

Risk score based on expression of five novel genes predicts survival in soft tissue sarcoma

Hui-Yun Gu¹, Chao Zhang², Jia Guo³, Min Yang¹, Hou-Cheng Zhong¹, Wei Jin¹, Yang Liu², Li-Ping Gao⁴, Ren-Xiong Wei¹

¹Department of Spine and Orthopedic Oncology, Zhongnan Hospital of Wuhan University, Wuhan, China

²Center for Evidence-Based Medicine and Clinical Research, Taihe Hospital, Hubei University of Medicine, Shiyan, China

³Department of Plastic Surgery, Zhongnan Hospital of Wuhan University, Wuhan, China

⁴The Third Clinical School, Hubei University of Medicine, Shiyan, China

Correspondence to: Ren-Xiong Wei; **email:** wls0821@126.com

Keywords: soft tissue sarcoma, least absolute shrinkage and selection operator regression analysis, biomarker, prognostic model, nomogram

Received: November 9, 2019

Accepted: February 4, 2020

Published: February 21, 2020

Copyright: Gu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

In this study, The Cancer Genome Atlas and Genotype-Tissue Expression databases were used to identify potential biomarkers of soft tissue sarcoma (STS) and construct a prognostic model. The model was used to calculate risk scores based on the expression of five key genes, among which MYBL2 and FBN2 were upregulated and TSPAN7, GCSH, and DDX39B were downregulated in STS patients. We also examined gene signatures associated with the key genes and evaluated the model's clinical utility. The key genes were found to be involved in the cell cycle, DNA replication, and various cancer pathways, and gene alterations were associated with a poor prognosis. According to the prognostic model, risk scores negatively correlated with infiltration of six types of immune cells. Furthermore, age, margin status, presence of metastasis, and risk score were independent prognostic factors for STS patients. A nomogram that incorporated the risk score and other independent prognostic factors accurately predicted survival in STS patients. These findings may help to improve prognostic prediction and aid in the identification of effective treatments for STS patients.

INTRODUCTION

Soft tissue sarcomas (STS), which are derived from mesenchymal tissues [1], are highly clinically diverse and can originate from many sources, including muscle, adipose tissue, peripheral nerves, blood vessels, and connective tissue [2]. In the United States of America, STS accounts for 1% of cancer cases and 2% of cancer-related deaths [3]. In 2019, 12,750 new STS cases were diagnosed and 5,270 patients died from the disease [4].

Diagnosis and treatment of the multiple histological types of STS are challenging for physicians [5], and a multidisciplinary approach is often beneficial [6]. Because metastasis and disease recurrence are common

in STS, patients with localized and early-stage STS could benefit from early diagnosis and radical resection [2, 7–10]. Currently, imaging and biopsy are the primary methods recommended for diagnosing STS. Magnetic resonance imaging is the most effective method for identifying STS originating in the extremities, pelvis, and trunk, while standard radiography and computed tomography are typically used to detect STS in other areas [9]. In the era of precision medicine, identification of biomarkers for and molecular characterization of STS will likely play an increasingly prominent role in diagnosis, treatment, and prognosis prediction. An integrated genomic characterization analysis of 206 adult STS patients conducted by The Cancer Genome Atlas (TCGA)

Research Network has been crucial for improving treatment of STS [11]. Nevertheless, identification of additional specific biomarkers would further improve STS treatments.

The TCGA database contains genomic and clinical data from >20,000 primary cancer and matched normal samples representing 33 types of cancer [12]. In addition, the Genotype-Tissue Expression (GTEx) database contains openly available clinical data, including gene expression, quantitative trait locus, and histology images, from 1,000 healthy individuals [13, 14]. In this study, we conducted an integrated analysis of gene expression profiles and clinical data from these databases to identify common gene signatures associated with the development, pathological mechanisms, and prognosis of STS. In addition, we established a prognostic model and nomogram for STS based on clinical data obtained from the TCGA database.

RESULTS

Samples and clinical data

TCGA Sarcoma gene expression profiles and clinical data were downloaded from the UCSC Xena Hub datasets. Following data preprocessing, gene expression profiles from 263 STS samples and two matched controls and survival information and other clinical data from 256 patients were included in subsequent analysis. The 263 STS samples included 106 (40%) leiomyosarcomas, 58 (22%) dedifferentiated liposarcomas, 52 (18%) undifferentiated pleomorphic sarcomas, 25 (10%) myxofibrosarcomas, 10 (4%) malignant peripheral nerve sheath tumors, 10 (4%) synovial sarcomas, and two (0.8%) desmoid tumors. Patients ranged from 20 to 89 years in age; 54% were female and 46% were male. The matched controls obtained from the GTEx database included 396 muscle and 515 fat samples.

Identification and functional annotation of primary differentially expressed genes (DEGs)

DEGs were obtained from the three groups. A total of 2,290 and 1,301 genes were identified as DEGs in the muscle and fat control groups ($|\log_2 \text{FC}| > 2$, $p < 0.05$), respectively. 775 DEGs were identified in the matched controls group ($|\log_2 \text{FC}| > 1$, $p < 0.05$). Heatmaps of DEGs from the three groups are displayed in Supplementary Figure 1A–1C. 121 DEGs common to all three groups were identified as primary DEGs and examined in subsequent analyses. Functional annotations revealed that primary DEGs were mainly involved in mitotic nuclear division, sister chromatid

segregation, chromosome segregation, nuclear division, mitotic sister chromatid segregation, and nuclear chromosome segregation based on the top six terms identified in the GO analysis ($p < 0.05$; Figure 1A, Supplementary Figure 2A, 2B). KEGG analysis showed that primary DEGs were associated with cell cycle, DNA replication, cellular senescence, oocyte meiosis, progesterone-mediated oocyte maturation, and the p53 signaling pathway based on the top six terms ($p < 0.05$; Figure 1B, Supplementary Figure 2C, 2D). The results of these functional annotation suggest that the 121 primary DEGs are associated with the formation and development of tumors.

Identification of survival-related DEGs and establishment of a prognostic model

A total of 256 patients were randomly assigned to the training (128 patients) and test (128 patients) groups. A total of 44 DEGs were selected via univariate Cox regression analysis ($p < 0.05$). Following Lasso regression analysis, seven DEGs were selected for multivariate Cox regression analysis (optimal $\log(\text{Lambda})$: 7; Figure 2A, 2B). Finally, five key genes highly associated with survival were used to establish the prognostic model: $\text{risk score} = -0.201 \times \text{exp}(\text{TSPAN7}) + 0.284 \times \text{exp}(\text{MYBL2}) + 0.941 \times \text{exp}(\text{GCSH}) + 0.159 \times \text{exp}(\text{FBN2}) + 0.417 \times \text{exp}(\text{DDX39B})$ (Table 1). Patients were separated into high- and low-risk groups based on this risk score; those in the high-risk group had lower survival rates ($p = 6.691 \times 10^{-5}$, Figure 3A). As shown in Figure 3D, these five survival-related DEGs performed satisfactorily in predicting prognosis of STS patients (area under the curve: 0.781). There were obvious differences in expression of the key genes between patients with lower and higher risk scores (Figure 3G). Moreover, those with higher risk scores had shorter survival times. Similar results were obtained for the test group, which validated the prognostic model (Figure 3B, 3E, 3H). Finally, we confirmed the value of this prognostic model by testing it in the overall patient group, which included all patients in the study (Figure 3C, 3F, 3I).

Expression levels of the five key DEGs in STS patients

Among the five key genes, expression of MYBL2 and FBN2 was increased, and expression of TSPAN7, GCSH, and DDX39B was decreased, in STS patients ($p < 0.001$; Supplementary Figure 3A and 3B). Expression levels of all of these key genes except DDX39B, which was not indexed in the external dataset (GSE21122), were validated in Supplementary Figure 3C. The five key genes effectively discriminated

between STS patients and controls (Supplementary Figure 4). As shown in Supplementary Figure 5A and 5B, the five key genes could also distinguish between low- and high-risk patients and healthy individuals, further validating the accuracy of the prognostic model ($p < 0.001$). Moreover, there were significant differences in

expression of the five key genes between all histological types and controls (Supplementary Figure 6A and 6B). This result was also validated using the GSE21122 datasets (Supplementary Figure 6C), indicating these five key genes were suitable for predicting prognosis in all histological types included in this study.

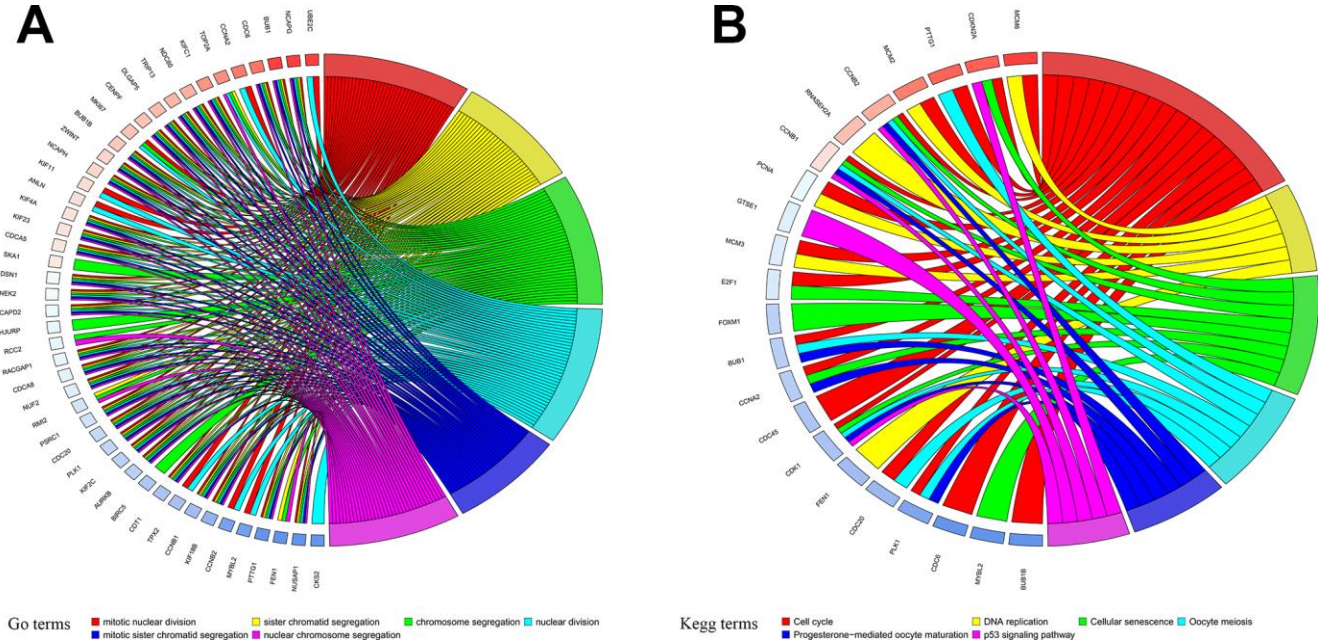


Figure 1. Functional annotation of primary differentially expressed genes (DEGs). (A) Gene Ontology (GO) functional annotation and (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments for DEGs.

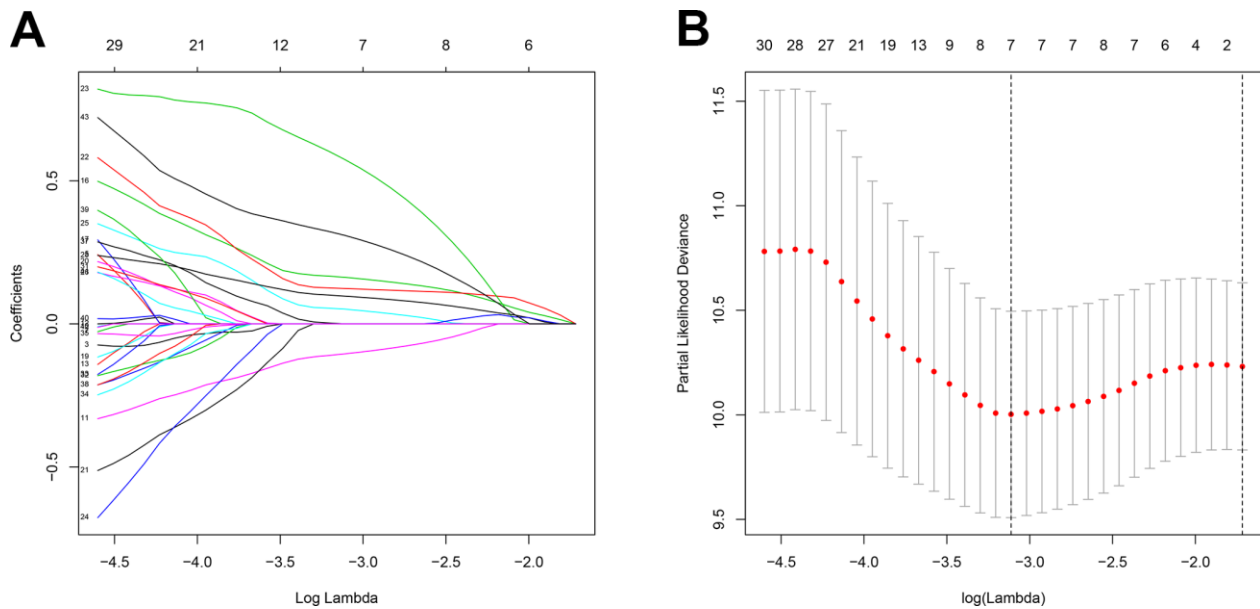


Figure 2. Feature selection using the Lasso regression model. (A) Lasso regression analysis coefficients. (B) Selection of tuning parameters in the Lasso regression analysis based on 1,000 cross-validations.

Table 1. Multivariate Cox regression analysis of key genes.

Genes	Overall survival				
	coef	HR	95% CI	p value	
TSPAN7	-0.201	0.818	0.652	1.026	0.082
MYBL2	0.284	1.328	1.057	1.669	0.015
GCSH	0.941	2.562	1.388	4.728	0.003
FBN2	0.159	1.172	1.006	1.366	0.041
DDX39B	0.417	1.518	1.071	2.150	0.019

Note: coef: coefficient; HR: hazard ratio; CI: confidence interval.

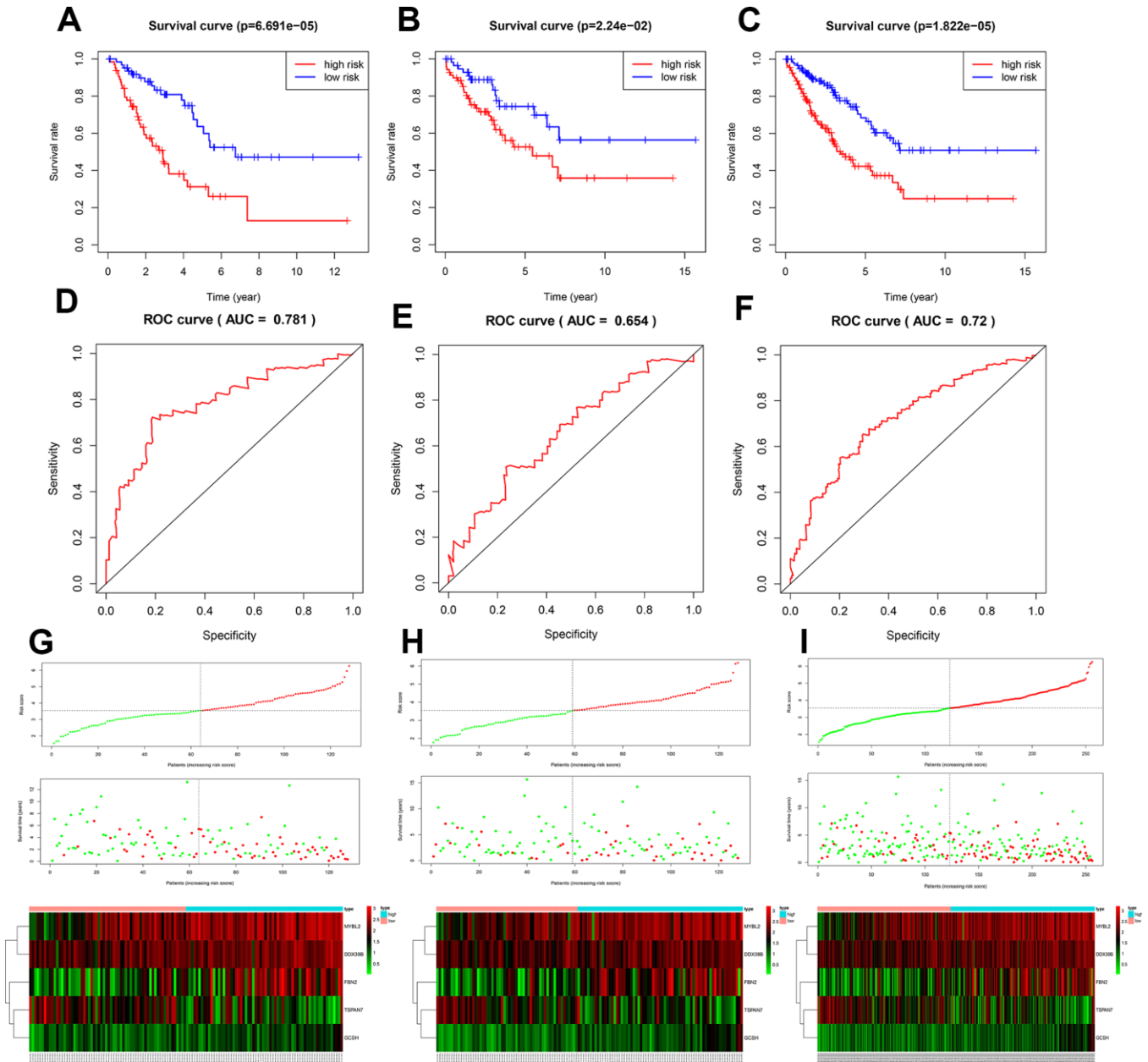


Figure 3. Assessment of the prognostic model. Survival analyses for the training (A), test (B), and overall (C) datasets. Receiver operating curves (ROC) of the prognostic model in the training (D), test (E), and overall (F) datasets. Differences in risk score, survival time, and gene expression between the high- and low-risk groups in the training (G), test (H), and overall (I) datasets.

Alterations in key genes in STS

Alterations of the key genes were explored using data obtained from the cBioPortal database. Of the 265 samples included, 73 (28%) had alterations in the key genes (Figures 4A, 4B); the network of key genes and most frequently altered neighbor genes is displayed in Figure 4C. Furthermore, patients with alterations in the key genes had shorter OS (Figure 4D).

Relationships between immune infiltration, risk scores, and gene expression

CD4+ T cell (correlation (cor)= -0.233, p=1.795e-04), CD8+ T cell (cor= -0.128, p=0.042), B cell (cor= -0.124, p=0.048), dendritic cell (cor= -0.217, p=4.923e-04), macrophage (cor= -0.250, p=5.745e-05), and neutrophil (cor= -0.218, p=4.535e-04) infiltration were negatively correlated with STS patient risk scores

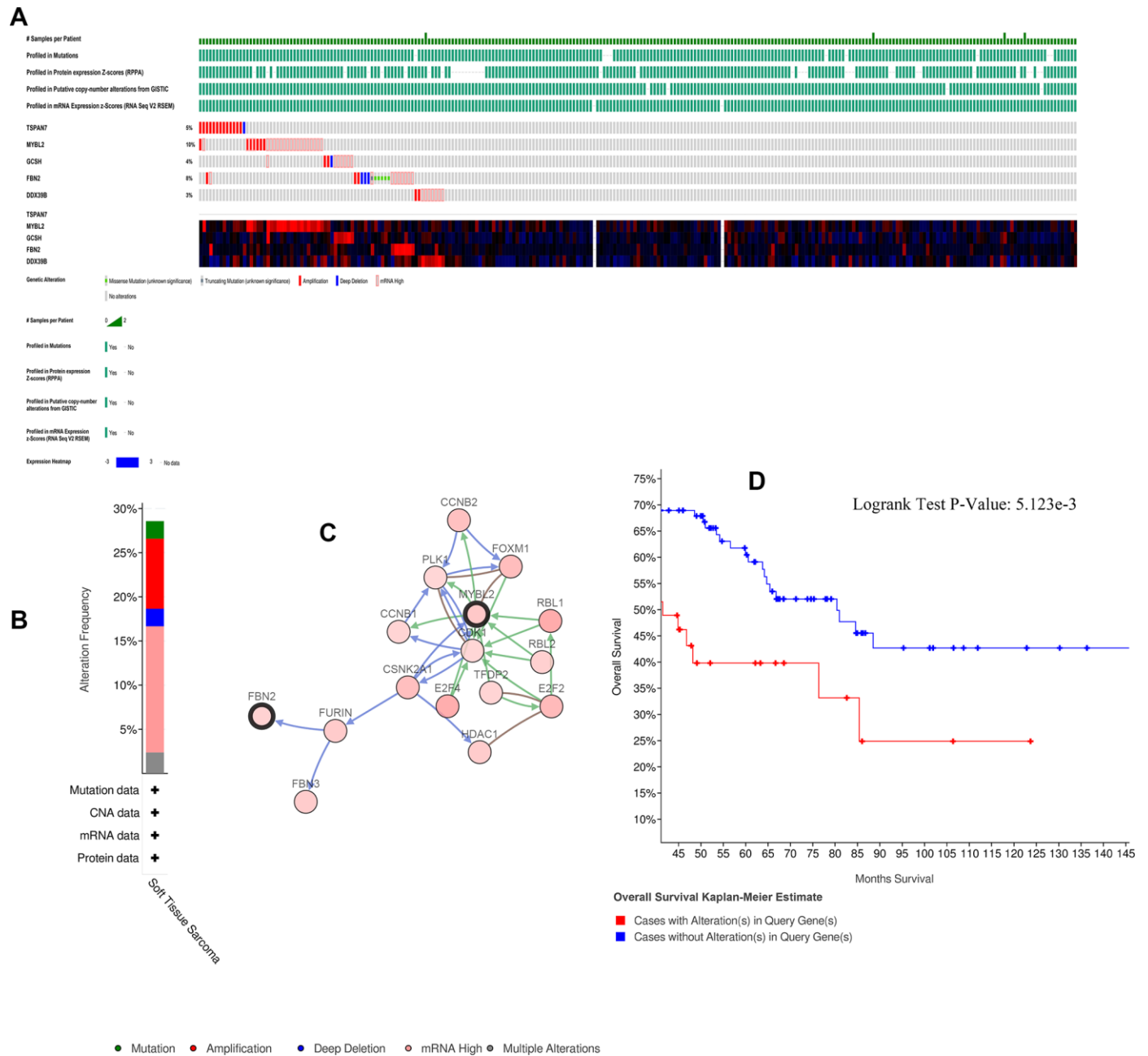


Figure 4. Alterations in expression of the five key genes. (A) 73 of 265 samples (28%) had alterations of the five key genes. **(B)** Frequencies of different alterations. **(C)** Network of key genes and most frequently altered neighbor genes. **(D)** Survival analysis for patients with and without alterations in the five key genes.

(Figure 5A). Similar negative correlations were observed between immune cells and FBN2 and DDX39B expression specifically (Figure 5D, 5E). TSPAN7 expression was negatively correlated with CD4⁺ T cell infiltration, but positively correlated with B-cell infiltration (Figure 5B). MYBL2 expression was positively correlated with CD4⁺ T cell infiltration, but negatively correlated with B-cell infiltration (Figure 5C). Moreover, low-risk patients had higher expression of immune checkpoint molecules PDCD1 and CD274 (Supplementary Figure 7). No differences in CTLA4 and LAG3 expression were observed between high- and low-risk patients (Supplementary Figure 7).

GSEA analysis for the five key genes

The top pathways for each key gene are illustrated in Supplementary Figure 8A–8E; overall, they were highly associated with cell cycle and cancer pathways. These results further validate the crucial roles these genes play in STS and tumors in general.

Identification of prognostic factors and nomogram construction

Univariate and multivariate Cox analyses indicated that age, margin status, presence of metastasis, and risk score had significant impacts on prognosis (Table 2). In

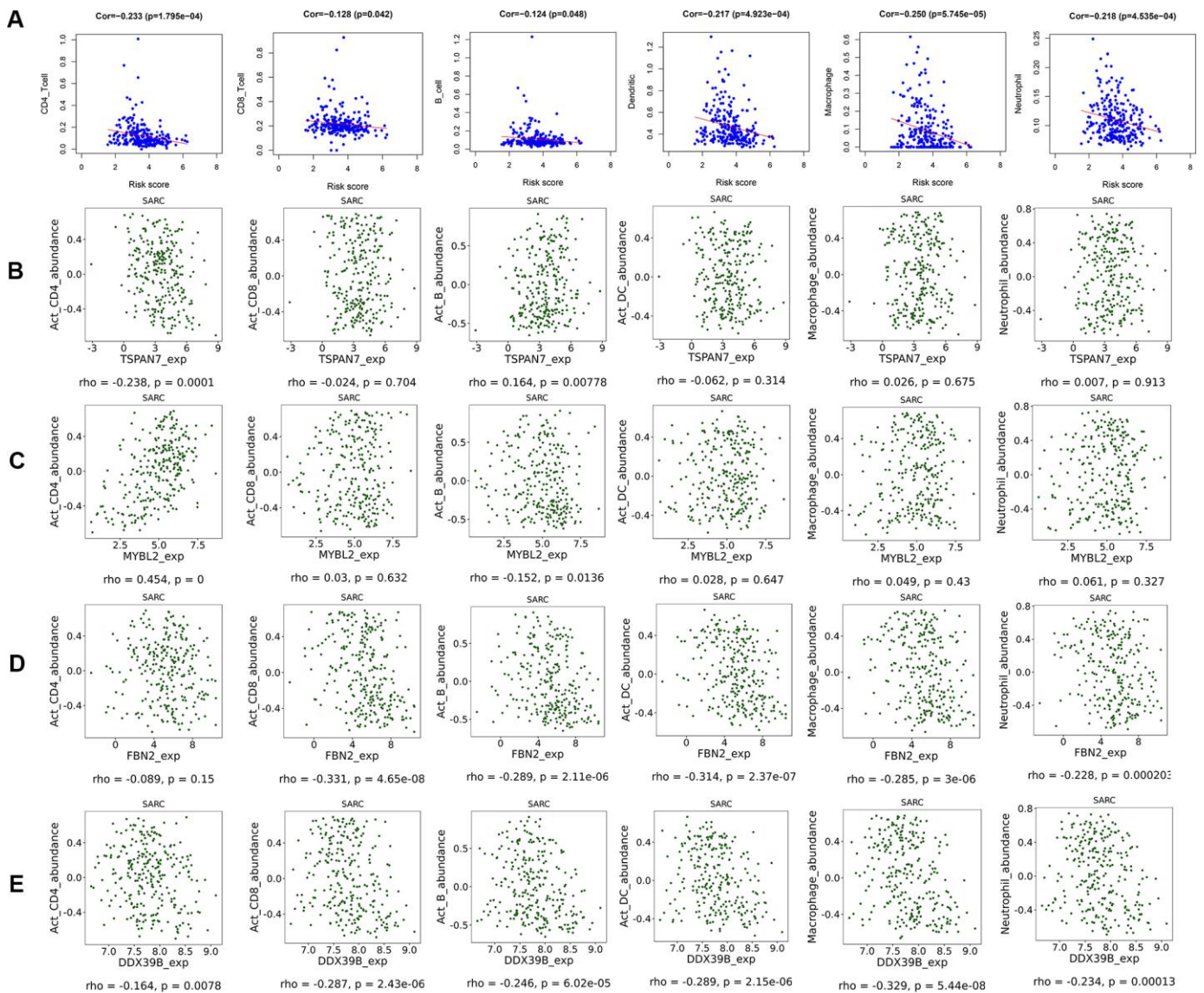


Figure 5. Scatter diagram of the relationship between immune cell infiltration, risk scores, and key gene expression. (A) Relationships between immune cell infiltration and risk scores. **(B)** Relationships between immune cell infiltration and expression of the TSPAN7 **(B)**, MYBL2 **(C)**, FBN2 **(D)**, and DDX39 **(E)** genes.

Table 2. Univariate and multivariate analysis of clinical factor and risk score.

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	P value	HR	95% CI	P value
Age	1.023	1.007 1.040	0.004	1.034	1.015 1.053	2.865E-04
Gender	0.853	0.571 1.275	0.438	NA	NA NA	NA
Race	1.257	0.591 2.671	0.553	NA	NA NA	NA
Sample weight	1.000	0.999 1.000	0.295	NA	NA NA	NA
Total necrosis percent	1.208	0.953 1.532	0.119	NA	NA NA	NA
Margin status	1.899	1.175 3.070	0.009	2.155	1.262 3.679	0.005
Metastatic diagnosis	2.939	1.788 4.831	2.123E-05	3.937	2.252 6.884	1.533E-06
Radiation therapy	0.604	0.280 1.301	0.198	NA	NA NA	NA
Histological type	0.972	0.865 1.093	0.636	NA	NA NA	NA
Risk score	1.908	1.506 2.417	8.518E-08	2.355	1.710 3.243	1.554E-07

Note: HR: hazard ratio; CI: confidence interval; NA: not available.

In addition to these four factors, sex and histological type were used to construct the nomogram for STS patients displayed in Figure 6. Internal validation showed that the predictive accuracy for STS as calculated using the C-index was 0.782. Actual survival rates and predictions obtained using the nomogram were largely concordant (Supplementary Figure 9A–9C).

Identification of small-molecule drugs

The 107 upregulated and 14 downregulated DEGs were mapped into the Connectivity Map database, which was then used to identify the top 10 small-molecule drugs most likely to be effective based on p-values < 0.05 (Table 3). Camptothecin, daunorubicin, 0175029-0000,

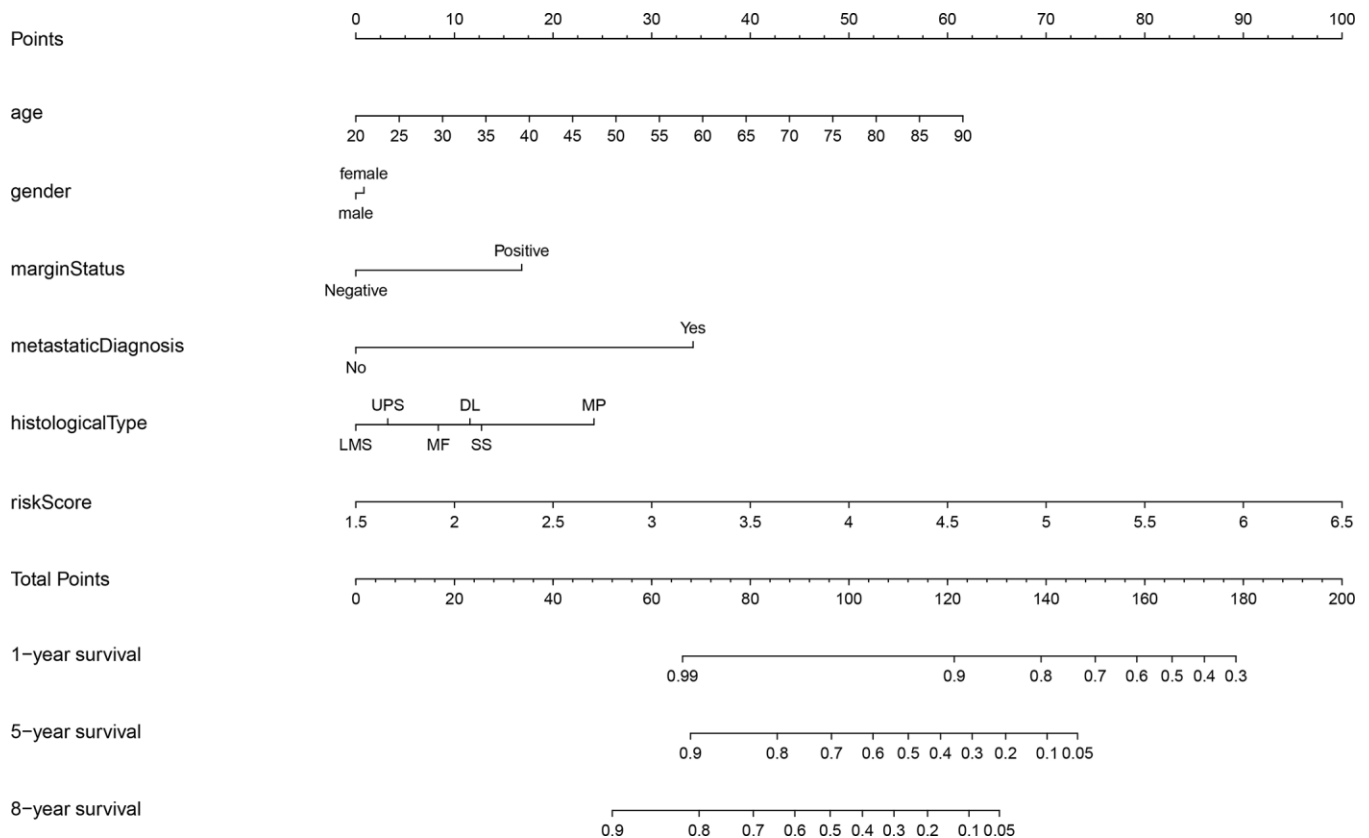


Figure 6. Nomogram for STS. STS: soft tissue sarcoma; LMS: leiomyosarcomas; UPS: undifferentiated pleomorphic sarcoma; MF: myxofibrosarcomas; DL: dedifferentiated liposarcomas; SS: synovial sarcomas; MP: malignant peripheral nerve sheath tumors.

Table 3. Potential small-molecule drugs identified using the Connectivity Map database.

Rank	Cmap name	Mean	n	Enrichment	P	Specificity	Percent non-null
1	camptothecin	-0.727	3	-0.985	0	0.0568	100
2	daunorubicin	-0.661	4	-0.952	0	0.0088	100
3	0175029-0000	-0.753	6	-0.952	0	0	100
4	resveratrol	-0.726	9	-0.836	0	0	100
5	trichostatin A	-0.419	182	-0.394	0	0.3663	84
6	GW-8510	-0.663	4	-0.927	0.00002	0.0916	100
7	colistin	0.567	4	0.873	0.00036	0	100
8	dipyridamole	-0.506	6	-0.756	0.0004	0.0123	100
9	fludrocortisone	0.269	8	0.669	0.00048	0.0177	50
10	apigenin	-0.619	4	-0.873	0.00056	0.0163	100

Note: camp: Connectivity Map.

resveratrol, and trichostatin A were the top molecules likely to act on the gene targets obtained in our comparison of tumor and normal tissues; these drugs might therefore be particularly useful for treating STS.

DISCUSSION

Because it encompasses a large and heterogenous group of malignant tumors with diverse origins, STS is difficult to diagnosis and treat effectively. In addition to conventional diagnosis and treatment, identification of novel specific biomarkers might improve outcomes for both early and advanced stage STS patients [15]. In this study, we attempted to identify potential biomarkers for risk stratification and prognosis prediction in STS patients, as well as small-molecule drugs that might aid in treatment by targeting these biomarkers.

The TCGA and the GTEx databases were utilized to identify differentially expressed genes (DEGs) that might serve as potential biomarkers for STS; univariate Cox regression analysis, Lasso regression analysis, and multivariate Cox regression analysis were then used to select key genes for inclusion in the prognostic model. The DEGs of interest were mainly enriched in mitotic nuclear division, sister chromatid segregation, chromosome segregation, and nuclear division, which are highly associated with tumorigenesis. They were also involved in cell cycle, DNA replication, and the p53 signaling pathway. The p53 signaling pathway is a well-established pathway associated with various types of cancer [16], including sarcomas [17]. A prognostic model was used to generate risk scores based on expression of the five key genes, among which MYBL2 and FBN2 were upregulated, while TSPAN7, GCSH, and DDX39B were downregulated, in STS; validation studies confirmed that this risk score was highly associated with survival. The model was effective in distinguishing between high- and low-risk STS patients

and between patients and healthy individuals, regardless of the histological type.

Changes in expression of the five key genes identified here have been reported in different types of cancer, including lung [18], renal [19], breast [20], colorectal [21], prostate [22], and multiple myeloma [23]. In an analysis of gene expression profiles from 13 primary and 15 metastatic uterine leiomyosarcoma cases, Davidson et al. [24] reported that TSPAN7 was overexpressed in primary uterine leiomyosarcoma; however, in our study, TSPAN7 was downregulated in all histological types of STS. This inconsistency might be due to the small sample size in the Davidson study. Consistent with our results, increased expression of TSPAN7, which can inhibit the development of multiple myeloma *in vivo* [23], was associated with longer survival time in clear-cell renal cell carcinoma [25]. FBN2 has been identified as a diagnostic biomarker in leiomyosarcoma and rhabdomyosarcoma [26, 27]. Additionally, aberrant methylation of FBN2 has been observed in breast cancer, non-small cell lung cancer, and esophageal squamous cell carcinoma [28–30]; FBN2 methylation might negatively impact STS prognosis as well. To our knowledge, the role of GCSH has not been examined in STS, but only in breast cancer and papillary thyroid cancer [20, 31]. MYBL2 is associated with poor prognosis in numerous cancers and plays a vital role in the regulation of cell proliferation, cell survival, and differentiation [32]. For example, MYBL2 was recently found to promote progression of Ewing sarcoma [33]. Here, overexpression of MYBL2 was associated with poor outcomes in STS patients. DDX39B, a DExD RNA helicase, is involved in pre-mRNA splicing and nuclear export of mRNAs [34]. Awasthi et al. [35] found that DDX39B could promote global translation and cell proliferation through upregulation of pre-ribosomal RNA, ultimately leading to oncogenesis. In addition, DDX39B is a crucial

contributor to Kaposi's sarcoma-associated herpesvirus intronless mRNA nuclear export and virus replication [36]. Because all histological types of STS were characterized by changes in the expression of these five key genes, they might be particularly useful as new prognostic biomarkers for STS. However, the specific roles of these genes in STS need to be examined in future studies.

In this study, we performed multilevel analyses to further explore associations between key genes in STS and immune infiltration, gene alterations, and GSEA pathways. Negative correlations between infiltration of six types of immune cells and risk scores indicated that increased immune cell infiltration contributed to better survival in STS, which is consistent with previous studies [11, 37]. The TCGA Research Network [11] reported that higher NK, T, and dendritic cell levels were associated with better outcomes. In contrast to our findings, Koirala et al. [38] found that increased dendritic cell (DC) and macrophage levels negatively impacted survival in human osteosarcoma. The absence of lymphatic vessels, and the resulting inhibition of antigen-presenting capacity, in human bone tissue might explain these detrimental effects of DCs [39]; this might also highlight important differences in immune infiltration between STS types containing lymphatic vessels and osteosarcoma. Conflicting results have been obtained regarding the association between macrophage infiltration and osteosarcoma prognosis [40], and additional studies are needed on this topic. In this study, we found that expression of two of the key genes, DDX39B and FBN2, was negatively correlated with infiltration of most immune cell types. MYBL2 expression was positively correlated with CD4+ T-cell infiltration, but negatively correlated with B-cell infiltration. TSPAN7 expression was negatively correlated with CD4+ T-cell infiltration and positively with B-cell infiltration. Finally, GCSH expression was not correlated with infiltration for any of the immune cell types examined. We also demonstrated that expression of PDCD1 and CD274 was higher in low-risk patients, suggesting that our prognostic model could potentially identify patients who would benefit from treatment with immune checkpoint inhibitors. Our use of five key genes together in a single model improved its prognostic value compared to the individual genes, based on the comparisons of risk scores with the individual gene expression in the correlations with immune cell infiltration. In addition, survival times were substantially reduced in patients with alterations in these key genes, indicating that they possess accurate prognostic power. Finally, a GSEA analysis revealed that the key genes promoted cell proliferation as well as cancer development and progression via different cell cycle, DNA replication,

mismatch repair, and cancer-associated pathways (e.g., phosphatidylinositol signaling system [41], basal cell carcinoma, transforming growth factor beta signaling pathway [42], WNT signaling pathway [43], and the p53 signaling pathway). These signaling pathways have also been reported as important regulators in osteosarcoma and STS [43, 44]. Finally, these key genes might also affect development and progression of STS through interactions with gene fusion products and miRNAs, which not only play important regulatory roles but can also act as therapeutic targets in sarcoma [45]. For example, EWSR1 fusion is common in Ewing sarcoma [46], and EWSR1-FLI1 regulates the expression of MYBL2 [33]. Additional studies of such mechanisms might also help improve diagnosis and treatment of STS.

Our prognostic model based on five key genes was able to stratify STS patients into clinically meaningful high- and low-risk groups which differed significantly in terms of survival. The accuracy of the prognostic model was validated using the test group. In addition, univariate and multivariate Cox regression analyses of our model demonstrated that age, margin status, diagnosis of metastasis, and risk score were independent prognostic factors for STS. Prognostic predictions for STS patients are currently based on the presence of metastases as well as tumor grade, size, and depth [47, 48]. While previous studies have constructed nomograms for prognostic predictions in STS [49], gene expression is not typically included. In this study, we incorporated our risk score together with multiple clinical factors (e.g., age, sex, margin status, diagnosis of metastasis, histological type) to generate a prognostic nomogram with high predictive accuracy for STS patients.

Finally, we identified small-molecule drugs that might improve STS treatment by targeting the key genes using the Camp database. One such drug, Camptothecin, is a topoisomerase inhibitor [50] that inhibited cell proliferation and showed anticancer activity in colon, lung, ovarian, breast, liver, pancreas, and stomach cancer [51] and in Ewing sarcoma [52]. Furthermore, combinations of immune checkpoint inhibitors and the small-molecule drugs might also help improve treatments for STS [53].

This integrated analysis of multiple databases enhanced the robustness and reproducibility of our conclusions; the inclusion of information from the GTEx database was especially helpful in compensating for the paucity of control samples in the TCGA database. In addition, our inclusion of both gene expression and clinical data might render our results especially applicable to STS patients in the clinical setting. However, analysis of additional clinical data is necessary to confirm these results.

CONCLUSION

This study demonstrated that five key genes (i.e., MYBL2, FBN2, DDX39B, TSPAN7, and GCSH) highly associated with prognosis, and immune infiltration, could promote STS via different signaling pathways. The prognostic model based on these five key genes demonstrated excellent performance in terms of risk stratification of patients and prediction of survival. Furthermore, the nomogram integrating multiple genes and clinical factors could provide specific predictions for the survival of individuals with STS. Based on the combination of gene and clinical data, this study may contribute to the management of STS.

MATERIALS AND METHODS

Data collection and preprocessing

Gene expression profile and clinical data from the TCGA and GTEx databases were downloaded from the University Of California Santa Cruz (UCSC) Xena Hub datasets (<https://xenabrowser.net/>). Gene expression data were downloaded from TCGA Sarcoma, which were obtained using the Illumina HiSeq 2000 RNA Sequencing platform by the University of North Carolina (Chapel Hill, NC, USA) TCGA genome characterization center. Data were obtained for 265 samples (263 samples with STS and two matched controls) and were $\log_2(x+1)$ transformed. We also downloaded gene expression profiles for muscle and fat tissues, which were the two most common histological types among TCGA samples, from the GTEx database for use as additional matched controls. GTEx gene expression data were also $\log_2(x+1)$ transformed to allow comparisons to TCGA data. Transformation of Ensembl identifiers and normalization of expression between the TCGA and GTEx datasets were also performed prior to subsequent differential analysis. GSE21122 gene expression profiles [54] for 149 STS samples and nine normal fat tissue samples were downloaded from the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>) and used as the validation dataset. If a database did not include an STS category, sarcoma sample data was used as STS data; most databases include only STS and not bone tumors in the sarcoma category.

Identification of differentially expressed genes (DEGs)

The DEG analyses were performed using the “limma” package [55] in R statistical software, version 3.53. DEGs were divided among three groups: “tumors” vs. “muscle tissues” + “matched controls” from TCGA Sarcomas (hereinafter “muscle controls”), “tumors” vs. “fat tissues” + “matched controls” (hereinafter “fat

controls”), and “tumors” vs. “matched controls”. DEGs were defined by $|\log_2 FC| > 2$ and $p < 0.05$ in the first two groups. Because the third group included only two matched controls, DEGs in that group were defined by $|\log_2 FC| > 1$ and $p < 0.05$ to increase the number of identified genes. Only DEGs that were identified in all three groups were included in subsequent analyses.

Functional annotation of DEGs

Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were performed in R using the “clusterProfiler” package [56]. Adjusted $p < 0.05$ was considered statistically significant in the GO and KEGG analyses. Bar plots, dot plots, and chord plots were constructed to visualize functional annotation results via the “GPlot” package [57] in R.

Identification of survival-related DEGs and establishment of a prognostic model

All available samples were randomly assigned to the training and test datasets. The training group was used to establish the prognostic model, which was then validated using the test group. First, the primary DEGs identified above were used in a univariate Cox regression analysis in the training dataset using the “survival” package in R. DEGs with $p < 0.05$ in the univariate Cox regression analysis were then selected for least absolute shrinkage and selection operator (Lasso) regression analysis using the “glmnet” and “survival” packages [58]; Lasso regression analysis, which detects the optimal lambda value based on 1,000 cross-validations, is particularly useful in high-dimension datasets [59]. The DEGs of interest identified in the Lasso regression analysis were then subjected to multivariate Cox regression analysis. The prognostic model was constructed based on the following equation:

risk score = $\sum_{i=1}^n \beta_i \times \exp(G_i)$; where n is the number of genes identified for the multivariate Cox regression model; $\exp(G_i)$ is the expression value of gene i ; and β_i refers to the coefficient for gene i . The DEGs included in the prognostic model were regarded as key genes associated with survival. Patients were divided into “high-risk” and “low-risk” groups based on the median risk score obtained using this prognostic model. To confirm the results, the prognostic model was also used to analyze the test datasets. Finally, analyses of overall survival (OS) were performed to evaluate differences in OS between high- and low-risk patients in the training and test datasets. The receiver operating characteristic (ROC) curve was plotted using the “survivalROC” package to determine the specificity and sensitivity of the risk model. We also analyzed OS and plotted ROC curves when all samples included in the study were combined in a single dataset.

Expression of the key genes in STS

Differences in expression of the key genes between the STS and normal groups and between the high-risk, low-risk, and normal groups were compared using the “ggstatsplot” package (<http://doi.org/10.5281/zenodo.2074621>); ROC curves for the five key genes were plotted for each group. Data from the TCGA and GTEx databases and the GSE21122 dataset were then used to investigate differences in expression of the key genes between various histological types and matched controls.

Alterations in the key genes in STS

The cBio Cancer Genomics Portal database (cBioPortal database), (<http://cbioportal.org>) containing multi-dimensional cancer genomics data sets from >5,000 tumor samples enables multilevel analysis for a diverse set of tumors [60]. This database was used to investigate associations between alterations in the key genes and survival in STS patients.

Immune infiltration analysis

Immune infiltration analysis was performed using the Tumor Immune Estimation Resource (TIMER) database, which includes molecular characterizations of six tumor-infiltrating immune subsets in 32 types of cancer (<https://cistrome.shinyapps.io/timer/>) [61, 62]. A matrix of six immune cell types (CD4+ T cells, CD8+ T cells, B cells, dendritic cells, macrophages and neutrophils) from the TIMER database estimation module were downloaded and used to explore the relationship between immune cells and risk scores. We also investigated associations between the five key genes and immune cells in TISIDB, an integrated repository portal for tumor-immune system interactions [63], to understand the impact of gene expression on the immune infiltration. Accumulating evidence has demonstrated that patients’ responses to immune checkpoint inhibitors are associated with the expression level of immune checkpoint molecules and immune cell infiltration [38, 64, 65]. Differences in the common immune checkpoint molecules PD-L1(CD274), PD1(PDCD1), CTLA4, and LAG3 between high-risk and low-risk patients were examined to determine whether the five key genes might be associated with immune checkpoint inhibitor efficacy. A $p < 0.05$ denoted statistical significance.

Gene Set Enrichment Analysis (GSEA) of the five key genes

Gene expression profiles for the 263 STS patients from the TCGA database were utilized for GSEA using

GSEA 3.0 software [66]. Patients were divided into the high- and low-expression groups based on the median of key gene expression. KEGG pathways for the key genes were determined based on the $p < 0.05$ or a false discovery rate < 0.05 , and the top terms were visualized using R.

Identification of the prognosis-related clinical factors

Univariate and multivariate Cox regression analyses were performed to explore the impact of clinical factors on the prognostic model. Because clinical data were not available for all samples, the training and test datasets were combined and analyzed as a single dataset in subsequent analyses. The following factors were included in the univariate Cox regression analysis: age, sex, race, sample weight, percent of total necrosis, margin status, diagnosis of metastasis, radiation therapy, histological type, and risk score. Factors identified as significant in that analysis were used in multivariate Cox regression analyses to identify independent prognostic factors. These analyses were conducted using the “survival” package.

Construction of the nomogram and internal validation

We used R to plot the nomogram for OS in STS patients. The nomogram incorporated sex, histological type, and the independent prognostic factors identified in multivariate Cox regression analysis. Bootstrap resampling was used to internally validate the nomogram based on concordance index (C-index) values, which were obtained through 1,000 resampling events, and calibration curves, which were plotted to evaluate concordance between actual and predicted survival rates [67].

Identification of small-molecule drugs

The Connectivity Map database contains expression data from cultured human cells treated with bioactive small molecules. These data can be used to explore associations between small-molecule drugs and genes of interest identified in patients, to identify potential mechanisms of action for these drugs, and to promote novel drug design [68]. Small-molecule drugs that might be useful for the treatment of STS were identified by mapping the primary upregulated and downregulated DEGs into the database.

Abbreviations

Soft tissue sarcoma: STS; The Cancer Genome Atlas: TCGA; Genotype-Tissue Expression: GTEx; University

of California Santa Cruz: UCSC; Differentially expressed genes: DEGs; Gene Ontology: GO; Kyoto Encyclopedia of Genes and Genomes: KEGG; Least absolute shrinkage and selection operator: Lasso; Overall survival: OS; Receiver operating characteristic: ROC; Gene Set Enrichment Analysis: GSEA; Concordance index: C-index.

AUTHOR CONTRIBUTIONS

HG, RW, CZ, and WJ conceived and designed the study. RW, HG, and JG performed the analysis. HG, RW, MY, HZ, and LG interpreted the results. HG, YL, CZ, and RW provided analysis tools. RW, HG, and CZ helped write the manuscript. All authors reviewed the manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. von Mehren M, Randall RL, Benjamin RS, Boles S, Bui MM, Ganjoo KN, George S, Gonzalez RJ, Heslin MJ, Kane JM 3rd, Keedy V, Kim E, Koon H, et al. Soft Tissue Sarcoma, Version 2.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2018; 16:536–63. <https://doi.org/10.6004/jccn.2018.0025> PMID:29752328
2. Popovich JR, Cassaro S. *Cancer, Sarcoma*. StatPearls. Treasure Island (FL): StatPearls Publishing StatPearls Publishing LLC; 2019.
3. Singhi EK, Moore DC, Muslimani A. Metastatic Soft Tissue Sarcomas: A Review Of Treatment and New Pharmacotherapies. *P T*. 2018; 43:410–29. PMID:30013298
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019; 69:7–34. <https://doi.org/10.3322/caac.21551> PMID:30620402
5. Luis AM, Aguilar DP, Martín JA. Multidisciplinary management of soft tissue sarcomas. *Clin Transl Oncol*. 2010; 12:543–53. <https://doi.org/10.1007/s12094-010-0552-2> PMID:20709652
6. Luis AM. Multidisciplinary management of soft tissue sarcomas: it's time to make a team! *Clin Transl Oncol*. 2010; 12:515–17. <https://doi.org/10.1007/s12094-010-0547-z> PMID:20709647
7. Ferrari A, Dirksen U, Bielack S. Sarcomas of Soft Tissue and Bone. *Prog Tumor Res*. 2016; 43:128–41. <https://doi.org/10.1159/000447083> PMID:27595362
8. Lohman RF, Nabawi AS, Reece GP, Pollock RE, Evans GR. Soft tissue sarcoma of the upper extremity: a 5-year experience at two institutions emphasizing the role of soft tissue flap reconstruction. *Cancer*. 2002; 94:2256–64. <https://doi.org/10.1002/cncr.10419> PMID:12001125
9. ESMO/European Sarcoma Network Working Group. Soft tissue and visceral sarcomas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2014 (Suppl 3); 25:iii102–12. <https://doi.org/10.1093/annonc/mdu254> PMID:25210080
10. Biau DJ, Ferguson PC, Chung P, Griffin AM, Catton CN, O'Sullivan B, Wunder JS. Local recurrence of localized soft tissue sarcoma: a new look at old predictors. *Cancer*. 2012; 118:5867–77. <https://doi.org/10.1002/cncr.27639> PMID:22648518
11. Cancer Genome Atlas Research Network. Electronic address: elizabeth.demicco@sinaihealthsystem.ca; Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017; 171:950–965.e28. <https://doi.org/10.1016/j.cell.2017.10.014> PMID:29100075
12. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015; 19:A68–77. <https://doi.org/10.5114/wo.2014.47136> PMID:25691825
13. Ardlie KG, Deluca DS, Segre AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M, Ward LD, Kheradpour P, Iriarte B, et al, and GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015; 348:648–60. <https://doi.org/10.1126/science.1262110> PMID:25954001
14. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segrè AV, Djebali S, et al, and GTEx Consortium. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015; 348:660–65. <https://doi.org/10.1126/science.aaa0355> PMID:25954002
15. Stefano S, Giovanni S. The PTEN Tumor Suppressor Gene in Soft Tissue Sarcoma. *Cancers (Basel)*. 2019; 11:E1169.

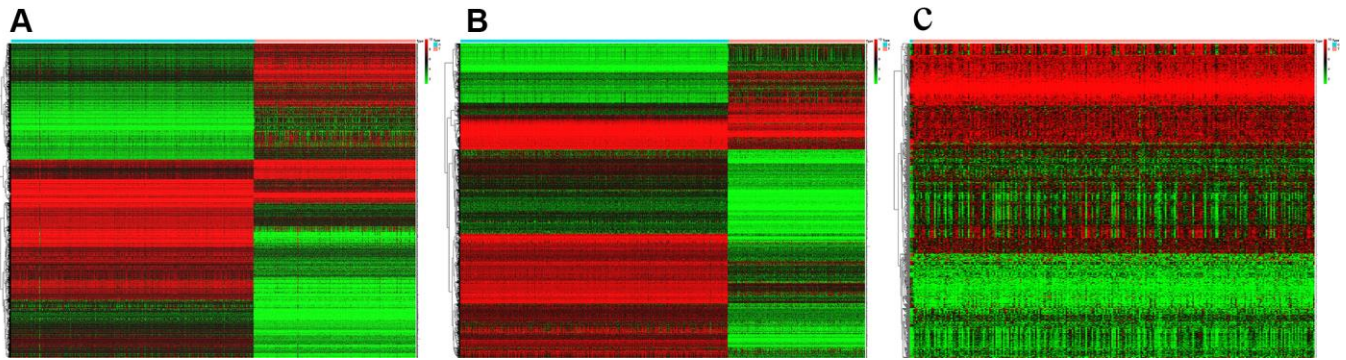
- <https://doi.org/10.3390/cancers11081169>
PMID:[31416195](https://pubmed.ncbi.nlm.nih.gov/31416195/)
16. Khoo KH, Verma CS, Lane DP. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nat Rev Drug Discov.* 2014; 13:217–36.
<https://doi.org/10.1038/nrd4236> PMID:[24577402](https://pubmed.ncbi.nlm.nih.gov/24577402/)
 17. Mathews JC, Pouryahya M, Moosmüller C, Kevrekidis YG, Deasy JO, Tannenbaum A. Molecular phenotyping using networks, diffusion, and topology: soft tissue sarcoma. *Sci Rep.* 2019; 9:13982.
<https://doi.org/10.1038/s41598-019-50300-2>
PMID:[31562358](https://pubmed.ncbi.nlm.nih.gov/31562358/)
 18. Wang X, Lin M, Zhao J, Zhu S, Xu M, Zhou X. TSPAN7 promotes the migration and proliferation of lung cancer cells via epithelial-to-mesenchymal transition. *Onco Targets Ther.* 2018; 11:8815–22.
<https://doi.org/10.2147/OTT.S167902> PMID:[30588007](https://pubmed.ncbi.nlm.nih.gov/30588007/)
 19. Gulati S, Martinez P, Joshi T, Birkbak NJ, Santos CR, Rowan AJ, Pickering L, Gore M, Larkin J, Szallasi Z, Bates PA, Swanton C, Gerlinger M. Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers. *Eur Urol.* 2014; 66:936–48.
<https://doi.org/10.1016/j.eururo.2014.06.053>
PMID:[25047176](https://pubmed.ncbi.nlm.nih.gov/25047176/)
 20. Adamus A, Müller P, Nissen B, Kasten A, Timm S, Bauwe H, Seitz G, Engel N. GCSH antisense regulation determines breast cancer cells' viability. *Sci Rep.* 2018; 8:15399. <https://doi.org/10.1038/s41598-018-33677-4>
PMID:[30337557](https://pubmed.ncbi.nlm.nih.gov/30337557/)
 21. Hibi K, Mizukami H, Saito M, Kigawa G, Nemoto H, Sanada Y. FBN2 methylation is detected in the serum of colorectal cancer patients with hepatic metastasis. *Anticancer Res.* 2012; 32:4371–74. PMID:[23060561](https://pubmed.ncbi.nlm.nih.gov/23060561/)
 22. Nakata D, Nakao S, Nakayama K, Araki S, Nakayama Y, Aparicio S, Hara T, Nakanishi A. The RNA helicase DDX39B and its paralog DDX39A regulate androgen receptor splice variant AR-V7 generation. *Biochem Biophys Res Commun.* 2017; 483:271–76.
<https://doi.org/10.1016/j.bbrc.2016.12.153>
PMID:[28025139](https://pubmed.ncbi.nlm.nih.gov/28025139/)
 23. Cheong CM, Chow AW, Fitter S, Hewett DR, Martin SK, Williams SA, To LB, Zannettino AC, Vandyke K. Tetraspanin 7 (TSPAN7) expression is upregulated in multiple myeloma patients and inhibits myeloma tumour development in vivo. *Exp Cell Res.* 2015; 332:24–38.
<https://doi.org/10.1016/j.yexcr.2015.01.006>
PMID:[25637218](https://pubmed.ncbi.nlm.nih.gov/25637218/)
 24. Davidson B, Abeler VM, Førsund M, Holth A, Yang Y, Kobayashi Y, Chen L, Kristensen GB, Shih IM, Wang TL. Gene expression signatures of primary and metastatic uterine leiomyosarcoma. *Hum Pathol.* 2014; 45:691–700.
<https://doi.org/10.1016/j.humpath.2013.11.003>
PMID:[24485798](https://pubmed.ncbi.nlm.nih.gov/24485798/)
 25. Wuttig D, Zastrow S, Füssel S, Toma MI, Meinhardt M, Kalman K, Junker K, Sanjmyatav J, Boll K, Hackermüller J, Rolle A, Grimm MO, Wirth MP. CD31, EDNRB and TSPAN7 are promising prognostic markers in clear-cell renal cell carcinoma revealed by genome-wide expression analyses of primary tumors and metastases. *Int J Cancer.* 2012; 131:E693–704.
<https://doi.org/10.1002/ijc.27419>
PMID:[22213152](https://pubmed.ncbi.nlm.nih.gov/22213152/)
 26. Grass B, Wachtel M, Behnke S, Leuschner I, Niggli FK, Schäfer BW. Immunohistochemical detection of EGFR, fibrillin-2, P-cadherin and AP2beta as biomarkers for rhabdomyosarcoma diagnostics. *Histopathology.* 2009; 54:873–79.
<https://doi.org/10.1111/j.1365-2559.2009.03303.x>
PMID:[19469909](https://pubmed.ncbi.nlm.nih.gov/19469909/)
 27. Wachtel M, Runge T, Leuschner I, Stegmaier S, Koscielniak E, Treuner J, Odermatt B, Behnke S, Niggli FK, Schäfer BW. Subtype and prognostic classification of rhabdomyosarcoma by immunohistochemistry. *J Clin Oncol.* 2006; 24:816–22.
<https://doi.org/10.1200/JCO.2005.03.4934>
PMID:[16391296](https://pubmed.ncbi.nlm.nih.gov/16391296/)
 28. Chen H, Suzuki M, Nakamura Y, Ohira M, Ando S, Iida T, Nakajima T, Nakagawara A, Kimura H. Aberrant methylation of FBN2 in human non-small cell lung cancer. *Lung Cancer.* 2005; 50:43–49.
<https://doi.org/10.1016/j.lungcan.2005.04.013>
PMID:[15951052](https://pubmed.ncbi.nlm.nih.gov/15951052/)
 29. Tsunoda S, Smith E, De Young NJ, Wang X, Tian ZQ, Liu JF, Jamieson GG, Drew PA. Methylation of CLDN6, FBN2, RBP1, RBP4, TFPI2, and TMEFF2 in esophageal squamous cell carcinoma. *Oncol Rep.* 2009; 21:1067–73.
<https://doi.org/10.3892/or.00000325> PMID:[19288010](https://pubmed.ncbi.nlm.nih.gov/19288010/)
 30. Kikuyama M, Takeshima H, Kinoshita T, Okochi-Takada E, Wakabayashi M, Akashi-Tanaka S, Ogawa T, Seto Y, Ushijima T. Development of a novel approach, the epigenome-based outlier approach, to identify tumor-suppressor genes silenced by aberrant DNA methylation. *Cancer Lett.* 2012; 322:204–12.
<https://doi.org/10.1016/j.canlet.2012.03.016>
PMID:[22433712](https://pubmed.ncbi.nlm.nih.gov/22433712/)
 31. Zhai T, Muhanhali D, Jia X, Wu Z, Cai Z, Ling Y. Identification of gene co-expression modules and hub genes associated with lymph node metastasis of papillary thyroid cancer. *Endocrine.* 2019; 66:573–84.
<https://doi.org/10.1007/s12020-019-02021-9>
PMID:[31332712](https://pubmed.ncbi.nlm.nih.gov/31332712/)

32. Musa J, Aynaud MM, Mirabeau O, Delattre O, Grünewald TG. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis.* 2017; 8:e2895. <https://doi.org/10.1038/cddis.2017.244> PMID:28640249
33. Musa J, Cidre-Aranaz F, Aynaud MM, Orth MF, Knott MM, Mirabeau O, Mazor G, Varon M, Hölting TL, Grossetête S, Gartlgruber M, Surdez D, Gerke JS, et al. Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes. *Nat Commun.* 2019; 10:4128. <https://doi.org/10.1038/s41467-019-12071-2> PMID:31511524
34. Shen H. UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. *BMB Rep.* 2009; 42:185–88. <https://doi.org/10.5483/BMBRep.2009.42.4.185> PMID:19403039
35. Awasthi S, Chakrapani B, Mahesh A, Chavali PL, Chavali S, Dhayalan A. DDX39B promotes translation through regulation of pre-ribosomal RNA levels. *RNA Biol.* 2018; 15:1157–66. <https://doi.org/10.1080/15476286.2018.1517011> PMID:30176153
36. Boyne JR, Colgan KJ, Whitehouse A. Recruitment of the complete hTREX complex is required for Kaposi's sarcoma-associated herpesvirus intronless mRNA nuclear export and virus replication. *PLoS Pathog.* 2008; 4:e1000194. <https://doi.org/10.1371/journal.ppat.1000194> PMID:18974867
37. Budczies J, Mechtersheimer G, Denkert C, Klauschen F, Mughal SS, Chudasama P, Bockmayr M, Jöhrens K, Endris V, Lier A, Lasitschka F, Penzel R, Dietel M, et al. PD-L1 (CD274) copy number gain, expression, and immune cell infiltration as candidate predictors for response to immune checkpoint inhibitors in soft-tissue sarcoma. *Oncoimmunology.* 2017; 6:e1279777. <https://doi.org/10.1080/2162402X.2017.1279777> PMID:28405504
38. Koirala P, Roth ME, Gill J, Piperdi S, Chinai JM, Geller DS, Hoang BH, Park A, Fremed MA, Zang X, Gorlick R. Immune infiltration and PD-L1 expression in the tumor microenvironment are prognostic in osteosarcoma. *Sci Rep.* 2016; 6:30093. <https://doi.org/10.1038/srep30093> PMID:27456063
39. Edwards JR, Williams K, Kindblom LG, Meis-Kindblom JM, Hogendoorn PC, Hughes D, Forsyth RG, Jackson D, Athanasou NA. Lymphatics and bone. *Hum Pathol.* 2008; 39:49–55. <https://doi.org/10.1016/j.humpath.2007.04.022> PMID:17904616
40. Ratti C, Botti L, Cancila V, Galvan S, Torselli I, Garofalo C, Manara MC, Bongiovanni L, Valenti CF, Burocchi A, Parenza M, Cappetti B, Sangaletti S, et al. Trabectedin Overrides Osteosarcoma Differentiative Block and Reprograms the Tumor Immune Environment Enabling Effective Combination with Immune Checkpoint Inhibitors. *Clin Cancer Res.* 2017; 23:5149–61. <https://doi.org/10.1158/1078-0432.CCR-16-3186> PMID:28600479
41. Dey N, Leyland-Jones B, De P. MYC-xing it up with PIK3CA mutation and resistance to PI3K inhibitors: summit of two giants in breast cancers. *Am J Cancer Res.* 2014; 5:1–19. PMID:25628917
42. Xie F, Ling L, van Dam H, Zhou F, Zhang L. TGF- β signaling in cancer metastasis. *Acta Biochim Biophys Sin (Shanghai).* 2018; 50:121–32. <https://doi.org/10.1093/abbs/gmx123> PMID:29190313
43. Pridgeon MG, Grohar PJ, Steensma MR, Williams BO. Wnt Signaling in Ewing Sarcoma, Osteosarcoma, and Malignant Peripheral Nerve Sheath Tumors. *Curr Osteoporos Rep.* 2017; 15:239–46. <https://doi.org/10.1007/s11914-017-0377-9> PMID:28647886
44. Chen Y, Zhang K, Li Y, He Q. Estrogen-related receptor α participates transforming growth factor- β (TGF- β) induced epithelial-mesenchymal transition of osteosarcoma cells. *Cell Adh Migr.* 2017; 11:338–46. <https://doi.org/10.1080/19336918.2016.1221567> PMID:27532429
45. Lim HJ, Yang JL. Regulatory roles and therapeutic potential of microRNA in sarcoma. *Crit Rev Oncol Hematol.* 2016; 97:118–30. <https://doi.org/10.1016/j.critrevonc.2015.08.001> PMID:26318097
46. Anderson ND, de Borja R, Young MD, Fuligni F, Rosic A, Roberts ND, Hajjar S, Layeghifard M, Novokmet A, Kowalski PE, Anaka M, Davidson S, Zarrei M, et al. Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science.* 2018; 361:eaam8419. <https://doi.org/10.1126/science.aam8419> PMID:30166462
47. Stojadinovic A, Leung DH, Hoos A, Jaques DP, Lewis JJ, Brennan MF. Analysis of the prognostic significance of microscopic margins in 2,084 localized primary adult soft tissue sarcomas. *Ann Surg.* 2002; 235:424–34. <https://doi.org/10.1097/0000658-200203000-00015> PMID:11882765
48. Casali PG, Abecassis N, Aro HT, Bauer S, Biagini R, Bielack S, Bonvalot S, Boukovinas I, Bovee JV, Brodowicz T, Broto JM, Buonadonna A, De Álava E, et al, and ESMO Guidelines Committee and EURACAN.

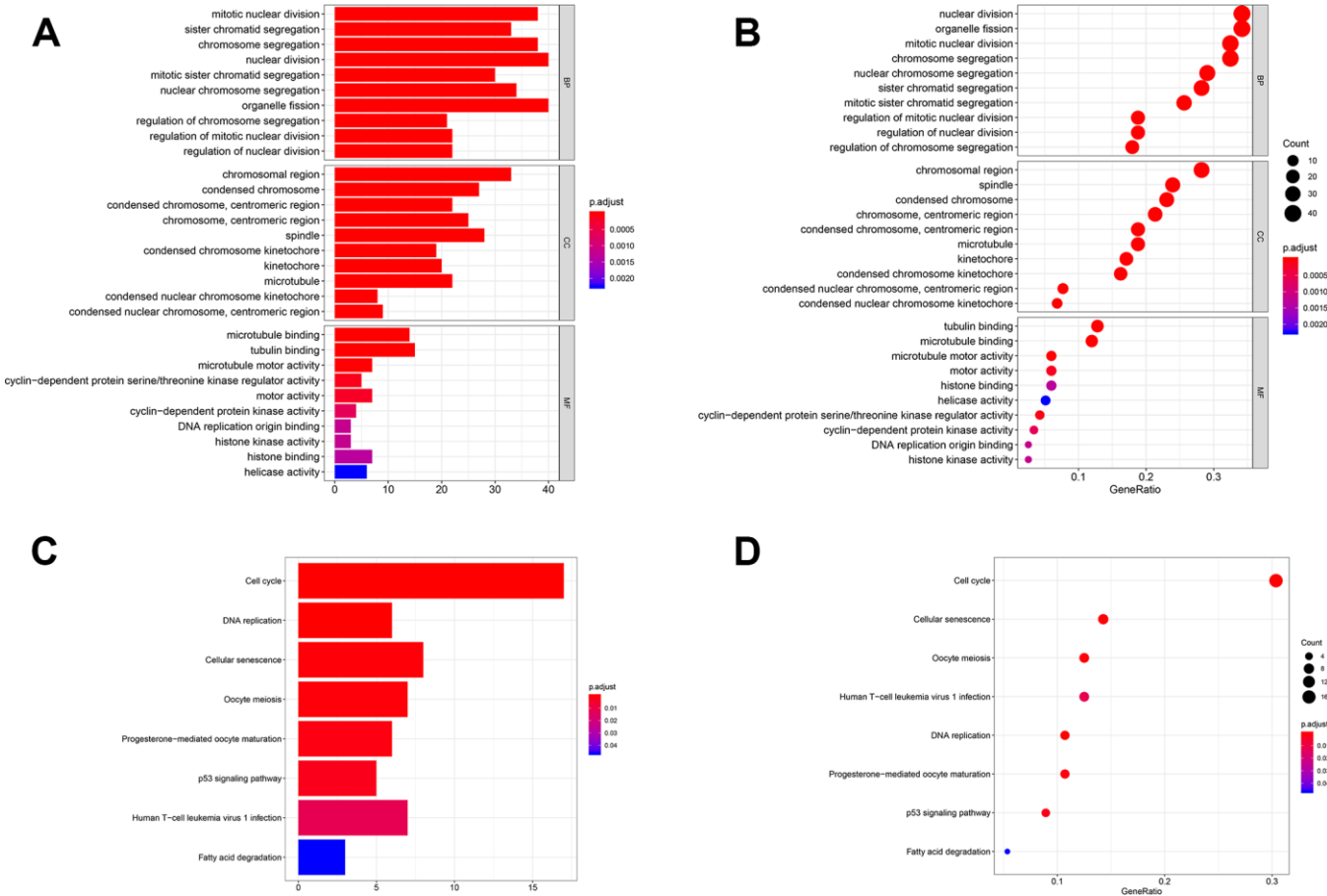
- Soft tissue and visceral sarcomas: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2018 (Suppl 4); 29:iv51–67. <https://doi.org/10.1093/annonc/mdy096> PMID:29846498
49. Callegaro D, Miceli R, Mariani L, Raut CP, Gronchi A. Soft tissue sarcoma nomograms and their incorporation into practice. *Cancer*. 2017; 123:2802–20. <https://doi.org/10.1002/cncr.30721> PMID:28493287
50. Veloso A, Biewen B, Paulsen MT, Berg N, Carmo de Andrade Lima L, Prasad J, Bedi K, Magnuson B, Wilson TE, Ljungman M. Genome-wide transcriptional effects of the anti-cancer agent camptothecin. *PLoS One*. 2013; 8:e78190. <https://doi.org/10.1371/journal.pone.0078190> PMID:24194914
51. Gokduman K, Strategies Targeting DN. Strategies Targeting DNA Topoisomerase I in Cancer Chemotherapy: Camptothecins, Nanocarriers for Camptothecins, Organic Non-Camptothecin Compounds and Metal Complexes. *Curr Drug Targets*. 2016; 17:1928–39. <https://doi.org/10.2174/1389450117666160502151707> PMID:27138759
52. Palmerini E, Jones RL, Setola E, Picci P, Marchesi E, Luksch R, Grignani G, Cesari M, Longhi A, Abate ME, Paioli A, Szucs Z, D'ambrosio L, et al. Irinotecan and temozolomide in recurrent Ewing sarcoma: an analysis in 51 adult and pediatric patients. *Acta Oncol*. 2018; 57:958–64. <https://doi.org/10.1080/0284186X.2018.1449250> PMID:29533113
53. Pappo AS, Dirksen U. Rhabdomyosarcoma, Ewing Sarcoma, and Other Round Cell Sarcomas. *J Clin Oncol*. 2018; 36:168–79. <https://doi.org/10.1200/JCO.2017.74.7402> PMID:29220292
54. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, Shah K, Succi ND, Weir BA, Ho A, Chiang DY, Reva B, Mermel CH, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat Genet*. 2010; 42:715–21. <https://doi.org/10.1038/ng.619> PMID:20601955
55. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47. <https://doi.org/10.1093/nar/gkv007> PMID:25605792
56. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–87. <https://doi.org/10.1089/omi.2011.0118> PMID:22455463
57. Walter W, Sánchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015; 31:2912–14. <https://doi.org/10.1093/bioinformatics/btv300> PMID:25964631
58. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010; 33:1–22. <https://doi.org/10.18637/jss.v033.i01> PMID:20808728
59. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J*. 2010; 52:70–84. <https://doi.org/10.1002/bimj.200900028> PMID:19937997
60. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2:401–04. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID:22588877
61. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res*. 2017; 77:e108–10. <https://doi.org/10.1158/0008-5472.CAN-17-0307> PMID:29092952
62. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS, Liu XS. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016; 17:174. <https://doi.org/10.1186/s13059-016-1028-7> PMID:27549193
63. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SS, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW, Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics*. 2019; 35:4200–02. <https://doi.org/10.1093/bioinformatics/btz210> PMID:30903160
64. Herbst RS, Soria JC, Kowanzet M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN, Kohrt HE, Horn L, Lawrence DP, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*. 2014; 515:563–67. <https://doi.org/10.1038/nature14011> PMID:25428504
65. Wimberly H, Brown JR, Schalper K, Haack H, Silver MR, Nixon C, Bossuyt V, Pusztai L, Lannin DR, Rimm DL. PD-

- L1 Expression Correlates with Tumor-Infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy in Breast Cancer. *Cancer Immunol Res.* 2015; 3:326–32.
<https://doi.org/10.1158/2326-6066.CIR-14-0133>
PMID:[25527356](https://pubmed.ncbi.nlm.nih.gov/25527356/)
66. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–50.
<https://doi.org/10.1073/pnas.0506580102>
PMID:[16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
67. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* 1982; 247:2543–46.
<https://doi.org/10.1001/jama.1982.03320430047030>
PMID:[7069920](https://pubmed.ncbi.nlm.nih.gov/7069920/)
68. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science.* 2006; 313:1929–35.
<https://doi.org/10.1126/science.1132939>
PMID:[17008526](https://pubmed.ncbi.nlm.nih.gov/17008526/)

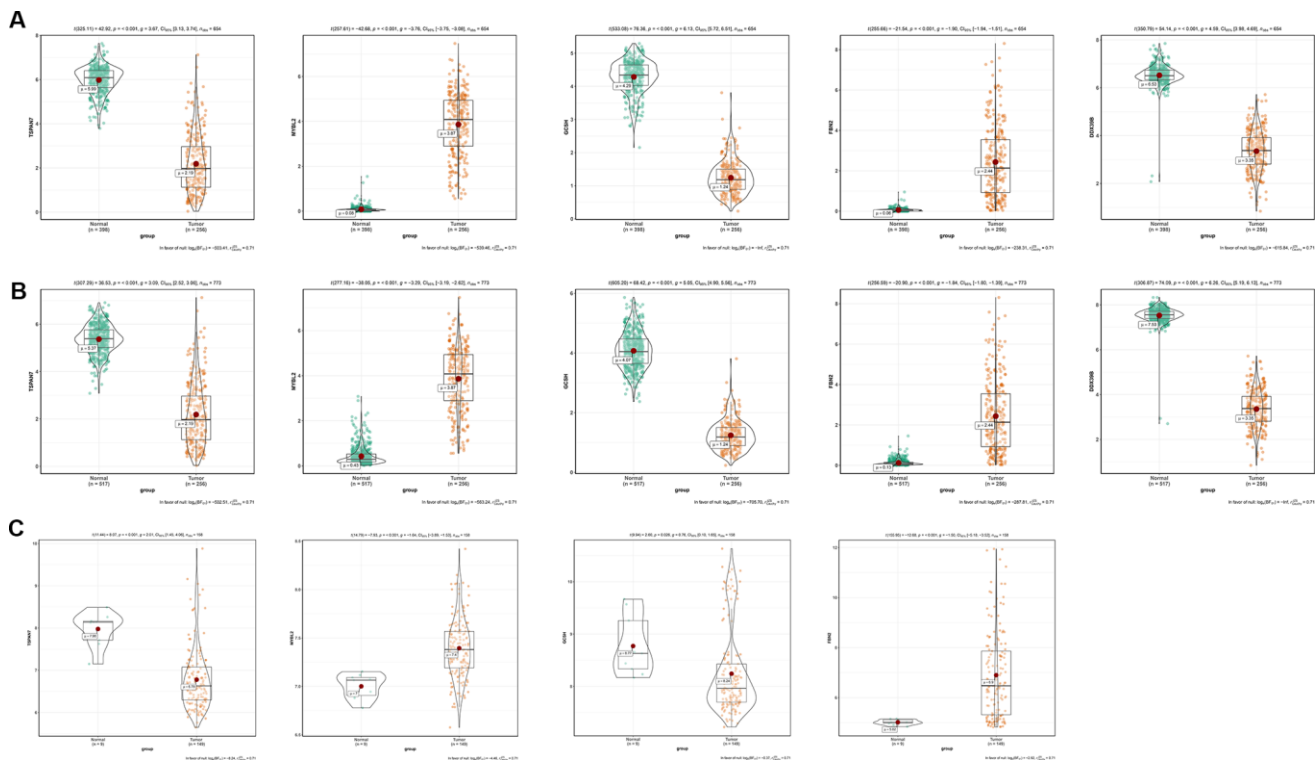
SUPPLEMENTARY MATERIALS



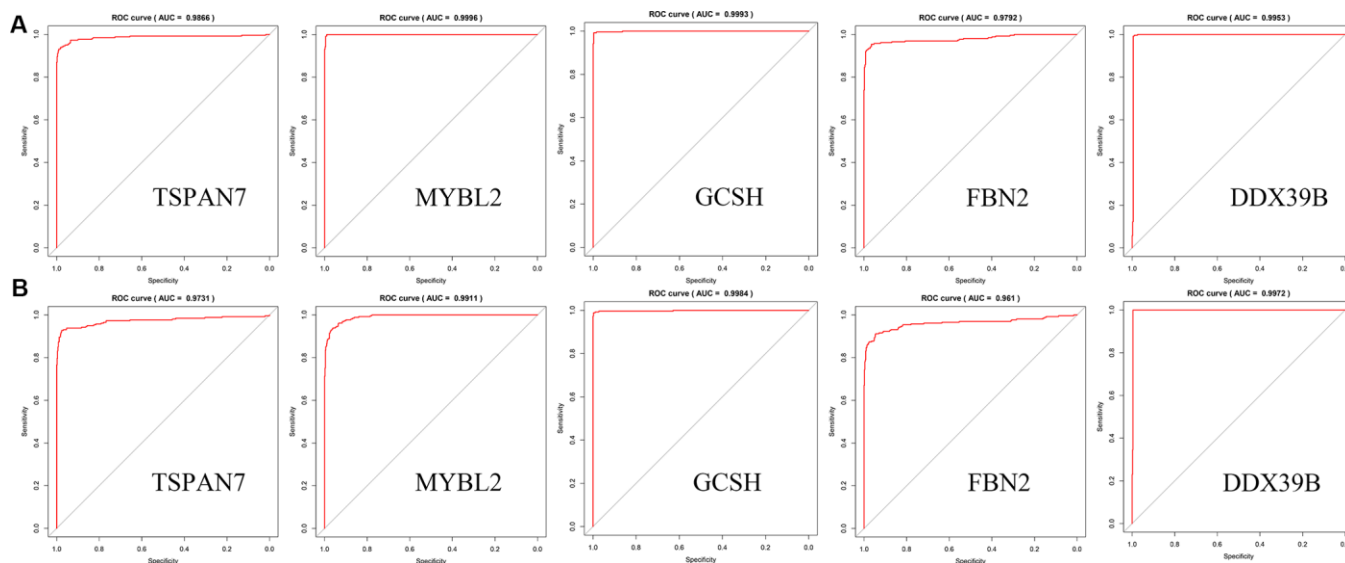
Supplementary Figure 1. Heatmaps of DEGs. Heatmaps of DEGs from “tumors” vs. “muscle controls” (A), “tumors” vs. “fat controls” (B), and “tumors” vs. “matched controls” (C).



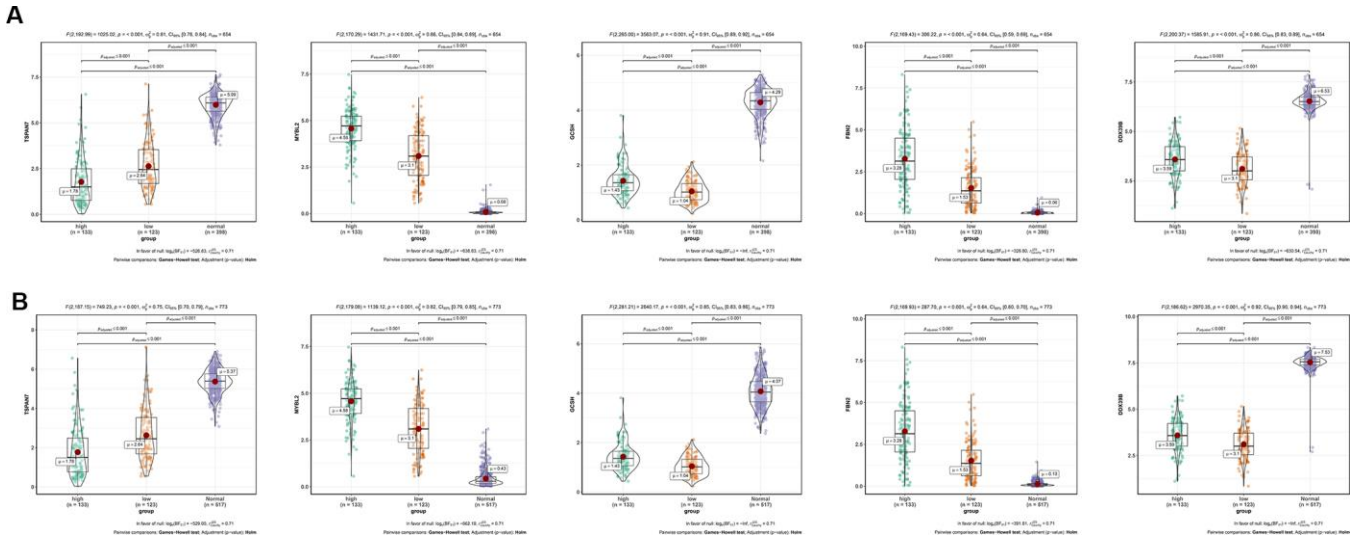
Supplementary Figure 2. Functional annotation of primary differentially expressed genes (DEGs). The Gene Ontology (GO) functional annotation (A and B) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments (C and D) of DEGs.



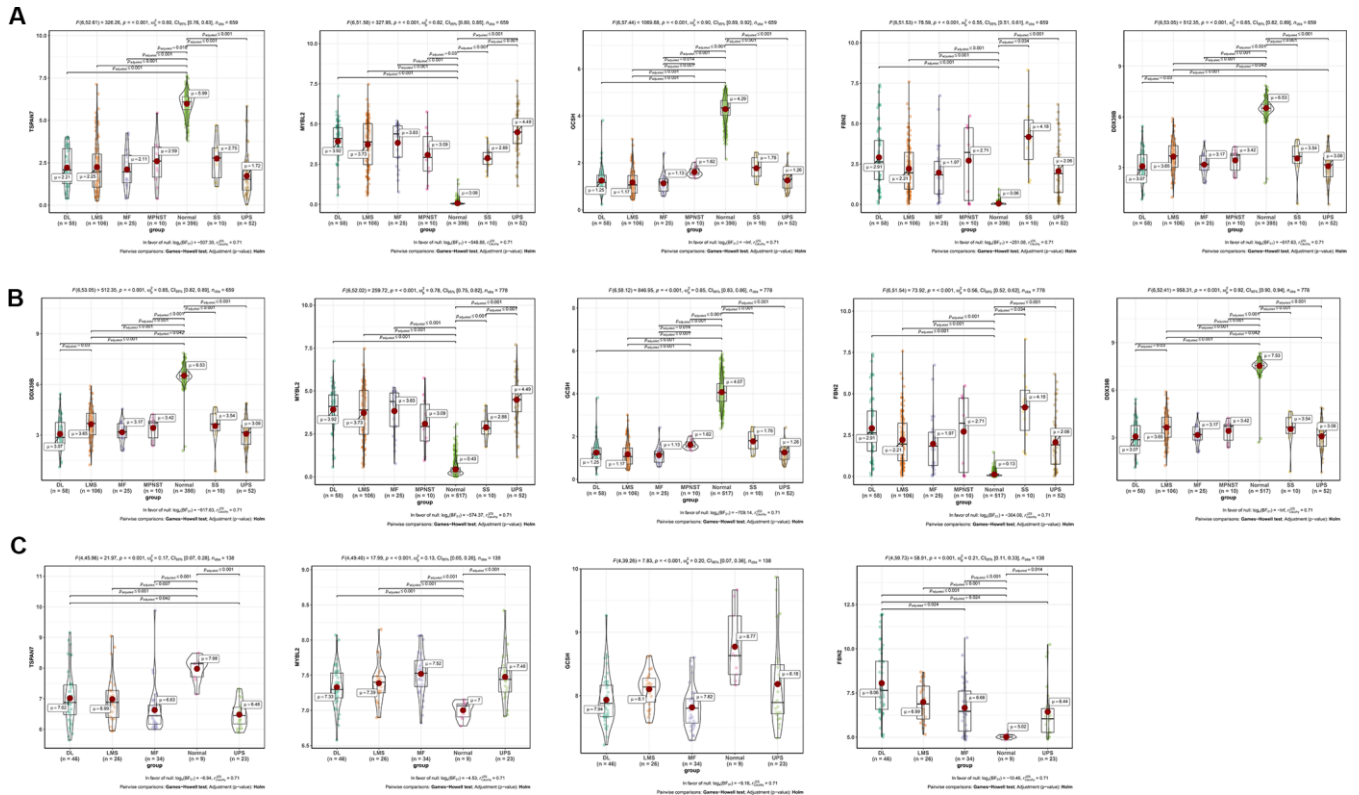
Supplementary Figure 3. Differences in the expression of key genes between soft tissue sarcoma (STS) and normal groups. (A) Differences in the expression of key genes between STS and muscle controls. (B) Differences in the expression of key genes between STS and fat controls. (C) Differences in the expression of five key genes between STS and normal groups in the GSE21122 datasets.



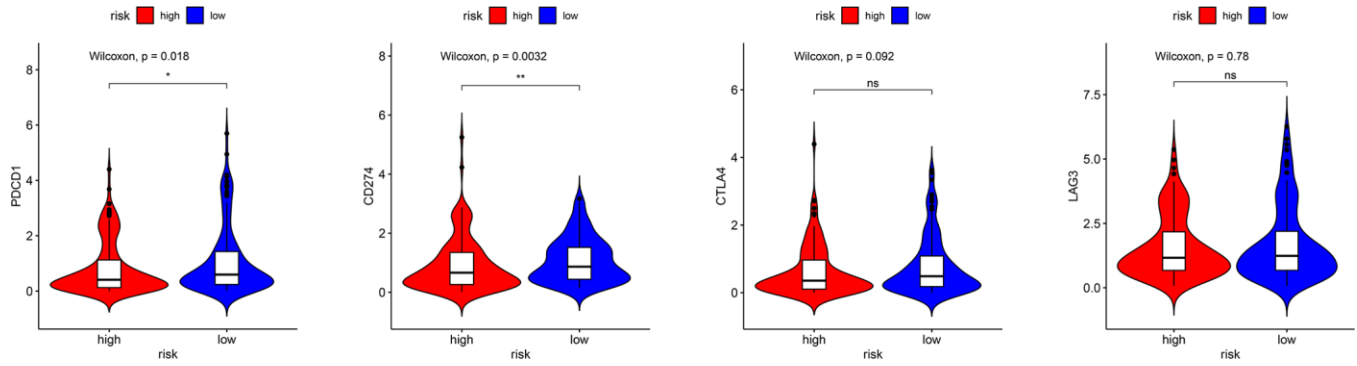
Supplementary Figure 4. Discrimination ability of key genes between STS and controls. (A) Discrimination ability of the five key genes between STS and muscle controls. (B) Discrimination ability of the five key genes between STS and fat controls.



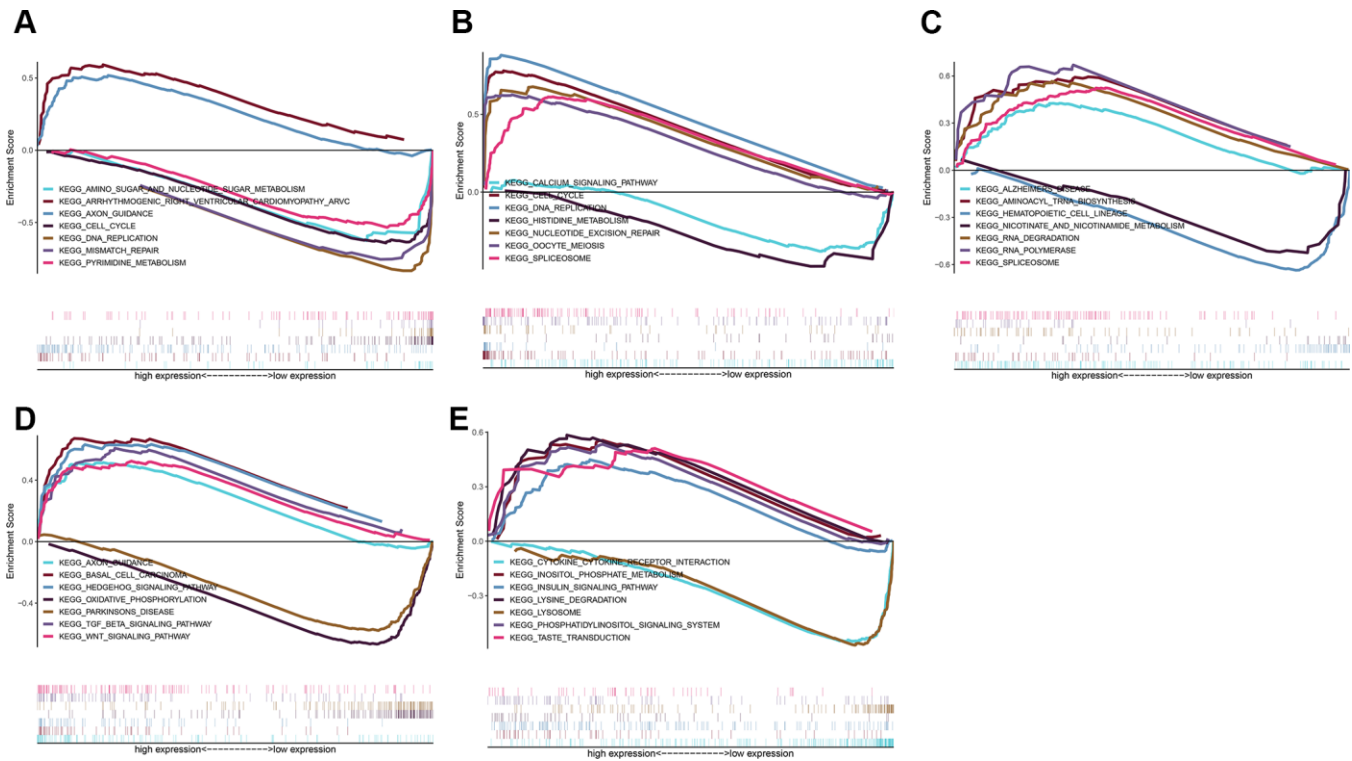
Supplementary Figure 5. Differences in the expression of key genes among high-risk and low-risk patients, and healthy (control) individuals. (A) Differences in the expression of key genes among high-risk and low-risk patients, and healthy (muscle controls) individuals. (B) Differences in the expression of key genes among high-risk and low-risk patients, and healthy (fat controls) individuals.



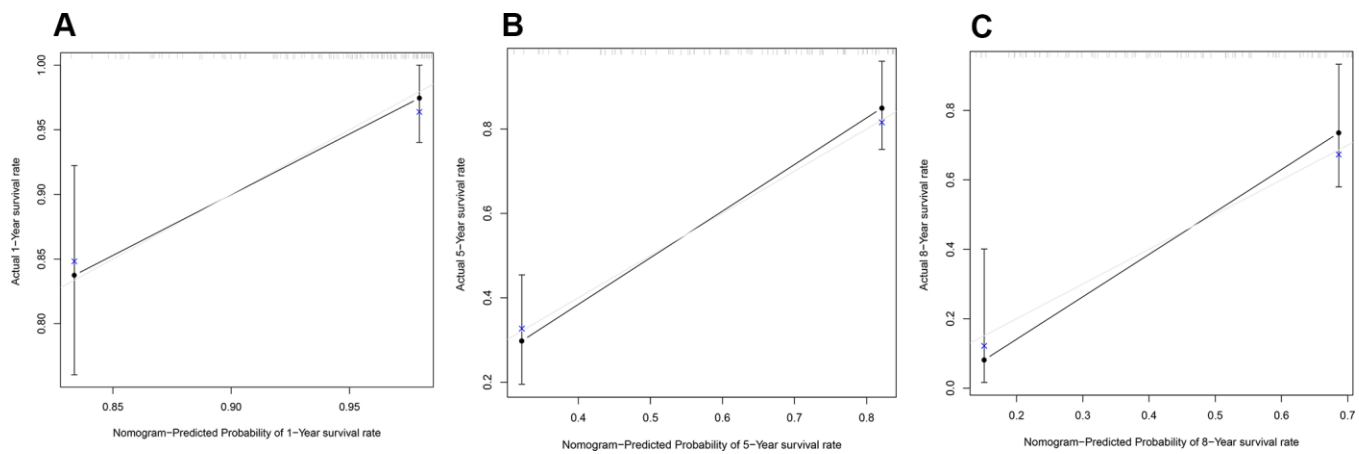
Supplementary Figure 6. Differences in the expression of key genes between all histological types and normal controls. (A) Differences in the expression of key genes between all histological types and muscle controls. (B) Differences in the expression of key genes between all histological types and fat controls. (C) Differences in the expression of key genes between all histological types and normal controls in the GSE21122 datasets.



Supplementary Figure 7. Differences in the common immune checkpoint molecules between high-risk and low-risk patients.



Supplementary Figure 8. Gene Set Enrichment Analysis (GSEA) of the five key genes. GSEA for TSPAN7 (A), MYBL2 (B), GCSH (C), FBN2 (D), and DDX39B (E).



Supplementary Figure 9. Accordance between actual survival rates and prediction of the nomogram.