**Abstract.** OBJECTIVE: To identify key genes associated with squamous lung cancer (SLC) through analyzing gene expression data with bioinformatic tools, which could be potential biomarkers for diagnosis and treatment.

MATERIALS AND METHODS: Gene expression dataset GSE3268 was downloaded from Gene Expression Omnibus, including 5 SLC samples and 5 healthy controls. Data pre-treatment and differential analysis were performed with packages of R. Cluster analysis was done based on gene expression values to globally present the difference between the two states. Differentially expressed genes (DEGs) were divided into up-regulated and down-regulated genes, and then underwent functional enrichment analysis with DAVID tools. WebGestalt was used to retrieve microRNAs for the DEGs and then a regulatory network was constructed. GENECODIS was selected for functional annotation for all the genes in the network.

RESULTS: A total of 537 DEGs were obtained. Functional enrichment analysis revealed that cell cycle was significantly enriched in up-regulated genes. Besides, two microRNAs (miRNAs), MIR-142-5p and miR-9, were retrieved, which were potential tools to regulate the expression of key genes.

CONCLUSION: These DEGs may be involved in pathogenesis of SLC and some of them could be potential biomarkers. Besides, MIR-142-5p and miR-9 may be utilized to treat SLC as they could modulate cell cycle.

Key Words: Squamous lung cancer, Differentially expressed gene, Function enrichment analysis, MicroRNA, Regulatory network.

**Abbreviations**

SLC = squamous lung cancer; DEGs = differentially expressed genes; EGFR = epidermal growth factor receptor; miRNAs = microRNAs; GEO = Gene Expression Omnibus; BH = Benjamini and Hochberg; FDR = False discovery rate; DAVID = Database for Annotation, Visualization and Integrated Discovery; WebGestalt = Web based Gene Set Analysis Toolkit; GO = Gene Ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes; NEK2 = NIMA-related kinase 2; FOXM1 = Forkhead box M1; PTTG1 = pituitary tumor-transforming 1; SMAD7 = SMAD family member 7; FOXF1 = Forkhead box F1.

**Introduction**

Squamous lung cancer (SLC) is a common form of lung cancer, which is the main cause of cancer-related death in the world. Accounting for 25%-30% of all lung cancers, SLC can spread to bones, adrenal glands, the liver, small intestine, or brain that introduce great threat to human life. SLC patients with clinical stage IA disease have a 5-year survival rate of about 60%, while the 5-year survival rate for clinical stage II to IV disease ranges from 40% to less than 5%. The poor outcome of SLC patients could be explained, in part, by the difficulty of early diagnosis and the poor prognosis.

Understanding the pathogenesis of SLC can improve the current diagnosis, treatment and prognosis of SLC. Both genetic and environmental factors contribute to the development of SLC. It is widely recognized that tobacco smoking is the main risk factor for lung cancer, especially for SLC and SCLC (small cell lung cancer). Now there is considerable information concerning molecular abnormalities involved in the pathogenesis of SLC. Focal fibroblast growth factor receptor 1 (FGFR1) amplification which was associated with tumor growth and survival was identified in SLC. EGFR mutations in lung cancer are associated with clinical response to gefitinib therapy. Besides, roles of ubiquitin and ubiquitin-like proteins as well as cytokines in development of lung cancer have also been indicated. Although many molecules have been identified, it is still need to explore new genes related to SLC.

MicroRNAs (miRNAs) are approximately 22-nucleotide non-coding RNAs that are known to
regulate gene expression by targeting miRNAs for cleavage or translational repression. It has been estimated that miRNAs can control the activity of about 30% of human genes. Through this function they participate in many physiological processes including stem cells development, cell differentiation, regulation of cell cycle, apoptosis and transformation. Thus, deregulation of miRNAs can induce a variety of human diseases. Indeed, the abnormal expression of miRNAs is closely related with development of cancers, so they can be good markers for diagnosis, treatment or prognosis for cancers. Therefore, miRNAs associated with potential key genes in SLC were also investigated in present study.

Microarray technology is a powerful tool to explore the global changes in the incidence and development of cancer. Therefore, gene expression profiles for SLC samples were compared with those for healthy controls to identify differentially expressed genes (DEGs). Combined with functional enrichment analysis and regulatory network analysis, our study would not only provide insights into the pathogenesis of SLC, but also discover potential biomarkers that may provide contribution to diagnosis, treatment or prognosis of SLC.

Materials and Methods

Microarray Data

Microarray data set GSE3268 was downloaded from Gene Expression Omnibus (GEO), including 5 SLC samples and 5 corresponding healthy controls. Each pair of samples represent a single patient with squamous lung cancer. One is derived from the cancer cells, and the other is from the normal cells. There are five patients, with two arrays for each patient. The annotation information of chip was downloaded from GPL96 [HG-U133A] Affymetrix Human Genome U133A Array.

Screening of Differentially Expressed Genes (DEGs)

Raw data was converted into recognizable format and missing values were imputed. After data normalization, multtest package in R was chosen for differential analysis. Multiple testing correction was applied with Benjamini-Hochberg method. FDR < 0.05 and |logFC| > 1 were set as the cut-offs to screen out DEGs.

Cluster Analysis

To globally present the difference in gene expression pattern between SLC and healthy control, cluster analysis was conducted for all the samples.
**Functional Enrichment Analysis for the DEGs**

Functional enrichment analysis is able to reveal disturbed biological functions based upon DEGs\(^\text{20}\). All DEGs were inputted into DAVID (The Database for Annotation, Visualization and Integrated Discovery)\(^\text{21}\) for functional enrichment analysis and FDR < 0.05 was selected as the threshold. DEGs were divided into up- and down-regulated genes before analysis.

**Retrieval of miRNAs and Construction of Regulatory Networks**

WebGestalt\(^\text{22,23}\) was chosen for retrieval of miRNAs interacting with the up- and down-regulated genes, as well as construction of regulatory networks.

**Functional Annotation for Genes in the Networks**

GENECODIS (http://genecodis.dacya.ucm.es/) is a web-based tool for functional annotation, which integrates information from GO, KEGG, SwissProt and etc\(^\text{24}\). Therefore, it was selected for functional annotation for all the genes in the regulatory networks and adj \(p < 0.05\) was set as the cut-off.

**Results**

**Differentially Expressed Genes**

Normalized gene expression data was shown in Figure 1A and a good normalization was acquired. A total of 537 DEGs were screened out, including 188 up-regulated genes and 349 down-regulated genes.

**Cluster Analysis Result**

The result was shown in Figure 1B. SLC samples could be easily distinguished from the healthy controls as obvious differences existed in the gene expression pattern.

**Functional Enrichment Analysis Result**

Cell cycle was significantly over-represented in up-regulated genes and blood vessel development was enriched in down-regulated genes (Figure 2).

![Figure 2](image_url)
Deregulation of cell cycle and blood vessel development was closely associated with tumorigenesis, indicating potential roles of these DEGs in SLC.

Relevant miRNAs and Regulatory Networks

Four miRNAs were retrieved for up-regulated genes and 2 ones for down-regulated genes (Table I). In these miRNAs, miR-142-5p and miR-9 were found to interact with both up- and down-regulated genes, and thus corresponding regulatory networks were constructed for the two miRNAs to further look into their biological functions (Figure 3).

Gene Functional Annotation Result

Eight functional terms were significantly over-represented for miR-142-5p regulatory network and 5 for miR-9 regulatory network (Table II). Positive regulation of cellular process and cell differentiation were the most significant term in each network.

Discussion

In recent years, with progress in researches of gene function and signal transduction, molecular targeted therapy becomes available and provides a new way for the treatment of tumors. In present study, by comparison of gene expression data between SLC and healthy control a range of DEGs were screened out, which not only offered insights into the molecular mechanisms of SLC, but also provided potential biomarkers for this tumor.

Table I. Relevant miRNAs for up- (A) and down-regulated genes.

<table>
<thead>
<tr>
<th>microRNA</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>hsa_ACTGCCT, miR-34B</td>
<td>O = 7; rawP = 4.30e-05; adjP = 0.0007</td>
</tr>
<tr>
<td>hsa_TTTGCA G, miR-518A-2</td>
<td>O = 6; rawP = 0.0003; adjP = 0.0021</td>
</tr>
<tr>
<td>hsa_ACTTTAT, miR-142-5p</td>
<td>O = 6; rawP = 0.0015; adjP = 0.0048</td>
</tr>
<tr>
<td>hsa_TAGCTTT, miR-9</td>
<td>O = 5; rawP = 0.0035; adjP = 0.0070</td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>hsa_ACTTTAT, miR-142-5p</td>
<td>O = 14; rawP = 9.39e-08; adjP = 9.70e-07</td>
</tr>
<tr>
<td>hsa_ACCAAAG, miR-9</td>
<td>O = 11; rawP = 0.0022; adjP = 0.0023</td>
</tr>
</tbody>
</table>

O: number of genes in the pathways; adjP: p values after multiple testing correction.

Table II. Functional annotation results for genes in the networks.

<table>
<thead>
<tr>
<th>GO-ID</th>
<th>corr. p-value</th>
<th>x</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) miR-142-5p</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48522</td>
<td>0.0009995</td>
<td>12</td>
<td>positive regulation of cellular process</td>
</tr>
<tr>
<td>48518</td>
<td>0.0009995</td>
<td>12</td>
<td>positive regulation of biological process</td>
</tr>
<tr>
<td>48731</td>
<td>0.0017834</td>
<td>12</td>
<td>system development</td>
</tr>
<tr>
<td>23033</td>
<td>0.002004</td>
<td>12</td>
<td>signaling pathway</td>
</tr>
<tr>
<td>48856</td>
<td>0.0023442</td>
<td>12</td>
<td>anatomical structure development</td>
</tr>
<tr>
<td>7275</td>
<td>0.0047838</td>
<td>12</td>
<td>multicellular organismal development</td>
</tr>
<tr>
<td>32501</td>
<td>0.0061241</td>
<td>14</td>
<td>multicellular organismal process</td>
</tr>
<tr>
<td>32502</td>
<td>0.0062943</td>
<td>12</td>
<td>developmental process</td>
</tr>
<tr>
<td>(B) miR-9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65008</td>
<td>0.039728</td>
<td>6</td>
<td>regulation of biological quality</td>
</tr>
<tr>
<td>30154</td>
<td>0.046665</td>
<td>6</td>
<td>cell differentiation</td>
</tr>
<tr>
<td>48869</td>
<td>0.046665</td>
<td>6</td>
<td>cellular developmental process</td>
</tr>
<tr>
<td>48731</td>
<td>0.035755</td>
<td>8</td>
<td>system development</td>
</tr>
<tr>
<td>48856</td>
<td>0.037195</td>
<td>8</td>
<td>anatomical structure development</td>
</tr>
<tr>
<td>7275</td>
<td>0.046665</td>
<td>8</td>
<td>multicellular organismal development</td>
</tr>
</tbody>
</table>

x: the total number of genes in each GO category.
According to the functional enrichment analysis result, cell cycle and other functions were significantly over-represented. Considering the close relationship between cell cycle and cancers, it is possible that these DEGs may be involved in cancers in a certain degree. In fact, many DEGs have been reported to take parts in pathogenesis of cancers. NIMA-related kinase 2 (NEK2) is a serine/threonine-protein kinase that is involved in mitotic regulation. Hayward et al. indicated that it is up-regulated in breast cancer and may contribute to chromosome instability. Forkhead box M1 (FOXM1) is a transcriptional activator involved in cell proliferation and it regulates the expression of several cell cycle genes, such as cyclin B1 and cyclin D1. It has been found to be involved in recurrence of non-small cell lung cancer with predictive potential. It promotes tumor metastasis and thus inhibition of FOXM1 can suppress cell proliferation and tumor growth. Pituitary tumor-transforming 1 (PTTG1), highly expressed in various tumors, has transforming activity in vitro and tumorigenic activity in vivo. Huang et al. reported that knockdown of PTTG1 with RNAi would suppress the proliferation and invasive potential of PC3 human prostate cancer cells. SMAD family member 7 (SMAD7) presents a protective role in colitis-associated cancer and low-level expression of Smad7 correlates with lymph node metastasis and poor prognosis in patients with pancreatic cancer. Forkhead box F1 (FOXF1) is a potential tumor suppressor gene that epigenetically silenced in breast cancer. Taken together, all above genes which are DEGs between SLC samples and controls have been identified to participate in the development of different cancers, so we deduced these DEGs may serve as potential biomarkers of SLC. However, more studies are needed to validate our result.

In addition, many investigations have indicated that miRNAs play important roles in development of cancers, such as miRNA-296, which can inhibit the growth of esophageal cancer cells via regulating cyclinD1 and p27. Takeshita et al. report that miRNA-16 can suppress the
growth of metastatic prostate tumors via down-regulation of multiple cell cycle genes. We retrieved relevant miRNAs for DEGs, among which miR-142-5p and miR-9 showed more potential relationship with SLC as they regulated both up- and down-regulated genes. Functional annotation further revealed their involvement in regulation of cellular process and cell differentiation. Many cellular processes, such as cell growth5, division40, differentiation41 and signal transduction42, are all involved in development of cancers.

In addition, many studies have reported the relationship of miR-142-5p or miR-9 with cancers. Zhang et al43 propose that hsa-miR-142-5p combined with hsa-miR-375 can act as a predictor for recurrence risk in gastric cancer patients following surgical resection. Liu et al44,45 have reported that miR-142-5p is repressed in transgenic lung cancers and overexpression of this miRNA would significantly inhibit lung cancer cell growth. These works suggest that miR-142-5p may be a potential target to treat SLC. When referred to miR-9, it can be activated by MYC/MYC N and regulates E-cadherin as well as cancer metastasis46 and thus regulates the development of cancers. A study have reported that it can be used to differentiate primary from metastatic brain tumors47. Besides, it’s also regarded as a potential biomarker for recurrent ovarian cancer18. Hildebrandt et al49 further point out that its methylation status is associated with cancer development and metastatic recurrence. The important roles of these two miRNAs in development of different cancers suggest that they may take a place in regulating the pathogenesis of SLC and also suggest the key positions of corresponding DEGs in the whole regulatory mechanisms.

Conclusions

Overall, a range of DEGs were obtained through comparing gene expression profiles of SLC with those of healthy controls. These genes may play important roles in the pathogenesis of SLC according to the functional enrichment analysis. Besides, miR-142-5p and miR-9 that regulated both up- and down-regulated genes were identified and they may act as potential biomarkers for SLC. Of course, more researches are needed to demonstrate this result and develop their potentials in clinical applications.

Conflict of Interest

The Authors declare that there are no conflicts of interest.

Reference

Key genes associated with squamous lung cancer


