BATF2 and PDK4 As Diagnostic Molecular Markers of Sarcoidosis and Their Relationship With Immune Infiltration

Xiaoyan Li
the First affiliated Hospital of Chengdu Medical College
Jing Zhou
the First affiliated Hospital of Chengdu Medical College
Jie He (tt851024238@126.com)
the First affiliated Hospital of Chengdu Medical College

Research Article

Keywords: sarcoidosis, immune cells, diagnostic, CIBERSOT

Posted Date: May 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-557951/v1

Abstract

Background: Sarcoidosis (SA) is an immune disorder disease featured with granulomas formation. The work purposed to uncover potential markers for sarcoidosis (SA) diagnosis and explore how immune cell infiltration contributes to the pathogenesis of SA.

Methods: Sarcoidosis GSE83456 samples and GSE42834 from Gene Expression Omnibus (GEO) were analyzed as the training and external validation sets, respectively. R statistical software was employed to uncover the differentially expressed genes (DEGs) of GSE83456. SVM algorithms and LASSO logistic regression were applied for screening and verification of the diagnostic markers for key module genes. The infiltration of immune cells in sarcoidosis patients' blood samples was assessed by CIBERSORT. The expression of serum BATF2 and PDK4 was detected by RT-qPCR method, and the value of BATF2 and PDK4 mRNA expression in the diagnosis of pulmonary sarcoidosis was analyzed.

Results: In total, 580 DEGs were identified from the key module. PDK4 (AUC=0.942) and BATF4 (AUC=0.980) were revealed as diagnostic markers of sarcoidosis. We found that monocytes, T cells regulatory (Tregs), mast cells, macrophages NK cells, and dendritic cells may contribute to sarcoidosis development. In addition, PDK4 and BATF4 were closely associated with these immune cells. BATF2 and PDK4 were highly expressed in pulmonary sarcoidosis. BATF2 and PDK4 combined to predict the area under the ROC curve of pulmonary sarcoidosis was 0.922.

Conclusions: PDK4 and BATF4 could be used as diagnostic markers of sarcoidosis, and immune cell infiltration severs an important role in sarcoidosis.

1. Introduction

Sarcoidosis, multisystem granulomatous disease with elusive etiology, is characterized histologically by non-caseating granulomas [1]. Sarcoidosis is, in most cases, affects the lungs, but all organs can be involved [2]. In the past 30 years, sarcoidosis-related mortality has elevated, whereas respiratory failure is highly linked to sarcoidosis-related deaths [3]. The severity range of pulmonary sarcoidosis is from imaging abnormalities accidentally found in patients presenting no symptoms to chronic diseases that are difficult to treat [4]. It is difficult to diagnose because it can mimic many other diseases, including lymphoproliferative diseases and granulomatous infections. There is no special examination to diagnose, depending on the correlation of clinical radiology and histopathological characteristics [5, 6]. For patients who require systemic treatment to manage their condition, clinicians, in most cases, administer corticosteroids as the first-line treatment. Antimetabolites are usually administered as alternative drugs for patients who do not respond to or are intolerant to corticosteroids [7]. In fact, corticosteroid treatment is related to toxic effects, and toxic effects are related to cumulative dose and treatment time [8]. The scarcity of reliable predictors for disease progression and truly effective therapies and individual patients poses great challenges in managing sarcoidosis [9]. Therefore, it is critical to identify early diagnostic biomarkers of Sarcoidosis.

A large number of studies recently indicated that sarcoidosis progression is highly linked to the infiltration of immune cells and inflammation. Studies on cellular players in the sarcoidosis mechanism include both innate and adaptive immune cells [10]. During granulomatous responses, several cytokines plus other mediators are released by T lymphocytes and activated macrophages [11]. Granulomas are known to be the pathological characteristic of sarcoidosis. They are tightly packed cell clusters, forming the primary core of multinucleated giant cells, epithelioid histiocytes, and macrophages enclosed in a lymphocyte collar [12]. Although CD4+ T-cells are majorly localized within the lymphocyte collar, a few B cells, CD8+ T-cells, fibroblasts, and plasma cells are present, too [13]. CD4+ T lymphocytes are crucial in sarcoidosis progression as they recruit leukocytes, eventually generating granulomas, whose interaction with B cells stimulates the production of antibodies [14]. Hence, from an immunological point of view, assessing the degree of immune cell infiltration and revealing how various infiltrating immune cell components are vital in determining the underlying molecular mechanism of sarcoidosis and the development of novel immunotherapeutic targets. Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) is a biological tool that adopts extensive deconvolution of data for the expression of genes and a sophisticated algorithm for quantifying various immune cells in different disease samples and substrates, in silico [15]. Whereas, to date, no previous studies have attempted to explore the infiltration of immune cells in sarcoidosis using CIBERSORT.

Herein, we retrieved the microarray dataset of sarcoidosis from the GEO database, then carried out WGCNA to find the key module. Then, the machine learning approaches were used to filtrate extensively and uncover the diagnostic molecular markers of sarcoidosis. Next, we applied CIBERSORT to evaluate the immune infiltration difference between sarcoidosis patients and normal people's blood samples in 22 immune cell subsets. Consequently, the association of markers with infiltrating immune cells was studied to comprehend the existing immune mechanisms in sarcoidosis

2. Material And Method

2.1 Data

Using the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/), we chose two datasets linked with sarcoidosis for subsequent analyses. Notably, the GSE83456 [16] based on the GPL10558 platform, including 49 sarcoidosis patients and 61 normal individual blood samples, was used to investigate the immunologic mechanism in sarcoidosis. Additionally, the GSE42834 [17] includes 61 sarcoidosis patients and 113 blood samples from normal individuals, which are also based on the GPL10558 platform, were used for the verification test. The arrays function within the limma package was explored for normalizing gene expression profiles in GSE83456 and GSE42834 [18]. Besides, we employed the impute package (http://bioconductor.org/packages/impute/) as supplementary to the absent data.

2.2 Differentially Expressed Genes (DEGs) Screening

To demonstrate the impact of inter-sample correction, we used a PCA cluster plot with 2 dimensions. DEGs between sarcoidosis and control was also revealed via lmFit and eBayes functions using limma package . The heatmap and the volcano map of DEGs were generated via the ggplot2 and pheatmap package was used in illustrating the differential expression of DEGs. DEGs showing P <0.05 upon adjustment using the false discovery rate (FDR), and $|\log_2$fold change $|>0.5$ were regarded as significant.

2.3 weighted gene coexpression network analysis (WGCNA)

WGCNA explores necessary modules and key genes of overlapping genes. It adopts the topological overlapping measurements to reveal the corresponding expression modules and describe the pattern of gene correlation between different samples[19]. The expression profile of DEGs, universally down-regulated or up-regulated in the sarcoidosis and control groups, were extracted to perform WGCNA in GSE83456. Firstly, we used hclust function for hierarchical clustering analysis. Secondly, we utilized the pick soft threshold function to screen for the soft thresholding power value when constructing the module. Using candidate power (1 to 20), we determined the average connectivity degrees of various modules in addition to their independent traits. For any degree of independence above 0.8, a suitable power value was chosen. The WGCNA R package was employed in establishing the co-expression network (modules), with the minimum-sized module set to 30, whereas we issued a unique color label to each module. In WGCNA, we identified gene significance (GS) as the relationship between genes and phenotypes. Then, a module membership (MM): MM (i)=cor(x i, ME) was highlighted to evaluate essential gene functions in the module. Notably, the highest correlation index between Module membership and gene significance was used to screen key module.

2.4 Functional Correlation Analysis of key module

The Database for Annotation, Visualization and Integrated Discovery(DAVID, http://david.abcc.ncifcrf.gov/) which incorporates a comprehensive biological knowledge base and a series of analytic tools employed when extracting biological themes for proteins or genes [20] was used for Gene Ontology(GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses on genes of key module. The Beniamini and Hochberg(BH) method was used for adjusting the p values, and the adjusted p-value <0.05 acted as the threshold for significant results. Visualization of results was compiled using the R ggplot2 package.

2.5 Screening and verifying diagnostic markers

The least absolute shrinkage and selection operator (LASSO) logistics regression and support vector machine recursive feature elimination (SVM-RFE) both were classic methods in machine learning which were executed to conduct feature selection for screening the diagnostic markers for sarcoidosis. Moreover, the GSE42834 was used to verify the diagnostic efficiency of the obtained diagnostic markers. We used the LASSO algorithm via the glmnet package. Additionally, SVM-RFE, which is a machine learning technique that relies on a support vector machine, was employed to reveal the optimal variations by eliminating SVM-generated eigenvectors [15]. SVM module was constructed to determine the

diagnostic value of molecular markers in sarcoidosis using e1071 and kernlab package. Consequently, we choose an interaction of genes derived from SVM-RFE or LASSO algorithm for subsequent analyses. We considered a two-sided p<0.05 to represent statistical significance.

2.6 Evaluating immune cell infiltration

With the uploading of the GSE83456 data to CIBERSORT. Those samples having p>0.05, were eliminated and we got the immune cell infiltration matrix data. Furthermore, ggplot2 package was used for the analysis of PCA clustering about immune cell infiltration quantitative data to generate a PCA clustering map with 2 dimensions. Next, we used corrplot package to create a correlation heatmap in identifying any relationship between 22 infiltrating immune cell types; ggpolt2 package was employed to picture violin plots to observe any difference in the immune cell infiltration.

2.7 relationship analysis of diagnostic molecular markers with infiltrating immune cells

The corrplot package of R software was adopted for Spearman correlation analysis of the infiltrating immune cells as well as the diagnostic markers. The threshold was set as |r | >0.3 and P<0.05. The former result was visualized by the ggplot2 package.

2.8 Clinical specimen verification

2.8.1 Specimen source

The serum samples of 60 patients with pulmonary sarcoidosis were collected from the Department of Respiratory and critical Care Medicine, the first affiliated Hospital of Chengdu Medical College from January 2016 to December 2020. At the same time, the serum samples of 60 healthy subjects were collected as the control group. The peripheral blood samples of all subjects' 10ml were collected by anticoagulant tube. After 30min was placed at 4 ℃, centrifuged for 15 minutes at room temperature at 3000r, the upper serum was retained and stored in the refrigerator at-80 ℃. Admission criteria: (1) Age greater than 18 years old, EBUS-TBNA biopsy, pathological diagnosis of pulmonary sarcoidosis; (2) patients with good compliance can cooperate with the examination. Exclusion criteria: (1) having been given glucocorticoid treatment; (2) complicated with infectious diseases, tumor diseases or immunodeficiency diseases. (3) severe cardiopulmonary insufficiency or blood coagulation disturbance. This study was approved by the Ethics Committee of the first affiliated Hospital of Chengdu Medical College, and the approval number is 2020CYFYIRB-BA-101. All the controls signed the relevant informed consent.2.8.1 Detection of mRNA transcription level in clinical tissue samples of BATF2 and PDK4 by RT-qPCR.

Total RNA was isolated using Trizol (Invitrogen, Grand Island, NY). According to the instructions of cDNA synthesis kit (TaKaRa Company, Japan), the reverse transcription conditions of cDNA, were as follows: 37 ℃, 30 min, 95 ℃, 5 min, -20 ℃. RT-qPCR preparation system: 10 μ L SYBR Green (Japan TaKaRa Co., Ltd.), upstream and downstream primers 0.8 μ L, 2 μ L cDNA, 6.4 μ L without enzyme water. The RT-qPCR reaction was carried out in ABI 8000 real-time quantitative PCR, with GAPDH as the internal reference, and

the reaction conditions were set as follows: pre-denaturation at 95 ℃, 30s at 95 ℃, 5s at 60 ℃, 34s at 60 ℃, 40 cycles and annealing at 60 ℃ for 30s. BATF2 upstream primer is 5’-CCTCTCCCGACAACCCTTC, downstream primer is GTGGACTTGAGCAGAGGAGA-3’; PDK4 upstream primer is 5’-TTGGCTGGTTTTGGTTACGG, downstream primer is CACCAGTCAGCCTCAGA-3’; GAPDH upstream primer is 5’-CCATCTTCCAGGAGCGAGAT, downstream primer is TGCTGATGATCTTGAGGCTG-3’. Gene expression levels were calculated relative to the housekeeping gene GAPDH. The ROC curve was drawn by ROCR package in R software, and the area (AUC) under ROC curve was used to evaluate the diagnostic value of serum BATF2 and PDK4 and their combination in pulmonary sarcoidosis.

3. Results

3.1 Differential expression analysis. After the 47230 probes in GSE83456 were preprocessed, 20936 genes were obtained. The result of PCA demonstrated that the clustering of the two sample groups was remarkable, suggesting the reliability of the sample source (Figure 1). After preprocessing the data, the R software was used to retrieve 764 differential genes from the GSE83456, as depicted in the volcano map and heatmap (Figure 2A, B).

3.2 Sarcoidosis-associated module

To reveal the key module most associated with sarcoidosis, WGCNA was conducted utilizing the expression profiles for the 764 DEGs. After merging the similar modules, we could identify a total of six modules and each module was displayed by distinct colors to distinguish different modules (Figure 3A).The blue module is most positively associated with sarcoidosis(correlation coefficient =0.92，P=3E-46;Figure 3B). Based on the absolute value of the GS in each module, the key module in sarcoidosis was selected (blue module) (Figure 3C), and the module membership in the blue module is positively correlated with GS for sarcoidosis(correlation coefficient =0.94，P<1E-200; Figure 3B). Thus, the blue module is a key module that contains 580 genes.

3.3 Functional correlation analysis of genes in the key module

The 580 genes of key module were grouped into BP (biological process: 8 BP terms were significantly enrichment)，MF (molecular function:9 MF terms were significantly enrichment) and CC (cellular component:12 CC terms were significantly enrichment) categories (Figure 4). The GO analysis result indicated that the genes were primarily associated with “immune response”, “type I interferon signaling pathway”, “immunological synapse”, “innate immune response”, and so on. KEGG pathway enrichment analysis (8 KEGG terms were statistically significant enrichment, Figure 5) showed that a major proportion of the genes were enriched in pathways, for example, “primary immunodeficiency”, “Cytokine-cytokine receptor interaction.” These findings suggest how immune response may play a significant role in the progression of sarcoidosis.

3.4 screening and validating the diagnostic markers

We identified 22 genes from the key module as diagnostic markers for sarcoidosis using LASSO logistic regression algorithm method (Figure 6A), and lambda value. min =0.00436. Twenty-five genes were screened from the key module employing the SVM-RFE algorithm as diagnostic markers (Figure 6B). Of note, we overlapped gene markers found by two algorithms and eventually confirmed two diagnostic related genes (PDK4 and BATF2) (Figure 6C). In GSE83456, the diagnostic power of BATF2 was 0.980(area under curve; AUC=0.980，95%CI=0.933-0.997), and the diagnostic power of PDK4 was 0.942(area under curve; AUC=0.942，95%CI=0.880-0.977)(Figure 7A) . The accuracy and efficiency of the two diagnostic markers were validated by GSE42834 as an external validation test. The diagnostic power of BATF2 was 0.896(area under the curve; AUC=0.896，95%CI=0.841-0.937), and the diagnostic power of PDK4 was 0.772(area under the curve; AUC=0.942，95%CI=0.703-0.832) (Figure 7B). The summary of the two diagnostic markers is shown in Table 1.

3.5 Immune cell infiltration

The PCA cluster analysis could be utilized to examine whether the biological repetition and the diversity of various groups concur. Herein, the PCA cluster assessment revealed a significant immune cell infiltration difference between the sarcoidosis samples and the control samples (Figure 8A). The heatmap of the 22 types of immune cells suggested that activated NK cells, eosinophils, and mast cells resting had a positive correlation (Figure 8B). In addition, the violin plot of the immune cell infiltration difference indicated significant differences of immune infiltration between sarcoidosis patients and the normal people' blood samples. Quite notably, monocytes, NK cells activated, macrophages M0, macrophages M1，dendritic cells activated, mast cells resting, T cells regulatory (Tregs), and mast cells activated infiltrated more in sarcoidosis patient than that in normal people. Conversely, T cell CD8，T cells CD4 naïve, T cells follicular helper and neutrophils infiltrated less in sarcoidosis patients than that in normal people (Figure 8C).

3.6 correlation analysis between PDK4, BATF2 and infiltrating immune cells.

We found that the BATF2 had a positive correlation with dendritic cells activated (r=0.378,P=0.007) , Macrophages M1 (r=0.376,P=0.008)，Monocytes(r=0.302,P=0.035) (Figure 9 A,B,C) and negatively correlated with Macrophages M0 (r=0.508,P=1.975E-04) (Figure 9D) ; PDK4 was negatively correlated with T cells CD4 naive (r=-0.304,P=0.034)，and positively correlated with dendritic cells resting(r=0.328,P=0.021)，Mast cells resting(r=0.307,P=0.032) (Figure 10 A, B, C)，NK cells activated(r=0.306,P=0.033) (Figure 10D).

3.7 Expression of PDK4 and BATF2 in in clinical serum specimens

Finally, there were 60 patients with pulmonary sarcoidosis (30 males and 30 females) with age of 46.2 ±10.4 years. A total of 60 patients (30 males and 30 females) were included in the control group, with age of 52.9 ±9.8 years. There was no significant difference in age and sex between the two groups. The serum levels of BATF2 and PDK4 in the two groups are shown in Figure 11. The levels of serum BATF2 and PDK4 in patients with pulmonary sarcoidosis were significantly higher than those in healthy controls

(BATF2: 6.37 ±5.44; PDK4:6.09 ±5.45). The area under ROC curve predicted by serum BATF2 was 0.909, and the best cut-off value was 11.412. The area under ROC curve predicted by PDK4 was 0.721, and the best cut-off value was 5.271. BATF2 and PDK4 combined to predict the area under the ROC curve of pulmonary sarcoidosis was 0.922. (Figure 12).

4. Discussion

Sarcoidosis is a granulomatous disease with multi-system and multi-organ involvement, which often injuries the chest, heart, liver, kidney, and central nervous system. The incidence of sarcoidosis in women is higher compared to that in men [21]. In recent years, due to the continuous deepening of clinicians' understanding of sarcoidosis, and the gradual improvement of examination techniques, the diagnosis rate of sarcoidosis is also significantly higher than before. Sarcoidosis diagnosis often is dependents on radiological and clinical imaging, which are related to the histology of epithelioid granulomas; nevertheless, granulomas are not unique pathognomonic for sarcoidosis [22]. In particular, the imaging features of pulmonary sarcoidosis are similar to those of pulmonary tuberculosis and mediastinal lymph node tuberculosis [23]. Therefore, pulmonary sarcoidosis might be easily missed in diagnosis or misdiagnosed. Remarkably, there have recently been studies suggesting that immune cell infiltration serves a vital role in sarcoidosis development [24, 25]. Thus, identifying special molecular biomarkers and exploring the immune cell infiltration pattern in sarcoidosis has become urgent, which may be valuable to improve sarcoidosis prognosis. With the rapid development of genome-sequencing technology today, bioinformatics also provides strong support for the screening of molecular biomarkers, and CIBERSORT tools also provide favorable conditions for exploring immune cell infiltration patterns for various diseases. Herein, we attempted to reveal blood biomarkers capable of diagnosing sarcoidosis. Meanwhile, the role of immune cell infiltration to sarcoidosis will be further explored.

In this paper, we obtained the microarray gene expression data from GEO database and identified 764 DEGs between sarcoidosis patients and 61 normal people' blood samples. Besides, Using WGCNA, we found the key module, the blue module, to be significantly correlated with sarcoidosis. Based on the blue module, 580 genes of the module were screened. The GO analysis result indicated that these 580 genes were mainly associated with “type I interferon signaling pathway”, “innate immune response”, “immune response”, “immunological synapse” and so on. KEGG pathway enrichment analysis demonstrated that the majority of these genes were enriched in pathways such as “Cytokine-cytokine receptor interaction” “primary immunodeficiency”. Taken together, these results strongly implicate that the immune response is essential for Sarcoidosis. Greaves SA et al. demonstrated that one of the defining features of pulmonary sarcoidosis is infiltration of activated CD4+ T cells in the lung tissue [24]. Previously, Grunewald J et al. reported that sarcoidosis is a systemic inflammatory disorder characterized by tissue infiltration of mononuclear phagocytes and lymphocytes which are associated with the formation of non-caseating granuloma [26]. The results of our analysis data were in line with the above previous studies, confirming the robustness and reliability of the present findings.

SVM-RFE (Support Vector Machine Recursive Feature Elimination), a wrapper technique employed in big data mining, utilizes a backward feature, which recursively eliminates insignificant traits from a larger subset [27]. It has been shown to be a powerful tool in the identification of potential alterations and thus, the classification of healthy and Sarcoidosis groups. LASSO logistic regression is a widely applied method for the regression of high-dimensional data. It imposes a constraint, λ, on the size of the regression coefficients β in ordinary least squares (ordinary least squares regression). Based on WGCNA analysis, we used two algorithms to screen feature variables and establish a reliable classification model. Then, PDK4 and BATF2 were identified as diagnostic markers of Sarcoidosis. Although PDK4 and BATF2 were just selected by combining SVM-RFE and LASSO, the diagnostic power of PDK4 and BATF2 were reliable in GSE42834 validation. These findings were also validated in clinical samples, suggesting that the PDK4 and BATF2 in peripheral blood has significant value in the diagnosis of SA.

Pyruvate dehydrogenase kinase 4 (PDK4), an important mitochondrial enzyme, impedes the acetyl-CoA production via selective inhibition of pyruvate dehydrogenase activity via phosphorylation. It can regulate the glycolysis pathway and affects cell metabolism, proliferation, apoptosis [28]. Some studies reported that inactivation of the pyruvate dehydrogenase complex by overexpression of PDK4 contributes to hyperglycemia. Therefore, the serious health problems associated with diabetes and PDK may play an essential role in hyper catecholamine-induced insulin resistance in the periadrenal adipose tissues of pheochromocytoma patients [29, 30]. Another study revealed that PDK4 is indispensable in dictating the fate of TNF/NF-κB-mediated hepatocyte apoptosis. Meanwhile, this pro-survival pathway switches to pro-apoptosis mediated by pyruvate dehydrogenase kinase 4 (PDK4)-deficiency [31]. TNF could exert significant effects in sarcoidosis development through the maintenance of chronic pro-inflammatory status in macrophages. Activated macrophages can be transformed into epithelioid cells and multinucleated macrophages, resulting in granuloma formation [32]. Given that PDK4 interacts with inflammatory factors, we propose the following hypothesis that PDK4 potentially contribute to the regulation of the pathological process of sarcoidosis. BATF2, a transcription factor belonging to the BATF family, has previously been characterized and reported to inhibit tumor growth by suppressing AP-1 activity [33]. Recent reports have highlighted BATF2 as a vital agent in innate immune responses and are essential as a transcription factor during gene regulation. It also exhibits effector functions in classical activation of macrophages [34]. Guler R et al. functional studies showed a predominant role of BATF2 in regulating Th2 cell functions and lineage development of T lymphocytes [35]. Kayama H et al. also reported that BATF2 in innate myeloid cells is a crucial molecule that suppresses IL-23/IL-17 pathway-mediated adaptive intestinal pathology. Studies have revealed several T-cell-associated cytokines in sarcoidosis immunopathogenesis; however, increasing reports show that IL-12 cytokine family members such as IL-12, IL-23, IL-27, and IL-35 are closely associated with sarcoidosis [36]. Findings from earlier studies show that PDK4 and BATF2 may be related to the development of Sarcoidosis, thus can be utilized as potential diagnostic markers for sarcoidosis. However, additional clinical reports are warranted to validate the diagnostic capability of PDK4 and BATF2.

Sarcoidosis is a chronic granulomatous disease with an aberrant immune response to undefined environmental or infectious triggers [7]. To further assess how immune cell infiltration is linked to

sarcoidosis, we employed CIBERSORT for in-depth exploration of Sarcoidosis immune infiltration. It was demonstrated that infiltration of T cells regulatory(Tregs), NK cells activated, monocytes, macrophages M0, macrophages M1，dendritic cells activated, mast cells resting, mast cells activated were elevated, and a decreased infiltration of T cell CD8, T cells CD4 naïve, T cells follicular helper and neutrophils may lead to sarcoidosis. Interleukin 33 ameliorates the regulatory effects of T cells in the occurrence of pulmonary sarcoidosis，Previous studies have shown that the pulmonary sarcoidosis exerts a Th1/Th17/regulatory T cells (Tregs) -driven inflammatory process in the lung, inducing noncaseating granulomas that comprise CD4+ T cells [14]. Tøndell A et al. found that healthy control subjects' natural killer T cell fractions of leucocyte were lower than that in sarcoidosis patients [37]. These conclusions are consistent with those of the present study. Additionally, Lepzien R et al. suggested that mast cells, monocytes, and dendritic cells (DCs)-are likely critical in sarcoidosis through the initiation and maintenance of T cell activation, thereby participates in granuloma formation driven via cytokine production [38]. By combining results obtained in this study, we propose that sarcoidosis was closely related to inflammation and immune dysregulation. Moreover, the present findings uncovered a detailed mechanism of 22 immune cell types in sarcoidosis. Activated NK cells and eosinophils infiltration are tightly linked to mast cells resting infiltration. However, the actual mechanisms underlying the highlighted correlations should be confirmed via experimental assessments. When we analyzed the correlation between PDK4, BATF2, and immune cells, notably, BATF2 was remarkably had a positive correlation with dendritic cells, macrophages M1, monocytes, whereas it had a negative correlation with macrophages M0. PDK4 showed a significantly negative correlation with T cells CD4 naive and significantly positively correlated with NK cells activated, dendritic cells resting, and mast cells resting. Researchers have found that dendritic cells, macrophages, monocytes, NK cells, mast cells are critically important in Sarcoidosis. Studies have shown that there was a significantly higher proportion of M1 in SA when compared with other ILD [39]. Roy S et al. illustrated that BATF2 (an activation marker gene for M1) mediates the regulation of genes in IFN-γ-activated classical macrophages and LPS/HKTB-induced macrophage modulation [40]. M1 macrophages polarization in the lung potentially aggravated the granuloma formation. As a result, we thought that BATF2 could raise dendritic cells and monocytes or reduces M0 macrophages cells, and PDK4 can raise dendritic cells, NK cells, mast cells, or reduces T cells naive to participate in sarcoidosis progression. However, these hypotheses need more researches to elucidate the unknown reciprocal relationship between immune cells, BATF2, and PDK4.

5. Conclusion

In conclusion, we found that PDK4 and BATF4 are diagnostic markers of Sarcoidosis. We also found that T cells regulatory (Tregs), NK cells activated, monocytes, macrophages, dendritic cells activated, mast cells resting and mast cells activated may contribute to the pathogenesis of sarcoidosis. Besides, PDK4 and BATF4 are closely related to the immune cells, which may have an important role in Sarcoidosis. Further exploration of the immune cells may determine new targets of Sarcoidosis immunotherapy and enhance the efficacy of immunomodulatory therapies for Sarcoidosis patients.

Declarations

Data availability statement

The datasets GSE83456 and GE42834 used and analyzed in the current study are available from the public database(http://www.ncbi.nlm.nih.gov/geo/).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationship that can influence the work reported in this study.

Acknowledgements

None

Funding

None.

References

1. M. A. Judson. "Environmental Risk Factors for Sarcoidosis," Front Immunol, vol.11, pp.1340, 2020.
2. N. Soto-Gomez, J. I. Peters, and A. M. Nambiar. "Diagnosis and Management of Sarcoidosis," Am Fam Physician, vol.93, no.10, pp.840–848, 2016.
3. E. Criado, M. Sánchez, J. Ramírez, et al., "Pulmonary sarcoidosis: typical and atypical manifestations at high-resolution CT with pathologic correlation," Radiographics, vol.30, no.6, pp.1567–1586, 2010.
4. P. Spagnolo, G. Rossi, R. Trisolini, et al., "Pulmonary sarcoidosis," Lancet Respir Med, vol.6, no.5, pp.389–402, 2018.
5. E. M. Carmona, S. Kalra, and J. H. Ryu. "Pulmonary Sarcoidosis: Diagnosis and Treatment," Mayo Clin Proc, vol.91, no.7, pp.946–954, 2016.
6. D. A. Culver. "Sarcoidosis," Immunol Allergy Clin North Am, vol.32, no.4, pp.487–511, 2012.
7. S. G. West. "Current management of sarcoidosis I: pulmonary, cardiac, and neurologic manifestations," Curr Opin Rheumatol, vol.30, no.3, pp.243–248, 2018.
8. R. P. Baughman and J. C. Grutters. "New treatment strategies for pulmonary sarcoidosis: antimetabolites, biological drugs, and other treatment approaches," Lancet Respir Med, vol.3, no.10, pp.813–822, 2015.
9. P. Korsten, K. Strohmayer, R. P. Baughman, et al., "Refractory pulmonary sarcoidosis - proposal of a definition and recommendations for the diagnostic and therapeutic approach," Clin Pulm Med, vol.23, no.2, pp.67–75, 2016.

10. S. Lee, D. Birnie, and G. Dwivedi. "Current perspectives on the immunopathogenesis of sarcoidosis," Respir Med, vol.173, pp.106161, 2020.
11. J. Talreja, C. Bauerfeld, E. Sendler, et al., "Derangement of Metabolic and Lysosomal Gene Profiles in Response to Dexamethasone Treatment in Sarcoidosis," Front Immunol, vol.11, pp.779, 2020.
12. J. Larsson, P. Graff, I. L. Bryngelsson, et al., "Sarcoidosis and increased risk of comorbidities and mortality in sweden," Sarcoidosis Vasc Diffuse Lung Dis, vol.37, no.2, pp.104–135, 2020.
13. E. R. Zhou and S. Arce. "Key Players and Biomarkers of the Adaptive Immune System in the Pathogenesis of Sarcoidosis," Int J Mol Sci, vol.21, no.19, 2020.
14. B. Zhang, F. Zhao, H. Mao, et al., "Interleukin 33 ameliorates disturbance of regulatory T cells in pulmonary sarcoidosis," Int Immunopharmacol, vol.64, pp.208–216, 2018.
15. Z. Liu, M. Mi, X. Li, et al., "A lncRNA prognostic signature associated with immune infiltration and tumour mutation burden in breast cancer," J Cell Mol Med, 2020.
16. S. Blankley, C. M. Graham, J. Turner, et al., "The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis," PLoS One, vol.11, no.10, pp. e0162220, 2016.
17. C. I. Bloom, C. M. Graham, M. P. Berry, et al., "Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers," PLoS One, vol.8, no.8, pp. e70630, 2013.
18. Y. J. Deng, E. H. Ren, W. H. Yuan, et al., "GRB10 and E2F3 as Diagnostic Markers of Osteoarthritis and Their Correlation with Immune Infiltration," Diagnostics (Basel), vol.10, no.3, 2020.
19. D. Pascut, M. Y. Pratama, F. Gilardi, et al., "Weighted miRNA co-expression networks analysis identifies circulating miRNA predicting overall survival in hepatocellular carcinoma patients," Sci Rep, vol.10, no.1, pp.18967, 2020.
20. T. Lin, Y. Qiu, W. Peng, et al., "Heat Shock Protein 90 Family Isoforms as Prognostic Biomarkers and Their Correlations with Immune Infiltration in Breast Cancer," Biomed Res Int, vol.2020, pp.2148253, 2020.
21. E. Niewiadomska, M. Kowalska, M. Skrzypek, et al., "Incidence and economic burden of sarcoidosis in years 2011–2015 in Silesian voivodeship, Poland," Sarcoidosis Vasc Diffuse Lung Dis, vol.37, no.1, pp.43–52, 2020.
22. M. Vijayaraj, P. A. Abhinand, P. Venkatesan, et al., "An ANN model for the differential diagnosis of tuberculosis and sarcoidosis," Bioinformation, vol.16, no.7, pp.539–546, 2020.
23. M. Węcławek, D. Ziora, and D. Jastrzębski. "Imaging methods for pulmonary sarcoidosis," Adv Respir Med, vol.88, no.1, pp.18–26, 2020.
24. S. A. Greaves, S. M. Atif, and A. P. Fontenot. "Adaptive Immunity in Pulmonary Sarcoidosis and Chronic Beryllium Disease," Front Immunol, vol.11, pp.474, 2020.
25. H. Matsuyama, T. Isshiki, A. Chiba, et al., "Activation of mucosal-associated invariant T cells in the lungs of sarcoidosis patients," Sci Rep, vol.9, no.1, pp.13181, 2019.

26. J. Grunewald, P. Spagnolo, J. Wahlström, et al., "Immunogenetics of Disease-Causing Inflammation in Sarcoidosis," Clin Rev Allergy Immunol, vol.49, no.1, pp.19–35, 2015.
27. H. Chang, X. Yang, K. You, et al., "Integrating multiple microarray dataset analysis and machine learning methods to reveal the key genes and regulatory mechanisms underlying human intervertebral disc degeneration," PeerJ, vol.8, pp. e10120, 2020.
28. J. Leem and I. K. Lee. "Mechanisms of Vascular Calcification: The Pivotal Role of Pyruvate Dehydrogenase Kinase 4," Endocrinol Metab (Seoul), vol.31, no.1, pp.52–61, 2016.
29. J. H. Jeon, T. Thoudam, E. J. Choi, et al., "Loss of metabolic flexibility as a result of overexpression of pyruvate dehydrogenase kinases in muscle, liver and the immune system: Therapeutic targets in metabolic diseases," J Diabetes Investig, 2020.
30. C. Wu, H. Zhang, X. Lin, et al., "Role of PDK4 in insulin signaling pathway in periadrenal adipose tissue of pheochromocytoma patients," Endocr Relat Cancer, vol.27, no.10, pp.583–589, 2020.
31. J. Wu, Y. Zhao, Y. K. Park, et al., "Loss of PDK4 switches the hepatic NF-κB/TNF pathway from pro-survival to pro-apoptosis," Hepatology, vol.68, no.3, pp.1111–1124, 2018.
32. V. Besnard, A. Calender, D. Bouvry, et al., "G908R NOD2 variant in a family with sarcoidosis," Respir Res, vol.19, no.1, pp.44, 2018.
33. H. Kayama, H. Tani, S. Kitada, et al., "BATF2 prevents T-cell-mediated intestinal inflammation through regulation of the IL-23/IL-17 pathway," Int Immunol, vol.31, no.6, pp.371–383, 2019.
34. G. Bondar, R. Togashi, M. Cadeiras, et al., "Association between preoperative peripheral blood mononuclear cell gene expression profiles, early postoperative organ function recovery potential and long-term survival in advanced heart failure patients undergoing mechanical circulatory support," PLoS One, vol.12, no.12, pp. e0189420, 2017.
35. R. Guler, S. Roy, H. Suzuki, et al., "Targeting Batf2 for infectious diseases and cancer," Oncotarget, vol.6, no.29, pp.26575–26582, 2015.
36. S. Ringkowski, P. S. Thomas, and C. Herbert. "Interleukin-12 family cytokines and sarcoidosis," Front Pharmacol, vol.5, pp.233, 2014.
37. A. Tøndell, A. D. Rø, M. Børset, et al., "Activated CD8 + T cells and natural killer T cells in bronchoalveolar lavage fluid in hypersensitivity pneumonitis and sarcoidosis," Sarcoidosis Vasc Diffuse Lung Dis, vol.31, no.4, pp.316–324, 2015.
38. R. Lepzien, G. Rankin, J. Pourazar, et al., "Mapping mononuclear phagocytes in blood, lungs, and lymph nodes of sarcoidosis patients," J Leukoc Biol, vol.105, no.4, pp.797–807, 2019.
39. P. Wojtan, M. Mierzejewski, I. Osińska, et al., "Macrophage polarization in interstitial lung diseases," Cent Eur J Immunol, vol.41, no.2, pp.159–164, 2016.
40. S. Roy, R. Guler, S. P. Parihar, et al., "Batf2/Irf1 induces inflammatory responses in classically activated macrophages, lipopolysaccharides, and mycobacterial infection," J Immunol, vol.194, no.12, pp.6035–6044, 2015.

Tables

Table 1. Diagnostic efficacy of the diagnostic markers

					95%CI	
Diagnostic markers	AUC	Youden index	Sensitivity	Specificity	Lower limit	Upper limit
PDK4(GSE83456)	0.980	0.824	88.52	93.88	0.880	0.977
BATF2(GSE83456)	0.942	0.906	96.72	93.88	0.933	0.997
PDK4(GSE42834)	0.772	0.447	84.07	60.66	0.703	0.832
BATF2(GSE42834)	0.896	0.702	96.46	73.77	0.841	0.937

Figures

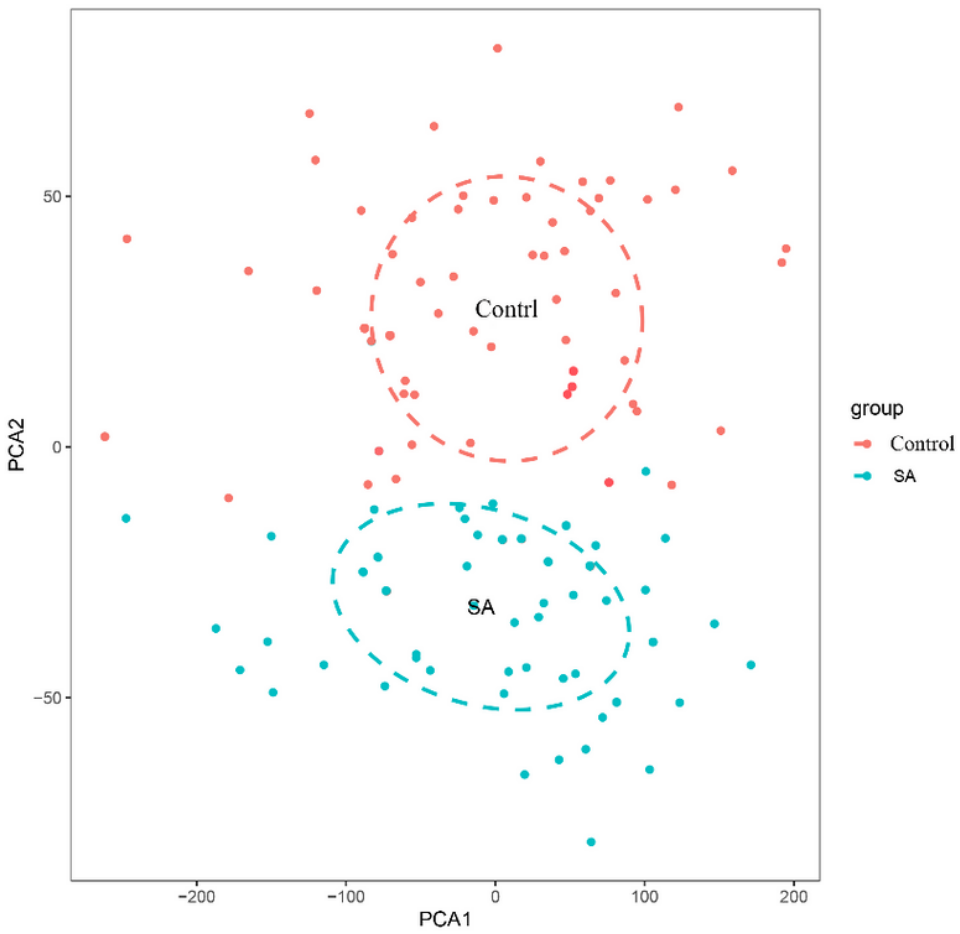

Figure 1

Two-dimensional PCA cluster plot of the GSE18781; blue denotes the Sarcoidosis group, and red denotes the normal group (control).

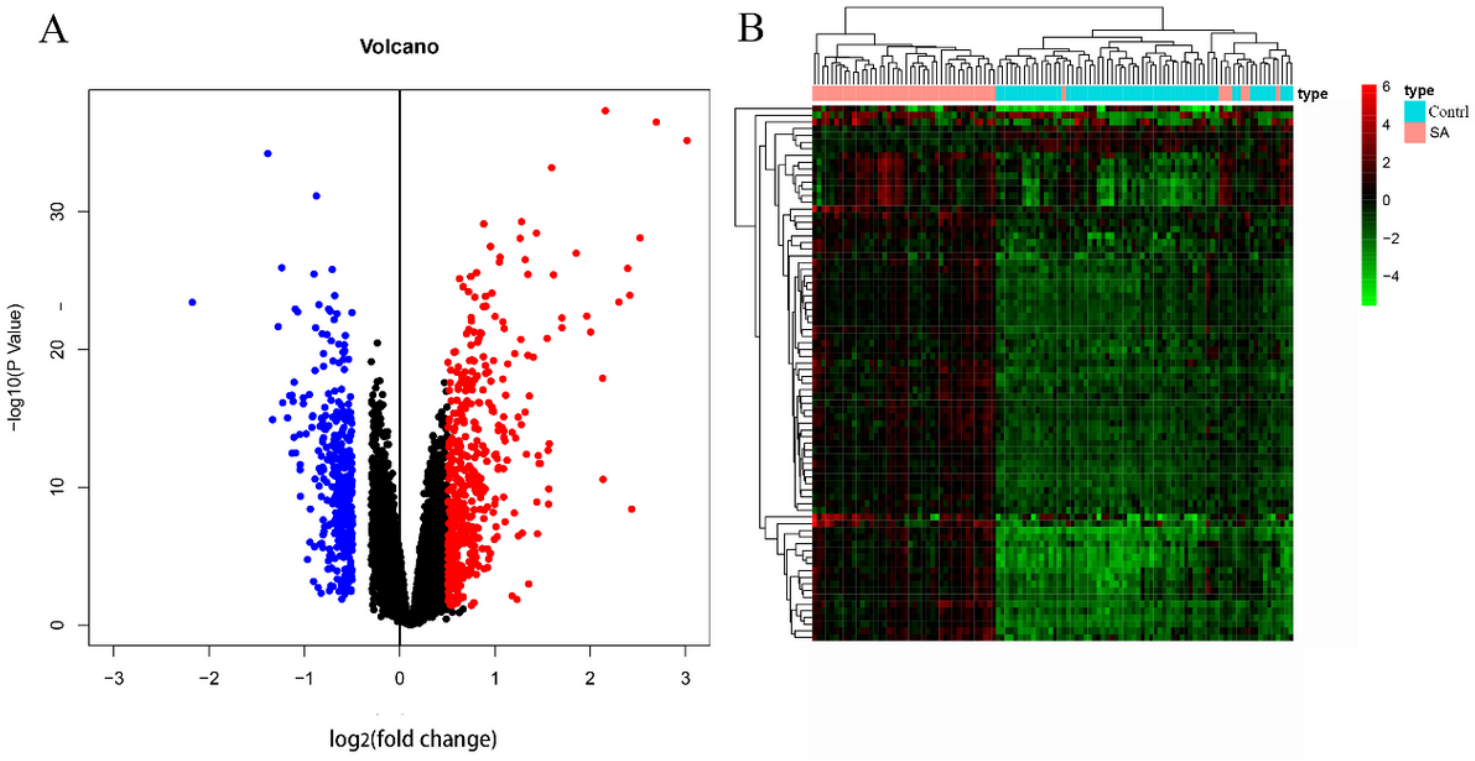

Figure 2

volcano map and heatmap of differential expressed genes. (A) Volcano map showing the DEGs; red denotes the up-regulation of differential genes, black represents non-significant differential genes, and blue denotes down-regulation of differential genes. (B) Heatmap map of DEGs; red represents the Sarcoidosis group, and blue represents the normal control group

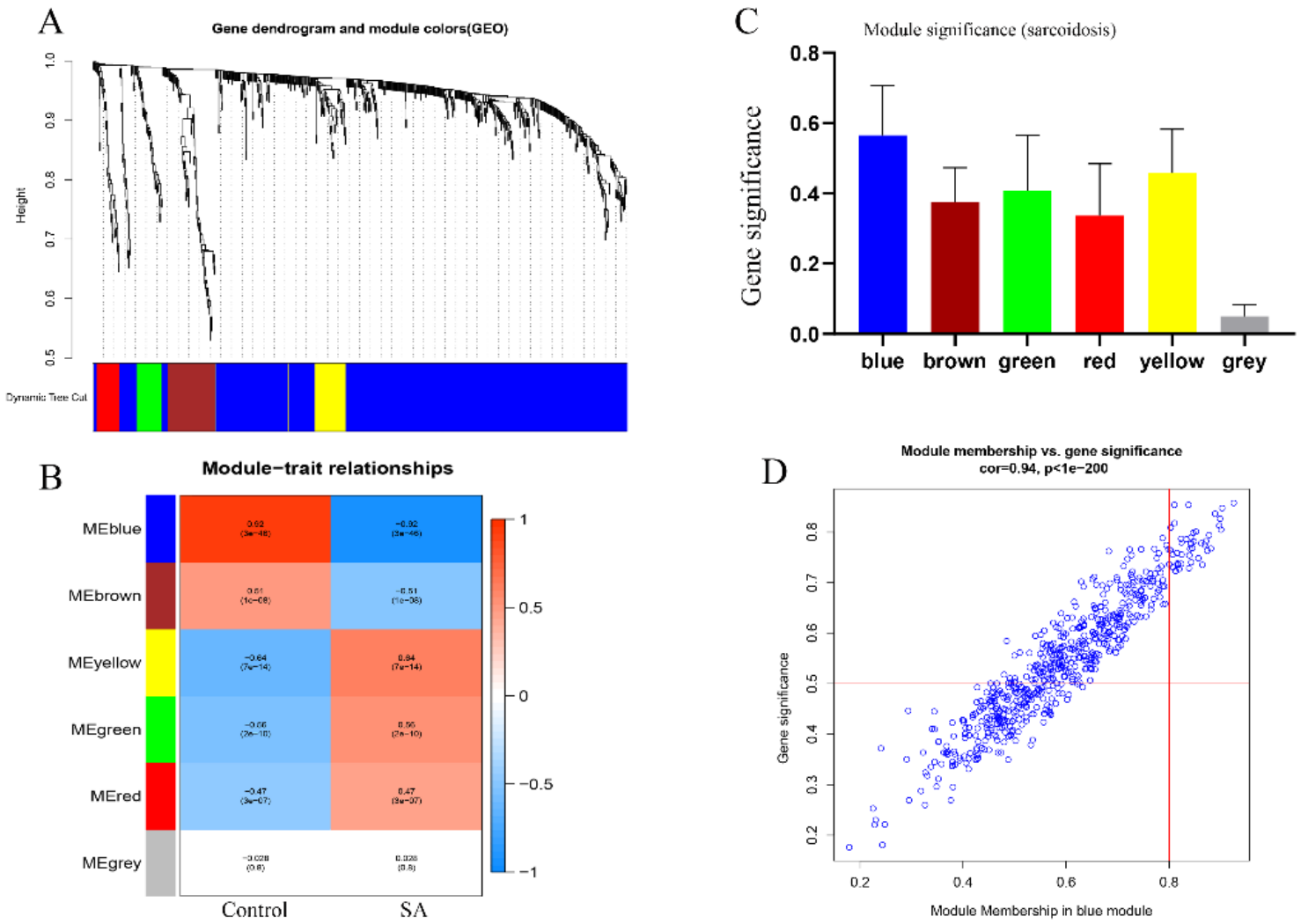

Figure 3

weighted correlation network analysis. (A) A recognition module, we gave every module a color to serve as identifiers for 6 distinct modules. (B) A correlation heatmap of gene modules and phenotypes, the red color shows a positive correlation with the phenotype; the color shows a negative correlation with the phenotype. (C) the gene significance for the modules in the sarcoidosis group. (D) the correlation between gene significance and module membership in the key modules.

Figure 4

biological process of the key module genes,

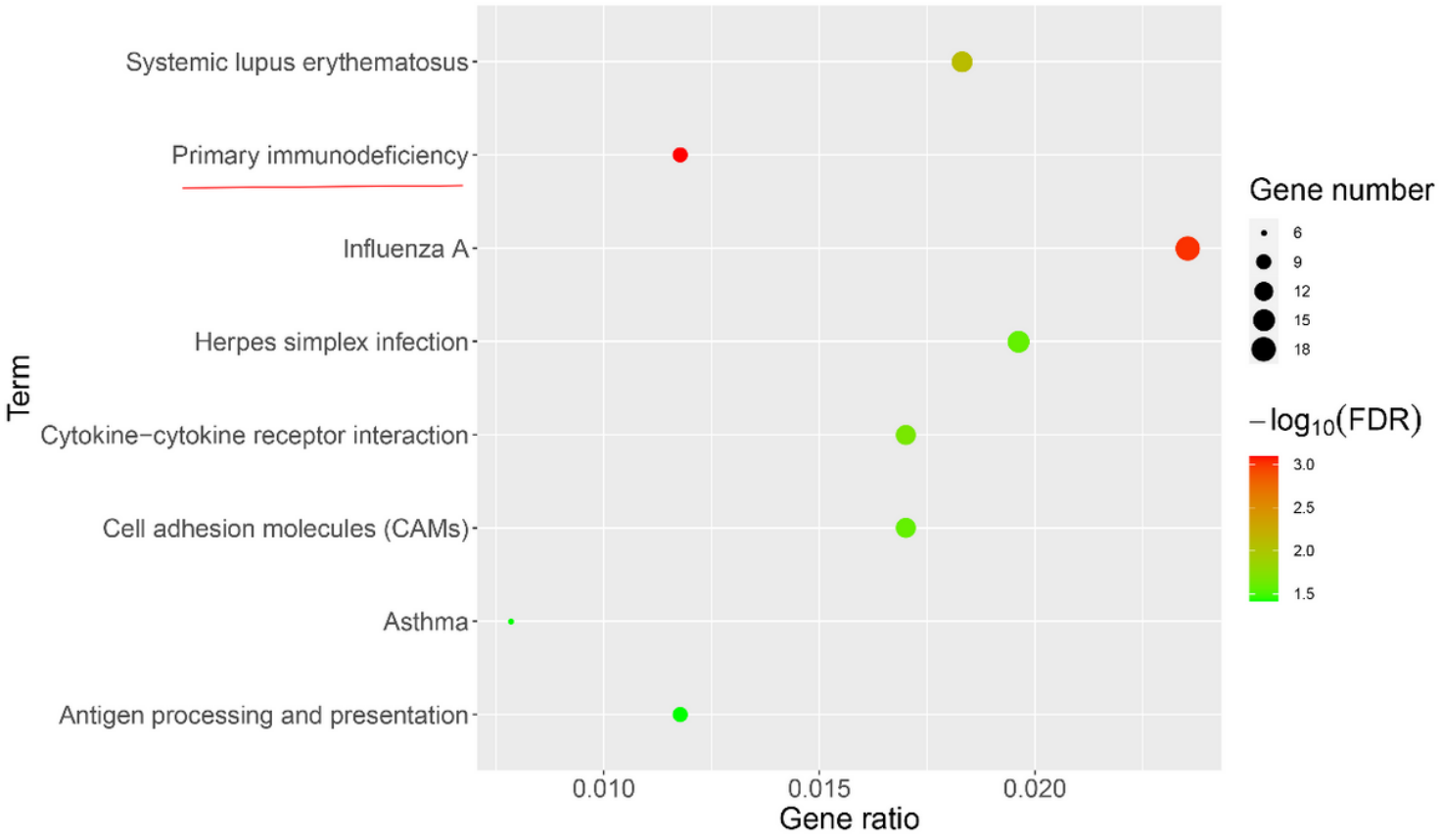

Figure 5

KEGG pathways analysis of the key module genes

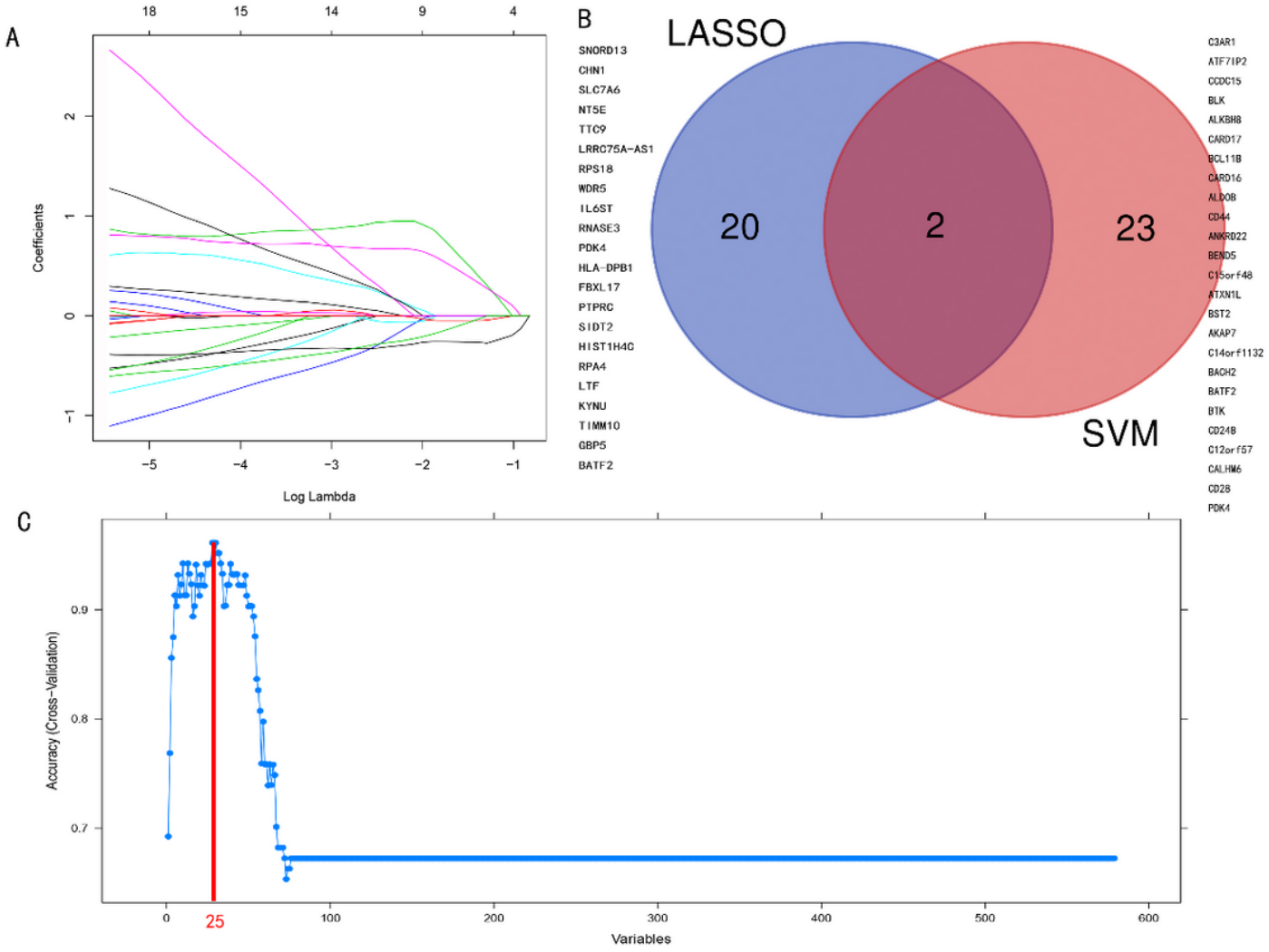

Figure 6

screening and validating the diagnostic markers. (A) least absolute shrinkage and selection operator (LASSO) logistic regression algorithm for screening the diagnostic markers. (B) SVM-RFE algorithm screening of the diagnostic markers. (C) Venn diagram displaying the intersection between diagnostic markers identified using the two algorithms

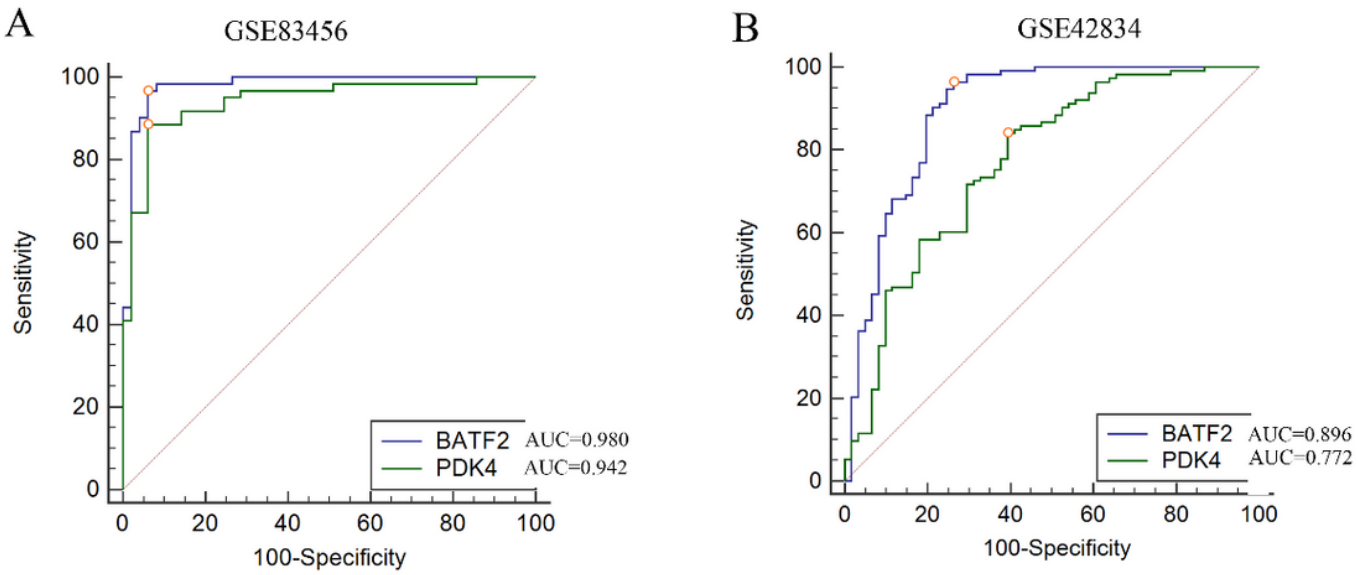

Figure 7

(A) A ROC curve demonstrating the diagnostic efficacy in GSE83456. (B) The ROC curve of the diagnostic efficacy in GSE42834

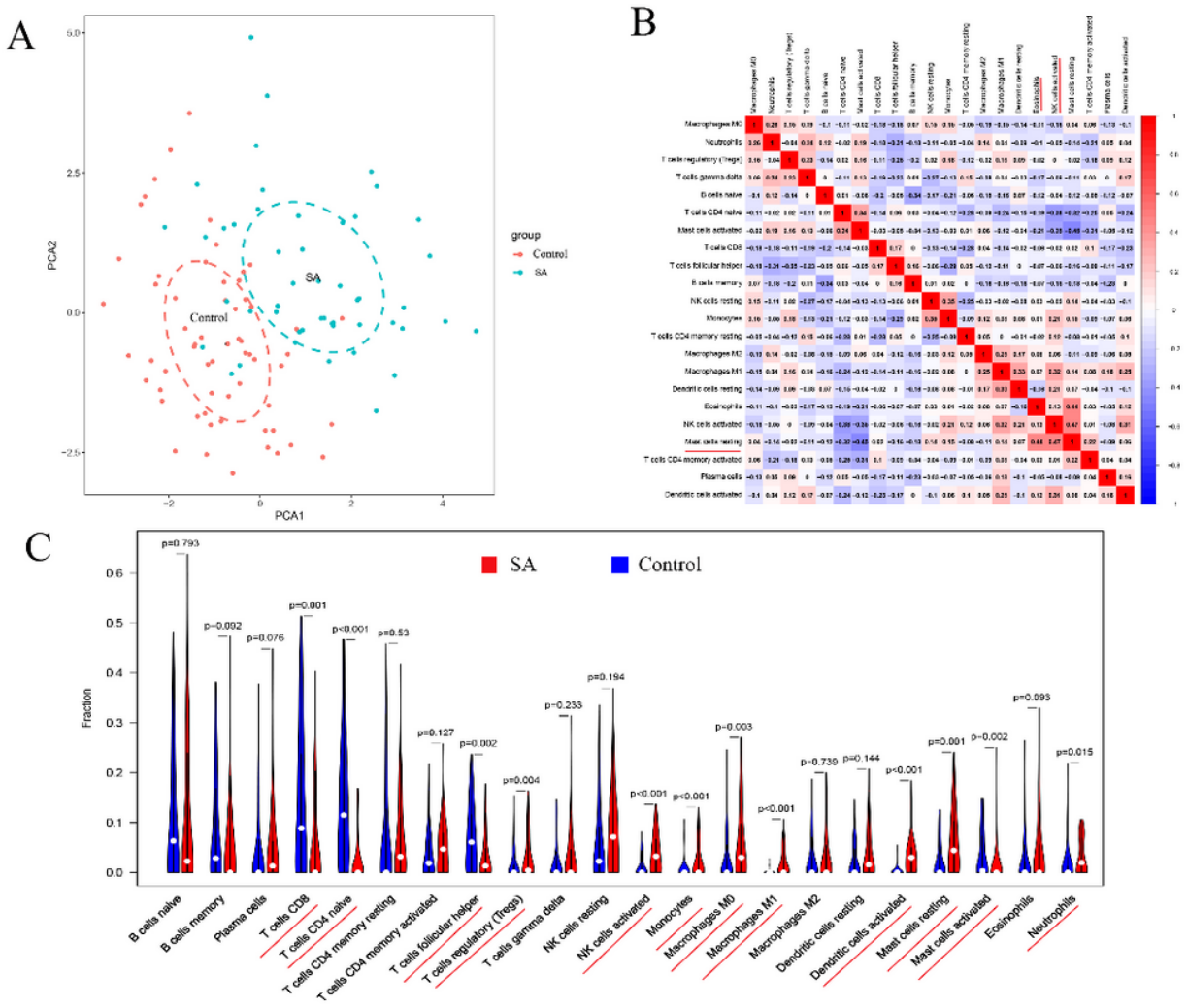

Figure 8

evaluation and visualization of immune cell infiltration. (A) PCA cluster plot of immune cell infiltration between SA and control samples. (B) Correlation heat map showing 22 kinds of immune cells. Colored squares highlight the correlation strength; red shows positive correlation; blue shows a negative correlation. (C) Violin diagram showing the proportion of 22 types of immune cells. Of note, red line marks show the different infiltration between the two sample groups.

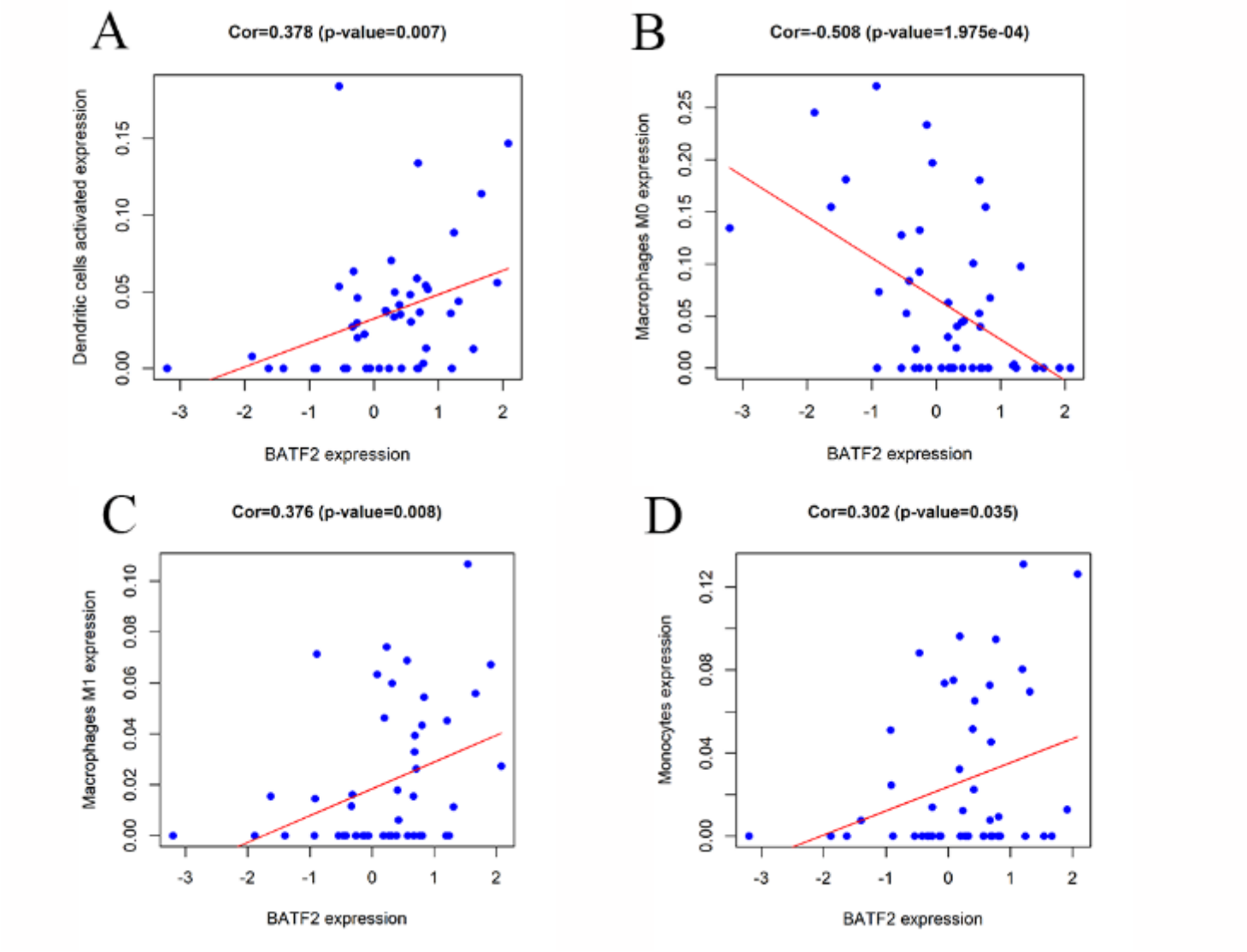

Figure 9

Correlation of BATF2 with infiltration immune cells

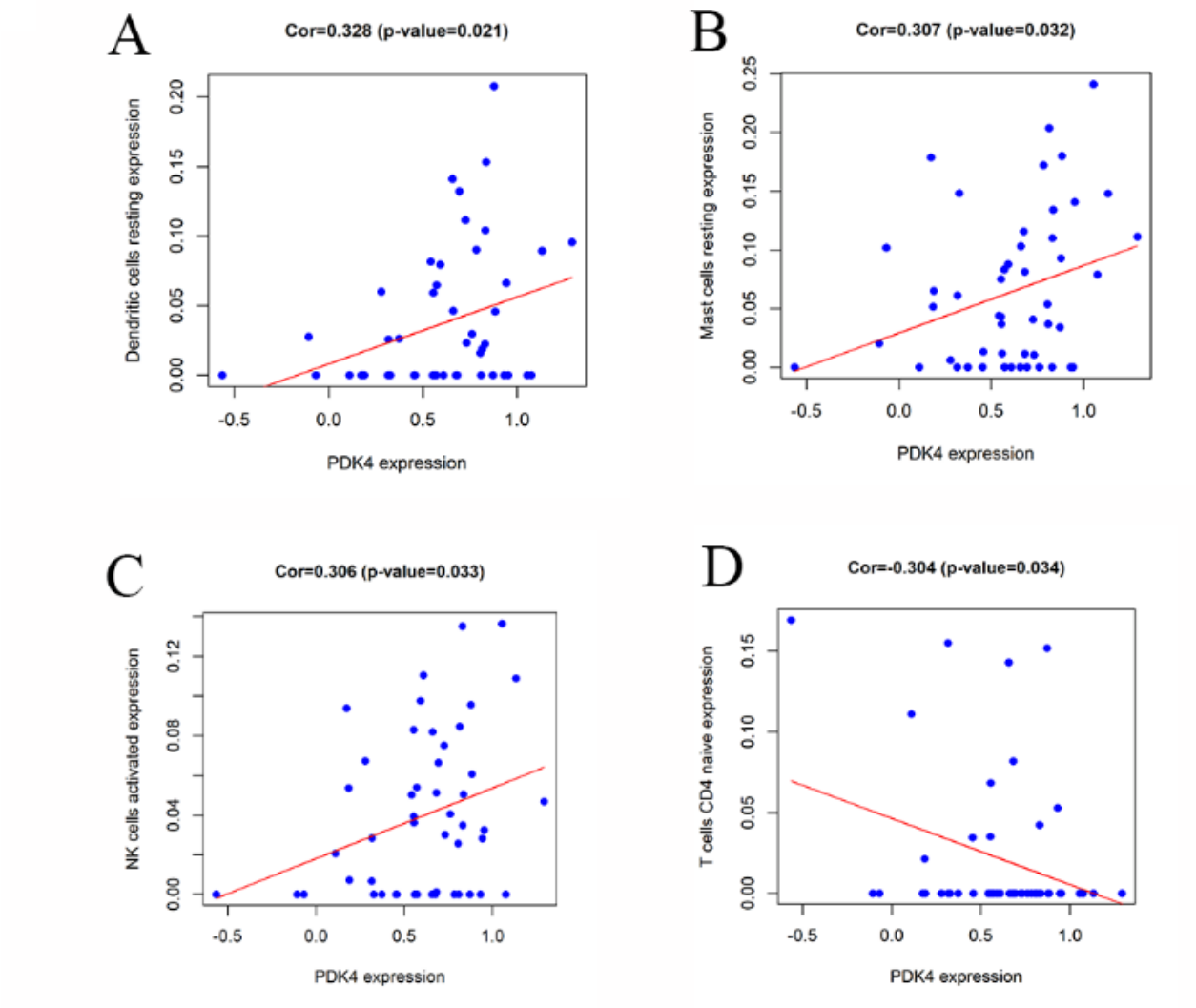

Figure 10

Correlation between PDK4 and infiltration immune cells

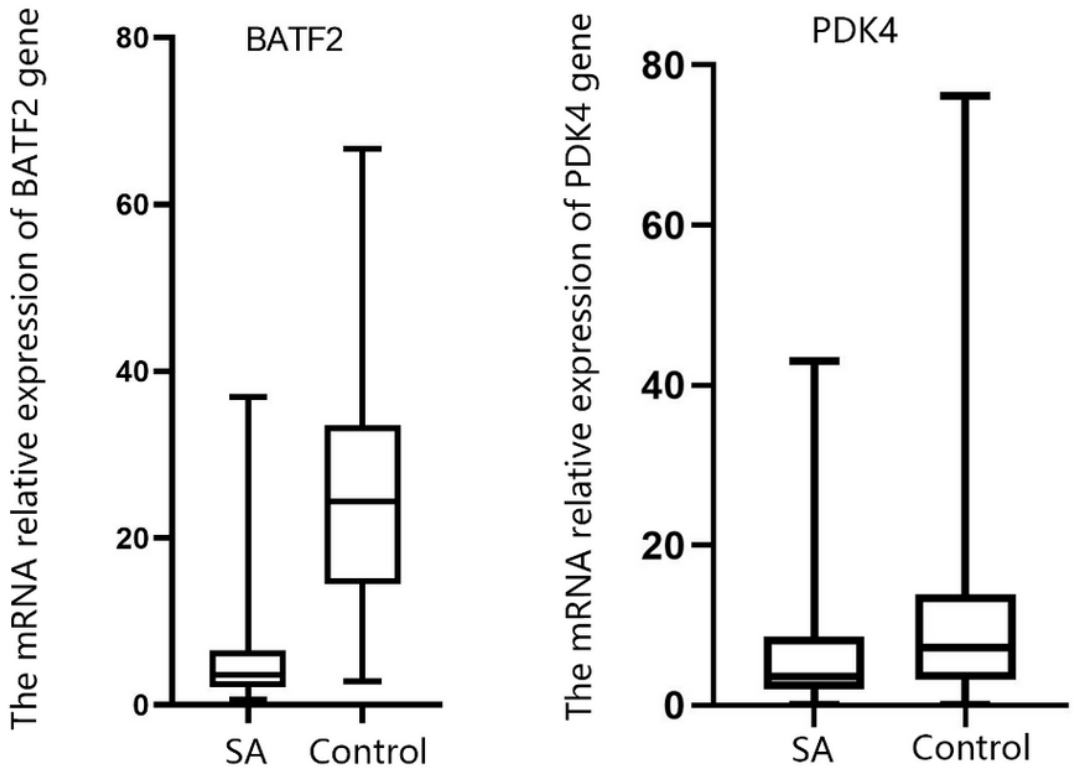

Figure 11

serum BATF2 and serum PDK4 in patients with pulmonary sarcoidosis and controls

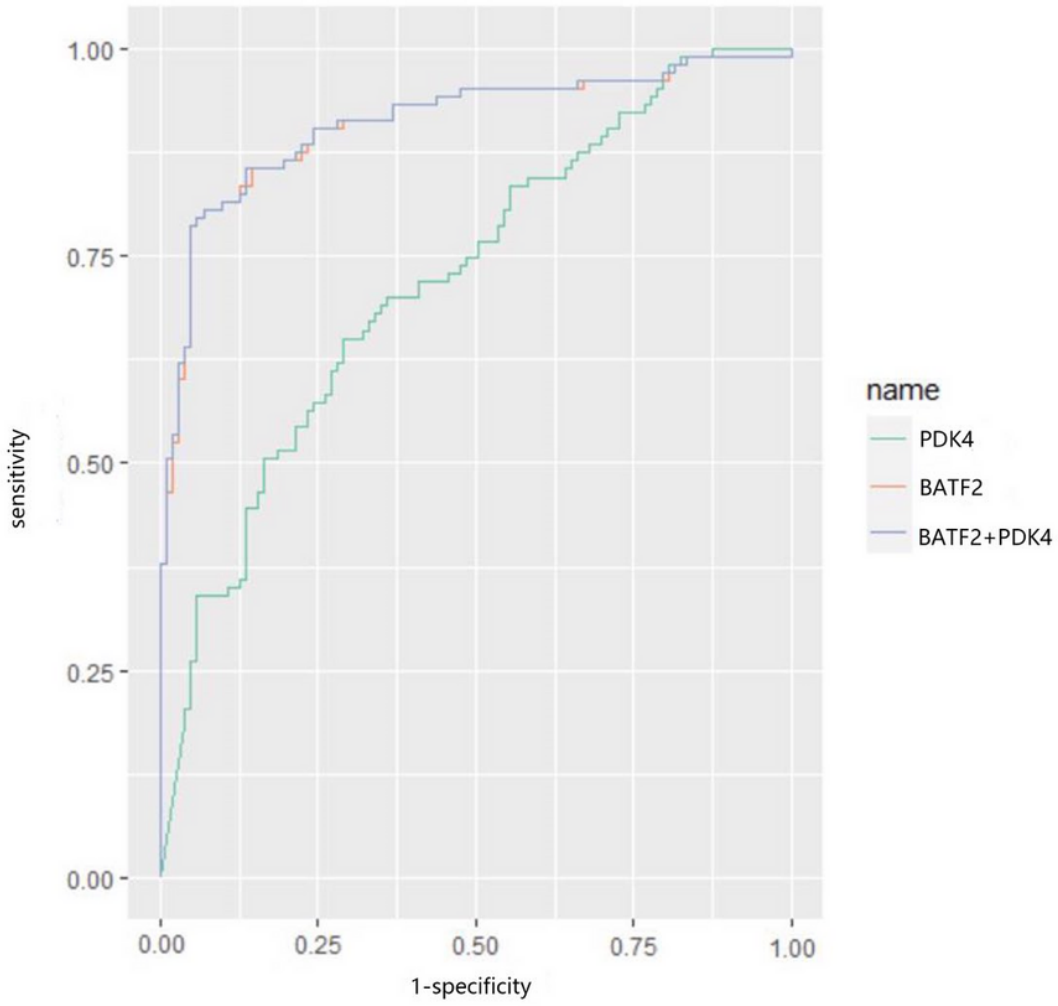

Figure 12

The diagnostic value of serum BATF2 and serum PDK4 for patients with pulmonary sarcoidosis