Research Paper

The construction and analysis of tumor-infiltrating immune cell and ceRNA networks in recurrent soft tissue sarcoma

Runzhi Huang[1,2,*], Tong Meng[2,3,*], Rui Chen[4,*], Penghui Yan[1], Jie Zhang[5], Peng Hu[1], Xiaolong Zhu[1], Huabin Yin[3], Dianwen Song[3], Zongqiang Huang[1]

[1]Department of Orthopaedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China
[2]Division of Spine, Department of Orthopedics, Tongji Hospital affiliated to Tongji University School of Medicine, Shanghai 200065, China
[3]Department of Orthopedics, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai 200080, China
[4]Department of Urology, Shanghai Changhai Hospital, Second Military Medical University, Shanghai 200433, China
[5]Shanghai East Hospital, Key Laboratory of Arrhythmias, Ministry of Education, Tongji University School of Medicine, Shanghai 200120, China
*Equal contribution and co-first author

Correspondence to: Zongqiang Huang, Huabin Yin, Dianwen Song; **email:** gzhuangzq@163.com, yinhuabin@aliyun.com, osongdianwen@126.com
Keywords: soft tissue sarcoma, bone tumor, recurrence, ceRNA, immune cell, prognosis
Received: August 3, 2019 **Accepted:** October 28, 2019 **Published:** November 18, 2019

ABSTRACT

Soft tissue sarcoma (STS) is one of the most challenging tumors for medical oncologists, with a high rate of recurrence after initial resection. In this study, a recurrent STS-specific competitive endogenous RNA (ceRNA) network including seven recurrence and overall survival (OS)-associated genes (LPP-AS2, MUC1, GAB2, hsa-let-7i-5p, hsa-let-7f-5p, hsa-miR-101-3p and hsa-miR-1226-3p) was established based on the gene expression profiling of 259 primary sarcomas and 3 local recurrence samples from the TCGA database. The algorithm "cell type identification by estimating relative subsets of RNA transcripts (CIBERSORT)" was applied to estimate the fraction of immune cells in sarcomas. Based on 5 recurrence and OS-associated immune cells (NK cells activated, dendritic cells resting, mast cells resting, mast cells activated and macrophages M1), we constructed a recurrent STS-specific immune cells network. Both nomograms were identified to have good reliabilities (Area Under Curve (AUC) of 5-year survival is 0.724 and 0.773, respectively). Then the co-expression analysis was performed to identify the potential regulation network among recurrent STS-specific immune cells and ceRNAs. Hsa-miR-1226-3p and MUC1 were significantly correlated and dendritic cells resting was related to hsa-miR-1226-3p. Additionally, the expression of MUC1 and dendritic cell marker CD11c were also verified by immunohistochemistry (IHC) assay and multidimensional databases. In conclusion, this study illustrated the potential mechanism of hsa-miR-1226-3p regulating MUC1 and dendritic cells resting might play an important role in STS recurrence. These findings might provide potential prognostic biomarkers and therapeutic targets for recurrent STS.

INTRODUCTION

Soft tissue sarcoma (STS) is a group of rare tumors including more than 50 different histological subtypes [1]. It accounts for approximately 1% of adult malignancies and 15 % of pediatric malignancies [2, 3]. STS is derived from mesenchymal cell and usually divided into two broad categories: sarcomas of the soft

tissues and sarcomas of the bone [2]. The extremities, viscera, retroperitoneum and trunk are the most frequent sites, accounting for 70% of all cases [3, 4]. A complete resection is recommended for STS, but anatomic constraints hinder such efforts and local recurrence rate is high [5, 6]. Even after radical surgeries, about 30% of patients would experience local recurrence (LR) within 10 years, which is the most common cause of death [7]. Thus, there is a pressing need to explore the underlying mechanism of STS recurrence, which may provide potential prognostic factors and therapeutic targets for its treatment in the clinic.

Both tumor cells and tumor-infiltrating immune cells participate in tumorigenesis and tumor progression [9] and have been confirmed to be associated with recurrence and overall survival (OS) [10, 11]. The crosstalk between the tumor cells and tumor-infiltrating immune cells is usually modulated by the competing endogenous RNA (ceRNA) networks, which are composed of mRNAs messenger RNAs (mRNAs), long non-coding RNAs (lncRNAs), and microRNAs (miRNAs) [12]. Increasing studies indicated that the ceRNA networks regulate the post-transcription of oncogenes and tumor suppressor genes, modulate interactions between protein and genes, and control the biological behaviors such as tumor invasion and metastasis [12]. However, no combined networks have been defined for predicting the prognosis of STS recurrence up to date. Therefore, a better understanding of the tumor-infiltrating immune cells and ceRNA networks is required.

In the current study, we identified the differential expressed ceRNAs involved in recurrent STSs based on their gene expression profiling available from the TCGA (The Cancer Genome Atlas) database and used the algorithm "CIBERSORT" to quantify the proportions of immune cells. In addition, prediction nomograms based on recurrence and OS-associated immune cells or ceRNAs were constructed to predict STS recurrence. Moreover, we assessed the relationships between recurrent STS-specific immune cells and ceRNA networks to identify the underlying immune gene signature.

RESULTS

Identification of significantly differentially expressed genes

Figure 1 summarizes the analysis process of this study. The baseline characteristics of all the patients available from the TCGA database are listed in

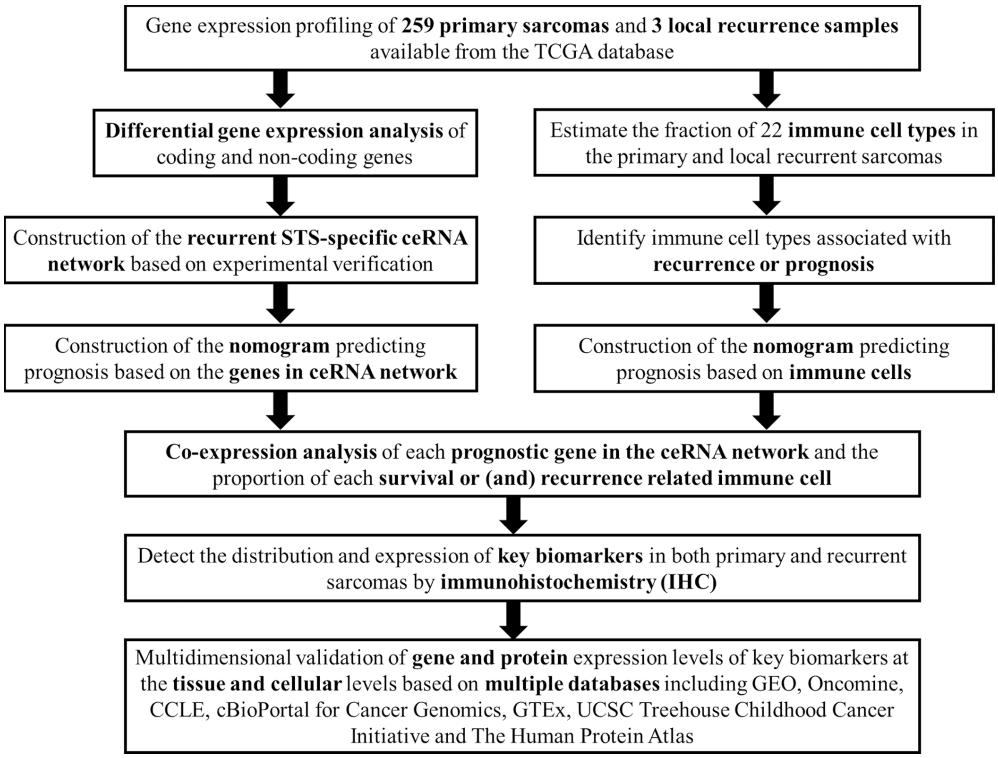


Figure 1. The flow chart of the analysis process. Abbreviations: TCGA: The Cancer Genome Atlas; STS: Soft tissue sarcoma; GEO: Gene Expression Omnibus; CCLE: Cancer Cell Line Encyclopedia; GTEx: Genotype-Tissue Expression; UCSC: University of California, Santa Cruz.

Supplementary Table 1. The Kaplan-Meier survival analysis revealed that recurrence was a significant predictor for poor prognosis of STSs (P = 0.001) (Supplementary Figure 1).

A total of 14,447 lncRNAs, 2,588 miRNAs and 19,660 mRNAs were found from the TCGA database. Among them, 148 differentially expressed protein-coding genes (143 downregulated and 5 upregulated) (Figure 2A–2C, 2E), 21 differentially expressed lncRNAs (downregulated) (Figure 2A, 2D) and 4 differentially expressed miRNAs (downregulated) were identified between primary and recurrent STSs using the cutoff of the log (fold-change) > 1.0 or < −1.0 and FDR < 0.05.

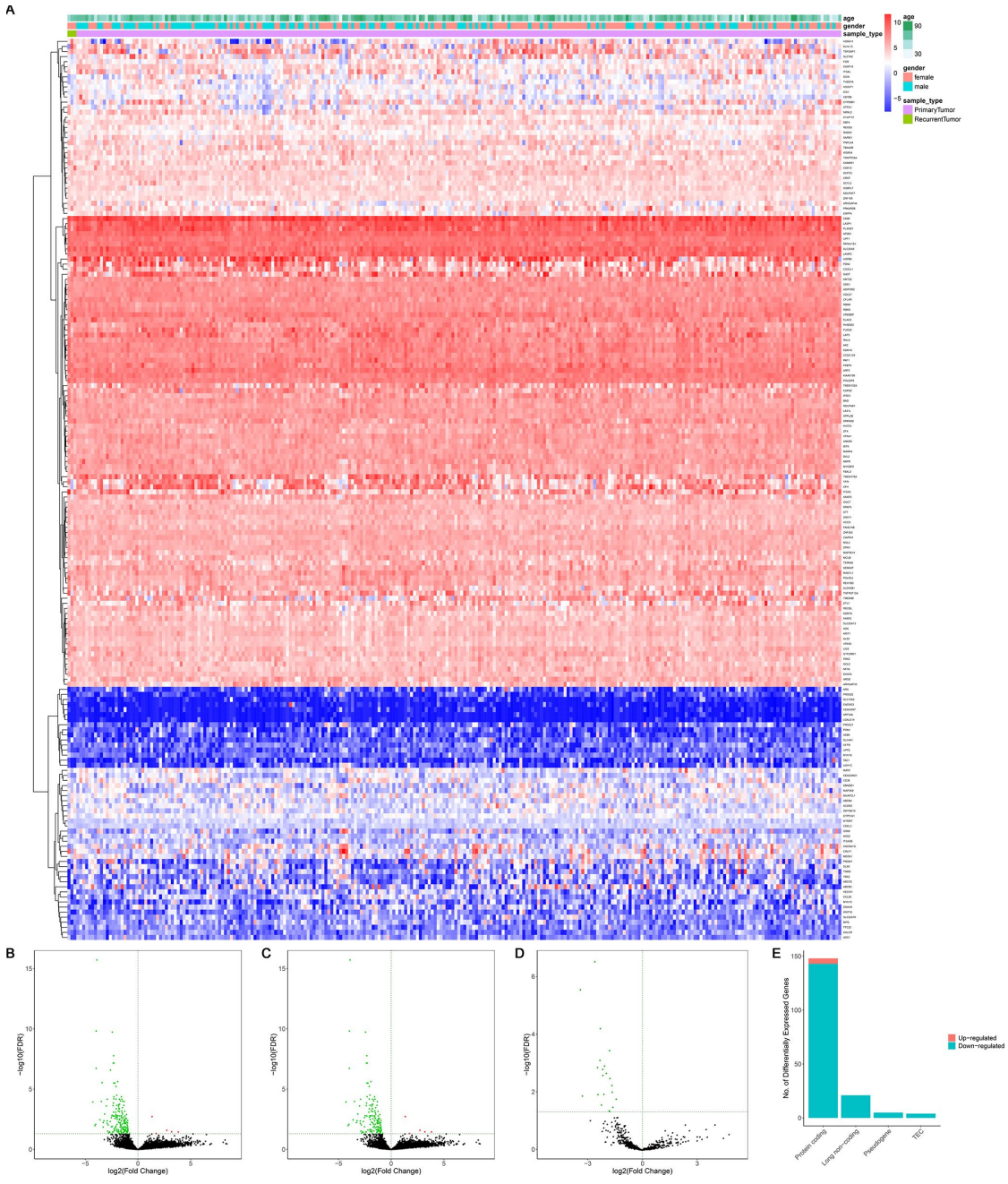


Figure 2. The differentially expressed genes between primary and recurrent STSs. (**A**) The heatmap and the volcano plot (**B**) of 178 differentially expressed genes between 259 primary and 3 recurrent STSs; (**C**) The volcano plot of 148 differentially expressed protein-coding genes between 259 primary and 3 recurrent STSs; The volcano Plot (**D**) of 21 differentially lncRNAs between 259 primary and 3 recurrent STSs; (**E**) The composition of differentially expressed genes. The log(fold-change) > 1.0 or < -1.0 and FDR < 0.05. Abbreviations: ceRNAs: competing endogenous RNAs; STSs: soft tissue sarcomas; LncRNA: long non-coding RNA.

Supplementary Table 2 summarizes the top 10 downregulated and top 10 upregulated genes in differential gene analysis.

Construction of the ceRNA network and survival analysis

A ceRNA network including 23 genes was established based on the interactions of 11 lncRNA-miRNA pairs and 12 miRNA-mRNA pairs (Figure 3A) (Table 1). The Cox regression and Kaplan-Meier method were applied to examine the relationship between the biomarkers in the recurrence-associated ceRNAs and OS. LPP-AS2 ($P = 0.039$), MUC1 ($P = 0.003$), GAB2 ($P = 0.049$), hsa-let-7i-5p ($P < 0.001$), hsa-let-7f-5p ($P = 0.025$), hsa-miR-101-3p ($P = 0.028$) and hsa-miR-1226-3p ($P = 0.001$) were significantly associated with survival in Kaplan-Meier analysis (Figure 3B–3G). Seven potential recurrence and OS-associated biomarkers were identified as key molecules in the ceRNA network and were integrated into a new multivariable model (Table 2). The results of the Lasso regression indicated that all seven genes were essential for modeling (Figure 4A, 4B). Additionally, the ROC and the calibration curves indicated decent accuracy (Area Under Curve (AUC) of 3-year survival: 0.731; AUC of 5-year survival: 0.724) and good discrimination (Figure 4C, 4E). Then, the nomogram was constructed based on the model (Figure 4D).

Composition of immune cells in sarcomas

Figure 5 illustrated the composition of immune cells estimated by the CIBERSORT algorithm in STSs. The fraction of the NK cells activated was consistently lower in the local recurrence tissue than in primary sarcomas, whereas the fractions of dendritic cells resting and the mast cells resting were higher in the local recurrence sarcoma tissue. Wilcoxon rank-sum test was then used and revealed that the fractions of dendritic cells resting ($P = 0.016$) and NK cells activated ($P = 0.036$) varied significantly between recurrent and primary tumors (Figure 5C).

Integrated analysis of immune cells, genes and prognosis

All immune cells were integrated into a Cox regression model. After the screening process of the Lasso regression, the fractions of NK cells activated ($P = 0.029$), dendritic cells resting ($P = 0.013$), mast cells resting ($P < 0.001$), mast cells activated ($P = 0.030$) and macrophages M1 ($P = 0.024$) were all considered as independent predictors in the final Cox model (Table 3). The results of the Lasso regression suggested that the model was not overfitting (Figure 6A, 6B). In addition, the calibration curve and the ROC demonstrated good discrimination and concordance (AUC of 3-year survival: 0.709; AUC of 5-year survival: 0.773) (Figure 6C, 6F). Similarly, the nomogram based on the multivariate analysis was constructed (Figure 6E). Lastly, immune cells and biomarkers significantly associated with OS were integrated into the nomogram (Supplementary Figure 2) for predicting the prognosis (AUC of 3-year survival: 0.789; AUC of 5-year survival: 0.822). With respect of the correlation analysis, significant co-expression patterns between fractions of immune cells and key molecules in the ceRNA network were identified, showing that hsa-miR-1226-3p was significantly associated with dendritic cells resting ($R = -0.19$, $P = 0.004$) (Figure 7). Additionally, according to the result of the Wilcoxon rank-sum test, hsa-miR-1226-3p was significant different between the sarcoma tissues of patients with and without recurrence ($P = 0.015$) (Supplementary Figure 3).

MUC1 and CD11c were associated with STS recurrence

We examined the expressions of MUC1 and CD11c in primary and recurrent leiomyosarcoma (LMS) and liposarcoma (LPS) specimens (Table 4). Of the 10 patients with recurrent LMS, the mean H-score of MUC1 was 2.25, which was significantly higher than that of patients with primary LMS ($P < 0.05$) (Figure 8A, 8A'). The results of CD11c were similar (Figure 8B, 8B'). Then we compared the H-score of MUC1 and CD11c in patients with primary or recurrent LPS. Both of them were significantly higher in patients with recurrent LPS (Figure 8C, 8C', 8D, 8D'). In addition, the results showed that the MUC1 and CD11c protein was predominantly localized in the membrane and extracellular matrix of LMS (Supplementary Figure 4A, 4B) and LPS cells (Supplementary Figure 4C, 4D).

Multidimensional validation

A dimensional validation was performed to explore the expressions of MUC1 and CD11c in the primary STS, normal soft tissue and cell lines (Table 5). First, MUC1 (Median rank 1,052, $P = 0.012$) was highly expressed in primary STS compared to normal tissue while CD11c (Median rank 10,499, $P = 0.952$) showed no difference in all of the 10 comparisons (Supplementary Figure 5). At cellular level, MUC1 was expressed in various STS cell lines while the expression of CD11c was low (Supplementary Figure 6). Besides, an analysis of genomics and clinical profiles with the cBioPortal suggested that MUC1 and CD11c were highly expressed in primary STS compared to other types of malignancies and both were in a co-expression relationship ($R = 0.28$, $P < 0.001$) (Supplementary Figure 7). Moreover, we

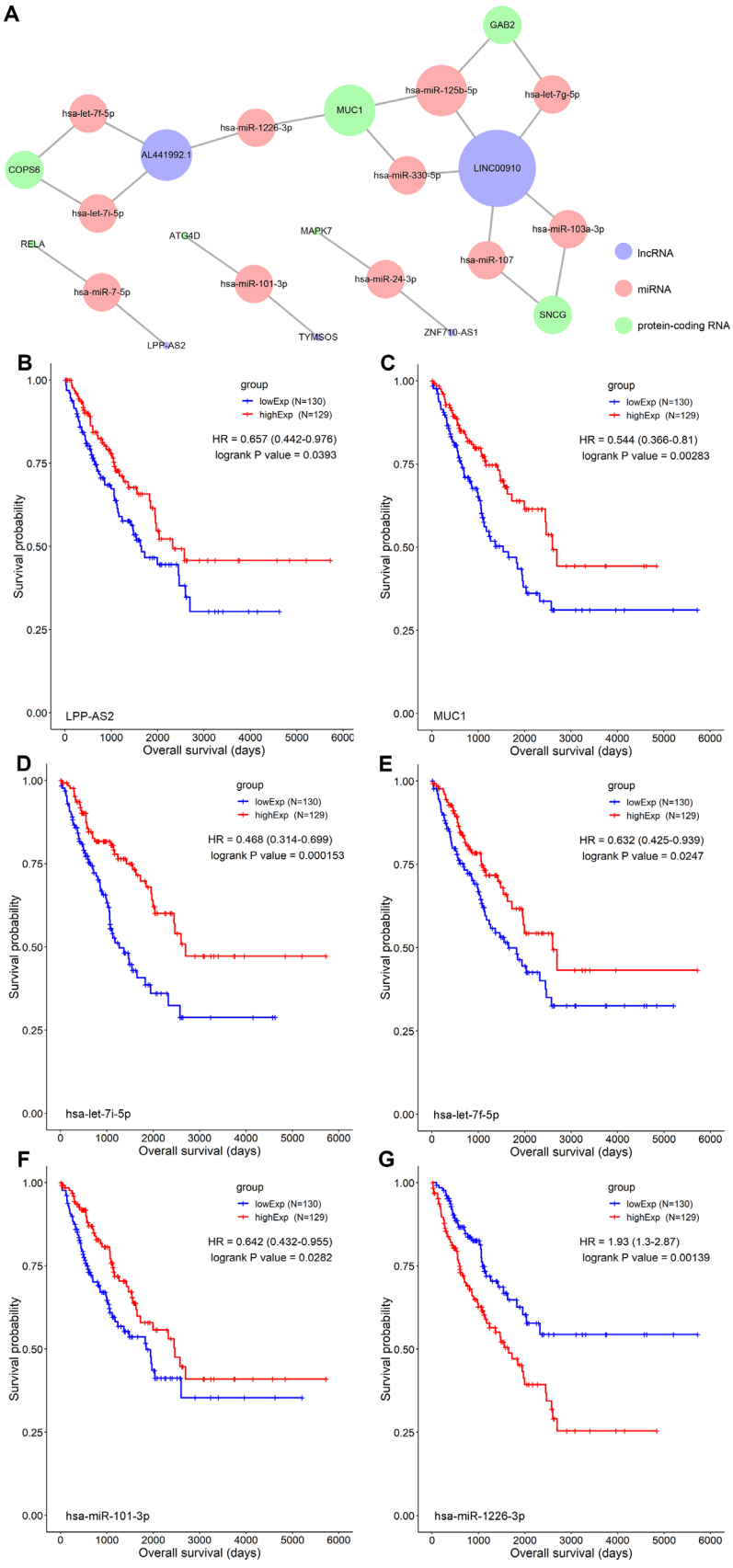


Figure 3. (**A**) The STS-recurrence related ceRNA network; The Kaplan-Meier survival curves of LPP-AS2 (**B**), MUC1 (**C**), hsa-let-7i-5p (**D**), hsa-let-7f-5p (**E**), hsa-miR-101-3p (**F**) and hsa-miR-1226-3p (**G**). Abbreviations: STSs: soft tissue sarcomas; ceRNAs: competing endogenous RNAs

Table 1. Hypergeometric testing and correlation analysis results of ceRNAs network.

LncRNA	Protein-coding RNA	MiRNAs	Correlation P	Hypergeometric test P
AL441992.1	MUC1	hsa-miR-1226-3p	2.81E-10	0.044178237
AL441992.1	COPS6	hsa-let-7f-5p,hsa-let-7i-5p	8.92E-07	0.000306228
LINC00910	GAB2	hsa-let-7g-5p,hsa-miR-125b-5p	0.002903258	0.005948459
LINC00910	MUC1	hsa-miR-125b-5p,hsa-miR-330-5p	4.29E-07	0.014261321
LINC00910	SNCG	hsa-miR-103a-3p,hsa-miR-107	4.46E-05	0.009708465
TYMSOS	ATG4D	hsa-miR-101-3p	3.68E-05	0.011207883
LPP-AS2	RELA	hsa-miR-7-5p	0.020065721	0.02784492
ZNF710-AS1	MAPK7	hsa-miR-24-3p	0.013123574	0.001872659

Abbreviations: ceRNAs: Competing endogenous RNAs; LncRNA: Long non-coding RNA; MiRNA: microRNA.
* P < 0.05

Table 2. Cox proportional hazards regression model including the key members of the ceRNA network for overall survival in patients with soft tissue sarcoma.

Gene	Hazard ratio	95%CI	P value
GAB2	0.81	(0.67 − 0.98)	0.034 *
RELA	0.57	(0.33 − 0.98)	0.041 *
MUC1	0.88	(0.79 − 0.97)	0.012 *
has-let-7f-5p	0.91	(0.73 − 1.13)	0.375
has-let-7i-5p	0.68	(0.50 − 0.92)	0.012 *
has-miR-1226-3p	1.28	(1.03 − 1.58)	0.024 *
LPP-AS2	0.61	(0.44 − 0.84)	0.003 **

Abbreviations: ceRNAs: Competing endogenous RNAs; CI: Confidence Interval.
*: P < 0.05; **:P < 0.010; ***:P < 0.001.
Note. In the variable selection process, first of all, the initial Cox models including all members of the ceRNA network were used to select potential prognostic genes. At the same time, the Lasso regression was performed based on all members of the ceRNA network. The results of the Lasso regression (Figure 4A, 4B) suggested that the eight genes were essential to modeling and ensuring not overfitting of the model. Eventually, the reduce Cox model shown in this table only including eight genes filtrating by the Lasso regression (The Cox model of immune cells was constructed in the same way).

extracted the RNA-seq data of 907 normal adipose or muscle tissues from the GTEx database and 509 sarcomas from the Treehouse for differential gene analysis. MUC1 (logFC = 4.10, P < 0.001) was identified in the inter-group differential expression while CD11c was not (Supplementary Figure 8). Additionally, the results from The Human Protein Atlas showed that the protein MUC1 and CD11c were almost not detected in normal adipose and smooth muscle tissue (Supplementary Figure 9). Finally, we also evaluated the prognostic value and relationship among MUC1 and twelve markers of dendritic cell. The results revealed that CD49d, CD304, CD209, CD11b and CD86 had a co-expression pattern with MUC1 and CD40, CD197, CD205 were associated with OS (Supplementary Figure 10).

DISCUSSION

STSs are one of the most challenging tumors for medical oncologists, with a high rate of relapse after initial resection [5]. During tumorigenesis and recurrence, molecular and cellular components played important roles and were often regarded as potential prognostic factors [13]. The significant genes that are aberrantly expressed in tumor and tumor-infiltrating immune cells attract our interest, however, very few studies of STS focused on them before. In the present study, we found out the significant tumor-infiltrating immune cells and ceRNAs between primary and recurrent STS. Two prediction nomograms with high efficacy were constructed based on these findings and thus both nomograms might assist clinical oncologists in evaluation of prognosis and recurrence. By comparing the correlation between the recurrence-associated ceRNAs and immune cells, we inferred a potential mechanism of STS recurrence and that was hsa-miR-1226-3p regulating MUC1 and dendritic cells resting. Then multidimensional validation of multiple databases confirmed the reliability of our results.

The ceRNA networks link the function of protein-coding mRNAs with ncRNAs, such as miRNAs and lncRNAs [12]. Previous studies revealed that miRNAs were able to bind to the 3' untranslated region (3' UTR) of the target mRNAs in a complementary base-pairing manner and take part in post-transcriptional regulation of oncogenes and antioncogenes [17, 18]. The lncRNAs not only regulated interactions between protein and genes, but also modulated transcription by recruiting

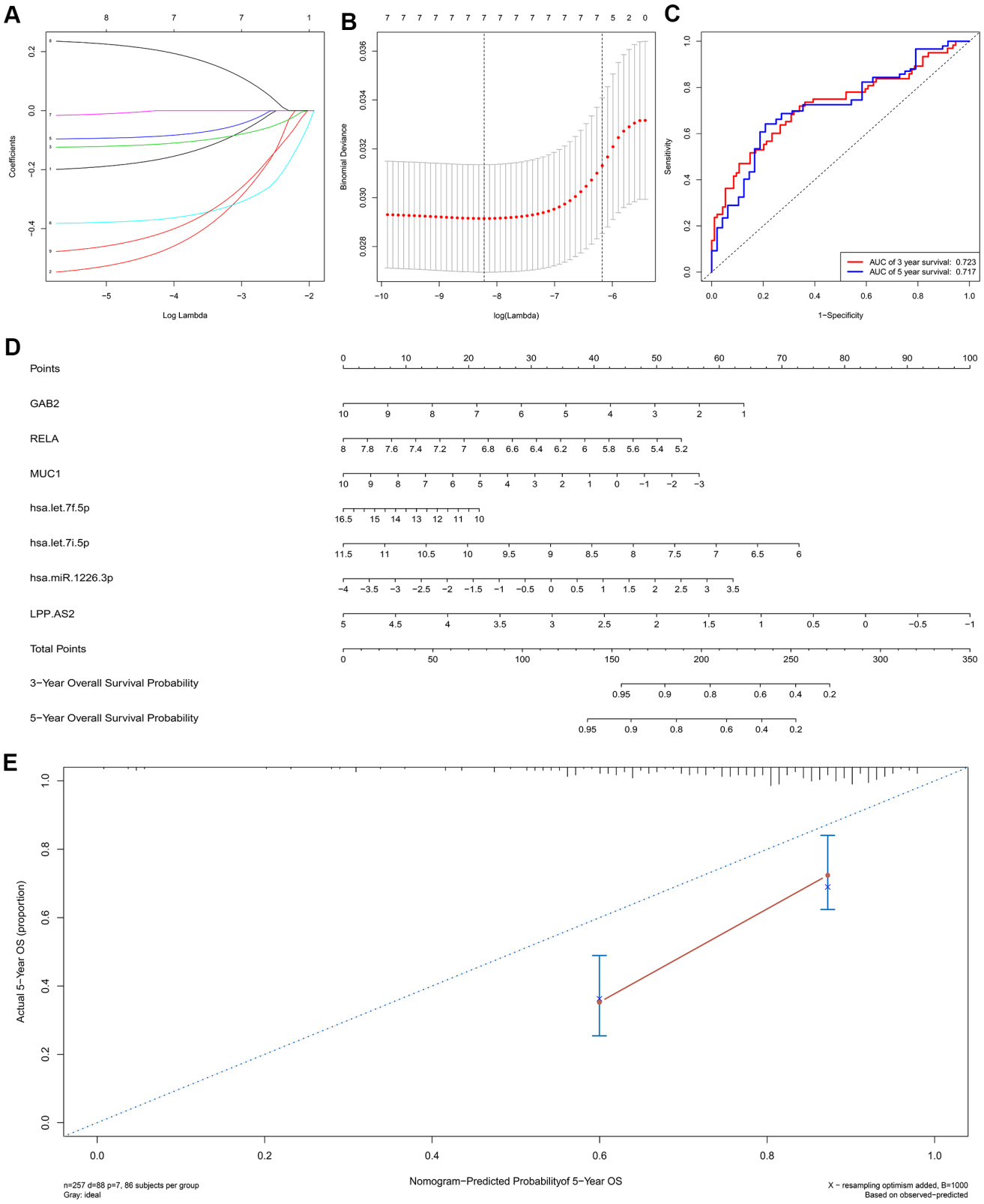


Figure 4. The results of the multivariate Cox regression, nomogram (**E**) and model diagnosis process (**B**, **C**, **D**, **F**) based on the key members in the ceRNA network. Seven potential prognosis-related ceRNAs were integrated into a new multivariable model. The results of the Lasso regression suggested that all seven genes were essential for modeling (**A**, **B**). The nomogram was constructed based on the model (**D**). The ROC and the calibration curves indicated acceptable accuracy (Area Under Curve (AUC) of 3-year survival: 0.731; AUC of 5-year survival: 0.724) and discrimination of the nomogram (**C**, **E**).

chromatin-modifying complexes [19, 20]. Emerging evidence demonstrated their potential roles in controlling the biological process including tumorigenesis, invasion and metastasis [20–22].

In this study, hypergeometric testing and correlation analysis results of the ceRNAs network revealed that hsa-miR-1226-3p (miRNA), MUC1 (protein-coding RNA) and AL441992.1 (lncRNA) were significantly

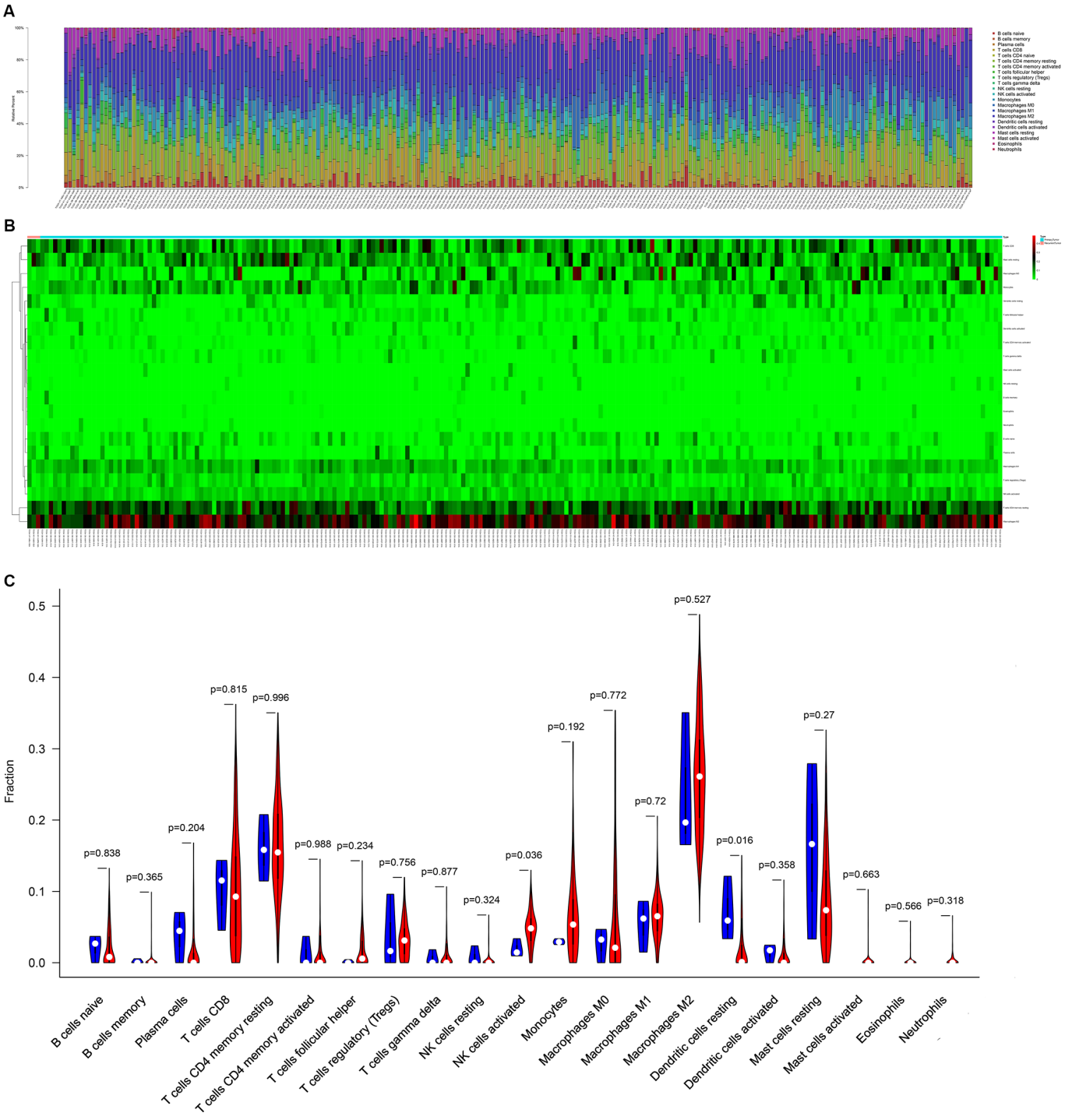
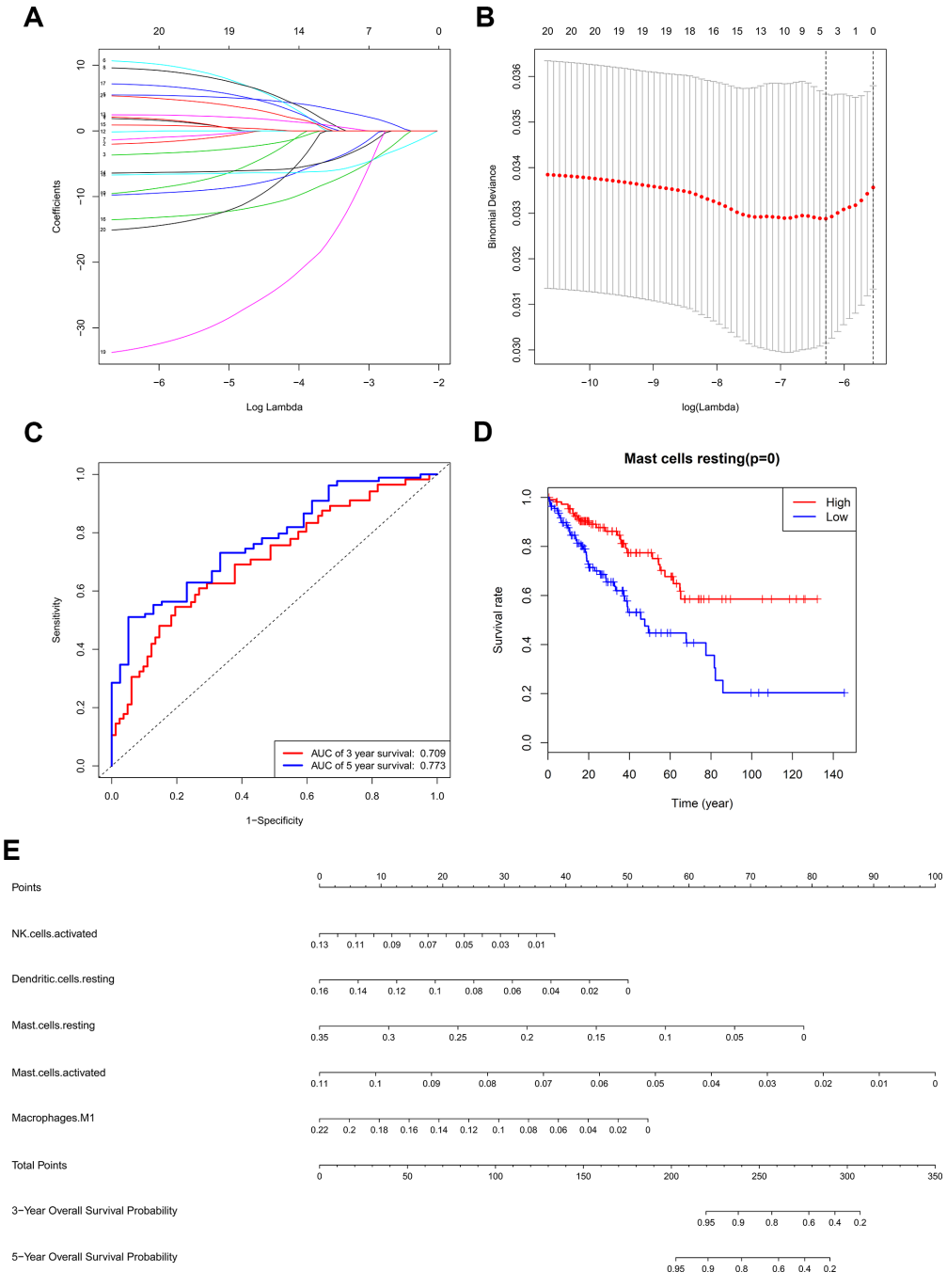


Figure 5. The composition (**A**) and heatmap (**B**) of immune cells estimated by CIBERSORT algorithm in sarcomas. (**C**) The violin plot of immune cells (The blue and red bar stand for recurrent tumor group and primary tumor group, respectively). Abbreviations: CIBERSORT: Cell type identification by estimating relative subsets of RNA transcripts.

Table 3. Cox proportional hazards regression model including the key immune cells for overall survival in patients with soft tissue sarcoma.

Immune cell	Hazard ratio	95%CI	P value
NK cells activated	9.6e−06	(3.0e−10 − 0.309)	0.029 *
Dendritic cells resting	4.4e−06	(2.7e−10 − 0.074)	0.013 *
Mast cells resting	1.4e−04	(1.9e−06 − 0.011)	< 0.001 ***
Mast cells activated	2.9e−16	(2.4e−30 − 0.034)	0.030 *
Macrophages M1	7.2e−05	(1.8e−08 − 0.280)	0.024 *

Abbreviations: CI: Confidence Interval.
*: P < 0.05; **:P < 0.010; ***:P < 0.001.



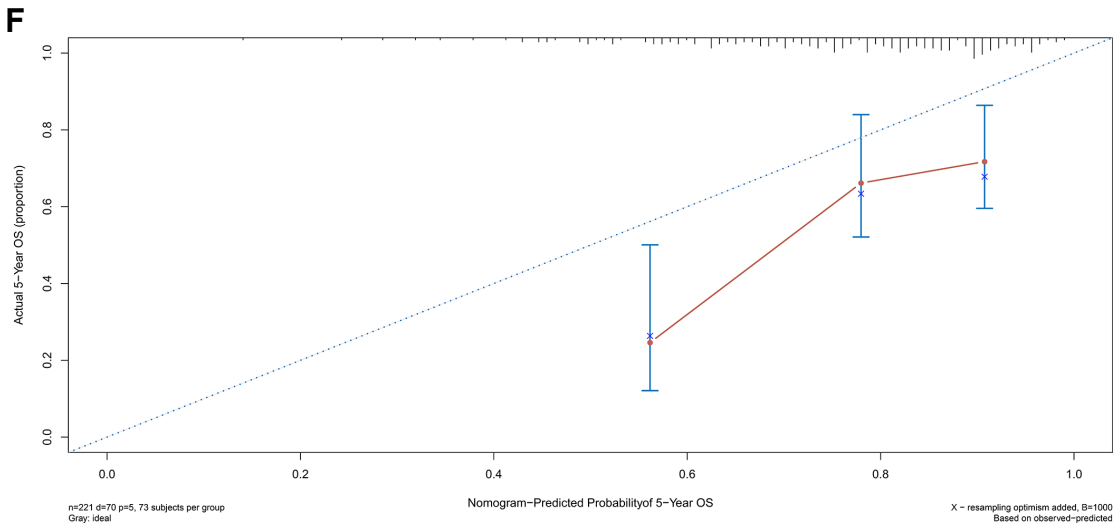


Figure 6. The results of the multivariate Cox regression, Lasso regression (**A**, **B**), Kaplan–Meier survival curve of (**D**), nomogram (**E**) and model diagnosis process (**C**, **F**) based on prognosis related immune cells. All immune cells were integrated into an initial Cox regression model. After the screening process of the Lasso regression, the results suggested that the model was not overfitting (**A**, **B**). The nomogram based on the multivariable model (**E**). The calibration curve and the ROC demonstrated good discrimination and concordance of the nomogram (AUC of 3-year survival: 0.709; AUC of 5-year survival: 0.773) (**C**, **F**).

Table 4. The mean H-score of MUC1 and CD11c in primary/recurrent LMS and LPS.

Biomarker	Primary LMS	Recurrent LMS	Primary LPS	Recurrent LPS
MUC1	1.05	2.55	1.15	2.05
CD11c	1.025	1.675	1.175	1.625
P value	<0.001	<0.001	0.004	0.42

Abbreviations: LMS: leiomyosarcoma; LPS: liposarcoma.

correlated. In the meanwhile, the correlation analysis also revealed that hsa-miR-1226-3p was significantly associated with dendritic cells resting (R= -0.190, P = 0.004). Thus, we inferred that the mechanism of hsa-miR-1226-3p regulating MUC1 and dendritic cells resting might play an important role in STS recurrence.

miR-1226 was reported to be involved in tumorigenesis, angiogenesis and drug resistance in breast cancer and non-small cell lung cancer [23, 24]. The correlation between hsa-miR-1226-3p and MUC1 has also been proved by the previous study which revealed that miR-1226 interacted with the MUC1 mRNA 3' UTR and induced downregulation of MUC1 [25].

Generally, MUC1 is overexpressed at mucosal surfaces and absent in the skin epithelium and mesenchymal cells [26, 27]. Aberrantly glycosylated MUC1 is often overexpressed in most human epithelial cancers, but not reported in mesenchymal cell originated STS. In the tumorigenesis, MUC1 was reported to induce the expression of growth factors such as connective tissue growth factor (CTGF), vascular endothelial growth factor-A (VEGF-A) and platelet-derived growth factor A (PDGF-A), that promote cell adhesion, angiogenesis and proliferation [28, 29]. During tumor metastases, it also induced epithelial to mesenchymal transition (EMT) by modulating the expression of miRNAs that promoted EMT-related gene expression [30, 31]. In this study, we suggested that MUC1 was highly expressed in recurrent LMS and LPS, suggesting a potential novel biomarker to predict STS recurrence. Additionally, as an extensively O-glycosylated and moderately N- glycosylated transmembrane protein on epithelial cells, many antibody-drug conjugates (ADC) have been explored for MUC1, such as HuHMFG1 [32, 33]. Thus, MUC1 can be regarded as a potential therapeutic target for recurrent STS.

Recently, MUC1 has also been designed to be the target of an anticancer vaccine. The MUC1 anticancer vaccine was equipped with covalently linked divalent mannose ligands and the mannose coupling also led to increasing numbers of macrophages, dendritic cells (DCs), and $CD4^+$ T cells [34]. MUC1 could be carried in extracellular microvesicles, which played a contradictory role in promoting both immunosuppression and tumor growth. The specific immune response was reported to positively impact DCs immunogenicity by reprogramming DC antigen processing machinery and intracellular signaling pathways [35].

DCs, also known as professional antigen-presenting cells (APC), are specialized in providing co-stimulation and cytokines to regulate tumor antigen-specific T cell immune response activation [14]. They interact with other immune cells, such as NK cells and B cells, and activate anti-tumor responses [15]. The diversity of DC populations, divided by localization and activity, makes its function specific. The former includes Langerhans cells, monocyte-derived DCs ($CD14^+$ DCs), myeloid DCs, plasmacytoid DCs (pDCs) [16]. The latter includes DCs activated and DCs resting [44].

In addition, both miR-1226 and MUC1 were reported to not only take effects in the intracellular environment but also be secreted to the extracellular environment which also provides the opportunity to regulate dendritic cells resting [35, 36].

In addition, correlation between hsa-let-7i-5p with dendritic cells resting was also significant (R = 0.200, P = 0.003). After a systematic literature review, we found no direct reports on hsa-let-7i-5p association with dendritic cells and tumor immunity. However, COP9

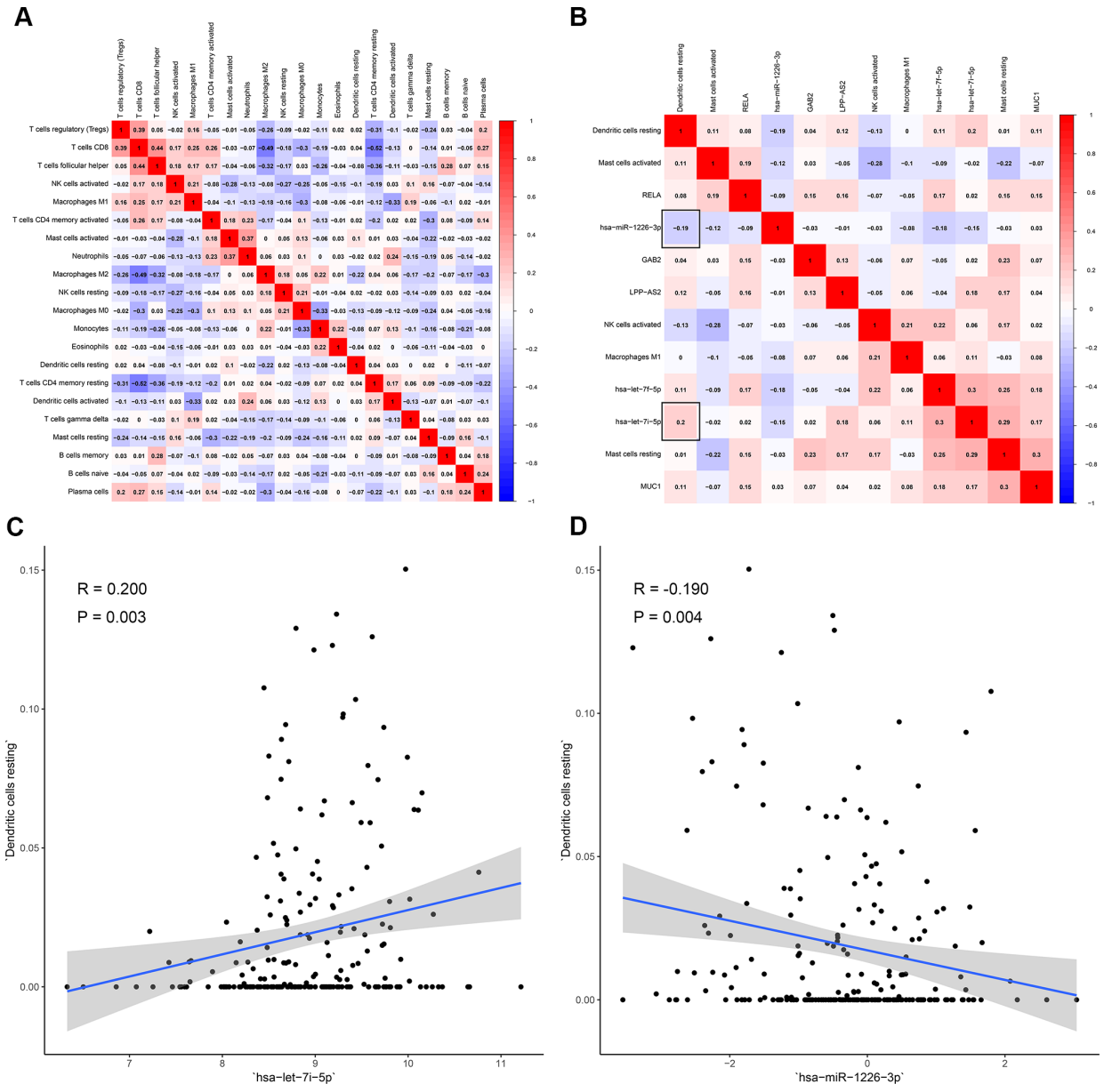


Figure 7. The co-expression patterns among fractions of immune cells and key members in the ceRNA network. (**A**) co-expression heatmap of all immune cells; (**B**) co-expression heatmap of prognostic immune cells and key members of ceRNA network; (**C**) has-let-7i-5p was significantly associated with dendritic cells resting (R = 0.200, P = 0.003); (D) hsa-miR-1226-3p was significantly associated with dendritic cells resting (R = -0.190, P = 0.004).

Signalosome Subunit 6 (COPS6) regulated by hsa-let-7i-5p had been proved by a previous study biological experiment [37]. COP9 Signalosome is a highly conserved protein complex, working as an important regulator in multiple signaling pathways. Especially, it had been reported to involve in the human immunodeficiency virus type 1 (HIV-1) regulating immune cell death [38–40]. Therefore, we speculated that hsa-let-7i-5p might affect dendritic cells by regulating its target gene COPS6, which was involved in immune regulation. We had such a conserved discussion because of the lack of literature. Subsequent studies are needed to explore the relationship between hsa-let-7i-5p and dendritic cells.

There are inevitably several limitations of our study that should be acknowledged. First, the amount of data released in publicly available datasets is limited, so that the clinicopathological parameters analyzed in this study are not comprehensive, which might lead to potential error or bias. And the sample size of the recurrent samples was very small, which might cause analysis bias.

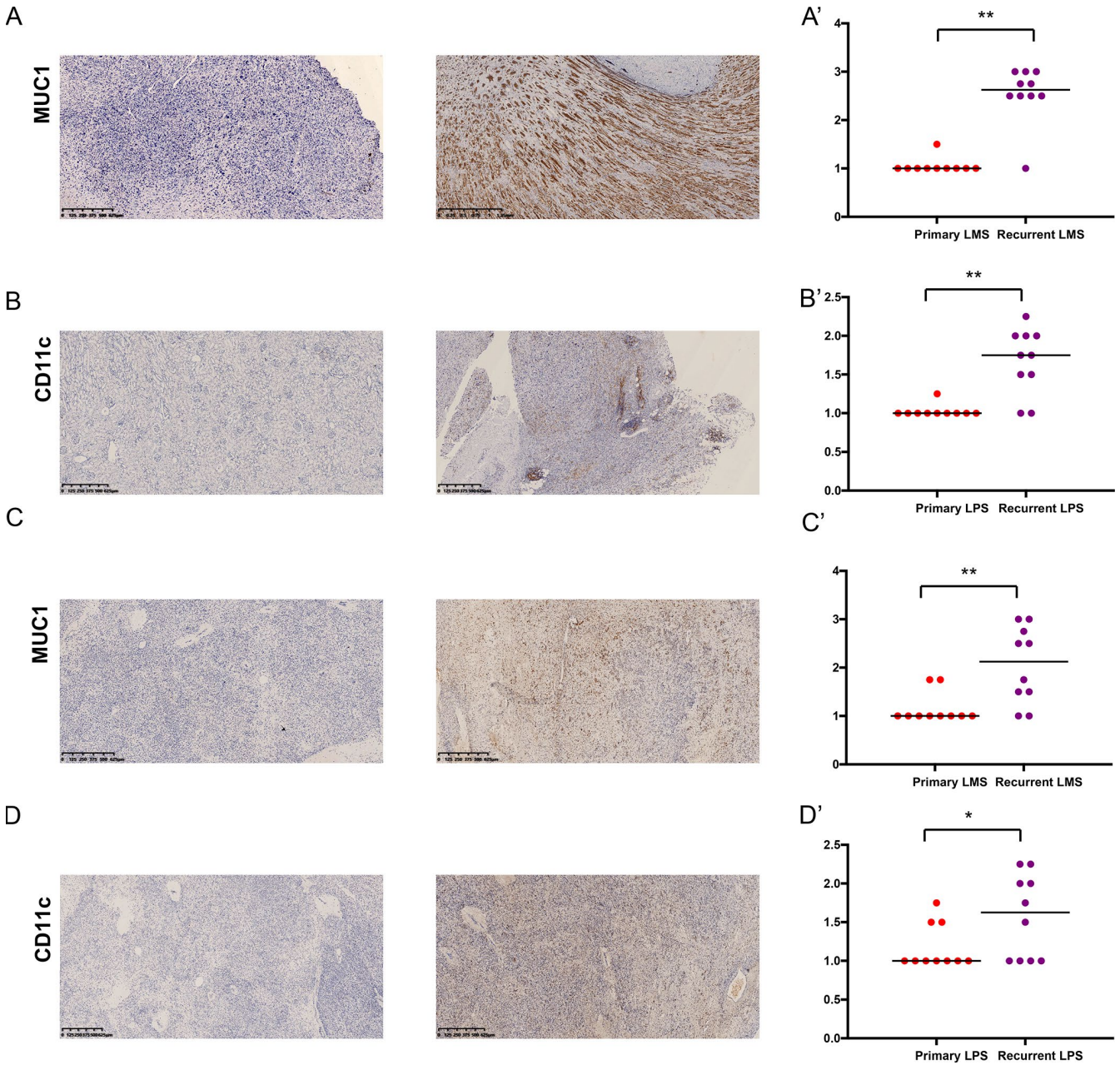


Figure 8. The expressions of MUC1 and CD11c proteins in primary/ recurrent leiomyosarcoma (LMS) (**A**, **B**) and liposarcoma (LPS) (**C**, **D**) specimens examined by immunohistochemistry (IHC) assay. The upper one is primary STS and under one is recurrent STS.

Table 5. Summary of multidimensional external validation results base on multiple databases.

Database	MUC1/MUC1		ITGAX/CD11c		Results
	Cancer	Normal	Cancer	Normal	
Oncomine	↑	↓	-	-	Across ten analysis, MUC1 was highly expressed in primary STS compared to normal tissue while CD11c (Gene symbol: ITGAX) showed no difference (Supplementary Figure 5).
CCLE	↑	NA	-	NA	At the cellular level, MUC1 was expressed in various STS cell lines while ITGAX expression was low STS cell lines (Supplementary Figure 6).
cBioPortal	↑	↓	-	-	MUC1 and ITGAX were highly expressed in primary STS compared to some other types of cancer and they had a significant co-expression pattern in STS (Supplementary Figure 7).
GTEx	NA	↓	NA	↓	MUC1 and ITGAX were lowly expressed in normal adipose tissue and smooth muscle tissue (Supplementary Figure 8).
UCSC Treehouse	↑	↓	-	-	MUC1 was highly expressed in primary STS compared to normal soft tissue while ITGAX did not (Supplementary Figure 8).
The Human Protein Atlas	NA	ND	NA	ND	Protein MUC1 and CD11c were almost not detected in normal adipose tissue and smooth muscle tissue (Supplementary Figure 9).

Note: "↑" was defined as a significantly upregulated gene; "↓" was defined as a significantly downregulated gene; "-" was defined as a a gene with no significant difference in expression; "NA" was defined as "Not available"; "ND" was defined as "Not detected";
Abbreviations: CCLE: Cancer Cell Line Encyclopedia; GTEx: Genotype-Tissue Expression; UCSC: University of California, Santa Cruz

Second, we have not considered the heterogeneity of the immune microenvironment related to the location of immune infiltration. And the heterogeneity of the histological subtypes could affect the accuracy and generalization of the prediction models. Third, all data series downloaded for establishment of the prediction nomograms were from Western countries; thus, caution should be exerted when applying the conclusion of this study to patients from Asian countries. And to minimize bias, multiple databases were used to detect gene and protein expression levels of key biomarkers at the tissue and cellular levels, showing the key biomarkers were significantly upregulated in common sarcoma tissues and cell lines and their proteins were not expressed in normal soft tissues (Supplementary Figure 5–9). Last but not least, this study is only a correlation study on multiple dimensions rather than a biological mechanism study. However, notwithstanding its limitations, this study firstly established the nomograms to predict the survival of STS patients based on recurrent STS-specific tumor-infiltrating immune cells and ceRNA networks and inferred that the mechanism of hsa-miR-1226-3p regulating MUC1 and dendritic cells resting might play an important role in STS recurrence. In the future, more data should be incorporated to improve the model. As our future directions, we would investigate the direct molecular biological mechanisms of the recurrent STS-specific ceRNAs and the intercellular communication between cancer cells and dendritic cells resting.

CONCLUSIONS

Our study constructed two nomograms to predict survival and recurrence of STS patients based on tumor-infiltrating immune cells and ceRNA networks, and demonstrated the utility by their high AUC values. The proposed prediction nomograms might provide much comprehensive clinical information for improving the personalized management of STS patients. Moreover, this study inferred that the mechanism of hsa-miR-1226-3p regulating MUC1 and dendritic cells resting might play an important role in STS recurrence.

MATERIALS AND METHODS

Data collection and differential gene expression analysis

The study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (No. 2019-KY-108). RNA profiles of the primary sarcomas and local recurrence samples were downloaded from

TCGA (https://tcga-data.nci.nih.gov/tcga/) database. These samples were taken from patients confirmed as soft tissue sarcomas by histopathological diagnosis. All patients in this database have a uniform ID. The .Xml (Extensible Markup Language) files containing all the matadata for each sample was downloaded and merged by Practical Extraction and Report Language (Perl) script to determine the grouping of the samples in this study. Both HTseq-count and fragments per kilobase of exon per million reads mapped (FPKM) profiles of 262 samples, comprising 259 primary sarcomas and 3 local recurrence samples were collected (The specimens used for analysis in each experiment were primary/recurrent STS, not primary STS/normal tissue, or primary STS in patients with/without recurrence). Demographic information and survival endpoint of each patient were also retrieved.

After filtering non-sarcoma specific expression genes (No expression was detected in both experimental group and control group), the edgeR method was used to identify differentially expressed mRNAs, lncRNAs, and miRNAs. With a false discovery rate (FDR) P value < 0.05, the log(fold-change) > 1.0 or < -1.0 was defined a downregulated or upregulated gene, respectively.

Construction of the ceRNA network

Before primary statistical analysis, the miRNA–mRNA interaction information based on experimental verification was download from miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/) [41] and while lncRNA–miRNA interaction information was download from lncbase v.2 Experimental Module (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbasev2%2Findex-experimental) [42]. Then, miRNAs regulated both lncRNAs and mRNAs showing significant results in hypergeometric testing and correlation analysis were selected for construction of the ceRNA network using Cytoscape v.3.5.1 [43].

Survival analysis and nomograms of key members in the ceRNA network

Kaplan–Meier survival analysis and Cox proportional hazards model were generated to identify the prognostic value of all biomarkers. All significant biomarkers were integrated into the Cox model and the Lasso regression was performed to ensure that the multifactor models were not overfitting. Eventually, we built a nomogram based on the multivariable models to predict the prognosis of patients with sarcomas. The calibration curves and receiver operating characteristic curves (ROC) were utilized to assess the discrimination and accuracy of the nomogram.

CIBERSORT estimation

In order to further explore the cytological causes of sarcoma-recurrence and molecular mechanism of the vital biomarkers in ceRNA network to some extent, the CIBERSORT algorithm [44] was used to estimate the fraction of 22 immune cell types in the primary and local recurrent sarcomas. Samples with a CIBERSORT output of $P < 0.05$ were considered to be eligible for further analysis. The Wilcoxon rank-sum test was implemented to find the immune cells, which had significant differences in the proportion between recurrent and primary tumors. Besides, Cox regression and Kaplan–Meier method were also applied to assess the relationship between the proportion of immune cells and sarcoma patients' overall survival. Pearson correlation analysis was done for each prognostic biomarker in the ceRNA network and the proportion of each survival related immune cell. Finally, immune cells and biomarkers that were significantly associated with overall survival were incorporated into a nomogram.

Immunohistochemistry (IHC)

Paraffin-embedded, formalin-fixed LMS and LPS specimens were used for IHC. Sections were incubated overnight in a humidified container at 4°C with the primary antibodies of MUC1 (1:100, ab109185; Abcam) and CD11c (1:100, ab52632, Abcam). After three times washing, tissue sections were incubated with the secondary antibody conjugated with streptavidin–horseradish peroxidases for 1 h at room temperature. The slides were stained with 3, 3-diaminobenzidine tetrahydrochloride (DAB) and the nuclei were counterstained with hematoxylin. Immunostaining on each slide was assessed by experienced pathologists to examine the percentage of MUC1 or CD11c positive tumor cells and presented as histochemistry score (H-score). H-score = $\Sigma pi(i+1)$ where i is the intensity score and pi is the percent of the cells with that intensity.

Multidimensional validation

To minimize bias, multiple databases including the Gene Expression Omnibus (GEO) (ID: GSE21050 [45], GSE21122 [45], GSE6481 [46]. These three data sets were used for multidimensional external validation in the online database), Oncomine [47], Cancer Cell Line Encyclopedia (CCLE) [48], cBioPortal for Cancer Genomics [49, 50] Genotype-Tissue Expression (GTEx) [51], UCSC Treehouse Childhood Cancer Initiative [52], The Human Protein Atlas [53], CellMarker [55] were used to detect gene and protein expression levels of key biomarkers at the tissue and cellular levels.

Statistical analysis

Only two-sided P value < 0.05 was thought to be statistical significance. All statistical analyses were enforced with R version 3.5.1 software (Institute for Statistics and Mathematics, Vienna, Austria; https://www.r-project.org) (Package: GDCRNATools [54], edgeR, ggplot2, rms, glmnet, preprocessCore, survminer, timeROC).

Abbreviations

STS: Soft tissue sarcoma; AUC: Area under curve; ceRNA: competitive endogenous RNA; mRNA: messenger RNA; lncRNA: long non-coding RNA; miRNA: microRNA; CIBERSORT: Cell type identification by estimating relative subsets of RNA transcripts; IHC: Immunohistochemistry; MUC1: Mucin 1; LR: Local recurrence; TCGA: The Cancer Genome Atlas; FPKM: Fragments per kilobase of exon per million reads mapped; FDR: False discovery rate; SD: Standard deviation; ROC: Receiver operating characteristic curves; DAB: 3: 3-diaminobenzidine tetrahydrochloride; DCs: Dendritic cells; APC: Antigen-presenting cells; pDCs: plasmacytoid DCs; 3' UTR: 3' untranslated region; CTGF: Connective tissue growth factor; VEGF-A: Vascular endothelial growth factor-A; PDGF-A: Platelet-derived growth factor A; EMT: Epithelial to mesenchymal transition.

AUTHOR CONTRIBUTIONS

Sheng Zhong prepared and compiled the draft for initial review and incorporated all suggested edits into the final draft. Yang Bai, Bo Wu, and Junliang Ge completed an initial review and provided significant edits and additional content before review and approval of other authors. All other authors reviewed, suggested edits, and approved the final manuscript.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

FUNDING

This study was supported in part by the National Natural Science Foundation of China (Grant No. 81702659; 81772856; 81501203). Youth Fund of Shanghai Municipal Health Planning Commission (No.2017YQ054); Henan medical science and technology research project (Grant No. 201602031).

REFERENCES

1. 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

2. Hoefkens F, Dehandschutter C, Somville J, Meijnders P, Van Gestel D. Soft tissue sarcoma of the extremities: pending questions on surgery and radiotherapy. Radiat Oncol. 2016; 11:136. https://doi.org/10.1186/s13014-016-0668-9 PMID:27733179

3. 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/jnccn.2018.0025 PMID:29752328

4. Honoré C, Méeus P, Stoeckle E, Bonvalot S. Soft tissue sarcoma in France in 2015: Epidemiology, classification and organization of clinical care. J Visc Surg. 2015; 152:223–30. https://doi.org/10.1016/j.jviscsurg.2015.05.001 PMID:26088366

5. Scheer M, Dantonello T, Hallmen E, Blank B, Sparber-Sauer M, Vokuhl C, Leuschner I, Münter MW, von Kalle T, Bielack SS, Klingebiel T, Koscielniak E, and Cooperative Weichteilsarkom Studiengruppe [CWS]. Synovial Sarcoma Recurrence in Children and Young Adults. Ann Surg Oncol. 2016 (Suppl 5); 23:618–26. https://doi.org/10.1245/s10434-016-5535-2 PMID:27638676

6. De Angelis F, Guy F, Bertaut A, Méjean N, Varbedian O, Hervieu A, Truc G, Thibouw D, Barra CC, Fraisse J, Burnier P, Isambert N, Causeret S. Limbs and trunk soft tissue sarcoma systematic local and remote monitoring by MRI and thoraco-abdomino-pelvic scanner: A single-centre retrospective study. Eur J Surg Oncol. 2019; 45:1274–80. https://doi.org/10.1016/j.ejso.2019.02.002 PMID:30765271

7. Trans-Atlantic RPS Working Group. Management of Recurrent Retroperitoneal Sarcoma (RPS) in the Adult: A Consensus Approach from the Trans-Atlantic RPS Working Group. Ann Surg Oncol. 2016; 23: 3531–40. https://doi.org/10.1245/s10434-016-5336-7 PMID:27480354

8. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, McGraw T, Mittal V. The lung microenvironment: an important regulator of tumour growth and metastasis. Nat Rev Cancer. 2019; 19:9–31. https://doi.org/10.1038/s41568-018-0081-9 PMID:30532012

9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–74. https://doi.org/10.1016/j.cell.2011.02.013 PMID:21376230

10. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoué F, Bruneval P, Cugnenc PH, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006; 313:1960–64. https://doi.org/10.1126/science.1129139 PMID:17008531

11. Geiger J, Hutchinson R, Hohenkirk L, McKenna E, Chang A, Mulé J. Treatment of solid tumours in children with tumour-lysate-pulsed dendritic cells. Lancet. 2000; 356:1163–65. https://doi.org/10.1016/S0140-6736(00)02762-8 PMID:11030299

12. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011; 146:353–58. https://doi.org/10.1016/j.cell.2011.07.014 PMID:21802130

13. Rodina A, Wang T, Yan P, Gomes ED, Dunphy MP, Pillarsetty N, Koren J, Gerecitano JF, Taldone T, Zong H, Caldas-Lopes E, Alpaugh M, Corben A, et al. The epichaperome is an integrated chaperome network that facilitates tumour survival. Nature. 2016; 538:397–401. https://doi.org/10.1038/nature19807 PMID:27706135

14. Tran Janco JM, Lamichhane P, Karyampudi L, Knutson KL. Tumor-infiltrating dendritic cells in cancer pathogenesis. J Immunol. 2015; 194:2985–91. https://doi.org/10.4049/jimmunol.1403134 PMID:25795789

15. Le Gall CM, Weiden J, Eggermont LJ, Figdor CG. Dendritic cells in cancer immunotherapy. Nat Mater. 2018; 17:474–75. https://doi.org/10.1038/s41563-018-0093-6 PMID:29795222

16. Tchekneva EE, Goruganthu MU, Uzhachenko RV, Thomas PL, Antonucci A, Chekneva I, Koenig M, Piao L, Akhter A, de Aquino MT, Ranganathan P, Long N, Magliery T, et al. Determinant roles of dendritic cell-expressed Notch Delta-like and Jagged ligands on anti-tumor T cell immunity. J Immunother Cancer. 2019; 7:95. https://doi.org/10.1186/s40425-019-0566-4 PMID:30940183

17. Huang Q, Jiang Z, Meng T, Yin H, Wang J, Wan W, Cheng M, Yan W, Liu T, Song D, Chen H, Wu Z, Xu W, et al. MiR-30a inhibits osteolysis by targeting RunX2 in giant cell tumor of bone. Biochem Biophys Res Commun. 2014; 453:160–65. https://doi.org/10.1016/j.bbrc.2014.09.076 PMID:25264196

18. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov. 2017; 16:203–22. https://doi.org/10.1038/nrd.2016.246 PMID:28209991

19. Bhan A, Soleimani M, Mandal SS. Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res. 2017; 77:3965–81. https://doi.org/10.1158/0008-5472.CAN-16-2634 PMID:28701486

20. Qi X, Lin Y, Chen J, Shen B. Decoding competing endogenous RNA networks for cancer biomarker discovery. Brief Bioinform. 2019. [Epub ahead of print]. https://doi.org/10.1093/bib/bbz006 PMID:30715152

21. Martens-Uzunova ES, Böttcher R, Croce CM, Jenster G, Visakorpi T, Calin GA. Long noncoding RNA in prostate, bladder, and kidney cancer. Eur Urol. 2014; 65:1140–51. https://doi.org/10.1016/j.eururo.2013.12.003 PMID:24373479

22. Zhao J, Li L, Han ZY, Wang ZX, Qin LX. Long noncoding RNAs, emerging and versatile regulators of tumor-induced angiogenesis. Am J Cancer Res. 2019; 9:1367–81. PMID:31392075

23. Zhou Q, Zeng H, Ye P, Shi Y, Guo J, Long X. Differential microRNA profiles between fulvestrant-resistant and tamoxifen-resistant human breast cancer cells. Anticancer Drugs. 2018; 29:539–48. https://doi.org/10.1097/CAD.0000000000000623 PMID:29557813

24. Zheng L, Li X, Gu Y, Lv X, Xi T. The 3'UTR of the pseudogene CYP4Z2P promotes tumor angiogenesis in breast cancer by acting as a ceRNA for CYP4Z1. Breast Cancer Res Treat. 2015; 150:105–18. https://doi.org/10.1007/s10549-015-3298-2 PMID:25701119

25. Jin C, Rajabi H, Kufe D. miR-1226 targets expression of the mucin 1 oncoprotein and induces cell death. Int J Oncol. 2010; 37:61–69. https://doi.org/10.3892/ijo_00000653 PMID:20514397

26. Hanisch FG, Müller S. MUC1: the polymorphic appearance of a human mucin. Glycobiology. 2000; 10:439–49. https://doi.org/10.1093/glycob/10.5.439 PMID:10764832

27. Zhang H, Ji J, Liu Q, Xu S. MUC1 downregulation promotes TNF-α-induced necroptosis in human bronchial epithelial cells via regulation of the RIPK1/RIPK3 pathway. J Cell Physiol. 2019; 234:15080–88. https://doi.org/10.1002/jcp.28148 PMID:30666647

28. de Boer HR, Pool M, Joosten E, Everts M, Samplonius DF, Helfrich W, Groen HJ, van Cooten S, Fusetti F, Fehrmann RS, de Vries EG, van Vugt MA. Quantitative proteomics analysis identifies MUC1 as an effect sensor of EGFR inhibition. Oncogene. 2019; 38:1477–88. https://doi.org/10.1038/s41388-018-0522-7 PMID:30305724

29. Nabavinia MS, Gholoobi A, Charbgoo F, Nabavinia M, Ramezani M, Abnous K. Anti-MUC1 aptamer: A potential opportunity for cancer treatment. Med Res Rev. 2017; 37:1518–39. https://doi.org/10.1002/med.21462 PMID:28759115

30. Nath S, Mukherjee P. MUC1: a multifaceted oncoprotein with a key role in cancer progression. Trends Mol Med. 2014; 20:332–42. https://doi.org/10.1016/j.molmed.2014.02.007 PMID:24667139

31. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest. 2009; 119: 1420–28. https://doi.org/10.1172/JCI39104 PMID:19487818

32. Gendler SJ. MUC1, the renaissance molecule. J Mammary Gland Biol Neoplasia. 2001; 6:339–53. https://doi.org/10.1023/A:1011379725811 PMID:11547902

33. Adil Butt M, Pye H, Haidry RJ, Oukrif D, Khan SU, Puccio I, Gandy M, Reinert HW, Bloom E, Rashid M, Yahioglu G, Deonarain MP, Hamoudi R, et al. Upregulation of mucin glycoprotein MUC1 in the progression to esophageal adenocarcinoma and therapeutic potential with a targeted photoactive antibody-drug conjugate. Oncotarget. 2017; 8:25080–96. https://doi.org/10.18632/oncotarget.15340 PMID:28212575

34. Glaffig M, Stergiou N, Hartmann S, Schmitt E, Kunz H. A Synthetic MUC1 Anticancer Vaccine Containing Mannose Ligands for Targeting Macrophages and Dendritic Cells. ChemMedChem. 2018; 13:25–29. https://doi.org/10.1002/cmdc.201700646 PMID:29193802

35. Dionisi M, De Archangelis C, Battisti F, Rahimi Koshkaki H, Belleudi F, Zizzari IG, Ruscito I, Albano C, Di Filippo A, Torrisi MR, Benedetti Panici P, Napoletano C, Nuti M, Rughetti A. Tumor-Derived Microvesicles Enhance Cross-Processing Ability of Clinical Grade Dendritic Cells. Front Immunol. 2018; 9:2481. https://doi.org/10.3389/fimmu.2018.02481 PMID:30455687

36. Di Mauro S, Ragusa M, Urbano F, Filippello A, Di Pino A, Scamporrino A, Pulvirenti A, Ferro A, Rabuazzo AM, Purrello M, Purrello F, Piro S. Intracellular and extracellular miRNome deregulation in cellular models of NAFLD or NASH: clinical implications. Nutr Metab Cardiovasc Dis. 2016; 26:1129–39. https://doi.org/10.1016/j.numecd.2016.08.004 PMID:27756518

37. Leppert U, Henke W, Huang X, Müller JM, Dubiel W. Post-transcriptional fine-tuning of COP9 signalosome subunit biosynthesis is regulated by the c-Myc/Lin28B/let-7 pathway. J Mol Biol. 2011; 409:710–21. https://doi.org/10.1016/j.jmb.2011.04.041 PMID:21530537

38. Muthumani K, Choo AY, Premkumar A, Hwang DS, Thieu KP, Desai BM, Weiner DB. Human immunodeficiency virus type 1 (HIV-1) Vpr-regulated cell death: insights into mechanism. Cell Death Differ. 2005 (Suppl 1); 12:962–70. https://doi.org/10.1038/sj.cdd.4401583 PMID:15832179

39. Mahalingam S, Ayyavoo V, Patel M, Kieber-Emmons T, Kao GD, Muschel RJ, Weiner DB. HIV-1 Vpr interacts with a human 34-kDa mov34 homologue, a cellular factor linked to the G2/M phase transition of the mammalian cell cycle. Proc Natl Acad Sci USA. 1998; 95:3419–24. https://doi.org/10.1073/pnas.95.7.3419 PMID:9520381

40. Ramanathan MP, Curley E 3rd, Su M, Chambers JA, Weiner DB. Carboxyl terminus of hVIP/mov34 is critical for HIV-1-Vpr interaction and glucocorticoid-mediated signaling. J Biol Chem. 2002; 277:47854–60. https://doi.org/10.1074/jbc.M203905200 PMID:12237292

41. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. 2018; 46:D296–302. https://doi.org/10.1093/nar/gkx1067 PMID:29126174

42. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, Zagganas K, Tsanakas P, Floros E, Dalamagas T, Hatzigeorgiou AG.

DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic Acids Res. 2016; 44:D231–38. https://doi.org/10.1093/nar/gkv1270 PMID:26612864

43. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504. https://doi.org/10.1101/gr.1239303 PMID:14597658

44. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015; 12:453–57. https://doi.org/10.1038/nmeth.3337 PMID:25822800

45. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, Shah K, Socci 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

46. Nakayama R, Nemoto T, Takahashi H, Ohta T, Kawai A, Seki K, Yoshida T, Toyama Y, Ichikawa H, Hasegawa T. Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol. 2007; 20:749–59. https://doi.org/10.1038/modpathol.3800794 PMID:17464315

47. Elfilali A, Lair S, Verbeke C, La Rosa P, Radvanyi F, Barillot E. ITTACA: a new database for integrated tumor transcriptome array and clinical data analysis. Nucleic Acids Res. 2006; 34:D613–16. https://doi.org/10.1093/nar/gkj022 PMID:16381943

48. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, Barretina J, Gelfand ET, Bielski CM, Li H, Hu K, Andreev-Drakhlin AY, Kim J, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019; 569:503–08. https://doi.org/10.1038/s41586-019-1186-3 PMID:31068700

49. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6:pl1. https://doi.org/10.1126/scisignal.2004088 PMID:23550210

50. 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

51. Consortium G, 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

52. Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, Haussler D, Zhu J. The UCSC Cancer Genomics Browser: update 2015. Nucleic Acids Res. 2015; 43:D812–17. https://doi.org/10.1093/nar/gku1073 PMID:25392408

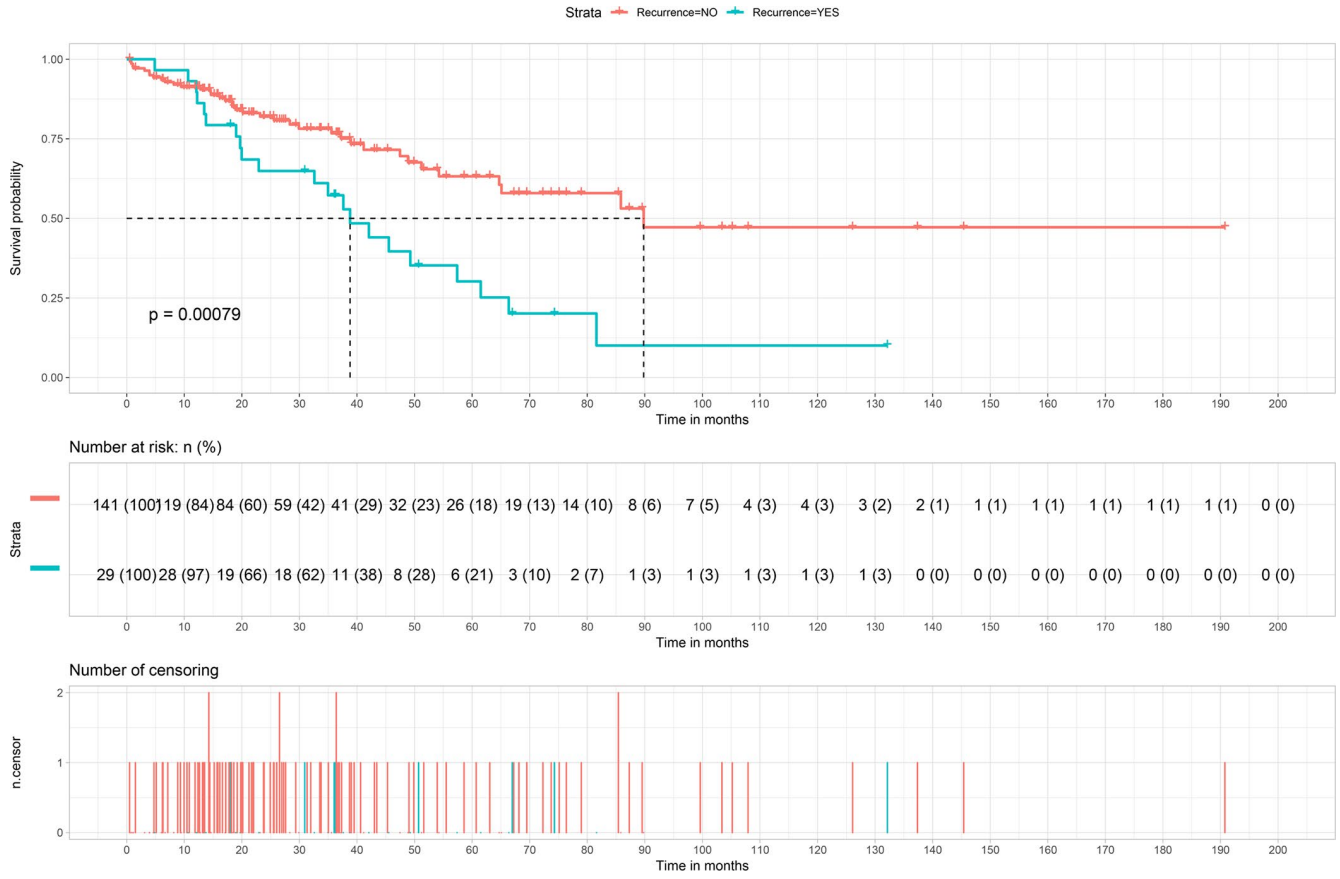
53. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015; 347:1260419. https://doi.org/10.1126/science.1260419 PMID:25613900

54. Li R, Qu H, Wang S, Wei J, Zhang L, Ma R, Lu J, Zhu J, Zhong WD, Jia Z. GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. Bioinformatics. 2018; 34:2515–17. https://doi.org/10.1093/bioinformatics/bty124 PMID:29509844

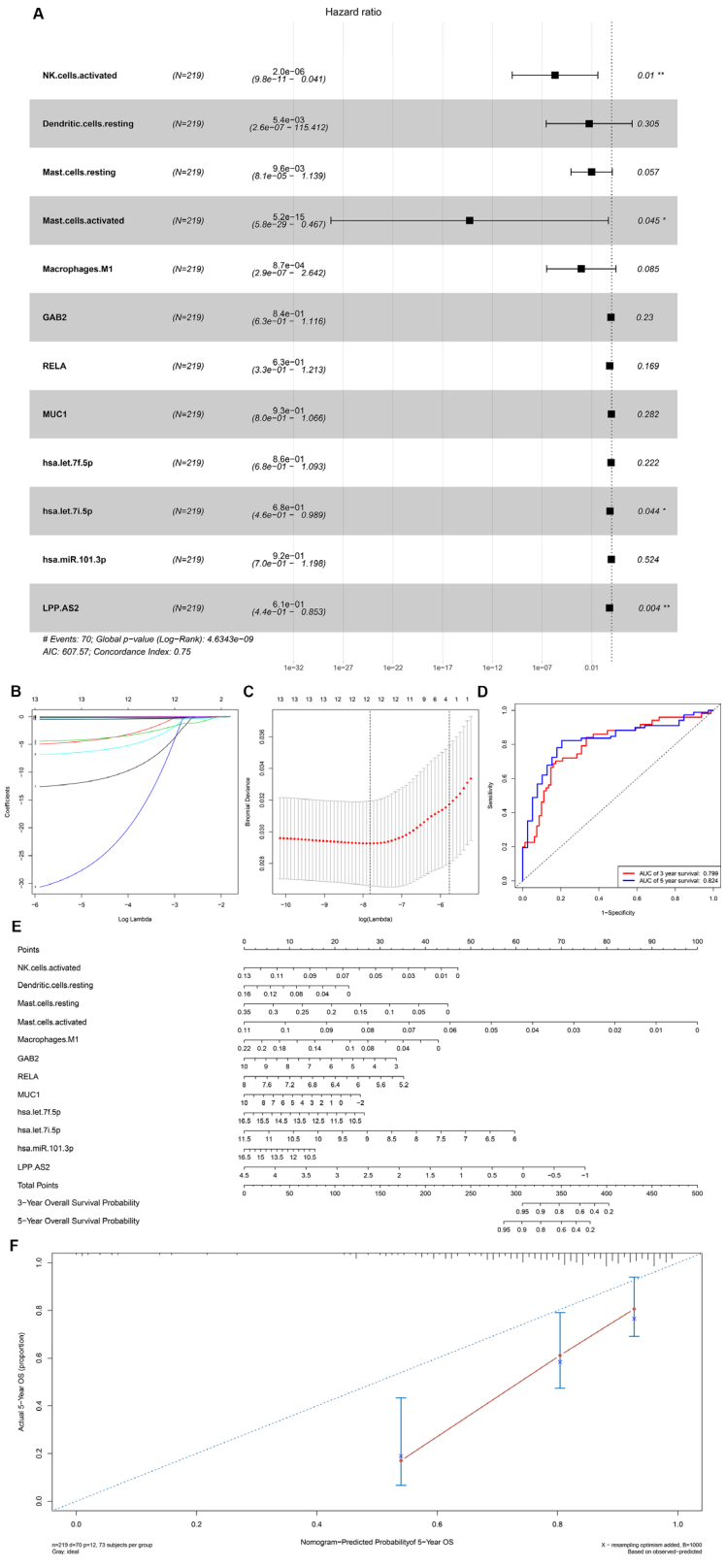
55. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. 2019; 47:D721–28. https://doi.org/10.1093/nar/gky900 PMID:30289549

SUPPLEMENTARY MATERIALS

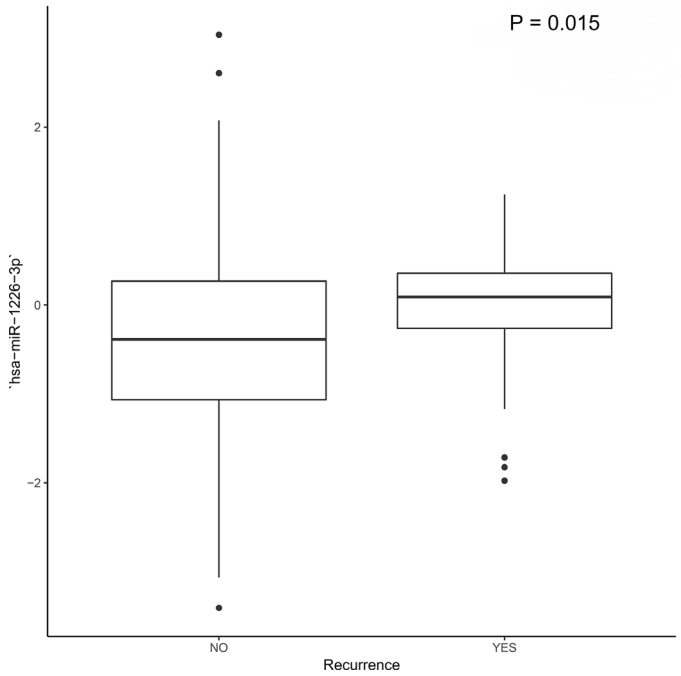
Supplementary Figures



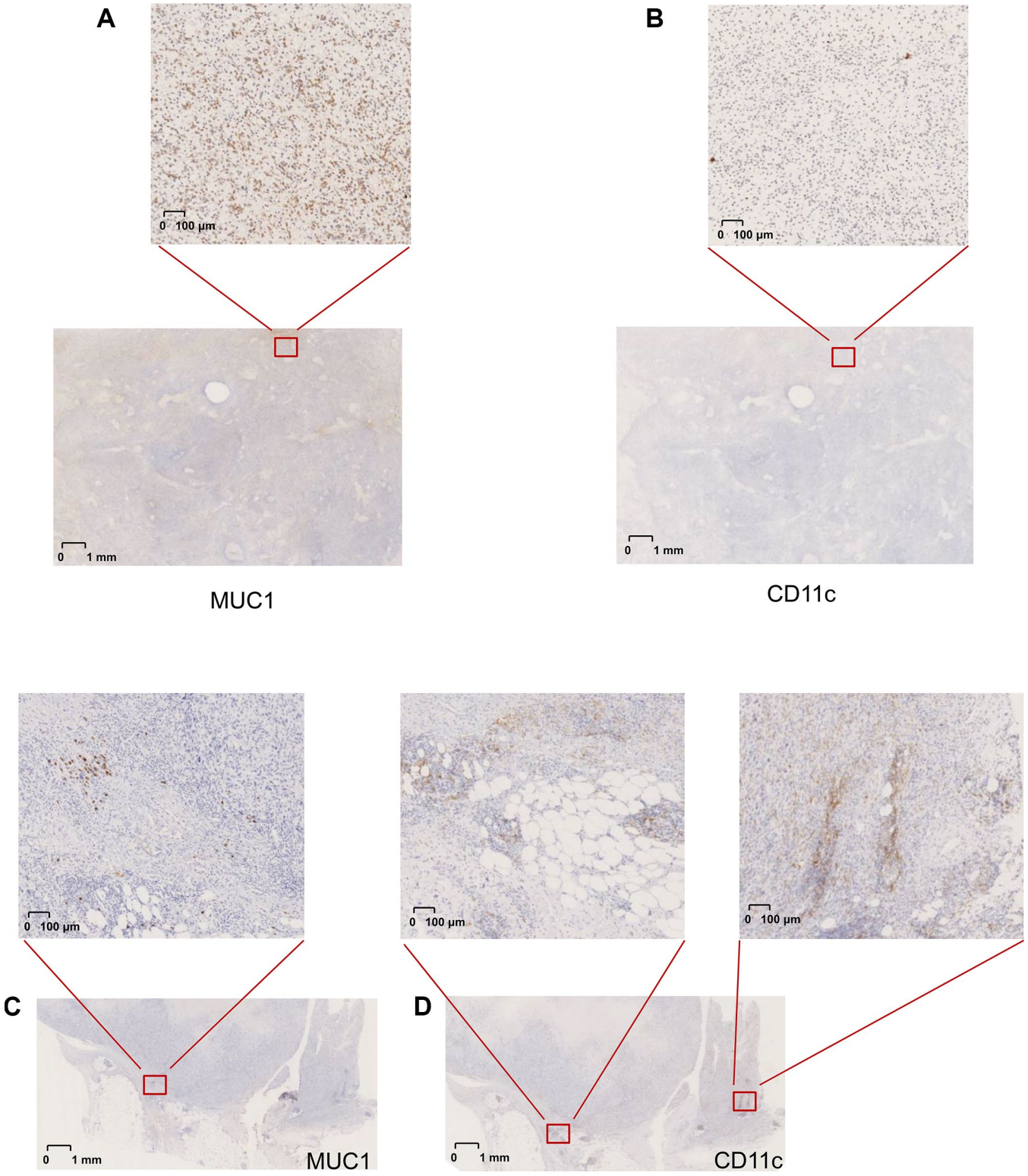
Supplementary Figure 1. **The Kaplan–Meier survival curve revealed that recurrence was a significant risk indicator for poor prognosis of STSs (P = 0.001).** Why is there a higher survival probability in first 10 months in those cases with STS recurrence? In general, solid tumor recurrence refers that the tumor reappears during the follow-up of patients after the primary operation. So, what this means is that patients who have had a recurrence of soft tissue sarcomas have had the primary operation. By reanalysis of the data, we found that most of the patients with survival time less than 10 months had tumor-bearing status (survive with tumor). Therefore, we speculated that these patients might have no chance to perform the primary operation, leading to poor prognosis. Tumor recurrence cannot happen to a patient who do not have the primary operation. Hence the phenomenon.



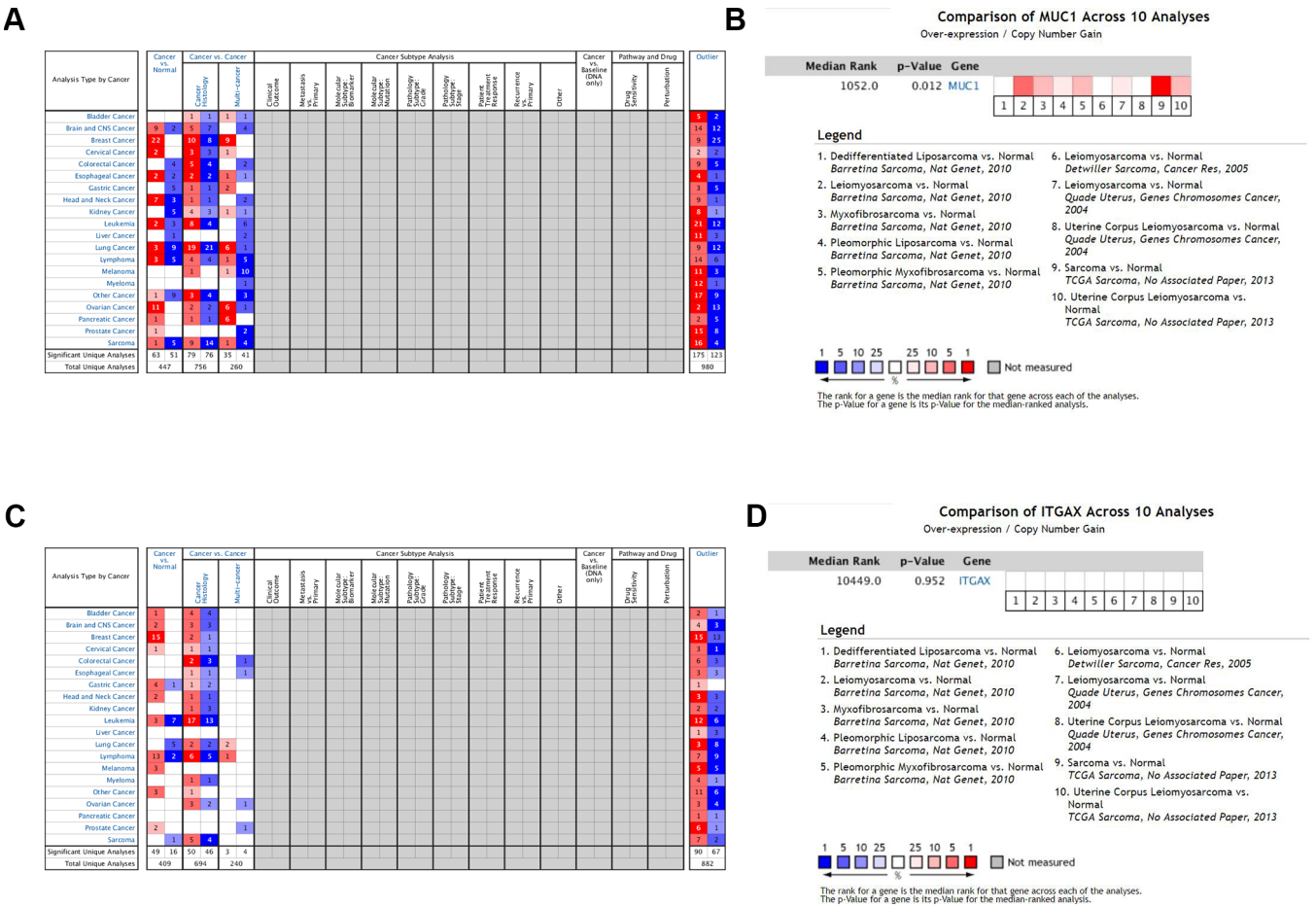
Supplementary Figure 2. The results of multivariable model and the nomogram integrating immune cells and biomarkers significantly associated with overall survival. Key members of the ceRNA network and immune cells were integrated into one Cox regression model (**A**) After the screening process of the Lasso regression, the results suggested that the model was not overfitting (**B**, **C**). The calibration curve and the ROC demonstrated good discrimination and concordance of the multivariable model (AUC of 3-year survival: 0.799; AUC of 5-year survival: 0.824) (**D**, **F**). The nomogram was constructed based on the multivariable model (**E**).



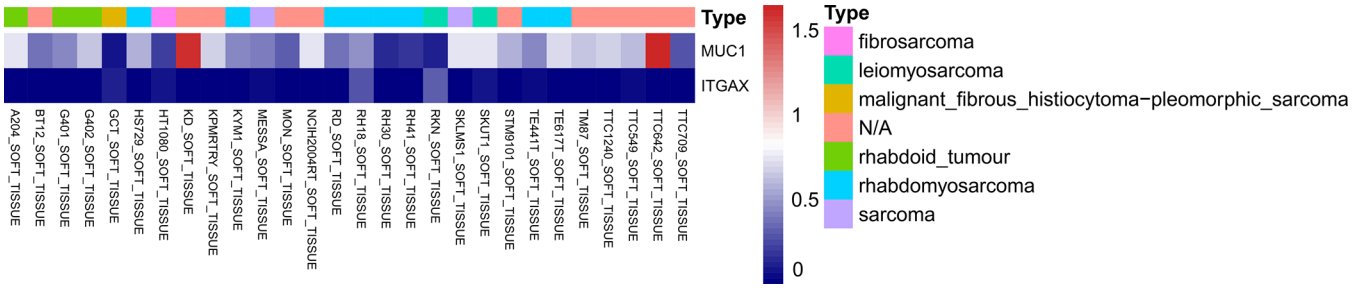
Supplementary Figure 3. The result of the Wilcoxon rank-sum test suggesting that the hsa-miR-1226-3p expression level is significant differences between the primary sarcoma tissues of patients with and without recurrence (P = 0.015).



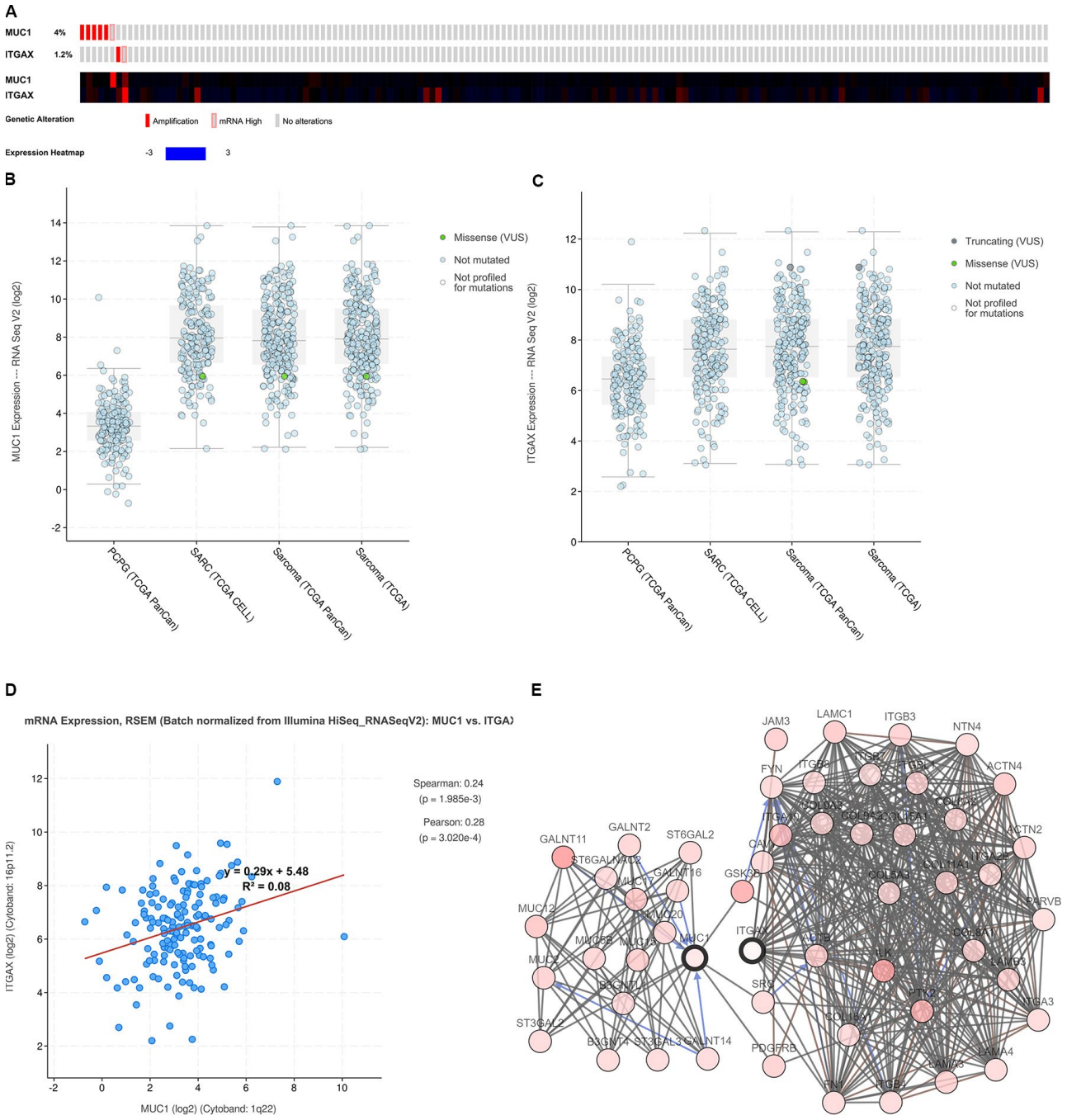
Supplementary Figure 4. The results of the immunohistochemistry (IHC) stain showing that MUC1 and CD11c were correlated with STS recurrence and predominantly localized in the membrane and extracellular matrix of leiomyosarcoma and liposarcoma cells. MUC1 (**A**) and CD11c (**B**) protein was predominantly localized in the membrane and extracellular matrix of LMS. MUC1 (**C**) and CD11c (**D**) protein was predominantly localized in the membrane and extracellular matrix of LPS. Abbreviations: LMS: leiomyosarcoma; LPS: liposarcoma.



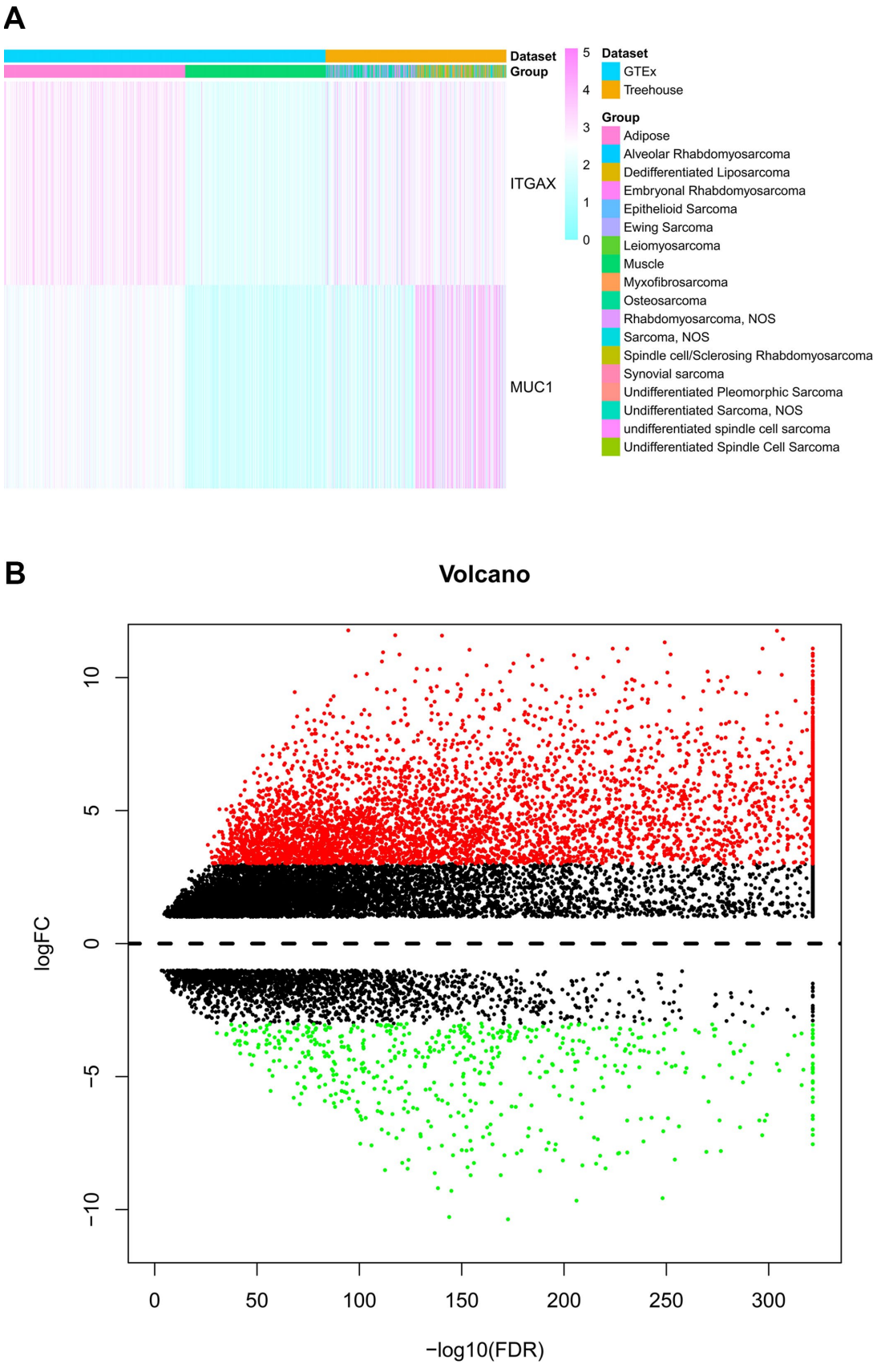
Supplementary Figure 5. Validation of MUC1 (**A**, **B**) and CD11c (**C**, **D**) on a transcriptional level in multiple cancer types and multiple studies using the Oncomine database.



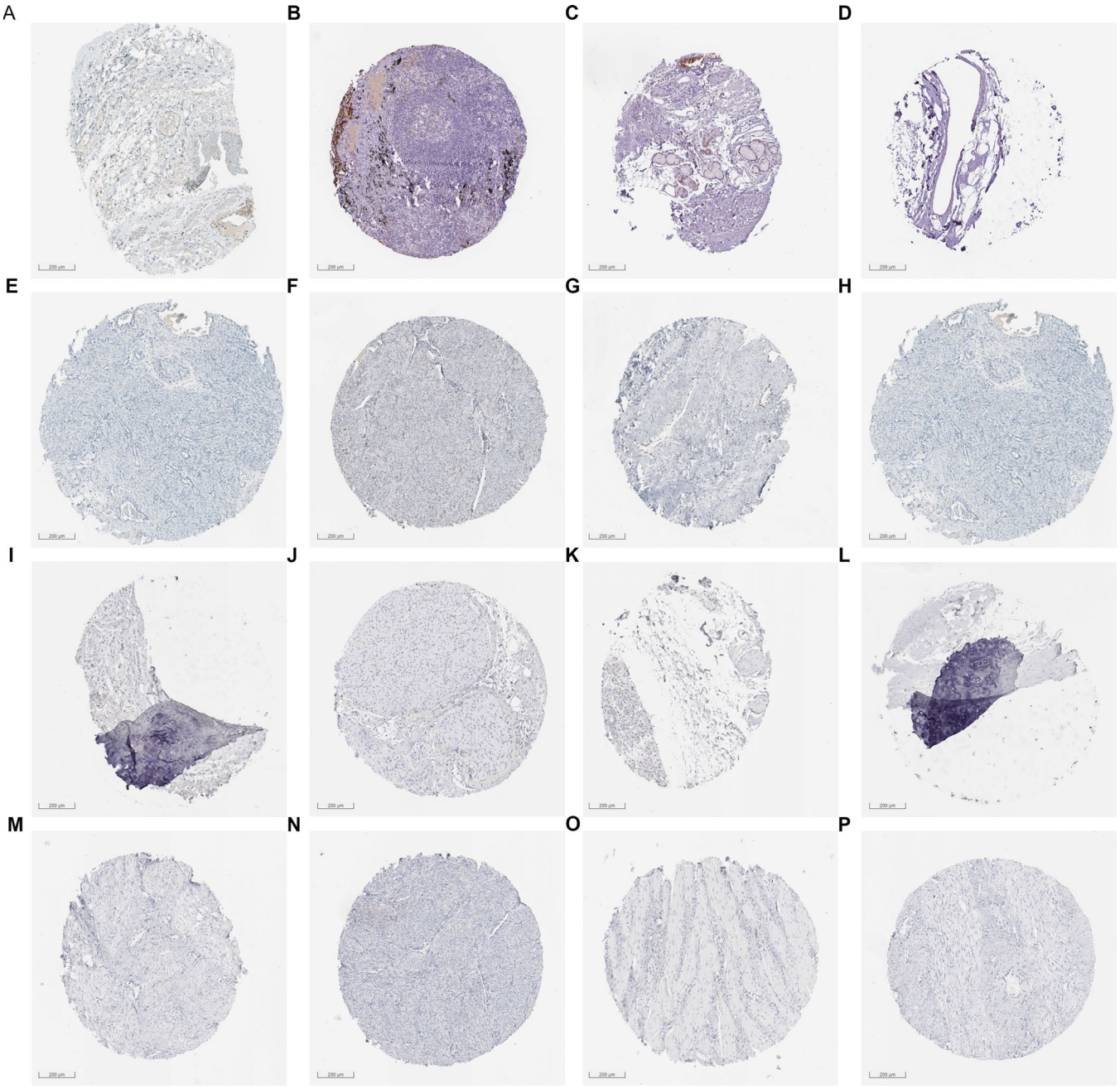
Supplementary Figure 6. The expression levels of MUC1 and CD11c in various soft tumor cell lines in Cancer Cell Line Encyclopedia (CCLE).



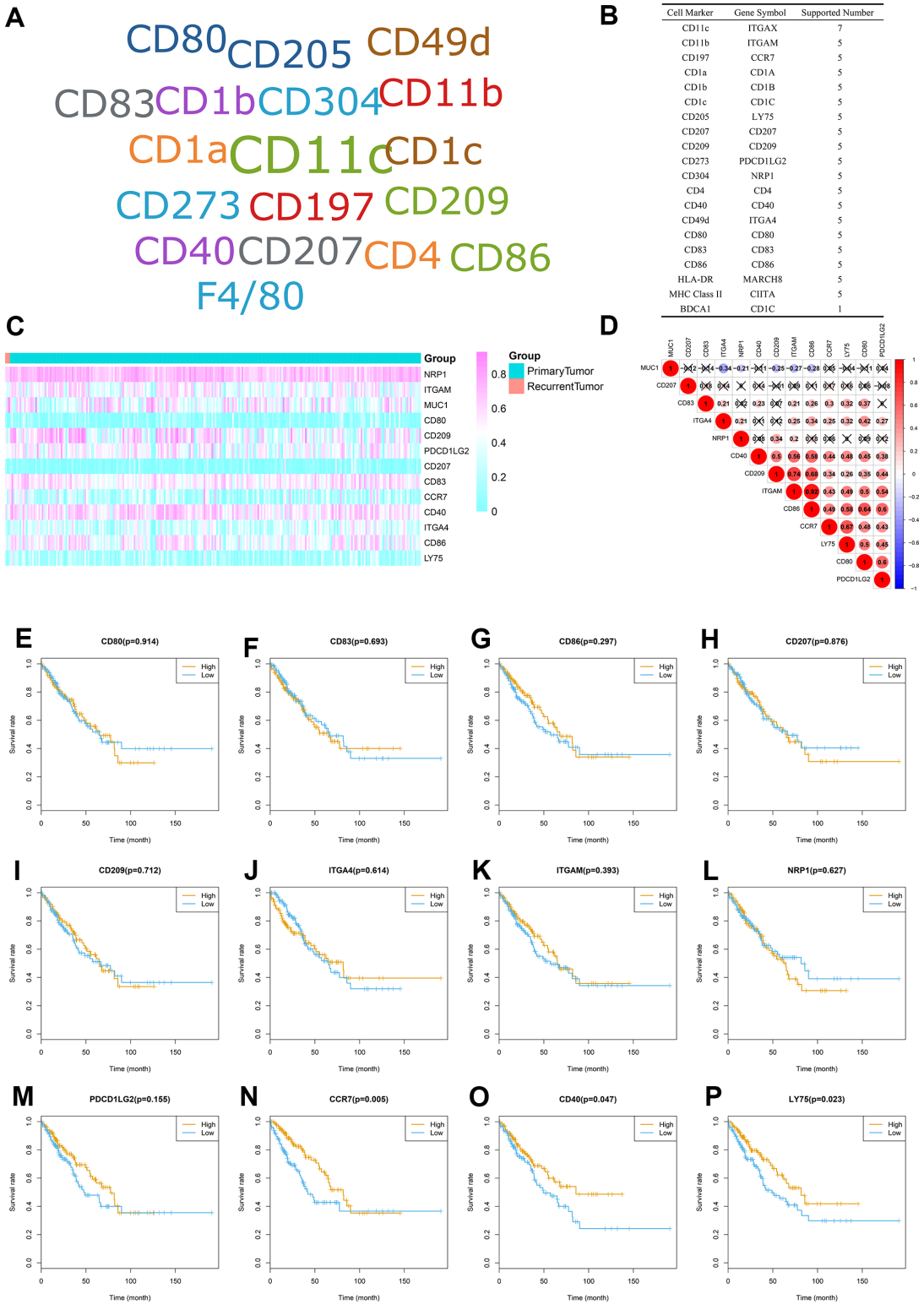
Supplementary Figure 7. **Integrative analysis of genomics and clinical profiles using the cBioPortal database.** (**A**) Alteration frequency of MUC1 and CD11c; (**B**, **C**) MUC1 and CD11c were highly expressed in primary STS compared to some other types of cancer; (**D**) The co-expression between MUC1 and CD11c. (**E**) The Protein-Protein Interaction (PPI) network of MUC1 and CD11c.



Supplementary Figure 8. Differential gene analysis heatmap of (**A**) MUC1 and CD11c and volcano plot (**B**) using data from the GTEx database and the Treehouse database.



Supplementary Figure 9. Validation of MUC1 (Adipose: **A**–**D**; Smooth muscle: **E**–**H**) and CD11c (Adipose: **I**–**L**; Smooth muscle: **M**–**P**) on a translational level using the Human Protein Atlas database. The results of data mining of The Human Protein Atlas showed that the protein MUC1 and CD11c were almost not detected in normal adipose tissue and smooth muscle tissue.



Cell Marker	Gene Symbol	Supported Number
CD11c	ITGAX	7
CD11b	ITGAM	5
CD197	CCR7	5
CD1a	CD1A	5
CD1b	CD1B	5
CD1c	CD1C	5
CD205	LY75	5
CD207	CD207	5
CD209	CD209	5
CD273	PDCD1LG2	5
CD304	NRP1	5
CD4	CD4	5
CD40	CD40	5
CD49d	ITGA4	5
CD80	CD80	5
CD83	CD83	5
CD86	CD86	5
HLA-DR	MARCH8	5
MHC Class II	CIITA	5
BDCA1	CD1C	1

Supplementary Figure 10. Evaluation prognostic value and relationship with MUC1 of all specific surface markers of dendritic cell. On the basis of detecting CD11c, we determined all surface markers of dendritic cell reported for more than 5 times in previous studies by CellMarker database. After removing nonspecific surface markers, CD11b, CD197, CD205, CD207, CD209, CD273, CD304, CD40, CD49d, CD80, CD83, CD86 were integrated into further validation. (**A**, **B**) All surface markers of dendritic cell in CellMarker database; (**C**) Heatmap of twelve markers and MUC1; (**D**) Co-expression heatmap of twelve markers and MUC1; (**E**–**P**) Kaplan–Meier survival curves of twelve markers.

Supplementary Tables

Supplementary Table 1. Baseline information of 261 patients diagnosed with soft tissue sarcoma.

Variables	Total Patients (N = 261)
Age, years	
Mean ± SD	60.87 ± 14.62
Gender	
Female	142 (54.41%)
Male	119 (45.59%)
Race	
Asian	6 (2.30%)
Black or African American	18 (6.90%)
White	228 (87.36%)
Unknown	9 (3.44%)
Histological subtype	
Dedifferentiated Liposarcoma	59 (22.61%)
Leiomyosarcoma	105 (40.23%)
Undifferentiated Pleomorphic Sarcoma	21 (8.05%)
Malignant Peripheral Nerve Sheath Tumors	9 (3.44%)
Myxofibrosarcoma	25 (9.58%)
Undifferentiated Pleomorphic Sarcoma	29 (11.11%)
Synovial Sarcoma	10 (3.83%)
Desmoid Tumor	2 (0.77%)
Undifferentiated Pleomorphic Sarcoma With Giant Cells	1 (0.38%)
Recurrence	
Yes	29 (11.11%)
No	144 (55.17%)
Unknown	88 (33.72%)

Abbreviations: SD: Standard deviation.

Supplementary Table 2. The list of top 10 downregulated and top 10 upregulated genes in differential gene analysis.

Gene	Type	LogFC	P value	FDR
CEND1	protein coding	-4.28656	2.09E-07	0.000116
CHRDL2	protein coding	-4.09075	7.90E-05	0.010565
PGR	protein coding	-4.0102	5.88E-05	0.008329
S100A3	protein coding	-3.96647	9.09E-11	1.88E-07
NT5M	protein coding	-3.96283	2.10E-14	1.52E-10
PODXL	protein coding	-3.90095	1.32E-20	1.91E-16
ESR1	protein coding	-3.60849	1.85E-05	0.003478
CAMK2N2	protein coding	-3.47569	1.73E-07	0.000104
AC105277.1	long non coding	-3.4071	2.01E-09	2.90E-06
RHOBTB3	protein coding	3.825303	0.000402	0.035656
PDGFRB	protein coding	3.196464	0.000375	0.034536
MTSS1	protein coding	2.728508	0.00025	0.025802
HMGN1	protein coding	1.728704	0.000592	0.048322
SNX14	protein coding	1.340009	8.38E-06	0.001893
RELA	protein coding	-0.78947	0.000562	0.046364
ARMC10	protein coding	-0.94507	0.000287	0.028824
PDCD7	protein coding	-0.95058	0.000204	0.022307
ZNF3	protein coding	-0.96601	0.000162	0.018492
PMPCB	protein coding	-0.96741	2.15E-05	0.003875

Abbreviations: FC: Fold change; FDR: False Discovery Rate.