A Death-Related Gene Signature for Prognosis with Osteosarcoma

Bin Xie

weihaiwei renmin hospital

Shiyong Tan

weihaiwei renmin hospital

Chao Li

weihaiwei renmin hospital

Junyang Liang (✉ whwjzkw@163.com)

Weihaiwei renmin hospital <https://orcid.org/0000-0002-7527-2049>

Research

Keywords: Bioinformatics, Osteosarcoma, Death, Gene, Prognostic Signature

Posted Date: October 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-982144/v1>

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

[Read Full License](#)

Abstract

Purpose: Osteosarcoma is one of the most prevalent malignancies, and despite significant advances in its treatment, patient prognosis remains poor and survival rates are low. It is undoubtedly important to explore the possible reasons for the low survival rates of patients and to reveal the differences.

Methods and Results: We obtained RNA-Seq (HT seq) and clinical characteristics of osteosarcoma patients from the TCGA database and divided them into survival group and death group. We defined the differentially expressed genes (DEGs) between the two groups as death-related genes (DRGs) and used them to construct a prognostic signature for overall survival of patients with osteosarcoma. The results of the validation demonstrated satisfactory accuracy and predictive prognostic value of the model. In addition, we performed a series of bioinformatic analyses that identified two key genes and the regulatory networks they constituted that may play a role in the progression of osteosarcoma.

Conclusion: Our DRGs signature represents a novel and clinically useful prognostic biomarker for patients with osteosarcoma, helping to aid clinical decision-making.

1. Introduction

Osteosarcoma (OS) is one of the most common malignancies, with an incidence of one to four per million[1, 2]. It is most often seen in adolescents and originates in bone tissue, commonly in the epiphysis of bones. Osteosarcoma of the limbs can be cured in about 75% of cases with current treatments[3]. Osteosarcoma in the spine, however, has a relatively low cure rate due to the special anatomy of the spine. Moreover, osteosarcoma is highly malignant and aggressive, and spreads and metastasizes easily through the bloodstream at an early stage. Pulmonary metastases from osteosarcoma are the leading cause of death in patients[4]. Therefore, although many advances have been made in the diagnosis and treatment of osteosarcoma, its survival rate is low and its prognosis is poor.

In order to improve the prognosis of patients with osteosarcoma, it has become imperative to find the real factors that influence prognosis. A number of factors affecting the overall survival of OS patients have been reported, including the size and location of the tumor, the surgical option, and the responsiveness of the tumor to chemotherapy[5–7]. However, the predictive power of these factors is inadequate due to the heterogeneity of the sample and the diversity of clinical strategies. Therefore, it is necessary to construct more accurate, adaptable and better predictive models.

With the rise of bioinformatics and advances in molecular biology techniques, it has become possible to address these issues. Bioinformatics methods and tools are increasingly being used to explore and analyze biomolecules such as target genes or proteins. A number of studies have used bioinformatics methods to reveal the mechanisms of tumor development, including lung cancer and breast cancer. In addition, high-throughput sequencing technologies have revolutionized tumor genetics, providing comprehensive molecular characterization of a wide range of tumors, including osteosarcoma[8]. A number of epigenetic abnormalities in osteosarcoma have been identified from high-throughput

sequencing data and experimentally confirmed to be associated with the progression of osteosarcoma[9, 10]. Honglai Tian et al. used WGCNA (Weighted Gene Co-expression Network Analysis) to predict the gene expression modules associated with the mechanism of OS occurrence[4]. Hongwu Fan et al. identified several genes closely associated with osteosarcoma metastasis, including MMP3, by bioinformatics analysis[11].

Therefore, in this study, data on OS patients were obtained from the TCGA database, and various bioinformatics tools were used to obtain DRGs, analyze their functions and successfully construct a DRGs signature and a nomogram to predict the prognosis of patients with OS.

2. Materials And Methods

2.1 Acquisition of genes expression dataset and clinical features and screening of death-related genes (DRGs)

We obtained RNA-Seq (HT seq) data of OS from TARGET program in TCGA database. The clinical features of OS patients were also downloaded from TCGA. We included data from 84 cases after excluding patients with unknown survival status. 55 patients were placed in the survival group and 29 patients were placed in the death group. Then the data were analyzed using the limma package[12], $P < 0.05$ and $|\log_2FC| > 1$ was considered statistically significant when the DRGs were obtained.

2.2 Functional analysis and identification of hub genes

To reveal the functional roles of the DRGs, the clusterProfiler package[13] was used to perform GO (Gene Ontology) enrichment analysis and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis and visualize results. PPI (Protein-Protein Interaction) network was constructed by STRING database successively and the MCC algorithm in Cytohubba was then used to identify the top 50 most core genes. Cytoscape software (version 3.8.2) was utilized to visualize the network. The significant level was set as combined score > 0.4 .

2.3 Construction and validation of the DRG-Based prognostic signature

We used lasso regression via glmnet package[14] to obtain the genes associated with prognosis. In order to improve the prediction accuracy of the signature, we used more stringent criteria to screen genes and $P < 0.001$ was statistically significant level. A DRG-based prognostic signature was constructed by stepwise

multivariate Cox regression. The risk score was calculated as follows: $\text{Risk score} = \sum_i^n \text{Exp}_i \text{Coe}_i$ (Exp = expression level; Coe = regression coefficient). Patients were divided separately into low-risk group and high-risk group based on the median risk score value. We assessed differences in overall survival between two groups by KM survival analysis. A nomogram consisting of clinical features and risk score was then constructed and calibration plots were used as method of assessing the accuracy of it. Then

ROC curve analysis was carried out to compute the sensitivity and specificity of the DRG signature in predicting the survival of osteosarcoma patients using the survivalROC package[15]. The ggplot2 package[16] was used for visualizing.

2.4 Regulatory network of key genes

We took the intersection of the genes that make up the prognostic signature with the top 50 genes in the PPI and obtained two key genes. Then two databases (Targetscan and miRcode) were used separately to identify the target miRNAs for the two key genes. Following this, we used Starbase to find lncRNAs that interacted with miRNAs. Finally, the regulatory network of key genes in OS was established. In addition, the top 20 KEGG pathways in which miRNAs were mainly involved were shown.

2.5 Analysis of osteosarcoma immune infiltration patterns

The percentage of various immune cells in all OS samples was analyzed using CIBERSORT. Wilcoxon rank sum test was performed to detect the differences in immune cell proportions between the two groups and the statistical significance level was $P < 0.05$. In addition, we used difference test and correlation tests to analyze the association between key genes and the proportion of immune cell infiltration, respectively. In the difference test, all tumor samples were divided into high and low expression groups according to the median expression of key genes, and then compared the difference in the proportion of various immune cell infiltrations between the two groups. The correlation test was a method that used the Pearson test to directly determine the correlation between the expression of key genes and the proportion of immune cell infiltration. Finally, the intersection of the results obtained by the two methods was considered to be statistically significant.

2.6 Overall survival analysis of tumor microenvironment scores

We used the estimate package[17] to score the tumor microenvironment and divided the entire set into low-score group and high-score group based on the median value of the score. KM (Kaplan-Meier) survival analysis was performed to assess differences in overall survival between two groups. In addition, Wilcoxon rank sum test was used to analysis the differences of correlation between clinical characteristics and tumor microenvironment scores ($P < 0.05$).

3. Results

3.1 Identification of death-related genes (DRGs) and functional analysis

After screening the two groups of data, we obtained 561 differentially expressed genes (DEGs) (Fig. 1). The top 50 up-regulated and down-regulated DEGs were located in Table 1. Then, we defined these DEGs as death-related genes (DRGs). Fig. 2A shows the results of KEGG pathway enrichment analysis, DRGs were mainly involved in neuroactive ligand-receptor interaction. The GO analysis showed that DGRs were

mainly enriched in immune response, Fc receptor-related pathways, etc (Fig. 2B, D). In terms of CC (cellular component), blood mircoparticle and extracellular matrix were enriched with more DGRs (Fig. 2C). We successfully constructed the PPI of DRGs using Cytoscape and demonstrated the top 50 core genes in the network (Fig. 2E).

Table 1	
Screening of DRGs from two groups, including top 50 up-regulated and down-regulated genes	
DRGs	Gene symbol
Up-regulated	SLC38A4 MAGEC2 LINC01234 XAGE1A MKRN3 MAGEA12 IGHG1 IGHG2 GNGT1 LMO3 MAGEA3 PAGE2 AC006206.2 CSAG1 IGLC2 GJA5 ACTG2 SSX2B SSX2 PRAME XAGE1B EPYC MFAP4 IGKC IGHG3 IGKV3-11 AC117402.1 PAGE5 MAGEB1 LINC01980 IGLV2-14 MAGEA6 ALX1 LINC01667 KERA MAGEB2 MAGEA4 DSCR8 MAGEA1 MAGEA10 IGHA1 CXCL13 AL136537.2 MUC15 SSX1 AF279873.3 IGHV4-39 MAGEA11 CSAG3 GRID2
Down-regulated	AL161909.1 LINC01517 OR7E11P GAL KCNJ3 CCDC26 AP000851.2 CGREF1 AL512330.1 CLDN11 AL133410.1 DMRT2 CRYBA2 KCNU1 AC004862.1 CORT CYP2C8 AC004936.1 TERT KIF25 AC010609.1 ELFN1-AS1 FRMD1 CKMT2 LINC02568 LINC00837 COL13A1 GALNT14 TAC4 LGR6 AC005274.1 SLC38A5 GNG4 ANXA13 MB CACNA1E GLB1L2 MTNR1B LPAR3 UNC5B-AS1 AC125616.1 AP003174.1 COLEC10 AC078925.2 ARX SMPDL3B SOST AC233723.1 AC073862.5 MT1A

3.2 Construction and validation of the DRG-Based prognostic signature

After a series of statistical analyses ($P < 0.001$), we obtained 14 hub genes. This led us to the risk score formula. Detail information of 14 hub genes used to construct the signature was located in Table 2. Finally, a fourteen-DRG prognostic signature for osteosarcoma patients was successively constructed. We divided the patients into a high-risk group and a low-risk group using the median risk score as a cut-off (Fig. 3C).

Table 2
Detail information of 14 hub genes used to construct the signature

Gene symbol	Coefficient	LogFC	Average expression	Regulation
APBB1IP	-0.070	1.042	9.110	Up
CAMK4	-0.159	1.164	6.907	Up
CCDC42	-0.236	1.320	3.328	Up
CDK6	-0.232	1.468	10.696	Up
CSAG1	-0.040	3.100	5.031	Up
EPHB6	-0.126	1.452	6.689	Up
GBP1	-0.060	1.722	10.081	Up
GJA5	-0.005	1.968	7.561	Up
MAGEA12	-0.049	3.183	5.304	Up
PPARG	-0.022	1.245	8.520	Up
PTPRG.AS1	-0.117	1.167	5.400	Up
TAC4	0.056	-1.629	4.576	Down
TTC9B	-0.039	1.332	3.668	Up

The survival times of patients with different risk scores are shown in Fig. 3D. The survival rate of patients in the low-risk group was significantly higher than that in the high-risk group ($P < 0.001$) (Fig. 3F).

A nomogram consisting of age, gender and risk score was made to predict the probability of patient survival at 1, 3 and 5 years (Fig. 4A). The calibration plots demonstrated the high accuracy of this nomogram (Fig. 4B - D). The forest plots showed that risk score was associated with patient prognosis in both univariate COX regression and multivariate COX regression ($P < 0.001$) (Fig. 4E, F). Finally, we plotted the time-dependent ROC curve to evaluate the predictive power of the fourteen-DRG signature. As shown, the AUC was 0.887, 0.941 and 0.949 at 1 year, 3 years and 5 years (Fig. 3G). Fig. 4H showed that risk score was a better predictor of prognosis compared to other clinical features.

3.3 Regulatory network of key genes

We obtained two key genes (CSAG1 and MAGEA12) by intersecting the 14 genes that made up the signature with the top 50 core genes in the PPI of the DGRs (Fig. 5A). Two databases were then used to find common target miRNAs for key genes, a key miRNA — miR-193a-3p was obtained (Fig. 5B). Subsequently, we had put together a regulatory network with miRNA's targeting lncRNAs, miR-193a-3p, CSAG1 and MAGEA12 (Fig. 5C). In addition, Fig. 5D showed that miR-193a-3p was mainly involved in KEGG pathways such as Pathways in cancer and MAPK signaling pathway.

3.4 Immune infiltration patterns

We analyzed the proportion of various immune cells in all OS samples and found that three immune cells, Macrophages M0, Macrophages M2, T cells CD4 memory resting, were present in high proportions in most samples (Fig. 6A). In addition, we obtained the degree of correlation between various immune cells (Fig. 6B). Furthermore, by comparing the proportions of various immune cells in the survival group and death group, the proportion of T cells CD8 was the only statistically significant difference (Fig. 6C). We noted that the P value for T cells CD4 naive was 0.07 which was within the borderline significant interval, so we believed that there was a tendency for differences in the proportions of T cells CD4 naive. Finally, the immune cells that we derived from both tests that were associated with both key genes was T cells CD4 naive (Fig. 6D - I). More detailed data were shown in Table 3.

Table 3
Immune cells obtained by correlation test and difference test for two key genes

Gene symbol	Analysis	Cell	P value
CASG1	Difference test	T cells CD4 naive	0.006
	Correlation test	T cells CD4 naive	0.019
		T cells CD8	0.018
		Monocytes	0.048
MAGEA12	Difference test	T cells CD4 naive	0.024
		T cells CD8	0.046
	Correlation test	T cells CD4 naive	0.028
		T cells CD8	0.031
		Monocytes	0.049
		NK cells resting	0.035

3.5 Overall survival analysis of tumor microenvironment scores

We analyzed all OS samples using the estimate package and obtained three types of scores: StromalScore, ImmuneScore and ESTIMATEScore. Patients were then divided into high and low subgroups based on the median of the scores. KM survival curves show that survival rates were significantly higher in the high score group than in the low score group for all three kinds of scores ($P < 0.05$) (Fig. 7A - C). Analysis of clinical characteristics showed that ESTIMATEScore was significantly higher in female patients than in males ($P < 0.05$) (Fig. 7I). However, there was no statistically significant correlation between the remaining clinical characteristics and the scores (Fig. 6D - H).

4. Discussion

Osteosarcoma (OS) is the most common type of bone malignancy, originating from mesenchymal cells and characterized by the production of bone-like tissue or bone tissue by the tumor cells. And osteosarcoma is commonly found in the long bones of the limbs, particularly the epiphysis, and is characterized by strong localized aggressiveness and early haematogenous metastasis. In recent years, significant advances have been made in the diagnosis and treatment of osteosarcoma, including neoadjuvant chemotherapy and improvements in surgical techniques. Although early detection and timely treatment have largely improved the survival rate of the disease, osteosarcoma currently remains a high mortality rate among malignancies in children and adolescents. Therefore, it is necessary to search for markers with high predictive value for patient prognosis during OS development, so that more precise treatment can be implemented. There is no doubt that this has important implications for patients and clinicians alike.

With advances in computer technology and molecular biology, a new interdisciplinary discipline, bioinformatics, has emerged. The analytical tools of bioinformatics can identify those that are meaningful from the vast amount of biological data. The advent of bioinformatics has transformed disease research, with advanced methods and tools used to explore the mechanisms involved in the onset and progression of disease, which also includes various tumors. At the same time, the application of high-throughput sequencing technologies has enabled a comprehensive molecular characterization of tumors, both temporally and spatially. In 2019, Yizhe Xi et al. conducted a study in which they analyzed several hundred circ RNAs that were differentially expressed between osteosarcoma and paraneoplastic tissues and analyzed their potential functions[18]. Kun-Peng Zhu et al.'s study combined bioinformatics analysis and experimentation to construct an RNA regulatory network and explore the underlying mechanisms of OS chemoresistance[19]. In addition to this, there is a large body of literature that has examined OS using bioinformatics and has found valuable information. However, there is no literature that correlates the possible reasons for the poor prognosis of OS patients. Therefore, in the present study, the potential causes of low survival in OS patients were focused on. We grouped patients according to their survival status and conducted a series of studies.

In our study, we identified several hundred genes that were differentially expressed between the survival and death groups and defined these genes as death-related genes (DRGs). Then a functional analysis of these genes was performed. KEGG enrichment analysis revealed a pathway related to intercellular signaling. We therefore hypothesize that the OS cell-to-cell signaling was more active in the dead group, and that the OS tissue was more "active" and more prone to metastasis and invasion, resulting in a lower survival rate.

Our main concern was how to predict the prognosis of patients more accurately. The fourteen-DGRs signature obtained after a series of analyses helped us to solve this problem to some extent, and both the

KM survival curve and the ROC curve showed the high value of the signature. We have analyzed the two key genes obtained and established a regulatory network. Both CSAG1 and MAGEA12 are related to cancer/testis antigen family. There are several studies have been reported on the role of two key genes in a variety of human tumors. MAGEA12 has been shown to act as a prognostic-related gene in gastric cancer[20]. In addition, overexpression of MAGEA12 is involved in the pathogenesis of human cutaneous squamous cell carcinoma and can also promote invasion of breast cancer cells[21, 22]. The study by Chuzhao Lin et al. reported that cancer/testis antigen CSAGE was concurrently expressed with MAGE in chondrosarcoma[23]. In our study, both CSAG1 and MAGEA12 were positively associated with T cells CD8 (Fig. 6E, H). And the percentage of T cells CD8 was higher in the death group than in the survival group ($P < 0.05$). These results suggest that MAGEA12 and CSAG1 may act through some mechanism on T cell CD8 to promote the progression of osteosarcoma, which is the direction of our future research.

In addition, we also noted that of the fourteen DGRs that comprised the signature, only TAC4 expression was relatively low in the death group ($\log_2FC = -1.629$). This gene is a member of the tachykinin family of neurotransmitter-encoding genes. And the products of this gene preferentially activate tachykinin receptor 1, and are thought to regulate peripheral endocrine and paracrine functions including blood pressure, the immune system, and endocrine gland secretion[24–26]. Unfortunately, no studies have reported a relationship between TAC4 and human tumors. We believe that the relationship between TAC4 and tumors is of interest to explore.

The interaction between immune function and tumor has received increasing attention in recent years and a large number of studies have been reported in the literature. Immunotherapy has a long history in osteosarcoma, with inactivated bacteria being used to treat osteosarcoma for over 100 years, but with controversial results[27]. Mifamurtide, as an immune adjuvant, has been at the forefront of osteosarcoma treatment, but there is a lack of large-scale clinical trials to demonstrate its efficacy and safety, so Maya Kansara et al. suggest that the understanding of the relationship between bone tumors and immune function is still in its infancy[1]. Our GO enrichment data show that DRGs are focused on a number of immune-related pathways. Our analysis of the immune cell composition of OS revealed that Macrophages M0, Macrophages M2, and T cells CD4 memory resting were the three most predominant. When comparing the two groups of OS samples, the proportion of T cells CD8 was significantly higher in the death group than in the survival group ($P < 0.05$). There is no doubt that $CD8^+$ T cells are important for their protective immune role against pathogens and tumors. In the case of chronic infections or tumors, the "load" on the immune system is significantly increased. Excessive antigenic and/or inflammatory signals continue to act on $CD8^+$ T cells and this leads to a progressive deterioration of T cell function, T cell depletion in RE and ultimately to hypo- and even loss of immune function. Therefore, we believe that the reason for the low survival rate in OS patients is most likely related to the reduced function of the immune system.

As with the immune system, the tumor microenvironment has been one of the focal points of research in recent years. Osteosarcoma grows in a complex and specific bone microenvironment composed of a wide range of cells. Of these cells that make up the microenvironment, stromal cells and immune cells

have received the most attention. Notably, the "Mr. Tumor" will adapt the local microenvironment to its own heterogeneity in a way that suits it. We therefore scored stromal cells and immune cells separately and derived an estimatescore based on these two scores. As shown in the figure, patients with higher scores had significantly higher overall survival ($P < 0.05$). We believe that the scores we obtained are useful as a guide to clinical decision-making, although the process of performing the scoring is more complex.

However, there are some limitations to this study. Firstly, the number of cases we obtained was not sufficiently large and a larger sample is needed to validate the efficacy of the fourteen-DGR signature. In addition, the mechanism by which key genes affect patient's survival needs to be further explored.

Conclusions

We successfully constructed a fourteen-DRG signature to predict the prognosis of patients. In addition, a nomogram was obtained. Both of them can help clinical decision making and thus implement more precise interventions for patients.

Declarations

Ethics approval and consent to participate:

As this study did not involve human participants, ethical approval and informed consent were not applicable.

Consent for publication:

All authors grant BioMed Central a license to publish the article and identify itself as the original publisher.

Authors' contributions:

Bin Xie wrote the manuscript, analyzed the data.

Shiyong Tan analyzed a portion of the data.

Chao Li plotted the graphs.

Junyang Liang designed and guided the implementation of this study.

Acknowledgements:

Not applicable.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest:

The author(s) declare that they have no competing interests.

Availability of data and material and code availability:

The datasets supporting the conclusions of this article are available in the TCGA repository (<https://portal.gdc.cancer.gov/>).

References

1. Kansara, M., et al., *Translational biology of osteosarcoma*. Nat Rev Cancer, 2014. **14**(11): p. 722-35.
2. Gianferante, D.M., L. Mirabello, and S.A. Savage, *Germline and somatic genetics of osteosarcoma - connecting aetiology, biology and therapy*. Nat Rev Endocrinol, 2017. **13**(8): p. 480-491.
3. Huang, R., et al., *Evaluating and Predicting the Probability of Death in Patients with Non-Metastatic Osteosarcoma: A Population-Based Study*. Med Sci Monit, 2019. **25**: p. 4675-4690.
4. Tian, H., D. Guan, and J. Li, *Identifying osteosarcoma metastasis associated genes by weighted gene co-expression network analysis (WGCNA)*. Medicine (Baltimore), 2018. **97**(24): p. e10781.
5. Chen, Y., et al., *Chemotherapy in head and neck osteosarcoma: Adjuvant chemotherapy improves overall survival*. Oral Oncol, 2017. **73**: p. 124-131.
6. Harting, M.T., et al., *Age as a prognostic factor for patients with osteosarcoma: an analysis of 438 patients*. J Cancer Res Clin Oncol, 2010. **136**(4): p. 561-70.
7. Iwata, S., et al., *Prognostic factors in elderly osteosarcoma patients: a multi-institutional retrospective study of 86 cases*. Ann Surg Oncol, 2014. **21**(1): p. 263-8.
8. Zheng, D., et al., *A Novel Six Metastasis-Related Prognostic Gene Signature for Patients With Osteosarcoma*. Front Cell Dev Biol, 2021. **9**: p. 699212.
9. Zhu, S.T., Wang, X., Wang, J. Y., Xi, G. H., and Liu, Y, *Downregulation of miR-22 contributes to epithelial-mesenchymal transition in osteosarcoma by targeting Twist1*. Front. Oncol, 2020. **10**:406. doi: 10.3389/fonc.2020.00406.
10. Wu, W., Jing, D., Meng, Z., Hu, B., Zhong, B., Deng, X., et al., *FGD1 promotes tumor progression and regulates tumor immune response in osteosarcoma via inhibiting PTEN activity*. . Theranostics, 2020. **10**, 2859–2871.
11. Fan, H., et al., *Identification of critical genes associated with human osteosarcoma metastasis based on integrated gene expression profiling*. Mol Med Rep, 2019. **20**(2): p. 915-930.
12. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Research, 2015.

13. Guangchuang Yu, L.-G.W., Yanyan Han and Qing-Yu He, *clusterProfiler: an R package for comparing biological themes among gene clusters*. A Journal of Integrative Biology, 2012. **16(5):284-287**.
14. Jerome Friedman, T.H., Robert Tibshirani, *Regularization Paths for Generalized Linear Models via Coordinate Descent*. Journal of Statistical Software, 2010. **33(1), 1-22**. URL <https://www.jstatsoft.org/v33/i01/>.
15. Patrick J. Heagerty, S.-C., *survivalROC: Time-dependent ROC curve estimation from censored survival data*. <https://CRAN.R-project.org/package=survivalROC>, 2013.
16. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
17. Kosuke Yoshihara, H.K.a.R.G.V., *estimate: Estimate of Stromal and Immune Cells in Malignant Tumor Tissues from Expression Data*. <https://R-Forge.R-project.org/projects/estimate/>, 2016.
18. Xi Y, F.M., Liu Y, Wu H, He M, Zhao J., *Differential expression and bioinformatics analysis of circRNA in osteosarcoma*. Biosci Rep, 2019. **May 17;39(5):BSR20181514**. doi: **10.1042/BSR20181514**. PMID: **31036604**; PMCID: **PMC6522716**.
19. Zhu KP, Z.C., Ma XL, Hu JP, Cai T, Zhang L, *Analyzing the Interactions of mRNAs and ncRNAs to Predict Competing Endogenous RNA Networks in Osteosarcoma Chemo-Resistance*. Mol Ther, 2019. **Mar 6;27(3):518-530**. doi: **10.1016/j.ymthe.2019.01.001**. Epub 2019 Jan 7. PMID: **30692017**; PMCID: **PMC6401193**.
20. Wu J, W.J., Shen W. Identification of MAGEA12 as a prognostic outlier gene in gastric cancers. Neoplasma. 2017;64(2):238-243. doi: 10.4149/neo_2017_210. PMID: 28043151.
21. Zhao G, B.J., Zheng Z, Park HS, Chung KY, Roh MR, Jin Z. Overexpression and Implications of Melanoma-associated Antigen A12 in Pathogenesis of Human Cutaneous Squamous Cell Carcinoma. Anticancer Res. 2019 Apr;39(4):1849-1857. doi: 10.21873/anticancer.13292. PMID: 30952725.
22. Oh C, K.H., Oh S, Ko JY, Kim Y, Kang K, Yang Y, Kim J, Park JH, Roe JS, Yoo KH. Epigenetic Upregulation of MAGE-A Isoforms Promotes Breast Cancer Cell Aggressiveness. Cancers (Basel). 2021 Jun 25;13(13):3176. doi: 10.3390/cancers13133176. PMID: 34202157; PMCID: PMC8268034.
23. Lin C, M.S., Meitner PA, Wolf JM, Bluman EM, Block JA, Terek RM. Cancer/testis antigen CSAGE is concurrently expressed with MAGE in chondrosarcoma. Gene. 2002 Feb 20;285(1-2):269-78. doi: 10.1016/s0378-1119(02)00395-5. PMID: 12039054.
24. Cunningham S, O.D.C., Patterson C, McDonnell G, Hawkins S, Marrosu MG, Vandenbroeck K. The neuropeptide genes TAC1, TAC3, TAC4, VIP and PACAP(ADCYAP1), and susceptibility to multiple sclerosis. J Neuroimmunol. 2007 Feb;183(1-2):208-13. doi: 10.1016/j.jneuroim.2006.11.002. Epub 2006 Dec 18. PMID: 17175032.
25. Ravina CG, S.M., Pinto FM, Orea A, Fernández-Sánchez M, Pintado CO, Cadenas ML. A role for tachykinins in the regulation of human sperm motility. Hum Reprod. 2007 Jun;22(6):1617-25. doi: 10.1093/humrep/dem069. Epub 2007 Apr 16. PMID: 17437961.
26. Kurtz MM, W.R., Clements MK, Cascieri MA, Austin CP, Cunningham BR, Chicchi GG, Liu Q. Identification, localization and receptor characterization of novel mammalian substance P-like

peptides. Gene. 2002 Aug 21;296(1-2):205-12. doi: 10.1016/s0378-1119(02)00861-2. PMID: 12383518.

27. Il, C.W., *Contribution to the Knowledge of Sarcoma.* . Ann Surg., 1891. Sep;14(3):199-220. doi: 10.1097/00000658-189112000-00015. PMID: 17859590; PMCID: PMC1428624.

Figures

Figure 1

Screening of DRGs from two groups. The volcano plot of DRGs (A). The heatmap showed the expression profile of DRGs (B).

Figure 2

Functional enrichment analysis of DRGs. KEGG (Kyoto Encyclopaedia of Genes and Genomes) bubble plot showed the enriched pathways of DRGs (A). The results of GO (GeneOntology) pathway including BP (Biological Process), CC (Cellular Component) and MF (Molecular Function) of DRGs (B - D). The protein-protein interaction (PPI) network constructed from the top 50 hub genes obtained by CytoHubba, and from yellow to red indicating progressively higher MCC scores (E).

Figure 3

Construction of the signature The lasso regression analysis identified several genes associated with prognosis (A). The optimal values of the penalty parameter were defined by 1,000-round cross-validation (B). The risk score distribution of patients (C). The survival status of osteosarcoma patients (D). The expression profiles of the 14 DRGs (E) . Kaplan-Meier survival analysis of patients (F).

Figure 4

Construction and validation of the nomogram. The nomogram consisted of age, gender, and risk score based on the DRGs signature (A). The calibration curve for validation of the nomogram for estimating the survival of osteosarcoma patients at 1, 3 and 5 years (B - D). Forest plot showed risk score was an independent prognostic factor, with green indicated univariate COX regression and red indicated multivariate COX regression (E - F). ROC curves demonstrated the predictive prognostic value of risk score (G - H).

Figure 5

Construction of regulatory network of two key genes Two key genes obtained by the intersection of the top 50 hub genes from Cytohubba and 14 genes from the prediction signature (A). Intersection of target miRNAs for two key genes identified by Targetscan and miRcode (B). Regulatory network consisting of two key genes, miRNA and miRNA-targeted lncRNA (C). Top 20 KEGG pathways obtained from Starbase for miR-193a-3p (D).

Figure 6

Immuno-infiltration analysis Barplot showed the proportion of 22 kinds of immune cells in osteosarcoma samples and column names of plot were sample ID (A). Violin plot showed the ratio differentiation of 19 kinds of immune cells between survival group and death group (B). The correlation coefficient between immune cells (C). Immune cells associated with CSAG1 expression derived from difference test and correlation test (D - F). Immune cells associated with MAGEA12 expression obtained from two kinds of analysis methods (G - I).

Figure 7

Overall survival analysis of tumour microenvironment scores. The KM survival curves plots showed the difference in survival rates between the two groups (A - C). The correlation between clinical characteristics and tumour microenvironment scores, and Wilcoxon rank sum was used for the significance test (D - I).