

Identification of surrogate prognostic biomarkers for allergic asthma in nasal epithelial brushing samples by WGCNA

Zhaoyu Liu^{1,2}  | Ming Li^{1,2} | Xiangming Fang¹ | Lu Shen¹ | Wenxia Yao² | Zhiyuan Fang² | Jitao Chen^{1,2} | Xiao Feng² | La Hu² | Zicheng Zeng² | Chunyi Lin¹ | Jinsheng Weng² | Yuxiong Lai² | Gao Yi^{1,2}

¹Department of Respiratory Medicine, State Key Laboratory of Respiratory Disease, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

²Department of Center Laboratory, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

Correspondence

Gao Yi, Jinsheng Weng and Yuxiong Lai, Department of Center Laboratory, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou 510700, China.

Email: GYWYer@126.com (G.Y.);

jweng35@yahoo.com (J.W.);

laiyuxiong163@163.com (Y.L.)

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 81400013, 81570189, 81873419; Science and Technology Program of Guangzhou, China,

Grant/Award Number: 201804010018; Science and Technology Planning Project of Guangdong Province, China, Grant/Award Number: 2014A20212329; Department of education of Guangdong Province, China, Grant/Award Number: 2016KTSCX110

Abstract

Background: Allergic asthma is a lower respiratory tract disease of Th2 inflammation with multiple molecular mechanisms. The upper and lower airways can be unified by the concept of a united airway and, as such, gene expression studies of upper epithelial cells may provide effective surrogate biomarkers for the prognostic study of allergic asthma.

Objective: To identify surrogate biomarkers in upper airway epithelial cells for the prognostic study of allergic asthma.

Methods: Nasal epithelial cell gene expression in 40 asthmatic and 17 healthy control subjects were analyzed by weighted gene coexpression network analysis (WGCNA) to identify gene network modules and profiles in allergic asthma. Functional enrichment analysis was performed on the coexpression genes in certain highlighted modules.

Results: A total of 13 coexpression modules were constructed by WGCNA from 2804 genes in nasal epithelial brushing samples of the 40 asthmatic and 17 healthy subjects. The number of genes in these modules ranged from 1086 (Turquoise module) to 45 (Salmon). Eight coexpression modules were found to be significantly correlated ($P < 0.05$) with two clinic traits, namely disease status, and severity. Four modules were positively correlated ($P < 0.05$) with the traits and these, therefore, contained genes that are mostly overexpressed in asthma. Contrastingly, the four other modules were found to be negatively correlated with the clinic traits. Functional enrichment analysis of the positively correlated modules showed that one (Magenta) was mainly enriched in mast cell activation and degranulation; another (Pink) was largely involved in immune cell response; the third (Yellow) was predominantly enriched in transmembrane signal pathways; and the last (Blue) was mainly enriched in substructure components of the cells. The hub genes in the modules were *KIT*, *KITLG*, *GATA2*, *CD44*, *PTPRC*, and *CFTR*, and these were confirmed as having significantly higher expression in the nasal epithelial cells. Combining the six hub genes enabled a relatively high capacity for discrimination

Zhaoyu Liu and Ming Li contributed equally to this work.

between asthmatics and healthy subjects with an area under the receiver operating characteristic (ROC) curve of 0.924.

Conclusions: Our findings provide a framework of coexpression gene modules from nasal epithelial brushing samples that could be used for the prognostic study of allergic asthma.

KEYWORDS

allergic asthma, gene expression, WGCNA

1 | INTRODUCTION

Asthma is a common chronic inflammatory disease of the lower airway characterized by airflow obstruction, bronchial hyperresponsiveness, and airway inflammation.¹ Pathologically, a mixture of wheezing, cough, dyspnea, and chest tightness are typical asthmatic symptoms.²⁻⁴ Although a range of preventive and therapeutic strategies have been applied in the treatment of asthma, its prevalence has rapidly increased across the world, especially in children and young adults.^{5,6} Airway epithelial cells, located on the surface of the respiratory tract, are thought to play a vital role in the development of allergic asthma.^{7,8} Gene expression studies of brushed bronchial epithelial cells in the lower airway have identified different asthmatic phenotypes based on observed expression profiles. 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.⁹ In another study, gene expression in patients with mildly asthmatic and healthy control subjects found that upregulated expression of interleukin-13 (IL-13)/T2 signature genes in lower airway epithelial cells was associated with atopy and eosinophilic airway inflammation.^{10,11} Notwithstanding these findings, gene expression differences in upper airway nasal epithelial cells from asthmatics remain poorly understood.

Weighted gene coexpression network analysis (WGCNA) is a relatively recent methodology for exploring the complex relationships between gene expression profiles and phenotypes.¹² The WGCNA transforms gene expression data into networks, known as modules, which contain groups of similarly behaving genes with a common biological relationship or function, thus providing modules that may be responsible for phenotypic traits of interest. It has been successfully used to study various biological processes such as cancer,¹³ severe asthma,¹⁴ and chronic obstructive pulmonary disease (COPD),^{15,16} proving to be quite useful for the identification of

potential biomarkers and therapeutic targets. The WGCNA can provide useful information not only for comparing differentially expressed genes but also for investigating interactions between genes in different coexpression modules.

In this study, we aimed to construct coexpression modules using nasal epithelial brushings from allergic asthmatics. The WGCNA was used to associate coexpression modules with numerous clinic traits including asthma status, severity, age, and gender. The biological function was determined for each module and those that were significantly associated with status and severity were highlighted. Gene ontology (GO) enrichment analysis was performed on the highlighted modules and “hub” genes were further identified using STRING (Search Tool for the Retrieval of Interacting Genes) which helped clarify the dominant genes in each module. Our study provides a framework of coexpression gene modules in nasal epithelial cells, which could be used to assess disease status and severity in allergic asthma.

2 | MATERIALS AND METHODS

RNA sequence data of nasal epithelium from allergic asthmatics and patient clinic traits were obtained from the National Center for Biotechnology Information (NCBI) GEO DataSets with access number GSE41861. The combined dataset contained 57 nasal epithelial samples from participants with ($n=40$) and without ($n=17$) asthma. The samples were obtained from the tissue bank at Translational Research Sciences, Hoffmann-La Roche, Nutley, New Jersey. Before implementing the WGCNA algorithms, the gene IDs were respectively mapped to the microarray probes using annotation information. Probes matching more than one gene were eliminated from the dataset, and the average expression value for genes measured by multiple probes was then calculated. The threshold value was based on the number of probes with different thresholds of expression.

2.1 | Identification of differentially expressed genes (DEGs)

The raw data used for analysis were presented as CEL files by Affymetrix platform GPL570, and analysis was performed using the R programming language (version 3.3.3) and Bioconductor. Raw data were normalized using a robust multiarray average expression measure. Probesets were annotated using the annotation available in Bioconductor and those that mapped to nonunique locations in the genome were flagged and ignored in the subsequent analysis. Differentially expressed genes (DEGs) were calculated using DESeq, an R software package.

2.2 | Construction of coexpression modules

The soft-thresholding power β was calculated in the construction of each module using the pickSoftThreshold function of the WGCNA. This method provides a suitable power value for network construction by calculating the scale-free topology fit index for a set of candidate powers, ranging from 1 to 30. In this way, the appropriate power was determined when the index value for the reference dataset exceeded 0.8. Once the soft-thresholding power value was set, the WGCNA algorithms were used to construct the coexpression modules.

Coexpression modules were constructed using the WGCNA algorithms implemented in the R software. A one-step network construction method was used to identify coexpression modules using the blockwiseModules function. The minimum number of genes for each module was set at 40, and the strength of interaction (correlation) between each module was analyzed and visualized using the heatmap tool package.

2.3 | Construction of module-clinic trait relationships

The relationships between modules and clinic traits were assessed using the correlations between the module

eigengene and the clinic trait, allowing the identification of modules that are correlated with external traits and searching for the most significant associations. For each module, the quantitative measure of module membership MM was defined as the correlation of the module eigengene and the gene expression profile; gene significance (GS) was defined as the absolute value of the correlation between the gene and the clinic trait. Genes with high significance for a clinic trait as well as high module membership (MM) were identified using GS and MM measures.

2.4 | Functional enrichment analysis of coexpression modules

The constructed modules were arranged by a number of genes and functional enrichment analysis was then performed on the genes in those modules. The corresponding gene annotation information was mapped to the Database for Annotation, Visualization, and Integrated Discovery. GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were conducted using the ClusterProfiler function of the R software; a *P* value of less than or equal to 0.05 after correction was considered statistically significant. If more than 10 records were returned, the top 10 were extracted. Modules of interest were visualized using STRING, and only experimentally validated interactions with a combined score of more than 0.4 were selected as significant. Using node degree and interaction score as the key topological parameters, the maximally connected genes were informally referred to as “hub” genes.

2.5 | Statistical analysis

Statistical analysis was performed using SPSS 22.0 and R software 3.3.3. Graphs were generated using Graph-Pad Prism 6.0. Student *t* test and the Mann-Whitney U test analyses of variance were used to evaluate statistical differences between asthmatic and healthy

TABLE 1 Subject characteristics

	Upper airway (nasal)		Lower airway (bronchial)	
	Control	Asthma	Control	Asthma
Number, N	17	40	30	51
Age, y	38.4 (22-66)	33.1 (19-58)	38.8 (21-66)	32.8 (19-58)*
Male/female, N	7/10	21/19	11/19	27/24
Cell type	Epithelium	Epithelium	Epithelium	Epithelium
Severity (mild/moderate/severe)	–	19/16/5	–	25/21/5

Data are presented as mean values with Min-Max ranges unless stated otherwise. The Wilcoxon test method was used to test differences between the groups.

**P* < 0.05 for the difference between asthma (*n* = 51) and healthy control (*n* = 30) in the low airway (bronchial) epithelial cells.

subjects. Receiver operating characteristic (ROC) curves were generated to evaluate the diagnostic accuracy of each hub gene, and the area under the curve (AUC) was used to evaluate sensitivity and specificity. All P values were two-sided and $P < 0.05$ was considered to be significant.

3 | RESULTS

3.1 | Subject characteristics

The clinical characteristics of subjects in this study are presented in Table 1; 57 nasal epithelial samples and 81 lower airway epithelial samples from the GSE41861 cohort were included. The nasal group contained 17 healthy participants and 40 asthmatics, and the lower airway group involved 30 normal controls and 51 subjects with asthma. The asthmatics in the lower airway epithelial group were somewhat younger than the normal controls. Overall, patients with asthma and the non-asthma controls did not differ in age or gender.

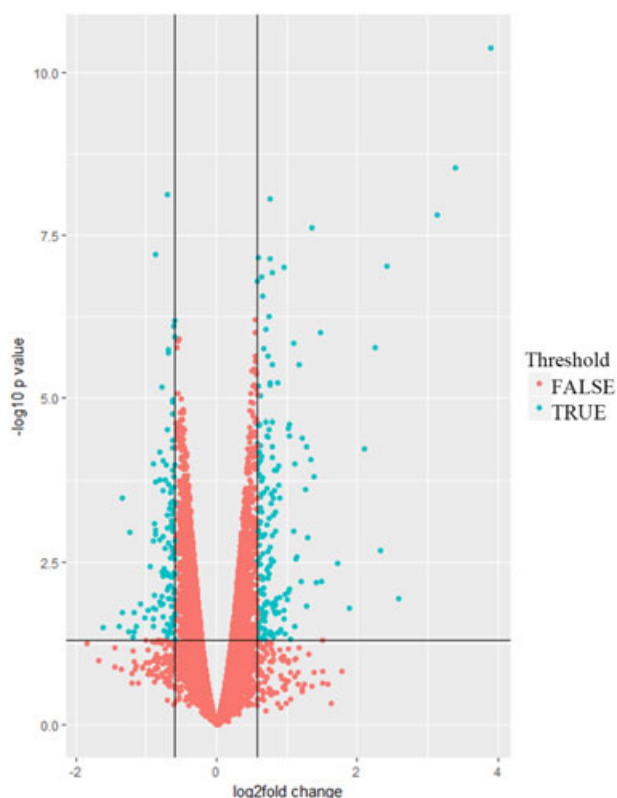


FIGURE 1 The volcano plot shows the upregulated and downregulated genes in nasal epithelial cells from asthma. The horizontal axis represents the fold change between health and asthmatics small airway epithelium. The vertical axis represents the P value of t test for the differences between health and asthmatics small airway epithelium. The genes most relevant for asthma and this manuscript are highlighted in blue

3.2 | Identification of DEGs in nasal epithelial brushings from allergic asthmatics

The nasal epithelial samples from cohort GSE41861 were analyzed using R software and its extension packages. Using an absolute fold change of equal to or greater than 1.5 and with $P < 0.05$, a total of 307 genes were identified of which 190 were upregulated and 117 were downregulated. By using a lower cut-off of a 1.2-fold change and $P < 0.05$, 1183 upregulated and 1621 downregulated genes were identified (Figures 1 and 2). A volcano plot of all probesets is shown in Figure 1 and a heatmap of the 50 most changed genes is shown in Figure 2, including the 25 most upregulated and downregulated genes. The complete list of gene and statistic symbols is available in online Supporting Information Supplementary Table S1.

3.3 | Construction of coexpression modules

As a result, the expression values of the 2804 genes exhibiting a 1.2-fold change in the 57 nasal epithelial brushing samples were used to construct coexpression modules using the WGCNA algorithms. Hierarchical

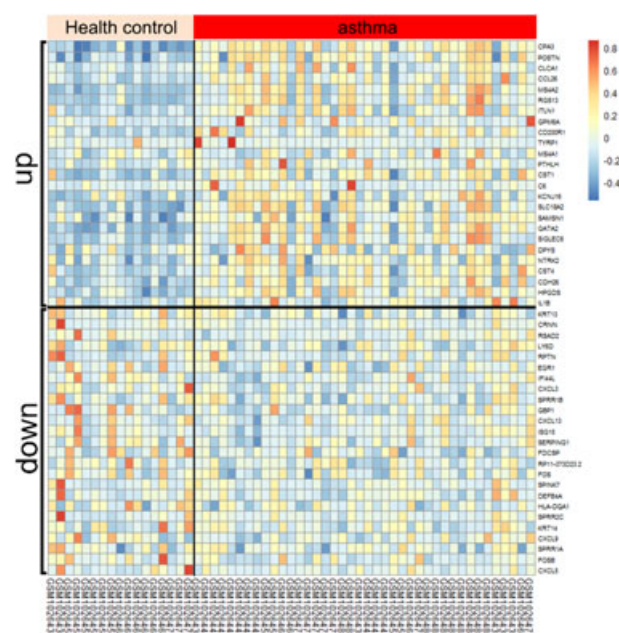


FIGURE 2 Heatmap of differentially expressed genes in nasal epithelial cells from asthmatics. The heatmap shows differential gene expression ($P < 0.05$) between health and asthma nasal epithelium from GSE41861. The colored, yellow, and red bars above the heatmap indicate the different groups of small airway epithelium. Genes with a higher expression in allergic asthma small airway epithelium are shown in the upper part of the heatmap, genes with a lower expression in the lower part. Blue represents lower and red higher relative expression

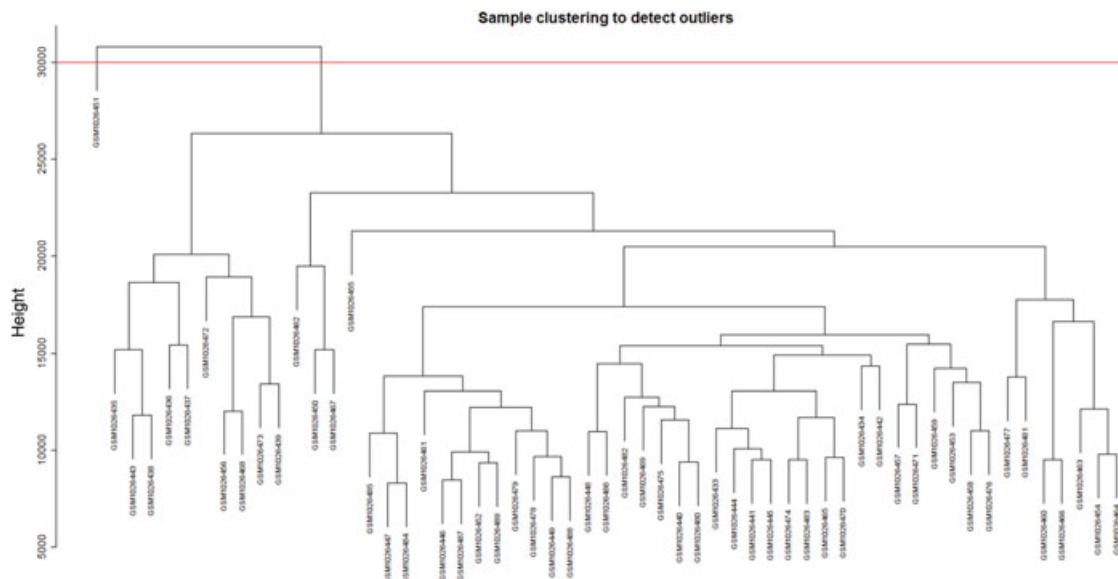


FIGURE 3 Sample clustering to detect outliers. There was obviously one outlier (GSM1026451), and it was removed from the sample cohort before the next step analysis

clustering analysis was performed with the flashClust function and the results are presented in Figure 3. We found one obvious outlier (GSM1026451) which was removed from the cohort. The other 56 samples mainly yielded two clusters, where GSM1026435, GSM1026436, GSM1026437, GSM1026438, GSM1026439, GSM1026443, GSM1026456, GSM1026468, GSM1026472, and GSM1026473 became one group while the remaining samples formed the other. Before the WGCNA, we

performed network topology analysis to determine candidate power values for relative, balanced scale independence, and mean connectivity in the WGCNA. As shown in Figure 4, a power value of 9 was the lowest power for which scale independence was below 0.8, and this was selected to produce a hierarchical clustering tree of the 2804 genes (Figure 5). We identified 13 modules ranging in size from 1086 genes (Turquoise) to 45 genes (Salmon). In addition, an improper module (Gray) of

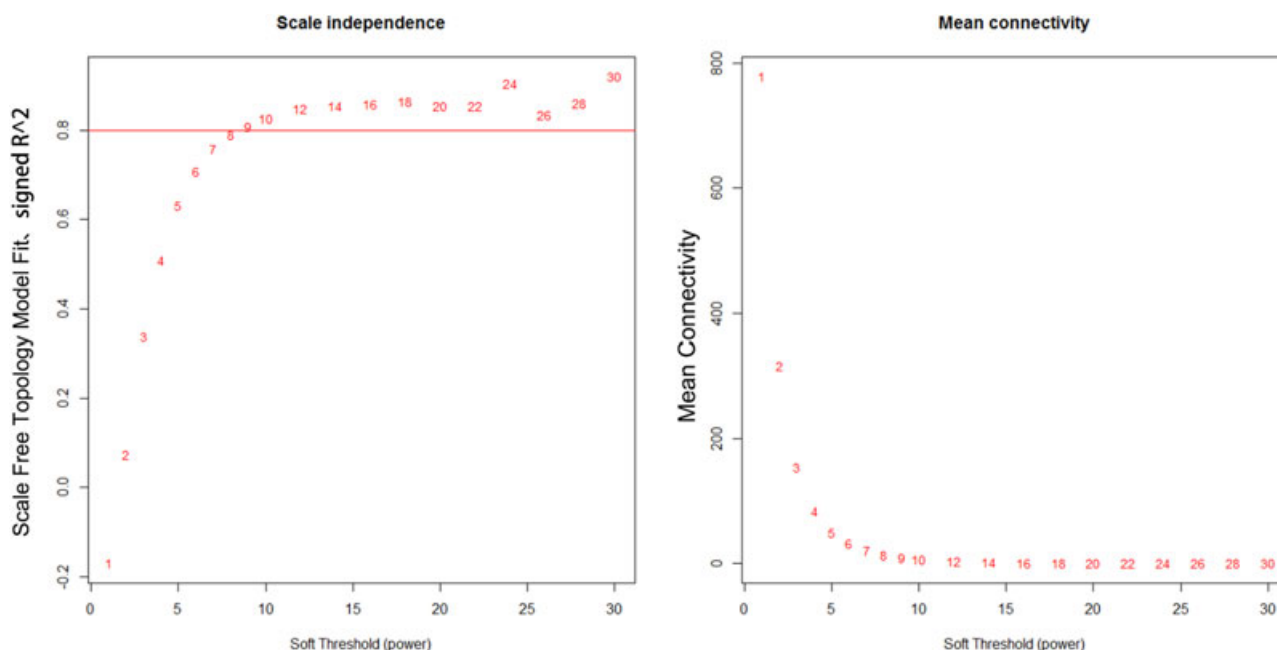


FIGURE 4 Analysis of network topology for a set of soft-thresholding powers. The left picture displays the scale free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right picture shows the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis)

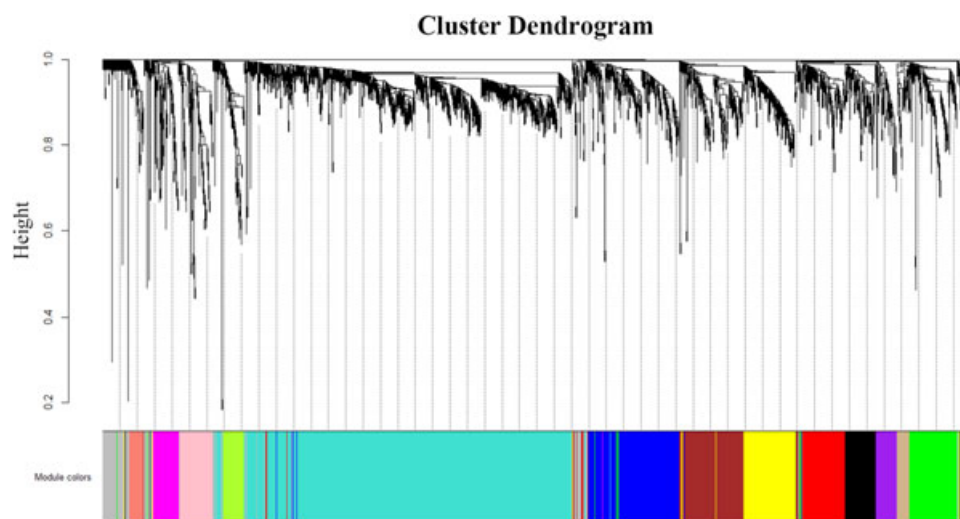


FIGURE 5 Clustering dendrograms of genes, with dissimilarity based on topological overlap, together with assigned module colors. As a result, 13 coexpression modules were constructed and were shown in distinctive color

genes not belonging to any of the other 13 modules was defined (Table 2). Interactions between the 13 modules were then analyzed (Figure 6).

3.4 | Quantifying module-clinic trait associations

The corresponding clinic traits dataset was downloaded from the NCBI database, and the associations between them and the modules were identified based on the correlation between module eigengene and clinic trait (Figure 7). In addition, the eigengene

dendrogram and heatmap were used to quantify module similarity by eigengene correlation (Figure 8). The results indicate that eight modules were significantly correlated with asthma clinic traits; of these, four (magenta, pink, yellow, and blue) were positively correlated to two clinic traits, namely disease status and severity, therefore including genes that are mostly overexpressed in asthma. In contrast, four modules (turquoise, salmon, red, and green-yellow) were found to negatively correlate with asthma traits, meaning that genes in these modules are predominantly under-expressed in asthma cases.

TABLE 2 Correlation of module eigengene with allergic asthma disease trait

Module colors	Genes	Asthma status		Asthma severity	
		Correlation	<i>P</i> value	Correlation	<i>P</i> value
Black	124	−0.46	3e−04	−0.21	0.1
Blue	285	0.58	2e−06	0.43	8e−04
Brown	216	0.42	0.001	0.14	0.3
Green	169	0.44	7e−04	0.14	0.3
Green-yellow	69	−0.43	9e−04	−0.33	0.01
Gray	159	0.7	2e−09	0.4	0.002
Magenta	90	0.49	1e−04	0.34	0.01
Pink	111	0.41	0.002	0.4	0.002
Purple	70	−0.36	0.006	−0.051	0.7
Red	151	−0.54	2e−05	−0.28	0.04
Salmon	45	−0.46	3e−04	−0.39	0.003
Tan	47	0.51	5e−05	0.24	0.07
Turquoise	1086	−0.55	1e−05	−0.51	7e−05
Yellow	182	−0.43	9e−04	0.26	0.05

The modules that were found by the WGCNA in the dataset are listed together with the number of genes they contain (shown in the second column). Eight coexpression modules were significantly correlated to asthma status, severity, and the correlation coefficients and *P* value are listed in the following columns.

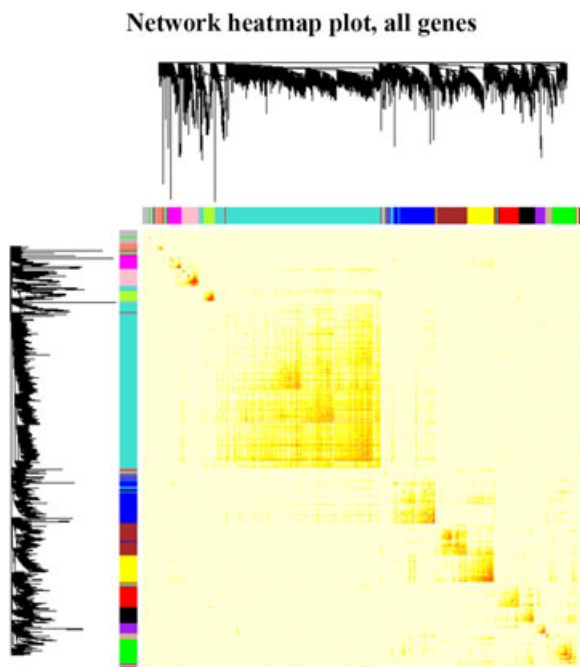


FIGURE 6 Visualizing the gene network using a heatmap plot. The heatmap describes the Topological Overlap Matrix (TOM) among all DEGs in the analysis. Light color represents low overlap and progressively darker red color indicates higher overlap. Blocks of darker colors along the diagonal are the coexpression modules. The gene dendrogram and module assignment are also shown along the left side and the top. DEGs, differentially expressed genes

Subsequently, the WGCNA was further used to calculate GS vs MM in the positively correlated modules. We found that the genes in the magenta,

pink, yellow, and blue modules that are most significantly associated with asthma characteristic traits (GS) were also the most important elements of the modules (MM), as demonstrated by the upper right genes in the plot of Figure 9. In this way, genes such as *GATA2*, *PTGS1*, *POSTN*, and *CD44* can be observed in the upper right section of Figure 9, indicating that they are not only key components in the underlying biological function of asthma but are also highly correlated to its clinical traits.

3.5 | Functional enrichment analysis

GO enrichment analysis was used for genes in the four positively correlated modules (Figure 10 and Table 3). Genes in magenta were mainly enriched in GO:0035162 (embryonic hemopoiesis), GO:0045576 (mast cell activation), and GO:2000116 (regulation of cysteine-type endopeptidase). Genes in pink were mainly enriched in GO:0002250 (adaptive immune response), GO:0045576 (mast cell activation), and GO:0002366 (leukocyte activation involved in immune response). Genes in blue were mainly enriched in GO:0005885 (Arp2/3 protein complex) and GO:0098562 (cytoplasmic side of the membrane). Finally, genes in yellow were mainly involved in membrane signal transduction, including GO:0031594 (neuromuscular junction), GO:0005765 (lysosomal membrane), and GO:0098852 (lytic vacuole membrane).

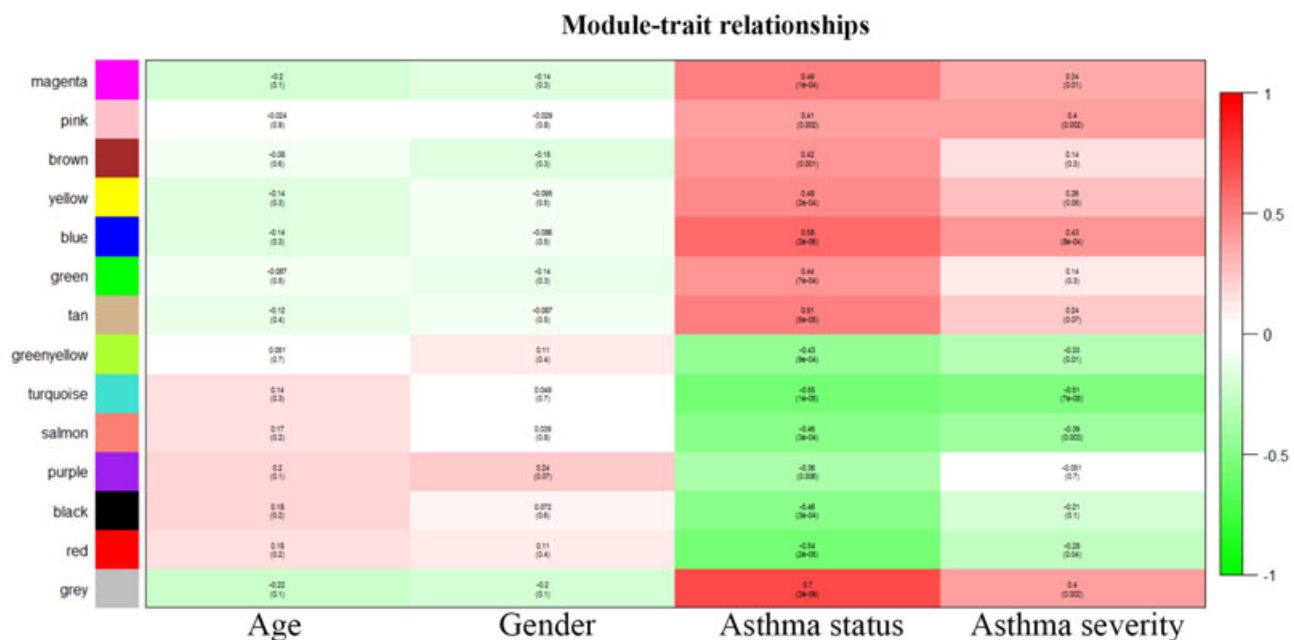


FIGURE 7 Module-trait relationships. Each row presents a module eigengene, column to a clinic trait. Each cell contains the corresponding correlation and *P* value. The table is color-coded by correlation according to the color legend

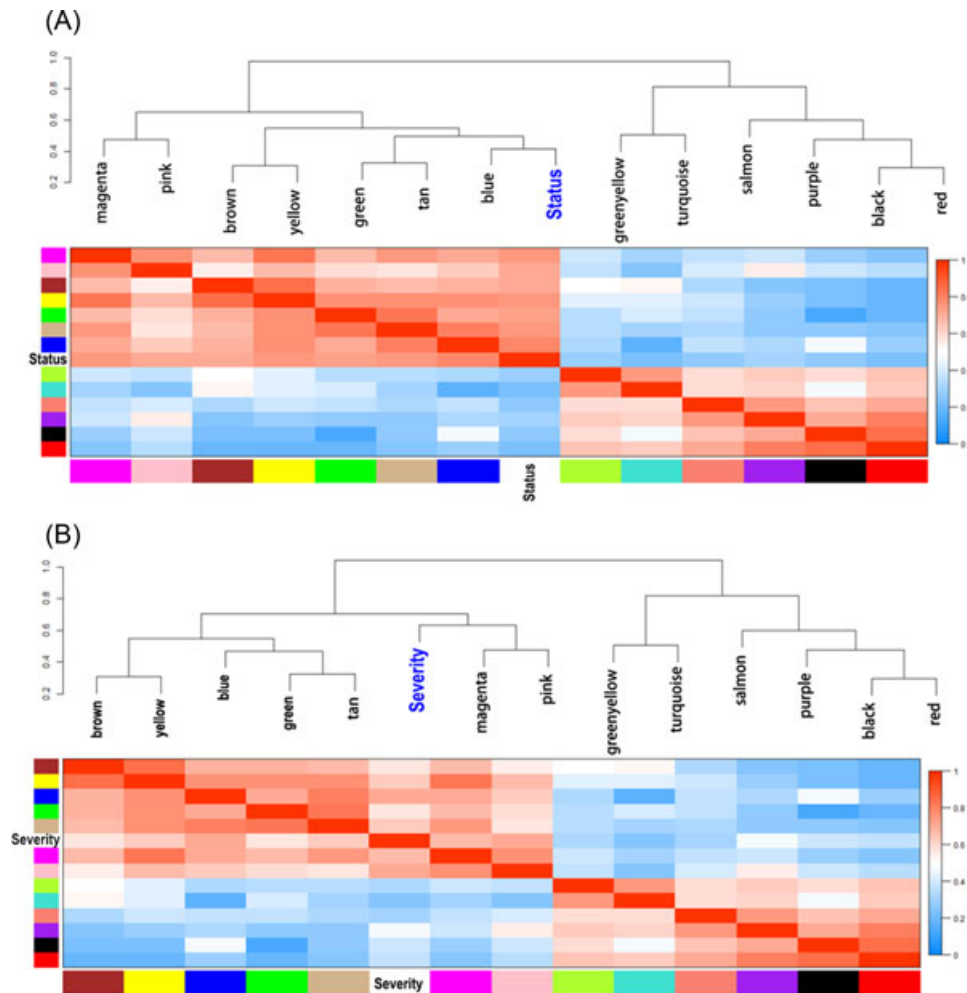


FIGURE 8 The eigengene dendrogram and heatmap identify groups of correlated eigengenes termed metamodules. As a result, A, the upper panel indicates magenta, pink, brown, yellow, green, tan, and blue modules are highly related to patient's disease status. B, The under panel represents that brown, yellow, blue, green, tan, magenta, and pink modules are highly involved in disease severity

3.6 | Module visualization and hub genes

The four positively correlated modules were visualized using the STRING database and only experimentally validated interactions with a combined score greater than 0.4 were selected as significant. Intramodular connectivity was calculated for each gene and those with high intramodular connectivity were considered hub genes. The hub genes *KIT*, *KITLG*, *GATA2*, *CD44*, *PTPRC*, and *CFTR* are shown in Figure 11.

3.7 | Verification of hub gene expression and ROC curve analysis

Analysis of hub gene expression was performed on the upper and lower airway epithelial cell samples and the results are presented in Figure 12. We found significantly higher *KIT*, *KITLG*, *GATA2*, *CD44*,

PTPRC, and *CFTR* expression in the 40 asthmatic nasal epithelial cells than in the 17 healthy controls. In addition, we found that *KIT*, *GATA2*, and *CD44* had significantly higher expression in the lower airway epithelial cells of 51 patients with asthma and 30 healthy subjects. However, there was no difference in *KITLG*, *PTPRC*, or *CFTR* expression levels in the asthmatic lower airway epithelial cells. Subsequently, ROC curve analysis was used to evaluate the diagnostic prediction value of the hub genes for asthma. This analysis revealed that the area under the ROC curve for *KIT* was 0.691 ($P = 0.023$). At the optimal cut-off value of 0.619, sensitivity and specificity were 41.2% and 77.5%, respectively. Similar results for *KITLG*, *GATA2*, *CD44*, *PTPRC*, and *CFTR* are presented in Figure 13 and Table 4. Combining these hub genes possesses a moderate ability for discriminating between asthmatics and healthy subjects with an area under the ROC curve of 0.924.

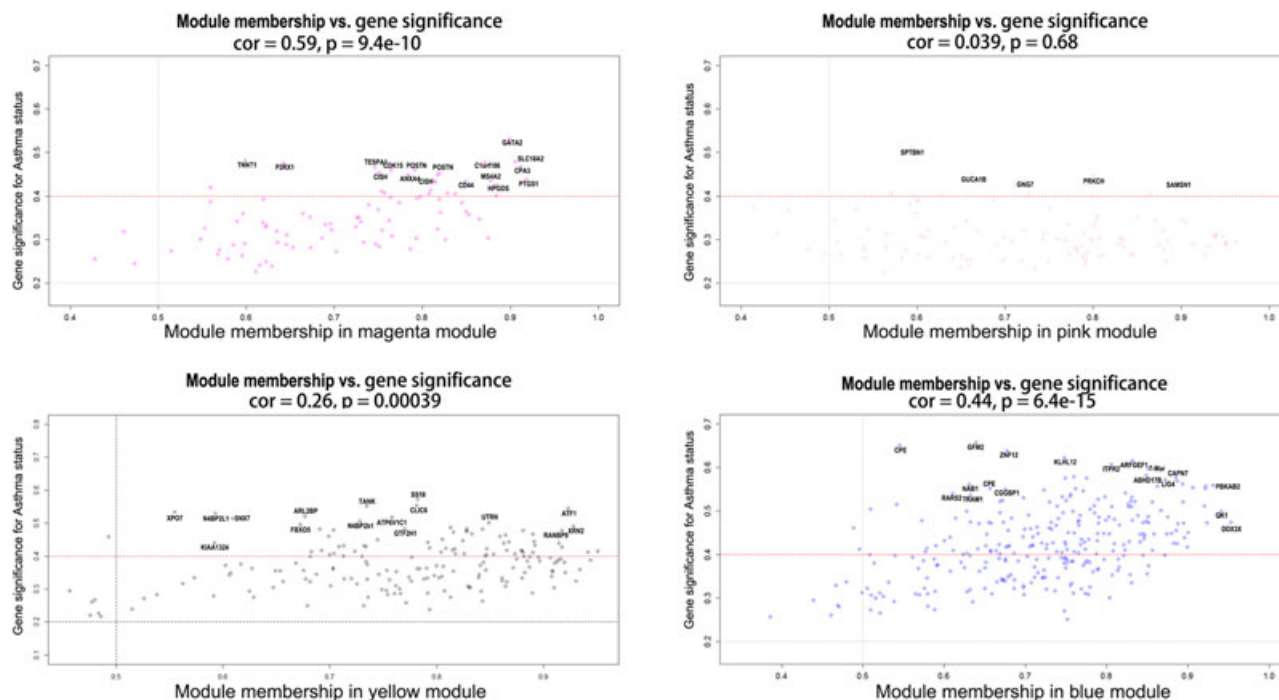


FIGURE 9 A scatterplot of gene significance (GS) for asthma status vs module membership (MM) in magenta, pink, yellow, and blue modules. The corresponding correlation coefficient and *P* value are listed above the scatterplot

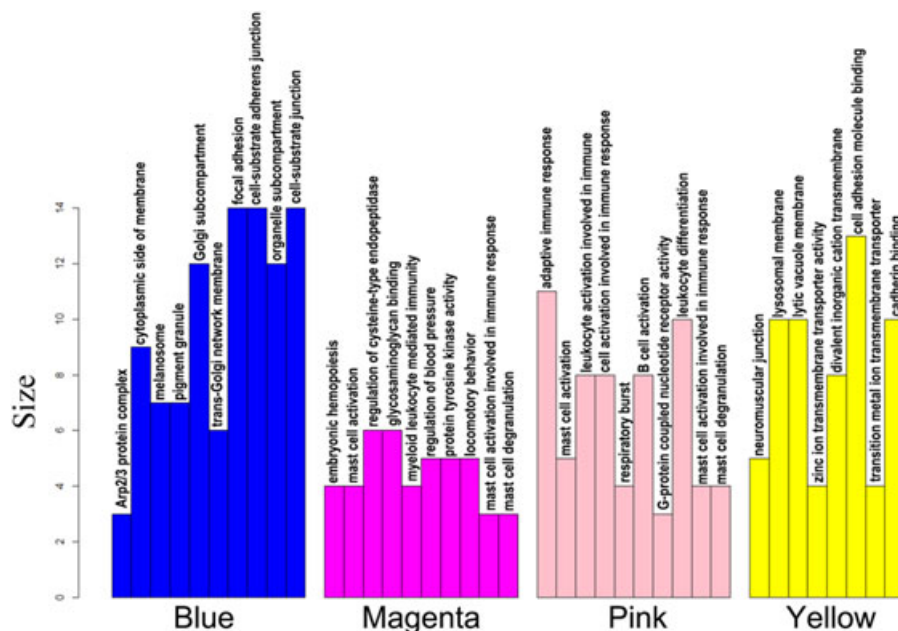


FIGURE 10 GO enrichment analysis of genes in blue, magenta, pink, and yellow modules. GO, gene ontology

4 | DISCUSSION

Although allergic asthma is a severe inflammatory disease in the lower airway, clinical manifestation is not limited to the lower respiratory tract.^{17,18} The interrelation between the lower and upper airways has been argued repeatedly in the past; they can be unified by the concept of a “united airway.”^{19–21} Clinical, epidemiological, and immunological relationships between the two areas have

been verified at the experimental level and published official documents of scientific research.^{22,23} Much research has demonstrated that most patients with allergic asthma have concomitant nasal symptoms as well as sinus mucosal disease,^{24,25} and chronic rhinitis has also been proven to be a risk factor for asthma development.²⁶ More evidence of the united airway is provided in evaluations of the correlation between asthma, airway infections, and nasal disease, and the

TABLE 3 GO enrichment analysis of genes in coexpression modules

Module	Term ID	Term ontology	nmod genes	P value	Term name
Blue	GO:0005885	CC	3	0.000397	Arp2/3 protein complex
Blue	GO:0098562	CC	9	0.00061	Cytoplasmic side of membrane
Blue	GO:0042470	CC	7	0.000676	Melanosome
Blue	GO:0048770	CC	7	0.000676	Pigment granule
Blue	GO:0098791	CC	12	0.000721	Golgi subcompartment
Blue	GO:0032588	CC	6	0.00084	Trans-golgi network membrane
Blue	GO:0005925	CC	14	0.001055	Focal adhesion
Blue	GO:0005924	CC	14	0.001135	Cell-substrate adherens junction
Blue	GO:0031984	CC	12	0.00118	Organelle subcompartment
Blue	GO:0030055	CC	14	0.00125	Cell-substrate junction
Magenta	GO:0035162	BP	4	6.38E-07	Embryonic hemopoiesis
Magenta	GO:0045576	BP	4	2.37E-05	Mast cell activation
Magenta	GO:2000116	BP	6	6.10E-05	Regulation of cysteine-type endopeptidase activity
Magenta	GO:0005539	MF	6	6.36E-05	Glycosaminoglycan binding
Magenta	GO:0002444	BP	4	8.88E-05	Myeloid leukocyte mediated immunity
Magenta	GO:0008217	BP	5	0.000175	Regulation of blood pressure
Magenta	GO:0004713	MF	5	0.000303	Protein tyrosine kinase activity
Magenta	GO:0007626	BP	5	0.000305	Locomotory behavior
Magenta	GO:0002279	BP	3	0.000366	Mast cell activation involved in immune response
Magenta	GO:0043303	BP	3	0.000366	Mast cell degranulation
Pink	GO:0002250	BP	11	6.46E-07	Adaptive immune response
Pink	GO:0045576	BP	5	3.65E-06	Mast cell activation
Pink	GO:0002366	BP	8	4.71E-06	Leukocyte activation involved in immune response
Pink	GO:0002263	BP	8	5.04E-06	Cell activation involved in immune response
Pink	GO:0045730	BP	4	7.81E-06	Respiratory burst
Pink	GO:0042113	BP	8	9.96E-06	B cell activation
Pink	GO:0001608	MF	3	2.34E-05	G-protein coupled nucleotide receptor activity
Pink	GO:0002521	BP	10	3.07E-05	Leukocyte differentiation
Pink	GO:0002279	BP	4	4.24E-05	Mast cell activation involved in immune response
Pink	GO:0043303	BP	4	4.24E-05	Mast cell degranulation
Yellow	GO:0031594	CC	5	0.000118	Neuromuscular junction
Yellow	GO:0005765	CC	10	0.000232	Lysosomal membrane
Yellow	GO:0098852	CC	10	0.000232	Lytic vacuole membrane
Yellow	GO:0005385	MF	4	4.65E-05	Zinc ion transmembrane transporter activity
Yellow	GO:0072509	MF	8	0.000186	Divalent inorganic cation transmembrane transporter activity
Yellow	GO:0050839	MF	13	0.000332	Cell adhesion molecule binding
Yellow	GO:0046915	MF	4	0.000463	Transition metal ion transmembrane transporter activity
Yellow	GO:0045296	MF	10	0.000541	Cadherin binding

Abbreviation: BP, biological process; CC, cellular component; GO, Gene ontology; MF, molecular function.
GO enrichment analysis of genes in coexpression modules;

likelihood of occupational asthma development in subjects with occupational nasal symptoms.²⁷

Accordingly, we speculate that assessing gene expression in the upper airway may provide potential biomarkers for the prognostic study of lower airway asthma. Using gene expression data of nasal epithelial brushings

from participants with ($n=40$) and without ($n=17$) allergic asthma, we constructed a total of 13 coexpression modules of 2804 genes using a WGCNA method. Eight modules were identified as significantly correlated to asthma status; four of these were positively correlated with asthma traits and GO enrichment analysis showed

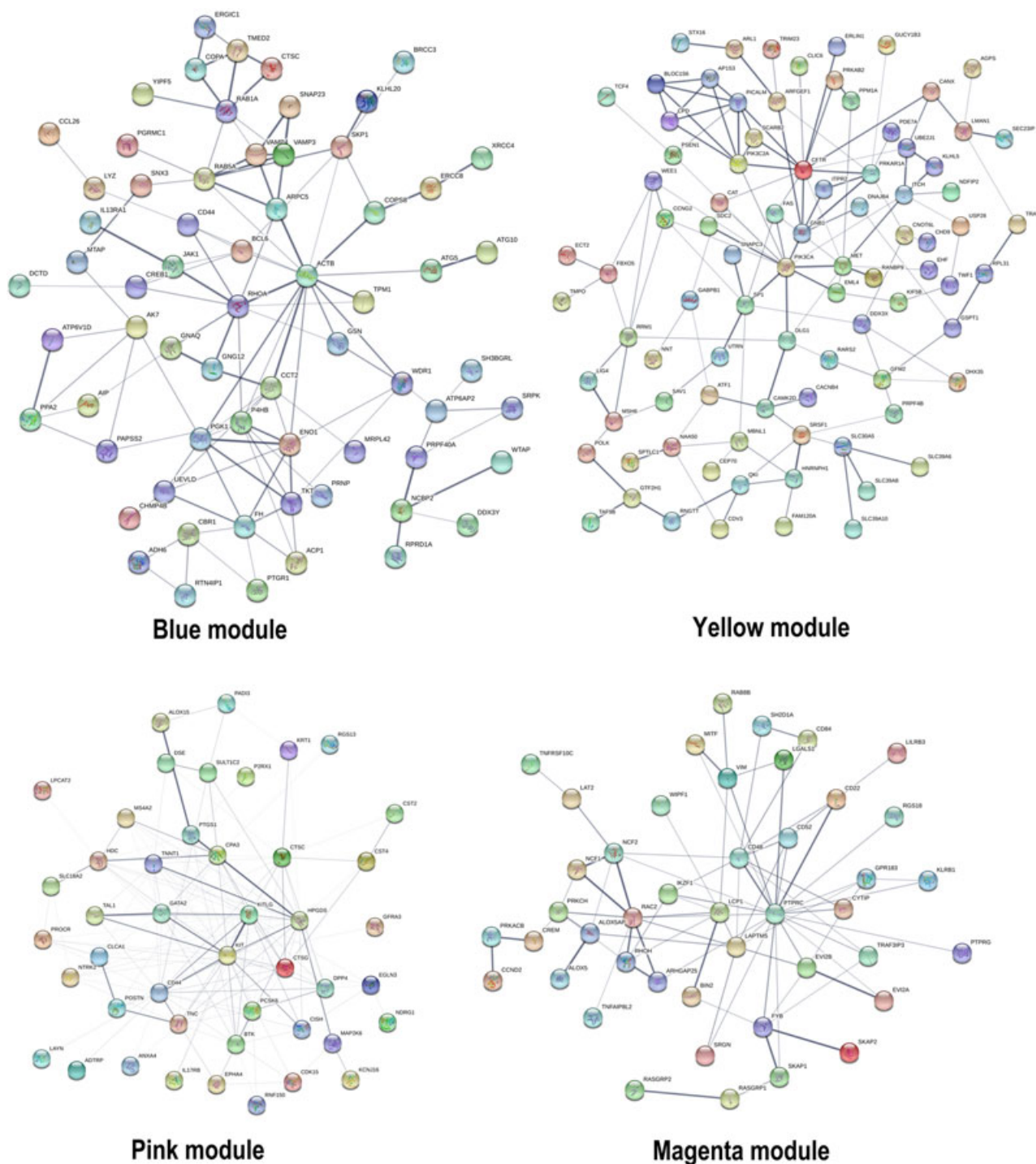


FIGURE 11 The visualization of different coexpression modules: blue module is arranged in the upper left; yellow, pink and magenta modules are arranged in the upper right, under left, and under right, respectively

these to be mainly enriched in mast cell activation, mast cell degranulation, immune cell response, transmembrane signal pathways, and substructure components of the cells. The most central hub genes of these four modules were *KIT*, *KITLG*, *GATA2*, *CD44*, *PTPRC*, and *CFTR*, which may directly or indirectly regulate this asthma-associated gene expression network and may

potentially serve as surrogate biomarkers for the detection and treatment of allergic asthmas.

Our results are somewhat consistent with those of previous research. Several reports have identified a typical Th2 feature in allergic asthma, characterized by elevated Th2-type inflammation in the lower airway assessed through gene expression data from bronchial brushings.^{14,28}

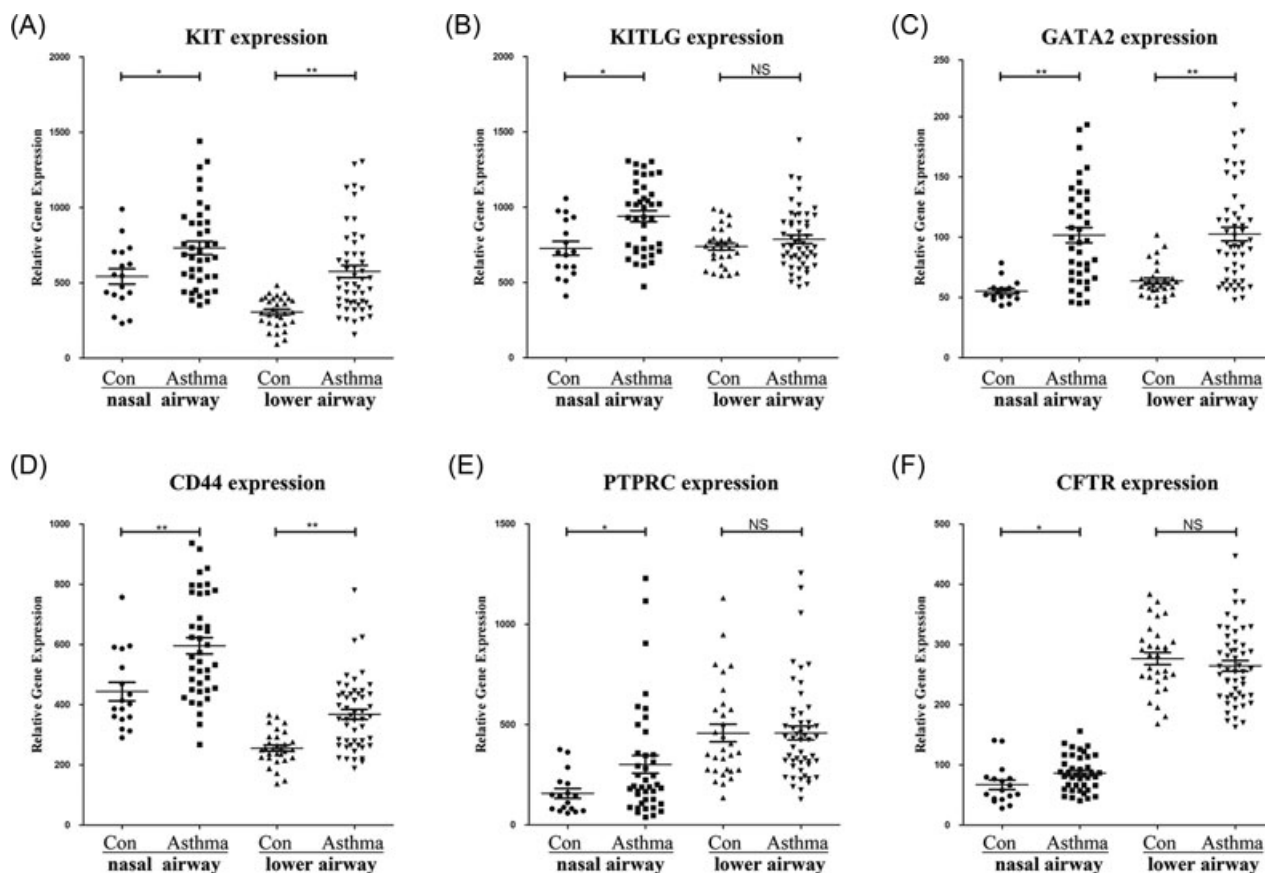


FIGURE 12 The gene expression level of six hub genes in upper airway epithelial cells and lower airway epithelial cells

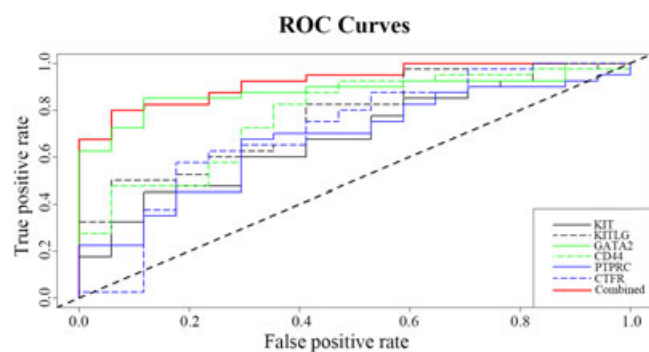


FIGURE 13 Receiver operating characteristics curve analyses of the sensitivity and the specificity of the six hub genes for the diagnosis of asthma

Interestingly, we also detected two novel modules positively correlated to asthma severity: Magenta ($\text{cor} = 0.49$, $P = 1\text{e} - 04$) and Pink ($\text{cor} = 0.41$, $P = 2\text{e} - 3$). The magenta module contained 90 genes encoding proteins that relate to mast cell activation and mast cell degranulation; these included *KIT*, *NDRG1*, *BTK*, and *GATA2*. The pink module contained 111 genes encoding proteins that are involved in the adaptive immune response, including B and T cell

activity and activation of other leukocytes; these included *PTPRC*, *SH2D1A*, *RAC2*, and *LCP1*. The reasons for the distinct expression profiles between the lower and upper airways are unclear but possible explanations include the proportional difference between the lower and upper airways. For example, the epithelium in the upper airways is pseudostratified, prismatic, and ciliated, while in the lower airways, the major cell types are ciliated, undifferentiated columnar, secretory, and basal, as well as a set of less common types such as mucous glands and neuroendocrine cells.^{29,30}

The four positively correlated modules were further analyzed using a GO enrichment method. The results show the significant distinction between the interactions among different modules, mainly related to their different functions. Genes in magenta were mainly enriched in pathways relating to mast cell activation and mast cell degranulation. The pink module was found to be mainly enriched in adaptive immune response. Yellow module genes were mainly enriched in transmembrane signal pathways, and those in blue were found to be mainly enriched in substructure components of the cells. Since the pathogenetic mechanism of allergic asthma is considered to relate

TABLE 4 AUC ROC values of different hub genes in asthmatics

Hub genes	AUC ROC	P value	Cut-off	Sensitivity	Specificity
<i>KIT</i>	0.691	0.023	0.619	41.2	77.5
<i>KITLG</i>	0.768	0.002	0.573	52.9	82.5
<i>GATA2</i>	0.881	6.25e−06	0.465	64.7	87.5
<i>CD44</i>	0.779	9.19e−04	0.581	58.8	85.0
<i>PTPRC</i>	0.676	0.036	0.600	41.2	77.5
<i>CFTR</i>	0.709	0.013	0.611	47.1	80.0
Combined	0.924	5.07e−07	0.477	70.6	90.0

Abbreviation: AUC, area under the curve; ROC, receiver operating characteristic.

to an imbalance of Th1/Th2 and abnormal mast cell activation,^{31,32} we regard the magenta and pink modules observed here as the most important in allergic asthma pathogenesis.

The enriched pathways were further visualized by constructing a protein interaction network and the hub genes *KIT*, *KITLG*, *GATA2*, *CD44*, *PTPRC*, and *CFTR* were identified. The hub genes in the magenta module were *KIT* (CD117; stem cell growth factor receptor) and *KITLG* which are indispensable for the growth, differentiation, survival, and homing of mast cells.³³ The point mutation in *KIT* or *KITLG* leads to a deficiency in mast cells in mice; while the activating mutation in *KIT* of aspartic acid is replaced by valine at residue 816 always leads to mast cell hyperplasia in bone marrow and peripheral tissue in both adults and children.^{33–35} *GATA2*, a zinc finger transcription factor of the GATA family, is critical in driving the differentiation of prebasophil and mast cell progenitors into basophils and mast cells, respectively, and also plays a role in the maintenance of basophil and mast cell repertoires.^{36,37} The hub gene *CD44*, an adhesion molecule, has been associated with asthma and plays a key role in mediating the adhesion of mast cells to bronchial smooth muscle.^{38,39}

Further, the hub gene in the pink module was *PTPRC*. The protein encoded by this gene is a tyrosine phosphatase receptor that is involved in autoimmune disorders.⁴⁰ Recently, it has been reported that *PTPRC* is a biomarker associated with disease severity in asthma.⁴¹ The yellow hub gene was *CFTR* which encodes the cystic fibrosis transmembrane conductance regulator; mutation on the copies of the *CFTR* gene has been known to lead to cystic fibrosis.⁴² Several studies have delivered interesting results about asthma development and *CFTR* gene variation, with some research showing *CFTR* as positively related to asthma and others showing either protective or no correlation.⁴³

The hub genes are mainly involved in mast cell activation and asthma development and we, therefore, speculate that they are potential surrogate biomarkers in the upper airway for asthma.

5 | CONCLUSIONS

In summary, we have constructed a framework of four coexpression modules that can be used for the prognostic study of allergic asthma using data from nasal epithelial cells. In addition, we have identified the hub genes *KIT*, *KITLG*, *GATA2*, *CD44*, *PTPRC*, and *CFTR* as potential prognostic biomarkers for asthma. Our study provides a basis for the future development of clinical strategies against allergic asthma.

ORCID

Zhaoyu Liu  <http://orcid.org/0000-0002-0142-9883>

REFERENCES

- Loftus PA, Wise SK. Epidemiology of asthma. *Curr Opin Otolaryngol Head Neck Surg*. 2016;24:245–249.
- Fahy JV. Type 2 inflammation in asthma—present in most, absent in many. *Nat Rev Immunol*. 2015;15:57–65.
- Braunstaal GJ, Fokkens WJ, Overbeek SE, KleinJan A, Hoogsteden HC, Prins JB. Mucosal and systemic inflammatory changes in allergic rhinitis and asthma: a comparison between upper and lower airways. *Clin Exp Allergy*. 2003;33:579–587.
- Chapman DG, Tully JE, Nolin JD, Janssen-Heininger YM, Irvin CG. Animal models of allergic airways disease: where are we and where to next? *J Cell Biochem*. 2014;115:2055–2064.
- Calhoun KH. Asthma treatments: new and emerging therapies. *Int Forum Allergy Rhinol*. 2015;5(Suppl 1):S76–S81.
- Croissant S. Epidemiology of asthma: prevalence and burden of disease. *Adv Exp Med Biol*. 2014;795:17–29.
- Hammad H, Lambrecht BN. Barrier epithelial cells and the control of type 2. *Immunity*. 2015;43:29–40.
- Hirose K, Iwata A, Tamachi T, Nakajima H. Allergic airway inflammation: key players beyond the Th2 cell pathway. *Immunol Rev*. 2017;278:145–161.
- Modena BD, Tedrow JR, Milosevic J, et al. Gene expression in relation to exhaled nitric oxide identifies novel asthma phenotypes with unique biomolecular pathways. *Am J Respir Crit Care Med*. 2014;190:1363–1372.

10. G. woodruff P. Subtypes of asthma defined by epithelial cell expression of messenger RNA and microRNA. *Ann Am Thorac Soc.* 2013;10(Suppl.):S186-S189.
11. Woodruff PG, Boushey HA, Dolganov GM, et al. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci USA.* 2007;104:15858-15863.
12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
13. Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp Eye Res.* 2018;166:13-20.
14. Modena BD, Bleecker ER, Busse WW, et al. Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease. *Am J Respir Crit Care Med.* 2017;195:1449-1463.
15. Morrow JD, Qiu W, Chhabra D, et al. Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. *BMC Med Genomics.* 2015;8:1.
16. Yi G, Liang M, Li M, et al. A large lung gene expression study identifying IL1B as a novel player in airway inflammation in COPD airway epithelial cells. *Inflamm Res.* 2018;67:539-551.
17. Feng CH, Miller MD, Simon RA. The united allergic airway: connections between allergic rhinitis, asthma, and chronic sinusitis. *Am J Rhinol Allergy.* 2012;26:187-190.
18. Bourdin A, Gras D, Vachier I, Chanez P. Upper airway x 1: allergic rhinitis and asthma: united disease through epithelial cells. *Thorax.* 2009;64:999-1004.
19. Fasano MB. Combined airways: impact of upper airway on lower airway. *Curr Opin Otolaryngol Head Neck Surg.* 2010;18:15-20.
20. Khan DA. Allergic rhinitis and asthma: epidemiology and common pathophysiology. *Allergy Asthma Proc.* 2014;35:357-361.
21. Burgel PR. United airway diseases. Should we add upper airway inflammatory disorders to the list of chronic obstructive pulmonary disease comorbidities. *Ann Am Thorac Soc.* 2015;12:968-970.
22. Cruz AA, Popov T, Pawankar R, et al. Common characteristics of upper and lower airways in rhinitis and asthma: ARIA update, in collaboration with GA(2)LEN. *Allergy.* 2007;62(Suppl 84):1-41.
23. Licari A, Castagnoli R, Denicolò CF, Rossini L, Marseglia A, Marseglia GL. The nose and the lung: united airway disease? *Front Pediatr.* 2017;5:44.
24. Kariyawasam HH, Rotiroti G. Allergic rhinitis, chronic rhinosinusitis and asthma: unravelling a complex relationship. *Curr Opin Otolaryngol Head Neck Surg.* 2013;21:79-86.
25. Rosati MG, Peters AT. Relationships among allergic rhinitis, asthma, and chronic rhinosinusitis. *Am J Rhinol Allergy.* 2016;30:44-47.
26. Rönmark EP, Ekerljung L, Mincheva R, et al. Different risk factor patterns for adult asthma, rhinitis and eczema: results from West Sweden Asthma Study. *Clin Transl Allergy.* 2016;6:28.
27. Moscato G, Siracusa A. Rhinitis guidelines and implications for occupational rhinitis. *Curr Opin Allergy Clin Immunol.* 2009;9:110-115.
28. Bhakta NR, Christenson SA, Nerella S, et al. IFN-stimulated gene expression, type 2 inflammation, and endoplasmic reticulum stress in asthma. *Am J Respir Crit Care Med.* 2018;197:313-324.
29. Minnich DJ, Mathisen DJ. Anatomy of the trachea, carina, and bronchi. *Thorac Surg Clin.* 2007;17:571-585.
30. Scudamore CL, McInnes EF. Histopathological analysis of the respiratory tract. *Curr Protoc Mouse Biol.* 2014;4:181-196.
31. Endo Y, Hirahara K, Yagi R, Tumes DJ, Nakayama T. Pathogenic memory type Th2 cells in allergic inflammation. *Trends Immunol.* 2014;35:69-78.
32. Foster PS, Maltby S, Rosenberg HF, et al. Modeling TH2 responses and airway inflammation to understand fundamental mechanisms regulating the pathogenesis of asthma. *Immunol Rev.* 2017;278:20-40.
33. Dahlin JS, Hallgren J. Mast cell progenitors: origin, development and migration to tissues. *Mol Immunol.* 2015;63:9-17.
34. Arock M, Sotlar K, Akin C, et al. KIT mutation analysis in mast cell neoplasms: recommendations of the european competence network on mastocytosis. *Leukemia.* 2015;29:1223-1232.
35. Grimbaldston MA, Chen CC, Piliponsky AM, Tsai M, Tam SY, Galli SJ. Mast cell-deficient W-sash c-kit mutant Kit W-sh/W-sh mice as a model for investigating mast cell biology in vivo. *Am J Pathol.* 2005;167:835-48.
36. Li Y, Qi X, Liu B, Huang H. The STAT5-GATA2 pathway is critical in basophil and mast cell differentiation and maintenance. *J Immunol.* 2015;194:4328-4338.
37. Ohmori S, Moriguchi T, Noguchi Y, et al. GATA2 is critical for the maintenance of cellular identity in differentiated mast cells derived from mouse bone marrow. *Blood.* 2015;125:3306-3315.
38. Girodet PO, Ozier A, Triantafyllidis T, et al. Mast cell adhesion to bronchial smooth muscle in asthma specifically depends on CD51 and CD44 variant 6. *Allergy.* 2010;65:1004-1012.
39. Rothenberg ME. CD44--a sticky target for asthma. *J Clin Invest.* 2003;111:1460-1462.
40. Tchilian EZ, Beverley PCL. Altered CD45 expression and disease. *Trends Immunol.* 2006;27:146-153.
41. Baos S, Calzada D, Cremades L, et al. Biomarkers associated with disease severity in allergic and nonallergic asthma. *Mol Immunol.* 2017;82:34-45.
42. Guggino WB, Stanton BA. New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nat Rev Mol Cell Biol.* 2006;7:426-436.
43. Maurya N, Awasthi S, Dixit P. Association of CFTR gene mutation with bronchial asthma. *Indian J Med Res.* 2012;135:469-478.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Liu Z, Li M, Fang X, et al. Identification of surrogate prognostic biomarkers for allergic asthma in nasal epithelial brushing samples by WGCNA. *J Cell Biochem.* 2018;1-14. <https://doi.org/10.1002/jcb.27790>