

Received: 2020.04.26
Accepted: 2020.05.11
Available online: 2020.05.28
Published: 2020.07.23

Identification of Biomarkers for Sarcoidosis and Tuberculosis of the Lung Using Systematic and Integrated Analysis

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCEF 1 **Min Zhao**
BCE 1 **Xin Di**
CDE 2 **Xin Jin**
BF 1 **Chang Tian**
CF 1 **Shan Cong**
BF 1 **Jiangyin Liu**
ABDEG 1 **Ke Wang**

1 Department of Respiratory and Critical Care Medicine, The Second Hospital of Jilin University, Changchun, Jilin, P.R. China
2 Department of Oncology and Hematology, The Second Hospital of Jilin University, Changchun, Jilin, P.R. China

Corresponding Author: Ke Wang, e-mail: wke@jlu.edu.cn

Source of support: This study was supported by the Technology Research Funds of Jilin Province (201903031625F), the Natural Science Foundation of Jilin Province (201801011031C), and the Medical and Health Project Funds of Jilin Province (20191102012YY)

Background: Sarcoidosis (SARC) is a multisystem inflammatory disease of unknown etiology and pulmonary tuberculosis (PTB) is caused by *Mycobacterium tuberculosis*. Both of these diseases affect lungs and lymph nodes and share similar clinical manifestations. However, the underlying mechanisms for the similarities and differences in genetic characteristics of SARC and PTB remain unclear.





Material/Methods: Three datasets (GSE16538, GSE20050, and GSE19314) were retrieved from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) in SARC and PTB were identified using GEO2R online analyzer and Venn diagram software. Functional enrichment analysis was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) and R packages. Two protein-protein interaction (PPI) networks were constructed using Search Tool for the Retrieval of Interacting Genes database, and module analysis was performed using Cytoscape. Hub genes were identified using area under the receiver operating characteristic curve analysis.

Results: We identified 228 DEGs, including 56 common SARC-PTB DEGs (enriched in interferon-gamma-mediated signaling, response to gamma radiation, and immune response) and 172 SARC-only DEGs (enriched in immune response, cellular calcium ion homeostasis, and dendritic cell chemotaxis). Potential biomarkers for SARC included *CBX5*, *BCL11B*, and *GPR18*.

Conclusions: We identified potential biomarkers that can be used as candidates for diagnosis and/or treatment of patients with SARC.

MeSH Keywords: **Biological Markers • Gene Expression Profiling • Sarcoidosis • Tuberculosis**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/925438>

 3000  7  5  40



Background

Sarcoidosis (SARC) is a multisystem inflammatory disease of unknown etiology, and primarily affects the lungs and thoracic lymph nodes and manifests with noncaseating epithelioid granulomas (aggregates of lymphocytes, macrophages, epithelioid, and giant cells) [1]. Granulomatous diseases result from the continued presentation of a poorly degradable antigen [2]. Among these, the infectious agents *Propionibacterium* and *Mycobacterium tuberculosis* (Mtb) are currently the hotspots for research [3]. SARC has been considered to be associated with tuberculosis since it was first described [4]. Pulmonary tuberculosis (PTB) is an infectious disease caused by Mtb that mainly affects the lungs and lymph nodes. SARC and sputum-negative PTB share similar clinical, radiological, and histopathological characteristics, thereby making the differential diagnosis of sputum-negative TB and SARC challenging [5]. The blood profiles of patients with PTB and SARC predominantly have similar indications of inflammation [6]. However, gene expression of the tissues involved in the pathogenesis of SARC and PTB remains to be understood in detail.

Microarray technology provides an all-in-one system to simultaneously scan hybridization signals from tens of thousands of gene probes on a chip and quantitatively analyze the transcriptome of samples [7]. Transcriptomic analysis can be used to understand various aspects of SARC and PTB. These profiles are stored in genomic databases that need to be analyzed to understand the characteristics of diseases, such as SARC and PTB. This enables the identification of potential candidates that can be used for diagnosis and treatment. In this study, we compared the expression profiles of lung tissues from patients and healthy individuals to identify disease signatures and understand the pathogenesis of SARC. We used the GSE16538, GSE20050, and GSE19314 expression microarray data from the Gene Expression Omnibus (GEO) database to identify differentially expressed genes (DEGs) and their interaction patterns using a protein–protein interaction (PPI) network. DEG functions were determined using Gene Ontology (GO) and pathway enrichment analysis. A number of genes were putative (novel) candidates for use as diagnostic, therapeutic, and prognostic markers for SARC. This will help develop new array-based diagnostic tools to distinguish patients from healthy individuals and differentiate between diseases of similar pathology. Differential biomarker profiles can also provide insight into the common biological processes affected by similar diseases, such as SARC and PTB.

Material and Methods

Data download and pre-processing

NCBI-GEO is a free public database of microarray/genomic data repository. The gene expression profiles in lung tissues from SARC patients, lung tissues from PTB patients, and peripheral blood from SARC patients were obtained from the GSE16538, GSE20050, and GSE19314 datasets, respectively. Microarray data of GSE16538 and GSE19314 were obtained using GPL570 Platforms ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) that included 6 tissue samples from patients with SARC and healthy individuals each, and 37 and 20 blood samples from patients with SARC and healthy individuals, respectively. Microarray data from GSE20050 was obtained using the GPL1352 platform ([U133_X3P] Affymetrix Human X3P Array) comprising 3 PTB and 2 healthy lung tissue samples (Table 1).

Data processing and identification of DEGs

DEGs in the lung samples from patients with SARC and PTB and healthy individuals were identified using the GEO2R online tools with fold change ($|\log FC|$) >1 and adjusted P value <0.05 , respectively. Subsequently, the raw TXT data were analyzed using the Venn online software (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to detect the common DEGs between SARC and PTB. DEGs with $\log FC <0$ and $\log FC >0$ were categorized as downregulated and upregulated genes.

GO and pathway enrichment analysis

GO is commonly used to determine the function(s) of genes during the analysis of high throughput transcriptomic or genomic data. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of databases associated with genomes, diseases, biological pathways, drugs, and chemical materials. DAVID (Database for Annotation, Visualization and Integrated Discovery) is an online bioinformatics tool that is used to identify the functions of genes or proteins. We used DAVID to determine the functions of DEGs in the biological process (BP), cellular component (CC), molecular function (MF), and pathways ($P < 0.05$).

PPI networks and module analysis

PPI networks help understand the molecular mechanisms underlying biological progresses. PPI networks was generated by Search Tool for the Retrieval of Interacting Genes (STRING) (<https://string-db.org/>) using a threshold interaction score of 0.4. Subsequently, we used the MCODE app in Cytoscape to analyze PPI network modules (degree cut-off=2, max; Depth=100; k-core=2; and node score cut-off=0.2).

Table 1. The detail information of 3 GEO datasets.

ID	Source	Platform	Normal	Test
GSE16538	Tissues	GPL570	6	6 (SARC)
GSE20050	Tissues	GPL1352	2	3 (PTB)
GSE19314	PBMC	GPL570	20	37 (SARC)

GEO – Gene Expression Omnibus; SARC – sarcoidosis; PTB – pulmonary tuberculosis; PBMC – peripheral blood mononuclear cells.

Table 2. SARC-PTB common DEGs.

DEGs	Genes name
Upregulated	SOD2 ST8SIA4 YME1L1 BCL2 OTUD6B CARD8 CRIPT GNB4 NKTR UBQLN2 ZNF107 PRKACB GNA13 STAT1 TNRC6B GPR155 HS3ST3B1 TBL1XR1 ZBTB10 PPA2 ATXN1 KLHL5 LAP3 WIPF1 FGL2 MDFIC RECQL HCP5 ADAMDEC1 ABCA1 TAP1 PLXNC1 NR3C1 HPS3 GBP2 SLC4A7 ZNF292 GIMAP2 SCYL2 WDR11 MGAT2 GZMA LY75 TRIM13 ZBTB1 CXCL9 STK4 DR1 ITM2A ANKRD12 VCAM1 IFNGR1 YIPF5 TRBC1 ETNK1
Downregulated	WDFY2

DEGs – differentially expressed genes; SARC – sarcoidosis; PTB – pulmonary tuberculosis.

Table 3. SARC-only DEGs.

DEGs	Genes name
Upregulated	U2SURP LCORL CACNA2D1 ZEB1 TMCC3 OSBPL8 CCDC34 PRKCB INTS6 MANEA CCL19 EIF1AX PPM1A FCN1 FYB UBXN7 KLHL28 BICD1 FOXN2 ZCCHC10 DMXL1 LOC101928361//ARHGAP15 PAPD4 GNG2 CDH11 GIMAP6 BCL11B GPR18 CHIT1 GBP4 LOC101060835//LOC100996809//HLA DRB6//HLA-DRB5//HLA-DRB4//HLA-DRB3//HLA-DRB1//HLA-DQB1 MMP12 CCL3L3//CCL3L1//CCL3 MYBL1 RB1CC1 NRN1 FMR1 FAM171B TAP2 PAPOLG LOC101927027 GBP1 LOC101927933//LRRC8C ANKRD44 SLC30A1CAPS2FRG1KP//LOC101930531//LOC101930278//FRG1CP//LOC100134091//FRG1BP//FRG1 RSNB1 RASGRP3 RUNX2 SLC25A36 CCL5 MAF ZNF207 ELK3 DNAJB14 RBMS1 KLRB1 NLRC3 SOS1 DCK EXOC5 SH2D1A SAMD3 FAR5B DGKH TMEM267 DICER1 BTN3A2 KBTBD6 HAUS6 N4BP2 MIR15A//DLEU2//DLEU2 CCNG2 AEBP2 CXCL11 ZNF366 CSTF2T DNAJC3 UHMK1 DYRK2 PARP11 ZNF92 CDK6 CCR2 DDX60 SGTB ZNF518A SOS1-IT1 ROBO1 CYP7B1 STXBP5 ZNF217 ZNF101 IL15 TRMT13 RBM12B SERPINB4//SERPINB3 TRERF1 GOPC PLEKHF2 RNF138 ZBED8 WDFY1 ZNF267 FAM26F UBE2D1 ITGA4 THAP2 PRKAR2B IKBIP MARCKS SACS ARHGAP18 INPP4B PCGF5 GPR27 GUCY1A2 DENND1B ZNF302 FGFR10P2 DCP2 CD48 ZEB2 7-Sep CSNK1G3 LPAR6 ZNF654 SLFN5 BCL11A SAMD9 CBX5 KCTD12 C18orf54 MEX3C LOC100134445 TBC1D9 C3orf38 GATM TMF1 IL7 SEMA3C
Downregulated	CAPN8 MIR612//NEAT1 GDF15 SVIL MFSD2A GLDN PGC LDLR EIF3B ALOX15B ZNF431 RBP4 LOC100507487 CFI AQP3 EIF1 SLC16A5 C14orf1 SFN LMNA HKDC1 WSB1 MICALCL SNX25 CCDC91 SLC11A1 RASEF ALPL PVR MYADM

DEGs – differentially expressed genes; SARC – sarcoidosis.

Identification of hub gene(s)

The hub genes in the candidate modules were identified using Cytoscape plugin cytoHubba (top 10 nodes ranked by MCC, MNC, Radiality, Stress, Betweenness, BottleNeck, Closeness, ClusteringCoefficient, Degree, DMNC, EcCentricity, and EPC). We used the expression data of GSE19314 database to generate the receiver operating characteristic (ROC) curves of the hub genes by R/Bioconductor. The area under the ROC curve (AUC) for these genes were calculated to test their potential in diagnosing SARC using. AUC >0.5 was considered statistically significant.

Results

Identification of DEGs in SARC and PTB

We identified 228 and 2913 DEGs from the GSE16538 and GSE20050 databases, respectively. Using Venn diagrams, we categorized the DEGs related to SARC that were shared with PTB or characteristic of SARC (SARC-PTB common and SARC-only DEGs). There were 56 SARC-PTB common DEGs, including 1 downregulated gene (logFC <0) and 55 upregulated genes (logFC >0; Table 2). We identified 172 SARC-only DEGs, including 30 downregulated genes and 142 upregulated genes (Table 3, Figure 1).

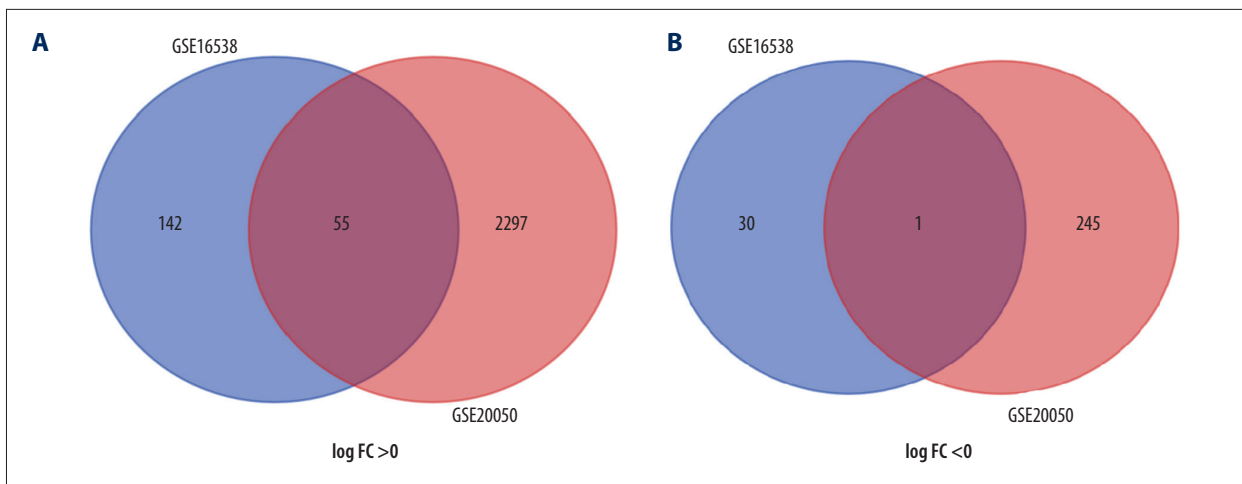


Figure 1. Identification of 56 common SARC-PTB DEGs in the 2 datasets (GSE16538 and GSE20050), and 172 SARC-only DEGs in GSE16538 through Venn diagrams software. Different color meant different datasets. **(A)** 55 DEGs were upregulated in the 2 datasets, 142 DEGs were in GSE16238 (logFC >0). **(B)** 1 DEG was downregulated in 2 datasets, 30 DEGs in GSE16238 (logFC <0). Different color meant different datasets. DEGs – differentially expressed genes; SARC – sarcoidosis; PTB – pulmonary tuberculosis.

GO and KEGG analysis

The 56 SARC-PTB common DEGs were analyzed by DAVID followed by the R package to analyze their biological functions and associated pathways. DEGs were enriched in BP involved in interferon-gamma-mediated signaling pathway, response to gamma radiation, immune response, response to hydrogen peroxide, defense response, response to nutrient, signal transduction, apoptotic process, regulation of interferon-gamma-mediated signaling pathway, and intrinsic apoptotic signaling pathway in response to oxidative stress; in the CC comprising the membrane; and MF in protein homodimerization activity, signal transducer activity, and anion transmembrane transporter activity (Table 4). Table 5 shows the KEGG analysis highlighting the enrichments of DEGs in pathways related to cancer and chemokine signaling pathway ($P < 0.05$).

The 172 SARC-only upregulated DEGs were enriched in BP involved in immune response, cellular calcium ion homeostasis, dendritic cell chemotaxis, chemokine-mediated signaling pathway, positive regulation of Rho protein signal transduction, signal transduction, cell maturation, RNA processing, regulation of transcription from RNA polymerase II promoter, negative regulation of chemokine-mediated signaling pathway, regulation of T cell differentiation in thymus, and response to virus; the CC of neuronal ribonucleoprotein granule and histone deacetylase complex; and in MF involved in metal ion binding, nucleic acid binding, RNA polymerase II core promoter proximal region sequence-specific DNA binding, ATP binding, protein binding, chemokine activity, and manganese ion binding. The downregulated DEGs were enriched in BP involved in response to vitamin D, CC including the extracellular space, extracellular

exosome, and endosome membrane, and MF in transporter activity (Figure 2A–2F). Table 6 shows the KEGG analysis highlighting that upregulated DEGs were particularly enriched in chemokine signaling pathway, microRNAs in cancer, and cytokine–cytokine receptor interaction; no KEGG pathway was associated with downregulated DEGs ($P < 0.05$).

PPI network and module analysis

PPI networks reflect the spatiotemporal relationship of macromolecules within the cell that helps understand the molecular mechanisms involved in physiological and pathological processes. To explore the molecular mechanism(s) underlying SARC progression, we used the STRING database to construct the PPI network. An interaction score of 0.4 (highest confidence) was set as threshold. The PPI network for SARC-PTB common DEGs consisted of 31 nodes and 31 edges (Figure 3A) and that of SARC-only DEGs comprised 95 nodes with 139 edges (Figure 4A). Subsequently, we identified the key modules using the MCODE plugin. We detected one functional cluster module and related genes in the PPI network of SARC-PTB common DEGs (Figure 3B). KEGG analysis of the genes in the module revealed the association with Toll-like receptor signaling pathway (hsa04620). A total of 3 functional clusters of modules and related genes were detected in the PPI network for SARC-only DEGs (Figure 4B–4D). KEGG analysis of the module genes revealed their association with cytokine-cytokine receptor interaction (hsa04060), chemokine signaling pathway (hsa04062), and RNA transport (hsa03013).

Table 4. GO analysis of SARC-PTB common DEGs.

Category	Term	Count	%	P-value	FDR
GOTERM_BP_DIRECT	GO: 0060333~interferon-gamma-mediated signaling pathway	4	0.05	0.00138	1.94609
GOTERM_BP_DIRECT	GO: 0010332~response to gamma radiation	3	0.03	0.00413	5.71893
GOTERM_BP_DIRECT	GO: 0006955~immune response	6	0.07	0.00956	12.77981
GOTERM_BP_DIRECT	GO: 0042542~response to hydrogen peroxide	3	0.03	0.01088	14.42335
GOTERM_BP_DIRECT	GO: 0006952~defense response	3	0.03	0.01829	23.10506
GOTERM_BP_DIRECT	GO: 0007584~response to nutrient	3	0.03	0.02204	27.18070
GOTERM_BP_DIRECT	GO: 0007165~signal transduction	9	0.10	0.02538	30.63884
GOTERM_BP_DIRECT	GO: 0006915~apoptotic process	6	0.07	0.03045	35.60432
GOTERM_BP_DIRECT	GO: 0060334~regulation of interferon-gamma-mediated signaling pathway	2	0.02	0.03953	43.67413
GOTERM_BP_DIRECT	GO: 0008631~intrinsic apoptotic signaling pathway in response to oxidative stress	2	0.02	0.04548	48.43701
GOTERM_CC_DIRECT	GO: 0016020~membrane	13	0.15	0.02196	22.02011
GOTERM_MF_DIRECT	GO: 0042803~protein homodimerization activity	7	0.08	0.02208	23.23086
GOTERM_MF_DIRECT	GO: 0004871~signal transducer activity	4	0.05	0.02368	24.71270
GOTERM_MF_DIRECT	GO: 0008509~anion transmembrane transporter activity	2	0.02	0.02687	27.57188

Go – Gene Ontology; DEGs – differentially expressed genes; SARC – sarcoidosis; PTB – pulmonary tuberculosis; FDR – false discovery rate.

Table 5. KEGG pathway analysis of SARC-PTB common DEGs.

Pathway ID	Name	Count	%	P-value	Genes
hsa05200	Pathways in cancer	6	0.07	0.00849	GNA13, BCL2, GNB4, PRKACB, STAT1, STK4
hsa04062	Chemokine signaling pathway	4	0.05	0.02314	CXCL9, GNB4, PRKACB, STAT1

KEGG – Kyoto Encyclopedia of Genes and Genomes; DEGs – differentially expressed genes; SARC – sarcoidosis; PTB – pulmonary tuberculosis.

Hub gene analysis

The top 10 genes of each sub-plugin in the PPI network of SARC-only DEGs were identified by the cytoHubba plugin (Table 7). The 13 genes with the highest frequency in all sub-plugins were classified as hub genes, including IL7, DICER1, CCL19, CBX5, CDK6, CXCL11, IL15, GNG2, BCL11B, CCR2, CCL5, GPR18, and ZEB1; all these genes were upregulated. KEGG analysis of the hub genes showed their involvement in cytokine-cytokine receptor interaction (hsa04060), chemokine signaling pathway (hsa04062), and microRNAs in cancer (hsa05206). We used AUC analysis to evaluate their potential in diagnosing SARC using the expression data in the GSE19314 database. ROC-AUCs were higher than 0.50 for 12 of the genes (except CCL5). AUC >0.5 was considered statistically significant. The AUCs for CBX5, BCL11B, and GPR18 were 0.81, 0.91, and 0.87, respectively;

these genes are of higher diagnostic potential since they were all higher than 0.80 (Figure 5).

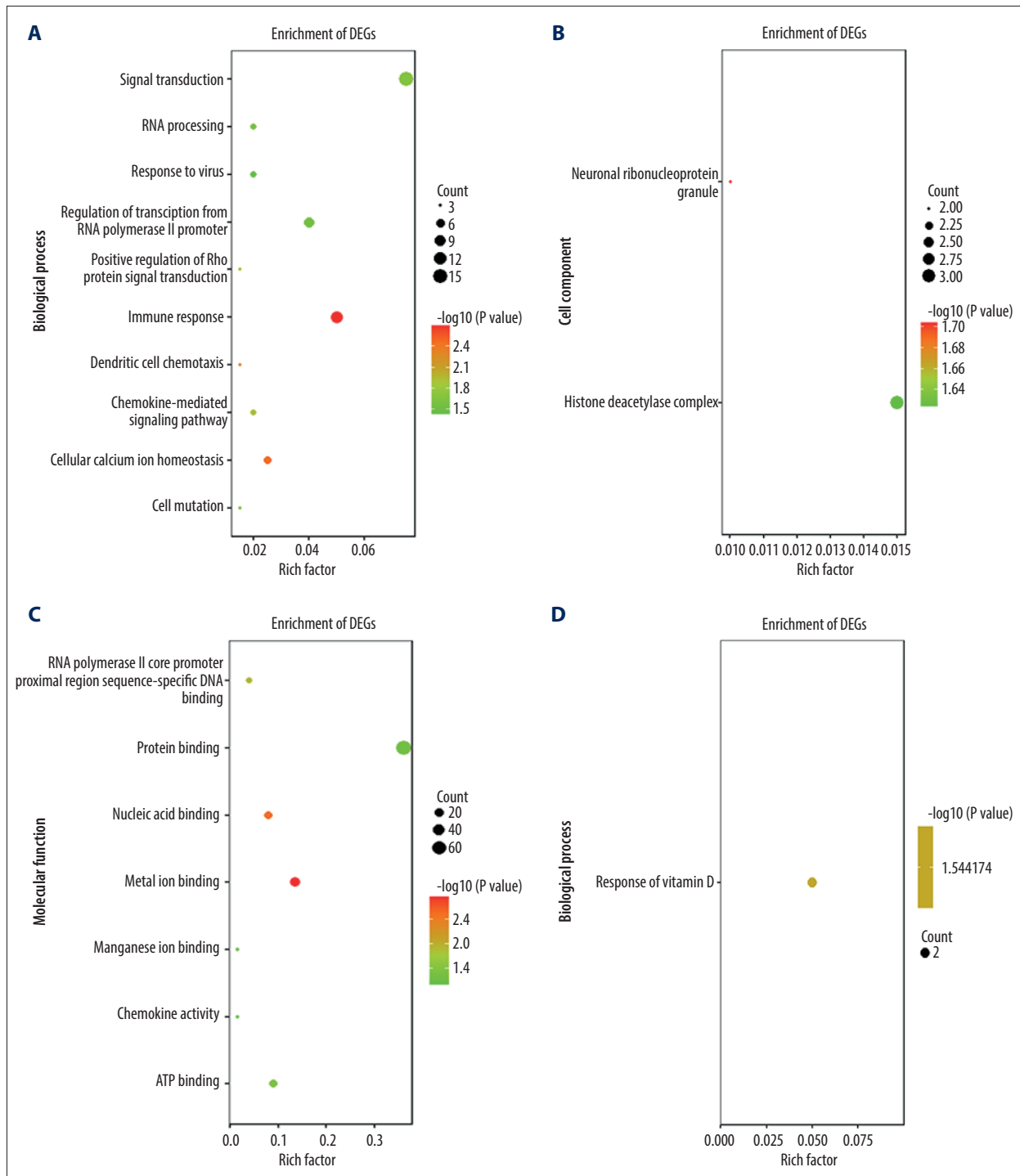
Discussion

Multiple lines of evidence show that Mtb is a predisposing factor for SARC. SARC and PTB share histopathological characteristics, such as the presence of granulomas; thus, histopathological evidence can be used as the gold standard for diagnosis. Studying the gene expression profiles of tissues involved in pathogenesis is a direct tool for understanding the role of Mtb in SARC. Thus, we employed an integrated bioinformatics approach using 2 GEO datasets (GSE16538 and GSE20050) to identify 228 DEGs associated with SARC, including 56 SARC-PTB common DEGs and 172 SARC-only DEGs. There are more

upregulated genes and less downregulated genes among these DEGs. Although the current etiology of sarcoidosis is not yet clear, sarcoidosis is closely related to the immune response. Excessive activation of the immune response to unknown inhaled antigens is considered to be one of the pathogenesis of sarcoidosis [8]. Most of the genes related to sarcoidosis that we have obtained are also related to the immune response,

and most of DEGs are upregulated genes. We believe that the complex relationship of these immune-related upregulated genes may lead to excessive immune responses.

GO analysis revealed that the pathways enriched for SARC-PTB common DEGs were primarily associated with immune responses, especially interferon-gamma-mediated signaling.



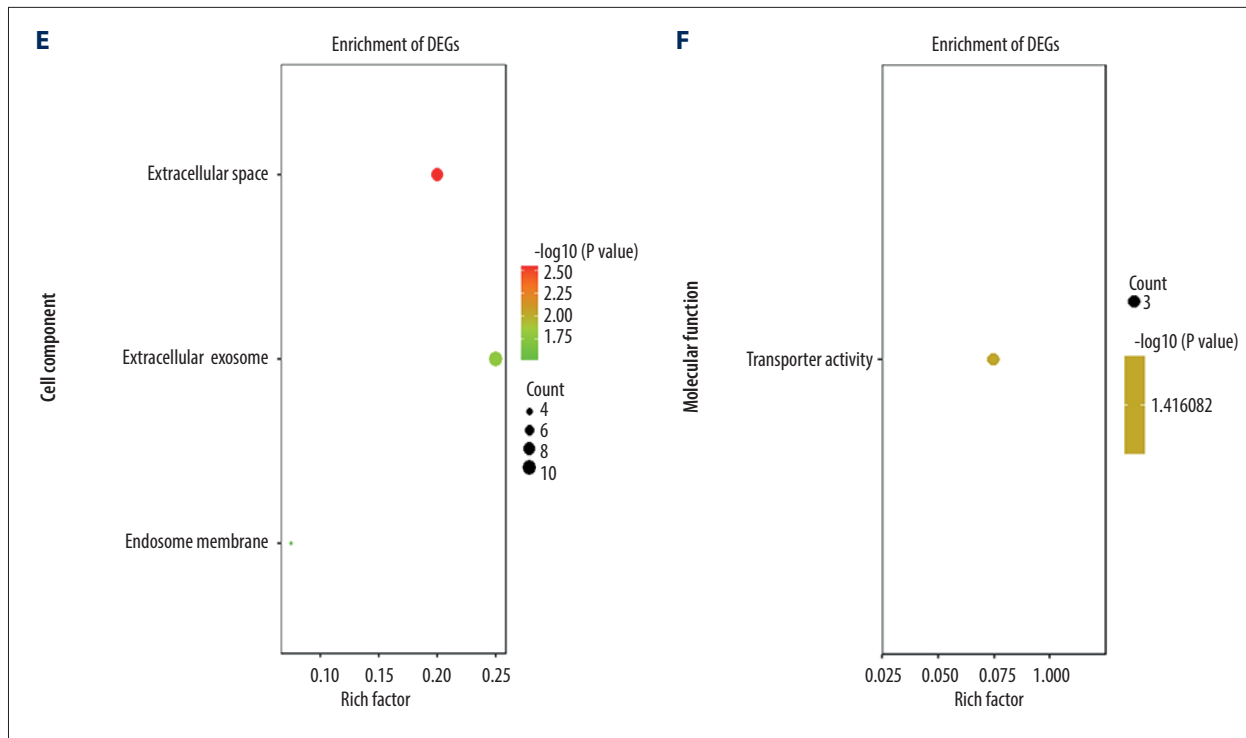


Figure 2. GO terms of overlapped SARC-only DEGs. (A–C) Upregulated genes. (D–F) Downregulated genes. (A, D), biological process; (B, E), cellular component; (C, F), molecular function. GO – Gene ontology; DEGs – differentially expressed genes; SARC – sarcoidosis.

Table 6. KEGG pathway analysis of SARC-only DEGs.

Pathway ID	Name	Count	%	P-value	Genes
hsa04062	Chemokine signaling pathway	6	0.03	0.00915	SOS1, CCR2, CCL19, GNG2, CCL5, CXCL11
hsa05206	MicroRNAs in cancer	7	0.04	0.01384	SOS1, DICER1, ZEB2, CDK6, MARCKS, ZEB1, PRKCB
hsa04060	Cytokine-cytokine receptor interaction	6	0.03	0.02626	IL7, CCR2, CCL19, IL15, CCL5, CXCL11

KEGG – Kyoto Encyclopedia of Genes and Genomes; DEGs – differentially expressed genes; SARC – sarcoidosis.

Unlike other chronic infectious diseases, Mtb promotes a powerful systemic immune response that damages local tissues and enhances the extended survival and spread of Mtb [9]. When the body is infected with an exogenous antigen, γ -interferon (IFN- γ) is released, following which local macrophages phagocytose, process and present antigens, and activate T cells. Subsequently, blood-derived lymphocytes and monocytes accumulate at the site of infection to form granulomas [10]. KEGG analysis demonstrated that SARC-PTB common DEGs were enriched in pathways related to cancer and the chemokine signaling pathway. This highlights the molecular interactions that are common to the pathogenesis of SARC and PTB. Some studies have shown a positive correlation between patient history of TB with increased risk for lung cancer and associated mortality. A Korean prospective cohort study showed

the correlation between TB and risk of lung cancer in a high-income country [11]. Local chronic inflammation and oxidative stress are mechanisms that increase the risk of lung cancer in patients with a history of lung disease. Fibrosis is crucial in maintaining inflammation [12–16]. Similar studies have revealed a link between SARC and lung cancer [17–21]. The immunologic abnormalities in patients with SARC may be a result of the lack of immune response against tumors or oncogenic viruses. Conversely, SARC-related chronic inflammation may lead to the development of cancer [22].

GO and KEGG analysis of SARC-only DEGs showed enrichment in immune response, suggesting the involvement of immunity in the pathogenesis of SARC. These genes were also associated with BP related to viruses and vitamin D-related biological

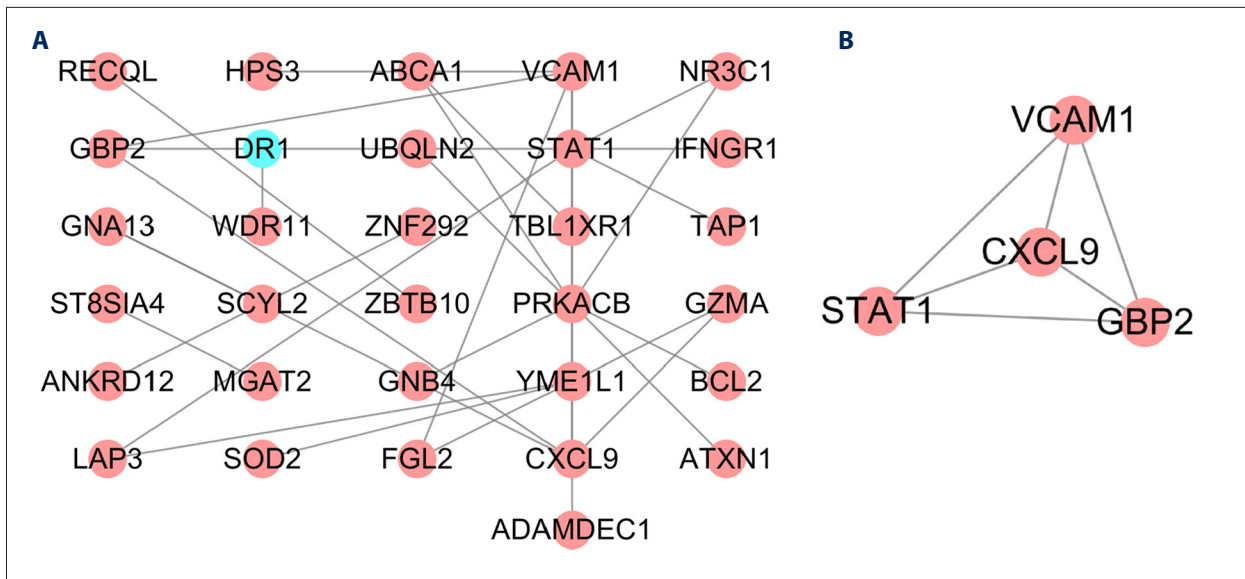


Figure 3. PPI network of SARC-PTB common DEGs and modules analysis. (A) PPI network of SARC-PTB common DEGs; (B) the only one module of PPI network. The red nodes represent the upregulated DEGs. The blue nodes represent the downregulated DEGs. The nodes meant proteins; the edges meant the interaction of proteins. PPI – protein–protein interaction; DEGs – differentially expressed genes; SARC – sarcoidosis; PTB – pulmonary tuberculosis.

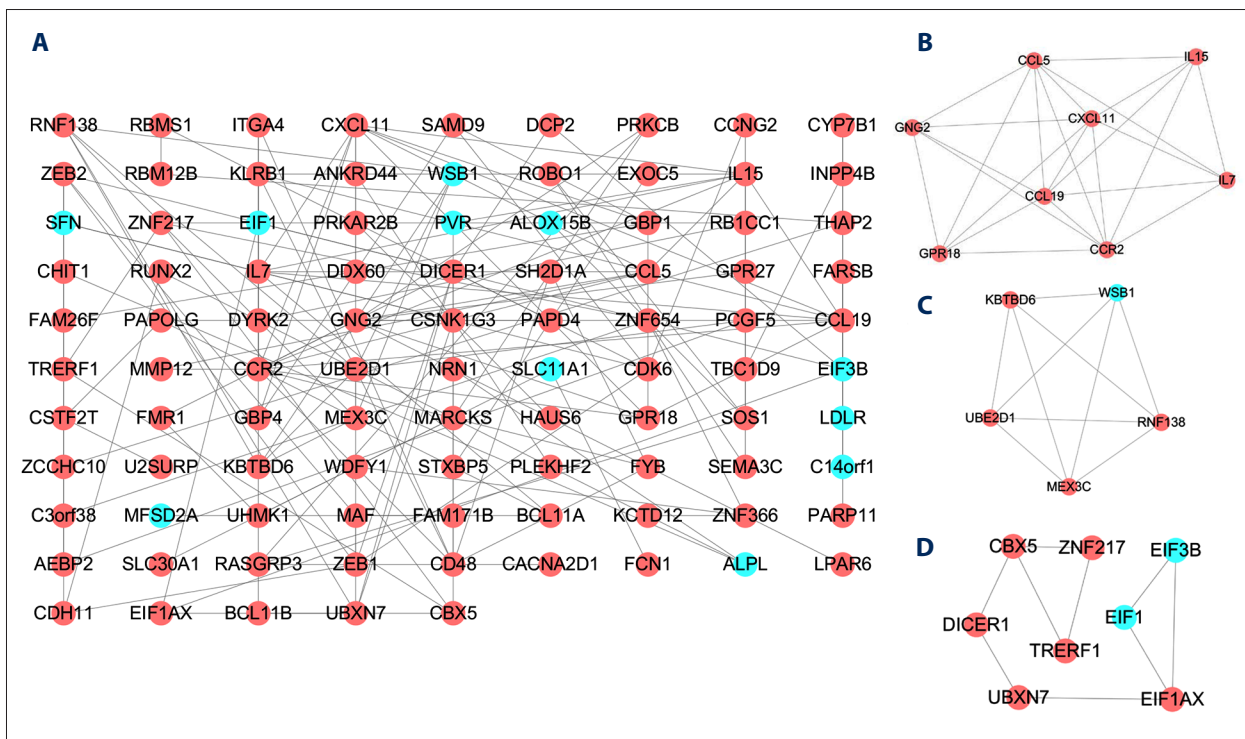


Figure 4. PPI network of SARC-only DEGs and modules analysis. (A) PPI network of SARC-only DEGs; (B–D) the top 3 modules of PPI network. The red nodes represent the upregulated DEGs. The blue nodes represent the downregulated DEGs. The nodes meant proteins; the edges meant the interaction of proteins. PPI – protein–protein interaction; DEG – differentially expressed genes; SARC – sarcoidosis.

Table 7. The top 10 genes in each sub-plugin of SARC-only DEGs by cytoHubba.

Sub-plugins	Genes
EPC	IL7 CCL19 CXCL11 IL15 GNG2 CCR2 CCL5 GPR18 CD48 KLRB1
MCC	IL7 CCL19 CXCL11 IL15 GNG2 CCR2 CCL5 GPR18 RNF138 UBE2D1
MNC	IL7 CCL19 CXCL11 IL15 GNG2 CCR2 CCL5 GPR18 ZEB1 RNF138
DMNC	IL7 CCL19 IL15 GNG2 GPR18 RNF138 MEX3C KBTBD6 WSB1 KLRB1
Stress	IL7 DICER1 CBX5 CDK6 GNG2 BCL11B UBXXN7 BCL11A SAMD9 CDH11
Degree	IL7 DICER1 CCL19 CXCL11 IL15 GNG2 CCR2 CCL5 CD48 UBE2D1
Radiality	IL7 DICER1 CBX5 CDK6 BCL11B ZEB1 SAMD9 ZEB2 MAF PAPD4
Closeness	IL7 DICER1 CCL19 CBX5 CDK6 CXCL11 IL15 CCR2 CCL5 ZEB1
BottleNeck	IL7 DICER1 CBX5 CDK6 CXCL11 BCL11B ZEB1 UBXXN7 SAMD9 GBP1
Betweenness	IL7 DICER1 CBX5 CDK6 BCL11B UBXXN7 BCL11A CDH11 ZNF654 ZNF217
EcCentricity	IL7 DICER1 CBX5 CDK6 BCL11B BCL11A ZEB2 MAF RBMS1 SFN
ClusteringCoefficient	CCL19 GPR18 MEX3C WSB1 KBTBD6 MMP12 EIF1 TRERF1 GBP4 SH2D1A

DEGs – differentially expressed genes; SARC – sarcoidosis.

reactions. SARC may be caused by a variety of pathogens. In addition to *Mtb* and *Propionibacterium acnes*, these pathogens include *Borrelia*, HHV-8 and Rickettsia, *Chlamydia pneumoniae*, and Epstein-Barr virus [23–26]. However, a 2016 meta-analysis comprising more than 6000 patients across several countries showed that none of these other pathogens were significantly associated with SARC [27]. Two studies reported a significant correlation between *Borrelia* and SARC in regions endemic for Lyme disease [28]. In contrast, 4 similar studies were performed in non-endemic areas where this correlation was not significant [29]. BP involved in the pathogenesis of SARC, including viruses, depend on the local epidemiological background. The pathogenesis of SARC is related to the metabolism of vitamin D. The active form of vitamin D, 1,25-(OH)-2-D3, regulates immune response via cytokine expression, including IFN- γ and IL-2 [30]. The expression of vitamin D receptors in diseased BALT cells increases, and nodular granulomatous macrophages overproduce 1,25-dihydroxyvitamin D; hypercalcemia also manifests in these patients (10%) [31].

In accordance with our KEGG analysis, among the 13 hub genes identified by the PPI network and modules analysis, IL7, CCR2, CCL19, IL15, and CXCL11 were associated with cytokine-cytokine receptor interaction. ROC curves demonstrated the diagnostic potential for these upregulated hub genes. IL7 is produced locally by intestinal epithelial and epithelial goblet cells and regulates intestinal mucosal lymphocytes. Knockout studies in mice have shown that this cytokine plays an essential role in lymphoid cell survival; it is important for B and T cell development. However, there is no report on the correlation between IL7 and SARC. CCR2 is expressed on the surface of macrophages, dendritic cells, and T lymphocytes and is a receptor for chemotactic proteins involved in inflammatory cell

recruitment [32]. The research showed that CCR2 haplotype 2 is an independent factor that predisposes patients with SARC to Löfgren's syndrome. CCR2 and its ligands have been shown to regulate cellular proliferation during granuloma formation in mice [33,34]. CCR2-deficient mice exhibit defective macrophage recruitment and impaired type-1 cytokine production. CCL19 is a CC-cytokine gene located on the p-arm of chromosome 9. Cytokines are a family of secreted proteins involved in immunoregulation and inflammation. Studies have shown that the overexpression of CCL19 in fluid from alveolar lavage obtained from patients with SARC promotes the development of alveolitis in the lungs of these patients by stimulating CD4⁺ lymphocyte recruitment [35,36]. IL15 is a cytokine that regulates T and natural killer cell activation and proliferation. IL15 and IL2 have similar biological functions. Tissues from patients with SARC possess a higher abundance of IL15⁺ cells as compared to that in asthmatic patients and healthy individuals. CXCL11 is a C-X-C motif chemokine that induces a chemotactic response in activated T cells and is the primary ligand for CXCR3. Studies shown that CXCL11 increases in the serum of patients with SARC as compared to that in healthy controls. Moreover, serum CXCL11 correlates with the severity of symptoms in patients with SARC, specifically organ involvement, respiratory symptoms, and abnormalities in pulmonary function [37]. Thus, CXCL11 may also be useful in identifying patients at risk for rapid pulmonary dysfunction.

ROC curve analysis showed high diagnostic potential (AUC >0.080) of CBX5, BCL11B, and GPR18. To the best of our knowledge, this is the first report to correlate these hub genes with SARC. CBX5 encodes a highly conserved non-histone member of the heterochromatin protein family. CBX5 is overexpressed in gastric cancer tissues and can be downregulated

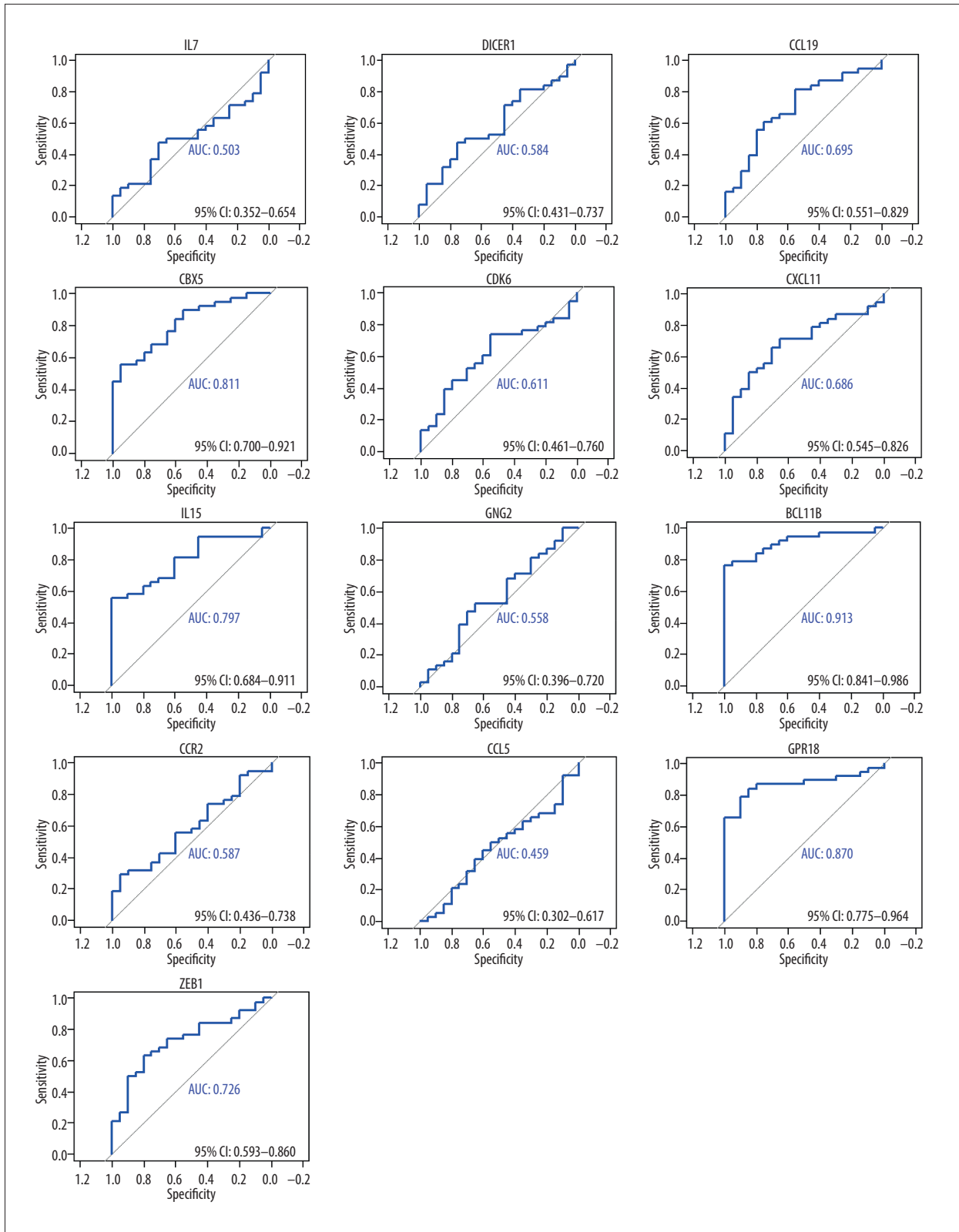


Figure 5. The diagnosis information of the 13 hub genes. The ROC curves were used to identify the diagnosis information of the 13 hub genes and, except CCL5. *CBX5*, *BCL11B* and *GPR18* had a significantly diagnostic value (AUC >0.80). ROC curve – receiver operating characteristic curve; AUC – the area under the ROC curve.

using vorinostat, indicating the oncogenic role of CBX5 in gastric cancer. MiR-758-3p targets CBX5 and inhibits the progression of gastric cancer [38]. Sarcoidosis is related to the occurrence of lung cancer, so it is possible that sarcoidosis-induced lung cancer is related to CBX5. BCL11B encodes a C2H2-type zinc finger protein that is closely related to BCL11A. A recent study based on HAART (the treatment regimen used for patients with HIV) showed that increased expression of BCL11B and its related chromatin modifiers contribute to the transcriptional silencing of HIV-1 and alternating BCL11B levels result in abnormal transcription and inflammation [39]. BCL11B is closely related to the immune response, so we speculate that there may be BCL11B participation in the pathogenesis of sarcoidosis, and more follow-up experiments are needed to demonstrate. GPR18, or N-arachidonyl glycine receptor (NAGly receptor), is a class A G-protein coupled receptor that, along with GPR119 and GPR55, comprises a novel cannabinoid receptor that consists of 331 amino acids. GPR18 is overexpressed in lymphoid tissues, such as the spleen and thymus, with moderate expression in the testes, ovaries, brain, and lungs. Current research, like in a rat model of inflammatory pain, an animal model of neuropathic pain (sciatic nerve injury), and an inflammatory pain model induced by intraplantar injection of complete Freund's adjuvant (CFA), suggested GPR18 has an important role in mediating the analgesic properties of NAGly in different pain models [40]. GPR18 is highly expressed in lymphoid tissues and lungs, but the relationship between specific GPR18 and the respiratory system, including sarcoidosis, is rarely reported and requires further exploration. The molecular mechanism employed the potential hub gene candidates in patients with SARC need to be studied in detail.

References:

1. Landi C, Bargagli E, Carleo A et al: A functional proteomics approach to the comprehension of sarcoidosis. *J Proteomics*, 2015; 128: 375–87
2. Perez RL, Rivera-Marrero CA, Roman J: Pulmonary granulomatous inflammation: From sarcoidosis to tuberculosis. *Semin Respir Infect*, 2003; 18(1): 23–32
3. Gupta D, Agarwal R, Aggarwal AN, Jindal SK: Molecular evidence for the role of mycobacteria in sarcoidosis: A meta-analysis. *Eur Respir J*, 2007; 30(3): 508–16
4. Sharma OP: Murray Kornfeld, American College of Chest Physician, and sarcoidosis: A historical footnote: 2004 Murray Kornfeld Memorial Founders Lecture. *Chest*, 2005; 128(3): 1830–35
5. Li QH, Zhang Y, Zhao MM et al: Simultaneous amplification and testing method for *Mycobacterium tuberculosis* rRNA to differentiate sputum-negative tuberculosis from sarcoidosis. *Am J Physiol Lung Cell Mol Physiol*, 2019; 316(3): L519–24
6. Maertzdorf J, Weiner J 3rd, Mollenkopf HJ et al: Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci USA*, 2012; 109(20): 7853–58
7. Guo T, Ma H, Zhou Y: Bioinformatics analysis of microarray data to identify the candidate biomarkers of lung adenocarcinoma. *Peer J*, 2019; 7: e7313
8. Landi C, Carleo A, Cillis G, Rottoli P: Sarcoidosis: Proteomics and new perspectives for improving personalized medicine. *Expert Rev Proteomics*, 2018; 15(10): 829–35
9. Kim MJ, Wainwright HC, Lockett M et al: Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *Embo Mol Med*, 2010; 2(7): 258–74
10. Katti MK: Assessment of serum IL-1, IL-2 and IFN-gamma levels in untreated pulmonary tuberculosis patients: Role in pathogenesis. *Arch Med Res*, 2011; 42(3): 199–201
11. Hong S, Mok Y, Jeon C et al: Tuberculosis, smoking and risk for lung cancer incidence and mortality. *Int J Cancer*, 2016; 139(11): 2447–55
12. Alberg AJ, Brock MV, Ford JG et al: Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*, 2013; 143(5 Suppl.): e15–e295
13. O'Callaghan DS, O'Donnell D, O'Connell F, O'Byrne KJ: The role of inflammation in the pathogenesis of non-small cell lung cancer. *J Thorac Oncol*, 2010; 5(12): 2024–36
14. Chang SH, Mirabolfathinejad SG, Katta H et al: T helper 17 cells play a critical pathogenic role in lung cancer. *Proc Natl Acad Sci USA*, 2014; 111(15): 5664–69
15. Sato T, Arai E, Kohno T et al: Epigenetic clustering of lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue: Its correlation with smoking history and chronic obstructive pulmonary disease. *Int J Cancer*, 2014; 135(2): 319–34
16. Chaturvedi AK, Caporaso NE, Katki HA et al: C-reactive protein and risk of lung cancer. *J Clin Oncol*, 2010; 28(16): 2719–26

In this study, we have identified a set of candidate biomarkers that can be exploited to study the pathogenesis, diagnosis, and prognosis of SARC.

Conclusions

In conclusion, we employed an integrated bioinformatics approach to identify the SARC-PTB common and SARC-only DEGs. A total of 13 hub genes from SARC-only DEGs were selected from the PPI network, among which 12 genes had the potential to be diagnostic biomarkers. CBX5, BCL11B, and CPR18 had the highest diagnostic potential. These genes are associated with cytokine-cytokine receptor interaction. However, this study had some limitations. This study was based on bioinformatics analysis of published data and lacks experimental verification. Thus, we could not determine the presence of a causal relationship between the differential expression of the hub genes and disease progression. Moreover, although we analyzed 3 GEO datasets, the sample size used was relatively small. This can lead to potential bias. Therefore, subsequent bioinformatics analyses and experimental verification with large sample sizes are necessary in the future to understand the molecular mechanism(s) employed by these novel hub genes in the pathogenesis of SARC and PTB. Furthermore, relevant models can also be constructed based on these genes to diagnose SARC and differentiate SARC from sputum negative PTB.

Conflict of interest

None.

17. Kumar S, Baghdadi S, Cale AR: Concurrence of sarcoidosis and lung cancer: A diagnostic dilemma. *Thorax*, 2006; 61(12): 1100
18. McNeill M, Zanders TB, Morris MJ: A 49-year-old man with concurrent diagnoses of lung cancer, sarcoidosis, and multiple regions of adenopathy on positron emission tomography. *Chest*, 2009; 135(2): 546–49
19. Tokuyasu H, Izumi H, Mukai N et al: Small cell lung cancer complicated by pulmonary sarcoidosis. *Intern Med*, 2010; 49(18): 1997–2001
20. Kachalia AG, Ochieng P, Kachalia K, Rahman H: Rare coexistence of sarcoidosis and lung adenocarcinoma. *Respir Med Case Rep*, 2014; 12: 4–6
21. Iijima Y, Sugiyama Y, Sawahata M et al: Clinical features of pulmonary sarcoidosis complicated by lung cancer. *Intern Med*, 2017; 56(15): 1957–60
22. Tsuchiya K, Karayama M, Sato T et al: Simultaneous occurrence of sarcoidosis and anti-neutrophil cytoplasmic antibody-associated vasculitis in a patient with lung cancer. *Intern Med*, 2019; 58(22): 3299–304
23. Svendsen CB, Milman N, Nielsen HW et al: A prospective study evaluating the presence of *Rickettsia* in Danish patients with sarcoidosis. *Scand J Infect Dis*, 2009; 41(10): 745–52
24. Planck A, Eklund A, Grunewald J, Vene S: No serological evidence of *Rickettsia helvetica* infection in Scandinavian sarcoidosis patients. *Eur Respir J*, 2004; 24(5): 811–13
25. Biberfeld P, Petren AL, Eklund A et al: Human herpesvirus-6 (HHV-6, HBLV) in sarcoidosis and lymphoproliferative disorders. *J Virol Methods*, 1988; 21(1–4): 49–59
26. Moravvej H, Vesal P, Abolhasani E et al: Comorbidity of leishmania major with cutaneous sarcoidosis. *Indian J Dermatol*, 2014; 59(3): 316
27. Esteves T, Aparicio G, Garcia-Patos V: Is there any association between Sarcoidosis and infectious agents?: A systematic review and meta-analysis. *BMC Pulm Med*, 2016; 16(1): 165
28. Derler AM, Eisendle K, Baltaci M et al: High prevalence of ‘*Borrelia-like*’ organisms in skin biopsies of sarcoidosis patients from Western Austria. *J Cutan Pathol*, 2009; 36(12): 1262–68
29. Martens H, Zollner B, Zissel G et al: Anti-*Borrelia burgdorferi* immunoglobulin seroprevalence in pulmonary sarcoidosis: A negative report. *Eur Respir J*, 1997; 10(6): 1356–58
30. Niimi T, Tomita H, Sato S et al: Vitamin D receptor gene polymorphism in patients with sarcoidosis. *Am J Respir Crit Care Med*, 1999; 160(4): 1107–9
31. Stjepanovic MI, Mihailovic-Vucinic V, Spasovski V et al: Genes and metabolic pathway of sarcoidosis: Identification of key players and risk modifiers. *Arch Med Sci*, 2019; 15(5): 1138–46
32. Riddle EL, Schwartzman RA, Bond M, Insel PA: Multi-tasking RGS proteins in the heart: the next therapeutic target? *Circ Res*, 2005; 96(4): 401–11
33. van Dijk N, Sprangers MA, Boer KR et al: Quality of life within one year following presentation after transient loss of consciousness. *Am J Cardiol*, 2007; 100(4): 672–76
34. Huang YJ, Zhou ZW, Xu M et al: Alteration of gene expression profiling including GPR174 and GNG2 is associated with vasovagal syncope. *Pediatr Cardiol*, 2015; 36(3): 475–80
35. Gibejova A, Mrazek F, Subrtova D et al: Expression of macrophage inflammatory protein-3 beta/CCL19 in pulmonary sarcoidosis. *Am J Respir Crit Care Med*, 2003; 167(12): 1695–703
36. Kriegova E, Tsyruynyk A, Arakelyan A et al: Expression of CCX CKR in pulmonary sarcoidosis. *Inflamm Res*, 2006; 55(10): 441–45
37. Arger NK, Ho M, Woodruff PG, Koth LL: Serum CXCL11 correlates with pulmonary outcomes and disease burden in sarcoidosis. *Respir Med*, 2019; 152: 89–96
38. Guo J, Zhang Z, Pan L, Zhou Y: Identification of miR-758-3p as potential modulator of CBX5 expression in gastric cancer. *Technol Cancer Res Treat*, 2018; 17: 1533033818816061
39. Wang J, Yang Z, Wu NP, Yang J: Increased expression of BCL11B and its recruited chromatin remodeling factors during highly active antiretroviral therapy synergistically represses the transcription of human immunodeficiency virus type 1 and is associated with residual immune activation. *Arch Virol*, 2020; 165(2): 321–30
40. Nourbakhsh F, Atabaki R, Roohbakhsh A: The role of orphan G protein-coupled receptors in the modulation of pain: A review. *Life Sci*, 2018; 212: 59–69