CYP1B1, VEGFA, BCL2, and CDKN1A Affect the Development of Chronic Obstructive Pulmonary Disease

Danlei Yang¹
Ying Yan²
Fen Hu¹
Tao Wang¹

¹Department of Respiratory and Critical Care Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, People’s Republic of China; ²Department of Respiratory and Critical Care Medicine, Ningxia People’s Hospital, Yinchuan 750002, People’s Republic of China

Purpose: Chronic obstructive pulmonary disease (COPD) is a progressive lung disease characterized by poor airflow. The purpose of this study was to explore the mechanisms involved in the development of COPD.

Patients and Methods: The mRNA expression profile GSE100281, consisting of 79 COPD and 16 healthy samples, was acquired from the Gene Expression Omnibus database. The differentially expressed genes (DEGs) between COPD samples and healthy samples were analyzed using the limma package. Functional enrichment analysis for the DEGs was carried out using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) tool. Furthermore, DEG-compound pairs were predicted using the Comparative Toxicogenomics Database. The KEGG metabolite IDs corresponding to the compounds were also obtained through the MetaboAnalyst pipeline. Based on the diffusion algorithm, the metabolite network was constructed. Finally, the expression levels of key genes were determined using quantitative PCR (qPCR).

Results: There were 594 DEGs identified between the COPD and healthy samples, including 242 upregulated and 352 downregulated genes. A total of 696 DEG-compound pairs, such as BCL2-C00469 (ethanol) and BCL2-C00389 (quercetin) pairs, were predicted. CYP1B1, VEGFA, BCL2, and CDKN1A were included in the top 10 DEG-compound pairs. Additionally, 57 metabolites were obtained. In particular, hsa04750 (inflammatory mediator regulation of TRP channels)-C00469 (ethanol) and hsa04152 (AMPK signaling pathway)-C00389 (quercetin) pairs were found in the metabolite network. The results of qPCR showed that the expression of CYP1B1, VEGFA, BCL2, and CDKN1A was consistent with that predicted using bioinformatic analysis.

Conclusion: CYP1B1, VEGFA, BCL2, and CDKN1A may play important functions in the development and progression of COPD.

Keywords: chronic obstructive pulmonary disease, differentially expressed genes, enrichment analysis, disease metabolites, metabolite network

Introduction

Chronic obstructive pulmonary disease (COPD), which manifest as either chronic bronchitis or emphysema, is characterized by poor airflow.¹ The primary symptoms of COPD are cough, shortness of breath, and expectoration.²⁻⁵ COPD is a progressive disease that may further develop into cor pulmonale and respiratory failure.⁴⁻⁵ Tobacco smoking and air pollution are the main risk factors of COPD, while metabolic disturbances have a critical influence on this disease.⁶ COPD is also associated with increased chronic inflammatory responses to toxic particles or gases in the airways and
In 2015, COPD has impacted 174.5 million people that lead to 3.2 million deaths globally. Thus, the pathogenesis of COPD should be further investigated to improve COPD treatment strategies.

The expression of pleomorphic adenoma gene like-2 (PLAGL2) is significantly higher in COPD patients than that in healthy smokers, suggesting that PLAGL2 over-expression may be involved in the mechanisms underlying the development and progression of COPD. Further, over-expression of nuclear factor-kB2 (NF-κB2) is related to the expression of cyclooxygenase-2 (COX-2), interleukin-1 beta (IL-1β), and interleukin-8 (IL-8), which are also associated with the pathogenesis of COPD. Prostaglandin E2 (PGE2) is an important cyclic component that is responsible for fibroblast senescence and inflammation in COPD patients. Regulation of the PGE2 signaling pathway may suppress fibroblast senescence and inflammation during disease progression. Moreover, the interleukin-4/interleukin-13 (IL-4/IL-13) receptor complex induces the overproduction of mucus and enhances the viscosity of the airway surface liquid in patients with COPD and bronchial asthma by promoting pendrin expression. However, despite the previous reports, the mechanisms involved in the development of COPD have not been fully elucidated.

Willis-Owen et al generated the COPD-related dataset GSE100281 to analyze global gene transcription. In their study, the differential gene expression between healthy and COPD samples was analyzed using limma package, while the patterns of transcriptional coordination were analyzed using weighed gene co-expression network analysis (WGCNA). They found 1826 transcripts showing COPD-related variations, where 18 transcripts exhibited more than 2-fold change in expression. In the current study, the GSE100281 dataset was analyzed again. However, the study differed from their analyses in that enrichment analysis followed the screening for the differentially expressed genes (DEGs) between COPD and healthy samples. Additionally, prediction of disease metabolites and metabolite network analysis were also performed, which were not mentioned in the study by Willis-Owen et al. The findings of our study might help establish the theoretical foundation for developing novel therapies for COPD.

Materials and Methods

Data Source

The mRNA expression profile GSE100281 (sequencing platform: GPL11532) [HuGene-1_1-st] Affymetrix Human Gene 1.1 ST Array [transcript (gene) version]) was extracted from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. GSE100281 is composed of mRNA expression profiles from 79 COPD and 16 healthy samples.

Data Preprocessing and Differential Expression Analysis

The raw data expression matrix was obtained using the R Affy package (version 1.56.0, http://bioconductor.org/packages/release/bioc/html/affy.html). Data preprocessing, including background adjustment, quantile normalization, final summarization, and log base 2 scaling, was conducted using the robust multi-array average (RMA) method. Combined with the platform annotation files, probe IDs were mapped to the genes symbols. Probes without corresponding gene symbols were removed. For several probes that matched with only one gene symbol, the mean expression value of the probes was calculated and considered as the final gene expression value.

Using the R limma package (version 3.34.9, http://bioconductor.org/packages/release/bioc/html/limma.html), differential expression analysis was performed on COPD and healthy samples. The p-values obtained using t-test were adjusted according to the Benjamini & Hochberg method. Genes with adjusted p-values < 0.05 were considered as differentially expressed.

Functional Enrichment Analysis

To identify the Gene Ontology (GO) functional terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways involved with COPD, the DEGs were analyzed using Database for Annotation, Visualization, and Integrated Discovery (DAVID) tool (version 6.8, https://david.ncifcrf.gov/). GO terms and KEGG pathways with p-value < 0.05 and count > 1 were considered significant.

Prediction of Disease Metabolites

Using “Pulmonary Disease” and “Chronic Obstructive” as key words, the genes and compounds directly correlated with COPD were searched from the Comparative Toxicogenomics Database (http://ctd.mdibl.org/). The DEGs-associated compounds with Reference.Count > 1 were screened. The KEGG metabolite IDs corresponding to the compounds were also obtained using the Compound ID Conversion tool in the MetaboAnalyst database (http://www.metaboanalyst.ca).
Metabolite Network Analysis
The R package was used for pathway enrichment analysis of the identified disease metabolites. Utilizing KEGG database as the background, a differential pathway network composed of pathways, compounds, reactions, enzymes, and KEGG modules was built according to the results of the functional enrichment analysis. Based on the diffusion algorithm, heat flows from a given metabolite node and diffuses to the differential pathway network. Heat diffusion caused the change of node scores in the differential pathway network. The p-values of the diffusion scores were obtained using the displacement test. The top 100 nodes with p-value < 0.05 were selected for constructing the gene-metabolite network.

The formula for calculating the diffusion scores was $T = -K_I - 1^*G$, where G is the inputted metabolite (ie if it is the inputted metabolite, G is 1; otherwise, G is 0) and KI is the matrix.

Validation with CTD Database
The genes directly associated with COPD were extracted from the CTD database. The parameter was set with an inference score > 5. The inference score reflects the degree of similarity between the CTD chemical–gene–disease networks and a similar scale-free random network. A high score means a high probability that the inference network has atypical connectivity. Furthermore, Venn diagram analysis was performed to identify the genes shared between the CTD databases and the identified DEGs.

Quantitative PCR (qPCR)
Based on the bioinformatics analysis results, the expression of key genes were verified in six CAD and five normal blood samples. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer’s instructions. The purity and quality of the extracted RNA were determined using the Infinite M100 PRO microplate reader (Tecan, Männedorf, Switzerland). Complementary DNA (cDNA) was synthesized from high quality total RNA using PrimeScript™ RT Master Mix (No. RR036A, Takara Bio USA, Mountain View, CA, USA). Real-time qPCR was performed to validate gene expression using Power SYBR™ Green PCR Master Mix (No. A25742, Thermo Fisher Scientific, Waltham, MA, USA) on the 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA), with the following thermal cycling conditions: 50 °C for 2 min, 95 °C for 10 min, followed by 40 cycles at 95 °C for 10 s and 60 °C for 30 s. The melting curve was analyzed from 60 °C to 95 °C at an increment of 0.5 °C/10 s. The primer sequences for the genes were as follows: CYP1B1, forward 5′-GGCAGAATGGATCAGGG-3′; reverse 5′-TGTGGTACGCCAAGACAGAAGG-3′; VEGFA, forward 5′-GAGGGCAAGATCATCACAGG-3′; reverse 5′-CACCAAGGCTCAGATTGGA-3′; BCL2, forward 5′-GACTTCCGGAGATGTTCCAG-3′; reverse 5′-GGTGCCGTTTCAGGTACTCA-3′; CDKN1A, forward 5′-TGGCACTCACCCTGCTCTG-3′; reverse 5′-CGGGTGGTGGAGTGGTAGA-3′; GAPDH, forward 5′-TGGACACTTGTGTATCCGGAGG-3′; reverse 5′-AGGAGGATGATGTTCTTGGGA-3′. The relative quantification of gene expression was based on the comparative CT ($2^{-\Delta\Delta CT}$) method. Statistical analysis was performed using GraphPad Prism 5 (GraphPad Software, San Diego, CA, USA). Data were presented as mean ± standard deviation (SD), with a statistical significance of P < 0.05.

Results
Differential Expression Analysis
Box plots showed that the medians of data after normalization were at the same level (Figure 1). Therefore, data preprocessing improved the quality of in silico data used for bioinformatics analysis. There were 594 DEGs, including 242 upregulated and 352 downregulated genes, between the COPD and healthy samples. The heatmap and volcano plot displaying the DEGs are presented in Figure 2.

Functional Enrichment Analysis
The upregulated and downregulated genes were separately subjected to functional enrichment analysis. The top 5 results of GO terms and KEGG pathway analysis are shown in Figure 3. The upregulated genes were mainly implicated in the endocytosis pathway, with top GO terms being associated with signal transduction (GO biological process, GO_BP), membrane (GO cellular component, GO_CC), and protein binding (GO molecular function, GO_MF). However, the downregulated genes were involved in the oxidative phosphorylation pathway, with top GO terms being associated with mitochondrial electron transport, NADH to ubiquinone (GO_BP), mitochondrial (GO_CC), and NADH dehydrogenase (ubiquinone) activity (GO_MF).
Prediction of Disease Metabolites

From the 21,555 genes and 4900 compounds directly related to COPD, a total of 123 compounds, 243 DEGs, and 696 DEG-compound pairs were screened. The top 10 DEG-compound pairs, including cytochrome P450 family 1 subfamily B member 1 (CYP1B1), vascular endothelial growth factor A (VEGFA), B cell leukemia/lymphoma 2 (BCL2), and cyclin dependent kinase inhibitor 1A (CDKN1A), are presented in Table 1. Furthermore, 57 compounds with corresponding KEGG metabolite IDs were obtained. In particular, both C00469 (ethanol) and C00389 (quercetin) were found to be related with BCL2.

Metabolite Network Analysis

Based on the differential pathway network, there were 9994 nodes and 33,828 edges identified. Among the 57 metabolites, 23 metabolites matched with nodes and edges in the differential pathway network. In addition, a gene-metabolite network composed of 77 nodes and 106 edges was constructed (Figure 4). Particularly, the hsa04750 (inflammatory mediator regulation of TRP channels)-C00469 (ethanol) and hsa04152 (AMPK signaling pathway)-C00389 (quercetin) pairs were discovered in the network.

Validation with CTD Database

A total of 12,455 COPD-associated genes were obtained from the CTD database. The results of Venn diagram analysis are shown in Figure 5 and Supplemental Table 1. There were 221 shared genes, including CYP1B1, VEGFA, BCL2, and CDKN1A, between the COPD-associated genes and the 243 DEGs.

Expression Levels of Key Genes

Based on in silico analysis, the key genes CYP1B1, VEGFA, BCL2, and CDKN1A were selected for validation using qPCR. As shown in Figure 6, the expression levels of CYP1B1 and CDKN1A were significantly upregulated (P < 0.05) in COPD.
Figure 3 The Gene Ontology (GO) biological process (BP), GO cellular component (CC), and GO molecular function (MF) functional terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways identified after functional enrichment analysis of the differentially expressed genes (DEGs). (A) Top 5 results for the upregulated genes. (B) Top 5 results for the downregulated genes.
samples than in the control. By contrast, the expression levels of **BCL2** and **VEGFA** were significantly downregulated (P < 0.05) in COPD samples than in the control. These results were consistent with the results previously obtained from the bioinformatics analysis.

**Discussion**

In the present study, 594 DEGs consisting of 242 upregulated genes and 352 downregulated genes between COPD samples and healthy samples were screened and functionally analyzed. A total of 696 DEG-compound pairs, including **CYP1B1**, **VEGFA**, **BCL2**, and **CDKN1A**, and 57 metabolites were obtained. Moreover, both C00469 (ethanol) and C00389 (quercetin) were found to be associated with **BCL2**. In the gene-metabolite network, the hsa04750 (inflammatory mediator regulation of TRP channels)-C00469 (ethanol) and hsa04152 (AMPK signaling pathway)-C00389 (quercetin) pairs were also identified.

**Table 1** The Top 10 Differentially Expressed Gene (DEG)-Compound Pairs

<table>
<thead>
<tr>
<th>Chemical Name</th>
<th>Gene Symbol</th>
<th>Reference Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetrachlorodibenzodioxin</td>
<td>CYP1B1</td>
<td>243</td>
</tr>
<tr>
<td>Oxygen</td>
<td>VEGFA</td>
<td>60</td>
</tr>
<tr>
<td>Cobaltous chloride</td>
<td>VEGFA</td>
<td>47</td>
</tr>
<tr>
<td>Tetrachlorodibenzodioxin</td>
<td>CYP1B1</td>
<td>44</td>
</tr>
<tr>
<td>Resveratrol</td>
<td>BCL2</td>
<td>32</td>
</tr>
<tr>
<td>Quercetin</td>
<td>BCL2</td>
<td>29</td>
</tr>
<tr>
<td>Resveratrol</td>
<td>CDKN1A</td>
<td>24</td>
</tr>
<tr>
<td>Cobaltous chloride</td>
<td>VEGFA</td>
<td>20</td>
</tr>
<tr>
<td>Resveratrol</td>
<td>CDKN1A</td>
<td>19</td>
</tr>
</tbody>
</table>

**Figure 4** Constructed gene-metabolite network. The red, purple, blue, and orange circles and the green inverted triangle represent the pathway, module, enzyme, reaction, and disease metabolite, respectively.
The levels of CYP1B1 mRNA are known to be upregulated in alveolar epithelial type II (ATII) cells relative to that in hepatocytes. In addition, increased CYP2C19 and reduced CYP2J2 levels have been detected in the ATII cells of COPD patients relative to that in smokers without COPD. CYP1A1, CYP1B1, CYP2J2, CYP2E1, CYP2B6, and CYP3A5 encode proteins that are localized in the lungs; thus, these proteins may activate the COPD-associated compounds. The transcription factor T-box (TBX) and its target CDKN2A are involved in the pathogenesis of COPD by affecting senescence. The expression of CDKN1A, CDKN2A, and caveolin 1 (CAV1) is known to be higher in the peripheral lungs of COPD patients. CDKN1A,

Figure 5 Results of the Venn diagram analysis.

Figure 6 Expression levels of CYP1B1, VEGFA, BCL2, and CDKN1A quantified using qPCR. *P<0.05.
a biomarker of cell cycle arrest and premature ageing, is upregulated in COPD patients, which is related to skeletal muscle wasting.\textsuperscript{31,32} Skeletal muscle dysfunction is common in COPD patients. A previous study identified 18 genes, including CDKN1A, that were upregulated in the quadriceps of COPD patients relative to that in healthy controls.\textsuperscript{14} These results suggest that CYP1B1 and CDKN1A might be implicated in the pathogenesis of COPD.

Decreased miR-503 function promotes the release of VEGF from lung fibroblasts, thereby mediating vascular homeostasis in patients with COPD.\textsuperscript{33} VEGF can be used to diagnose COPD in healthy donors (HD), with better overall accuracy and Youden’s index (YI), while IL-8 can be used to diagnose cancer in both HD and COPD patients.\textsuperscript{34} The serum levels of VEGF and its soluble receptor sVEGF R2 are higher in COPD patients than that in the control; therefore, VEGF and sVEGF R1 may be involved in aberrant pulmonary vascular remodeling in COPD patients.\textsuperscript{35} VEGF overexpression promotes the development of Th2 inflammatory disorders such as asthma, while VEGF downregulation can affect the mechanisms of viral disorders including COPD.\textsuperscript{36} Therefore, VEGF may also be involved in the development of COPD.

BCL2 is implicated in apoptotic regulation and lung function, which may be correlated with the development and progression of COPD.\textsuperscript{37} The BCL2 family members mediate cell apoptosis through maintenance of mitochondrial membrane potential, which promotes the development of COPD and affects its severity.\textsuperscript{38} Through the AMPK/mTOR signaling pathway, β-arrestin2 reduces the expression of inflammatory cytokines in the BEAS-2B bronchial epithelial cells by suppressing autophagy.\textsuperscript{39} Thus, BCL2 may also be associated with the progression of COPD through the has04750 (inflammatory mediator regulation of TRP channels)-C00469 (ethanol)-BCL2 and has04152 (AMPK signaling pathway)-C00389 (quercetin)-BCL2 pairs.

Conclusion
A total of 594 DEGs between COPD and healthy samples were identified. The key genes CYP1B1, VEGFA, BCL2, and CDKN1A may affect the mechanisms underlying the development and progression of COPD. Specifically, BCL2 may be involved in the development of COPD via inflammation-mediated regulation of TRP channels and AMPK signaling pathway. However, in-depth experimental studies are still needed to confirm these results.

Funding
This study was supported by the National Natural Science Foundation of China (81470252, 81170049, 81570325).

Disclosure
The authors declare no conflicts of interest in this work.

References
9. Ni K, Zhao ZH, Zhang JQ, et al. Stud...