SARCOIDOSIS VASCULITIS AND DIFFUSE LUNG DISEASES 2025; 42 (4): 17074 DOI: 10.36141/svdld.v42i4.17074

Bioinformatics analysis identifies candidate biomarkers associated with sarcoidosis

Linling Jin[1], Chunfeng Pan[2]*, Yujie Sun[1], Jiayi Zhang[1], Weiping Xie[1], Mengyu He[1]*

[1]Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital with Nanjing Medical University, Nanjing, Jiangsu, China; [2]Department of Thoracic Surgery, The First Affiliated Hospital with Nanjing Medical University, Nanjing, Jiangsu, China

*These authors contributed equally to the work

Abstract. *Background and aim:* Sarcoidosis is a complex inflammatory disorder characterized by the formation of non-caseating granulomas, which can affect multiple systems, with a predominant impact on the lungs and thoracic lymph nodes. The aim of the study is to identify potential biomarkers and explore the infiltration of immune cells associated with sarcoidosis. *Methods:* The datasets of GSE19314, GSE83456, and GSE37912 were obtained from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were screened by comparing sarcoidosis samples to healthy controls in GSE19314 and GSE83456. Functional enrichment analyses and protein-protein interactions (PPIs) network construction were performed to elucidate the functional roles of the DEGs. Hub genes were identified through PPI network analysis. The GSE37912 dataset was used as a validation set. Immune cell infiltration in sarcoidosis patients was investigated. Furthermore, the trehalose 6,6′-dimycolate-granuloma (TDM)-induced granuloma experimental model was employed to resemble human sarcoid granulomas, and the expression of hub genes in lung tissues was detected by qRT-PCR. *Results:* A total of 71 common DEGs, including 53 upregulated DEGs and 18 downregulated DEGs, were identified between sarcoidosis patients and controls. The signal transducer and activator of transcription 1 (STAT1), C-X-C motif chemokine ligand 10 (CXCL10) and basic leucine zipper ATF-like transcription factor 2 (BATF2) were considered as hub genes, showing promising diagnostic potential for sarcoidosis. Immune infiltration analysis indicated that compared with the controls, sarcoidosis samples exhibited increased infiltration of activated NK cells, monocytes, macrophage, activated dendritic cells and resting mast cells. In the TDM-induced lung granuloma model, STAT1, CXCL10 and BATF2 expression was upregulated in sarcoidosis lung tissue. *Conclusions:* Bioinformatics analysis indicated that STAT1, CXCL10 and BATF2 may as candidate biomarkers associated with sarcoidosis.

Key words: sarcoidosis, differentially expressed genes, biomarkers, immune infiltration

Introduction

Sarcoidosis is a multisystem granulomatous disorder characterized by non-caseating granulomatous inflammation, capable of affecting virtually any organ system. The annual incidence ranges from 1 to 15 cases per 100000 individuals depending on geographic region, sex and ethnicity (1,2). While pulmonary involvement dominates clinical presentations, approximately 10% of patients progress to advanced fibrocystic lung disease, which is a critical prognostic determinant strongly correlated with mortality. Globally, the 5-year mortality rate for sarcoidosis is estimated at approximately 7% (3). Histopathologically, the disease hallmark manifests as tightly clustered epithelioid histiocytes and

Received: 15 April 2025
Accepted: 18 May 2025
Correspondence: Weiping Xie, Professor, M.D., Ph.D. & Mengyu He, M.D.
Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital with Nanjing Medical University, 300 Guangzhou Road, Nanjing, Jiangsu 210029, China
E-mail: wpxie@njmu.edu.cn & myhe@njmu.edu.cn

multinucleated giant cells encircled by lymphocytes, fibroblasts, and plasma cells, forming granulomas (4). Despite these defining features, accurate recognition of sarcoidosis still remains difficult due to absence of disease-specific biomarkers, ambiguous clinical presentations, and its ability to masquerade as other diseases. Consequently, there is a high rate of misdiagnosis and delayed diagnosis of sarcoidosis. Accumulating evidence has shown that both innate and adaptive immune responses are critically involved in the pathogenesis of sarcoidosis (5). Antigen-presenting cells (APCs), such as macrophages and dendritic cells (DCs), initiate granuloma formation by processing persistent presenting antigens and activating $CD4^+$ T helper (Th) cells. A dysregulated Th1/Th17 response, characterized by up-regulation of Th1-specific transcription factors, promotes macrophage aggregation and differentiation into epithelioid and multinucleated giant cells. Concurrently, regulatory T cells (Tregs) fail to suppress this inflammatory cascade, while B cells contribute via autoantibody production and cytokine signaling (6). Recent research also implicates that loss of immunoregulatory functions mediated by NKT-like may contribute to $CD4^+$ lymphocyte overactivity, thereby facilitating granuloma formation (7). Collectively, these findings highlight a strong association between sarcoidosis progression and immune cell infiltration. The novel bioinformatics method, Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT), was employed to investigate the discrepancy in 22 human immune cell subtypes between sarcoidosis patients and healthy controls (8). In the present study, sarcoidosis-related common differentially expressed genes (DEGs) were obtained by analyzing publicly available microarray datasets comparing sarcoidosis and healthy samples. Then, we conducted bioinformatics systems biology analysis to investigate the underlying gene interactions and molecular mechanisms. The diagnostic biomarkers of sarcoidosis were selected and further to be validated. CIBERSORT was utilized to study the roles of different immunocytes in the development of sarcoidosis. Additionally, experimental models were applied to validate signal transducer and activator of transcription 1 (STAT1), C-X-C motif chemokine ligand 10 (CXCL10) and basic leucine zipper ATF-like transcription factor 2 (BATF2) as key regulators and potential therapeutic targets for sarcoidosis.

Materials and methods

Data collection

Using the NCBI-Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/), three datasets associated with sarcoidosis were chosen and downloaded. Microarray data of the GSE19314 (9,10), GSE83456 (11) and GSE37912 datasets (12) were obtained. The detailed information and functions of the datasets were presented in the Table 1. Only the samples of sarcoidosis and healthy individuals were collected for subsequent analysis.

2.2 Identification of DEGs

To screen the disparities between the sarcoidosis group and the control group, we utilized the R package "limma" to standardize datasets and uncover the DEGs in GSE19314 and GSE83456 (13). Genes with fold change (|logFC|) >0.5 and adjusted *P* value <0.05 using the false discovery rate (FDR) were identified as common DEGs. Volcano maps depicting the DEGs were generated using the "ggplot2" package. Heatmaps illustrating the top 20 DEGs with the highest and lowest differential expressions were created by the "ComplexHeatmap" package.

Table 1. Characteristics of the included datasets

ID	Source	Platform	No. of sample	Usage
GSE19314	PBMC	GPL570	20 healthy control, 37 sarcoidosis	Identification of hub genes
GSE83456	PBMC	GPL10558	61 healthy control, 49 sarcoidosis	Identification of hub genes, verification of diagnostic markers and CIBERSORT analysis
GSE37912	PBMC	GPL5175	35 healthy control, 39 sarcoidosis	Verification of hub genes and diagnostic markers

Abbreviation: PBMC, peripheral blood mononuclear cells

The significant DEGs in the above datasets were displayed through column charts.

Functional enrichment analysis

Gene Ontology (GO) functional analysis, including the biological process (BP), cellular component (CC), molecular function (MF) enrichment, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using online tool Metascape (https://metascape.org/) (14) with P <0.01 as the cutoff value.

Protein-protein interaction (PPI) network analysis

Network analysis is an important part of systems biology and is used to understand the interactions between molecules and proteins. To investigate the relations of DEGs, Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org/) online database was used with a threshold interaction score of 0.4. The PPI network was imported into the Cytoscape software (version 3.9.1) for visualization.

Hub genes identification

The hub genes were screened out using Cytoscape plugin tool CytoHubba. The top ten nodes were considered as hub genes according to the topological algorithms of degree. Molecular Complex Detection (MCODE) was applied to identify highly interconnected portions within the PPI network. Expression data from GSE37912 were extracted to validate the statistical significance of hub genes. Statistical analysis including t-test and non-parametric tests were performed using GraphPad Prism (Version 6.0.1). Statistical significance was considered as $P < 0.05$.

Screening and verifying diagnostic markers

We used the expression data of GSE37912 and GSE83456 datasets to construct the receiver operating characteristic (ROC) curves of the verified hub genes. The normalized expression values of each hub gene (and their multigene combination) were entered as covariates in a binary logistic regression model with diagnostic status in SPSS (version 26). The area under the ROC curve (AUC) were used to assess the diagnostic efficiency. AUC > 0.5 was considered statistically significant.

Evaluating immune cell infiltration

CIBERSORT algorithm was conducted to assess immune cell infiltration in the microenvironment, containing 22 kinds of immunocytes. The data from GSE83456 were submitted into the CIBERSORT to get the relative proportions of infiltrated immune cells. A histogram was used to display the percentage of each immune cells in the samples and comparison of immunocyte composition between sarcoidosis patients and healthy individuals was assessed using SangerBox (sangerbox.com) (15). Significance was defined as $P < 0.05$.

Experimental protocol

The trehalose 6.6′-dimycolate (TDM) (Enzo Life Sciences GmbH, Germany) extracted from Mycobacterium tuberculosis was prepared in a water-in-oil emulsion using incomplete Freund's Adjuvant (Sigma Aldrich, America). C57BL/6J mice, aged between six to twelve weeks, were injected with 1 μg of TDM in 10 μl water-in-oil emulsion per gram of body weight via the tail vein. To evaluate the development of pulmonary granulomas, lung samples were collected seven days after TDM injection (n=8 per group). Lung tissues were fixed with 4% polyformaldehyde for 24 hours and embedded in paraffin, which subsequently stained with hematoxylin and eosin (H&E). All pictures were captured using a Leica 2500 microscope (Wetzlar, Germany).

cDNA synthesis and quantitative polymerase chain reaction (qRT-PCR)

Trizol reagent (Gibco BRL, Grand Island, NY) was used to extract total RNA from lung tissues. Reverse transcription was used with 500 ng of total RNA with SYBR®Premix Ex Taq™ (TaKaRa, Shiga, Japan). qRT-PCR was carried out using an ABI Prism 7500 FAST apparatus (Applied Bio-systems, Foster City, CA, USA), with β-actin as the housekeeping gene. PCR primers of target genes were listed in Table 2. The data were presented as mean ± standard deviation (SD) and analyzed using Student's t-test with GraphPad Prism 6 (GraphPad Inc., La Jolla, CA, USA). Statistical significance was confirmed at P <0.05.

Table 2. The primer sequences of targeted genes

Gene primer	Species		Sequence (5' to 3')
STAT1	Mouse	Forward	TCACAGTGGTTCGAGCTTCAG
		Reverse	CGAGACATCATAGGCAGCGTG
CXCL10	Mouse	Forward	CCAAGTGCTGCCGTCATTTTC
		Reverse	GGCTCGCAGGGATGATTTCAA
BATF2	Mouse	Forward	GAAGCACACCAGTAAGGCG
		Reverse	GCACAGGCGTTCATGCAAG
β-actin	Mouse	Forward	GATTACTGCTCTGGCTCCTAGC
		Reverse	GACTCATCGTACTCCTGCTTGC

Abbreviations: BATF2, basic leucine zipper ATF-like transcription factor 2; CXCL10, C-X-C motif chemokine ligand 10; STAT1, signal transducer and activator of transcription 1.

Results

Identification of DEGs

In the analysis of differential expression, sarcoidosis was regarded as the case group. There were 173 identified DEGs including 73 upregulated and 100 downregulated genes in the GSE19314 dataset (Figure 1A); while 575 DEGs were screened out including 374 upregulated and 201 downregulated genes in the GSE83456 dataset (Figure 1C). The top 20 DEGs with the highest significance were shown in Figure 1B and 1D. Following the intersection of DEGs between the two datasets, 53 upregulated DEGs and 18 downregulated DEGs were obtained for subsequent analysis (Figure 2A a nd 2B).

GO and KEGG enrichment analysis

Metascape was applied to perform function enrichment analysis, and the top 5 items for BP, CC, and MF pathways of GO analysis were plotted (Figure 3A). DEGs were significantly enriched in the innate immune response as well as defense response to protozoan in BP subsets. Primary lysosome and specific granule pathways were enriched in CC subsets. Pathways involved in protein homodimerization activity and heparin binding were enriched in MF subsets. KEGG pathway enrichment analysis indicated that the DEGs were primarily associated with nucleotide-binding oligomerization domain (NOD)-like receptor (NLR) signaling, acute myeloid leukemia, and thyroid hormone signaling pathways (Figure 3B).

PPI network construction and hub gene identification

Common DEGs were uploaded to the STRING database to detect their interactions. The PPI network, as shown in Figure 4A, consisted of 48 nodes and 249 edges. Subsequently, the top 10 hub genes according to degree analytical methods were determined, including STAT1, CXCL10, GBP5, GBP1, GBP4, BATF2, RSAD2, PARP9, IFI44 and EPSTI1 (Table 3 and Figure 4B). In addition, we used Cyotscape plugin MCODE to obtain a highly clustered region of PPI networks. The clustering network contained all 10 hub genes (Figure 4C). After validation with GSE37912 dataset, only genes of STAT1, CXCL10 and BATF2 exhibited statistical significance (Figure 4D).

Screening and validating the diagnostic markers

The diagnostic performance of STAT1, CXCL10, and BATF2 was assessed using AUC analysis. In GSE37912, STAT1, CXCL10, and BATF2 individually achieved AUCs of 0.652 [95% confidence interval (CI): 0.523-0.782], 0.573 (95% CI: 0.649-0.871), and 0.760 (95% CI: 0.649-0.871), respectively (Figure 5A). The logistic regression model combining all three markers increased the AUC of 0.803 (95% CI: 0.700-0.906) (Figure 5C). Furthermore, GSE83456 dataset was used to verify the accuracy and efficiency of the hub

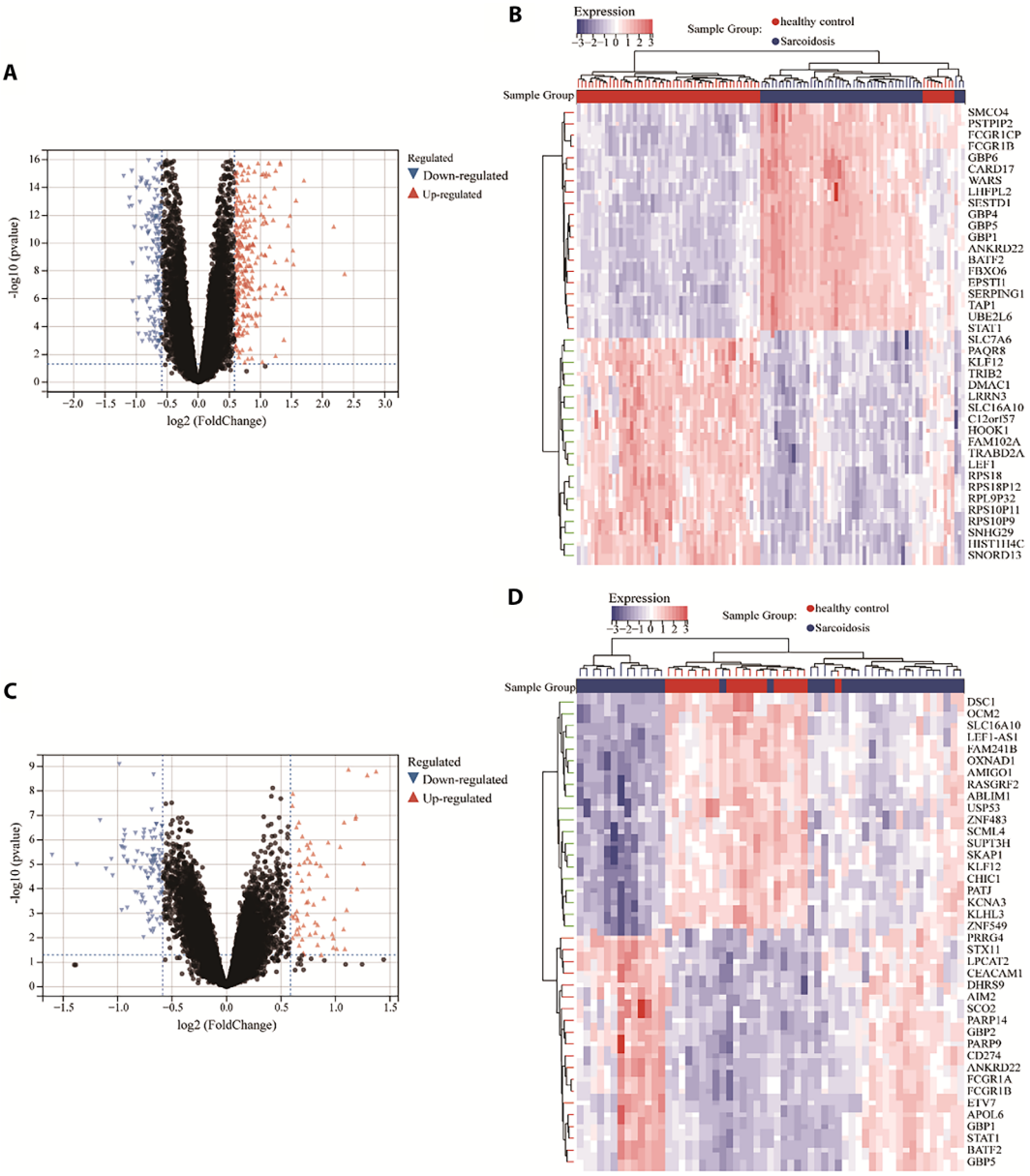

Figure 1. Volcano map and heatmap of DEGs. (A) Volcano map showing the DEGs of GSE19314; (C) Volcano map showing the DEGs of GSE83456; red indicates the up-regulation of DEGs, blue indicates down-regulation of DEGs. (B) Heatmap of DEGs in GSE19314; (D) Heatmap of DEGs in GSE83456; red represents the healthy control group, and blue represents the sarcoidosis group.

gene diagnostic capabilities. The results showed that the AUCs of STAT1, CXCL10 and BATF2 individually were 0.974 (95% CI: 0.949-0.998), 0.942 (CI: 0.902-0.983), and 0.980 (CI: 0.956-1.000), respectively, with the combined marker model yielding an AUC of 0.980 (95% CI: 0.955-1.000) (Figure 5B, D).

Immune cell infiltration analysis

The abundances of immunocytes in both health controls and sarcoidosis samples were measured by CIBERSORT (Figure 6A). The results showed that activated NK cells, monocytes, macrophage

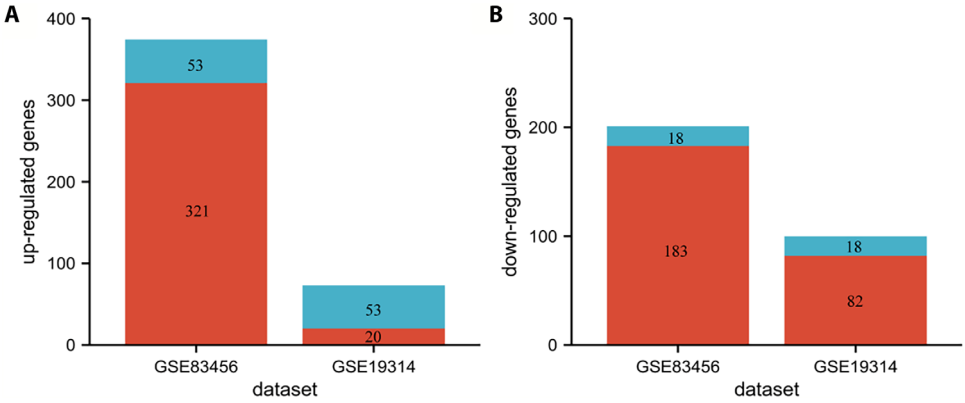

Figure 2. Column chart showing the number of DEGs, including shared (blue) and dataset-specific (red) up-regulated (A) and down-regulated (B) genes between datasets GSE83456 and GSE19314.

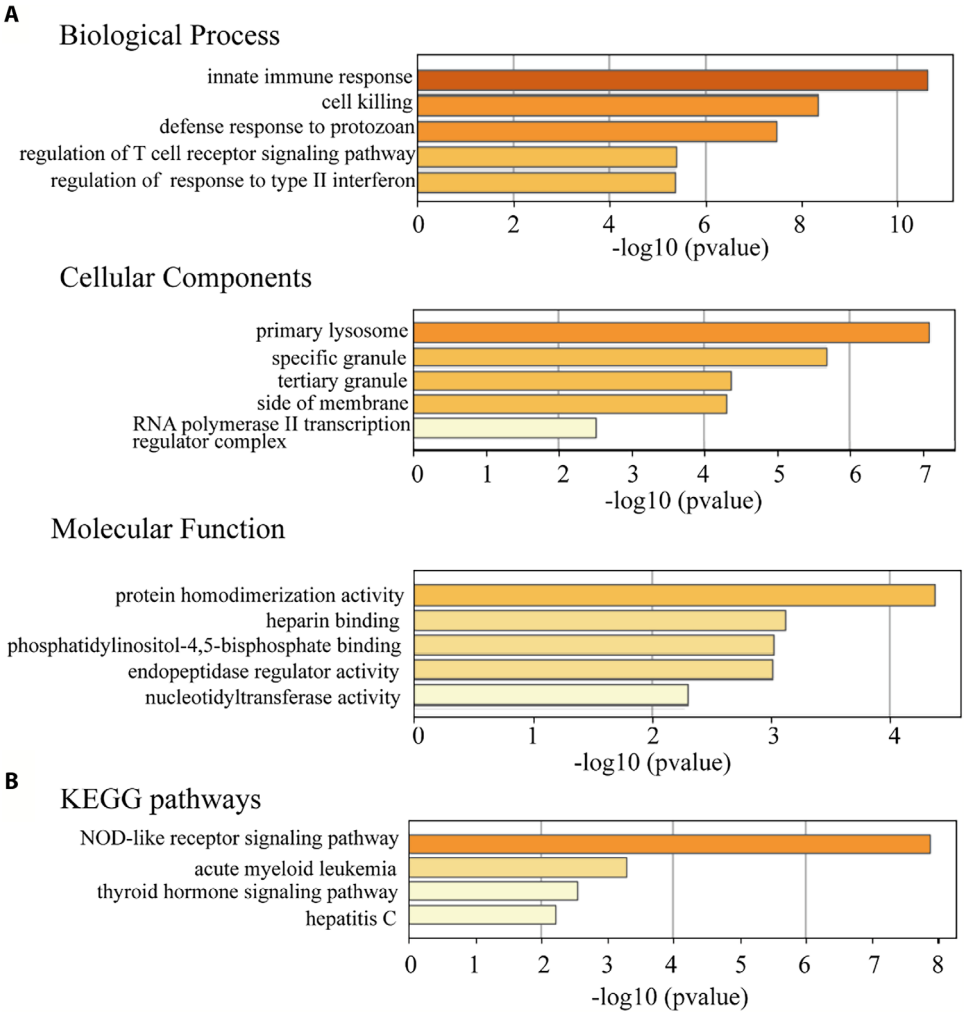

Figure 3. Gene ontologies and pathways for shared DEGs based on combined score. (A) Go enrichment analysis histogram. (B) KEGG pathway analysis histogram.

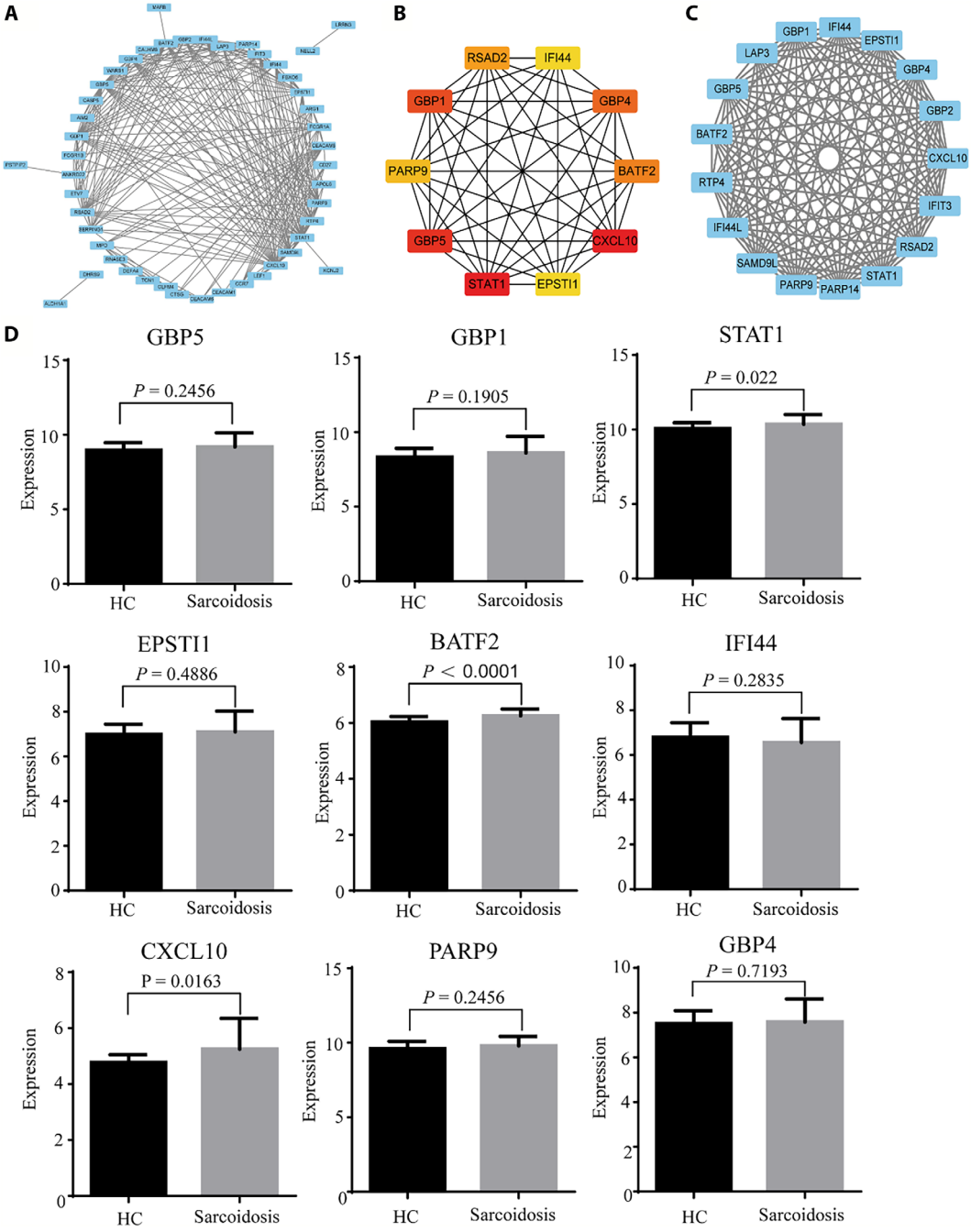

Figure 4. PPI network and hub genes identification for common DEGs. (A) PPI network for shared DEGs generated by STRING. (B) The hub genes identified using CytoHubba plugin based on the degree value. The score is reflected in color, and darker colors indicate hub genes ranked higher. (C) Cluster analysis network extracted from PPI network using MCODE plugin. (D) The expression of hub genes in the validation dataset. HC, health controls.

(M0, M1, and M2), activated DCs and mast cell resting were higher infiltrated in sarcoidosis, while $CD8^+$ T cells, naive $CD4^+$ T cells and T cells follicular helper (T_{fh}) were found to be lower infiltrated in sarcoidosis (Figure 6B).

Expression of STAT1, CXCL10, and BATF2 in lung tissues.

The TDM-induced lung granuloma in mice showed similarities to the granulomas in human

Table 3. Degree of top 10 genes in top module

Gene ID	Gene name	Degree
STAT1	signal transducer and activator of transcription 1	27
CXCL10	C-X-C motif chemokine ligand 10	27
GBP5	guanylate binding protein 5	26
GBP1	guanylate binding protein 1	24
GBP4	guanylate binding protein 4	21
BATF2	basic leucine zipper ATF-like transcription factor 2	20
RSAD2	radical S-adenosyl methionine domain containing 2	19
PARP9	poly (ADP-ribose) polymerase family member 9	18
IFI44	interferon induced protein 44	17
EPSTI1	epithelial stromal interaction 1	17

Abbreviations: BATF2, basic leucine zipper ATF-like transcription factor 2; CXCL10, C-X-C motif chemokine ligand 10; EPSTI1, epithelial stromal interaction 1; GBP1, guanylate binding protein 1; GBP4, guanylate binding protein 4; GBP5, guanylate binding protein 5; IFI44, interferon-induced protein 44; PARP9, poly (ADP-Ribose) polymerase family member 9; RSAD2, radical S-Adenosyl methionine domain containing 2; STAT1, signal transducer and activator of transcription 1.

sarcoidosis (Figure 7A). Compared with control group, the expression levels of STAT1, CXCL10 and BATF2 were significantly upregulated in the sarcoidosis group (Figure 7B).

Discussion

Immune cell infiltration plays a central role in sarcoidosis pathogenesis, contributing to granulomas formation and tissue damage (16). Current diagnosis relies on radiological and clinical imaging, confirmed by invasive biopsy for histopathology. Recent increases in sarcoidosis diagnoses reflect greater awareness, improved imaging, and minimally invasive techniques. However, sarcoidosis diagnose remains challenging due to its varied clinical presentation and similarities to other diseases, such as tuberculosis and malignancies, resulting in misdiagnosis and delayed treatment (17). Thus, identifying specific biomarkers is essential to enhance diagnostic accuracy and guide patient management. In this study, bioinformatics tools were utilized to help screen the potential biomarkers and provide in-depth insights into the roles of immune cells in the development of sarcoidosis. The current study identified 53 upregulated and 18 downregulated sarcoidosis-related DEGs from GEO database containing 81 healthy people and 86 sarcoidosis patients. GO and associated pathways were further detected. It showed that innate immune response was one of the most prominent pathways in biological processes. KEGG analysis revealed that the most significantly enriched pathway was NLR signaling. Traditionally, sarcoidosis has been regarded as a Th1/Th17-mediated disorder (18). Recent findings indicate that animal models completely lacking adaptive immunity can still generate epithelioid granulomas stimulating by mycobacterium strains, suggesting a vital contribution by innate immune system in driving the granulomas formation (19). Activation of inflammasomes, overproduction of pro-inflammatory cytokines, M1/M2 macrophage polarization and reduced other innate/innate-like immune cells were involved in sarcoidosis development (20,21). NLRs are cytoplasm-localized pattern-recognition molecules (PRMs) that are crucial in the innate immune response. Aberrant sensing through NOD1 evoked a sustained mitogen-activated protein kinase (MAPK) phosphorylation resulting to sarcoidosis (22). In addition, activated NLRs are closely related to the onset of sarcoidosis due to triggering the formation of inflammasomes (23). Our analysis aligns with prior studies, strongly implicating that dysregulated immune responses are fundamental in the development of sarcoidosis. PPI network was conducted and visualized using STRING and Cytoscape software, from which ten hub genes were identified. Further verification has confirmed the association of STAT1, CXCL10 and BATF2 with the pathogenesis of sarcoidosis. Moreover, using the previously described TDM-induced granulomas method (20), we established a murine sarcoidosis model that resembles human granuloma formation and confirmed the increased expression of STAT1, CXCL10 and BATF2 *in vivo*. In additionally, these genes have been identified as potential diagnostic markers for sarcoidosis. Previous work has found that STAT1 and the downstream chemokine CXCL10 are upregulated in both lung tissue and lymph node of patients with sarcoidosis (24). The STAT proteins are a family of transcription factors that play crucial roles in regulating genes involved in inflammatory responses. The activation of IFN-γ signaling is partially relied on the Janus kinase (JAK)/STAT pathway, in which STAT1 mainly mediates the accumulation and activation of macrophages in

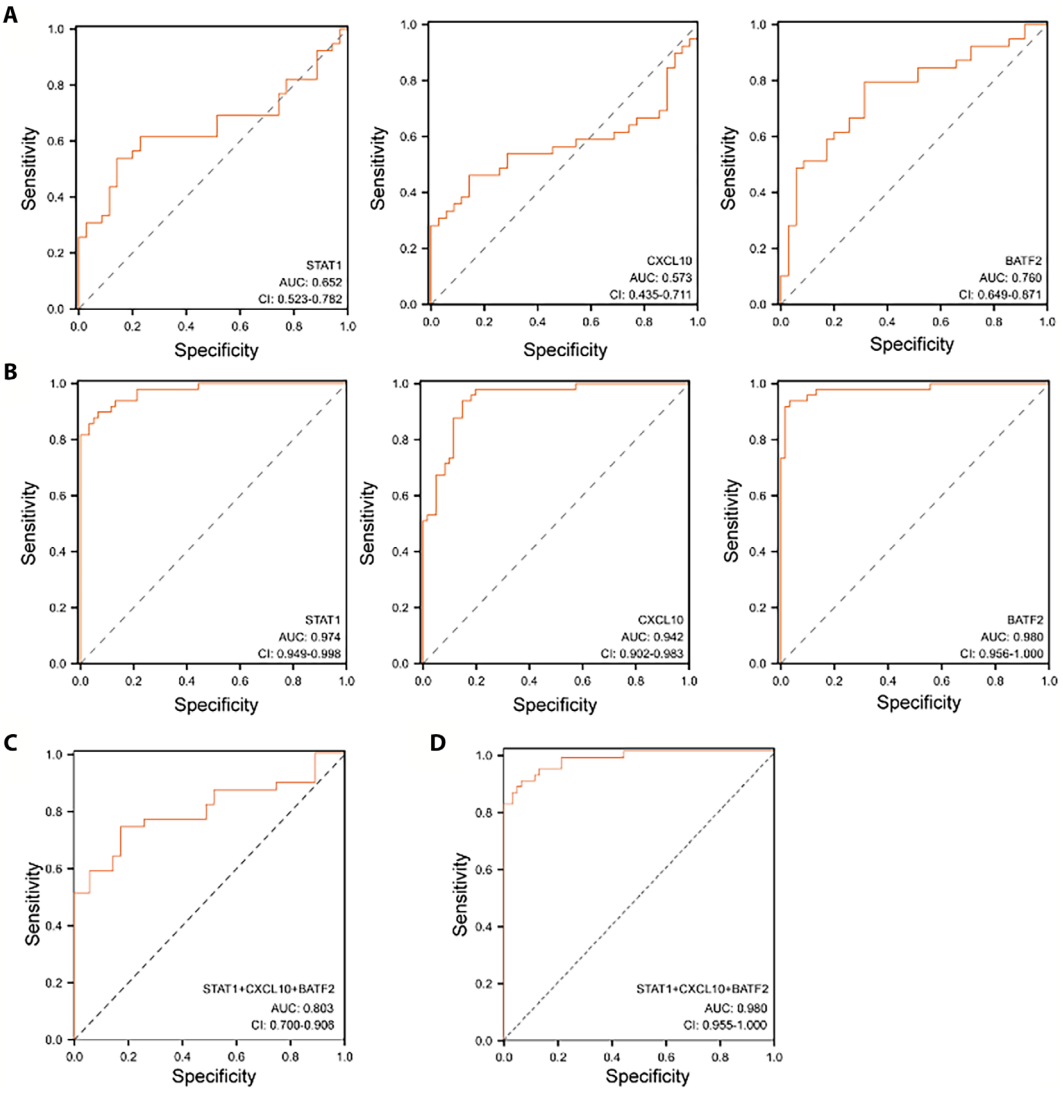

Figure 5. The receiver operating characteristic (ROC) curve of identified diagnostic markers. (A and C) The ROC curve of the diagnostic efficacy in GSE37912. (B and D) The ROC curve of the diagnostic efficacy in GSE83456.

sarcoidosis granuloma (25). CXCL10, a STAT1-dependent chemokine, correlates with worse respiratory outcomes in sarcoidosis, with higher level linked to reduced lung function and greater dyspnea scores (26). Therefore, STAT1 and CXCL10 may be used to evaluate not only the diagnosis but also the prognosis of sarcoidosis. Research on the relationship between BATF2 and sarcoidosis has remained limited. He et al. reported that serum BATF2 levels were elevated in pulmonary sarcoidosis patients compared with healthy controls (27). BATF2 is regarded as an important transcription regulator of the innate immune system. BATF2 knockdown inhibited the expression of immune regulatory genes, such as Tnf, Cxcl9, Cxcl11 and Nos2, in IFNγ- and LPS-stimulated macrophages (28). It also contributed to the early host defense during infection with pulmonary Klebsiella pneumoniae (Kp) (29). Moreover, BATF2 plays a predominate role in regulating Th1 and Th17 cells differentiation and activation (30). Given that, BATF2 may contribute to the pathological process of sarcoidosis by regulating abnormal inflammatory responses. Interestingly, we observed that the diagnostic accuracy of our candidate DEGs was substantially higher in GSE83456 than in GSE37912. Several factors may underlie this discrepancy. First,

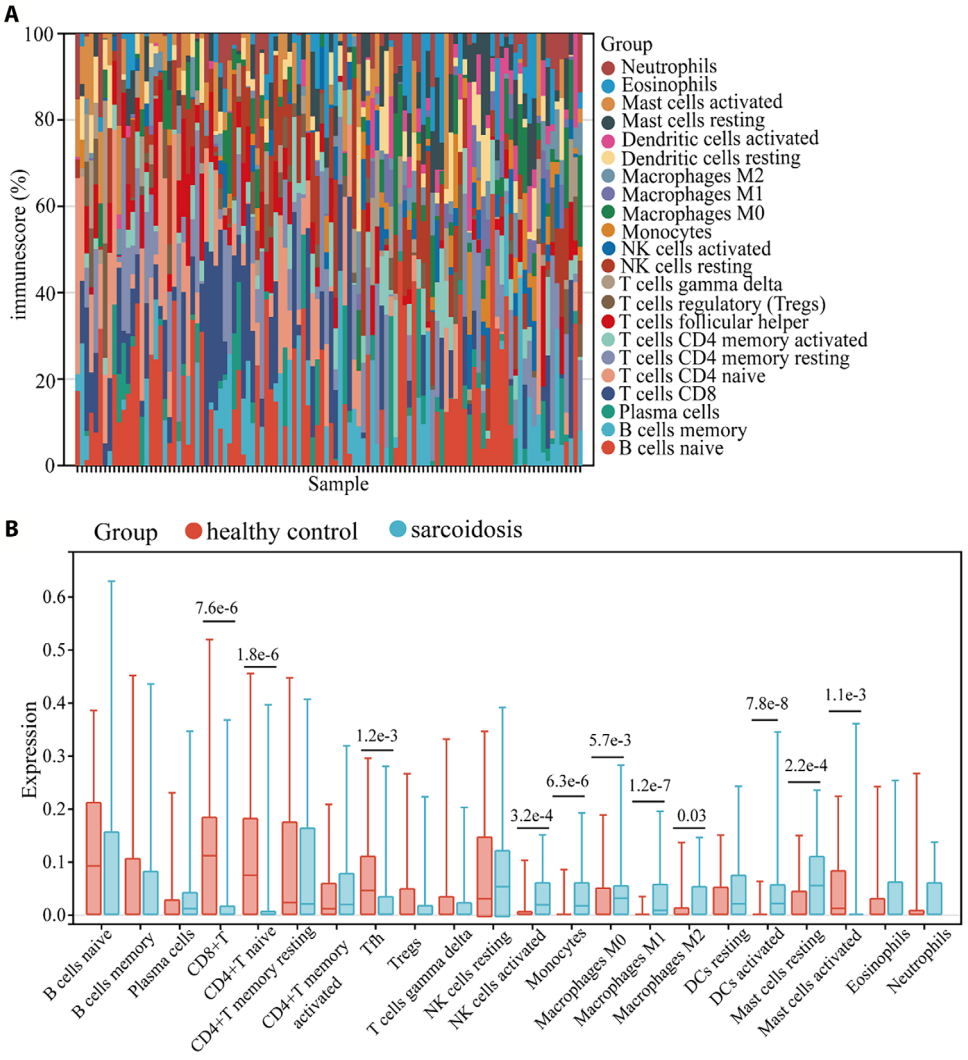

Figure 6. Evaluation and visualization of immune cell infiltration. (A) A histogram displayed the percentage of each immune cells in the samples of sarcoidosis patients and healthy individuals. (B) Violin diagram revealing the proportion of 22 types of immune cells between healthy controls and sarcoidosis patients.

although both datasets profiled peripheral blood from sarcoidosis patients and healthy controls, they were generated on different microarray platforms (Illumina HumanHT-12 V4.0 in GSE83456 versus Affymetrix in GSE37912), which may lead to potential platform-specific biases and batch effects. In addition, the clinical and demographic characteristics of the two cohorts differ: GSE37912 comprised 39 patients, 56.4% of whom had complicated sarcoidosis (mean age 50.8 ± 11.2 years; 45% male), whereas GSE83456 included patients exclusively with mediastinal involvement (mean age 47.8 ± 13.0 years; 59% male). These differences in disease stage and demographic profiles may alter gene expression patterns and thus impact the measured diagnostic performance. Emerging evidence highlights the critical role of infiltrating immune cells and their interactions in sarcoidosis pathogenesis. CIBERSORT analysis showed that compared with healthy controls, activated NK cells, monocytes, macrophage (M0, M1, M2), activated dendritic cells and mast cell resting were increased infiltrated in sarcoidosis, while $CD8^+$ T cells, naive $CD4^+$ T cells and T_{fh} were lower in sarcoidosis. This pronounced immune-cell imbalance underlies disease development. Environment antigens were captured by APCs and presented via human leukocyte antigen (HLA) molecules to naive T lymphocytes, which then differentiated into

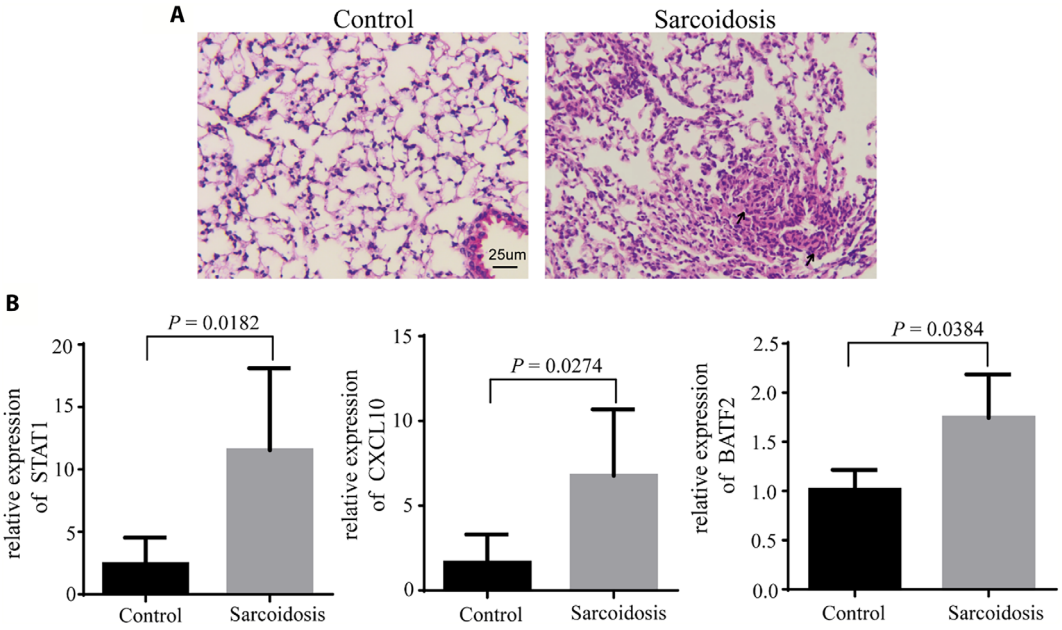

Figure 7. TDM-induced pulmonary granuloma in mice. (A) H&E staining of lung tissue harvested from TDM-treated mice on day 7. (B) Comparison of STAT1, CXCL10, and BATF2 mRNA expression in lung tissue of mice in sarcoidosis and control groups.

$CD4^+$, $CD8^+$, Treg and Th17 cells, proliferated and accumulated at the inflammatory site (31). Previous studies have shown an increased infiltrated of M1 macrophages in the airspace of sarcoidosis samples (32); while the M2 macrophages were found within fibrotic areas in muscle tissue from systemic sarcoidosis patients (33). Moreover, bronchoalveolar lavage fluid (BALF) from sarcoidosis patients shows elevated NK and NKT-like cell counts (34). Additionally, Lepzien et al. indicated that increased monocytes/monocyte-derived cells in peripheral blood and BALF contributed to inflammatory responses in sarcoidosis via promoting TNF production (35). Together, these data revealed that immune dysregulation, particularly the altered phenotypes and functions of 22 immune cell types, may drive sarcoidosis pathology and offer deep insight into its underlying mechanisms. Some limitations of our study need to be acknowledged. First, while we identified hub genes distinguishing sarcoidosis from healthy controls, the analysis did not assess their diagnostic specificity against clinically relevant mimics including granulomatous interstitial lung diseases, autoimmune disorders, and infectious granulomatosis such as tuberculosis. Second, our study was largely descriptive and did not elucidate the mechanistic roles of these genes in sarcoidosis. In conclusion, STAT1, CXCL10, and BATF2 were identified as potential biomarkers associated with sarcoidosis pathogenesis, underscoring the central role of immune dysregulation in its development and progression.

Conflict of Interest: Each author declares that he or she has no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangement etc.) that might pose a conflict of interest in connection with the submitted article.

Data Availability: The datasets used and analyzed in the current study are available from the corresponding author upon reasonable request.

Ethics Statement: Institutional Review Board approval of this study was obtained from the Ethics Committee of Nanjing Medical University First Affiliated Hospital (2025-SR-433). All experimental procedures involving animals were approved by the Institutional Animal Care and Use Committee of Nanjing Medical University (NJMU/IACUC-2403057).

Author Contributions: LJ, CP, YS, JZ, WX, and MH designed the research. LJ, CP, YS, JZ, and MY performed the experiments and data analysis. LJ, CP, YS, JZ, MY, and WX contributed to the manuscript drafting.

Funding: This work was funded by the National Natural Science Foundation of China (82000061), the Key Project of National Science & Technology for Infectious Diseases of China (2018ZX10722301), and the Jiangsu Province Capability Improvement Project through Science, Technology and Education;

Jiangsu Provincial Medical Innovation Center (CXZX202206), and the Young Scholars Fostering Fund of the First Affiliated Hospital of Nanjing Medical University (No. PY2022012).

References

1. Rossides M, Darlington P, Kullberg S, Arkema EV. Sarcoidosis: epidemiology and clinical insights. J Intern Med. 2023;293(6):668–80. doi:10.1111/joim.13629
2. Vij O, Dey M, Morrison K, Kouranloo K. Incidence, management and prognosis of new-onset sarcoidosis post COVID-19 infection. Sarcoidosis Vasc Diffuse Lung Dis. 2024;41(1):e2024004. doi:10.36141/svdld.v41i1.15027
3. Belperio JA, Shaikh F, Abtin FG, et al. Diagnosis and treatment of pulmonary sarcoidosis: a review. JAMA. 2022;327(9):856–67. doi:10.1001/jama.2022.1570
4. Nunes H, Brillet PY, Bernaudin JF, Gille T, Valeyre D, Jeny F. Fibrotic pulmonary sarcoidosis. Clin Chest Med. 2024;45(1):199–212. doi:10.1016/j.ccm.2023.08.011
5. Lee S, Birnie D, Dwivedi G. Current perspectives on the immunopathogenesis of sarcoidosis. Respir Med. 2020;173:106161. doi:10.1016/j.rmed.2020.106161
6. Drent M, Crouser ED, Grunewald J. Challenges of sarcoidosis and its management. N Engl J Med. 2021;385(11):1018–32. doi:10.1056/NEJMra2101555
7. Bergantini L, Cameli P, d'Alessandro M, et al. NK and NKT-like cells in granulomatous and fibrotic lung diseases. Clin Exp Med. 2019;19(4):487–94. doi:10.1007/s10238-019-00578-3
8. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773–82. doi:10.1038/s41587-019-0114-2
9. Koth LL, Solberg OD, Peng JC, Bhakta NR, Nguyen CP, Woodruff PG. Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. Am J Respir Crit Care Med. 2011;184(10):1153–63. doi:10.1164/rccm.201106-1143OC
10. Su R, Li MM, Bhakta NR, et al. Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes. Eur Respir J. 2014;44(4):985–93. doi:10.1183/09031936.00039714
11. Blankley S, Graham CM, Turner J, et al. The transcriptional signature of active tuberculosis reflects symptom status in extra-pulmonary and pulmonary tuberculosis. PLoS One. 2016;11(10):e0162220. doi:10.1371/journal.pone.0162220
12. Zhou T, Zhang W, Sweiss NJ, et al. Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. PLoS One. 2012;7(9):e44818. doi:10.1371/journal.pone.0044818
13. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi:10.1093/nar/gkv007
14. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523. doi:10.1038/s41467-019-09234-6
15. Shen W, Song Z, Zhong X, et al. Sangerbox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform. iMeta. 2022;1(3):e36. doi:10.1002/imt2.36
16. Greaves SA, Atif SM, Fontenot AP. Adaptive immunity in pulmonary sarcoidosis and chronic beryllium disease. Front Immunol. 2020;11:474. doi:10.3389/fimmu.2020.00474
17. Brennan M, Breen D. Sarcoidosis in the older person: diagnostic challenges and treatment consideration. Age Ageing. 2022;51(9):afac203. doi:10.1093/ageing/afac203
18. Facco M, Cabrelle A, Teramo A, et al. Sarcoidosis is a Th1/Th17 multisystem disorder. Thorax. 2011;66(2):144–50. doi:10.1136/thx.2010.140319
19. Davis JM, Clay H, Lewis JL, Ghori N, Herbomel P, Ramakrishnan L. Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. Immunity. 2002;17(6):693–702. doi:10.1016/s1074-7613(02)00475-2
20. Huppertz C, Jager B, Wieczorek G, et al. The NLRP3 inflammasome pathway is activated in sarcoidosis and involved in granuloma formation. Eur Respir J. 2020;55(3):00119-2019. doi:10.1183/13993003.00119-2019
21. Locke LW, Crouser ED, White P, et al. IL-13-regulated macrophage polarization during granuloma formation in an in vitro human sarcoidosis model. Am J Respir Cell Mol Biol. 2019;60(1):84–95. doi:10.1165/rcmb.2018-0053OC
22. Rastogi R, Du W, Ju D, et al. Dysregulation of p38 and MKP-1 in response to NOD1/TLR4 stimulation in sarcoid bronchoalveolar cells. Am J Respir Crit Care Med. 2011;183(4):500–10. doi:10.1164/rccm.201005-0792OC
23. Kim YK, Shin JS, Nahm MH. NOD-like receptors in infection, immunity, and diseases. Yonsei Med J. 2016;57(1):5–14. doi:10.3349/ymj.2016.57.1.5
24. Rosenbaum JT, Pasadhika S, Crouser ED, et al. Hypothesis: sarcoidosis is a STAT1-mediated disease. Clin Immunol. 2009;132(2):174–83. doi:10.1016/j.clim.2009.04.010
25. Kraaijvanger R, Janssen Bonas M, Vorselaars ADM, Veltkamp M. Biomarkers in the diagnosis and prognosis of sarcoidosis: current use and future prospects. Front Immunol. 2020;11:1443. doi:10.3389/fimmu.2020.01443
26. Arger NK, Ho ME, Allen IE, Benn BS, Woodruff PG, Koth LL. CXCL9 and CXCL10 are differentially associated with systemic organ involvement and pulmonary disease severity in sarcoidosis. Respir Med. 2020;161:105822. doi:10.1016/j.rmed.2019.105822
27. He J, Li X, Zhou J, Hu R. BATF2 and PDK4 as diagnostic molecular markers of sarcoidosis and their relationship with immune infiltration. Ann Transl Med. 2022;10(2):106. doi:10.21037/atm-22-180
28. Roy S, Guler R, Parihar SP, et al. Batf2/Irf1 induces inflammatory responses in classically activated macrophages, lipopolysaccharides, and mycobacterial infection. J Immunol. 2015;194(12):6035–44. doi:10.4049/jimmunol.1402521
29. van der Geest R, Penaloza HF, Xiong Z, et al. BATF2 enhances proinflammatory cytokine responses in macrophages and improves early host defense against pulmonary Klebsiella pneumoniae infection. Am J Physiol Lung Cell Mol Physiol. 2023;325(5):L604–16. doi:10.1152/ajplung.00441.2022
30. Li C, Liu M, Liu K, et al. BATF2 balances the T cell-mediated immune response of CADM with an anti-MDA5 autoantibody. Biochem Biophys Res Commun. 2021;551:155–60. doi:10.1016/j.bbrc.2021.02.128
31. Baughman RP, Culver DA, Judson MA. A concise review of pulmonary sarcoidosis. Am J Respir Crit Care Med. 2011;183(5):573–81. doi:10.1164/rccm.201006-0865CI
32. Wojtan P, Mierzejewski M, Osinska I, Domagala-Kulawik J. Macrophage polarization in interstitial lung diseases. Cent Eur J Immunol. 2016;41(2):159–64. doi:10.5114/ceji.2016.60990
33. Prokop S, Heppner FL, Goebel HH, Stenzel W. M2 polarized macrophages and giant cells contribute to myofibrosis in neuromuscular sarcoidosis. Am J Pathol. 2011;178(3):1279–86. doi:10.1016/j.ajpath.2010.11.065
34. Manika K, Domvri K, Kyriazis G, Kontakiotis T, Papakosta D. BALF and blood NK cells in different stages of pulmonary sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis. 2022;38(4):e2021039. doi:10.36141/svdld.v38i4.10810
35. Lepzien R, Liu S, Czarnewski P, et al. Monocytes in sarcoidosis are potent tumour necrosis factor producers and predict disease outcome. Eur Respir J. 2021;58(1):03468-2020. doi:10.1183/13993003.03468-2020