Identification of Metastasis-Related Signature For Predicting Survival In Osteosarcoma

Song Shi

Affiliated Hospital of Nantong University

Shuaijie Yang

Affiliated Hospital of Nantong University

Zhenyu Zhou

Affiliated Hospital of Nantong University

Kai Sun

Affiliated Hospital of Nantong University

Ran Tao

Affiliated Hospital of Nantong University

Fuyin Wan (✉ jshzwfy@126.com)

Affiliated Hospital of Nantong University

Research Article

Keywords: osteosarcoma, prognosis, risk score, immune subtype

Posted Date: July 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-683582/v1>

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

[Read Full License](#)

Abstract

Background

RNA sequencing has become a powerful tool for exploring tumor recurrence or metastasis mechanisms. In this study, we aimed to develop a signature to improve the prognostic predictions of osteosarcoma.

Materials and methods

By comparing the expression profiles between metastatic and non-metastatic samples, we obtained 57 metastatic-related gene signatures. Then we constructed a 3-gene signature to predict the prognostic risk of osteosarcoma patients by the Cox proportional hazards regression model. The risk score derived from this signature could successfully stratify osteosarcoma patients into subgroups with different survival outcomes.

Results

Patients in the low-risk group showed more prolonged overall survival than those in the high-risk group. And the performance was validated with another independent dataset. Multivariate cox regression revealed that the risk score served as an independent risk factor. Besides, we found that patients with low-risk scores had higher expression levels of immune-related signatures, suggesting an active immune status in those patients. Using the CIBERSORT database, we further systematically analyzed the relationships between the risk score and immune cell infiltration levels, as well as the immune activation markers. Higher infiltration of immune cells (CD8 T cells, monocytes, M2 macrophages, and memory B cells) and higher levels of immune cytotoxic markers (GZMA, GMZB, IFNG, and TNF) were observed in patients in the low-risk group.

Conclusions

In summary, this 3-gene signature could be a reliable marker for prognostic evaluation and help clinicians identify high-risk patients.

Introduction

Osteosarcoma (OS) is one of the most common malignancies of bone, which mainly affects adolescents and young adults¹. OS is characterized by early metastasis and poor prognosis. Patients with localized OS at diagnosis had about 80% 5-year survival rate, whereas only 20% of patients with metastatic disease are becoming long-term survivors². Metastasis was the primary cause of death in patients with OS, and pulmonary metastases constitute a majority³. OS is often treated with a combination of therapies that includes surgery, chemotherapy, and radiation therapy^{4,5}. Such standard treatment achieved survival benefit in approximately 60% of patients with low-grade OS and 20–30% for patients with metastases⁴. Developing methods for early detection of disease and metastasis would primarily

benefit OS patients. However, for metastatic OS patients, neither prognostic factors nor appropriate treatment strategies have been well established.

Identifying prognostic biomarkers for OS patients, especially for metastatic patients, is desperately needed for improving overall survival^{1,6}. The integration of genome-scale genomic and transcriptomic data enabled researchers to identify critical genes related to the development and pathogenesis of OS^{7,8}. Li et al. reported that high expression of matrix metalloproteinase 9 (MMP-9) was associated with the poor prognosis of patients with OS⁹. Other studies found HIF and VEGF were up-regulated in patients with OS. High expression of HIF and VEGF was associated with poor overall survival of OS patients¹⁰⁻¹². Beta-galactoside binding proteins of the lectin family, such as galectin 1 and galectin 3, were also reported to promote the migration and invasion of OS cells¹³⁻¹⁵. Apart from these proteins, some miRNAs have been reported to be markers for OS. Previous studies identified that compared with healthy donors, the expression levels of miR-21, miR-34, miR-140, miR-143, miR-199a-3p, and miR-132 were higher in the serum of OS patients^{16,17}, whereas miR-199a-3p, miR-195, and miR-143 were lower in the serum of patients with OS¹⁸. Such studies provided scientific rationale for the potential use of these biomarkers for the early detection and monitoring of the progression of OS. However, very few biomarkers could be reliably used by the cancer caregiver in the clinical settings to make a precise diagnosis due to the lack of clinical evidence.

Apart from the biomarker identification, multiple cooperative clinical trials aimed at improving outcomes for patients with OS have been performed. One clinical trial explored the potentiation of adding high-dose ifosfamide and etoposide to chemotherapy for poor histologic responders with surgical resection¹⁹. Unfortunately, the addition of high-dose ifosfamide and etoposide had limited success. No statistical significance was observed between these two arms, which implied that the addition of high-dose ifosfamide and etoposide could not improve the outcome for OS patients¹⁹. Another study attempted to investigate the benefit of adding interferon alfa-2b to methotrexate, doxorubicin, and cisplatin (MAP) therapy in patients with resectable high-grade OS²⁰. Interferon alfa-2b has been proved to elicit anti-proliferative, anti-angiogenic, and pro-apoptotic effects in multiple cancer models, including OS²¹. However, the results were disappointing, with no significant difference between MAP plus interferon alfa-2b versus MAP alone. In a randomized phase 3 trial INT-0133, the authors compared the efficacy for liposomal muramyl tripeptide phosphatidylethanolamine (MTP-PE) plus chemotherapy versus chemotherapy alone for patients with newly diagnosed metastatic osteosarcoma²². MTP-PE is an analog of muramyl dipeptide and served as a nonspecific immune modulator²³. The addition of MTP-PE to chemotherapy has been shown to improve five-year event-free survival (EFS) in patients with non-metastatic OS. Moreover, the combination therapy increased the time to distant metastasis and the decreased risk of metastasis in the surgical margin positive group. These clinical trials provided new possibilities for the treatment of OS. However, improving the prognosis of metastatic OS patients is still a relatively unexploited field.

In our study, we integrated multiple datasets to identify metastatic-related genes. Then cox regression was used to identify genes with prognostic prediction value among these metastatic-related genes. Based on such analysis, we constructed a risk model that could predict the survival of OS patients and might benefit the prognosis evaluation of OS.

Materials And Methods

Data collection and normalization

Gene expression data in this research was derived from GEO (<http://www.ncbi.nlm.nih.gov/geo/>) and National Cancer Institute TARGET Data Matrix (<https://ocg.cancer.gov/programs/target/data-matrix>). After excluding the samples without metastatic tissues, 193 OS patients in five datasets were used in this study, including GSE9508 (n = 17), GSE14359 (n = 9), GSE14827 (n = 27), GSE21257 (n = 53), TARGET (n = 101).

R program (version 3.5.1; <http://www.r-project.org>) was used for data analysis. For the microarray data, the raw “.CEL” files were downloaded. The “affy”²⁴ R package was applied for the data processing, and then the Robust Multi-array Average (RMA) method²⁵ was used for background correction, normalization, and probe-set summarization.

For RNA-Seq data, the raw data were aligned to the human reference genome using the STAR aligner. Then RSEM was applied to map aligned reads and to generate a gene count matrix²⁶. The expression matrix was normalized using the quantile normalization method.

Identification of metastasis-related genes

The “limma”²⁷ R package was used to identify differentially expressed genes (DEGs) between metastatic and non-metastatic tissues. The P-value of each DEG was calculated and then adjusted using the Benjamini and Hochberg (BH) method. In this study, the threshold used to select genes that were significantly differentially expressed was |fold change| > 1.2 and FDR < 0.1. Genes that were differentially expressed in three of the four datasets were selected and used for further analysis. The “MAGeCKFlute”²⁸ R package was used for gene ontology and gene set enrichment analysis.

Development and validation of the risk signatures

Datasets with both expression data and complete clinical information were selected for survival analysis. After excluding the samples without survival information, 87 OS samples from the TARGET cohort and 53 OS samples from the GSE21257 dataset were used for further analysis. Since the TARGET cohort contained more samples, we used this dataset as the training data. The association between the gene expression and patients' overall survival was assessed by univariate cox regression analysis. Based on the differential analysis, a list of 57 differentially expressed genes between metastatic and non-metastatic tissues was screened out. Then, cox analysis with the least absolute shrinkage and selection

operator (LASSO)²⁹ was performed using the “glmnet” R package to identify genes with the best predictive value. The optimal values of the penalty parameter lambda were determined by 10-times cross-validations at 1 SE beyond the minimum partial likelihood deviance in the training dataset. Prognostic genes with non-zero coefficients were selected from the candidate genes. The risk score of each sample was calculated based on lasso cox regression coefficients derived from the training dataset and the corresponding expression. The risk score was calculated using the following formula:

$$\text{Risk score} = \sum_{i=1}^n \text{coef}_i * \text{exp}$$

By applying the “survminer” R package, OS patients were divided into low- and high-risk groups based on the median risk score. The Kaplan-Meier curve and non-parametric log-rank test were employed to compare the survival difference between the low and high-risk groups. Univariate and multivariate cox proportional hazards regression models were used to evaluate the prognostic relationship between risk score and other clinical factors. In multivariate analysis, we adjusted for age, sex, and grade. Receiver operating characteristic (ROC) analysis was performed to detect the accuracy of the prediction using the “pROC” package of R software.

Functional annotation and gene set enrichment analysis

Gene set enrichment analysis for DEGs was performed using the “MAGeCKFlute”²⁸ R package. Kyoto Encyclopedia of Genes and Genomes (KEGG; www.genome.ad.jp/kegg/) pathway analysis and Gene Ontology (GO; www.geneontology.org) enrichment analysis were used for the enrichment annotation of DEGs. GO terms and KEGG pathways with adjust p-value < 0.05 were considered as statistically significant pathways.

Correlation analysis of immune cell infiltration

Immune cell infiltration level was estimated by the CIBERSORT³⁰ method (<https://cibersort.stanford.edu/>) to estimate the fractions of 22 tumor-infiltrating leukocytes subsets. For each tumor sample, the correlation between risk scores and immune infiltration was calculated by Spearman correlation.

Statistical analysis

All the statistical analyses were performed by R software (version 3.5.1, <http://www.R-project.org/>). The student's t-test was employed in the two-group comparisons. Multiple hypothesis testing corrections were applied where multiple hypotheses were tested and were indicated by the use of FDR. For all the statistical analyses, including t-test, cox regression analysis, and the log-rank test, p < 0.05 was defined as a significant difference.

Results

Identification of metastasis-related genes

To identify metastasis-related genes in OS, we compared the gene expression between metastatic and non-metastatic samples in four public datasets. Genes that were differentially expressed in three of the four datasets (fold change < 1.2 and FDR < 0.1) were selected as metastasis-related genes (Fig. 1A). Functional annotation of these genes revealed that genes related to embryonic placenta development, regulation of cell differentiation, phosphatidylinositol-4,5-bisphosphate 3-kinase signaling, cellular response to nitric oxide, and positive regulation of protein binding were up-regulated in patients with metastasis (Fig. 1B). In contrast, pathways associated with regulation of mitotic cell cycle, tertiary granule lumen, glycolysis, glucose metabolism, cellular response to peptide, co-stimulation by the CD28 family, and cysteine and methionine metabolism were up-regulated in non-metastatic tissues (Fig. 1B).

Generating signature genes for survival prediction

To further characterize genes with the prognostic value among these metastasis-related genes, we performed Cox proportional hazards regression. Seven genes with p-value < 0.05 were selected as prognostic related genes. Next, the LASSO cox regression model analysis was applied to select genes with predictive value by shrinking a large number of coefficients of the candidate genes towards zero (Fig. 2A and 2B). The prognostic prediction model was constructed based on the expression of genes included in the constructed signature and the corresponding coefficients. The equation was as follows: risk score = 0.05* MCAM – 0.08* PLD3–0.249*SNX2

Evaluation and validation of 3-gene signature for survival prediction

To verify whether the 3-gene signature could provide an accurate prediction of overall survival in OS patients, we first calculated a risk score for each sample in the training dataset. Then OS patients were divided into high- and low-risk groups using the median risk score as the cutoff. As expected, patients with high-risk scores had poor prognosis than those with low-risk scores ($p < 0.0001$) (Fig. 2C). And the model effectively provided prognostic classifications with an AUC of 0.7099 (Fig. 2D). Next, we assessed the reproducibility of the 3-gene signature prediction model using an independent GSE21257 cohort. Similarly, the overall survival of patients in the low-risk group was significantly better than that of the high-risk group in the validation dataset ($p = 0.016$) (Fig. 2E). ROC analysis showed that the AUC for overall survival was up to 0.8167 (Fig. 2F). Thus, the 3-gene signature model effectively provided prognostic classifications in both training and validation datasets.

To assess the independence of the 3-gene signature model in clinical applications, we performed univariate and multivariate cox regression analysis. We systematically analyzed the patients' clinical information as recorded in GSE21257 datasets, including their age, gender, disease stage, and the risk score. The univariate Cox regression analysis revealed that age, gender, and disease stage were not significantly correlated with overall survival (Fig. 3A). The risk score we generated was the only factor that significantly associated with the overall survival (HR = 0.353, 95% CI = 0.1469–0.8486, $P = 0.02$) (Fig. 3A). The corresponding multivariate cox regression analysis also found that only risk score (HR = 0.3631, 95% CI = 0.14119–0.934, $P = 0.04$) had clinical independence to associate with overall survival. (Fig. 3B).

These data suggested that the risk score is a prognostic factor independent of other clinical factors and exhibits independent predictive performance upon clinical application.

Molecular features of low-risk and high-risk osteosarcoma patients

To characterize the molecular signature difference between low-risk and high-risk groups, we compared the gene expression patterns between patients in low-risk and high-risk groups. Patients with low-risk scores had increased expression of immune-related pathways, such as chemotaxis, defense response, toll-like receptor signaling pathway, phagocytosis, and complement activation (Fig. 4A). Patients with high-risk scores had increased expression of genes associated with DNA replication, mitotic cell cycle, RNA processing, chromatin remodeling, and ribosome biogenesis (Fig. 4B). Functional annotation was further performed by using the KEGG dataset. Immune-related pathways, such as antigen-processing and presentation, toll-like signaling pathway, intestinal immune network for IgA production, phagosome, and natural killer cell-mediated cytotoxicity, were up-regulated in patients with low-risk scores (Fig. 4C). Patients in the high-risk group had a higher expression level of genes associated with cell cycle, mismatch repair, RNA transport, DNA replication, and spliceosome (Fig. 4D). Altogether, these data suggested that the molecular expression pattern, especially patients' immune status in the low-risk group, was quite different from that in the high-risk group.

Correlation between the 3-gene signature with the infiltration of immune cell subtypes

Given that the immune status in patients with low-risk score and high-risk score was quite different, we further explored the potential reason for this difference. We analyzed the correlation between the prognostic signature and the infiltration of immune cells. The immune cell infiltration level was estimated by CIBERSORT. The expression of the 3-gene signature was negatively correlated with the infiltration of CD8 T cells, neutrophils, monocytes, resting mast cells, M2 macrophages, and memory B cells (Fig. 5A), while positively correlated with activated mast cells, naïve B cells, and M0 macrophages (Fig. 5A). These data suggested that patients with low-risk scores might have higher infiltration levels of CD8 T cells, monocytes, M2 macrophages, and memory B cells. Notably, the risk score was negatively correlated with the overall infiltration level of immune cells as estimated by the absolute score in CIBERSORT (Fig. 5A), indicating patients with low-risk scores had higher general infiltration level of immune cells. Interestingly, the risk score was significantly negatively correlated with markers for leukocyte mediated cytotoxicity, such as GZMA, GZMB, IFNG, and TNF (Fig. 5B), which implying effective anti-tumor cytotoxicity in patients with the low-risk score. Altogether, these data implied that patients in the low-risk group had a higher level of immune cell infiltration and anti-tumor immune activation. This could somehow explain why patients in the low-risk group had better prognosis than those in the high-risk group.

Discussion

OS is the most common malignancy of bone³¹. It is most prevalent in teenagers and young adults³². The prognosis of OS is very poor, and metastasis is the most adverse factor at diagnosis among known

prognostic factors. Many patients develop lung metastases within one year. The 5-year survival rate of metastatic patients is only 15%². Although outcomes of patients with OS have improved since the introduction of chemotherapy, outcomes of metastatic OS are still unsatisfactory³³. Therefore, further studies on prognosis prediction are warranted. In this study, we identified metastasis-related genes by comparing the expression pattern between metastatic and non-metastatic tissues. Next, we applied the cox regression model to identify genes with the prognostic value among these metastasis-related genes. Then the LASSO cox regression model was used to calculate the coefficient of each factor. The risk score for each sample was calculated based on the prognostic genes' expression and the corresponding coefficients. By applying the 3-gene signature to the TARGET (n = 87) training set, a significant distinction was observed between low and high-risk patients. And the prediction value of our signature was also validated in the validation set, indicating the broad applicability of this signature in predicting OS patients' prognosis. Patients in the low-risk group had significantly better overall survival than those in the high-risk group. Univariate and multivariate Cox regression analyses showed that the signature was an independent factor significantly associated with the overall survival. Besides, we found that the expression of our signature gene was negatively correlated with the infiltration of some immune cells and the expression of genes associated with immune-cell mediated cytotoxicity. The differences in the immune landscape might be the reason for the different prognosis of these two groups.

The biological functions of genes included in the 3-gene signature have been reported in previous research; however, the function of these genes in OS was largely unexploited. MCAM plays a role in cell adhesion and cohesion of the endothelial monolayer at intercellular junctions in vascular tissue. MCAM had a higher expression level in OS cell lines than in normal osteoblast cells³⁴. PLD3 encodes an enzyme that catalyzes the hydrolysis of membrane phospholipids³⁵. PLD3 regulates inflammatory cytokine responses via the degradation of nucleic acids by reducing ssDNA concentration to stimulate TLR9³⁶. SNX2 encodes a member of the sorting nexin family, a phosphoinositide binding domain, and is involved in intracellular trafficking³⁷. The role of these genes in the osteosarcoma was poorly understood, and additional detailed experiments are needed to fully elucidate the specific functions of these genes in osteosarcoma.

Since this 3-gene signature can effectively screen out OS patients with poor prognosis, it is necessary to explore the underlying mechanism. The integrated analysis revealed that patients with low-risk scores had higher expression levels of immune-related genes and higher infiltration of immune cells, such as CD8 + T cells, macrophages, and neutrophils, implying a higher level of immune activation in those patients. CD8 + T cells have central roles in suppressing tumors by releasing cytotoxic molecules, such as IFN- γ , granzyme, and perforin³⁸. Higher infiltration levels of CD8 + T cells were associated with better prognosis in multiple cancer types³⁹. Innate immune cells, such as monocytes, macrophages, and neutrophils, also played important role in regulating anti-tumor immunity by initiating the inflammatory response and interacting with adaptive immune cells⁴⁰. Previous studies reported that macrophages work as mediators of bone homeostasis and osteoimmunology⁴¹. Higher infiltration level of tumor-associated macrophages (TAM) was associated with reduced metastasis and improved survival in high-grade OS

patients⁴². Macrophage activator, Muramyl tripeptide (MTP), has been reported to increase levels of TNF- α and IL-6^{22,43}. Adding MTP to conventional chemotherapy improved 6-year overall survival from 70–78% (P = 0.03) in patients with non-metastatic OS⁴⁴. Thus, high infiltration levels of cytotoxic immune cells and higher levels of tumor-killing might be the two reasons for good prognosis for patients with low-risk scores.

There are some limitations in our work. Firstly, due to the limited number of available public expression data with corresponding clinical information, only two datasets with limited numbers of patients were included in our analysis. More datasets, especially independent cohorts from multiple centers, were needed to further evaluate the reliability and the clinical applicability of this model. Secondly, our analysis is entirely based on bioinformatics analysis. More carefully defined preclinical studies and properly controlled randomized clinical trials are needed to accumulate more comprehensive evidence. Finally, more experiments should be conducted to further validate the biological function of our 3-gene signature in OS development.

Conclusion

In summary, our study successfully created a 3-gene signature as a prognostic marker that can efficiently and reliably predict the prognosis of patients with OS. The 3-gene signature for OS was associated with the infiltration of immune cell subtypes and immune activation markers. This 3-gene signature can be used as a powerful tool for prognostic evaluation and help clinicians identify high-risk patients.

Abbreviations

OS: Osteosarcoma;

DEGs: Differentially expressed genes;

RMA: Robust Multi-array Average;

LASSO: least absolute shrinkage and selection operator;

ROC: Receiver operating characteristic;

GO: Gene Ontology;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed in this study are available from the corresponding author on reasonable request.

Competing interests

All authors declare no competing interests.

Funding

This research was supported by the Municipal Science and technology Plan of Nantong City (JCZ19037).

Author's contributions

Fuyin Wan and Ran Tao are the corresponding authors, and are responsible for research supervision and conceptualization. Song Shi is the first author. Shuaijie Yang the co-first author. Song Shi, Shuaijie Yang collected the transcriptomic data and performed data analysis. All authors discussed and reviewed the manuscript.

Acknowledgements

The authors wish to thank the Gene Expression omnibus (GEO) and National Cancer Institute TARGET Data Matrix database for making their processed data publicly available.

References

1. Zamborsky R, Kokavec M, Harsanyi S, Danisovic L. Identification of Prognostic and Predictive Osteosarcoma Biomarkers. *Med Sci (Basel)*. 2019;7(2).
2. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results Program. *Cancer*. 2009;115(7):1531-1543.
3. Durfee RA, Mohammed M, Luu HH. Review of Osteosarcoma and Current Management. *Rheumatol Ther*. 2016;3(2):221-243.
4. Bielack SS, Hecker-Nolting S, Blattmann C, Kager L. Advances in the management of osteosarcoma. *F1000Res*. 2016;5:2767.
5. Mirabello L, Troisi RJ, Savage SA. International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. *Int J Cancer*. 2009;125(1):229-234.

6. Kong C, Hansen MF. Biomarkers in Osteosarcoma. *Expert Opin Med Diagn.* 2009;3(1):13-23.
7. Lowe WL, Jr., Reddy TE. Genomic approaches for understanding the genetics of complex disease. *Genome Res.* 2015;25(10):1432-1441.
8. Gorlick R. Current concepts on the molecular biology of osteosarcoma. *Cancer Treat Res.* 2009;152:467-478.
9. Li H, Zhang K, Liu LH, et al. A systematic review of matrix metalloproteinase 9 as a biomarker of survival in patients with osteosarcoma. *Tumour Biol.* 2014;35(6):5487-5491.
10. Wu Q, Yang SH, Wang RY, Ye SN, Xia T, Ma DZ. [Effect of silencing HIF-1alpha by RNA interference on expression of vascular endothelial growth factor in osteosarcoma cell line SaOS-2 under hypoxia]. *Ai Zheng.* 2005;24(5):531-535.
11. Mizobuchi H, Garcia-Castellano JM, Philip S, Healey JH, Gorlick R. Hypoxia markers in human osteosarcoma: an exploratory study. *Clin Orthop Relat Res.* 2008;466(9):2052-2059.
12. Bajpai J, Sharma M, Sreenivas V, et al. VEGF expression as a prognostic marker in osteosarcoma. *Pediatr Blood Cancer.* 2009;53(6):1035-1039.
13. Zhou X, Jing J, Peng J, et al. Expression and clinical significance of galectin-3 in osteosarcoma. *Gene.* 2014;546(2):403-407.
14. Lei P, He H, Hu Y, Liao Z. Small interfering RNA-induced silencing of galectin-3 inhibits the malignant phenotypes of osteosarcoma in vitro. *Mol Med Rep.* 2015;12(4):6316-6322.
15. Miao JH, Wang SQ, Zhang MH, et al. Knockdown of galectin-1 suppresses the growth and invasion of osteosarcoma cells through inhibition of the MAPK/ERK pathway. *Oncol Rep.* 2014;32(4):1497-1504.
16. Zhou G, Lu M, Chen J, et al. Identification of miR-199a-5p in serum as noninvasive biomarkers for detecting and monitoring osteosarcoma. *Tumour Biol.* 2015;36(11):8845-8852.
17. Ouyang L, Liu P, Yang S, Ye S, Xu W, Liu X. A three-plasma miRNA signature serves as novel biomarkers for osteosarcoma. *Med Oncol.* 2013;30(1):340.
18. Cai H, Zhao H, Tang J, Wu H. Serum miR-195 is a diagnostic and prognostic marker for osteosarcoma. *J Surg Res.* 2015;194(2):505-510.
19. Marina NM, Smeland S, Bielack SS, et al. Comparison of MAPIE versus MAP in patients with a poor response to preoperative chemotherapy for newly diagnosed high-grade osteosarcoma (EURAMOS-1): an open-label, international, randomised controlled trial. *Lancet Oncol.* 2016;17(10):1396-1408.
20. Bielack SS, Smeland S, Whelan JS, et al. Methotrexate, Doxorubicin, and Cisplatin (MAP) Plus Maintenance Pegylated Interferon Alfa-2b Versus MAP Alone in Patients With Resectable High-Grade

- Osteosarcoma and Good Histologic Response to Preoperative MAP: First Results of the EURAMOS-1 Good Response Randomized Controlled Trial. *J Clin Oncol*. 2015;33(20):2279-2287.
21. Whelan J, Patterson D, Perisoglou M, et al. The role of interferons in the treatment of osteosarcoma. *Pediatr Blood Cancer*. 2010;54(3):350-354.
22. Chou AJ, Kleinerman ES, Krailo MD, et al. Addition of muramyl tripeptide to chemotherapy for patients with newly diagnosed metastatic osteosarcoma: a report from the Children's Oncology Group. *Cancer*. 2009;115(22):5339-5348.
23. Meyers PA, Chou AJ. Muramyl tripeptide-phosphatidyl ethanolamine encapsulated in liposomes (L-MTP-PE) in the treatment of osteosarcoma. *Adv Exp Med Biol*. 2014;804:307-321.
24. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20(3):307-315.
25. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249-264.
26. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
27. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
28. Wang B, Wang M, Zhang W, et al. Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc*. 2019;14(3):756-780.
29. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J*. 2010;52(1):70-84.
30. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457.
31. Luetke A, Meyers PA, Lewis I, Juergens H. Osteosarcoma treatment - where do we stand? A state of the art review. *Cancer Treat Rev*. 2014;40(4):523-532.
32. Gianferante DM, Mirabello L, Savage SA. Germline and somatic genetics of osteosarcoma - connecting aetiology, biology and therapy. *Nat Rev Endocrinol*. 2017;13(8):480-491.
33. Carrle D, Bielack SS. Current strategies of chemotherapy in osteosarcoma. *Int Orthop*. 2006;30(6):445-451.
34. Schiano C, Grimaldi V, Casamassimi A, et al. Different expression of CD146 in human normal and osteosarcoma cell lines. *Med Oncol*. 2012;29(4):2998-3002.

35. Fazzari P, Horre K, Arranz AM, et al. PLD3 gene and processing of APP. *Nature*. 2017;541(7638):E1-E2.
36. Gavin AL, Huang D, Huber C, et al. PLD3 and PLD4 are single-stranded acid exonucleases that regulate endosomal nucleic-acid sensing. *Nat Immunol*. 2018;19(9):942-953.
37. Zhong Q, Lazar CS, Tronchere H, et al. Endosomal localization and function of sorting nexin 1. *Proc Natl Acad Sci U S A*. 2002;99(10):6767-6772.
38. Tsukumo SI, Yasutomo K. Regulation of CD8(+) T Cells and Antitumor Immunity by Notch Signaling. *Front Immunol*. 2018;9:101.
39. Lanitis E, Dangaj D, Irving M, Coukos G. Mechanisms regulating T-cell infiltration and activity in solid tumors. *Ann Oncol*. 2017;28(suppl_12):xii18-xii32.
40. Cersosimo F, Lonardi S, Bernardini G, et al. Tumor-Associated Macrophages in Osteosarcoma: From Mechanisms to Therapy. *Int J Mol Sci*. 2020;21(15).
41. Gu Q, Yang H, Shi Q. Macrophages and bone inflammation. *J Orthop Translat*. 2017;10:86-93.
42. Buddingh EP, Kuijjer ML, Duim RA, et al. Tumor-infiltrating macrophages are associated with metastasis suppression in high-grade osteosarcoma: a rationale for treatment with macrophage activating agents. *Clin Cancer Res*. 2011;17(8):2110-2119.
43. Kleinerman ES, Jia SF, Griffin J, Seibel NL, Benjamin RS, Jaffe N. Phase II study of liposomal muramyl tripeptide in osteosarcoma: the cytokine cascade and monocyte activation following administration. *J Clin Oncol*. 1992;10(8):1310-1316.
44. Meyers PA, Schwartz CL, Krailo MD, et al. Osteosarcoma: the addition of muramyl tripeptide to chemotherapy improves overall survival—a report from the Children's Oncology Group. *J Clin Oncol*. 2008;26(4):633-638.

Figures

Figure 1

Identification of metastasis-related genes. (A) Heat map showing the Log2 fold change of differentially expressed genes between metastatic and non-metastatic tissues. (B) Gene ontology analysis for differentially expressed genes between metastatic and non-metastatic samples

Figure 2

Development and validation of the risk signatures. (A) The tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. (B) LASSO coefficient profiles of the prognostic candidates. (C) Overall survival of patients harboring high (above the group median) or low (below the group median) levels of risk score for the TARGET dataset. Log-rank test was used to compute the p-value in survival analysis. (D) ROC curves analysis of 3-gene signature for the TARGET dataset.

ROC curves and AUC statistics are used to evaluate the reliability of the 3-gene signature in predicting survival. (E) Overall survival of patients harboring high (above the group median) or low (below the group median) levels of risk score for the GSE21257 dataset. Log-rank test was used to compute the p-value in survival analysis. (F) ROC curves analysis of 3-gene signature for the GSE21257 dataset. ROC curves and AUC statistics are used to evaluate the reliability of the 3-gene signature in predicting survival.

Figure 3

Univariate and multivariate cox regression analyses. (A) Univariate cox regression analyses of the risk score, tumor stage, gender, and age on the GSE21257 dataset. (B) Multivariate Cox regression analyses of the risk score, tumor stage, gender, and age on the GSE21257 dataset.

Figure 4

Gene ontology for differentially expressed genes between high-risk and low-risk group. (A) Gene sets enriched in the low-risk group. GOBP dataset was used to calculate the enrichment. (B) Gene sets enriched in the high-risk group. GOBP dataset was used to calculate the enrichment. (C) Gene sets enriched in the low-risk group. KEGG dataset was used to calculate the enrichment. (D) Gene sets enriched in the high-risk group. KEGG dataset was used to calculate the enrichment.

Figure 5

Correlation analysis between risk score and immune cell infiltration or immune cytotoxic marker. (A). Heat map showing the correlation between the risk score and immune cell infiltration level. Spearman correlation was used for the correlation analysis. (B) Correlation analysis for the risk score and cytotoxic markers (GZMA, GZMB, IFNG, TNF).