

Bioinformatics analysis of differentially co-expressed genes and immune cell infiltration features associated with pulmonary arterial hypertension

Wenshi Liu^{1,2}, Dongdong Zheng¹, Wenjing Lv¹, Ying Hua¹, Rong Huang¹, Xiaofei Li¹

¹Department of Cardiology, Affiliated Hospital of Nantong University, Nantong 226001, China

²Department of Cardiology, The Fifth Hospital of Xiamen, Xiamen, 361101, China.

Submitted: 9 June 2021; **Accepted:** 9 September 2021

Online publication: 17 September 2021

Arch Med Sci

DOI: <https://doi.org/10.5114/aoms/142102>

Copyright © 2022 Termedia & Banach

Corresponding author:

Xiaofei Li

Department of Cardiology

Affiliated Hospital of

Nantong University

Nantong 226001,

China

e-mail:

lixiaofei1991@163.com

Abstract

Introduction: Gene expression and immune cell infiltration are essential in the etiopathogenesis of pulmonary arterial hypertension (PAH). Our study attempted to identify novel differentially co-expressed genes and to investigate the features of immune cell infiltration in PAH.

Material and methods: The GSE113439 and GSE117261 datasets were acquired from the Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) between the PAH and control groups were identified based on the GSE117261 dataset. Weighted gene co-expression network analysis (WGCNA) was used to analyze the pre-processed data. Functional enrichment analysis was then carried out to explore the biological functions of these gene modules. The differentially co-expressed key gene modules were verified in-depth by GEO2R analysis. The immune infiltration in PAH was investigated by cell type identification by estimating relative subsets of RNA transcripts (CIBERSORT).

Results: The WGCNA analysis found 15 differentially co-expressed gene modules, with the blue module indicating the strongest positive link to PAH and the green module indicating the strongest negative association with PAH. The Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis showed that the genes in the blue module were largely enriched in lysosome, complement, and coagulation cascades, and others, whereas the genes in the green module were primarily enriched in the chemokine signaling pathway, platelet activation, etc. Integrin subunit alpha M (ITGAM) was identified as the differentially co-expressed key gene. Immune infiltration analysis by CIBERSORT showed the differences between the PAH and control groups or between the PAH subgroups.

Conclusions: ITGAM was considered a promising biomarker to discriminate PAH from the controls. Obvious differences were observed in immune infiltration between patients with the PAH and normal groups, providing new insight into understanding the molecular mechanisms on PAH's pathogenesis.

Key words: pulmonary arterial hypertension, differentially co-expressed genes, weighted gene co-expression network analysis, functional enrichment analysis, immune infiltration, cell type identification by estimating relative subsets of RNA transcripts.

Introduction

Pulmonary arterial hypertension (PAH) is considered a progressive cardiopulmonary disease in which increased pulmonary artery pressure

causes right ventricle failure and premature death. Pulmonary arterial hypertension is an extremely burdensome disease with a poor long-term clinical prognosis. Despite advances in treatment, PAH is still an incurable disease with high mortality and morbidity [1]. Recent research has tried to reveal more novel pathogenic mechanisms to reverse PAH. Currently, the underlying etiology of PAH and the potential key genes involved in this disease remain poorly understood even though some genetic factors have been found to participate in the PAH pathogenesis [2]. Moreover, mounting evidence has revealed that inflammation has a fundamental role in the pathology of PAH [3]. Pulmonary arterial hypertension patients presented gathered immune cells, which consist of dendritic cells (DCs), B and T lymphocytes, etc, across blood. Also, some reports showed that the circulating levels of several chemokines and cytokines were increased [4]. Therefore, it is urgent to explore the key genes with specific methods to uncover their molecular mechanisms. Studies on the immune landscape may also be valuable for developing novel approaches to treat PAH.

Currently, bioinformatics analysis has been extensively applied to investigate the molecular mechanisms of various diseases by analyzing the potential changes in gene expression. Here, we screened the differentially expressed genes (DEGs) between the PAH and control groups and further explored their potential biological functions. Considering that genes with similarly expressed patterns may display potential relations related to function, exploring the expression profile of the related genes can contribute to understanding their hidden details [5, 6]. Weighted gene co-expression network analysis (WGCNA) has been applied to unearth gene modules that were closely related to systemic sclerosis-associated PAH (SSC-PAH) and idiopathic PAH (IPAH) [7–9]. Based on this, we selected different gene chips to recognize more types of PAH and expanded the sample size by merging the chip data. We further analyzed the data using the WGCNA to find gene modules closely related to PAH. This is a comprehensive study utilizing the WGCNA method to establish a co-expressed network to explore the underlying causes of PAH, providing a theoretical foundation for elucidating the occurrence and development of PAH.

The Cell type identification by estimating relative subsets of RNA transcripts (CIBERSORT) analysis tool has been employed to estimate the composition of immune cells in different samples using standardized gene expression data, with the advantages of high resolution and quantification towards multiple types of immune cells verified in a variety of malignant neoplasms [10]. Immune in-

filtration analysis by CIBERSORT was carried out in PAH. For the first time, we adopted the CIBERSORT algorithm to systematically investigate the immune infiltration of PAH.

Materials and methods

Microarray data preprocessing

The GSE113439 and GSE117261 datasets were downloaded from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>). The GSE113439 dataset contained 26 lung tissue samples, including 6 IPAH, 8 associated PAH (APAH), 1 chronic thromboembolic pulmonary hypertension (CTEPH), and 11 normal controls. The GSE117261 dataset contained 83 lung tissue samples, including 32 IPAH, 17 APAH, 9 other PAH, and 25 normal controls.

The GSE113439 and GSE117261 datasets were based on the GPL6244 (Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]) platform. After removing the samples that did not meet the requirements of this study, the two microarray data were combined. The “SVA” R package was used to eliminate the batch effects and other adverse changes, and the “limma” R package was used to normalize the microarray data.

Differentially expressed genes identification

R software (version 3.5.1, <https://www.r-project.org>) and the “limma” packages were exploited to select the DEGs between the PAH and control samples. Adjusted p -value < 0.05 and $|\log_2$ fold change (FC)| ≥ 0.5 were considered as the selection thresholds. The DEGs with \log_2 FC < 0 were considered down-regulated, whereas those with \log_2 FC > 0 were considered upregulated.

Construction of WGCNA network and identification of modules

The “WGCNA” R package was applied to build a co-expressed network [11]. Firstly, the samples were clustered to determine whether evident outliers existed. Secondly, the automatic network construction function was applied to construct a gene co-expressed network. The R function pickSoft Threshold was employed to determine the soft thresholding power β , to which the co-expression similarity was heightened to obtain adjacency. Thirdly, the hierarchical clustering and dynamic tree cutting function detection module was used. Finally, gene significance (GS) and module identity (MM) were calculated to associate the modules with the clinical features. Genes with $|GS| > 0.5$ and $|MM| > 0.8$ in the gene module network were highly correlated with clinical phenotype as module

hub genes. The corresponding module genes were extracted for further analysis. The cytoHubba plug-in in Cytoscape was used for further screening of the key genes. Finally, the eigengene network was visualized.

Functional enrichment analysis

Gene Ontology (GO) enrichment analysis is generally utilized in describing gene functions, including molecular function (MF), biological process (BP), or cellular component (CC) [12]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was conducted to identify the molecular interactions and related networks [13]. The p -value < 0.05 meant there was a significant statistical difference.

Identification and verification of differentially co-expressed key genes

We defined the genes with $|GS| > 0.5$ and $|MM| > 0.8$ in the gene module network that were highly correlated with the clinical phenotype as the module hub genes. The cytoHubba plug-in in Cytoscape was used for further screening key genes. The differentially co-expressed key genes were obtained by complementing the DEG analysis results. Through the GEO2R online analysis tool on the GEO website, the GSE15197 and GSE38267 datasets were analyzed to verify the differentially co-expressed key genes.

Immune infiltration analysis

The gene expression data from the GSE117261 dataset were normalized by the CIBERSORT algorithm. The ratio of 22 immune cells was then assessed and all samples were screened with the p -value < 0.05 . The percentages of the individual types of immune cells in the samples were determined. The difference in the immune cell infiltration between the PAH and control samples was determined by principal component analysis (PCA). The “vioplot” R package was used to compare the levels of 22 immune cells between PAH and control groups or between PAH subgroups. A correlation heatmap, depicting the cross-relation of 22 immune cells, was drawn by the “corrplot” package [14] of R software. The plots were visualized by the “ggplot2” package [15].

Results

Data preprocessing and differentially expressed genes identification

The top 25% of the genes in the variance were screened for subsequent evaluation. First, the missing values of the genes and samples were examined, and all genes passed the cut-off value.

Second, the samples were clustered to determine whether there were obvious outliers. The cut-off height was set to 65, and samples that did not meet the standard were removed. Finally, after excluding the unwanted samples, the remaining samples were re-clustered hierarchically and combined with the clinical information to draw a heatmap (Supplementary Figure S1). The results suggested that PAH and normal lung tissues were basically clustered reasonably and could be used for subsequent analysis.

Based on previous criteria of $|\log_2 FC| > 0.5$ and adjusted p -value < 0.05 , we totally screened and identified 318 DEGs, including 191 genes with up-regulation and 127 genes with down-regulation in the PAH tissues compared with normal tissues. All the DEGs are presented as the volcano map (Supplementary Figure S2).

Construction of genes co-expressed modules

To construct the gene co-expressed network, we need to select soft thresholding power β to increase the co-expression similarity to this threshold to calculate adjacency. The soft threshold was selected by approaching the criterion of scale-free topology, that is, R^2 of the scale-free topology model should be close to 0.9 and reach the platform stage (Figure 1 A), and the mean connectivity should be relatively high (Figure 1 B). Therefore, 6 was selected as the soft thresholding power β value for subsequent analysis.

The one-step network construction function of the “WGCNA” R package was employed to set up the gene co-expressed network and identify the gene modules. The minimum number of genes for each module was set to 50, and the deep split was set to 2 (indicating medium sensitivity). The dynamic splicing method was utilized to merge the gene modules with a height below 0.25 (Supplementary Figure S3). Finally, 15 gene modules were chosen and constructed (Figure 2).

We analyzed the relationships between the established modules and mapped them. The topological overlap matrix (TOM) of all genes used for the WGCNA analysis was shown on the heatmap (Supplementary Figure S4). The lighter the color, the lower the degree of overlap, whereas the darker the color, the higher the degree of overlap. The results of the analysis showed that gene expression was comparatively independent amongst the modules.

We investigated in depth the association between pairwise gene co-expressed modules. We clustered analysis of eigengenes. The results indicated that 15 modules could be divided into two clusters and that four combinations (cyan and salmon modules, pink and turquoise mod-

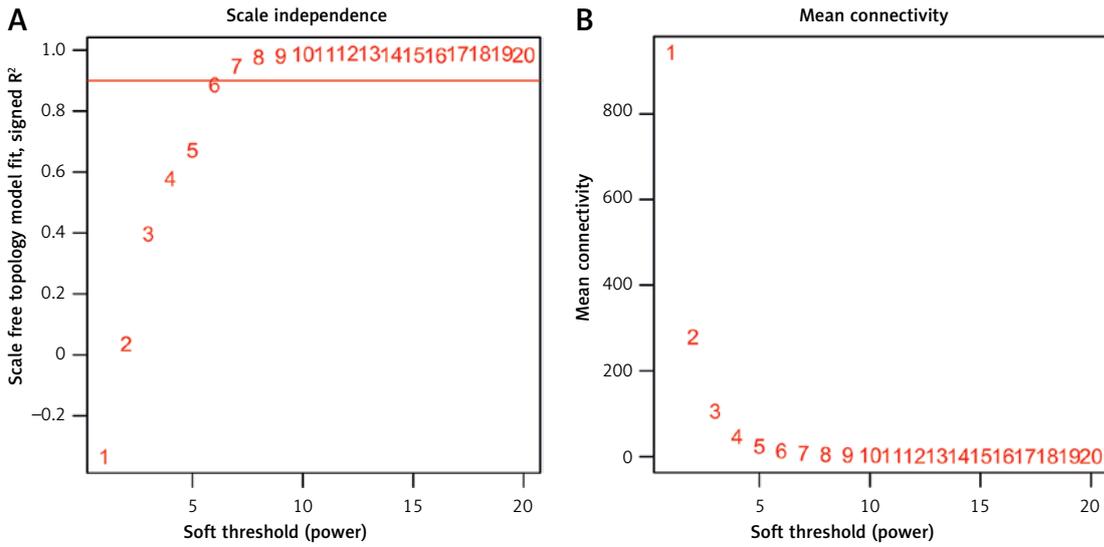


Figure 1. Analysis of the scale-free topology model fit index R^2 (A) and the mean connectivity (B) of the soft thresholding power β values

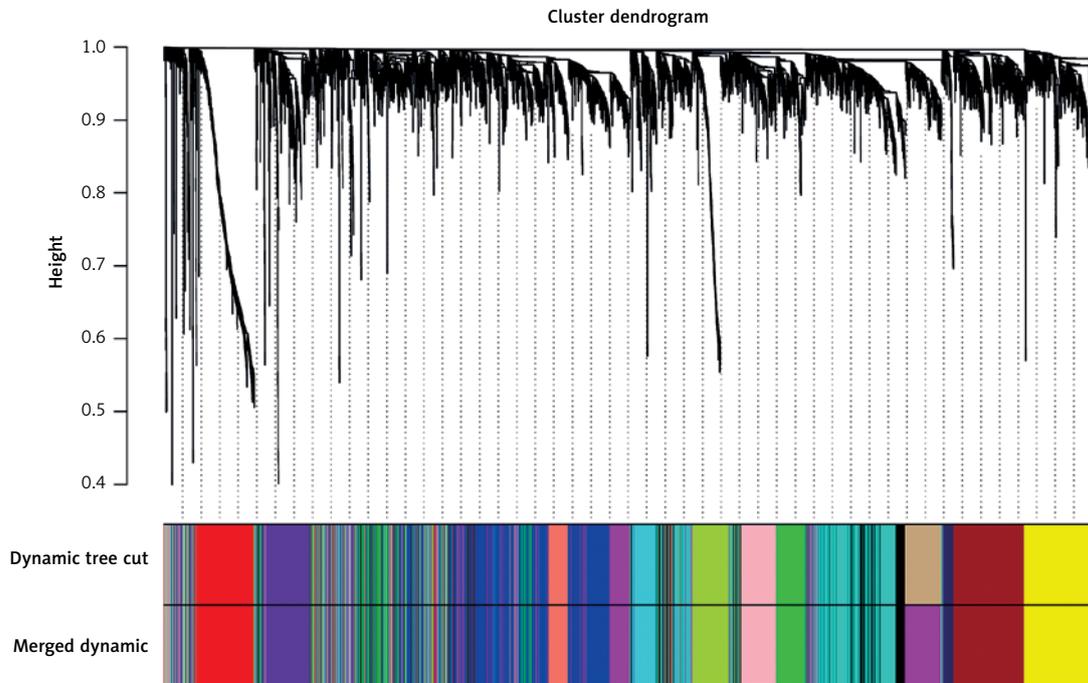


Figure 2. The cluster dendrogram of genes. Each branch represents one gene. Each color represents one co-expressed module

ules, blue and magenta modules, black and yellow modules) had a high degree of interactive connectivity (Supplementary Figure S5).

Identifying key modules

Through the correlation analysis between the gene modules and clinical information, we discovered gene modules related to the traits of the research object. The results showed that the blue module presented the highest positive correlation

with PAH ($r = 0.58$, p -value < 0.05), and the green module had the highest negative correlation with PAH ($r = -0.59$, p -value < 0.05) (Figure 3).

Functional analysis of the key module

GO and KEGG analyses were performed on the genes in the blue module and the green module, respectively. GO enrichment histograms showed the top enriched functions of the blue module and green genes. The GO enrichment analysis showed

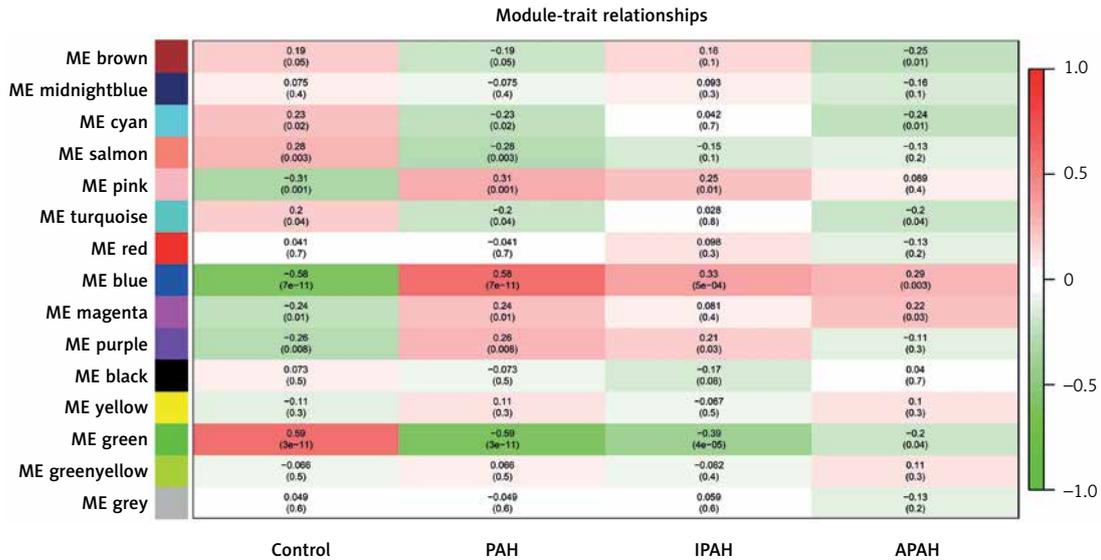


Figure 3. Heatmap of the correlation between gene modules and clinical information. The *p*-value is displayed in each color unit (red and green indicate positive and negative correlations, respectively)

PAH – pulmonary arterial hypertension, IPAH – idiopathic pulmonary arterial hypertension, APAH – associated pulmonary arterial hypertension

that the blue module genes were mainly associated with the maintenance of location in the cell, maintenance of location, glial cell migration, regulation of macrophage migration, macrophage migration, and sensory perception of taste in the BP terms; cell adhesion molecule binding, phosphoric ester hydrolase activity, and actin-binding in the MF terms; and apical part of the cell, basolateral plasma membrane, actin-based cell projection, and filopodium in the CC terms (Figure 4 A). The GO enrichment analysis of the green module genes showed that they were primarily involved in neutrophil activation, neutrophil activation involved in immune response, neutrophil degranulation, neutrophil mediated immunity and leukocyte migration in the BP terms; immune receptor activity, carbohydrate binding, cytokine receptor activity and cytokine binding in the MF terms; secretory granule, tertiary granule, specific granule, cytoplasmic vesicle lumen, secretory granule lumen, and vesicle lumen in the CC terms (Figure 4 B). It was found that maintenance of location was the most abundant BP, the apical part of the cell was the most abundant CC, while cell adhesion molecule binding was the most abundant MF regarding blue module genes (Figure 4 A). The GO analysis of green module genes showed that the most abundant biological process was neutrophil activation, the most abundant cellular component was secretory granules, while the most abundant molecular function was immune receptor activity (Figure 4 B). The KEGG pathway analysis showed that blue module genes were significantly enriched in herpes simplex virus 1 infection lysosome, taste transduction, complement and coagulation cascades (Figure 5 A).

KEGG pathway analysis showed that the green module genes were significantly enriched in osteoclast differentiation, B cell receptor signaling pathway, chemokine signaling pathway, phagosome, tuberculosis, leishmaniasis, lysosome, pathogenic *Escherichia coli* infection, and so on (Figure 5 B). The KEGG pathway analysis showed that the genes in the blue module and green module were mainly involved in lysosome, and complement and coagulation cascades.

Screening and identifying the differentially co-expressed key genes

We applied GS to determine the degree of association between genes and external information. We screened 12 blue module hub genes (Figure 6 A, Table I) and 12 green module hub genes (Figure 6 B, Table II). The cytoHubba plugin was utilized to identify the top 10 genes of degree in the blue module and the green module (Figure 6 C, D). We defined the common genes of the module hub gene and the top ten genes of degree in the modules as the key genes for co-expression in the module. Two co-expressed key genes in the blue modules were S-phase cyclinA associated protein residing in the endoplasmic reticulum (SCAPER), and luc-7 like protein 3 (LUC7L3). Five co-expressed key genes in the green modules were integrin subunit alpha M (ITGAM), beta-arrestin 2 (ARRB2), leukocyte immunoglobulin-like receptor B3 (LILRB3), glucose-6-phosphate dehydrogenase (G6PD), and leucine rich repeat-containing 25 (LRRC25).

The VENN online website was used to determine the differentially co-expressed key genes. Finally, we obtained 3 differentially co-expressed

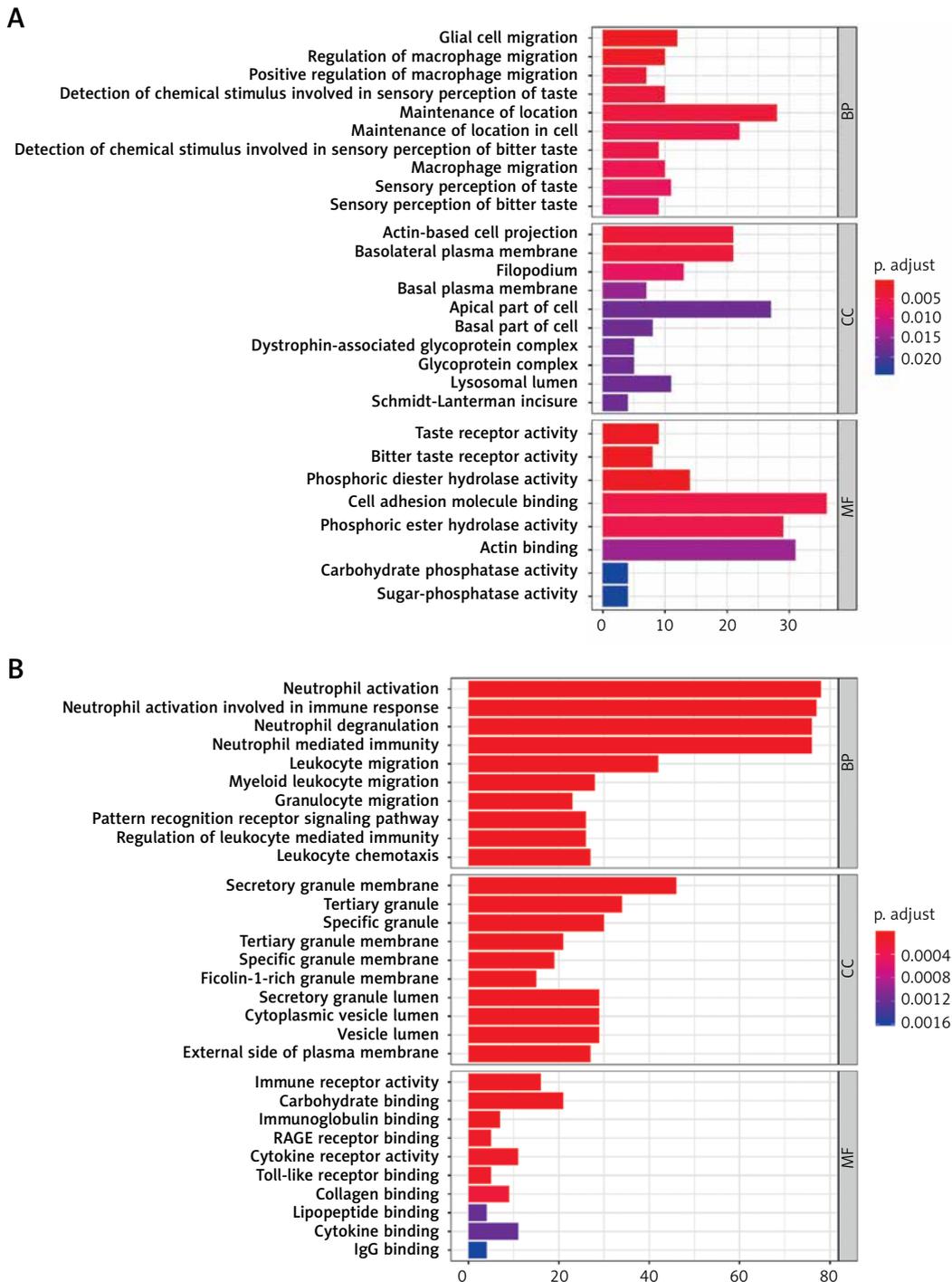


Figure 4. Gene ontology (GO) enrichment analysis of genes in the gene co-expressed modules. The horizontal axis is the number of genes enriched to each GO entry **A** – blue module genes. **B** – green module genes

key genes in the green modules: ITGAM, ARRB2, LILRB3 (Figure 6 E).

GSE15197 and GSE38267 were further used to verify the three differentially co-expressed key genes. We selected the lung tissue from 18 PAH patients and 13 normal controls in GSE15197 chip data for GEO2R online analysis (Supplementary Figure S6). We found the expression of the ITGAM gene in the GSE15197 chip data with log₂ FC <

−0.5 (log₂ FC = −0.57, 344) and *p*-value < 0.05). Then, we selected peripheral blood from 13 PAH patients and 28 normal controls in the GSE38267 chip data for GEO2R online analysis (Supplementary Figure S7) and found the expression of the ITGAM gene in GSE38267 with log₂ FC < −0.5 (log₂ FC = −0.51, 242) and *p*-value < 0.05.

The ITGAM mRNA levels in the PAH lung tissue and peripheral blood were lower than that of the

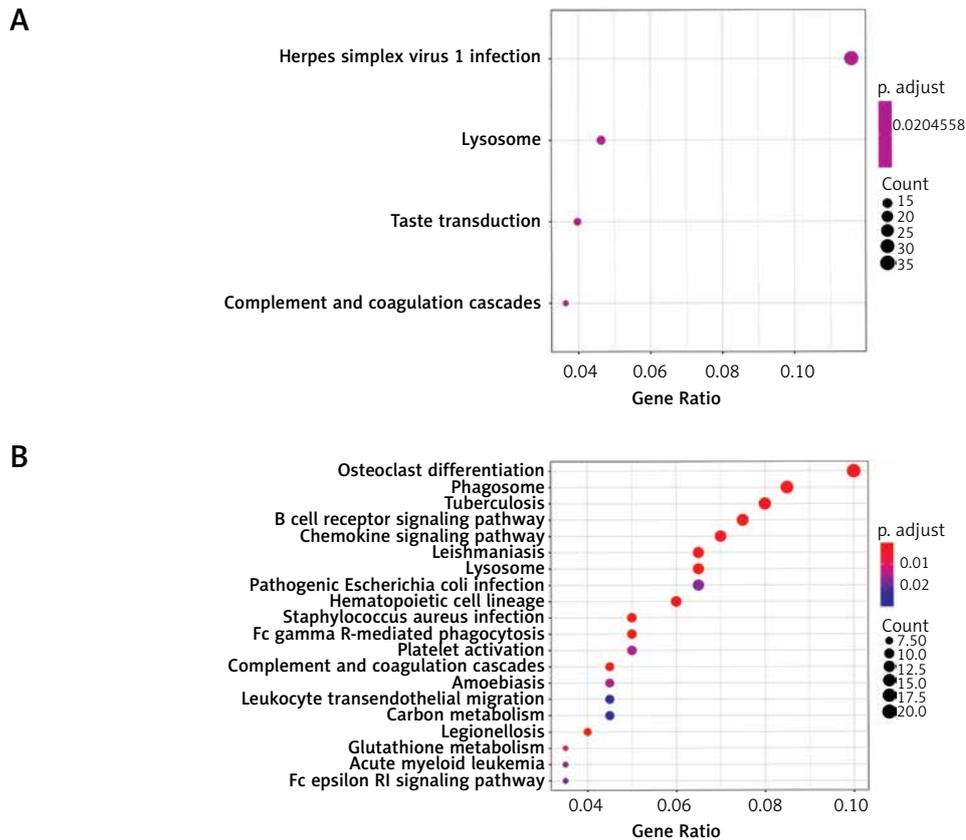


Figure 5. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the genes in the gene co-expressed modules. Each circle size represents the gene amount. Each circle color means different *p*-value. Gene ratio is the number of genes belonging to the pathway divided by the number of all included genes. The top 4 KEGG enrichment pathways associated with genes in blue module (A) and the top 20 KEGG enrichment pathways associated with genes in green module (B)

healthy controls, implying that ITGAM may have a potential role in the PAH's occurrence and development. Taking the evidence together, we identified ITGAM as the differentially co-expressed key gene that which was lowly expressed in both the lung tissue and peripheral blood of PAH patients.

Immune infiltration analysis

Due to technical limitations, PAH immune infiltration is still unclear, particularly in PAH subpopulations with low cell abundance. Utilizing the CIBERSORT algorithm, we explored the differences of 22 immune cell subpopulations in lung tissue between the PAH and control groups. The general outline of the analysis results from 35 controls and 71 PAH patients is shown in Figure 7 A. Principal component analysis analysis suggested that the immune cell proportion from the lung tissues of the PAH patients and controls exhibited distinct group-bias clustering and individual differences (Figure 7 B). Compared with normal lung tissues, PAH lung tissues presented a higher proportion of CD8⁺ T cells, resting memory CD4⁺ T cells, gamma delta T cells, activated DCs, resting mast cells, and

eosinophils, while they possessed a lower proportion of naive CD4⁺ T cells, activated memory CD4⁺ T cells, resting nature killer (NK) cells, monocytes, and neutrophils (Figure 7 C, *p*-value < 0.05). Further subgroup analysis revealed that the proportion of CD8⁺ T cells in the IPAH group was higher than that in the APAH group (Figure 7 D, *p*-value < 0.05).

Discussion

PAH is a serious cardiovascular disease with a complex pathogenesis and its molecular mechanism has not been fully elucidated. Genetic dysfunction is generally considered as the underlying pathogenesis of PAH, and immune disorders also play an important role in disease progression. In recent years, the research on gene dysfunction and the immune landscape of PAH has attracted unprecedented attention, which may be conducive to ameliorate the poor treatment effect. Here, we specifically focused on the gene expression and immune landscape between PAH and control samples, giving a new hint of exploring the molecular mechanism of PAH and discovering the key factors affecting its occurrence and development.

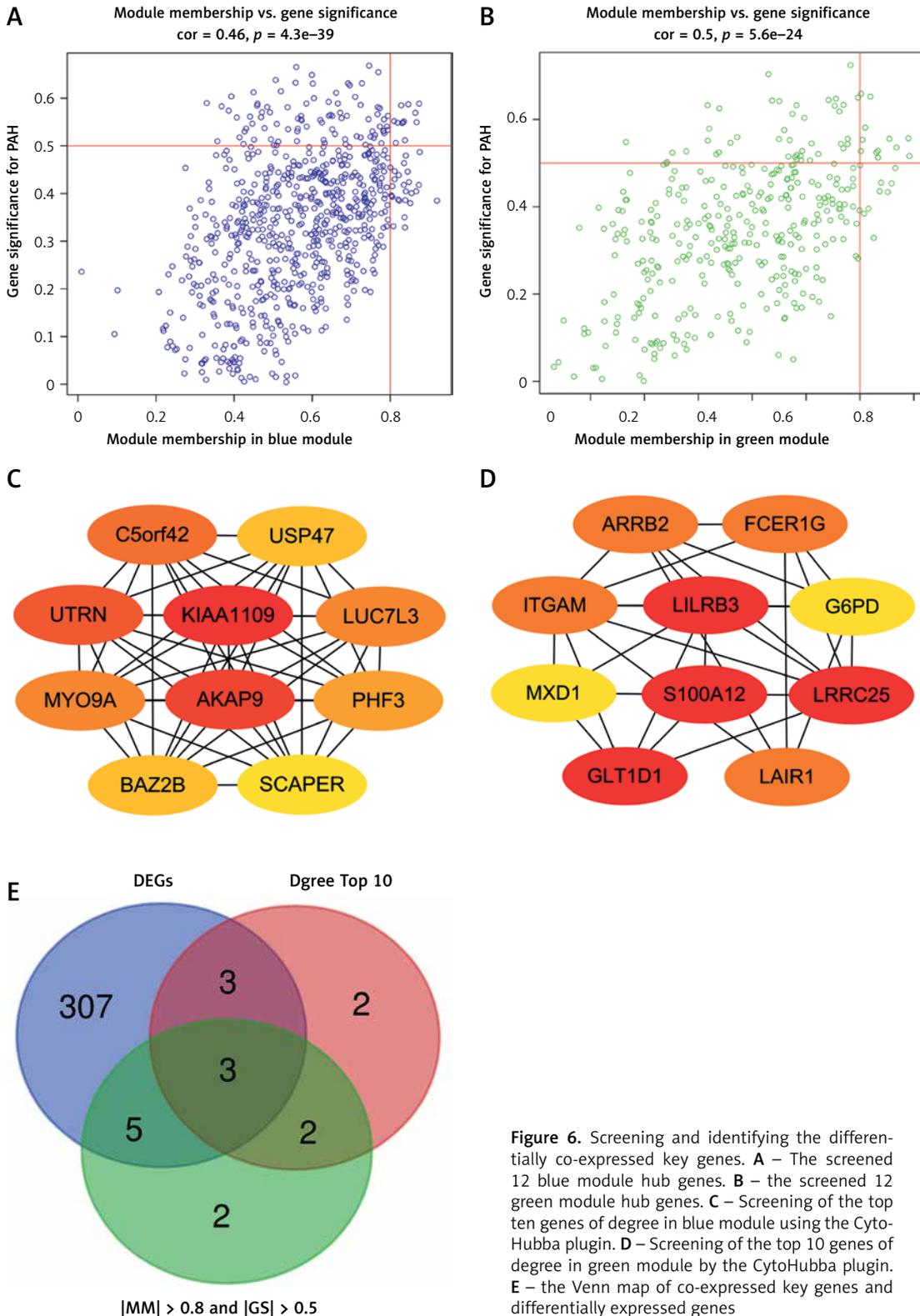


Figure 6. Screening and identifying the differentially co-expressed key genes. **A** – The screened 12 blue module hub genes. **B** – the screened 12 green module hub genes. **C** – Screening of the top ten genes of degree in blue module using the CytoHubba plugin. **D** – Screening of the top 10 genes of degree in green module by the CytoHubba plugin. **E** – the Venn map of co-expressed key genes and differentially expressed genes

Here, we used a systematic bioinformatics method to investigate the working mechanisms of PAH. WGCNA is widely utilized to illustrate the pairwise relationship between gene transcriptomes. Unlike other network analysis methods that focus on un-

weighted networks, WGCNA is more comprehensive and qualified to create weighted or unweighted association networks [11]. The CIBERSORT algorithm uses 547 genes to distinguish 22 different types of immune cells with high sensitivity and

Table I. The 12 hub genes in blue module

Probes	Module color	GS.PAH	p.GS.PAH	MMblue	p.MMblue
VPS13A	Blue	0.547811	1.22E-09	0.830276	3.71E-28
AGTRAP	Blue	-0.5512	9.20E-10	-0.83476	1.04E-28
NBEAL1	Blue	0.551959	8.63E-10	0.8099	7.71E-26
SCAPER	Blue	0.549586	1.05E-09	0.867366	2.72E-33
DOPEY1	Blue	0.506172	3.11E-08	0.814168	2.66E-26
RHOG	Blue	-0.58842	3.29E-11	-0.87393	2.31E-34
FLOT2	Blue	-0.58497	4.56E-11	-0.82804	6.89E-28
MPDU1	Blue	-0.54869	1.13E-09	-0.81861	8.55E-27
LUC7L3	Blue	0.569328	1.91E-10	0.864249	8.38E-33
COPE	Blue	-0.58714	3.71E-11	-0.84621	3.39E-30
PNISR	Blue	0.511706	2.07E-08	0.837581	4.59E-29
BRWD1	Blue	0.548787	1.13E-09	0.832459	2.01E-28

Table II. The 12 hub genes in green module

Probes	Module color	GS.Control	p.GS.Control	MMgreen	p.MMgreen
LILRB3	Green	0.556664	5.79E-10	0.889993	2.96E-37
G6PD	Green	0.581212	6.48E-11	0.809718	8.06E-26
ITGAM	Green	0.535771	3.26E-09	0.867758	2.36E-33
STXBP2	Green	0.527496	6.25E-09	0.80182	5.37E-25
SIGLEC9	Green	0.516197	1.48E-08	0.891822	1.30E-37
ARRB2	Green	0.652382	3.56E-14	0.819444	6.88E-27
TALDO1	Green	0.659029	1.59E-14	0.802223	4.89E-25
MYO1F	Green	0.555813	6.22E-10	0.8547	2.22E-31
NCF4	Green	0.553257	7.73E-10	0.804626	2.77E-25
TLR8	Green	0.528193	5.92E-09	0.833341	1.56E-28
APBB1IP	Green	0.524127	8.12E-09	0.830303	3.68E-28
LRRC25	Green	0.512548	1.95E-08	0.833617	1.44E-28

specificity, which is superior to other algorithms. In our study, we applied the WGCNA method to obtain the gene co-expressed network, screen the differentially co-expressed key genes, and verify them using other gene chips. The CIBERSORT algorithm was then used to probe for the immune infiltration between the PAH and control groups or between the PAH subgroups.

Through the in-depth and systematic analysis of the merged and preprocessed datasets of GSE113439 and GSE117261, we found that the blue module had the strongest positive correlation with PAH, and the green module had the strongest negative correlation with PAH.

GO analysis of the blue module genes indicated that cell adhesion molecule binding was involved in PAH's occurrence. The cell adhesion molecule serves as the molecular basis in a variety of important physiological and pathological processes, such as immune response, inflammation, coagulation, tumor metastasis, and wound healing. GO analysis

of the green module genes indicated that neutrophil activation occurred in the process of PAH. Studies have shown that monocrotaline -induced PAH in rats was related to increased activation and interaction of platelets and leukocytes *in vivo*, and these changes were maintained from the early stage to the late stage of PAH [16]. Recent experiments have revealed the link between pulmonary vascular remodeling and the formation of neutrophil extracellular traps (NETs), which displayed important roles in lung inflammation and autoimmune diseases [17]. In addition, neutral granulocytes could modulate the formation of bronchus-associated lymphoid tissue (BALT), and the formation of BALT was related to the progression of PAH [18]. KEGG analysis showed that lysosome, complement and coagulation cascades were important pathways for PAH. Studies have shown that chloroquine can inhibit the autophagic pathway and increase the expression of bone morphogenetic protein type II receptor (BMPRII) through inhibiting lysosome

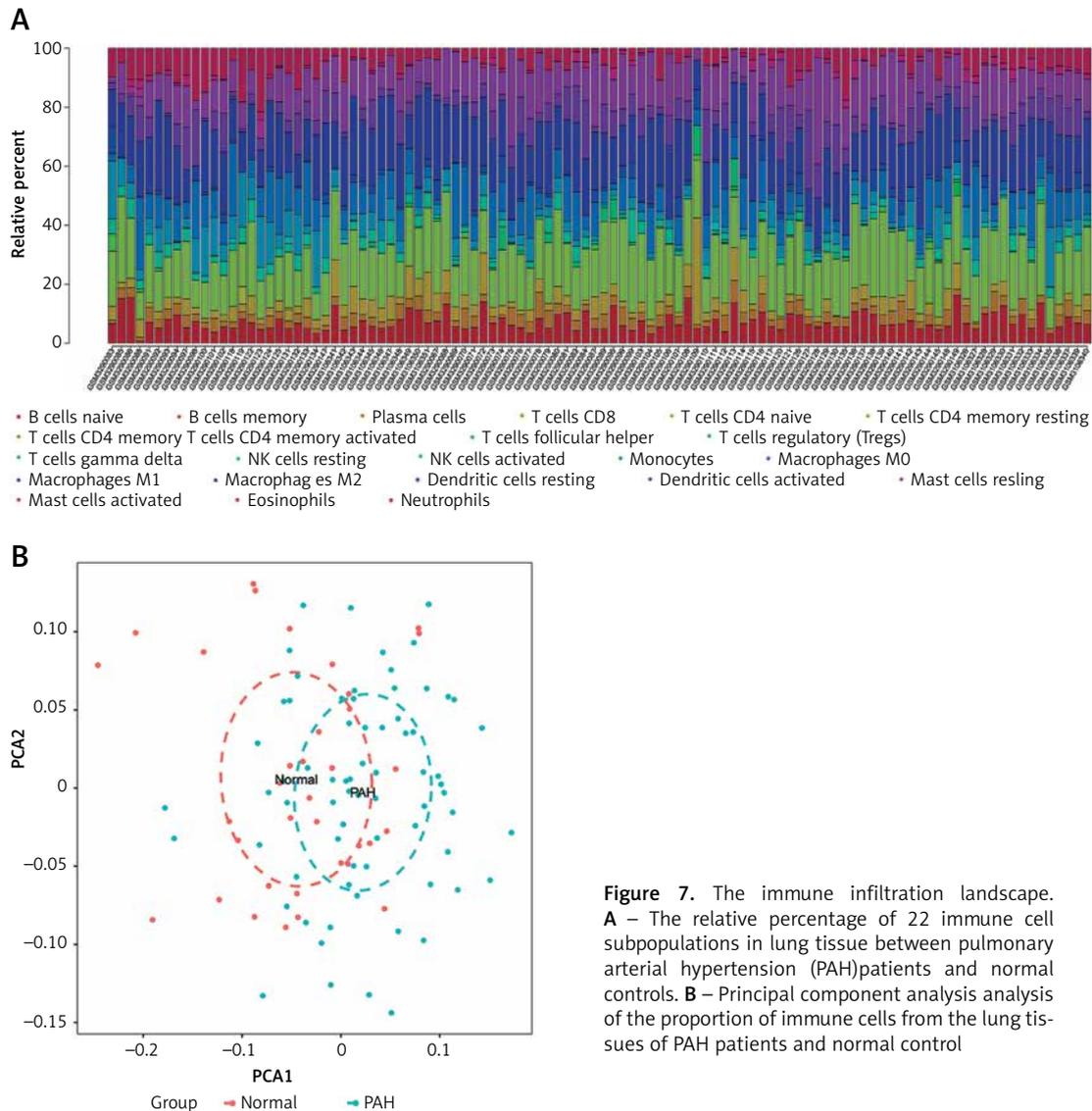


Figure 7. The immune infiltration landscape. **A** – The relative percentage of 22 immune cell subpopulations in lung tissue between pulmonary arterial hypertension (PAH) patients and normal controls. **B** – Principal component analysis analysis of the proportion of immune cells from the lung tissues of PAH patients and normal control

activity to prevent the progress of experimental pulmonary hypertension [19]. Another study found that the damage from endothelin-1 (ET-1) degradation caused by dysfunctional lysosomes may be associated with the occurrence of some pulmonary hypertension [20]. The over-activation of blood coagulation and complement was evident in many diseases, including atherosclerosis, diabetes, venous thromboembolic diseases, thrombotic microangiopathy, arthritis, carcinoma, and infectious diseases. A proteomics study on the effect of osthole on the treatment of pulmonary hypertension supported that complement and coagulation cascades may be involved in the process of PAH [21]. KEGG analysis also showed that the chemokine signaling pathway and platelet activation are important pathways of PAH. Chemokine signalings and their receptor families such as CXCL12/CXCR4 [22, 23], CXCL10/CXCR3 [24, 25], CCL2/CCR2 [26], CX3CL1/CX3CR1 [27, 28], and others are widely in-

involved in the occurrence and development of PAH. On the one hand, it allows immune cells to migrate to the inflammation site and mediates the process of pulmonary vascular remodeling, including boosting the proliferation, migration of pulmonary artery cells, and the expression of cytokines. On the other hand, chemokines also actively participate in the upstream and downstream signaling pathways of BMPR-II, accelerating disease progression [29]. Studies have shown that the in-situ thrombosis caused by the activation of platelets may be associated with the progressive pulmonary vascular changes in patients with falcate disease and hemolytic associated pulmonary hypertension [30, 31]. Platelet abnormalities and abnormal platelet activation are important factors in the progression of PAH because they all contribute to pulmonary artery thrombosis and cause vascular remodeling [32].

We discovered a differentially co-expressed key gene related to PAH: ITGAM. The ITGAM

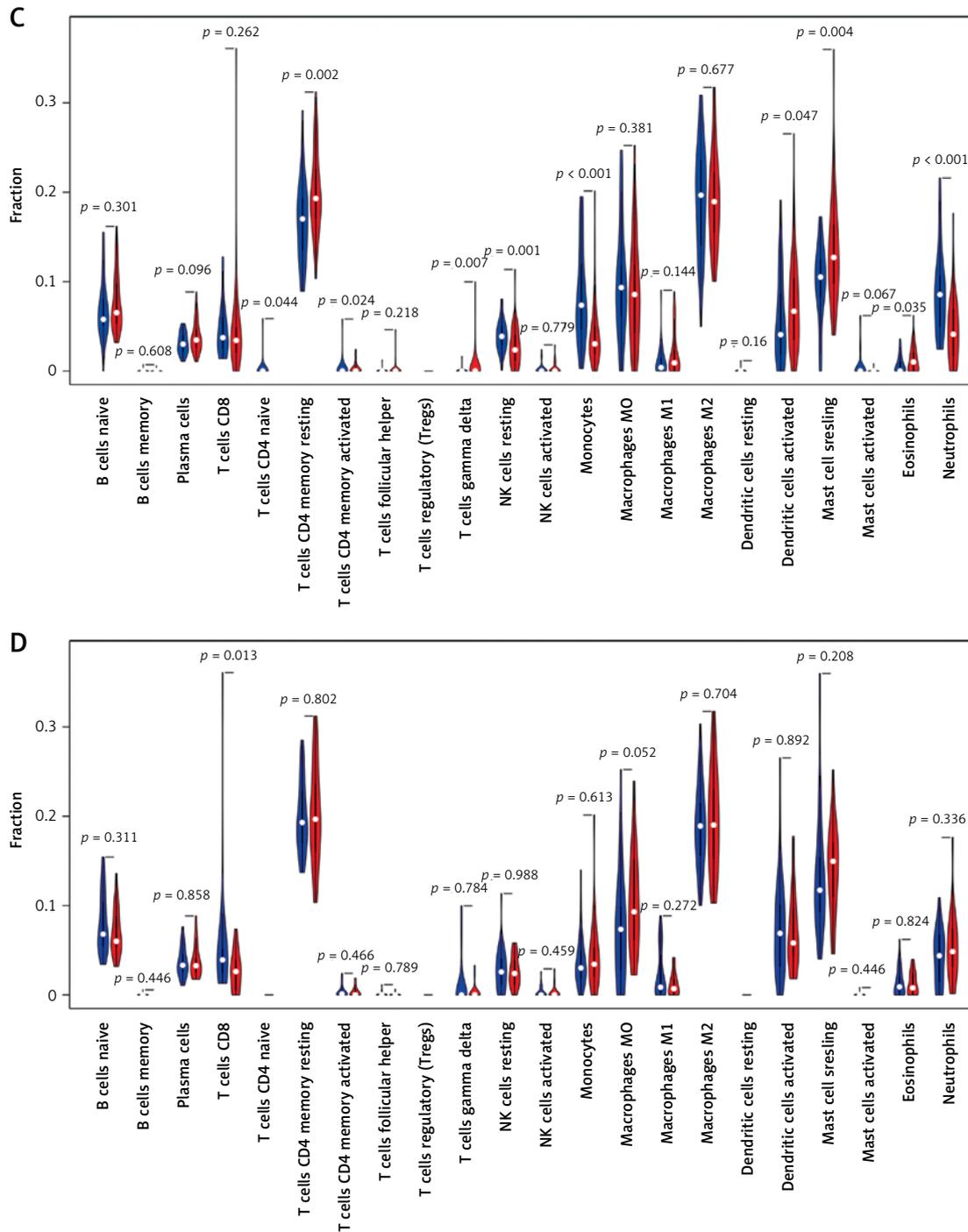


Figure 7. Cont. **C** – Different immune cell infiltration between normal lung tissue and pulmonary arterial hypertension (PAH) lung tissue. Blue color and red color indicate the control group and the PAH group, respectively. **D** – Different immune cell infiltration between idiopathic pulmonary arterial hypertension (IPAH) group and associated pulmonary arterial hypertension (APAH) group. Blue color and red color represent the IPAH group and APAH group, respectively. *P*-value < 0.05 means statistically significant difference

gene encodes the integrin α M chain. Integrin α M β 2 plays an important role in the adhesion of neutrophils and monocytes to stimulated endothelial cells and in the phagocytosis of complement coated particles. Studies have shown that ITGAM was linked to various autoimmune

diseases, such as systemic lupus erythematosus (SLE) [33, 34], systemic sclerosis (SSc) [35], female psoriasis [36], IgA nephropathy [37], and so on. ITGAM deficiency may up-regulate the production of IL-6 by antigen presenting cell (APC) cells and induce the preferential differentiation of

naïve T cells into Th17 cells, which may lead to the occurrence of certain autoimmune diseases [38, 39]. A study showed that the ITGAM mRNA level in SLE patients was greatly lower than that in healthy controls, suggesting that ITGAM has a potential role in SLE's pathogenesis and development [40]. PAH is a complication of connective tissue diseases (CTDs), such as SSc, SLE, and mixed CTDs. A rare missense mutation identified in ITGAM (p.E1071D) was related to persistent pulmonary hypertension (PPHN) in newborns. A previous study showed that during pulmonary hypertension, integrins were differentially modulated in pulmonary artery smooth muscle cells [41]. Nevertheless, it was not well understood how this integrin was involved in PAH or PPHN or whether the mutation of ITGAM was a protective or detrimental factor. As Jiang *et al.* described, platelet-mediated mesenchymal stem cell homing to the lung can result in the reduction of monocrotaline-induced rat pulmonary hypertension [42]. The latest study showed that ITGAM could regulate thrombosis through the interplay with platelet GPIb [43], implying that it perhaps exerts a protective function against thrombosis.

The immune landscape provides a deeper understanding of the inflammatory components of the pathogenesis of PAH. As previously described, the peripheral blood of IPAH patients displayed fewer CD8⁺ T cells and more T regulatory cells compared with the healthy controls. Additionally, compared with the control group, the number of mast cells has higher in the pulmonary arteries of IPAH, and these cells were probably involved in vascular remodeling [41]. Our study showed that PAH patients had a higher level of CD8⁺ T cells, resting memory CD4⁺ T cells, gamma delta T cells, activated DCs, and resting mast cells, and a lower level of naïve CD4⁺ T cells, monocytes, M0 macrophages, activated mast cells, and neutrophils compared with the control people. Unlike the previous results, this could be due to the CIBERSORT algorithm, which is prone to systematically overestimate or underestimate the cell type, despite with a fairly lower estimation bias.

There are several limitations to this study. The sample size is a little small, so we should enlarge the sample size in the future. In addition, molecular biology methods should be used to verify the findings, and further *in vivo* and *in vitro* studies are needed to consolidate our conclusion. Despite the current limitations, our discovery highlights potential genes and sheds new light on understanding the mechanisms of PAH.

Conclusions

This is a study using the WGCNA method to acquire a co-expression network to explore the

underlying causes of PAH. Here, we identified the differentially co-expressed key gene that can be further studied to determine the molecular mechanism of PAH. This study, for the first time, employed the CIBERSORT algorithm to comprehensively discuss the immune infiltration of PAH. Our findings shed new light on the mechanisms of PAH.

Acknowledgments

This study was supported by a grant from the Nantong Science and Technology Bureau (No. JCZ19107).

Conflict of Interest

The authors declare no conflict of interest.

References

1. Stearman RS, Bui QM, Speyer G, et al. Systems analysis of the human pulmonary arterial hypertension lung transcriptome. *Am J Respir Cell Mol Biol* 2019; 60: 637-49.
2. Zeng H, Liu X, Zhang Y. Identification of potential biomarkers and immune infiltration characteristics in idiopathic pulmonary arterial hypertension using bioinformatics analysis. *Front Cardiovasc Med* 2021; 8: 624714.
3. Xu J, Yang Y, Yang Y, Xiong C. Identification of potential risk genes and the immune landscape of idiopathic pulmonary arterial hypertension via microarray gene expression dataset reanalysis. *Genes (Basel)* 2021; 12. doi: 10.3390/genes12010125.
4. Zhao E, Xie H, Zhang Y. Identification of differentially expressed genes associated with idiopathic pulmonary arterial hypertension by integrated bioinformatics approaches. *J Comput Biol* 2021; 28: 79-88.
5. Horvath S, Dong J. Geometric interpretation of gene co-expression network analysis. *PLoS Comput Biol* 2008; 4: e1000117. doi: 10.1371/journal.pcbi.1000117.
6. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; 4: Article17. doi: 10.2202/1544-6115.1128.
7. Tang Y, Zha L, Zeng X, Yu Z. Identification of biomarkers related to systemic sclerosis with or without pulmonary hypertension using co-expression analysis. *J Comput Biol* 2020; 27: 1519-31.
8. Zheng JN, Li Y, Yan YM, et al. Identification and validation of key genes associated with systemic sclerosis-related pulmonary hypertension. *Front Genet* 2020; 11: 816. doi: 10.3389/fgene.2020.00816.
9. Wang T, Zheng X, Li R, et al. Integrated bioinformatic analysis reveals YWHAB as a novel diagnostic biomarker for idiopathic pulmonary arterial hypertension. *J Cell Physiol* 2019; 234: 6449-62.
10. Ali HR, Chlon L, Pharoah PD, Markowitz F, Caldas C. Patterns of immune infiltration in breast cancer and their clinical implications: a gene-expression-based retrospective study. *PLoS Med* 2016; 13: e1002194. doi: 10.1371/journal.pmed.1002194.
11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559. doi: 10.1186/1471-2105-9-559.
12. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003; 4: P3.

13. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; 45(D1): D353-D61. doi: 10.1093/nar/gkw1092.
14. Friendly M. Corrrgrams: exploratory displays for correlation matrices. *Am Stat* 2002; 56: 316-24.
15. Wilkinson L. *ggplot2: elegant graphics for data analysis* by WICKHAM, H. Wiley Online Library; 2011.
16. Hu XS, Du CQ, Zeng CL, et al. Systemic evaluation of platelet and leukocyte activation and interaction in a rat model of pulmonary arterial hypertension. *Cardiology* 2010; 117: 44-53.
17. Vengethasamy L, Hautefort A, Tielemans B, et al. BMPRII influences the response of pulmonary microvascular endothelial cells to inflammatory mediators. *Pflugers Arch* 2016; 468: 1969-83.
18. Colvin KL, Cripe PJ, Ivy DD, Stenmark KR, Yeager ME. Bronchus-associated lymphoid tissue in pulmonary hypertension produces pathologic autoantibodies. *Am J Respir Crit Care Med* 2013; 188: 1126-36.
19. Long L, Yang X, Southwood M, et al. Chloroquine prevents progression of experimental pulmonary hypertension via inhibition of autophagy and lysosomal bone morphogenetic protein type II receptor degradation. *Circ Res* 2013; 112: 1159-70.
20. Recla S, Hahn A, Aplitz C. Pulmonary arterial hypertension associated with impaired lysosomal endothelin-1 degradation. *Cardiol Young* 2015; 25: 773-6.
21. Yao L, Yang Y, He G, Ou C, Wang L, Liu KJSR. Global proteomics deciphered novel-function of osthole against pulmonary arterial hypertension. *Sci Rep* 2018; 8: 5556.
22. Wei L, Zhang B, Cao W, et al. Inhibition of CXCL12/CXCR4 suppresses pulmonary arterial smooth muscle cell proliferation and cell cycle progression via PI3K/Akt pathway under hypoxia. *J Recept Signal Transduct Res* 2015; 35: 329-39.
23. Dai Z, Ming L, Wharton JJC. Prolyl-4 hydroxylase 2 (PHD2) deficiency in endothelial cells and hematopoietic cells induces obliterative vascular remodeling and severe pulmonary arterial hypertension in mice and humans through hypoxia-inducible factor-2 α . *Circulation* 2016; 133: 2447-58.
24. Heresi GA, Aytakin M, Newman J, Dweik RAJL. CXC-chemokine ligand 10 in idiopathic pulmonary arterial hypertension: marker of improved survival. *Lung* 2010; 188: 191-7.
25. Farkas D, Thompson A, Bhagwani A, et al. Toll-like receptor 3 is a therapeutic target for pulmonary hypertension. *Am J Respir Crit Care Med* 2019; 199: 199-210.
26. Sanchez O, Marcos E, Perros F, et al. Role of endothelium-derived CC chemokine ligand 2 in idiopathic pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2007; 176: 1041-7.
27. Balabanian K, Foussat A, Dorfmüller P, et al. CX(3)C chemokine fractalkine in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2002; 165: 1419-25.
28. Zhang J, Yang W, Luo B, Hu B, Maheshwari A, Fallon MB. The role of CX3CL1/CX3CR1 in pulmonary angiogenesis and intravascular monocyte accumulation in rat experimental hepatopulmonary syndrome. *J Hepatol* 2012; 57: 752-8.
29. Tielemans B, Delcroix M, Belge C, Quarck R. TGF β and BMPRII signalling pathways in the pathogenesis of pulmonary arterial hypertension. *Drug Discov Today* 2019; 24: 703-16.
30. Hu W, Jin R, Zhang J, et al. The critical roles of platelet activation and reduced NO bioavailability in fatal pulmonary arterial hypertension in a murine hemolysis model. *Blood* 2010; 116: 1613-22.
31. Villagra J, Shiva S, Hunter LA, Machado RF, Gladwin MT, Kato GJ. Platelet activation in patients with sickle disease, hemolysis-associated pulmonary hypertension, and nitric oxide scavenging by cell-free hemoglobin. *Blood* 2007; 110: 2166-72.
32. Berger G, Azzam ZS, Hoffman R, Yigla M. Coagulation and anticoagulation in pulmonary arterial hypertension. *Isr Med Assoc J* 2009; 11: 376-9.
33. Martin J, Fonseca C. The genetics of scleroderma. *Curr Rheumatol Rep* 2011; 13: 13-20.
34. Assassi S, Mayes MD, Arnett FC, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheuma* 2010; 62: 589-98.
35. Carmona FD, Simeon CP, Beretta L, et al. Association of a non-synonymous functional variant of the ITGAM gene with systemic sclerosis. *Ann Rheum Dis* 2011; 70: 2050-2.
36. Hruska P, Kuruczova D, Vasku V, Bienertova-Vasku J. MiR-21 binding site SNP within ITGAM associated with psoriasis susceptibility in women. *PLoS One* 2019; 14: e0218323. doi: 10.1371/journal.pone.0218323.
37. Shi D, Zhong Z, Xu R, et al. Association of ITGAX and ITGAM gene polymorphisms with susceptibility to IgA nephropathy. *J Hum Genet* 2019; 64: 927-35.
38. Ehirchiou D, Xiong Y, Xu G, Chen W, Shi Y, Zhang L. CD11b facilitates the development of peripheral tolerance by suppressing Th17 differentiation. *J Exp Med* 2007; 204: 1519-24.
39. Garrett-Sinha LA, John S, Gaffen SL. IL-17 and the Th17 lineage in systemic lupus erythematosus. *Curr Opin Rheumatol* 2008; 20: 519-25.
40. Zhou M, Li LH, Peng H, et al. Decreased ITGAM and Fc γ RIIIA mRNA expression levels in peripheral blood mononuclear cells from patients with systemic lupus erythematosus. *Clin Exp Med* 2014; 14: 269-74.
41. Montani D, Perros F, Gambaryan N, et al. C-kit-positive cells accumulate in remodeled vessels of idiopathic pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2011; 184: 116-23.