Comprehensive Characterization of Prognostic Long Noncoding RNAs in Osteosarcoma

Hua Gao, Yuanyuan Guo, Miaomiao Zhang, and Zuqiang Yi

Department of Oncology, Henan Provincial Hospital, Zhengzhou Airport Economy Zone, Zhengzhou City, Henan Province, China 450000
Department of Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou City, Henan Province, China 450000
Department of Osteology, Henan Provincial Hospital, Zhengzhou City, Henan Province, China 450000

Correspondence should be addressed to Yuanyuan Guo; guoyuanyuan@gs.zzu.edu.cn and Zuqiang Yi; 15936275945@163.com

Received 27 May 2020; Revised 17 July 2020; Accepted 22 July 2020; Published 25 August 2020

Guest Editor: Tao Huang

Copyright © 2020 Hua Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The molecular mechanism of osteosarcoma (OS) based on protein-coding genes has largely been studied in the past decades. However, much remains to be explored when it comes to the role that long noncoding RNAs (lncRNAs) play in the pathogenesis and progression of OS and how they are associated with OS metastasis. In the present study, we collected RNA-seq-based gene expression data of 82 OS samples from the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) database, along with their clinical information. We found that 50 lncRNAs were significantly associated with patients’ survival by univariable Cox regression model. Moreover, we built multivariable Cox regression model based on 7 lncRNAs and successfully stratified patients into two risk groups, which exhibited significantly different prognostic outcomes. Significantly enriched Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways detected by differential expression analysis on DEGs between the two groups with different prognostic outcomes were both immune-related, indicating that such GO terms and pathways are critical for OS survival. Among the seven lncRNA signatures, AC011442.1 was predicted to act as an oncogenic driver in OS by correlation analysis of copy number alteration (CNA) and lncRNA expression, and it was predicted to regulate AMPK and hedgehog signaling pathways. In summary, the identification of novel prognostic lncRNAs in OS could not only improved our understanding of the lncRNAs involved in OS tumorigenesis or progression but also assist the diagnosis and development of molecularly targeted therapies for OS, which in turn benefit patients’ survival.

1. Introduction

Osteosarcoma (OS) is among the most prevalent malignancies in children and adolescents [1]. According to previous study, it takes up approximately 20% of all bone cancers, which also makes it one of the most common primary skeletal tumors [2, 3]. For example, according to the American Cancer Society, the estimated number of newly diagnosed cases of skeletal malignancies in 2017 would reach 3,260 in the United States, among them there would be roughly over 600 OS patients [4]. Unfortunately, over one-fifth of osteosarcoma patients exhibit lung metastasis at the time of diagnosis, which often results in unsatisfactory prognosis [5]. No significant improvement in 10-year overall survival of OS patients has been observed since the 1990s [6]. The magical effect of traditional tumor resection surgery and chemotherapy seems to encounter a bottleneck as they had once improved overall 10-year survival of OS from 30% to about 50% in the 1970s, and with the advances in molecular biology and related techniques, molecularly targeted therapies have since emerged as a new option in the management strategy of various cancers, including OS.

It is crucial for the development of molecularly targeted therapies to identify metastatic-related biomarkers and underlying mechanism in OS, in order to deliver a more accurate prognosis prediction and therapeutic decisions [7]. Long noncoding RNAs (lncRNAs) attract researchers’ keen attention worldwide as they play a critical role through...
82 OS samples with clinical characteristics

Univariable Cox model

50 prognostic lncRNAs

Random forest algorithm

Top 20 lncRNAs

Multivariable Cox model

Seven-lncRNA model

Risk stratification

High vs. low risk groups

Functional enrichment analysis

CNA-related oncogenic driver lncRNA

Differentially expressed immune markers

Figure 1: The study design and the expression profiles of the 50 prognostic lncRNAs in OS. (a) The workflow of the present study. (b) The lncRNAs were clustered by hierarchical clustering algorithm, and the samples were ordered by survival status and survival time.
epigenetic, transcriptional, and posttranscriptional mechanisms in diverse biological processes, such as tumor initiation, growth, and metastasis [8]. Though lncRNAs are not to be translated into proteins, they can function as key regulators through interacting with miRNAs, mRNAs, and proteins [8]. Many lncRNAs are identified to exert oncogenic or tumor suppressor functions in OS, such as ZEB1-AS1 [9], SPRY4-IT1 [10], BCAR4 [11], and MFI2 [12]. For example, previous studies have reported that lncRNA DANCR could function as a competitive endogenous RNA in OS, thereby promoting ROCK1-mediated proliferation and metastasis [13]. CEBPA-AS1, an antisense RNA of CEBPA, has the capability of inhibiting proliferation and migration and promoting apoptosis in OS via Notch signaling [14]. These studies demonstrate that lncRNAs can regulate the progression, metastasis, and prognosis of OS [15].

In this study, RNA-seq data and clinical information of patients with osteosarcoma from the TARGET database were processed with univariable Cox regression and random forest algorithm, and we selected seven long noncoding RNAs (lncRNAs); all of them have the potential to affect the survival of osteosarcoma patients and to construct a prognosis risk model. Based on the stratification offered by our model, the corresponding biological differences among osteosarcoma patients and how these characteristics would result in varied prognostic outcomes were further explored and explained.

## 2. Materials and Methods

### 2.1. Data Resources

We downloaded RNA-seq-based gene expression data (TPM, transcript per million), somatic copy number alteration (SCNA) data, and clinical data of 82 corresponding osteosarcoma patients from the TARGET (Therapeutically Applicable Research to Generate Effective Treatments) database [16]. The segmented SCNA was annotated by Ensembl gene annotation v37.75 [17]. The SCNA status for each was called as gain or loss only if the log2 ratio (tumor/normal copy numbers) was more than 0.6 or less than -0.6. To meet the requirement for data analysis, we only collected 82 osteosarcoma samples with matched SCNA, gene expression, and clinical data.

### 2.2. Selection of lncRNAs in OS for Prognostic Risk Model Construction

First, based on 9 biotypes for lncRNAs (which were 3prime_overlapping_ncRNA, antisense, lincRNA, macro_lncRNA, non_coding, sense_intronic, sense_overlapping, bidirectional_promoter_lncRNA, and retained intron) in Ensembl, we obtained a total of 3,159 lncRNAs that exhibited TPM (transcript per million) > 0.1 in more than half of the samples. The expression status of lncRNAs were firstly classified into high and low expression, respectively, based on the median of the expression levels. Combined with the clinical information, univariable Cox regression analysis was then performed with package Survival v3.1-11 in R v3.6.3 to pick up lncRNAs significantly related to the survival of the patients (log-rank test, \( P < 0.05 \)). Utilizing the random forest algorithm in R package randomForestSRC with default options, we evaluated ranked those lncRNAs and built a multivariable Cox model based on the top 20 prognostic lncRNAs. Subsequently, we only retained the prognostically insignificant lncRNAs in the initial multivariable Cox model (\( P > 0.05 \)) and built the optimal multivariable Cox model based on these prognostic lncRNAs.

### 2.3. Model Construction for Evaluating Osteosarcoma Prognostic Risk

Taking into consideration the expression of qualified lncRNAs in each patient and the patient’s survival status, we applied multivariable Cox regression with survival package in R v3.6.3 to build our osteosarcoma prognosis risk model, and lncRNAs with significant contribution to the model were selected. These lncRNAs were used to construct a risk-scoring method, which assigned a score that reflected the risk of death to each osteosarcoma patient. The patients were then divided by the median score into the high-risk and low-risk groups, accordingly. We visualized the survival curves of the two groups of patients by the Kaplan-Meier method and assessed the differences between the two groups by log-rank test.

### 2.4. Functional Enrichment Analysis of the Dysregulated Genes in the Two Risk Groups

As osteosarcoma patients were categorized, their gene expression profiles fell into two groups, accordingly. Utilizing the screening criteria of log2 (fold change) > 1 and \( P \) value < 0.05, genes with significant differential expression between the two groups were selected. Subsequently, Gene Ontology (GO) [18] and Kyoto
Figure 2: Continued.
Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed on identified differentially expressed genes with the package clusterProfiler v3.12.0 in R v3.6.3.

2.5. Estimation of Immune Cell Infiltrating Levels. The infiltrating levels of immune cells were estimated based on the gene expression profiles and marker genes of immune cells. Single-sample gene set enrichment analysis (ssGSEA) was
Table 2: The comparative analysis of the risk score with other clinical factors in univariable and multivariable Cox regression models.

<table>
<thead>
<tr>
<th>Features</th>
<th>Univariable Cox regression</th>
<th>Multivariable Cox regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P value</td>
<td>HR</td>
</tr>
<tr>
<td>Risk score</td>
<td>6.82E-12</td>
<td>19.7</td>
</tr>
<tr>
<td>Gender (female/male)</td>
<td>0.30</td>
<td>0.68</td>
</tr>
<tr>
<td>Race (white/other)</td>
<td>0.23</td>
<td>0.64</td>
</tr>
<tr>
<td>Age</td>
<td>0.82</td>
<td>1</td>
</tr>
</tbody>
</table>

employed in this study. This analysis was implemented in R GSVA v1.32.0 package [20].

3. Results

3.1. Identification of Prognostic IncRNAs in OS. As shown in Figure 1(a), the present study conducted a series of data analysis to build a predictive model for OS risk. The gene expression and clinical information of 84 osteosarcoma patients were obtained using the TARGET database, among which, two samples were excluded due to a lack of overall survival time. Based on the gene annotation from the Ensembl database and criteria regarding TPM, we selected 3,159 noncoding RNA for later establishment of the prognostic risk model (see Materials and Methods). Among these 3,159 IncRNAs, we identified 50 IncRNAs significantly associated with patients’ overall survival by univariable Cox regression model (log-rank test, P values < 0.05, Supplementary Table S1). As illustrated in Figure 1(b), the expression of the prognostic IncRNAs were significantly differentially expressed between the alive and deceased OS patients. These results indicated that the prognostic IncRNAs identified by the univariable Cox regression model may be essential for OS tumorigenesis and/or progression.

3.2. Construction of IncRNA-Based Multivariable Cox Model for Risk Prediction in OS Patients. To build an IncRNA-based Cox regression model for OS risk prediction, we first ranked the prognostic IncRNAs by random forest algorithm, and the top 20 IncRNAs were considered candidates for the construction of an OS prognostic risk model. We then built our model with multivariable Cox regression on the samples with clinical information and expression data of these IncRNAs and obtained seven IncRNAs that significantly contributed to the model (Table 1). Based on the multivariable Cox model, the OS patients were divided into two risk groups using the median risk score. As shown in Figure 2(a), the proportion of deceased samples in the high-risk group (high-risk) was much greater than that in the low-risk group (low-risk) (25/41 vs. 4/41, test of proportion, P < 0.05). Moreover, compared with the low-risk group, patients in the high-risk group exhibited significantly lower overall survival time (33.6 vs. 68.7, log-rank test, P = 1.54E-5). Furthermore, patients were then divided into the high-/low-expression groups based on the expression profiles of these seven IncRNAs, respectively. The Kaplan-Meier curves showed a significant association of the seven IncRNAs with overall survival of patients with OS (Figures 2(b)–2(h)). Consistently, the risk score was observed to have a higher statistical significance than any of the seven prognostic IncRNAs (Figure 2(i)).

In addition, to assess the independence of this scoring in predicting patients’ prognosis, we performed both univariable and multivariable Cox regression for samples using the precalculated risk scores and their clinical information such as gender, race, and age. We found that this risk score was an independent indicator for OS patients’ survival (Table 2), further suggesting that the risk score by the seven-IncRNA-based Cox model had the potential to predict the risk of OS patients.

3.3. Functional Characterization of Dysregulated Genes in High-Risk and Low-Risk Groups. To investigate dysregulated genes in the two risk groups, we compared the gene expressions of these two risk groups. With thresholds at |log2 (fold change)| >1 and P value < 0.05, we identified 864 significant differentially expressed genes (DEGs), and when compared with the low-risk group, the expression of 728 gene was significantly upregulated in the high-risk group, and the expression of another 136 genes was downregulated (Figure 3(a)).

The GO and KEGG pathway enrichment analyses proved that the immune microenvironment of osteosarcoma patients played a crucial role in OS progression. It can be learned that the top 10 GO terms exhibited close association with immunity, including inflammatory responsive response T cell activation, humoral immune response, lymphocyte-mediated immunity, axonemal dynein complex assembly, positive regulation of T cell activation, and regulation of leukocyte cell-cell adhesion (Figure 3(b)), suggesting that the varied immune environment between the high- and low-risk groups may result in their prognostic differences. What is more, from the KEGG pathway enrichment analysis, we observed that a majority of the pathways, where these differentially expressed genes were significantly enriched, consisted of immune-related ones, such as NK cell-mediated cytotoxicity, staphylococcus aureus infection, Th1 and Th2 cell differentiation, antigen processing, and presentation (Figure 3(c)). The consistence between the GO and KEGG enrichment analyses further demonstrated the immune-related biological process may play a key role in OS progression.

3.4. AC011442.1 May Act as an Oncogenic Driver IncRNA in OS. As IncRNAs upregulated or downregulated by copy number alterations (CNA) probably acted as driver IncRNAs in cancer, we performed correlation analysis of the expression level and the corresponding copy number status for the seven prognostic IncRNAs in the multivariable Cox
Figure 3: Continued.

- Axonemal dynein complex assembly
- Positive regulation of T cell activation
- Acute inflammatory response
- Regulation of leukocyte cell-cell adhesion
- Lymphocyte mediated immunity
- Adaptive immune response based on somatic recombination of immune receptors
- Humoral immune response
- Immune response-regulating cell surface receptor signaling pathway
- T cell activation
- GeneRatio

(a) 

(b)
We observed that AC011442.1 was highly upregulated in samples with CNA as compared with wild-type samples \((P < 0.001\), Figure 4(a)). Notably, the copy numbers of the four genes were frequently gained in OS samples (frequency > 10%).

To further investigate the biological function of the four lncRNAs, we conducted gene set enrichment analysis on protein-coding genes that highly correlated with identified lncRNAs. We found that AC011442.1 was significantly and positively correlated with genes involved in the AMPK signaling pathway and hedgehog signaling pathway, respectively (Figures 4(b) and 4(c), \(P\) value < 0.05). These results indicated that AC011442.1 may enhance the activities of the AMPK signaling pathway and hedgehog signaling pathway.

### 3.5. The Immune Markers Associated with OS Prognosis

To further explore the immune cells and related markers associated with OS prognosis, we first examined the expression patterns of the immune markers. Specifically, the immune inhibitory genes such as BTN3A1, CD48, HAVCR2, LAG3, and TIGIT were significantly upregulated in the low-risk group (Figure 5(a), \(P < 0.01\)), suggesting that the anticancer activity of the immune cells might be suppressed by these inhibitory genes. Furthermore, we also observed that the relative infiltrating levels of CD8 T cells and activated natural killer cells were attenuated in the high-risk group (Figure 5(b), \(P < 0.01\)), suggesting that the worse survival in the high-risk group of OS may be caused by the lack of CD8 and NK cells. Consistently, the marker genes of CD8 and NK cells, CD8A, CD8B, GZMA, and NCR3 were also downregulated in the high-risk group. These findings indicated that the immune cells and related markers were highly associated with OS prognosis.

### 4. Discussion

The molecular mechanism of OS based on protein-coding genes has largely been studied in the past decades. Despite extensive researches about the molecular mechanism of OS, there is still a lack of understanding of the lncRNAs’ role in OS tumorigenesis, progression, and metastasis. Meanwhile, the identification of the prognostic lncRNAs involved in OS can facilitate the development of new diagnostic or therapeutic biomarkers.
Figure 4: The oncogenic driver lncRNA AC011442.1 and its functionality. (a) The expression patterns of AC011442.1 in OS patients with CNAs and wild type. (b) The predicted pathways that the AC011442.1 might participate in.
Figure 5: Continued.
In the present study, we collected 82 OS samples with RNA-seq-based gene expression data and their clinical information from the TARGET database. We found that 50 lncRNAs were significantly associated with patients’ survival by a univariable Cox regression model (P values < 0.05). Through the selection of prognostic lncRNAs, we identified 7 lncRNAs with significant performance in OS survival prediction, built multivariable Cox regression model under the 7 lncRNAs, and successfully stratified patients into different risk groups with distinctive survival outcomes. Notably, DDN-AS1, one of the seven lncRNAs used by the multivariable Cox regression model, has been reported to act as competing endogenous RNA (ceRNA) that promoted the expression of TCF3 through competitively binding miR-15a and miR-16 [21], suggesting that DDN-AS1 may promote OS progression in a similar manner. We further analyzed gene expression profiles of patients in different risk groups and obtained a list of DEGs. Functional enrichment analysis revealed that significantly enriched GO terms and pathways were associated with many aspects of immunity, indicating that immune-related functions are critical for OS survival, which is consistent with previous studies [22].

As dysregulated lncRNAs caused by copy number alterations (CNA) may act as driver lncRNAs in cancer, correlation analysis of the expression level and the corresponding copy number status for the seven prognostic lncRNAs was performed to identify the driver lncRNAs. Notably, AC011442.1 was also predicted as one of the four-driver lncRNA (P < 0.001). Interestingly, AC011442.1 was predicted to participate in cancer-related pathways [23–25], including the AMPK signaling pathway and hedgehog signaling pathway. Hedgehog signaling pathways have been frequently observed to drive tumorigenesis and metastasis of OS [26]. These results further demonstrated that the driver lncRNAs played a key role in OS, which could be used for further research of molecular mechanism.

As the exploration into varied molecular patterns between the two risk groups revealed that the immune-related pathways were enriched by DEGs in OS, we then examined whether the abundance of immune cells and markers were associated with OS prognosis. Specifically, the immune inhibitors such as BTN3A1, CD48, HAVCR2, LAG3, and TIGIT; CD8 T and activated NK cells; and related markers were significantly downregulated in the high-risk group. Particularly, CD48, HAVCR2, LAG3, and TIGIT were identified as novel immunotherapeutic targets of several cancers [27–30], suggesting that the low-risk OS patients might benefit from their candidate inhibitors.

In addition, the limitations of this study should be pointed out. Firstly, the multivariable Cox regression model needs an independent gene expression data for the validation of its robustness. Secondly, though a list of dysregulated lncRNAs associated with OS survival was identified, but future experimental verification is still needed. Moreover, detailed molecular functions of identified dysregulated lncRNAs had not been thoroughly discussed in this study. We hope that, when validation datasets become available in the near future, we can further confirm our findings and perform experimental validation. In summary, the identification of novel prognostic lncRNAs in OS would not only improve our understanding of the lncRNAs involved in OS tumorigenesis or progression but also assist the prediction of OS survival and development of molecularly targeted therapies to some extent, which in turn benefit patients’ survival.

Data Availability

Previously reported gene expression and clinical data were used to support this study and are available at TARGET (Therapeutically Applicable Research To Generate Effective Treatments) database (https://ocg.cancer.gov/programs/target/data-matrix). These prior studies (and datasets) are cited at relevant places within the text as references [14].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Y.G. and Z.Y. designed this study. H.G., Y. G., and M. Z. conducted the experiments. H.G., Y. G., and Y. Z. contributed to the writing of the paper and setting of figures. Hua Gao and Yuanyuan Guo contributed equally to this work.
Supplementary Materials

Supplementary Table S1: the 50 prognostic long noncoding RNAs and statistical significance (P value) in osteosarcoma. (Supplementary Materials)

References