



Identification of Molecular Subtypes of Chronic Obstructive Pulmonary Disease by Gene Expression Profiling

Zhang Pingan¹, Gao Na², Li Xiaoning¹, Ji Guochao¹, Wu Jianjun^{1*}

¹Department of Respiratory Disease, The Third Affiliated Hospital of Beijing University of Traditional Chinese Medicine, Beijing, China;

²Department of Respiratory Disease, Henan University Of Traditional Chinese Medicine, Zhengzhou City, Henan Province, China

ABSTRACT

Background: Chronic Obstructive Pulmonary Disease (COPD) has become the fourth most lethal disease in the world and is expected to rise to the third most lethal disease in the world after 2030. COPD is complex and has clinical heterogeneity. However, identifying the subgroup characteristics of chronic obstructive pulmonary disease has become a challenge.

Objectives: In order to delay the progress of COPD patients and improve their quality of life, we can find patients with different treatment goals and formulate different targeted treatment schemes by studying the differences between different subgroups.

Methods: We obtained the relevant gene chip by searching the Gene Expression Omnibus (GEO) database. 151 patients with COPD obtained from GEO database were divided into three subgroups by consensus clustering. In order to study the differential gene expression patterns between different subgroups, five subgroup specific weighted gene coexpression analysis modules were determined by Weighted Gene Coexpression Analysis (WGCNA).

Results: The characteristics of WGCNA module showed that subjects in subgroup I showed airway remodeling characteristics; Subjects in subgroup II showed metabolic activity; Subjects in subgroup III showed inflammatory characteristics.

Conclusion: This study obtained the clinical subgroup classification of chronic obstructive pulmonary disease through consensus clustering, and found that patients in different subgroups may have unique gene expression patterns, which can help researchers explore new treatment strategies for COPD according to the characteristics of clinical subgroups.

Keywords: COPD; Gene expression profile; WGCNA module; Subgroup analysis; Inflammatory characteristics

INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD), referred to as COPD for short, is a chronic airway inflammatory disease characterized by persistent respiratory symptoms and airflow restriction. Macrophages, neutrophils and inflammatory cells including Tc1, Th1, Th17 and ilc3 lymphocytes in peripheral airways, lung parenchyma and pulmonary vessels increase significantly and release a variety of inflammatory mediators [1]. COPD has been more than 250 million worldwide, which the incidence rate is over 8.6%, 13.7% and 27.4% in 20, 40, and 60 years old. It has become the main cause of global morbidity and mortality with be expected that it will become the third largest cause of death in 2030 [2-4]. Since the classic research of Fletcher C and Peto R, spirometry

has been introduced into the diagnosis of diseases. In clinical practice, FEV1/FVC<70% after inhalation of bronchodilators is used to judge the existence of continuous airflow restriction, which is used as the pulmonary function standard for the diagnosis of COPD [5,6]. FEV1 is often used to evaluate the severity of airflow restriction in clinic and FEV1 accounted for 80%, 50% and 30% of the expected value is corresponding to mild, moderate, severe and very severe COPD, after using bronchodilators [7,8]. The study found that smoking can lead to the frequent and acute exacerbation of COPD. With the increase of smoking history, the increase of smoking times and the increase of "package year" index, the frequency of bronchial obstruction in patients with chronic obstructive pulmonary disease increases, which can be used as a

Correspondence to: Wu Jianjun, Department of Respiratory Disease, The Third Affiliated Hospital of Beijing University of Traditional Chinese Medicine, Beijing, China, E-mail: awusi59@163.com

Received: 25-Apr-2022, Manuscript No. JIDD-22-16319; **Editor assigned:** 27-Apr-2022, PreQC No. JIDD-22-16319 (PQ); **Reviewed:** 16-May-2022, QC No. JIDD-22-16319; **Revised:** 23-May-2022, Manuscript No. JIDD-22-16319 (R); **Published:** 30-May-2022, DOI: 10.35248/2576-389X.22.07.186.

Citation: Pingan Z, Na G, Xiaoning L, Guochao J, Jianjun W (2022) Identification of Molecular Subtypes of Chronic Obstructive Pulmonary Disease by Gene Expression Profiling. J Infect Dis Diagn. 7:186.

Copyright: © 2022 Pingan Z, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

predictor of the increase of mortality of such diseases [9]. Moreover, Smoke Stimulation (CSE) can promote the progress of COPD by down regulating Growth Differentiation Factor 11 (GDF11) and activating the expression of Akt signal transduction pathway [10]. However, in the study of disease progression and adverse prognosis of oligomyosis and COPD, it was found that the incidence of COPD in the low BMI group (7.6%) was significantly higher than that in other groups (3.4%-4.1%, $P < 0.0001$), while the incidence of COPD in the low BMI group (20.1%) was higher than that in other groups (8.4%-12.4%) among participants with smoking history ≥ 30 years, Therefore, it can be found that BMI index is significantly correlated with the risk of COPD occurrence and death [11].

High Throughput Sequencing (HTS) is a representative innovative technology in the emerging biological field in recent years. It is used to study the biomarkers of genes and proteins in human tissue or blood, and can reflect the progress of diseases at the level of genome, epigenome, transcriptome, proteome and metabolome [12,13]. Through the transmission process of genetic data such as transcription, translation and protein modification, high throughput gene sequencing can also analyze the disease risk and response to treatment which is the phenotype of the disease [14,15]. Through genome sequencing, it was found that the expression of IL-1 β , TNF α and IL-17 of neutrophils and eosinophils in different subgroups of COPD was different [16]. Genome Wide Association Analysis (GWAS) found that 37 genetic variants were associated with COPD ($P < 0.05$), in which the C allele of the synonymous variant rs8040868 decreased the cholinergic gene receptor $\alpha 5$ (CHRNA5) expression and was associated with nicotine addiction in chronic obstructive pulmonary disease [17,18]. Genome Wide Association Analysis (GWAS) found that rs2013701 can specifically regulate the allele of FAM13A in 16HBE cells, which is related to the decline of lung function, and the expression of FAM13A can increase the risk of COPD [19]. However, through the difference analysis between COPD and normal control group, we can only find the susceptibility genes of COPD patients, and cannot analyze the heterogeneity between different COPD patients. Studies have shown that the heterogeneity between different tumor patients can be found by dividing the gene expression patterns of tumor patients into subgroups, so as to predict the clinical endpoint and guide the treatment [20]. Therefore, we studied the heterogeneity between different subgroups by dividing the gene expression profiles of COPD cases into subgroups, and used the Kyoto Encyclopedia of Genes and Genomes Database (KEGG) path to annotate the corresponding expression modules to screen the functions of specific genes related to subgroups [21].

MATERIALS AND METHODS

Data collection

Download data: We used R/Bio conductor package GEOquery (201) to extract "Gene Expression Omnibus" (GEO) objects. Searched the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), the search term was "Chronic obstructive pulmonary disease", the gene chip that meets the requirements was included in this study, and the platform files and sequence probe matrix were downloaded separately document. According to the annotation information of the platform file, converted the probe matrix into a gene matrix. Extracted the information about the clinical features in the probe matrix file into the newly created EXCEL as the clinical data file for the research.

Inclusion and exclusion criteria: In inclusion criteria patients with

chronic obstructive pulmonary disease, including clinical gender, age, smoking time, BMI index, FEV1 value and other basic clinical information. The type of study was contralateral gene expression profile. The input type is series. The study included samples from the COPD group and the normal control group. In exclusion criteria, there was no comparison between the disease group and the normal group, and the sample size of each group is less than 20.

Removal of batch effect

Firstly, "limma" package and "SVA" package in R/Bioconductor are used to merge the expression data and perform batch correction [22]. When the data of log 1 is converted, only the data of log 2 is taken as the mean value, and the data of Log 1 is retained. When there is differential batch effect in the data, combat SEQ can obtain better statistical ability and control the false positive rate compared with other available methods. Therefore, we choose the combat method to eliminate the batch effect between platforms. Finally, the R/ggplot2 package is used for principal component analysis to evaluate whether the batch effect is removed.

Consensus clustering

Firstly, the "limma" package and "consensus cluster plus" package in R/Bioconductor package were used for consensus clustering [23], and the included COPD cases were divided into different subgroups. K-means algorithm with Spearman distance is used for clustering. Set the maximum cluster number to 10, and the final cluster number is determined by the consistency matrix and cluster consistency score (> 0.8).

Compare the clinical characteristic subgroups of the three groups

The paired Wilcoxon rank sum test was performed on the data in the three subgroups to detect whether there were differences in the clinical indexes of COPD patients between the age and the number of years of smoking in each subgroup (* represents $P < 0.05$, ** represents $P < 0.01$, *** represents $P < 0.001$).

Extraction of specific up-regulated genes in subtypes

Subgroup specific up-regulated genes were determined by comparing cases in a specific subgroup with cases in other subgroups. It should be noted that Wilcoxon is adopted S-sum rank test was used to test the differential expression. After correction, the threshold of P value was less than 0.05 and the absolute value of mean difference was more than 0.2. For a given gene, the difference in the mean is calculated by subtracting the expression mean of the normal control group from the cases of a specific subgroup.

Gene set enrichment analysis

Whether there are differences in gene enrichment between each gene set and the normal gene set was also observed by gene analysis. Gene set enrichment analysis is implemented in GSEA 4.1.0 in GSEA pre rank mode. Gene set files (genotyping specific gene files) are composed of subgroup specific up-regulated genes. The gene list of each subgroup is sorted by the p value of t-test, which is calculated by comparing the COPD cases of each subgroup with the normal control group.

Weighted gene coexpression network analysis

Weighted Gene Co-Expression Analysis (WGCNA) was used to analyze typing specific genes in subgroups. WGCNA has been proved to be an effective method to detect multiple coexpression

modules and can be used to find clustering (modules) of highly related genes [24]. Find the optimal power value through the power value scatter diagram and calculate the distance between genes. In addition, the average method and dynamic method are used for hierarchical cluster analysis, the cluster diagram is established respectively, the genes are classified, and the similar modules are combined. We finally determine five functional modules, calculate the core function by using the Spearman method in WGCNA, and calculate the Spearman correlation coefficient and its corresponding relationship, and the p value between clinical characteristics and functional modules. At the same time, the labeled heat map function option in WGCNA package is used to draw the heat map, visualize the data, and set the low expression of gene to blue, the middle expression to white, and the high expression to red.

KEGG enrichment analysis

Each functional module of WGCNA is enriched and analyzed in the Kyoto Encyclopedia of Genes and Genomes Database (KEGG), which is a reference knowledge base for the biological interpretation of genome sequences and other high-throughput data. The genome of KEGG pathway is downloaded from MSigDB, and the gene species are selected as human [25]. During KEGG enrichment analysis, set the p-value filtering condition to <0.05 . Visualize the enrichment results and draw bubble diagram.

RESULTS

Microarray data characteristics

Three independent microarray information were included in this study, involving three independent clinical trials, which were taken from geo database, including 277 samples of GSE37768, GSE76925 and GSE130928 (including 151 patients with chronic obstructive pulmonary disease and 126 healthy subjects). All provided clinical information such as age, lung function, smoking and body mass index.

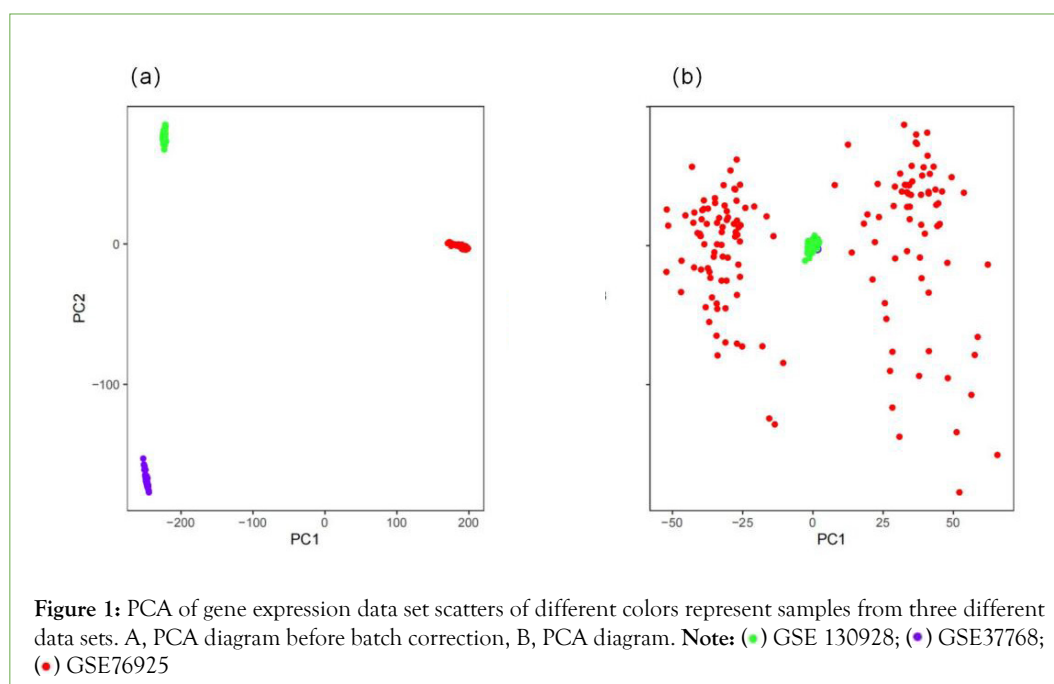
Eliminate batch effect through cross platform standardization

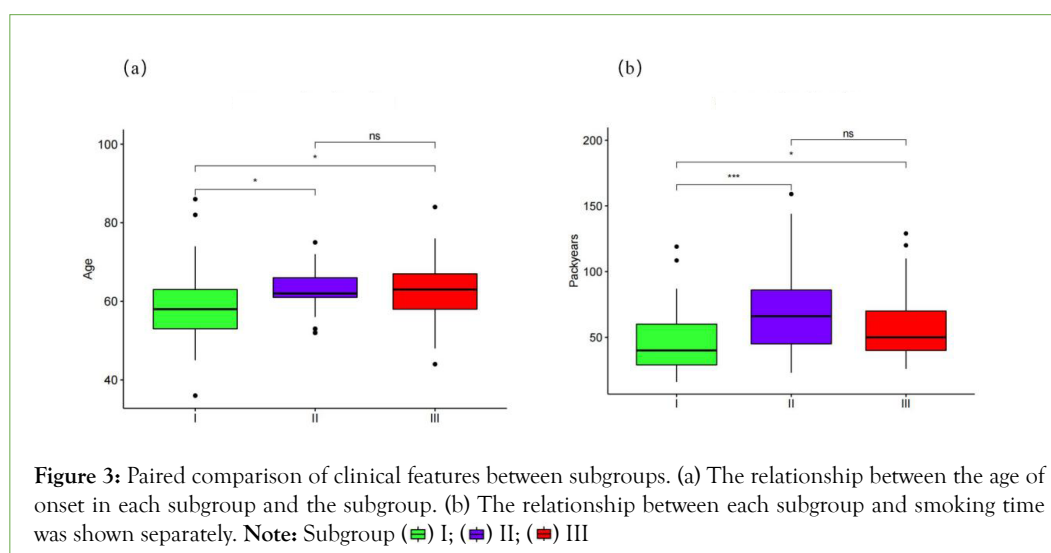
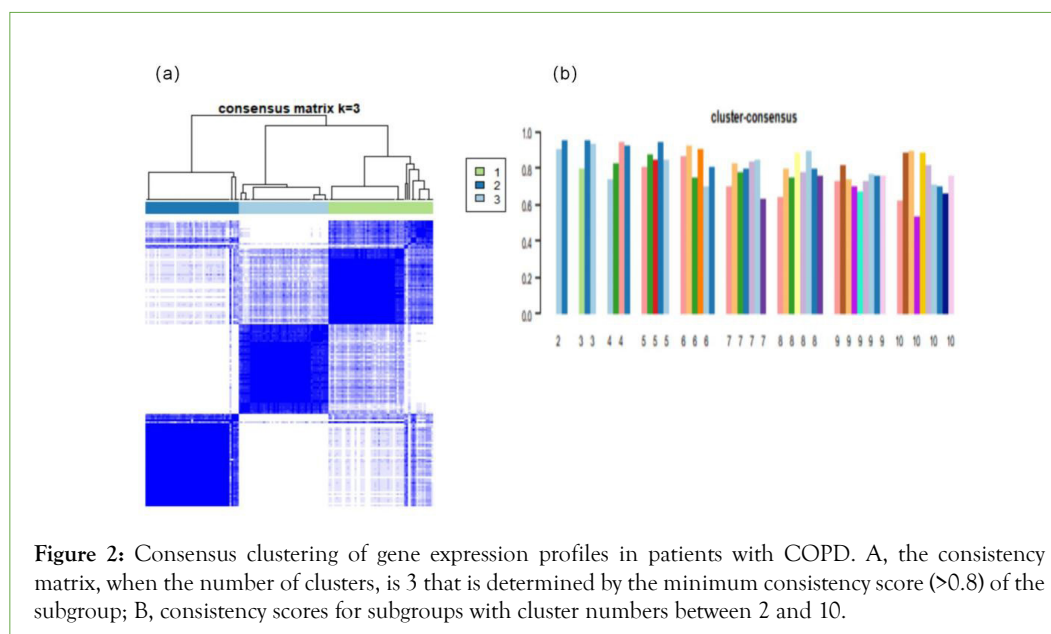
The combat method is used to process data sets from different

platforms and different batches to eliminate the batch effect between different data sets. A total of 15326 genes were detected in three different geo data sets included in this study. Before eliminating the batch effect, the two main components (PCS) of non-standardized expression values were clustered (Figure 1a). In order to eliminate the batch effect; the scatter diagram is drawn after the principal component standardization of the data. The results show that the batch effect of different platforms has been eliminated (Figure 1b).

Consensus clustering of COPD cases

The gene expression matrix after eliminating the batch effect and the sample information of chronic obstructive pulmonary disease were analyzed by cluster analysis. The results showed that 151 patients with chronic obstructive pulmonary disease were divided into different subgroups. According to the consistency cluster analysis of gene expression profile, the results were divided into three subgroups. The number of cases in subgroups I, II and III in the three subgroups were 55, 49 and 47 respectively, and the gene expression patterns were different among the three subgroups (Figure 2a). In contrast, the results based on consistency matrix analysis showed that the gene expression patterns in each subgroup were highly similar. In general, the higher the consistency score, the more stable the grouping. In this study, when the groups are divided into 2 and 3 groups, the clustering scores of each subgroup are greater than 0.8, indicating that the results of these groups are more stable. Therefore, 151 patients with COPD were divided into 3 groups (Figure 2b). By describing the clinical characteristics of the three subgroups and analyzing the data of 151 cases of chronic obstructive pulmonary disease, it was found that the age of subgroup I was younger (Figure 3a), and the age of subgroup II and III was older. Smoking time in subgroups II and III was significantly longer than that in subgroup I (Figure 3b), indicating that subjects in subgroups II and III were heavier in subgroup I, while subgroup I developed earlier. However, no differences were found in age and smoking time in subgroups II and III.





Identification of gene coexpression modules of subgroups

In order to reveal the gene differences between COPD subgroups, WGCNA analysis was carried out under the expression level of specifically up-regulated genes in each subgroup. Through pairwise differential expression analysis between each two subgroups, it was determined that 1076, 4271 and 2339 genes were specifically up-regulated in subgroups I, II and III respectively (the corrected threshold was $p < 0.05$, and the absolute difference of the mean value was >0.2). Differential expression analysis was performed by comparing the gene expression profiles of each subgroup with those of the normal control group. The results of GSEA enrichment analysis showed that there was also significant difference between the specific up-regulated genes in the subtypes and the normal samples (Figures 4a-4c, $FDR < 0.05$). It is worth noting that subgroup I has the least subgroup specific up-regulated genes compared with other subgroups, and the number of differentially expressed genes is relatively small by comparing the cases of each subgroup with the normal control group.

Based on the expression levels of 7686 specifically up-regulated genes in the subgroup, a gene heat map expression network was constructed, which identified five WGCNA modules (Figure 4d

and Table 1). By analyzing the relationship between WGCNA module and corresponding subgroups and KEGG enrichment, it was found that PPAR signaling pathway, regulation of actin cytoskeleton and platelet activation pathway were significantly enriched only in brown module (Figure 5a), and brown module was significantly enriched in sub I, indicating that subjects in sub group I showed airway remodeling characteristics; Glycerol phospholipid metabolism, peroxisome and neuroactive ligand receptor interaction are only significantly enriched in the blue module, trap interaction in vesicle transport and bacterial invasion of epithelial cells are only significantly enriched in the Yellow module, vascular smooth muscle contraction is only significantly enriched in the blue and yellow modules, and yellow and blue modules are significantly enriched in subgroup II, The results showed that the subjects in subgroup II showed the characteristics of metabolic activity; JAK-STAT signal pathway, PI3K Akt signal pathway and HIF-1 signal pathway are significantly enriched in the green module, and the green module is significantly enriched in subgroup III, indicating that the subjects in subgroup III show inflammatory characteristics. However, the gray module did not find significant enrichment in the corresponding subgroups. These findings suggest that each subgroup has its specific functional modules or pathways that can regulate the occurrence or progression of COPD.

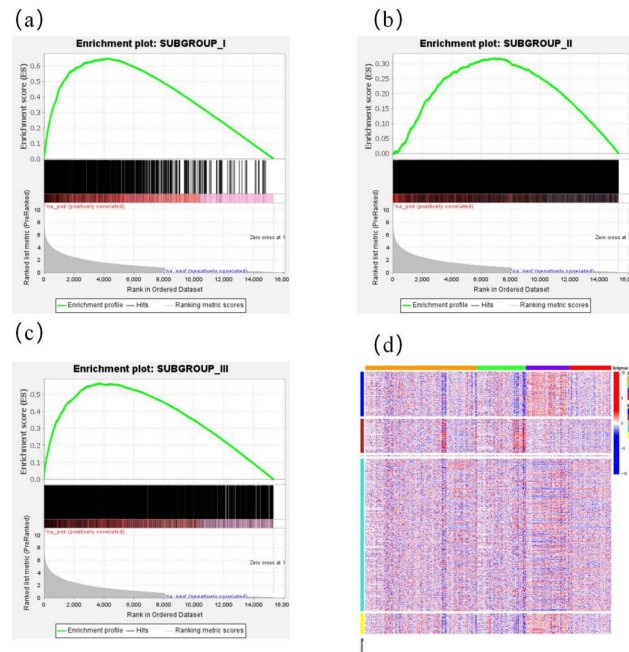


Figure 4: The expression patterns of subgroup-specific up regulated genes. The enrichment plots of (a), (b), and (c) illustrate the subgroup-specific up regulated genes are also expressed higher in the corresponding subgroup than the normal controls. (d) The scaled expression values of genes that comprise each of the five weighed gene co-expression network analysis modules are displayed in the heat-map. **Note:** (■) Enrichment profile; (■) Hits; (■) Ranking metric scores

Table 1: The number of differentially expressed genes by case-control and case-case comparisons and weighted gene co-expression analysis modules in each subgroup.

Transcriptome classification	#Of up regulated genes by case-control comparison	#Of down regulated genes by case-control comparison	#Of subgroup-specific up regulated genes by case-control comparison	Modules
I	74	5220	1076	Brown
II	418	1257	4271	Blue and Yellow
III	562	28	2339	Turquoise

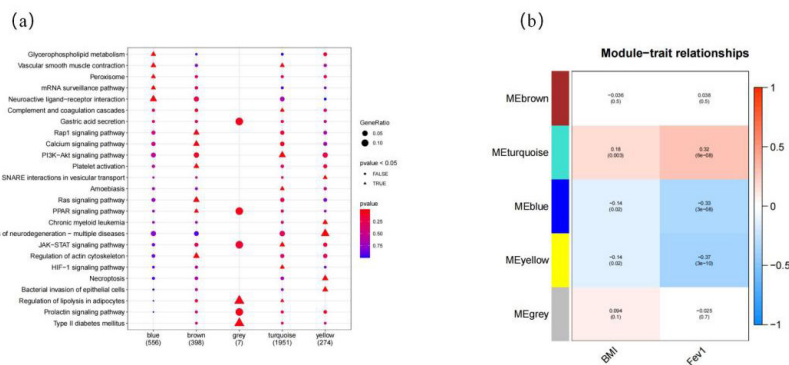


Figure 5: Functional characterization and clinical association of WGCNA modules. (a) The heat-map shows the negative log₁₀ p-value significance of top 20 of the significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for each WGCNA module. Over representation of WGCNA modules in KEGG pathways was evaluated by overrepresentation enrichment analysis. (b) The negative correlation coefficients of positive and WGCNA modules with clinical characteristics, BMI index and FEV1 were expressed in red and green, WGCNA: Weighted Gene Coexpression Network Analysis.

Clinical features and WGCNA module discussion

Through the correlation module, correlation coefficient and corresponding p value of clinical features and WGCNA, the relationship between the characteristic genes of each module and BMI index or FEV1 is analyzed and calculated. It is found that the characteristic genes are represented by the characteristic vector of gene expression matrix. Brown and gray modules were not associated with BMI index or FEV1 (Figures 5a and 5b), indicating that the airway remodeling process in subgroup I was not significantly correlated with BMI index or FEV1. In contrast, blue and yellow modules were negatively correlated with BMI index or FEV1, and the difference was statistically significant, indicating that patients with low BMI index and high FEV1 were metabolically active in subgroup II; The green module was positively correlated with BMI index or FEV1, and the difference was statistically significant, indicating that patients in subgroup III showed a continuously enhanced inflammatory response with the increase of BMI index and the decrease of FEV1. Our results further suggest that the WGCNA module is associated with some clinical features, such as BMI index or FEV1.

DISCUSSION

In recent years, subgroup analysis has been widely used to screen biomarkers related to clinical traits in cancer, which is conducive to analyze the molecular function of biomarkers in different subgroups in cancer. In addition, the association between gene differences between subgroups and internal or external factors can also be analyzed. For example, subgroup analysis found that HNF4 α can activate the classical gene expression program and inhibit the expression of SIX4 and SIX1 to inhibit tumor growth in the growth and molecular subtype of pancreatic ductal carcinoma [26]. In addition to cancer research, other diseases can also be analyzed by clinical variables and transcriptional differences or subtypes, such as preeclampsia, psoriasis, coronary heart disease and other diseases [27,28]. Although these studies have some limitations, they have improved our new understanding of disease research methods. The clinical manifestations of COPD are very complex and have obvious clinical heterogeneity. Different from the previous studies that simply analyzed the gene expression of COPD and normal control group, this study further divided the cases of COPD into subgroups. The results showed that different subgroups showed different clinical characteristics. For example, according to the comparison of onset age and smoking time, it is found that patients in subgroup I have an earlier onset, while patients in subgroup II and III have a later onset. However, the smoking time of subgroups II and III was longer, and the decline of FEV1 was faster, suggesting that the patients in subgroups II and III were faced with severe disease. In addition, we not only proved the subgroup specific functional modules and pathways and the regulatory pathways related to COPD, but also constructed the relationship between specific pathways and specific subgroups. For example, Peroxisome Proliferator Activated Receptor (PPAR) is a receptor in the nucleus, and its family member PPAR γ Agonists can inhibit the proliferation of human airway smooth muscle cells and reduce airway remodeling [29]. In our study, PPAR signaling pathway was most significantly enriched in subgroup I. In addition, PPAR- γ can also inhibit the activation of TGF- β 1-smad2/3 pathway, which plays an important role in epithelial mesenchymal transformation and airway remodeling [30]. Studies have shown that artesunate can significantly intervene airway inflammation induced by CSE through PPAR- γ /TGF- β 1/Smad2/3 signaling

pathway to inhibit cell proliferation and improve airway remodeling in rats *in vivo* and *in vitro* [31]. Pioglitazone (PGZ), an agonist of PPAR γ , can regulate the expression of its downstream protein G protein signal regulator and inhibit airway remodeling induced by ovalbumin in asthmatic mice, accompanied by IL-4, IL-13, IL-17 and IFN- γ and the decrease of serum IgE level [32]. At the same time, the regulatory signal pathway of actin cytoskeleton was also significantly enriched in subgroup I. FAM13A is a modified gene of cystic fibrosis lung disease phenotype, which can regulate RhoA activity, actin cytoskeleton dynamics and epithelial mesenchymal transformation to change the remodeling of cytoskeleton in many lung diseases such as COPD, asthma and pulmonary hypertension, which is related to the assembly of focal adhesion and F-actin stress fibers [33,34]. However, subgroup II showed significant enrichment of glycerol phospholipid metabolism, peroxisome, neuroactive ligand receptor interaction and other pathways, and showed the characteristics of active metabolism. Glycerol phospholipids (mainly including LPA, LPS, LPG and LPC) are the main components of cell membrane and play an important role in cell growth, migration, signal recognition and apoptosis. Recent studies have shown that glycerophosphate is involved in the pathogenesis of lung diseases such as pulmonary infection, asthma and COPD, and is related to the disorder of lipid metabolism of alveolar surfactant. Cigarette smoke can induce alveolar surfactant dysfunction, alveolar epithelial cell apoptosis and emphysema [35,36]. The study found that lysophosphatidylcholinyltransferase (lpcat) can participate in the conversion of LPC and acyl CoA into PC and free CoA, thereby reducing the concentration of endogenous LPC, while lpcat gene is highly expressed in COPD patients, which is related to the severity of FEV1% PRED lung function [37]. In subgroup III, we found that JAK-STAT signal pathway, PI3K Akt signal pathway and HIF-1 signal pathway were significantly enriched and showed significant inflammatory characteristics. Among them, phospholipase A2 receptor 1 (pla2r1) in COPD patients can promote inflammatory response and accelerate the formation of emphysema, pulmonary fibrosis and pulmonary hypertension through Janus Kinase (JAK)/ Signal Transducer And Activator Of Transcription (STAT) signal transduction. Rusotinib, an inhibitor of JAK1, can alleviate this process [38]. Chemokine (CCR1) can regulate JAK/STAT3/NF- κ B signaling pathway in mediating the inflammatory response caused by smoking and can induce IL-8, IL-6, TNF- α and other chronic inflammatory factors [39]. Recent studies have shown that HIF-1 α overexpressed in the serum of patients with COPD can raise the expression of inflammatory factors by activating EGFR/PI3K/Akt pathway. Meanwhile, pulmonary inflammation can activate EGFR/PI3K/Akt pathway to feedback up regulate HIF-1 α thus aggravate the inflammatory response of COPD [40]. Finally, studies have demonstrated that environmental particles (PM) can regulate Amphiregulin (areg) and Epidermal Growth Factor Receptor (EGFR) through PI3K-Akt pathway to promote inflammation and mucus hyper secretion in COPD and asthma. Combined with the analysis of clinical characteristics, it can be found that the subjects in subgroup I smoked for a short time, had a young age of onset, and showed the characteristics of early airway remodeling, Subjects in subgroups II and III smoked for a longer time, had a later age of onset, showed active metabolism and inflammatory characteristics, and could cause more severe COPD subgroups.

In summary, through the above results, this paper further proves that different subgroups represent different development stages and internal biological characteristics of COPD. We look forward

to large sample prospective trials in the future.

CONCLUSION

In conclusion, inspired by the study of cancer subgroups, we adopted a similar strategy to reveal the molecular subgroups of COPD. The ultimate goal of studying these different subgroups is to find patient groups with unique treatment characteristics and formulate targeted treatment plans. Our study suggests that patients in different subgroups may have unique gene expression patterns. This new classification method is helpful for researchers to explore new treatment strategies for COPD according to the characteristics of clinical subgroups, improve the prognosis of the disease and improve the quality of life of patients.

ACKNOWLEDGEMENT

Not applicable

AUTHOR CONTRIBUTIONS

PAZ conceived the study and contributed to data collation, original draft preparation and draft review. Ng and XNL contributed to writing, commenting, editing and supervision. Gcj and JJW contributed to writing the first draft. All the authors participated in the experimental design, writing manuscripts, manuscript modification and translation.

FUNDING

The work was supported by the Beijing Natural Science Foundation (7182100).

DECLARATION

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

REFERENCES

- Polosukhin VV, Richmond BW, Du RH. Secretory IgA deficiency in individual small airways is associated with persistent inflammation and remodeling. *Am J Respir Crit Care Med.* 2017;195(8):1010-1021.
- Lareau SC, Fahy B, Meek P, Wang A. Chronic obstructive pulmonary disease (COPD). *Am J Respir Crit Care Med.* 2019;199(1):P1-P2.
- Wang C, Xu J, Yang L. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): A national cross-sectional study. *Lancet* 2018;391(2):1706-1717.
- Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 2006;3(11):e442.
- BR Celli, W MacNee. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J.* 2004;23(6):932-946.
- Bhatt SP, Balte PP, Schwartz JE. Discriminative accuracy of FEV1: FVC thresholds for COPD-related hospitalization and mortality. *JAMA.* 2019;321(24):2438-2447.
- Zhou M, Wang H, Zeng X. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2019;94(10204):1145-1158.
- Boiko OO, Rodionova VV. The effect of smoking on nutritional status severity of the disease and the development of systemic effects in patients with chronic obstructive pulmonary disease. *Wiad Lek.* 2021;74(1):52-56.
- Tang F, Ling C, Liu J. Reduced expression of growth differentiation factor 11 promoted the progression of chronic obstructive pulmonary disease by activating the AKT signaling pathway. *Biomed Pharmacother.* 2018;103(3):691-698.
- Park HJ, Cho JH, Kim HJ, Park JY, Lee HS, Byun MK. The effect of low body mass index on the development of chronic obstructive pulmonary disease and mortality. *J Intern Med.* 2019;286(5):573-582.
- Haberland V, Browne F, Taggart L, Zheng H, Parkes E, Blayney JK. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Brief Bioinform.* 2019;20(5):1795-1811.
- NCI. Definition of personalized medicine-National Cancer Institute Dictionary of Cancer Terms. 2017.
- Alberts B, Johnson A, Lewis J (2002) *Molecular Biology of the Cell.* (4th edn), Garland Science, New York, United States.
- Gibney ER, Nolan CM. Epigenetics and gene expression. *Heredity.* 2010;105(1):4-13.
- Wang Z, Locantore N, Haldar K, Ramsheh MY, Beech AS, Ma W. Inflammatory end type-associated airway micro biome in chronic obstructive pulmonary disease clinical stability and exacerbations: A multicohort longitudinal analysis. *Am J Respir Crit Care Med.* 2021;203(12):1488-1502.
- Matsson H, Soderhall C, Einarsdottir E, Lamontagne M, Gudmundsson S, Backman H. Targeted high-throughput sequencing of candidate genes for chronic obstructive pulmonary disease. *BMC Pulm Med.* 2016;16(1):146.
- Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP. Variants in FAM13A is associated with chronic obstructive pulmonary disease. *Nat Genet.* 2010;42:200-202.
- Castaldi PJ, Guo F, Qiao D, Du F, Naing ZC, Li Y. Identification of functional variants in the FAM13A chronic obstructive pulmonary disease genome-wide association study locus by massively parallel reporter assays. *Am J Respir Crit Care Med.* 2019;199(1):52-61.
- Giordano TJ. The cancer genome atlas research network: A sight to behold. *Endocr Pathol.* 2014;25(4):362-365.
- Davis S, Meltzer PS. GEO query: A bridge between the Gene Expression Omnibus (GEO) and bio conductor. *Bioinformatics.* 2007;23(14):1846-1847.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28(6):882-883.
- Wilkerson MD, Hayes DN. Consensus Cluster Plus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010;26(12):1572-1573.
- Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* 2008;9(4):559.
- Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 2019;47(D1):D590-D595.
- Camolotto SA, Belova VK, Torre-Healy L, Vahrenkamp JM, Berrett KC, Conway H. Reciprocal regulation of pancreatic ductal adenocarcinoma growth and molecular subtype by HNF4 and SIX1/4. *Gut.* 2021;70(5):900-914.

26. Peng XY, Wang Y, Hu H, Zhang XJ, Li Q. Identification of the molecular subgroups in coronary artery disease by gene expression profiles. *J Cell Physiol.* 2019;234(9):16540-16548.
27. Aibar S, Abaigar M, Laborie C, Santos S. Identification of expression patterns in the progression of disease stages by integration of transcriptomic data. *BMC Bioinform.* 2016;17(Suppl 15):432.
28. Liu L, Pan Y, Zhai C, Zhu Y, Ke R, Shi W. Activation of peroxisome proliferation-activated receptor- γ inhibits transforming growth factor- β 1-induced airway smooth muscle cell proliferation by suppressing Smad-miR-21 signaling. *J Cell Physiol.* 2018;234(1):669-681.
29. Zhao C, Chen W, Yang L, Chen L, Stimpson SA, Diehl AM. PPAR gamma agonists prevent TGFbeta1/Smad3-signalling in human hepatic stellate cells. *Biochem Biophys Res Commun.* 2006;350(2):385-391.
30. Pan K, Lu J, Song Y. Artesunate ameliorates cigarette smoke-induced airway remodelling via PPAR- γ /TGF- β 1/Smad2/3 signalling pathway. *Respir Res.* 2021;22(1):91.
31. Meng X, Sun X, Zhang Y, Shi H, Deng W, Liu Y. PPAR γ Agonist PGZ Attenuates OVA-Induced Airway Inflammation and Airway Remodeling via RGS4 Signaling in Mouse Model. *Inflammation.* 2018;41(6):2079-2089.
32. A Nusrat, M Giry, Turner JR, Colgan SP, Parkos CA. Rho protein regulates tight junctions and perijunctional actin organization in polarized epithelia. *Proc Natl Acad Sci USA.* 1995;92(23):10629-10633.
33. Corvol H, Rousselet N, Thompson KE, Berdah L, Cottin G, Foussigniere T. FAM13A is a modifier gene of cystic fibrosis lung phenotype regulating rhoa activity, actin cytoskeleton dynamics and epithelial-mesenchymal transition. *J Cyst Fibros.* 2018;17(2):190-203.
34. Telenga ED, Hoffmann RF, Kindt R, Hoonhorst SJ, Willemse BW. Untargeted lipidomic analysis in chronic obstructive pulmonary disease uncovering sphingolipids. *Am J Respir Crit Care Med.* 2014;190:155-164.
35. Cruickshank-Quinn CI, Jacobson S, Hughes G, Powell RL, Petrache I, Kechris K. Metabolomics and transcriptomics pathway approach reveals outcome-specific perturbations in COPD. *Sci Rep.* 2018;8:17132.
36. Wang B, Tontonoz P. Phospholipid remodeling in physiology and disease. *Annu Rev Physiol.* 2019;81:165-188.
37. Beaulieu D, Attwe A, Breau M, Lipskaia L, Marcos E, Born E. Phospholipase A2 receptor 1 promotes lung cell senescence and emphysema in obstructive lung disease. *Eur Respir J.* 2021;58(2):2000752.
38. Zhao K, Dong R, Yu Y, Tu C, Li Y, Cui Y. Cigarette smoke-induced lung inflammation in COPD mediated via CCR1/JAK/STAT/NF- κ B pathway. *Aging (Albany NY).* 2020;12(10):9125-9138.
39. Zhang HX, Yang JJ, Zhang SA, Zhang SM, Wang JX, Xu ZY. HIF-1 α promotes inflammatory response of chronic obstructive pulmonary disease by activating EGFR/PI3K/AKT pathway. *Eur Rev Med Pharmacol Sci.* 2018;22(18):6077-6084.
40. Wang J, Zhu M, Wang L, Chen C, Song Y. Amphiregulin potentiates airway inflammation and mucus hypersecretion induced by urban particulate matter via the EGFR-PI3K/AKT/ERK pathway. *Cell Signal.* 2019;53:122-131.