Prediction of clinical outcome and survival in soft-tissue sarcoma using a ten-lncRNA signature

Rong-quan He[1,*], Qing-jun Wei[2,*], Rui-xue Tang[3], Wen-jie Chen[4], Xia Yang[3], Zhi-gang Peng[1], Xiao-hua Hu[1], Jie Ma[1] and Gang Chen[3]

[1]Department of Medical Oncology, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, P. R. China

[2]Department of Orthopaedics Trauma and Hand Surgery, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, P. R. China

[3]Department of Pathology, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, P. R. China

[4]Department of Thoracic and Cardiovascular Diseases, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, P.R. China

[*]These authors are contributed equally to this work

Correspondence to: Jie Ma, **email:** majie086@163.com
                Gang Chen, **email:** chen_gang_triones@163.com

ABSTRACT

The prognostic value of long non-coding RNAs (lncRNAs) in patients with soft-tissue sarcoma has rarely been unraveled. The aim of the study was to find a lncRNA signature to predict the clinical outcome and survival in soft-tissue sarcoma based on the high-throughput RNA-seq data from The Cancer Genome Atlas (TCGA) database. The lncRNAs which closely correlated with overall survival in 258 soft-tissue sarcoma patients were identified with Cox proportional regression model. Ten lncRNAs, including RP11-560J1.2, AP001432.14, RP4-665J23.1, LINC00680, AC006129.2, RP11-230G5.2, BACH1-IT2, RP11-274B21.9, RP11-504A18.1 and RP11-713P17.3, were selected to calculate a risk score. The risk score could effectively predict patients' outcome, such as the status of mitotic count of tumor cells, person neoplasm cancer and residual tumor. More inspiringly, the risk score generated from the 10-lncRNA signature was an independent prognostic indicator for soft-tissue sarcoma patients. Overall, this 10-lncRNA signature gains the potential as an effective prognostic tool for soft-tissue sarcoma as part of the integrated clinical RNA-seq program.

INTRODUCTION

A carcinoma is a type of cancer arising from epithelial cells. On the contrary, a sarcoma is a malignant tumor originating from all classes of mesenchymal tissues, including cancellous bony, cartilaginous, adipose, muscular, vascular, fibrous or synovial tissues. Hence, the pathology of these sarcomas is particularly varied, which has more than seventy identified subtypes [1–4], Sarcomas are generally divided into two broad categories: osteosarcoma and soft-tissue sarcoma [5, 6].

Unlike carcinoma, sarcomas are rather rare. Basically, sarcomas only account for around 1% of all malignancies in adults and 15% in children. In 2016, a total of 12, 310 cases are estimated to be diagnosed with soft-tissue sarcoma in the United States, with nearly 4, 990 deaths. The lungs are the most common organ for the metastasis of soft-tissue sarcomas, which progresses extremely fast. The tremendous aggressiveness of the soft-tissue sarcomas could partially explain the high mortality rates [3, 7–12].

Though sarcomas are now histopathologically categorized, they have been also molecularly classified into different groups. The molecular advancement has improved the understanding of pathogenesis of soft-tissue sarcomas, as well as provided beneficial targets

for clinical settings including diagnosis and treatments [13–19]. However, the pathogenesis of soft-tissue sarcomas is complex due to multiple molecular events being involved. Nowadays, growing evidence has shown that long non-coding RNAs (lncRNAs) play essential roles in the regulation of multiple cellular processes, including tumorigenesis and metastasis of sarcomas; however, the studies available mostly focused on a single lncRNA in osteosarcoma [20–28]. Furthermore, the clinical role of lncRNAs based on high-throughput data has scarcely been investigated. To the best of our knowledge, only Li et al. [29] examined the contributions of lncRNAs to osteosarcoma with microarray analysis. However, no lncRNAs have been studied in soft-tissue sarcomas. Additionally, no investigation on lncRNAs is based on high-throughput sequencing data of soft-tissue sarcomas. To this end, we, for the first time, investigated the clinical value, especially the prognostic role of a series of lncRNAs based on the high-throughput sequencing data from the cancer genomic atlas (TCGA) with 258 cases of soft-tissue sarcomas.

RESULTS

Identification of lncRNAs related to overall survival (OS)

According to the exclusion criterion, 258 cases of soft-tissue sarcomas were included in the prognosis analysis. The univariate Cox proportional hazards regression method revealed that a total of 50 lncRNAs gained significant prognostic value for soft-tissue sarcomas with P value less than 0.01. Subsequently, multivariate Cox proportional hazards regression analysis was performed to further verify the prognostic roles of these 50 lncRNAs and eventually, a total of 10 lncRNAs (RP11-560J1.2, AP001432.14, RP4-665J23.1, LINC00680, AC006129.2, RP11-230G5.2, BACH1-IT2, RP11-274B21.9, RP11-504A18.1 and RP11-713P17.3) were identified to be independent prognostic indicators of soft-tissue sarcoma. Among these 10 lncRNAs, LINC00680, AC006129.2, RP11-274B21.9 and RP11-713P17.3 were protective biomarkers due to HR being less than one. On the contrary, the other six were risky lncRNAs, since their HRs were all more than one (Table 1).

The risk score generated from the 10-lncRNA signature as an independent indicator to predict soft-tissue sarcoma prognosis

The risk score for predicting the OS of soft-tissue sarcoma prognosis was generated with the formula based on the 10 lncRNAs above according to previous reports [30–32]. The risk score generated from the 10-lncRNA signature = expression of RP11-560J1.2 * 0.582+ expression of AP001432.14 * 0.828 + expression of RP4-665J23.1 * 0.348 + expression of LINC00680 * −0.681 + expression of AC006129.2 * −0.236 + expression of RP11-230G5.2 * 0.495 + expression of BACH1-IT2 * 0.423 + expression of RP11-274B21.9 * −0.557 + expression of RP11-504A18.1 * 0.466 + expression of RP11-713P17.3 * -0.249. Each soft-tissue sarcoma patient had a score based on the formula above and all patients were then divided into two groups: low-risk (n = 129) and high-risk (n = 129) on the basis of the median point of the prognostic risk score (Figure 1A). The survival status of each patient was shown in Figure 1B and a heatmap was drawn to display the expression level of the top 10 lncRNAs for each patient (Figure 1C). More importantly, the risk score could act as an independent indicator for the OS of soft-tissue sarcomas, as the HR was 1.445 (95% CI: 1.321–1.581, $P < 0.001$) assessed by the univariate Cox regression analysis and K-M method (Table 2, Figure 2).

The prognostic role of the risk score was also compared to the classical clinicopathological parameters of soft-tissue sarcoma (Table 2). Among 199 patients with complete clinical data, eight (4.0%) achieved complete response (CR), 64 (32.2%) partial response (PR), 124 (62.3%) stable disease (SD) and three (1.5%) progressive disease (PD). The median age for all patients was 60 years old. The histopathological subtypes of these 258 soft-tissue sarcomas included 103 leiomyosarcomas (LMS) (39.9%), 58 dedifferentiated liposarcomas (22.5%), 51 undifferentiated pleomorphic sarcomas (19.8%), 25 myxofibrosarcomas (9.7%), 10 synovial sarcomas (3.9%), 9 malignant peripheral nerve sheath tumors (MPNST) (3.5%) and 2 desmoid tumors (0.8%). Univariate cox regression analysis of OS showed that several clinicopathological parameters could predict the survival of soft-tissue sarcomas, including the histological subtype, local disease recurrence, metastasis, invasion of contiguous organ, margin status, person neoplasm cancer status, residual tumor and treatment completion success outcome (Table 2, Figure 3). Not surprisingly, the "person neoplasm cancer status" was the most remarkable parameter to predict survival and patients with tumor were 9.532 times more prone to suffer from death than those who were tumor free.

The risk score of the 10-lncRNA-signature related to the progression and treatment outcome of soft-tissue sarcomas

The relationship between the risk score of the 10-lncRNA-signature and different clinical features was also analyzed. The student's t test disclosed that the risk score clearly varied between two groups of different ages (P = 0.048), contiguous organ invaded status (P = 0.026), margin status (P = 0.038), mitotic count (P = 0.001), metastasis status (P = 0.045), person neoplasm cancer status ($P < 0.001$) and residual tumor (P = 0.001, Supplementary Table 1, Figure 4). Furthermore, the ROC curves indicated that the risk score

Table 1: The detailed information of ten prognostic lncRNAs significantly associated with overall survival in 258 sarcoma patients

lncRNA	Ensemble ID	Location	β (Cox)	HR (95%CIs)	P
RP11-560J1.2	ENSG00000271888	Chromosome 6: 15,243,923-15,245,000	0.582	1.789 (1.269,2.522)	0.001
AP001432.14	ENSG00000242553	Chromosome 21: 37,221,419-37,237,744	0.828	2.288 (1.502,3.480)	0.000
RP4-665J23.1	ENSG00000233593	Chromosome 1: 90,782,983-90,851,657	0.348	1.417 (1.019,1.970)	0.038
LINC00680	ENSG00000215190	Chromosome 6: 57,946,074-57,961,501	−0.681	0.506 (0.310,0.827)	0.007
AC006129.2	ENSG00000268027	Chromosome 19: 41,545,192-41,555,462	−0.236	0.790 (0.657,0.949)	0.012
RP11-230G5.2	ENSG00000250748	Chromosome 12: 65,466,820-65,642,372	0.495	1.641 (1.193,2.256)	0.002
BACH1-IT2	ENSG00000228817	Chromosome 21: 29,370,497-29,373,709	0.423	1.527 (1.046,2.228)	0.028
RP11-274B21.9	ENSG00000271344	Chromosome 7: 128,690,451-128,691,717	−0.557	0.573 (0.420,0.781)	< 0.001
RP11-504A18.1	ENSG00000260971	Chromosome 1: 56,248,294-56,258,571	0.466	1.594 (1.126,2.257)	0.009
RP11-713P17.3	ENSG00000204241	Chromosome 11: 134,032,272-134,046,849	−0.249	0.780 (0.621,0.979)	0.032

**Table 2: Univariate Cox regression analysis of overall survival in each cohort**

Clinical features	Number	P	HR	95%CIs of HR	
Risk score (High-risk/Low-risk)	129/129	< 0.001	1.445	1.321	1.581
Gender(male/female)	118/140	0.439	1.172	0.784	1.751
Age(< = 60/> 60)	128/130	0.134	1.357	0.91	2.023
Leiomyosarcoma histologic subtype(poorly differentiated or pleomorphic or epithelioid leiomyosarcoma/conventional leiomyosarcoma/ well-differentiated leiomyosarcoma (resembling leiomyoma)	34/64/4	0.025	0.592	0.303	0.924
leiomyosarcoma major vessel involvement(yes/no)	12/78	0.483	0.685	0.239	1.968
New neoplasm event type(new primary tumor /distant metastasis/locoregional recurrence)	5/41/29	0.668	0.885	0.506	1.548
Local disease recurrence(yes/no)	29/143	0.002	2.316	1.37	3.916
Metastatic diagnosis(yes/no)	56/119	< 0.001	3.009	1.831	4.946
Tumor depth(deep/superficial)	184/21	0.069	2.924	0.921	9.285
Contiguous organ invaded(yes/no)	14/43	0.012	2.809	1.261	6.261
Margin status(positive/negative)	73/136	0.013	1.84	1.138	2.974
Person neoplasm cancer status(with tumor/tumor free)	124/124	< 0.001	9.532	5.07	17.92
Residual tumor(yes/no)	154/78	< 0.001	2.553	1.668	3.909
Tumor total necrosis percent(0/< 10%/> = 10%,< 50%	12/61/38	0.121	1.207	0.952	1.532
Radiation therapy(yes/no)	73/179	0.508	0.882	0.566	1.376
Treatment completion success outcome(SD/PD/CR/ PR)	124/3/8/64	< 0.001	2.386	1.795	3.171

could largely predict the status of mitotic count, person neoplasm cancer and residual tumor with the AUCs being 0.666, 0.616 and 0.622, respectively (Figure 5). The Spearman Correlation test also confirmed the close relationships between risk score and mitotic count ($r = 0.297$, $P = 0.005$), person neoplasm cancer status ($r = 0.201$, $P = 0.001$) or residual tumor ($r = 0.2$, $P = 0.002$, Supplementary Table 1).

The risk score was notably different among all soft-tissue sarcomas of different histology, as well as among distinct differentiation levels in leiomyosarcoma (Supplementary Table 1). When the patients were stratified into different subgroups based on differentiation, the risk score of the 10-lncRNA signature remained to be an independent prognostic indicator for the subgroups of dedifferentiated liposarcoma and leiomyosarcoma (Table 3,

Table 3: Prognostic value of risk score stratified by histological type and treatment modality

Parameters	Number	HR	95% CIs		P
Histological type					
Leiomyosarcoma (LMS)	103	2.901	1.538	5.473	0.001
Dedifferentiated liposarcoma	58	3.952	1.529	10.211	0.005
Undifferentiated pleomorphic sarcoma	51	2.116	0.679	6.596	0.196
Treatment modality					
Radiation therapy	178	3.865	2.267	6.588	< 0.001
Non-radiation therapy	73	3.082	1.298	7.318	0.011
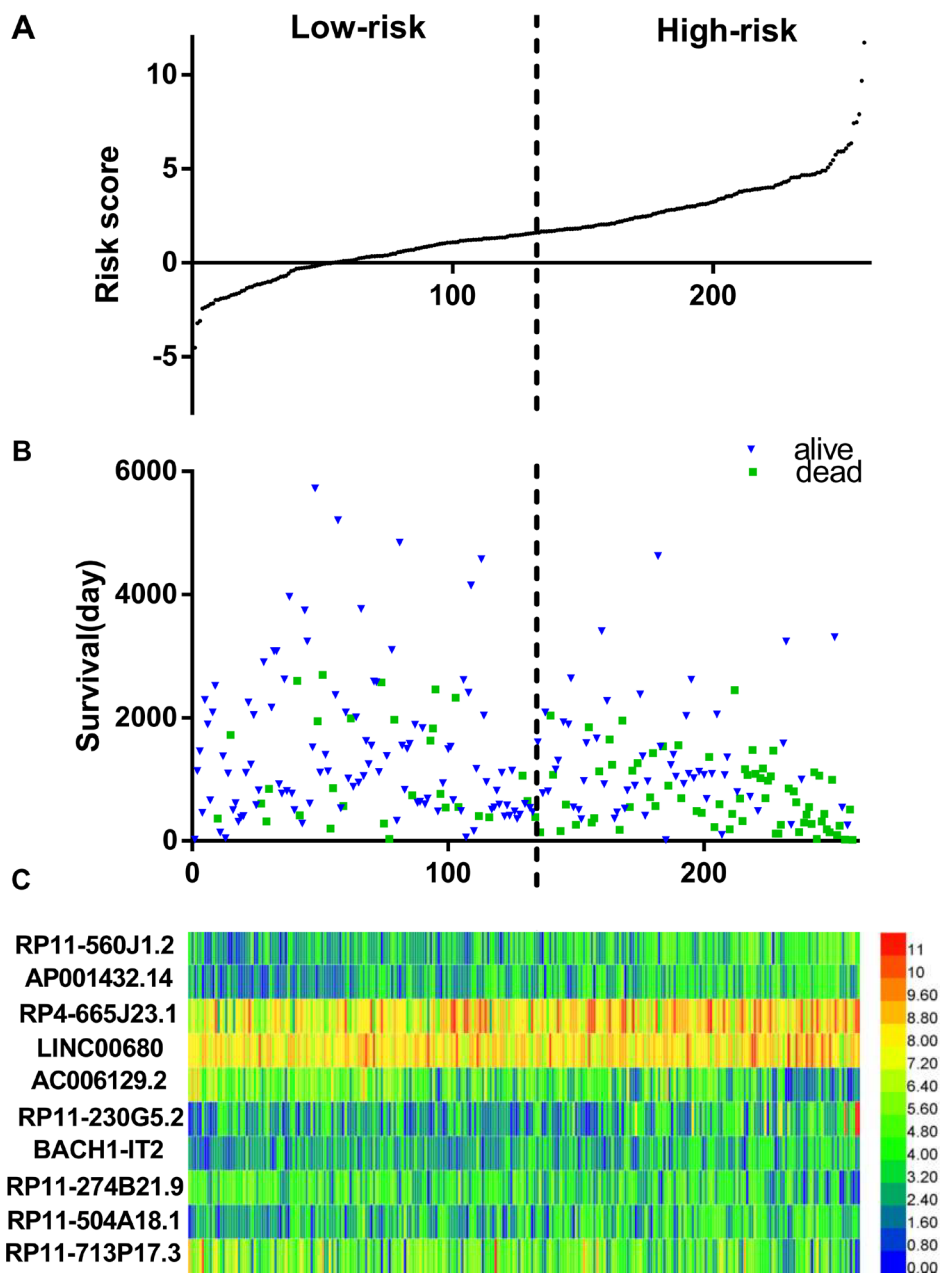
Figure 1: LncRNA predictive risk-score analysis of 258 soft-tissue sarcoma patients in TCGA cohort. (**A**) LncRNA risk-score distribution; (**B**) Patients survival; (**C**) Heatmap of lncRNA expression profiles of sarcoma patients. The black dotted line represents the median signature cutoff dividing patients into low-risk and high-risk groups.

Figure 6). According to the National Comprehensive Cancer Network (NCCN) Guidelines (Soft Tissue Sarcoma, https://www.nccn.org), radiation therapy influences the survival of soft-tissue sarcoma. In consideration of this, we also adjusted radiation therapy as a parameter to better reveal the prognostic value of the 10-lncRNA-based risk score for overall survival. No matter radiation therapy was received or not, the 10-lncRNA signature was a stable prognostic marker for soft-tissue sarcoma patients (Table 3, Figure 7).

Since the most life-threatening aspect of soft-tissue sarcomas is their capacity to disseminate hematogenously. We were also interested in the effect of each lncRNA on the metastasis. However, there was no significant difference of the level of these lncRNAs between metastatic and non-metastatic sarcomas. Regarding to ROC analysis, only RP4-665J23.1 conferred a mild diagnostic value for metastasis in soft-tissue sarcomas (AUC=0.615, $P = 0.0146$, Supplementary Table 2 in Supplementary Information).

Potential functional assessment of the lncRNAs by multi experiment matrix

Among these ten lncRNAs (four protective lncRNAs: LINC00680, AC006129.2, RP11-274B21.9
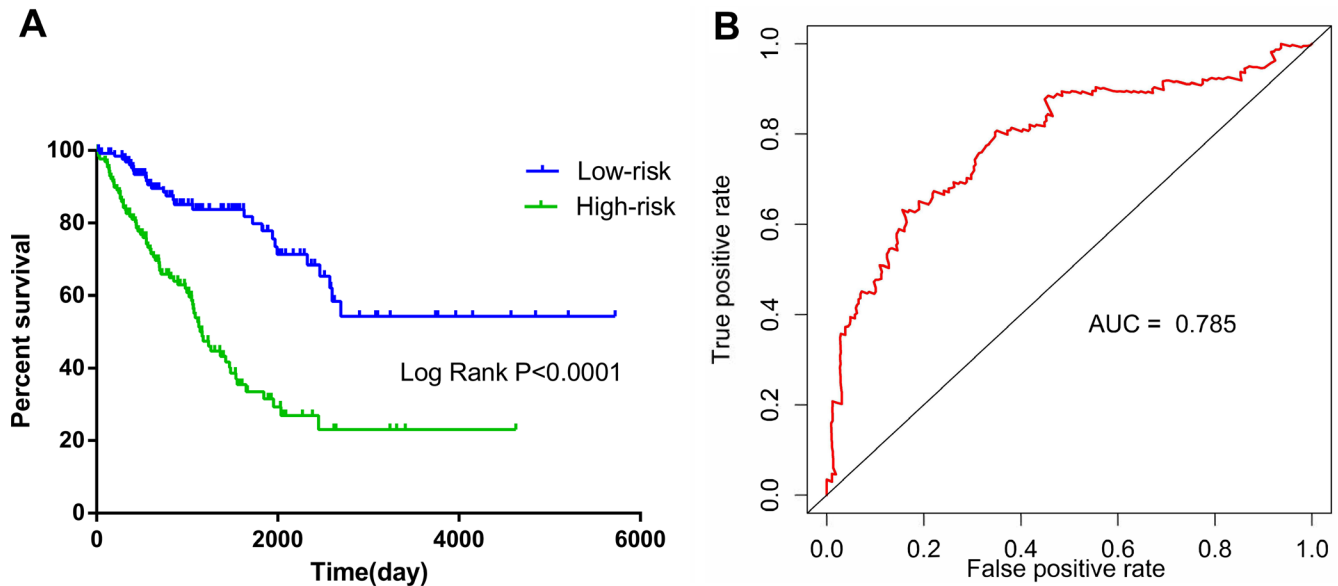

Figure 2: Kaplan–Meier and survival ROC curves for the ten-lncRNA signature in TCGA soft-tissue sarcoma cohort. (**A**) The Kaplan–Meier curves for high-risk and low-risk group sarcoma patients from the TCGA cohort divided by the median cutoff point. (**B**) The ROC curve had an AUC of 0.785.
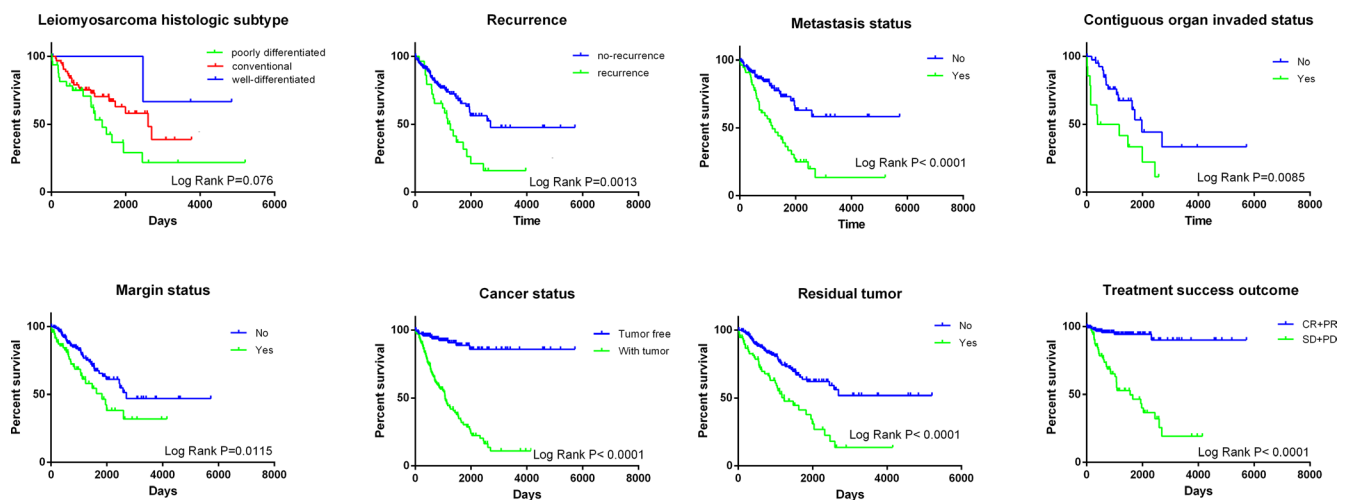

Figure 3: Kaplan–Meier curves of clinical features for the overall survival of soft-tissue sarcoma patients.

and RP11-713P17.3 and six risky lncRNAs: RP11-560J1.2, AP001432.14, RP4-665J23.1, RP11-230G5.2, BACH1-IT2, and RP11-504A18.1), the relevant genes of key lncRNAs were evaluated by MEM, and the co-expression networks of lncRNAs was visualized by MEM (Figure 8, Supplementary Figures 1–6 in Supplementary Information). Due to the absence of valid probe sets for four lncRNAs (RP11-560J1.2, AC006129.2, BACH1-IT2, RP11-274B21.9) on Affymetrix Gene Chip Human Genome U133 Plus 2.0 Array platform, we could obtain only six of ten lncRNAs from this online tool (AP001432.14, RP4-665J23.1, LINC00680, RP11-230G5.2, RP11-504A18.1, RP11-713P17.3), thus only these six lncRNAs were presented. We performed additional protein–protein interaction analysis for all the genes significantly co-expressed with the lncRNAs and found out three hub genes (RANBP2, POLR1B, GMPS) (Figure 9). The correlative genes of each lncRNA were
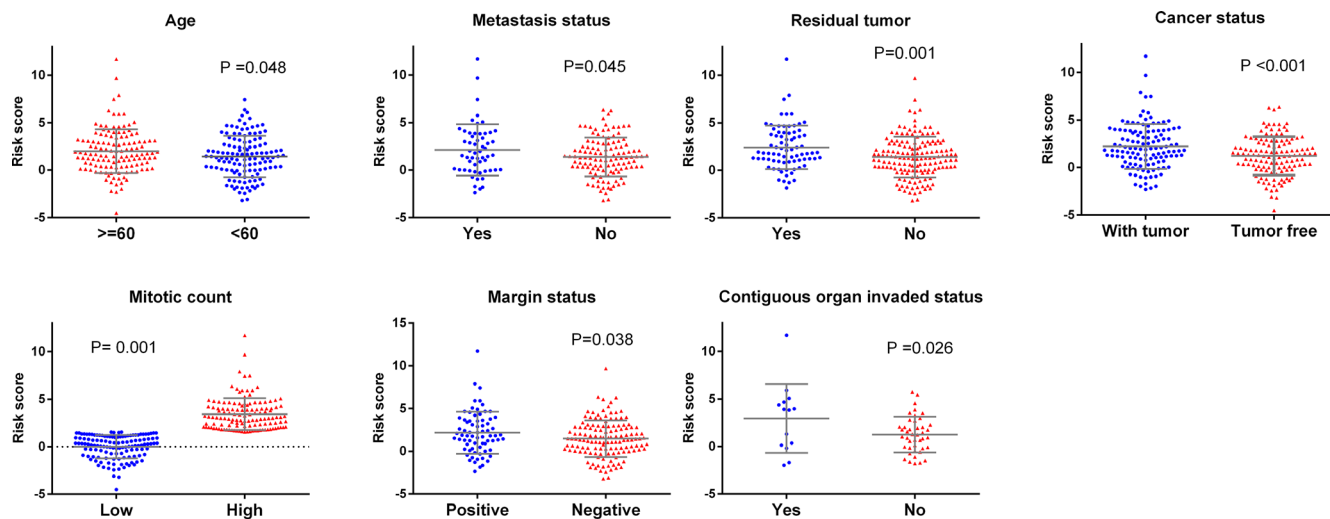
Figure 4: Relationship between clinical features and risk score. (**A**) Age; (**B**) Metastasis status (**C**) Residual tumor (**D**) Tumor status (**E**) Mitotic count; (**F**) Margin status; (**G**) Contiguous organ invaded.
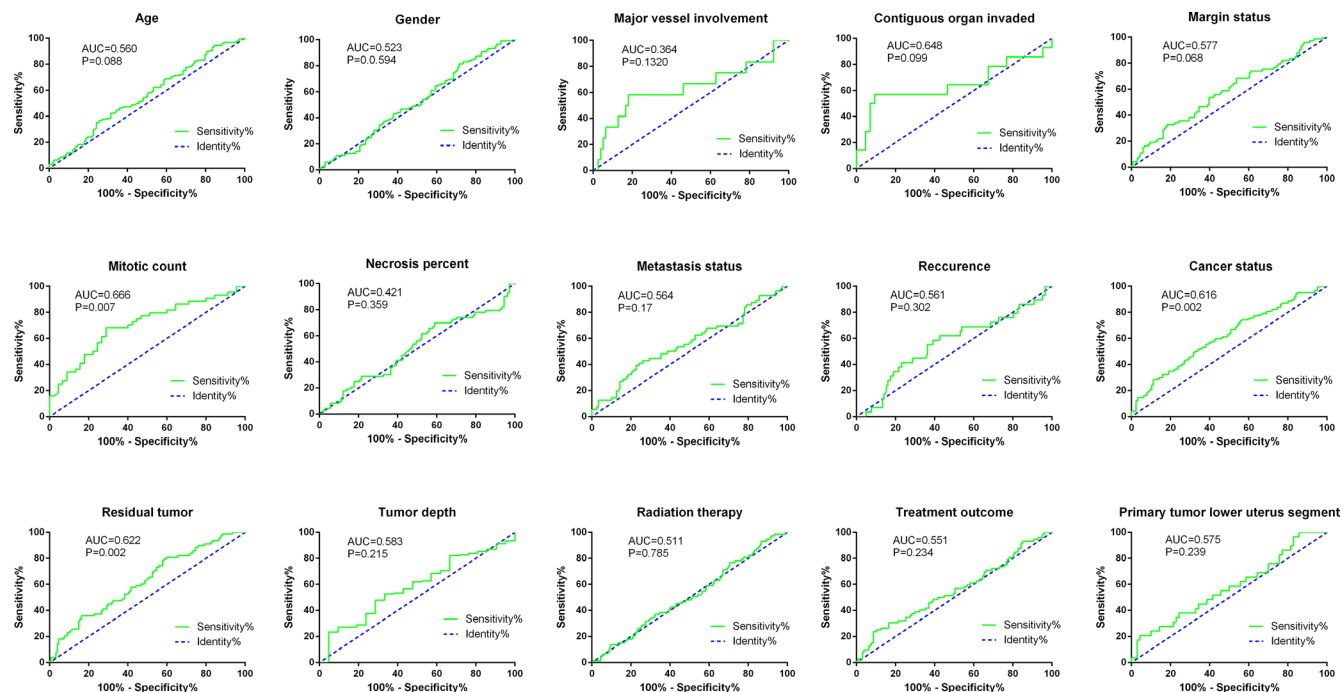
**Figure 5: ROC curves of risk score for clinical features in soft-tissue sarcoma patients.**

enriched in multiple pathways assessed by GO and KEGG analyses (Supplementary Table 3).

DISCUSSION

Soft-tissue sarcoma represents a rare cancer entity, accounting for less than 1% of adult malignancies. The keystone of curative intent treatment is surgery with free margins. However, multimodal methodologies which often entail radiation therapy have substituted extensive surgical procedures to preserve functionality while maintaining sufficient local control [33–37]. In terms of the aggressiveness and shortness of specific biomarkers for soft tissue sarcoma patients, there is a crucial requirement for trustworthy prognostic indicators identifying a subclass of patients with poor prognosis, who would hence benefit from extra treatment possibilities [38, 39]. Recently, large-scale genomic analyses have shown a set of molecular events were closely related to the tumorigenesis or progression of soft-tissue sarcomas. Subramanian et al. [40] conducted a comprehensive investigation into microRNA expression signatures of 27 cases of sarcomas, 5 cases of normal smooth muscle and 2 cases of normal skeletal muscle tissues with microarray and small RNA cloning approaches. The identification of exceptional microRNA signatures in every sarcoma type may specify their function in pathogenesis of soft-tissue sarcomas.

However, all existing studies on the relationship between lncRNA and soft-tissue sarcomas focused on single lncRNA and single disease. For example, higher expression of lncRNA HOTAIR was detected in both primary and metastatic sarcoma tissues. High level expression of HOTAIR in the primary sarcoma was also significantly related to metastasis. HOTAIR expression level tend to be correlated with the necrotic status in various sarcoma tissues [41]. But no studies have been performed so far to investigate the clinical role of an lncRNA signature in soft-tissue sarcoma based on high-throughput data. In the current study, a comprehensive investigation of lncRNA expression based on RNA-seq data was conducted with a large cohort of soft-tissue sarcoma patients from the data of TCGA database, including leiomyosarcoma, dedifferentiated liposarcoma, undifferentiated pleomorphic sarcoma, myxofibrosarcoma synovial sarcoma and malignant peripheral nerve sheath tumors. Ten lncRNAs showed the most potent prognostic
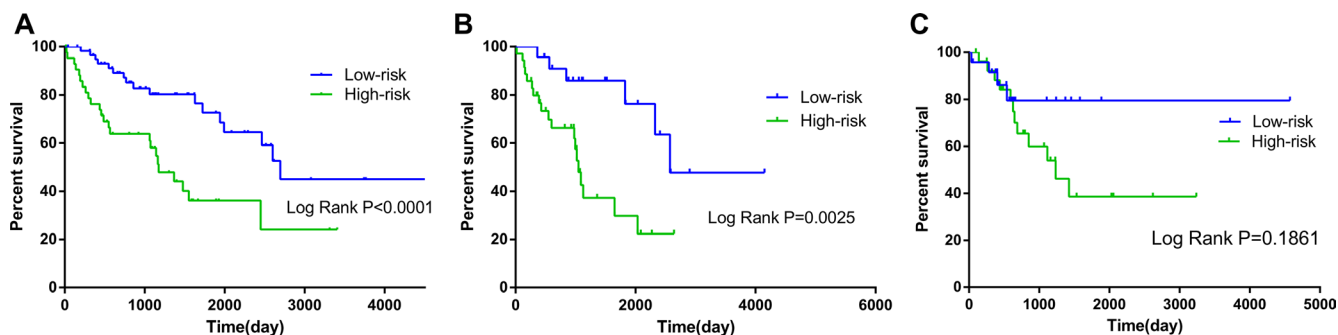
Figure 6: Kaplan–Meier curves for prognostic value of risk-score signature according to histologic subtypes. (**A**) Dedifferentiated liposarcoma (**B**) Leiomyosarcoma (LMS); (**C**) Undifferentiated sarcoma.
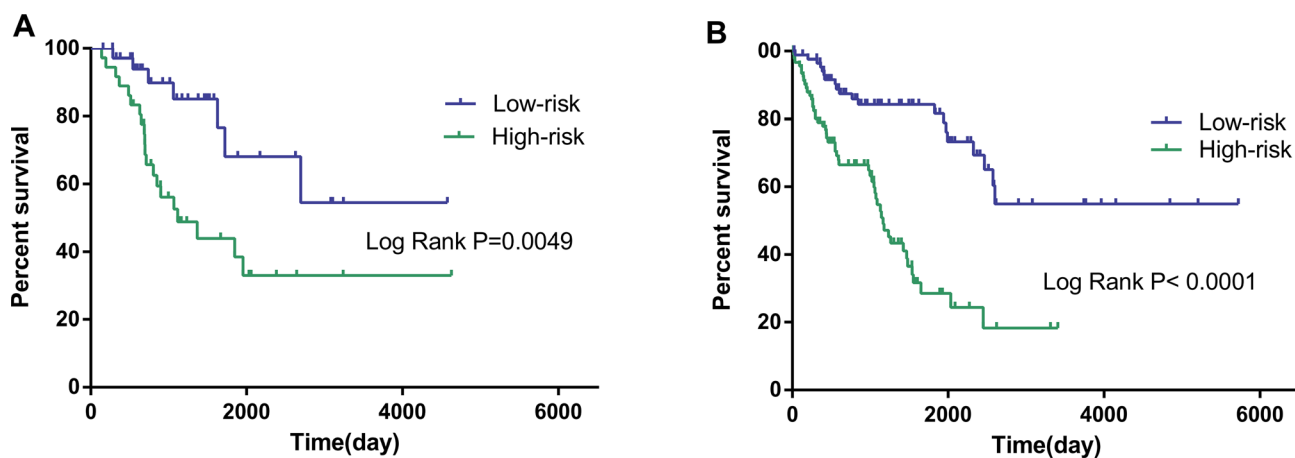

Figure 7: Kaplan–Meier curves for prognostic value of risk-score signature for the patients divided by treatment modalities. (**A**) Radiation therapy (**B**) Non-radiation therapy.

value, including four protective lncRNAs (LINC00680, AC006129.2, RP11-274B21.9 and RP11-713P17.3) and six risky lncRNA: RP11-560J1.2, AP001432.14, RP4-665J23.1, RP11-230G5.2, BACH1-IT2, and RP11-504A18.1). More importantly, the risk score calculated by the 10-lncRNA signature was capable to predict the outcome and overall survival of soft-tissue sarcoma independent of other clinicopathological features, which was assessed by Cox regression analysis. This novel 10-lncRNA signature could be an independent prognostic indicator for soft-tissue sarcoma and this study discovered the potency of a combined lncRNA signature to properly predict the survival of soft-tissue sarcoma.

We attempted to validate the current findings from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/gds) and ArrayExpress (http://www.ebi.ac.uk/arrayexpress/). "soft-tissue sarcoma" OR "soft tissue sarcoma" OR leiomyosarcoma OR liposarcoma OR "undifferentiated pleomorphic sarcoma" OR myxofibrosarcoma OR "synovial sarcoma" OR "peripheral nerve sheath tumor" were searched and altogether 278 series were achieved (data not shown). Next, we searched our 10 lncRNAs in these series and unfortunately, only GSE21050 could be included with the data of AC006129.2 and RP11-504A18.1. But no survival data was available and we could only assess the relationship between the level of lncRNAs and metastasis.

No significant correlations between these 2 lncRNAs and metastasis could be noted (Supplementary Figures 7, 8 in Supplementary Information). So our findings from TCGA need to be verified by other experiments in the future.

Among all the 10 lncRNAs to calculate the risk score, none of researchers has reported in literatures so far. Hence, the clinical role or biological function of any of them remains absolutely unknown. Previous studies have proposed that lncRNAs may take part in different biological processes by interrelating with correlative genes. To this end, MEM was used to gather the correlative genes of each lncRNA and also to inspect the rough molecular mechanism. The results suggested that the six of these 10 prognostic lncRNAs might be related to a number of genes and multiply signaling pathways involved in soft-tissue sarcoma. Even these lncRNAs have consistent prognostic value for soft-tissue sarcoma, they may share absolutely different molecular mechanisms. Interestingly, from KEGG analysis, the Neuroactive ligand-receptor interaction pathway and Calcium signaling pathway are the top 2 pathways for both AP001432.14 and RP11-230G5.2, indicating that these 2 lncRNAs may have similar functions via targeting comparable signaling pathways; however, validation with additional *in vitro* and *in vivo* experiments is requisite to uncover the underlying molecular mechanism of lncRNAs in soft-tissue sarcoma.
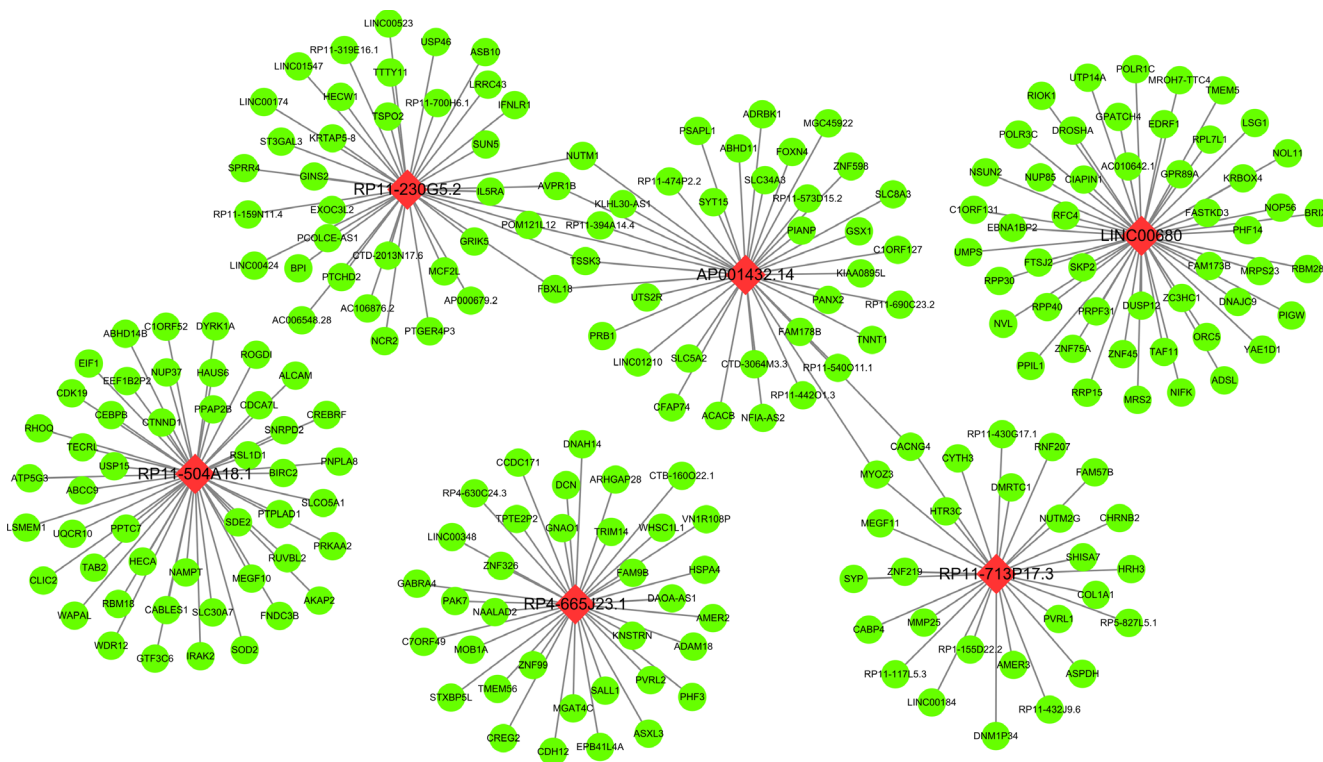
Figure 8: Regulation network of each key lncRNA by multi experiment matrix. This network was established based on the top 50 target genes for the lncRNA by utilizing the Multi Experiment Matrix. The green balls present the target genes and the red diamonds show the key lncRNAs.

By mining the high throughput RNA-seq and clinical data from TCGA database, we identified a specific 10-lncRNA signature closely related to patient survival in soft-tissue sarcoma that can provide a potent prognostic tool for this class of tumors. However, the current finding based on a single cohort of TCGA needs further validation by other detecting methods, for instance, real time RT-qPCR and fluorescence *in situ* hybridization. Furthermore, the biological function and molecular mechanism of these lncRNAs remain unexplored, which need to be explored in-depth in the future.

MATERIALS AND METHODS

Patient information and lncRNA expression profiles

The mRNA expression data (Level 3) and clinical data for soft-tissue sarcoma patients (up to January 22, 2017) were achieved from TCGA data portal. The expression of 7549 lncRNAs in soft-tissue sarcoma samples was analyzed on IlluminaHiSeq mRNA Seq platform. Those samples without lncRNA sequence data or clinical data were omitted. Therefore, a sum of 258 soft-tissue sarcoma patients were involved in the current study, with matching clinical data including gender, age, histology, differentiation, leiomyosarcoma major vessel involvement, new neoplasm event type, local disease recurrence, metastatic diagnosis, tumor depth, contiguous organ invaded, margin status, person neoplasm cancer status, residual tumor, tumor total necrosis percent, radiation therapy and treatment completion success outcome (SD/PD/CR/PR) were involved in the current study (Table 2, Supplementary Table 1). The end-point was OS in this study for the soft-tissue sarcoma. As the data were downloaded from TCGA, additional approval by ethics committee was not obligatory. Data was processed according to the TCGA human subject protection and data access policies.
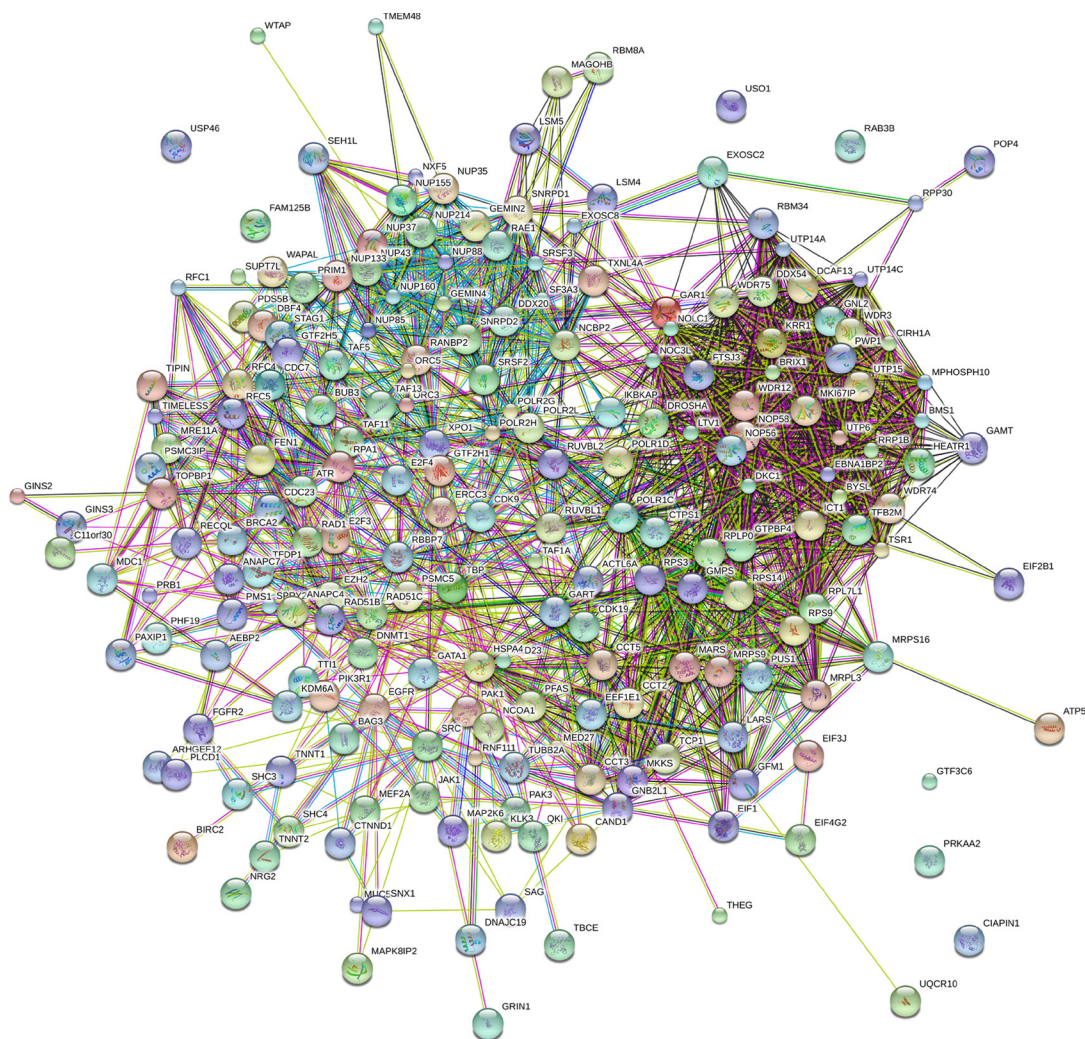
Figure 9: The protein protein interaction of correlative genes for the lncRNAs. The protein-to-protein network analysis was performed using STRING (version: STRING 10.0).

Statistical analysis

The expression level of 7549 lncRNAs was shown as raw count lncRNA mapped data. The lncRNAs which were less than 1 raw count in exceeded 10% of all subjects were eliminated using R language. The expression level of each lncRNA was log2 transformed for further analysis. The univariate Cox proportional hazards regression with significance level set as 0.01 was performed to find out the lncRNAs evidently associated with OS. The filtered lncRNAs were divided into risky (with a hazard ratio (HR) for death greater than 1) and protective (based on a HR for death less than 1) types. A risk score formula for predicting OS was developed based on a linear combination of the expression level that multiplied regression coefficient derived from the multivariate cox stepwise regression model (β): risk score = expression of gene 1*β gene 1 + expression of gene 2*β gene 2 + …expression of gene n*β gene n. The "β" value is the estimated regression coefficient of lncRNA derived from the multivariate Cox stepwise regression analysis. By utilizing the median risk score as the cutoff point, the soft-tissue sarcoma patients were divided into two subgroups of high score and low score. We also used the R package "survivalROC" to assess the predictive accuracy of prognostic model for time dependent disease outcomes within 5 years as the defining point.

Univariate Cox proportional hazards regression analyses were performed to explore the effects of clinical features and the risk score on OS of soft-tissue sarcoma patients. Each predictor identified via univariate analysis was further assessed by multivariate cox proportional hazards regression analysis. Survival curves were generated by the Kaplan–Meier and log-rank method. The difference of risk score between diverse groups according to various clinical parameters was assessed by student t test or ANOVA test. ROC curves were used to evaluate the predictive values of risk score for different parameters representative of patients' outcome and survival. The Spearman Correlation test was conducted to evaluate the correlation between risk score and the progression of soft-tissue sarcoma. The student t-test, ROC analysis as well as Spearman correlation test were also performed to examine the relationship between these novel lncRNAs and metastasis individually, since metastasized soft-tissue sarcomas generally indicate a poorer prognosis. Statistical significance was defined as a two-sided P value < 0.05. The statistical analyses were performed with SPSS22.0 software.

Potential functional assessment of the lncRNAs by multi experiment matrix

Then, we also explored the co-expressed genes for the key lncRNAs by Multi Experiment Matrix (MEM, http://biit.cs.ut.ee/mem/index.cgi) based on Affymetrix Gene Chip Human Genome U133 Plus 2.0 Array platform [42, 43]. After the identification of weighted correlation, Cytoscape 3.4.0 was used to show the network between lncRNAs and their related genes. GO and KEGG analyses were also performed based on the Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/). The protein-to-protein network analysis was conducted by STRING (version: STRING 10.0).

CONFLICTS OF INTEREST

None.

REFERENCES

1. Jo VY, Doyle LA. Refinements in Sarcoma Classification in the Current 2013 World Health Organization Classification of Tumours of Soft Tissue and Bone. Surg Oncol Clin N Am. 2016; 25:621–43.

2. Yamaguchi K, Katagiri H, Takahashi M, Ishida Y, Ono A, Takahashi T, Ohshima K, Mochizuki T, Urakami K, Muramatsu K, Kameya T, Ito I, Nakajima T. ProGRP is a possible tumor marker for patients with Ewing sarcoma. Biomed Res. 2015; 36:273–77.

3. Dancsok AR, Asleh-Aburaya K, Nielsen TO. Advances in sarcoma diagnostics and treatment. Oncotarget. 2017; 8:7068–93. https://doi.org/10.18632/oncotarget.12548.

4. Hong SH, Tilan JU, Galli S, Izycka-Swieszewska E, Polk T, Horton M, Mahajan A, Christian D, Jenkins S, Acree R, Connors K, Ledo P, Lu C, et al. High neuropeptide Y release associates with Ewing sarcoma bone dissemination - *in vivo* model of site-specific metastases. Oncotarget. 2015; 6:7151–65. https://doi.org/10.18632/oncotarget.3345

5. Wang W, Zhang L, Zheng K, Zhang X. miR-17-5p promotes the growth of osteosarcoma in a BRCC2-dependent mechanism. Oncol Rep. 2016; 35:1473–82.

6. Paulis YW, Huijbers EJ, van der Schaft DW, Soetekouw PM, Pauwels P, Tjan-Heijnen VC, Griffioen AW. CD44 enhances tumor aggressiveness by promoting tumor cell plasticity. Oncotarget. 2015; 6:19634–46. https://doi.org/10.18632/oncotarget.3839

7. Bergerat S, Barthelemy P, Mouracade P, Lang H, Saussine C, Lindner V, Jacqmin D. Primary CIC-DUX4 round cell sarcoma of the kidney: A treatment-refractory tumor with poor outcome. Pathol Res Pract. 2017; 213:154–60.

8. Jiang Q, Zhang Z, Li S, Wang Z, Ma Y, Hu Y. Defective heat shock factor 1 inhibits the growth of fibrosarcoma derived from simian virus 40/T antigen-transformed MEF cells. Mol Med Rep. 2015; 12:6517–26.

9. López-Pousa A, Martin Broto J, Martinez Trufero J, Sevilla I, Valverde C, Alvarez R, Carrasco Alvarez JA, Cruz Jurado J, Hindi N, Garcia Del Muro X. SEOM Clinical Guideline of management of soft-tissue sarcoma (2016). Clin Transl Oncol. 2016; 18:1213–20.

10. Chen JC, Chen YJ, Lin CY, Fong YC, Hsu CJ, Tsai CH, Su JL, Tang CH. Amphiregulin enhances alpha6beta1 integrin expression and cell motility in human chondrosarcoma cells through Ras/Raf/MEK/ERK/AP-1 pathway. Oncotarget. 2015; 6:11434–46. https://doi.org/10.18632/oncotarget.3397.

11. Ozcelik M, Seker M, Eraslan E, Koca S, Yazilitas D, Ercelep O, Ozaslan E, Kaya S, Hacibekiroglu I, Menekse S, Aksoy A, Taskoylu BY, Varol U, et al. Evaluation of prognostic factors in localized high-grade undifferentiated pleomorphic sarcoma: report of a multi-institutional experience of Anatolian Society of Medical Oncology. Tumour Biol. 2016; 37:5231–37.

12. Richardson K, Potter M, Damron TA. Image intensive soft tissue sarcoma surveillance uncovers pathology earlier than patient complaints but with frequent initially indeterminate lesions. J Surg Oncol. 2016; 113:818–22.

13. Chen Q, Zhou Z, Shan L, Zeng H, Hua Y, Cai Z. The importance of Src signaling in sarcoma. Oncol Lett. 2015; 10:17–22.

14. Toda-Ishii M, Akaike K, Kurisaki-Arakawa A, Arakawa A, Mukaihara K, Suehara Y, Takagi T, Kaneko K, Yao T, Saito T. Sarcomatous transformation of EGFR and TP53 mutation-positive metastatic adenocarcinoma of the lungs, masquerading as a primary pleomorphic sarcoma of the proximal femur. Int J Clin Exp Pathol. 2015; 8:3270–78.

15. Tanaka K, Joyama S, Chuman H, Hiraga H, Morioka H, Yoshikawa H, Hosaka M, Takahashi M, Kubo T, Hatano H, Kaya M, Toguchida J, Nishida Y, et al. Feasibility and efficacy of gemcitabine and docetaxel combination chemotherapy for bone and soft tissue sarcomas: multi-institutional retrospective analysis of 134 patients. World J Surg Oncol. 2016; 14:306.

16. Nakamura T, Matsumine A, Asanuma K, Matsubara T, Sudo A. The role of C-reactive protein in predicting post-metastatic survival of patients with metastatic bone and soft tissue sarcoma. Tumour Biol. 2015; 36:7515–20.

17. Lazo de la Vega L, Hovelson DH, Cani AK, Liu CJ, McHugh JB, Lucas DR, Thomas DG, Patel RM, Tomlins SA. Targeted next-generation sequencing of CIC-DUX4 soft tissue sarcomas demonstrates low mutational burden and recurrent chromosome 1p loss. Hum Pathol. 2016; 58:161–70.

18. Xing P, Zhang J, Yan Z, Zhao G, Li X, Wang G, Yang Y, Zhao J, Xing R, Teng S, Ma Y, Liao Z, Ren Z, et al. Recombined humanized endostatin (Endostar) combined with chemotherapy for advanced bone and soft tissue sarcomas in stage IV. Oncotarget. 2017; 8:36716-36727. https://doi.org/10.18632/oncotarget.13545.

19. De Sanctis R, Marrari A, Santoro A. Trabectedin for the treatment of soft tissue sarcomas. Expert Opin Pharmacother. 2016; 17:1569–77.

20. Liu C, Lin J. Long noncoding RNA ZEB1-AS1 acts as an oncogene in osteosarcoma by epigenetically activating ZEB1. Am J Transl Res. 2016; 8:4095–105.

21. Tian ZZ, Guo XJ, Zhao YM, Fang Y. Decreased expression of long non-coding RNA MEG3 acts as a potential predictor biomarker in progression and poor prognosis of osteosarcoma. Int J Clin Exp Pathol. 2015; 8:15138–42.

22. Luo W, He H, Xiao W, Liu Q, Deng Z, Lu Y, Wang Q, Zheng Q, Li Y. MALAT1 promotes osteosarcoma development by targeting TGFA via MIR376A. Oncotarget. 2016; 7:54733–43. https://doi.org/10.18632/oncotarget.10752

23. Uzan VR, Lengert A, Boldrini É, Penna V, Scapulatempo-Neto C, Scrideli CA, Filho AP, Cavalcante CE, de Oliveira CZ, Lopes LF, Vidal DO. High Expression of HULC Is Associated with Poor Prognosis in Osteosarcoma Patients. PLoS One. 2016; 11:e0156774.

24. Li F, Cao L, Hang D, Wang F, Wang Q. Long non-coding RNA HOTTIP is up-regulated and associated with poor prognosis in patients with osteosarcoma. Int J Clin Exp Pathol. 2015; 8:11414–20.

25. Li W, He X, Xue R, Zhang Y, Zhang X, Lu J, Zhang Z, Xue L. Combined over-expression of the hypoxia-inducible factor 2α gene and its long non-coding RNA predicts unfavorable prognosis of patients with osteosarcoma. Pathol Res Pract. 2016; 212:861–66.

26. Chen F, Mo J, Zhang L. Long noncoding RNA BCAR4 promotes osteosarcoma progression through activating GLI2-dependent gene transcription. Tumour Biol. 2016; 37:13403–12.

27. Zhao H, Hou W, Tao J, Zhao Y, Wan G, Ma C, Xu H. Upregulation of lncRNA HNF1A-AS1 promotes cell proliferation and metastasis in osteosarcoma through activation of the Wnt/β-catenin signaling pathway. Am J Transl Res. 2016; 8:3503–12.

28. Yin Z, Ding H, He E, Chen J, Li M. Overexpression of long non-coding RNA MFI2 promotes cell proliferation and suppresses apoptosis in human osteosarcoma. Oncol Rep. 2016; 36:2033–40.

29. Li JP, Liu LH, Li J, Chen Y, Jiang XW, Ouyang YR, Liu YQ, Zhong H, Li H, Xiao T. Microarray expression profile of long noncoding RNAs in human osteosarcoma. Biochem Biophys Res Commun. 2013; 433:200–06.

30. Zhou X, Huang Z, Xu L, Zhu M, Zhang L, Zhang H, Wang X, Li H, Zhu W, Shu Y, Liu P. A panel of 13-miRNA signature as a potential biomarker for predicting survival in pancreatic cancer. Oncotarget. 2016; 7:69616–24. https://doi.org/10.18632/oncotarget.11903

31. Niyazi M, Pitea A, Mittelbronn M, Steinbach J, Sticht C, Zehentmayr F, Piehlmaier D, Zitzelsberger H, Ganswindt U,

Rödel C, Lauber K, Belka C, Unger K. A 4-miRNA signature predicts the therapeutic outcome of glioblastoma. Oncotarget. 2016; 7:45764–75. https://doi.org/10.18632/oncotarget.9945.

32. Shi KQ, Lin Z, Chen XJ, Song M, Wang YQ, Cai YJ, Yang NB, Zheng MH, Dong JZ, Zhang L, Chen YP. Hepatocellular carcinoma associated microRNA expression signature: integrated bioinformatics analysis, experimental validation and clinical significance. Oncotarget. 2015; 6:25093–108. https://doi.org/10.18632/oncotarget.4437

33. Roeder F, Krempien R. Intraoperative radiation therapy (IORT) in soft-tissue sarcoma. Radiat Oncol. 2017; 12:20.

34. Astone M, Pizzi M, Peron M, Domenichini A, Guzzardo V, Töchterle S, Tiso N, Rugge M, Meyer D, Argenton F, Vettori A. A GFP-Tagged Gross Deletion on Chromosome 1 Causes Malignant Peripheral Nerve Sheath Tumors and Carcinomas in Zebrafish. PLoS One. 2015; 10:e0145178.

35. Maekawa A, Kohashi K, Kuda M, Iura K, Ishii T, Endo M, Nakatsura T, Iwamoto Y, Oda Y. Prognostic significance of FOXM1 expression and antitumor effect of FOXM1 inhibition in synovial sarcomas. BMC Cancer. 2016; 16:511.

36. Okuda K, Yano M, Moriyama S, Haneda H, Tatematsu T, Suzuki A, Oda R, Nakanishi R. A case of mediastinum undifferentiated high grade pleomorphic sarcoma. Int J Clin Exp Med. 2015; 8:19566–70.

37. Feng D, Xu F, Wang M, Gu X, Ma Z. Anesthetic management of a patient with giant retroperitoneal liposarcoma: case report with literature review. Int J Clin Exp Med. 2015; 8:19530–34.

38. Nassif NA, Tseng W, Borges C, Chen P, Eisenberg B. Recent advances in the management of liposarcoma. F1000 Res. 2016; 5:2907.

39. Childhood Soft Tissue Sarcoma Treatment. (PDQ(R)): Health Professional Version. PDQ Cancer Information Summaries. (Bethesda (MD). 2002.

40. Subramanian S, Lui WO, Lee CH, Espinosa I, Nielsen TO, Heinrich MC, Corless CL, Fire AZ, van de Rijn M. MicroRNA expression signature of human sarcomas. Oncogene. 2008; 27:2015–26.

41. Milhem MM, Knutson T, Yang S, Zhu D, Wang X, Leslie KK, Meng X. Correlation of MTDH/AEG-1 and HOTAIR Expression with Metastasis and Response to Treatment in Sarcoma Patients. J Cancer Sci Ther. 2011; S5:S5.

42. Adler P, Kolde R, Kull M, Tkachenko A, Peterson H, Reimand J, Vilo J. Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods. Genome Biol. 2009; 10:R139.

43. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012; 28:573–80.