



# Large Scale Identification of Osteosarcoma Pathogenic Genes by Multiple Extreme Learning Machine

Zhipeng Zhao<sup>1†</sup>, Jijun Shi<sup>2†</sup>, Guang Zhao<sup>3</sup>, Yanjun Gao<sup>3</sup>, Zhigang Jiang<sup>4</sup> and Fusheng Yuan<sup>3\*</sup>

<sup>1</sup> Department of Basic Medical Sciences, Taizhou University, Taizhou, China, <sup>2</sup> Department of Orthopedics, Songyuan Central Hospital, Songyuan, China, <sup>3</sup> Department of Orthopedics, The Fourth Affiliated Hospital of China Medical University, Shenyang, China, <sup>4</sup> Department of Hand Surgery, Changchun Central Hospital, Changchun, China

## OPEN ACCESS

### Edited by:

Lei Deng,  
Central South University, China

### Reviewed by:

Ningyi Zhang,  
Harbin Institute of Technology, China  
Hong Ju,  
Heilongjiang Vocational College  
of Biology Science and Technology,  
China

### \*Correspondence:

Fusheng Yuan  
yuanfs07@mails.jlu.edu.cn

<sup>†</sup> These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular and Cellular Pathology,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

**Received:** 09 August 2021

**Accepted:** 02 September 2021

**Published:** 27 September 2021

### Citation:

Zhao Z, Shi J, Zhao G, Gao Y,  
Jiang Z and Yuan F (2021) Large  
Scale Identification of Osteosarcoma  
Pathogenic Genes by Multiple  
Extreme Learning Machine.  
*Front. Cell Dev. Biol.* 9:755511.  
doi: 10.3389/fcell.2021.755511

At present, the main treatment methods of osteosarcoma are chemotherapy and surgery. Its 5-year survival rate has not been significantly improved in the past decades. Osteosarcoma has extremely complex multigenomic heterogeneity and lacks universally applicable signal blocking targets. Osteosarcoma is often found in adolescents or children under the age of 20, so it is very important to explore its genetic pathogenic factors. We used known osteosarcoma-related genes and computer algorithms to find more osteosarcoma pathogenic genes, laying the foundation for the treatment of osteosarcoma immune microenvironment-related treatments, so as to carry out further explorations on these genes. It is a traditional method to identify osteosarcoma related genes by collecting clinical samples, measuring gene expressions by RNA-seq technology and comparing differentially expressed gene. The high cost and time consumption make it difficult to carry out research on a large scale. In this paper, we developed a novel method “RELM” which fuses multiple extreme learning machines (ELM) to identify osteosarcoma pathogenic genes. The AUC and AUPR of RELM are 0.91 and 0.88, respectively, in 10-cross validation, which illustrates the reliability of RELM.

**Keywords:** Index Term-osteosarcoma, pathogenic genes, fuses multiple extreme learning machine, machine learning, large scale identification

## INTRODUCTION

Osteosarcoma is the most common malignant bone tumor in clinic (Marko et al., 2016), which is mostly seen in children and adolescents. Although surgery combined with neoadjuvant chemotherapy significantly improves the 5-year survival rate of patients with local tumors (Yang et al., 2018), most patients with osteosarcoma will metastasize, and the 5-year survival rate of patients with metastatic osteosarcoma is only 20 ~ 30% (Murakami et al., 2017). At present, osteosarcoma is still the second leading cause of cancer-related death in adolescents (Chen et al., 2020). Considering the complex intra- and inter tumor heterogeneity, a suitable specific target for osteosarcoma has not been found. However, based on previous studies on the heterogeneity of other tumors, the immune microenvironment may have relatively low heterogeneity and become a more appropriate direction of intervention (Koirala et al., 2016). Therefore, identifying genes

related to osteosarcoma immune microenvironment may provide a robust and effective target for clinical application (Mirabello et al., 2020).

According to the age at which osteosarcoma occurs suddenly increases with the onset of puberty, and its largest growth site is shown to be related to the rapid proliferation of bones, it indicates that osteosarcoma is significantly related to the rapid growth of bones (Ho et al., 2017). At the same time, exposure to alkylating agents may also promote the development of osteosarcoma (Zhang et al., 2021). In addition, radiotherapy is one of the few identified environmental risk factors for osteosarcoma. Studies have shown that increasing the radiation dose of primary cancer is linearly related to the risk of secondary osteosarcoma. Another study based on American adults also found that radiotherapy is significantly associated with an increased risk of osteosarcoma diagnosis in the future (Wu et al., 2012).

Whole-exome and whole-genome sequencing analysis of the germline DNA of patients with osteosarcoma showed that the prevalence of pathogenic variants in genes associated with known cancer susceptibility syndromes was higher than expected (Gianferante et al., 2017). Chromosomal abnormalities, pathogenic variants of tumor suppressor genes, transcription factors and growth factors, and abnormalities of WWOX and miRNA all play an important role in the occurrence and development of osteosarcoma (Lin et al., 2017). The frequency of osteosarcoma in individuals with mutations in the RB1 gene is higher than that in the population. Studies have shown that there is an interaction between primary inheritance and genes in the pathogenesis of the disease (Spritz, 2007). A 2016 study found that among individuals with pathogenic mutations in the germline tumor suppressor gene TP53, the cumulative incidence of osteosarcoma reached 5–11% (Mai et al., 2016). Transforming growth factor  $\beta$  (TGF- $\beta$ ) protein affects cell growth and metabolism, and the expression of TGF- $\beta$ 1 is significantly increased in highly malignant osteosarcoma. Insulin growth factors IGF-I and IGF-II can bind to the corresponding receptors to play a role, and they are overexpressed in osteosarcoma. The overexpression of CCN3 in osteosarcoma is related to its poor prognosis. Parathyroid hormone (PTH), parathyroid hormone related peptide (PTHrP) and parathyroid hormone receptor (PTHr1) have been shown to be related to the progression and metastasis of osteosarcoma (Berdiaki et al., 2010). Various molecular changes and genomes closely related to the occurrence and progress of osteosarcoma have been identified. These changes include gene amplification, deletion and germline mutation, overexpression and RTK activation, abnormal cell proliferation, metastasis, apoptosis, drug tolerance genes and miRNAs (Saraf et al., 2018). Osteosarcoma is characterized by complex and unbalanced karyotypes and abnormal gene expression profiles. Abnormalities of chromosome structure and value can be detected in most osteosarcoma (Isakoff et al., 2015). Common chromosome numerical abnormalities include germline mutation, deletion, polyploidy, aneuploidy, duplication and unbalanced ectopic errors (Morrow and Khanna, 2015). TP53 tumor gene and retinoblastoma tumor suppressor gene RB1 are the most prominent genes of germline mutation (Oliveira et al., 2005). They are the key detection sites of

mitosis and the root cause of chromosome instability. Most osteosarcoma contains inactivation of both p53 and Rb pathways (Levine and Fleischli, 2000). In essence, the main causes of osteosarcoma are the inactivation of tumor suppressor gene expression and the abnormal doubling of oncogenes (Orr and Compton, 2013). Common oncogenes, such as avian cell homolog Myc, purine / pyrimidine exonuclease 1 (APEX1), action associated vascular endothelial growth factor A (VEGFA) and RecQ protein analog 4 (RecQL4). These amplified genes are closely related to the biological processes of osteosarcoma cell proliferation, growth and angiogenesis. Liu et al. (2019) identified 125 genes which are related to osteosarcoma and can be used to predict survival of osteosarcoma. Deng et al. (2021) used univariate, Lasso, and machine learning algorithm-iterative Lasso Cox regression analyses to predict survival of osteosarcoma by lncRNAs.

At present, there are two common biological methods for discovering disease-related genes. First, collect disease samples and health samples, respectively, conduct RNA-seq sequencing to obtain the expression of genes in different health states, and then obtain the genes significantly differentially expressed in disease and health populations through differential expression analysis (Zhao et al., 2021b). Second, through genome-wide association analysis, collect a large number of disease and healthy people, sequence the whole genome, and then compare the sequences to obtain sites with significant differences in mutation frequency (Peng and Zhao, 2020; Zhao et al., 2020c). However, both of them need a large number of samples to support in order to ensure the accuracy, which results in a large consumption of time and money (Bhakta and Tsukahara, 2020). With the continuous accumulation of biological data and the continuous improvement of calculation methods, bioinformatics experts find biological laws through calculation methods, and then infer more biological conclusions (Chen et al., 2019; Liu et al., 2020; Zhao et al., 2020a). The calculation methods have identified disease-related genes and drugs on a large scale (Tianyi et al., 2020; Zhao et al., 2020b). Although some conclusions are not completely accurate, they greatly reduce the scope of research and save time and money (Wu et al., 2021). Moreover, the models constructed by deep learning and machine learning can be used for reference by other research problems (Zhao et al., 2021a). Therefore, we developed a machine learning method to identify osteosarcoma-related genes in this paper. Using the idea of random forest for reference, we fused multiple Extreme Learning Machines (ELM) to build a model through the known osteosarcoma related genes to predict more genes potentially associated with osteosarcoma.

## MATERIALS AND METHODS

### Workflow

Firstly, we obtained 2,339 genes which are reported to be related to osteosarcoma in DisGeNET (Piñero et al., 2020). Then, we constructed gene interaction network based on these genes. More genes are included in this network since many genes can interact with these 2,339 genes. We extracted the features of this network by random walk and used Random Extreme Learning Machine

(RELM) to identify osteosarcoma-related genes. The way of constructing RELM is to build multiple ELM models and the output of each model is attached with weight, and the final result is obtained by voting. The whole workflow is shown in **Figure 1**.

### Extreme Learning Machine

The calculation process of single hidden layer neural network is as follows:

1. The input value is multiplied by the weight value
2. Add bias value
3. Calculation of activation function
4. Repeat steps 1 to 3 for each layer
5. Calculate output value
6. Error back propagation
7. Repeat steps 1 to 6.

Extreme learning machines improves it by removing step 4 and replacing step 6 with a primary matrix inverse operation and removing step 7.

The process of ELM is to construct the formula (1):

$$f_L(x) = \sum_{i=1}^L \beta_i g_i(x) = \sum_{i=1}^L \beta_i g(w_i * x_j + b_j), j = 1, \dots, N \quad (1)$$

L is the number of hidden units. N is the number of training samples.  $\beta_i$  is the weight between *i*th hidden layer and output.  $w_i$  is the weight between input and output.  $g(x)$  is activation function. b is bias and x is the input. Since ELM only has one hidden layer, i is 1 in our model.

The calculation process of the extreme learning machine is very similar to the standard back-propagation neural network, but the weight matrix between the hidden layer and the output is a pseudo-inverse matrix. The above formula can be abbreviated as:

$$T = H\beta \quad (2)$$

$$H = \begin{bmatrix} g(w_1 * x_1 + b_1) & \dots & g(w_L * x_L + b_L) \\ \vdots & & \vdots \\ g(w_1 * x_N + b_1) & \dots & g(w_L * x_N + b_L) \end{bmatrix}_{N \times L} \quad (3)$$

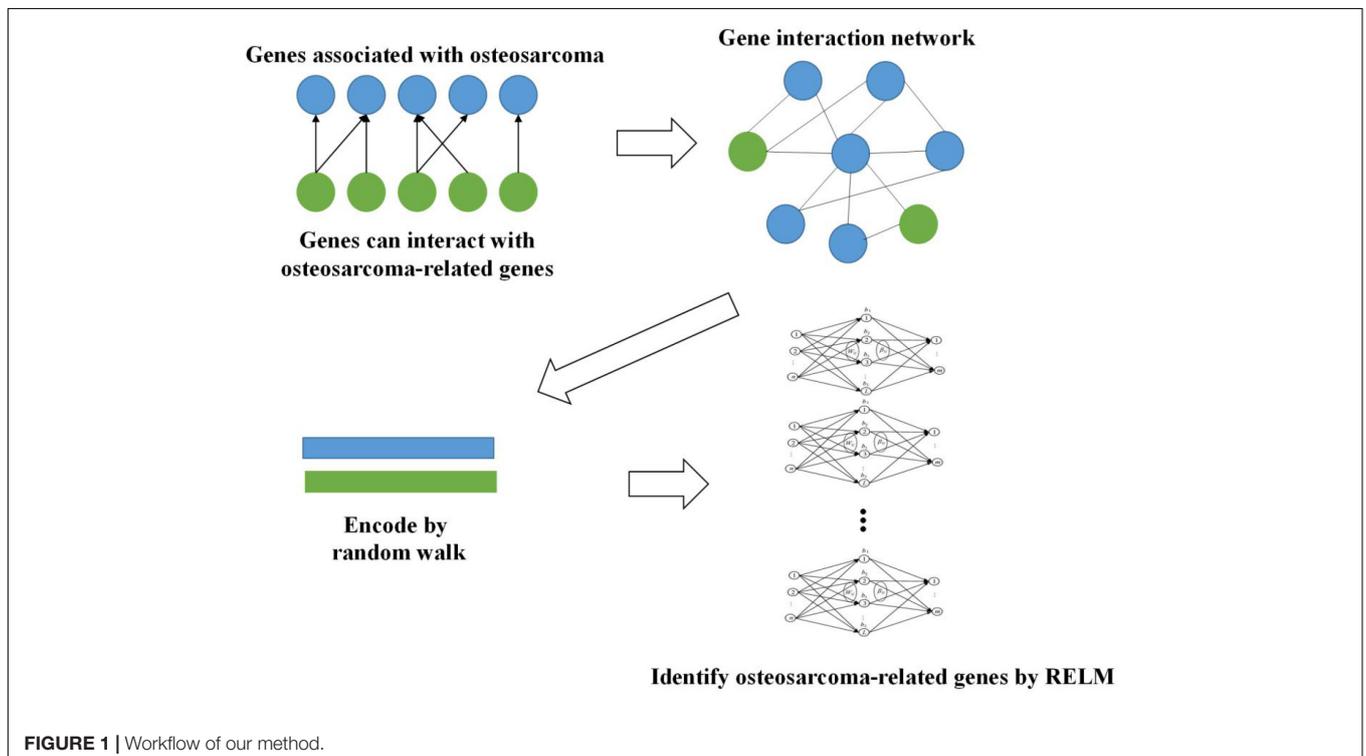
m is the number of outputs; H is the hidden layer output matrix; T is the target matrix of the training set.

### Random Extreme Learning Machine (RELM)

Extreme learning machines is a special artificial neural network with only one hidden layer, which causes its accuracy to be low. However, the calculation speed of ELM is extremely fast. Therefore, we can use this advantage to build multiple ELM models and use weighted voting to improve accuracy.

Random extreme learning machine draws on the idea of random forest (RF), regards ELM as a simple decision tree, and trains multiple ELMs to form an ELM forest to achieve the goal of improving accuracy.

The idea of RELM is to randomly extract the multi-dimensional features of genes, and then randomly extract the training set to form a simple ELM. Through repeated extraction with replacement, new ELMs are continuously trained. After



**FIGURE 1** | Workflow of our method.

getting enough ELM models, the final result is obtained by weighting and averaging the output results of the 500 models.

The number of features for each ELM model is selected as (Zhao et al., 2017):

$$n = \sqrt{N} \tag{4}$$

N is the whole dimension of features. n is the number of features for each ELM model.

In the meanwhile, we randomly selected samples for each ELM model too. After each modeling, we will also put the sample back. We choose one-tenth of the samples for modeling each time.

## RESULTS

### Selection of Extreme Learning Machine Model Number

We should construct multiple ELM models to obtain RELM, but the number of ELM models is not sure. Therefore, we tried 10, 20, 50, 100, 200 ELM models and used 10-cross validation to obtain the final number.

The AUC curves of 10, 20, 50, 100, 200 ELM models are shown in **Figure 2**. The AUC values of these models are 0.66, 0.72, 0.82, 0.92, 0.92, respectively. The AUC of 100 models and 200 models are similar.

The PR curves of 10, 20, 50, 100, 200 ELM models are shown in **Figure 3**. The AUPR values of these models are 0.46, 0.54, 0.72, 0.88, 0.88, respectively. The AUPR of 100 models and 200 models are similar too. Therefore, we chose 100 ELM models to construct RELM.

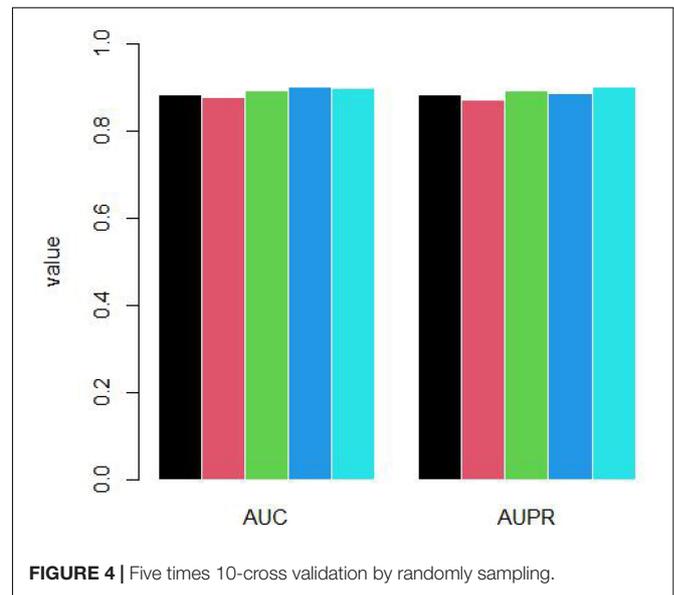
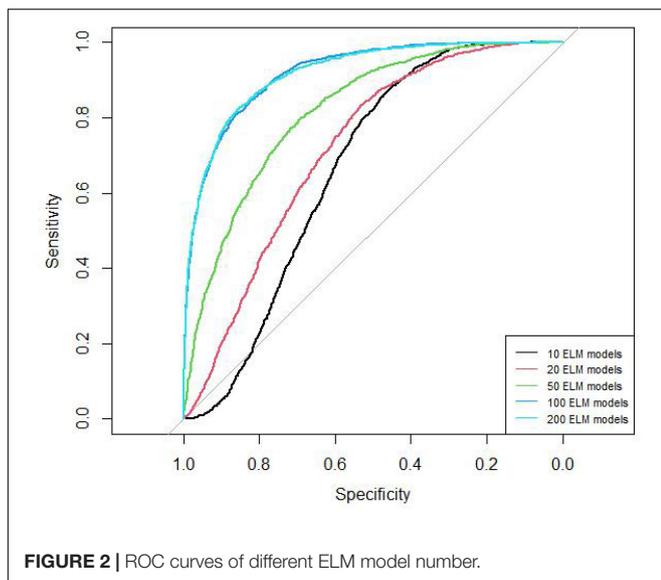
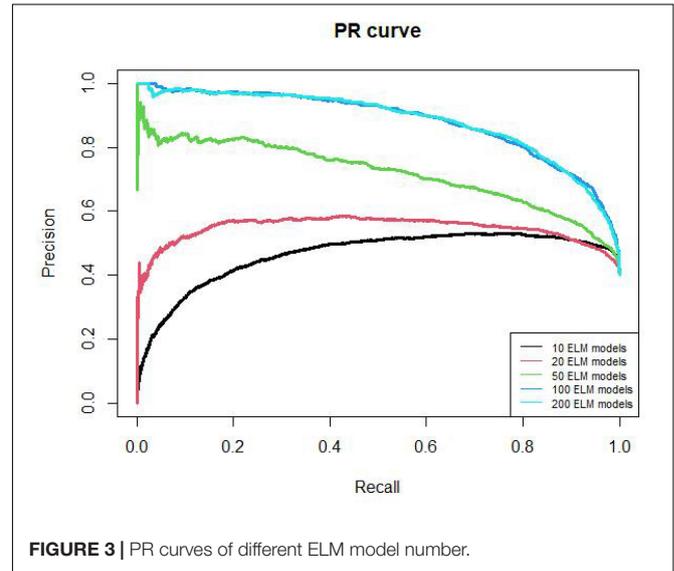
### Performance of Random Extreme Learning Machine

Because the unknown genes are far more than known osteosarcoma-related genes, we randomly selected negative samples to build RELM model. For each time, the number of

negative samples is as same as positive samples. We repeated to select negative samples 5 times and did 10-cross validation for each time. The AUC and AUPR is shown as **Figure 4**.

The mean AUC is 0.889 and standard deviation is 0.009. The mean AUPR is 0.887 and standard deviation is 0.011.

In order to further explore the advantages of RELM, we compared RELM with ELM, RSVM, RANN. RSVM is to replace



**TABLE 1 |** Comparison result.

Method	AUC	AUPR
RELM	0.889	0.887
ELM	0.761	0.684
RSVM	0.865	0.841
RANN	0.876	0.849

the ELM of RELM by SVM and RANN is to replace ELM of RELM with ANN. The experiments results are shown in **Table 1**.

As we can see in **Table 1**, RELM performed best among these method. SVM is more suitable for small sample modeling and ANN needs large sample set to build a precise model. Therefore, these two methods are not suitable for our case.

## CONCLUSION

Whole-exome and whole-genome sequencing analysis of the germline DNA of patients with osteosarcoma showed that the prevalence of pathogenic variants in genes associated with known cancer susceptibility syndromes was higher than expected. Osteosarcoma is highly aggressive and progresses rapidly. In all age groups, as many as 25% of patients have metastasized at the time of diagnosis, so its early diagnosis is necessary for the long-term prognosis of patients. At present, the diagnosis of osteosarcoma is still based on the patient's clinical manifestations, imaging examinations and biopsy. Gene therapy includes tumor suppressor gene therapy, antisense gene therapy, suicide gene therapy and combined gene therapy. Although the research of gene therapy has made great progress and it has good therapeutic prospects, the clinical application of gene therapy still has a long way to go. In recent years, with continuous research on the key genes of osteosarcoma, its application value as a gene therapy target has gradually revealed.

To identify osteosarcoma-related genes in large scale, in this paper, we developed an ELM-based method for identifying osteosarcoma-related genes. 100 ELM models have been constructed to build a final RELM model. By constantly randomly selecting negative sets, we performed five times of 10-cross validation. The accuracy of RELM is stable and high in all experiments.

## REFERENCES

- Berdiaki, A., Datsis, G. A., Nikitovic, D., Tsatsakis, A., Katonis, P., Karamanos, N. K., et al. (2010). Parathyroid hormone (PTH) peptides through the regulation of hyaluronan metabolism affect osteosarcoma cell migration. *IUBMB Life* 62, 377–386. doi: 10.1002/iub.320
- Bhakta, S., and Tsukahara, T. (2020). Artificial RNA editing with ADAR for gene therapy. *Curr. Gene Ther.* 20, 44–54. doi: 10.2174/1566523220666200516170137
- Chen, L., Wang, M., Lin, Z., Yao, M., Wang, W., Cheng, S., et al. (2020). Mild microwave ablation combined with HSP90 and TGF- $\beta$ 1 inhibitors enhances the therapeutic effect on osteosarcoma. *Mol. Med. Rep.* 22, 906–914. doi: 10.3892/mmr.2020.11173
- Chen, X. G., Shi, W. W., and Deng, L. (2019). Prediction of disease comorbidity using hetsim scores based on multiple heterogeneous networks. *Curr. Gene Ther.* 19, 232–241. doi: 10.2174/1566523219666190917155959
- Deng, Y., Yuan, W., Ren, E., Wu, Z., Zhang, G., and Xie, Q. (2021). A four-methylated LncRNA signature predicts survival of osteosarcoma patients based on machine learning. *Genomics* 113, 785–794. doi: 10.1016/j.ygeno.2020.10.010
- Gianferante, D. M., Mirabello, L., and Savage, S. A. (2017). Germline and somatic genetics of osteosarcoma—connecting aetiology, biology and therapy. *Nat. Rev. Endocrinol.* 13, 480–491. doi: 10.1038/nrendo.2017.16
- Ho, X. D., Phung, P., Le, V. Q., Nguyen, V. H., Reimann, E., Prans, E., et al. (2017). Whole transcriptome analysis identifies differentially regulated

Overall, we purposed a reliable method for identifying osteosarcoma-related genes in large-scale. This method could help understand the pathogenesis of osteosarcoma and develop drug targets.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: gene-disease associations: <https://www.disgenet.org/> and gene interaction: <http://www.inetbio.org/humannet>.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

ZZ, JS, and FY conceived and designed this study. ZZ, JS, GZ, YG, and ZJ analyzed the data. ZZ, JS, and GZ wrote the manuscript. All authors read and approved the final version of the manuscript.

## FUNDING

This work was supported by the Natural Science Foundation of China (NSFC) (No. 82060458).

- networks between osteosarcoma and normal bone samples. *Exp. Biol. Med.* 242, 1802–1811. doi: 10.1177/1535370217736512
- Isakoff, M. S., Bielack, S. S., Meltzer, P., and Gorlick, R. (2015). Osteosarcoma: current treatment and a collaborative pathway to success. *J. Clin. Oncol.* 33, 3029–3035. doi: 10.1200/jco.2014.59.4895
- Koirala, P., Roth, M. E., Gill, J., Piperdi, S., Chinai, J. M., Geller, D. S., et al. (2016). Immune infiltration and PD-L1 expression in the tumor microenvironment are prognostic in osteosarcoma. *Sci. Rep.* 6:30093.
- Levine, R., and Fleischli, M. (2000). Inactivation of p53 and retinoblastoma family pathways in canine osteosarcoma cell lines. *Vet. Pathol.* 37, 54–61. doi: 10.1354/vp.37-1-54
- Lin, W., Zhu, X., Yang, S., Chen, X., Wang, L., Huang, Z., et al. (2017). MicroRNA-203 inhibits proliferation and invasion, and promotes apoptosis of osteosarcoma cells by targeting Runt-related transcription factor 2. *Biomed. Pharmacother.* 91, 1075–1084. doi: 10.1016/j.biopha.2017.05.034
- Liu, F., Xing, L., Zhang, X., and Zhang, X. (2019). A four-pseudogene classifier identified by machine learning serves as a novel prognostic marker for survival of osteosarcoma. *Genes* 10:414. doi: 10.3390/genes10060414
- Liu, Y. B., Zhang, X., and Yang, L. (2020). Genetic engineering of AAV capsid gene for gene therapy application. *Curr. Gene Ther.* 20, 321–332. doi: 10.2174/1566523220666200930105521
- Mai, P. L., Best, A. F., Peters, J. A., DeCastro, R. M., Khincha, P. P., Loud, J. T., et al. (2016). Risks of first and subsequent cancers among TP53 mutation carriers in the National cancer institute Li-Fraumeni syndrome cohort. *Cancer* 122, 3673–3681. doi: 10.1002/cncr.30248

- Marko, T. A., Diessner, B. J., and Spector, L. G. (2016). Prevalence of metastasis at diagnosis of osteosarcoma: an international comparison. *Pediatr. Blood Cancer* 63, 1006–1011. doi: 10.1002/pbc.25963
- Mirabello, L., Zhu, B., Koster, R., Karlins, E., Dean, M., Yeager, M., et al. (2020). Frequency of pathogenic germline variants in cancer-susceptibility genes in patients with osteosarcoma. *JAMA Oncol.* 6, 724–734. doi: 10.1001/jamaoncol.2020.0197
- Morrow, J. J., and Khanna, C. (2015). Osteosarcoma genetics and epigenetics: emerging biology and candidate therapies. *Crit. Rev. Oncogen.* 20, 173–197. doi: 10.1615/critrevoncog.2015013713
- Murakami, T., Igarashi, K., Kawaguchi, K., Kiyuna, T., Zhang, Y., Zhao, M., et al. (2017). Tumor-targeting *Salmonella typhimurium* A1-R regresses an osteosarcoma in a patient-derived xenograft model resistant to a molecular-targeting drug. *Oncotarget* 8, 8035–8042. doi: 10.18632/oncotarget.14040
- Oliveira, A. M., Ross, J. S., and Fletcher, J. A. (2005). Tumor suppressor genes in breast cancer: the gatekeepers and the caretakers. *Pathol. Patterns Rev.* 124(Suppl. 1), S16–S28. doi: 10.1309/5XW3L8LU445QWGQR
- Orr, B., and Compton, D. A. (2013). A double-edged sword: how oncogenes and tumor suppressor genes can contribute to chromosomal instability. *Front. Oncol.* 3:164. doi: 10.3389/fonc.2013.00164
- Peng, J., and Zhao, T. (2020). Reduction in TOM1 expression exacerbates Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* 117, 3915–3916. doi: 10.1073/pnas.1917589117
- Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855. doi: 10.1093/nar/gkz1021
- Saraf, A. J., Feng, J. M., and Roberts, R. D. (2018). Osteosarcoma: accelerating progress makes for a hopeful future. *Front. Oncol.* 8:4. doi: 10.3389/fonc.2018.00004
- Spritz, R. A. (2007). The genetics of generalized vitiligo and associated autoimmune diseases. *Pigment Cell Res.* 20, 271–278. doi: 10.1111/j.1600-0749.2007.00384.x
- Tianyi, Z., Yang, H., Valsdottir, L. R., Tianyi, Z., and Jijie, P. (2020). Identifying drug–target interactions based on graph convolutional network and deep neural network. *Brief. Bioinform.* 22:bbaa044. doi: 10.1093/bib/bbaa044
- Wu, L. C., Kleinerman, R. A., Curtis, R. E., Savage, S. A., and de González, A. B. (2012). Patterns of bone sarcomas as a second malignancy in relation to radiotherapy in adulthood and histologic type. *Cancer Epidemiol. Prevent. Biomark.* 21, 1993–1999. doi: 10.1158/1055-9965.epi-12-0810
- Wu, Q., Nasoz, F., Jung, J., Bhattarai, B., Han, M. V., Greenes, R. A., et al. (2021). Machine learning approaches for the prediction of bone mineral density by using genomic and phenotypic data of 5130 older men. *Sci. Rep.* 11:4482. doi: 10.1038/s41598-021-83828-3
- Yang, Y., Han, L., He, Z., Li, X., Yang, S., Yang, J., et al. (2018). Advances in limb salvage treatment of osteosarcoma. *J. Bone Oncol.* 10, 36–40.
- Zhang, Z., Wu, X., Han, Q., and Huang, Z. (2021). Downregulation of long non-coding RNA UCA1 represses tumorigenesis and metastasis of osteosarcoma via miR-513b-5p/E2F5 axis. *Anti Cancer Drugs* 32, 602–613. doi: 10.1097/cad.0000000000001034
- Zhao, T., Hu, Y., and Cheng, L. (2020a). Deep-DRM: a computational method for identifying disease-related metabolites based on graph deep learning approaches. *Brief. Bioinform.* 22:bbaa212. doi: 10.1093/bib/bbaa212
- Zhao, T., Hu, Y., Peng, J., and Cheng, L. (2020b). DeepLGP: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics* 36, 4466–4472. doi: 10.1093/bioinformatics/btaa428
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2020c). MRTFB regulates the expression of NOMO1 in colon. *Proc. Natl. Acad. Sci. U.S.A.* 117, 7568–7569. doi: 10.1073/pnas.2000499117
- Zhao, T., Liu, J., Zeng, X., Wang, W., Li, S., Zang, T., et al. (2021a). Prediction and collection of protein–metabolite interactions. *Brief. Bioinform.* 22:bbab014. doi: 10.1093/bib/bbab014
- Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., Wei, Z., et al. (2021b). SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res.* 49, D1413–D1419. doi: 10.1093/nar/gkaa838
- Zhao, T., Zhang, N., Zhang, Y., Ren, J., Xu, P., Liu, Z., et al. (2017). A novel method to identify pre-microRNA in various species knowledge base on various species. *J. Biomed. Semant.* 8(Suppl. 1):30. doi: 10.1186/s13326-017-0143-z

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhao, Shi, Zhao, Gao, Jiang and Yuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.