Pathology & Oncology Research (2020) 26:1041–1048
https://doi.org/10.1007/s12253-019-00645-w

ORIGINAL ARTICLE

Identification of Pathogenic Genes and Transcription Factors in Osteosarcoma

Chenggang Yang[1,2] · Di Huang[1,2] · Cui Ma[1,2] · Jing Ren[2] · Lina Fu[2] · Cheng Cheng[2] · Bangling Li[3] · Xiaofeng Shi[1,2]

Received: 17 December 2018 / Accepted: 19 March 2019 / Published online: 13 April 2019

Abstract

Osteosarcoma (OS) is an aggressive malignant tumor of the bones. Our study intended to identify and analyze potential pathogenic genes and upstream regulators for OS. We performed an integrated analysis to identify candidate pathogenic genes of OS by using three Gene Expression Omnibus (GEO) databases (GSE66673, GSE49003 and GSE37552). GO and KEGG enrichment analysis were utilized to predict the functional annotation and potential pathways of differentially expressed genes (DEGs). The OS-specific transcriptional regulatory network was established to study the crucial transcriptional factors (TFs) which target the DEGs in OS. From the three GEO datasets, we identified 759 DEGs between metastasis OS samples and non-metastasis OS samples. After GO and KEGG analysis, 'cell adhesion' (FDR = 1.27E-08), 'protein binding' (FDR = 1.13E-22), 'cytoplasm' (FDR = 5.63E-32) and 'osteoclast differentiation' (FDR = 0.000992221) were significantly enriched pathways for DEGs. HSP90AA1 exhibited a highest degree (degree = 32) and was enriched in 'pathways in cancer' and 'signal transduction'. BMP6, regulated by Pax-6, was enriched in the 'TGF-beta signaling pathway'. We indicated that BMP6 may be downregulated by Pax-6 in the non-metastasis OS samples. The up-regulated HSP90AA1 and down-regulated BMP6 and 'pathways in cancer' and 'signal transduction' were deduced to be involved in the pathogenesis of OS. The identified biomarkers and biological process in OS may provide foundation for further study.

Keywords Osteosarcoma · Transcription factors · DEGs · Integrated analysis

Introduction

Osteosarcoma (OS), characterized with fast growth, high metastatic potential, and local aggressiveness, is a mesenchymal malignancy in skeletal system affecting mainly children and adolescents [1, 2]. It is still a challenge for current therapeutic strategies to effectively cure osteosarcoma due to the unknown molecular mechanisms of pathogenesis. The treatment methods of OS patients are chemotherapy and complete surgical resection of cancer tissue. In recent years, despite tremendous progress in early diagnostic and therapeutic techniques for OS, resistance to chemotherapy and the recurrence of disease remain the two roadblocks in the therapy of this tumor [3]. The overall 5-year survival rate is around 20% to 30% in OS patients with metastatic disease at diagnosis. Therefore, it is necessary to explore the candidate pathogenic genes of OS in order to improve therapeutic treatments [4].

Transcription factors (TFs) are proteins that bind to specific DNA sequences of the target gene promoter and enhance or inhibit gene transcription [5]. Transcription of many genes was regulated by TFs, including cytokines, apoptosis-inducing molecules, growth factors and intercellular adhesion molecules [6]. Some researchers have detected abnormally expressed TFs in the progression of many diseases in human and animals. TFs may regulate the expression levels of crucial genes and modulate pathologic biological pathways as endogenous regulators. However, the function of TFs in the pathogenesis of OS remains unclarified. Hence, there is an increasing urgency to target candidate proteins or pathways involved in pathogenesis of disease in OS [7].

We integrated three datasets to identify DEGs between OS metastasis group and non-metastasis group. Functional

✉ Xiaofeng Shi
shi.xiaofeng@medintell.com

[1] Gu'an Bojian Bio-Technology Co., LTD, Langfang, China

[2] Department of BigData, Beijing Medintell Bioinformatic Technology Co., LTD, No. 1, Shanyuan Street, Haidian District, Beijing 100081, China

[3] School of biotechnology, Jiangnan University, Wuxi, China

annotation of DEGs, related TFs and target genes were also identified, which may be involved in the progression of OS. Furthermore, the transcriptional regulatory network was constructed to clarify the possible mechanism in OS.

Methods

Microarray Expression Profiling in GEO Datasets

The high-throughput microarray GSE66673, GSE49003 and GSE37552 datasets of OS metastasis group and non-metastasis group were get from the GEO database (http://www.ncbi.nlm.nih.gov/geo), which was sequenced on the GPL13607Agilent-028004 SurePrint G3 Human GE 8x60K Microarray, GPL6947 Illumina Human HT-12 V3.0 expression beadchip and GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, respectively [8]. The samples of metastasis group vs non-metastasis group were 12:12, 6:6 and 2:2 of the three datasets, respectively. The following key search terms were used: "osteosarcoma", AND "*Homo sapiens*" AND "gse". The selection standards were: (1) The dataset should be genome-wide mRNA transcriptome data. (2) These data must be from osteosarcoma cell lines [9].

Screening of Differentially Expressed Genes

Based on the three databases of GSE66673, GSE49003 and GSE37552, we utilized log2 transformation to normalize the data. The DEGs between these two groups were identified via the limma package in R (www.bioconductor.org/packages/release/bioc/ html/ limma.html). The Limma package in R was used to calculate *p*-values by two-tailed Student's t test. MetaMA package in R was used to combine p-values, and the false discovery rate (FDR) was obtained from multiple comparisons using the Benjamini and Hochberg method [10]. The DEGs with criterion of FDR < 0.05 were screened out.

Functional Annotation of DEGs

We performed GO and KEGG pathway enrichment analysis to study the characteristic biological functions and potential pathway of DEGs [11]. The gene ontology functions of the DEGs were determined, which include biological process, molecular functions and cellular components. In addition, based on the KEGG database, pathway enrichment analysis was obtained [12].

Integration of Protein-Protein Interaction (PPI) Network

According to the data from BioGRID (http://thebiogrid.org/), we constructed PPI networks of significantly DEGs to obtain candidate genes related to OS. Based on the subsistent data of protein interaction in BioGRID database, Cytoscape was used to find top 100 up-regulated and top 100 down-regulated DEGs [13]. We drew the protein network interaction map after removing the non-differentially expressed genes.

Screening TFs of the Top 20 DEGs and Constructing Regulation Network

On the UCSC website, we downloaded the 2 kb upstream promoter region for the top 20 DEGs. Then we analyzed TFs capable of binding to the promoter region of the DEGs by using TRANSFAC website's match tool. Based on the criterion of FDR < 0.001, altered expression of genes in OS were revealed by TFs. The match tool on the TRANSFAC website was used to analyze transcription factors that bind to the promoter regions of these DEGs. After that, the OS-specific transcriptional regulatory network was constructed by the Cytoscape software (http://www.cytoscape.org/).

Electronic Validation of DEGs in GEO Database

We utilized the Gene Expression Omnibus GSE87624 (GEO: GSE87624) database to validate the expression of selected OS related DEGs. The expression levels of these DEGs were compared between metastasis group and non-metastasis group. The expression of 6 genes (HSP90AA1, BMP6, Pax-6, FOXA2, SNPH and RBP1) was reported. The different expression levels of these DEGs were displayed by box-plots.

Results

Differential Expression Analysis of Genes in Metastasis Group Compared to Non-metastasis Group

Three gene expression microarray datasets (GSE66673, GSE49003 and GSE37552) were enrolled. Compared with the non-metastasis group, 759 DEGs in metastasis group were obtained with the criterion $P < 0.05$, among which, the expression of 352 genes were increased and 407 genes were decreased. The top 10 up- and down-regulated DEmRNAs between metastasis group and non-metastasis group were shown in the Table 1. Based on the three datasets, after cluster analysis, the heatmap of top 50 DE genes is shown in Fig. 1.

Functional Annotation

In Fig. 2, GO enrichment showed that the DEGs were significantly enriched in biological processes such as 'cell adhesion' (FDR = 1.27E-08), 'signal transduction' (FDR = 4.31E-07), 'multicellular organismal

Table 1 Top 10 up- and down-regulated DEmRNAs between metastasis group and non-metastasis group

DEmRNAs	*p* value	fdrp	Regulation
SNPH	1.54E-07	0.002585307	down
RBP1	3.40E-07	0.002862441	down
SLC7A10	1.33E-06	0.006034765	down
SDF2	2.47E-06	0.007026406	down
NPTX2	2.51E-06	0.007026406	down
MOXD1	4.62E-06	0.011101994	down
ANKRD30B	7.90E-06	0.01328414	down
SELENOM	9.65E-06	0.013530842	down
RGMA	1.10E-05	0.01391482	down
QPRT	1.16E-05	0.01391482	down
FOXA2	1.44E-06	0.006034765	up
BMP6	6.62E-06	0.01328414	up
FRMD3	7.71E-06	0.01328414	up
ANTXR2	8.86E-06	0.013530842	up
KLF2	2.39E-05	0.016578998	up
EMP1	2.74E-05	0.016578998	up
C12orf56	3.13E-05	0.016578998	up
ZRANB2	4.25E-05	0.018827855	up
CNIH3	4.51E-05	0.019449218	up
F2RL1	4.64E-05	0.019506684	up

development' (FDR = 7.28E-06). DEGs were significantly enriched in cellular components: 'cytoplasm' (FDR = 5.63E-32), 'plasma membrane' (FDR = 1.18E-19) and molecular functions: 'protein binding' (FDR = 1.13E-22), 'nucleotide binding' (FDR = 1.97E-08), 'calcium ion binding' (FDR = 1.01E-06). Furthermore, in Fig. 3, the results of KEGG pathway enrichment showed that DEGs were enriched in 'pathways in cancer' (FDR = 0.000974355), 'osteoclast differentiation' (FDR = 0.000992221), 'cytokine-cytokine receptor interaction' (FDR = 0.000150798).

PPI Network Analysis of DEGs

The result of PPI network was shown in Fig. 4. The network was consisted of 169 nodes and 165 edges. The nodes were represented the proteins and the lines were represented the interaction between them. Among them, the top 10 genes with higher degree were HSP90AA1 (degree = 32), NEDD4L (degree = 10), TUBB3 (degree = 7), PDGFRB (degree = 7), MYH9 (degree = 6), CDH1 (degree = 6), ACTA2 (degree = 6), FOS (degree = 6), NOTCH3 (degree = 6), MANSC1 (degree = 6).

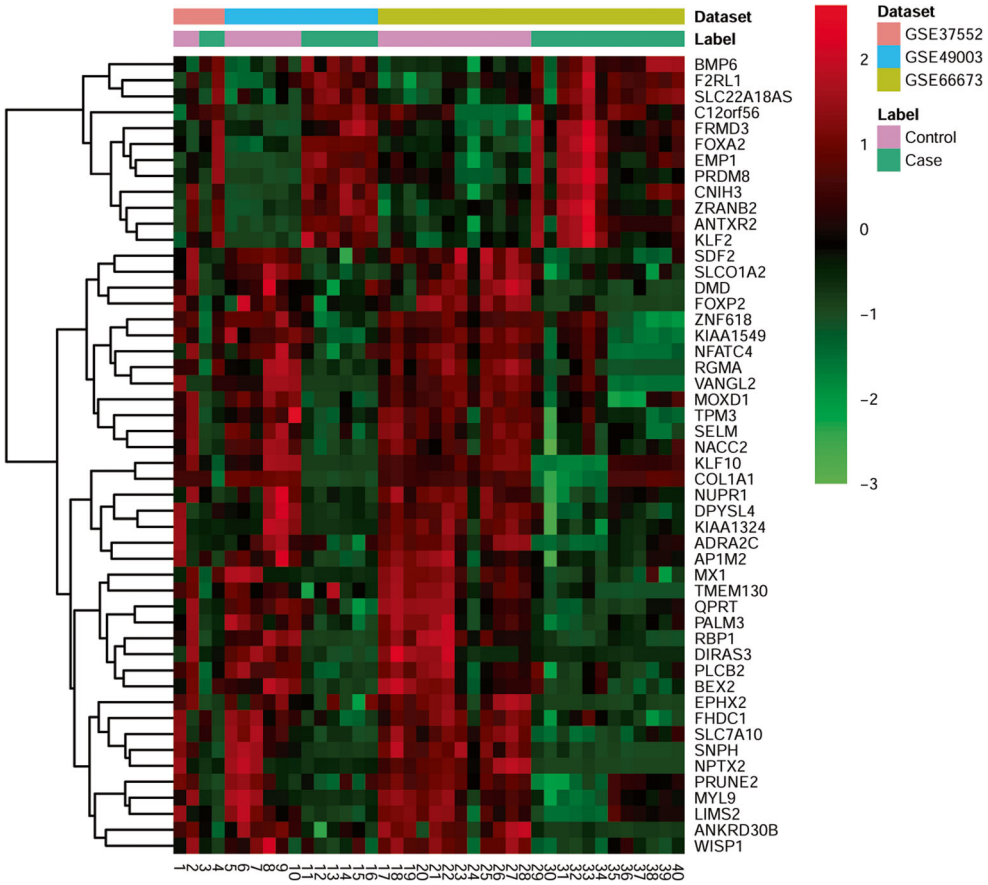

Fig. 1 Heatmap image displaying top 50 genes that were significantly up-regulated or down-regulated (*P* value<0.05) in the metastasis OS samples compared with the non-metastasis OS samples

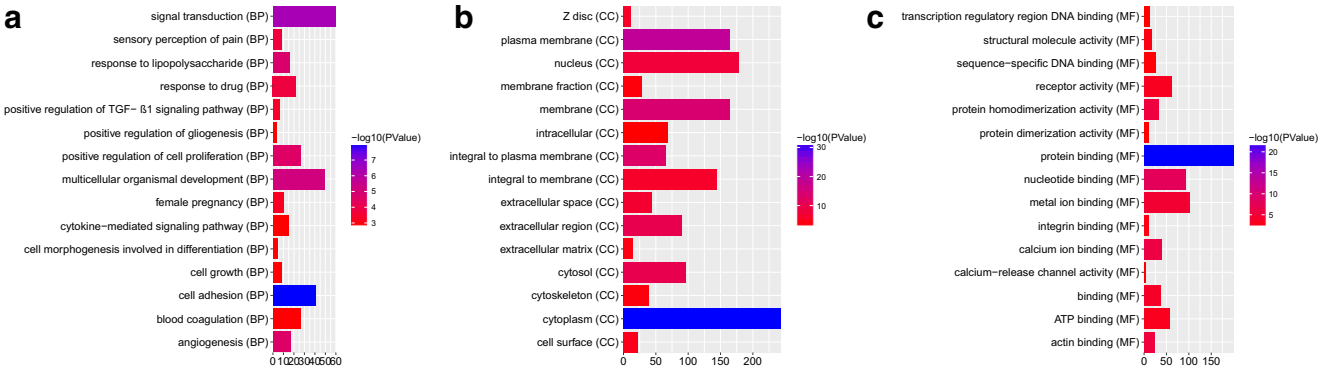

Fig. 2 Go functional enrichments of DEGs (FDR < 0.05). **a** Biological process, **b** Cellular components, **c** Molecular functions

TFs of the Top 20 DEGs and Regulatory Network

A total of 378 TFs-target genes were obtained from the regulatory network, which include 39 TFs in the binding relationships. In Table 2, the top 8 TFs with the most downstream genes include Pax-4, Elk-1, 1-Oct, Nkx2–5, myogenin, Pax-6, AP-1 and HNF-4. In Fig. 5, there were 59 nodes and 123 edges in the regulatory network. Among which, the top 7 TFs with highest degree were FRMD3 (degree = 15), Pax-4 (degree = 12), Nkx2–5 (degree = 10), RGMA (degree = 10), 1-Oct (degree = 10), Pax-6 (degree = 8) and myogenin (degree = 8).

Validation of DEGs in GEO GSE87624 Dataset

We searched the online GSE87624 dataset to define the key genes that play important role in OS. In Fig. 6, the expression of FOX2, Pax6 and RBP1 were up-regulated in metastasis group compared to the non-metastasis group, Inversely, the expression of BMP6, HSP90AA1 and SNPH were down-regulated in metastasis group compared to the non-metastasis group. Among them, the expression of FOX2, Pax6 and SNPH was generally consistent with the integrated analysis in GSE66673, GSE49003 and GSE37552 datasets.

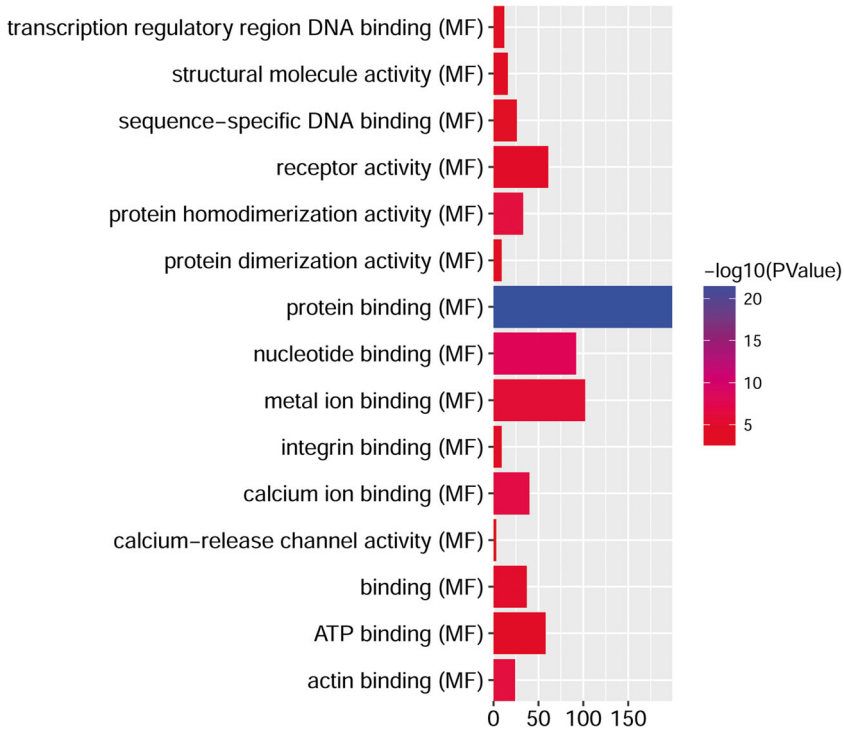

Fig. 3 KEGG analysis of DEGs in the metastasis OS samples compared with the non-metastasis OS samples

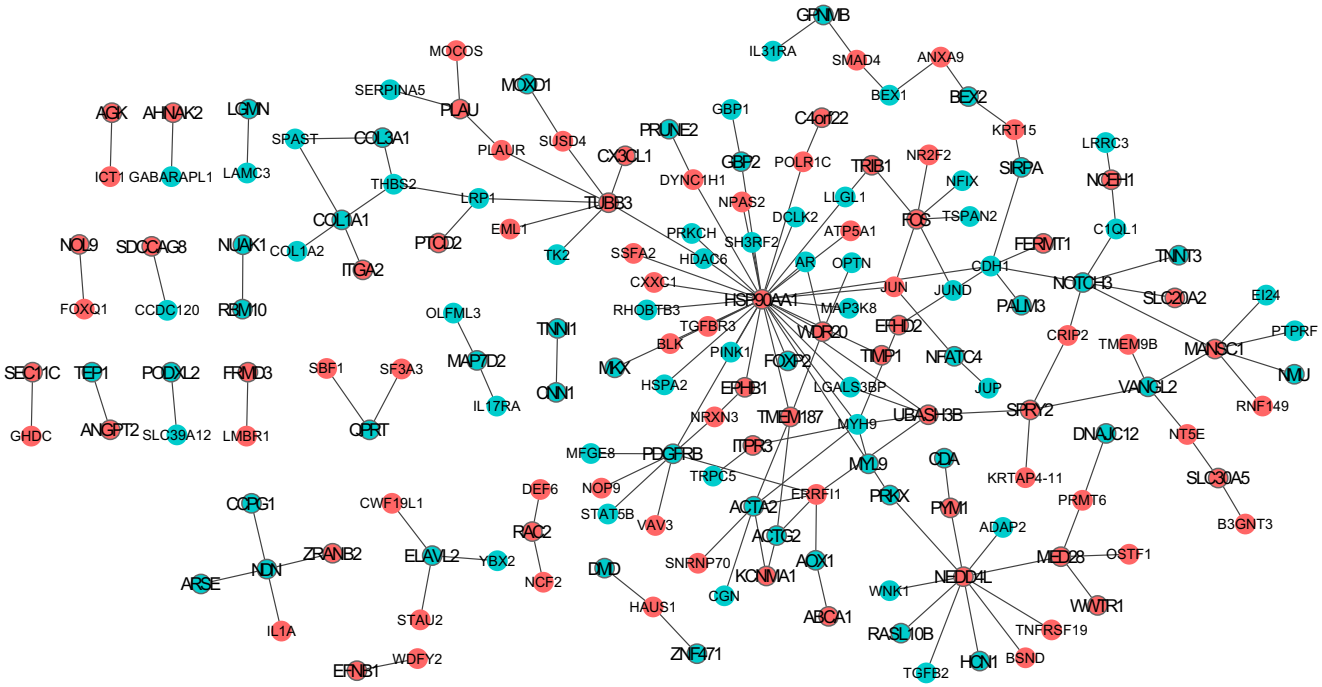

Fig. 4 The PPI network of top 100 significantly DEGs. The green circles were represented the proteins encoded by down-regulated DEGs and the red circles were represented the proteins encoded by up-regulated DEGs

Discussion

OS often occurs in the metaphyseal region of tubular long bones, which is the most common type of primary bone cancer [14, 15]. Though researchers have reported that many genes are involved in the pathogenesis of OS, it is still obscure how genes are modulated by other molecular [3, 4]. In the current study, a total of 759 DEGs were identified in the metastasis OS samples compared with the non-metastasis OS samples, which include 352 upregulated and 407 downregulated genes. In the KEGG analysis, DEGs were enriched in 'pathways in cancer' (FDR = 0.000974355), 'osteoclast differentiation' (FDR = 0.000992221). HSP90AA1 was upregulated in the metastasis OS samples compared with the non-metastasis OS samples. In the PPI network constructed for the DEGs, HSP90AA1 exhibited a highest degree (degree = 32) and therefore was identified to be highly interconnected with other proteins. Enrichment analysis revealed that HSP90AA1 was enriched in pathways in cancer and signal transduction.

HSP90AA1 is a 90-kDa heat shock protein related to numerous proteins that are highly expressed in many cancer cells [16]. Several cancer-related client proteins including PIM1, AKT, and HIF1A were stabilized by HSP90AA1, which are crucial for tumor progression [16]. Thus, HSP90AA1 is a remarkable target for cancer therapy. Shen et.al demonstrated that HSP90AA1 was downregulated by 2-24a/Cu in cancer cells, which is a crucial protein for cancer cell survival. Coskunpinar et al. reported that HSP90AA1 polymorphisms might be related to an increased risk of lung cancer [17]. Chu et al. suggested that HSP90AA1

Table 2 The top8 TFs with the most downstream regulatory genes and their target genes

Factor.name	Number of regulated genes	Regulated genes
Pax-4	12	CNIH3,F2RL1,FRMD3,MOXD1,NPTX2,QPRT,RBP1,SLC7A10,SNPH,ZRANB2
Elk-1	11	ANTXR2,C12orf56,EMP1,FRMD3,MOXD1,SDF2,ZRANB2
1-Oct	10	FOXA2
Nkx2-5	10	FRMD3
myogenin	8	NPTX2
Pax-6	8	BMP6,EMP1,SNPH,ZRANB2
AP-1	7	CNIH3
HNF-4	5	FOXA2,FRMD3

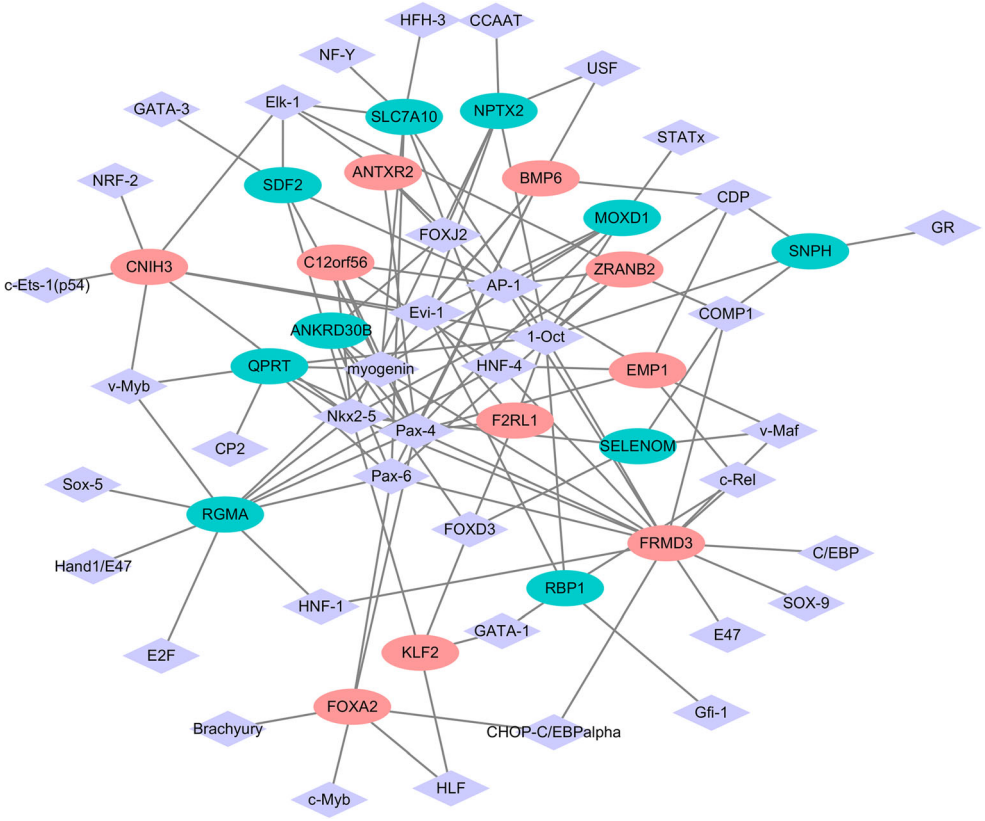

Fig. 5 The transcription factors regulation network diagram. Purple rhombus were represented transcription factors, ellipses were represented top20 genes, the red ellipse were represented the up-regulated DEGs and the green ellipse were represented the down-regulated DEGs

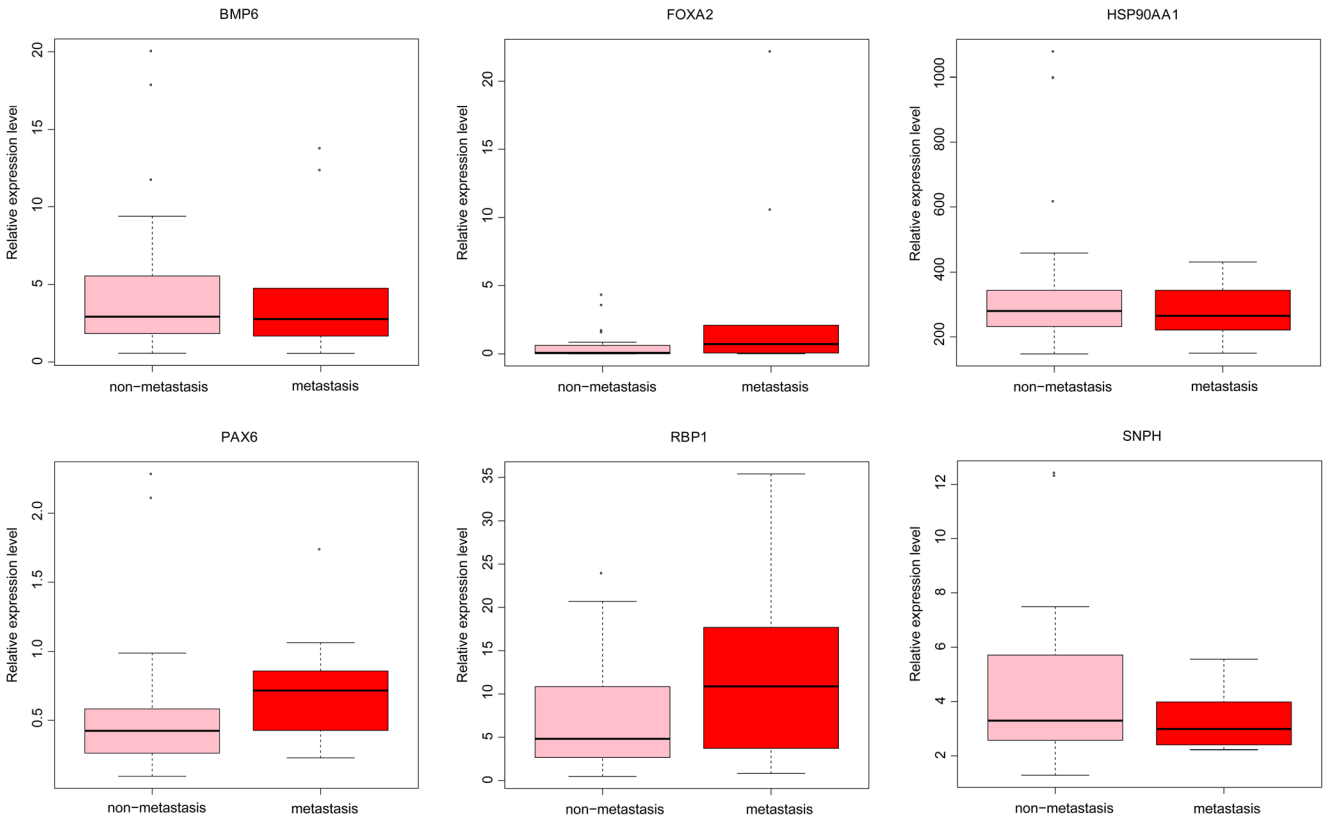

Fig. 6 Validation of the expression levels of selected DEGs in OS based on GEO database. The x-axis shows case and normal groups and y-axis shows expression reads counts/gene expression level: BMP6, FOXA2, HSP90AA1, PAX6, RBP1, SNPH

can function as factor in ovarian cancer cells and promotes chemoresistance to cisplatin [18].

Bone morphogenetic proteins (BMPs) initiate new bone formation in vivo and promote the growth and differentiation of cells in the osteoblastic lineage, which belong to the transforming growth factor-β (TGF-β) superfamily [19]. Bone morphogenetic protein 6 (BMP6) regulates cell growth, differentiation and apoptosis in many types of tumor [20]. Shi et al. reported that BMP6 was upregulated in patients with cancer-related anemia compared with non-anemia cancer group [21]. Hu et al. indicated that BMP-6 suppressed breast cancer metastasis by modulating the secretion of matrix metalloproteinase (MMPs) in the tumor microenvironment [22]. Liu et al. indicated that hypermethylation modifications can regulate the expression of BMP6 and caused an epithelial-mesenchymal transition phenotype of breast cancer [23].

Pax-6, acting as a member of the paired box (Pax) family, plays an important role in oncogenesis. Pax-6 was identified to be involved in pa thogenesis of glioblastoma, bladder cancer and prostate cancer, indicating that Pax-6 may function as a tumor suppressor and serve as a molecular biomarker for cancer development. Xiangyun Zong et al. [24] reported that Pax6 facilitates important regulatory roles in breast cancer cell proliferation and tumor progression, and could serve as a diagnostic marker for clinical investigation. Shyr CR et al. [25] reported PAX6 expression was higher in normal epithelial cells than cancer cells in prostate cancer tissues. In the TFs regulation network, Pax-6 is the transcription factor that binding to the promoter of BMP6. In the KEGG pathway enrichment, BMP6 was enriched in the TGF-beta signaling pathway. So the BMP6 may be downregulated by Pax-6 in the non-metastasis OS samples.

Conclusion

To conclude, from three GEO datasets, we identified a total of 759 DEGs between the metastatic and non-metastatic samples. After GO and KEGG analysis, 'cell adhesion', 'protein binding', 'cytoplasm' and 'osteoclast differentiation' were significantly enriched pathways for DEGs. HSP90AA1 exhibited a highest degree and was enriched in pathways in cancer and signal transduction. BMP6 was enriched in the TGF-beta signaling pathway which was regulated by Pax-6. We assumed that BMP6 may be downregulated by Pax-6 in the non-metastasis OS samples. The identified biomarkers and pathways in OS may provide references for further study.

Authors' Contributions CY and XS designed and performed the train of thought, DH and CM analyzed the resulting data, JR, LF, CC and BL contributed the analysis tools. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Ethics Approval and Consent to Participate Not applicable.

Consent to Publication All authors consented to publication.

Competing Interests All authors declare that they have no conflicts of interest.

References

1. Zhou W, Hao M, Du X, Chen K, Wang G, Yang J (2014) Advances in targeted therapy for osteosarcoma. Discov Med 17(96):301–307
2. Bernardini, G., Geminiani, M., Gambassi, S., Orlandini, M., Petricci, E., & Marzocchi, B., et al. (2017). Novel smoothened antagonists as anti-neoplastic agents for the treatment of osteosarcoma. J Cell Physiol
3. Yan H, Zhang B, Fang C, Chen L (2018) Mir-340 alleviates chemoresistance of osteosarcoma cells by targeting zeb1. Anti-Cancer Drugs:1
4. Pang Y, Zhao J, Fowdur M, Liu Y, Wu H, He M (2018) To explore the mechanism of the grm4 gene in osteosarcoma by rna sequencing and bioinformatics approach. Med Sci Monit Basic Res 24:16–25
5. Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific dna binding proteins. Science 245(4916):371–378
6. Zhu M, Liu CC, Cheng C (2013) Reactin: regulatory activity inference of transcription factors underlying human diseases with application to breast cancer. BMC Genomics 14(1):504
7. Yang L, Feng S, Yang Y (2016) Identification of transcription factors (tfs) and targets involved in the cholangiocarcinoma (cca) by integrated analysis. Cancer Gene Ther 23(12):439–445
8. Diao C, Xi Y, Xiao T (2018) Identification and analysis of key genes in osteosarcoma using bioinformatics. Oncol Lett 15(3): 2789–2794
9. Liu HY, Zhang CJ (2017) Identification of differentially expressed genes and their upstream regulators in colorectal cancer. Cancer Gene Ther 24:244–250
10. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate-a practical and powerful approach to multiple testing. J R Stat Soc 57(1):289–300
11. Li JJ, Wang BQ, Fei Q, Yang Y, Li D (2016) Identification of candidate genes in osteoporosis by integrated microarray analysis. Bone Joint Res 5(12):594–601
12. Wang F, Wang R, Li Q, Qu X, Hao Y, Yang J, Zhao H, Wang Q, Li G, Zhang F, Zhang H, Zhou X, Peng X, Bian Y, Xiao W (2017) A transcriptome profile in hepatocellular carcinomas based on integrated analysis of microarray studies. Diagn Pathol 12(1):4
13. Lee YS, Kim JK, Ryu SW, Bae SJ, Kwon K, Noh YH, Kim SY (2015) Integrative meta-analysis of multiple gene expression profiles in acquired gemcitabine-resistant cancer cell lines to identify novel therapeutic biomarkers. Asian Pac J Cancer Prev 16(7):2793–2800
14. Luetke A, Meyers PA, Lewis I, Juergens H (2014) Osteosarcoma treatment - where do we stand? A state of the art review. Cancer Treat Rev 40(4):523–532
15. Guan D, Tian H (2017) Integrated network analysis to explore the key genes regulated by parathyroid hormone receptor 1 in osteosarcoma. World J Sugr Oncol 15(1):177
16. Taipale M, Jarosz DF, Lindquist S (2010) Hsp90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol 11(7):515–528

17. Coskunpinar E, Akkaya N, Yildiz P, Oltulu YM, Aynaci E, Isbir T, Yaylim I (2014) The significance of hsp90aa1, hsp90ab1 and hsp90b1 gene polymorphisms in a turkish population with non-small cell lung cancer. Anticancer Res 34(2):753–757
18. Chu SH, Liu YW, Zhang L, Liu B, Li L, Shi JZ, Li L (2013) Regulation of survival and chemoresistance by hsp90aa1 in ovarian cancer skov3 cells. Mol Biol Rep 40(1):1–6
19. Mckay, W., Boden, S., & Yoon, S. (2006). METHODS OF INDUCING THE EXPRESSION OF BONE MORPHOGENETIC PROTEINS (BMPs) AND TRANSFORMING GROWTH FACTOR-beta-PROTEINS (TGF-betas) IN CELLS. CA, EP1629106
20. Honda Y, Knutsen R, Strong DD, Sampath TK, Baylink DJ, Mohan S (1997) Osteogenic protein-1 stimulates mrna levels of bmp-6 and decreases mrna levels of bmp-2 and -4 in human osteosarcoma cells. Calcif Tissue Int 60(3):297–301
21. Shi, Y. J., & Pan, X. T. (2016). Bmp6 and bmp4 expression in patients with cancer-related anemia and its relationship with hepcidin and s-hjv. Genet Mol Res Gmr, 15(1)
22. Hu F, Zhang Y, Li M, Zhao L, Chen J, Yang S et al (2015) Bmp-6 inhibits the metastasis of mda-mb-231 breast cancer cells by regulating mmp-1 expression. Oncol Rep 35(3)
23. Liu G, Liu YJ, Lian WJ, Zhao ZW, Yi T, Zhou HY (2014) Reduced bmp6 expression by dna methylation contributes to emt and drug resistance in breast cancer cells. Oncol Rep 32(2): 581–588
24. Zong X, Yang H, Yu Y, Zou D, Ling Z, He X, Meng X (2011) Possible role of pax-6 in promoting breast cancer cell proliferation and tumorigenesis. BMB Rep 44(9):595–600
25. Shyr CR, Tsai MY, Yeh S, Kang HY, Chang YC, Wong PL et al (2010) Tumor suppressor pax6 functions as androgen receptor co-repressor to inhibit prostate cancer growth. Prostate 70(2):190–199