

Identification of a disease-specific gene expression profile of children with acute asthma by weighted gene co-expression network analysis

Yan Luo*, Jing Wang, Wei Lu, Yang Liu, Yun Huang and Dichun Luo

Department of Preventive Health Care, Affiliated Hospital of Zunyi Medical University,
Zunyi, Guizhou 563003, China

(Received 18 June 2020, accepted 17 September 2020; J-STAGE Advance published date: 28 February 2021)

Asthma is one of the most common diseases, with a high prevalence among children. To date, systemic co-expression analysis for this disease has not been undertaken to explain its pathogenesis. Here we identified differentially expressed genes (DEGs) in 87 samples, and then constructed co-expression modules via weighted gene co-expression network analysis (WGCNA) and investigated the functional enrichment of co-expressed genes in terms of Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG). Meanwhile, protein–protein interaction (PPI) network and miRNA–transcription factor–target (miRNA–TF–target) regulatory network analyses were performed to screen hub genes. As a result, 3,469 DEGs were identified in this study, of which 1,860 genes were up-regulated and 1,609 genes were down-regulated. Using WGCNA, we identified two key modules, named MEbrown and MEblue, that may play important roles in asthma. Functional enrichment analysis revealed that MEbrown was enriched in 37 KEGG pathways and 472 biological processes (BPs), while MEblue was enriched in 16 KEGG pathways and 449 BPs. From PPI and miRNA–TF–target regulatory network analysis, a total of 31 TFs, seven miRNAs and 28 nodes were identified. Our findings should provide a framework of therapeutic targets for treating children with acute asthma.

Key words: asthma, children, hub gene, modules, network

INTRODUCTION

Asthma is a common chronic disease, globally affecting about 300 million people of all ages, and its prevalence is still rising (Masoli et al., 2004). Pediatric asthma is especially concerning, with up to 9.5% of children having been affected (Akinbami et al., 2016). Asthma in children places a heavy burden not only on the children's families, but also on the health care system of the whole society (Asher and Pearce, 2014).

Pediatric asthma is usually classified into four stages:

acute onset, chronic persistent, clinical remission and other (Hoch et al., 2019). The acute onset stage is also known as the “attack” period. Children with acute asthma exhibit increasing chest tightness, shortness of breath, coughing and/or wheezing, along with dyspnea. They may even experience respiratory failure, putting their lives at risk (Pavord et al., 2018). Thus, identifying robust biomarkers that can aid asthma diagnosis or enable better prognostic prediction may help to improve prognosis and children's quality of life.

With the development of bioinformatic approaches, systematic description and high-throughput data analysis are now widely used to provide important clues for molecular mechanism studies (Nam et al., 2009). Moreover, bioinformatic analysis can also reveal a systematic biological process that traverses different functional networks in different disease models (Chasman et al., 2016). In asthma-related research, however, only a few studies have investigated gene expression profiles. For example, Katayama et al. (2020) have identified a unique acute wheeze-specific gene module in blood from preschool wheezers. This module was associated with vitamin D

Edited by Hie Lim Kim

* Corresponding author. qasdz09YU@163.com

DOI: <https://doi.org/10.1266/ggs.20-00031>



Copyright: ©2020 The Author(s). This is an open access article distributed under the terms of the Creative Commons BY 4.0 International (Attribution) License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits the unrestricted distribution, reproduction and use of the article provided the original source and authors are credited.

levels and asthma medication. Another study from Aoki et al. (2009) showed expression profiles of genes related to asthma exacerbations. There were 137/16 genes that were significantly up-/down-regulated during an asthma exacerbation, of which 62 were also differentially expressed during upper respiratory infection. However, little is known about the gene expression profile of children with acute asthma. Here, we used a novel tool, weighted gene co-expression network analysis (WGCNA), to explore the molecular interaction mechanism and correlation networks of this disease.

As an effective bioinformatics method, WGCNA is able to clarify synergetic expressed modules at the transcriptome level. More importantly, it can be used to identify relationships between gene networks (Langfelder and Horvath, 2008). WGCNA has been successfully applied in different disease models to generate correlation networks, further to identifying candidate biomarkers or therapeutic targets (Pei et al., 2017). In the present study, based on datasets from the Gene Expression Omnibus (GEO), we used WGCNA to analyze differentially expressed genes (DEGs) from 87 samples. Key gene modules associated with DEGs were then identified. Further, the biological functions and pathways of genes in the key modules were analyzed. These informative genes identified in our study offer new insights into the pathogenesis of asthma; moreover, our findings may be beneficial to clinical treatment of children with acute asthma.

MATERIALS AND METHODS

Data collection and preprocessing Datasets related to childhood asthma were obtained from the NCBI GEO (<http://www.ncbi-nlm-nih-gov.ezproxy.lb.polyu.edu.hk/geo>) with accession number GSE103166. GSE103166 consists of 87 samples, comprising 56 samples of children with acute asthma/wheeze and 31 control samples. The platform used was GPL23961 [HuGene-2_1-st] Affymetrix Human Gene 2.1 ST Array [HuGene21st_Hs_ENTREZG_20.0.0]. The GEO database download has been standardized. The limma package in R (<http://www.bioconductor.org/packages/release/bioc/html/limma.html>) was used to identify DEGs (children with asthma versus control) from the candidate genes. $P < 0.05$ was the cut-off criterion.

Screening children with acute asthma-related modules based on WGCNA WGCNA (Langfelder and Horvath, 2008) is a typical biological algorithm for constructing gene co-expression networks (Bacchetti and Leung, 2002; Yin et al., 2018). We used WGCNA to analyze the expression values of the DEGs obtained in the previous screening in each group, and screened the modules and genes associated with acute asthma in children.

First, the *trait-based node significance measure* was calculated:

$$S_{mn} = \left| \text{cor}_{(m,n)} \right|$$

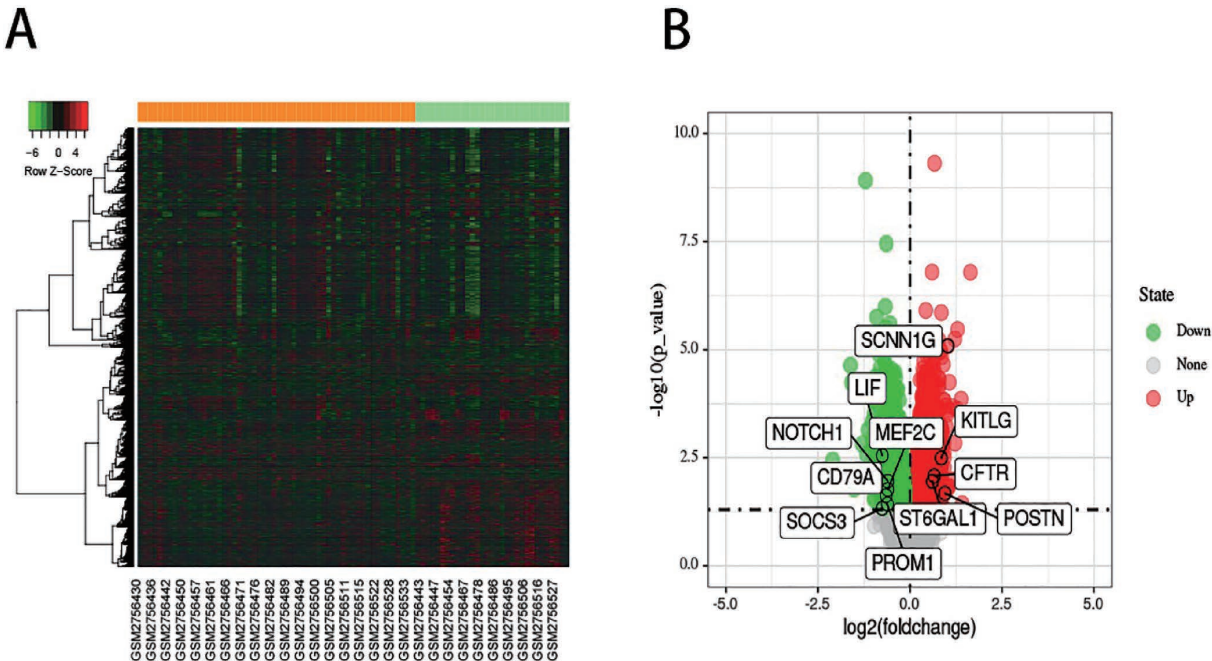


Fig. 1. Gene expression heatmap (A) and volcano plot (B) of children with acute asthma. Red represents up-regulated genes and green represents down-regulated genes compared with control.

where m and n are the expressed genes m and n , respectively. cor is the Pearson coefficient of these two vectors. This transition improved the reliability of the co-expression network by giving more weight to the strong connections and reducing the importance of the weak connections in the predicted co-expression network.

Next, a *power function* was applied to correlate the adjacency of genes, where β is soft-thresholding powers:

$$a_{mn} = \text{power}_{(S_{mn}, \beta)}$$

The adjacency matrix was then converted to a topological matrix using the following formula:

$$w_{mn} = \frac{l_{mn} + a_{mn}}{\min\{k_m, k_n\} 3333333334e + 1 - a_{mn}}$$

l_{mn} represents the sum of the adjacency coefficient area of the nodes connected to genes m and n , and k_m is the sum of the adjacency numbers of the nodes individually connected to gene m . The topological properties were also confirmed. Module identification was then accomplished with the dynamic tree cut method. Modules with high similarities were identified and merged together with a height cut-off of 0.95. Furthermore, P -values of gene expression difference between the asthma and control groups were evaluated with Student's t -test, and log-transformed P -values were used to determine statistical significance. The mean value of gene significance for all genes in a particular module was defined as module significance.

Functional enrichment analysis Gene Ontology (GO)

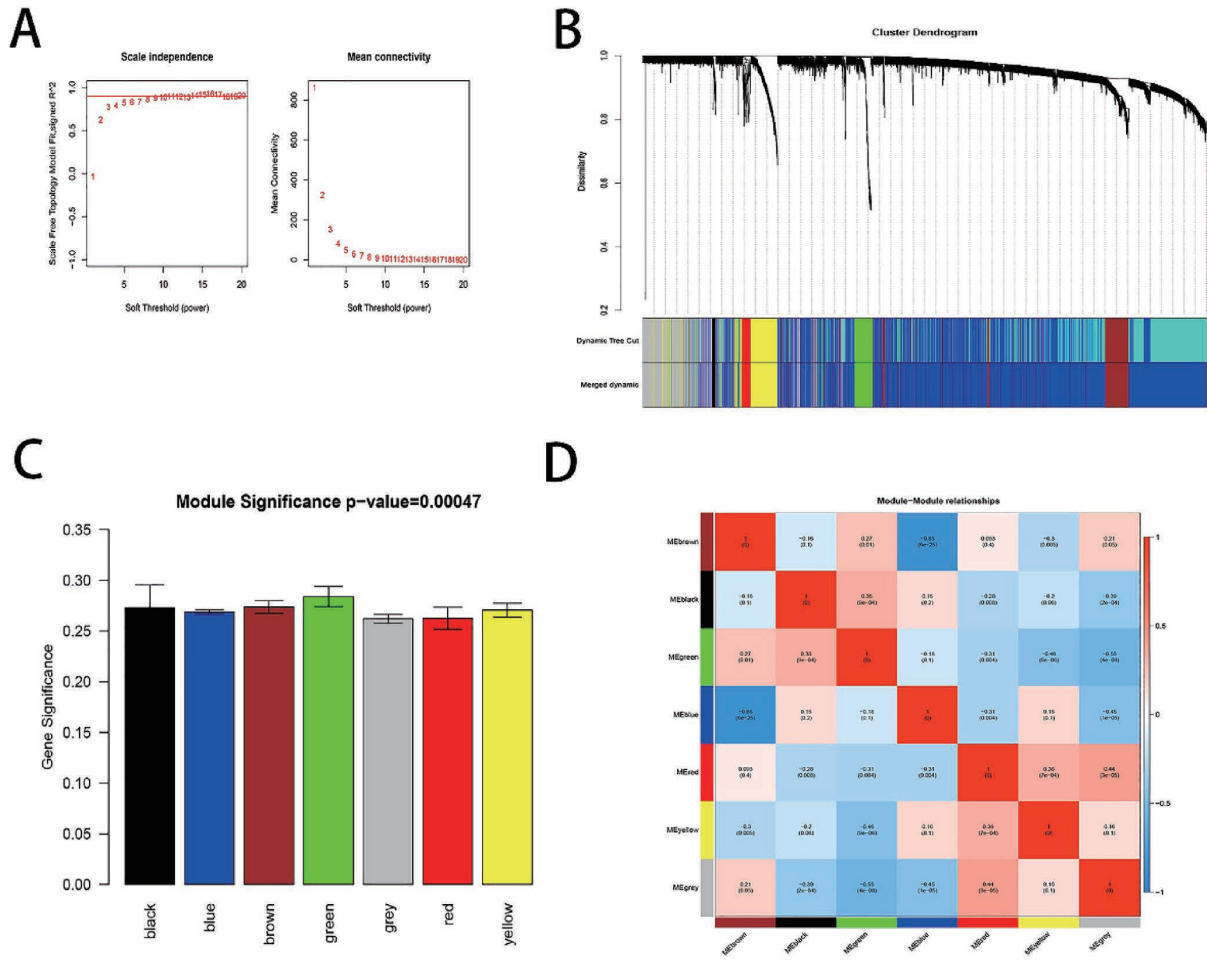


Fig. 2. WGCNA analysis based on DEGs and identification of key modules. (A) Analysis of the scale-free fit index for various soft-thresholding powers (β). The left panel displays the influence of soft-thresholding power (x-axis) on scale-free fit index (y-axis). The right panel shows the influence of soft-thresholding power (x-axis) on mean connectivity (degree, y-axis). (B) Cluster dendrogram of genes, with dissimilarity based on topological overlap, together with assigned module colors. Dynamic tree cutting was applied to identify modules by dividing the dendrogram at significant branch points. Modules are displayed with different colors in the horizontal bars below the dendrogram, with gray representing unassigned genes. (C) Diagram of correlation of module colors with asthma. Colors represent modules and the y-axis represents gene significance. (D) Heatmap plot of the adjacencies in the hub gene network. In the heatmap, red represents high adjacency (positive correlation), while blue represents low adjacency (negative correlation).

and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using the R package clusterProfiler for significant modules. The gene lists of modules were uploaded, and we obtained the results for biological process (BP) and KEGG pathways. $P < 0.05$ was regarded as significant.

Protein-protein interaction (PPI) network analysis To construct a PPI network of DEGs, we first used the STRING (Szklarczyk et al., 2015) (version: 10.0, <http://www.string-db.org/>) database to predict and analyze the interactions between proteins encoded by DEGs. Cytoscape (Shannon et al., 2003) (version: 3.2.0, <http://www.cytoscape.org/>) software was then used to conduct PPI network analysis.

MicroRNA-transcription factor-target (miRNA-TF-target) co-regulation network analysis In the PPI network, we used the overrepresentation enrichment analysis (ORA) enrichment method in WebGestalt (<http://>

Table 1. Correlation-values and P -values of different modules

Module	Correlation	P -value
MEbrown	0.36	5.19E-04
MEblack	-0.36	6.71E-04
MEgreen	-0.44	1.91E-05
MEblue	-0.41	7.12E-05
MEred	0.36	6.41E-04
MEyellow	0.39	1.88E-04
MEgray	0.62	1.92E-10

Data for key modules are shown in bold type.

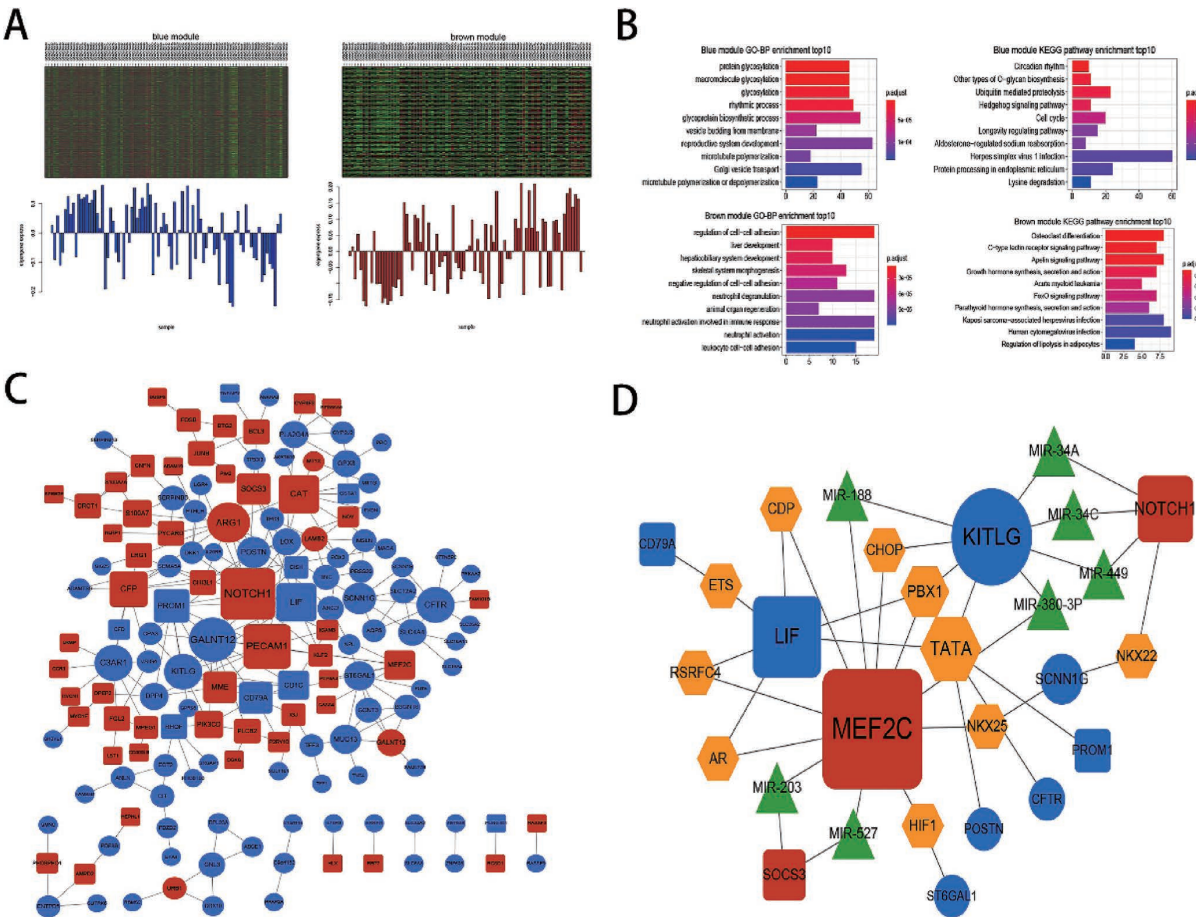


Fig. 3. Functional enrichment analysis, PPI network analysis and miRNA-TF-target regulatory network. (A) Heatmap plot of gene expression in MEblue (left panel) and MEbrown (right panel). Red and green represent up- and down-regulated genes, respectively, compared with control. (B) Enrichment results for MEblue (upper panels) and MEbrown module genes (lower panels) for BP-top10 (bottom left panels) and KEGG-top10 (right panels). (C) PPI network analysis for MEbrown and MEblue. Circles and squares represent up- and down-regulated genes, respectively, while brown and blue mark genes in MEbrown and MEblue. Node size is based on the PPI degree value. (D) miRNA-TF-target regulatory network. Circles and squares indicate up- and down-regulated genes, respectively. Brown and blue indicate genes in MEbrown and MEblue. Yellow hexagons indicate TFs and green triangles indicate miRNAs incorporated into this network. Node size is based on the degree value.

www.webgestalt.org/) to perform miRNA–TF–target enrichment prediction for genes with degree values > 5. The miRNA–TF–target co-regulation network was constructed by Cytoscape (Shannon et al., 2003).

RESULTS

Identification of DEGs between children with acute asthma and healthy samples To identify DEGs, gene microarray data of 87 samples were obtained from the GEO database (<http://www.ncbi-nlm-nih-gov.ezproxy.lb.polyu.edu.hk/geo>). After standardization and

gene annotation, the gene expression distribution of each sample was depicted as S1. A total of 3,469 DEGs were identified, among which 1,860 genes were up-regulated and 1,609 genes were down-regulated in the asthma group. A gene expression heatmap and volcano plot are shown in Fig. 1A and Fig. 1B, respectively.

Identification of key modules based on WGCNA To clarify asthma-related key modules, WGCNA was performed on the 3,469 DEGs. Prior to WGCNA, in order to have relatively balanced scale independence and mean connectivity of the WGCNA, we analyzed network topol-

Table 2. KEGG pathway and GO enrichment analysis of MEblue and MEbrown

Module	ID	Description	Count	P-value
KEGG	hsa04710	Circadian rhythm	10	3.97E-04
	hsa00514	Other types of O-glycan biosynthesis	11	3.21E-03
	hsa04120	Ubiquitin-mediated proteolysis	23	4.44E-03
	hsa04340	Hedgehog signaling pathway	11	6.34E-03
	hsa04110	Cell cycle	20	1.28E-02
	hsa04211	Longevity regulating pathway	15	1.99E-02
	hsa04960	Aldosterone-regulated sodium reabsorption	8	2.07E-02
	hsa05168	Herpes simplex virus 1 infection	60	2.40E-02
	hsa04141	Protein processing in endoplasmic reticulum	24	2.45E-02
	hsa00310	Lysine degradation	11	2.73E-02
Blue	GO:0006486	Protein glycosylation	46	5.47E-06
	GO:0043413	Macromolecule glycosylation	46	5.47E-06
	GO:0070085	Glycosylation	46	1.68E-05
	GO:0048511	Rhythmic process	49	2.02E-05
	GO:0009101	Glycoprotein biosynthetic process	54	5.46E-05
	GO:0006900	Vesicle budding from membrane	22	9.02E-05
	GO:0061458	Reproductive system development	63	1.04E-04
	GO:0046785	Microtubule polymerization	18	1.09E-04
	GO:0048193	Golgi vesicle transport	55	1.30E-04
	GO:0031109	Microtubule polymerization or depolymerization	23	1.39E-04
KEGG	hsa04640	Hematopoietic cell lineage	3	1.30E-03
	hsa05146	Amoebiasis	2	2.15E-02
	hsa04668	TNF signaling pathway	2	2.56E-02
	hsa05418	Fluid shear stress and atherosclerosis	2	3.81E-02
Brown	GO:0072073	Kidney epithelium development	6	5.68E-09
	GO:0001822	Kidney development	7	9.89E-09
	GO:0072001	Renal system development	7	1.42E-08
	GO:0001655	Urogenital system development	7	3.22E-08
	GO:0072009	Nephron epithelium development	5	8.94E-08
	GO:0035850	Epithelial cell differentiation involved in kidney development	4	1.26E-07
	GO:0061005	Cell differentiation involved in kidney Development	4	2.93E-07
	GO:0072006	Nephron development	5	3.35E-07
	GO:0072160	Nephron tubule epithelial cell differentiation	3	4.74E-07
	GO:0032835	Glomerulus development	4	5.13E-07

ogy for various soft-thresholding powers. In Fig. 2A, we took power as 8 when the correlation index reached 0.90. The hierarchical cluster dendrogram of the 3,469 genes was then analyzed (Fig. 2B). We set the MEDis-sThres as 0.25 to merge similar modules, and seven modules were generated (Fig. 2C), which are shown in Table 1. Interaction-based relationships for the seven modules are displayed as a heatmap in Fig. 2D, and the results revealed that there is a high level of independence among the modules.

Taken together, the brown module (MEbrown) and blue module (MEblue) were considered to be the key modules. Gene significance versus module membership plots of these two models are shown in Supplementary Fig. S1.

Functional enrichment analysis of genes in the MEbrown and MEblue modules For functional enrichment analysis, we used KEGG pathway enrichment analysis and GO-BP methods separately. First, the genes included in MEbrown and MEblue modules were displayed in a heatmap plot (Fig. 3A). A total of 326 and 2,170 genes were identified in MEbrown and MEblue modules, respectively. In enrichment analysis, MEbrown was enriched in 37 KEGG pathways and 472 BPs, while MEblue was enriched in 16 KEGG pathways and 449 BPs. The results are shown in Fig. 3B, with detailed information provided in Table 2.

Table 3. Top 20 nodes based on PPI-degree

Node	Description	Degree
NOTCH1	down-brown	16
GNL3	up-blue	13
PECAM1	down-brown	13
ARG1	up-brown	11
CAT	down-brown	10
LIF	down-blue	10
KITLG	up-blue	9
C3AR1	up-blue	9
CFP	down-brown	9
CFTR	up-blue	9
PROM1	down-blue	8
SCNN1G	up-blue	7
MME	down-brown	7
SOCS3	down-brown	7
CD79A	down-blue	7
POSTN	up-blue	7
MUC13	up-blue	6
ST6GAL1	up-blue	6
MEF2C	down-brown	6
CD1C	down-blue	6

PPI network analysis of MEbrown and MEblue modules Next, PPI network analysis was used to predict and analyze the interactions between proteins encoded by DEGs in MEbrown and MEblue. As shown in Fig. 3C, there were 151 nodes and 231 interaction pairs in the whole network. The top 20 nodes of this network are listed in Table 3.

miRNA-TF-target regulatory network To determine whether miRNAs and TFs are responsible for the observed altered gene expression in the MEbrown and MEblue modules, the miRNA-TF-target regulatory network was obtained. As shown in Fig. 3D, a total of 31 TFs, seven miRNAs and 28 nodes were identified. Moreover, there were 38 pairs of regulatory relationships in this network, of which five were up-regulated (round shape, including KITLG, SCNN1G, CFTR, POSTN and ST6GAL1) while six were down-regulated (square shape, including NOTCH1, CD79A, MEF2C, PROM1, LIF and SOCS3).

DISCUSSION

Acute asthma in children is considered to be a disease with a high prevalence. To achieve early diagnosis and treatment of pediatric asthma, more knowledge about the disease mechanism and gene expression profile is urgently needed. In the present study, by searching the GEO database, we have identified 3,469 DEGs, of which 1,860 were up-regulated and 1,609 were down-regulated (Supplementary Fig. S2). We then applied WGCNA to identify the key modules and regulators of gene expression among DEGs.

Most existing bioinformatic tools focus only on unweighted networks, whereas WGCNA can be used for both weighted and unweighted correlation networks, and to further explore the module (cluster) structure in a network. Moreover, it also can be used to rank genes or modules in independent data sets (Langfelder and Horvath, 2008). So far, this approach has been widely used with different disease-related datasets. For example, in glioblastoma multiforme, Yang et al. (2018) performed WGCNA to analyze 1,913 DEGs between a glioblastoma group and a control group and found five modules that are mainly enriched in the inflammatory response. Based on WGCNA, Yin et al. (2018) identified key pathways and genes in different stages of hepatocellular carcinoma. In the modules they identified were genes enriched in energy metabolism, protein ubiquitination and ephrin receptor signaling pathways. In our study, the first step was to identify the key modules based on WGCNA. To this end, network topology analysis for various soft-thresholding powers (Fig. 2A) and dynamic tree-cutting analysis (Fig. 2B) were conducted. This yielded seven modules (Fig. 2D), among which two mod-

ules, the brown (MEbrown) and blue module (MEblue), were considered to be the key modules for further analysis (Fig. 2D). A total of 326 and 2,170 genes were included in MEbrown and MEblue modules, respectively (Fig. 3A).

To investigate the networks of these two key modules, functional enrichment analysis was performed using KEGG pathway enrichment analysis and GO-BP methods. MEbrown was enriched in 37 KEGG pathways and 472 BPs, while MEblue was enriched in 16 KEGG pathways and 449 BPs. Genes related to protein glycosylation, macromolecule glycosylation and circadian rhythm were most enriched in MEblue, while genes related to regulation of cell–cell adhesion, liver development, osteoclast differentiation and the C-type lectin receptor signaling pathway were more enriched in MEbrown (Fig. 3B). PPI network analysis was subsequently used to predict and analyze the interactions between proteins encoded by DEGs in MEbrown and MEblue. PPI networks are tools to understand cell functions, disease machinery and drug design/repositioning, because they can be used to map the relationships between species through conserved pathways and protein complexes (Athanasios et al., 2017; Miryala et al., 2018). In our study, we identified 151 nodes and 231 interaction pairs in the whole PPI network (Fig. 3C). Finally, based on previous analysis, a miRNA–TF–target regulatory network was generated (Fig. 3D), with 31 TFs, seven miRNAs and 28 nodes identified. Moreover, for the hub genes, five genes (KITLG, SCNN1G, CFTR, POSTN and ST6GAL1) were up-regulated, and six genes (NOTCH1, CD79A, MEF2C, PROM1, LIF and SOCS3) were down-regulated. A previously published study demonstrated that the Notch1–GATA3 signaling pathway is involved in inflammation in asthma (Chong et al., 2014). The authors showed that Notch1 and Notch2 receptors play an important role in regulating allergic airway inflammation. These receptors might act as therapeutic options in the treatment of allergic asthma.

In summary, our study used WGCNA network analysis to identify co-expression modules for genes of children with acute asthma. Two modules, named MEbrown and MEblue, were found to be highly enriched for genes involved in many pathways. Furthermore, we identified 11 hub genes that are potential biomarkers for diagnosis and treatment of children with acute asthma.

REFERENCES

- Akinbami, L. J., Simon, A. E., and Rossen, L. M. (2016) Changing trends in asthma prevalence among children. *Pediatrics* **137**, 1–7.
- Aoki, T., Matsumoto, Y., Hirata, K., Ochiai, K., Okada, M., Ichikawa, K., Shibasaki, M., Arinami, T., Sumazaki, R., and Noguchi, E. (2009) Expression profiling of genes related to asthma exacerbations. *Clin. Exp. Allergy* **39**, 213–221.
- Asher, I., and Pearce, N. (2014) Global burden of asthma among children. *Int. J. Tuberc Lung. Dis.* **18**, 1269–1278.
- Athanasios, A., Charalampos, V., Vasileios, T., and Ashraf, G. M. (2017) Protein-protein interaction (PPI) network: recent advances in drug discovery. *Curr. Drug Metab.* **18**, 5–10.
- Bacchetti, P., and Leung, J. M. (2002) Sample size calculations in clinical research. *Anesthesiology* **97**, 1028–1029; author reply 1029–1032.
- Chasman, D., Fotuhi Siahpirani, A., and Roy, S. (2016) Network-based approaches for analysis of complex biological systems. *Curr. Opin. Biotechnol.* **39**, 157–166.
- Chong, L., Zhang, W., Nie, Y., Yu, G., Liu, L., Lin, L., Wen, S., Zhu, L., and Li, C. (2014) Protective effect of curcumin on acute airway inflammation of allergic asthma in mice through Notch1-GATA3 signaling pathway. *Inflammation* **37**, 1476–1485.
- Hoch, H. E., Houin, P. R., and Stillwell, P. C. (2019) Asthma in children: a brief review for primary care providers. *Pediatr. Ann.* **48**, e103–e109.
- Katayama, S., Stenberg Hammar, K., Krjutskov, K., Einarsdottir, E., Hedlin, G., Kere, J., and Söderhäll, C. (2020) Acute wheeze-specific gene module shows correlation with vitamin D and asthma medication. *Eur. Respir. J.* **55**, 1901330.
- Langfelder, P., and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.
- Masoli, M., Fabian, D., Holt, S., and Beasley, R. (2004) The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy* **59**, 469–478.
- Miryala, S. K., Anbarasu, A., and Ramaiah, S. (2018) Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* **642**, 84–94.
- Nam, H.-J., Jeon, J., and Kim, S. (2009) Bioinformatic approaches for the structure and function of membrane proteins. *BMB Rep.* **42**, 697–704.
- Pavord, I. D., Beasley, R., Agusti, A., Anderson, G. P., Bel, E., Brusselle, G., Cullinan, P., Custovic, A., Ducharme, F. M., Fahy, J. V., et al. (2018) After asthma: redefining airways diseases. *Lancet* **391**, 350–400.
- Pei, G., Chen, L., and Zhang, W. (2017) WGCNA application to proteomic and metabolomic data analysis. *Methods Enzymol.* **585**, 135–158.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., et al. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452.
- Yang, Q., Wang, R., Wei, B., Peng, C., Wang, L., Hu, G., Kong, D., and Du, C. (2018) Candidate biomarkers and molecular mechanism investigation for glioblastoma multiforme utilizing WGCNA. *Biomed Res. Int.* **2018**, 4246703.
- Yin, L., Cai, Z., Zhu, B., and Xu, C. (2018) Identification of key pathways and genes in the dynamic progression of HCC based on WGCNA. *Genes (Basel)* **9**, 92.