Exploring the association mechanism between metastatic osteosarcoma and non-metastatic osteosarcoma based on dysfunctionality module

Jun Xiong, Tao Li

Department of Sports Medicine, Puai Hospital Affiliated to Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei 430000, P.R. China.

Summary

**Purpose:** Osteosarcoma (OS) is the primary malignant tumor which is common in children and adolescents. The treatment effect is still poor, though the treatment strategy has been dramatically improved.

**Methods:** Differentially expressed genes in metastatic osteosarcoma and non-metastatic osteosarcoma were obtained first. Secondly, co-expression analysis has been processed for differentially expressed genes, and it is necessary to figure the gene drive of each module. Furthermore, both GO function and KEGG pathway enrichment analysis were performed on the module genes. Comprehensively, the module gene set which was predicted according to hypergeometric testing was importantly regulated by both transcription factors (TFs) and non-coding RNAs (ncRNAs).

**Results:** Conclusively, 16 co-expression modules were obtained. ACAT1 and ATBF1 would actively regulate in dysfunction modules, and thus they are identified as osteosarcoma-driven genes. Enrichment results showed that the module genes were significantly involved in transcription factor activity, specific DNA binding of the RNA polymerase II proximal promoter sequence, DNA-binding transcriptional activator activity, ubiquitin-like protein transferase activity, and another biological process. Moreover, module genes significantly regulates FcγR-mediated phagocytosis, MAPK signaling pathway, phagocytosis, PI3K-Akt signaling pathway and others. Finally, we identified pivot ncRNAs (including CRNDE, miR-106a-5p, miR-181a-5p, etc) and pivot TFs (including NFKB1, STAT6, PPARG, RELA, etc) that significantly regulate dysfunction modules.

**Conclusion:** Overall, this work deciphered a co-expression network of common core pathogenic genes including metastatic osteosarcoma and non-metastatic osteosarcoma. It helps to identify core dysfunction modules and potential regulatory factors of the disease and improves understanding the underlying molecular association mechanisms between the two diseases.

**Key words:** osteosarcoma, co-expression, module, driver gene, enrichment, regulatory factors

Introduction

Osteosarcoma (OS) is a malignant tumor that primarily affects long bones, but may also involve other bones in the body. Its incidence has a bimodal distribution that peaks in the second decade of life and later in adulthood [1]. The most common site of origin is the distal femur, proximal humerus, usually characterized by pain, swelling, local swelling of the extremities, and occasionally pathological fractures [2]. Since the lung is a common site of metastatic spread of OS, accordingly concurrent bilateral spontaneous pneumothorax (SBSP) is a rare clinical sequel of this malignancy [3]. Excitingly, some human genetic diseases and familial cancer syndromes, such as Li-Fraumeni syndrome are associated with an increased risk of OS [4].

In contrast to many other malignancies, a variety of genetic and environmental factors will lead to the development of OS [5]. In genetics, fibroblast
growth factor 2 (FGF2) SNP rs11737764 is significantly associated with increased susceptibility to OS in a Chinese population [6]. Meanwhile, the p53 gene is the most important tumor suppressor and is related to the risk of many tumors (including OS). Its rs1295925 and rs3787547 play an essential role in the etiology of OS and are potential genetic modifiers in the development of OS [7].

Until now, scientists have interpreted OS from various angles, and it has been linked to success. LncRNA SOX2-OT has been tested and verified to regulate the proliferation, migration, invasion, and expression of cancer stem cell biomarkers in OS cells. Additionally, it has become a prognostic biomarker for OS patients [8]. Overexpression of DANCR increases proliferation, migration, and invasion of OS cells in vitro, while promotes xenograft tumor growth and in vivo lung metastasis [9]. SNHG15 can promote cell proliferation, migration, and autophagy in OS by negatively regulating miR-141, providing new potential targets and prognostic biomarkers for OS treatment [10]. MALAT 1 may inhibit tumor growth and metastasis through the PI3K/AKT signaling pathway and may be considered as a therapeutic target for human OS [11].

On the other hand, miR-133b regulates the growth and mechanism of OS cells through the miR-133b/Sirt1 axis, and the low expression of miR-133b is significantly associated with tumor size, distant metastasis, and advanced clinical stage, and can be used as potential therapeutic targets of OS in the future [12]. Regarding treatment, MAP regimens (methotrexate, doxorubicin, and cisplatin) have become the preferred treatment [13]. More importantly, chemotherapy for OS is one of the most important chemotherapies in solid tumors, and treatment of primary tumors is associated with a certain degree of permanent disability in a large proportion of patients [14]. These findings have deepened our understanding of the pathogenesis of OS and guided us in the direction of further research. In addition, the survival status of metastatic OS is different from that of non-metastatic OS [15]. To comprehensively and deeply explore the potential association between metastatic and non-metastatic OS, we conducted a systematic modular analysis to determine the dysfunction modules and core modules between them.

As a consequence, we can explore OS further concerning knowing the driving genes of OS cell metastasis. In sum, our work radically introduces the relationship between multifactor-mediated dysfunction modules in metastatic and non-metastatic OS. It helps identify potential therapeutic targets and related biological processes, which may assist to build a better understanding of OS and provide improved treatments for this disease. This research offers abundant resources for future experimental validation and drug repositioning. Equally important, the study offers theoretical guidance to future biological research on OS.

**Methods**

**Data resource**

The NCBI Gene Expression Omnibus database (GEO Dataset) [15] with a broad classification of high-throughput experimental data has single-channel and dual-channel microarray-based assays for mRNA abundance, genomic DNA and protein molecules. Besides, it includes data from non-array-based functional genomics with high-throughput and proteomics technologies. We first collected a set of gene expression profiling for non-metastatic and metastatic OS from GEO, numbered GSE9508 [16]. The dataset included 13 non-metastatic osteosarcomas, 21 metastatic osteosarcomas, and 5 normal samples. Then, we screened the ncRNA-mRNA interaction pairs with score ≥0.5 from the RAID v2.0 database [17], including 5431 ncRNAs and 451937 interaction pairs. The RAID v2.0 database enrolls more than 5.27 million RNA-related interactions, including more than 4 million RNA-RNA interactions as well as there are more than 1.2 million RNA-protein interactions between them.

Further, 60 species and 150,000 RNA/protein are involved. It can help observe various RNA-related interactions comprehensively. At the same time, human transcription factor target data were downloaded and used in the general database TRUST v2 database in transcription studies [18]. There are a total of 2,492 transcription factors, and 9,396 pairs.

**Differential expression analysis**

To explore the potential key genes for OS cell metastasis, we respectively performed a differential analysis of non-metastatic OS samples and normal samples, and also metastatic OS and normal samples. The differential expression analysis of the gene expression profiling data was performed using the R language limma package [19-21]. First, we used the correct background function to perform both background correction and normalization of the data. Next, the control probe and the low expression probe were filtered using the Normalize Between Arrays function quantize normalization. Then, the differentially expressed genes of the data set were identified based on the lmFit and eBayes functions, using default parameters.

**Co-expression analysis**

We used weighted gene co-expression network analysis (WGCNA) [22] to analyze the expression of differentially expressed genes and to find out gene modules for synergistic expression. First, the correlation coefficient weighting value was used, that is, the gene correlation coefficient which was taken to the power of N to
work out the correlation coefficient (Pearson Coefficient) between any two genes which are calculated. The connections between genes in the network are subject to scale-free networks, making the algorithm more meaningful biologically. Then, a hierarchical clustering tree is constructed by correlation coefficients between genes. Different branches of the clustering tree represent different gene modules. Meanwhile, different colors represent different modules.

**Function and pathway enrichment**

Exploring the functions and signaling pathways which are involved in gene organization often help to study the molecular mechanisms of disease. Meanwhile, the enrichment analysis of features and pathways of genes in dysfunctional modules are useful means to explore the underlying mechanisms of OS. Therefore, we performed the function and KEGG pathway enrichment analysis on the 16 modular genes of OS using the R language Cluster profiler package [23]. Cluster Profiler is a Bioconductor software package that provides both statistical analysis and visualization of functional clustering of gene sets or gene clusters. In addition, we used Cytoscape’s ClueGO application to perform biological pathway analysis on the integrated module network.

**Transcription factors and non-coding RNAs that regulate dysfunctional modules**

Transcriptional and post-transcriptional regulation of genes is often driven by transcription factors (TF) and non-coding genes (ncRNA). Therefore, we have scientifically predicted and tested its role in the occurrence and development of OS. The pivot regulator is defined as a regulator that has significant regulatory implications for the module, including ncRNA and TF. We required that there are more than two regulatory connections between each regulator and each module, and the significance of the enriched target in each module calculated based on the hypergeometric test is $p<0.01$.

**Results**

**Identification of differentially expressed genes in metastatic and non-metastatic osteosarcoma**

Biologists have conducted many experiments and studies on the pathogenesis of OS, and thus they have identified potential pathogenic genes for OS metastasis. However, the complex molecular connections and overall effects of these genes are unclear. To observe possible molecular changes in
OS cell metastasis, we performed differential expression analysis of metastatic, standard sample microarray data, non-metastatic microarray data and standard microarray data to obtain differentially expressed genes (DEGs) in metastatic and non-metastatic OS, appropriately we can identify genes that may affect OS metastasis. Finally, disease-related differential genes were obtained and we believe that there may be genes in these differential genes that affect the migration of OS cells.

Identification of functional disorders of metastatic osteosarcoma associated with non-metastatic osteosarcoma

Modularity is a process that deals with globally complex systems and decomposes them into more detailed and organized subsystems, while each of which exhibits its characteristics. For each elemental gene, the module is a collection of genes with a synergistic expression relationship, and the genes of the same module have consistent expression behavior. On the other hand, each module also had a specific interaction relationship. The overall effect of these interactions represents the global characteristics, and it is a bridge for each element gene to play a role in the worldwide network. According to the expression behavior of related genes in diseases, they are clustered into modules, which helps observe the complex synergy between these genes from the perspective of expression behavior. To start with, based on the patient gene expression profile, 4,067 differential genes, and their interaction gene expression profile matrices were constructed. Then, using weighted gene co-expression network analysis (WGCNA), we observed that these genes exhibited significant group co-expression in disease samples, resulting in 16 functional disorder modules (Figure 1 and 2). Based on the phenotypic

![Module-trait relationships](image)

**Figure 3.** Each row represents a module, while each column represents a phenotype. The corresponding correlation coefficient maps the color of each cell, and the value is from -1 to 1; the color transitions from blue to white, and then transitions to red.
Figure 4. Extraction of GO function enrichment analysis of modular genes. The color increases from blue to purple, representing an increase in the enrichment significantly. The larger the circle, the more significant the proportion of the gene in the module that accounts for the GO function.
Figure 5. Extraction of KEGG pathway enrichment analysis of modular genes. The color increases from blue to purple, representing a significantly increase in the enrichment. The larger the circle, the more significant the proportion of the gene in the KEGG pathway entry.
association analysis of the module genes, it can be seen that MEred is associated with non-metastatic OS, whereas turquoise is associated with metastatic OS (Figure 3).

Functions and pathways involved in the gene of interest

Function and pathway are essential to the physiological response of the disease. Exploring both the features and pathways which are involved in the dysfunctional module gene, it not only helps determine the upstream and downstream relationship between different genes in the same biological pathway in the module. However, it also helps establish a molecular bridge between the module and the disease in system biology. More importantly, it deepens the understanding of the underlying molecular mechanism of the disease. We performed GO function and KEGG pathway enrichment analysis on 16 modules further, and we obtained 33,980 biological processes (BP), 4,278 cell components (CC), 6,362 molecular functions (MF) and 1,872 KEGG pathways (Figure 4 and 5). The analysis revealed that the tasks involved in these modules are mainly focused on transcription factor activity, RNA polymerase II proximal promoter sequence-specific DNA binding, DNA-binding transcriptional activator activity, RNA polymerase II specificity, ubiquitin-like protein transferase activity, etc. The significant biological pathways include FcγR-mediated phagocytosis, MAPK signaling pathway, phagocytosis, and PI3K-Akt signaling pathway. We believe that the genes which are with the most function and biological pathway would play a critical role in the module. Looking back at the whole, we integrated 16 module networks and used ClueGO for biological pathway analysis (Figure 6).

Pivot regulator that may affect disease metastasis

From the perspective of systems biology and systems genetics, transcriptional and post-transcriptional regulation of genes have already been realized as the primary regulatory factors of disease development. Meanwhile, transcription factors and ncRNAs are universal regulators of ex-
pression and function. Although many biologists have valued the regulation of the progression of metastatic OS by single or several TFs and ncRNAs, few studies have focused on their overall global effects on dysfunctional mechanisms. In this study, we performed a pivotal analysis of the co-expression module based on the targeted regulatory relationship between TF and ncRNA on the module gene. It is to explore the pivot regulators and potential regulatory mechanisms that may affect OS metastasis progression. A total of 19 crucial transcription factors were involved in the analysis of 19 TF-module target pairs, and 587 ncRNAs which touch upon 844 ncRNA-module regulatory pairs (Figure 7). Among them, ncRNAs (CRNDE, miR-106a-5p, miR-181a-5p, etc.) and TFs (NFκB1, STAT6, PPARG, RELA, etc.) were the most regulated dysfunctional modules and they were identified as potential dysfunctional molecules in the process of OS metastasis. Besides, they may regulate the operation of OS cell metastasis by mediating a dysfunctional module. Notably, PPARG regulates key gene ACAT1 from module 10, and MALAT1 and FENDRR regulate key gene VPS15A from module 10. Besides, CRNDE, NFκB1 and RELA regulate the negative regulation of leukocyte-mediated immunity, while CRNDE, NFκB1, RELA, PPARG, and STAT6 regulate vesicle-mediated trafficking. In addition, both CRNDE and STAT6 get involved in the hematopoietic lineage signaling pathway as well as CRNDE regulates the signaling pathway of Staphylococcus aureus infection.

Discussion

OS is a primary malignant bone tumor and the most common malignant bone tumor with poor prognosis with local recurrence, metastasis, chemotherapy resistance and so on. Its 5-year survival rate is 70% [24,25]. Although several researchers studied OS in various aspects, the pathogenesis of the disease remains unclear. In this study, we collected gene expression profiling of metastatic and non-metastatic OS on the NCBI Gene Expression Omnibus database (GEO Dataset). Accordingly, we identified differentially expressed genes which are disease-related. Integrating transcriptional and post-transcriptional regulatory data, and analyzing the co-expression modules of differentially expressed genes in various aspects, was intended to understand the molecular mechanism of OS cell metastasis thoughtfully. At the module level, module gene is significantly involved in transcription factor activity, RNA polymerase II proximal promoter sequence-specific DNA binding, DNA-binding transcriptional activator activity, RNA polymerase II specificity, ubiquitin-like protein transferase activity, and other biological processes. At the same time, it also significantly involved in FcγR-mediated phagocytosis, MAPK signaling pathway, PI3K-Akt signaling pathway, and other signaling pathways. Among them, studies have shown that downregulation of RPS9 leads to inactivation of MAPK signaling pathway and inhibits OS cell growth [26]. The high expression of PLA2G16 activates the MAPK pathway to enhance OS metastasis, and maybe it is the new therapeutic target for these cancers [27]. Moreover, more and more evidence proves that the PI3K/Akt signaling pathway is frequently overactivated in esophageal cancer and promotes the development and progression of the disease, including tumorigenesis, proliferation, invasion and cell cycle progression, inhibition of apoptosis, angiogenesis, metastasis and chemical resistance [28]. On the other hand, at the molecular level, we have extracted 16 core genes such as ACAT1 and ATBF1 by co-expression analysis. These core genes are not only significantly differentially expressed, but they also play an essential regulatory role in the dysfunction module. Among them, studies have shown that TLR4 siRNA inhibits the proliferation and invasion of colorectal cancer cells by down-regulating ACAT1 expression [29]. The ATBF1 gene may promote the development of HCC through transcription down-regulation of mRNA expression [30]. In breast cancer, low expression of ATBF1 promotes cell proliferation [31]. Although there is no association between OS and existing studies on core genes, our analysis shows that these core gene-driven dysfunction modules are the fundamental causes of OS metastasis and can be used as one of the future research focus. Furthermore, we predicted 587 ncRNAs and 19 transcription factors which are involved in the metastasis of OS cells through a mediator module. Besides, as reported by differential expression analysis, it was demonstrated that degrees of abnormal expression in OS are different. Based on statistical analysis, we determined that CRNDE, miR-106a-5p, miR-181a-5p, miR-200a-3p and miR-520D-3P all significantly regulated 5 dysfunctional modules. Among them, LncRNA CRNDE promotes proliferation, invasion, and migration of OS cells by regulating Notch 1 signaling and epithelial-mesenchymal transition [32]. MiR-106-5p inhibits C2C12 myogenesis by targeting PIK3R1 and modulating the PI3K/AKT signaling pathway [33]. Also, miR-181A-5P significantly inhibits the proliferation and migration of gastric cancer cells by regulating MEG2 and further inhibits the growth of tumor cells in vitro and in vivo [34]. The miR-182-5p improves the viability, mitosis, migration, and invasion of human gastric
cancer cells by downregulating RAB27A [35]. MiR-200a-3p inhibits tumor proliferation and induces apoptosis by targeting SPAG9 in renal cell carcinoma [36]. LncRNA MIAT and EPHA2 signaling contribute to the dysregulation of miR-520d-3p. Thus it promotes the development of hepatocellular carcinoma [37]. No studies involving OS have been reported involving miR-106a-5p, miR-181a-5p, miR-182-5p, miR-200a-3p, and miR-520d-3p, but our analysis revealed the above ncRNA. It has a regulatory relationship with the disease dysfunction module, which can be used as a candidate for further molecular experiments. Besides, PPARG regulates the key gene ACAT1 from module 10, and MALAT1 and FENDRR regulate the key gene VPS13A from module 10. Functional analysis of these regulators revealed that CRNDE, NFKB1, and RELA regulate the negative regulation of leukocyte-mediated immunity. Additionally, CRNDE, NFKB1, RELA, PPARG, and STAT6 regulate vesicle-mediated transport. Further, CRNDE and STAT6 are involved in the hematopoietic lineage signaling pathway, and CRNDE regulates the signaling path of Staphylococcus aureus infection. Recent studies have shown that the PPARG c.1547C > T polymorphism is associated with cancer risk [38]. When MALAT1 is knocked down in vitro by siRNA will significantly inhibit cell proliferation and migration. Besides, it is supposed to induce cell cycle arrest and apoptosis in OS cells. MALAT1 may be a promising therapeutic target for patients with OS [39]. At the same time, MALAT1 regulates HMGB1 through miR-142-3p and miR-129-5p to promote the development of OS [40]. Besides, overexpression of FENDRR which inhibits doxorubicin resistance and G2/M phase cell cycle can both promote apoptosis of OS cells in vitro and tumor growth in vivo [41]. NFKB1 has been shown to play a vital role in the pathogenesis of OS, and OS is associated with the NFKB1 promoter-94ins/del ATTG polymorphism [42]. The expression of p65 which is a nuclear factor-kB (NF-kB) subunit may beneficially be directed toward the survival of cells which contribute to tumor expression by inducing multiple antiapoptotic genes [43]. Importantly, serine/threonine kinase 35, as a target gene for STAT3, regulates proliferation and apoptosis of OS cells [44].

**Conflict of interests**

The authors declare no conflict of interests.

**References**

17. Endo-Munoz L, Cumming A, Rickwood D et al. Loss of osteoclasts contributes to development of osteosar-
Association mechanism between metastatic osteosarcoma and non-metastatic osteosarcoma

1. JHUO 2020; 25(3): 1578


29. Ding H, Chen Y, Qiu H et al. PPARγ c.1347C>T polymorphism is associated with cancer susceptibility: from a case-control study to a meta-analysis. Oncotarget 2017;8:102277-90.


35. Liu Z, Sun F, Hong Y et al. MEG2 is regulated by miR-181a-5p and functions as a tumour suppressor gene to suppress the proliferation and migration of gastric cancer cells. Mol Cancer 2017;16:135.


39. Ding H, Chen Y, Qiu H et al. PPARγ c.1347C>T polymorphism is associated with cancer susceptibility: from a case-control study to a meta-analysis. Oncotarget 2017;8:102277-90.


