Identification and Validation of a Novel Immune-related Five-lncRNA Signature for Osteosarcoma

Tianqun Huo

Changzhou Municipal Hospital of Traditional Chinese Medicine <https://orcid.org/0000-0002-9181-5330>

Jun Li

Changzhou Municipal Hospital of Traditional Chinese Medicine

Baofang Liu

Anhui Medical College: Anhui Medical University

Yu Zhai

Changzhou Municipal Hospital of Traditional Chinese Medicine

Xingbiao Wu

Changzhou Municipal Hospital of Traditional Chinese Medicine

Zhiguo Fu

Changzhou Municipal Hospital of Traditional Chinese Medicine

Hongming Tao

Changzhou Municipal Hospital of Traditional Chinese Medicine

Hong Zhao (✉ hongzhao2022@163.com)

Changzhou Municipal Hospital of Traditional Chinese Medicine <https://orcid.org/0000-0003-2997-0185>

Research Article

Keywords: Osteosarcoma, Prognosis, Immune infiltration, Immune-related lncRNA

Posted Date: February 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1291091/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Osteosarcoma (OS) is the most common malignant tumor in primary bone cancer. It often occurs in children or adolescents. It has the characteristics of strong invasiveness, early blood metastasis, and poor prognosis. Recent studies have shown that the tumor microenvironment (TME) and immune-related lncRNAs (IR-lncRNAs) play an important role in tumor prognosis. Therefore, exploring prognostic biomarkers that reflect the bioheterogeneity of OS can provide better interventions for patients.

Methods: The gene expression data and corresponding clinical data of OS patients were downloaded from the GEO (n=102) and TARGET (n=101) databases. We applied single-sample gene set enrichment analysis (ssGSEA) to the transcriptomes of these samples, generating the three immune subgroups. The ESTIMATE algorithm, CIBERSORT, and Kaplan–Meier analysis were used to explore the TME and tumor-infiltrating immune cells (TIICs) among the 3 immune subgroups. The immune gene set in the ImmPort database was used to screen out immune-related mRNAs (IR-mRNAs) in the two databases, and IR-lncRNAs were selected through coexpression with IR-mRNAs. Based on these groupings, we identified differentially expressed immune-related lncRNAs (IRD-lncRNAs) between the high- and low-immune subgroups. Then, an IRD-lncRNA prognostic signature was established by least absolute shrinkage and selection operator (LASSO) regression.

Results: We found that as the immune status of the immune subgroups increased, the prognosis significantly improved ($P<0.05$). Moreover, the stromal, immune, and ESTIMATE scores increased, while tumor purity decreased ($P<0.01$). The TIIC results showed that CD8+ T cells had a positive correlation with the prognosis of OS ($P<0.01$). After using the ImmPort database and coexpression analysis, we screened 3767 IR-lncRNAs and identified 117 IRD-lncRNAs between the immune-H and immune-L subgroups. Finally, we identified five IRD-lncRNAs as a prognostic signature for OS. Kaplan–Meier, ROC curve and univariate and multivariate Cox regression analyses further confirmed that the prognostic signature was an innovative and significant independent prognostic factor.

Conclusion: The study results may help clarify the impact of the TME on OS and provide a method to predict the prognosis of OS patients using the five-IR-lncRNA signature as a potential biomarker.

Background

Osteosarcoma (OS) is the most common tumor in primary bone cancer and frequently occurs in children and adolescents, with an annual incidence of 5/100 million people[1]. It has a very aggressive course characterized by early hematogenous metastasis and a poor prognosis[2]. Currently, comprehensive treatment methods for OS include surgery, chemotherapy, radiotherapy, and targeted therapy. Among them, surgical resection combined with chemotherapy has increased the long-term survival rate of patients with local OS to 60-70%. [3, 4]. However, the five-year survival rate is still less than 20% in malignant OS[5]. Therefore, it is essential to identify new biomarkers to improve the prognosis, thereby ensuring the effective treatment of OS patients.

The tumor microenvironment (TME) is a kind of cellular environment that contains all factors that are needed for tumors, including noncellular components (such as microvessels and extracellular matrix) and cellular components (such as tumor cells, immune cells, and stromal cells)[6, 7]. In the occurrence and development of tumors, tumor cells induce various types of stress (including hypoxia, oxidative stress, acidosis, etc.) to pathologically change the TME [8]. These pathological effects accelerate the necrosis and metastasis of neighboring immune cells and stromal cells[9]. Studies have shown that the type, location, and density of tumor-infiltrating immune cells (TIICs) present in the TME can predict the overall survival time of patients and their response to treatment[10, 11]. Recent research shows that TIICs are highly related to tumor treatment response and clinical efficacy, may have antitumor or tumor-promoting functions and are significantly related to prognosis[12, 13]. Considering the importance of TIICs in OS prognosis, immunotherapy is a promising treatment option for malignant OS[14]. Many studies have shown that immunotherapy has achieved good effects on some cancers, such as breast cancer and hepatocellular carcinoma[15, 16]. Due to the rarity and heterogeneity of OS, it is urgent to better understand the TME of OS to improve patient survival outcomes [17]. In addition, the different types of TIICs have different effects on tumor development and progression[18]. Exploring the TME can help elucidate the role of immune cells and establish a good signature for predicting tumor prognosis.

In mammalian genomes, approximately 80% of the genomes are transcribed in a cell-specific manner, especially noncoding regions. Only a small part of the genome is transcribed into protein-encoding mRNA, and most of them produce large amounts of long noncoding RNAs (lncRNAs)[19]. lncRNAs are a type of transcript with a length of more than 200 nucleotides. Their genes are ordinarily located in the intergenic sequence of protein-coding genes or overlapping antisense transcripts[20]. Recent studies show that lncRNAs regulate the development and activation of a diversity of immune cells by controlling the dynamic interpretation of transcription programs and are involved in carcinogenesis[21, 22]. In addition, many studies have identified lncRNAs, such as the lncRNA LOC100129620 and the lncRNA LINC00426, as oncogenes or tumor suppressors in OS development and progression[23, 24]. Although many studies have explored lncRNAs related to OS, they mainly focused on the prognosis and predictive ability of a single molecule, and the occurrence and development of tumors are often the result of the synergistic effect of multiple molecules. In addition, with the importance of immune-related lncRNAs (IR-lncRNAs), they have become biomarkers of prognostic risk in many cancers[25-27]. Although IR-lncRNAs are increasingly being recognized as important regulators in the occurrence and development of malignant tumors, their role in OS has not been investigated thoroughly.

Therefore, the present study analyzed the gene expression profiles and clinical data of OS patients downloaded from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and Gene Expression Omnibus (GEO) databases. Patients with OS were divided into high-, medium-, and low-immune subgroups using single-sample gene set enrichment analysis (ssGSEA), and the differences in TIICs were verified by ESTIMATE, CIBERSORT, and Kaplan-Meier (K-M) analysis. Then, coexpression analysis was used to screen IR-lncRNAs through immune-related mRNAs (IR-mRNAs), which were identified from the two databases. Differentially expressed immune-related lncRNAs (IRD-lncRNAs) were identified between the immune-H and immune-L subgroups. Next, univariate Cox, multivariate Cox, and

least absolute shrinkage and selection operator (LASSO) regression analyses were used and identified 5 IRD-lncRNAs related to prognosis. Finally, we evaluated the accuracy of the 5-IRD-lncRNA prognostic signature. Overall, the results of the present study not only explain the underlying mechanism of OS but also enhance the ability to predict the prognosis of patients.

Materials And Methods

Data acquisition and processing

The TARGET database (<https://ocg.cancer.gov/programs/target>) uses a comprehensive genomic approach to identify the molecular changes that lead to the occurrence and development of childhood cancer. It aims to promote the discovery of treatment targets in childhood cancer and to promote the rapid translation of these discoveries into clinical applications. The GEO database is currently the world's largest and most authoritative public gene expression database, created and maintained by the NCBI American Biotechnology Information Center. It mainly includes uploaded microarray data and other forms of high-throughput gene expression data from research institutions around the world and provides detailed information and annotations of complete data. A total of 101 samples with OS-related gene expression data and related clinical information were downloaded from the TARGET database (95 samples with clinical survival data). The gene expression levels of GSE14827, GSE21257, and GSE32981 were obtained from the GEO database as TARGET data verification. Using the "SVA" R package data standardization and batch-to-batch difference removal, a total of 102 tumor samples were integrated with the GEO gene expression profiles. The clinical information of patients with OS was obtained from TARGET.

Clustering and validation of immune subgroups and survival analysis

ssGSEA was applied to 29 immune-associated gene sets, which included immune cell types, immune-related functions, and immune-related pathways, to explore the different infiltration degrees in the OS expression profiles of the TARGET and GEO databases using the R package "GSVA". Using a hierarchical clustering algorithm, the OS samples were divided into three clusters according to immune-associated gene sets (cutoff = 1.0). According to the ssGSEA results, the OS samples were divided into a high-immune subgroup (immunity-H), a medium-immune subgroup (immunity-M), and a low-immune subgroup (immunity-L) using the "sparcl" package. To validate the immune subgroups, the ESTIMATE algorithm was used to compare the immune score, stromal score, and tumor purity between the subgroups. In addition, to validate the relationship between the immune subgroups and the prognosis of OS, target data set survival analysis was performed using the "survival" package.

Estimation of TIICs using CIBERSORT

CIBERSORT is an algorithm used to characterize the cell composition of complex tissues from their gene expression profiles. To explore the relationship between the immune subgroups and immune cell infiltration, CIBERSORT analysis was performed on the two databases of gene expression profiles to derive the relative proportions of 22 infiltrating immune cells. The TARGET data set was used as the

training cohort, and the GEO data set was used as the verification cohort. Samples with a CIBERSORT output of $P < 0.05$ were further analyzed. The Wilcoxon rank-sum test was used to identify the differences in the content of different immune cells in the 3 immune subgroups. At the same time, the COR function was used to compare the correlation between immune cells. In addition, K–M analysis was used to evaluate the relationship between immune cells and OS patients' overall survival.

Identification of IRD-lncRNAs between the immune-H and immune-L subgroups

To obtain the expression levels of IR-mRNAs in the GEO and TARGET databases, we used the IR-mRNAs list, which was downloaded from the ImmPort database (<https://import.niaid.nih.gov/home>). The “VennDiagram” package was used to create a Venn diagram, which was used to identify the overlapping IR-mRNAs in the two databases. To obtain the key IR-mRNAs, we used the STRING database (<https://cn.string-db.org>) to construct a protein–protein interaction (PPI) network diagram and further used the MCODE plug-in in Cytoscope to screen the closest IR-mRNA cluster.

We next extracted the lncRNA expression matrix from the TARGET database and calculated the correlation between lncRNAs and key IR-mRNAs using Pearson coefficients. lncRNAs with Pearson coefficients $> \pm 0.4$ and $P < 0.001$ were considered IR-lncRNAs. lncRNA expression in the TARGET database is in transcripts per million (TPM) values, which were standardized by log2 conversion and offset by 1 to avoid errors in lncRNA expression analysis. The “limma” R package (version 4.1.0) from Bioconductor 3.8 was used for quantile normalization of the data to identify IRD-lncRNAs between the Immunity-H and Immunity-L subgroups, as previously described[28, 29]. The IRD-lncRNAs were evaluated using the Wilcox test, and the P values were corrected for the false discovery rate (FDR) by the Benjamini-Hochberg (BH) procedure. Only IR-lncRNAs with a $|\log_2 \text{fold change (FC)}| > 1$ and an FDR < 0.05 were selected as IRD-lncRNAs.

Screening and confirmation of an IRD-lncRNA prognostic signature for OS

According to the IRD-lncRNA expression and survival data, a prognostic signature for OS was identified through univariate Cox regression, multivariate Cox regression and LASSO Cox regression analyses. $P < 0.05$ was considered statistically significant. A formula was constructed for the prognostic signature, and we calculated the prognostic risk score of each patient. The formula is as follows:

$$\text{Risk score} = \sum \text{IRD-lncRNA}_{\text{Cox coefficient}} \times \text{IRD-lncRNA}_{\text{expression level}}$$

Based on the median risk value, the patients were divided into high-risk and low-risk groups. The K-M survival curves and forest maps were constructed by the software package “survminer” (version 0.4.9). The K–M log-rank test was used to evaluate the predictive ability of the signature in the total cohort, the training cohort, and the verification cohort; the total cohort was split at a ratio of 1:1 to obtain the training and verification cohorts using the “sampling” package. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) were used to evaluate the stability of the prognostic signature.

Pearson correlation analysis was used to explore the correlation between the 5 IRD-lncRNA signature and TIICs.

Statistical analysis

All statistical analyses were performed using R software version 4.1.0. K–M analysis and log-rank tests were used to assess survival and compare survival differences between immune subgroups and risk groups. K-M curves were used to evaluate the relationship between TIICs and overall survival with the R package “survival”. The Wilcoxon signed-rank test was used to compare the different genes between the immune subgroups. $P < 0.05$ was considered statistically significant.

Result

Construction and verification of OS immune subgroups

We obtained 101 OS genome-wide data points and 95 corresponding clinical sample data points from the TARGET database. The clinical characteristics are shown in Table 1. In addition, 3 OS data sets were screened from the GEO database, and a total of 102 samples were integrated after normalization with the R “SVA” package and the removal of interbatch differences. Immune cell infiltration was evaluated in the TARGET-OS and GEO-OS databases using the ssGSEA method. The OS samples were divided into 3 subgroups based on the 29 immune-related gene set scores by applying an unsupervised hierarchical clustering algorithm (cutoff = 1.0), including the immunity-H (n= 14), immunity-M (n=65), and immunity-L (n=22) subgroups in TARGET-OS and the immunity-H (n= 13), immunity-M (n=68), and immunity-L (n=21) subgroups in GEO-OS. Next, we validated the clustering results in TARGET-OS and GEO-OS using the ESTIMATE algorithm. The results were the same in the two databases, which showed that as the immune subgroups increased, the stromal score, immune score, and ESTIMATE score increased, while tumor purity decreased. The visualization results are shown in the heatmap and violin map (Fig. 2a-d). The results showed that among the three subgroups, the survival rate of the immune-H subgroup was significantly higher than that of the immune-L group ($P < 0.05$) (Fig. 2e). Thus, the research results show that the prognosis of OS is closely related to immunity and that increasing the expression of immune-related gene set scores can significantly improve the prognosis of OS. All these results indicate that the OS cluster analysis method can be used for further research and analysis.

Table 1
Clinical characteristics of OS patients in the TARGET database.

Parameters	Number	Ratio (%)
Age(y)		
<16	56	58.9
>16	39	41.1
Gender		
Male	55	57.9
Female	40	42.1
Metastasis		
Positive	23	24.2
Negative	72	75.8
Survival status		
Alive	57	60.0
Death	38	40.0

Differences in the composition of TIICs in the immune subgroups

To understand the infiltration of immune cells in the immunity subgroups, we calculated the percentages of the 22 immune cells in each OS sample through the CIBERSORT database. Among the 95 OS samples in the training cohort and 102 OS samples in the verification cohort, with the increasing order of immune subgroups, the infiltration abundance of CD8+ T cells and M1 macrophages also increased. In addition, naive CD4 T cells and M0 macrophages continuously decreased (Fig. 3a, b) ($P < 0.001$). We further analyzed the correlation of the infiltration abundances of 22 immune cells in the training cohort and verification cohort. Among them, CD8+ T cells had significant negative correlations with CD4-naive T cells and M0 macrophages. At the same time, CD8+ T cells had a significant positive correlation with M1 macrophages, which was the same result in the verification cohort (Fig. 3c, d). Furthermore, we used K-M analysis to explore the relationship between immune cell infiltration and the prognosis of OS patients in the training cohort. The results showed that patients with high infiltration of activated mast cells, resting NK cells, and plasma cells had a significantly poorer survival rate, while those with high infiltration of CD8+ T cells, follicular helper T cells, and activated NK cells had a significantly better survival rate (Fig. 4a-f). The above results indicate that these immune cells are significantly related to the prognosis of OS and can be used as potential targets for the immunotherapy of OS.

Screening IRD-lncRNAs through coexpression analysis and differential gene analysis between immune subgroups

A total of 2,483 IR-mRNAs were downloaded from the ImmPort database. The IR-mRNAs were extracted from the two OS databases and intersected through a Venn diagram (Fig. 5a). A total of 654 IR-mRNA profiles were identified after data normalization and the removal of batch differences. Furthermore, using the STRING database, we constructed a network diagram of the PPIs of IR-mRNAs. A total of 131 key IR-mRNAs in the closest gene cluster were screened through the MCODE plug-in in Cytoscope (Fig. 5b-d). A total of 3,767 IR-lncRNAs were screened from the TARGET-OS database through the 131 key IR-mRNAs using the coexpression analysis method. We next screened the immunologically significant lncRNAs, and a total of 117 IRD-lncRNAs were screened between the Immunity-H and Immunity-L subgroups and visualized by volcano maps and heatmaps (Fig. 6a, b).

Construction and assessment of the 5 IRD-lncRNA prognostic signature for OS

In this study, using 95 OS samples with complete survival analysis data, univariate Cox regression analysis was performed on 117 IRD-lncRNAs, and 17 IRD-lncRNAs related to prognosis were identified. Furthermore, five IRD-lncRNAs were identified as a prognostic signature with multivariate cox regression analysis, including RP1-120G22.11, RP11-445L13_B.3, RP11-525G13.2, AC092484.1, and RP11-733018.1, which were significantly correlated with the survival and prognosis of OS (Fig. 7a-d). Each sample was scored for risk according to the aforementioned formula. The samples were divided into low-risk and high-risk groups based on the median value of the risk score. To further verify the reliability of the prognostic signature, comparing the risk scores and the survival status in the training cohort, the verification cohort, and total cohort, it was found that with the continuous improvement of the risk score, the greater the positive correlation with death status (Fig. 8a, b). The survival analysis had the same results in the training cohort, the verification cohort, and the total cohort, showing that the survival rate of patients in the high-risk group was lower than that of patients in the low-risk group (Fig. 8c). In addition, the time-dependent ROC curve analysis in the training cohort showed that the AUCs for overall survival in the first year, third year, and fifth year were 0.795, 0.769, and 0.761, respectively; those in the verification cohort were 0.724, 0.712, and 0.727; and those in the total cohort were 0.610, 0.754, and 0.753 (Fig. 8d). Clinical independent prognostic analysis showed that compared with other clinicopathologic characteristics, the risk value can be used as an independent prognostic factor ($P < 0.01$), and it is positively correlated with prognosis (Fig. 9a, b). Among them, metastasis and prognosis were also positively correlated, and the comparison of metastasis and the risk score showed that patients with metastasis had higher risk scores, which was statistically significant. All the results indicate that the prognostic signature based on the IRD-lncRNAs is an effective tool to judge the prognosis of OS patients and has important clinical value.

Correlation analysis between the prognostic signature and infiltrating immune cells

To further understand the relationship between the prognostic signature and immune cell infiltration, we analyzed the relationship between immune cell infiltration and the five IRD-lncRNAs in the prognostic

signature through Spearman correlation analysis. The results showed that the five IRD-lncRNAs are closely related to immune cells (Fig. 10).

Discussion

OS is derived from mesenchymal stem cells and is the most common malignant bone tumor. It is highly aggressive and metastatic and has a very poor prognosis[30]. Therefore, there is an urgent need to identify effective biomarkers for OS-specific prognosis to improve the survival rate of patients. Recent studies have shown that the heterogeneity of OS not only occurs in genotype and phenotype but also has macroscopic and microscopic aspects of the microenvironment of heterogeneous tumors[3, 17]. In OS tissue, in addition to tumor cells, osteoblasts, fibroblasts, immune cells and stromal cells are also mixed and interact to form a complex microenvironment[31, 32]. Considering that James P. Allison and Tasuku Honjo found that cancer can be cured by inhibiting negative immunomodulation, the research and application of TIICs in tumors has reached a new height. In addition, studies have shown that the abnormal expression of lncRNAs is related to the occurrence, progression, and metastasis of OS[23]. Moreover, lncRNAs are related to many aspects of tumor immunity, including the recognition and killing of cancer cells, cell migration, and immune cell infiltration[33]. Therefore, this study first explored the immune infiltration and heterogeneity of OS as well as the interaction between TIICs and tumor cells. Based on the immune subgroups, we further explored and analyzed the IR-lncRNA prognostic signature. This is of great significance for studying the mechanism of OS occurrence and development, as well as for developing new prognostic diagnostic and treatment methods.

In this study, we obtained mRNA expression profiles from the TARGET-OS and GEO-OS databases and obtained lncRNA expression profiles from TARGET-OS. ssGSEA clustered the OS samples into 3 immune subgroups through immune infiltration degrees, and the subgroups were verified by ESTIMATE and CIBERSORT to explore the relationship between the immune subgroups and TIICs in the two databases. The results showed that the tumor purity, ESTIMATE score, immune score, and stromal score in the 3 immune subgroups were significantly different, and the results in the two databases showed the same trend. The results of survival analysis showed that with the improvement of immune subgroups, the prognosis of patients also continued to improve. The CIBERSORT algorithm and verification of the correlation between immune cells showed that with the improvement of the immune subgroups, the infiltration abundance of CD8+ T cells and M1 macrophages continued to increase, and M1 macrophages had a positive correlation with CD8+ T cells. Naive CD4 T cells and M0 macrophages continued to decrease with the improvement of the immune subgroups, and there was a negative correlation with CD8+ T cells. In addition, the survival analysis showed that patients with high CD8+ T cell infiltration of OS has a good prognosis compared with those with low CD8+ T cell infiltration. Many studies have shown that CD8+ T cells mainly generate energy through oxidative phosphorylation and then use aerobic glycolysis when activated. The amount of CD8+ T cell infiltration in tumors is significantly related to prognosis[34-36]. In HPV-driven cervical cancer and oropharyngeal squamous cell carcinoma research reports, the number of infiltrating CD8+ T cells and the appearance of HPV E6/E7 epitopes are significantly correlated with improved prognosis[37]. In a prognostic study of

nasopharyngeal carcinoma, it was shown that a high density of CD8+ T cells in the tumor and/or stroma was significantly associated with a decrease in mortality [38]. In addition, a growing number of studies have also shown that naive T cells, M0 macrophages, and M1 macrophages are closely related to the development and prognosis of tumors[39-43].

Furthermore, we used the ImmPort database, Venn diagram, and PPI network analysis to identify the overlapping key IR-mRNAs in TARGET-OS and GEO-OS. Next, we used coexpression analysis to identify IR-lncRNAs in TARGET-OS. Then, we used the “limma” package to select IRD-lncRNAs between the high-immunity group and the low-immunity group. Univariate Cox regression and stepwise multivariate Cox regression were used to build a five IRD-lncRNA prognostic signature. To increase the credibility of our results, the risk score was calculated using the coefficients of the five IRD-lncRNAs. The patients were classified into low-risk and high-risk groups based on the median risk score. The K–M overall survival curve of the low-risk group was significantly longer than that of the high-risk group in the training cohort, the verification cohort, and the total cohort. The ROC curve, score plot, and survival status also showed that the prognostic model of the IRD-lncRNAs had favorable predictive ability in the training, verification, and total cohorts.

In recent years, most studies have focused on the prognosis and predictive ability of a single molecule when exploring lncRNAs related to cancer, but the occurrence and development of tumors are often the result of multiple molecules synergistically. We identified a 5-IRD-lncRNA signature (including RP11-445L13_B.3, RP1-120G22.11, RP11-733O18.1, AC092484.1, and RP11-525G13.2) that performed well in predicting the prognosis of OS patients. Although the IRD-lncRNAs in the prognostic signature have not been functionally annotated or fully clarified, there have been many studies on the tumor prognosis of the IRD-lncRNAs regulated mRNAs in this signature. The most relevant to RP11-733O18.1 is TNFRSF1B (TNFR2). Studies have reported that TNFRSF1B can not only promote tumor cell growth but also play an important role in tumor development by suppressing the immune response. Many studies have shown that antagonistic anti-TNFRSF1B antibodies can inhibit Treg proliferation and directly kill tumor cells with high levels of TNFRSF1B-expressing tumor cells, indicating that targeting TNFRSF1B may be an effective antitumor therapy[44, 45]. Regarding amphiregulin (AREG), which is most related to RP11-445L13_B.3, a study showed that AREG is packed into multiple myeloma-derived exosomes and implicated in osteoclast differentiation through osteoblast mechanism mediation, which affects the prognosis of multiple myeloma patients[46]. Regarding MMP9, which has the greatest correlation with AC092484.1, tumor-associated macrophages and tumor-associated neutrophils, which play important roles in the TME, are mainly derived from MMP9, which is promoted by its extracellular matrix degradation properties angiogenesis[47] Ptpn11 has the greatest correlation with RP11-525G13.2. Studies have reported that Ptpn11 gene deletion greatly inhibits tumors in pancreatic ductal adenocarcinoma (PDAC) and non-small cell lung cancer (NSCLC) in mutant KRAS-driven mouse models[48]. PSMB8 has the greatest correlation with RP1-120G22.11. Studies have reported that PSMB8 is closely related to the migration, proliferation, and apoptosis of glioma cells[49]. Other studies have reported that inducible knockdown of PSMB8 reduces the expression of vascular endothelial growth factor (VEGF), VEGF receptor and CD31 and the progression of human glioblastoma[50]. The above

evidence shows that the 5 IRD-lncRNAs in this signature play important roles in the occurrence and development of tumors, and their introduction may identify new molecular targets for OS immunotherapy.

Inevitably, this study has certain limitations that need to be addressed. First, our research is a retrospective study based on an open online database. Second, in the current cohort, although IR-mRNAs were obtained from two databases, IR-lncRNAs were obtained from a single database, and there were some other data sets, but their sample sizes were too small. To improve the research quality, we excluded other databases. Therefore, further experimental evidence is needed to verify the molecular functions of OS biomarkers, which will be the focus of our future research.

Conclusion

In this study, we divided the OS patients in the two major databases into three immune subgroups based on ssGSEA and applied unsupervised consensus cluster analysis. The three subgroups differed in survival outcomes, tumor microenvironment, and immune cell infiltration. It has also been determined that CD8⁺ T cells have a significant correlation with improving prognosis. Furthermore, we performed coexpression analysis to screen out IR-lncRNAs, as well as IR-lncRNAs with differences between the high- and low-immune groups, and successfully established a new 5-IR-lncRNA signature to predict the prognosis of OS patients. These results may provide new ideas for future studies and related molecular mechanisms of targeted therapy for OS.

Abbreviations

OS: Osteosarcoma; TME: Tumor microenvironment; TIICs: Tumor-infiltrating immune cells; lncRNAs: Long noncoding RNAs; IR-lncRNAs: Immune-related lncRNAs; IR-mRNAs: Immune-related mRNAs; IRD-lncRNAs: Differentially expressed immune-related lncRNAs; TARGET: Therapeutically Applicable Research to Generate Effective Treatments; GEO: Gene Expression Omnibus; ssGSEA: Single-sample gene set enrichment analysis; LASSO: least absolute shrinkage and selection operator; ROC: The receiver operating characteristic; AUC: The area under the curve; AREG: Regarding amphiregulin; PDAC: Pancreatic ductal adenocarcinoma; NSCLC: Non-small cell lung cancer; VEGF: Vascular endothelial growth factor;

Declarations

Acknowledgements

Not applicable

Authors' contribution

HZ designed the present study. TH, JL and BL performed the experiments. YZ, XW and ZF analyzed the data and prepared the figures. TH and JL drafted the initial manuscript. TH and BL reviewed and revised the manuscript. All authors read and approved the manuscript and agree to be accountable for all

aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

The present study was supported by the Youth Program of National Natural Science Foundation of China (81801213), China Postdoctoral Science Foundation, No.65 General Fund(2019M651967), Jiangsu Planned Projects for Postdoctoral Research Funds(2018K176C) and China Postdoctoral start-up fund (2018107007)

Availability of data and materials

The datasets that support the findings of this study are available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) and TARGET database(<https://ocg.cancer.gov/programs/target>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest regarding this work.

References

1. Chen, T. and L. Zhao, *Patrolling monocytes inhibit osteosarcoma metastasis to the lung*. Aging (Albany NY), 2020. **12**(22): p. 23004.
2. Yao, F., et al., *PODN is a prognostic biomarker and correlated with immune infiltrates in osteosarcoma*. Cancer Cell International, 2021. **21**(1): p. 1-16.
3. Fernandes, I., et al., *Osteosarcoma Pathogenesis Leads the Way to New Target Treatments*. International Journal of Molecular Sciences, 2021. **22**(2): p. 813.
4. Zhou, Y., et al., *Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma*. Nature communications, 2020. **11**(1): p. 1-17.
5. Luetke, A., et al., *Osteosarcoma treatment—where do we stand? A state of the art review*. Cancer treatment reviews, 2014. **40**(4): p. 523-532.
6. Reina-Campos, M., J. Moscat, and M. Diaz-Meco, *Metabolism shapes the tumor microenvironment*. Current opinion in cell biology, 2017. **48**: p. 47-53.

7. Sharma, V.P., et al., *Live tumor imaging shows macrophage induction and TMEM-mediated enrichment of cancer stem cells during metastatic dissemination*. Nature communications, 2021. **12**(1): p. 1-24.
8. Yang, B., et al., *Identification of prognostic biomarkers associated with metastasis and immune infiltration in osteosarcoma*. Oncology Letters, 2021. **21**(3): p. 1-1.
9. Lambrechts, D., et al., *Phenotype molding of stromal cells in the lung tumor microenvironment*. Nature medicine, 2018. **24**(8): p. 1277-1289.
10. Lei, X., et al., *Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy*. Cancer letters, 2020. **470**: p. 126-133.
11. O'Donnell, J.S., M.W. Teng, and M.J. Smyth, *Cancer immunoediting and resistance to T cell-based immunotherapy*. Nature reviews Clinical oncology, 2019. **16**(3): p. 151-167.
12. Wu, T. and Y. Dai, *Tumor microenvironment and therapeutic response*. Cancer letters, 2017. **387**: p. 61-68.
13. Malka, D., et al., *Immune scores in colorectal cancer: Where are we?* European Journal of Cancer, 2020. **140**: p. 105-118.
14. Huang, Q., et al., *Immune-Related LncRNAs Affect the Prognosis of Osteosarcoma, Which Are Related to the Tumor Immune Microenvironment*. Frontiers in cell and developmental biology, 2021: p. 2847.
15. Emens, L.A., *Breast cancer immunotherapy: facts and hopes*. Clinical Cancer Research, 2018. **24**(3): p. 511-520.
16. Iñarrairaegui, M., I. Melero, and B. Sangro, *Immunotherapy of hepatocellular carcinoma: facts and hopes*. Clinical Cancer Research, 2018. **24**(7): p. 1518-1524.
17. Xiao, B., et al., *Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma*. Frontiers in oncology, 2020. **10**: p. 2885.
18. Fridman, W.H., et al., *The immune contexture in human tumours: impact on clinical outcome*. Nature Reviews Cancer, 2012. **12**(4): p. 298-306.
19. Ali, T. and P. Grote, *Beyond the RNA-dependent function of LncRNA genes*. elife, 2020. **9**: p. e60583.
20. Peng, W.-X., P. Koirala, and Y.-Y. Mo, *LncRNA-mediated regulation of cell signaling in cancer*. Oncogene, 2017. **36**(41): p. 5661-5667.
21. Huang, D., et al., *NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death*. Nature immunology, 2018. **19**(10): p. 1112-1125.
22. Zhang, M., et al., *LncRNA GATA3-AS1 facilitates tumour progression and immune escape in triple-negative breast cancer through destabilization of GATA3 but stabilization of PD-L1*. Cell proliferation, 2020. **53**(9): p. e12855.
23. Chen, Y., et al., *LncRNA LOC100129620 promotes osteosarcoma progression through regulating CDK6 expression, tumor angiogenesis, and macrophage polarization*. Aging (Albany NY), 2021. **13**(10): p. 14258.

24. Wang, L., et al., *Long non-coding RNA LINC00426 contributes to doxorubicin resistance by sponging miR-4319 in osteosarcoma*. Biology direct, 2020. **15**(1): p. 1-11.
25. Li, Y., et al., *Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers*. Nature communications, 2020. **11**(1): p. 1-13.
26. Hong, W., et al., *Immune-related lncRNA to construct novel signature and predict the immune landscape of human hepatocellular carcinoma*. Molecular Therapy-Nucleic Acids, 2020. **22**: p. 937-947.
27. Zhang, Y., et al., *Immune-related long noncoding RNA signature for predicting survival and immune checkpoint blockade in hepatocellular carcinoma*. Journal of cellular physiology, 2020. **235**(12): p. 9304-9316.
28. Diboun, I., et al., *Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma*. BMC genomics, 2006. **7**(1): p. 1-14.
29. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data*. Biostatistics, 2003. **4**(2): p. 249-264.
30. Chen, C., et al., *Immunotherapy for osteosarcoma: Fundamental mechanism, rationale, and recent breakthroughs*. Cancer Letters, 2021. **500**: p. 1-10.
31. Şenbabaoğlu, Y., et al., *Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures*. Genome biology, 2016. **17**(1): p. 1-25.
32. Zhang, T., et al., *Identification of immune-related prognostic genes and lncRNAs biomarkers associated with osteosarcoma microenvironment*. Frontiers in oncology, 2020. **10**: p. 1109.
33. Statello, L., et al., *Gene regulation by long non-coding RNAs and its biological functions*. Nature Reviews Molecular Cell Biology, 2021. **22**(2): p. 96-118.
34. Luo, L., et al., *Regulating immune memory and reversing tumor thermotolerance through a step-by-step starving-photothermal therapy*. Journal of nanobiotechnology, 2021. **19**(1): p. 1-18.
35. Noble, F., et al., *Tumour infiltrating lymphocytes correlate with improved survival in patients with oesophageal adenocarcinoma*. Cancer Immunology, Immunotherapy, 2016. **65**(6): p. 651-662.
36. Welters, M.J., et al., *Intratumoral HPV16-specific T cells constitute a type I-oriented tumor microenvironment to improve survival in HPV16-driven oropharyngeal cancer*. Clinical Cancer Research, 2018. **24**(3): p. 634-647.
37. Reeves, E., et al., *HPV epitope processing differences correlate with ERAP1 allotype and extent of CD8+ T-cell tumor infiltration in OPSCC*. Cancer immunology research, 2019. **7**(7): p. 1202-1213.
38. Wang, Y.-Q., et al., *Prognostic value of immune score in nasopharyngeal carcinoma using digital pathology*. Journal for immunotherapy of cancer, 2020. **8**(2).
39. Pan, Y., et al., *Tumor-associated macrophages in tumor immunity*. Frontiers in immunology, 2020. **11**: p. 3151.

40. Pucci, M., et al., *Tumor-Derived Small Extracellular Vesicles Induce Pro-Inflammatory Cytokine Expression and PD-L1 Regulation in M0 Macrophages via IL-6/STAT3 and TLR4 Signaling Pathways*. International Journal of Molecular Sciences, 2021. **22**(22): p. 12118.
41. Yang, W., et al., *ACE2 correlated with immune infiltration serves as a novel prognostic biomarker in clear cell renal cell carcinoma: implication for COVID-19*. International journal of biological sciences, 2021. **17**(1): p. 20.
42. Bernsmeier, C., S. van der Merwe, and A. Périanin, *Innate immune cells in cirrhosis*. Journal of hepatology, 2020. **73**(1): p. 186-201.
43. Vitale, I., et al., *Macrophages and metabolism in the tumor microenvironment*. Cell metabolism, 2019. **30**(1): p. 36-50.
44. Li, H. and S.K. Anderson, *Association of TNFRSF1B promoter polymorphisms with human disease: Further studies examining T-regulatory cells are required*. Frontiers in immunology, 2018. **9**: p. 443.
45. Vanamee, É.S. and D.L. Faustman, *TNFR2: a novel target for cancer immunotherapy*. Trends in molecular medicine, 2017. **23**(11): p. 1037-1046.
46. Raimondo, S., et al., *Multiple myeloma-derived exosomes are enriched of amphiregulin (AREG) and activate the epidermal growth factor pathway in the bone microenvironment leading to osteoclastogenesis*. Journal of hematology & oncology, 2019. **12**(1): p. 1-15.
47. Mizuno, R., et al., *The role of tumor-associated neutrophils in colorectal cancer*. International journal of molecular sciences, 2019. **20**(3): p. 529.
48. Ruess, D.A., et al., *Mutant KRAS-driven cancers depend on PTPN11/SHP2 phosphatase*. Nature medicine, 2018. **24**(7): p. 954-960.
49. Yang, B.-y., et al., *PSMB8 regulates glioma cell migration, proliferation, and apoptosis through modulating ERK1/2 and PI3K/AKT signaling pathways*. Biomedicine & Pharmacotherapy, 2018. **100**: p. 205-212.
50. Chang, H.H., et al., *PSMB8 inhibition decreases tumor angiogenesis in glioblastoma through vascular endothelial growth factor A reduction*. Cancer science, 2020. **111**(11): p. 4142.

Figures

Figure 1

Flowchart of the study involving the identification of IR-mRNAs and the construction of IR-lncRNA-based immune subgroups.

Figure 2

Construction and verification of OS subgroups. **a, b** The OS samples were divided into 3 subgroups (immune-H, immune-M, and immune-L) based on the 29 immune-related gene set scores by applying an unsupervised hierarchical clustering algorithm in two databases. **c, d** The violin map shows the difference in the ESTIMATE score, immune score, stromal score, and tumor purity among the subgroups. **e** There was a positive correlation between the overall survival rate and subgroups in TARGET-OS. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; ns, no statistical significance.

Figure 3

Differences in TIICs among the immune subgroups. **a, b, c, d** The boxplot shows the TIICs in the 3 subgroups using the CIBERSORT algorithm and the correlations between immune cells in the two databases.

Figure 4

Survival curves of 22 immune cells were created using the Kaplan–Meier method in TARGET-OS; red indicates high expression of immune cells, and blue indicates low expression. high infiltration of **a** activated mast cells, **b** plasma cells, **c** resting NK cells and had a significantly poorer survival rate, while those with high infiltration of **a** activated NK cells, **b** CD8+ T cells and **c** f follicular helper T had a significantly better survival rate.

Figure 5

Analysis of IR-mRNA. **a** The Venn diagram shows a total of 654 IR-mRNAs in TARGET-OS and GEO-OS. **b** PPI network diagram of IR-mRNAs. **c, d** The MCODE plug-in in Cytoscape selects the closest IR-mRNA cluster.

Figure 6

Identification of IRD-lncRNAs. **a** The heatmaps and **b** volcano maps **b** show the 117 IRD-lncRNAs between the Immunity-H and Immunity-L subgroups.

Figure 7

Construction of the IRD-IncRNA prognostic signature. **a** LASSO Cox analysis identified 17 IRD-IncRNAs with prognostic value. **b** Partial likelihood deviation of the LASSO coefficient distribution. **c** The five IRD-IncRNAs as the prognostic signature, with the HR and p value calculated by multivariate Cox regression analysis. **d** Coexpression relationship between the five IRD-IncRNAs of the prognostic signature and IR-mRNAs.

Figure 8

Verification the reliability of the prognostic signature. The risk curve of each sample was reordered by the risk score in the training, validation, and total cohorts. **b** Scatter plot of the survival status of the sample in the training, validation, and total cohorts. **c** The overall survival rate of patients in the high-risk group was lower than that of patients in the low-risk group in the training, validation, and total cohorts. **d** Time-dependent ROC curve for 1-year, 3-year, and 5-year overall survival predictions in the training, validation, and total cohorts.

Figure 9

Evaluation of independent prognostic value. **a** Multivariate Cox regression analysis of risk score, age, sex, and metastasis. **b** AUC based on ROC curve transfer model.

Figure 10

Correlation between prognostic signature of the five IRD-lncRNA and TIICs. ***p <0.001, **p <0.01, *p <0.05; ns, no statistical significance.