Regulation network analysis in the esophageal squamous cell carcinoma

J. Xiang, W. Zang*, J. Che, K. Chen, J. Hang

Department of Thoracic Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
*Department of Biology, Shanghai Jiaotong University, Shanghai, China

Abstract. – BACKGROUND: The incidence of esophageal squamous cell carcinoma (ESCC) has high regional selectivity. The molecular mechanisms of ESCC are complex and involve multiple oncogenes, tumor suppressor genes, receptor tyrosine kinases, cytoplasmic enzymes, and tumor interstitial elements.

AIM: Here we used bioinformatics to obtain some important genes and pathways involved in ESCC.

MATERIALS AND METHODS: In this article, we did Affymetrix microarray data collection from three big databases, and then selected all the differentially expressed genes (DEGs) according to some principles. On this basis, we carried out regulation network analysis and pathway enrichment analysis, obtaining ESCC related regulation network analysis, after which we selected significant pathways on regulation network and established TF-pathway regulation network.

RESULTS: In the transcription factors (TFs) regulation network we found SP1, E2F1, USF2 and SP3 form a local network which suggested that these TFs might play a more important role in ESCC. Some key pathways were also identified, such as P53 signaling pathway, melanoma and prostate cancer pathways.

CONCLUSIONS: The identification of crucial molecular pathways involved in ESCC would ultimately improve therapeutic effects and facilitate the development of new treatment strategies.

Key Words: Regulation network, Esophageal squamous cell carcinoma, Pathway enrichment analysis.

Introduction

Esophageal squamous cell carcinoma (ESCC), one of the six common malignant tumors, has high incidence in China1–2, especially in southern areas of Taihang Mountains on the borders of Henan, Shanxi and Hubei provinces, its morbidity and fatality rank first in the world3. All of this shows that this type of carcinoma is affected by environmental factors. Now, it is universally acknowledged that smoking and alcohol intake are mainly responsible for ESCC tumorigenesis, which causes death due to cancer deep invasion and metastasis4. However, the tumorigenesis of ESCC is a gradual process with multiple steps, and effective treatment relies on early diagnosis and therapy5. Although big progress has been made in detection and treatment, the five-year survival rate of this cancer still remains quite low right now. Related research focuses on the pathogenesis, aiming at to look for molecular markers that will be used in early diagnosis, as well as gene targets for therapy6–7. Recently, plenty of gene mutation sites have been found in cancer cells of ESCC, such as epidermal growth factor receptor (EGFR) and p53 mutations, which promoted the proliferation and infiltration of cancer cells by microenvironment changes8. Meanwhile, some reports proved that hepatocyte growth factor (HGF) can boost cancer cell infiltration while cancer-associated fibroblasts (CAF s) could affect tumorigenesis in extracellular matrices. Beside, several genes have been demonstrated to associate with bad prognosis of ESCC. For example, high expression of CC chemokine receptor 7 (CCR7)9, cyclooxygenase-2 (COX2)10, Beclin-111, Toll-like receptor 9 (TLR9)12, and FOXC213 were all significantly correlated with ESCC invasion, stage, tumor depth, lymph node metastasis and poor survival rates. Low claudin-4 expression was found to be significantly associated with histological differentiation, invasion depth and lymph node metastasis. But unfavorable influences were present on disease-free survival and overall survival14. Phosphatase regenerating liver-1 (PRL-1) protein expression closely correlated with the stage of ESCC, 79.4% in stage III ESCC while 33.3% in stage I ESCC15.

These factors all take great important roles in the regulation network of ESCC, through activating some signal pathways in ESCC or inhibiting them. While, multiple molecular pathways are in-
volved in this process and interact with each other. DNA microarray analysis as a global approach is applied to investigate physiological mechanisms in health and disease. Many researches and analysis have been done on the transcription factors (TF) and pathways in ESCC, to find out some potential diagnostic and therapeutic target for this disease and determine the specific relationship between some TF and pathway. It was demonstrated that knockdown of protein kinase C iota (PKCι) in suspended ESCC cells caused a decrease in S-phase kinase-associated protein 2 (SKP2) that had been reported to promote resistance to anoikis via the PI3K/AKT (phosphatidylinositol 3 kinase/protein kinase B) pathway. However, much more details still remained unclear.

The molecular mechanisms of ESCC are complex and involve multiple oncogenes, tumor suppressor genes, receptor tyrosine kinases, cytoplasmic enzymes, and tumor interstitial elements. Better theoretical foundation will be supplied to early diagnosis and targeted therapy of ESCC by research and exploration of the relationship of these pathways and transcription factors. In this research, through the regulation network analysis and pathway enrichment analysis, we try to find potential candidate genes associated with ESCC and explore their potential importance in the future.

**Data and Methods**

**Affymetrix Microarray Data**

The transcription profile of GSE17351 was deposited with Gene Expression Omnibus at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo) which is based on the Affymetrix Human Genome U133A Plus 2.0 Array. Esophageal tissues were procured via surgery at the Hospital of the Uni-

![Figure 1](image.png)

**Figure 1.** Regulation network in ESCC. The yellow triangles stand for TF and the pink ellipses stand for target genes while the red arrows represent for activation and the blue arrows represent for repression.
versity of Pennsylvania through the Cooperative Human Tissue Network. All were pathologically diagnosed as ESCC. Five paired tumors and adjacent normal tissues, including 10 chips.

**Differentially Expressed Genes (DEGs) Selection**

For the GSE17351 dataset, the limma method was used to identify DEGs. The original expression datasets at all conditions were extracted into expression estimation, and then the linear model construction. Only DEGs with a fold change value > 2 and \( p < 0.05 \) were selected.

To demonstrate the potential regulation relationships, the Pearson Correlation Coefficient (PCC) was calculated for all pair-wise comparisons of gene-expression values between TFs and the DEGs. The relationships whose absolute PCC are larger than 0.6 were considered as significant.

**Regulation Network Analysis**

Transcriptional Regulatory Element Database (TRED) has been built in response to the increasing needs of an integrated repository for both cis- and trans- regulatory elements in mammals. TRED has done the curation for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. The curation is currently focusing on target genes of 36 cancer-related TF families. In total 774 pairs of regulatory relationship between 219 transcription factors (TFs) and 265 target genes were collected from TRANSFAC (http://www.gene-regulation.com/pub/databases.html). Total 5722 pairs of regulatory relationship between 102 transcription factors (TFs) and 2920 target genes were collected from TRED (http://rulai.cshl.edu/TRED/).

Combining the two regulation datasets together, total 6328 regulatory relationships between 276 TFs and 3002 target genes were collected.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Description</th>
<th>Count</th>
<th>( p )-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04614</td>
<td>Renin-angiotensin system</td>
<td>3</td>
<td>0.001929</td>
<td>0.116273</td>
</tr>
<tr>
<td>hsa05215</td>
<td>Prostate cancer</td>
<td>4</td>
<td>0.004763</td>
<td>0.141693</td>
</tr>
<tr>
<td>hsa05214</td>
<td>Glioma</td>
<td>3</td>
<td>0.024869</td>
<td>0.415649</td>
</tr>
<tr>
<td>hsa04115</td>
<td>p53 signaling pathway</td>
<td>3</td>
<td>0.02867</td>
<td>0.372135</td>
</tr>
<tr>
<td>hsa05218</td>
<td>Melanoma</td>
<td>3</td>
<td>0.031057</td>
<td>0.332245</td>
</tr>
</tbody>
</table>

FDR: Benjamin corrected p.

For the publicly available microarray dataset GSE17351 were used to find the Differentially Expressed Genes (DEGs) in those ESCC related. After microarray analysis, DEGs with a fold change value > 2 and \( p < 0.05 \) were selected. Totally 578 genes were selected as DEGs from GSE17351. To get the regulatory relationships, the PCC \( \geq 0.6 \) was chosen as the threshold which resulted in a total of 35 regulatory relationships between 11 TFs and their 28 DEGs. By integration of these regulatory relationships above, a regulation network of ESCC was built between TFs and its target genes by Cytoscape (Figure 1). In this network, TFs with higher regulating degrees, SP1, E2F1, USF2 and SP3 form a local network which suggested that these TFs might play a more important role in ESCC. Besides, the CYP27B1 regulated by SP1 and POU2F1 and the AOX1 target gene regulated by SP1 and SP3 was also observed in our network.

**Pathway Enrichment Analysis**

DAVID, a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. Use the DAVID to identify over-represented pathways based on the hypergeometric distribution.

**TF-Pathway Network (TPN)**

To further investigate the regulatory relationships between TFs and pathways, we mapped DEGs in the regulation network to significant pathways and got a TF-Pathway Network (TPN) between TFs and pathways. The TFs regulated the pathways which the DEGs were mapped.

**Results**

**Regulation Network Analysis**

The publicly available microarray dataset GSE17351 were used to find the Differentially Expressed Genes (DEGs) in those ESCC related. After microarray analysis, DEGs with a fold change value > 2 and \( p < 0.05 \) were selected. Totally 578 genes were selected as DEGs from GSE17351. To get the regulatory relationships, the PCC \( \geq 0.6 \) was chosen as the threshold which resulted in a total of 35 regulatory relationships between 11 TFs and their 28 DEGs. By integration of these regulatory relationships above, a regulation network of ESCC was built between TFs and its target genes by Cytoscape (Figure 1). In this network, TFs with higher regulating degrees, SP1, E2F1, USF2 and SP3 form a local network which suggested that these TFs might play a more important role in ESCC. Besides, the CYP27B1 regulated by SP1 and POU2F1 and the AOX1 target gene regulated by SP1 and SP3 was also observed in our network.

**Significant Pathway Based on Regulation Network**

To identify the relevant pathways changed in ESCC, we used hypergeometric distribution ap-
Figure 2. TF-pathway regulation networks. The yellow triangles stand for the TFs and the green ellipses stand for pathways while the red arrows represent for activation and the blue arrows represent for repression.

Discussion

From enrichment regulation network analysis, we could find that two TFs, MYB and MYBL2 regulate four of the five enriched significant pathways. Many researchers have proved that MYB is essential for mammary tumorigenesis, especially in breast cancer. For example, it has been proved that this oncogene upregulation is associated with estrogen receptor (ER)-positive breast cancer and TGF-β-induced c-Myb affects the expression of EMT-associated genes and promotes invasion of ER+ breast cancer cells. As to MYBL2, which has been suggested to play a very important role in the proliferation and differentiation of several cell types and is a gene we have found is commonly regulated in several systems of colon cell maturation. MYBL2 is also suspected to be a biomarker in cervical cancer while MYBL2 activation is crucial for human hepatocellular carcinoma (HCC) progression. It is very interesting and confusing to find that the two TFs have reverse functions in the five pathways. Maybe that is because of the complexity and uncertainty of the regulation network, but answering this question is our aim in the future.

Besides, the five enriched pathways also have significant importance in the tumorigenesis of many mammary cancers, ESCC of course included. It has been demonstrated that a local renin-angiotensin system (RAS) in the musculature of the distal esophagus and that Angiotensin II is a potent stimulator of esophageal contractions via the angiotensin (AT) receptor. The prostate cancer pathway, by whose name the function is very obvious, has attracted much attention for its important role in the regulation of cell prolifera-
Regulation network analysis in the esophageal squamous cell carcinoma

tion and differentiation in prostate cancer. Conditions are just the same for the pathways of Glioma and Melanoma, the former of which could promote the tumorigenesis of glioma and is considered to have the same or similar effects on some other neuron related tumors. As to the latter, the Melanoma pathway is also demonstrated to play vital role in the pathogenesis, diagnosis, inhibition and resistance of melanoma.

While in p53 signaling pathway, which is closely related to a TF of great importance in cell proliferation, the activation of this pathway in cancer cells could result in cell cycle arrest, cellular senescence or apoptosis. It has been reported that focal adhesion kinase (FAK) deletion promotes p53-mediated induction of p21, DNA-damage responses and radio-resistance in advanced squamous cancer cells. According to an analysis research on Aberrant Signaling Pathways in Squamous Cell Lung Carcinoma, many molecules in p53 mediated signaling pathway were changed and their over expression could increase the apoptosis of both preneoplastic and carcinomatous mammary epithelial cells as well as repressing tumor growth through a combination of reduced angiogenesis and increased apoptosis.

During the multistep process of ESCC carcinogenesis, to date no single molecular marker came out that is able to predict who will and who will not develop cancer in the esophagus. Therefore, panels of markers need to be identified in future that could be used to indicate disease progression. In brief, the identification of crucial molecular pathways involved in ESCC would ultimately improve therapeutic effects and facilitate development of new treatment strategies.

References


