Observational Study

Medicine®

OPEN

# CD74, a novel predictor for bronchopulmonary dysplasia in preterm infants

Junyan Gao, MD[a], Mingfu Wu, MM[a], Fudong Wang, MM[a], Lijun Jiang, MD[a], Rui Tian, MM[b], Xueping Zhu, MD[c], Shan He, MD[c,d,*]

**Abstract**

Bronchopulmonary dysplasia (BPD) remains a major complication and accounts for high morbidity and mortality of preterm infants. The present study aimed to identify the key genes in the development of BPD and to provide some new insights into the pathogenesis of BPD. The GSE108754 dataset was downloaded from Gene Expression Omnibus database containing 5 samples of BPD patients and 6 of non-BPD infants. The differentially expressed genes (DEGs) between BPD and non-BPD patients were identified by R software. The pathway and function enrichment analyses were performed through Database for Annotation Visualization and Integrated Discovery website. The protein-protein interaction network for DEGs was established by Cytoscape software and the most highly connected module was selected through MCODE plugin. Furthermore, the clinical sample verification among 25 BPD patients and 10 non-BPD infants was carried out in our center. Finally, based on the results above, the gene set enrichment analysis focusing on CD74 upregulated status was employed. Totally, 189 DEGs including 147 upregulated genes and 42 downregulated genes between BPD and non-BPD patients were screened out. The pathway and function enrichments revealed these DEGs were mainly enriched in asthma, intestinal immune network for IgA production, antigen processing and presentation and immune response. Thirteen DEGs (CD74, HLA-DMA, HLA-DRA, HLA-DMB, HLA-DOB, HLA-DQA1, HLA-DRB5, HLA-DPA1, HLA-DOA, HLA-DPB1, HLA-DQB2, HLA-DQA2, and HLA-DQB1) were determined as hub genes. The mRNA expression levels of the 13 hub genes were tested by quantitative real-time polymerase chain reaction among our clinical samples. Eventually, CD74 was confirmed to be the most significant highly expressed in BPD samples ($P < .001$) and its expression level was negatively correlated with gestational age ($r = -0.653$) and birth weight ($r = -0.675$). The gene set enrichment analysis results showed the gene sets associated with lupus erythematosus, viral myocarditis, immune network for IgA production, graft versus host disease, cell adhesion molecules and so no were differentially enriched with the phenotype of high-expression CD74. In conclusion, CD74 may serve to predict the BPD development and provide a new therapeutic target for BPD.

**Abbreviations:** BPD = bronchopulmonary dysplasia, BW = birth weight, DEGs = differentially expressed genes, GA = gestational age, GO = gene ontology, GSEA = gene set enrichment analysis, KEGG = Kyoto Encyclopedia, MHC = major histocompatibility complex, PPI = protein-protein interaction, qRT-PCR = quantitative real-time PCR.

**Keywords:** bionomics, bronchopulmonary dysplasia, hub genes, pathway enrichment analyses, preterm infants, quantitative real-time PCR

## 1. Introduction

Bronchopulmonary dysplasia (BPD) remains a leading cause of respiratory illness and results in significant morbidity and mortality of neonates specifically of preterm infants.[1] The incidence of BPD varies country to country, based on the different risk factors and care practices as well as the inconsistence in the clinical definitions of BPD.[2,3] Neonates with birth weight (BW) <1250 g contribute 97 percent of the cases of BPD. For extremely preterm infants (defined as gestational age [GA] <28 weeks), the incidence of BPD is approximately 40%, and the risk increases with the decreasing GA.[3]

In recent decades, as the great progress has been achieved in neonatal-perinatal medicine, the survival rate of extremely preterm infants is considerable higher than it used to be. However, BPD stands out a frustrating clinical problem. So far, the exact mechanism regarding BPD pathogenesis is still ambiguous and multiple factors are thought to be involved in BPD development including preterm delivery, long duration of oxygen supplementation, mechanical ventilation, and so on.

It has been confirmed that genetic influences play a crucial role in the pathogenesis of BPD.[4] Moreover, in some twin studies,[5,6] it revealed that the heritability component is estimated to be between

*Editor: Jing Liu.*

*The authors have no conflicts of interest to disclose.*

*The datasets generated during and/or analyzed during the current study are publicly available.*

*[a] Department of Pediatrics, Affiliated Hospital of Yangzhou University, Yangzhou, Jiangsu, [b] Department of Pediatrics, The First People's Hospital of Kunming City, Kunming, Yunnan, [c] Department of Neonatology, Children's Hospital of Soochow University, Suzhou, Jiangsu, [d] Department of Pediatrics, The First People's Hospital of Yunnan Province, Kunming, Yunnan, China.*

*[*] Correspondence: Shan He, Department of Neonatology, Children's Hospital of Soochow University, Suzhou, Jiangsu 215025, China, Department of Pediatrics, The First People's Hospital of Yunnan Province, Kunming, Yunnan 650031, China (e-mail: 543939290@qq.com).*



*How to cite this article: Gao J, Wu M, Wang F, Jiang L, Tian R, Zhu X, He S. CD74, a novel predictor for bronchopulmonary dysplasia in preterm infants. Medicine 2020;99:48(e23477).*

*Received: 6 May 2020 / Received in final form: 29 September 2020 / Accepted: 31 October 2020*

*http://dx.doi.org/10.1097/MD.0000000000023477*

50% and 80%, indicating that BPD has a significant genetic background. Despite this, no specific genes or heritable factors have been successfully identified yet.

Thanks to the development of microarray technology, the differentially expressed genes (DEGs) between BPD patients and non-BPD infants could be analyzed through the gene database. In the present study, we aim to screen and identify the potential key genes of BPD pathogenesis and hope to assist the early diagnosis and targeted treatment of BPD.

The gene expression matrix of BPD patients and non-BPD neonates were downloaded from Gene Expression Omnibus database[7] and analyzed. The DEGs regarding BPD versus no-BPD group were screened out. Furthermore, the Kyoto Encyclopedia of Genes and Genomes (KEGG)[8] pathway and Gene Ontology (GO)[9] enrichments were analyzed by means of Database for Annotation, Visualization, and Integrated Discovery website[10] subsequently. The topological analyses based on protein to protein interaction (PPI) make the selection of hub genes feasible. In our study, several key genes derived from the common network representing both PPI and risk sub-pathway information were identified. After key genes determined, further verification of expressions of these genes in BPD patients and non-BPD infants were conducted by quantitative real-time polymerase chain reaction (qRT-PCR). Furthermore, the gene set enrichment analysis (GSEA) focusing on CD74 upregulated status was conducted based on the results of clinical verification.

These results may provide a new insight into the understanding of BPD pathogenesis and these key genes may be served as potential therapeutic targets for BPD.

## 2. Materials and methods

### 2.1. Microarray data

The dataset of GSE108754[11] (GPL13497 platform, Agilent-026652 Whole Human Genome Microarray 4x44K v2) was downloaded from Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) which contained 5 and 6 blood samples from BPD patients and non-BPD infants, respectively. The probes were converted into the corresponding gene symbol based on the annotation information of the platform.

### 2.2. DEGs identification

The DEGs between BPD and non-BPD infants were identified by using Lima package of R software (version 3.6.2) and the further selective criteria were |logFC(fold-change)| >1 plus adj. *P*-value < .01.

### 2.3. KEGG and GO enrichment analyses of DEGs

The Database for Annotation, Visualization, and Integrated Discovery website (https://david.ncifcrf.gov/, version 6.8) was adopted to perform KEGG pathway and GO term analyses of DEGs and the enrollment criteria were *P*-value < .05 and DEG counts ≥2.

### 2.4. PPI network construction and module analysis

The Search Tool for the Retrieval of Interacting Genes (version 11.0, http://string-db.org) website was used for building the PPI network of aforementioned DEGs and the PPI score >0.4 was considered as the inclusion criteria. After the network established, the nodes without any connections were excluded and the data of the remaining network was imported in Cytoscape software (version 3.7.2) for further module analysis and data visualization.

### 2.5. Gene-disease network construction

The gene-disease association information of DEGs was collected form the DisGeNET database (https://www.disgenet.org/, version 6.0) which is a discovery platform containing one of the largest publicly available collections of genes and variants associated to human diseases.[12] The threshold was set at a degree of >1.0 and the network was visualized by R software.

### 2.6. Samples collection and ethics statement

Thirty-five blood samples were collected from 25 BPD patients and 10 non-BPD infants who were hospitalized from June 2018 to June 2019 in our center. Patients who were enrolled in have been informed and signed the informed consent form. This study was approved by the institutional ethics committee of Children's Hospital of Soochow University. Patients' clinical characteristics were summarized in Table 1.

### 2.7. qRT-PCR

Total RNA from peripheral blood mononuclear cells were isolated using TRIzol reagent (Invitrogen [Carlsbad CA, United States]) immediately and stored at −80°C. Total RNA of each sample was reverse transcribed into cDNAs by using the reverse transcription kit (Takara, Japan). The qRT-PCR was employed to measure the levels of mRNAs using the comparative Ct method. GAPDH was considered as the normalization control for mRNA. All the primers for qRT-PCR were listed in Table 2.

**Table 1**

**The clinical characteristics of 25 BPD patients and 10 non-BPD infants.**

| Variables | BPD | Control | *P*-value |
|---|---|---|---|
| Patients' characteristics | | | |
| Gender, male N (%) | 20 (80.0) | 6 (60.0) | .393 |
| Gestational age (wk) median (range) | 32.3 (25.7–36.6) | 35.3 (30.6–39.4) | .009 |
| Birth weight (kg) median (range) | 1.44 (0.75–2.79) | 2.38 (1.37–4.08) | .003 |
| Respiratory distress syndrome N (%) | 20 (80.0) | 0 (0) | <.001 |
| Sepsis N (%) | 18 (72.0) | 2 (20.0) | .008 |
| Pulmonary surfactant N (%) | 17 (68.0) | 0 (0) | <.001 |
| Mechanical ventilation N (%) | 16 (64.0) | 0 (0) | .001 |
| Perinatal steroids N (%) | 13 (52.0) | 4 (40.0) | .711 |
| Oxygen supplementation (d) median (range) | 37 (12–69) | 3.5 (0–14) | <.001 |
| Delivery mode | | | |
| Caesarean section N (%) | 9 (36.0) | 3 (30.0) | 1.000 |
| Perinatal asphyxia N (%) | 11 (44.0) | 3 (30.0) | 1.000 |
| Maternal conditions | | | |
| Maternal age >35 yr N (%) | 10 (40.0) | 3 (30.0) | .709 |
| Preeclampsia/eclampsia N (%) | 2 (8.0) | 1 (10.0) | 1.000 |
| Smoking during pregnacy N (%) | 2 (8.0) | 0 (0) | 1.000 |

BPD = bronchopulmonary dysplasia.

**Table 2**

**The sequences of hub gene primers.**

| Gene symbol | | Primer sequence |
|---|---|---|
| CD74 | Forward | 5′-CCAGCGAGGAGCAGAGTCAC-3′ |
| | Reverse | 5′-TTATCTCCAACAATGAGCAACT-3′ |
| HLA-DRA | Forward | 5′-TCCCTGAAGCTCCTACTCCAA-3′ |
| | Reverse | 5′-CTGTGTGGCAAGAAGGTATG-3′ |
| HLA-DMB | Forward | 5′-GAGCAGGTGGCTTCGTGGC-3′ |
| | Reverse | 5′-CATCTTTACAGAGCAGAGCAT-3′ |
| HLA-DMA | Forward | 5′-TCCCTGAAGCTCCTACTCCAA-3′ |
| | Reverse | 5′-CTGTGTGGCAAGAAGGTATG-3′ |
| HLA-DOB | Forward | 5′-ATCTGACCCGACTGGATTCCT-3′ |
| | Reverse | 5′-GCACCTTTTCTGTCCCGTTG-3′ |
| HLA-DQA1 | Forward | 5′-AGATGAGCAGTTCTACGTGGA-3′ |
| | Reverse | 5′-ACGGGAGACTTGGAAAACACT-3′ |
| HLA-DRB5 | Forward | 5′-AGGCAGCATTGAAGTCAGGTG-3′ |
| | Reverse | 5′-GAGAGGGCTTGTCACGCTT-3′ |
| HLA-DPA1 | Forward | 5′-ATGCGCCCTGAAGACAGAATG-3′ |
| | Reverse | 5′-ACACATGGTCCGCCTTGATG-3′ |
| HLA-DOA | Forward | 5′-CCTACGGACCCGCCTTCTA-3′ |
| | Reverse | 5′-GGCCTCGCTTTTCTTCAGG-3′ |
| HLA-DPB1 | Forward | 5′-TGTCCACCAACCTGATCCGTA-3′ |
| | Reverse | 5′-CCACGGTGACAGGACTATCCA-3′ |
| HLA-DQB2 | Forward | 5′-GTGTGCAGACACAACTACGAGG-3′ |
| | Reverse | 5′-TCACTGAGCAGACCAGCAGGTT-3′ |
| HLA-DQB1 | Forward | 5′-AGACTCTCCCGAGGATTTCGT-3′ |
| | Reverse | 5′-GAAGTAGCACATGCCCTTAAACT-3′ |
| HLA-DQA2 | Forward | 5′-TGCCTCCTATGGTGTGAACTT-3′ |
| | Reverse | 5′-AGACAGTCTCTTTCGTCTCCAG-3′ |

### 2.8. Gene set enrichment analysis

GSEA is a computational method that assesses whether a set of priori defined genes shows statistically significant and concordant differences between 2 biological states. To investigate the role of CD74 in BPD, GSEA was conducted to analyze the enrichment of datasets between high-expression CD74 (defined as the mRNA level higher than the median one) and low-expression CD74 (defined as the mRNA level lower than the median one) groups. False discovery rate <25% and nominal *P*-value < 5% were set as the cut-off criteria.

### 2.9. Statistical analysis

Continuous variables are expressed as median (range) and categorical variables are expressed as number (percentage). The independent Student *t* test was utilized to compare normal distributional variables while the Mann–Whitney *U* test was used to compare skewed distributional variables. The categorical variables were analyzed using Chi-square or Fisher Exact Test, as appropriate. Correlation between variables with normal distribution was assessed using Pearson correlation test, while skewed distributions were assessed using Spearman Rho test. SPSS 26.0 software was employed for data processing. Prism 8.0 software was served for results visualization. $P < .05$ was considered to be statistically significant.

## 3. Results

### 3.1. Identification of the DEGs

The gene expression matrix was standardized by log2 conversion (Fig. 1) and according to the inclusion criteria, eventually, there were 189 DGEs were identified (Fig. 2) consisting of 147 upregulated genes and 42 downregulated genes between BPD and non-BPD samples.

### 3.2. KEGG and GO enrichment analyses of DEGs

The KEGG pathway results revealed that the DEGs were mainly enriched in intestinal immune network for IgA production, *Staphylococcus aureus* infection and hematopoietic cell lineage (Fig. 3). While the GO term analysis showed the DEGs were mainly enriched in antigen processing and presentation, interferon-gamma-mediated signaling pathway, antigen receptor-mediated signaling pathway, and immune response (Fig. 4).

### 3.3. Gene-disease network

The significant genes from previous analysis are mapped to the corresponding molecular interaction database. The results showed that the genes mainly connected with diseases were: CNR1, HLA-DPB1, CNR2, CD74, and HLA-DQB1, while the majority of diseases that related to these genes were: schizophrenia, autosomal recessive predisposition, and liver cirrhosis (Fig. 5)

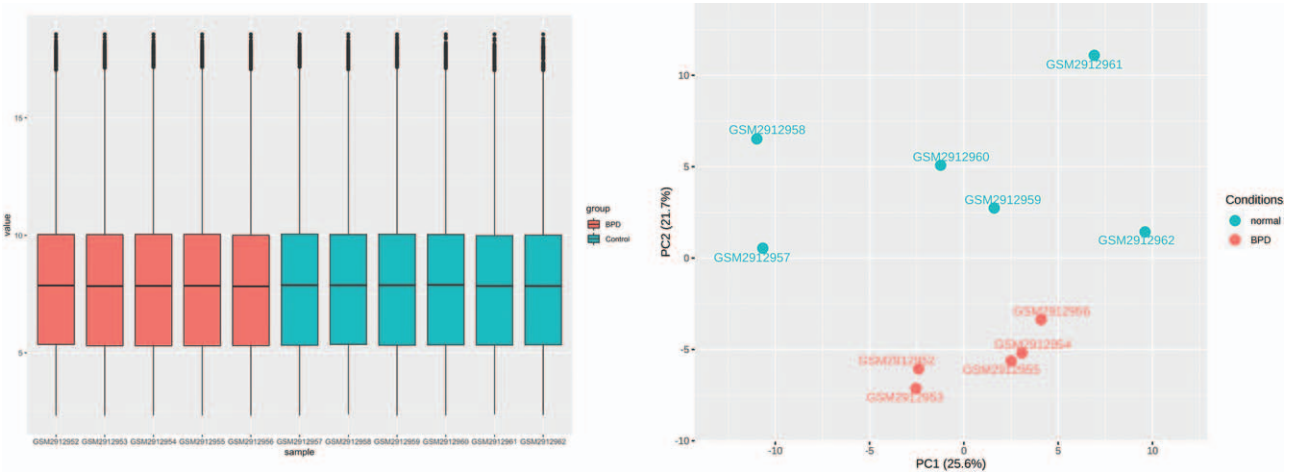


**Figure 1.** (A) Data normalized by log2 conversion. (B) The principal component analysis of samples.

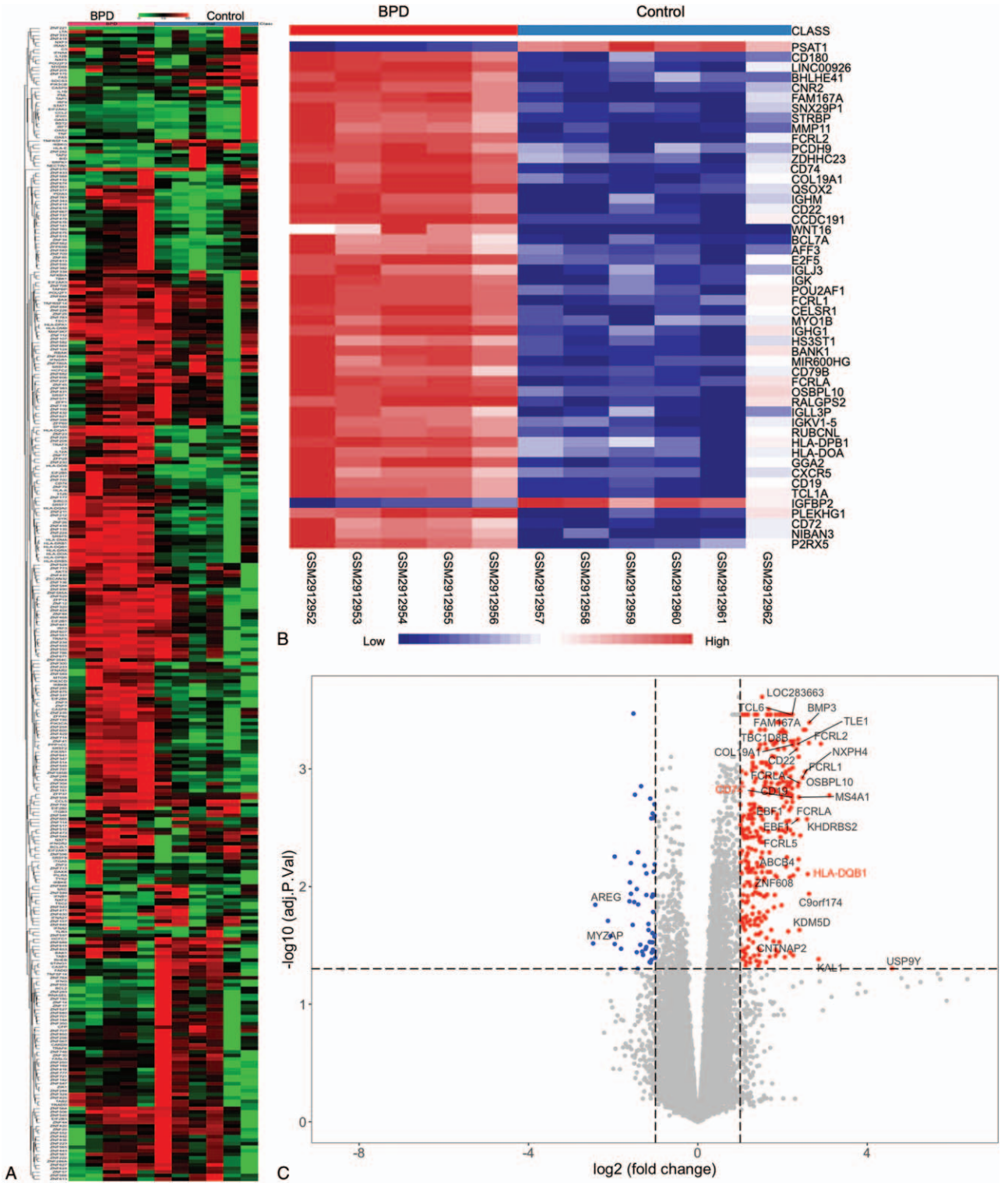


**Figure 2.** DEGs demonstrated as heatmap and volcano plot. (A) The global heatmap of DEGs. (B) The heatmap of 50 most significant DEGs. (C) The volcano plot of DEGs, red dots represent the upregulated genes while blue dots represent the downregulated genes. DEGs = differentially expressed genes.

### *3.4. PPI network construction and the hub genes identification*

The PPI network of DEGs was constructed (Fig. 6A) containing 104 nodes and 279 edges, of which, the most significant clusters were obtained by using plug-in MCODE of Cytoscape which contains 13 nodes and 78 edges. These 13 nodes were considered as the hub genes (CD74, HLA-DMA, HLA-DRA, HLA-DMB, HLA-DOB, HLA-DQA1, HLA-DRB5, HLA-DPA1, HLA-DOA, HLA-DPB1, HLA-DQB2, HLA-DQA2, and HLA-DQB1) and were ranked by using the plug-in cytoHubba of Cytoscape (Fig. 6B). The depth of color represented the different scores

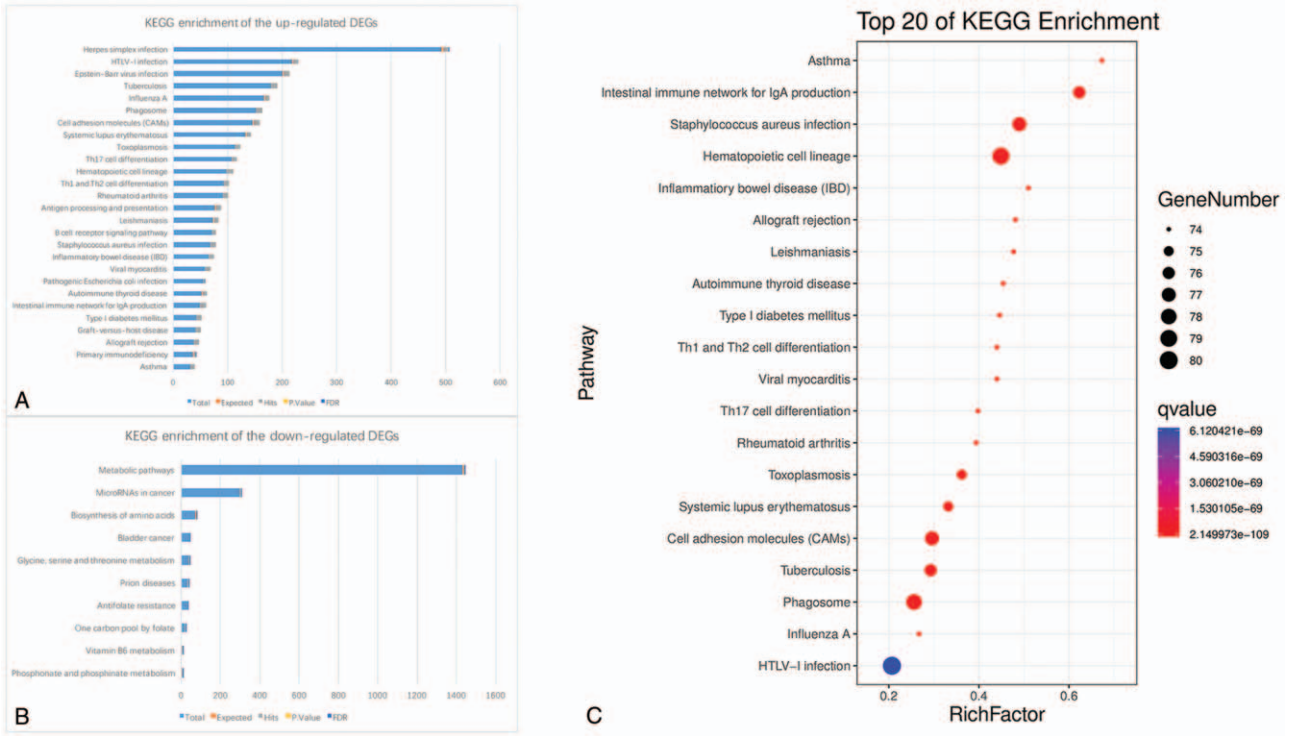


**Figure 3.** The KEGG enrichment analysis. (A) The KEGG enrichment of the upregulated genes. (B) The KEGG enrichment of the downregulated genes. (C) The top 20 of KEGG enrichment. KEGG = Kyoto Encyclopedia.

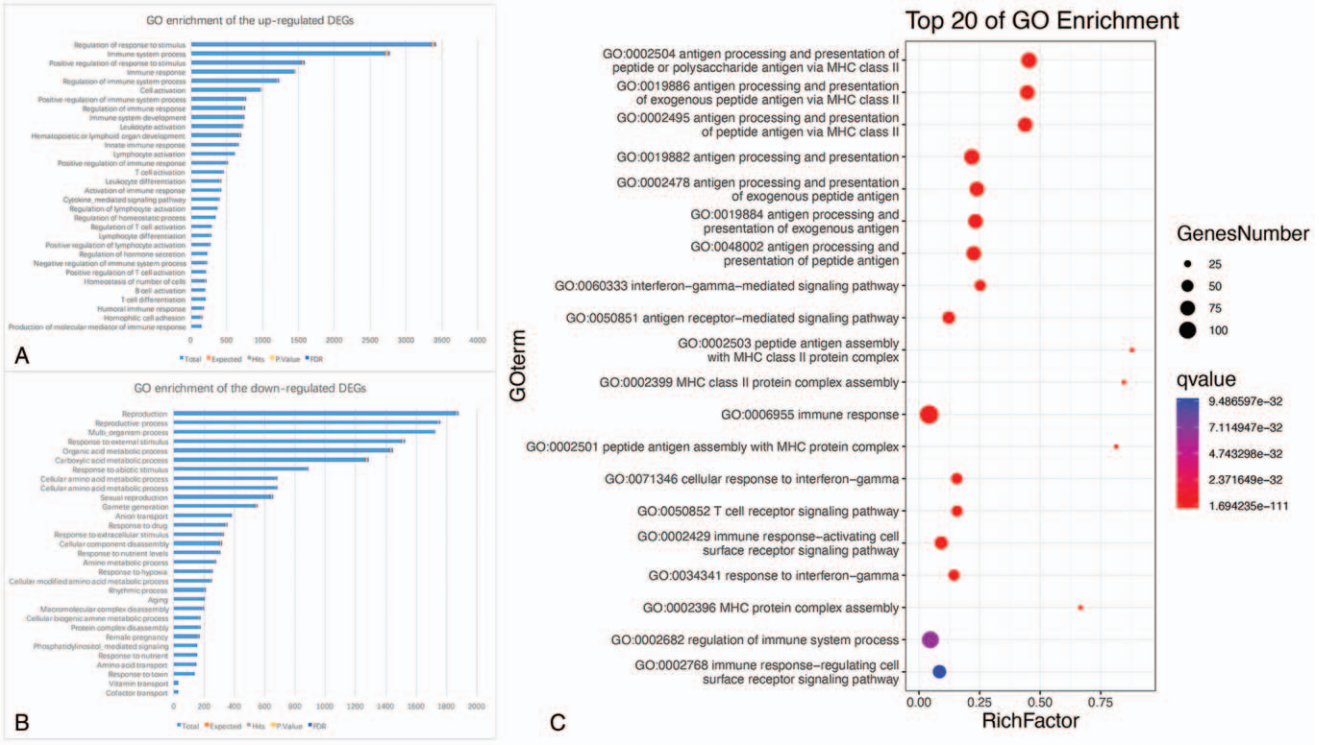


**Figure 4.** The GO enrichment analysis. (A) The GO enrichment of the upregulated genes. (B) The GO enrichment of the downregulated genes. (C) The top 20 of GO enrichment. GO = gene ontology.

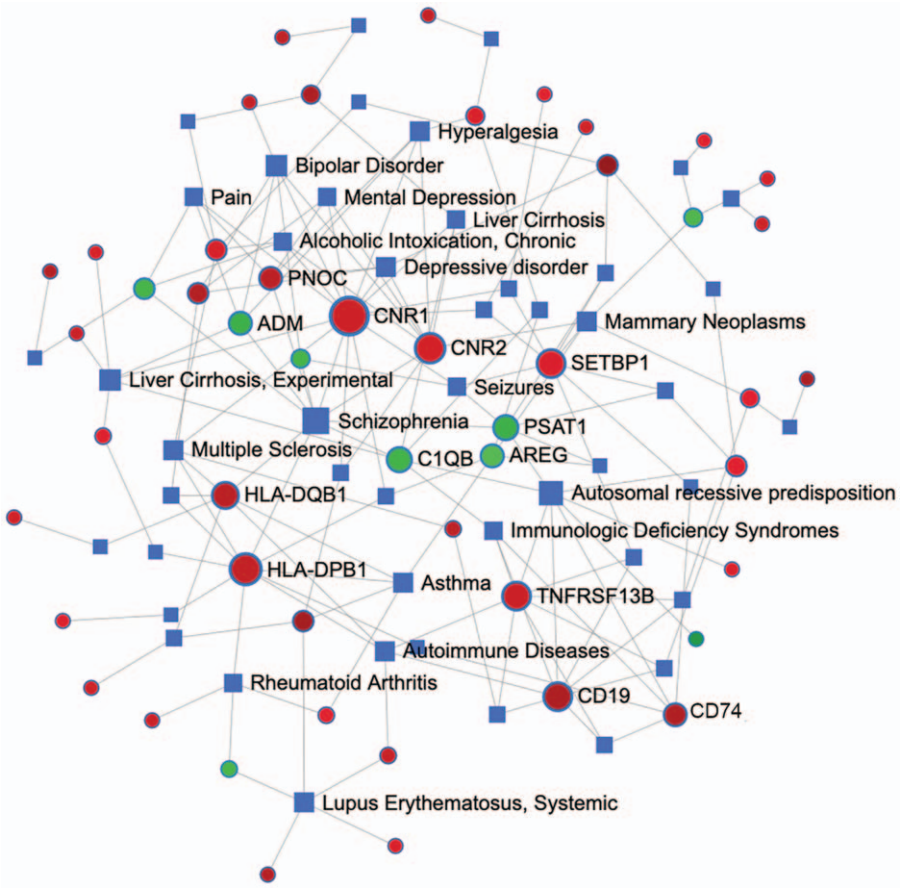


**Figure 5.** Gene-disease network was constructed through the DisGeNET database.

calculated by cytoHubba. The darker of the node, the higher score was. The names, abbreviations, and functions for these 13 hub genes were shown in Table 3.

### 3.5. Expressions of key genes in BPD patients

The 13 genes were further verified by qRT-PCR in our clinical samples. CD74 was the highest expressed in BPD patients with a $P$-value $<.001$, followed by HLA-DPB1 ($P=.017$), HLA-DOB ($P=.023$), HLA-DMA ($P=.025$), HLA-DRA ($P=.038$), and HLA-DQA1 ($P=.038$). Whereas, the remaining DEGs showed no significant difference between groups (Fig. 7).

### 3.6. Patients' clinical features and linear correlation analyses between GA, BW, and hub genes

There was no significant difference of the gender distribution, maternal age, incidence of chorioamnionitis, smoking during pregnancy, maternal preeclampsia/eclampsia, caesarean section rate, incidence of perinatal asphyxia, and the antenatal steroids administration rate between the 2 groups. The differences of GA, BW, incidence of sepsis, respiratory distress syndrome, mechanical ventilation requirement, the duration of oxygen therapy, and the pulmonary surfactant usage rate were significant between the 2 groups (Table 1). In addition, CD74, HLA-DOB, and HLA-DPB1 were negatively correlated with GA and BW (Fig. 8).

### 3.7. GSEA of CD74

To better understand the function of CD74 and signaling pathways activated in BPD, we applied GSEA comparing the datasets of low-expression and high-expression of CD74. The results showed the gene sets associated with lupus erythematosus, viral myocarditis, immune network for IgA production, graft versus host disease, cell adhesion molecules, B cell receptor signaling pathway, auto immune thyroid disease, asthma, and antigen processing and presentation pathway were differentially enriched with the phenotype of high-expression CD74 (Fig. 9).

## 4. Discussion

BPD is a type of chronic lung disease due to its disruption of pulmonary development and repeatedly damage-repairment of pulmonary tissue mainly in the population of preterm infants. It mostly affects preterm infants with GA <32 weeks and/or BW <1250 g and such neonates account for 97 percent of the cases of BPD.[13] Despite great efforts have been done on BPD, the etiology of BPD is still not fully understood and there is still no specific treatment for it. The mainly strategies are supportive care and symptomatic treatments such as steroids, oxygen therapy, diuretics, and so on. Thus, it is meaningful to explore the mechanism of BPD, and the knowledge obtained might provide great help for further diagnoses and therapeutic strategies. What is more, BPD has a strong genetic background which has been confirmed in recent decades. Therefore, figuring out which gene

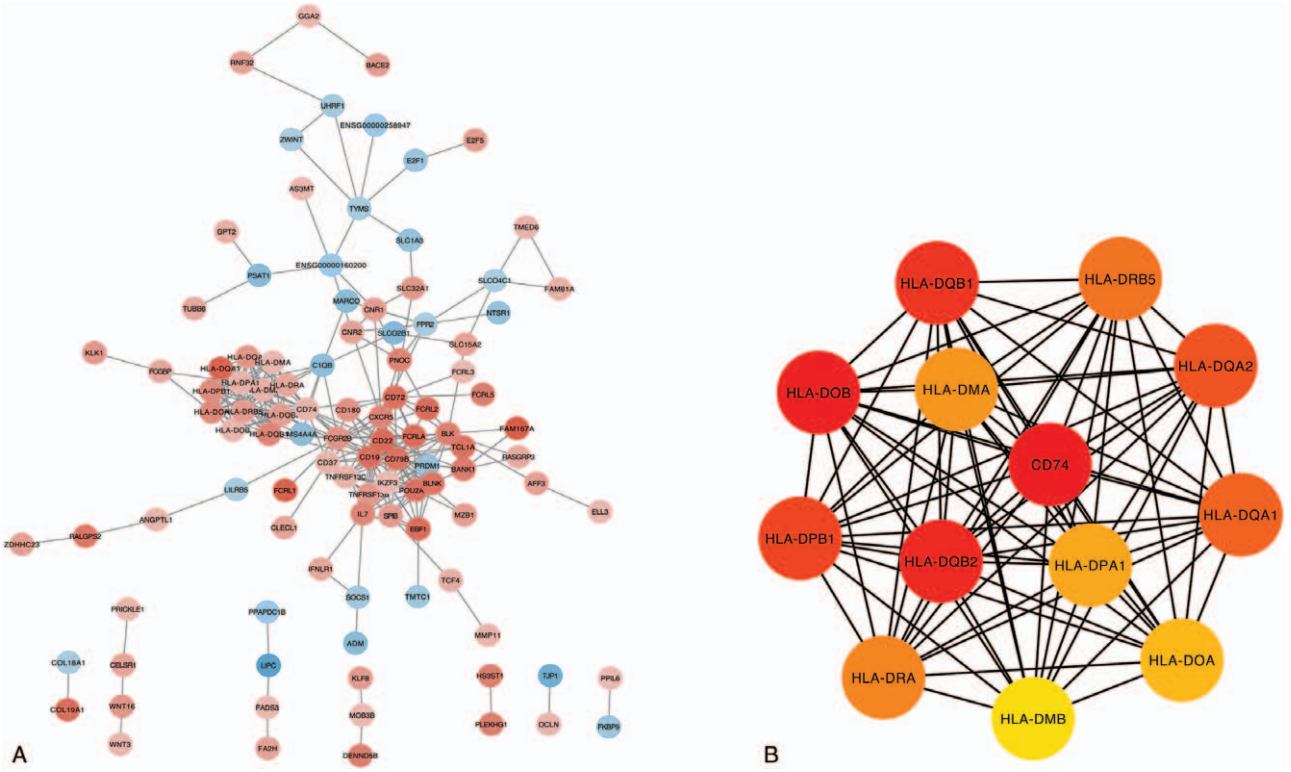


**Figure 6.** The PPI network of DEGs. (A) The global PPI network of DEGs. The red dots represent the upregulated genes while the blue dots represent the downregulated genes. The darker the dot, the more significant is. (B) The hub genes extracted from the global network by the plug-in MCODE. DEGs = differentially expressed genes, PPI = protein-protein interaction.

or genome involved in the pathogenesis of BPD may shed a light on diagnosis and treatment of it.

As mentioned above, the pathogenesis of BPD has a strong genetic background, therefore, in the recent decade, many researchers devote to explore the susceptibility genes even non-coding RNAs which may get involved in regulating the development of BPD using bioinformical methods. Bhattacharya et al demonstrated that the gene Ahr as one driver of gene dysregulation in the acute neonatal hyperoxia model of BPD-like pathology through high-throughput sequencing-based transcriptional profiling.[14] Another research performed by Nitkin. C. R, et al found that FOSL1 is a novel mediator of endotoxin/lipopolysaccharide-induced pulmonary angiogenic signaling which may result in BPD development.[15] In the present study, a total of 189 DEGs including 147 upregulated genes and 42 downregulated genes between BPD and non-BPD samples have

**Table 3**

**The names, abbreviations and main functions of the hub genes.**

| No. | Gene symbol | Full name | Function |
|---|---|---|---|
| 1 | CD74 | CD74 Molecule | Important chaperone that regulates antigen presentation for immune response. |
| 2 | HLA-DQB1 | Major Histocompatibility Complex, Class II, DQ Beta 1 | Plays a central role in the immune system by presenting peptides derived from extracellular proteins |
| 3 | HLA-DQA1 | Major Histocompatibility Complex, Class II, DQ Alpha 1 | Plays a central role in the immune system by presenting peptides derived from extracellular proteins |
| 4 | HLA-DPA1 | Major Histocompatibility Complex, Class II, DP Alpha 1 | Plays a central role in the immune system by presenting peptides derived from extracellular proteins |
| 5 | HLA-DQB2 | Major Histocompatibility Complex, Class II, DQ Beta 2 | Plays a central role in the immune system by presenting peptides derived from extracellular proteins |
| 6 | HLA-DRA | Major Histocompatibility Complex, Class II, DR Alpha | Plays a central role in the immune system by presenting peptides derived from extracellular proteins |
| 7 | HLA-DRB5 | Major Histocompatibility Complex, Class II, DR Beta 5 | Plays a central role in the immune system by presenting peptides derived from extracellular proteins |
| 8 | HLA-DPB1 | Major Histocompatibility Complex, Class II, DP Beta 1 | Plays a central role in the immune system by presenting peptides derived from extracellular proteins |
| 9 | HLA-DOB | Major Histocompatibility Complex, Class II, DO Beta | Suppresses peptide loading of MHC class II molecules by inhibiting HLA-DM |
| 10 | HLA-DOA | Major Histocompatibility Complex, Class II, DO Alpha | Regulates HLA-DM-mediated peptide loading on MHC class II molecules |
| 11 | HLA-DQA2 | Major Histocompatibility Complex, Class II, DQ Alpha 2 | Plays a central role in the peptide loading of MHC class II molecules by helping to release the CLIP molecule from the peptide binding site |
| 12 | HLA-DMB | Major Histocompatibility Complex, Class II, DM Beta | Plays a central role in the peptide loading of MHC class II molecules by helping to release the CLIP molecule from the peptide binding site |
| 13 | HLA-DMA | Major Histocompatibility Complex, Class II, DM Alpha | Plays a central role in the peptide loading of MHC class II molecules by helping to release the CLIP molecule from the peptide binding site |

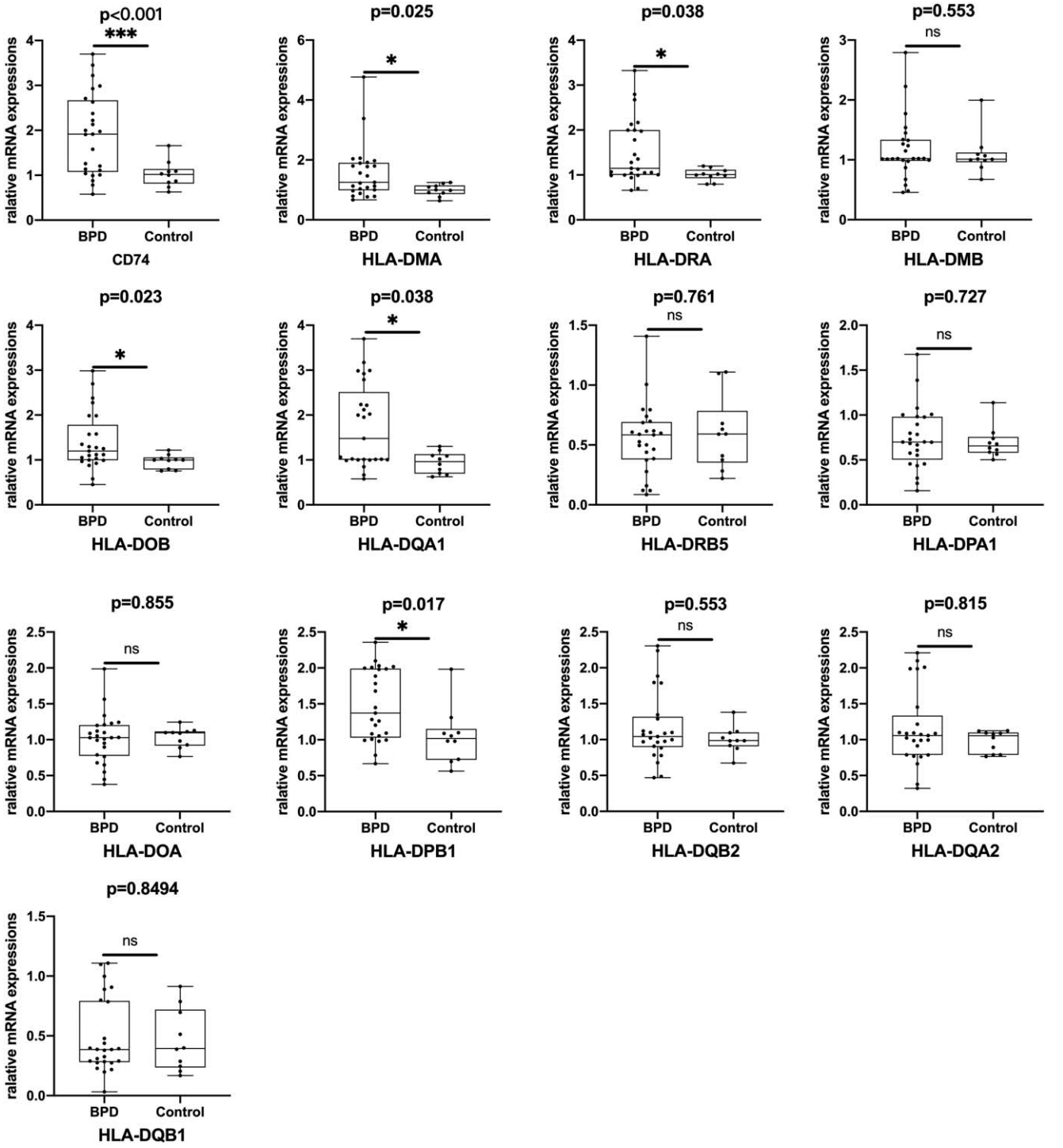


**Figure 7.** The expressions of 13 hub genes of 2 group samples by qRT-PCR verification. qRT-PCR = quantitative real-time PCR.

been identified. The pathway and function enrichments were assessed, the results summarized above revealed the DEGs were mainly enriched in immune system and response to viral infection. The pathogenesis of BPD is complicated and the inappropriate immunological reaction to inflammatory insult is a crucial cause. Those DEGs yielded from the enrichments mentioned above may probably be involved in the development of BPD. Furthermore, in terms of the PPI network we established, thirteen hub genes were identified to be probably associated with BPD pathogenesis. These genes belong to class II major histocompatibility complex (MHC) family regulating antigen presentation for immune response and the inappropriate immune response to inflammatory impairment occurring in the immature lungs is a critical factor of BPD development. However, up to now, researches focusing on the linkage between MHC II family and BPD pathogenesis are so rare. A research conducted in fetal

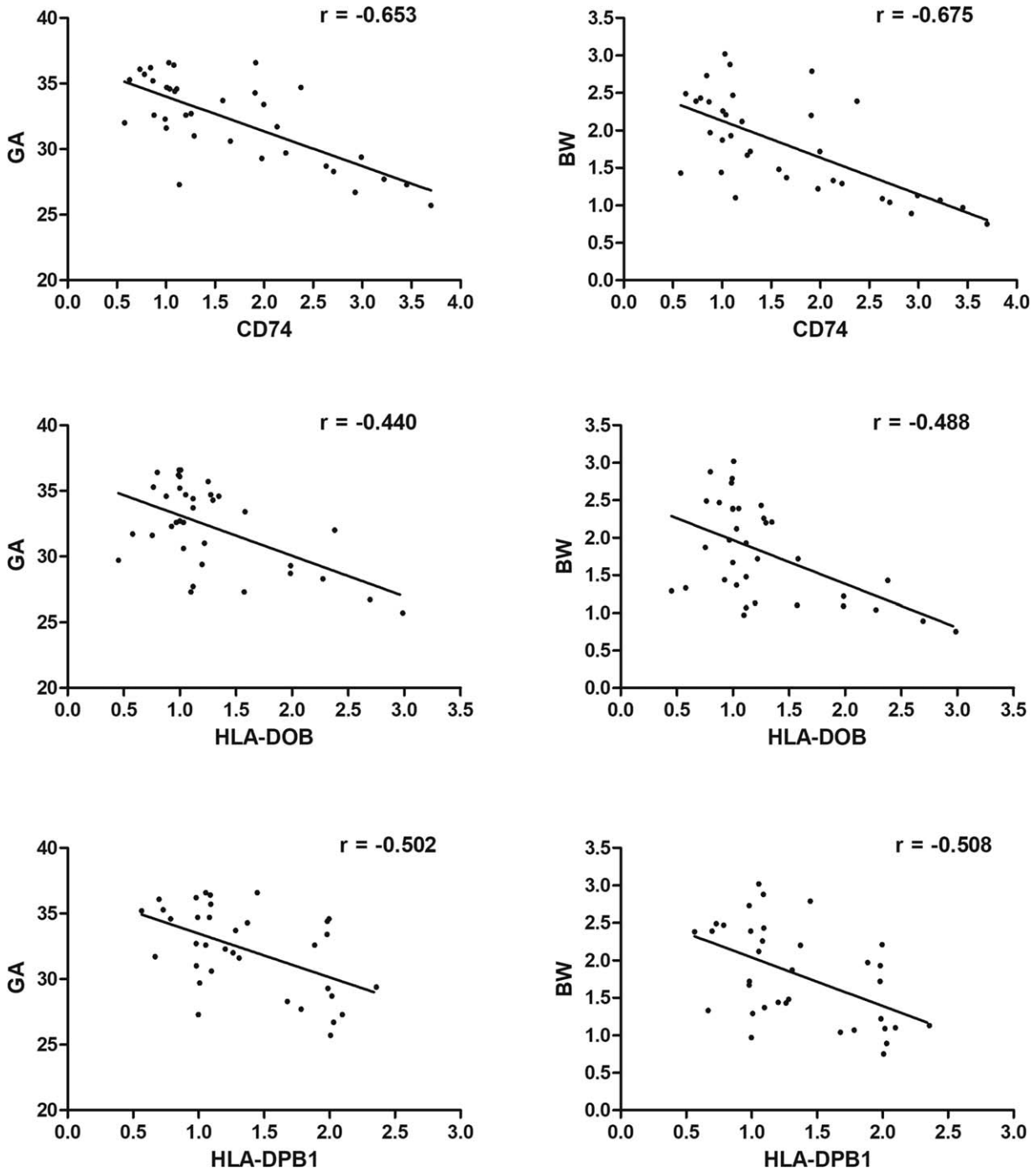


**Figure 8.** The linear analysis of hub genes associated with GA and BW. BW = birth weight, GA = gestational age.

lambs using *Ureaplasma parvum* serovar 3 to induce an inflammatory response in fetal lungs which mimic the pathology of BPD showing that 14 days after being exposed to *U parvum* serovar 3, the MHC class II-positive cells increased, although macrophage levels were not elevated.[16] This phenomenon indicates that the MHC II family genes may get involved in BPD development. In our work, among aforementioned 13 genes, CD74, HLA-DPB1, HLA-DOB, HLA-DMA, HLA-DRA, and HLA-DQA1 were the highest expressed with $P < .05$ in BPD samples comparing with non-BPD samples. By univariate analysis of the clinical features, the most significant risk factors of BPD occurrence were low birth weight and small gestational age. Through the correlation analyses of aforementioned 6 hub genes with BW and GA, the results showed that CD74, HLA-DOB, and HLA-DPB1 were negatively correlated with GA and BW with statistical significance. In addition, CD74 showed the strongest linkage with BW and GA with r coefficients of $-0.653$ and $-0.675$, respectively.

CD74 is a protein coding gene that associates with major histocompatibility complex class II (MHCII) regulating antigen

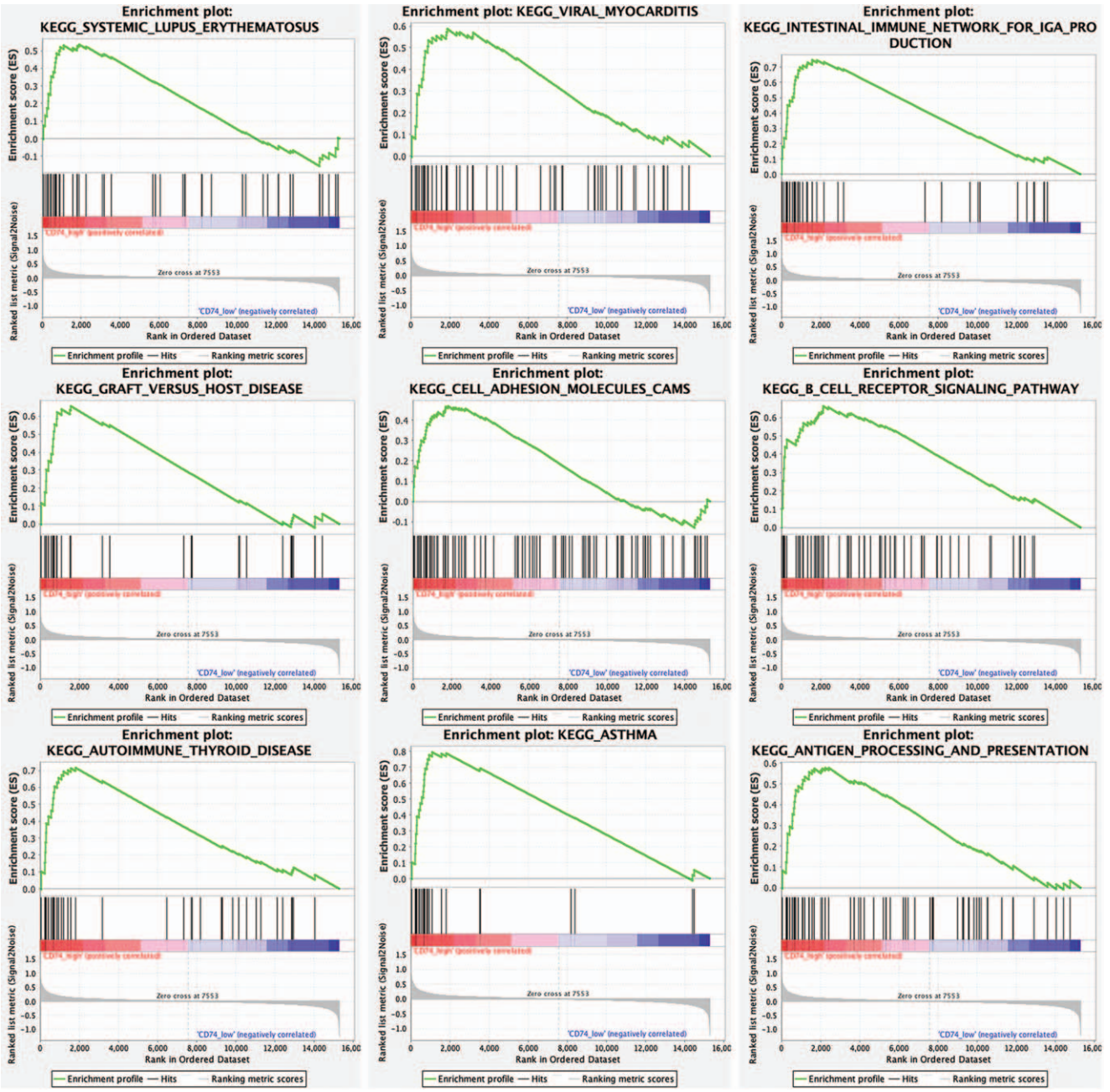


**Figure 9.** The GSEA analysis of the status when CD74 upregulated. GSEA = gene set enrichment analysis.

presentation for immune response.[17,18] Furthermore, the protein encoded by CD74 is a type of cell surface receptor for the cytokine macrophage migration inhibitory factor bounding, initiates survival pathways, and cell proliferation. This protein also interacts with amyloid precursor protein and suppresses the production of amyloid beta. CD74 also facilitates the export of class II MHC from the endoplasmic reticulum to the Golgi apparatus, and then to the late endosome compartment. During this stage, CD74 is broken down by proteases in stages called cathepsins, leaving only a small fragment known as CLIP (class II associated invariant chain peptide), which maintains blockage of the peptide binding cleft on the MHC molecule[19] (Fig. 10). A flood of previous researches regarding CD74 are mostly focused on oncology and diseases related to CD74 mainly are mucinous lung adenocarcinoma and undifferentiated pleomorphic sarcoma.[20–22] However, as a highly immune-related gene, it also participates the pathogenesis in many inflammatory diseases such as infectious diseases and rheumatic diseases.[23–25] In terms of the gene-disease network analysis demonstrated in Figure 5 showing that CD74 was linked to immunologic deficiency syndromes and autoimmune diseases. BPD pathogenesis, in turn, has been confirmed to be multifactorial and the inappropriate immune response to the pulmonary damage plays an essential role in it. To sum up, CD74 may be one of the critical genes that participate in the development of BPD through immunologic pathway. However, the gene-disease

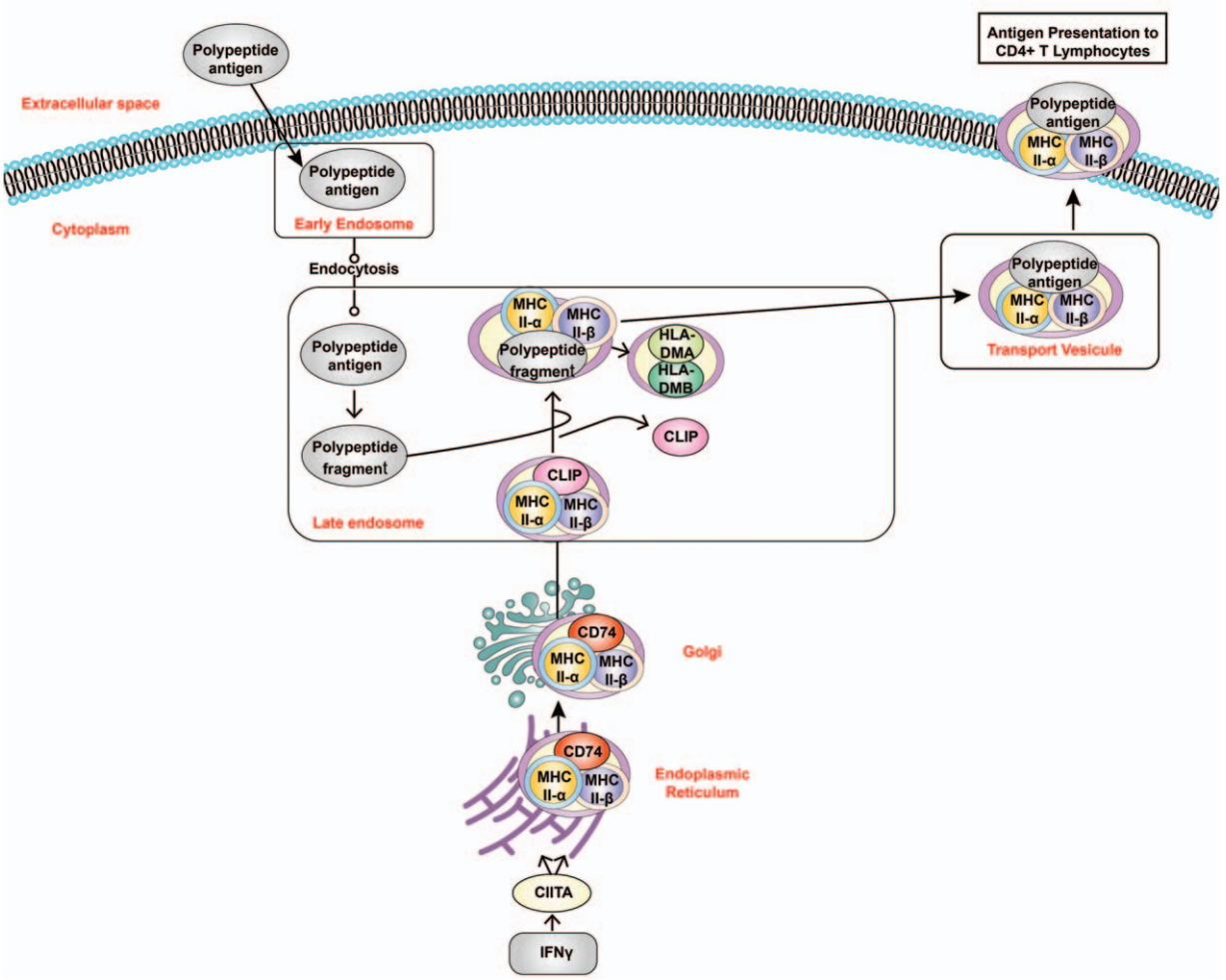


**Figure 10.** Major histocompatibility class II related molecules and antigen presentation process. The MHC class II network was generated through the use of IPA (Ingenuity Systems, www.ingenuity.com). MHC = major histocompatibility complex.

network analysis is a method that based on big data mining, the results yielded from which could only offer some kind of the potential relationships between genes and diseases, therefore the laboratorial research and verification are still needed. Cazalis et al found that decreased invariant chain/CD74 mRNA expression significantly predict 28-day mortality in septic shock patients alive at day 3.[26] BPD is a kind of inflammatory disease due to the improper damage-repairment of lung tissues. Thus, in terms of the so many significant roles that CD74 play, it is reasonable to believe that CD74 may also be important in BPD. In our clinical verifications and the correlation analyses, CD74 is demonstrated to be significantly high expressed in BPD patients. In addition, CD74 expression level is negatively correlated with BW and GA while both of BA and GA are well known as the risk factors of pathogenesis of BPD. Then, from the GSEA enrichment analysis focused on CD74, we can also see that many immune-related diseases were involved in the status of CD74 upregulation.

The other 2 genes, HLA-DOB and HLA-DPB1, are both members of HLA class II β chain paralogues. HLA class II molecule is a heterodimer consisting of an alpha (DPA) and a beta chain (DPB), both located on the membrane of the antigen presenting cells which presents peptides coming from extracellular proteins.[27–30] They are closely related to CD74 because MHC class II gene transcription is regulated by class II transactivator A and factors of the Regulatory Factor X family. During their synthesis, the α and β chains are complexed with the invariant chain (that is CD74) resulting the prevention of the binding of newly synthesized MHC II α/β heterodimers with the numerous polypeptides.[19]

HLA-DMA, HLA-DRA, and HLA-DQA1 all belong to the HLA class II alpha chain paralogues either. These class II molecules are heterodimer consisting of an alpha (DMA) and a beta chain (DMB), both anchored in the membrane. It is located in intracellular vesicles. DM plays a central role in the peptide loading of MHC class II molecules by helping to release the CLIP (which is released by CD74, mentioned before) molecule from the peptide binding site. Class II molecules are expressed in antigen presenting cells (APC: B lymphocytes, dendritic cells, macrophages). The alpha chain is approximately 33 to 35 kDa and its gene contains 5 exons. Exon one encodes the leader peptide, exons 2 and 3 encode the 2 extracellular domains, exon 4 encodes the transmembrane domain and the cytoplasmic tail.

In summary, among MHC class II family, CD74 serves as a bridge connecting other molecules, and such relationship between CD74 and other MHC II molecules were well demonstrated in Figure 10. Indeed, during PCR verification,

CD 74 showed the highest expression compared to others ($P < .001$).

To our knowledge, this is the first time to speculate and explore the relationship between CD74 and the pathogenesis of BPD. BPD is a chronic lung disease induced by inflammatory insult, which can lead to the impairment of the epithelial cells reside in the distal alveolar region of the lung. These cells are crucial that not only for respiration but also for its critical functions of the physical barriers and pulmonary host defenses. The damage to aforementioned epithelial barrier might be one of the most important pathogenesis of BPD. In a study regarding the neonatal necrotizing enterocolitis conducted by Laura Farr et al,[17] CD74 signaling is proved to be involved in intestinal epithelial cell regeneration and it also promotes mucosal healing. During inflammation, CD74 could be activated and protect the host by promoting epithelial cell regeneration, healing, and maintaining mucosal barrier integrity. In necrotizing enterocolitis patients, CD74 expressions were significant enhanced. The damage-repair mechanism of epithelial cells in lungs resembles that of intestinal epithelial cells. Therefore, it is reasonable to deduce that CD74 may also play an important role in BPD pathogenesis. Indeed, in present study, during qRT-PCR verification stage, CD74 was significantly higher expressed in BPD samples than that of non-BPD controls.

Taken together, CD74 may be served as a novel biomarker for BPD early evaluation and diagnosis. Even more, it may also be considered as a new therapeutic target of BPD.

There are also some limitations of our work that should be considered. Firstly, the sample size of the dataset we downloaded is too small, however, since the microarray profile of BPD in human samples are so rare, through the quality control, this dataset is the best one among others. Secondly, as a bioinformical study, we paid more attention on data mining and less on experimental verification. The sample size of clinical verification is small either. Based on the encouraging result of the present study, we will expand the sample size in the future, and further laboratorial verification will be performed.

In summary, our study has identified the common biological pathways involved in the progression of BPD through the integrated bioinformatics analysis, laboratorial experiments, and clinical research. CD74 is one of the most significant gene that highly upregulated in BPD and CD74 may have the potential for predicting the development of BPD in preterm infants. Our study may provide some innovative strategies for dealing with this disease.

## 5. Conclusion

CD74 may have the potential of predicting BPD development in preterm infants and may also be considered as a novel therapeutic target for this disease

## Author contributions

**Conceptualization:** Shan He.
**Data curation:** Mingfu Wu, Fudong Wang, Rui Tian.
**Formal analysis:** Junyan Gao, Rui Tian, Shan He.
**Funding acquisition:** Xueping Zhu.
**Investigation:** Fudong Wang.
**Resources:** Lijun Jiang, Xueping Zhu.
**Software:** Mingfu Wu, Xueping Zhu.
**Visualization:** Lijun Jiang, Shan He.
**Writing – original draft:** Junyan Gao.
**Writing – review & editing:** Shan He.

## References


[1] Kair LR, Leonard DT, Anderson JM. Bronchopulmonary dysplasia. Pediatr Rev 2012;33:255–63.
[2] Poindexter BB, Feng R, Schmidt B, et al. Comparisons and limitations of current definitions of bronchopulmonary dysplasia for the prematurity and respiratory outcomes program. Ann Am Thorac Soc 2015;12: 1822–30.
[3] Fanaroff AA, Stoll BJ, Wright LL, et al. Trends in neonatal morbidity and mortality for very low birthweight infants. Am J Obstet Gynecol 2007;196:147.e1–8.
[4] Bhandari V, Gruen JR, Jang KL, et al. Genetics of bronchopulmonary dysplasia: when things do not match up, it is only the beginning. J Pediatr 2019;208:298–9.
[5] Parker RA, Lindstrom DP, Cotton RB. Evidence from twin study implies possible genetic susceptibility to bronchopulmonary dysplasia. Semin Perinatol 1996;20:206–9.
[6] Bhandari V, Bizzarro MJ, Shetty A, et al. Familial and genetic susceptibility to major neonatal morbidities in preterm twins. Pediatrics 2006;117:1901–6.
[7] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7.
[8] Kanehisa M. The KEGG database. Novartis Found Symp 2002;247: 91–101.
[9] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
[10] Dennis GJr, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:3.
[11] Gong XQJ, Ma L, Qiu G, et al. Differentially expressed genes in blood of BPD and normal preterm neonates. Gene Expression Omnibus (GEO) Database 2018;GSE10854.
[12] Pinero J, Bravo A, Queralt-Rosinach N, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 2017;45:D833–9.
[13] Walsh MC, Szefler S, Davis J, et al. Summary proceedings from the bronchopulmonary dysplasia group. Pediatrics 2006;117:S52–6.
[14] Bhattacharya S, Zhou Z, Yee M, et al. The genome-wide transcriptional response to neonatal hyperoxia identifies Ahr as a key regulator. Am J Physiol Lung Cell Mol Physiol 2014;307:L516–523.
[15] Nitkin CR, Xia S, Menden H, et al. FOSL1 is a novel mediator of endotoxin/lipopolysaccharide-induced pulmonary angiogenic signaling. Sci Rep 2020;10:13143.
[16] Collins JJ, Kallapur SG, Knox CL, et al. Inflammation in fetal sheep from intra-amniotic injection of Ureaplasma parvum. Am J Physiol Lung Cell Mol Physiol 2010;299:L852–60.
[17] Farr L, Ghosh S, Jiang N, et al. CD74 signaling links inflammation to intestinal epithelial cell regeneration and promotes mucosal healing. Cell Mol Gastroenterol Hepatol 2020.
[18] Song H, Zhu Z, Zhou Y, et al. MIF/CD74 axis participates in inflammatory activation of Schwann cells following sciatic nerve injury. J Mol Histol 2019;50:355–67.
[19] Choi NM, Majumder P, Boss JM. Regulation of major histocompatibility complex class II genes. Curr Opin Immunol 2011;23:81–7.
[20] Klein S, Mauch C, Wagener-Ryczek S, et al. Immune-phenotyping of pleomorphic dermal sarcomas suggests this entity as a potential candidate for immunotherapy. Cancer Immunol Immunother 2019; 68:973–82.
[21] Murayama T, Nakaoku T, Enari M, et al. Oncogenic fusion gene CD74-NRG1 confers cancer stem cell-like properties in lung cancer through a IGF2 autocrine/paracrine circuit. Cancer Res 2016;76:974–83.
[22] Lim SM, Kim EY, Kim HR, et al. Genomic profiling of lung adenocarcinoma patients reveals therapeutic targets and confers clinical benefit when standard molecular testing is negative. Oncotarget 2016; 7:24172–8.
[23] Bozza MT, Lintomen L, Kitoko JZ, et al. The role of MIF on eosinophil biology and eosinophilic inflammation. Clin Rev Allergy Immunol 2020;58:15–24.
[24] Herrero LJ, Sheng KC, Jian P, et al. Macrophage migration inhibitory factor receptor CD74 mediates alphavirus-induced arthritis and myositis

in murine models of alphavirus infection. Arthritis Rheum 2013; 65:2724–36.
[25] Zaidi T, Reidy T, D'Ortona S, et al. CD74 deficiency ameliorates Pseudomonas aeruginosa-induced ocular infection. Sci Rep 2011; 1:58.
[26] Cazalis MA, Friggeri A, Cave L, et al. Decreased HLA-DR antigen-associated invariant chain (CD74) mRNA expression predicts mortality after septic shock. Crit Care 2013;17:R287.
[27] Yin L, Trenh P, Guce A, et al. Susceptibility to HLA-DM protein is determined by a dynamic conformation of major histocompatibility complex class II molecule bound with peptide. J Biol Chem 2014; 289:23449–64.
[28] Lenormand C, Bausinger H, Gross F, et al. HLA-DQA2 and HLA-DQB2 genes are specifically expressed in human Langerhans cells and encode a new HLA class II molecule. J Immunol 2012;188:3903–11.
[29] Hernandez-Guerrero JC, Jimenez-Farfan D, Castel-Rodriguez A, et al. Immunoexpression of the HLA-DR class II molecule in cells from inflamed dental pulp. Arch Med Res 2006;37:461–4.
[30] Douek DC, Altmann DM. HLA-DO is an intracellular class II molecule with distinctive thymic expression. Int Immunol 1997;9:355–64.