Comprehensive analysis of gene-expression profile in chronic obstructive pulmonary disease

Lei Wei1,*
Dong Xu2,**
Yechang Qian1
Guoyi Huang1
Wei Ma1
Fangying Liu1
Yanhua Shen1
Zhongfu Wang1
Li Li1
Shanfang Zhang1
Yafang Chen1

1Department of Respiratory Disease, Baoshan District Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai, 2Medical College of Soochow University, Suzhou, People’s Republic of China

*These authors contributed equally to this work

Objective: To investigate the gene-expression profile of chronic obstructive pulmonary disease (COPD) patients and explore the possible therapeutic targets.

Methods: The microarray raw dataset GSE29133, including three COPD samples and three normal samples, was obtained from Gene Expression Omnibus. After data preprocessing with the Affy package, Student’s t-test was employed to identify the differentially expressed genes (DEGs). The up- and downregulated DEGs were then pooled for gene-ontology and pathway-enrichment analyses using the Database for Annotation, Visualization and Integrated Discovery (DAVID). The upstream regulatory elements of these DEGs were also explored by using Whole-Genome rVISTA. Furthermore, we constructed a protein–protein interaction (PPI) network for DEGs. The surfactant protein D (SP-D) serum level and HLA-A gene frequency in COPD patients and healthy controls were also measured by enzyme-linked immunosorbent assay (ELISA) and real-time polymerase chain reaction, respectively.

Results: A total of 39 up- and 15 downregulated DEGs were screened. Most of the upregulated genes were involved in the immune response process, while the downregulated genes were involved in the steroid metabolic process. Moreover, we also found that HLA-A has the highest degree in the PPI network. The SP-D serum level and HLA-A gene frequency in COPD patients were significantly higher than those in healthy controls (13.62±2.09 ng/mL vs 10.28±2.86 ng/mL; 62.5% vs 12.5%; P<0.05).

Conclusion: Our results may help further the understanding of the mechanisms of COPD. The identified DEGs, especially HLA-A, may serve as diagnosis markers for COPD.

Keywords: chronic obstructive pulmonary disease, differentially expressed genes, gene-ontology analysis, protein–protein interaction

Introduction

Chronic obstructive pulmonary disease (COPD) is an inflammatory disease distinguished by progressive airflow limitation and parenchyma destruction.1 It 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 2030.2

Tobacco smoking is by far the most important risk factor involved in the development of COPD. Patients with COPD often experience abnormal pulmonary inflammation. This is characterized by increased number of inflammatory cells (neutrophils, macrophages, and T-lymphocytes) and the release of multiple inflammatory mediators (lipids, chemokines, cytokines, and growth factors).3,4 The increase in the inflammatory response cells and subsequently released mediators further
amplify the normal inflammatory response to tobacco smoking in COPD disease.

In addition, the involvement of genetic factors in the pathogenesis of COPD has also been proven through 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. However, it is proving difficult to identify specific biomarkers to aid diagnosis of the disease. Better understanding the pathogenesis of COPD will help us explore novel treatments for COPD.

Gene-expression profiling of human diseased tissues may provide new insights into the molecular mechanisms of human disease and eventually lead to the identification of novel therapeutic targets. A high-throughput microarray experiment was designed to analyze genetic expression patterns and identify potential target genes for COPD. The identification of potential differentially expressed genes (DEGs) may assist in improved COPD diagnosis.

In this study, we analyzed the gene-expression profiles of lung-tissue samples collected from COPD patients and healthy controls to identify DEGs. Then, gene-ontology functional and pathway-enrichment analyses were conducted for the DEGs. Furthermore, transcription-factor analysis of the constructed protein–protein interaction (PPI) network was performed. The surfactant protein D (SP-D) serum level and HLA-A gene frequency in COPD patients and healthy controls were also measured.

Materials and methods

Affymetrix® microarray data

One transcription profile of GSE29133 included six samples of alveolar epithelial type II cells from the lung tissues of three COPD and three non-COPD patients was obtained from a public functional genomics data repository, the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). The transcriptome of the samples was determined using Affymetrix HG U133A GeneChips (Affymetrix, Cleveland, OH, USA).

Identification of DEGs

Background correction and quantile normalization were performed for the raw array data using the Affy package (http://www.bioconductor.org/) and robust multi-array average (RMA).11 The probes without corresponding gene symbols were filtered. The average value of a gene symbol with multiple probes was calculated. Then the gene-expression signal intensities were recalculated using custom chip description files (Brainarray lab, http://brainarray.ncbi.nlm.nih.gov/) by the median polish method. Student’s t-test was employed to obtain the P-values. DEGs were selected with P-values less than 0.05 and fold-change values larger than two.

Gene-ontology and pathway-enrichment analyses

Gene-ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were enriched for DEGs by the Functional Annotation Clustering Tool in the Database for Annotation, Visualization and Integrated Discovery (DAVID; v 6.7, http://david.abcc.ncifcrf.gov). The major GO terms associated with biological processes and cellular components were manually summarized based on gene-term enrichment buttons provided for each functional group at a P-value of <0.1.

Regulatory elements analysis

DEGs were divided into up- and downregulated genes. The promoter region was defined as the 1 kb regions upstream of transcription start site of the gene. The Whole-Genome rVISTA (WGRV) was used to perform the motif finding on the promoter regions. The P-value of motif-finding results should be less than 0.05.

Construction of protein–protein interaction network

DEGs were submitted to Search Tool for the Retrieval of Interacting Genes (STRING; v 9.0) and PPIs of COPD signature genes were retained. In our analysis, only interactions with higher probabilistic confidence scores from the text mining, databases, or experiments were retained. Then, the obtained DEG interactions were used to construct a PPI network that was visualized by Cytoscape software (version 2.8).

Measurement of surfactant protein D serum level

Eight COPD patients were randomly selected from patients without other respiratory disorders (such as asthma or bronchiectasis) or undertaking chemotherapy treatment in our department. Meanwhile, eight healthy controls were enrolled from the health examination center of our hospital. The demographics of these eight patients and eight healthy controls are listed in Table 1. The age, sex, and smoking history of the COPD patients and healthy controls were
matched. This study was approved by the Ethics Committee at Baoshan District Hospital of Integrated Traditional Chinese and Western Medicine.

Whole blood (5 mL) samples were collected from the eight COPD patients and eight healthy controls. The serum samples obtained by centrifugation were used for detection of SP-D with an enzyme-linked immunosorbent assay (ELISA) kit (Shanghai Biivol Technologies Co, Shanghai, People’s Republic of China) according to the manufacturer’s instructions. Briefly, microtiter wells coated with SP-D antibodies were incubated with 100 μL diluted serum and calibrator samples at 4°C for 24 hours after blocking nonspecific binding. The wells were then washed with phosphate-buffered saline (PBS) and incubated at 37°C for 1 hour with 100 μL horseradish peroxidase-conjugated anti-SP-D antibody. To the washed wells was then added 100 μL tetramethylbenzene and the wells were incubated at room temperature for 15 minutes. Finally, 50 μL of NaN₃ (2 mM) was added to terminate the peroxidase reaction and the absorbance was measured at 450 nm. All assays were performed in triplicate.

Detection of HLA-A by real-time polymerase chain reaction
The total RNA was extracted from whole blood samples of the eight COPD patients and eight healthy controls using a blood extraction kit (Tiangen Biotech Co, Beijing, People’s Republic of China) according to the manufacturer’s protocol. The RNA concentration was determined by ultraviolet spectrophotometry, and the purity was assessed by the ratio of A260/A280. Then, RNA (50 ng) was reverse transcribed into complementary DNA (cDNA) using the TIAN Script Kit (Tiangen Biotech Co). The following primers (Sangon Biotech Co, Beijing, People’s Republic of China) were used: 5’-CAC TCC ATG AGG TAG TTT TTAT-3′ (forward) and 5’-CTC CAG GTA GGC TCT CAA-3′ (reverse) for HLA-A; and 5’-AAG GTC GGA GTC AAC GG-3′ (forward) and 5’-ATC TCG GGC GGG AAT AG-3′ (reverse) for the internal control glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The polymerase chain reaction (PCR) was undertaken in a 25 μL mixture system containing 5.5 μL 10× Ex Taq™ buffer, 1 μL of primers (10 μM), 2 μL deoxyribonucleotide (dNTP), 2.5 μL cDNA, and 0.5 μL Ex Taq polymerase (TaKaRa). The amplification was performed under the following conditions: 2 minutes at 95°C; five cycles of 35 seconds at 95°C, 40 seconds at 65°C, 75 seconds at 72°C; ten cycles of 35 seconds at 95°C, 40 seconds at 55°C, 90 seconds at 72°C; 5 minutes of extension at 72°C. Finally, agarose gel electrophoresis and imaging were conducted for PCR products.

Statistical analysis
Data were expressed as mean ± standard deviation. Changes in the values between the COPD and control samples were analyzed by paired t-test using SPSS software (v17.0; IBM, Armonk, NY, USA). A P-value <0.05 was considered to be statistically significant.

Results
Differentially expressed genes of COPD
A total of 39 upregulated and 15 downregulated DEGs in COPD samples were identified (Figure 1). The dysregulated
genes may play a major role in COPD. The top ten up- and downregulated DEGs are listed in Table 2.

### Gene-ontology functional and pathway-enrichment analyses of the DEGs

The upregulated genes were involved in the immune response, antigen processing and presentation, positive regulation of immune response, and negative regulation of caspase activity, while the downregulated genes were significantly related to the functions of response to drugs, regulation of transcription, and especially steroid metabolic process (Table 3).

Furthermore, KEGG pathway analysis revealed that upregulated genes, including TAPI, HLA-A, HLA-DOB, and HLA-F, were strongly associated with the antigen processing and presentation pathway ($P=0.001065854$). The downregulated genes such as HMGC2 and Fabp6 were associated with the peroxisome proliferator-activated receptor (PPAR) signaling pathway ($P=0.05319786220472053$).

Moreover, we also found that the upregulated genes APOL3, TNFSF13B, CFB, C1R, C1S, and SFN were located in the extracellular region (Table 4) and that all of these except for SFN were found to be involved in the immune-response process. In addition, the downregulated genes INHB2, GRP, PLIN2, and TGFBI also belonged to the extracellular protein genes (Table 4).

#### Protein–protein interaction network of the DEGs

A PPI network consisting of up- and downregulated genes/proteins was identified by STRING. In our network (Figure 2), human class I histocompatibility antigen (HLA-A) had the highest degree (=10), suggesting that this gene may play an important role in COPD. A total of 13 other genes were also involved in this network. We also performed transcription-factor analysis on these 13 genes by WRVG and found that the hepatocyte nuclear factor 4 alpha (HNF4A) motif was presented in the genes of NFM, C1R, IFI6, C1S, and TAPI, with a $P$-value of $<0.05$. Nuclear factor kappa-light-chain-enhancer of activated B-cells (NF-$\kappa$B) motif was found in HLA-A gene with a $P$-value of $<0.005$. However, there was no significant difference in the expression levels of these two genes between the COPD and control samples (Figure 3).

### Table 2 The top ten up- and downregulated differentially expressed genes (DEGs)

<table>
<thead>
<tr>
<th>DEG</th>
<th>Log2FC</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downregulated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRP</td>
<td>-3.150417021</td>
<td>0.0270184001</td>
</tr>
<tr>
<td>CYP3A5</td>
<td>-2.899486791</td>
<td>0.024383033</td>
</tr>
<tr>
<td>PEG3</td>
<td>-2.04046247</td>
<td>0.014618135</td>
</tr>
<tr>
<td>PLIN2</td>
<td>-1.763811105</td>
<td>0.028428862</td>
</tr>
<tr>
<td>TGFBI</td>
<td>-1.736106108</td>
<td>0.025442174</td>
</tr>
<tr>
<td>GPM6A</td>
<td>-1.590094551</td>
<td>0.018970538</td>
</tr>
<tr>
<td>CES1P2</td>
<td>-1.4256446</td>
<td>0.02441444</td>
</tr>
<tr>
<td>NA</td>
<td>-1.364292778</td>
<td>0.03443264</td>
</tr>
<tr>
<td>GPM6A</td>
<td>-1.334776799</td>
<td>0.023643482</td>
</tr>
<tr>
<td>Upregulated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC283070</td>
<td>2.817720183</td>
<td>0.029000411</td>
</tr>
<tr>
<td>CAMK1D</td>
<td>2.804673002</td>
<td>0.018217518</td>
</tr>
<tr>
<td>IFI27</td>
<td>2.159277226</td>
<td>0.018217517</td>
</tr>
<tr>
<td>IFI44L</td>
<td>1.79267042</td>
<td>0.002548303</td>
</tr>
<tr>
<td>CFB</td>
<td>1.730128584</td>
<td>0.023090984</td>
</tr>
<tr>
<td>NCRNA00185</td>
<td>1.660861629</td>
<td>0.01899832</td>
</tr>
<tr>
<td>PSMB9</td>
<td>1.640735827</td>
<td>0.00282925</td>
</tr>
<tr>
<td>HLA-L</td>
<td>1.585143624</td>
<td>0.01230165</td>
</tr>
<tr>
<td>HCP5</td>
<td>1.583656356</td>
<td>0.010858103</td>
</tr>
<tr>
<td>SPTLC3</td>
<td>1.55797381</td>
<td>0.04496727</td>
</tr>
</tbody>
</table>

### Table 3 Gene-ontology functional enrichment analysis of the differentially expressed genes

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upregulated genes in COPD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0006955–immuneresponse</td>
<td>15</td>
<td>8.37E-11</td>
<td>HLA-L, CFB, HLA-A, C1R, IFI44L, C1S, PSMB8, HLA-F, PSMB9, TNFSF13B, TAPI, HLA-DOB, PAG1, IFI6, GBP1</td>
</tr>
<tr>
<td>GO:0019882–antigen processing and presentation</td>
<td>7</td>
<td>2.89E-08</td>
<td>HLA-L, HLA-A, C1R, IFI6, C1S, HLA-DOB, PSMB8, PSMB9, HLA-F</td>
</tr>
<tr>
<td>GO:0050778–positive regulation of immune response</td>
<td>4</td>
<td>0.004351951</td>
<td>TNFSF13B, CFB, C1R, C1S</td>
</tr>
<tr>
<td>GO:0006952–defense response</td>
<td>6</td>
<td>0.012168745</td>
<td>APOL3, CFB, TAPI, HCP5, C1R, C1S</td>
</tr>
<tr>
<td>GO:0002253–activation of immune response</td>
<td>3</td>
<td>0.019497772</td>
<td>CFB, C1R, C1S</td>
</tr>
<tr>
<td>GO:0043086–negative regulation of catalytic activity</td>
<td>4</td>
<td>0.025015212</td>
<td>SFN, PSMB8, IFI6, PSMB9</td>
</tr>
<tr>
<td>GO:0043154–negative regulation of caspase activity</td>
<td>2</td>
<td>0.040479035</td>
<td>SFN, IFI6</td>
</tr>
<tr>
<td>GO:0008202–steroid metabolic process</td>
<td>3</td>
<td>0.009226796</td>
<td>CYP3A5, HMGC2, FABP6</td>
</tr>
<tr>
<td>GO:004293–response to drug</td>
<td>2</td>
<td>0.148717527</td>
<td>PLIN2, HMGC2</td>
</tr>
<tr>
<td>GO:0045449–negative regulation of transcription</td>
<td>2</td>
<td>0.881878371</td>
<td>BNC2, PEG3</td>
</tr>
</tbody>
</table>

**Abbreviations:** COPD, chronic obstructive pulmonary disease; GO, gene-ontology.
Surfactant protein D serum level and the HLA-A gene in COPD patients and healthy controls

The ELISA measurement results for SP-D serum level showed that the SP-D serum level in the eight COPD patients was significantly higher than that in the eight healthy controls (13.62±2.09 ng/mL vs 10.28±2.86 ng/mL, P<0.05, Figure 4A). What's more, the HLA-A gene was detected in the blood samples of one healthy control and five COPD patients (Figure 4B). The HLA-A gene frequency in the eight healthy controls (1/8, 12.5%) was obviously lower than that in the eight COPD patients (5/8, 62.5%) (P<0.05).

Discussion

In our study, nine upregulated and 15 downregulated DEGs were identified. Fujino et al found that 156 and 82 were up- and downregulated, respectively, in COPD alveolar epithelial type II cells which might be due to the methods for DEG identification being different. In particular, the probe with the maximum value of genes corresponding to more than one probe was adopted in their study, while we calculated the average value of gene symbols using multiple probes for further analysis. Of the screened DEGs, four genes (TAP1, HLA-A, HLA-DOB, and HLA-F) related to immune response and the antigen processing and presentation pathway were upregulated significantly in the COPD compared with in the control samples, while HMGCS2 and FABP6, which participate in the PPAR inflammatory pathway, were downregulated. In the study of Fujino et al, functional annotation revealed that upregulated genes, such as TAP1, TAP2, PSMB8, PSMB9, PSMB10, HLA-B, and HLA-C, were associated with antigen processing and presentation. Therefore, the results of our study are similar to those of the study of Fujino et al to some extent. Moreover, in our study, HLA-A with the highest degree in the PPI network was indicated as playing an important role in COPD.

According to a previous study, the genes TAP1, HLA-A, HLA-DOB, and HLA-F are involved in the major histocompatibility complex (MHC) class I pathway. Once a cytotoxic cluster of differentiation (CD)8 T-cells recognizes specific antigens loaded on the MHC class I molecule, the CD8 T-cells will kill the somatic cells using proteolytic enzymes, including granzyme A or B and perforin. The numbers of CD8 T-cells in the airways and alveoli of patients with COPD are correlated with airflow limitation. Furthermore, the CD8 T-cells in the lungs of patients with COPD exhibit oligoclonal expansion, suggesting that the infiltrated CD8 T-cells are antigen specific. Moreover, Park et al revealed an association of the HLA-A gene with diffuse panbronchiolitis in Korean patients. Their finding was further proved by Maruyama et al, who suggested that anti-HLA class I antibodies alone could induce obliteratorive airway

Table 4 The differentially expressed genes involved in extracellular region

<table>
<thead>
<tr>
<th>Gene count</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upregulated genes</td>
<td>6</td>
</tr>
<tr>
<td>in COPD</td>
<td></td>
</tr>
<tr>
<td>Downregulated genes</td>
<td>4</td>
</tr>
<tr>
<td>in COPD</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: COPD, chronic obstructive pulmonary disease.

Figure 2 Protein–protein interaction network in COPD patients. Nodes with colors from yellow to red represent different degrees. The lowest degree, 1, is represented by yellow and the highest degree, 10, by red. Abbreviation: COPD, chronic obstructive pulmonary disease.

Figure 3 (A) The expression level of nuclear factor kappa-light-chain-enhancer of activated B-cells (NF-κB) and hepatocyte nuclear factor 4 alpha (HNF4A) in COPD and non-COPD. (B) Motif finding results of NF-κB and HNF4A. Abbreviation: COPD, chronic obstructive pulmonary disease.
disease in heterotopic murine tracheal allografts by inducing growth-factor production, apoptosis, and the chemotaxis of inflammatory cells.22

Faner et al investigated the distribution of HLA class II alleles (DR and DQ) in the blood samples of COPD patients and found that the frequency of HLA class II allele DRB1*14 was significantly higher in COPD patients.23 However, as far as we are aware, few studies have investigated differences in the frequency of HLA-A between COPD patients and healthy controls. Therefore, the HLA-A gene frequency in the blood samples of COPD patients and healthy controls was measured in our study. We also found that the HLA-A gene frequency in the blood samples of COPD patients (62.5%) was significantly higher than in the healthy controls (12.5%). Therefore, the pathological processes involved in COPD might be mediated by HLA-A.

“SP-D” is a lung-derived protein that has been suggested as a potential biomarker for inflammatory lung disease, and many studies have revealed that the level of serum SP-D in COPD patients is significantly elevated.24–26 The level of serum SP-D detected in our study confirms the reliability of the blood samples collected from COPD patients.

**Conclusion**

Our results may help to further understanding of the mechanism of COPD. The identified multiple pathways will also provide novel avenues in the treatment of COPD. HLA-A may serve as a diagnosis marker for COPD. However, the sample size in our study was relative small, and the DEGs among different stages of COPD were not identified due to the sample size. In our further studies, the results will be confirmed using a larger sample size, and samples will be collected from COPD patients with different stages to study the development of the COPD.

**Acknowledgment**

This study was supported by the Shanghai Construction Project of Integrated Traditional Chinese and Western Medicine (zxb2012–06).

**Disclosure**

The authors declare no conflicts of interest in this work.

**References**