Mining TCGA to reveal immunotherapy-related genes for soft tissue sarcoma

Ruixin Li, MS[a], Fan Yao, MD[a],*[iD], Yijin Liu, MS[a], Xiaodan Wu, MS[a], Peng Su, MD[a], Tianran Li, MD[b], Nan Wu, MS[a]

Abstract

Immunotherapy of soft tissue sarcoma is considered an important development direction for the future. Bioinformatics analysis of genetic changes in tumors and the immune microenvironment around tumors has proven to be a mature and reliable method for predicting tumor prognosis. By mining the Cancer Genome Atlas Program database, we found immunotherapy targets of soft tissue sarcoma and analyzed their biological behavior. The data of 265 samples were downloaded to analyze the expression profile of soft tissue sarcomas. This included calculating tumor purity through the estimation of stromal and immune cells in malignant tumors using expression data, acquisition of differential genes as prognostic factors, and enrichment analysis of the differential genes. Survival analysis showed longer overall survival times for patients with higher immune scores. We obtained 83 survival-related differential genes through survival analysis, and 23 genes that could be used as independent risk factors for the prognosis of soft tissue sarcoma were obtained by multiple regression analysis of the differential genes and other recognized risk factors. Gene set enrichment analysis of the differential genes obtained immune and inflammatory gene ontology terms and signal pathways, including regulation of the T-cell apoptotic process and leukocyte transendothelial migration. After validation in an independent data set of the Gene Expression Omnibus database, 12 genes were confirmed as a result. We believe that these differential genes will be new targets for sarcoma immunotherapy and key genes for the prognosis of soft tissue sarcoma.

Abbreviations: BP = biological processes, CC = cellular components, David = the database for annotation visualization and integrated discovery, DEG = differential genes, DFS = disease-free survival, ESTIMATE = Estimation of Stromal and Immune cells in Malignant Tumors using Expression data, FC = fold change, FDR = false discovery rate, GEO = gene expression omnibus, GO = gene ontology, HR = hazard ratio, KEGG = Kyoto Encyclopedia of Genes and Genomes, K-M curve = Kaplan–Meier survival curves, LMS = leiomyosarcoma, MCODE = molecular complex detection, MF = molecular functions, MFH = malignant fibrous histiocytoma, MPNST = malignant peripheral nerve sheath tumor, NADPH = nicotinamide adenine dinucleotide phosphate, OS = overall survive, PPI = protein–protein interaction network, SPSS = statistical product and service solutions, STRING = search tool for the retrieval of interacting genes/proteins, TCGA = the cancer genome atlas program, UPS = undifferentiated pleomorphic sarcoma.

Keywords: ESTIMATE, immunotherapy, predictive gene signature, soft tissue sarcoma, TCGA, tumor microenvironment

1. Introduction

Soft tissue sarcoma is a malignant tumor whose etiology is still unknown. Its classification is variable, and depending on its source, its histological and biological characteristics also differ. The treatment of soft tissue sarcoma involves surgery, radiotherapy, and chemotherapy, and it is necessary to select high-risk patients for preoperative radiotherapy and chemotherapy in clinical studies. However, the classification of soft tissue sarcomas is complex, and conventional chemotherapy regimens are not effective for all types of tumors.[1] Numerous studies[2,3] have shown that soft tissue sarcomas are highly related to the immune system and can promote tumor growth by inhibiting immune infiltration in the tumor microenvironment. The relationship between the relevant genes and the tumor microenvironment has been studied extensively,[4–7] and it has been demonstrated that the tumor microenvironment can affect gene expression in tumor tissues, and the high expression of some genes in tumor tissues can lead to changes in the tumor microenvironment.[5,8–11] For example, undifferentiated pleomorphic sarcoma (UPS) and leiomyosarcoma have been shown to highly express antigen-presenting genes PD-1 and PD-L1, as well as T-cell infiltration-related genes.[10,12,13] The strong inflammatory expression in the immune microenvironment makes it possible to treat sarcomas by inducing an immune response.[2,14,15]

The ESTIMATE algorithm, based on the TCGA database, can be used to evaluate the purity of tumors.[4,16] Based on gene expression data, it involves the numbers of immune cells and stromal cells whose expression is highest in the tumor microenvironment. Thus, we can directly study the genes that have the highest impact using the immune and stromal scores, and the value of these genes in predicting the prognosis and survival rate of soft tissue sarcoma can be evaluated.[9,17,18] In this study, the ESTIMATE algorithm was employed to analyze data from the sarcoma group in the TCGA database and collect the immune, stromal, and estimate scores. These 3 scores were combined to identify genes associated with the prognosis of soft tissue sarcomas. We also confirmed that some of these differentially expressed genes (DEGs) can be used as independent risk factors for the prediction of the prognosis of soft tissue sarcoma. Through a subsequent confirmation based on the Gene Expression Omnibus (GEO) database, we obtained 12 genes that are meaningful in both the TCGA and GEO databases.

2. Materials and methods

2.1. Immune and stromal scores

The data of 265 samples were downloaded from the TCGA database to analyze the expression profile of soft tissue sarcomas. Soft tissue sarcoma dataset from the TCGA database include both sarcomas with simple genomic alteration and sarcomas with complex genomics.[19] The gene expression data of 265 cases were also downloaded from the TCGA database. Gene expression data used to calculate immune and stromal scores were derived from TCGA, as were clinical data, such as overall survival, age at first diagnosis, and residual tumor data. The independent dataset used to verify differential genes was derived from the GEO database, GSE30929 dataset (September 2, 2011).

2.2. Methods for calculating DEGs

Data analysis was conducted using the "limma" R package analysis, set $\log(FC)2 = 2$ and $P < .05$. This function will automatically calculate the differential genes for microarray data. The samples were divided into high and low scores groups based on median, and we preliminarily obtained downregulated DEGs and upregulated DEGs of each stromal, immune, and estimate groups by "limma" package. The intersection of downregulated and upregulated genes in the 3 groups was taken by Veen diagram. We combined these downregulated and upregulated DEGs for further survival analysis and selected the statistically significant genes.

2.3. Graphics

Volcano maps and differential gene survival curves were produced using the "SangerBox" tool. Kaplan–Meier survival curves of immunological and stromal scores and sarcoma typing scores were obtained using the GraphPad tool. The Venn diagram was drawn using the web tool "Venny." The overall protein–protein interaction (PPI) network was derived from the STRING database and analyzed using the "Cytoscape" tool. Molecular complex detection (MCODE) was used to divide the network into 3 modules, and only genes with more than 5 edges were selected as the key genes.

2.4. Gene enrichment analysis

Results of gene enrichment analysis were obtained from the Database for Annotation, Visualization, and Integrated Discovery (DAVID), including GO analysis results of biological processes (BPs), molecular functions, cellular components, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and a false discovery rate of <0.05 was set as the cutoff.

2.5. Statistical analysis methods

Univariate and multivariate regression analyses were conducted using Cox analysis and Kaplan–Meier survival analysis in the Statistical Product and Service Solutions (SPSS) tool, and only genes with a significant P-value of .05 in univariate and multivariate analyses were retained.

3. Results

3.1. 265 cases of soft tissue sarcoma classification

There were 59 cases of dedifferentiated liposarcoma (22.3%), 2 cases of desmoid tumor (0.8%), and 52 cases of undifferentiated pleomorphic sarcoma (UPS) (including pleomorphic, malignant fibrous histiocytoma (MFH), giant cell MFH/undifferentiated pleomorphic sarcoma with giant cells, and undifferentiated pleomorphic sarcoma) (19.6%),[20] 107 cases of leiomyosarcoma (LMS) (40.4%), 10 cases of malignant peripheral nerve sheath (MPNST) (3.8%), 25 cases of myxofibrosarcoma (9.4%), and 10 cases of synovial sarcoma (including synovial sarcoma-biphasic, synovial sarcoma-monophasic, synovial sarcoma-poorly differentiated) (3.77%).[21]

3.2. Immune and stromal scores

The range of estimate scores based on the ESTIMATE algorithm was from –2445.31 to 13,152.62. Stromal scores ranged from –740.76 to 6339.22. The range of immune scores was from –1890.07 to 7914.02.

The gene expression data were divided into 2 groups – high expression and low expression – based on the median immune and stromal scores. Kaplan–Meier survival curves were drawn, which showed that in the high immune score group, the total survival time was generally longer than that in the low immune score group (Fig. 1A). The median of overall survival times in the high and low immune score groups were 2464 days and 1722 days, respectively (log-rank test $P < .05$). Similarly, although the result was not statistically significant (log-rank test $P > .05$), the survival curve obviously indicated that the total survival time of the group with high stromal scores was higher than that of the group with low stromal scores, with a median total survival time of 2448 days for the high score group compared with 1722 days for the low score group (Fig. 1B).

3.3. Differentially expressed gene results

As shown in Figure 1, patients with higher immune/stromal scores had longer overall survival times. Therefore, we calculated differential genes based on immune/stromal scores and took the intersection of differential genes to search for the genes most strongly associated with survival. To further narrow the range of differential genes, we also included the estimated score as the stratification basis, since the estimated score is the sum of the immune score and the stromal score. In order to compare the difference in gene expression between the high and low score groups, we draw volcano plots to indicate differences in gene expression among cases with different immune, stromal, and estimate scores(Figs. 2A–C). We used the limma algorithm, with the setting, $\log (fc) 2 = 2$ and false discovery rate (FDR) = 0.05, to calculate the most significant DEGs for the low stromal, immune, and estimate score groups relative to the high group. In this step, we got 91 upregulated genes and 479 downregulated genes in the immune score group; 157 upregulated genes and 281 downregulated genes in the stromal score group; 100 upregulated genes and 482 downregulated genes in
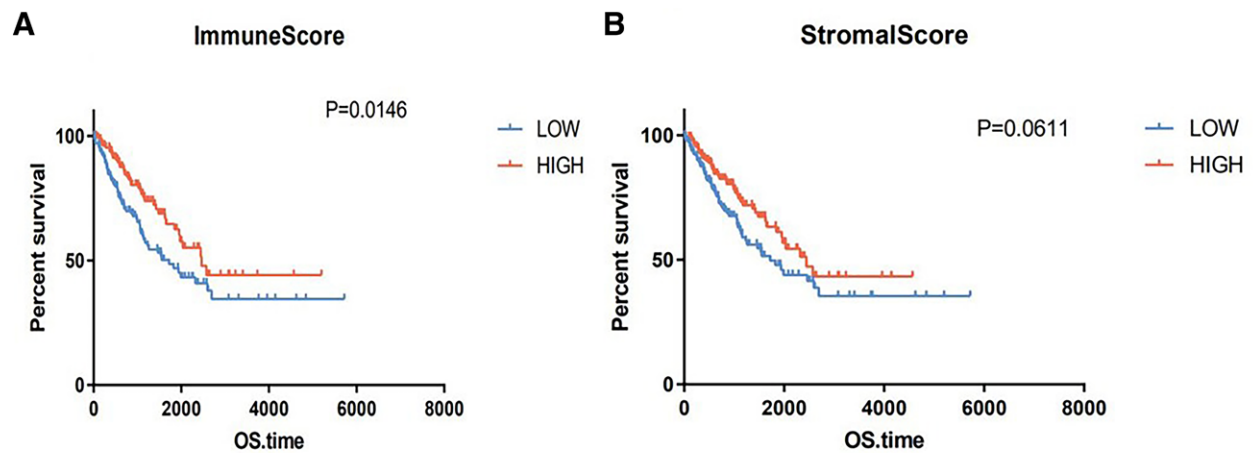
Figure 1. There is a relationship between immune and stromal scores and overall survival time. (A) People with high immune scores had significantly longer overall survival times. (B) A higher stromal score indicated a longer overall survival time, although this was not statistically significant.
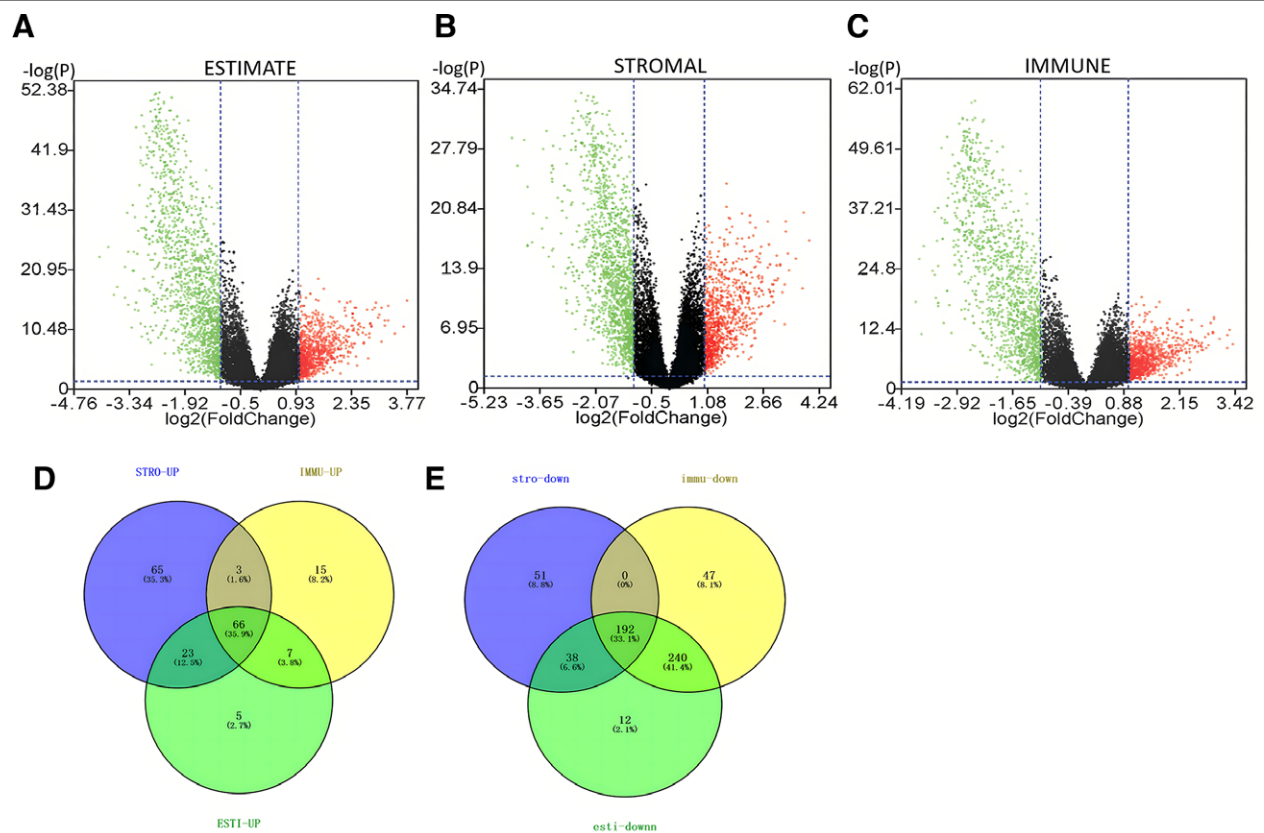

Figure 2. Number of upregulated and downregulated differentially expressed genes: (A–C) In the volcano map, green represents downregulated genes, red represents upregulated genes, and black represents genes with the same expression level in the high and low score groups. The parts with log2(FC) > 2 were selected as genes with significant differences to draw the Venn diagram. (D,E) Venn diagrams show genes that are generally upregulated or downregulated in all 3 scores.

the estimate score group. Finally, we used the Venn diagram to obtain the intersection of the 3 group with 66 upregulated genes and 192 downregulated genes (Fig. 2D–E).

3.4. Functional enrichment analysis of DEGs

A total of 18 cellular components, 26 molecular functions, and 95 BPs were obtained by GO analysis of 258 DEGs. Analysis of the top 10 terms resulted in the minimum *P*-values for immune and inflammatory responses, plasma membrane, receptor activity,

and chemokine activity. Similarly, the results of the KEGG pathway analysis were mainly related to immunity, among which the highest proportions included Staphylococcus aureus infection, accounting for 29.17%, Leishman disease, accounting for 25%, and viral protein interactions with cytokine and cytokine receptors, accounting for 12.5% (Fig. 3).

The enrichment analysis of upregulated genes and downregulated genes suggested that the downregulated genes mainly related to immune-associated biological process, but the upregulated genes related to signal transduction (Table 1).

3.5. Relationship between DEG and overall survival results

To explore the different roles of the 258 DEGs in survival, we set the median as the boundary to draw a Kaplan–Meier survival curve for each DEG. A total of 83 genes were statistically significant predictors of overall survival (log-rank $P < .05$) (Table 2). All 83 genes were used to construct a protein co-expression network, and multivariate regression analysis was used to further narrow down the key genes.

3.6. Construction of protein co-expression network results

A total of 15 key node genes were selected because they were most closely related to other genes. Cluster 1 had 14 nodes and 39 edges; the key nodes were LILRB2, AOAH, CCL5, NCF4,[22] and CYBB. Cluster 2 had 8 nodes and 16 edges; the key nodes were ITGAMCD2, IL2RG, and CD163. Cluster 3 had 16 nodes and 33 edges; the key nodes were VAV1, CXCR3, HCK, MS4A6A, LY86, and IDO1 (Fig. 4). Enrichment analysis was performed on these key genes, and the GO analysis results were as follows:

- Positive regulation of T-cell apoptotic process
- Negative regulation of T-cell apoptotic process
- Inflammatory response

The KEGG results included:

- Chemokine signaling pathway
- Leukocyte transendothelial migration

These results are all associated with the immune system (only results of $P < .05$ are included).

3.7. Multivariate regression analysis of DEG results (Table 3)

Using clinical data from the TCGA database, univariate regression analysis was performed to obtain "age at initial pathologic diagnosis" and "residual cancer" as prognostic factors
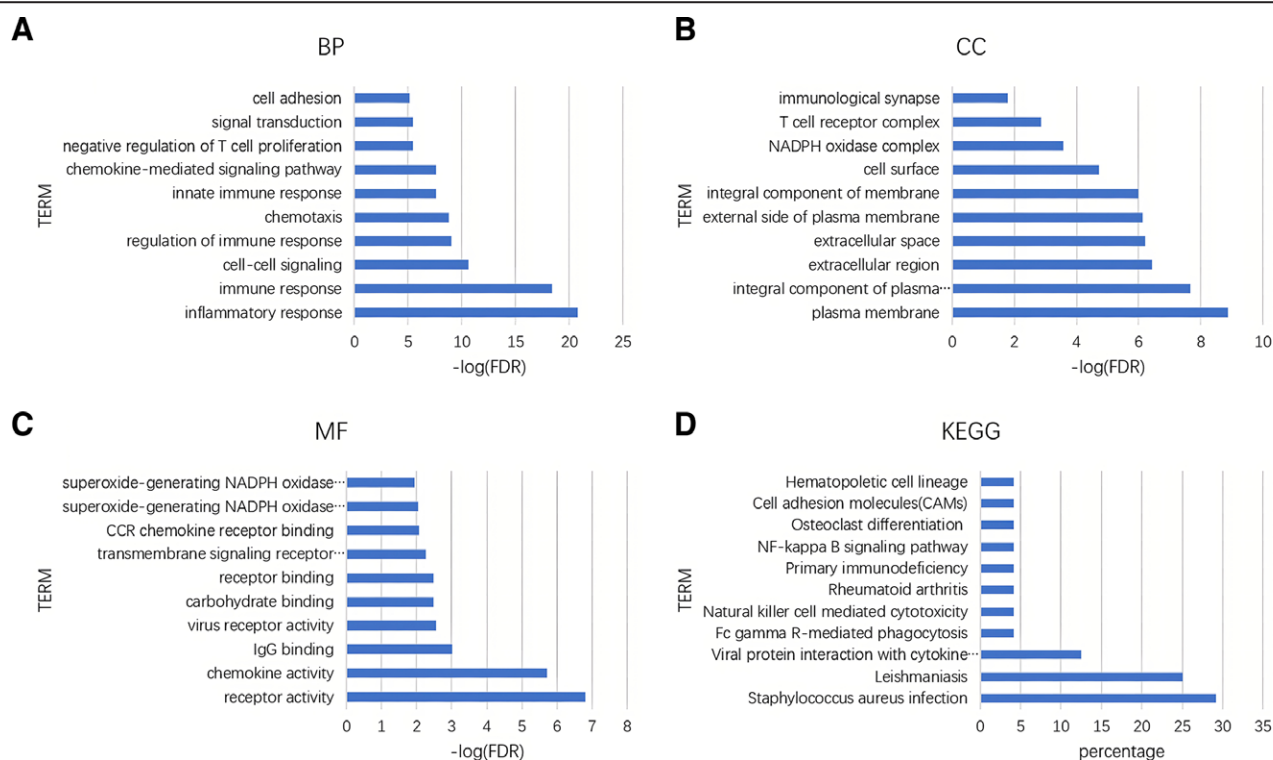
Figure 3. Top 10 terms of GO analysis results: (A) biological processes, (B) cellular components, (C) molecular functions, and (D) KEGG pathways. GO = gene ontology, KEGG = kyoto encyclopedia of genes and genomes.

Table 1

GO analysis outcomes of DEGs.

	Upregulated genes	Downregulated genes
BP	Collagen-activated tyrosine kinase	Inflammatory response
	Smooth muscle cell differentiation	Immune response
	Extracellular matrix organization	Cell surface receptor signaling pathway
	Retina layer formation	Cell-cell signaling
	Actomyosin structure organization	Innate immune response
	Positive regulation of excitatory postsynaptic potential	Chemotaxis
	Regulation of exocytosis	Negative regulation of t-cell proliferation
	Regulation of synaptic vesicle exocytosis	Neutrophil chemotaxis
	Positive regulation of transforming growth factor beta receptor signaling pathway	Chemokine-mediated signaling pathway
		Complement receptor mediated signaling pathway

DEG = differential genes, GO = gene ontology.

for sarcoma. Multivariate regression analysis of the differential genes was conducted to obtain a gene list that could predict the prognosis of sarcoma, including risk genes (HR > 1) and benign genes (HR < 1).

3.8. Verification of DEGs in the GEO database results

To verify the significance of the obtained differential genes, we used an independent dataset from the GEO database for validation. The dataset included sequencing data on gene expression levels in 140 liposarcoma patients and clinical data on disease-free survival.

In the GEO dataset, we divided the obtained DEGs into high expression and low expression groups according to the median. Disease-free survival time and survival outcomes were cited as variables for survival analysis(Fig. 5). A total of 12 prognostic DEGs with statistical significance (P < .05) were identified: CLEC10A,[23] IDO1,[24] FCER2,[25] CCL13,[26] KLRB1,[27,28] P2RY13,[4] PYHIN1,[29,30] DOK2,[31] PLA2G2D,[32,33] MARCO,[34] MYH11,[5] and NCF4. Nine of the DEGs caught our attention, because they have not been reported to be associated with the prognosis of sarcoma: CLEC10A, FCER2, CCL13, KLRB1, P2RY13, PYHIN1, PLA2G2D, MARCO, and NCF4.

4. Discussion

Soft tissue sarcoma is a type of tumor with complicated classification and high malignancy, and its immunotherapy is a hot topic. To identify relevant targets for immunotherapy for soft tissue sarcomas, we performed a series of data analyses using samples from the TCGA database. By calculating the immune and stromal scores, we confirmed that the immunological/stromal scores differed among the different types of soft tissue sarcomas. As shown in previous studies, UPS expressed more T-cell permeation-related genes, while synovial sarcoma expressed significantly fewer. Therefore, UPS is a kind of soft tissue sarcoma suitable for immunotherapy.[13,35] The overall survival time was longer in the high immune group than in the low immune group, indicating that the high expression of some immune-related genes was beneficial in prolonging the overall survival time.

To further understand which genes most greatly affected immune scores, we performed a differential analysis and obtained differential genes. Our results showed that most genes were downregulated when the low score group was compared to the high score group. According to the Venn diagrams, we chose 258 genes with differential expression in 3 scores. These differential genes were analyzed by functional enrichment. The top 10 results of GO analysis suggested that these DEGs had

Table 2

DEGs with survival significance.

Genes	Log-rank P	HR
ADAM6	.004	>1
ADAMDEC1	.035	>1
AOAH	.018	>1
ADAP2	.047	<1
BATF	.036	>1
C1QC	.028	>1
C1QL4	.018	<1
C13orf33	.019	>1
CADM3	.024	>1
CCL5	.043	>1
CCL13	.036	>1
CCL18	.026	>1
CCL19	.002	>1
CD2	.019	>1
CD8B	.006	>1
CD38	.010	>1
CD53	.016	>1
CD163	.047	>1
CLEC4G	.008	>1
CLEC4GP1	.046	>1
CLEC10A	.000	>1
CSDC2	.025	<1
CSF1R	.014	>1
CXCR3	.008	>1
CYBB	.037	>1
DOK2	.005	>1
EMILIN3	.004	<1
FCER2	.000	>1
GGTA1	.020	>1
GPR34	.010	>1
GZMK	.032	>1
GAL3ST3	.045	>1
HCK	.029	>1
IDO1	.001	>1
IGJ	.010	>1
IL10RA	.005	>1
IL18	.008	>1
ILR2G	.005	>1
ITGAM	.014	>1
KLRB1	.001	>1
KLHL23	.013	>1

(Continued)

Table 2

(Continued)

| Genes	Log-rank P	HR
LATR1	.037	>1
LAIR1	.047	<1
LCK	.018	>1
LILRB1	.016	>1
LILRB2	.010	>1
LILRB5	.010	>1
LRRC25	.021	>1
LTB	.002	>1
LY86	.030	>1
LYVE1	.043	>1
MARCO	.018	>1
MGC29506	.004	>1
MPEG1	.033	>1
MS4A6A	.040	>1
MS4A7	.027	>1
MYH11	.023	>1
MYOC	.016	>1
NCF1	.005	>1
NCF1B	.037	>1
NCF1C	.032	>1
NCF4	.039	>1
NKG7	.038	>1
P2RY13	.003	>1
PLA2G2A	.011	>1
PLA2G2D	.016	>1
PYHIN1	.021	>1
RARRES1	.022	>1
S100A9	.045	>1
SAMSN1	.002	>1
SASH3	.002	>1
SH2D1A	.021	>1
SIGLEC11	.044	>1
SIGLEC14	.016	>1
SLA	.043	>1
SLAMF6	.037	>1
TBXAS1	.025	>1
TLR8	.006	>1
TNFRSF1B	.047	>1
TNFAIP8L2	.020	<1
VAV1	.009	>1
WAS	.023	>1
WISP2	.010	>1

DEG = differential genes.
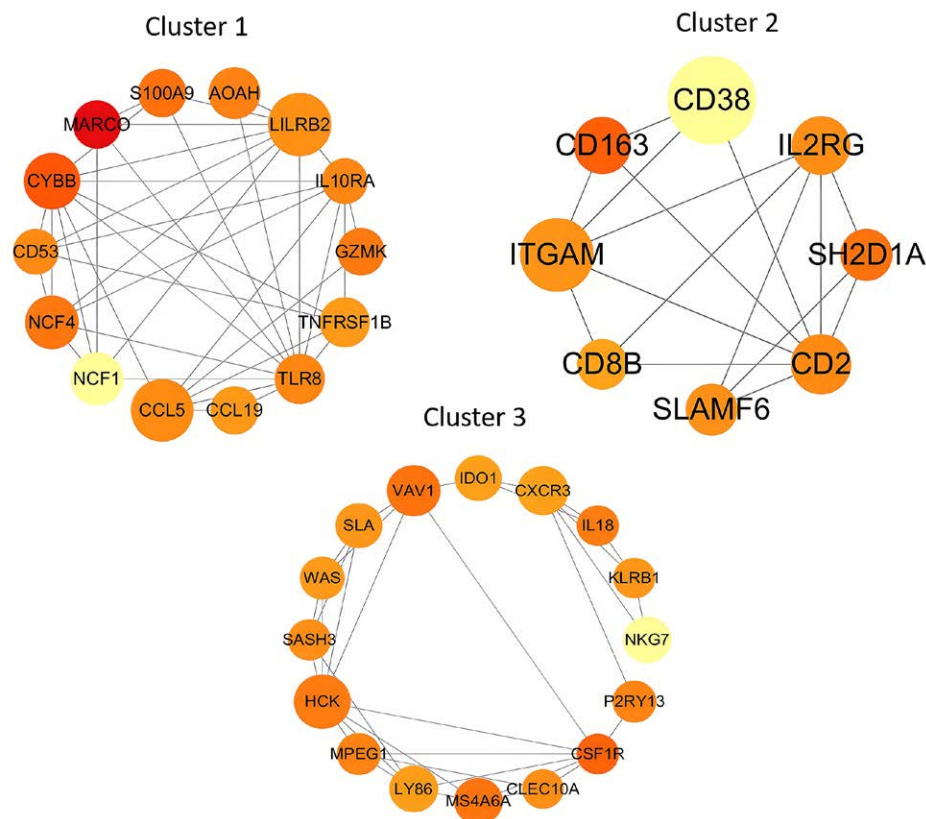
Figure 4. The PPI network of 83 different genes was created to discover the relationships, including Clusters 1, 2, and 3. A protein expression network of 83 DEGs was obtained using the STRING web tool and divided into 3 clusters with 64 nodes and 326 edges using cystoscope analysis. The depth of the color corresponds to the multiple gene expression differences, and the size of the circle corresponds to the correlation between proteins and the surrounding network.

Table 3

Univariate regression analysis of DEGs.

Group	Significant value	HR	Confidence interval 95.0%	
age_at_initial_pathologic_diagnosis	.000	1.082	1.035	1.130
residual_tumor	.015	3.536	1.279	9.777
ADAM6	.012	5.874	1.480	23.307
CSF1R	.001	0.042	0.007	0.266
CYBB	.003	10.208	2.159	48.262
MPEG1	.049	0.253	0.064	0.996
NCF4	.000	69.966	8.793	556.733
TBXAS1	.019	4.680	1.282	17.084
C13orf33	.008	0.608	0.421	0.878
GPR34	.044	3.385	1.034	11.079
ITGAM	.019	0.165	0.037	0.741
DOK2	.030	0.219	0.056	0.862
CD163	.015	0.186	0.048	0.726
TLR8	.041	2.663	1.043	6.799
LYVE1	.001	2.269	1.415	3.638
BATF	.000	0.121	0.045	0.323
P2RY13	.030	0.434	0.204	0.922
SLAMF6	.037	3.636	1.080	12.236
S100A9	.000	0.373	0.220	0.630
CD38	.023	0.412	0.192	0.884
FCER2	.033	0.669	0.463	0.968
CADM3	.013	0.823	0.706	0.959
WISP2	.049	1.232	1.001	1.516
MYH11	.000	0.565	0.448	0.712

DEG = differential genes.

immune-related functions. So it was in the KEGG analysis. As we know, sarcoma tumor cells inhibit the immune response in the tumor microenvironment to promote its growth and metastasis. Therefore, some treatments use trace amounts of Staphylococcus aureus infection to stimulate the immune response of the body to treat sarcoma.[12,14] People with low
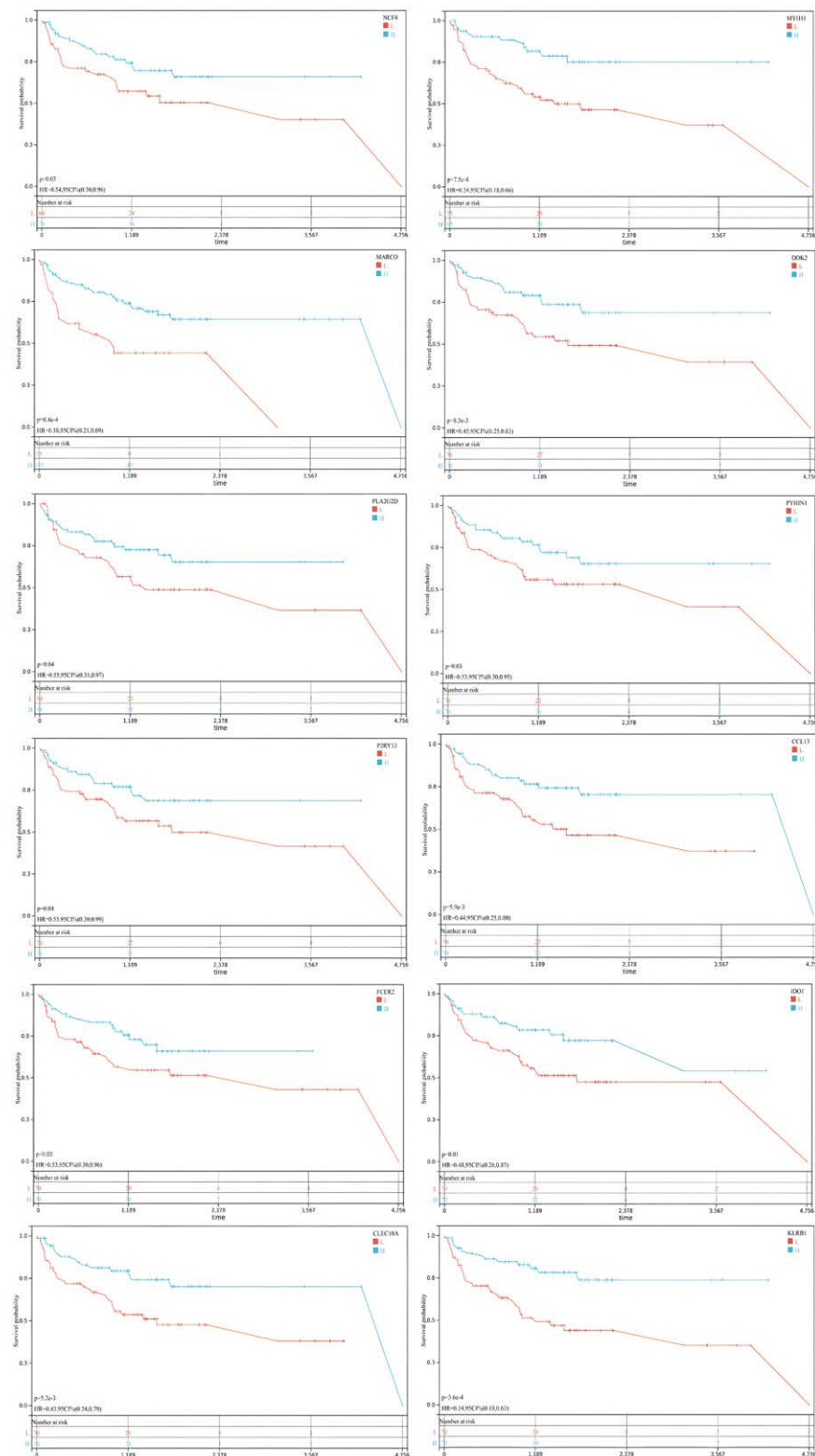
Figure 5. Kaplan–Meier survival curve of DEGs in both TCGA and GEO dataset. A total of 12 DEGs with survival significance were obtained. This figure shows the relationship between DFS and gene expression as calculated from clinical data in the GEO database.

immunity are susceptible to infection with the Leishmania parasite and often develop Kaposi sarcoma. Kaposi sarcoma, which is highly prevalent in AIDS patients, is greatly reduced with the use of cocktail therapy. This shows that the occurrence and development of sarcoma are directly related to the stability of the patient's internal immune environment.[14] As in previous

studies, we also identified immune-related pathways such as NF-κB signaling and NK cell mediated cytotoxicity.[36]

By plotting survival curves for individual genes, we disproved the obtained relationship between immune/stromal scores and overall survival time. AOAH, for example, is a downregulated gene; that is, AOAH gene expression was low in the low score

group compared to the high score group. The higher the expression of AOAH, the higher the corresponding immune/stromal score, and the longer the overall survival time. This corresponded to a longer overall survival rate in the high immune/stromal rating group. In total, there were 83 survival-related DEGs. We performed enrichment analyses on the 83 DEGs; the outcomes included 51 BP, 9 CC, and 11MF that were identified as significant. The top terms were immune and inflammatory responses, NADPH oxidase complex,[36] integral component of plasma membrane, superoxide-generating NADPH oxidase activator activity, and superoxide-generating NADPH oxidase activity. After comparing the enrichment analysis results of upregulated genes and downregulated genes, we found that downregulated genes were mainly involved in immune-related BPs, while upregulated genes were mainly involved in signal transduction-related BPs, which provided us with ideas for finding new gene targets.

In further study, we constructed a protein expression network of 83 genes, which was divided into 4 modules, among which the main 3 modules included 15 key gene nodes because they are most closely related to the surrounding genes: LILRB2, AOAH, CCL5, NCF4, CYBB, ITGAMCD2, IL2RG, CD163, VAV1, CXCR3, HCK, MS4A6A, LY86, and IDO1. Through the PPI network, the relationship between DEGs was figured out, and the key proteins that affected survival were identified. We found that these genes all had immune-related functions using functional enrichment analysis, and we confirmed that these DEGs were significant in immunotherapy. Among them, IDO1 has been shown to be related to the growth and immunosuppression of various tumors,[10,24,35] and in previous studies, CCL5 has demonstrated the ability to recruit CD8+ T cells, leading to a better prognosis,[37] which verifies the results of our study.

To confirm that DEGs can be used as prognostic indicators, a proportional hazards model was performed on the 83 DEGs that were significant for survival, and significant genes were obtained ($P < .05$), including 10 genes that were indicators of malignancy – ADAM6, CYBB, NCF4,[22] TBXAS1, GPR34, TLR8, LYVE1, SLAMF6, WISP2 and KLHL23(HR > 1) – and 13 genes (HR < 1) that were good indicators: CSF1R, MPEG1, C13orf33, ITGAMDOK2,[31] CD163, BATF, P2RY13,[4] S100A9, CD38, FCER2,[25] CADM3, and MYH11. NCF4, CYBBITGAM, and CD163 are also key nodes in PPI. A cluster analysis confirmed the importance of the MYH11-ADM regulatory network in STS patients,[38] which further supports the results of this study.

Finally, we used an independent group of 140 samples from the GEO database to verify the roles of these genes in survival and confirmed that 12 genes were meaningful in this independent group. Nine of the genes are involved in immunity and have been reported to be involved in the development or metastasis of different types of tumors, while none have been reported to be associated with the development and progression of sarcomas. This means that these genes are likely to play a role in the treatment of sarcoma as immunotherapeutic targets. Compared with the sarcoma-related mutant genes previously revealed by gene sequencing, this study identified other genes and related pathways previously unrelated to the disease from the perspective of the tumor microenvironment.[7] We believe that these genes will be new targets for sarcoma immunotherapy and key genes for the analysis of sarcoma prognosis. This will address a key step in tumor immunotherapy, and effective tumor cell surface antigens can be developed so that immune agents can attack tumor cells directly.

Among these, CLEC10A aroused our interest. CLEC10A is a specific marker of CD1c DC.[39] When it binds to the receptor, it can enhance the secretion of TNF+, IL-8, and IL-10 induced by TLR 7/8 stimulation,[23,39] thereby promoting humoral immunity and inhibiting tumor progression. It has been shown that dendritic cells are independent risk factors for the recurrence of soft tissue sarcomas and are directly related to tumor immunity.[11] In addition, CLEC10A can be used as a marker of damaged and dead cells. After binding with ligands, CLEC10A is preferentially internalized by macrophages to attack tumor cells from the perspective of cellular immunity. In the treatment of breast cancer, we found that female hormones have an inhibitory effect on the CLEC10A ligand, thus demonstrating the induction of the CLEC10A ligand by the estrogen receptor antagonist tamoxifen,[23] which inspires us to explore endocrine therapy for sarcoma.

In conclusion, we believe that these genes will be new targets for sarcoma immunotherapy and key genes for the analysis of sarcoma prognosis, and provide gene targets for immunotherapy for soft tissue sarcoma.

Author contributions

Data curation: Xiaodan Wu.
Funding acquisition: Fan Yao.
Project administration: Peng Su.
Supervision: Peng Su.
Validation: Nan Wu.
Writing – original draft: Ruixin Li, Yijin Liu.
Writing – review & editing: Fan Yao, Tianran Li.

References

[1] Hall KS, Bruland OS, Bjerkehagen B, et al. Preoperative accelerated radiotherapy combined with chemotherapy in a defined cohort of patients with high risk soft tissue sarcoma: a Scandinavian Sarcoma Group study. Clin Sarcoma Res. 2020;10:22.

[2] Francescutti V, Skitzki JJ. Sarcomas and the immune system: implications for therapeutic strategies. Surg Oncol Clin N Am. 2012;21:341–55.

[3] Maki RG. Soft tissue sarcoma as a model disease to examine cancer immunotherapy. Curr Opin Oncol. 2001;13:270–4.

[4] Fan T, Zhu M, Wang L, et al. Immune profile of the tumor microenvironment and the identification of a four-gene signature for lung adenocarcinoma. Aging (Albany NY). 2020;13:2397–417.

[5] Hu C, Chen B, Huang Z, et al. Comprehensive profiling of immune-related genes in soft tissue sarcoma patients. J Transl Med. 2020;18:337.

[6] Zhu N, Hou J. Assessing immune infiltration and the tumor microenvironment for the diagnosis and prognosis of sarcoma. Cancer Cell Int. 2020;20:577.

[7] Barretina J, Taylor BS, Banerji S, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. Nat Genet. 2010;42:715–21.

[8] Zhang B, Yang L, Wang X, Fu D. Identification of a survival-related signature for sarcoma patients through integrated transcriptomic and proteomic profiling analyses. Gene. 2021;764:145105.

[9] Li F, Hu W, Zhang W, Li G, Guo Y. A 17-gene signature predicted prognosis in renal cell carcinoma. Dis Markers. 2020;2020:8352809.

[10] Toda Y, Kohashi K, Yamada Y, et al. PD-L1 and IDO1 expression and tumor-infiltrating lymphocytes in osteosarcoma patients: comparative study of primary and metastatic lesions. J Cancer Res Clin Oncol. 2020;146:2607–20.

[11] Huang R, Meng T, Chen R, et al. The construction and analysis of tumor-infiltrating immune cell and ceRNA networks in recurrent soft tissue sarcoma. Aging (Albany NY). 2019;11:10116–43.

[12] Toulmonde M, Penel N, Adam J, et al. Use of PD-1 targeting, macrophage infiltration, and IDO pathway activation in sarcomas: a phase 2 clinical trial. JAMA Oncol. 2018;4:93–7.

[13] Pollack SM, He Q, Yearley JH, 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.

[14] Burgess M, Gorantla V, Weiss K, Tawbi H. Immunotherapy in sarcoma: future horizons. Curr Oncol Rep. 2015;17:52.

[15] Mesiano G, Leuci V, Giraudo L, et al. Adoptive immunotherapy against sarcomas. Expert Opin Biol Ther. 2015;15:517–28.

[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.

[17] Lan H, Zeng J, Chen G, Huang H. Survival prediction of kidney renal papillary cell carcinoma by comprehensive LncRNA characterization. Oncotarget. 2017;8:110811–29.

[18] Peille AL, Brouste V, Kauffmann A, et al. Prognostic value of PLAGL1-specific CpG site methylation in soft-tissue sarcomas. PLoS One. 2013;8:e80741.

[19] Widemann BC, Italiano A. Biology and management of undifferentiated pleomorphic sarcoma, myxofibrosarcoma, and malignant peripheral nerve sheath tumors: state of the art and perspectives. J Clin Oncol. 2018;36:160–7.

[20] von Mehren M, Kane JM, Agulnik M, et al. Soft tissue sarcoma, version 2.2022, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2022;20:815–33.

[21] Nakayama R, Nemoto T, Takahashi H, et al. Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol. 2007;20:749–59.

[22] Ryan BM, Zanetti KA, Robles AI, et al. Germline variation in NCF4, an innate immunity gene, is associated with an increased risk of colorectal cancer. Int J Cancer. 2014;134:1399–407.

[23] Kurze AK, Buhs S, Eggert D, et al. Immature O-glycans recognized by the macrophage glycoreceptor CLEC10A (MGL) are induced by 4-hydroxy-tamoxifen, oxidative stress and DNA-damage in breast cancer cells. Cell Commun Signal. 2019;17:107.

[24] Nafia I, Toulmonde M, Bortolotto D, et al. IDO targeting in sarcoma: biological and clinical implications. Front Immunol. 2020;11:274.

[25] Sharma V, Michel S, Gaertner V, et al. A role of FCER1A and FCER2 polymorphisms in IgE regulation. Allergy. 2014;69:231–6.

[26] Mendez-Enriquez E, García-Zepeda EA. The multiple faces of CCL13 in immunity and inflammation. Inflammopharmacology. 2013;21:397–406.

[27] Zhang G, Liu Y, Dong F, Liu X. Transcription/expression of KLRB1 gene as a prognostic indicator in human esophageal squamous cell carcinoma. Comb Chem High Throughput Screen. 2020;23:667–74.

[28] Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21:938–45.

[29] Massa D, Baran M, Bengoechea JA, Bowie AG. PYHIN1 regulates pro-inflammatory cytokine induction rather than innate immune DNA sensing in airway epithelial cells. J Biol Chem. 2020;295:4438–50.

[30] Tong Y, Song Y, Deng S. Combined analysis and validation for DNA methylation and gene expression profiles associated with prostate cancer. Cancer Cell Int. 2019;19:50.

[31] Ohsugi T. Effects of expressing human T-cell leukemia virus type 1 (HTLV-I) oncoprotein Tax on DOK1, DOK2 and DOK3 gene expression in mice. J Vet Med Sci. 2017;79:935–8.

[32] Miki Y, Kidoguchi Y, Sato M, et al. Dual roles of group IID phospholipase A2 in inflammation and cancer. J Biol Chem. 2016;291:15588–601.

[33] Miki Y, Yamamoto K, Taketomi Y, et al. Lymphoid tissue phospholipase A2 group IID resolves contact hypersensitivity by driving antiinflammatory lipid mediators. J Exp Med. 2013;210:1217–34.

[34] Xiao Y, Chen B, Yang K, et al. Down-regulation of MARCO associates with tumor progression in hepatocellular carcinoma. Exp Cell Res. 2019;383:111542.

[35] Toulmonde M, Lucchesi C, Verbeke S, et al. High throughput profiling of undifferentiated pleomorphic sarcomas identifies two main subgroups with distinct immune profile, clinical outcome and sensitivity to targeted therapies. EBioMedicine. 2020;62:103131.

[36] Hu C, Wang Y, Liu C, et al. Systematic profiling of alternative splicing for sarcoma patients reveals novel prognostic biomarkers associated with tumor microenvironment and immune cells. Med Sci Monit. 2020;26:e924126.

[37] Martner A, Aydin E, Hellstrand K. NOX2 in autoimmunity, tumor growth and metastasis. J Pathol. 2019;247:151–4.

[38] Morales E, Olson M, Iglesias F, Dahiya S, Luetkens T, Atanackovic D. Role of immunotherapy in Ewing sarcoma. J ImmunoTher Cancer. 2020;8:e000653.

[39] Heger L, Balk S, Lühr JJ, et al. CLEC10A is a specific marker for human CD1c+ dendritic cells and enhances their toll-like receptor 7/8-induced cytokine secretion. Front Immunol. 2018;9:744.