Molecular characterization of metastatic osteosarcoma: Differentially expressed genes, transcription factors and microRNAs

LISONG HENG1,2*, ZHEN JIA3*, JIE BAI1, KUN ZHANG1, YANGJUN ZHU1, JIANBING MA1, JUN ZHANG1 and HONGHAO DUAN1

1Department of Orthopedics and Traumatology, Honghui Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710054; 2Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Xi'an Jiaotong University, Xi'an, Shaanxi 710061; 3Department of Endocrinology, Xi'an No. 1 Hospital, Xi'an, Shaanxi 710002, P.R. China

Received December 8, 2015; Accepted January 9, 2017

DOI: 10.3892/mmr.2017.6286

Abstract. The present study aimed to understand the molecular mechanisms underlying osteosarcoma metastasis. Microarray dataset GSE49003 was downloaded from the Gene Expression Omnibus database and used for analysis. Raw expression data were preprocessed using the preprocessCore, impute and aggregate packages in R. Differentially expressed genes (DEGs) between metastatic and non-metastatic osteosarcoma cell lines were screened using the limma package following exclusion of DEGs with a higher significance in intra-groups compared with inter-groups using the genefilter package. Enrichment analysis was performed on DEGs using TargetMine, followed by identification of transcription factors (TFs) and microRNAs (miRNAs). Regulatory networks were constructed using Cytoscape software. A total of 248 upregulated and 208 downregulated genes were obtained. The upregulated genes were significantly enriched in the following pathways: Downregulation of transforming growth factor β (TGF-β) receptor signaling activates SMADs; these upregulated genes included protein phosphatase 1, regulatory subunit 15A, transforming growth factor, β receptor II and ubiquitin carboxyl-terminal hydrolase L5. In addition, some upregulated genes were enriched in lung cancer disease ontology, including epidermal growth factor receptor (EGFR), insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3), runt-related transcription factor 3 (RUNX3) and secreted frizzled-related protein 1 (SFRP1). Conversely, the downregulated genes were significantly enriched in extracellular matrix-associated pathways or functions, such as collagen, type XII, α 1; collagen, type I, α 1; collagen, type IV, α 1; and collagen, type V, α 1. In addition, some downregulated genes were significantly enriched in the TGF-β signaling pathway, including bone morphogenetic protein 4, inhibitor of DNA binding 3 and SMAD family member 6. A total of 10 TFs and 84 miRNAs (e.g. miR-21-5p) were deemed to be associated with DEGs. In conclusion, DEGs enriched in the downregulation of TGF-β receptor signaling, TGF-β receptor signaling activates SMADs and TGF-β signaling pathways, as well as the extracellular matrix may be implicated in the progression of osteosarcoma metastasis. Dysregulated EGFR, IGF2BP3, RUNX3 and SFRP1 may contribute to the metastasis of osteosarcoma to the lungs. In addition, screened TFs and miR-21-5p may be associated with metastasis via target genes.

Introduction

Osteosarcoma is characterized as a primary bone malignancy, which originates from bone-forming mesenchymal cells and commonly occurs in children and young adults, with a second incidence peak occurring in adulthood (1,2). The pulmonary metastasis of osteosarcoma is responsible for the majority of cases of mortality in patients, due to the lack of effective therapeutic methods (3). Previous studies have aimed to develop specific therapeutic approaches targeting metastatic progression, and to understand the mechanisms underlying metastasis (4,5). Cysteine-rich angiogenic inducer 61 has been reported to promote the epithelial to mesenchymal transition of osteosarcoma (6). In addition, Ezrin expression has been reported to be necessary for metastasis of osteosarcoma (7), and high expression levels of transforming growth factor-β (TGF-β) have been observed in osteosarcoma cell lines (8). However, these previous findings are insufficient and gene expression profiling is required to comprehensively analyze the key genes associated with osteosarcoma metastasis.
Transcription factors (TFs) (9) and microRNAs (miRNAs) (10) are the two critical regulators targeting protein-coding genes involved in cancer initiation and progression. For example, runt-related transcription factor 2 may support the cell adhesion and motility of osteosarcoma cells (11). In addition, it has previously been reported that overexpression of SRY (sex determining region Y)-box 9, Wnt family member 1 and frizzled class receptor 1 may be associated with an advanced clinical stage (12), whereas inhibition of Wnt may markedly reduce the lung metastasis of osteosarcoma (13). miRNAs also contribute to osteosarcoma and have potential applications as therapeutic targets (14). For example, miRNA (miR)-143 downregulation may result in matrix metalloproteinase 13 upregulation, and thus promote lung metastasis of human osteosarcoma (15). Furthermore, miR-199a-3p and miR-34a may inhibit osteosarcoma cell growth and migration via downregulation of target genes, such as Met, mTOR and Stat3 (16,17). Therefore, the identification of these TFs and miRNAs may aid in the determination of regulatory network alterations over the progression of osteosarcoma.

The present study aimed to improve the understanding of the molecular mechanisms underlying osteosarcoma metastasis. The differentially expressed genes (DEGs) between metastatic osteosarcoma and non-metastatic osteosarcoma cells were determined by analyzing the GSE49003 public dataset deposited in the Gene Expression Omnibus (GEO), providers of which have not used the data to publish an article at present. Subsequently, identification of the associated TFs and miRNAs of DEGs was conducted using bioinformatics methods. Enrichment analysis was also performed on the DEGs to determine the metastasis-associated molecular mechanisms.

Materials and methods

Microarray data. Microarray data with accession number GSE49003 was downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) (18). This dataset contained data obtained from KHOS and KRB metastatic osteosarcoma cell lines, and HOS and U2OS non-metastatic osteosarcoma cell lines. Each of the four cell lines was analyzed in triplicate using Illumina HumanHT-12 V3.0 expression BeadChip (GPL6947; Illumina, Inc., San Diego, CA, USA).

Data processing. The downloaded microarray data were processed through missing value filtration with the criterion of probe intensity values P>0.05 in >3 of the 12 expression profiles. Kernel and nearest neighbor averaging methods (19) were used to impute the missing values using the impute package (20) in R (version 3.13; http://www.r-project.org/). Quantile normalization was performed to obtain standardized microarray data using preprocessCore package (21) in R. The probes were transformed into gene symbols based on the annotation information of platform GPL6947. Subsequently, the gene expression level of each gene was calculated by determining the average of the expression levels of probes corresponding to the same gene using aggregate package in R (22).

Screening of DEGs. In order to account for the innate differences among various cell lines, one-way analysis of variance was performed to exclude DEGs with higher significance in the intra-groups compared with the inter-groups using the genefilter package (23) in R. Student’s t-test was performed to identify differential expression between metastatic and non-metastatic cell lines using the limma package (24) in R. The Benjamini and Hochberg method was applied to adjust the P-value with false discovery rate (25). The DEGs were identified under the thresholds of log2 fold change (FC)>1(FC magnitude >2) and adjusted P<0.05.

Functional enrichment analysis. In order to determine the biological processes and pathways associated with the DEGs, the online tool TargetMine (http://targetmine.nibio.go.jp) (26) was used to perform Gene Ontology (GO) function enrichment analysis based on GO (http://www.genontology.org/) and UniProtKB GOA (http://www.ebi.ac.uk/GOA) databases; pathway enrichment analysis based on Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) and Reactome (http://www.reactome.org/) databases; and Disease Ontology (http://disease-ontology.org/) enrichment analysis.

Construction of integrated regulatory networks. In order to determine the regulatory networks of DEGs, Target Mine (26) was used to predict the TFs from the DEGs based on OregAnno (http://nar.oxfordjournals.org/content/early/2015/11/16/nar.gkv1203.abstract) and AMADEUS (http://acgt.cs.tau.ac.il/amadeus/) databases; and upstream miRNAs of DEGs based on miRBase (http://www.mirbase.org) and miRTarBase (http://mihtarbase.mbc.nctu.edu.tw/) databases. Construction of the integrated regulatory network between TFs and target genes, and the network between miRNAs and DEGs, was conducted using Cytoscape software (version 3.0.0) (27).

Results

Screened DEGs. Under the thresholds of log2, FC>1 and adjusted P<0.05, a total of 456 DEGs were identified in the metastatic osteosarcoma cell lines in comparison with the non-metastatic osteosarcoma cell lines, consisting of 248 upregulated and 208 downregulated genes.

Significantly enriched terms. The upregulated genes were significantly enriched in various pathways, including downregulation of TGF-β receptor signaling and TGF-β receptor signaling activates SMADS; genes enriched in these pathways included protein phosphatase 1, regulatory subunit 15A (PPP1R15A), transforming growth factor, β receptor II (TGFB2R) and ubiquitin carboxyl-terminal hydrolase L5 (UCHL5). The upregulated genes were also enriched in various biological processes and different diseases, including Barrett's esophagus, esophageal disease and lung cancer; genes enriched in these disease ontology terms included epidermal growth factor receptor (EGFR), insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3), runt-related transcription factor 3 (RUNX3) and secreted frizzled-related protein 1 (SFRP1). The top 5 significantly enriched terms are presented in Table I.

The downregulated genes were significantly enriched in various pathways, including extracellular matrix organization,
and collagen biosynthesis and modifying enzymes; genes enriched in these pathways included collagen, type XII, \( \alpha_1 \) (COL12A1), collagen, type I, \( \alpha_1 \) (COL1A1), collagen, type IV, \( \alpha_1 \) (COL4A1) and collagen, type V, \( \alpha_1 \) (COL5A1). Downregulated genes were also enriched in the TGF-\( \beta \) signaling pathway, including bone morphogenetic protein 4 (BMP4), inhibitor of DNA binding 3 (ID3) and SMAD family member 6 (SMAD6), as well as in biological processes, such as extracellular matrix organization and extracellular structure organization (e.g. COL12A1, COL1A1, COL4A1 and COL5A1). The top 5 significantly enriched terms are presented in Table II.

**Analysis of regulatory networks.** Based on the regulatory network between TFs and target genes, 7 TFs targeting 30 genes were screened from the upregulated genes, including FOS-like antigen 1 (FOSL), forkhead box A2 (FOXA2) and peroxisome proliferator-activated receptor \( \gamma \) (PPARG). In addition, 3 TFs targeting 19 genes were obtained from the downregulated genes, including jun D proto-oncogene (JUND). Notably, PPARG and JUND had a regulatory effect on the expression of downstream spermidine/spermine N1-acetyltransferase 1 (SAT1) (Fig. 1).

The regulatory network between miRNA and DEGs consisted of 26 upregulated genes, 32 downregulated genes
Table II. Significantly enriched terms of downregulated genes.

### A, Pathways

<table>
<thead>
<tr>
<th>Enriched term</th>
<th>P-value</th>
<th>Count</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular matrix organization [REACT_118779]</td>
<td>3.29x10^{-5}</td>
<td>13</td>
<td>BMP4, CASK, COL12A1, COL1A1, COL4A1, COL5A1, FBLN1, FBN2, PLOD2, SDC4, SERPINH1, THBS1, TIMP2</td>
</tr>
<tr>
<td>Collagen biosynthesis and modifying enzymes [REACT_121139]</td>
<td>1.77x10^{-4}</td>
<td>6</td>
<td>COL12A1, COL1A1, COL4A1, COL5A1, PLOD2, SERPINH1</td>
</tr>
<tr>
<td>Syndecan-1-mediated signaling events [syndecan_1_pathway]</td>
<td>3.12x10^{-4}</td>
<td>5</td>
<td>CASK, COL12A1, COL1A1, COL4A1, COL5A1</td>
</tr>
<tr>
<td>TGF-β signaling pathway [hsa04350]</td>
<td>5.99x10^{-4}</td>
<td>6</td>
<td>BMP4, FST, ID3, PPP2CB, SMAD6, THBS1</td>
</tr>
<tr>
<td>Legionellosis [hsa05134]</td>
<td>7.22x10^{-3}</td>
<td>5</td>
<td>CXCL8, EEF1A2, HSPA2, IL18, NFkBIA</td>
</tr>
</tbody>
</table>

### B, Gene ontology: Biological process

<table>
<thead>
<tr>
<th>Enriched term</th>
<th>P-value</th>
<th>Count</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular matrix organization [GO:0030198]</td>
<td>4.94x10^{-6}</td>
<td>15</td>
<td>APP, BMP4, CASK, COL12A1, COL1A1, COL4A1, COL5A1, FBLN1, FBN2, GAS6, PLOD2, SDC4, SERPINH1, THBS1, TIMP2</td>
</tr>
<tr>
<td>Extracellular structure organization [GO:0043062]</td>
<td>5.15x10^{-6}</td>
<td>15</td>
<td>APP, BMP4, CASK, COL12A1, COL1A1, COL4A1, COL5A1, FBLN1, FBN2, GAS6, PLOD2, SDC4, SERPINH1, THBS1, TIMP2</td>
</tr>
<tr>
<td>Positive regulation of cellular component movement [GO:0051272]</td>
<td>6.48x10^{-6}</td>
<td>13</td>
<td>BMP4, COL12A1, CXCL16, CXCL8, DAB2, GAS6, GLIPLR2, GPER1, HSPA5, PDPN, SCARB1, SEMA4D, THBS1</td>
</tr>
<tr>
<td>Multicellular organism development [GO:0007275]</td>
<td>1.55x10^{-5}</td>
<td>54</td>
<td>ABLIM1, APP, ARID5B, AXIN2, BASP1, BMP4, CACNA1H, CX2, COL12A1, COL1A1, COL4A1, COL5A1, CRBP2, CTGF, CXCL8, DAB2, DYSPL2, DUSP1, EBP, EPB41L3, EYA2, FBN2, FES, FLOT2, FST, GAS6, GLIPLR2, GPER1, HOXC6, HSPA5, ID3, IGSF3, IL18, JUP, MDK, NDN, NINJ1, NPY, PDPN, PIM1, PLXNB2, PQBP1, PRKCH, RAS, SEMA4D, SFN, SHROOM3, SMAD6, TAGLN, THBS1, TPM1, TRO, TUBB2B, VASN</td>
</tr>
<tr>
<td>Regulation of cellular component movement [GO:0051270]</td>
<td>1.98x10^{-5}</td>
<td>18</td>
<td>BMP4, COL1A1, CXCL16, CXCL8, DAB2, FES, GAS6, GLIPLR2, GPER1, HSPA5, JUP, PDPN, PLXNB2, RCC2, SCARB1, SEMA4D, THBS1, TPM1</td>
</tr>
</tbody>
</table>

### C, Disease ontology terms

<table>
<thead>
<tr>
<th>Enriched term</th>
<th>P-value</th>
<th>Count</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryoma [DOID:4766]</td>
<td>4.67x10^{-5}</td>
<td>22</td>
<td>AMPH, BMP4, CALD1, CHD4, CTGF, CXCL8, DAB2, E2F2, GSTT1, HSPA5, IL18, KRT8, MDK, NUA1, PDPN, RBP1, SEMA4D, SHC1, SKP2, TIMP2, TK1, TRO</td>
</tr>
<tr>
<td>Embryonal cancer [DOID:688]</td>
<td>5.20x10^{-5}</td>
<td>22</td>
<td>AMPH, BMP4, CALD1, CHD4, CTGF, CXCL8, DAB2, E2F2, GSTT1, HSPA5, IL18, KRT8, MDK, NUA1, PDPN, RBP1, SEMA4D, SHC1, SKP2, TIMP2, TK1, TRO</td>
</tr>
</tbody>
</table>
| Cell type cancer [DOID:0050687]      | 6.41x10^{-5} | 33    | AMPH, APP, BMP4, CALD1, CCNA2, CHD4, CTGF, CXCL8, DAB2, E2F2, FES,
and 84 miRNAs (Fig. 2). From that network, two miRNAs with a high degree were further investigated. miR-21-5p had a regulatory effect on 5 DEGs, including tropomyosin 1 (TPM1), EGFR and TGFBR2; and miR-155-5p regulated 3 DEGs, including E2F transcription factor 2 (E2F2).

**Discussion**

The present study presented similar research type to a related study (28). The similarities between the two studies were that the present study used the same microarray data of GSE49003

Table II. Continued.

<table>
<thead>
<tr>
<th>C, Disease ontology terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enriched term</td>
</tr>
<tr>
<td>Reproductive organ cancer [DOID:193]</td>
</tr>
<tr>
<td>Liver disease [DOID:409]</td>
</tr>
</tbody>
</table>

Figure 1. Regulatory network between transcription factors screened from the differentially expressed genes and their downstream genes. Dark grey represents downregulation and light grey represents upregulation. Rhombus represents transcription factors and rectangle represents target genes.
to identify the DEGs in the metastatic osteosarcoma samples and analyze the DEGs enriched function and pathways, and the miRNA-target network. Additionally, different results were obtained which may be due to the different analytical tools used. The present study identified 248 upregulated and 208 downregulated DEGs in metastatic cell lines compared with in non-metastatic cell lines by analyzing a dataset deposited in the GEO, in order to determine the molecular mechanisms underlying osteosarcoma metastasis. The dysregulated genes were enriched in TGF-β signaling, downregulation of TGF-β receptor signaling and TGF-β receptor signaling activates SMADs pathways, as well as in the lung cancer disease ontology term. A cluster of TFs and miRNAs were also identified to be associated with the DEGs.

Pathway enrichment analysis revealed that the upregulated \textit{PPPIR15A}, \textit{TGFBR2} and \textit{UCHL5} genes were significantly associated with the downregulation of TGF-β receptor signaling and TGF-β receptor signaling activates SMADs pathways, whereas the downregulated \textit{BMP4}, \textit{ID3} and \textit{SMAD6} were associated with the TGF-β signaling pathway. Altered TGF-β expression was previously observed in patients with metastatic osteosarcoma compared with in patients without metastasis (29). In addition, TGF-β may increase the metastatic potential of human osteosarcoma cells by triggering the malignant phenotype presentation of hyaluronan and versican (30). Overexpressed \textit{SMAD7} may have an inhibitory effect on the TGF-β/SMAD signaling pathway and subsequently protect against tumor metastasis of osteosarcoma (31). \textit{PPPIR15A} and \textit{UCHL5} have previously been implicated in the control of TGF-β signaling via interaction with \textit{SMAD7} (32,33). Furthermore, the TGF-β growth factor family member \textit{BMP4} may suppress breast cancer metastasis (34). \textit{ID} genes are the downstream targets of the TGF-β pathway, which contribute to cell migration (35). \textit{SMAD6}, instead of \textit{SMAD7}, may negatively regulate TGF-β-induced activation of the TRAF6-TAK1-P38 MAPK/JNK pathway (36). Therefore, it is possible that TGF-β and TGF-β receptor signaling may contribute to osteosarcoma metabolism.
metastasis, which may be mediated by the dysregulated genes enriched in these pathways.

Disease ontology enrichment analysis revealed that the upregulated DEGs EGFR, IGF2BP3, RUNX3 and SFRP1 were associated with various diseases, including lung cancer. Osteosarcoma has the potential to metastasize to the lungs (3). As previously reported, EGFR may be associated with osteosarcoma metastasis to the lungs (37). In addition, IGF2BP3 may promote cell invasiveness and tumor metastasis of pancreatic cancer (38). Therefore, the upregulation of EGFR and IGF2BP3 observed in the metastatic cells may indicate the metastatic potential of osteosarcoma cells to the lung. Conversely, low expression of RUNX3 may contribute to lymph node metastasis and invasion in pancreatic carcinoma tissues (39). SFRP1 may inhibit TGF-β-induced epithelial to mesenchymal transition phenotype and thus inhibit metastasis (40). Therefore, it is possible that the upregulation of RUNX3 and SFRP1 may exhibit antitumor activity and have an opposite effect on the metastasis of osteosarcoma to the lungs. Therefore, it is possible that those four genes may be involved in the metastasis of osteosarcoma to lung cancer. In addition, the downregulated DEGs COL12A1, COL1A1, COL4A1 and COL5A1 were associated with the extracellular collagen biosynthesis pathway. In accordance with the enriched pathway, those four genes were also enriched in biological processes associated with extracellular matrix structure and organization. Increased collagen degradation may promote the formation of primary osteosarcoma tumors and metastasis to the lungs (41). The downregulation of the four genes responsible for collagen synthesis may imply an impaired extracellular matrix, which may contribute to tumor invasion in an indirect manner.

A total of 10 TFs were screened from the DEGs, including upregulated FOSL1, FOXA2 and PPARG, and downregulated JUND. Based on the regulatory network of TFs, SAT1 was targeted by TFs PPARG and JUND. SAT1 deletion may lead to a significant reduction in BRCA1, DNA repair associated expression (42), which may have a directly suppressive effect on tumor metastasis-associated epithelial to mesenchymal transition (43). It is possible that upregulated SAT1 may be associated with osteosarcoma metastasis. Activation of the upstream PPARG may exert pro-tumorigenic effects on cancer progression and metastasis in myeloid cells (44); therefore, PPARG may be involved in the metastasis of osteosarcoma through regulation of SAT1 expression. The upstream JUND inhibition may suppress the migration, invasion and metastasis of osteosarcoma (45). Therefore, it is possible that upregulated PPARG and downregulated JUND may exert opposite effects on osteosarcoma metastasis via competitively regulating the expression of SAT1. In addition, upregulation of the FOSL1 oncogene is involved in breast cancer migration and invasion (46). FOXA2 may function as a suppressor of lung cancer metastasis by inhibiting TGF-β-induced epithelial to mesenchymal transition (47). The dysregulation of these 10 TFs may suggest an involvement of TFs in metastasis.

The miRNA signature in osteosarcoma has been extensively investigated using global microarray analyses of miRNAs and mRNAs in osteosarcoma cell lines (48). Specific miRNAs targeting DEGs involved in metastasis were identified in the present study. Notably, two miRNAs with a high degree among these miRNAs were detected. miR-21-5p had a regulatory effect on TPM1, EGFR and TGFBR2, whereas miR-155-5p was able to regulate E2F2. In a previous study, miR-155-5p was predicted to be associated with the metastatic capacity of osteosarcoma (49). It has previously been reported that E2F2 has a key role in mediating tumor metastasis of breast cancer (50). Therefore, miR-155-5p may be involved in osteosarcoma metastasis via targeting E2F2. In addition, miR-21-5p may contribute to osteosarcoma metastasis via its target genes. EGFR has been demonstrated to be associated with metastasis of osteosarcoma to the lungs (37), whereas TGFBR2 has been associated with metastasis of gastric cancer cells (51). In addition, suppressive TPM1 may alter TGF-β tumor suppressor function and thus promote metastasis of tumor cells (52). Therefore, it is possible that miR-21-5p exerts a regulatory effect on metastasis via regulation of the expression of its target genes.

In conclusion, upregulated PPIA, TGFBR2 and UCHL5, which are enriched in the downregulation of TGF-β receptor signaling and TGF-β receptor signaling activates SMADs pathways, may contribute to the progression of osteosarcoma metastasis. In addition, the downregulated DEGs BMP4, ID3 and SMAD6, which are enriched in the TGF-β signaling pathway, may also be involved in osteosarcoma metastasis. EGFR, IGF2BP3, RUNX3 and SFRP1 were associated with metastasis to lung cancer. Furthermore, 10 TFs screened from DEGs, and various miRNAs (e.g. miR-21-5p), may be associated with metastasis via their target genes. The full understanding of the complex regulatory network of DEGs associated with metastatic osteosarcoma may aid in improving the production of novel metastasis-targeted therapeutic strategies.

References


Heng et al. MOLECULAR CHARACTERIZATION OF OSTEOSARCOMA METASTASIS


