Identifying Key Genes and Functionally Enriched Pathways in Th2-high Asthma by Weighted Gene Co-Expression Network Analysis

Yao CAO
  West China Second University Hospital of Sichuan University

Yi WU
  West China Second University Hospital of Sichuan University

Li LIN
  West China Second University Hospital of Sichuan University

Lin YANG
  West China Second University Hospital of Sichuan University

Xin PENG
  West China Second University Hospital of Sichuan University

Lina CHEN (✉ chenlina_78@163.com)
  West China Second University Hospital of Sichuan University

---

Research Article

Title Page

Identifying Key Genes and Functionally Enriched Pathways in Th2-high Asthma by Weighted Gene Co-Expression Network Analysis

Yao Cao[1], Yi Wu[1], Li Lin[1], Lin Yang[1], Xin Peng[1], Lina Chen[1*]

1. Division of pediatric pulmonology and immunology; Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, P.R China; NHC Key Laboratory of Chronobiology (Sichuan University)

*Correspondence to: Lina Chen; E-mail: chenlina_78@163.com

Identifying Key Genes and Functionally Enriched Pathways in Th2-high Asthma by Weighted Gene Co-Expression Network Analysis

Abstract

Background: Asthma is a chronic lung disease characterized by reversible inflammation of the airways. The imbalance of Th1/Th2 cells plays a significant role in the mechanisms of asthma. The aim of this study was to identify asthma-related key genes and functionally enriched pathways in a Th2-high group by using weighted gene coexpression network analysis (WGCNA).

Methods: The gene expression profiles of GSE4302, which included 42 asthma patients and 28 controls, were selected from the Gene Expression Omnibus (GEO). A gene network was constructed, and genes were classified into different modules using WGCNA. Gene ontology (GO) was performed to further explore the potential function of the genes in the most related module. In addition, the expression profile and diagnostic capacity (ROC curve) of hub genes of interest were verified by dataset GSE67472.

Results: In dataset GSE4302, subjects with asthma were divided into Th2-high and Th2-low groups according to the expression of the SERPINB2, POSTN and CLCA1 genes. A weighted gene coexpression network was constructed, and genes were classified into 7 modules. Among them, the red module was most closely associated with Th2-high asthma, which contained 60 genes. These genes were significantly enriched in different biological processes and molecular functions. A total of 8 hub genes (TPSB2, CPA3, ITLN1, CST1, SERPINB10, CEACAM5, CHD26 and P2RY14) were identified, and the expression levels of these genes (except TPSB2) were confirmed in dataset GSE67472. ROC curve analysis validated that the expression of these 8 genes exhibited excellent diagnostic efficiency for Th2-high asthma and Th2-low asthma.

Conclusions: The study provides a novel perspective on Th2-high asthma by WGCNA, and the hub genes and potential pathways involved may be beneficial for the diagnosis and management of Th2-high asthma.

Keywords: Asthma, weighted gene coexpression network analysis (WGCNA), hub gene, gene ontology

Background

Asthma is a common chronic inflammatory disease of the lower airway characterized by airflow obstruction, bronchial hyperresponsiveness, and airway inflammation. [1] The CD4 T-helper cell type 2 (Th2)-mediated pathway in the airway epithelium is critical in allergic asthma. [2] Drugs targeting Th2 cytokines (IL-4, IL-5, and IL-13) are in development. However, the etiology and progression of asthma are still under investigation.

Although multiple studies have shown differences in cytokine expression between Th2-high and Th2-low asthma[2,3], it is not known whether the gene signature of bronchial airway epithelium can identify Th2-high and Th2-low endotypes. Weighted gene coexpression network analysis (WGCNA) is a useful tool to identify functional pathways and candidate biomarkers and has recently been used to explore the complex relationships between gene network signatures and complicated phenotypes. [4] Meanwhile, WGCNA has been successfully adopted to investigate the gene-network signature, coexpression modules, and hub genes in respiratory diseases, such as severe asthma[5-7] and chronic obstructive pulmonary disease (COPD)[8-10].

In this study, we analyzed gene expression in asthmatic patients and healthy control subjects and measured the expression of genes related to type 2 inflammation. We aimed to identify hub genes through WGCNA to further explore the potential mechanism underlying Th2-high asthma.

Materials and Methods

Microarray Data Collection

We downloaded mRNA expression profiles GSE4302 and GSE67472 from the Gene Expression Omnibus (GEO). Specifically, airway epithelial brushings for microarray analysis of 42 subjects with asthma (not on inhaled steroids) and 28 healthy controls were obtained from GSE4302. Airway epithelial brushings for microarray analysis in GSE67472 were obtained from 62 subjects with mild-to-moderate asthma (not on inhaled steroids) and 43 healthy controls. Asthma subjects in the GSE4302 microarray dataset were divided into Th2-high and Th2-low asthma based on their expression of a three-gene signature of type 2 inflammation: SERPINB2, POSTN and CLCA1[2,10].

WGCNA

The coefficient of variation (CV) of each gene in the GSE4302 microarray dataset was calculated, and genes with CV >5% (4133 genes) were used to construct a gene coexpression network using the WGCNA package (version 1.69)[4] in R. In this study, we constructed a scale-free network ($R^2 = 0.9$) based on the criteria that soft-thresholding power β was set as 10, and 30 was chosen as the minimum number of genes in modules. A value of 0.25 was defined as the threshold for cut height to merge possible similar modules.

Functional Enrichment Analysis

To obtain further insights into the function of the genes in the module most related to Th2-high asthma, Gene Ontology (GO) analysis was performed and visualized with the R package "Clusterprofiler" (version 3.18.0)[11]. The ontology contains three categories: molecular function (MF), biological process (BP), and cellular component (CC). A Q value <0.05 was set as the cutoff.

Identification and Validation of Hub Genes

In the module-trait correlation analysis, genes with gene significance greater than 0.4 and high within-module connectivity of the modules (cor.geneModuleMembership > 0.8) were considered hub genes. The hub genes were further validated in datasets GSE67472. We screened the differentially expressed genes (DEGs) from Th2-high asthma vs. Th2-low asthma and Th2-high asthma vs. healthy controls using the "limma" R package. DEGs were considered genes with a false discovery rate (FDR) <0.05 and absolute log2 fold change (FC) ≥1. Among the hub genes, eight genes (TPSB2, CPA3, ITLN1, CST1, SERPINB10, CEACAM5, CHD26, P2RY14) were further validated in another dataset GSE67472. The AUC was calculated, and an ROC curve was plotted with the "ROCR" package to evaluate the capability of hub genes to distinguish Th2-high asthma and Th2-low asthma.

Results

Asthma Phenotyping Based on Epithelial Markers

Seventy subjects (28 healthy control subjects and 42 asthma subjects) in datasets GSE4302
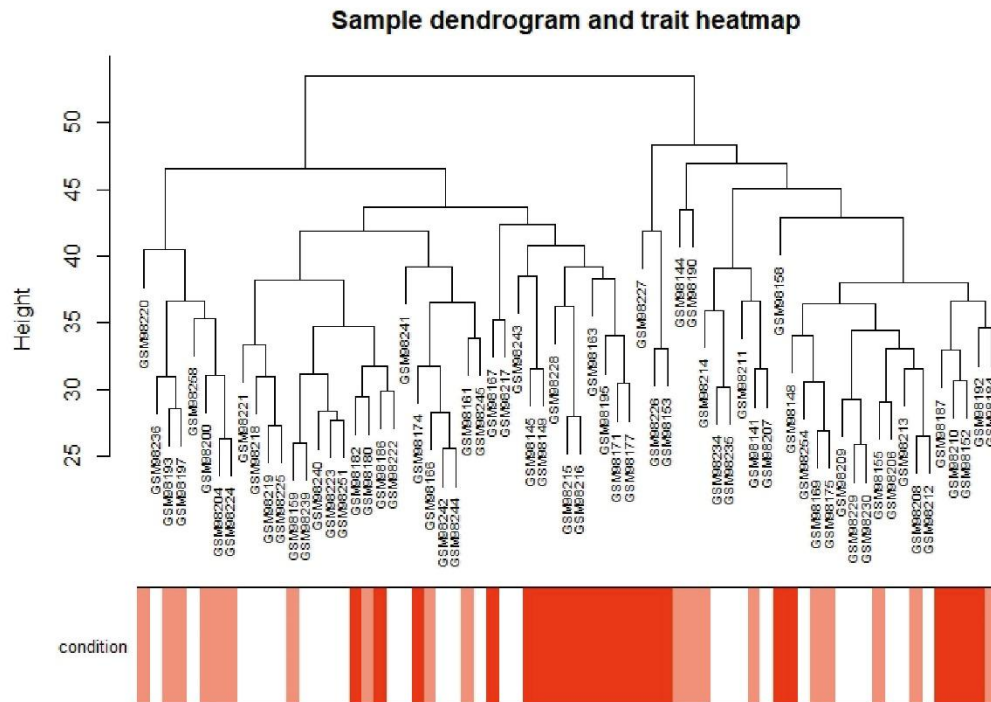
underwent unsupervised hierarchical clustering based on the microarray expression levels of SERPINB2, POSTN and CLCA1 (Figure 1). Approximately half of the asthma subjects (n = 22) with consistently high expression of a signature of Type 2 inflammation (SERPINB2, POSTN and CLCA1) were identified as Th2-high asthma, and 20 asthma subjects were identified as Th2-low asthma.
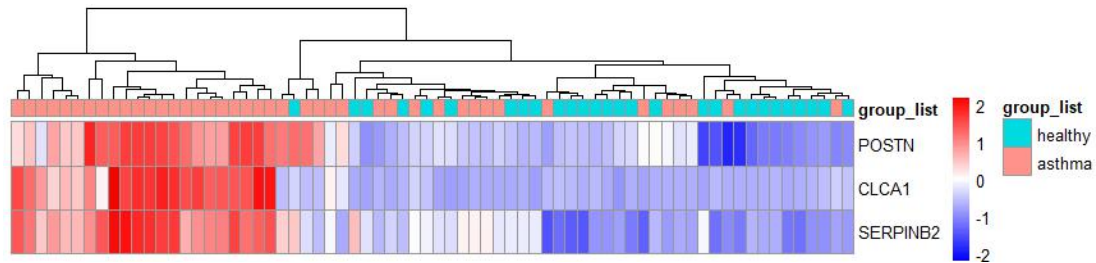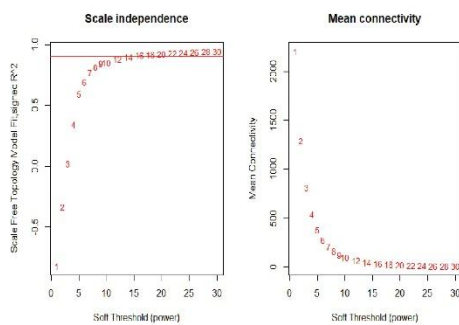

Figure 1. **Heatmap showing unsupervised hierarchical clustering of POSTN, CLCA1, and SERPINB2 expression levels in bronchial epithelium. Red represents a high expression level, and green represents a low expression level.**

Weighted Coexpression Network Construction

The GSE4302 samples were clustered using hierarchical agglomerative clustering with average linkage, and one outlier sample (GSM98201) was removed (Figure 2A). In this study, GSE4302 was used to construct coexpression networks. The power of $\beta = 10$ (scale-free $R2 = 0.9$) was selected as the correlation coefficient threshold to ensure a scale-free network (Figure 2B). A total of 7 modules were identified through WGCNA (Figure 2C).
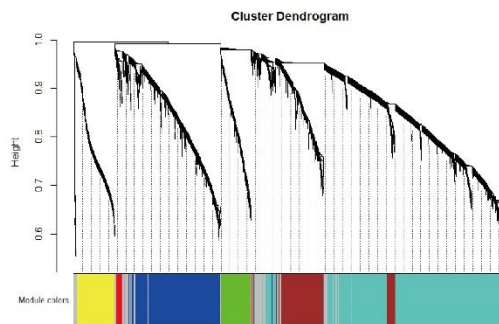
Figure 2. (a) Clustering dendrogram of 49 samples. (b) Determination of soft-thresholding power in the WGCNA. Analysis of the scale-free fit index for various soft-thresholding powers (β) and analysis of the mean connectivity for various soft-thresholding powers. (c) Clustering dendrogram of differentially expressed genes based on a dissimilarity measure (1-TOM).

Key Modules and Hub Genes Identification

Module-trait relationship analyses showed that multiple modules were related to Th2-high asthma, and the red module, which included 60 genes, was most significantly associated with Th2-high asthma (Figure 3A). This module was identified as the clinically significant module for further analysis. Figure 3B shows the significance of these genes in the red module, and 15 genes with high connectivity in the clinically significant module were identified as hub genes based on the cutoff criteria (|GS| > 0.4 and |MM| > 0.8).
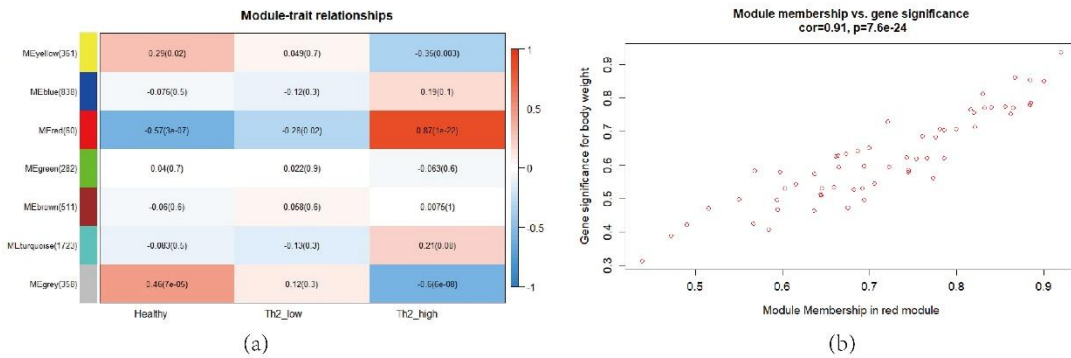
Figure 3. **(a) Heatmap of the correlation between module eigengenes and clinical traits (each cell contained the correlation coefficient and corresponding P value). (b) The gene significance for Th2-high asthma in the red module (one dot represents one gene in the red module).**

Gene Ontology of the Key Co-Expression Module

GO functional enrichment analysis of the genes in the clinically significant module showed that the genes were mainly enriched in BP and were involved in negative regulation of hydrolase activity, T cell activation and negative regulation of peptidase and endopeptidase activity (Figure 4). The genes in the MF category were mainly enriched in endopeptidase/peptidase inhibitor activity and endopeptidase/peptidase regulator activity (Figure 4). However, the genes in the CC group showed no enrichment results. These results indicated that these genes could be related to multiple biological pathways orchestrating Th2-high asthma development.
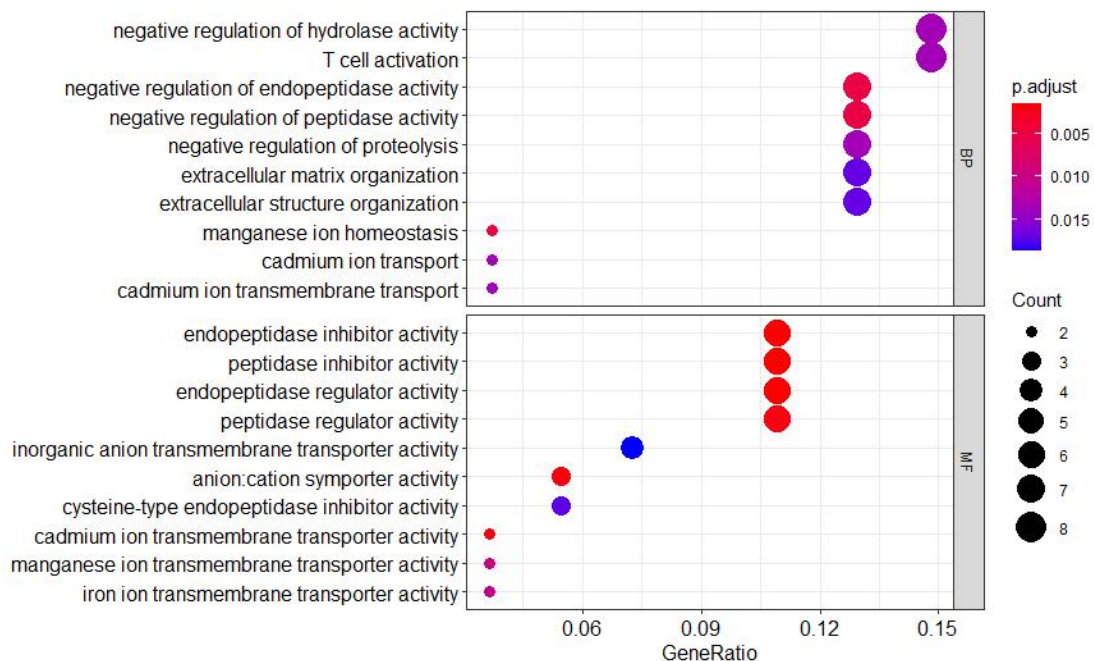
**Figure 4.** **Gene ontology of red module genes including biological process (BP) analysis and molecular function (MF) analysis.**

Validation of Hub Genes

In dataset GSE4302, 27 DEGs were identified in the group of Th2-high asthma vs. Th2-low

asthma (Figure 5A), and 38 DEGs were identified in the Th2-high asthma vs. healthy group (Figure 5B). The expression of 8 hub genes (CST1, P2RY14, SERPINB10, CEACAM5, CDH26, CPA3, TPSB2, and ITLN1) was significantly upregulated in Th2-high asthmatic subjects compared to Th2-low asthmatic subjects and healthy subjects (Figure 5A). Moreover, the expression levels of the seven hub genes were investigated in another dataset, GSE67472 (the gene TPSB2 expression data were not found in GSE67472). The seven genes were also significantly upregulated in the airway epithelium of Th2-high asthma patients compared to Th2-low asthma patients and controls in dataset GSE67472 (Figure 5B). In addition, the ROC curve indicated that these hub genes exhibited excellent diagnostic efficiency for Th2-high asthma and Th2-low asthma in datasets GSE4302 (Figure 5C) and GSE67472 (Figure 5D).
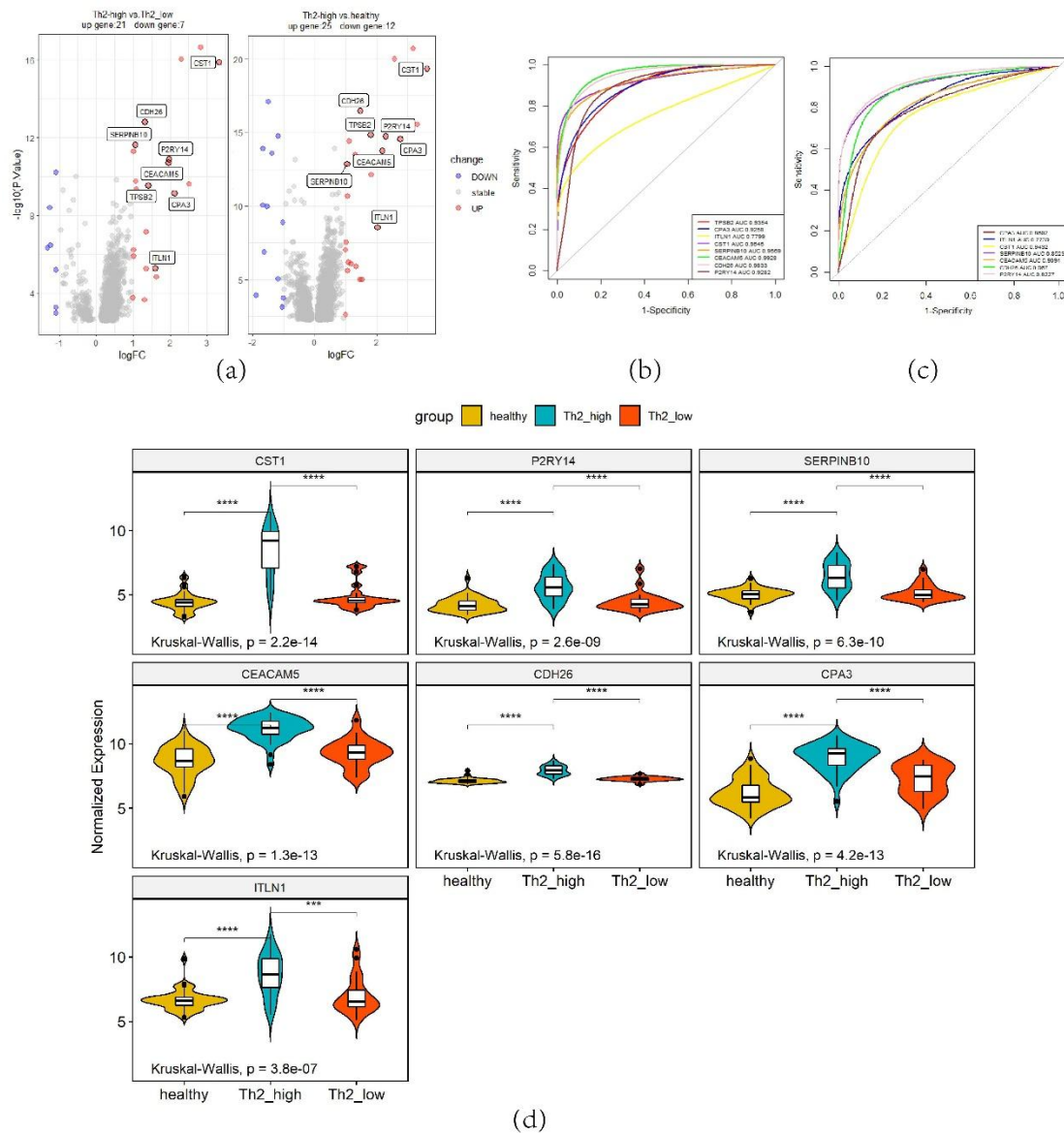
Figure 5. Validation of hub genes in datasets GSE4302 and GSE67472. (a) Volcano plot for differential gene expression in dataset GSE4302. ROC curve of hub genes in dataset (b) GSE4302 and dataset (c) GSE67472. (d) The expression levels of CST1, P2RY14, SERPINB10, CEACAM5, CDH26, CPA3 and ITLN1 were significantly upregulated in Th2-high asthma patients. The Kruskal–Wallis test was used to evaluate the statistical

significance of differences.

Discussion

Bronchial asthma a highly heterogenous disease. The etiology and progression of asthma are still unclear to date. Airway epithelial cells are thought to play a vital role in the development of allergic asthma. [12,13] In a study of asthmatic and healthy individuals, gene expression analysis from freshly brushed bronchial epithelial cells identified five phenotypes of asthma, each with unique biological pathways and clinical characteristics. [14] To date, the most common phenotypes identified include allergic asthma (extrinsic), nonallergic asthma (intrinsic), asthma with fixed airflow obstruction, infection-related asthma, late-onset asthma, and asthma associated with obesity, and two endotypes include type 2 asthma and nontype 2 asthma. [15,16]

Th2 cells and cytokines (an imbalance of Th1/Th2) play a major role in type 2 inflammation, which is considered a major pathophysiological mechanism of asthma. [2,4] Several reports have identified a typical Th2 feature characterized by elevated Th2-type inflammation in the lower airway. [5,17] In another study, gene expression in patients with mild asthma and healthy control subjects found that Th2 signature genes in lower airway epithelial cells were associated with atopy and eosinophilic airway inflammation. [18,19]

There is growing interest in seeking biomarkers to distinguish Th2-high and Th2-low phenotypes and to predict responsiveness to treatment. [20] At present, several biomarkers, such as fractional exhaled nitric oxide (FeNO), serum IgE, blood or sputum eosinophils and serum periostin, have been adopted. [21] Nirav R Bhakta et al. reported that the three-gene mean of periostin, CLCA1 and serpinB2 in airway epithelial brushings identifies Th2-high and Th2-low populations and predicts FEV1 improvement with inhaled corticosteroid (ICS). [22]

WGCNA can provide a deep understanding of gene networks, hub genes and pathogenesis associated with asthma. In this study, we speculated that assessing gene expression in airway epithelial cells using WGCNA may identify the gene-network signature, provide potential biomarkers of Th2-high asthma and further predict treatment responses.

Seven modules were identified as significantly correlated with asthma status. The red module observed in this study was positively correlated with asthma traits and regarded as the most important module in Th2-high asthma pathogenesis. It was further analyzed using a GO enrichment method and the results showed it to be mainly enriched in endopeptidase and peptidase inhibitor/regulator activity. Then, key genes of TPSB2, CPA3, ITLN1, CST1, SERPINB10, CEACAM5, CHD26, and P2RY14 were identified.

The above 8 hub genes were validated in another dataset GSE4302. We found that their expression levels (except TPSB2, which had not been studied in dataset GSE4302) were also significantly upregulated in Th2-high asthma patients compared to controls in dataset GSE4302. Although the mechanisms underlying asthma are unclear, some of these hub genes are under investigation.

For example, CST1 is a member of the type 2 cystatin (CST) superfamily, which is known to inhibit the proteolytic activities of cysteine proteases. Increased CPA3 expression in intraepithelial mast cells among Th2-high asthma patients was observed. [23] Zhanfeng Yan et al. discussed the coexpressed DEGs of CST1 and CPA3 and found that there were significant correlations of the CST1 and CPA3 genes with novel biomarkers involved in the comorbidity of rhinitis and asthma. [24] In addition, Natasha A Winter et al. found that the gene expression of TPSAB1/TPSB2 and

CPA3 correlated with sputum mast cells/basophils, and in severe asthma, TPSB2 and CPA3 were associated with eosinophilic airway inflammation and blood eosinophils. [25]

The hub gene ITLN1, which is expressed in airway epithelial cells and involved in inflammatory pathways downstream of IL-13, was found to be significantly upregulated in the bronchial brushings of patients with asthma compared with controls. [26,27] It has been reported that ITLN1 is a biomarker associated with disease susceptibility in asthma. [28] A study on ERPINB10-knockdown mice found that these groups of mice had diminished numbers of Th2 cells and were more susceptible to apoptosis, indicating that SERPINB10 may contribute to allergic inflammation and the Th2 response of asthma by inhibiting the apoptosis of Th2 cells. [29]

Nevertheless, the expression of these identified hub genes still needs to be validated in more datasets, and the exact mechanisms also need more exploration. Furthermore, the Th2-high subgroup of asthma is corticosteroid sensitive and benefits from ICS therapies. [2,30] In this respect, future research on the function of these genes not only helps promote our understanding of Th2-high and Th2-low endotypes of asthma but may also help to predict responsiveness to ICSs and guide treatment strategies for successful outcomes.

Conclusion

In summary, our study performed WGCNA to provide a framework of coexpression gene modules of Th2-high asthma and led to the identification of some hub genes that may help us to better understand the underlying pathogenesis and guide treatment decisions for the Th2-high subgroup of asthma. The exact molecular mechanisms of the hub genes and functional pathways in Th2 asthma still need to be further explored.

Availability of data and materials

Our data can be found in the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/, GSE4302 and GSE67472) database.

Abbreviations

WGCNA: weighted gene coexpression network analysis

GEO: Gene Expression Omnibus

GO: Gene ontology

Th2: T-helper cell type 2

COPD: chronic obstructive pulmonary disease

CV: coefficient of variation

MF: molecular function

BP: biological process

CC: cellular component

DEGs: differentially expressed genes

FDR: false discovery rate

FC: fold change

FeNO: fractional exhaled nitric oxide

ICS: inhaled corticosteroid

Acknowledgments

Funding

Author information

Yao Cao and Yi Wu are the co-first authorship.

Affiliations

Division of pediatric pulmonology and immunology; Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, P.R China; NHC Key Laboratory of Chronobiology (Sichuan University)

Yao Cao, Yi Wu, Li Lin, Lin Yang, Xin Peng & Lina Chen

Contributions

Yao Cao, Lina Chen and Yi Wu are responsible for the design of the study and prepared for the manuscript. Li Lin, Lin Yang and Xin Peng collected the data and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Corresponding authors

Correspondence to Lina Chen.

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest.

References

1.   Liu Z, Li M, Fang X, Shen L, Yao W, Fang Z, et al. Identification of surrogate prognostic biomarkers for allergic asthma in nasal epithelial brushing samples by WGCNA. Journal of Cellular Biochemistry. 2018;120(4):5137-50.

2.   Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, et al. T-helper Type 2–driven Inflammation Defines Major Subphenotypes of Asthma. American Journal of Respiratory and Critical Care Medicine. 2009;180(5):388-95.

3.   Peters MC, Mekonnen ZK, Yuan S, Bhakta NR, Woodruff PG, Fahy JV. Measures of gene expression in sputum cells can identify TH2-high and TH2-low subtypes of asthma. Journal of Allergy and Clinical Immunology. 2014;133(2):388-94.e5.

4.   Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1).

5.   Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, et al. Gene Expression Correlated with Severe Asthma Characteristics Reveals Heterogeneous Mechanisms of Severe Disease. American Journal of Respiratory and Critical Care Medicine. 2017;195(11):1449-63.

6.   Zhang Z, Wang J, Chen O. Identification of biomarkers and pathogenesis in severe asthma by coexpression network analysis. BMC Medical Genomics. 2021;14(1).

7.   He L-L, Xu F, Zhan X-Q, Chen Z-H, Shen H-H. Identification of critical genes associated with the development of asthma by co-expression modules construction. Molecular Immunology. 2020;123:18-25.

8.   Morrow JD, Qiu W, Chhabra D, Rennard SI, Belloni P, Belousov A, et al. Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. BMC Medical Genomics. 2015;8(1).

9.   Yi G, Liang M, Li M, Fang X, Liu J, Lai Y, et al. A large lung gene expression study identifying IL1B as a novel player in airway inflammation in COPD airway epithelial cells. Inflammation Research. 2018;67(6):539-51.

10.   Christenson SA, Steiling K, van den Berge M, Hijazi K, Hiemstra PS, Postma DS, et al. Asthma–COPD Overlap. Clinical Relevance of Genomic Signatures of Type 2 Inflammation in Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine. 2015;191(7):758-66.

11.   Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology. 2012;16(5):284-7.

12.   Hammad H, Lambrecht Bart N. Barrier Epithelial Cells and the Control of Type 2 Immunity. Immunity. 2015;43(1):29-40.

13.   Hirose K, Iwata A, Tamachi T, Nakajima H. Allergic airway inflammation: key players beyond the Th2 cell pathway. Immunological Reviews. 2017;278(1):145-61.

14.   Modena BD, Tedrow JR, Milosevic J, Bleecker ER, Meyers DA, Wu W, et al. Gene Expression in Relation to Exhaled Nitric Oxide Identifies Novel Asthma Phenotypes with Unique Biomolecular Pathways. American Journal of Respiratory and Critical Care Medicine. 2014;190(12):1363-72.

15.   Santus P, Saad M, Damiani G, Patella V, Radovanovic D. Current and future targeted therapies for severe asthma: Managing treatment with biologics based on phenotypes

and biomarkers. Pharmacological Research. 2019;146:104296.

16. Lötvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: A new approach to classification of disease entities within the asthma syndrome. Journal of Allergy and Clinical Immunology. 2011;127(2):355-60.

17. Bhakta NR, Christenson SA., Nerella S, Solberg OD, Nguyen CP, Choy DF, et al. IFN-stimulated Gene Expression, Type 2 Inflammation, and Endoplasmic Reticulum Stress in Asthma. American journal of respiratory and critical care medicine. 2018;197(3), 313–324.

18. G. Woodruff P. Subtypes of Asthma Defined by Epithelial Cell Expression of Messenger RNA and MicroRNA. Annals of the American Thoracic Society. 2013;10(Supplement):S186-S9.

19. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, et al. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. Proceedings of the National Academy of Sciences. 2007;104(40):15858-63.

20. Robinson D, Humbert M, Buhl R, Cruz AA, Inoue H, Korom S, et al. Revisiting Type 2-high and Type 2-low airway inflammation in asthma: current knowledge and therapeutic implications. Clinical & Experimental Allergy. 2017;47(2):161-75.

21. Diamant Z, Tufvesson E, Bjermer L. Which Biomarkers Are Effective for Identifying Th2-Driven Inflammation in Asthma? Current Allergy and Asthma Reports. 2013;13(5):477-86.

22. Bhakta NR, Solberg OD, Nguyen CP, Nguyen CN, Arron JR, Fahy JV, et al. A qPCR-based metric of Th2 airway inflammation in asthma. Clinical and translational allergy. 2013;3(1):24.

23. Dougherty RH, Sidhu SS, Raman K, Solon M, Solberg OD, Caughey GH, et al. Accumulation of intraepithelial mast cells with a unique protease phenotype in TH2-high asthma. Journal of Allergy and Clinical Immunology. 2010;125(5):1046-53.e8.

24. Yan Z, Liu L, Jiao L, Wen X, Liu J, Wang N. Bioinformatics Analysis and Identification of Underlying Biomarkers Potentially Linking Allergic Rhinitis and Asthma. Medical Science Monitor. 2020;26.

25. Winter NA, Qin L, Gibson PG, McDonald VM, Baines KJ, Faulkner J, et al. Sputum mast cell/basophil gene expression relates to inflammatory and clinical features of severe asthma. Journal of Allergy and Clinical Immunology. 2021;148(2):428-38.

26. Kuperman DA, Lewis CC, Woodruff PG, Rodriguez MW, Yang YH, Dolganov GM, et al. Dissecting asthma using focused transgenic modeling and functional genomics. Journal of Allergy and Clinical Immunology. 2005;116(2):305-11.

27. Watanabe T, Chibana K, Shiobara T, Tei R, Koike R, Nakamura Y, et al. Expression of intelectin-1 in bronchial epithelial cells of asthma is correlated with T-helper 2 (Type-2) related parameters and its function. Allergy, Asthma & Clinical Immunology. 2017;13(1).

28. Pemberton AD, Rose-Zerilli MJ, Holloway JW, Gray RD, Holgate ST. A single-nucleotide polymorphism in intelectin 1 is associated with increased asthma risk. J Allergy Clin Immunol. 2008;122(5):1033-4.

29. Mo Y, Ye L, Cai H, Zhu G, Wang J, Zhu M, et al. SERPINB10 contributes to asthma by inhibiting the apoptosis of allergenic Th2 cells. Respiratory Research. 2021;22(1).

30.   Fahy JV. Type 2 inflammation in asthma — present in most, absent in many. Nature Reviews Immunology. 2014;15(1):57-65.