Int J Clin Exp Med 2021;14(8):2242-2249
www.ijcem.com /ISSN:1940-5901/IJCEM0128895

Original Article

Establishment of key genes and associated outcomes in osteosarcoma patients using bioinformatics methods

Peng Fu[1], Yu Shi[3], Gang Chen[1], Zhenzhen Gao[2]

[1]*Department of Orthopedic, The Second Affiliated Hospital of Jiaxing University, Jiaxing 314000, Zhejiang, China;* [2]*Department of Clinical Oncology, The Second Affiliated Hospital of Jiaxing University, Jiaxing 314000, Zhejiang, China;* [3]*Department of Radiotherapy, Affiliated Hospital of Nantong University, Nantong 226000, Jiangsu, China*

Received December 24, 2020; Accepted June 8, 2021; Epub August 15, 2021; Published August 30, 2021

Abstract: Background: Osteosarcoma (OS) is a highly malignant bone tumor with a poor prognosis and it mainly occurs in children and adolescents. This condition results from a series of molecular events, including somatic mutations, epigenetic alterations, and aberrant protein expression. Patients and methods: In this study, we obtained RNA sequencing and clinical characteristic data from OS patients from the TARGET database. EdgeR package was applied to identify differentially expressed RNAs. Thereafter, a COX regression model was established to identify the key genes correlated to survival. Based on the results of the multi-COX regression model, the risk model was established to assign OS patients to high-risk and low-risk groups. Lastly, the risk model was verified by the receiver operating characteristic curves involved in the development and progression of OS. Results: RNAs that were differentially expressed among 69 metastatic samples and 210 non-metastatic samples were evaluated using the edgeR model. The COX regression model (with high sensitivity and specificity) and Kaplan-Meier (K-M) analyses were used to establish a model of survival. Thereafter, 131 differentially expressed RNAs (DEGs) were detected in this study. Single and multiple factor COX analyses showed that 10 genes were involved in the prognosis of OS. The K-M analysis identified a total of six genes (*ARX*, *DDN*, *MYC*, *NNAT*, *TAC4*, and *TRPM5*) that were closely related to the prognosis of OS patients. Conclusion: This study found six genes related to the prognosis of OS through R Software analysis and further verified by COX regression models. The results of this study provide important information for further research in the direction of OS treatment.

Keywords: Osteosarcoma, TARGET database, differentially expressed genes, kaplan-meier, COX regression models

Introduction

Osteosarcoma (OS) is the most common primary bone tumor. This condition mostly occurs in adolescents, is highly malignant, and has a poor prognosis. In recent years, the treatment of OS has been gradually improving. The existing treatment methods for OS mainly include surgery, neoadjuvant chemotherapy, and radiotherapy; however, its 5 year survival rate is still low (approximately 25%) [1]. In combination with the biological behavior of OS, approximately 20% of patients with OS show metastasis at the time of diagnosis and the 5 year survival rate of patients with metastatic OS is significantly lower than that of patients without metastasis [2-4]. Considering the high degree of malignancy, ease of metastasis, and limited means of treatment, the molecular mechanisms underlying OS, especially OS metastasis, must be further studied.

Genetic and micro-environmental factors play a significant role in carcinogenesis [5]. Multiple existing strategies are available for the detection of OS, which can primarily be classified into non-invasive and invasive techniques. Although numerous molecular biomarkers for OS have been studied, recent studies indicate that these biomarkers are lacking in adequate sensitivity and specificity to be used in OS screening [5, 6]. Therefore, more comprehensive studies are necessary to elucidate the molecular and genetic aspects of OS and obtain novel and reliable biomarkers for the early detection of OS. Although surgical resection is the first choice of treatment for OS worldwide, other therapeutic strategies, such as chemotherapy, radiation

Table 1A. General characteristics of patients

Characteristic	frequency	Percentage
Gender		
Male	148	75%
Female	121	25%
NA	10	
Age		
< 30	267	23%
≥ 30	12	77%
Transition state		
Non-metastasis	210	
Metastasis	69	
Tumor site		
Arm/hand	30	3%
Calvarium	2	58%
Hip	2	39%
Leg/foot	232	
Pelvis	6	
NA	7	

therapy, or a combination of both, are also used to prolong patient survival [7]. However, not all patients benefit from these strategies, and patients in advanced stages of the disease have poor prognosis. Therefore, formulating a suitable management plan for patients with metastasis and post-operative cancer recurrence is critical [8].

An increasing number of studies have shown that the abnormal expression of genes and the regulatory network formed by the RNA upstream to or interacting with the genes modified in OS play a significant role in its occurrence and development [9-11]. In this study, we extracted information (69 metastatic and 210 non-metastatic OS samples) from the TARGET database to generate differentially expressed mRNAs (DEmRNAs) and differentially expressed microRNAs (DEmiRNAs).

Materials and methods

General information

The RNA sequencing of raw data was acquired from the TARGET database (http://ocg.cancer.gov/programs/target), and 279 diagnosed OS cases met the inclusion criteria. The clinical data of 210 OS non-metastatic samples and 69 OS metastatic samples were downloaded from the TARGET database. Ethical committee assessment was not required for this study. The clinical information of OS patients is listed in **Table 1A**, **1B**.

Table 1B. Results of specific measurement data

characteristic	Mean ± sd	Range
Age	16.39±0.57 (year)	3-89
DFS	78.10±5.32 (m)	3.25-441.83
OS	3.00±1.26 (cm)	2.30-192

Data processing and differential expression analysis

The miRNA and mRNA data were corrected and normalized. Expression was calculated using the edgeR package in the R statistical software program (version 3.4.4) on Bioconductor (http://www.bioconductor.org/). The mRNAs and miRNAs that were differentially expressed between non-metastatic and metastatic samples were determined. Moreover, $|\log2FC| \geq 2$ and $P < 0.001$ were considered statistically significant. Volcano plots were obtained using the ggplot2 packages in R.

Establishment of the COX regression model of the OS

A COX regression model was established to detect the independent prognostic factors for OS, and the risk score for OS was listed as follows: Risk score = $exp_{mRNA1} * \beta_{mRNA1} + exp_{mRNA2} * \beta_{mRNA2} + ... + exp_{mRNAn} * \beta_{mRNAn}$ ("exp" implies the expression level of DEmRNAs, and "β" is the regression coefficient acquired from the multivariate COX regression model analysis) [12]. OS patients were divided into two groups based on risk scores: high-risk and low-risk groups. The "Survival ROC" package in R software (version 3.4.4) was used for all statistical analyses to measure the risk prediction rate of specific mRNAs between the two groups.

Functional enrichment analysis

Gene ontology (GO) functional analyses and KEGG pathway enrichment analyses were conducted to detect the function of the DEmRNAs in the ceRNA networks with the R clusterProfiler package. Fisher's test was used to generate the notable terms, and $P < 0.01$ was considered to be statistically significant. The afore-

mentioned test was used to generate the notable GO terms, and GO varieties with $P < 0.01$ were considered to be statistically significant.

Data processing and survival analysis

The Kaplan-Meier (K-M) method and log-rank test were used to detect the relationships among 10 DEmRNAs and DEmiRNAs. The prognosis and overall survival curves of CC patients showed significant results ($P < 0.05$).

Results

Data processing and identification of DEmRNAs and DEmiRNAs

DEmRNAs and DEmiRNAs between OS metastatic and non-metastatic tissues were generated, with $P < 0.001$ and $|logFC| > 2$ as the cutoff criteria for DEmRNAs and $P < 0.01$ and $|logFC| > 2$ as the cutoff criteria for DEmiRNAs. In summary, 268 upregulated DEmRNAs and 106 downregulated DEmRNAs were generated between OS metastatic and non-metastatic tissues (Supplementary Table 2). The three DEmiRNAs were downregulated (Supplementary Table 1).

Functional enrichment analysis

We performed GO analyses and KEGG pathway enrichment analyses on the upregulated and downregulated mRNAs and selected meaningful GO entries. The results of GO analysis are shown in **Figure 1A** and **1B**. The KEGG pathway analysis revealed that the upregulated mRNAs were enriched in cytokine-cytokine receptor interaction and arachidonic acid metabolism, while the downregulated mRNAs were not significantly enriched (Supplementary Figure 1). Then, GO analyses were performed for the target genes. The *P*-value < 0.01 and gene count r the target genes. The edhed in cyto. The genes of the upregulated mRNAs participated in the humoral immune response, proteinaceous extracellular matrix, and extracellular region, and those of the downregulated ones were involved in negative regulation of multicellular organismal process, striated muscle contraction, and negative regulation of response to external stimulus.

Establishment of prognostic models in the relationship between DEmRNAs and overall survival

In this study, the COX risk regression model was further established to find DEmRNAs related to OS prognosis, and the risk score was calculated for each patient according to the calculation formula for the risk index. The patients were then divided into high-risk and low-risk groups. The result showed that the regression model was ideal, and the single and multiple factor COX regression analyses were performed on the DEmRNAs. The results of single factor regression analysis are listed in Supplementary Table 3. The multifactor COX analysis generated 10 DEmRNAs associated with prognosis, specifically for the *MYC*, *TRABD2A*, *AOC3*, *RPL22L1*, *ARX*, *NNAT*, *TAC4*, *TRPM5*, *RP11*, and *NCR2* (**Table 2**). The prognosis analysis revealed that the mortality of high-risk patients was significantly higher than that of patients with low risk (**Figure 2A-C**).

The prognosis of the high-risk group was significantly worse than that of the low-risk group (Supplementary Figure 2A). Moreover, the 5 year survival rate related to these 10 genes were analyzed on the basis of the ROC curve analysis, and the corresponding AUC value was calculated [13]. Supplementary Figure 2B shows that the area under the ROC curve was high (AUC = 0.884), and the sensitivity and specificity were high.

Relationship between DEmRNAs and prognosis

The RNAs were analyzed through the K-M method and log-rank test, and no prognostic miRNAs were found. A possible reason could be that the miRNAs in the database were detected through PCRs by using small sample sizes. Further experiments are needed to find OS-related miRNAs. The 10 prognostic mRNAs detected in this study were further analyzed and narrowed down. Six molecules related to the prognosis of OS were finally identified among DEmRNAs: *ARX*, *DDN*, *MYC*, *NNAT*, *TAC4*, and *TRPM5* (**Figure 3**).

Discussion

OS is a highly invasive malignant tumor with a high incidence of distant metastasis. The study of biomarkers related to the metastasis of OS can provide a basis for the diagnosis and treatment of OS. RNAs interacting with genes mainly include long noncoding RNAs and microRNAs (miRNAs) [9, 14-17]. Studies found that the expression of the oncogene *MYC* is significantly increased in the samples of metastatic OS, and

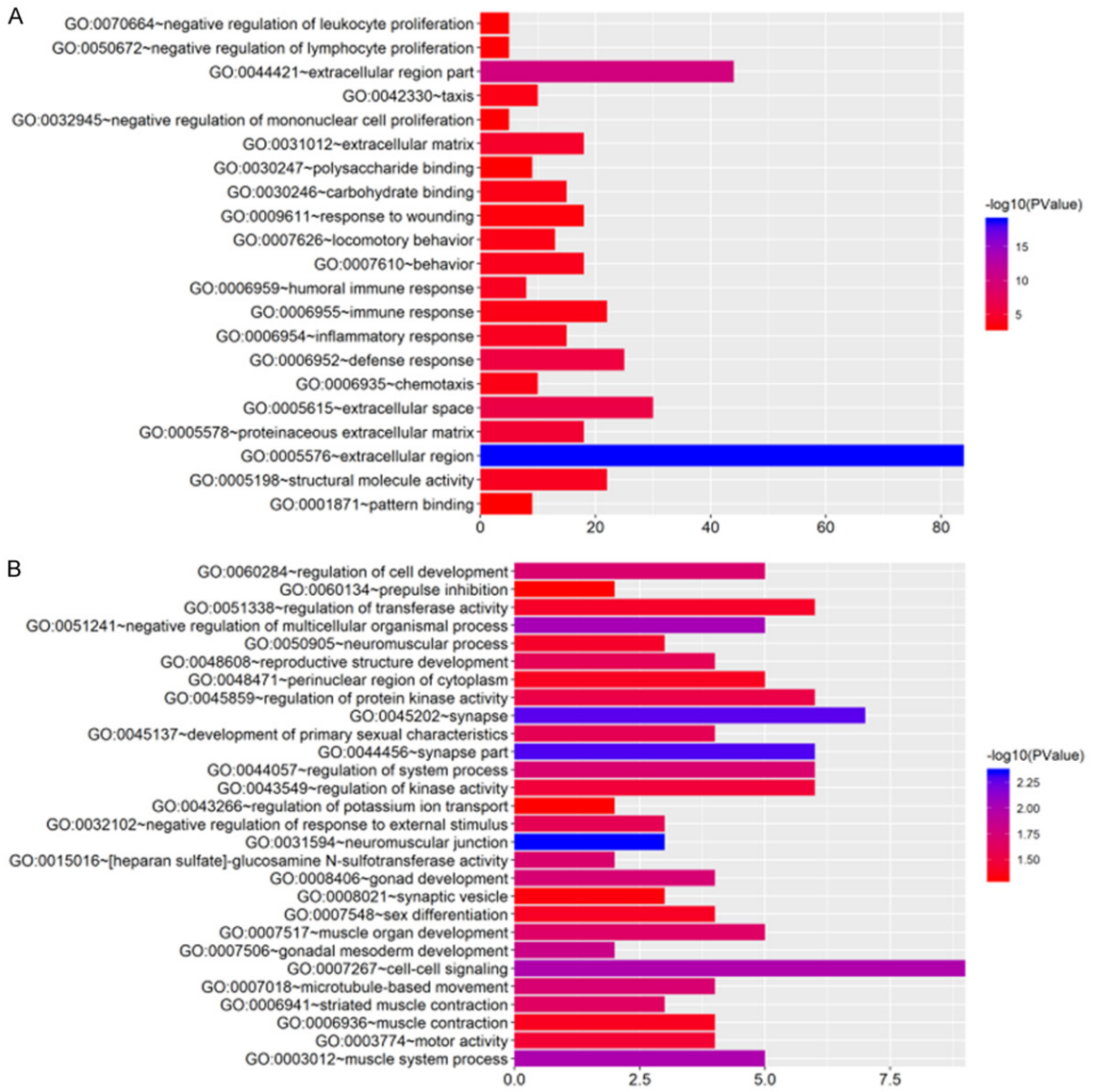


Figure 1. GO analysis of up-regulated mRNAs and down-regulated mRNAs. A. GO results for the target mRNAs of the up-regulated mRNAs, and the bar plot indicates the enrichment scores of the significant GO terms (*P*-value < 0.01, gene count 01, r t. B. GO results for the target mRNAs of the down-regulated mRNAs (*P*-value < 0.05, gene count 05, r t).

that *MYC* can promote the proliferation, metastasis, and invasion of OS cells by regulating the super-enhancer signaling pathway [18, 19]. MiRNA is a type of endogenous noncoding RNA, which can regulate the post-transcriptional expression of genes by blocking or changing the stability of mRNAs. Some studies have shown that mir-221-3p and mir-222-3p are related to the prognosis of OS and the occurrence of distant metastasis [20]. The study of abnormal mRNA and miRNA expression in metastatic OS is of great significance for the in-depth understanding of the metastatic mechanism of OS and the guiding of clinical treatment. Therefore, this study introduces mRNAs and miRNAs from two perspectives.

Studies have shown that tumors occur due to the abnormal expression of related genes [21-23]. However, a large sample of clinical data and the results of genetic testing are difficult to obtain due to the low incidence of OS. Most research sample size is limited, and the control group consisted of normal cell lines. In addi-

Table 2. Multivariate COX regression analysis of DEmRNAs

id	coef	Exp (coef)	Se (coef)	z	Pr (> |z|)
RP11-598P20.5	0.3179	1.374239	0.109915	2.892225	0.003825
AOC3	0.39861	1.489752	0.170748	2.334494	0.01957
NNAT	0.584394	1.793904	0.125376	4.661118	3.14E-06
TRPM5	0.184897	1.203094	0.107998	1.71203	0.086891
RPL22L1	0.432955	1.541807	0.201674	2.146812	0.031808
MYC	0.497304	1.644283	0.207039	2.401986	0.016306
TAC4	0.166896	1.181631	0.088592	1.883859	0.059584
TRABD2A	-0.61741	0.53934	0.194399	-3.17599	0.001493
ARX	0.289262	1.335441	0.119234	2.426002	0.015266
NCR2	0.32669	1.386372	0.130573	2.501976	0.01235

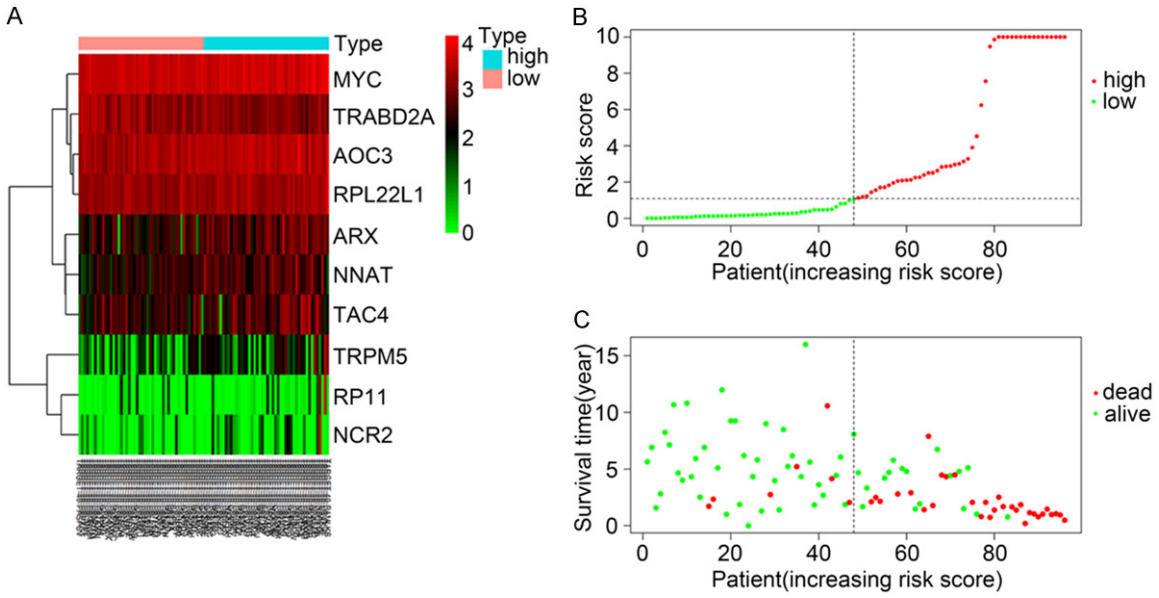


Figure 2. Six-mRNAs signatures predicted overall survival in TARGET-OS cohort. A. Heat map of the six-mRNA expression signature in OS patients. B. Risk-score distribution. Red pots demonstrating higher expression while green pots representing lower expression. C. Survival status with green indicating dead and red standing for alive.

tion, the role of oncogenes in OS is still being debated, making the results biased. In this study, we extracted the information and gene map of 279 patients with OS from the TARGET database, making this study the largest sample size known to us. The experimental and control groups in this study are humans, and the result bias is small, allowing the experimental results to be objectively evaluated.

In this study, we found 131 abnormally expressed genes by comparing data from the non-metastatic OS patients and those from patients with metastatic OS. Six confirmed genes, namely, *ARX*, *DDN*, *MYC*, *NNAT*, *TAC4*, and *TRPM5*, were found to be closely related to patient prognosis. The six genes were rendered as having a high importnace. The survival rate of patients had a downward trend with the increase in quantity of these gene expressions. The results of *MYC* in this study were consistent with those of previous studies. Ken et al. [23] discovered the *ARX* cluster through short conserved biosynthetic gene sequences, in which *ARX* showed the potential to fight the proliferation of human tumor cells. They also found that *ARX* was highly expressed in ependymoma of the spinal cord; however, the relationship between *ARX* and OS had not been reported to date. Lindenmeyer [24] found that

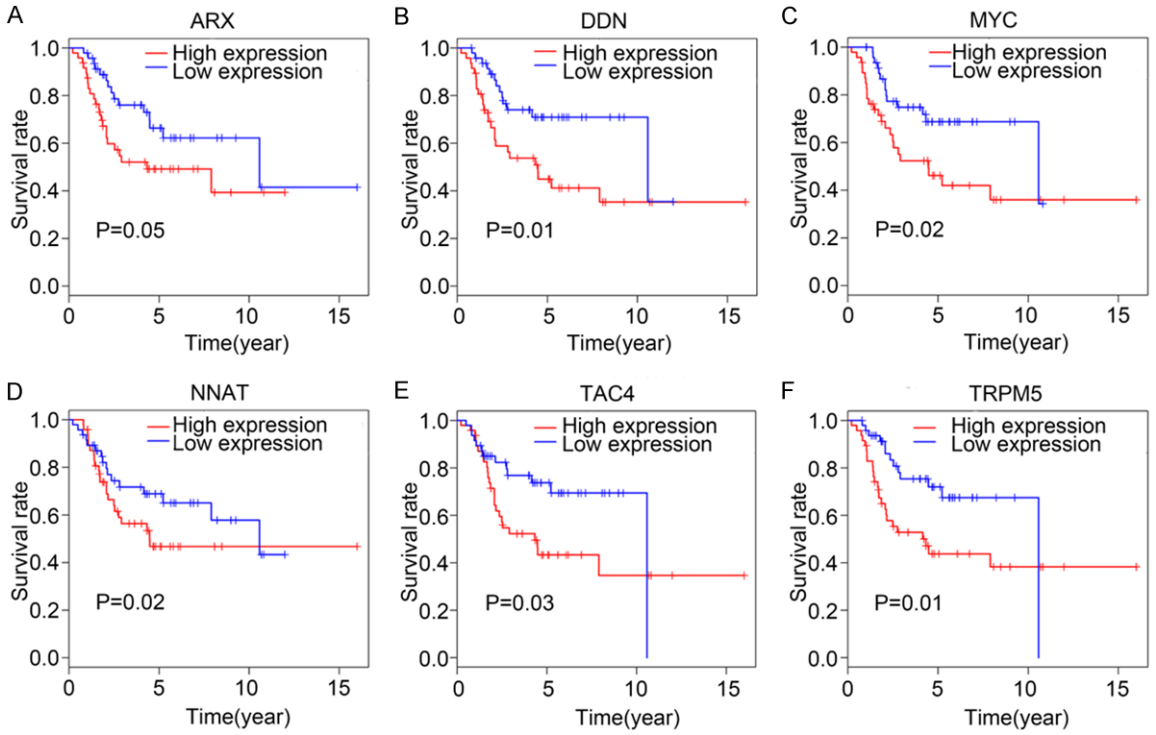


Figure 3. Kaplan-Meier survival curves for ARX, DDN, MYC, NNAT, TAC4, and TRPM5 of OS patients.

DDN was associated with glomerular function. Callahan [25] found that *DDN* was often activated in breast cancer induced by a mouse mammary tumor virus MMTV, although the relationship between *DDN* and human tumors was not yet clear. Nestheidez [11] found that *NNAT* was highly methylated and downregulated in transcription in the cell line of primary Ewing sarcoma; and Ocko-Wojciechowska [10] found that upregulated NNAT was related to the prognosis in MEN2A-like mutated medullary thyroid cancer. Hajna's [26] study, which used pneumonia-induced mice model, found that *TAC4* Hemokinin-1 plays a promoting role in the inflammation in the lung. The distant metastasis of OS *TAC4* found in this study was related to metastatic OS. A lower survival rate in patients was related with high *TAC4* expression, indicating that *TAC4* may be associated with pulmonary metastasis of OS. However, the specific mechanism for this phenomenon needs further research. Maeda's [27] results showed that *TRPM5* was associated with the poor prognosis of melanoma and gastric cancer, and that *TRPM5* inhibitors could significantly reduce the incidence of lung metastasis of melanoma in mice. However, the relationship between *TRPM5* and osteosarcoma has not been studied. Our experimental results show that the expressions of *ARX*, *DDN*, *NNAT*, *TAC4*, and *TRPM5* are increasing in metastatic OS and are related to a decreasing survival rate of patients. This study is the first to discover that these five genes are related to the metastasis and prognosis of OS, providing a new direction for the research and treatment of OS metastasis. The expression of miRNA and its target genes are extremely important in the occurrence of tumors; thus, EdgR was also used in this study to identify miRNA related to OS metastasis. Only three miRNAs had been found to be highly expressed in metastatic OS due to the limited data in the database; however, further application of the KM method and the COX analysis suggests that they were not related to the prognosis.

In summary, EdgR was used to find the differentially expressed genes in patients with metastatic OS and those without metastatic OS from a large sample database. Furthermore, the univariate and multivariate COX regression models identified six genes associated with metastasis, one of which has been previously

confirmed. The final K-M method suggested that the high expression of these six genes predicted a worse prognosis of patients.

Conclusion

The comparison of the information of patients with metastatic OS and patients with non-metastatic OS in the TARGET database indicated that six genes were related to the metastasis and prognosis of OS. *ARX*, *DDN*, *NNAT*, *TAC4*, and *TRPM5* were reported in OS for the first time, providing new gene targets for the study of the mechanism and treatment of OS metastasis.

Acknowledgements

This study was supported by the Jiaxing Science and Technology Project (No. 2020AD30084, No. 2021AD3017).

Disclosure of conflict of interest

None.

Address correspondence to: Zhenzhen Gao, Department of Oncology, The Second Affiliated Hospital of Jiaxing University, 1518 Huanchen Road, Jiaxing 314000, Zhejiang, China. Tel: +86-13736845410; E-mail: sarrah0112@163.com

References

[1] Babalyan KA, Sultanov R, Generozov EV, Zakharzhevskaya NB, Sharova EI, Peshkov MN, Vasilev AO, Govorov AV, Pushkar DY, Prilepskaya EA, Danilenko SA, Babikova EA, Larin AK and Govorun VM. Genome-wide analysis of DNA methylation in prostate cancer using the technology of Infinium HumanMethylation450 BeadChip (HM450). Vopr Onkol 2016; 62: 122-132.

[2] Zanotti S and Canalis E. Notch signaling and the skeleton. Endocr Rev 2016; 37: 223-253.

[3] Schoenebeck JJ and Ostrander EA. Insights into morphology and disease from the dog genome project. Annu Rev Cell Dev Biol 2014; 30: 535-560.

[4] Zhou Q, Zuo MZ, He Z, Li HR and Li W. Identification of circulating microRNAs as diagnostic biomarkers for ovarian cancer: a pooled analysis of individual studies. Int J Biol Markers 2018; 1724600818766500.

[5] Lin YH, Jewell BE, Gingold J, Lu L, Zhao R, Wang LL and Lee DF. Osteosarcoma: molecular pathogenesis and iPSC modeling. Trends Mol Med 2017; 23: 737-755.

[6] Kansara M, Teng MW, Smyth MJ and Thomas DM. Translational biology of osteosarcoma. Nat Rev Cancer 2014; 14: 722-735.

[7] Gianferante DM, Mirabello L and Savage SA. Germline and somatic genetics of osteosarcoma - connecting aetiology, biology and therapy. Nat Rev Endocrinol 2017; 13: 480-491.

[8] Zhang Y, Zvi YS, Batko B, Zaphiros N, O'Donnell EF, Wang J, Sato K, Yang R, Geller DS, Koirala P, Zhang W, Du X, Piperdi S, Liu Y, Zheng D, Roth M, Gill J, Zhang J, Ren T, Gorlick R, Zi X and Hoang BH. Down-regulation of Skp2 expression inhibits invasion and lung metastasis in osteosarcoma. Sci Rep 2018; 8: 14294.

[9] Ram Kumar RM, Boro A and Fuchs B. Involvement and clinical aspects of microrna in osteosarcoma. Int J Mol Sci 2016; 17: 877.

[10] Oczko-Wojciechowska M, Swierniak M, Krajewska J, Kowalska M, Kowal M, Stokowy T, Wojtas B, Rusinek D, Pawlaczek A, Czarniecka A, Szpak-Ulczok S, Gawlik T, Chmielik E, Tyszkiewicz T, Nikiel B, Lange D, Jarzab M, Wiench M and Jarzab B. Differences in the transcriptome of medullary thyroid cancer regarding the status and type of RET gene mutations. Sci Rep 2017; 7: 42074.

[11] Nestheide S, Bridge JA, Barnes M, Frayer R and Sumegi J. Pharmacologic inhibition of epigenetic modification reveals targets of aberrant promoter methylation in Ewing sarcoma. Pediatr Blood Cancer 2013; 60: 1437-1446.

[12] Yang R, Xing L, Wang M, Chi H, Zhang L and Chen J. Comprehensive analysis of differentially expressed profiles of lncRNAs/mRNAs and miRNAs with associated ceRNA networks in triple-negative breast cancer. Cell Physiol Biochem 2018; 50: 473-488.

[13] Li Y, Wang Z, Li J and Sang X. Diallyl disulfide suppresses FOXM1-mediated proliferation and invasion in osteosarcoma by upregulating miR-134. J Cell Biochem 2018; [Epub ahead of print].

[14] Li D, Yang J, Ma H, Sun C and Feng R. Inositol polyphosphate-4-phosphatase type II and rucaparib treatment inhibit the growth of osteosarcoma cells dependent on phosphoinositide 3-kinase/protein kinase B pathway. J Cell Biochem 2018; 119: 9899-9909.

[15] Lehen'kyi V and Prevarskaya N. Oncogenic TRP channels. Adv Exp Med Biol 2011; 704: 929-945.

[16] Kennedy MI and LaPrade RF. Editorial commentary: is the anterolateral ligament always injured in knee dislocations? Does it matter? Arthroscopy 2018; 34: 1898-1899.

[17] He X, Gao Z, Xu H, Zhang Z and Fu P. A meta-analysis of randomized control trials of surgical methods with osteosarcoma outcomes. J Orthop Surg Res 2017; 12: 5.

[18] Chen D, Zhao Z, Huang Z, Chen DC, Zhu XX, Wang YZ, Yan YW, Tang S, Madhavan S, Ni W, Huang ZP, Li W, Ji W, Shen H, Lin S and Jiang YZ. Super enhancer inhibitors suppress MYC driven transcriptional amplification and tumor progression in osteosarcoma. Bone Res 2018; 6: 11.

[19] Dong B, Wang G, Yao J, Yuan P, Kang W, Zhi L and He X. Predicting novel genes and pathways associated with osteosarcoma by using bioinformatics analysis. Gene 2017; 628: 32-37.

[20] Xu B, Yang T, Wang Z, Zhang Y, Liu S and Shen M. CircRNA CDR1as/miR-7 signals promote tumor growth of osteosarcoma with a potential therapeutic and diagnostic value. Cancer Manag Res 2018; 10: 4871-4880.

[21] Palmini G, Marini F and Brandi ML. What is new in the miRNA world regarding osteosarcoma and chondrosarcoma? Molecules 2017; 22: 417.

[22] Luo SS, Liao XW and Zhu XD. Prognostic value of excision repair cross-complementing mRNA expression in gastric cancer. Biomed Res Int 2018; 2018: 6204684.

[23] Kim KT, Lee CH, Chung CK and Kim JH. Is NF2 a key player of the differentially expressed gene between spinal cord ependymoma and intracranial ependymoma? World Neurosurg 2018; 118: e906-e917.

[24] Lindenmeyer MT, Eichinger F, Sen K, Anders HJ, Edenhofer I, Mattinzoli D, Kretzler M, Rastaldi MP and Cohen CD. Systematic analysis of a novel human renal glomerulus-enriched gene expression dataset. PLoS One 2010; 5: e11545.

[25] Callahan R, Mudunur U, Bargo S, Raafat A, McCurdy D, Boulanger C, Lowther W, Stephens R, Luke BT, Stewart C, Wu X, Munroe D and Smith GH. Genes affected by mouse mammary tumor virus (MMTV) proviral insertions in mouse mammary tumors are deregulated or mutated in primary human mammary tumors. Oncotarget 2012; 3: 1320-1334.

[26] Hajna Z, Borbely E, Kemeny A, Botz B, Kereskai L, Szolcsanyi J, Pinter E, Paige CJ, Berger A and Helyes Z. Hemokinin-1 is an important mediator of endotoxin-induced acute airway inflammation in the mouse. Peptides 2015; 64: 1-7.

[27] Maeda T, Suzuki A, Koga K, Miyamoto C, Maehata Y, Ozawa S, Hata RI, Nagashima Y, Nabeshima K, Miyazaki K and Kato Y. TRPM5 mediates acidic extracellular pH signaling and TRPM5 inhibition reduces spontaneous metastasis in mouse B16-BL6 melanoma cells. Oncotarget 2017; 8: 78312-78326.

Supplementary Table 1. Differentially expressed microRNAs (DEmiRNAs)

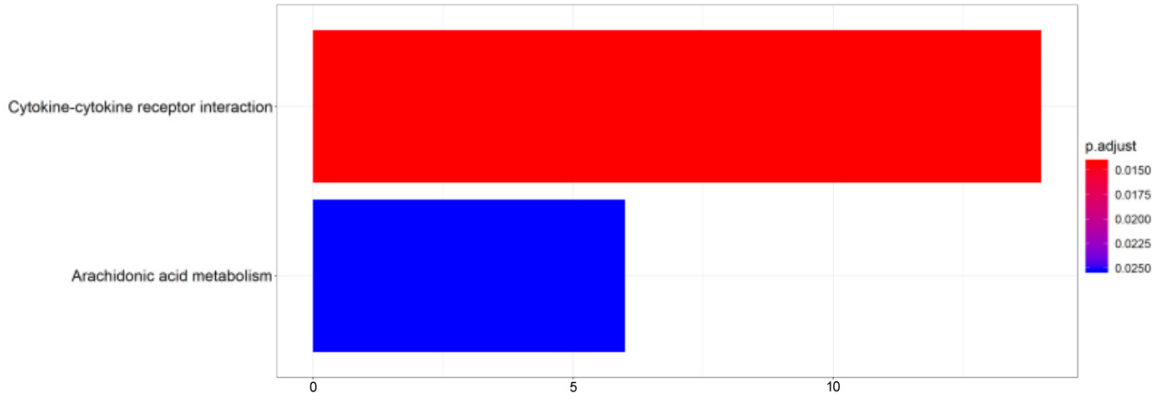
gene	Non-metastatic Mean	Metastatic Mean	logFC	*p* Value
hsa-miR-106a-4395280	0.69	0.1467	-2.2354	0.01
hsa-miR-17-4395419	0.64	0.1305	-2.2960	0.02
hsa-miR-19b-4373098	0.51	0.0922	-2.4794	0.03

Supplementary Table 2. Differentially expressed message RNAs (DEmRNAs)

	logFC	logCPM	*P* Value	FDR
SCGB3A1	6.504169	4.27811	4.43E-35	7.80E-31
FIGF	5.214485	1.827453	3.20E-22	2.81E-18
INMT	3.568238	3.03606	1.33E-20	7.80E-17
PGC	5.536455	-0.05616	2.85E-19	1.12E-15
SLC34A2	4.265108	2.952162	3.19E-19	1.12E-15
MUC17	5.063942	1.451461	9.07E-19	2.66E-15
SCGB1A1	8.248815	3.001746	6.30E-18	1.58E-14
SFTPB	5.118472	5.386341	7.64E-18	1.68E-14
TCF21	5.31579	1.13571	2.03E-17	3.97E-14
NAPSA	3.775943	2.061817	3.37E-16	5.92E-13
ADAMTS8	3.978746	0.823767	1.66E-15	2.66E-12
RP1-37E16.12	7.195868	-2.59078	1.86E-15	2.73E-12
KRT83	4.839869	-2.64328	5.26E-15	7.11E-12
SFTPA2	8.290663	4.72455	6.05E-15	7.60E-12
ADH1B	3.922775	2.420327	9.63E-15	1.13E-11
HPSE2	4.4419	0.226669	1.38E-14	1.52E-11
SFTPD	4.245484	1.106065	6.69E-14	6.92E-11
MAP1LC3C	2.885403	0.580404	1.09E-13	1.07E-10
PTGER1	3.052288	1.207533	1.81E-13	1.68E-10
SCGB3A2	2.902414	1.888381	3.03E-13	2.67E-10
SFTPC	7.349726	5.244315	3.88E-13	3.25E-10
C4BPA	3.89878	0.443418	6.15E-13	4.72E-10
PRG4	3.033902	1.205573	6.21E-13	4.72E-10
CXCL17	4.370957	-0.6404	6.44E-13	4.72E-10
PIGR	2.729291	1.637135	1.25E-12	8.57E-10
CCR9	3.858688	-1.72956	1.27E-12	8.57E-10
CRYBA1	3.376132	-2.02999	1.58E-12	1.03E-09
JCHAIN	3.819069	4.376299	1.69E-12	1.06E-09
TUBA4B	2.676953	-2.73048	2.80E-12	1.70E-09
STAB2	3.075031	1.231427	3.23E-12	1.89E-09
KRTAP19-6	6.010323	-1.87131	3.40E-12	1.93E-09
SLC22A12	4.842667	-2.6682	3.63E-12	1.99E-09
RPRML	3.422452	-0.40385	5.55E-12	2.96E-09
SFTPA1	8.114813	4.149211	6.49E-12	3.36E-09
TMEM190	3.922198	-1.15074	8.34E-12	4.19E-09
C20orf85	5.121539	-1.73556	1.41E-11	6.87E-09
S100A14	3.107978	-1.4941	1.65E-11	7.73E-09
CFAP57	2.404651	-0.33237	1.67E-11	7.73E-09
FMO2	3.643196	2.259354	2.08E-11	9.40E-09
MMRN1	2.550097	2.347179	2.65E-11	1.17E-08
WDR38	2.600632	-1.90751	3.82E-11	1.64E-08

PI16	3.462455	0.200841	6.29E-11	2.63E-08
AGR2	4.03327	-0.8158	1.09E-10	4.47E-08
PCSK5	2.392097	6.993292	1.17E-10	4.66E-08
TMEM130	2.665706	1.836857	1.29E-10	5.06E-08
KCNA5	2.993625	-0.90438	1.61E-10	6.17E-08
PPP1R14A	2.344435	2.314045	2.25E-10	8.43E-08
C1orf87	3.33643	-0.69472	2.83E-10	1.04E-07
SFTA2	3.99412	-0.88503	3.18E-10	1.14E-07
RSPH1	2.277106	-0.57945	4.03E-10	1.42E-07
ERAS	4.28815	1.55697	4.17E-10	1.44E-07
AGER	1.868709	2.752906	4.32E-10	1.46E-07
C1orf116	3.079883	-0.31831	4.87E-10	1.62E-07
CYP4B1	2.716223	0.46713	1.83E-09	5.96E-07
SYT16	2.163863	1.243663	4.22E-09	1.35E-06
SFTA3	3.463821	-0.89702	6.72E-09	2.11E-06
SERPINB3	4.284251	-1.37102	9.76E-09	3.01E-06
ADRB3	2.228174	-1.18906	1.07E-08	3.24E-06
CFAP100	2.600542	-2.3064	1.27E-08	3.80E-06
RP13-1032I1.10	3.917276	-1.1254	1.82E-08	5.32E-06
SCN7A	2.393107	1.182559	2.09E-08	6.02E-06
RP11-598P20.5	5.165679	-1.17792	2.26E-08	6.40E-06
AGR3	4.458271	-1.92444	2.48E-08	6.92E-06
MMP1	2.256652	3.254618	2.54E-08	6.98E-06
CYP4F8	1.958046	1.367274	3.50E-08	9.47E-06
FBL	1.248893	7.386184	3.70E-08	9.87E-06
ACPP	1.867134	1.140971	3.80E-08	9.97E-06
KRT14	3.234002	0.820791	4.34E-08	1.12E-05
TNFRSF13B	2.900369	0.27584	5.64E-08	1.44E-05
ALDH1A1	2.076697	4.057225	6.11E-08	1.53E-05
C10orf10	1.805004	5.324286	6.29E-08	1.56E-05
CLEC4G	2.499008	-0.00624	7.43E-08	1.82E-05
PRR15L	2.255984	-1.75116	8.01E-08	1.93E-05
CGB5	3.506678	-1.2611	9.85E-08	2.34E-05
FOXI1	3.766873	-1.55618	1.18E-07	2.76E-05
TNXB	1.409443	3.814158	1.34E-07	3.10E-05
GPI	1.325771	9.214929	2.19E-07	4.99E-05
DCC	2.173719	0.678481	2.27E-07	5.11E-05
GABRB3	2.009213	1.90354	2.39E-07	5.32E-05
CFAP73	1.883921	-1.29251	2.42E-07	5.32E-05
RAB26	1.514052	0.074503	2.45E-07	5.32E-05
AOC3	1.329264	5.411164	2.49E-07	5.33E-05
CFAP45	1.483386	0.443026	2.63E-07	5.57E-05
MYOC	2.759768	-2.18453	3.05E-07	6.38E-05
MUCL1	3.400534	-2.63272	3.68E-07	7.61E-05
GJC2	1.250637	1.362764	4.46E-07	9.12E-05
TSLP	2.423988	-0.50692	5.97E-07	0.000121
FNDC7	2.50728	-1.37293	6.51E-07	0.00013
FAM216B	2.325743	-1.58278	6.72E-07	0.000133
DLL1	1.489549	3.710405	7.77E-07	0.000152
MAP3K8	1.364869	2.014094	9.77E-07	0.000189

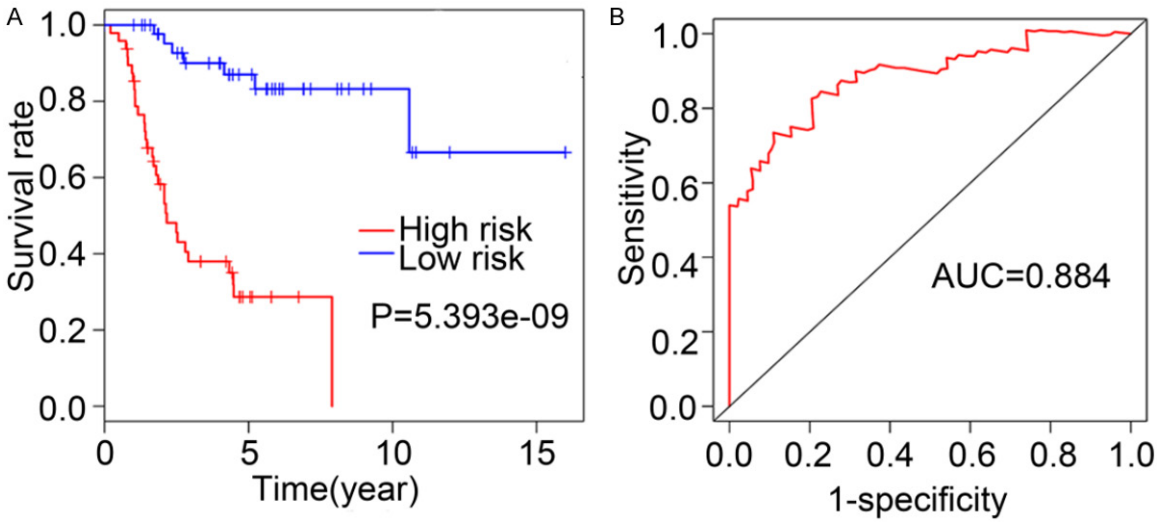
TACSTD2	1.861716	1.094718	1.02E-06	0.000195
C19orf33	2.05122	-1.42548	1.05E-06	0.000199
RAB25	2.249974	-1.91512	1.06E-06	0.000199
WNT16	2.153828	2.585792	1.16E-06	0.000213
TFF3	2.308778	0.417743	1.18E-06	0.000215
TMEM125	2.993684	-0.40134	1.61E-06	0.000288
EIF3K	1.119487	7.253977	1.69E-06	0.000301
SPEF1	1.670424	-1.6009	1.73E-06	0.000303
AK8	1.612393	-0.1861	1.76E-06	0.000307
DRC1	1.874086	-1.62792	2.40E-06	0.000411
SFN	1.913194	-0.94434	2.41E-06	0.000411
DPEP3	-4.68831	0.822968	2.50E-06	0.000421
KRT7	2.251233	1.362676	2.74E-06	0.000454
CHIT1	1.686592	-0.25986	2.94E-06	0.000483
IGF2	-2.52306	10.25083	3.01E-06	0.000491
TSPAN1	1.615139	0.637707	3.10E-06	0.000496
KRT19	2.671882	0.398799	3.23E-06	0.000512
NNAT	1.658318	0.796915	3.40E-06	0.00053
KRT15	1.696104	-0.77139	3.41E-06	0.00053
ROPN1L	1.532188	-1.6596	3.55E-06	0.000548
WNT7B	1.653408	3.66718	3.64E-06	0.000556
PRSS8	1.66583	-0.2068	3.94E-06	0.000598
PCDHGB3	1.420837	0.918215	4.12E-06	0.000619
TJP3	1.860072	-1.31994	4.30E-06	0.00064
DNAI2	1.852086	-1.891	4.48E-06	0.000662
PDCD2L	1.215968	2.961734	4.63E-06	0.000679
ASTN1	-3.35075	1.475476	4.69E-06	0.000679
PRB2	-8.01043	6.256728	4.71E-06	0.000679
GAL	2.014627	2.973661	4.88E-06	0.000698
RERGL	2.896863	-0.64952	4.99E-06	0.000706
IL33	1.827406	1.116518	5.01E-06	0.000706
NEFM	-4.23031	2.377707	5.15E-06	0.000719
MSLN	2.763824	-0.12204	5.42E-06	0.000739
GPRC5A	1.428022	4.370173	5.97E-06	0.000807
WISP2	1.648382	3.85593	6.06E-06	0.000814
SLC6A3	2.737521	0.032396	6.40E-06	0.000853
BHMT	1.79711	-0.11249	6.67E-06	0.000883
SMIM10	1.026795	4.441679	7.25E-06	0.000944
FNDC8	2.024877	-2.29873	7.25E-06	0.000944
TSPY2	-6.65338	-2.02266	7.73E-06	0.000999



Supplementary Figure 1. KEGG pathway enrichment analysis. KEGG results for the up-regulared mRNAs (*P*-value < 0.01 and gene count ≥ 3). None enrichment for down-regulated mRNAs.

Supplementary Table 3. Univariate COX regression analysis of prognostic mRNAs

gene	HR	z	*p* value
MYC	2.047228	4.372698	1.23E-05
DDN	1.389636	3.646746	0.000266
FAM166B	1.249489	3.593461	0.000326
TRPM5	1.408126	3.534797	0.000408
ARX	1.409588	3.464121	0.000532
AOC3	1.61605	3.432277	0.000599
TAC4	1.323188	3.378801	0.000728
TMEM125	1.228702	3.318877	0.000904
DLL1	1.538084	3.274146	0.00106
GAL	1.286871	3.257301	0.001125
RPL22L1	1.631039	3.254148	0.001137
RP11-598P20.5	1.336689	3.238673	0.001201
NNAT	1.386732	3.094145	0.001974
TRABD2A	0.681189	-2.84275	0.004473
NCR2	1.294242	2.696542	0.007006
ST8SIA6	0.759671	-2.6314	0.008503
FOXI1	1.241073	2.618778	0.008825
TNFRSF21	0.689175	-2.59973	0.00933
GABRA5	1.181843	2.587899	0.009656



Supplementary Figure 2. The mRNAs signature of OS for the outcome. A. The survival diference between the high-risk group and low-risk group. B. ROC curves indicated that the area under the curve (AUC) is 0.884.