Abstract. – BACKGROUND: Pulmonary disease has become one of the major health problems.

AIM: To investigate the regulation mechanism of Chronic Obstructive Pulmonary Disease (COPD) on gene expression and pathway level.

MATERIALS AND METHODS: We mapped the differentially expressed genes to a regulation network and pathways, using transcriptome profiles and regulation data. We constructed a TF-target gene regulation network, TF-pathway regulation network, and pathway crosstalk network.

RESULTS: STAT1, NFKB1, SMAD4, and STAT3 played an important role in COPD through participating in a number of pathways. Although these related pathways all have been demonstrated associated with COPD in previous reports, the detail mechanism may be not very clear.

CONCLUSIONS: Our results may help to further understanding the mechanism of COPD. And Identified multiple pathways will also provide novel avenues in the treatment of COPD.

Key words: Chronic obstructive pulmonary disease, Gene regulatory networks, Pathway analysis.

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a major and increasing global health problem, which is predicted to become the fourth leading cause of death and the fifth commonest cause of disability in the world by 2030. COPD is characterized by chronic irreversible airflow limitation.

COPD is caused by noxious particles or gas, most commonly from cigarette smoke (CS), which triggers an abnormal inflammatory response in the lung. This is characterized by increased numbers of neutrophils, macrophages, T-lymphocytes, and the release of multiple inflammatory mediators (lipids, chemokines, cytokines, growth factors). Neutrophil recruitment to the airways and parenchyma involves adhesion to endothelial cells in the airways of COPD patients. Neutrophils can secrete serine proteases, including neutrophil elastase, cathepsin G and proteinase-3, as well as matrix metalloproteinase (MMP)-8 and MMP-9. There is a marked increase in the numbers of macrophages in airways, lung parenchyma, bronchoalveolar lavage (BAL) fluid and sputum in patients with COPD. Macrophages may be activated by CS extract to release TNF-α, IL-8, other CXC chemokines, monocyte chemotactic peptide (MCP)-1, LTB4 and reactive oxygen species (ROS). Dendritic cells (DCs) may migrate from the airways to regional lymph nodes and stimulate proliferation of CD8+ and CD4+ T-lymphocytes. T-lymphocytes in peripheral airways of COPD patients show increased expression of CXCR3, a receptor activated by IP-10, Mig and I-TAC. In a word, the increasing of the inflammatory response cell and subsequently mediators release further amplify the normal inflammatory response to CS in COPD disease.

Proof that genetic factors are involved in the pathogenesis of COPD comes from the observation that individuals with severe deficiency for alpha-1-antitrypsin, a major inhibitor of serine proteases, have an increased risk of developing COPD. Individuals with a severe deficiency for alpha-1-antitrypsin tend to develop more severe COPD at an earlier age. Gene expression profiling of human diseased tissues may provide insights into the molecular mechanisms of human disease and may eventually lead to the identification of novel therapeutic targets. A high-throughput microarray experiment was designed to analyze genetic expression patterns and identify potential genes to target for COPD. The identification of potential differen-
tionally expressed genes may assist for improved COPD diagnosis.

In this study, we analyzed gene expression profiles that distinguish COPD patients from healthy control subjects. Furthermore, the relevant transcription factors, target genes and pathways in the network are used to construct regulation network and explain potential regulation mechanisms in COPD.

**Materials and Methods**

**Affymetrix microarray data**

One transcription profile of GSE16972 was obtained from a public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/).

Bronchoalveolar lavage (BAL) fluid samples were collected from 5 COPD patients and 5 healthy controls, with all of them being smokers. Besides, the control and COPD groups were homogenous regarding age and smoking habits, but only differed in lung function parameters. All subjects were between 40 and 65 years old. In COPD patients, percent of predicted one-second forced expiratory volume (FEV$_1$) (FEV$_1$%) was < 80% and FEV$_1$ to percent of predicted forced vital capacity (FVC) (FVC%) was < 70%; in control patients, FEV$_1$% was ≥ 80% and FEV$_1$/FVC% was ≥ 70%. COPD patients all belong to stage 2-3 according to the Global Initiative for Chronic Obstructive Lung Disease. Alveolar macrophages were isolated using Percoll gradient centrifugation, isolated using CD14+ magnetic beads, and total RNA was extracted using Qiagen RNeasy Kit (Qiagen, Hilden, Germany). Total RNA was labeled with biotin and individual samples were hybridized to Affymetrix HG U133A GeneChips (Affymetrix, Cleveland, OH, USA).

**Pathway data**

Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals. The PATHWAY database records networks of molecular interactions in the cells, and variants of them specific to particular organisms. Total 130 pathways, involving 2287 genes, were collected from KEGG.

**Differentially expressed genes (DEGs) analysis**

For the GSE16972 dataset, the limma method was used to identify differentially expressed genes (DEGs). The original expression datasets from all conditions were processed into expression estimates using the RMA (Robust Multi-array Average) method with the default settings implemented in Bioconductor, and then construct the linear model. The DEGs only with the fold change value larger than 1.5 and $p$-value less than 0.05 were selected.

**Co-expression analysis**

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

<table>
<thead>
<tr>
<th>Database</th>
<th>Regulationship</th>
<th>TFs</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRANSFAC</td>
<td>774</td>
<td>219</td>
<td>265</td>
</tr>
<tr>
<td>TRED</td>
<td>5722</td>
<td>102</td>
<td>2920</td>
</tr>
<tr>
<td>Total</td>
<td>6328</td>
<td>276</td>
<td>3002</td>
</tr>
</tbody>
</table>
Gene Ontology analysis
DAVID\(^\text{16}\) is a high-throughput and integrated data-mining environment analyzing gene lists derived from high-throughput genomic experiments. Use the DAVID to identify over-represented GO categories in biological process.

**Regulation and PPI network construction**
Using the regulation data that have been collected from TRANSCRIP database and TRED database, we matched the relationships between differentially expressed TFs and its differentially expressed target genes.

Base on the above two regulation datasets and the pathway relationships of the target genes, we build the regulation networks by Cytoscape\(^\text{17}\). Base on the significant relationships (PCC \(> 0.6\) or PCC \(< -0.6\)) between TFs and its target genes, 17 putative regulatory relationships were predicted between 12 TFs and 16 target genes.

Pathway crosstalk analysis
To investigate the relationship between pathways which were regulated by the same TF in the TF to pathway regulation network, we used the PPI-network approach to find the crosstalk of pathways.

Here the crosstalk pathways are defined as those pathways which have the overlapping genes and edges with each other. The overlapping genes mean both of the two pathways included and the overlapping edges mean both of the two pathways included the PPI interaction edges.

To determine the co-expressed significance of a gene pair in disease cases, we used the PCC test to calculate the p-value.

Map those p-values to the nodes and edges in the PPI network collected from the human protein reference database (HPRD)\(^\text{18}\) and BIOGRID\(^\text{19}\) database. The following formula is used to define a function as the combination of statistical significance of an interaction by a scoring scheme. The detail description could be seen in Liu et al\(^\text{20}\).

\[
S(e) = f(\text{diff}(x), \text{cor}(x, y), \text{diff}(y)) = -2\sum_{t=1}^{k} \log(p_t)
\]

The \text{diff}(x) and \text{diff}(y) are differential expression assessments of gene x and gene y, respectively. \text{cor}(x, y) represents their correlation between gene x and gene y. f is a general data integration method that can handle multiple data sources differing in statistical power. Where \(k = 3\), p1 and p2 are the p-values of differential expression of two nodes, p3 is the p-value of their co-expression.

The detailed analysis of crosstalk of relationships among pathways is then investigated, especially that with overlap of two significant pathway analysis results.

To define the interaction significance between pathways, we summarize all the scores of edges \(S(e)\) of all non-empty overlaps. Specifically, the interaction score between two pathways is estimated by their overlapping status of weighted pathways in the following formula:

\[
C(pi, pj) = \sum_{e \in Oij} S(e)
\]

To estimate the significance of the overlapping between different pathways, we random sample \(10^5\) times of the same size two pathways in the edges of pathway network and calculate their overlapping scores. The frequency larger than C is regarded as the interaction significance p-value. At last the pathway crosstalk with the p-value < 0.05 as the significant crosstalk.

**Results**

**Regulation network**
To get DEGs of COPD, we obtained publicly available microarray data sets GSE16972 from GEO. 207 DEGs with the fold change value larger than 1.5 and p-value less than 0.05 were selected. To get the regulatory relationships, 0.6 is as the co-expressed value (PCC) threshold. Finally, 17 regulatory relationships including 12 different expressed TFs and their 16 differently expressed target genes were selected. By integrating the regulatory relationships above, a regulation network of COPD was built between TFs and its target genes (Figure 1). In our network, signal transducers and activators of transcription (STAT)1, NFKB1 (nuclear factor K B1) and CDKN1A with higher degrees form a local network which suggesting that these genes may play an important role in COPD. STAT1 activates 3 target genes and NFKB1 activates 4 target genes. The target CDKN1A was regulated by both SMAD4 and STAT3 was observed in this network.

**GO analysis of the regulation network**
Several Gene Ontology (GO) categories were enriched among these genes in the regulatory network, including immune response, inflamma-
Table II. GO enrichment analysis of biological process

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Counts</th>
<th>p-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>GO:0006955–immune response</td>
<td>10</td>
<td>8.43E-09</td>
<td>1.15E-05</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0006954–inflammatory response</td>
<td>7</td>
<td>7.65E-07</td>
<td>0.001042</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0006952–defense response</td>
<td>8</td>
<td>1.81E-06</td>
<td>0.002469</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0009611–response to wounding</td>
<td>7</td>
<td>1.30E-05</td>
<td>0.017701</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0006935–chemotaxis</td>
<td>4</td>
<td>6.65E-04</td>
<td>0.902068</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0042330–taxis</td>
<td>4</td>
<td>6.65E-04</td>
<td>0.902068</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0006954–positive regulation of immune system process</td>
<td>4</td>
<td>0.002092</td>
<td>2.810571</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0007626–locomotory behavior</td>
<td>4</td>
<td>0.003122</td>
<td>4.166664</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0007267–cell-cell signaling</td>
<td>5</td>
<td>0.003536</td>
<td>4.707006</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0007050–cell cycle arrest</td>
<td>3</td>
<td>0.005651</td>
<td>7.423657</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0002696–positive regulation of leukocyte activation</td>
<td>3</td>
<td>0.005975</td>
<td>7.833621</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0007155–cell adhesion</td>
<td>5</td>
<td>0.006138</td>
<td>8.039181</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0022610–biological adhesion</td>
<td>5</td>
<td>0.006169</td>
<td>8.078366</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0050867–positive regulation of cell activation</td>
<td>3</td>
<td>0.006534</td>
<td>8.536473</td>
</tr>
</tbody>
</table>

BP: biological process
NFκB1 is a transcription regulator that is activated by various stimuli such as cytokines, oxidant-free radicals, ultraviolet irradiation, etc. Inappropriate activation of NFκB has been associated with a number of inflammatory diseases. Study showed a significant reduction in the percentage of sputum neutrophils undergoing spontaneous apoptosis in subjects with COPD compared to non-smokers. And a significant increase was observed in the expression of both the p50 ($p = 0.006$) and p65 ($p = 0.006$) subunits of NF-kB in neutrophils from COPD subjects compared to non-smokers. These results give an effective support that NFKB signaling could delay constitutive neutrophil apoptosis, which contribute to the ineffective resolution of inflammation in COPD\textsuperscript{22}.

SMAD4 gene encodes a member of the smad family of signal transduction proteins. Smad/TGFβ1 (transforming growth factor-β1) is an important signaling pathway to mediate fibrosis stimulating the secretion of extracellular ma-
CDKN1A (p21) gene encodes a potent cyclin-dependent kinase inhibitor which binds to and inhibits the activity of cyclin-CDK2 or -CDK4 complexes, in turn inhibiting cell proliferation so as to allow cells to repair the DNA damage. Previous study about the relationship between CS and COPD inflammation revealed that p21 was also an important modifier of lung inflammation, and genetic ablation of p21 (p21−/−) could attenuate CS-mediated lung inflammation.

Further, regulation network between TFs and pathways was constructed. As we expected, STAT1, NFKB1, SMAD4, and STAT3 were still as hub nodes linked to lots of COPD related pathways. For example, STAT1 regulated the antigen processing and presentation, Primary immunodeficiency and ABC transporters pathway; SMAD4 and STAT3 could regulate the bladder cancer, ErbB (also know as EGFR) signaling pathway and pathways in cancer; NFKB1 regulated natural killer cell mediated cytotoxicity pathway, apoptosis pathway, cytokine-cytokine receptor interaction pathway and chemokine signaling pathway.

Figure 3. Pathway crosstalk analysis.
Expression and regulation analysis of pulmonary disease

These results were in agreement with our subsequent crosstalk analysis. All of these pathways have been demonstrated involved in COPD in direct or indirectly manner in previous reports.

Antigen processing and presentation have been reported related with COPD progression. The mature lymphoid collections are rarely observed in the lungs of nonsmokers, but they are present in the airways of smokers with COPD. This increase has been attributed to the large antigen load associated with bacterial colonization and more frequent infection of the lower respiratory tract. The epithelium covering the surface of the mucosal lymphoid follicles contains specialized M cells that transport antigens from the epithelial surface to the lamina propria in the mucosal immune system. Subsequently, antibody or effector T cells were induced to complete immune response.

Primary immunodeficiency may also contribute to persistent airway inflammation and progressive airway remodeling in COPD. Air-liquid interface epithelial cell cultures revealed that complete epithelial differentiation was required for normal pIgR expression and IgA transcytosis. However, areas of bronchial epithelial remodeling was found reduced pIgR expression, localized selective IgA deficiency, and increased CD4(+) and CD8(+) lymphocyte infiltration in patients with COPD. These results indicated that epithelial structural abnormalities lead to localized selective IgA deficiency (most common primary immunodeficiency) in COPD airways.

As an ATP-binding cassette (ABC) transporter, multidrug resistance-associated protein 1 (MRP1) expression was found diminished in bronchial epithelium of COPD patients (ex-smokers) and that lower expression is related to worse lung function. In addition, bronchial MRP1 expression was higher during smoking than after one year of smoking cessation, indicating functional activity of MRP1 in the lung may play an important role in the antioxidant defence against toxic compounds generated by cigarette smoke.

COPD was investigated associated with a significantly worse survival in cancer patients at the time of cancer diagnosis, being about 15% in elderly patients with bladder cancer. Therefore, closer involvement of pulmonologists and COPD nurses in cancer patients might be warranted. A number of pathways in cancer have been suggested underlying mechanism to understand of how lung cancer is associated with COPD. Such as downregulation of NF-kB activation may improve the efficacy of first-line therapy in both COPD and lung cancer. In addition, immune dysfunction, altered adhesion signaling pathways, epithelial-to-mesenchymal transition, and oxidant/inflammation-driven extracellular matrix (ECM) degradation were all suggested associated with COPD.

There was some evidence that ErbB signaling pathway involved in the pathophysiology of COPD. Mucus hypersecretion from hyperplastic airway goblet cells is a hallmark of COPD. Activation of epidermal growth factor receptors (EGFR, also know as ErbB) is responsible for mucin production after inhalation of cigarette smoke in airways. In the airway epithelial cell line and rat airway model, exposure to CS upregulated the EGFR mRNA expression and induced activation of EGFR-specific tyrosine phosphorylation, resulting in upregulation of MUC5AC mRNA and protein production, effects that were inhibited completely by selective EGFR tyrosine kinase inhibitor.

Natural killer mediated cell cytotoxic activity was found to be significantly decreased in patients with COPD compared with levels in healthy volunteers. This defect could be partially rescued by glycoposphosphate.

Recently, an increasing number of data suggest apoptosis of structural cells in the lung might possibly be an important upstream event in the pathogenesis of COPD. There is an increase in apoptotic alveolar epithelial and endothelial cells in the lungs of COPD patients. Since this is not counterbalanced by an increase in proliferation of these structural cells, the net result is destruction of lung tissue and the development of emphysema.

Conclusions

We used the regulation network approach to identify the most robust and consistent set of pathways...
genes for discrete and quantitative COPD phenotypes. This strategy used microarray data extraction methods and regulation datasets. The genes STAT1, NFKB1, STAT3 and CDKN1A had been proved high related to COPD with previous studies. Although SMAD4 had not been proved to be responded to the COPD directly, our results suggest that it involve in. Importantly, many pathways have been linked to those genes, which may help to understanding the mechanism of COPD. Identified multiple pathways will also provide novel avenues in the treatment of COPD.

Conflict of Interest
None.

Reference


3) Barnes PJ. Mechanisms in COPD: differences from asthma. Chest 2000; 117(2 Suppl): 10S-14S.


Expression and regulation analysis of pulmonary disease


