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

Rong-quan He1,*, Qing-jun Wei2,*, Rui-xue Tang3, Wen-jie Chen4, Xia Yang3, Zhi-gang Peng1, Xiao-hua Hu1, Jie Ma1 and Gang Chen3

1Department of Medical Oncology, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, P. R. China
2Department of Orthopaedics Trauma and Hand Surgery, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, P. R. China
3Department of Pathology, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, P. R. China
4Department 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

Keywords: soft-tissue sarcoma, lncRNA, prognosis, RNA-seq, TCGA

Received: January 23, 2017    Accepted: May 10, 2017    Published: May 22, 2017

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

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
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 2).

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

<table>
<thead>
<tr>
<th>IncRNA</th>
<th>Ensemble ID</th>
<th>Location</th>
<th>(\beta) (Cox)</th>
<th>HR (95% CIs)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP11-560J1.2</td>
<td>ENSG00000271888</td>
<td>Chromosome 6: 15,243,923-15,245,000</td>
<td>0.582</td>
<td>1.789 (1.269, 2.522)</td>
<td>0.001</td>
</tr>
<tr>
<td>AP001432.14</td>
<td>ENSG00000242553</td>
<td>Chromosome 21: 37,221,419-37,237,744</td>
<td>0.828</td>
<td>2.288 (1.502, 3.480)</td>
<td>0.000</td>
</tr>
<tr>
<td>RP4-665J23.1</td>
<td>ENSG00000233593</td>
<td>Chromosome 1: 90,782,983-90,851,657</td>
<td>0.348</td>
<td>1.417 (1.019, 1.970)</td>
<td>0.038</td>
</tr>
<tr>
<td>LINC00680</td>
<td>ENSG00000215190</td>
<td>Chromosome 6: 57,946,074-57,961,501</td>
<td>−0.681</td>
<td>0.506 (0.310, 0.827)</td>
<td>0.007</td>
</tr>
<tr>
<td>AC006129.2</td>
<td>ENSG00000268027</td>
<td>Chromosome 19: 41,545,192-41,555,462</td>
<td>−0.236</td>
<td>0.790 (0.657, 0.949)</td>
<td>0.012</td>
</tr>
<tr>
<td>RP11-230G5.2</td>
<td>ENSG00000250748</td>
<td>Chromosome 12: 65,466,820-65,642,372</td>
<td>0.495</td>
<td>1.641 (1.193, 2.256)</td>
<td>0.002</td>
</tr>
<tr>
<td>BACH1-IT2</td>
<td>ENSG00000228817</td>
<td>Chromosome 21: 29,370,497-29,373,709</td>
<td>0.423</td>
<td>1.527 (1.046, 2.228)</td>
<td>0.028</td>
</tr>
<tr>
<td>RP11-274B21.9</td>
<td>ENSG00000271344</td>
<td>Chromosome 7: 128,690,451-128,691,717</td>
<td>−0.557</td>
<td>0.573 (0.420, 0.781)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>RP11-504A18.1</td>
<td>ENSG00000260971</td>
<td>Chromosome 1: 56,248,294-56,258,571</td>
<td>0.466</td>
<td>1.594 (1.126, 2.257)</td>
<td>0.009</td>
</tr>
<tr>
<td>RP11-713P17.3</td>
<td>ENSG00000204241</td>
<td>Chromosome 11: 134,032,722-134,046,849</td>
<td>−0.249</td>
<td>0.780 (0.621, 0.979)</td>
<td>0.032</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>Number</th>
<th>(P)</th>
<th>HR</th>
<th>95% CIs of HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk score (High-risk/Low-risk)</td>
<td>129/129</td>
<td>&lt; 0.001</td>
<td>1.445</td>
<td>1.321</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>118/140</td>
<td>0.439</td>
<td>1.172</td>
<td>0.784</td>
</tr>
<tr>
<td>Age (&lt; 60/ &gt; 60)</td>
<td>128/130</td>
<td>0.134</td>
<td>1.357</td>
<td>0.910</td>
</tr>
<tr>
<td>Leiomyosarcoma histologic subtype (poorly differentiated or pleomorphic or epithelioid leiomyosarcoma/conventional leiomyosarcoma/well-differentiated leiomyosarcoma (resembling leiomyoma))</td>
<td>34/64/4</td>
<td>0.025</td>
<td>0.592</td>
<td>0.303</td>
</tr>
<tr>
<td>leiomyosarcoma major vessel involvement (yes/no)</td>
<td>12/78</td>
<td>0.483</td>
<td>0.685</td>
<td>0.239</td>
</tr>
<tr>
<td>New neoplasm event type (new primary tumor/distant metastasis/locoregional recurrence)</td>
<td>5/41/29</td>
<td>0.668</td>
<td>0.885</td>
<td>0.506</td>
</tr>
<tr>
<td>Local disease recurrence (yes/no)</td>
<td>29/143</td>
<td>0.002</td>
<td>2.316</td>
<td>1.37</td>
</tr>
<tr>
<td>Metastatic diagnosis (yes/no)</td>
<td>56/119</td>
<td>&lt; 0.001</td>
<td>3.009</td>
<td>1.831</td>
</tr>
<tr>
<td>Tumor depth (deep/superficial)</td>
<td>184/21</td>
<td>0.069</td>
<td>2.924</td>
<td>0.921</td>
</tr>
<tr>
<td>Contiguous organ invaded (yes/no)</td>
<td>14/43</td>
<td>0.012</td>
<td>2.809</td>
<td>1.261</td>
</tr>
<tr>
<td>Margin status (positive/negative)</td>
<td>73/136</td>
<td>0.013</td>
<td>1.84</td>
<td>1.138</td>
</tr>
<tr>
<td>Person neoplasm cancer status (with tumor/tumor free)</td>
<td>124/124</td>
<td>&lt; 0.001</td>
<td>9.532</td>
<td>5.07</td>
</tr>
<tr>
<td>Residual tumor (yes/no)</td>
<td>154/78</td>
<td>&lt; 0.001</td>
<td>2.553</td>
<td>1.668</td>
</tr>
<tr>
<td>Tumor total necrosis percent (0/&lt; 10%/&gt; = 10%,&lt; 50%)</td>
<td>12/61/38</td>
<td>0.121</td>
<td>1.207</td>
<td>0.952</td>
</tr>
<tr>
<td>Radiation therapy (yes/no)</td>
<td>73/179</td>
<td>0.508</td>
<td>0.882</td>
<td>0.566</td>
</tr>
<tr>
<td>Treatment completion success outcome (SD/PD/CR/PR)</td>
<td>124/3/86/64</td>
<td>&lt; 0.001</td>
<td>2.386</td>
<td>1.795</td>
</tr>
</tbody>
</table>
Table 3: Prognostic value of risk score stratified by histological type and treatment modality

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Number</th>
<th>HR</th>
<th>95% CIs</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histological type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leiomyosarcoma (LMS)</td>
<td>103</td>
<td>2.901</td>
<td>1.538</td>
<td>5.473</td>
</tr>
<tr>
<td>Dedifferentiated liposarcoma</td>
<td>58</td>
<td>3.952</td>
<td>1.529</td>
<td>10.211</td>
</tr>
<tr>
<td>Undifferentiated pleomorphic sarcoma</td>
<td>51</td>
<td>2.116</td>
<td>0.679</td>
<td>6.596</td>
</tr>
<tr>
<td>Treatment modality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiation therapy</td>
<td>178</td>
<td>3.865</td>
<td>2.267</td>
<td>6.588</td>
</tr>
<tr>
<td>Non-radiation therapy</td>
<td>73</td>
<td>3.082</td>
<td>1.298</td>
<td>7.318</td>
</tr>
</tbody>
</table>

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.](image)

(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.](image)
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 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 IncRNA and soft-tissue sarcomas focused on single IncRNA and single disease. For example, higher expression of IncRNA 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 IncRNA signature in soft-tissue sarcoma based on high-throughput data. In the current study, a comprehensive investigation of IncRNA 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, synovial sarcoma and malignant peripheral nerve sheath tumors. Ten IncRNAs showed the most potent prognostic value.

**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 ($\beta_i$): risk score = expression of gene $1^*\beta_1$ gene 1 $+$ expression of gene $2^*\beta_2$ gene 2 $+$ …expression of gene $n^*\beta_n$ gene n. The “$\beta$” 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).

ACKNOWLEDGMENTS AND FUNDING

This work was supported by the fund of China National Natural Science Foundation (NO.81560371) and Guangxi Medical University Training Program for Distinguished Young Scholars (2017). The authors thank the data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

CONFLICTS OF INTEREST

None.

REFERENCES


