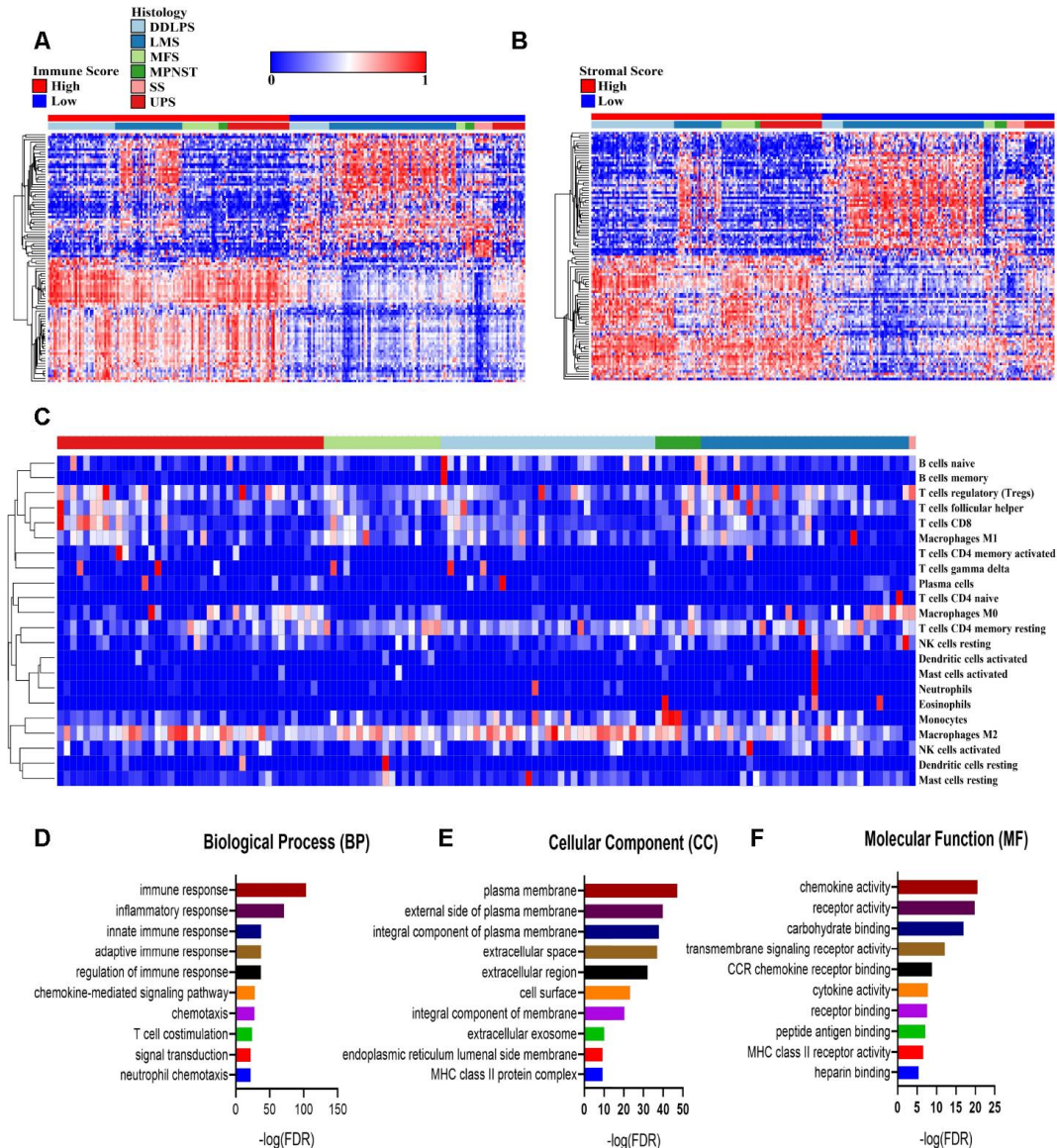
Figure 2. Comparisons of differentially expressed genes (DEGs) with immune scores and stromal scores. (**A**, **B**) The 100 top differentially expressed genes of high- and low- immune scores and stromal scores. The cohort was divided into high- and low-score groups based on the median. (**C**) The proportion of immune cells estimated by the CIBERSORT algorithm with outputs in which *P*<0.05 in 119 sarcoma samples. (**D**–**F**) Enrichment analysis of differentially expressed genes between groups with high and low immune scores. DDLPS: dedifferentiated liposarcoma, LMS: leiomyosarcoma, MFS: myxofibrosarcoma, MPNST: malignant peripheral nerve sheath tumor, SS: synovial sarcoma and UPS: undifferentiated pleomorphic sarcoma.

networks using the STRING tool. The networks were mainly made up of six modules, which included 514 nodes and 1,635 edges.

The top three significant modules were selected for further study. In the cell chemotaxis network (Figure 4A), CCL13, CXCL9, C3, CCL5 and CXCL13 were the core nodes since they had the highest expressions. For the immune response network (Figure 4B), most nodes were correlated with the response to interferon-gamma, including ICAM1, HLA-B, HLA-F, and HLA-D family. For the leukocyte mediated immunity network (Figure 4C), BTK, TYROBP, CTSS, CTSC and LYZ had higher degree values and were enriched in leukocyte degranulation and myeloid leukocyte activity.

The prognostic model of immune checkpoint modulators in sarcomas

In sarcoma patients, the efficacy of emerging immune checkpoint blockade therapies (such as PD1 inhibitors)

has been linked to the expression of immune checkpoint signaling molecules and immune cell infiltration fractions [22]. We performed a univariate Cox regression analysis of 20 immunomodulators on six pathological subtypes of sarcoma patients. B2M and CD40 were associated with the good prognostic value among DDLPS patients (Figure 5A). ICOS, TIGIT, CD274, CD276, CD47, IDO1, CD27 and PDCD1LG2 predicted good outcomes for LMS (Figure 5B). No immunomodulators had prognostic value for MFS, MPNST and SS (Figure 5C–5E). Moreover, LAG3, IDO1, TNFRSF14, PDCD1LG2, CD86, B2M, CD40 and HAVCR2 were associated with favorable outcomes in UPS patients (Figure 5F).

The prognostic factors identified by univariate Cox analysis were further analyzed in a multivariate Cox model for LMS and UPS. The results indicated that CD276 was an independent risk factor (Figure 5G). According to the immune risk score, the immuno-modulator model for LMS was established based on the combination of TIGIT, CD276, CD47 and IDO1 after
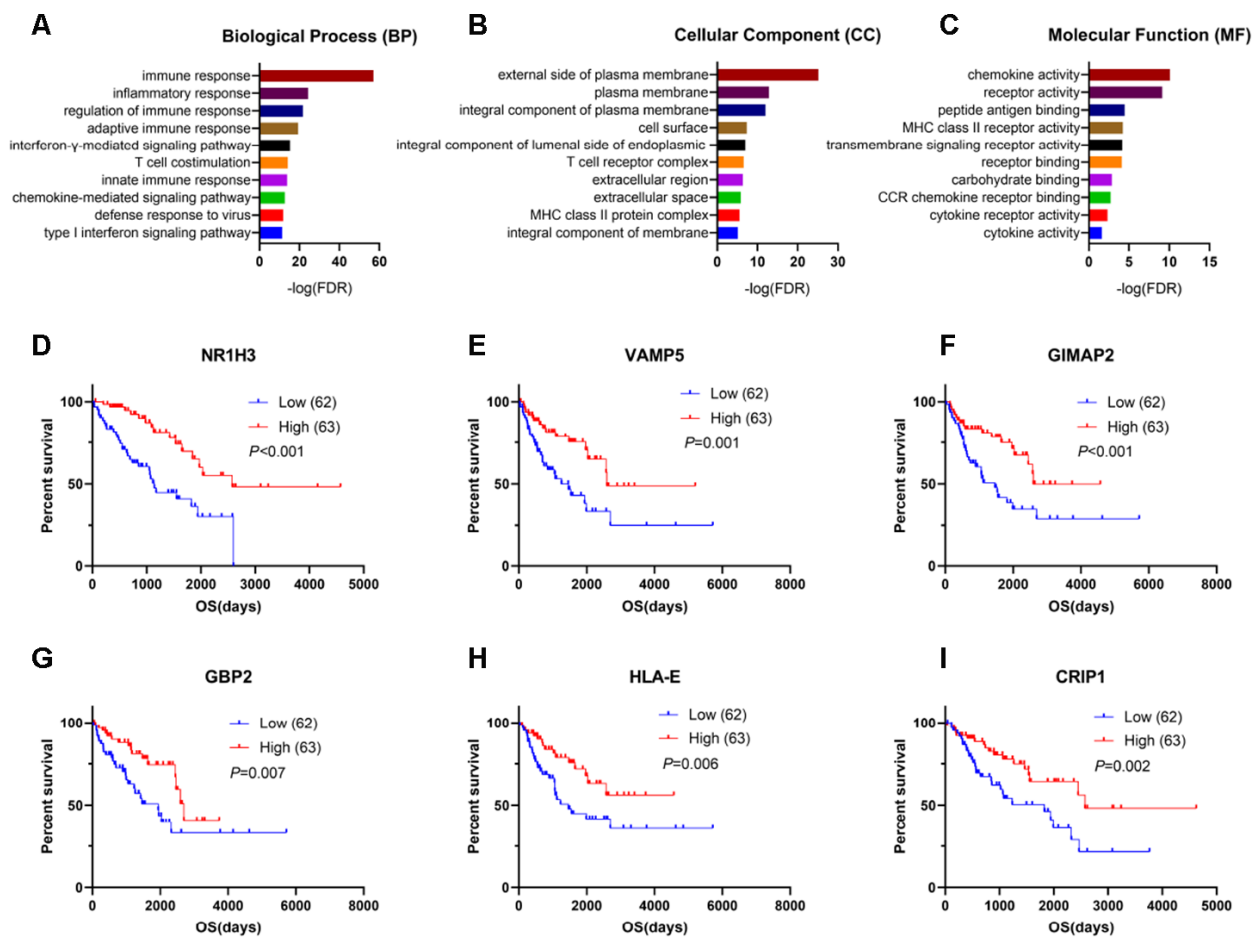
Figure 3. Genes for survival prediction among sarcoma patients. (A–C) Enrichment analysis of good survival-related genes among sarcoma patients. (D–I) The Kaplan-Meier survival curves for sarcoma patients further separated into the high and low expression groups based on the quartiles of the NR1H3, VAMP5, GIMAP2, GBP2, HLA-E and CRIP1 mRNA levels, separately.

the stepwise selection, and the prognosis index was calculated as follows:

Figure 4. Network analysis results of the top 3 protein-protein interaction networks for DEGs significantly associated with overall survival. (**A**–**C**) The cell chemotaxis network, immune response network and leukocyte mediated immunity network. The gradient color scale indicates the log value of expression fold change (log2 (FC)), and node sizes reflect the number of interactions identified for each protein. FC: fold change.

Risk score = -0.3529 × $\exp_{TIGIT}$ + 4.4598 × $\exp_{CD276}$ − 3.0041 × $\exp_{CD47}$ − 0.6090 × $\exp_{IDO1}$

where exp indicates the log2(mRNA expression).

Similarly, a predictive model for UPS was also established according to the regression coefficients. The results indicated that IDO1 and CD40 could be independent risk factors (Figure 5J). The model had the power to calculate the prognostic risk score by the following formula:

Risk score = -1.3995 × $\exp_{IDO1}$ − 8.8754 × $\exp_{CD40}$
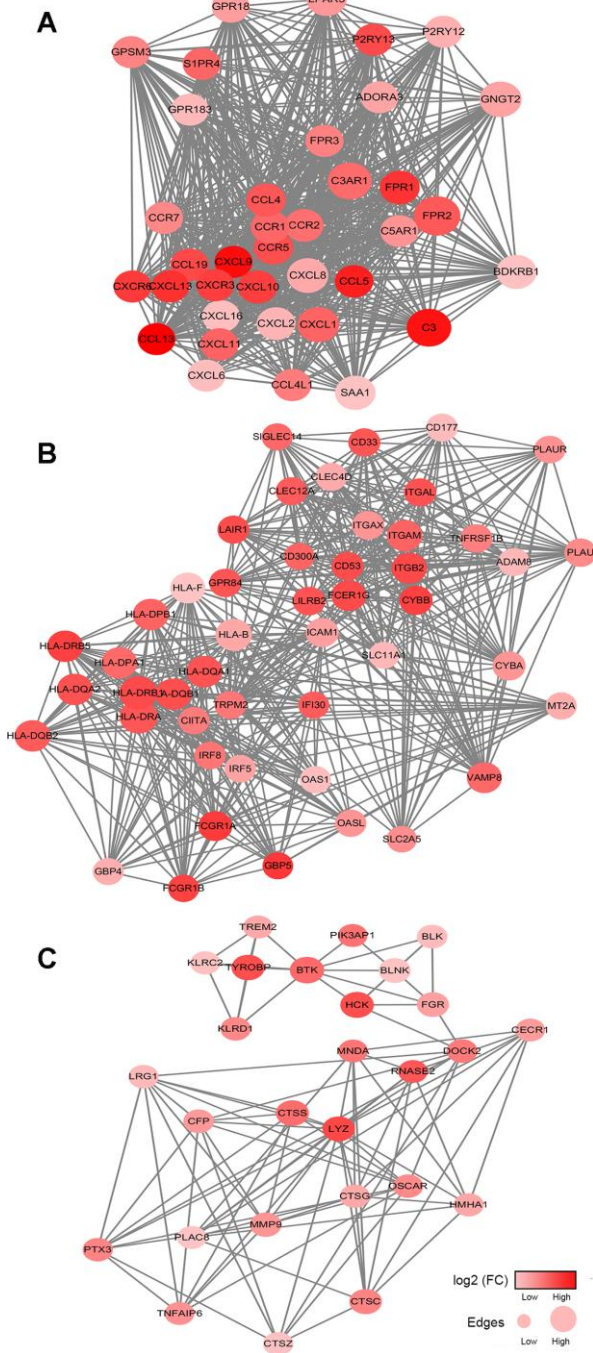
where exp indicates the log2(mRNA expression).

Patients with a high-risk score in the immunomodulator model tended to have unfavorable outcomes. The Kaplan-Meier curves showed that a high-risk score was significantly associated with poor prognoses in LMS and UPS (Figure 5H, 5K). The area under the curve (AUC) values of the receiver operating characteristic (ROC) were 0.735 and 0.839, respectively, which indicated that these models had great value in estimating patient survival (Figure 5I, 5L).

Prognostic value of infiltration immune cells in sarcoma

To assess the predictive value of immune cells in the TME, we examined the relationship between overall survival and different immune cell distribution patterns. For each patient, the relative infiltration scores of the 28 immune cell subpopulations were calculated using ssGSEA (Supplementary Figure 3). As shown in Figure 6, Kaplan-Meier survival analyses showed that activated B cells, effector memory CD4+ T cells, effector memory CD8+ T cells, immature B cells, immature dendritic cells, mast cells, monocyte, natural killer cells, plasmacytoid dendritic cells, T follicular helper cells and Th1 cells were significantly associated with improved prognoses.

DISCUSSION

There is increasing awareness that sarcomas are one of the most challenging obstacles when trying to promote immune responses to cancer. The current understanding of the sarcoma TME is limited, but previous studies have demonstrated that sarcomas might be cold tumors and that suppressor cells, including tumor-associated macrophages and myeloid-derived suppressor cells (MDSC), constitute the immune infiltration rather than exhausted CD8+ T cells [23, 24]. Moreover, PD-L1 expression is relatively low in sarcomas, and the mutation burden is also relatively low [19, 25]. To improve the therapeutic treatments for sarcomas, research

on immunomodulators and the infiltration of immune cells would be the appropriate future efforts.

First, by comparing the DEGs in the two groups of samples with high and low immune scores, we found that the high-score group, with better prognoses, was enriched in immune-related pathways such as T cell costimulation, immune receptor binding, and antigen processing via MHC-I and cytokine activity. After performing validation in both TCGA dataset and GEO
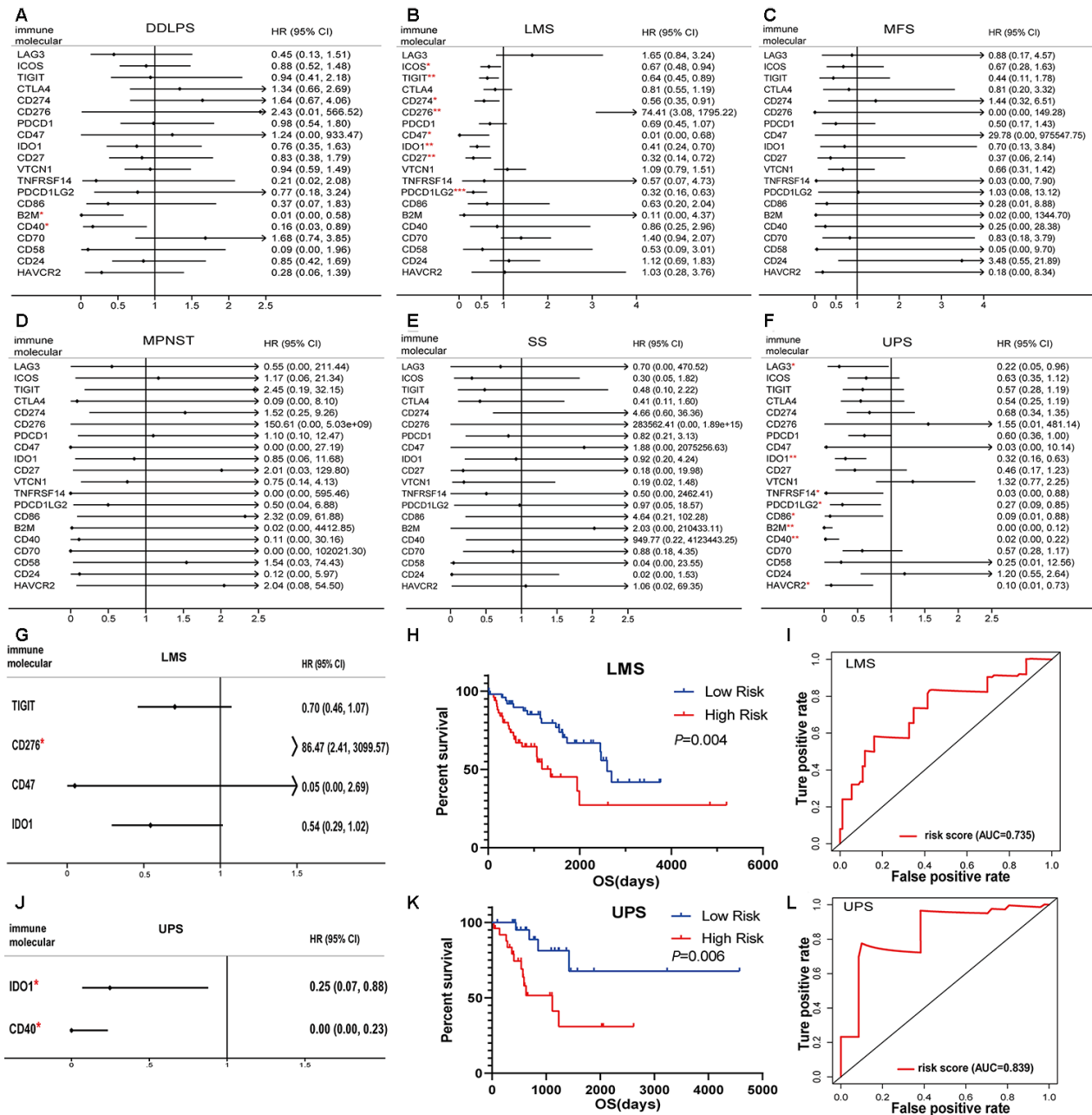

Figure 5. Immunomodulators significantly associated with prognosis in sarcoma. Forest plots of univariate Cox-regression analysis for 20 immunomodulators in the (**A**) DDLPS, (**B**) LMS, (**C**) MFS, (**D**) MPNST, (**E**) SS and (**F**) UPS. Forest plot of multivariate Cox analysis for the immunomodulators model in the (**G**) LMS and (**J**) UPS. The hazard ratio with 95% CI and *P*-values were illustrated in the figure (* p < 0.05, ** p < 0.01, *** p < 0.001). The immunomodulator model significantly predicts OS in (**H**) LMS and (**K**) UPS, and the high- and low- risk groups were divided based on the median. Receiver-operating characteristic (ROC) analysis of one-year survival prediction by the immunomodulator model in (**I**) LMS and (**L**) UPS. AUC: area under the curve, DDLPS: dedifferentiated liposarcoma, LMS: leiomyosarcoma, MFS: myxofibrosarcoma, MPNST: malignant peripheral nerve sheath tumor, SS: synovial sarcoma, UPS: undifferentiated pleomorphic sarcoma and HR: hazard ratio.

dataset GSE17679, we extracted a group of immune-related genes, including NR1H3, VAMP5, GIMAP2, GBP2, HLA-E and CRIP1 that were most significant to predict good outcomes in sarcoma patients. These immune-related genes were rarely studied in tumors but has certain therapeutic value. NR1H3 promotes the monocyte to macrophage transition and prolongs the length of survival in nasopharyngeal carcinoma [26]. GTPases of Immunity-Associated Proteins (GIMAPs) are related to the regulation of apoptosis in lymphocytes [27]. Guanylate-binding proteins (GBPs) mediate a broad scope of innate immune functions in response to several pro-inflammatory cytokines [28]. Similarly, VAMP5, HLA-E and CRIP1 have been confirmed to be involved in immune regulation [29, 30]. Extracting the potential of these molecules in anti-tumor response seems to be necessary for future work.

Besides, we extracted six protein-protein interaction modules, all of which are related to immune responses. MHC molecules present antigens for cells in the adaptive immune system, including cytotoxic CD8+ T cells. Loss or downregulation of class I MHC was observed in a large percentage of soft tissue sarcomas

patients [31–33]. For sarcoma patients with low MHC-I expression, their overall survival and event-free survival significantly worsened compared with those in the high-expression group [34]. Meanwhile, many cytokines have been employed for several sarcoma subtypes with intriguing results. Tumor necrosis factor α and melphalan (TNF-ILP) have benefited patients with locally advanced primary and recurrent extremity sarcoma in multimodal treatments [35]. Finding antigen-presenting molecules and cytokines related to favorable prognoses will help to design new clinical trials for sarcomas.

We further assessed the prognostic value of immunomodulators in sarcomas. Immune checkpoint inhibition is an encouraging approach for immuno-therapy, namely the removal of the "brakes" of the immune system [36]. Building on the limited clinical success of immune checkpoint agents for sarcomas, immunomodulators are expected to treat cancer effectively and durably [11, 37–40]. Tawbi et al. reported that in a single-arm phase II study of pembrolizumab among 40 patients with soft-tissue sarcoma, seven patients (18%) showed an objective
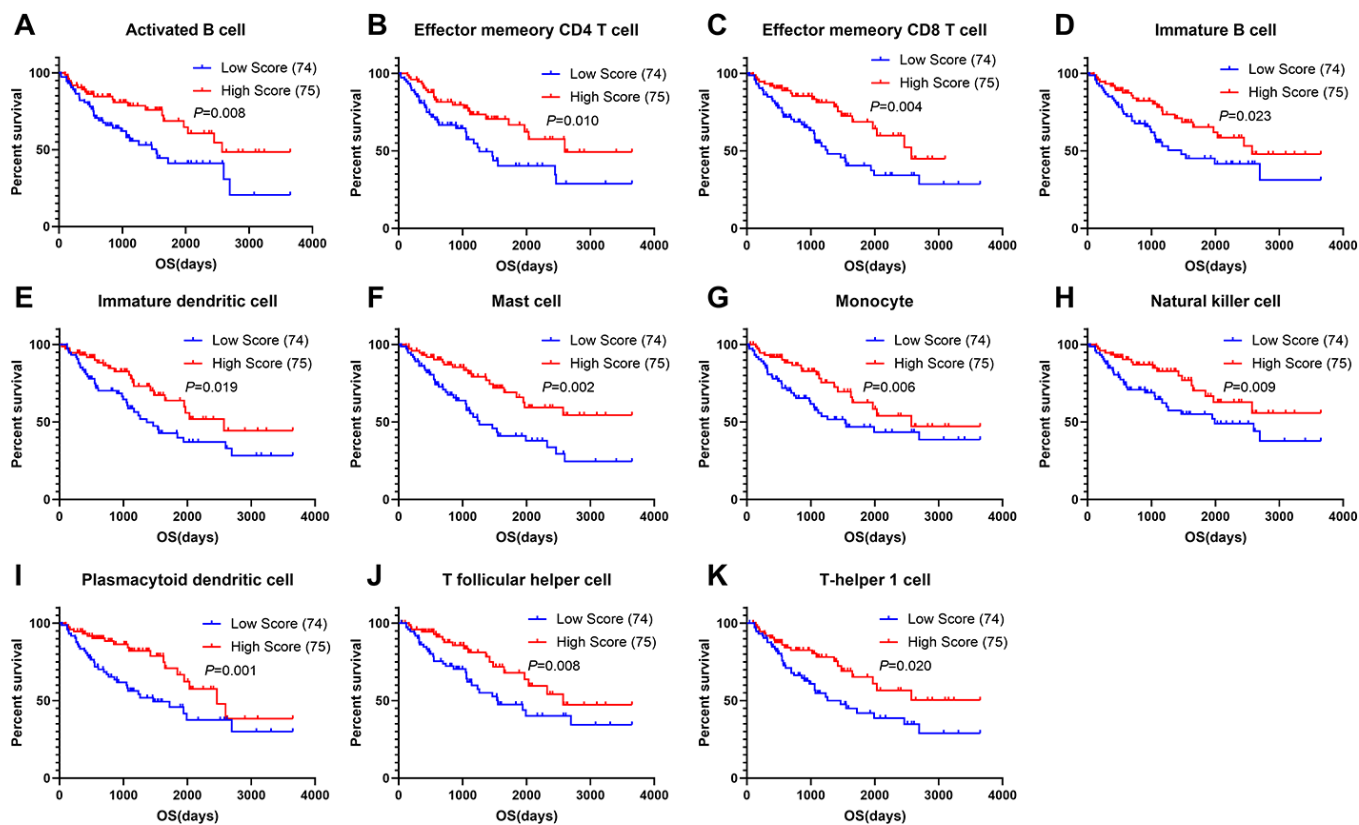
Figure 6. Correlations of immune cell scores and overall survival in sarcoma. (**A**–**K**) Infiltrating immune cells significantly associated with improved prognosis. The high- and low-score groups were divided based on the top 30% and the bottom 30% infiltrating scores calculated by ssGSEA algorithm, respectively)

response, including four of 10 patients (40%) with UPS, two of 10 patients (20%) with liposarcoma, and one of 10 patients (10%) with SS [11]. The initial results of the CTLA-4 blockade ipilimumab and dasatinib study showed no response per the Response Evaluation Criteria in Solid Tumors (RECIST) or Immune-Related Response Criteria [39]. Thus, extracting new specific checkpoints has become necessary. The checkpoint molecules analyzed in this study represented immune response signals, including costimulatory, coinhibitory and situation-dependent signals. Two immune checkpoint modulators (e.g., B2M and CD40) were associated with good prognostic value for DDLPS, seven modulators (e.g., ICOS, TIGIT, CD274, CD276, CD47, IDO1, CD27 and PDCD1LG2) were identified as prognostic molecules for LMS, and eight modulators were significantly correlated with UPS outcomes in the univariate Cox regression model. In the immuno-modulator risk score model, the combination of TIGIT, CD276, CD47 and IDO1 had a certain value in estimating survival of LMS patients, while IDO1 and CD40 had a higher predictive value for UPS, as indicated by the AUC. These immune response signals may be future targets in sarcoma treatment to further rev up or release the brakes of the immune system [41].

Finally, we evaluated the prognostic value of infiltrating immune cells in sarcoma. The results showed that 11 kinds of highly infiltrating immune cells were predictors of longer overall survival, most of which have been confirmed in previous studies. After B cell activation by T helper cells initiates the humoral immune response to most protein antigens, T follicular helper (Tfh) cells provide signals to B cells, including cytokines and cell surface ligands, to direct isotype switching and activate germinal center formation, somatic hypermutation, and affinity maturation [42–44]. Tfh cells regulate humoral immunity and cell-mediated anti-tumor responses and improve overall survival [45, 46]. Evaluating the functions of Tfh cells and exploring their interactions with B cells and T helper cells in sarcoma therefore represents a potential therapeutic strategy. Our results were in accordance with previous research, showing that Tfh cells, T helper cells and B cells infiltration predicts good prognoses in several cancers.

In conclusion, this study revealed distinct immune infiltration patterns and indicated that immunomodulators are essential determinants of sarcoma prognoses. Moreover, we extracted a list of tumor microenvironment related genes that could be useful for outlining the prognoses of sarcoma patients. The number of cohorts included in the study was small. Larger scale sequencing data will be tested for future work. The evolution of this comprehensive investigation will lead to the elucidation of the immunological mechanisms that affect sarcoma progression and the development of immunotherapeutic sarcoma trials.

MATERIALS AND METHODS

Database and differentially expressed genes analysis

The Level 3 RNA sequencing (RNAseq) data, somatic mutation data and clinical information of 254 sarcoma patients were downloaded from The Cancer Genome Atlas (TCGA) hub by the University of California, Santa Cruz, Xena browser (https://xenabrowser.net/datapages/). TCGA RNA sequencing data show the gene-level transcription estimates, as in $\log2(x + 1)$ transformed RSEM normalized counts. 64 cases from the Gene Expression Omnibus (GEO) database GSE17679 (https://www.ncbi.nlm.nih.gov/geo) were enrolled as a validation set. RNA expressions data for GSE17679 were using Affymetrix Human Genome U133 Plus 2.0 Array. The pan-cancer analysis was performed on the Gene Expression Profiling Interactive Analysis 2 (GEPIA2) webserver (http://gepia2.cancer-pku.cn/), which integrated RNA expression and clinical data from 33 TCGA cohorts. Differentially expressed genes (DEGs) were obtained using the R package edgeR [47]. Fold change>2 and $P<0.05$ were set as the cut-offs to screen for DEGs.

Immune infiltration algorithm

To infer the fractions of stromal and immune cells in tumor samples, we applied the ESTIMATE algorithm to calculate the immune and stromal scores [18]. In addition, a deconvolution approach, the CIBERSORT algorithm [20], was introduced to estimate the fractions of 22 immune cell types. After 100 permutations of calculations, samples with outputs in which $P<0.05$ were considered to be available for further analysis. Finally, the resulting CIBERSORT values were defined as the immune cell infiltration fractions of each sample. Moreover, ssGSEA [21] was used to quantify the relative infiltration scores of 28 immune cell types in the TME. Feature gene panels for each immune cell type were collected from a recent publication [48]. The enrichment score in ssGSEA represented the relative abundance of each immune cell type.

Functional enrichment analysis and PPI network

For functional analysis, gene ontology (GO) analyses were conducted via DAVID (The Database for Annotation, Visualization and Integrated Discovery) [49]. False discovery rates (FDR)<0.05 were considered significant. The protein-protein interaction (PPI) networks were retrieved from the STRING database [50].

The DEGs list was used for protein-protein interaction (PPI) analysis in Cytoscape (v3.7.1, National Resource for Network Biology, https://cytoscape.org/). Samples with outputs with interaction scores>0.9 were considered in further study. Only individual networks with 10 or more nodes were included for Molecular COmplex DEtection (MCODE), which could explore clusters based on subnetwork module analysis to find densely connected components.

Statistical analysis

The Kaplan–Meier survival curves with the log-rank test were used to estimate the correlation between DEGs or immune cell types and overall survival. The Mantel-Cox test was performed in the pan-cancer survival analysis for patients from 33 disease cohorts. Multivariate analyses applied stepwise selection and Cox proportional hazard regression model. Statistical analyses were performed using R version 3.6.0. For all tests, $P<0.05$ was considered to be statistically significant.

Code availability

R and other custom codes for analyzing data are available upon request to the authors.

AUTHOR CONTRIBUTIONS

W.J. and L.J.C. designed the study; C.H.M. collected the RNA-seq and clinical data; C.H.M. and S.Y.J. performed the statistical analysis; C.H.M., D.C.Z. and X.Y.Y. drafted the manuscript. W.J., L.J.C., X.H.Y., Z.X.J., S.G.H. and T.Q.L. critically reviewed the manuscript and approved the final draft.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

REFERENCES

1. Schaefer IM, Cote GM, Hornick JL. Contemporary sarcoma diagnosis, genetics, and genomics. J Clin Oncol. 2018; 36:101–10.
   https://doi.org/10.1200/JCO.2017.74.9374
   PMID:29220288

2. Casali PG, Abecassis N, Aro HT, Bauer S, Biagini R, Bielack S, Bonvalot S, Boukovinas I, Bovee JVMG, Brodowicz T, Broto JM, Buonadonna A, De Álava E, et al. Soft tissue and visceral sarcomas: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018; 29:iv51–iv67.
   https://doi.org/10.1093/annonc/mdy096
   PMID:29846498

3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020; 70:7–30.
   https://doi.org/10.3322/caac.21590 PMID:31912902

4. Saito T, Oda Y, Sakamoto A, Kawaguchi K, Tanaka K, Matsuda S, Tamiya S, Iwamoto Y, Tsuneyoshi M. APC mutations in synovial sarcoma. J Pathol. 2002; 196:445–49.
   https://doi.org/10.1002/path.1066 PMID:11920741

5. Mitchell G, Ballinger ML, Wong S, Hewitt C, James P, Young MA, Cipponi A, Pang T, Goode DL, Dobrovic A, Thomas DM, and International Sarcoma Kindred Study. High frequency of germline TP53 mutations in a prospective adult-onset sarcoma cohort. PLoS One. 2013; 8:e69026.
   https://doi.org/10.1371/journal.pone.0069026
   PMID:23894400

6. Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett. 2017; 387:61–68.
   https://doi.org/10.1016/j.canlet.2016.01.043
   PMID:26845449

7. Pitt JM, Marabelle A, Eggermont A, Soria JC, Kroemer G, Zitvogel L. Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. Ann Oncol. 2016; 27:1482–92.
   https://doi.org/10.1093/annonc/mdw168
   PMID:27069014

8. Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging targeted and immune-based therapies in sarcoma. J Clin Oncol. 2018; 36:125–35.
   https://doi.org/10.1200/JCO.2017.75.1610
   PMID:29220291

9. Wilky BA. Immune checkpoint inhibitors: the linchpins of modern immunotherapy. Immunol Rev. 2019; 290:6–23.
   https://doi.org/10.1111/imr.12766 PMID:31355494

10. D'Angelo SP, Mahoney MR, Van Tine BA, Atkins J, Milhem MM, Jahagirdar BN, Antonescu CR, Horvath E, Tap WD, Schwartz GK, Streicher H. Nivolumab with or without ipilimumab treatment for metastatic sarcoma (alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. Lancet Oncol. 2018; 19:416–26.
    https://doi.org/10.1016/S1470-2045(18)30006-8
   PMID:29370992

11. Tawbi HA, Burgess M, Bolejack V, Van Tine BA, Schuetze SM, Hu J, D'Angelo S, Attia S, Riedel RF, Priebat DA, Movva S, Davis LE, Okuno SH, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. Lancet Oncol. 2017; 18:1493–501.
https://doi.org/10.1016/S1470-2045(17)30624-1 PMID:28988646

12. Broto JM, Hindi N, Grignani G, Trufero JM, Redondo A, Valverde C, Pousa AL, Stacchiotti S, Palmerini E, de Alava E. IMMUNOSARC: A collaborative Spanish (GEIS) and Italian (ISG) sarcoma groups phase I/II trial of sunitinib plus nivolumab in advanced soft tissue and bone sarcomas: Results of the phase II-soft-tissue sarcoma cohort. Annals of Oncology. 2019; 30:v684.
https://doi.org/10.1093/annonc/mdz283.002

13. Wilky BA, Kumthekar P, Wesolowski R, Hwang JJ, Park SI, Proscurshim I, Yuan G, Dupont CD, Shebanova O, Cuillerot JM, Dow E, Raizer JJ, Gentry C, et al. Phase I open-label, ascending dose trial of AGEN1884, an anti-CTLA-4 monoclonal antibody, in advanced solid malignancies: Dose selection for combination with PD-1 blockade. Journal of Clinical Oncology. 2018; 36:3075.
https://doi.org/10.1200/JCO.2018.36.15_suppl.3075

14. Wang J, Li Z, Gao A, Wen Q, Sun Y. The prognostic landscape of tumor-infiltrating immune cells in cervical cancer. Biomed Pharmacother. 2019; 120:109444.
https://doi.org/10.1016/j.biopha.2019.109444 PMID:31562978

15. Anurag M, Zhu M, Huang C, Vasaikar S, Wang J, Hoog J, Burugu S, Gao D, Suman V, Zhang XH, Zhang B, Nielsen T, Ellis MJ. Immune checkpoint profiles in luminal B breast cancer (alliance). J Natl Cancer Inst. 2020; 112:737–46.
https://doi.org/10.1093/jnci/djz213 PMID:31665365

16. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging (Albany NY). 2018; 10:592–605.
https://doi.org/10.18632/aging.101415 PMID:29676997

17. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M, West RB, Plevritis SK, Alizadeh AA. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015; 21:938–45.
https://doi.org/10.1038/nm.3909 PMID:26193342

18. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612.
https://doi.org/10.1038/ncomms3612 PMID:24113773

19. Cancer Genome Atlas Research Network. Electronic address: elizabeth.demicco@sinaihealthsystem.ca, and Cancer Genome Atlas Research Network. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. Cell. 2017; 171:950–65.e28.
https://doi.org/10.1016/j.cell.2017.10.014 PMID:29100075

20. 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; 12:453–57.
https://doi.org/10.1038/nmeth.3337 PMID:25822800

21. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009; 462:108–12.
https://doi.org/10.1038/nature08460 PMID:19847166

22. Petitprez F, de Reyniès A, Keung EZ, Chen TW, Sun CM, Calderaro J, Jeng YM, Hsiao LP, Lacroix L, Bougoüin A, Moreira M, Lacroix G, Natario I, et al. B cells are associated with survival and immunotherapy response in sarcoma. Nature. 2020; 577:556–60.
https://doi.org/10.1038/s41586-019-1906-8 PMID:31942077

23. D'Angelo SP, Shoushtari AN, Agaram NP, Kuk D, Qin LX, Carvajal RD, Dickson MA, Gounder M, Keohan ML, Schwartz GK, Tap WD. Prevalence of tumor-infiltrating lymphocytes and PD-L1 expression in the soft tissue sarcoma microenvironment. Hum Pathol. 2015; 46:357–65.
https://doi.org/10.1016/j.humpath.2014.11.001 PMID:25540867

24. Pollack SM, He Q, Yearley JH, Emerson R, Vignali M, Zhang Y, Redman MW, Baker KK, Cooper S, Donahue B, Loggers ET, Cranmer LD, Spraker MB, et al. T-cell infiltration and clonality correlate with programmed cell death protein 1 and programmed death-ligand 1 expression in patients with soft tissue sarcomas. Cancer. 2017; 123:3291–304.
https://doi.org/10.1002/cncr.30726 PMID:28463396

25. Raj S, Bui M, Gonzales R, Letson D, Antonia S. Impact of PDL1 expression on clinical outcomes in subtypes of sarcoma. Annals of Oncology. 2014; 25:iv 498.
https://doi.org/10.1093/annonc/mdu354.10

26. Chen YP, Yin JH, Li WF, Li HJ, Chen DP, Zhang CJ, Lv JW, Wang YQ, Li XM, Li JY, Zhang PP, Li YQ, He QM, et al. Single-cell transcriptomics reveals regulators

underlying immune cell diversity and immune subtypes associated with prognosis in nasopharyngeal carcinoma. Cell Res. 2020; 30:1024–42.
https://doi.org/10.1038/s41422-020-0374-x
PMID:32686767

27. Schwefel D, Daumke O. GTP-dependent scaffold formation in the GTPase of immunity associated protein family. Small GTPases. 2011; 2:27–30.
https://doi.org/10.4161/sgtp.2.1.14938
PMID:21686278

28. Praefcke GJ. Regulation of innate immune functions by guanylate-binding proteins. Int J Med Microbiol. 2018; 308:237–45.
https://doi.org/10.1016/j.ijmm.2017.10.013
PMID:29174633

29. Kong HJ, Moon JH, Han YH, Nam BH, Kim YO, Kim WJ, Kim DG, Kim HS, Kim JH, Kim BS, Lee SJ. PoCRIP1, paralichthys olivaceus cysteine-rich intestinal protein 1: molecular characterization, expression analysis upon edwardsiella tarda challenge and a possible role in the immune regulation. Fish Shellfish Immunol. 2011; 30:917–22.
https://doi.org/10.1016/j.fsi.2011.01.017
PMID:21288490

30. Xiong Z, Xiong Y, Liu H, Li C, Li X. Identification of purity and prognosis-related gene signature by network analysis and survival analysis in brain lower grade glioma. J Cell Mol Med. 2020; 24:11607–12.
https://doi.org/10.1111/jcmm.15805 PMID:32869484

31. Tsukahara T, Kawaguchi S, Torigoe T, Asanuma H, Nakazawa E, Shimozawa K, Nabeta Y, Kimura S, Kaya M, Nagoya S, Wada T, Yamashita T, Sato N. Prognostic significance of HLA class I expression in osteosarcoma defined by anti-pan HLA class I monoclonal antibody, EMR8-5. Cancer Sci. 2006; 97:1374–80.
https://doi.org/10.1111/j.1349-7006.2006.00317.x
PMID:16995877

32. Meissner M, König V, Hrgovic I, Valesky E, Kaufmann R. Human leucocyte antigen class I and class II antigen expression in Malignant fibrous histiocytoma, fibrosarcoma and dermatofibrosarcoma protuberans is significantly downregulated. J Eur Acad Dermatol Venereol. 2010; 24:1326–32.
https://doi.org/10.1111/j.1468-3083.2010.03644.x
PMID:20337820

33. Berghuis D, de Hooge AS, Santos SJ, Horst D, Wiertz EJ, van Eggermond MC, van den Elsen PJ, Taminiau AH, Ottaviano L, Schaefer KL, Dirksen U, Hooijberg E, Mulder A, et al. Reduced human leukocyte antigen expression in advanced-stage ewing sarcoma: implications for immune recognition. J Pathol. 2009; 218:222–31.
https://doi.org/10.1002/path.2537 PMID:19274709

34. Nada OH, Ahmed NS, Abou Gabal HH. Prognostic significance of HLA EMR8-5 immunohistochemically analyzed expression in osteosarcoma. Diagn Pathol. 2014; 9:72.
https://doi.org/10.1186/1746-1596-9-72
PMID:24667142

35. Jakob J, Hohenberger P. Role of isolated limb perfusion with recombinant human tumor necrosis factor α and melphalan in locally advanced extremity soft tissue sarcoma. Cancer. 2016; 122:2624–32.
https://doi.org/10.1002/cncr.29991 PMID:27197621

36. Pardoll DM. Immunology beats cancer: a blueprint for successful translation. Nat Immunol. 2012; 13:1129–32.
https://doi.org/10.1038/ni.2392 PMID:23160205

37. Merchant MS, Wright M, Baird K, Wexler LH, Rodriguez-Galindo C, Bernstein D, Delbrook C, Lodish M, Bishop R, Wolchok JD, Streicher H, Mackall CL. Phase I clinical trial of ipilimumab in pediatric patients with advanced solid tumors. Clin Cancer Res. 2016; 22:1364–70.
https://doi.org/10.1158/1078-0432.CCR-15-0491
PMID:26534966

38. Ben-Ami E, Barysauskas CM, Solomon S, Tahlil K, Malley R, Hohos M, Polson K, Loucks M, Severgnini M, Patel T, Cunningham A, Rodig SJ, Hodi FS, et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: results of a phase 2 study. Cancer. 2017; 123:3285–90.
https://doi.org/10.1002/cncr.30738
PMID:28440953

39. D'Angelo SP, Shoushtari AN, Keohan ML, Dickson MA, Gounder MM, Chi P, Loo JK, Gaffney L, Schneider L, Patel Z, Erinjeri JP, Bluth MJ, Sjoberg A, et al. Combined KIT and CTLA-4 blockade in patients with refractory GIST and other advanced sarcomas: a phase ib study of dasatinib plus ipilimumab. Clin Cancer Res. 2017; 23:2972–80.
https://doi.org/10.1158/1078-0432.CCR-16-2349
PMID:28007774

40. Toulmonde M, Penel N, Adam J, Chevreau C, Blay JY, Le Cesne A, Bompas E, Piperno-Neumann S, Cousin S, Grellety T, Ryckewaert T, Bessede A, Ghiringhelli F, et al. Use of PD-1 targeting, macrophage infiltration, and IDO pathway activation in sarcomas: a phase 2 clinical trial. JAMA Oncol. 2018; 4:93–97.
https://doi.org/10.1001/jamaoncol.2017.1617
PMID:28662235

41. Nathenson MJ, Conley AP, Sausville E. Immunotherapy: a new (and old) approach to treatment of soft tissue and bone sarcomas. Oncologist. 2018; 23:71–83.
https://doi.org/10.1634/theoncologist.2016-0025
PMID:28935774

42. Crotty S. T follicular helper cell biology: a decade of discovery and diseases. Immunity. 2019; 50:1132–48. https://doi.org/10.1016/j.immuni.2019.04.011 PMID:31117010

43. Roco JA, Mesin L, Binder SC, Nefzger C, Gonzalez-Figueroa P, Canete PF, Ellyard J, Shen Q, Robert PA, Cappello J, Vohra H, Zhang Y, Nowosad CR, et al. Class-switch recombination occurs infrequently in germinal centers. Immunity. 2019; 51:337–50.e7. https://doi.org/10.1016/j.immuni.2019.07.001 PMID:31375460

44. Kawabe T, Naka T, Yoshida K, Tanaka T, Fujiwara H, Suematsu S, Yoshida N, Kishimoto T, Kikutani H. The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation. Immunity. 1994; 1:167–78. https://doi.org/10.1016/1074-7613(94)90095-7 PMID:7534202

45. Amé-Thomas P, Le Priol J, Yssel H, Caron G, Pangault C, Jean R, Martin N, Marafioti T, Gaulard P, Lamy T, Fest T, Semana G, Tarte K. Characterization of intratumoral follicular helper T cells in follicular lymphoma: role in the survival of Malignant B cells. Leukemia. 2012; 26:1053–63. https://doi.org/10.1038/leu.2011.301 PMID:22015774

46. Gu-Trantien C, Migliori E, Buisseret L, de Wind A, Brohée S, Garaud S, Noël G, Dang Chi VL, Lodewyckx JN, Naveaux C, Duvillier H, Goriely S, Larsimont D, Willard-Gallo K. CXCL13-producing TFH cells link immune suppression and adaptive memory in human breast cancer. JCI Insight. 2017; 2:e91487. https://doi.org/10.1172/jci.insight.91487 PMID:28570278

47. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–40. https://doi.org/10.1093/bioinformatics/btp616 PMID:19910308

48. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017; 18:248–62. https://doi.org/10.1016/j.celrep.2016.12.019 PMID:28052254

49. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. https://doi.org/10.1038/nprot.2008.211 PMID:19131956

50. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43:D447–52. https://doi.org/10.1093/nar/gku1003 PMID:25352553
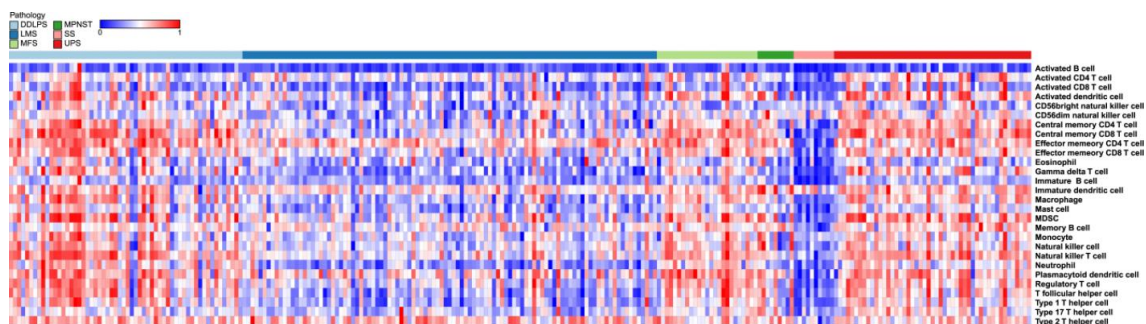
Supplementary Figures
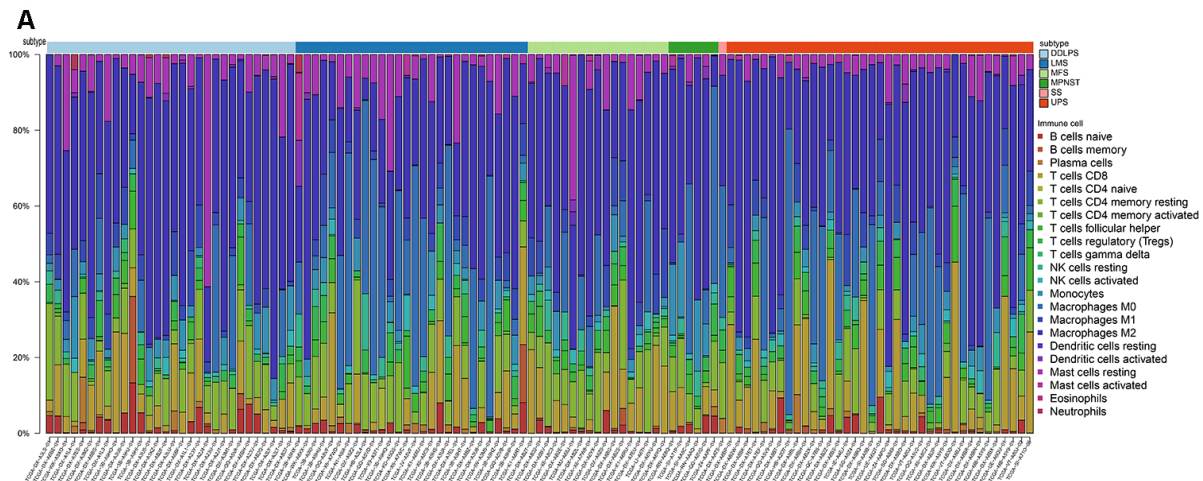
Supplementary Figure 1. Infiltration of immune cell subpopulations in sarcoma samples. (**A**) Stacked bar chart of immune cells estimated by the CIBERSORT algorithm in sarcomas samples.
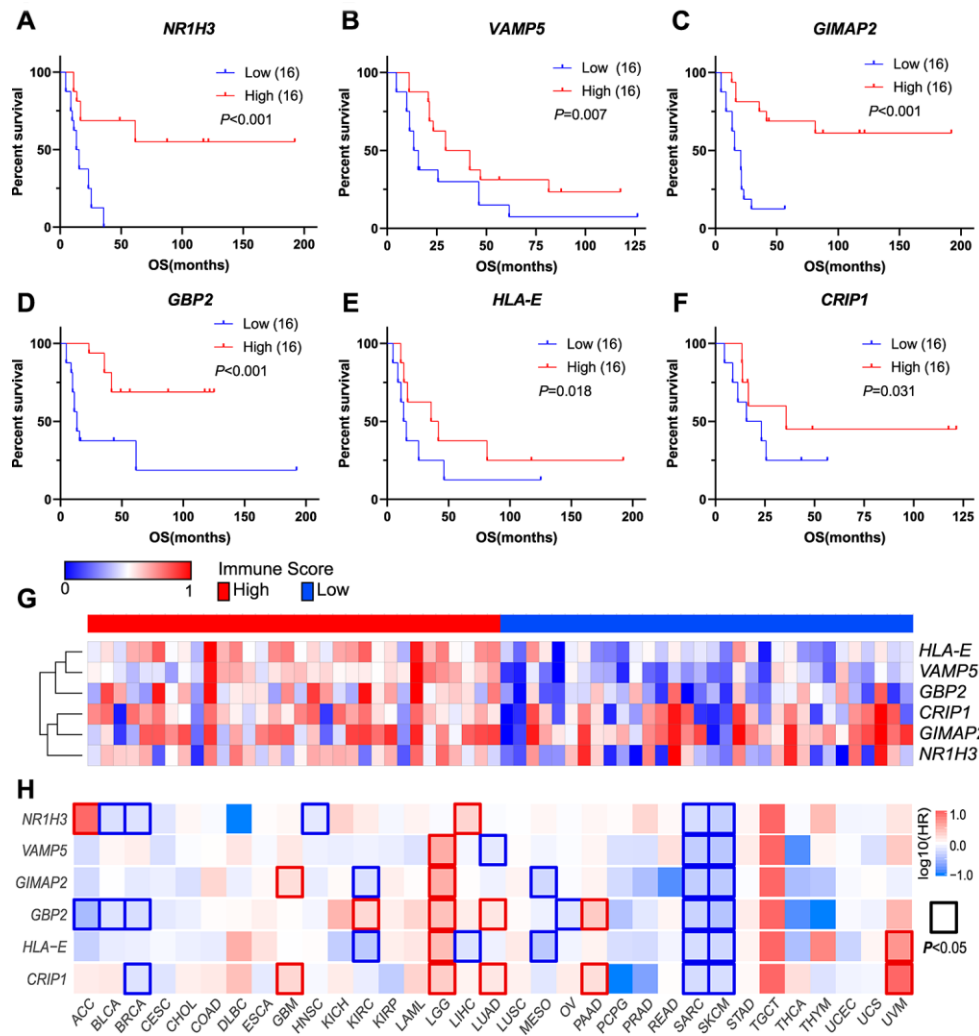
Supplementary Figure 2. Validation of the DEG correlation extracted from the database of TCGA with overall survival in GSE17679. The Kaplan-Meier survival curves for sarcoma patients further separated into high and low expression groups based on the quartiles of NR1H3 (**A**), VAMP5 (**B**), GIMAP2 (**C**), GBP2 (**D**), HLA-E (**E**) and CRIP1 (**F**), separately. (**G**) Heatmap of the mRNA levels of prognosis-related genes in dataset GSE17679. (**H**) The Mantel-Cox tests compare the survival contribution of NR1H3, VAMP5, GIMAP2, GBP2, HLA-E and CRIP1 mRNA levels (the top 30% and the bottom 30%, respectively) in patients from 33 cancer cohorts. The HR was restricted to the range −10 to 10 by replacing all values > 10 with 10 and all values < −10 with −10. HR: Hazard ratio.

Supplementary Figure 3. Scores of immune cell subpopulations in sarcoma samples. Single-sample gene set enrichment analysis identified the relative infiltration of immune cell subpopulations for 254 sarcoma samples. The relative infiltration of each cell type was normalized by min-max normalization.

Supplementary Table

Supplementary Table 1. Genes associated with good prognoses identified in TCGA.

Interferon-γ-mediated signaling pathway	CIITA, ICAM1, HLA-DQB2, HLA-A, HLA-C, OAS1, HLA-B, OAS2, HLA-E, TRIM22, HLA-DQA2, HLA-F, B2M, OASL, CD44, IRF5, IFNG, IRF8, IRF1, HLA-DPA1, HLA-DPB1, GBP2, GBP1, HLA-DRA
Antigen processing and presentation via MHC class II	HLA-DQB2, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DQA2, HLA-DOB, HLA-DRA
Antigen processing and presentation via MHC class I	TAP1, HLA-A, HLA-C, HLA-B, HLA-E, TAPBP, HLA-F, B2M
Macrophage chemotaxis	CCL2, LGALS3, CCL5
Neutrophil chemotaxis	CCL2, LGALS3, S100A9, CCL8, CCL5, VAV1, CCL18, CCL17, CCL22, CCL13, CCL23, IFNG, CSF3R, XCL1, XCL2, SYK
Dendritic cell chemotaxis	CCR7, CCR5, CCR2, CCL19, CCL5
Positive regulation of T cell chemotaxis	CCR2, TNFSF14, CCL5, XCL1
B cell activation	IKZF3, CXCR5, LAX1, ZAP70, CD79A, CD40, PRKCB, BTK