Screening pathogenic genes in oral squamous cell carcinoma based on the mRNA expression microarray data

YANG DING1*, PENGFEI LIU2*, SHENSHENG ZHANG1, LIN TAO1 and JIANMIN HAN3

1Digestive Disease Center, Beijing Hospital of Traditional Chinese Medicine Affiliated to Capital Medical University, Beijing 100010; 2Department of Lymphoma, Sino-US Center of Lymphoma and Leukemia, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin 300060; 3Dental Materials Laboratory, National Engineering Laboratory for Digital and Material Technology of Stomatology, Peking University School and Hospital of Stomatology, Beijing 100081, P.R. China

Received June 6, 2016; Accepted January 22, 2018

DOI: 10.3892/ijmm.2018.3514

Correspondence to: Dr Jianmin Han, Dental Materials Laboratory, National Engineering Laboratory for Digital and Material Technology of Stomatology, Peking University School and Hospital of Stomatology, Zhong-Guan-Cun South Avenue 22, Hai Dian, Beijing 100081, P.R. China
E-mail: hanjianmin@bjmu.edu.cn

*Contributed equally

Key words: oral squamous cell carcinoma, the pathogenic gene, the miRNA-gene regulated network, mRNA expression microarray data

Abstract. Oral squamous cell carcinoma (OSCC) is one of the most common malignancies and its survival rate has barely improved over the past few decades. The purpose of this study was to screen pathogenic genes of OSCC via microarray analysis. The mRNA expression microarray datasets (GSE2280 and GSE3524) were downloaded from the Gene Expression Omnibus (GEO) database. In GSE2280, there were 22 OSCC samples without metastasis and 5 OSCC samples with lymph node metastasis. In GSE3524, there were 16 OSCC samples and 4 normal tissue samples. The differentially expressed genes (DEGs) in OSCC samples with lymph node metastasis compared with those without metastasis (named as DEGs-1), and the DEGs in OSCC samples compared with normal tissue samples (named as DEGs-2), were obtained via limma package. The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to perform the functional enrichment analyses of DEGs-1 and DEGs-2. The miRNA-gene pairs of overlaps among DEGs were screened out with the TargetScan database, and the miRNA-gene regulated network was constructed by Cytoscape software. A total of 233 and 410 DEGs were identified in the sets of DEGs-1 and DEGs-2, respectively. DEGs-1 were enriched in 188 Gene Ontology (GO) terms and 8 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and DEGs-2 were enriched in 228 GO terms and 6 KEGG pathways. In total, 126 nodes and 135 regulated pairs were involved in the miRNA-gene regulated network. Our study indicated that transglutaminase 2 (TGM2) and Islet 1 (ISL1) may be biomarkers of OSCC and their metastases. Moreover, it provided some potential pathogenic genes (e.g. P2RY2 and RAPGEL1) in OSCC.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is one of the leading cancer types by incidence worldwide, with ~500,000 new cases each year worldwide and a five-year survival rate of 40-50% (1). Oral squamous cell carcinoma (OSCC) is the most prevalent malignancy in oral cavity and ranks sixth among the most common cancers worldwide (2,3). Furthermore, OSCC is prevalent particularly in developing countries, such as Indian subcontinent, and mainly a problem of older men, accounting for 90% in the over 45 year-old group (4). With characteristics of rapid progression and worse outcome, OSCC is a deadly and particularly risky because it progresses without producing pain or symptoms that may be readily recognized by the patient in its early stages (5). It is usually discovered when the cancer has metastasized to the lymph nodes of the neck (6). The etiology of OSCC has not yet been well illustrated, and some risk factors may be associated with it. Tobacco and alcohol consumption are the most important risk factors, and tobacco smoking and alcohol intake have a strong interactive effect on the risk of OSCC (7,8). Other factors in OSCC include dietary factors, immunodeficiency and viral infections such as chronic candidosis and herpes simplex virus (8-10). Besides, the mutagen sensitivity is related to the progression of OSCC (11-13). From relative risk factors, it has been estimated that 75% of all oral cancers are preventable. However, in the remaining 25% of patients who are not exposed to these substances, the causes of their tumors remain unknown (14). In this study, the gene expression microarray data of OSCC samples both with lymph nodes metastasis and without metastasis were investigated via microarray analysis, in order to screen some potential pathogenic genes of OSCC and provide some clues for the diagnose and treatment.
Materials and methods

**mRNA expression microarray data.** The mRNA expression microarray datasets GSE2280 (15) and GSE3524 (16) were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. The microarray dataset GSE2280 contained 22 OSCC samples without metastasis and 5 OSCC samples with lymph node metastasis. In GSE3524, there were 16 OSCC samples and 4 normal tissue samples. The former was detected with GPL96 [HG-U133A] Affymetrix Human Genome U133A array platform, and the latter with GPL96 [HG-U133A] Affymetrix Human Genome U133A array platform.

**Data pre-processing and identification of differentially expressed genes.** The original data were converted into the recognizable format by R, and the Robust Multi Array (RMA) of the affy (17) package was used for the background correction and normalization. After the data pre-processing, the differentially expressed genes (DEGs) in OSCC samples with lymph node metastasis compared with those without metastasis (named as DEGs-1), regarding DEGs in IOSCC samples compared with normal tissue samples (named as DEGs-2), were selected out via the limma (18) package of R according to the criteria: P-value <0.05 and |log(fold2change)| >1. Besides, the two-way cluster analysis of the 2 sets DEGs was conducted via the gplots package in R, and their overlapped genes were found.

**Functional enrichment analysis.** Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs-1 and DEGs-2 were performed via Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/) (19). The GO terms and KEGG pathways with P<0.05 were screened out.

**Construction of the miRNA-gene regulated network.** The known and predictable miRNA regulating the overlapped genes were selected via the TargetScan (20) database, and afterwards, the miRNA-gene regulated pairs were obtained. Ultimately, the miRNA-gene regulated network was constructed and visualized by Cytoscape (21) software. The nodes were screened out in the network when the degree of node attributes was ≥1, and ‘degree’ represented the connections with other nodes.

**Results**

**DEGs.** A total of 233 DEGs (133 up- and 100 downregulated) were identified in the sets of DEGs-1, and 410 (99 up- and 313 downregulated) in the sets of DEGs-2. The two-way cluster graph is shown in Fig. 1. Fourteen overlapped genes of the 2 set DEGs were found, and the heatmap of the overlapped genes is shown in Fig. 2.

**GO terms and KEGG pathways.** DEGs-1 were enriched in 188 GO terms and 8 KEGG pathways, and the top 10 GO terms and all the KEGG pathways are shown in Tables IA and IIA. DEGs-2 were enriched in 228 GO terms and 6 KEGG pathways, and the top 10 GO terms and all the KEGG pathways are shown in Tables IB and IIB.
Table I. The top 10 GO terms of DEGs-1 and DEGs-2.

A. The top 10 GO terms of DEGs-1

<table>
<thead>
<tr>
<th>Category</th>
<th>GO ID</th>
<th>GO name</th>
<th>Gene no.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>GO:0043292</td>
<td>Contractile fiber</td>
<td>25</td>
<td>1.57E-21</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0030016</td>
<td>Myofibril</td>
<td>24</td>
<td>3.94E-21</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0030017</td>
<td>Sarcomere</td>
<td>22</td>
<td>1.06E-19</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0044449</td>
<td>Contractile fiber part</td>
<td>23</td>
<td>1.26E-19</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0006936</td>
<td>Muscle contraction</td>
<td>22</td>
<td>7.37E-15</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0003012</td>
<td>Muscle system process</td>
<td>22</td>
<td>4.98E-14</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0006941</td>
<td>Striated muscle contraction</td>
<td>14</td>
<td>6.77E-14</td>
</tr>
<tr>
<td>MF</td>
<td>GO:0008307</td>
<td>Structural constituent of muscle</td>
<td>12</td>
<td>5.45E-12</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0015629</td>
<td>Actin cytoskeleton</td>
<td>23</td>
<td>1.54E-11</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0005865</td>
<td>Striated muscle thin filament</td>
<td>8</td>
<td>4.69E-10</td>
</tr>
</tbody>
</table>

B. The top 10 GO terms of DEGs-2

<table>
<thead>
<tr>
<th>Category</th>
<th>GO ID</th>
<th>GO name</th>
<th>Gene no.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>GO:0008544</td>
<td>Epidermis development</td>
<td>29</td>
<td>9.63E-15</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0007398</td>
<td>Ectoderm development</td>
<td>30</td>
<td>1.01E-14</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0009913</td>
<td>Epidermal cell differentiation</td>
<td>17</td>
<td>1.57E-11</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0001533</td>
<td>Cornified envelope</td>
<td>11</td>
<td>1.63E-11</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0030855</td>
<td>Epithelial cell differentiation</td>
<td>22</td>
<td>2.01E-11</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0018149</td>
<td>Peptide cross-linking</td>
<td>11</td>
<td>2.75E-10</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0030216</td>
<td>Keratinocyte differentiation</td>
<td>15</td>
<td>5.79E-10</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0005792</td>
<td>Microsome</td>
<td>23</td>
<td>2.31E-08</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0042598</td>
<td>Vesicular fraction</td>
<td>23</td>
<td>3.91E-08</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0060429</td>
<td>Epithelium development</td>
<td>23</td>
<td>4.65E-08</td>
</tr>
</tbody>
</table>

GO, Gene Ontology; DEGs, differentially expressed genes; BP, biological process; CC, cellular component; MF, molecular foundation.

Figure 2. (A and B) The heatmap graph of the overlapped genes.
The miRNA-gene regulated network. In total, 116 miRNAs regulating the overlapped genes were screened out, and then 135 miRNA-gene regulated pairs were obtained. Ultimately, the miRNA-gene regulated network was constructed and is shown Fig. 3. The network of 126 nodes were selected, and the top 20 are listed in Table III.

Discussion

Two sets of DEGs were identified in this study, namely DEGs in OSCC samples with lymph node metastasis compared with those without (DEGs-1), and DEGs in OSCC samples compared with normal tissue samples (DEGs-2). The two-way cluster analysis was performed, and it was obvious that only one OSCC sample with metastasis gathered in the OSCC samples without metastasis (Fig. 1A), and none of OSCC samples gathered in normal tissue samples (Fig. 1B). The result indicated that the identified DEGs, both DEGs-1 and DEGs-2, were comparatively accurate. Furthermore, 14 overlapped genes were obtained after comparison of the 2 sets of DEGs. Fig. 2 shows that TGM2 was overexpressed not only in OSCC samples but also in OSCC samples with lymph node metastasis, while ISL1 expression was low. TGM2 encoded TGM2, which was the most diverse and ubiquitously expressed member of the oncostatin-M receptor (OSMR) family. It was reported that OSMR is directly affected by the increasing of cell migration and invasiveness (22). TGM2 is a multifunctional protein and has both enzymatic and non-enzymatic functions. It was closely related to its subcellular location and depended on the pathophysiological context (23). TGM2 was overexpressed in a range of cancer types, where it was associated with metastasis and decreased overall patient survival (24,25). Miyoshi et al (26) confirmed that TGM2 was a novel marker for prognosis and therapeutic target in colorectal cancer. Besides, ISL1 encoded ISL1, a LIM-homeodomain transcription factor, which was essential for promoting pancreatic islets proliferation and maintaining endocrine cells survival in embryonic and postnatal pancreatic islets (27). In 2008, Cheung et al (28) explored biomarkers of neuroblastoma via microarray analysis and found that ISL1 was overexpressed in stage IV, which was related to the overall
survival rate and the degree of tumor progression. Another study reported that ISL1 was a reliable marker of pancreatic endocrine tumors and metastases thereof (29). Thus, it was indicated that \textit{TGM2} and \textit{ISL1} may be biomarkers of OSCC and their metastases.

In this study, DEGs-1 and DEGs-2 were enriched in only 8 and 6 KEGG pathways (Tables IIA and IIB) respectively, which was a small amount and convenient to experimental study. DEGs of OSCC samples with lymph metastasis were mainly enriched in cardiomyopathy-related pathways (such as viral myocarditis, hypertrophic cardiomyopathy and dilated cardiomyopathy) and immune-related pathways (such as B cell receptor signaling pathway, leukocyte transendothelial migration and primary immunodeficiency). Nevertheless, DEGs of OSCC samples compared with normal tissue samples were all enriched in drug metabolism or other metabolic processes of organic compounds (e.g. retinol metabolism, arachidonic acid metabolism, linoleic acid metabolism and metabolism of xenobiotics by cytochrome p450). A report verified that it was similar in patients between with lung squamous cell carcinoma and dilated cardiomyopathy (30). Besides, immunodeficiency and other immune reactions were critical in the occurrence and development of tumors. Although more explorations are necessary to excavate relationships of these pathways and OSCC, it was suspected that these cardiomyopathy or immune related pathways may be associated with the metastasis of OSCC. Similarly, these metabolic processes may be related to the emergence of OSCC.

\textit{RAPGEFL1} and \textit{P2RY2} were the top two nodes with the highest degree in the miRNA-gene regulated network. In 2013, Takahashi et al (31) reported that \textit{RAPGEFL1} was highly methylated in some esophageal squamous cell carcinoma (ESCC) cell lines and it could be used to estimate the fraction of cancer cells in tumor DNA. However, another study screened aberrant methylation profile in ESCC, and results showed that \textit{RAPGEFL1} was not involved in any biological processes (32). In this study, we found that \textit{RAPGEFL1} was not enriched in any GO terms or KEGG pathways, but it could be regulated by most miRNAs (Fig. 3). \textit{P2RY2} was a member of purinergic receptors (P2-receptors), which is considered associating with both growth inhibition and programmed cell death (33-35). Besides, extracellular ATP could inhibit growth and induced apoptosis of various tumors by activating specific P2-receptors (36-38). P2Y2-receptors were considered as promising target proteins for innovative approaches in esopha-

![Figure 3. The miRNA-gene regulated network for the overlapped genes.](image-url)
geal cancer therapy (39). Therefore, RAPGEFL1 and P2RY2 may be the potential pathogenic genes for OSCC.

In conclusion, this study indicated that TGM2 and ISL1 may be the biomarkers of OSCC and their metastases. Moreover, it also provided some other potential pathogenic genes (e.g. P2RY2 and RAPGEFL1) in OSCC.

Acknowledgements

Not applicable.

Funding

This study was supported by the Beijing Natural Science Foundation (no. 7164265) and the National Natural Science Foundation (no. 81400560).

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors’ contributions

JH designed the experiments. YD and PL performed data analysis. YD and SZ wrote the main manuscript text and prepared all the figures. JH and LT discussed the results and revised the manuscript. All authors contributed to discussions regarding the results and the manuscript. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References