Identification of biomarkers associated with clinical severity of chronic obstructive pulmonary disease

Jie Zhang Equal first author, 1, Changli Zhu Equal first author, 1, Hong Gao 1, Xun Liang 2, Xiaqian Fan 3, Yulong Zheng 1, Song Chen Corresponding Authors, 4, Yufeng Wan Corresponding Author 1

1 Department of Respiratory Diseases, the Affiliated Hua’ian Hospital of Xuzhou Medical University, Hua’ian, Jiangsu, China
2 College of Nursing and Midwifery, Jiangsu College of Nursing, Hua’ian, Jiangsu, China
3 Department of Emergency Medicine, Suqian First Hospital, Suqian, Jiangsu, China
4 Institute of Medicinal Biotechnology, Jiangsu College of Nursing, Hua’ian, Jiangsu, China

Corresponding Authors: Song Chen, Yufeng Wan
Email address: biocs@163.com, ggwanyufeng@163.com

We sought to identify the biomarkers related to the clinical severity of stage I to stage IV chronic obstructive pulmonary disease (COPD). Gene expression profiles from the blood samples of COPD patients at each of the four stages were acquired from the Gene Expression Omnibus Database (GEO, accession number: GSE54837). Genes showing expression changes among the different stages were sorted by soft clustering. We performed functional enrichment, protein-protein interaction (PPI), and miRNA regulatory network analyses for the differentially expressed genes. The biomarkers associated with the clinical classification of COPD were selected from logistic regression models and the relationships between TLR2 and inflammatory factors were verified in clinical blood samples by qPCR and ELISA. Gene clusters demonstrating continuously rising or falling changes in expression (clusters 1, 2, and 7 and clusters 5, 6, and 8, respectively) from stage I to IV were defined as upregulated and downregulated genes, respectively, and further analyzed. The upregulated genes were enriched in functions associated with defense, inflammatory, or immune responses. The downregulated genes were associated with lymphocyte activation and cell activation. TLR2, HMOX1, and CD79A were hub proteins in the integrated network of PPI and miRNA regulatory networks. TLR2 and CD79A were significantly correlated with clinical classifications. TLR2 was closely associated with inflammatory responses during COPD progression. Functions associated with inflammatory and immune responses as well as lymphocyte activation may play important roles in the progression of COPD from stage I to IV. TLR2 and CD79A may serve as potential biomarkers for the clinical severity of COPD. TLR2 and CD79A may also serve as independent biomarkers in the clinical classification in COPD. TLR2 may play an important role in the inflammatory responses of COPD.
Identification of biomarkers associated with clinical severity of chronic obstructive pulmonary disease

Jie Zhang¹#, Changli Zhu¹#, Hong Gao¹, Xun Liang², Xiaqian Fan³, Yulong Zheng¹, Song Chen⁴*, Yufeng Wan¹*

¹Department of Respiratory Diseases, the Affiliated Huai’an Hospital of Xuzhou Medical University, Huai’an 223002, China.
²College of nursing and midwifery, Jiangsu College of Nursing, Huaian 223300, China.
³Department of Emergency Medicine, Suqian First Hospital, Suqian 223899, China.
⁴Institute of Medicinal Biotechnology, Jiangsu College of Nursing, Huaian 223300, China.

#These authors contributed equally to this work.

*Corresponding authors:
Yufeng Wan
No.60, Huaihai South Road, Huai’an, Jiangsu, 223002, China
Email address: ggwanyufeng@163.com

Song Chen
No.2, Huanghe West Road, Huai’an, Jiangsu, 223300, China
Email address: biocs@163.com

Abstract
We sought to identify the biomarkers related to the clinical severity of stage I to stage IV chronic obstructive pulmonary disease (COPD). Gene expression profiles from the blood samples of COPD patients at each of the four stages were acquired from the Gene Expression Omnibus Database (GEO, accession number: GSE54837). Genes showing expression changes among the different stages were sorted by soft clustering. We performed functional enrichment, protein-protein interaction (PPI), and miRNA regulatory network analyses for the differentially expressed genes. The biomarkers associated with the clinical classification of COPD were selected from logistic regression models and the relationships between TLR2 and inflammatory factors were verified in clinical blood samples by qPCR and ELISA. Gene clusters demonstrating continuously rising or falling changes in expression (clusters 1, 2, and 7 and clusters 5, 6, and 8, respectively) from stage I to IV were defined as upregulated and downregulated genes, respectively, and further analyzed. The upregulated genes were enriched in functions associated with defense, inflammatory, or immune responses. The downregulated genes were associated with lymphocyte activation and cell activation. TLR2, HMOX1, and CD79A were hub proteins in the integrated network of PPI and miRNA regulatory networks. TLR2 and CD79A were significantly correlated with clinical classifications. TLR2 was closely associated with inflammatory responses during COPD progression. Functions associated with inflammatory and immune responses as well as lymphocyte activation may play important roles in the progression of COPD from stage I to IV. TLR2 and CD79A may serve as potential biomarkers for the clinical severity of COPD. TLR2 and CD79A may also serve as independent biomarkers in the clinical classification in COPD. TLR2
may play an important role in the inflammatory responses of COPD.

Introduction

Chronic obstructive pulmonary disease (COPD) is a heterogeneous condition characterized by the progressive, irreversible limitation of a patient's airflow (Agusti 2014; Vestbo et al. 2013). It is a leading cause of chronic morbidity and it has been predicted to be the third leading cause of mortality worldwide by 2030 (Celli et al. 2015; Rodriguez-Roisin et al. 2017). COPD is a progressive disease that typically worsens over time (Rabe et al. 2013). However, the molecular mechanisms underlying its progression are poorly understood (Singh et al. 2018). It has been reported that an early diagnosis and individualized treatments can reduce mortality and the socioeconomic burden caused by the disease. Damage to the airways and lungs may be delayed but not effectively reversed (PAUWELS et al. 2004).

Spirometry is used to diagnose the severity of COPD by determining the degree of airflow limitation. Spirometry is based on the forced expiratory volume in one second (FEV1) expressed as a percentage of the predicted normal value for a person's gender, age, weight, and height (Vestbo et al. 2013). COPD is divided into four stages based on FEV1 according to the GOLD guidelines (Vestbo et al. 2013), and COPD exacerbations are classified as episodes of worsening symptoms from stage I to IV (Wedzicha & Seemungal 2007). The treatment recommendations for European and American practitioners are based on FEV1 (Qaseem et al. 2011) but FEV1 does not directly reflect the systemic manifestation of symptoms in COPD patients (Alotaibi & Ansari 2016). Several studies have demonstrated that there is a genetic component to COPD susceptibility (Hersh et al. 2011; McCLOSKEY et al. 2001). Thus, the identification of these genetic markers may help to develop diagnostic and therapeutic targets for the treatment of COPD. Previous genome-wide association studies have identified several genetic loci related to COPD susceptibility (Cho et al. 2011; Cho et al. 2014). A recent study investigated the gene expression profiles among patients with frequent COPD exacerbations and identified three genes (ARHGEF10, LAF4, and B3GNT) as predictors of exacerbations (Singh et al. 2014). However, that study did not investigate whether those genes could be used as biomarkers for the clinical classification of COPD.

We downloaded the aforementioned gene expression profile (GSE54837) (Singh et al. 2014) and investigated the genes showing changes in expression among different stages of COPD. We identified key genes using bioinformatics analyses and then further selected the biomarkers associated with the clinical classification of COPD using logistical regression models.

Materials and methods

Microarray dataset

We downloaded the gene expression profiles with accession number GSE54837 (Singh et al. 2014) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54837) from the Gene Expression Omnibus (GEO) database. These profiles included different clinical stage data, and were obtained based on the Affymetrix Human Genome U133 Plus 2.0 Array. We obtained 226 blood samples from 90 stage I patients, 58 stage II patients, 55 stage III patients, and 13 stage IV patients. The age and sex of each patient is shown in Supplementary Table 1.
Data preprocessing

The original CEL data were preprocessed (background adjustment, quantile normalization, final summarization, and probe ID to gene symbol) using the RMA method (Irizarry et al. 2003) in the Affy package (Gautier et al. 2004). The expression values of the probes were averaged as the final gene expression value when the different probes corresponded to the same gene symbol. After the expression matrix of all genes in each sample was obtained, the gene expression values of each sample group with different clinical grades were averaged according to the clinical classification and the final gene expression differential matrix was obtained.

Gene clustering analysis

The differential matrix of samples was subjected to noise-robust soft clustering using the fuzzy c-means algorithm in the Mfuzz package (Futschik & Carlisle 2005; Kumar & Futschik 2007). The Fuzzy C-Means (Futschik & Carlisle 2005) clustering method was adopted for clustering analysis on the genes of samples using varying times and changes in expression levels. Multiple clustering results were obtained, and the results were divided according to the trends into two parts: rising and falling. The parameters used were minSTD = 0.1, acore = 0.5, and cOptimal = 10.

Functional enrichment analyses

The gene sets with significant rising or falling trends were examined using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al. 2008). Enrichment results were identified in biological processes (BP), cellular components (CC), molecular functions (MF), and pathways.

Protein-Protein Interaction (PPI) analysis and miRNA-target gene regulatory relation analysis

PPI pairs were predicted based on the genes with rising and falling trends using the Search Tool for the Retrieval of Interacting Genes (STRING) (Mering et al. 2003). The required confidence of (combined score) > 0.4 was set as the threshold value for PPI relation prediction. The regulatory miRNAs of genes with PPI relations were predicted using Webgestalt (Zhang et al. 2005) based on the hypergeometric test with a threshold of enriched gene count ≥ 3. The p value was adjusted using the Benjamini-Hochberg (BH) method (Benjamini & Hochberg 1995), and the adjusted p value < 0.05 was set as the threshold.

Integration of PPI and miRNA-target gene regulatory relations

The PPI pairs and miRNA-target gene regulatory relation pairs were integrated to construct a network using Cytoscape (Shannon et al. 2003). The degree of connectivity in the network was analyzed and the hub protein (the node with a high degree of connectivity) was selected.

Influencing factors of the clinical grade of COPD
We performed logistical regression analysis after obtaining the hub proteins in the network based on the age ($\leq 60 = 0$, $> 60 = 1$), gender (female $= 0$, male $= 1$), and the gene corresponding to the hub protein as potential factors influencing the clinical grade to further determine whether the expression of the genes could be used as biomarkers of clinical grade. Univariate logistic regression analysis was conducted to preliminarily screen the factors with p value $< 0.05$. Multivariate logistic regression analysis was then conducted to further select the factors with p value $< 0.05$.

**Clinical data collections and detections**

Investigators obtained written informed consent before enrolling participants in the clinical trial. We selected patients who underwent pulmonary function testing at the respiratory function laboratory of the Affiliated Huai’an Hospital of Xuzhou Medical University between January 1, 2019 and May 31, 2019. Patients with complications including asthma, pulmonary interstitial fibrosis, bronchial dilatation, other pulmonary diseases, or malignant tumors were excluded. COPD was staged according to the GOLD guidelines: GOLD I, FEV1 $\geq 80\%$ predicted; GOLD II, FEV1 $< 80\%$ and $\geq 50\%$ predicted; GOLD III, FEV1 $< 50\%$ and $\geq 30\%$ predicted and GOLD IV, FEV1 $< 30\%$ predicted. Patients selected for the study were divided into the following four groups: no related diseases non-smoker, no related diseases smoker, COPD stage I-II, and COPD III-IV. Peripheral venous blood samples were collected from all subjects and placed in anticoagulation centrifuge tubes. Mononuclear cells were separated by density gradient centrifugation. CD14$^+$ immunomagnetic beads (Miltenyi, Germany) were sorted by the automatic magnetic bead sorter. TLR2 gene expression in the monocytes was detected using real-time quantitative PCR according to the instructions of the Real-time PCR kit (Takara, Tokyo, Japan). The amplification specificity was confirmed by melting curves and fluorescence was determined at 60°C. PCR conditions were as follows: predegeneration at 95°C for 30 s, followed by 40 cycles for 5 s at 95°C and 31 s at 60°C. qRT-PCR reactions were performed in a total volume of 20 µL. The primer sequences for TLR2 were as follows: F: 5’-TCGGAGTTCTCCAG TTCTCT-3’, R: 5’-TCCAGTGCTTCAACCCACAA -3’. The internal reference was β-actin (F: 5’-CCTGGCACCCAGCACAAT-3’, R:5’-GGGCCGGACTCGTCATAC3’. Additional blood samples were taken and centrifuged at 3,000 r/min for 10 min, after which the serum was separated and stored in a refrigerator at −80 °C for testing. All experiments were performed in triplicate. Relative quantitative expression levels were calculated using the $2^{-ΔΔCt}$ method. The levels of IL-6, IL-8, TNF-α, and IFN-γ in the serum samples were detected and compared by an ELISA Kit (Sangon Biotech, Shanghai, China). All experiments were approved by the Ethics Committee of the Affiliated Huai’an Hospital of Xuzhou Medical University. The clinical characteristics of all the subjects are shown in Supplementary Table 2.
**Statistical analysis**

We used SPSS 19.0 and Graphpad Prism 7.0 for statistical analysis. The expression levels of TLR2 and inflammatory factors were analyzed using a one-way ANOVA, followed by the least significant difference multiple comparison post-hoc test, when appropriate. We analyzed the correlation between TLR2 expression levels and inflammatory factor levels in all detected samples using Pearson Correlation Analysis. All statistical analyses were conducted with a significance level of $\alpha=0.05$ (P < 0.05).

**Results**

**Gene clustering analysis**

Opposing trends were observed among clusters 1, 2, and 7 and clusters 5, 6, and 8 as the clinical stage progressed from stage I to stage IV (Figure 1). Clusters 5, 6, and 8 showed significantly increased expression, while clusters 1, 2, and 7 displayed significantly decreased expressions from stage I to IV. The changes in expression in clusters 3, 4, 9, and 10 were disordered. We selected clusters 1, 2, 5, 6, 7, and 8 for further analysis. A total of 78, 42, and 77 genes were included in clusters 5, 6, and 8, respectively, and were defined as upregulated genes. 59, 44, and 56 genes were included in clusters 1, 2, and 7, respectively, and were defined as downregulated genes.

**Functional enrichment analyses**

The upregulated genes were significantly related to GO functions including defense, inflammatory, wounding, and immune responses, the positive regulation of biosynthetic processes, system lupus erythematosus pathways, the toll-like receptor signaling pathway, and cytokine-cytokine receptor interactions (Table 1). The downregulated genes were involved in GO functions including lymphocyte, cell, and leukocyte activation, the external side of the plasma membrane, type I diabetes mellitus pathway signaling, Jak-STAT signaling, and T cell receptor signaling (Table 2).

**PPI and miRNA-target gene regulatory relations prediction**

100 pairs were predicted from upregulated genes and 40 PPI pairs were predicted from downregulated genes using STRING. 64 and 92 miRNA-target gene regulatory relation pairs were predicted from upregulated and downregulated genes, respectively (Table 3).

**Network integration analysis**

Two integrated networks were constructed based on the obtained PPI and miRNA-target gene regulatory relation pairs (Figure 2). A constructed network of upregulated genes displayed the following five nodes with degrees $\geq 6$: toll like receptor 2 (TLR2), matrix metallopeptidase 9 (MMP9), heme oxygenase 1 (HMOX1), and C-C motif chemokine receptor 1 (CCR1) (Figure 2A). In the network of downregulated genes, the CD79a molecule (CD79A) and phospholipase C gamma 1 (PLCG1) had degrees $\geq 6$ and were considered to be hub nodes (Figure 2B).
Influencing factors of the clinical grade of COPD

Univariate logistic regression analysis results revealed that the factors of age, TLR2, MMP9, CCR1, CD79A, and PLCG1 were significantly associated with clinical classification (Table 4). These factors were further analyzed by multivariate logistic regression analysis and TLR2 was found to be significantly positively correlated and CD79A was significantly negatively correlated with clinical classification, suggesting that they could be used as independent biomarkers of clinical classification (Table 5).

The relationships between TLR2 and inflammatory factors in COPD progression

TLR2 can trigger signal transduction and lead to the release of inflammatory mediators. COPD is defined as a chronic inflammatory lung disease. We chose IL-6, IL-8, TNF-α, and IFN-γ to assess the severity of COPD as they are known inflammatory factors in COPD progression. We found that the levels of IL-6, IL-8, TNF-α, and IFN-γ in the smoker and COPD III-IV groups were significantly higher than those in the non-smoker and COPD I-II groups, and TLR2 in mononuclear cells of the peripheral blood was significantly correlated with the clinical classification of the patients (Figure 3A-E). We also found that TLR2 and the expressions of inflammatory markers were more significant in the GOLDIII/VI group than those in the healthy smoker group. We used the ROC curves to identify the sensitivity and specificity of the expression levels. The results showed that the five markers, IL-6, IL-8, TNF-α, IFN-γ and TLR2, have diagnostic value for COPD. The areas under the curve (AUC) were 0.7782, 0.8857, 0.9318, 0.9276 and 0.7647, respectively (P<0.0001, figures 3F-3J). The correlations between the TLR2 expression and inflammatory factor levels was analyzed by Pearson correlation analysis. The expression levels of TLR2 were correlated with the serum levels of IL-6 (standardized regression coefficient = 0.5594, P < 0.0001, Figure 3K), IL-8 (standardized regression coefficient = 0.6498, P < 0.0001, Figure 3L), TNF-α (standardized regression coefficient = 0.5126, P < 0.0001, Figure 3M), and IFN-γ (standardized regression coefficient = 0.4973, P < 0.0001, Figure 3N).

Discussion

COPD is a growing global health problem. As the population ages, life expectancy among the elderly is expected to rise in relation to the number of smokers and the prevalence of COPD. Researchers are working to identify and diagnose COPD using easily available biomarkers to facilitate the early detection of fluid build-up in COPD and improve diagnostic accuracy. Previous studies focused on finding new biomarkers in the blood for the early diagnosis of COPD and many of these potential diagnostic markers need to be verified in additional studies. We studied gene clusters with continuously rising (clusters 5, 6, and 8) or falling (clusters 1, 2, and 7) trends in expression changes from stage I to stage IV. The upregulated genes, such as TLR2, and HMOX1, were significantly enriched in functions associated with defense, inflammatory, or immune responses, while the downregulated genes, such as CD79A, were associated with the activation of lymphocytes, cells, and leukocytes. TLR2, HMOX1, and CD79A were hub proteins...
in the integrated network. Logistic regression analysis showed that TLR2 was positively correlated
and CD79A was negatively correlated with clinical classification.

Our results showed that the upregulated genes, such as TLR2, and HMOX1, were significantly
associated with the inflammatory and immune responses. Exposure to inhaled pollutants can result
in chronic airway inflammation in COPD by activating structural and inflammatory cells within
the lungs (Rovina et al. 2013). COPD exacerbation is defined as an acute worsening of the
symptoms that are implicated in increased systemic and airway inflammation (Wedzicha &
Seemungal 2007). The inflammatory and immune responses are known to play a critical role in
COPD. Dave Singh, et al. (2014b) found that systemic immune function is associated with the
exacerbation of COPD. Our study further suggested the important role of inflammatory and
immune responses in the progression of COPD from stage I to stage IV. IL-6, IL-8, TNF-α and
IFN-γ are important inflammatory mediators during the progression of COPD and IL-6 participates
in the inflammatory response. It can further aggravate the oxidative stress and inflammatory
responses by promoting the proliferation and maturation of T lymphocytes and B lymphocytes.
IL-8 is an inflammatory mediator that promotes the activation of neutrophils and the release of
various enzymes that damage the bronchi. Excessive levels of IL-8 can aggravate the inflammatory
response and destroy the lung tissue (Farahi et al. 2017; Nakamoto et al. 2019). By acting on
neutrophils, TNF-α enhances the expression of leukocyte adhesion molecules and activates other
factors involved in the inflammatory response. Higher levels of TNF-α and IFN-γ in COPD also
indicate increased inflammation (Xu et al. 2019; Zhang et al. 2016). Our findings about these
inflammatory factors in COPD were consistent with previous studies.

TLR2 was among the top two hub proteins in the network. TLR2 is a member of the toll-like
receptor family, which plays a fundamental role in the activation of innate immunity (Da Silva et
al. 2008). Haw et al. (2018) found that TLR2 mRNA expression was increased in the epithelium
and parenchyma of the mouse airway when chronically exposed to cigarette smoke. These results
were replicated in human COPD patients. Our results showed that increased TLR2 was an
independent clinical classification biomarker in COPD. The results of some studies do not support
this research. A previous study reported that the alveolar macrophages in smokers and COPD
patients presented an equally decreased surface expression of TLR2 compared to non-smokers
(Droemann et al. 2005). The role of TLR2 in the pathogenesis of COPD is also controversial in
previous literature (Freeman et al. 2013; Haw et al. 2018; Simpson et al. 2013; von Scheele et al.
2011). These conflicting results may be due to differences in experimental methods (e.g. peripheral
blood monocytes vs. macrophages) or cohorts of patients with varying medical backgrounds.

We found that the downregulated genes were involved in functions associated with lymphocyte
activation, which may suggest that lymphocyte activation plays a role in COPD development.
COPD has been thought to possess an autoimmune component (Agusti et al. 2003; Lee et al. 2007),
but the antigenic stimulation responsible for lymphocyte activation is unclear. Previous studies
demonstrated that the numbers of B lymphocytes and CD8+ T cells are increased in the airways
of COPD patients compared with healthy controls (Gosman et al. 2006; Saetta et al. 1999). The
number of lymphocytes has been shown to increase with the severity of the disease (Hogg et al.
2004). We found that the hub protein, CD79A, was enriched with lymphocyte activation. CD79A
has been demonstrated to play diverse roles in the development and function of B lymphocytes
(Pelanda et al. 2002) and may be downregulated in severe COPD when compared with controls
(Cockayne et al. 2011). Our results showed that CD79A was negatively correlated with the clinical
classification of COPD. Thus, lymphocyte activation and CD79A may be associated with the
clinical severity of COPD.
Our study is limited because we analyzed the differentially expressed genes in COPD patients using a single set of GEO data. The sample capacity was small and more GEO data should be used in further studies. Secondly, by extracting the total RNA content in blood monocytes, we found that the Ct value was relatively large, indicating that the expression of TLR2 in blood was lower, and the micro expression levels may have caused experimental errors. We performed each test at least three times to improve the repeatability of the results. Finally, we only studied the expression of TLR2 in the clinical population but its specific mechanism and inconsistency in previous studies was well explained and requires additional experimentation. Our results show that the inflammatory and immune responses, as well as lymphocyte activation, may play important roles in COPD development from stage I to stage IV. TLR2, and CD79A may serve as potential biomarkers in the exacerbation of COPD, and TLR2 and CD79A may also serve as independent biomarkers for the clinical classification of COPD.

Acknowledgements

We wish to thank Mr. Lizhang Xun for assisting with the collection of clinical blood samples and Prof. Song Chen for their assistance with the experiments.

References

Agusti A. 2014. The path to personalised medicine in COPD. Thorax 69:857-864.
Agusti A, MacNee W, Donaldson K, and Cosio M. 2003. Hypothesis: does COPD have an autoimmune component? : BMJ Publishing Group Ltd.
Alotaibi N, and Ansari K. 2016. The Body Mass Index, Airflow Obstruction, Dyspnea and Exercise Capacity (BODE) Index in Chronic Obstructive Pulmonary Disease for Saudi Population. J Pulm Respir Med 6:2.
Benjamini Y, and Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society 57:289-300.
Celli BR, Decramer M, Wedzicha JA, Wilson KC, Agusti A, Criner GJ, Macnee W, Make BJ, Rennard SI, and Stockley RA. 2015. An Official American Thoracic Society/european Respiratory Society Statement: Research questions in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 191:e4-e27.
Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, and Ziniti JP. 2011. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Human molecular genetics 21:947-957.
Cho MH, McDonald M-LN, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, DeMeo DL, Sylvia JS, Ziniti J, and Laird NM. 2014. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. The lancet Respiratory medicine 2:214-225.
Cockayne DA, Cheng DT, Waschki B, Hilton H, Kourteva G, Sridhar S, Ravindran P, Bitter H, Visvanathan S, and Mueller K-C. 2011. Whole Blood Gene Expression And Serum Biomarker Profiling In A COPD Cohort Identifies Markers Related To Neutrophilic Inflammation, Wound Healing And Altered B Cell Signaling. C22 PATHOBIOLOGY OF COPD: LESSONS FROM INFLAMMATORY MECHANISMS AND GENOMIC STUDIES: Am
Thoracic Soc, A4091-A4091.

Da Silva CA, Hartl D, Liu W, Lee CG, and Elias JA. 2008. TLR-2 and IL-17A in chitin-induced macrophage activation and acute inflammation. *The Journal of Immunology* 181:4279-4286.

Droemann D, Goldmann T, Tiedje T, Zabel P, Dalhoff K, and Schaaf B. 2005. Toll-like receptor 2 expression is decreased on alveolar macrophages in cigarette smokers and COPD patients. *Respiratory research* 6:68.

Farahi N, Paige E, Balla J, Prudence E, Ferreira RC, Southwood M, Appleby SL, Bakke P, Gulsvik A, Litonjua AA, Sparrow D, Silverman EK, Cho MH, Danesh J, Paul DS, Freitag DF, and Chilvers ER. 2017. Neutrophil-mediated IL-6 receptor trans-signaling and the risk of chronic obstructive pulmonary disease and asthma. *Hum Mol Genet* 26:1584-1596. 10.1093/hmg/ddx053

Freeman CM, Martinez FJ, Han MK, Washko GR, Jr., McCubbrey AL, Chensue SW, Arenberg DA, Meldrum CA, McCloskey L, and Curtis JL. 2013. Lung CD8+ T cells in COPD have increased expression of bacterial TLRs. *Respir Res* 14:13. 10.1186/1465-9921-14-13

Futschik ME, and Carlisle B. 2005. Noise-robust soft clustering of gene expression time-course data. *Journal of bioinformatics and computational biology* 3:965-988.

Gautier L, Cope L, Bolstad BM, and Irizarry RA. 2004. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307-315. 10.1093/bioinformatics/btg405

Gosman MM, Willemse BW, Jansen DF, Lapperre TS, van Schadewijk A, Hiemstra PS, Postma DS, Timens W, and Kerstjens H. 2006. Increased number of B-cells in bronchial biopsies in COPD. *European Respiratory Journal* 27:60-64.

Haw TJ, Starkey MR, Pavlidis S, Fricker M, Arthurs AL, Nair PM, Liu G, Hanish I, Kim RY, Foster PS, Horvat JC, Adcock IM, and Hansbro PM. 2018. Toll-like receptor 2 and 4 have opposing roles in the pathogenesis of cigarette smoke-induced chronic obstructive pulmonary disease. *Am J Physiol Lung Cell Mol Physiol* 314:L298-L317. 10.1152/ajplung.00154.2017

Hersh CP, Hokanson JE, Lynch DA, Washko GR, Make BJ, Crapo JD, Silverman EK, and Investigators C. 2011. Family history is a risk factor for COPD. *Chest* 140:343-350.

Hogg JC, Chu F, Uтокапарч S, Woods R, Elliott WM, Buzatu L, Cherniack RM, Rogers RM, Sciurba FC, and Coxson HO. 2004. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *New England Journal of Medicine* 350:2645-2653.

Huang DW, Sherman BT, and Lempicki RA. 2008. Systematic and integrative analysis of large gene lists using DAVID biominformatics resources. *Nat Protocols* 4:44-57. [http://www.nature.com/nprot/journal/v4/n1/suppinfo/nprot.2008.211_S1.html](http://www.nature.com/nprot/journal/v4/n1/suppinfo/nprot.2008.211_S1.html)

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, and Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249-264. 10.1093/biostatistics/4.2.249

4/2/249 [pii]

Kumar L, and Futschik ME. 2007. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* 2:5.

Lee S-H, Goswami S, Grudo A, Song L-z, Bandi V, Goodnight-White S, Green L, Hacken-Bitar J, Huh J, and Bakaee F. 2007. Antielastin autoimmunity in tobacco smoking–induced emphysema. *Nature medicine* 13:567-569.

McCLOSKEY SC, Patel BD, Hinchliffe SJ, Reid ED, Wareham NJ, and Lomas DA. 2001. Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. *American journal of...*
respiratory and critical care medicine 164:1419-1424.

Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, and Snel B. 2003. STRING: a database of predicted functional associations between proteins. Nucleic acids research 31:258-261.

Nakamoto K, Watanabe M, Sada M, Inui T, Nakamura M, Honda K, Wada H, Ishii H, and Takizawa H. 2019. Pseudomonas aeruginosa-derived flagellin stimulates IL-6 and IL-8 production in human bronchial epithelial cells: A potential mechanism for progression and exacerbation of COPD. Exp Lung Res 45:255-266.

doi:10.1080/01902148.2019.1665147

PAUWELS RA, BUIST AS, CALVERLEY PMA, JENKINS CR, and HURD SS. 2004. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: GOLD Executive Summary Updated 2003. Cptd-journal of Chronic Obstructive Pulmonary Disease 1:105.

Pelanda R, Braun U, Hobeika E, Nussenzweig MC, and Reth M. 2002. B cell progenitors are arrested in maturation but have intact VDJ recombination in the absence of Ig-α and Ig-β. The Journal of Immunology 169:865-872.

Qaseem A, Wilt TJ, Weinberger SE, Hanania NA, Criner G, van der Molen T, Marciñuk DD, Denberg T, Schünemann H, Wedzicha W, MacDonald R, Shekelle P. 2011. Diagnosis and management of stable chronic obstructive pulmonary disease: A clinical practice guideline update from the american college of physicians, american college of chest physicians, american thoracic society, and european respiratory society. Annals of internal medicine 155:179-191. 10.7326/0003-4819-155-3-201108020-00008

Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguezroisin R, and Weel CV. 2013. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. American Journal of Respiratory & Critical Care Medicine 176:532.

Rodriguez-Roisin R, Rabe KF, Vestbo J, Vogelmeier C, and Agustí A. 2017. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 20th anniversary: a brief history of time. Eur Respiratory Soc.

Rovina N, Koutsoukou A, and Koulouris NG. 2013. Inflammation and immune response in COPD: where do we stand? Mediators of inflammation 2013.

Saetta M, Baraldo S, Corbino L, Turato G, Braccioni F, Rea F, CAVALLESCO G, Tropeano G, Mapp CE, and Maestrelli P. 1999. CD8+ ve cells in the lungs of smokers with chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 160:711-717.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research 13:2498-2504. 10.1101/gr.1239303

Simpson JL, McDonald VM, Baines KJ, Oreo KM, Wang F, Hansbro PM, and Gibson PG. 2013. Influence of age, past smoking, and disease severity on TLR2, neutrophilic inflammation, and MMP-9 levels in COPD. Mediators Inflamm 2013:462934. 10.1155/2013/462934

Simpson JL, McDonald VM, Baines KJ, Oreo KM, Wang F, Hansbro PM, and Gibson PG. 2013. Influence of age, past smoking, and disease severity on TLR2, neutrophilic inflammation, and MMP-9 levels in COPD. Mediators Inflamm 2013:462934. 10.1155/2013/462934

Singh D, Fox SM, Tal-Singer R, Bates S, Riley JH, and Celli B. 2014. Altered gene expression in blood and sputum in COPD frequent exacerbators in the ECLIPSE cohort. PloS one 9:e107381.

Singh S, Verma S, Kumar S, Ahmad M, Nischal A, Singh S, and Dixit R. 2018. Correlation of severity of chronic obstructive pulmonary disease with potential biomarkers. Immunology letters 196:1-10.

Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, and Nishimura M. 2013. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. American journal of respiratory and critical care medicine 187:347-365.

von Scheele I, Larsson K, Dahlén B, Billing B, Skedinger M, Lantz AS, and Palmberg L. 2011. Toll-like receptor
expression in smokers with and without COPD. *Respir Med* 105:1222-1230. 10.1016/j.rmed.2011.02.012

Wedzicha JA, and Seemungal TA. 2007. COPD exacerbations: defining their cause and prevention. *Lancet* 370:786-796.

Xu W, Li R, and Sun Y. 2019. Increased IFN-γ-producing Th17/Th1 cells and their association with lung function and current smoking status in patients with chronic obstructive pulmonary disease. *BMC Pulm Med* 19:137. 10.1186/s12890-019-0899-2

Zhang B, Kirov S, and Snoddy J. 2005. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic acids research* 33:W741-W748.

Zhang L, Gu H, Gu Y, and Zeng X. 2016. Association between TNF-α -308 G/A polymorphism and COPD susceptibility: a meta-analysis update. *Int J Chron Obstruct Pulmon Dis* 11:1367-1379. 10.2147/copd.s105394
Figure 1

Results of soft clustering for genes.

Changes in colors (red-blue-green) indicate the coincidence degree of a gene change with the central variation of the cluster. Red indicates a high degree and green indicates a low degree of coincidence. Cluster 1-10 (A-J).
Figure 2

The integrated network of protein-protein interactions and miRNA-target gene regulatory relations

The integrated network of protein-protein interactions and miRNA-target gene regulatory relations for upregulated genes (A) and downregulated genes (B). Red circles indicate upregulated genes, green circles indicate downregulated genes, and blue diamonds indicate miRNA.
The relationships between TLR2 and inflammatory factors in COPD progression.

The levels of IL-6 (A), IL-8 (B), TNF-α (C), and IFN-γ (D) in the serum and TLR2 in the peripheral blood mononuclear cells of the patients (E). The sensitivity and specificity of IL-6 (F), IL-8 (G), TNF-α (H), IFN-γ (I) and TLR2 (J) in the prediction of COPD. The TLR2 expression levels were correlated with the serum levels of IL-6 (K), IL-8 (L), TNF-α (M), and IFN-γ (N).
**Table 1** (on next page)

Upregulated genes with enriched functions of Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG)
| Category | Term     | Description                                      | Count | P value       |
|----------|----------|--------------------------------------------------|-------|---------------|
| BP       | GO:0006952 | defense response                                 | 22    | 3.84E-07      |
| BP       | GO:0006954 | inflammatory response                            | 16    | 5.18E-07      |
| BP       | GO:0009611 | response to wounding                             | 20    | 7.08E-07      |
| BP       | GO:0006955 | immune response                                  | 22    | 2.45E-06      |
| CC       | GO:0009891 | positive regulation of biosynthetic process      | 35    | 0.002989929   |
| CC       | GO:0051173 | positive regulation of nitrogen compound metabolic process | 22    | 0.004929606   |
| CC       | GO:0031328 | positive regulation of cellular biosynthetic process | 22    | 0.006361208   |
| CC       | GO:0045935 | positive regulation of nucleobase, nucleoside, nucleotide, and nucleic acid metabolic process | 51    | 0.006484061   |
| KEGG     | hsa05322  | Systemic lupus erythematosus                     | 8     | 4.11E-04      |
| KEGG     | hsa04620  | Toll-like receptor signaling pathway             | 6     | 0.012939      |
| KEGG     | hsa04060  | Cytokine-cytokine receptor interaction           | 9     | 0.028464      |

BP: biological process; CC: cellular component
**Table 2 (on next page)**

Downregulated genes with enriched functions of Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG)
| Category | Term         | Description                  | Count | P value     |
|----------|--------------|------------------------------|-------|-------------|
| BP       | GO:0046649   | lymphocyte activation        | 11    | 1.89E-06    |
| BP       | GO:0001775   | cell activation              | 12    | 7.79E-06    |
| BP       | GO:0045321   | leukocyte activation         | 11    | 1.08E-05    |
| BP       | GO:0048534   | lymphoid organ development   | 10    | 1.19E-04    |
| BP       | GO:0002520   | immune system development    | 10    | 1.87E-04    |
| BP       | GO:0030097   | hemopoiesis                  | 9     | 3.33E-04    |
| BP       | GO:0030098   | lymphocyte differentiation    | 6     | 9.42E-04    |
| BP       | GO:0042110   | T cell activation            | 6     | 0.002303352 |
| BP       | GO:0002521   | leukocyte differentiation    | 6     | 0.0027276   |
| BP       | GO:0042113   | B cell activation            | 4     | 0.018182557 |
| BP       | GO:0030217   | T cell differentiation       | 3     | 0.081968615 |
| CC       | GO:0009897   | external side of plasma膜    | 4     | 0.120963049 |
| CC       | GO:0009986   | cell surface                 | 5     | 0.235650749 |
| CC       | GO:0044459   | plasma membrane part         | 24    | 0.033231996 |
| CC       | GO:0031226   | intrinsic to plasma membrane | 14    | 0.088759042 |
| CC       | GO:0005886   | plasma membrane              | 34    | 0.099440061 |
| CC       | GO:0005887   | integral to plasma membrane  | 13    | 0.136364905 |
| KEGG     | hsa04940     | Type I diabetes mellitus     | 4     | 0.001194    |
| KEGG     | hsa04630     | Jak-STAT signaling pathway   | 5     | 0.007356    |
| KEGG     | hsa05330     | Allograft rejection          | 3     | 0.014236    |
| KEGG     | hsa05332     | Graft-versus-host disease    | 3     | 0.016588    |
| KEGG     | hsa04660     | T cell receptor signaling pathway | 4 | 0.016985 |

BP: biological process; CC: cellular component
Table 3 (on next page)

Regulatory microRNAs of gene sets
| sequence     | microRNA                                          | count |
|--------------|--------------------------------------------------|-------|
| Hsa-TGGTGCT  | miR-29a, miR-29b, miR-29c                       | 6     |
| Hsa-AAGGGAT  | miR-188                                          | 3     |
| Hsa-GTG CCTT | miR-506                                          | 7     |
| Hsa-ACACTAC  | miR-142-3p                                       | 3     |
| Hsa-TGC ACTG | miR-148a, miR-152, miR-148b                     | 4     |
| Hsa-ACTGAAA  | miR-30a-3p, miR-30e-3p                           | 3     |
| Hsa-GGGACCA  | miR-133a, miR-133b                              | 3     |
| Hsa-CTTTGCA  | miR-527                                          | 3     |
| Hsa-CATT TCA | miR-203                                          | 3     |
| Hsa-TTTGTAG  | miR-520d                                         | 3     |
| Hsa-TAGCTTT  | miR-9                                            | 4     |
| Hsa-AAGCA CA | miR-218                                          | 4     |
| Hsa-TGGTGCT  | miR-29a, miR-29b, miR-29c, miR-17-5p, miR-20a, miR-106a, miR-106b, miR-20b, miR-519d | 4     |
| Hsa-GCAC TT T | miR-519a, miR-519b                              | 3     |
| Hsa-TGC ACT T | miR-519c, miR-519b, miR-519a                    | 3     |
| Hsa-CTTT GTA, | miR-524                                          | 3     |
| Hsa-TGAA GTT | miR-181a, miR-181b, miR-181c, miR-181d           | 3     |
| Hsa-TGCTGCT  | miR-15a, miR-16, miR-15b, miR-195, miR-424, miR-497 | 3     |
Table 4 (on next page)

Results of univariate logistic regression analysis
| Factor   | Coefficient | P value       |
|----------|-------------|---------------|
| Age      | 0.2796      | 0.0389 *      |
| Gender   | 0.1545      | 0.2383        |
| TLR2     | 0.9265      | 2.1068E-07 ***|
| IL-10    | 0.3250      | 0.1524        |
| MMP9     | 0.2702      | 0.0059 **     |
| HMOX1    | 0.3163      | 0.0713        |
| CCR1     | 0.3503      | 0.0298 *      |
| CD79A    | -0.3443     | 0.0002 ***    |
| PLCG1    | -0.7284     | 0.0037 **     |

Factor: factors that may affect clinical grading; coefficient: regression coefficient (> 0, positive correlation; < 0, negative correlation).
Table 5 (on next page)

Results of multivariate logistic regression analysis
| Factor      | Coefficients | P value |
|-------------|--------------|---------|
| Age         | 0.1069       | 0.4220  |
| TLR2        | 0.6894       | 0.0027 **|
| MMP9        | 0.0374       | 0.7310  |
| CCR1        | 0.0282       | 0.8695  |
| CD79A       | -0.1930      | 0.0450 *|
| PLCG1       | -0.1267      | 0.6488  |

Factor: factors that may affect clinical grading; coefficient: regression coefficient (> 0, positive correlation; < 0, negative correlation).