# Gene network in pulmonary tuberculosis based on bioinformatic analysis

**Lili Li**
Wuhan University Renmin Hospital https://orcid.org/0000-0003-3550-2820
**Jian Lv**
Wuhan University Renmin Hospital
**Yuan He**
Wuhan University Renmin Hospital
**Zhihua Wang** ( zhihuawang@whu.edu.cn )
https://orcid.org/0000-0002-1800-2384

**Research article**

**Keywords:** Pulmonary tuberculosis, Mycobacterium tuberculosis, Protein-Protein Interaction, Hub genes, Bioinformatic analysis

**Posted Date:** February 4th, 2020

**DOI:** https://doi.org/10.21203/rs.2.22571/v1



**Version of Record:** A version of this preprint was published on August 18th, 2020. See the published version at https://doi.org/10.1186/s12879-020-05335-6.

# Abstract

Background: Pulmonary tuberculosis (PTB) is one of the serious infectious diseases worldwide; however, the gene network involved in the host response remain largely unclear.

Methods: This study integrated two cohorts profile datasets GSE34608 and GSE83456 to elucidate the potential gene network and signaling pathways in PTB. Differentially expressed genes (DEGs) were obtained for Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis using Metascape database. Protein-Protein Interaction (PPI) network of DEGs was constructed by the online database the Search Tool for the Retrieval of Interacting Genes (STRING). Modules were identified by the plug-in APP Molecular Complex Detection (MCODE) in Cytoscape. GO and KEGG pathway of Module 1 were further analyzed by STRING. Hub genes were selected for further expression validation in dataset GSE19439. The gene expression level was also investigated in the dataset GSE31348 to display the change pattern during the PTB treatment.

Results: Totally, 180 shared DEGs were identified from two datasets. Gene function and KEGG pathway enrichment revealed that DEGs mainly enriched in defense response to other organism, response to bacterium, myeloid leukocyte activation, cytokine production, etc. Seven modules were clustered based on PPI network. Module 1 contained 35 genes related to cytokine associated functions, among which 14 genes, including chemokine receptors, interferon-induced proteins and Toll-like receptors, were identified as hub genes. Expression levels of the hub genes were validated with a third dataset GSE19439. The signature of this core gene network showed significant response to Mycobacterium tuberculosis (Mtb) infection, and correlated with the gene network pattern during anti-PTB therapy.

Conclusions: Our study unveils the coordination of causal genes during PTB infection, and provides a promising gene panel for PTB diagnosis. As major regulators of the host immune response to Mtb infection, the 14 hub genes are also potential molecular targets for developing PTB drugs.


# Background

Pulmonary tuberculosis (PTB) is one of the serious infectious disease that causes the most deaths in the world. It is estimated that one third of the world's population are infected with Mycobacterium tuberculosis (Mtb) as the latent infections, 5 to 10 percent of the patients with latent infections developed into active tuberculosis (TB) [1, 2]. According to the World Health Organization (WHO) report, there were 9 million new cases of PTB disease and 1.5 million deaths worldwide in 2013. TB remains one of the highest mortality infectious diseases. PTB is caused by various strains of mycobacteria, mostly observed in human is caused by Mtb. PTB is contagious with airborne infection. Once infected, Mtb primarily attacks the lungs. Most people who develop symptoms of a TB infection once they are infected. Roughly 10% of the latent infections can progress to active TB. Therefore, efficient treatment is of great importance factors to control the spread of PTB and reduce mortality. It needs more effort to clarify the molecular mechanism underlying PTB development and progression, and to find potential drugs and

diagnostic biomarkers. Although there are extensive studies on the mechanism in PTB progression, the molecular progress for PTB are still unclear. Due to high morbidity and mortality in PTB, identifying essential pathways in PTB would be critically important and highly demanded for the diagnosis, prevention and treatment of this disease [3, 4].

Gene expression analysis based on microarray technology is a gene detection technique that has been widely used for differentially expressed genes (DEGs) screening. With the wide application of gene chips, a large number of data have been produced, and most of the data have been deposited and stored in public databases. Integrating and re-analyzing these data can provide valuable clues for new research. Many microarray data profiling studies have been carried out on PTB in recently years [5]. With bioinformatic analysis, DEGs and functional pathways were identified [6]. However, these results are always limited or inconsistent due to tissue or sample heterogeneity in independent studies, or the results are generated from a single cohort study. Thus, no reliable biomarkers have been identified in PTB. However, the integrated bioinformatics methods combining with expression profiling data will be innovative and might solve the disadvantages.

In the present study, two microarray datasets GSE34608 and GSE83456 from human whole blood samples were downloaded, including 53 health controls and 79 PTB samples. In this study, we aimed to use different bioinformatics methods to identify the DEGs between the two kinds of samples. Based on the DEGs, we performed Gene Ontology, pathway enrichment, Protein-Protein Interaction (PPI) network construction to reveal the function of hub genes in PTB. Findings of this study may help to explore essential diagnostic signatures for PTB.

## Methods

# Gene expression microarray data acquisition

NCBI Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo) is a public functional genomics database with high throughput gene expression sequencing data and microarrays data. Two gene expression datasets GSE34608 [7] and GSE83456 [6], were downloaded from GEO. GSE34608 contained 8 PTB samples and 18 control samples, which is based on GPL6480 platform (Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F). The GSE83456 dataset contained 45 PTB tissue samples and 61 control samples. It is based on GPL10558 platform (Illumina HumanHT-12 V4.0 expression beadchip). Another two datasets GSE19439 and GSE31348 were used for hub gene validation. GSE19439 contained 12 health and 13 PTB samples were used as validation dataset [8]. GSE19439 is based on GPL6947 platform (Illumina HumanHT-12 V3.0 expression beadchip). GSE31348 contained 27 subjects (135 samples) in five time point: diagnosis, treatment for 1, 2, 4 and 26 weeks, which is based on GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) [9].

# Identification of DEGs

Based on the microarray platform annotation, probe sets were converted into the corresponding gene symbol for the following analysis. Probe sets without corresponding gene symbols were removed. The DEGs between control and PTB samples were analyzed using limma (linear models for microarray data) package in R. |logFC (fold change)| > 1 and adj. P-value < 0.05 were considered as statistically significant threshold for the DEGs selection of GSE34608. |logFC| > 0.585 and adj. P-value < 0.05 were considered as statistically significant threshold for the DEGs selection of GSE83456. The co-existed DEGs were identified by drawing the venn diagram of DEGs of GSE34608 and GSE83456.

# KEGG and GO enrichment analyses of DEGs

Metascape (a gene annotation & analysis resource; http://metascape.org/) is online gene functional annotation tool to provide a comprehensive set of biological information of genes and proteins [10]. To understand the function of DEGs, Gene Oncology (GO) analysis, including biological process (BP), cellular components (CC), molecular function (MF), and KEGG pathway enrichment were performed using Metascape.

# PPI network construction and module analysis

In the present study, the PPI network was predicted using the Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) (version 11.0) online database [11]. The cut off value for STRING analysis is 0.04. Analyzing the functional interactions between proteins may provide insights into the biological mechanisms of action. PPI network were further visualized and analyzed with Cytoscape (version 3.4.0) plug-in APP Molecular Complex Detection (MCODE), which is used for clustering a given network based on topology. The most critical modules in the PPI network could be identified. The genes in top one module was displayed in this study. The hub gene selection criteria were as follows: MCODE scores > 10, degree > 20, neighborhood connectivity > 10.

# Hub genes analysis

The GO function, pathway, and protein domains of the top module was analyzed using STRING. The expression level of hub genes were further investigated in datasets GSE19439 and GSE31348.

## Results

# Identification of DEGs

Gene expression profile of GSE34608 and GSE83456 were downloaded from GEO database. The microarray data GSE34608 contains 18 control and 8 PTB patients. The GSE83456 data contains 61 control samples and 45 PTB samples. PCA plots of both datasets indicate the distinction expression of control and PTB samples (Fig. 1. A and B). 2214 and 1025 DEGs were identified from GSE34608 and GSE83456 datasets, respectively (Fig. 2A). Venn diagram was demonstrated that 180 DEGs were obtained by merging 51 up regulated genes in both datasets, and 129 down regulated genes in both datasets (Fig. 2A).

# GO enrichment and KEGG pathway analyses

Candidate DEGs functional Gene Ontology (GO) and pathway enrichment analyses were performed with Metascape. It showed that changes of DEGs were significantly enriched in defense response to other organism, response to bacterium, myeloid leukocyte activation, cytokine production, positive regulation of defense responses, cytokine-mediated signaling pathway, interferon signaling, etc. (Fig. 2B).

The subset of representative terms of gene function analysis were converted into a network layout in Metascape, as shown in Fig. 2C. Based on gene function analysis, all the significant terms were hierarchically clustered into a tree based on Kappa-statistical similarities. Each term is represented by a circle node, where its size is proportional to the number of input genes fall into that term. The color represents its cluster identity (Fig. 2C). Terms with a kappa score > 0.3 are linked by an edge. The statistically significant range of the node is marked by color range (Fig. 2C).

# PPI network construction and module analysis

The PPI network of 180 DEGs was constructed using the STRING online database, and further analyzed using app MCODE in Cytoscape software. Totally, seven modules were identified shown in Table 1. Module 1 from the PPI network complex containing 35 genes, indicating the core functional gene panel. GO analysis of 35 genes showed that the functions related to defense response and cytokine related pathway (Fig. 3A). PPI network of module 1 was redrawn by STRING (Fig. 3B). The expression level of 35 genes in dataset GSE34608 were shown in Fig. 3C. Genes CD27, CCR7, CD19, and CXCR3 were significantly down expressed in PTB samples, where other genes were up expressed. This result was consistent with the gene expression levels in GSE83456 (Fig. 3D). Furthermore, function analysis from STRING database were shown in Table 2. GO function was significant related to defense response and immune system response (biological process function), chemokine binding and chemokine receptor activity (molecular function), and external side of plasma membrane and cell surface (cellular component). KEGG and Reactome pathways indicated these 35 genes were significantly regulated in cytokine-cytokine receptor interaction, Toll-like receptor signaling pathway, immune system, and cytokine signaling in immune system. TIR domain, leucine rich repeat, and chemokine receptor family were the three important features revealed by PFAM and INTERPRO protein domains analysis (Table 2).

Table 1

Seven modules were identified by MCODE based on the 180 DEGs.

| Module | Gene symbol |
|---|---|
| Module 1 | TLR2 IL1B TLR8 TLR1 IFIH1 TLR5 IFIT1 CD19 IFIT2 CCR7 MPO CXCR3 IFI44 DDX60 FCGR2A CD163 IFI44L GBP2 TNFSF10 CD274 CCR2 XAF1 IFI16 IFITM1 IDO1 HERC5 SAMD9L EIF2AK2 RTP4 CCR1 CD27 PLSCR1 TNFSF13B PARP9 EPSTI1 |
| Module 2 | CXCL10 GBP5 ELANE AIM2 LCN2 DEFA4 HP NLRC4 MMP8 LTF TCN1 HPSE |
| Module 3 | STAT1 LCK FAS |
| Module 4 | CEACAM1 GPR84 BST1 |
| Module 5 | FAM26F SPPL2A USP25 |
| Module 6 | DBP TLE2 AES |
| Module 7 | CAMP CEACAM8 S100A12 RNASE3 |

Table 2

Function analysis of the 35 genes in module 1.

| Term | Description | Count | FDR |
|---|---|---|---|
| Biological Process (GO) | | | |
| GO:0006952 | Defense response | 28 of 1234 | 2.39E-24 |
| GO:0002376 | Immune system process | 31 of 2370 | 1.2E-21 |
| GO:0051707 | Response to other organism | 22 of 835 | 4.56E-19 |
| GO:0006955 | Immune response | 25 of 1560 | 9.48E-18 |
| GO:0002682 | Regulation of immune system process | 24 of 1391 | 1.52E-17 |
| Molecular Function (GO) | | | |
| GO:0019956 | Chemokine binding | 4 of 22 | 0.0000233 |
| GO:0004950 | Chemokine receptor activity | 4 of 27 | 0.0000243 |
| GO:0019957 | C-C chemokine binding | 3 of 11 | 0.000088 |
| GO:0016493 | C-C chemokine receptor activity | 3 of 13 | 0.00011 |
| GO:0003725 | Double-stranded RNA binding | 4 of 70 | 0.00028 |
| Cellular Component (GO) | | | |
| GO:0009897 | External side of plasma membrane | 9 of 223 | 2.93E-08 |
| GO:0009986 | Cell surface | 11 of 690 | 0.0000015 |
| GO:0035354 | Toll-like receptor 1-Toll-like receptor 2 protein complex | 2 of 2 | 0.00064 |
| GO:0005887 | Integral component of plasma membrane | 11 of 1564 | 0.0017 |
| GO:0044459 | Plasma membrane part | 13 of 2651 | 0.0083 |
| KEGG Pathways | | | |
| hsa04060 | Cytokine-cytokine receptor interaction | 8 of 263 | 0.00000133 |
| hsa04620 | Toll-like receptor signaling pathway | 5 of 102 | 0.0000409 |
| hsa05162 | Measles | 5 of 133 | 0.0000955 |
| hsa05168 | Herpes simplex infection | 5 of 181 | 0.00031 |
| hsa05134 | Legionellosis | 3 of 54 | 0.0019 |
| Reactome Pathways | | | |
| HSA-168256 | Immune system | 22 of 1925 | 3.84E-12 |

| Term | Description | Count | FDR |
|---|---|---|---|
| HSA-1280215 | Cytokine signaling in immune system | 12 of 654 | 5.29E-08 |
| HSA-913531 | Interferon signaling | 7 of 189 | 0.00000192 |
| HSA-168249 | Innate immune system | 12 of 1012 | 0.0000032 |
| HSA-909733 | Interferon alpha/beta signaling | 5 of 66 | 0.0000039 |
| PFAM Protein Domains | | | |
| PF01582 | TIR domain | 4 of 22 | 0.0000057 |
| PF13855 | Leucine rich repeat | 4 of 187 | 0.0051 |
| PF13676 | TIR domain | 2 of 11 | 0.0051 |
| PF13306 | Leucine rich repeats (6 copies) | 3 of 88 | 0.0051 |
| PF01463 | Leucine rich repeat C-terminal domain | 2 of 12 | 0.0051 |
| INTERPRO Protein Domains and Features | | | |
| IPR035897 | TIR domain superfamily | 4 of 26 | 0.00000766 |
| IPR000355 | Chemokine receptor family | 4 of 21 | 0.00000766 |
| IPR000157 | TIR domain | 4 of 22 | 0.00000766 |
| IPR024644 | Interferon-induced protein 44 family | 2 of 2 | 0.00031 |
| IPR000483 | Cysteine-rich flanking region, C-terminal | 4 of 83 | 0.00031 |

# Hub genes analysis

A total of 14 genes were selected as hub genes based on criteria MCODE (scores > 10, degree > 20, neighborhood connectivity > 10) in Table 3. All the hub genes were belonging to the module 1. These hub genes were significantly associated with Toll-like receptors, interferon-induce proteins, and chemokine receptors (Table 3). Among them, two genes were upregulated, whereas as others were downregulated (Fig. 3C). The expression levels were further validated in dataset GSE19439 (Fig. 4). The expression levels were also significantly different between health and PTB patients, expect gene CD19 and CXCR3 (Fig. 4).

Table 3

The functions of 14 hub genes.

| Gene | Score* | Full name | Function# |
|---|---|---|---|
| CD163 | 12 | Scavenger receptor cysteine-rich type 1 protein M130 | Acute phase-regulated receptor involved in clearance and endocytosis of hemoglobin/haptoglobin complexes by macrophages and may thereby protect tissues from free hemoglobin-mediated oxidative damage. |
| TLR5 | 11.02941 | Toll-like receptor 5 | Participates in the innate immune response to microbial agents. |
| IFIT1 | 10.82353 | Interferon-induced protein with tetratricopeptide repeats 1 | Interferon-induced antiviral RNA-binding protein that specifically binds single-stranded RNA bearing a 5'-triphosphate group (PPP-RNA), thereby acting as a sensor of viral single- stranded RNAs and inhibiting expression of viral messenger RNAs. |
| IFIT2 | 10.82353 | Interferon-induced protein with tetratricopeptide repeats 2 | IFN-induced antiviral protein which inhibits expression of viral messenger RNAs lacking 2'-O-methylation of the 5' cap. |
| CCR7 | 10.7451 | C-C chemokine receptor type 7 | Receptor for the MIP-3-beta chemokine. Belongs to the G-protein coupled receptor 1 family. |
| CXCR3 | 11.2 | C-X-C chemokine receptor type 3 | Isoform 1- Receptor for the C-X-C chemokine CXCL9, CXCL10 and CXCL11 and mediates the proliferation, survival and angiogenic activity of human mesangial cells (HMC) through a heterotrimeric G-protein signaling pathway. |
| TLR1 | 10.7451 | Toll-like receptor 1 | Participates in the innate immune response to microbial agents. Specifically recognizes diacylated and triacylated lipopeptides. |
| FCGR2A | 10.63736 | Low affinity immunoglobulin gamma Fc region receptor II-a | Binds to the Fc region of immunoglobulins gamma. |
| CD19 | 10.7451 | B-lymphocyte antigen CD19 | Assembles with the antigen receptor of B-lymphocytes in order to decrease the threshold for antigen receptor-dependent stimulation. |
| IFIH1 | 10.82353 | Interferon-induced helicase C domain-containing protein 1 | Innate immune receptor which acts as a cytoplasmic sensor of viral nucleic acids and plays a major role in sensing viral infection and in the activation of a cascade of antiviral responses including the induction of type I interferons and proinflammatory cytokines. |

Score*: indicated MCODE score. Function#: obtained from NCBI (https://www.ncbi.nlm.nih.gov/).

| Gene | Score* | Full name | Function# |
|---|---|---|---|
| IFI44L | 10.82353 | Interferon-induced protein 44-like | Exhibits a low antiviral activity against hepatitis C virus. |
| TLR8 | 10.7451 | Toll-like receptor 8 | Key component of innate and adaptive immunity. |
| IFI44 | 10.82353 | Interferon-induced protein 44 | This protein aggregates to form microtubular structures. |
| TLR2 | 10.7451 | Toll-like receptor 2 | Cooperates with LY96 to mediate the innate immune response to bacterial lipoproteins and other microbial cell wall components. |

Score*: indicated MCODE score. Function#: obtained from NCBI (https://www.ncbi.nlm.nih.gov/).

# Gene expression level detection during PTB treatment

To figure out the expression level changes during PTB treatment process, dataset GSE31348 were used to evaluate the change level. GSE31348 contained the 27 PTB patients, 135 samples were taking from diagnosis, 1-, 2-, 4- and 26-weeks treatment. Heatmap showed that the expression level of genes related with the functions (Fig. 5A). The expression level of CCR7, CD19, and CXCR3 were significantly increased, whereas the expression level of Interferon-induced proteins, Toll-like receptors were decreased during the treatment (Fig. 5A). Among these 14 genes, the expression level of CXCR3 were significantly increased, and TLR2 and TLR5 were significantly decreased during the PTB treatment (Fig. 5B). These three genes might have a potential to evaluate PTB as a gene panel.

# Discussion

In this study, we co-analyzed two GEO datasets GSE34608 and GSE83456 to identify hub genes related with PTB disease. Totally, 180 DEGs were obtained by combining these two GEO datasets (Fig. 2). With STRING protein-protein interaction data, 14 hub genes were identified (Figs. 3 and 4). The function of these 14 hub genes were chemokine receptors, Interferon-induced proteins, and Toll-like receptors (Tables 2 and 3). The signature of hub genes showed significant correlation with PTB, as well as patterns of correlation during the time course of anti-PTB therapy (Fig. 5).

A key finding of this study is that a universal core gene exists to functionally response Mtb infection. The common host response profile exist in PTB patients revealed also by transcriptional responses during Mtb infection [12]. Our study showed that chemokine receptors, Interferon-induced proteins, and Toll-like receptors were involved in this core response profiling, and significantly changed following successful treatment.

Chemokines play a major role to response Mtb infection as they appear to be controlled by the formation and maintenance of quiescent granulomas and establishment of the TB granuloma. High concentrations of cytokines and chemokines were required for early protection against Mtb infections [13]. Plasma cytokines also could be biomarkers of disease severity, and relieve mycobacterial burden in PTB [14]. Our results consistent with report that PTB individuals displayed significantly elevated levels of CCL1, CCL3, CXCL1, CXCL10 and CXCL11 as bacterial burdens dependent. The chemokines were significantly reduced following successful treatment [14].

It has now been established that the activation of the type I interferon response pathway is crucial in the defense against viral pathogens but detrimental infection with mycobacteria [15]. Thus although the signaling axis through this pathway is identical regardless of the type of infection, the outcome is substantially different, suggesting that the type-1 IFNs and the related IFN-inducible genes are able to create a favorable or unfavorable intracellular milieu to promote or disrupt survival of the invading pathogen [16]. Previous study confirmed that five genes (IFN-γ, TNF-α, IL-2R, CXCL9, and CXCL10) could be used for the detection of Mtb infection, including active PTB disease and LTB. The sensitivity of each gene was above to 80% [17]. The gene network profile provided more potential for diagnostic biomarkers selection. The transcriptional signatures revealed in this study should be further tested in other infection disease, to figure out the specific signature related of PTB infection or shared signature profiling related to other infections.

The essential role of Toll-like receptors against mycobacterial infection has been revealed in vivo. Toll-like receptors play key roles in innate and adaptive immunity against Mtb, and are involved in the recognition of conserved microbial structures, leading to activation of an inflammatory response. Previous study showed that TLR3 and TLR5 were up regulated at the first 24, 48 and 72 h post infection of the A549 pulmonary epithelial cell line induced by Mtb [18], and the expression of neutrophil TLR2 is also increased in PTB patients [19]. Whole blood from patients had increased mRNA levels of TLR1, and TLR2 [20]. TLR2-deficient mice showed increased subsequent progression to PTB disease, the rapid death and higher Mtb burden [21]. TLR2 may function as a regulator of inflammation, and in its absence an exaggerated immune inflammatory response develops. TLR1 rs5743551 and rs5743618 polymorphisms significantly increased under the M. tb infection in 203 PTB patients, compared to 203 healthy subjects [21]. TLR8 polymorphisms rs3764879 and rs3764880 associated between specific TLR polymorphisms and disease caused by specific strains [22].

# Conclusion

In summary, we construct a refined gene network representing the transcriptome signature in response to Mtb infection and its treatment. The identified 14 hub genes are promising biomarkers for developing transcriptome-based PTB diagnostic or prognostic tests. As major regulators of the host immune response to Mtb infection, these genes are also potential molecular targets for developing PTB drugs.

# Declarations

- Ethics approval and consent to participate

This study conformed to ethical instructions of Renmin Hospital of Wuhan University.

- Consent to publish

All authors have read the manuscript and consent to its publication.

- Availability of data and materials

Datasets including GSE34608, GSE83456, GSE19439 and GSE31348 were downloaded from NCBI Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo).

- Competing interests

There is no confliction of interests.

- Funding

The study was supported by the National Natural Science Foundation of China (No. 81722007) and National Health Commission of China (No. 2017ZX10304402001-008).

- Authors' Contributions

Z.W. and L.L. conceived and designed the project. L.L., J.L., and Y.H. performed bioinformatics analyses. L.L. drafted the manuscript. All authors contributed to text revision and discussion.

- Acknowledgements

Not applicable.

- Authors' Information

L.L., J.L., Y.H., Z.W..

Central Laboratory, Renmin Hospital of Wuhan University, 95 Zhangzhidong Rd. Wuhan 430060, China

Z.W.

Department of Cardiology, Renmin Hospital of Wuhan University, 95 Zhangzhidong Rd. Wuhan 430060, China

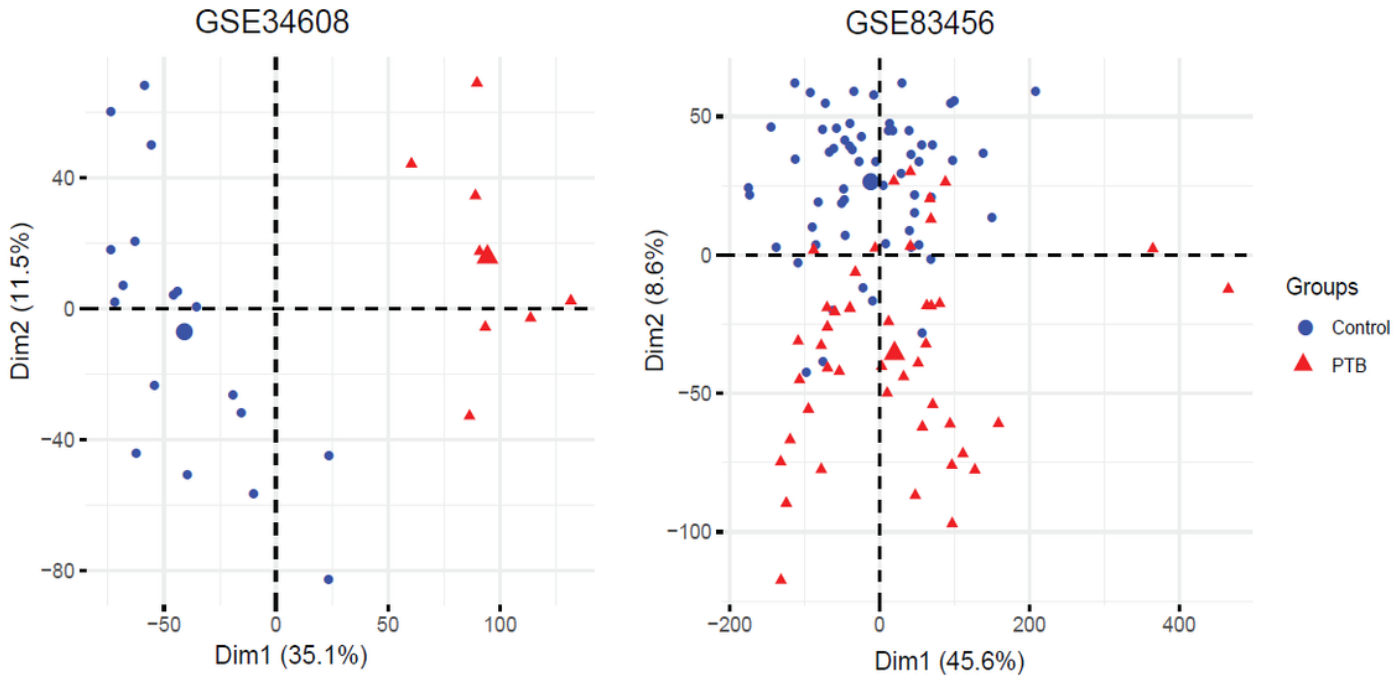
# References

1. Long B, Liang SY, Koyfman A, Gottlieb M: Tuberculosis: a focused review for the emergency medicine clinician. *Am J Emerg Med* 2019.

2. Tahan TT, Gabardo BMA, Rossoni AMO: Tuberculosis in childhood and adolescence: a view from different perspectives. *J Pediatr (Rio J)* 2019.
3. Yong YK, Tan HY, Saeidi A, Wong WF, Vignesh R, Velu V, Eri R, Larsson M, Shankar EM: Immune Biomarkers for Diagnosis and Treatment Monitoring of Tuberculosis: Current Developments and Future Prospects. *Front Microbiol* 2019, 10:2789.
4. Walzl G, McNerney R, du Plessis N, Bates M, McHugh TD, Chegou NN, Zumla A: Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *Lancet Infect Dis* 2018, 18(7):e199-e210.
5. Qin XB, Zhang WJ, Zou L, Huang PJ, Sun BJ: Identification potential biomarkers in pulmonary tuberculosis and latent infection based on bioinformatics analysis. *BMC Infect Dis* 2016, 16(1):500.
6. Blankley S, Graham CM, Turner J, Berry MP, Bloom CI, Xu Z, Pascual V, Banchereau J, Chaussabel D, Breen R *et al*: The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. *PLoS One* 2016, 11(10):e0162220.
7. Maertzdorf J, Weiner J, 3rd, Mollenkopf HJ, Network TB, Bauer T, Prasse A, Muller-Quernheim J, Kaufmann SH: Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci U S A* 2012, 109(20):7853-7858.
8. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ *et al*: An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010, 466(7309):973-977.
9. Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG, Ronacher K, King EC, Lukey PT, Duncan K, Van Helden PD *et al*: Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J Infect Dis* 2013, 207(1):18-29.
10. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019, 10(1):1523.
11. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017, 45(D1):D362-D368.
12. Shi C, Pamer EG: Monocyte recruitment during infection and inflammation. *Nat Rev Immunol* 2011, 11(11):762-774.
13. Slight SR, Khader SA: Chemokines shape the immune responses to tuberculosis. *Cytokine Growth Factor Rev* 2013, 24(2):105-113.
14. Kumar NP, Moideen K, Banurekha VV, Nair D, Babu S: Plasma Proinflammatory Cytokines Are Markers of Disease Severity and Bacterial Burden in Pulmonary Tuberculosis. *Open Forum Infect Dis* 2019, 6(7):ofz257.
15. McNab F, Mayer-Barber K, Sher A, Wack A, O'Garra A: Type I interferons in infectious disease. *Nat Rev Immunol* 2015, 15(2):87-103.

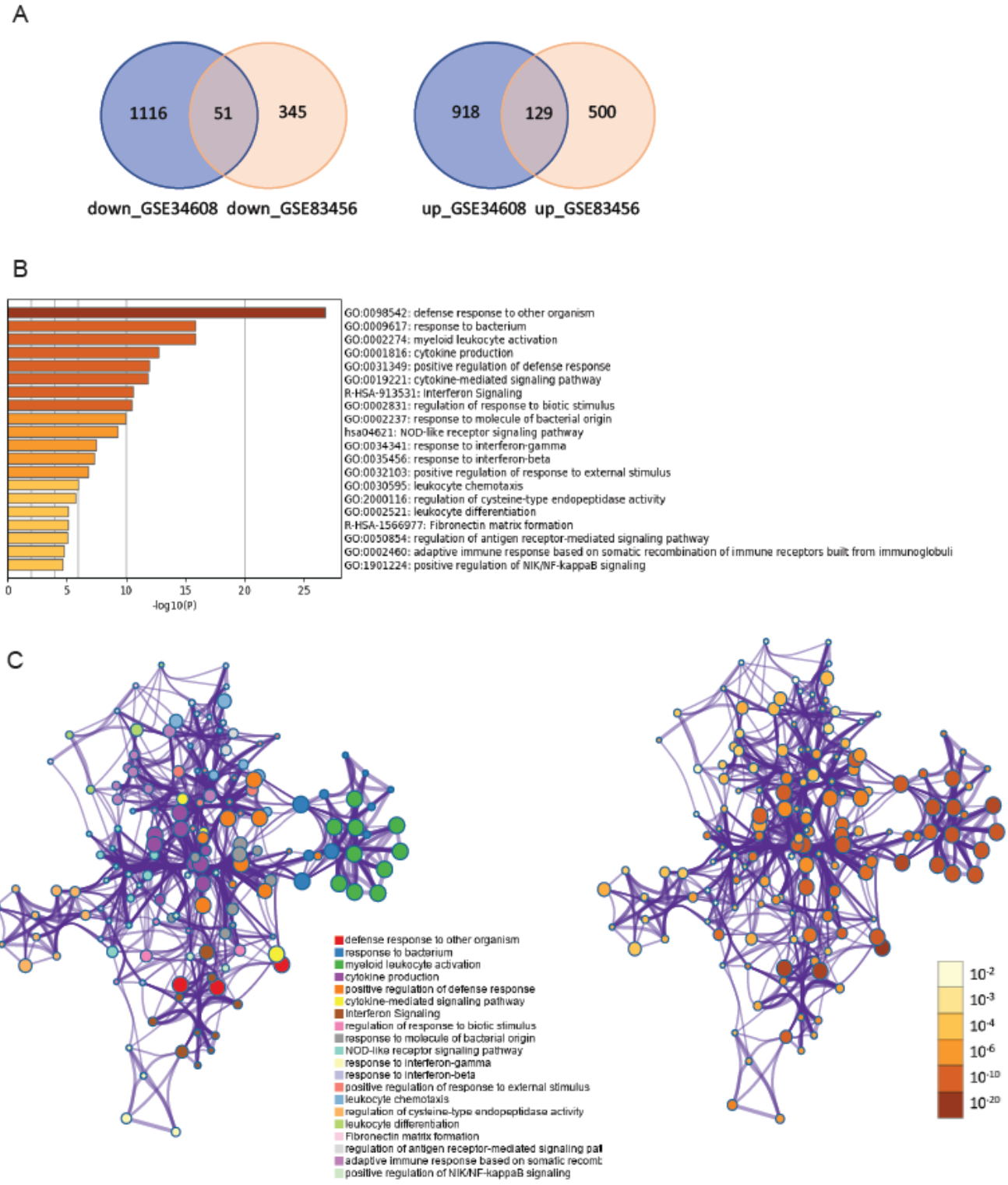
16. Desvignes L, Wolf AJ, Ernst JD: Dynamic roles of type I and type II IFNs in early infection with Mycobacterium tuberculosis. *J Immunol* 2012, 188(12):6205-6215.
17. Kim S, Lee H, Kim H, Kim Y, Cho JE, Jin H, Kim DY, Ha SJ, Kang YA, Cho SN *et al*: Diagnostic performance of a cytokine and IFN-gamma-induced chemokine mRNA assay after Mycobacterium tuberculosis-specific antigen stimulation in whole blood from infected individuals. *J Mol Diagn* 2015, 17(1):90-99.
18. Mvubu NE, Pillay B, McKinnon LR, Pillay M: Mycobacterium tuberculosis strains induce strain-specific cytokine and chemokine response in pulmonary epithelial cells. *Cytokine* 2018, 104:53-64.
19. Nancy Hilda J, Das S: Neutrophil CD64, TLR2 and TLR4 expression increases but phagocytic potential decreases during tuberculosis. *Tuberculosis (Edinb)* 2018, 111:135-142.
20. Chang JS, Huggett JF, Dheda K, Kim LU, Zumla A, Rook GA: Myobacterium tuberculosis induces selective up-regulation of TLRs in the mononuclear leukocytes of patients with active pulmonary tuberculosis. *J Immunol* 2006, 176(5):3010-3018.
21. Drennan MB, Nicolle D, Quesniaux VJ, Jacobs M, Allie N, Mpagi J, Fremond C, Wagner H, Kirschning C, Ryffel B: Toll-like receptor 2-deficient mice succumb to Mycobacterium tuberculosis infection. *Am J Pathol* 2004, 164(1):49-57.
22. Salie M, Daya M, Lucas LA, Warren RM, van der Spuy GD, van Helden PD, Hoal EG, Moller M: Association of toll-like receptors with susceptibility to tuberculosis suggests sex-specific effects of TLR8 polymorphisms. *Infect Genet Evol* 2015, 34:221-229.

# Figures



## Figure 1

Principal component analysis (PCA) to discriminate the gene expression levels between control and pulmonary tuberculosis (PTB). (A) PCA plot of dataset GSE34608. (B) PCA plot of dataset GSE83456.




## Figure 2

Selection and function of differentially expressed genes (DEGs). (A) Venn diagram of DEGs from the two datasets: GSE19439 and GSE34608. Totally 51 up-regulated and 129 down-regulated genes are shared between the two GSE datasets. (B and C) Functional annotation of DEGs using Metascape. The top 20 terms are displayed as bar plot based on P value (log10 scale) (B), and the network plot (C).

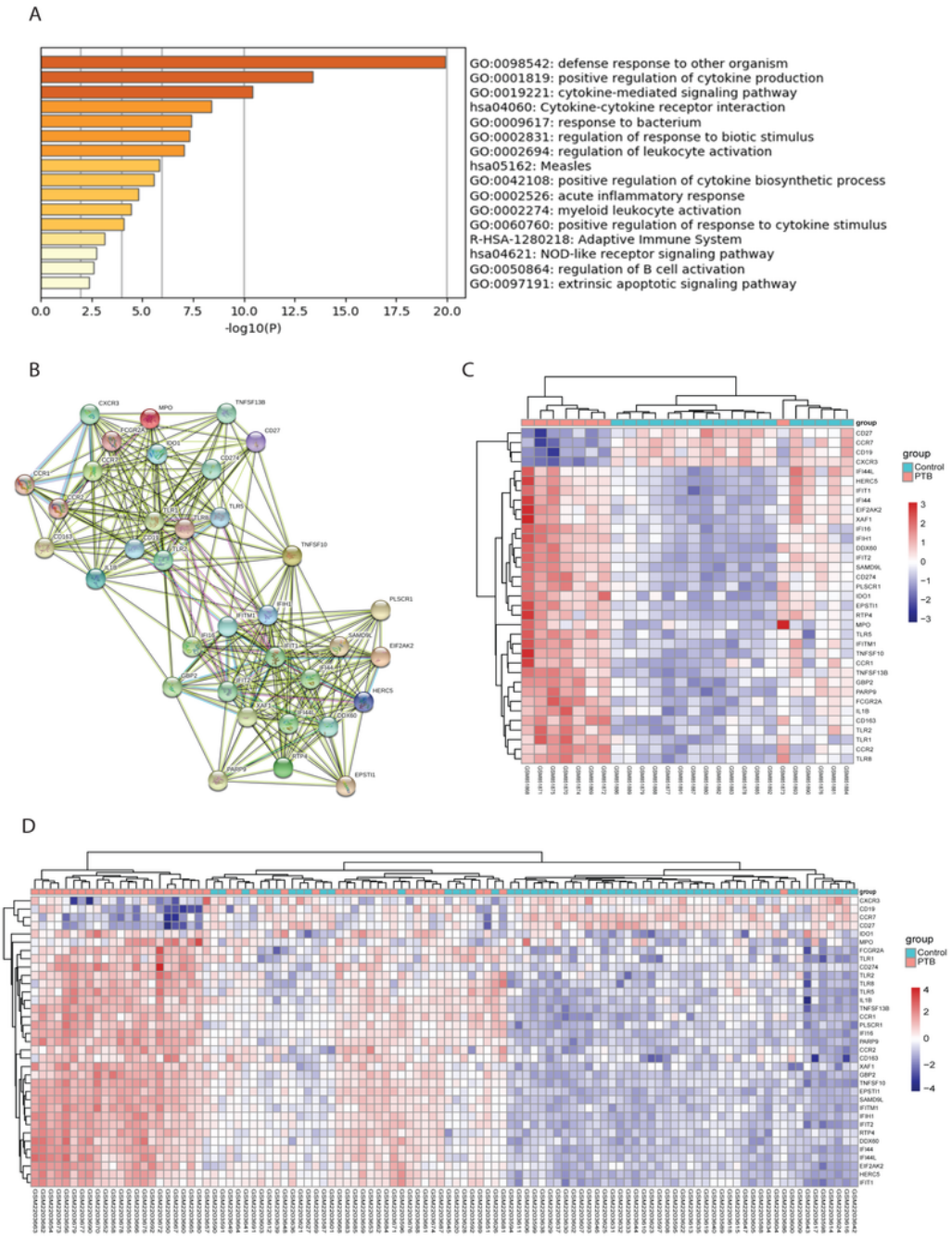
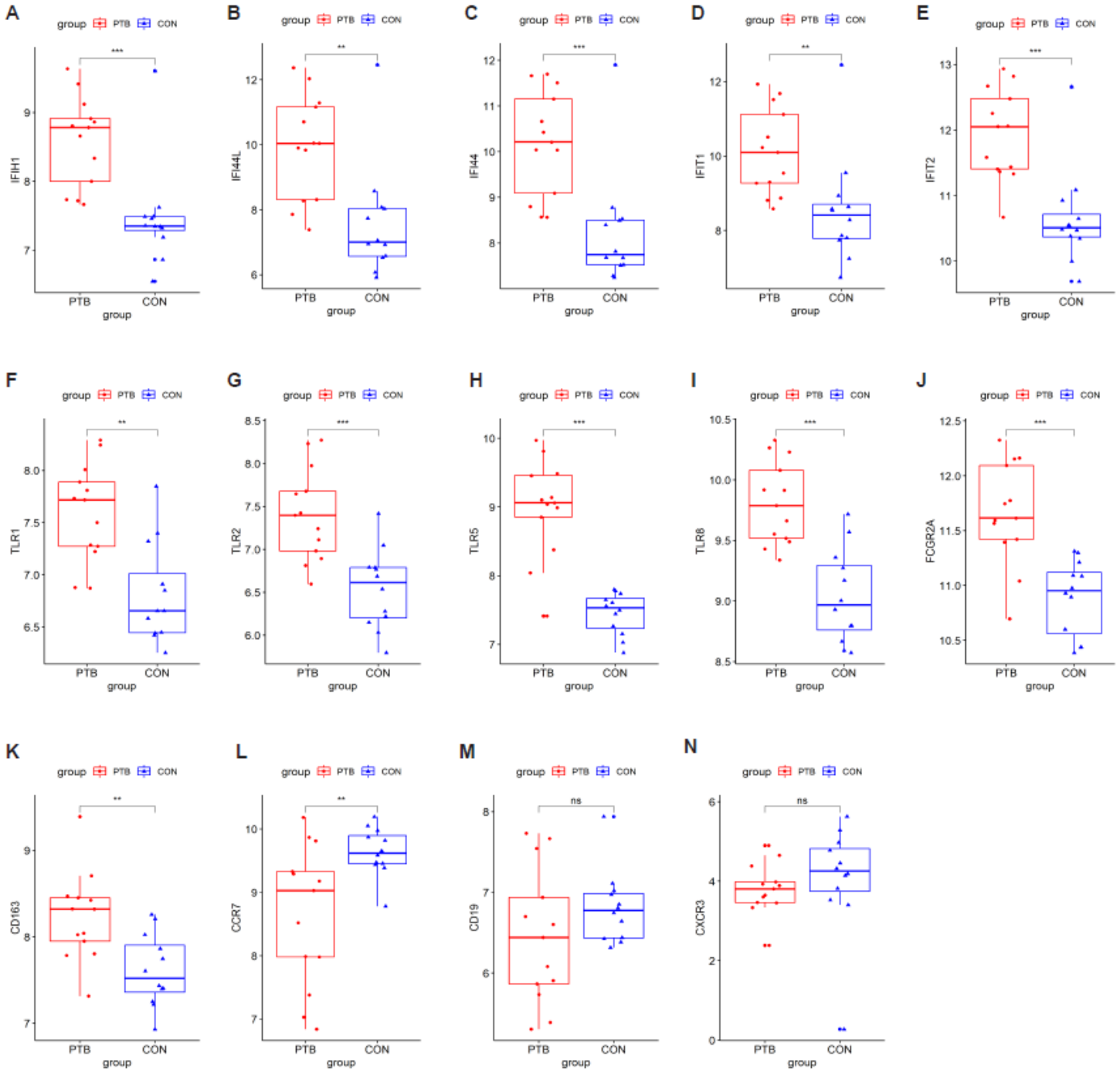



**Figure 3**

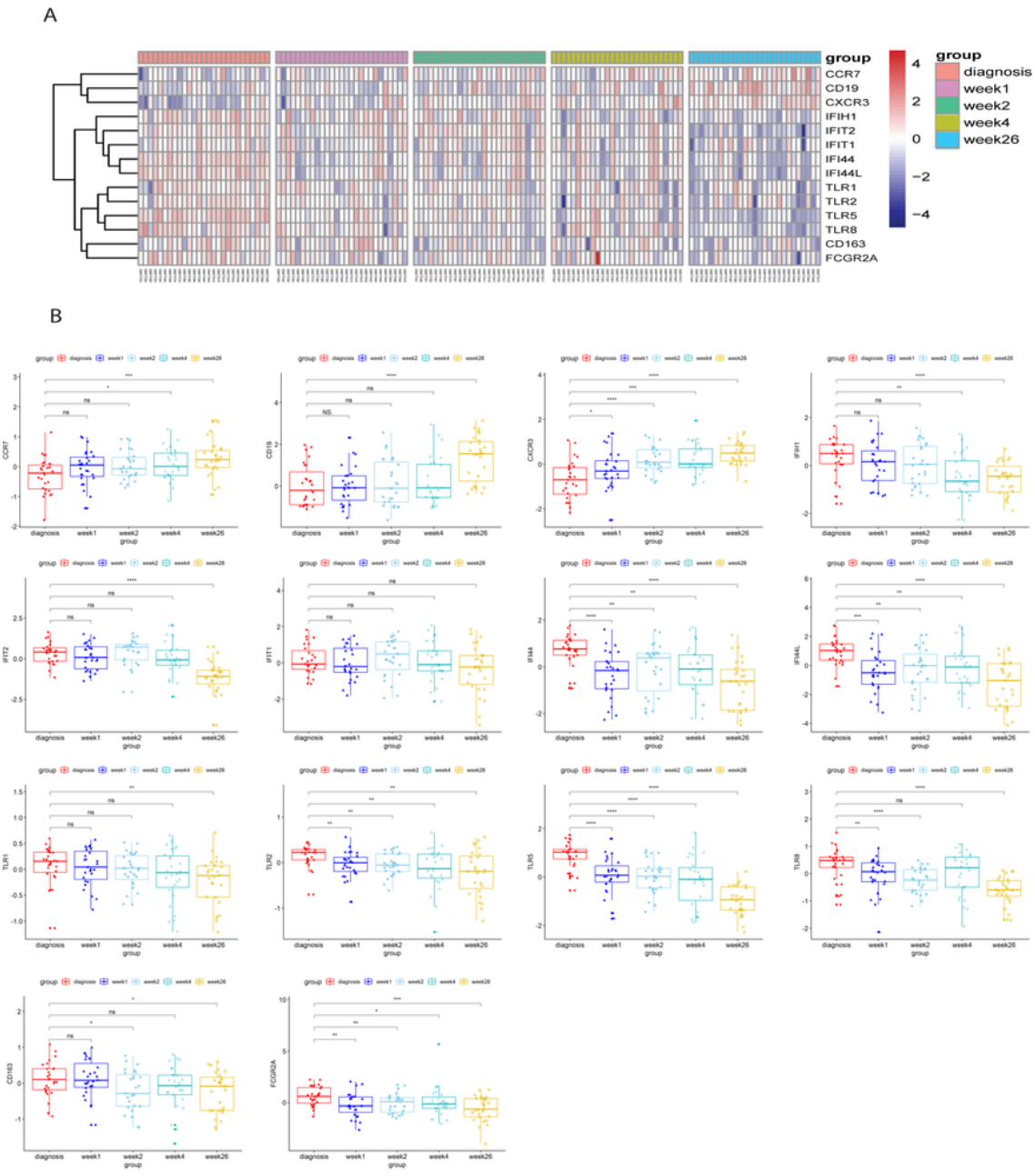
Functional analysis of 35 genes in Module 1. (A) GO analysis reveals that genes functionally related to defense response to other organism underlie PTB infection. (B) Network of Module 1 genes constructed by STRING. (C) Heatmap showing the gene expression of module 1 in individuals from dataset GSE34608. (D) Heatmap showing the gene expression of module 1 in individuals from dataset GSE83456. Each row represents a gene, and each column represents a sample.




## Figure 4

Expression validation of the hub genes in dataset GSE19439. (A-N) demonstrate the expression of IFIH1, IFI44L, IFI44, IFITI, IFIT2, TLR1, TLR2, TLR5, TLR8, FCCGR2A, CD163, CCR7, CD19, CXCR3 during PTB

infection, respectively. **, P value < 0.01; ***, P value < 0.001.




## Figure 5

The expression levels of hub genes during PTB treatment in five time points in dataset GSE31348. Heatmap (A) and box plot (B) of hub genes in dataset GSE31348. The central of rectangle indicates the medium. The ends of the whiskers represent the first quartile to the third quartile (the interquartile range or IQR). The ends of central line show the minimum or maximum value.