



# A network approach predicts NFKBIA and BIRC3 as pathogenic genes in childhood asthma

X.M. Gao

Department of Pediatrics, Shaanxi Provincial People's Hospital, Xi'an, China

Corresponding author: X.M. Gao

E-mail: gaoxiaomin369@163.com

Genet. Mol. Res. 15 (2): gmr.15028226

Received December 8, 2015

Accepted February 11, 2016

Published June 21, 2016

DOI <http://dx.doi.org/10.4238/gmr.15028226>

**ABSTRACT.** This study aimed to predict pathogenic genes of childhood asthma based on molecular interaction networks and gene expression data. Known pathogenic genes identified from the human protein-protein interaction network were denoted as seed genes, and were included in network A. We extracted sub-network B (pathogenic network), which consisted of genes that interacted with at least two seed genes. We assigned a weight to select the pathogenic genes from this network according to its interactions and co-expressions with the seed genes. We also conducted ClusterONE analysis for the pathogenic network, and determined the statistical significance of the predicted clusters through a significance score (SS). Lastly, we investigated the biological pathways of the seed and candidate genes based on information obtained from the KEGG database. In network A, we identified 172 interactions and 125 genes that interacted with seed genes. In the pathogenic network, we found 51 genes and 102 interactions. The top 10 candidate genes with high weight scores were recorded. The SS of the predicted clusters demonstrated that 3 gene clusters were statistically relevant for childhood asthma. Pathway analysis showed

that the seed genes and the top 10 candidate genes were significantly enriched in the same three biological processes. NFKBIA and BIRC3, which were involved in all three biological processes, may therefore be pathogenic genes. Using the network approach, we predicted that the pathogenic genes NFKBIA and BIRC3 are associated with childhood asthma. This information can provide guidelines for future experimental verification of childhood asthma.

**Key words:** Childhood asthma; Pathogenic genes; Seed genes; Pathogenic network

## INTRODUCTION

Childhood asthma is a chronic inflammatory disorder of the airways (Page et al., 2013). The prevalence of asthma in the 6-7-year-old group is 11.7%, while its prevalence in 13-14 years is 14.1% (Yamamoto and Nagano, 2015). Therefore, it is necessary to understand the pathogenesis of childhood asthma by examining the gene components involved in this disease. This may prevent or alter the harmful features of these pathogenic genes.

This study presents a novel systems biology approach to predict pathogenic genes for childhood asthma via integration of protein interaction maps and gene expression data. By predicting the underlying molecular biomarkers of childhood asthma, we can not only provide guidelines for future experimental verification, but also shed light on the pathogenesis of this disease.

## MATERIAL AND METHODS

### Collection and preprocessing of expression profile data

Based on the Affymetrix GeneChip Human Gene 1.0 ST Array platform (HuGene-1\_0-st-v1), the gene expression data of allergic asthmatic children and healthy controls (E-GEOD-40888) were obtained from the Array Express Archive of Functional Genomics Data (<http://www.ebi.ac.uk/arrayexpress/>). A total of 81 samples were obtained, including 41 children with allergic asthma and 40 healthy controls. Each probe was mapped to a single gene by the getSYMBOL function; the probe was discarded if it was not able to match to any genes. We were able to obtain 12,315 genes in the expression profile data.

### Detection of the DEGs

In the present study, the DEGs between allergic asthmatic children and healthy controls were screened by the Significance Analysis of Microarrays (SAM) package, which correlated a large number of features (e.g., genes) with an outcome variable (Tibshirani et al., 2011). SAM assigned a score to each gene on the basis of its expression change relative to the standard deviation of repeated measurements. Genes with scores greater than a set threshold were denoted to be potentially significant. The percentage of false positives identified by chance was the false-discovery rate (FDR). The tables of thresholds, cut points, and corresponding FDR for SAM analysis were calculated using the SAMR.compute.delta.table function (Tusher et al., 2001). The significant gene table was computed, starting with samr object "samr.obj" and

delta.table “delta.table” (Tusher et al., 2001). The “relative difference”  $d(i)$  in gene expression was calculated using the function:

$$d(i) = \frac{\bar{x}_I(i) - \bar{x}_U(i)}{s(i) + s_0} \quad (\text{Equation 1})$$

where  $\bar{x}_I(i)$  and  $\bar{x}_U(i)$  are defined as the average levels of expression for gene ( $i$ ) in states I and U, respectively. The “gene-specific scatter”  $s(i)$  is the standard deviation of repeated expression measurements. Genes were ranked to find significant changes in expression by magnitude of their  $d(i)$  values. The number of falsely significant genes corresponding to each permutation was computed by counting the number of genes that exceeded the horizontal cutoffs for induced and repressed genes. The DEGs were selected based on a delta value of 0.754.

### Identification of pathogenic network

We downloaded 31 genes that have been identified as pathogenic genes in childhood asthma from the OMIM database, which contained manually curated genes proven to affect the outcomes of pathogen-host interactions. Moreover, we selected intersection genes between the 12,315 genes in the expression profile data and the 31 pathogenic genes deposited in the OMIM database. We were able to obtain 27 intersection genes that were then defined as seed genes. All predicted human protein-protein interaction networks (PPINs) were downloaded from the database Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, <http://string-db.org/>) (von Mering et al., 2005). In this study, we selected high-confidence protein interactions from the PPIN, where the seed genes described above were mapped onto the PPIN. Network A was constructed based on the high-confidence protein interactions from PPIN, which consisted of genes that interacted with seed genes, as well as being included in the DEGs. Subsequently, sub-network B was extracted from network A, and consisted of genes that interacted with at least two seed genes. Furthermore, we regarded network B as the pathogenic network, where the genes were believed to be associated with disease onset. In addition, we predicted connected network clusters based on ClusterONE analysis.

### Ranking of pathogenic genes

To determine which genes were more pathogenic from the predicted pathogenic network, we assigned a weight to each gene according to its interactions and co-expressions with seed genes. The higher the weight score of a gene, the more it is involved in disease pathogenesis. Co-expression was evaluated by the Pearson correlation coefficient, which was a measure of the correlation between our predicted pathogenic gene and seed genes; this value ranged from -1 to +1 (Li et al., 2012). The weight  $w(x)$  of each gene  $x$  was computed as follows:

$$w(x) = \sum_{y \in S} PC(x, y) \times I(x, y) \quad (\text{Equation 2})$$

where  $S$  is the set of known pathogenic genes,  $PC(x, y)$  is the correlation coefficient between gene  $x$  and gene  $y$ , and  $I(x, y)$  is an indication function where  $I(x, y) \sim 1$  if protein  $x$  interacts

with protein  $y$ , and  $I(x, y) \sim 0$  otherwise. In this article, we selected pathogenic genes with weight value  $>0.7$  as candidate genes.

### Statistical analysis of predicted clusters

To illuminate the significance of the predicted clusters, we calculated a significance score (SS) for each cluster as the geometric mean of the P values of all nodes in this cluster. The P value of each node was obtained by the Mann-Whitney Wilcoxon test in accordance with gene expression data between the control group and disease group.

We performed permutation testing in order to identify the statistical significance of the predicted clusters. Firstly, the P value of each node in the cluster was randomly shuffled and redistributed. Secondly, the SS of each cluster was recalculated after the P value labels were shuffled and redistributed. Next, randomization was repeated 10,000 times. Lastly, the P value of each cluster was defined as the probability that each random cluster can be detected with smaller SS than that of our predicted cluster. This was used to determine whether the predicted clusters can be detected by chance ( $P < 0.05$ ).

### Pathway-enrichment analysis

The candidate genes in network B were expected to be involved in the same biological processes as the seed genes, which implied that they were potential pathogenic genes. To investigate the shared biological pathways between the seed genes and the candidate genes, we performed pathway analysis based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/>). We imported the seed genes and the candidate genes to the Database for Annotation, Visualization and Integrated Discovery (Huang et al., 2009) (DAVID, <http://david.abcc.ncifcrf.gov>), and obtained the same biological pathways with P value  $<0.1$ .

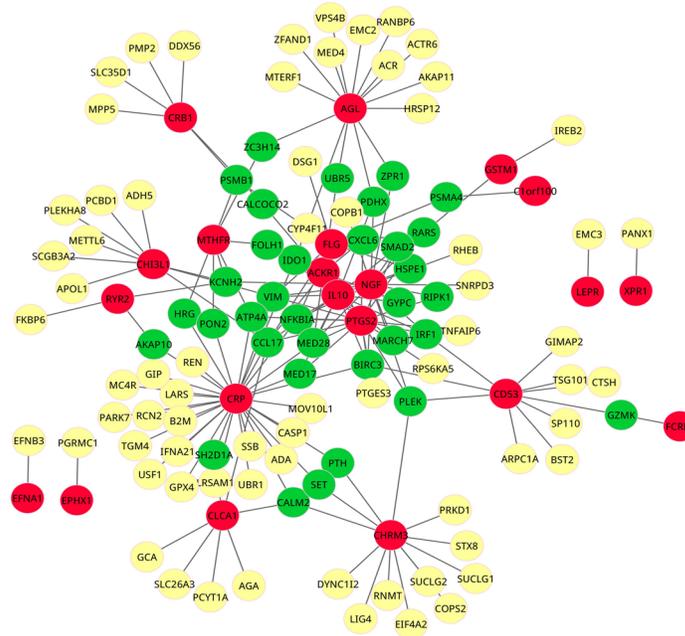
## RESULTS

### Detection of the DEGs

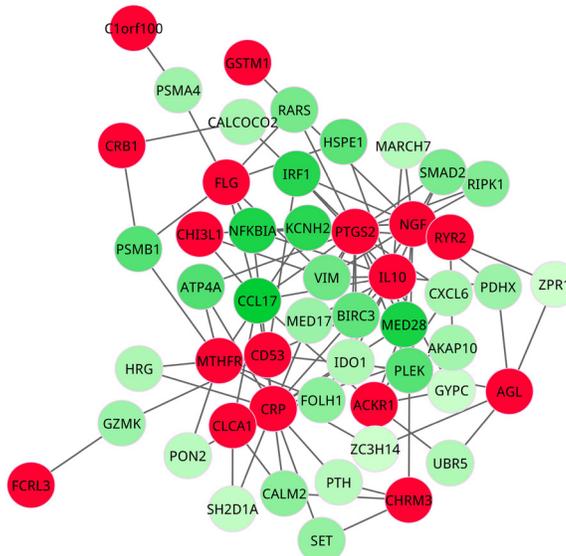
We found that certain genes were differentially expressed in the childhood asthma disease group. We found 1001 DEGs from the 12,315 genes included in the gene expression profile data of childhood asthma (delta value = 0.754). Since the DEGs may be related to the interactions between the pathogen and its host, they may act as pathogenic genes.

### Identification of pathogenic network

In this study, 31 pathogenic genes from OMIM-base were used as seed genes, among which 27 can be mapped to the childhood asthma interaction map. As a result, 125 genes in network A with 172 interactions were found to be linked to at least one seed gene; the results are shown in Figure 1. The genes in network B (pathogenic network), which interacted with at least two seed genes, are shown in Figure 2. These genes were more likely to be pathogenic due to their tight interactions with the seed genes. Result indicated that network B included 51 genes and 102 interactions.



**Figure 1.** Distribution results of the pathogenic network. The red vertices denote seed genes from OMIM, the green vertices denote genes that interacted with at least two seed genes, and the yellow vertices denote genes that interacted with only one seed gene.



**Figure 2.** Distribution results of the filtered pathogenic network. The red vertices denote seed genes, the green vertices are genes that interact with at least two seed genes. There are lots of the green vertices that the deeper the color represents the larger the weight was.

## Ranking of the pathogenic genes

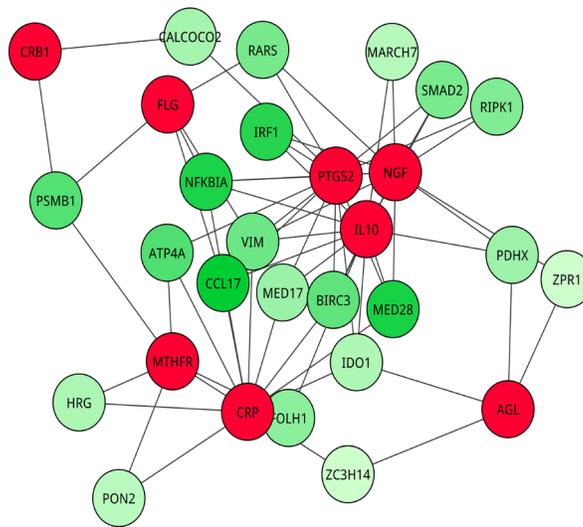
According to the interactions and co-expressions with seed genes, there were 34 candidate genes in total. The candidate genes with weight values  $>0.7$  are shown in Table 1. The candidate genes identified above with high weights were more likely to be pathogenic genes due to a greater number of interactions and higher correlations with known pathogenic genes.

**Table 1.** Candidate genes with weight values  $>0.7$ .

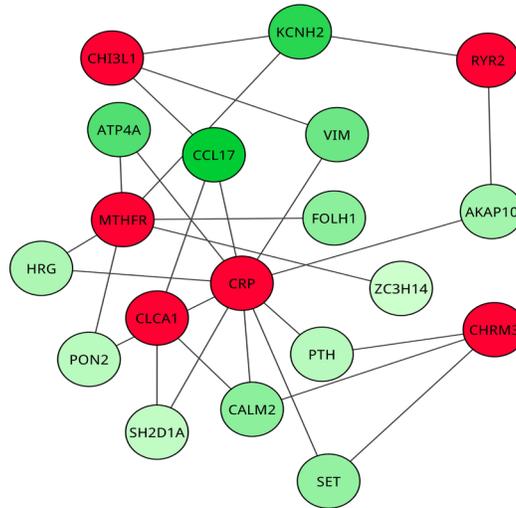
Row	Node	Weight
1	CCL17	1.359
2	MED28	1.208
3	NFKBIA	1.189
4	IRF1	1.128
5	KCNH2	1.106
6	ATP4A	0.857
7	PSMB1	0.842
8	PLEK	0.820
9	HSPE1	0.786
10	BIRC3	0.774

## Statistical analysis of prediction results

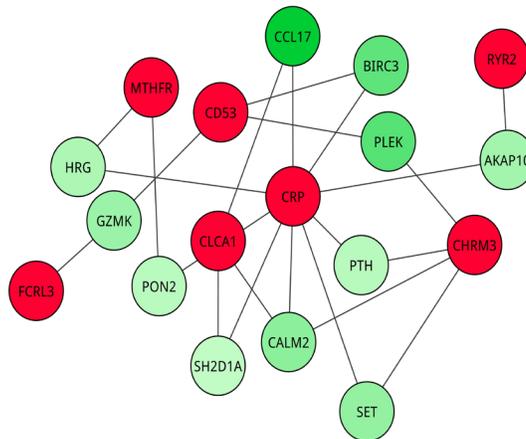
In this study, 3 clusters, shown in Figures 3, 4, and 5, were discovered to be associated with childhood asthma. We calculated the SS value of these clusters (0.014, 0.064, and 0.053). The P values of the clusters were  $0, 8.5 \times 10^{-3}$  and  $4.2 \times 10^{-3}$ , suggesting that these clusters were statistically significant results.



**Figure 3.** Distribution results of the cluster one. The red vertices denote seed genes from OMIM, the green vertices denote genes that interact with at least two seed genes.



**Figure 4.** Distribution results of the cluster two. The red vertices denote seed genes from OMIM, the green vertices denote genes that interacted with at least two seed genes.



**Figure 5.** Distribution results of the cluster three. The red vertices denote seed genes from OMIM, the green vertices denote genes that interact with at least two seed genes.

### Pathway-enrichment analysis

Pathway analysis showed that the seed genes and the candidate genes were significantly enriched in 3 biological processes: small cell lung cancer ( $P = 0.035$ ), apoptosis ( $P = 0.037$ ) and asthma ( $P = 0.098$ ). Among these terms, NFKBIA and BIRC3 belonged to the candidate genes, while the rest were all seed genes. Therefore, we hypothesized that NFKBIA and BIRC3 may be associated with childhood asthma.

## DISCUSSION

Studies show that there is a sharp increase in asthma worldwide, particularly in children; this is a major public health problem (Masoli et al., 2004; Asher and Pearce, 2014). Children with severe asthma have poor disease control despite high doses of inhaled corticosteroids and additional therapy. This leads to poor quality of life, early deterioration of lung function, and significant consumption of health care resources (Persson et al., 2015).

Identifying pathogenic genes can help to reduce the occurrence, development, and mortality of childhood asthma. In this study, we presented a novel network approach to predict pathogenic genes that interact with known pathogenic genes of childhood asthma. Through this technique, we were able to identify potential genes involved in the pathogenesis of childhood asthma. We hypothesized that if the candidate genes were involved in the same biological processes as the seed genes, the candidate genes may also be potential pathogenic genes. For further assessments of the overlaps between seed genes and the candidate gene pathways, we imported our findings to DAVID, and obtained 3 biological enrichment pathways that were shared between the seed genes and the candidate genes. Moreover, the genes NFKBIA and BIRC3 were candidate genes that belonged to all 3 biological pathways. Hence, it can be suggested that NFKBIA and BIRC3 may be pathogenic genes. Statistical analysis of the identified clusters and the predicted pathogenic genes demonstrated that our prediction results were statistically significant.

The NF- $\kappa$ B inhibitor alpha (NFKBIA) gene encodes the NF- $\kappa$ B inhibitor, I $\kappa$ B $\alpha$ . NFKBIA has been reported to be closely related to many diseases. For example, it was stated that a few deletions of the NFKBIA gene is associated with glioblastomas (Patanè et al., 2013). Han et al. (2015) reported that polymorphisms in the NFKBIA genes may modulate the risk of developing prostate cancer in the Chinese population. Moreover, the importance of negative innate immune regulators such as NFKBIA has been verified in pediatric lung diseases, and studies have begun to unravel certain common aspects between genetic predisposition to childhood asthma and lung diseases (Ali et al., 2013). Baculoviral IAP repeat-containing protein3 (also known as cIAP2) is a protein encoded by the BIRC3 gene in humans (Rothe et al., 1995; Liston et al., 1996). BIRC2/3-NIK signaling target effectors contribute directly to carcinogenesis (Yamato et al., 2015). It has been reported that chronic lymphocytic leukemia has BIRC3-associated abnormalities (Puiggros et al., 2014). Moreover, some researchers have suggested that Lynch syndrome and sporadic ovarian cancers differ by several deregulated genes such as BIRC3 (Jönsson et al., 2014).

In summary, the novel network approach presented here can open opportunities to predict pathogenic genes of childhood asthma based on protein interaction networks and gene expression data. Through mutual enrichment pathway analysis of both the seed genes and the candidate genes, we discovered that the candidate genes NFKBIA and BIRC3 may play important roles in childhood asthma.

## Conflicts of interest

The authors declare no conflict of interest.

## REFERENCES

- Ali S, Hirschfeld AF, Mayer ML, Fortuno ES, 3rd, et al. (2013). Functional genetic variation in NFKBIA and susceptibility to childhood asthma, bronchiolitis, and bronchopulmonary dysplasia. *J. Immunol.* 190: 3949-3958. <http://dx.doi.org/10.4049/jimmunol.1201015>
- Asher I and Pearce N (2014). Global burden of asthma among children. *Int. J. Tuberc. Lung Dis.* 18: 1269-1278. <http://dx.doi.org/10.5588/ijtld.14.0170>
- Han X, Zhang JJ, Yao N, Wang G, et al. (2015). Polymorphisms in NFKB1 and NFKBIA genes modulate the risk of developing prostate cancer among Han Chinese. *Med. Sci. Monit.* 21: 1707-1715. <http://dx.doi.org/10.12659/MSM.893471>
- Huang W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57. <http://dx.doi.org/10.1038/nprot.2008.211>
- Jönsson JM, Bartuma K, Dominguez-Valentin M, Harbst K, et al. (2014). Distinct gene expression profiles in ovarian cancer linked to Lynch syndrome. *Fam. Cancer* 13: 537-545. <http://dx.doi.org/10.1007/s10689-014-9728-1>
- Li M, Zhang H, Wang JX and Pan Y (2012). A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data. *BMC Syst. Biol.* 6: 15. <http://dx.doi.org/10.1186/1752-0509-6-15>
- Liston P, Roy N, Tamai K, Lefebvre C, et al. (1996). Suppression of apoptosis in mammalian cells by NAIP and a related family of IAP genes. *Nature* 379: 349-353. <http://dx.doi.org/10.1038/379349a0>
- Masoli M, Fabian D, Holt S and Beasley R; Global Initiative for Asthma (GINA) Program (2004). The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy* 59: 469-478. <http://dx.doi.org/10.1111/j.1398-9995.2004.00526.x>
- Page TF, Beck-Sague CM, Pinzon-Iregui MC, Cuddihy A, et al. (2013). Asthma in underserved schoolchildren in Miami, Florida: results of a school- and community-based needs assessment. *J. Asthma* 50: 480-487. <http://dx.doi.org/10.3109/02770903.2013.790416>
- Patanè M, Poratti P, Bottega E, Morosini S, et al. (2013). Frequency of NFKBIA deletions is low in glioblastomas and skewed in glioblastoma neurospheres. *Mol. Cancer* 12: 160. <http://dx.doi.org/10.1186/1476-4598-12-160>
- Persson H, Kwon AT, Ramilowski JA, Silberberg G, et al. (2015). Transcriptome analysis of controlled and therapy-resistant childhood asthma reveals distinct gene expression profiles. *J. Allergy Clin. Immunol.* 136: 638-648. <http://dx.doi.org/10.1016/j.jaci.2015.02.026>
- Puiggros A, Blanco G and Espinet B (2014). Genetic abnormalities in chronic lymphocytic leukemia: where we are and where we go. *BioMed Res. Int.* 2014: 435983. <http://dx.doi.org/10.1155/2014/435983>
- Rothe M, Pan MG, Henzel WJ, Ayres TM, et al. (1995). The TNFR2-TRAF signaling complex contains two novel proteins related to baculoviral inhibitor of apoptosis proteins. *Cell* 83: 1243-1252. [http://dx.doi.org/10.1016/0092-8674\(95\)90149-3](http://dx.doi.org/10.1016/0092-8674(95)90149-3)
- Tibshirani R, Chu G, Narasimhan B and Li J (2011). SAM: Significance Analysis of Microarrays. *R package, version 2.0*. URL: <https://cran.r-project.org/web/packages/samr/samr.pdf>
- Tusher VG, Tibshirani R and Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* 98: 5116-5121. <http://dx.doi.org/10.1073/pnas.091062498>
- von Mering C, Jensen LJ, Snel B, Hooper SD, et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33: D433-D437. <http://dx.doi.org/10.1093/nar/gki005>
- Yamamoto N and Nagano J (2015). Parental stress and the onset and course of childhood asthma. *Biopsychosoc. Med.* 9: 7. <http://dx.doi.org/10.1186/s13030-015-0034-4>
- Yamato A, Soda M, Ueno T, Kojima S, et al. (2015). Oncogenic activity of BIRC2 and BIRC3 mutants independent of nuclear factor- $\kappa$ B-activating potential. *Cancer Sci.* 106: 1137-1142. <http://dx.doi.org/10.1111/cas.12726>