Identification of molecular subtypes of chronic obstructive pulmonary disease by gene expression profiling

zhangpingan (1157401533@qq.com)
Na Gao
Xiaoning Li
Guochao Ji
Jianjun Wu

Research Article

Keywords: COPD, Gene Expression Profile, WGCNA module, Subgroup Analysis, Inflammatory Characteristics

Posted Date: April 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1540944/v1

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

The ultimate goal of studying these different subgroups is to find patients with unique treatment goals and formulate targeted treatment plans to improve the prognosis of the disease and improve the quality of life of patients.

Methods

We obtained the relevant gene chip by searching the gene expression omnibus (GEO) database. According to the gene expression profile, 151patients with chronic obstructive pulmonary disease were divided into three subgroups, which showed different expression patterns. Therefore, we used weighted gene coexpression analysis (WGCNA) to identify the differences between different subgroups and identified five subgroup specific weighted gene coexpression analysis modules.

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.

Conclusions

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.

1 Background

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, 60 yeras 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 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-1β、TNFα 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 receptorα 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, most studies only pay attention to the differential genes between COPD and control group and rarely pay attention to the differential genes between COPD patients.In cancer research, tumor samples are usually divided into several subtypes according to gene expression patterns, which can reveal the heterogeneity between tumors,predict clinical endpoints and guide treatment(20).In order to promote the understanding of the classification of chronic obstructive pulmonary disease, we divided the cases of chronic obstructive pulmonary disease into three subgroups according to the gene expression profile, and screened the functions of specific genes related to subgroups by annotating the corresponding

coexpression function modules with the path of Kyoto Encyclopedia of genes and genomes database (KEGG).

2 Materials And Methods

2.1 Data collection

2.1.1 Download data

We used R/Bioconductor package GEOqueryt(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.

2.1.2 inclusion and exclusion criteria

Inclusion criteria: (1) patients with chronic obstructive pulmonary disease, including clinical gender, age, smoking time, BMI index, FEV1 value and other basic clinical information; (2) The type of study was contralateral gene expression profile; (3) The input type is series; (4) The study included samples from the COPD group and the normal control group. Exclusion criteria: (1) there was no comparison between the disease group and the normal group; (2) The sample size of each group is less than 20.

2.2 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 Log1 is converted, only the data of log2 is taken as the mean value, and the data of Log1 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.

2.3 Consensus clustering

Firstly, the "limma" package and "consensus cluster plus" package in R / Bioconductor package were used for consensus clusteringt(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).

2.4 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).

2.5 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.

2.6 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 prerank 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.

2.7 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 genest(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 cor 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.

2.8 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.

3 Results

3.1 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.

3.2 Eliminate batch effect through cross platform standardization

In order to eliminate the batch effect from different platforms and different batches, we use the combat method to eliminate the batch effect between data sets. A total of 15326 genes were detected on three microarray platforms. Before eliminating the batch effect, the samples were clustered by batch according to the first two main components (PCS) of the non standardized expression value (Fig. 1a). In contrast, the scatter diagram is standardized based on principal component analysis. The results show that the samples of three batches are mixed together, indicating that the batch effect caused by different platforms has been thoroughly understood (Fig. 1b). The results show that cross platform standardization successfully eliminates the batch effect after batch correction.

3.3 Consensus clustering of COPD cases

The expression files corrected by batch effect and the sample information of disease group were used for cluster analysis. 151 patients with COPD were divided into subgroups.According to the consistency scoring in data statistics, the consistency cluster analysis of gene expression profile is divided into three subgroups, in which the number of cases in subgroups I, II and III are 55, 49 and 47 respectively, which have significantly different expression patterns (Fig. 2a).On the contrary, based on the consistency matrix, a high similarity of gene expression patterns was observed in each subgroup. Generally speaking, the higher the consistency score, the more stable the high score type is.In the results of this study, when divided into 2 groups and 3 groups, the cluster consistency score of each subgroup is higher than 0.8, which indicates that these classifications are more robust than other clusters, and it is better when considering more groups (Fig. 2b). Therefore, 151 patients with chronic obstructive pulmonary disease were divided into 3 subgroups.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 (Fig. 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 (Fig. 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.

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

Transciptome classification	#of upregulated genes by case-control comparison	#of downregulated genes by case-control comparison	#of subgroup-specific upregulated genes by case-control comparison	Modules
□	74	5220	1076	Brown
□	418	1257	4271	Blue and Yellow
□	562	28	2339	Turquoise

3.4 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 the paired differential expression analysis between each two subgroups, 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 average 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 (Fig. 4a-c, FDR < 0.05). It is worth noting that the number of up-regulated genes in each subgroup I is relatively less than that in the 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 (Fig. 4D□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 (Fig. 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.

3.5 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 (Fig. 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.

4. 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 HNF4α 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 It 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-1α 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-1α 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 hypersecretion in COPD and asthma (41).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.

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.

Declarations

Footnote:

Ethical approval and consent participation: in this study, relevant data were downloaded from GEO database for comprehensive analysis. It does not involve intervention in humans and animals; Therefore, there is no informed consent.

Agree to publish: all authors agree to publish this article.

Supporting data: the data provided in this study can be publicly obtained in one of the repositories, or at any time during submission or after publication, the representative of this journal can obtain it at the request of the corresponding author. Otherwise, all consequences of possible or future withdrawal will be borne by the corresponding author.

Usability competitive interest: the author declares that there is no conflict of interest in the publication of this paper.

Funding: this work is supported by the Beijing Natural Science Foundation (7182100).

Author's contribution: 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.

Thanks: thank the above-mentioned authors for their contributions to this article, especially Professor Wu Jianjun for his support.

Author's information: corresponding author, Wu Jianjun, awusi59@163.com

References

1. Polosukhin VV, Richmond BW, Du RH, et al. 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.
2. World Health Organization. Geneva, Switzerland: World Health Organization; 2017. Chronic obstructive pulmonary disease. [accessed 2019 Oct 6].
3. Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH]study): a national cross-sectional study. Lancet 2018;39l:1706–17.
4. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006;3(11):e442. doi:10.1371/journal.pmed.0030442.
5. B.R. Celli, W. MacNee Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper Eur. Respir. J., 23 (2004), pp. 932–946.
6. Bhatt SP, Balte PP, Schwartz JE, et al. Discriminative Accuracy of FEV1: FVC Thresholds for COPD-Related Hospitalization and Mortality[J]. JAMA, 2019, 321(24):2438-2447. DOI: 10.1001/jama.2019.7233.
7. Zhou M, Wang H, Zeng X, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. Lancet, 2019, 94(10204):1145–1158. DOI:10.1016/S0140-6736(19)30427-1
8. Chronic obstructive pulmonary disease group of respiratory branch of Chinese Medical Association, working committee of chronic obstructive pulmonary disease of respiratory branch of Chinese Medical Association Guidelines for the diagnosis and treatment of chronic obstructive pulmonary disease (revised in 2021) [J] Chinese Journal of tuberculosis and respiration, 2021,44 (03): 170–205.
9. 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. PMID: 33851587.
10. 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 Jul;103:691–698. doi: 10.1016/j.biopha.2018.04.091. Epub 2018 Apr 24. PMID: 29680737.
11. 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 Nov;286(5):573–582. doi: 10.1111/joim.12949. Epub 2019 Jul 4. PMID: 31215064.
12. Lightbody G, 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 Sep 27;20(5):1795–1811.
13. NCI. Definition of personalized medicine—National Cancer Institute Dictionary of Cancer Terms. NCI, 2017.

14. Alberts B, Johnson A, Lewis J, et al.Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002.
15. Gibney ER, Nolan CM.. Epigenetics and gene expression. Heredity 2010;105(1):4–13.
16. Wang Z, Locantore N, Haldar K, Ramsheh MY, Beech AS, Ma W, Brown JR, Tal-Singer R, Barer MR, Bafadhel M, Donaldson GC, Wedzicha JA, Singh D, Wilkinson TMA, Miller BE, Brightling CE. Inflammatory Endotype-associated Airway Microbiome in Chronic Obstructive Pulmonary Disease Clinical Stability and Exacerbations: A Multicohort Longitudinal Analysis. Am J Respir Crit Care Med. 2021 Jun 15;203(12):1488–1502.
17. Matsson H, Söderhäll C, Einarsdottir E, Lamontagne M, Gudmundsson S, Backman H, Lindberg A, Rönmark E, Kere J, Sin D, Postma DS, Bossé Y, Lundbäck B, Klar J. Targeted high-throughput sequencing of candidate genes for chronic obstructive pulmonary disease. BMC Pulm Med. 2016 Nov 11;16(1):146.
18. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet. 2010;42:200–2.
19. Castaldi PJ, Guo F, Qiao D, Du F, Naing ZZC, Li Y, Pham B, Mikkelsen TS, Cho MH, Silverman EK, Zhou X. 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 Jan 1;199(1):52–61.
20. Giordano TJ. The cancer genome atlas research network: a sight to behold. Endocr Pathol. 2014 Dec;25(4):362–5.
21. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007 Jul 15;23(14):1846–7.
22. 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 Mar 15;28(6):882–3.
23. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010 Jun 15;26(12):1572–3.
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9:559.
25. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019 Jan 8;47(D1):D590-D595.
26. Camolotto SA, Belova VK, Torre-Healy L, Vahrenkamp JM, Berrett KC, Conway H, Shea J, Stubben C, Moffitt R, Gertz J, Snyder EL. Reciprocal regulation of pancreatic ductal adenocarcinoma growth and molecular subtype by HNF4α and SIX1/4. Gut. 2021 May;70(5):900–914.
27. 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 Feb 25.
28. Aibar, S., Abaigar, M., Campos-Laborie, F. J., Sánchez-Santos, J. M., Hernandez-Rivas, J. M., & De Las Rivas, J. (2016). Identification of expression patternsin the progression of disease stages by

integration of transcriptomic data.BMC Bioinformatics, 17(Suppl 15), 432.

29. Liu L, Pan Y, Zhai C, Zhu Y, Ke R, Shi W, et al. Activation of peroxisome proliferation-activated receptor-γ inhibits transforming growth factor-β1-induced airway smooth muscle cell proliferation by suppressing Smad-miR-21 signalling. J Cell Physiol. 2018;234(1):669–681.
30. Zhao C, Chen W, Yang L, Chen L, Stimpson SA, Diehl AM. PPARgamma agonists prevent TGFbeta1/Smad3-signalling in human hepatic stellate cells. Biochem Biophys Res Commun. 2006;350(2):385–391.
31. Pan K, Lu J, Song Y. Artesunate ameliorates cigarette smoke-induced airway remodelling via PPAR-γ/TGF-β1/Smad2/3 signalling pathway. Respir Res. 2021 Mar 23;22(1):91.
32. Meng X, Sun X, Zhang Y, Shi H, Deng W, Liu Y, Wang G, Fang P, Yang S. PPARγ Agonist PGZ Attenuates OVA-Induced Airway Inflammation and Airway Remodeling via RGS4 Signaling in Mouse Model. Inflammation. 2018 Dec;41(6):2079–2089.
33. A. Nusrat, M. Giry, J.R. Turner, S.P. Colgan, C.A. Parkos, D. Carnes, et al.Rho protein regulates tight junctions and perijunctional actin organization in polarized epitheliaProc Natl Acad Sci U S A, 92 (23) (1995), pp. 10629–10633
34. Corvol H, Rousselet N, Thompson KE, Berdah L, Cottin G, Foussigniere T, Longchampt E, Fiette L, Sage E, Prunier C, Drumm M, Hodges CA, Boëlle PY, Guillot L. FAM13A is a modifier gene of cystic fibrosis lung phenotype regulating rhoa activity, actin cytoskeleton dynamics and epithelial-mesenchymal transition. J Cyst Fibros. 2018 Mar;17(2):190–203.
35. Telenga E. D., Hoffmann R. F., t'Kindt R., Hoonhorst S. J., Willemse B. W., van Oosterhout A. J., et al. (2014). Untargeted lipidomic analysis in chronic obstructive pulmonary disease uncovering sphingolipids.Am. J. Respir. Crit. Care Med. 190 155–164. 10.1164/rccm.201312-2210OC.
36. Cruickshank-Quinn C. I., Jacobson S., Hughes G., Powell R. L., Petrache I., Kechris K., et al. (2018). Metabolomics and transcriptomics pathway approach reveals outcome-specific perturbations in COPD. Sci. Rep. 8:17132. 10.1038/s41598-018-35372-w.
37. Wang B., Tontonoz P. (2019). Phospholipid remodeling in physiology and disease. Annu. Rev. Physiol. 81 165–188. 10.1146/annurev-physiol-020518-114444
38. Beaulieu D, Attwe A, Breau M, Lipskaia L, Marcos E, Born E, Huang J, Abid S, Derumeaux G, Houssaini A, Maitre B, Lefevre M, Vienney N, Bertolino P, Jaber S, Noureddine H, Goehrig D, Vindrieux D, Bernard D, Adnot S. Phospholipase A2 receptor 1 promotes lung cell senescence and emphysema in obstructive lung disease. Eur Respir J. 2021 Aug 12;58(2):2000752.
39. Zhao K, Dong R, Yu Y, Tu C, Li Y, Cui Y, Bao L, Ling C. Cigarette smoke-induced lung inflammation in COPD mediated via CCR1/JAK/STAT /NF-κB pathway. Aging (Albany NY). 2020 May 28;12(10):9125–9138.
40. Zhang HX, Yang JJ, Zhang SA, Zhang SM, Wang JX, Xu ZY, Lin RY. HIF-1α promotes inflammatory response of chronic obstructive pulmonary disease by activating EGFR/PI3K/AKT pathway. Eur Rev Med Pharmacol Sci. 2018 Sep;22(18):6077–6084.

41. 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-PI3Kα-AKT/ERK pathway. Cell Signal. 2019 Jan;53:122–131.

Figures

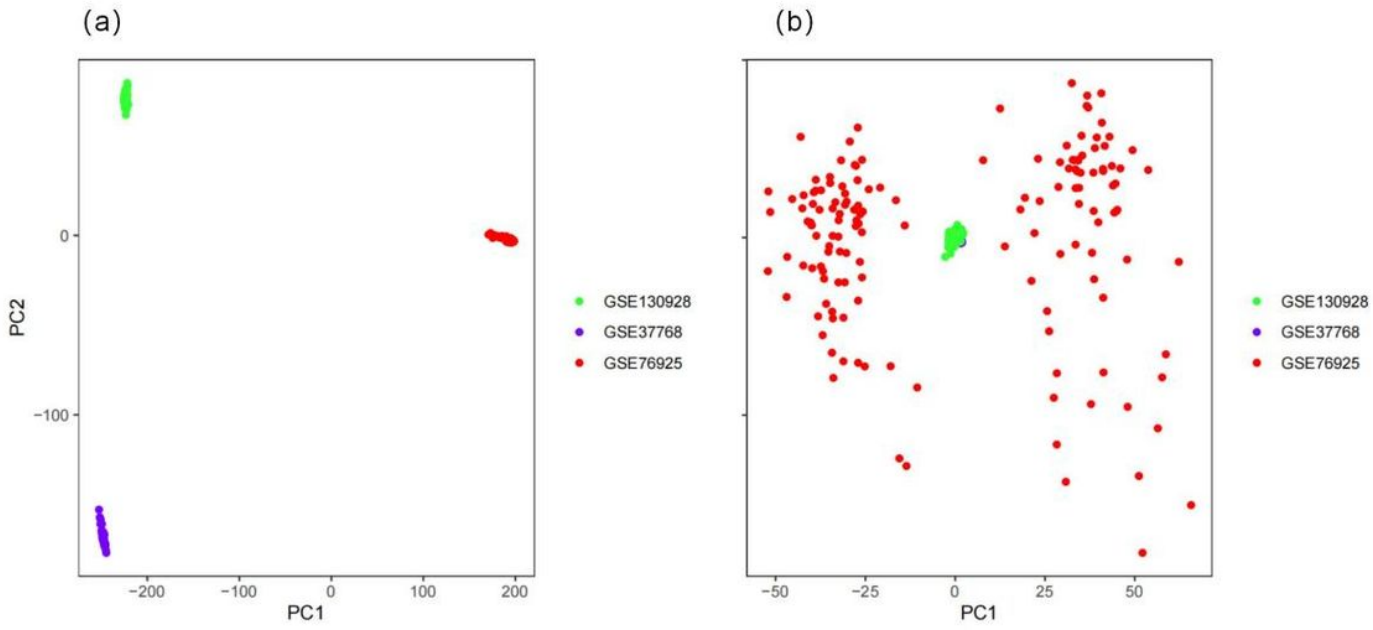

Figure 1

PCA of gene expression data set. scatters of different colors represent samples from three different data sets. A, PCA diagram before batch correction; B, PCA diagram

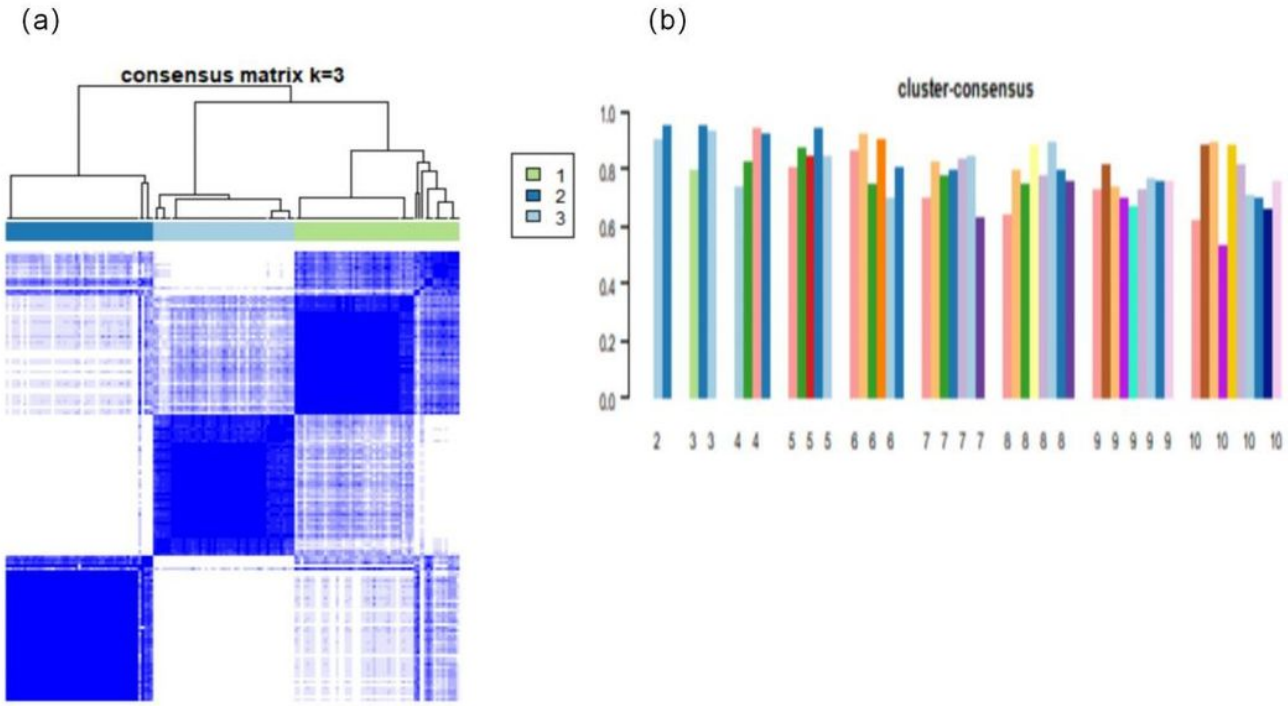

Figure 2

Consensus clustering of gene expression profiles in patients with COPD. A, the consistency matrix, when the number of clusters, is 3 that is determined by the minimum consistency score (> 0.8) of the subgroup; B, consistency scores for subgroups with cluster numbers between 2 and 10.

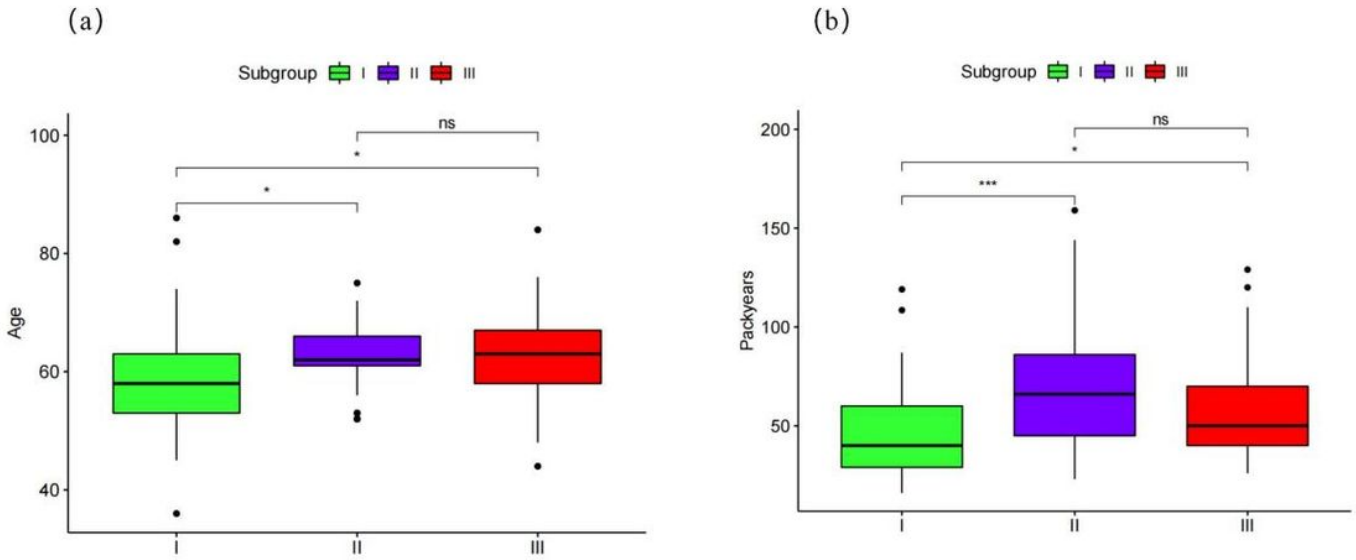

Figure 3

Paired comparison of clinical features between subgroups. (a) The relationship between the age of onset in each subgroup and the subgroup. (b) The relationship between each subgroup and smoking time was shown separately.

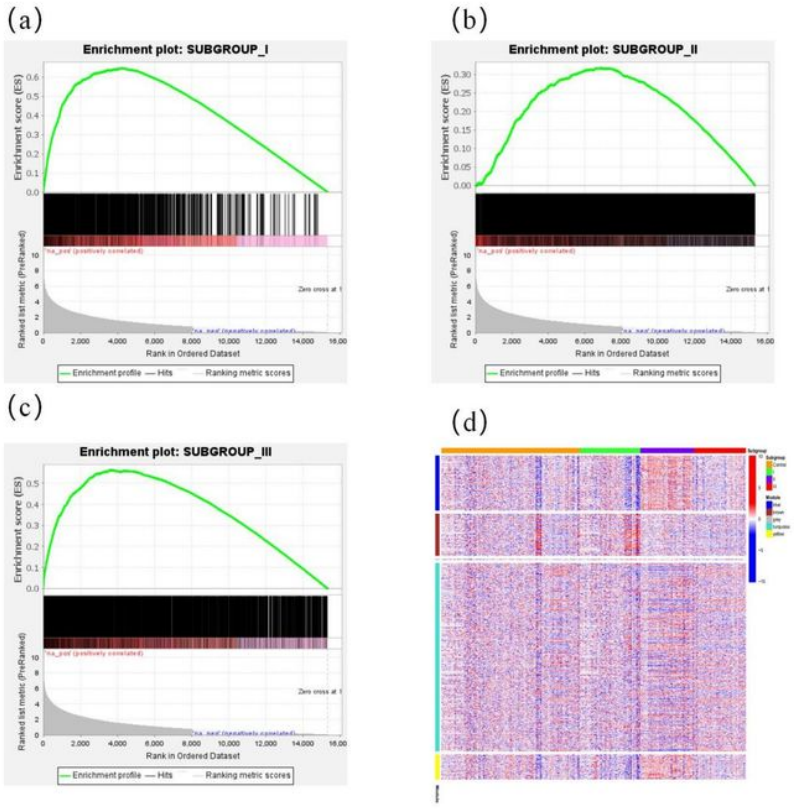

Figure 4

The expression patterns of subgroup-specific upregulated genes. The enrichment plots of (a), (b), and (c) illustrate the subgroup-specific upregulated 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.

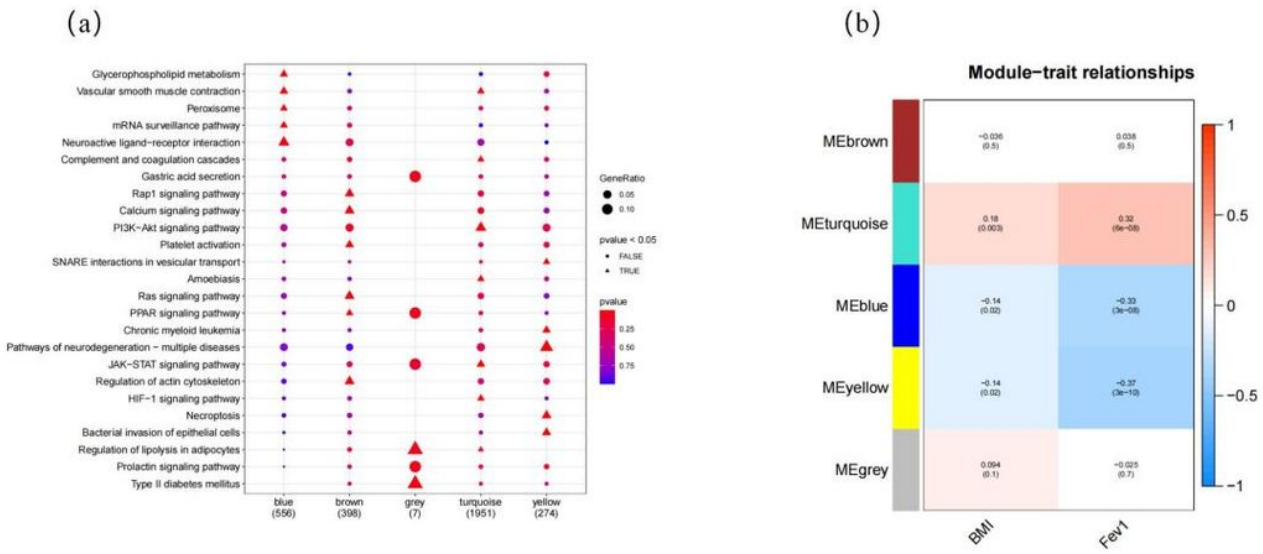

Figure 5

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. Overrepresentation 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.