# Screening and verification of prognostic lncRNA markers related to immune infiltration in the metastasis of osteosarcoma

**Chuan He, Xiaolong Wang, Linying Ni, Chunyu Song, Wei Mai**

Department of Orthopedics, Harbin Medical University Cancer Hospital, Harbin, China

*Contributions:* (I) Conception and design: C He; (II) Administrative support: W Mai; (III) Provision of study materials or patients: L Ni; (IV) Collection and assembly of data: C Song; (V) Data analysis and interpretation: X Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Wei Mai. Department of Orthopedics, Harbin Medical University Cancer Hospital, Harbin 150081, China.
Email: maiwei210@126.com

**Background:** We sought to screen and verify the long non-coding ribonucleic acids (lncRNAs) related to immune infiltration in metastatic osteosarcoma (OS).

**Methods:** We downloaded the RNA-sequencing expression data from The Cancer Genome Atlas (TCGA) database as the training data set. We downloaded the GSE39055 data set from the National Center for Biotechnology Information, Gene Expression Omnibus as the validation data set. The least absolute shrinkage and selection operator (LASSO) regression algorithm was used to screen the optimized lncRNA combinations. Kaplan-Meier curves were used to evaluate the associations between the lncRNAs and actual prognosis. The independent survival prognosis clinical factors were obtained by univariate and multivariate Cox analyses. A nomogram was established to explore the correlation between survival rate and risk information. The Tumor IMmune Estimation Resource was applied to estimate the composition of 6 subtypes of immune infiltration cells.

**Results:** In total, 1,398 lncRNAs and 14,631 messenger RNAs were screened from TCGA data set, and divided into the low and high immunity groups. The Estimation of STromal and Immune cells in MAlignant Tumour tissues using Expression data (ESTIMATE) scores differed significantly between the samples in the two groups. Additionally, 5 optimized lncRNA combinations were obtained using the LASSO algorithm. Risk factors including age, metastatic tumor, and risk-score (RS) were related to the prognosis of OS patients. The survival rates predicted by the nomogram model were consistent with the actual survival rates of OS patients. Finally, we found that RS was negatively correlated with the proportion of immune cells.

**Conclusions:** In total, 5 feature lncRNAs were identified as novel biomarkers for OS. Next, a RS nomogram model was constructed based on the 5 identified lncRNAs. This model predicted the survival rates and prognoses of OS patients well.

**Keywords:** Osteosarcoma (OS); metastasis; long non-coding ribonucleic acid (lncRNA); immune infiltration

## Introduction

Osteosarcoma (OS) is a common clinical malignant bone tumor (1,2). Despite advances in the treatment of OS, nearly 80% of patients retain the diseased limbs, the 5-year survival rate has increased from about 20% to nearly 80%, and >50% of patients still die from metastatic OS (3,4). In recent years, with the continuous improvement of multi-stage, multi-factor, and multi-step theories about the regulation of tumor metastasis factors, we have extended our understanding of the extremely complex pathological process of tumor metastasis, and discovered and confirmed that biomolecules are highly related to the process of tumor metastasis (5).

Biomolecules have become a hot spot in the research field of tumor metastasis, as research on biomolecules allows us to better understand the mechanisms underlying tumor metastasis (6). Further, and more importantly, these molecules and their products may become targets for anti-tumor metastasis or indicators for observing the prognosis and metastasis of tumor patients (7,8).

Among the various cell types related to the progression and development of tumors, the effect of tumor-infiltrating lymphocytes on prognosis has been extensively studied. Previous research has shown that the level extent of tumor infiltrating lymphocytes is a novel supplementary biomarker for predicting recurrence and mortality (9,10). In addition to lymphocytes, tumors usually contain different non-lymphocyte immune cells (11,12), which are regarded as having unique effects on prognosis in different types of cancer (13). However, traditional methods for examining tumor immune cell infiltration (e.g., immunohistochemistry or flow cytometry) do not completely assess the immune function of different types of cells. This is mainly because the existing technology is limited as to the number of immune signatures that can be detected.

Long non-coding ribonucleic acid (lncRNA) is a type of RNA molecule whose transcript length exceeds 200 nt (14), and is a new type of gene regulatory factor. Studies have shown that lncRNA is closely associated with the occurrence, development, and prognosis of human diseases, especially tumors (15,16). Further research has shown that the abnormal expression of lncRNA might be associated with epithelial-mesenchymal transition, cell apoptosis, metastasis, invasion, and the poor prognosis of OS patients (17). Yu *et al.* (18) identified 5 metastasis-associated lncRNAs (including LAMA5-AS1, RP5-894D12.4, RP11-231I13.2, RP11-128N14.5 and RP11-346L1.2) which are regarded as potential prognostic indicators for OS. Deng *et al.* (19) established a four-methylated lncRNA signature (including MAP3K14, SNHG12, DSCR8 and IGF2BP2-AS1) for determining OS patient prognosis. In recent years, the identification of lncRNAs that mediate the immune microenvironment of OS patients has been a research hotspot (20,21). Previous research has shown that there are many different immune cell infiltration types in OS, not only in the tumor matrix, but also in the cancer nest. Additionally, OS prognosis is associated with the type and the number of immune cell infiltrations around the tumor (22,23).

In the present study, we assessed the immune infiltration cells in OS tumor samples downloaded from The Cancer Genome Atlas (TCGA) based on a single-sample gene set enrichment analysis (ssGSEA) algorithm. We then divided the samples into a high immune cell infiltration group and a low immune cell infiltration group. After combining the sample metastasis information, we identified the RNAs associated with metastasis and immunity. The Cox regression analysis and least absolute shrinkage and selection operator (LASSO) algorithm were then used to identify the prognostic lncRNA markers. Finally, based on the prognostic lncRNA markers, we constructed and verified the survival prognosis prediction model. Compared with the previous studies, we adopted a larger sample size and the prediction performance of the prognostic model based on the five-lncRNA signature was quite good. Our study provided valuable lncRNAs as promising prognostic biomarkers for OS metastasis. We present the following article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-1926/rc).

## Methods

### Data source and preprocessing

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). We downloaded the RNA-sequencing expression data from TCGA database, including 265 samples from the Illumina HiSeq 2000 RNA Sequencing platform. After the 265 samples were matched with the clinical information of the OS samples, the samples with information on metastasis were retained. In total, 176 OS samples were included in the analysis, and these were used as the training data set. Additionally, the GSE39055 data set (24), which included 37 OS samples, was obtained from National Center for Biotechnology Information, Gene Expression Omnibus (25) (https://www.ncbi.nlm.nih.gov/geo/) based on the platform of Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip. The GSE39055 data set was used as the validation data set.

The lncRNAs and mRNAs in TCGA data set were annotated according to the annotation file of TCGA detection platform. Next, we obtained the Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip platform from the Ensembl genome browser 96 (http://asia.ensembl.org/index.html), including probe, gene symbol, RNA type, and other information. We then re-annotated the detection probes in the downloaded expression profile data set to screen the corresponding expressions of the mRNAs and lncRNAs.

### Construction and evaluation of OS immune grouping

A gene set variation analysis of the microarray and RNA-sequencing data (26) (version 1.36.3, http://www.bioconductor.org/packages/release/bioc/html/GSVA.html) based on the ssGSEA algorithm (27) was conducted to assess the type of immune infiltration in the OS samples. We grouped the immune infiltration types according to the results of the ssGSEA and named them the "Immunity_H" and "Immunity_L" groups. The estimate package in R (28) (version 3.6.1, http://127.0.0.1:29606/library/estimate/html/estimateScore.html) was applied to calculate the stromal, immune, and ESTIMATE scores. Next, the differences in the scores of the immune infiltration groups were compared to verify the correction of the immune infiltration groups. Additionally, CIBERSORT (29) (https://cibersort.stanford.edu/index.php) was used to calculate the proportions of various immune cells on the basis of the expression level of TCGA OS tumor samples. Next, the difference in the composition ratios of various immune cells between the immune infiltration groups was compared to verify the correction of the immune infiltration groups.

### Screening of significantly differentially expressed RNAs

According to the Immunity_H and Immunity_L grouping and the information as to whether or not there was tumor metastasis, limma package (30) (version 3.6.1, https://bioconductor.org/packages/release/bioc/html/limma.html) was used to screen the differentially expressed genes (DEGs) with a cutoff of false discovery rate (FDR) <0.05 and $|\log_2$ fold change (FC)| >0.263. We then compared the sets of DEGs filtered between the two comparison groups and selected the intersection genes for further analysis.

### Construction of prognostic model

A univariate Cox regression analysis was conducted using survival package of R (version 3.6.1, http://bioconductor.org/packages/survivalr/) to identify the lncRNAs related to survival prognosis in TCGA data set. The LASSO regression algorithm in the lars package (version 1.2) (31) of R (version 3.6.1, https://cran.r-project.org/web/packages/lars/index.html) was then used to perform the survival regression analysis to obtain the optimal lncRNAs based on the lncRNAs related to survival prognosis.

According to the LASSO coefficient of each element in the optimized lncRNAs and their expression in TCGA training data set, we constructed the risk-score (RS) model. The following RS formula was used:

$$RS = \sum Coef_{lncRNA} \times Exp_{lncRNA} \qquad [1]$$

where $Coef_{lncRNA}$ represented the LASSO coefficient of the target lncRNA, and $Exp_{lncRNA}$ represented the lncRNA expression in TCGA training data set.

We also calculated the RSs in TCGA training data set and the GSE39055 validation data set. Additionally, we separated TCGA training data set and the GSE39055 data set into high- and low-risk groups based on the RS median value. Next, Kaplan-Meier (KM) curves of survival package (32) in R (version 3.6.1) were used to evaluate the correlations between the high- and low-risk groups and the actual prognosis information. Additionally, based on the lncRNAs related to the RS model in TCGA training data set, the KM curves of survival package (32) (version 2.41-1) in R (version 3.6.1) were used to assess the associations between the different expressions of lncRNAs and actual prognosis information.

### Construction of nomogram survival rate model with the independent survival prognostic factors

The univariate and multivariate Cox regression analyses of survival package (version 2.41-1) in R (version 3.6.1) were used to screen the independent survival prognosis clinical factors with a cutoff of log-rank P value < 0.05. Next, to further study the prognostic independence between the clinical prognostic factors and RS factors, the rms package (version 5.1-2) in R3.6.1 (https://cran.r-project.org/web/packages/rms/index.html) (33) was used to construct nomogram 3- and 5-year survival rate prediction models on the basis of independent prognosis factors and risk information in the RS model. We then calculated the concordance index (C-index) coefficient of the nomogram prognostic model. The C-index is an evaluation index for assessing the predictive ability of the model (34), and was calculated using the survcomp package in R (version, 1.34.0 http://www.bioconductor.org/packages/release/bioc/html/survcomp.html) (35). A C-index >0.70 indicates that the model had good predictive ability.

### Construction and functional analysis of the characteristic lncRNA correlation network

The cor function in R 3.6.1 (http://77.66.12.57/R-help/cor.test.html) was used to calculate the Pearson correlation

coefficients (PCCs) between the lncRNAs in the RS model and the overlapping DEGs obtained by comparing the immunity_H and immunity_ L groups and the groups with and without tumor metastasis. Subsequently, we retained the pairs for which the absolute value of the PCC was >0.3 and the P value was <0.05 to construct the lncRNA-mRNA co-expression network, which was visualized using Cytoscape (36) (version 3.6.1, https://cytoscape.org/). Finally, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (37,38) (version 6.8, http://metascape.org/) to perform the Gene Ontology biological process (BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the DEGs in the co-expression network with a threshold of P value <0.05.

### Correlations between the prognostic signals of the feature RNAs and immune cell subtype infiltration

The Tumor IMmune Estimation Resource (TIMER) contains many conventional cancer data sets in TCGA data set, which can estimate the composition of 6 types of immune infiltrating cells (i.e., B cells, cluster of differentiation (CD)4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells). Thus, the TIMER (39) (https://cistrome.shinyapps.io/timer/) was used to analyze the immune cells related to OS based on TCGA training data set. We then calculated the correlation between each type of immune cell subtype and the RS value based on TCGA samples to observe the correlation between the prognostic signal based on each feature lncRNA and the infiltration of various immune cell subtypes.

### Statistical analysis

All the statistical analysis was performed by R (version 4.1.2), and P value <0.05 was considered to be statistically significant.

### Results

### OS immune group construction and evaluation

A total of 1,398 lncRNAs and 14,631 mRNAs were obtained from the annotation of the lncRNAs and mRNAs in TCGA training data set. Next, the ssGSEA algorithm was used to evaluate the immune infiltration type of each sample based on the expression level of the OS tumor samples. As *Figure 1A* shows, the sample was clearly divided into two clusters: cluster 1 (comprising 73 samples) and cluster 2 (comprising 103 samples).

Additionally, the immune score, including the ESTIMATE score, immune score, and stromal score, were obtained using the ESTIMATE algorithm (see *Figure 1B*). We found that the ESTIMATE scores differed significantly between the samples in the Immunity_H group and Immunity_L group. Further, the scores of the samples in the Immunity_H group were significantly higher than those in the Immunity_L group. Additionally, the evaluation of the immune cell types based on CIBERSORT (see *Figure 1C*) showed that important immune cell types, such as naive B cells, also differed significantly between the Immunity_H group and Immunity_L group based on the ssGSEA evaluation grouping. All these results indicated that the Immunity_H and Immunity_L groups obtained based on the ssGSEA evaluation grouping could be used for the subsequent grouping analysis.

### Screening of significant DEGs

Based on the Immunity_H and Immunity_L groupings obtained by ssGSEA, and according to whether the sample was metastatic or not, a total of 3465 and 646 DEGs were screened from both the Immunity_H and Immunity_L groups, and the with or without tumor metastasis groups, respectively (see *Figure 2A*). Next, we identified a total of 315 overlapping DEGs, including 53 lncRNAs and 262 mRNAs by comparing the two groups (see *Figure 2B*). The overlapping DEGs were used in the further analyses. Specifically, the lncRNAs were used to construct prognostic models, and the mRNAs were used in the functional analyses of the feature lncRNAs.

### Prognosis analysis and model construction

Based on the 53 lncRNAs from the overlapping DEGs, we identified 9 lncRNAs associated with prognosis using a univariate Cox regression analysis. Further, 5 optimal lncRNA combinations were identified using the LASSO algorithm. Additionally, we developed the following RS calculation formula based on the LASSO regression coefficient of the expression of 5 lncRNAs in TCGA training data set:
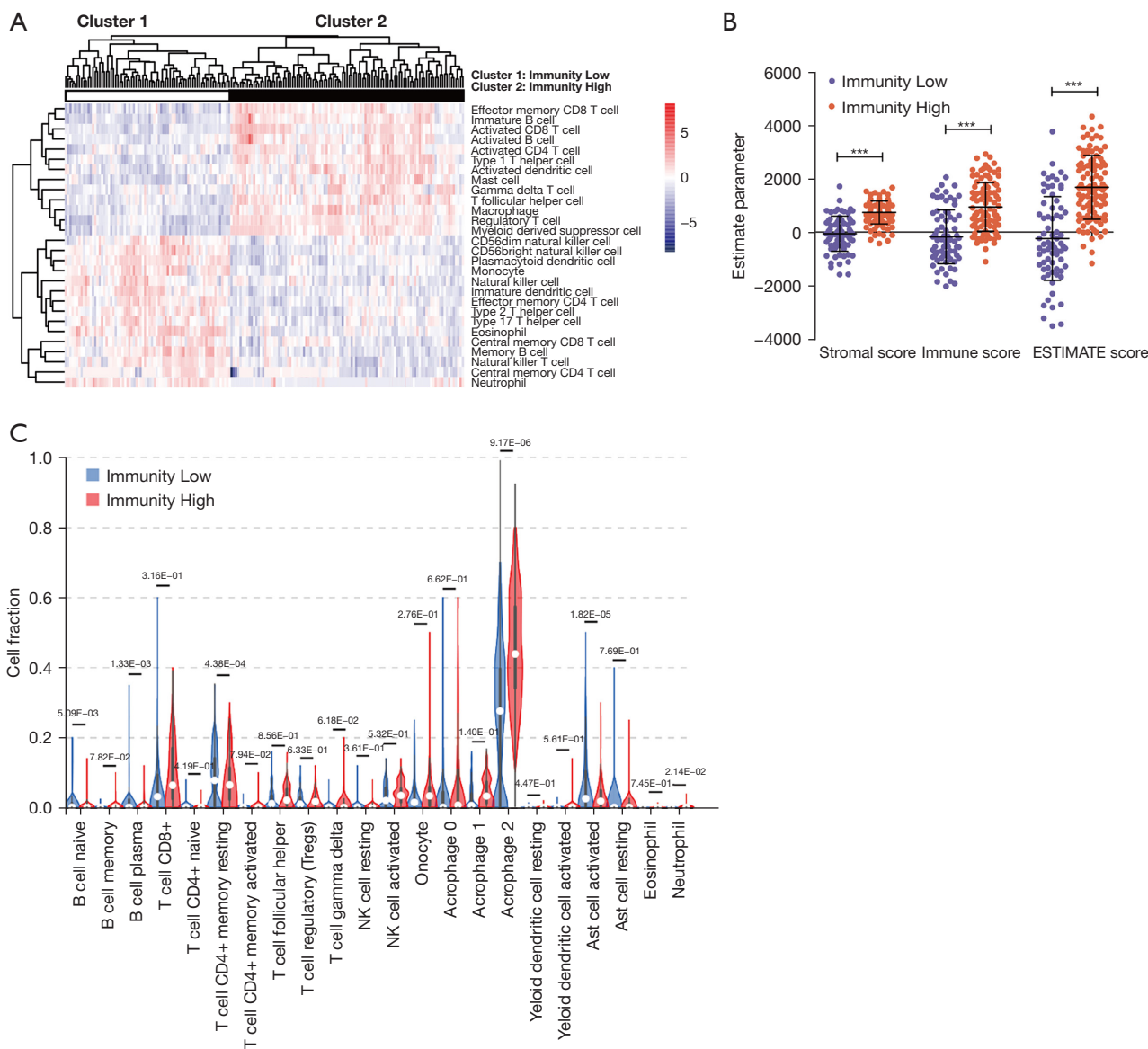
**Figure 1** OS immune group construction and evaluation. (A) Heatmap of the immune evaluation based on ssGSEA. (B) Scores obtained by the ESTIMATE algorithm in the Immunity_H and Immunity_L groups. The red and blue points represent the samples in the Immunity_H and Immunity_L groups, respectively. (C) Distribution of immune cell composition in Immunity_H and Immunity_L groups. ***, P<0.001. OS, osteosarcoma.

$$RS = (-0.090363017) \times Exp_{linc00243} + (-0.001107924) \times Exp_{linc00545}$$
$$+ (0.046224301) \times Exp_{linc00626} + (-0.047564194) \times Exp_{linc00672} \quad [2]$$
$$+ (-0.076902044) \times Exp_{linc00892}$$

We calculated the RSs of the samples in TCGA training data set and the GSE390555 validation data set. The

distribution of RSs, and the distribution of clinical survival information are shown in *Figure 3A,3B*.

Additionally, the 5 lncRNAs in the prognosis model were separated into high expression and low expression groups based on the median value in TCGA training data set. Next, KM curves were used to analyze the relationship
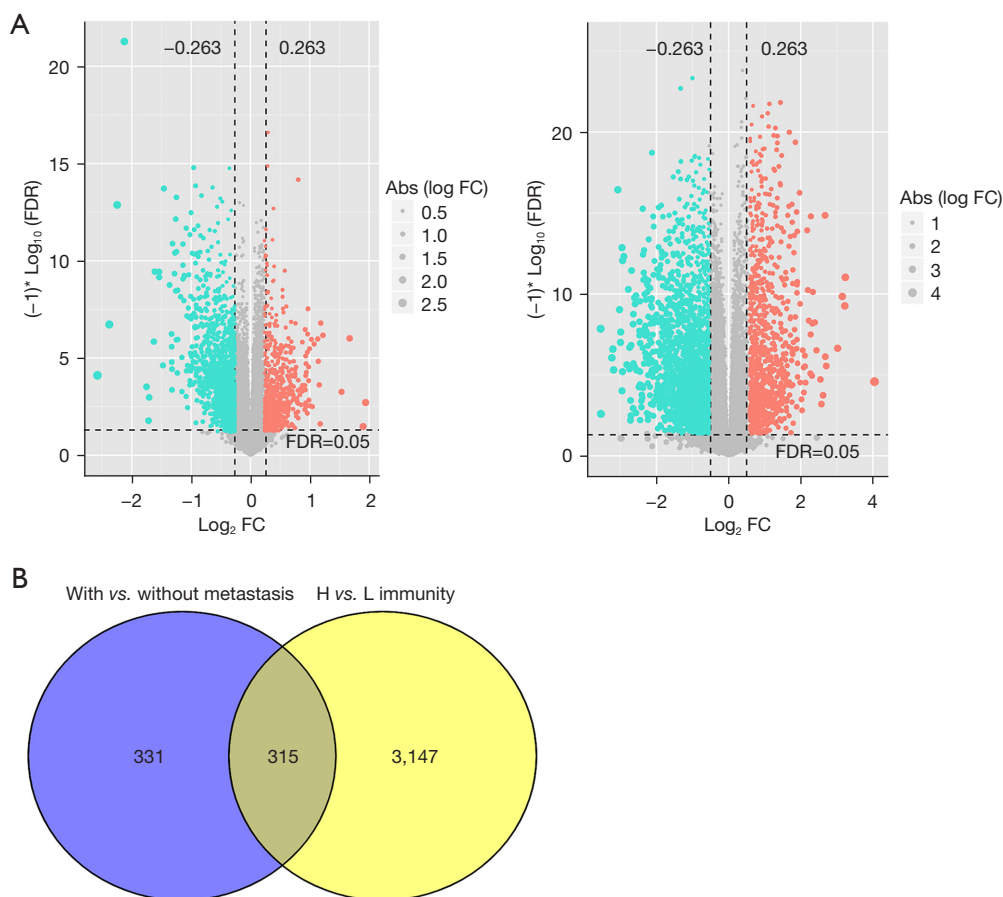
3240

He et al. Screening of lncRNAs for metastatic OS



**Figure 2** Screening of the significant DEGs. (A) Volcano map of DEGs in with *vs.* without tumor metastasis groups and in Immunity_H *vs.* Immunity_L groups; the red and green dots indicate significantly upregulated and downregulated RNAs, respectively; the horizontal dashed line indicates FDR <0.05; the 2 vertical dashed lines indicate |log2FC| >0.263. (B) Venn diagram of the comparison of the 2 sets of DEGs. DEGs, differentially expressed genes; FDR, false discovery rate; FC, fold change.

between the expression group and survival prognosis. As *Figure 4* shows, the high expression of linc00243 (P=0.029), linc00545 (P=0.00016), linc00672 (P=0.00015), and linc00892 (P=0.001) were significantly related to good survival. While the low expression of linc00626 was significantly related to good survival (P=0.014).

Additionally, KM curves were used to assess the correlations between the high and low risk groups and actual prognosis information in TCGA training data set and the validation data set, respectively. The results indicated that there was an obvious association between the different risk groups into which the samples had been divided based on the predictions of the RS model and the actual prognoses of patients in TCGA training data set and the GSE39055 validation data set (see *Figure 5A,5B*). Additionally, receiver

operating characteristic (ROC) curves were used to examine the prediction results of the 1-, 3-, and 5-year survival rates based on the RS prognosis model in TCGA training data set and the GSE39055 validation data set. The results indicated that the RS prognosis model predicted the survival rates of patients in TCGA training data set and the GSE39055 validation data set well (see *Figure 5C,5D*).

### Screening of independent clinical factors

In total, 3 independent clinical factors related to prognosis (i.e., age, metastatic, and RS model status) were screened by the univariate and multivariate Cox regression analyses (see *Table 1*).

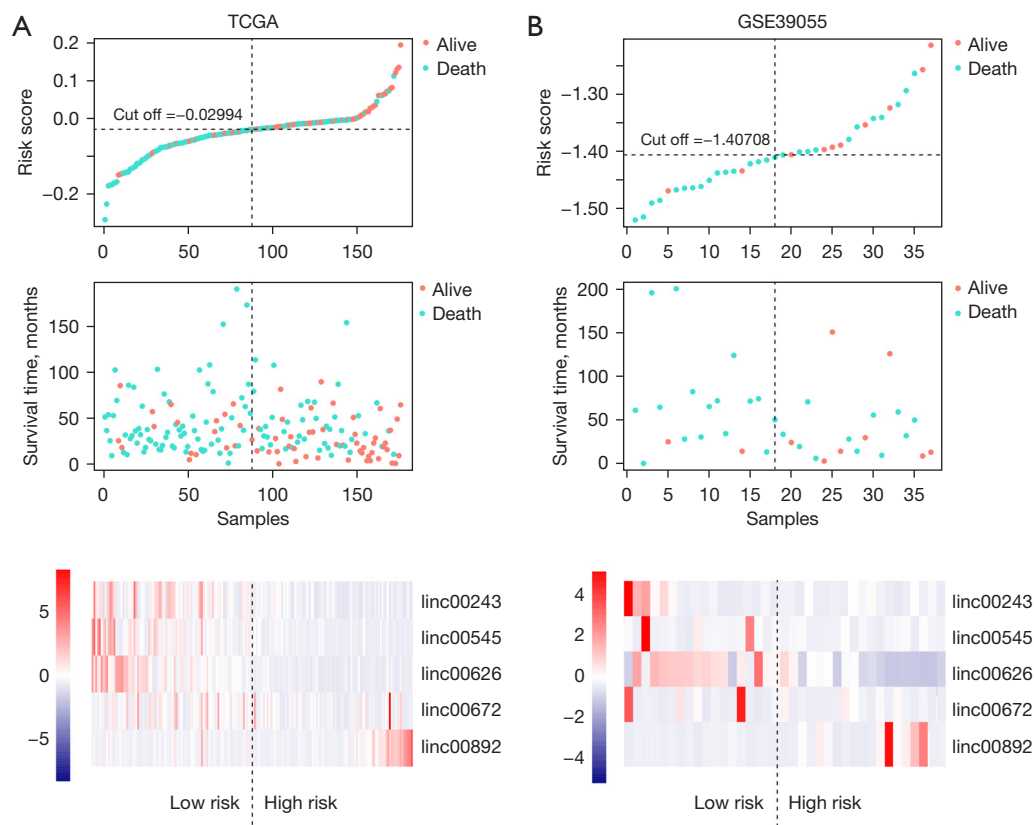Additionally, to analyze the correlation between

**Figure 3** Screening of the optimized lncRNA combinations. Distribution of RSs, survival prognostic information, and the heatmap of 5 characteristic lncRNAs in the samples in TCGA training set (A) and GSE39055 validation set (B). TCGA, The Cancer Genome Atlas; RSs, risk-scores; lncRNAs, long non-coding ribonucleic acids.

the independent clinical factors (i.e., age, a metastatic tumor, and RS model status) and survival prognosis, we conducted a nomogram survival rate model to examine the 3- and 5-year overall survival rates for the OS samples, and the results are shown in *Figure 6A*. Subsequently, the points of each indicator were summed, thereby predicting the survival probabilities at 3 and 5 years. A calibration curve was also constructed to assess the nomogram model results. As shown in *Figure 6B,6C*, the 3 and 5 years survival time with C-index of 0.7459 and 0.7328 indicated that the nomogram model could well predict the survival rates for patients with OS.

### Construction and functional analysis of the feature lncRNA correlation network

A total of 239 significant pairs related to OS with a threshold of PCC >3 and P value <0.05 were identified based on the expression levels of the lncRNAs in the RS

model and the overlapping DEGs in TCGA training data set. Subsequently, we constructed a co-expression network of the lncRNAs and mRNAs involved with the 5 feature lncRNAs (i.e., linc00243, linc00892, linc00626, linc00545, and linc00672) and 143 mRNAs (e.g., LILRB4, MMP13, and CHRD) (see *Figure 7*).

Next, we performed the DAVID-based BP and KEGG pathway enrichment analysis of the genes in the co-expression network, and a total of 16 significantly related BPs, such as signal transduction and positive regulation of interferon-gamma production, and 11 KEGG pathways, such as cytokine receptor interaction, were screened (see *Table 2*).

### Correlations between the feature lncRNA prognostic signals and immune cell subtype infiltration

The online TIMER tool was applied to analyze the proportion of immune cells associated with OS according
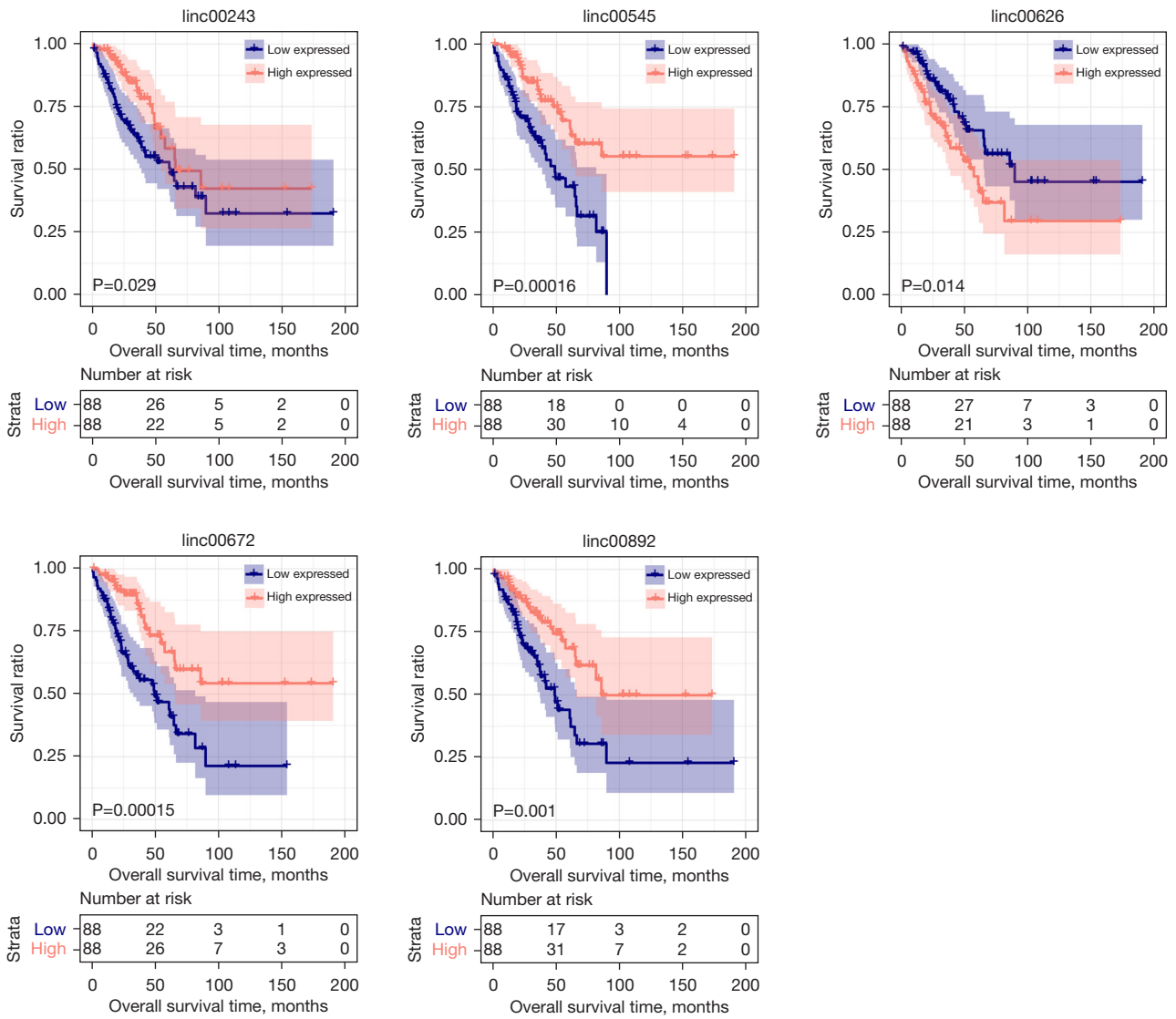
**Figure 4** The prognostic KM curves of lncRNA expression. The blue and red curves indicate the low and high expression sample groups, respectively. lncRNAs, long non-coding ribonucleic acids. KM, Kaplan-Meier.

to the expression levels of TCGA OS samples. Next, we calculated the correlation between the prognostic signal RS value of the characteristic lncRNAs and the subtypes of immune infiltration cells. The results indicated that the RSs were negatively correlated to the proportion of kinds of immune cells (see *Figure 8*).

## Discussion

OS is an aggressive malignant bone tumor with heterogeneous biology (40,41). A previous study indicated

that the heterogeneous biology of OS is related to the tumor microenvironment (15). OS tissue is not only made up of OS cells, but also contains a variety of other cells, including stromal cells, fibroblasts, and immune cells (4). The immune environment of osteosarcoma mainly consists of myeloid (macrophages, monocytes, dendritic cells) and lymphoid cells, as well as few B lymphocytes and mast cells. The recruitment and differentiation of immune infiltrating cells are controlled by OS cells, inducing a local immunosuppressive environment which contributes to the tumor growth and metastasis. Therefore, improving the
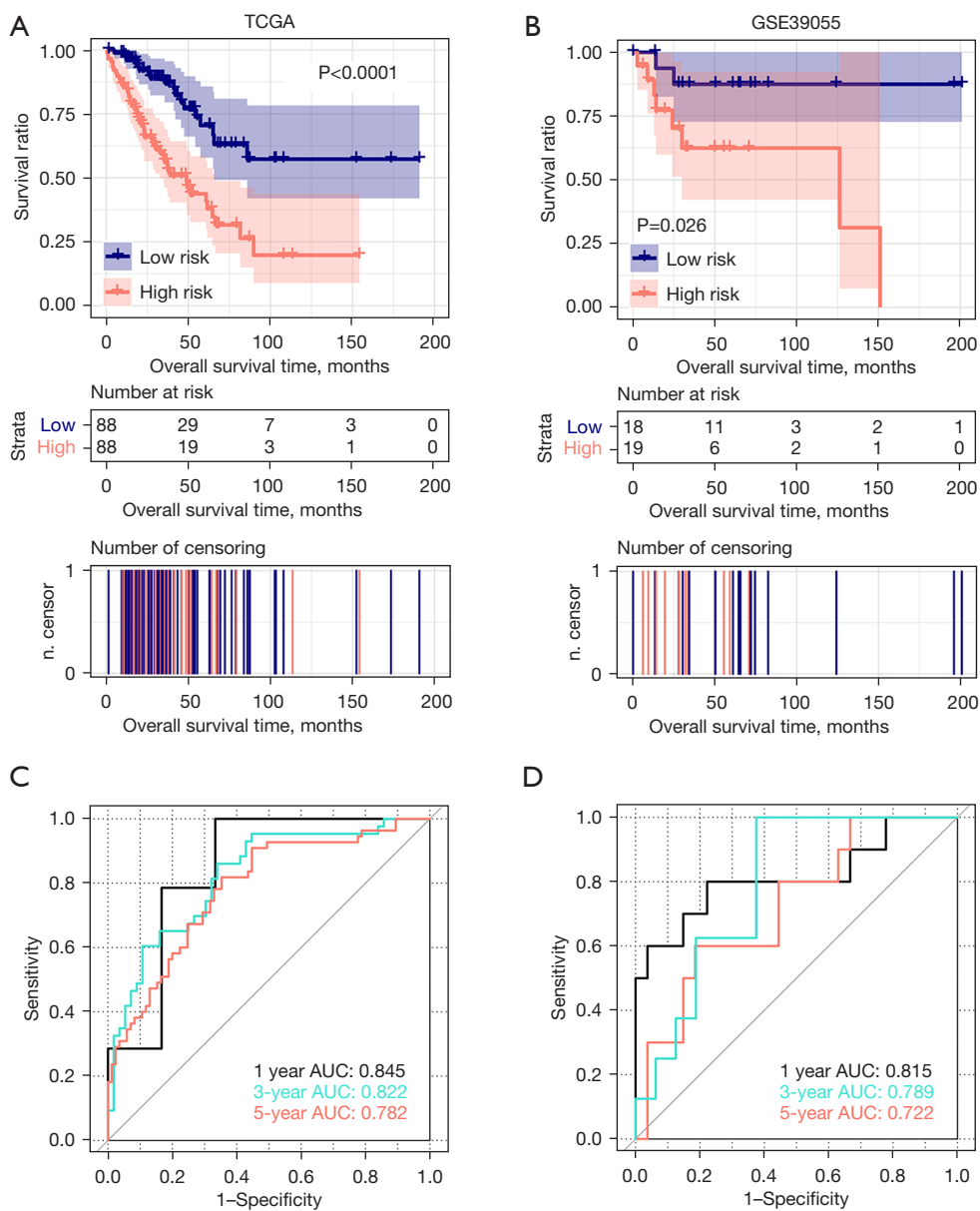
**Figure 5** Evaluation and comparison of the effectiveness of the prognostic risk prediction model. Prognosis-related KM curves of samples in TCGA (A) and GSE39055 (B). The blue and red curves represent the low- and high-risk samples, respectively. The 1-, 3-, and 5-year ROC curves of samples in TCGA (C) and GSE39055 (D). KM, Kaplan-Meier; TCGA, The Cancer Genome Atlas; AUC, Area Under Curve; ROC, receiver operating characteristic.

antitumor immune responses by means of reprogramming the immune microenvironment is a challenging opportunity to the osteosarcoma therapy (4). In our study, we explored the heterogeneity of OS and the relationship between tumor metastasis and immune infiltrating cells. We obtained the OS RNA-sequencing data and clinical information from

TCGA, and the results indicated that there were significant differences between the high immune cell infiltration group and the low immune cell infiltration group in terms of immune, stromal, and ESTIMATE scores.

Additionally, previous studies have reported that lncRNAs are involved in the occurrence, invasion, progression, and

3244

He et al. Screening of lncRNAs for metastatic OS

**Table 1** The information of clinical factors

| Clinical characteristics | TCGA (N=176) | Univariate Cox analysis | | | Multivariate Cox analysis | | |
|---|---|---|---|---|---|---|---|
| | | HR | 95% CI | P | HR | 95% CI | P |
| Age (years), mean ± SD | 61.10±15.21 | 1.018 | 1.001–1.036 | 3.99E-02 | 1.035 | 1.011–1.059 | 4.08E-03 |
| Gender (male/female) | 72/104 | 1.06 | 0.642–1.750 | 8.20E-01 | – | – | – |
| Pathologic tumor depth (cm), mean ± SD | 6.35±3.68 | 1.135 | 1.051–1.225 | 9.96E-04 | 0.997 | 0.867–1.145 | 9.63E-01 |
| Pathologic tumor length (cm), mean ± SD | 11.89±7.25 | 1.062 | 1.031–1.093 | 3.91E-05 | 1.089 | 0.965–1.228 | 1.66E-01 |
| Pathologic tumor width (cm), mean ± SD | 8.85±5.51 | 1.092 | 1.043–1.144 | 1.37E-04 | 0.996 | 0.853–1.163 | 9.58E-01 |
| Tumor recurrence (yes/no/–) | 28/141/7 | 2.603 | 1.533–4.422 | 2.38E-04 | 0.971 | 0.446–2.116 | 9.41E-01 |
| Metastatic tumor (yes/no) | 56/120 | 3.014 | 1.834–4.954 | 4.80E-06 | 2.992 | 1.516–5.904 | 1.58E-03 |
| Radiotherapy (yes/no/–) | 64/110/2 | 0.865 | 0.517–1.447 | 5.80E-01 | – | – | – |
| Tumor necrosis (no/slight/moderate/severe/–) | 61/35/59/11/10 | 1.182 | 0.923–1.513 | 1.83E-01 | – | – | – |
| PS model status (high/low) | 88/88 | 3.213 | 1.860–5.550 | 8.68E-06 | 4.107 | 1.916–8.802 | 2.81E-04 |

TCGA, The Cancer Genome Atlas; HR, hazard ratio; CI, confidence interval; SD, standard deviation.
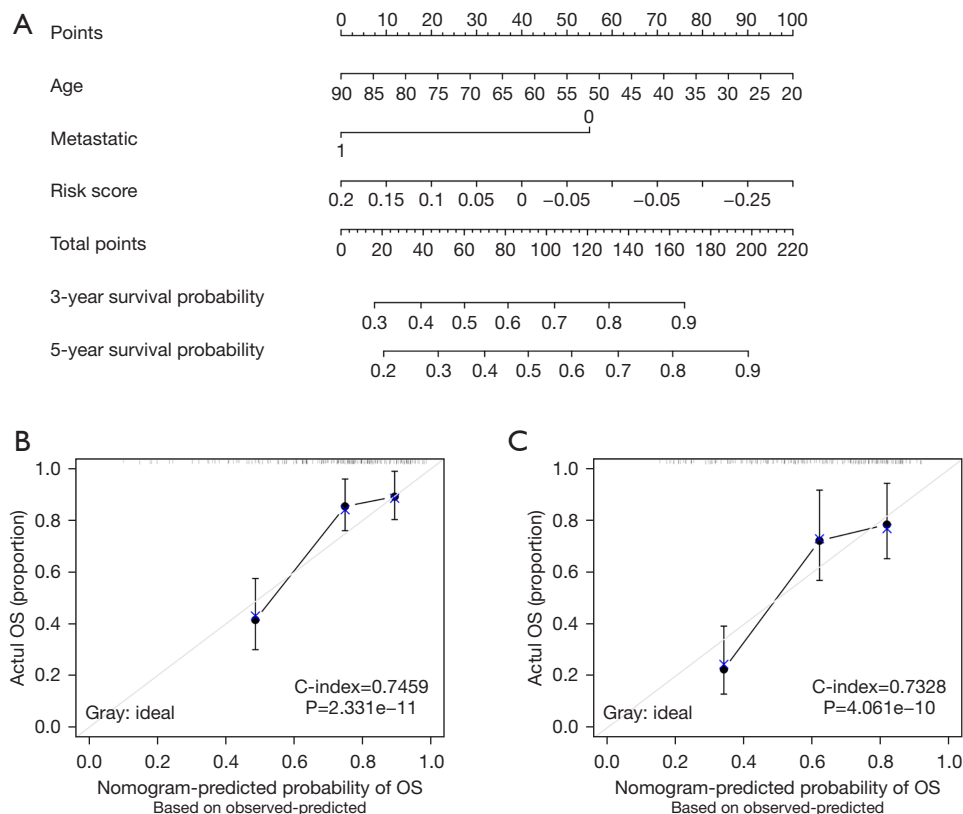


**Figure 6** Construction of nomogram survival rate model with independent prognostic factors. (A) Nomogram of survival rate prediction model. (B,C) Consistency line graph of nomogram-predicted and actual 3-year (B) and 5-year (C) survival rate; the horizontal axis represents the predicted survival rate, and the vertical axis represents the actual survival rate.

**Figure 7** The co-expression network of the feature lncRNAs and intersection mRNAs. The squares and circles represent the lncRNAs and intersection mRNAs, respectively. lncRNAs, long non-coding ribonucleic acids.

**Table 2** The significantly related BPs and KEGG pathways of genes contained in the co-expression network

| Category | Term | Count | P value | Gene |
|---|---|---|---|---|
| BP | GO:0032729~positive regulation of interferon-gamma production | 5 | 3.86E-04 | *IL18, IRF8, TNF* |
| | GO:0007165~signal transduction | 21 | 3.96E-04 | *ARHGAP9, IL15RA, ARRDC5,* |
| | GO:0006869~lipid transport | 5 | 2.56E-03 | *PTGES* |
| | GO:0032526~response to retinoic acid | 4 | 3.57E-03 | *CD38, AQP3, DKK* |
| | GO:0006955~immune response | 10 | 4.48E-03 | *C3, TMIGD2, CCL22* |
| | GO:0030514~negative regulation of BMP signaling pathway | 4 | 4.65E-03 | *HTRA3, CHRD, DKK1* |
| | GO:0002250~adaptive immune response | 6 | 5.17E-03 | *CD79A, CLEC10A, SH2D1B* |
| | GO:0006909~phagocytosis | 4 | 5.58E-03 | *CEBPE, CEACAM4, IRF8* |
| | GO:0006954~inflammatory response | 9 | 7.81E-03 | *C3, PSTPIP1, VNN1* |
| | GO:0050776~regulation of immune response | 6 | 1.10E-02 | *NCR1, C3, CD200R1* |
| | GO:0030593~neutrophil chemotaxis | 4 | 1.34E-02 | *CSF3R, CCL22, CCL17* |
| | GO:0045087~innate immune response | 9 | 1.58E-02 | |
| | GO:0019221~cytokine-mediated signaling pathway | 5 | 1.71E-02 | |
| | GO:0031295~T cell co-stimulation | 4 | 2.09E-02 | |
| | GO:0051897~positive regulation of protein kinase B signaling | 4 | 2.53E-02 | |
| | GO:0070374~positive regulation of ERK1 and ERK2 cascade | 5 | 4.29E-02 | |

**Table 2** (*continued*)

3246

He et al. Screening of lncRNAs for metastatic OS

**Table 2** (*continued*)

| Category | Term | Count | P value | Gene |
|---|---|---|---|---|
| KEGG pathway | hsa04060:Cytokine-cytokine receptor interaction | 11 | 1.20E-04 | |
| | hsa04662:B cell receptor signaling pathway | 5 | 4.66E-03 | |
| | hsa04640:Hematopoietic cell lineage | 5 | 1.05E-02 | |
| | hsa04630:Jak-STAT signaling pathway | 6 | 1.40E-02 | |
| | hsa04660:T cell receptor signaling pathway | 5 | 1.68E-02 | |
| | hsa04621:NOD-like receptor signaling pathway | 4 | 1.76E-02 | |
| | hsa04668:TNF signaling pathway | 5 | 2.11E-02 | |
| | hsa04650:Natural killer cell mediated cytotoxicity | 5 | 3.21E-02 | |
| | hsa04062:Chemokine signaling pathway | 6 | 3.62E-02 | |
| | hsa05133:Pertussis | 4 | 3.76E-02 | |
| | hsa04380:Osteoclast differentiation | 5 | 4.01E-02 | |

BP, biological process; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.
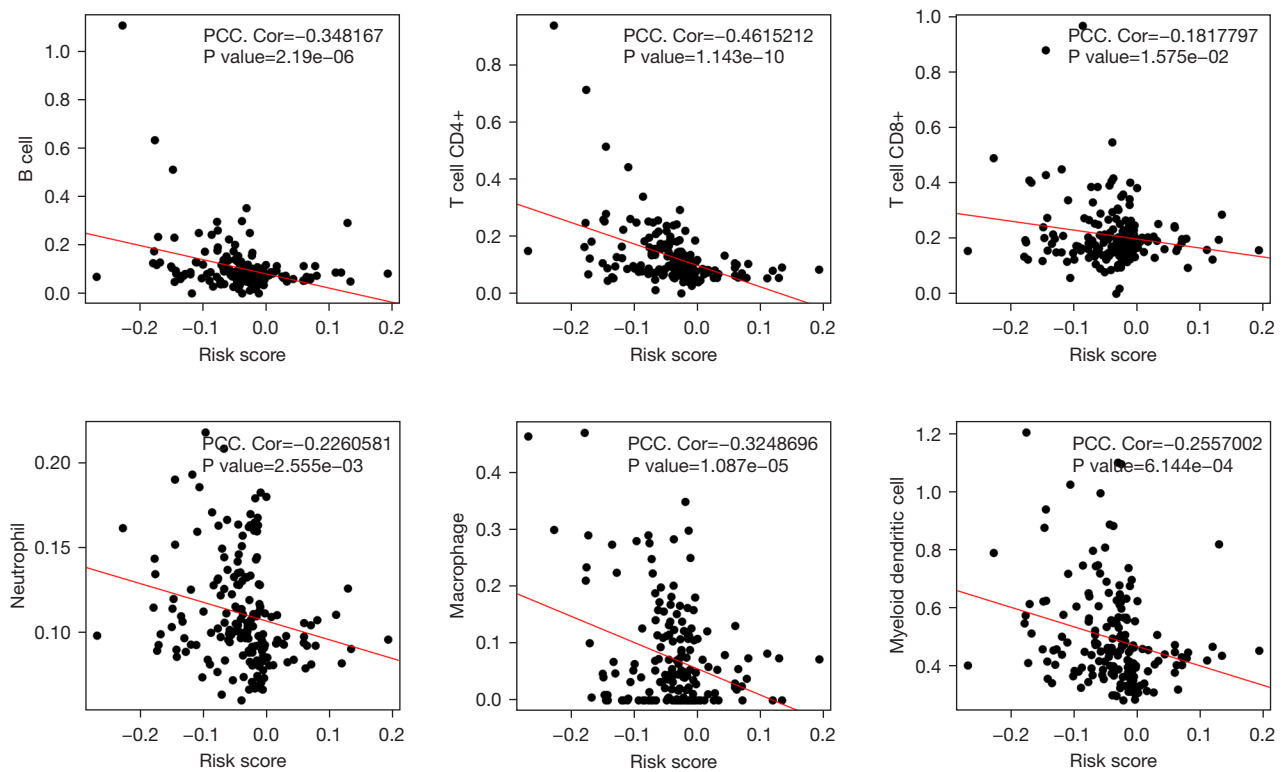


**Figure 8** Scatter plot of the correlations between RS and various types of immune cells in TIMER. RS, risk-score.

metastasis of OS. In our study, we identified 5 lncRNAs (i.e., linc00243, linc00892, linc00626, linc00545, and linc00672) that are related to the prognosis of OS patients. Similarly, Feng *et al.* (42) found linc00243 expression is significantly correlated to the prognosis and survival time of non-small cell lung cancer patients. Linc00892 is related to the microenvironment for bladder cancer patients, and is regarded as an important biomarker for bladder cancer metastasis and progression (43). Additionally, Linc00892 is related to some types of cancer, such as endometrial cancer (44,45), lung adenocarcinoma (46), and colorectal cancer (47), and is a novel biomarker for the prognosis and diagnosis of tumors.

We also constructed a RS model to verify the function of lncRNAs in the prognosis of OS patients, and found that the expression of lncRNAs is associated with OS patients' survival time. Further, univariate and multivariate Cox analyses were used to analyze the independent clinical factors, and then, a nomogram was constructed to predict the 3- and 5-year survival rates. We found that the survival rates predicted by the nomogram model were consistent with actual observations for OS patients at 3 and 5 years. All these results suggested that these 5 lncRNAs are key markers for treating OS in immunotherapy. However, further research on the functional verification of these key lncRNAs with clinical prognostic value and their involved multilayer genetic regulation networks is warranted to deeply understand the mechanism, which may contribute to its clinical application.

We also found that the 5 feature lncRNAs were negatively correlated to immune cells, including CD8$^+$ T cells, CD4$^+$ T cells, B cells, macrophages, neutrophils, and myeloid dendritic cells. The antigen-antibody complementarity determining regions of the T-cell and B-cell receptors have an important function in the recognition of tumor-specific antigens (48,49). Studying the sequence features of tumor infiltrative T-cell and B-cell surface receptors is useful in analyzing the interactions between immune cells and tumor cells, and provides a novel method for the diagnosis and treatment of cancer. After surgery, cancer tissues often include a number of immune infiltration cells, which cause the RNA sequencing data of the tumor tissue to be mixed with various information about the tumor immune microenvironment.

## Conclusions

Our research identified 5 feature lncRNAs as novel

biomarkers for OS. We also constructed a RS nomogram model based on the 5 feature lncRNAs. This model predicted the survival rates and prognoses of OS patients well. The 5 lncRNAs for OS were related to the infiltration of immune cell subtypes.

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-1926/rc

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-1926/coif). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

## References

1. Biazzo A, De Paolis M. Multidisciplinary approach to osteosarcoma. Acta Orthop Belg 2016;82:690-8.
2. Nørregaard KS, Jürgensen HJ, Gårdsvoll H, et al. Osteosarcoma and Metastasis Associated Bone Degradation-A Tale of Osteoclast and Malignant Cell Cooperativity. Int J Mol Sci 2021;22:6865.

3248

He et al. Screening of lncRNAs for metastatic OS

3. Durfee RA, Mohammed M, Luu HH. Review of Osteosarcoma and Current Management. Rheumatol Ther 2016;3:221-43.

4. Heymann MF, Lézot F, Heymann D. The contribution of immune infiltrates and the local microenvironment in the pathogenesis of osteosarcoma. Cell Immunol 2019;343:103711.

5. Li Y, Lin S, Xie X, et al. Highly enriched exosomal lncRNA OIP5-AS1 regulates osteosarcoma tumor angiogenesis and autophagy through miR-153 and ATG5. Am J Transl Res 2021;13:4211-23.

6. Han G, Guo Q, Ma N, et al. LncRNA BCRT1 facilitates osteosarcoma progression via regulating miR-1303/FGF7 axis. Aging (Albany NY) 2021;13:15501-10.

7. Li GB, Liu GY, Yang J, et al. Weighted gene correlation network analysis identifies the critical long non-coding RNAs participate in the progression of osteosarcoma. Gen Physiol Biophys 2021;40:173-82.

8. Sheng G, Gao Y, Yang Y, et al. Osteosarcoma and Metastasis. Front Oncol 2021;11:780264.

9. Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet 2018;391:2128-39.

10. Wang Y, Lin HC, Huang MY, et al. The Immunoscore system predicts prognosis after liver metastasectomy in colorectal cancer liver metastases. Cancer Immunol Immunother 2018;67:435-44.

11. Jang JE, Hajdu CH, Liot C, et al. Crosstalk between Regulatory T Cells and Tumor-Associated Dendritic Cells Negates Anti-tumor Immunity in Pancreatic Cancer. Cell Rep 2017;20:558-71.

12. Wang TT, Zhao YL, Peng LS, et al. Tumour-activated neutrophils in gastric cancer foster immune suppression and disease progression through GM-CSF-PD-L1 pathway. Gut 2017;66:1900-11.

13. Pulendran B, Davis MM. The science and medicine of human immunology. Science 2020;369:eaay4014.

14. Cai P, Otten ABC, Cheng B, et al. A genome-wide long noncoding RNA CRISPRi screen identifies PRANCR as a novel regulator of epidermal homeostasis. Genome Res 2020;30:22-34.

15. Zhang GZ, Wu ZL, Li CY, et al. Development of a Machine Learning-Based Autophagy-Related lncRNA Signature to Improve Prognosis Prediction in Osteosarcoma Patients. Front Mol Biosci 2021;8:615084.

16. Ghafouri-Fard S, Shirvani-Farsani Z, Hussen BM, et al. The critical roles of lncRNAs in the development of osteosarcoma. Biomed Pharmacother 2021;135:111217.

17. Tong CJ, Deng QC, Ou DJ, et al. LncRNA RUSC1-AS1 promotes osteosarcoma progression through regulating the miR-340-5p and PI3K/AKT pathway. Aging (Albany NY) 2021;13:20116-30.

18. Yu S, Shao F, Liu H, et al. A five metastasis-related long noncoding RNA risk signature for osteosarcoma survival prediction. BMC Med Genomics 2021;14:124.

19. Deng Y, Yuan W, Ren E, et al. A four-methylated LncRNA signature predicts survival of osteosarcoma patients based on machine learning. Genomics 2021;113:785-94.

20. Zhao A, Zhao Z, Liu W, et al. Carcinoma-associated fibroblasts promote the proliferation and metastasis of osteosarcoma by transferring exosomal LncRNA SNHG17. Am J Transl Res 2021;13:10094-111.

21. Wang Y, Ren X, Yuan Y, et al. Downregulated lncRNA GAS5 and Upregulated miR-21 Lead to Epithelial-Mesenchymal Transition and Lung Metastasis of Osteosarcomas. Front Cell Dev Biol 2021;9:707693.

22. Pagès F, Galon J, Dieu-Nosjean MC, et al. Immune infiltration in human tumors: a prognostic factor that should not be ignored. Oncogene 2010;29:1093-102.

23. Beltrán-Anaya FO, Romero-Córdoba S, Rebollar-Vega R, et al. Expression of long non-coding RNA ENSG00000226738 (LncKLHDC7B) is enriched in the immunomodulatory triple-negative breast cancer subtype and its alteration promotes cell migration, invasion, and resistance to cell death. Mol Oncol 2019;13:909-27.

24. Cusanovich DA, Billstrand C, Zhou X, et al. The combination of a genome-wide association study of lymphocyte count and analysis of gene expression data reveals novel asthma candidate genes. Hum Mol Genet 2012;21:2111-23.

25. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-10.

26. Li BL, Wan XP. Prognostic significance of immune landscape in tumour microenvironment of endometrial cancer. J Cell Mol Med 2020;24:7767-77.

27. Ye L, Zhang T, Kang Z, et al. Tumor-Infiltrating Immune Cells Act as a Marker for Prognosis in Colorectal Cancer. Front Immunol 2019;10:2368.

28. Hu D, Zhou M, Zhu X. Deciphering Immune-Associated Genes to Predict Survival in Clear Cell Renal Cell Cancer. Biomed Res Int 2019;2019:2506843.

29. Chen B, Khodadoust MS, Liu CL, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol 2018;1711:243-59.

30. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

31. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385-95.

32. Wang P, Wang Y, Hang B, et al. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget 2016;7:55343-51.

33. Zhao C, Lou Y, Wang Y, et al. A gene expression signature-based nomogram model in prediction of breast cancer bone metastases. Cancer Med 2019;8:200-8.

34. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.

35. Shan S, Chen W, Jia JD. Transcriptome Analysis Revealed a Highly Connected Gene Module Associated With Cirrhosis to Hepatocellular Carcinoma Development. Front Genet 2019;10:305.

36. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.

37. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.

38. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1-13.

39. Li T, Fan J, Wang B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res 2017;77:e108-10.

40. Wang X, Yu X, Long X, et al. MIR205 host gene (MIR205HG) drives osteosarcoma metastasis via regulating the microRNA 2114-3p (miR-2114-3p)/twist family bHLH transcription factor 2 (TWIST2) axis.

41. Angulo P, Kaushik G, Subramaniam D, et al. Natural compounds targeting major cell signaling pathways: a novel paradigm for osteosarcoma therapy. J Hematol Oncol 2017;10:10.

42. Feng X, Yang S. Long non-coding RNA LINC00243 promotes proliferation and glycolysis in non-small cell lung cancer cells by positively regulating PDK4 through sponging miR-507. Mol Cell Biochem 2020;463:127-36.

43. Lin G, Guo B, Wei Y, et al. Impact of Long Non-coding RNAs Associated With Microenvironment on Survival for Bladder Cancer Patients. Front Genet 2020;11:567200.

44. Li W, Li H, Zhang L, et al. Long non-coding RNA LINC00672 contributes to p53 protein-mediated gene suppression and promotes endometrial cancer chemosensitivity. J Biol Chem 2017;292:5801-13.

45. Dong P, Xiong Y, Yue J, et al. Exploring lncRNA-Mediated Regulatory Networks in Endometrial Cancer Cells and the Tumor Microenvironment: Advances and Challenges. Cancers (Basel) 2019;11:234.

46. Zhang X, Han J, Du L, et al. Unique metastasis-associated lncRNA signature optimizes prediction of tumor relapse in lung adenocarcinoma. Thorac Cancer 2020;11:728-37.

47. Mendelaar PAJ, Smid M, van Riet J, et al. Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features. Nat Commun 2021;12:574.

48. Sela-Culang I, Benhnia MR, Matho MH, et al. Using a combined computational-experimental approach to predict antibody-specific B cell epitopes. Structure 2014;22:646-57.

49. Wang Z, Wu H, Chen Y, et al. The Heterogeneity of Infiltrating Macrophages in Metastatic Osteosarcoma and Its Correlation with Immunotherapy. J Oncol 2021;2021:4836292.

Bioengineered 2021;12:1576-86.