Identification of six potential genes associated with the metastasis of osteosarcoma based on gene expression profile

Accepted 17th December, 2018

ABSTRACT

The purpose of our study was to explore the potential key genes related to the metastasis of osteosarcoma(OS) and provide utility biological information for further research of OS. The gene expression profile from different Sequencing platform including GPL570 (GSE37552,GSE85537,GSE14827), GPL96(GSE14359), GPL3307(GSE32981) and GPL6076(GSE9508), were downloaded from Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) from each platform were selected with R-studio software. Further, the functional enrichment analyses of DEGs and pathway enrichment analyses were conducted with the online analysis software-DAVID database. Thereafter, the protein–protein interaction (PPI) network was conducted based on the STRING database. Finally, we obtained the repeated gene whose frequency was more than twice. A total of 94 DEGs, such as SPP1, CXCL2, SERPINA1, ABCC3 and PIAS2, were screened out. Our study offered the research of general molecular mechanisms about metastasis and selected a series of repeated genes from different sequencing platform, which have the potential as a new targets for the treatment of OS. However, the article is still needed to be confirmed by experiments.

key words: Osteosarcoma, differentially expressed gene, enrichment analysis, network.

Abbreviations: OS, Osteosarcoma; DEGs, differentially expressed genes; PPI, protein–protein interaction; GO, gene oncology; KEGG, Kyoto encyclopedia of genes and genomes pathway; CC, cellular component; BP, biological process; MF, molecular function.

INTRODUCTION

Osteosarcoma is the most common primary tumor of bone in children and adolescents, which constitutes nearly 1% of newly diagnosed malignant neoplasm annually worldwide (Siege et al., 2014). At present, one of the most crucial clinical problems faced with is the early metastasis of OS to other part. Currently, despite the multimodal treatment of OS such as surgery and multi-agent chemotherapy, the 5-year overall survival rate of patients is increased from 60–70% (Friebele et al.,2015). However, the 5-year survival rates still keeps below 30%, once the patients present recurrence or metastasis (Smith et al., 2010). In the past several years, a variety of study on relevant gene and pathway to metastasis of OS has been conducted (Li et al., 2017; Guo et al., 2014; Pan et al.,2014). However, because of the unclear pathogenesis, there is lack of investigations on treatments, leading to research being at an early stage and as such, further research is needed. In fact, the majority of the studies cost lots of time on different molecular target of gene, ignoring that the metastasis involves abnormal expressions of multiple genes. The high metastasis rate and death rate have indicated that traditional therapeutic methods targeting different certain single gene are
inadequate to solve the problem of metastasis of OS. Thus, it is imperative to explore the metastasis-related multiple gene variations with genome-wide technologies, which is more precise to find out a novel therapeutic strategies.

Microarray is a high-throughput tool for efficiently performing global gene expression profiles, which has been widely used for exploring the mechanisms of etiology. Using the method of gene expression profile, these authors have identified the differentially expressed genes (DEGs) with two groups (metastasis and non-metastasis) with different platforms (Flores et al., 2012; Zhang and Zhu, 2010; Kobayashi et al., 2010; Raphaela and Guenther, 2010; Namlos et al., 2012; Endo-Munoz et al., 2010). However, the relevant molecular mechanisms of metastasis of OS is not clear. To better understand the mechanisms of metastasis and verify the statistical result, it is indispensable to carry out further research.

In the present study, depending on the expression profile data deposited, a further bioinformatic analysis was performed to identify DEGs from acquired metastasis and non-metastasis osteosarcoma samples. After screening of the DEGs, the functions of DEGs were further assessed by gene ontology (GO) annotation, KEGG pathway enrichment and protein–protein interaction network construction, which can provide valid biological information for further investigation.

METHODS

Identification of differentially expressed genes (DEGs) from public microarray data

To obtain DEGs in acquired samples, we downloaded the public gene expression profile from the gene expression omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) through diverse sequencing platform from different company. The Mesh term we used was metastasis and osteosarcoma. Thereafter, the organism from human was included. The data were deposited by researchers using various platform, containing differentially genes about metastasis and non-metastasis sample. The RMA-normalized data sets as text files were downloaded and the packages including LIMMA, affy, impute, xlsx, futile.logger were installed. Finally, the GPL570 (GSE37552, GSE85537, GSE14827), GPL96 (GSE14359), GPL3307(GSE32981), GPL6076(GSE9508) were included in our study. The included criterion: 1) Because the sample in the present study is insufficient, we set the threshold of fold change ≥2 and P value<0.05; 2) Removed the batch error between GSE37552, GSE85537 and GSE14827, because they were from the same platform-GPL570; 3) Obtained the maximize fold change when there was a repeated probe; 4) It was removed when the same probe corresponds to multiple genes; 5) Removed a probe that had no corresponding gene, because some probe only play an role in controlling the process of sequencing. 6) The adjust method-BH was employed. These data was analyzed with Rstudio software.

Functional enrichment analysis of DEGs

Database for annotation, visualization and integrated discovery (DAVID) (Huang et al., 2009), is a tool for function enrichment. Therefore, functional enrichment analyses of the DEGs, including gene ontology (GO) function analysis and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis, were conducted. In the GO analysis, the cellular component (CC), biological process (BP) and molecular function (MF) terms were included and P value <0.05 was regarded as statistically significance. In the KEGG pathways analysis, enriched pathway was employed according to the hypergeometric distribution with a P value <0.1

Protein–protein interaction network construction by STRING

Through the included official gene symbol of DEGs, these 94DEGs were analyzed using STRING to predict the interaction among them with score of not <0.4 (medium confidence score) and three important clusters were shown in PPI network. The hub gene can also be discovered.

RESULTS

Identification of DEGs between metastasis and non-metastasis osteosarcoma sample

The analysis of DEGs was conducted as shown in Figure 1. Based on the public microarray dataset from GPL570, GPL96, GPL3307 and GPL6076, the software R-studio was used to identify the DEGs with the above criterion. Consequently, a total of 94 DEGs were selected which emerged more than twice. Among them, the five genes including SPP1-down regulated, CXCL2-up regulated, SERPINA1-up regulated, ABCC3-up regulated and PIAS2-up regulated emerged three times. The results were listed using Venn plot (Figure 2 and Table 1).

Functional annotation and pathway enrichment of DEGs

Data were collected through gene ontology (GO) analysis in DAVID. A P value of <0.05 was considered statistical significant. The enriched GO terms, which included CC, BP and MF ontologies, are listed in Figure 3 AB and C, respectively. In the enriched CC GO terms, extracellular region(24 genes), basolateral plasma membrane (5 genes) and membrane (18 genes), were related with the cell
Figure 1: The flowchart of the bioinformatics analysis.

Figure 2: The Venn plot to show the intersection among various sequencing platform.
Table 1: The six crucial genes in various platform selected from 94 DEGs.

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>P.Value</th>
<th>logFC</th>
<th>Gene symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPL3307</td>
<td>0.0004</td>
<td>2.4795472</td>
<td>ABCC3</td>
</tr>
<tr>
<td>109767</td>
<td>0.0211</td>
<td>3.0260525</td>
<td>CXCL2</td>
</tr>
<tr>
<td>173408</td>
<td>0.0109</td>
<td>2.6905377</td>
<td>SERPINA1</td>
</tr>
<tr>
<td>GPL16076</td>
<td>0.0058</td>
<td>2.3981070</td>
<td>PIAS2</td>
</tr>
<tr>
<td>33519</td>
<td>0.0063</td>
<td>-2.3239240</td>
<td>SPP1</td>
</tr>
<tr>
<td>545</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPL96</td>
<td>0.0161</td>
<td>2.7007150</td>
<td>ABCC3</td>
</tr>
<tr>
<td>208161_s_at</td>
<td>0.008</td>
<td>2.3097815</td>
<td>CXCL2</td>
</tr>
<tr>
<td>209774_x_at</td>
<td>0.0000</td>
<td>3.6693861</td>
<td>PIAS2</td>
</tr>
<tr>
<td>214593_at</td>
<td>0.0139</td>
<td>-2.0692217</td>
<td>SPP1</td>
</tr>
<tr>
<td>209875_s_at</td>
<td>0.0000</td>
<td>3.9138315</td>
<td>SERPINA1</td>
</tr>
<tr>
<td>209189_at</td>
<td>0.0001</td>
<td>2.6982569</td>
<td>FOS</td>
</tr>
<tr>
<td>GPL570</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>209641_s_at</td>
<td>0.0161</td>
<td>2.7007150</td>
<td>ABCC3</td>
</tr>
<tr>
<td>209774_x_at</td>
<td>0.0228</td>
<td>3.0620995</td>
<td>CXCL2</td>
</tr>
<tr>
<td>1556492_a_at</td>
<td>0.0229</td>
<td>2.3442426</td>
<td>PIAS2</td>
</tr>
<tr>
<td>1568574_x_at</td>
<td>0.0015</td>
<td>-2.3574018</td>
<td>SPP1</td>
</tr>
<tr>
<td>211428_at</td>
<td>0.0201</td>
<td>2.4942902</td>
<td>SERPINA1</td>
</tr>
</tbody>
</table>

membrane. In the BP ontology, the most significant GO categories were complement activation (5 genes), immune response (9 genes), cell adhesion (7 genes). While in the MF ontology, the most significant GO categories were transcriptional activator activity (6 genes), cytokine activity (5 genes), glycoprotein binding (3 genes). Furthermore, the enrichment pathways was assessed by the KEGG pathway analysis, which includes Complement and coagulation cascades pathway (4 genes), ECM-receptor interaction (4 genes), and Pathways in cancer (6 genes) (Figure 3D).

Protein–protein interaction network construction

STRING is a tool to predict protein interactions among the DEGs. Firstly, the 94 DEGs were submitted to the STRING tool to obtain PPI data. Then, the PPIs, with minimum required interaction score: medium confidence 0.400 and 3 clusters were selected to construct PPI networks. As shown In the PPI networks, the FOS-up regulated in red cluster showed a strong association with other node proteins, suggesting that it may be a hub gene. And the duplicated SPP1, CXCL2, SERPINA1 were also associated with the hub gene (Figure 4).

DISCUSSION

The metastasis is still a tough problem we confront in clinic, which is a major cause for a high mortality (Smith et al., 2010). Therefore, it is very indispensable to explore the mechanisms of metastasis and exert feasible treatment strategy for it. Using gene expression profiling by microarray technology, the crucial genes associated with metastasis could be discovered, which could be further utilized to explore novel therapeutic strategies. In this study, we identified DEGs between metastasis and non-metastasis cell lines. Thereafter, we analyzed their functions by GO annotation and pathway enrichment. Finally, their interrelationship was explored using protein–protein interaction network and the software-cytoscape was used to explore the hub genes. As a result, a series of hub genes were selected, which might play crucial roles in metastasis and might be potential targets for treatment of OS.

The LIMMA package installed in software R-studio (https://www.rstudio.com/) is a powerful statistical tool for probe-level and high-level analysis of gene expression microarrays. Through this tool, we obtained 94 DEGs, which emerged more than twice between metastasis and non-metastasis samples. Among the DEGs, SPP1, CXCL2, SERPINA1, ABCC3 and PIAS2 emerged three times. Therefore, we hypothesized that these DEGs have the potential to be biomarkers and therapy targets for preventing metastasis of OS. The protein encoded by SPP1 is involved in the attachment of osteoclasts to the mineralized bone matrix. CXCL2 is part of a chemokine superfamily that encodes secreted proteins involved in
Figure 3: A: GO enrichment of DEGs in cellular component ontology; B: GO enrichment of DEGs in biological process ontology; C: GO enrichment of DEGs in molecular function ontology; D: KEGG pathway analysis.
immunoregulatory and inflammatory processes. The protein encoded by SERPINA1 is secreted and is a serine protease inhibitor whose targets include elastase, plasmin, thrombin, trypsin, chymotrypsin, and plasminogen activator. The protein encoded by ABCC3 is a member of the superfamily of ATP-binding cassette (ABC) transporters. PIAS2 encodes a member of the protein inhibitor of activated STAT family, which function as SUMO E3 ligases and play important roles in many cellular processes by mediating the sumoylation of target proteins. Studies have shown that these genes are related to the regulation and metastasis in various cancer (Li et al., 2017; Jia et al., 2016; Ye et al., 2016; Shakya et al., 2017; Liu et al., 2016; Gross et al., 2001). It should be noted that the DEGs might contribute to metastatic through complex mechanisms. Through GO annotation in DAVID database, we analyzed the biological function of the DEGs. First, in CC oncology, we found that the majority of the DEGs were enriched in extracellular region and other DEGs were enriched in the membrane. The results showed that metastasis might develop through complex cellular molecular mechanisms involved in extracellular and membrane structure. Previous research has suggested that cytosol (Yang and Wang, 2006) and membrane (Liang et al., 2015) are related to metastasis. In GO item of BP, our data showed that the most significant items were immune response and cell adhesion. It is reasonable to understand that metastasis could be interpreted as to alter the antigenicity of cancer through generation of variety of protein and change the original biological process. In response to the destruction of immune cell, cancer cells can develop by gene mutation and natural selection mechanism. In the MF portion, the most enriched oncologies are transcriptional activator activity and glycoprotein binding. In the evolution process, a series of genetic changes occurred in cancer cells, such as metastasis through different mechanisms such as generate variety protein. The DEGs might have impact on migration and invasion by participating in transcribed, cotranslational and regulation of cell proliferation. Considering the above, these results indicated that the DEGs might participate in various biological processes about the metastatic of OS. These data showed that the DEGs may influence the cell adhesion through membrane, then influence molecule mechanism and the product of protein in cancer cells. So, the DEGs might play crucial roles in metastasis through a chain of mechanisms, which is worth researching in further studies.

Pathway analysis could reveal more precise biological processes and functions of genes than GO analysis. In this study, ECM-receptor interaction pathway and TNF signaling pathway were included. These pathway have been proved relevant with cell proliferation and metastasis of tumor (Kang et al., 2015; Jacksonbernitsas et al., 2007). Through PPI network construction, a hub proteins (genes)-FOS has...
been shown in local network. The gene FOS encodes leucine zipper proteins that can dimerize with proteins of the JUN family, thereby forming the transcription factor complex AP-1. The FOS proteins have been implicated as regulators of cell proliferation, differentiation, and transformation (Barrett et al. 2016).

Taking into account what has been explored, these research supported the hypothesis that the six genes (SPP1, CXCL2, PIAS2, SERPINA1, ABCC3 and FOS) might play crucial roles in metastasis of osteosarcoma. However, it is unknown whether the selected genes can be used as targets for the treatment of OS to decline the mortality in clinic. The results of this study are superficial, and further research for the potential roles of crucial genes are needed in the future. In summary, the study provides preliminary research for the mechanisms of metastasis in OS. DEGs of metastasis and non-metastasis samples were screened out by computational bioinformatics methods. Then, the pathways of metastatic were identified. Moreover, several key DEGs were selected as potential biological targets of OS. The results of this study may give a valuable indication for both the basic research and clinical treatment of OS. However, in spite of all efforts, there are still many deficiencies, such as: 1) Random error in each sequencing platform; 2) The number of sample is insufficient; 3) The difference between every sample; 4) The random error about software. The most important is the lack of sufficient data to conduct survival analysis. Finally, lots of further investigation are still needed to confirm our hypothesis.

REFERENCES