



# Screening Biomarkers for Systemic Lupus Erythematosus Based on Machine Learning and Exploring Their Expression Correlations With the Ratios of Various Immune Cells

## OPEN ACCESS

### Edited by:

Nancy Agmon-Levin,  
Sheba Medical Center, Israel

### Reviewed by:

Noa Schwartz,  
Albert Einstein College of Medicine,  
United States  
Huo Zhenghao,  
Ningxia Medical University, China

### \*Correspondence:

Yong Dai  
daiyong22@aliyun.com;  
dai.yong@szhospital.com  
Donge Tang  
donge66@126.com  
Dongzhou Liu  
liu\_dz2001@sina.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Autoimmune and Autoinflammatory  
Disorders,  
a section of the journal  
Frontiers in Immunology

Received: 11 February 2022

Accepted: 13 May 2022

Published: 10 June 2022

### Citation:

Zhong Y, Zhang W, Hong X, Zeng Z,  
Chen Y, Liao S, Cai W, Xu Y, Wang G,  
Liu D, Tang D and Dai Y (2022)  
Screening Biomarkers for Systemic  
Lupus Erythematosus Based on  
Machine Learning and Exploring Their  
Expression Correlations With the  
Ratios of Various Immune Cells.  
*Front. Immunol.* 13:873787.  
doi: 10.3389/fimmu.2022.873787

Yafang Zhong<sup>1†</sup>, Wei Zhang<sup>1,2†</sup>, Xiaoping Hong<sup>1</sup>, Zhipeng Zeng<sup>1</sup>, Yumei Chen<sup>1</sup>,  
Shengyou Liao<sup>1</sup>, Wanxia Cai<sup>1</sup>, Yong Xu<sup>3</sup>, Gang Wang<sup>4</sup>, Dongzhou Liu<sup>1\*</sup>,  
Donge Tang<sup>1\*</sup> and Yong Dai<sup>1\*</sup>

<sup>1</sup> Clinical Medical Research Center, Guangdong Provincial Engineering Research Center of Autoimmune Disease Precision Medicine, Shenzhen Engineering Research Center of Autoimmune Disease, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, Shenzhen, China, <sup>2</sup> South China Hospital, Health Science Center, Shenzhen University, Shenzhen, China, <sup>3</sup> The First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital, Shenzhen, China, <sup>4</sup> Department of Nephrology, Shenzhen Hospital, University of Chinese Academy of Sciences, Shenzhen Guangming New District Hospital, Shenzhen, China

**Background:** Systemic lupus erythematosus (SLE) is an autoimmune illness caused by a malfunctioning immunomodulatory system. China has the second highest prevalence of SLE in the world, from 0.03% to 0.07%. SLE is diagnosed using a combination of immunological markers, clinical symptoms, and even invasive biopsy. As a result, genetic diagnostic biomarkers for SLE diagnosis are desperately needed.

**Method:** From the Gene Expression Omnibus (GEO) database, we downloaded three array data sets of SLE patients' and healthy people's peripheral blood mononuclear cells (PBMC) (GSE65391, GSE121239 and GSE61635) as the discovery metadata ( $n_{\text{SLE}} = 1315$ ,  $n_{\text{normal}} = 122$ ), and pooled four data sets (GSE4588, GSE50772, GSE99967, and GSE24706) as the validate data set ( $n_{\text{SLE}} = 146$ ,  $n_{\text{normal}} = 76$ ). We screened the differentially expressed genes (DEGs) between the SLE and control samples, and employed the least absolute shrinkage and selection operator (LASSO) regression, and support vector machine recursive feature elimination (SVM-RFE) analyze to discover possible diagnostic biomarkers. The candidate markers' diagnostic efficacy was assessed using the receiver operating characteristic (ROC) curve. The reverse transcription quantitative polymerase chain reaction (RT-qPCR) was utilized to confirm the expression of the putative biomarkers using our own Chinese cohort ( $n_{\text{SLE}} = 13$ ,  $n_{\text{normal}} = 10$ ). Finally, the proportion of 22 immune cells in SLE patients was determined using the CIBERSORT algorithm, and the correlations between the biomarkers' expression and immune cell ratios were also investigated.

**Results:** We obtained a total of 284 DEGs and uncovered that they were largely involved in several immune relevant pathways, such as type I interferon signaling pathway, defense response to virus, and inflammatory response. Following that, six candidate diagnostic

biomarkers for SLE were selected, namely ABCB1, EIF2AK2, HERC6, ID3, IFI27, and PLSCR1, whose expression levels were validated by the discovery and validation cohort data sets. As a signature, the area under curve (AUC) values of these six genes reached to 0.96 and 0.913, respectively, in the discovery and validation data sets. After that, we checked to see if the expression of ABCB1, IFI27, and PLSCR1 in our own Chinese cohort matched that of the discovery and validation sets. Subsequently, we revealed the potentially disturbed immune cell types in SLE patients using the CIBERSORT analysis, and uncovered the most relevant immune cells with the expression of ABCB1, IFI27, and PLSCR1.

**Conclusion:** Our study identified ABCB1, IFI27, and PLSCR1 as potential diagnostic genes for Chinese SLE patients, and uncovered their most relevant immune cells. The findings in this paper provide possible biomarkers for diagnosing Chinese SLE patients.

**Keywords:** machine learning, diagnostic biomarker, systemic lupus erythematosus, immune cell disturbance, CIBERSORT

## INTRODUCTION

Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease that primarily affects the adolescent and menopausal women (1). The clinical manifestations of SLE are heterogeneous, involving the blood, kidney, nerve and some other systems, making the treatment and management difficult and complicated (2). Currently, the Chinese hospitals often adopt European and American (EULAR/ACR) diagnostic guidelines for SLE diagnosis (3, 4), whereas the molecular profiling of SLE is highly heterogeneous among different races and locations (5, 6). Consequently, the specificity and sensitivity of diagnostic indicators recommended by the EULAR/ACR need to be validated further in Chinese population.

Nowadays, the diagnosis of SLE is primarily based on a series of clinical manifestations and laboratory indicators, such as skin erythema, arthralgia, complement C3, C4, and anti-dsDNA antibodies, etc. (7–11). SLE must be diagnosed based on clinical signs along with multiple immunological indicators. When patients were diagnosed as SLE, they often had occurred a certain degree of systemic involvements (12). Even an invasive biopsy is required to diagnose lupus nephritis (13). Because of the complexity, lag, and invasiveness of current SLE diagnostic methods, researchers are looking for new genetic biomarkers in the hopes of developing a simpler, faster, and more objective diagnostic “gold standard” than current markers/tests, without the need for clinical symptoms, especially in the Chinese population.

Until now, immune cells such as lymphocytes, dendritic cells, macrophages, basophils, and neutrophils have been identified to be disrupted in the course of SLE (14–17). There are differences in the disturbed cell types and proportions between SLE individuals (18–21). The identification of altered immune cells in SLE aids us in understanding the cellular results of the disease and establishing an appropriate diagnostic or treatment strategy.

In this study, we used the Gene Expression Omnibus (GEO) database to obtain three expression microarray datasets of SLE patients' and healthy people's peripheral blood mononuclear cells (PBMC). Then, we pooled the three datasets as a metadata cohort ( $n_{\text{SLE}} = 1315$ ,  $n_{\text{normal}} = 122$ ) and looked for the differentially

expressed genes (DEGs) between SLE and controls. Next, we identified the diagnostic biomarkers of SLE using different machine learning algorithms. Following that, we merged four GEO data sets as the validation data (GSE4588, GSE50772, GSE99967, and GSE24706;  $n_{\text{SLE}} = 146$ ,  $n_{\text{normal}} = 76$ ) and confirmed the expression of the identified diagnostic biomarkers, then used logistic regression to develop a diagnostic prediction model. Furthermore, we validated the expression of the candidate biomarkers using our own Chinese cohort ( $n_{\text{SLE}} = 13$ ,  $n_{\text{normal}} = 10$ ). Moreover, we used the CIBERSORT algorithm to quantify the proportion of 22 immune cells in PBMC of SLE patients and healthy people. Finally, we explored the relationship between the expression of the identified biomarkers and the ratios of immune cells in PBMC of SLE patients.

## MATERIALS AND METHODS

### Data Download and Processing

Seven expression matrix files of SLE PBMC samples were downloaded from the GEO database, namely GSE65391, GSE121239, GSE61635, GSE4588, GSE50772, GSE99967, and GSE24706. The GSE65391 dataset contained 924 SLE and 72 control samples, the GSE121239 dataset contained 292 SLE and 20 control samples, GSE61635 dataset contained 99 SLE and 30 control samples, the GSE4588 dataset contained 15 SLE and 19 control samples, the GSE50772 dataset contained 61 SLE and 20 control samples, the GSE99967 dataset contained 42 SLE and 17 control samples, and GSE24706 dataset contained 28 SLE and 20 control samples. Then, the three files GSE65391, GSE121239, and GSE61635 were pooled into a metadata cohort for the following analysis, after the batch effects were removed using the R package “SVA”. In addition, the four data sets (GSE4588, GSE50772, GSE99967, and GSE24706) were also merged as the verification data after normalization.

### DEGs Screening and Functional Analysis

The DEGs were screened out using the R package “limma” based on the metadata cohort's data set. The heat map showing the

expression of DEGs was depicted using the R package. Next, the DAVID database was used to analysis the functions of the DEGs based on gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. In addition, the R package “clusterProfiler” and “DOSE” were used for the Disease ontology (DO) enrichment analyses. The gene set enrichment analysis (GSEA) was used to distinguish the most significant functional items between SLE and controls. The gene set “c2.cp.kegg.v7.0.symbols.gmt” from the Molecular Signatures Database (MSigDB) (22) was selected as the reference gene set.

## Identification and Verification of the Diagnostic Markers

Two algorithms, namely least absolute shrinkage and selection operator (LASSO) logistic regression and support vector machine-recursive feature elimination (SVM-RFE), were used to screen the potential SLE diagnostic markers. The “glmnet” R package was used to perform the LASSO analysis to identify the optimal diagnostic markers in SLE, and the SVM-RFE was used to find the optimal variables. The candidate diagnostic markers were picked by intersecting the markers identified by these two algorithms. Then, the expression of the candidate diagnostic biomarkers were verified based on the merged dataset containing GSE4588, GSE50772, GSE99967, and GSE24706.

## The Diagnostic Efficacy of the Biomarkers in SLE

To examine the diagnostic efficacy of the candidate markers, the receiver operating characteristic (ROC) curves were drawn based on the discovery metadata and the validation data set.

## Patients

A total of 23 whole blood samples (including 13 SLE samples and 10 control samples) were collected from Shenzhen People’s Hospital. Patients diagnosed as SLE, with a SLE disease activity index (SLEDAI) score more than 5 were included. The clinical manifestations of all patients were shown in **Table 1**. All participants volunteered to participate in this research. This work was approved by the Ethics Committee of Shenzhen People’s Hospital (LL-KY-2021393). After collecting all of the whole blood samples, the PBMCs were isolated, and dissolved in Trizol (Beyotime, R-0016), and then stored in  $-80^{\circ}\text{C}$ .

## Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR)

The total RNA was isolated from the PBMC according to the manufacturer’s instructions. The transScript all-in-one first-strand cDNA synthesis superMix for qPCR (One-step gDNA removal) kit (TransGen Biotech, AT341-01) was used for the reverse transcription of mRNA. After that, the qPCR assays were conducted out with the PerfectStart Green qPCR SuperMix kit (TransGen Biotech, AQ601-02), with the following primers: GAPDH (Forward: 5-TGACTTCAACAGCGACACCCA-3. Reverse: 5-CACCCTGTTGCTGTAGCCAAA-3). ABCB1 (Forward: 5-GTCTGGACAAGCACTGAAAGATAAGA-3. Reverse: 5-CAACGGTTCGGAAGTTTTCTATTGC-3).

EIF2AK2 (Forward: 5-AGAAGGCGGAGCGTGAAGTAAAA G-3. Reverse: 5-ATCCATCCCAACAGCCATTGTAGTG-3). HERC6 (Forward: 5-CCTGCCAAGCCTAAACCTGAGAAG-3. Reverse: 5-TACAGAGCCAGTGGGAAAGGAAGG-3). ID3 (Forward: 5-GGACGACATGAACCACTGCT-3. Reverse: 5-TCAGTGGCAAAAGTCCTTT-3). IFI27 (Forward: 5-TCTG CAGTCACTGGGAGCAA-3. Reverse: 5-CCCAATGGAGC CCAGGAT-3), PLSCR1 (Forward: 5-CCTCAGTATCCACCG ACAGCATTC-3. Reverse: 5-ACTGGCTGATTTGGGACAGG AAAG-3). All the primers were synthesized by the Sangon Biotech Company. Among them, GAPDH was an internal reference gene. The gene relative expressions were calculated by the  $2^{-\Delta\Delta\text{CT}}$  method. A  $p$  value  $< 0.05$  was considered statistically significant.

## Immune Cell Composition

The CIBERSORT algorithm (23) was used to calculate the ratios of various immune cells in PBMC of SLE patients and healthy people based on the expression matrix, and the R package “vioplot” was used to visualize the proportions of 22 immune cells in SLE and control groups. The “corrplot” package was used to create the heat map displaying the quantitative correlation between distinct immune cells. In addition, the “ggplot2” R package was utilized to investigate the association between the expression of the diagnostic markers and the ratios of immune cells.

## RESULTS

### Screening the DEGs Between the SLE PBMC Versus Control PBMC

After batch effects were removed using the R package “SVA,” we combined three expression array data sets collected from the GEO database (GSE65391, GSE121239, and GSE61635) into one discovery data set, which included data from 1315 SLE patients and 122 healthy people. We used principal component analysis (PCA) to cluster all of the samples and discovered that each sample’s point was randomly distributed, indicating that the normalization was done correctly (**Supplementary Figure S1**). As a consequence, 284 genes were discovered to be differentially expressed in the PBMC of SLE patients versus healthy people ( $p$  value  $< 0.0001$ , fold change  $> 1.5$ ) (**Figure 1A**). We used the heat map showing the expression of the DEGs in each healthy people and SLE patient (**Figure 1B**). Through an enrichment analysis of the DEGs, we found that these DEGs were enriched in some typical autoimmune-disease-relevant pathways, such as type I interferon signaling pathway, innate immune response, inflammatory response, and systemic lupus erythematosus (**Figures 1C–F**). In addition, we conducted the DO enrichment analyses of the DEGs, and discovered that the DEGs were primarily involved in several immune related diseases, namely bacterial infectious disease, hematopoietic system disease, human immunodeficiency virus infectious disease, and rheumatic disease, etc. (**Figure 1G**). Furthermore, the GSEA results showed that the DEGs in SLE patients were mainly involved with chemokine signaling pathway, complement and coagulation cascades, cytoplasmic DNA sensing

**TABLE 1 |** The clinical information of the 13 SLE patients and 10 healthy people in our study.

	SLE (n = 13)	Healthy people (n = 10)
Gender (female/male)	11/2	9/1
Age (year), median (range)	38 (15-57)	31 (26-37)
C3(g/L), median (range)	0.76 (0.21-0.85)	N/A
C4(g/L), median (range)	0.14 (0.03-0.28)	N/A
ANA(+) (percentage)	13 (100%)	N/A
Anti-dsDNA(+) (percentage)	10 (76.9%)	N/A
ANuA(+) (percentage)	9 (69.2%)	N/A
Proteinuria (g/24 h), median (range)	0.363 (0.045-10.98)	N/A
SLEDAI-2K, median (range)	9 (6-21)	N/A
Rash (percentage)	3 (23.1%)	N/A
Alopecia (percentage)	2 (15.4%)	N/A
Fever (percentage)	4 (30.8%)	N/A
Pleurisy (percentage)	3 (23.1%)	N/A
Leukopenia (percentage)	4 (30.8%)	N/A
Pericarditis (percentage)	3 (23.1%)	N/A
Arthritis (percentage)	2 (15.4%)	N/A
Organic encephalopathy syndrome (percentage)	1 (7.7%)	N/A
Hematuria (percentage)	1 (7.7%)	N/A
Thrombocytopenia (percentage)	2 (15.4%)	N/A
Lupus headache (percentage)	1 (7.7%)	N/A
Vasculitis (percentage)	2 (15.4%)	N/A

SLE, systemic lupus erythematosus; ANA, anti-nuclear antibody; Anti-dsDNA, anti-dsDNA antibodies; ANuA, anti-nucleosome antibodies; C3, Complement C3; C4, Complement C4; N/A, not applicable; SLEDAI-2K, SLE disease activity index 2000.

pathway, RIG-I-like receptor signaling pathway, and Toll-like receptor signaling pathway (Figures 1H, I).

## Identification of the Potential Diagnostic Biomarkers for SLE Based on Machine Learning

Then, we adopted two machine learning algorithms to screen the biomarkers for SLE, namely LASSO regression and SVM-RFE algorithm. As a result, the LASSO regression algorithm revealed 70 probable biomarkers, while the SVM-RFE approach identified 37 (Figures 2A, B). After that, we made an intersection of the 70 and 37 probable biomarkers to arrive at 14 common biomarkers (Figure 2C; Supplementary Table S1). To further confirm the reliability of these biomarkers, we verified their expressions based on the validation data set ( $n_{\text{SLE}} = 146$ ,  $n_{\text{normal}} = 76$ ). The result showed that 10 genes harbored the similar expression trend with statistical significance in both discovery and validation data sets, including ABCB1, EIF2AK2, HERC6, PLSCR1, ID3, IFI27, SCRN1, CD160, HSP90AB1, and PCYOX1. Among these 10 genes, we selected ABCB1, EIF2AK2, HERC6, PLSCR1, ID3, and IFI27 for the following study because the statistical differences of their expression was most significant in both the discovery and validation sets (Figures 2D–I).

## The Diagnostic Efficacy of the Six Candidate Biomarkers for SLE

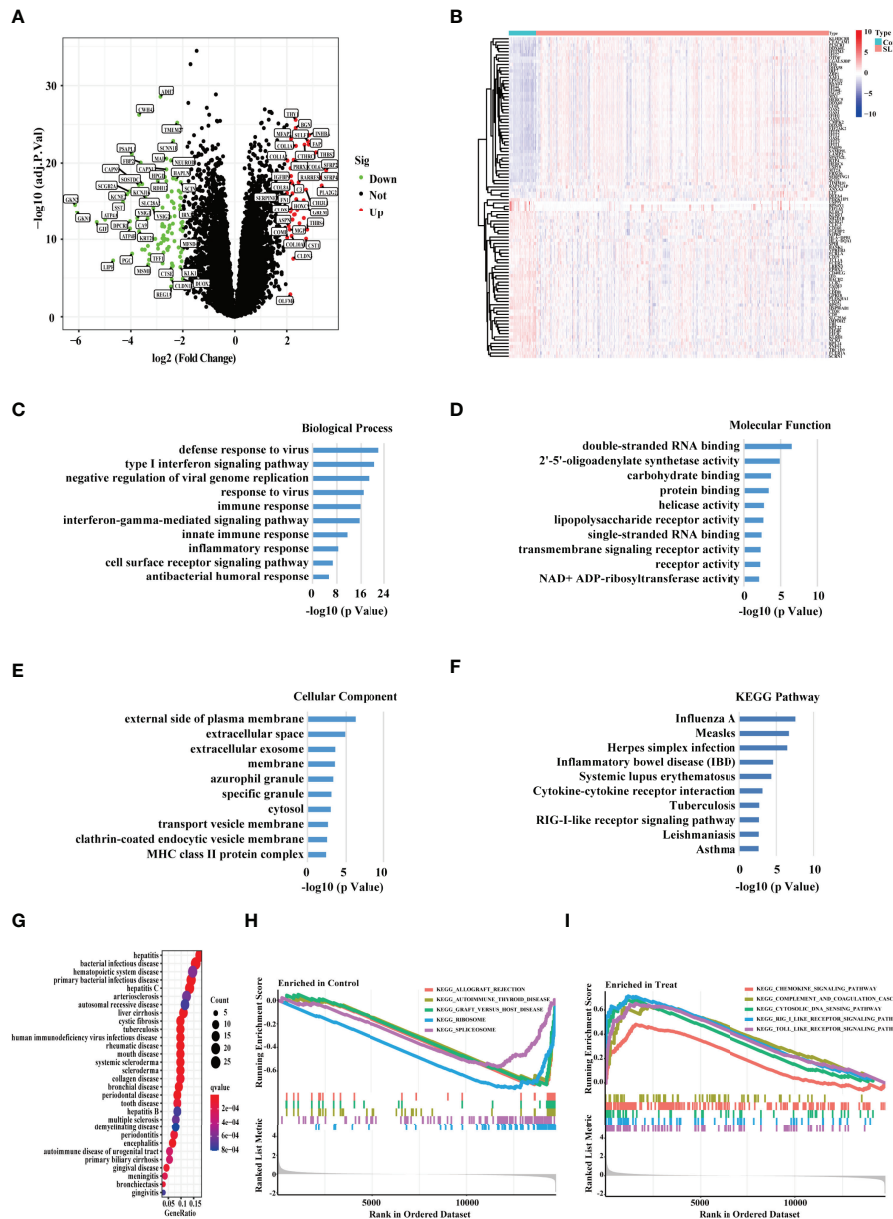
Subsequently, we plotted the ROC curves for the six candidate biomarkers, and found that ABCB1, EIF2AK2, HERC6, ID3, IFI27, and PLSCR1 all had a good diagnostic effect in the discovery data set, with an AUC value 0.839, 0.912, 0.894, 0.891, 0.902, and 0.907, respectively. When we used the six markers as one single signature, the AUC value reached 0.96 (Figure 3A). On the other hand, we also verified the diagnostic

effect of the six genes in the validation data set, and discovered that the AUC values of the six biomarkers, including ABCB1, EIF2AK2, HERC6, ID3, IFI27, and PLSCR1, were 0.754, 0.854, 0.81, 0.731, 0.875, and 0.851, respectively. When the six markers were combined as one single signature, the AUC value reached 0.913 (Figure 3B). The result suggests that the six genes have a good diagnostic power for SLE, and the power is higher when they are used together.

## The Validation of the Potential Diagnostic Markers Using Our Own Chinese Cohort Revealed that ABCB1, IFI27 and PLSCR1 Were More Likely to be SLE Biomarkers for the Chinese Population

As mentioned above, SLE is highly heterogeneous among different races and regions (24). To ensure that the biomarkers we screened could be applied in the Chinese population, we further performed the RT-qPCR assays on the expression of these six markers in the PBMC of 13 Chinese SLE patients and 10 Chinese healthy controls (Supplementary Table S1). Consistently, the mRNA expression level of ABCB1 in PBMC of SLE patients was significantly lower than that of healthy controls (Figure 4A). SLE patients had considerably greater levels of IFI27 and PLSCR1 expression than healthy controls (Figures 4B, C). Meanwhile, we found that the expression differences of EIF2AK2, HERC6 and ID3 in SLE patients and healthy controls followed the same pattern as the discovery and validation cohorts (though the difference was not statistically significant) (Figures 4D–F). Thus, these results indicate that our findings are reproducible, and ABCB1, IFI27 and PLSCR1 are more likely to be SLE biomarkers in the Chinese population. Notably, ABCB1 is a novel found biomarker for SLE that has not yet to be published, to the best of our knowledge.



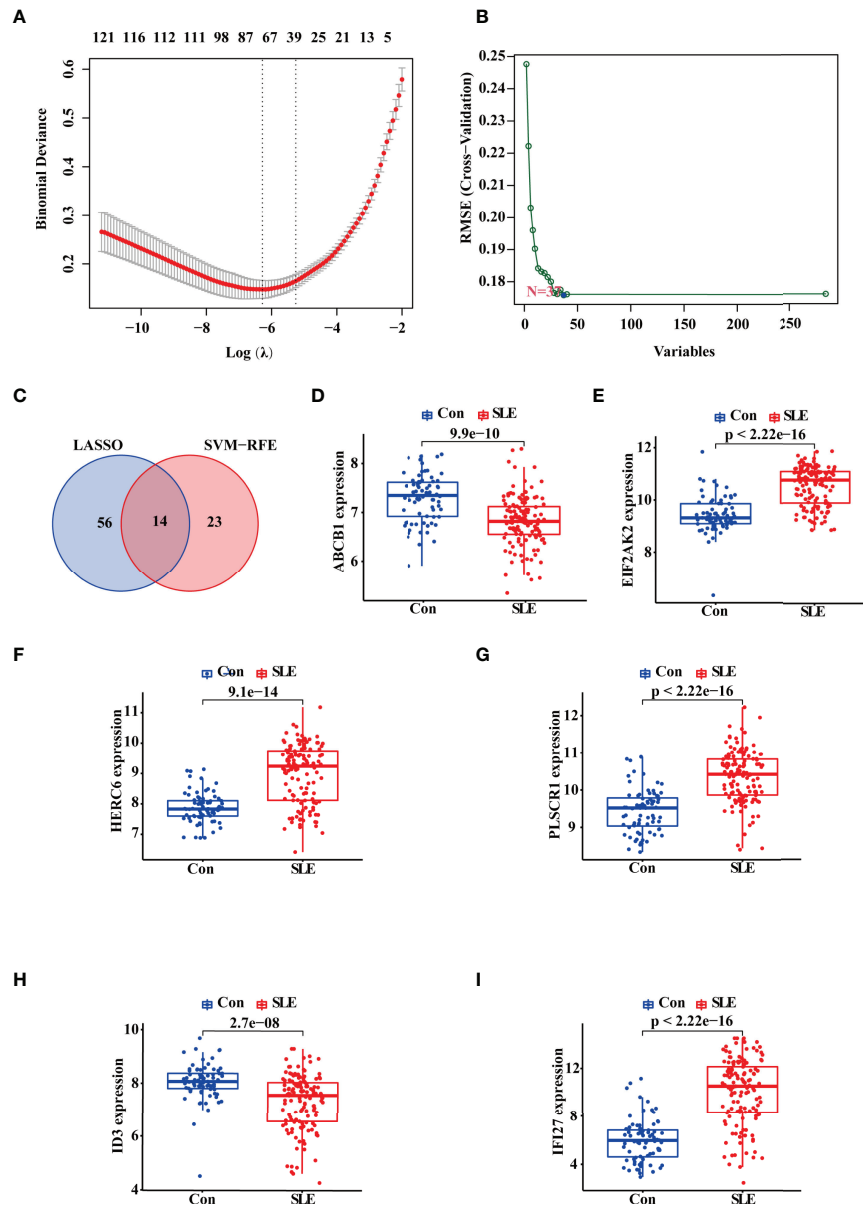


**FIGURE 1** | Screening the DEGs between the SLE PBMC versus Control PBMC. **(A)** Volcano plot showing the DEGs between the PBMC of SLE and control samples based on the metadata set including GSE65391, GSE121239, and GSE61635. Green represented down-regulation, and red represented up-regulation (SLE/Control). **(B)** The heat map showing the DEGs in each sample of SLE patients and healthy people. “Con” represented control samples, and “SLE” represented SLE patients. **(C–E)** The GO, **(F)** KEGG, and **(G)** DO enrichment analysis of the DEGs. The size of the dots represented the number of genes in each enriched module. The color of the dots represented the q value. **(H–I)** The GSEA analysis revealing the enriched pathways in the PBMC of SLE and control samples.

## The Ratio Changes of Immune Cells in SLE Patients, and Their Correlation With the Expression of ABCB1, IFI27 and PLSCR1

The onset of SLE causes changes in the proportion and function of a series of immune cells. To find the biomarkers whose expression were correlated with the proportions of immune cells, we first analyzed the ratio changes of 22 immune cells in

1315 SLE patients and 122 healthy people using the CIBERSORT algorithm. The results showed that compared with the control group, the proportions of naive B cells, CD8 T cells, CD4 memory resting T cells, CD4 memory activated T cells, and resting natural killer (NK) cells in SLE were significantly lower, while monocytes, macrophages M0, activated dendritic cells, neutrophils were higher (**Figure 5A**). Consistently, a number of previous studies have also demonstrated that the ratio of

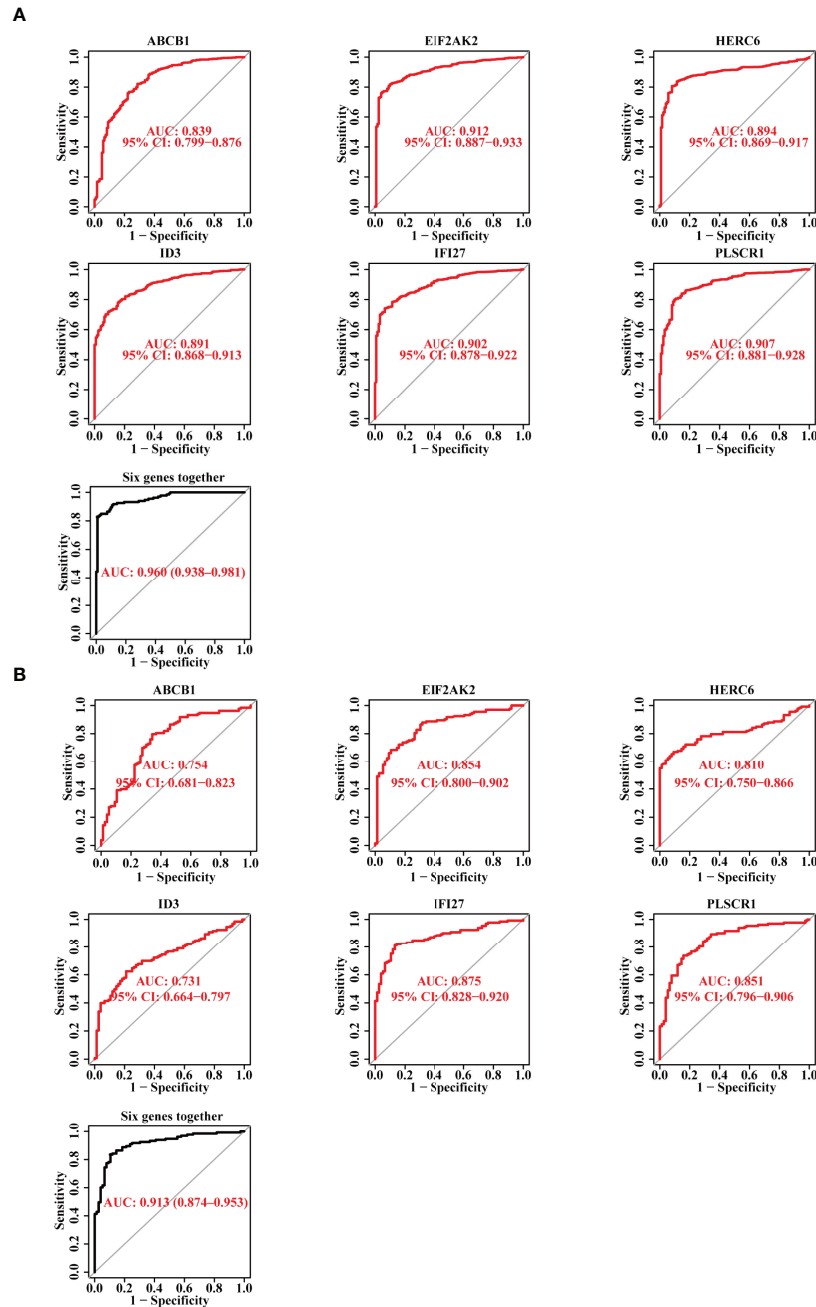


**FIGURE 2** | Identification and Validation of the Diagnostic Biomarkers for SLE. **(A, B)** LASSO logistic regression and SVM-RFE algorithm screening diagnostic biomarkers for SLE. **(C)** The Venn diagram showing the intersection of the diagnostic biomarkers screened by the two algorithms. **(D–I)** Validation of the expression of diagnostic biomarkers based on the validate data set including GSE4588, GSE50772, GSE99967, and GSE24706. “Con” represented the control samples, and “SLE” represented the SLE patients.

resting NK cells was decreased in SLE patients and the ratios of monocytes, neutrophils were increased in SLE patients (25–27).

Following that, we investigated the correlation between the ratios of the 22 types of immune cells in SLE patients, and discovered that the degrees of memory B cells and activated dendritic cells, the levels of regulatory T cells (Tregs) and memory B cells, the levels of Tregs and activated dendritic cells all had a strong positive link, respectively. Furthermore, the ratio of CD8 T cells was adversely linked with that of neutrophils, memory B cells, and monocytes (**Figure 5B**).

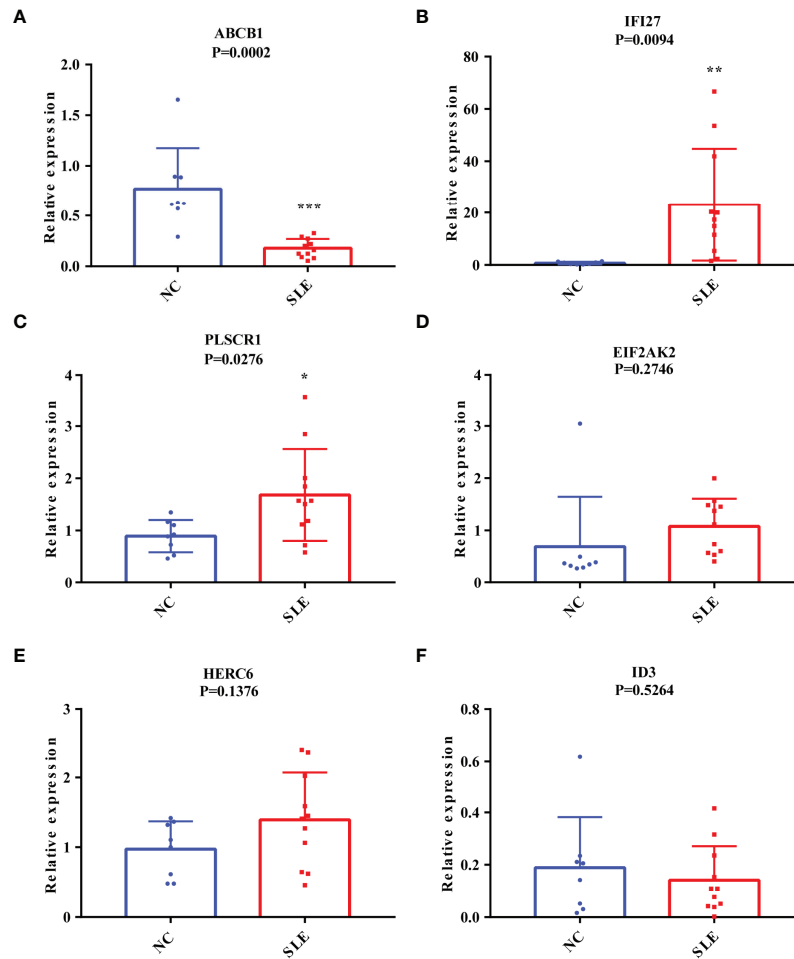
Finally, we looked at the relationship between the immune cell ratios and the expression of ABCB1, IFI27 and PLSCR1 in SLE patients. As a result, the ABCB1 expression was positively correlated with the levels of CD8 T cells, resting NK cells, CD4 memory resting T cells, naive B cells, CD4 memory activated T cells, and macrophages M2, and negatively correlated with the ratios of activated mast cells, resting mast cells, CD4 naive T cells, Tregs, memory B cells, plasma cells, macrophages M0, activated dendritic cells, neutrophils, and monocytes (**Figure 5C**). The IFI27 expression was positively linked with the degrees of



**FIGURE 3** | The Diagnostic Efficacy of the Selected Diagnostic Markers for SLE. **(A)** The ROC curve showing the AUC value of ABCB1, EIF2AK2, HERC6, ID3, IFI27, and PLSCR1 based on the data set of the discovery cohort. **(B)** The ROC curve showing the AUC value of ABCB1, EIF2AK2, HERC6, ID3, IFI27, and PLSCR1 based on the data set of the validation cohort.

activated dendritic cells, plasma cells, resting mast cells, activated NK cells, neutrophils, macrophages M0, memory activated cells CD4, and macrophages M1, and negatively linked with CD8 T cells, resting dendritic cells, naive B cells, and CD4 memory resting T cells (**Figure 5D**). The PLSCR1 expression was positively correlated with the ratios of neutrophils, activated dendritic cells, resting mast cells, macrophages M0, plasma cells,

monocytes, and activated NK cells, and negatively correlated with eosinophils, CD4 naive T cells, resting dendritic cells, resting NK cells, CD8 T cells, naive B cells, CD4 memory resting T cells (**Figure 5E**). All in all, the ABCB1 expression was most closely linked to the ratios of CD8 T cells and monocytes, the IFI27 expression to the levels of activated dendritic cells and CD4 memory resting T cells, and the



**FIGURE 4** | The Validation of the Putative Diagnostic Markers by RT-qPCR in Chinese population using Our Own Cohort. (A-F) The relative mRNA expression levels of ABCB1, IFI27, PLSCR1, EIF2AK2, HERC6, and ID3 in the PBMC of SLE patients and healthy people. \*\*\* represented  $P < 0.05$ , \*\*\*\* represented  $P < 0.01$ , and \*\*\*\*\* represented  $P < 0.001$ .

PLSCR1 expression to the degrees of neutrophils and CD4 memory resting T cells.

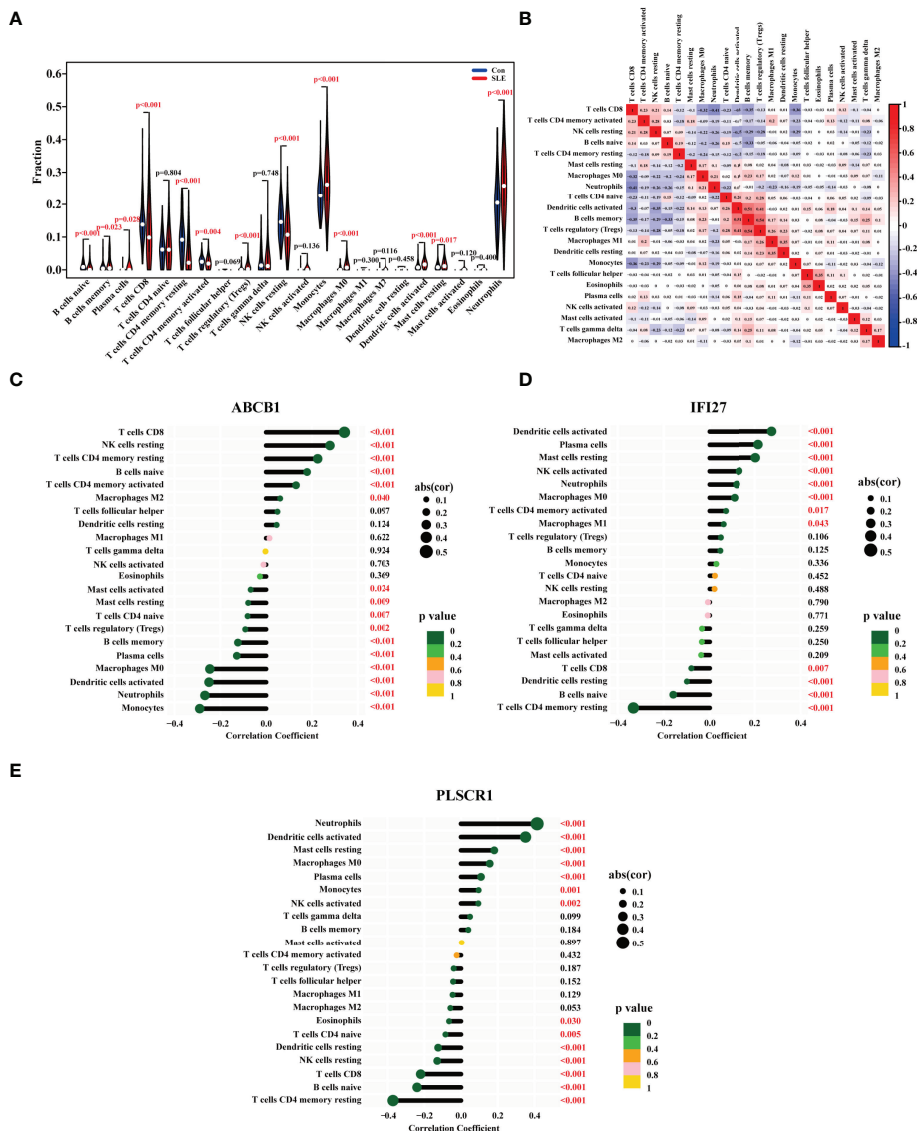
## DISCUSSION

SLE is a type of multi-system autoimmune disease, the pathophysiological mechanisms of which have not been thoroughly elucidated (27). In clinic, SLE is diagnosed according to the clinical features, such as injuries of skin, joints, kidneys, nervous system, as well as the serologic parameters, such as autoimmune antibody. Furthermore, the diagnosis of SLE must be based on the clinical manifestations (such as organ damages) together with multiple immunological indicators (12, 28). Some types of organ damage need to be diagnosed in an invasive way, causing great pain to the patients. Therefore, the diagnosis of SLE is a tedious, complex and time-consuming process, and it is of great significance to search for genetic markers for SLE diagnosis.

In this study, we screened the putative diagnostic biomarkers of SLE using machine learning. We all know that the conclusions drawn from a single data set are typically limited, unrepeatable, or inconsistent. To ensure the correctness of our findings, we retrieved three cohorts from the GEO database ( $n_{\text{SLE}} = 1315$ ,  $n_{\text{normal}} = 122$ ). Among the six putative biomarkers, we confirmed their expression using our own Chinese cohort ( $n_{\text{SLE}} = 13$ ,  $n_{\text{normal}} = 10$ ), and three of them were statistically significant, while the other three genes showed the same expression trend to the machine learning results. The reason for this phenomenon could be due to the small sample size and the racial heterogeneity. Thus, a bigger cohort may be necessary for the further validation in future research.

In 2021, Zhao X's team has revealed several potential SLE biomarkers based on a comprehensive bioinformatics analysis (27). Compared with this paper, we have the following advancements, extensions and deeper studies: First, we established a larger discovery cohort ( $n_{\text{SLE}} = 1315$ ,  $n_{\text{normal}} = 122$ ) to ensure that the markers screened out are more universal and accurate. Second,





**FIGURE 5 |** The Ratios of Immune Cells in the PBMC of SLE Patients and their Correlations with the Biomarkers' Expression. **(A)** Violin diagram showing the proportion of 22 types of immune cells in the PBMC of SLE patients and healthy people. The red marks represented  $p < 0.05$ . **(B)** The heat map showing the proportion correlation between the 22 types of immune cells in SLE patients. The size of the colored squares represented the strength of the correlation, and red represented a positive correlation, and blue represented a negative correlation. The correlation between **(C)** ABCB1, **(D)** IFI27, **(E)** PLSCR1 expression with the degrees of immune cells in SLE. The size of the dots represented the strength of the correlation. The color of the dots represented the p value. A  $p < 0.05$  was considered statistically significant.

we used two machine learning methods, LASSO and SVM, to screen diagnostic markers, whereas Zhao X's work screened biomarkers through overlapping the DEGs of six independent data sets. Compared with the intersection method, LASSO and SVM have stronger power in identifying biomarkers with higher distinguishing efficacy. Third, we validated the screened markers using our own Chinese cohort, thereby the markers had potential application for Chinese patients.

We also observed a link between the infiltration of certain immune cells and the expression of biomarker genes. However,

more evidence is needed to support the link in a wider range of situations, and the relevance of this correlation is unknown at this time. Furthermore, immune cell infiltration and marker gene expression may not be causally related. The systemic inflammation that occurs in lupus may cause changes in immune cell ratio, and the marker gene expression may be linked to that inflammation.

Among the six SLE candidate markers, ABCB1, HERC6 and ID3 are identified for the first time. Furthermore, we confirmed the expression of ABCB1 in the Chinese population using our own cohort. ABCB1, also known as P-glycoprotein (P-GP), is a

multidrug resistance protein mainly expressing in barrier organs such as brain, liver, kidney, and skin (29–32). Recently, ABCB1 has been reported being expressed in a variety of immune cells, such as monocytes, antigen presenting dendritic cells, T cells and B cells, and is involved in the efflux of inflammatory molecules (33). Therefore, it's possible that ABCB1 plays a role in SLE by participating in the inflammatory pathways. Besides, HERC6 has been revealed as a key determinant of cellular antiviral activity, and is also one constituent gene of type I interferon signaling pathway (34, 35), which was regarded as a central pathogenic pathway in SLE (36). Additionally, several studies have proclaimed that ID3 is a key regulator of IL-5 production and the homeostasis of B-1a B cells (37). ID3 helps the tuberculosis subunit vaccine to induce long-term immune memory, providing immune protection against M tuberculosis infection (38).

Machine learning, on the other hand, detected three recognized SLE biomarkers, including EIF2AK2, IFI27, and PLSCR1, confirming the algorithms' accuracy. The protein encoded by EIF2AK2, as one key component of the innate immune system, is increasingly expressed in T cells of SLE patients, and is likely to be associated with cellular translation and proliferation (39). Besides, EIF2AK2 selectively regulates the transcription of genes functioning in immune response in SLE (39). In proteasome-associated autoinflammatory cell models, the activation of EIF2AK2 is discovered responding to the decreased proteasome function (40). Numerous studies have demonstrated the indispensable role of IFI27 (Interferon (IFN)- $\alpha$ -inducible protein 27) in SLE (41–43). IFI27 is strongly correlated with the levels of T helper type 1 (Th1) cells, T helper type 2 (Th2) cells and activated dendritic cells (aDC), and the up-regulation of IFI27 is highly related to many inflammatory events induced by these immune cells (44, 45). Ultimately, PLSCR1 has been found an increased expression in multiple systemic autoimmune diseases, such as primary antiphospholipid syndrome, rheumatoid arthritis, idiopathic inflammatory myopathies and SLE (46, 47). Meanwhile, the PLSCR1 expression was discovered to be associated with type I interferon-stimulated genes (48), and was higher in neutrophils, dendritic cells, and macrophages (43), which is consistent with our results.

## CONCLUSIONS

Generally speaking, we identified six genes as prospective SLE biomarkers ( $n_{\text{SLE}} = 1315$ ,  $n_{\text{normal}} = 122$ ), including ABCB1, EIF2AK2, HERC6, ID3, IFI27, and PLSCR1, and demonstrated that ABCB1, EIFI27, and PLSCR1 might be suitable for the Chinese population. Meanwhile, we discovered the quantitative changes of 13 types of immune cells in SLE patients, as well as the link between the expression of ABCB1, IFI27, and PLSCR1 and

the ratios of different immune cells. Our findings provide potential biomarkers for Chinese SLE patients, and give the insight into the relationship between gene expression and immune cell ratios.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Shenzhen People's Hospital (LL-KY-2021393). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

WZ conceived and designed the project. YZ wrote the manuscript. SL provided data analysis support. XH and DL contributed to clinical sample collection. YC, ZZ, and WC performed the experiments. YX and GW provided conceptual guidance. DT and YD supervised the study. All authors read, revised and approved the final manuscript.

## FUNDING

This study was supported by grants from the Key Research and Development Program of Guangdong Province (No. 2019B020229001), the science and technology plan of Shenzhen (No. JCYJ20200109144218597), Shenzhen Guangming District Science and Technology Innovation Bureau (No. 2021R01020), Shenzhen Key Medical Discipline Construction Fund (No. SZXK011), and the Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515111071).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.873787/full#supplementary-material>

## REFERENCES

1. Kaul A, Gordon C, Crow MK, Touma Z, Urowitz MB, van Vollenhoven R, et al. Systemic Lupus Erythematosus. *Nat Rev Dis Primers* (2016) 2:16039. doi: 10.1038/nrdp.2016.39
2. Mohamed A, Chen Y, Wu H, Liao J, Cheng B, Lu Q. Therapeutic Advances in the Treatment of SLE. *Int Immunopharmacol* (2019) 72:218–23. doi: 10.1016/j.intimp.2019.03.010
3. Low ESH, Krishnaswamy G, Thumboo J. Comparing the 1997 Update of the 1982 American College of Rheumatology (Acr-97) and the 2012 Systemic

- Lupus International Collaborating Clinics (Slicc-12) Criteria for Systemic Lupus Erythematosus (Sle) Classification: Which Enables Earlier Classification of Sle in an Urban Asian Population? *Lupus* (2019) 28(1):11–8. doi: 10.1177/0961203318811599
4. Selvananda S, Kan SL. Performance of the 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus in a Multiethnic Malaysian Cohort. *Int J Rheumatic Dis* (2021) 25(2):131–9. doi: 10.1111/1756-185x.14269
  5. Minhas D, Marder W, Harlow S, Hassett AL, Zick SM, Gordon C, et al. Access and Cost-Related Nonadherence to Prescription Medications Among Lupus Patients and Controls: The Michigan Lupus Epidemiology and Surveillance Program. *Arthritis Care Res* (2021) 73(11):1561–7. doi: 10.1002/acr.24397
  6. Pons-Estel GJ, Ugarte-Gil MF, Alarcon GS. Epidemiology of Systemic Lupus Erythematosus. *Expert Rev Clin Immunol* (2017) 13(8):799–814. doi: 10.1080/1744666x.2017.1327352
  7. Kuhn A, Bonsmann G, Anders H-J, Herzer P, Tenbrock K, Schneider M. The Diagnosis and Treatment of Systemic Lupus Erythematosus. *Dtsch Arztebl Int* (2015) 112:423–32. doi: 10.3238/arztebl.2015.0423
  8. Fava A, Petri M. Systemic Lupus Erythematosus: Diagnosis and Clinical Management. *J Autoimmun* (2019) 96:1–13. doi: 10.1016/j.jaut.2018.11.001
  9. Fanouriakis A, Tziolos N, Bertisias G, Boumpas DT. Update Omicronn the Diagnosis and Management of Systemic Lupus Erythematosus. *Ann Rheum Dis* (2021) 80(1):14–25. doi: 10.1136/annrheumdis-2020-218272
  10. Qi S, Chen Q, Xu D, Xie N, Dai Y. Clinical Application of Protein Biomarkers in Lupus Erythematosus and Lupus Nephritis. *London England: SAGE Publications* (2018) . p:1582–90. doi: 10.1177/0961203318773643
  11. Capecchi R, Puxeddu I, Pratesi F, Migliorini P. New Biomarkers in SLE: From Bench to Bedside. *Rheumatology* (2020) 59:v12–v8. doi: 10.1093/rheumatology/keaa484
  12. Harden OC, Hammad SM. Sphingolipids and Diagnosis, Prognosis, and Organ Damage in Systemic Lupus Erythematosus. *Front Immunol* (2020) 11:586737. doi: 10.3389/fimmu.2020.586737
  13. Tunncliffe DJ, Singh-Grewal D, Kim S, Craig JC, Tong A. Diagnosis, Monitoring, and Treatment of Systemic Lupus Erythematosus: A Systematic Review of Clinical Practice Guidelines. *Arthritis Care Res (Hoboken)* (2015) 67(10):1440–52. doi: 10.1002/acr.22591
  14. Morel L. Immunometabolism in Systemic Lupus Erythematosus. *Nat Rev Rheumatol* (2017) 13(5):280–90. doi: 10.1038/nrrheum.2017.43
  15. Matta B, Barnes BJ. Coordination Between Innate Immune Cells, Type I Ifns and Irf5 Drives Sle Pathogenesis. *Cytokine* (2020) 132:154731. doi: 10.1016/j.cyt.2019.05.018
  16. Herrada AA, Escobedo N, Iruretagoyena M, Valenzuela RA, Burgos PI, Cuitino L, et al. Innate Immune Cells' Contribution to Systemic Lupus Erythematosus. *Front Immunol* (2019) 10:772. doi: 10.3389/fimmu.2019.00772
  17. Tsokos GC, Lo MS, Costa Reis P, Sullivan KE. New Insights Into the Immunopathogenesis of Systemic Lupus Erythematosus. *Nat Rev Rheumatol* (2016) 12(12):716–30. doi: 10.1038/nrrheum.2016.186
  18. Zhu Q, Li Y, Zhang L, Wang M, Chen Z, Shi J, et al. Patients With Systemic Lupus Erythematosus Show Increased Proportions of Cd19(+)Cd20(-) B Cells and Secretion of Related Autoantibodies. *Clin Rheumatol* (2021) 40(1):151–65. doi: 10.1007/s10067-020-05220-2
  19. Zecevic L, Mekic M, Subasic D, Hadziabulic M, Isak E, Subasic E, et al. Expression of Nk (Cd16+56+) and B Cells (Cd19) Receptor Molecules as a Reliable Clinical Response Biomarkers of Sle and Ra Patients Under the Rituximab Treatment. *Med Arch* (2019) 73(6):374–7. doi: 10.5455/medarch.2019.73.374-377
  20. Feng X, Wang D, Chen J, Lu L, Hua B, Li X, et al. Inhibition of Aberrant Circulating Tfh Cell Proportions by Corticosteroids in Patients With Systemic Lupus Erythematosus. *PLoS One* (2012) 7(12):e51982. doi: 10.1371/journal.pone.0051982
  21. Henning S, Lambers WM, Doornbos-van der Meer B, Abdulhad WH, Kroese FGM, Bootsma H, et al. Proportions of B-Cell Subsets Are Altered in Incomplete Systemic Lupus Erythematosus and Correlate With Interferon Score and Igg Levels. *Rheumatol (Oxford)* (2020) 59(9):2616–24. doi: 10.1093/rheumatology/keaa114
  22. Zhao E, Xie H, Zhang Y. Predicting Diagnostic Gene Biomarkers Associated With Immune Infiltration in Patients With Acute Myocardial Infarction. *Front Cardiovasc Med* (2020) 7:586871. doi: 10.3389/fcvm.2020.586871
  23. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust Enumeration of Cell Subsets From Tissue Expression Profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337
  24. Mok CC, Yap DY, Navarra SV, Liu ZH, Zhao MH, Lu L, et al. Overview of Lupus Nephritis Management Guidelines and Perspective From Asia. *Int J Rheum Dis* (2013) 16(6):625–36. doi: 10.1111/1756-185x.12212
  25. Burbano C, Vasquez G, Rojas M. Modulatory Effects of Cd14+Cd16++ Monocytes on Cd14++Cd16- Monocytes: A Possible Explanation of Monocyte Alterations in Systemic Lupus Erythematosus. *Arthritis Rheumatol* (2014) 66(12):3371–81. doi: 10.1002/art.38860
  26. Tay SH, Celhar T, Fairhurst AM. Low-Density Neutrophils in Systemic Lupus Erythematosus. *Arthritis Rheumatol* (2020) 72(10):1587–95. doi: 10.1002/art.41395
  27. Zhao X, Zhang L, Wang J, Zhang M, Song Z, Ni B, et al. Identification of Key Biomarkers and Immune Infiltration in Systemic Lupus Erythematosus by Integrated Bioinformatics Analysis. *J Transl Med* (2021) 19(1):35. doi: 10.1186/s12967-020-02698-x
  28. Lambers WM, Westra J, Bootsma H, de Leeuw K. From Incomplete to Complete Systemic Lupus Erythematosus; a Review of the Predictive Serological Immune Markers. *Semin Arthritis Rheumatism* (2021) 51(1):43–8. doi: 10.1016/j.semarthrit.2020.11.006
  29. Weng HJ, Tsai TF. Abcb1 in Dermatology: Roles in Skin Diseases and Their Treatment. *J Mol Med (Berl)* (2021) 99(11):1527–38. doi: 10.1007/s00109-021-02105-y
  30. Sita G, Hrelia P, Tarozzi A, Morroni F. P-Glycoprotein (Abcb1) and Oxidative Stress: Focus on Alzheimer's Disease. *Oxid Med Cell Longev* (2017) 2017:7905486. doi: 10.1155/2017/7905486
  31. Zhou HH, Chen X, Cai LY, Nan XW, Chen JH, Chen XX, et al. Erastin Reverses Abcb1-Mediated Docetaxel Resistance in Ovarian Cancer. *Front Oncol* (2019) 9:1398. doi: 10.3389/fonc.2019.01398
  32. Luo X, Teng QX, Dong JY, Yang DH, Wang M, Dessie W, et al. Antimicrobial Peptide Reverses Abcb1-Mediated Chemotherapeutic Drug Resistance. *Front Pharmacol* (2020) 11:1208. doi: 10.3389/fphar.2020.01208
  33. Liu X. Abc Family Transporters. *Adv Exp Med Biol* (2019) 1141:13–100. doi: 10.1007/978-981-13-7647-4\_2
  34. Paparisto E, Woods MW, Coleman MD, Moghadasi SA, Kochar DS, Tom SK, et al. Evolution-Guided Structural and Functional Analyses of the Herc Family Reveal an Ancient Marine Origin and Determinants of Antiviral Activity. *J Virol* (2018) 92(13):e00528-18. doi: 10.1128/jvi.00528-18
  35. Marfell BJ, O'Brien R, Griffin JF. Global Gene Expression Profiling of Monocyte-Derived Macrophages From Red Deer (*Cervus Elaphus*) Genotypically Resistant or Susceptible to *Mycobacterium Avium* Subspecies Paratuberculosis Infection. *Dev Comp Immunol* (2013) 40(2):210–7. doi: 10.1016/j.dci.2013.02.004
  36. Wu L, Qin Y, Xia S, Dai M, Han X, Wu Y, et al. Identification of Cyclin-Dependent Kinase 1 as a Novel Regulator of Type I Interferon Signaling in Systemic Lupus Erythematosus. *Arthritis Rheumatol* (2016) 68(5):1222–32. doi: 10.1002/art.39543
  37. Perry HM, Oldham SN, Fahl SP, Que X, Gonen A, Harmon DB, et al. Helix-Loop-Helix Factor Inhibitor of Differentiation 3 Regulates Interleukin-5 Expression and B-1a B Cell Proliferation. *Arteriosclerosis Thrombosis Vasc Biol* (2013) 33(12):2771–9. doi: 10.1161/atvbaha.113.302571
  38. Han J, Ma Y, Ma L, Tan D, Niu H, Bai C, et al. Id3 and Bcl6 Promote the Development of Long-Term Immune Memory Induced by Tuberculosis Subunit Vaccine. *Vaccines* (2021) 9(2):126. doi: 10.3390/vaccines9020126
  39. Ge L, Zhang Y, Zhao X, Wang J, Zhang Y, Wang Q, et al. Eif2ak2 Selectively Regulates the Gene Transcription in Immune Response and Histones Associated With Systemic Lupus Erythematosus. *Mol Immunol* (2021) 132:132–41. doi: 10.1016/j.molimm.2021.01.030
  40. Davidson S, Yu CH, Steiner A, Ebstein F, Baker PJ, Jarur-Chamy V, et al. Protein Kinase R Is an Innate Immune Sensor of Proteotoxic Stress Via Accumulation of Cytoplasmic Il-24. *Sci Immunol* (2022) 7(68):eabi6763. doi: 10.1126/sciimmunol.abi6763
  41. Toukap AN, Galant C, Theate I, Maudoux AL, Lories RJU, Houssiau FA, et al. Identification of Distinct Gene Expression Profiles in the Synovium of Patients With Systemic Lupus Erythematosus. *Arthritis Rheumatism* (2007) 56(5):1579–88. doi: 10.1002/art.22578

42. Gao J, Zhu X, Wu M, Jiang L, Wang F, He S. Ifi27 May Predict and Evaluate the Severity of Respiratory Syncytial Virus Infection in Preterm Infants. *Hereditas* (2021) 158(1):3. doi: 10.1186/s41065-020-00167-5
43. Bing PF, Xia W, Wang L, Zhang YH, Lei SF, Deng FY. Common Marker Genes Identified From Various Sample Types for Systemic Lupus Erythematosus. *PLoS One* (2016) 11(6):e0156234. doi: 10.1371/journal.pone.0156234
44. Huang H, Lv J, Huang Y, Mo Z, Xu H, Huang Y, et al. Ifi27 Is a Potential Therapeutic Target for Hiv Infection. *Ann Med* (2022) 54(1):314–25. doi: 10.1080/07853890.2021.1995624
45. Nagasawa Y, Okuzaki D, Muso E, Yamamoto R, Shinzawa M, Iwasaki Y, et al. Ifi27 Is a Useful Genetic Marker for Diagnosis of Immunoglobulin a Nephropathy and Membranous Nephropathy Using Peripheral Blood. *PLoS One* (2016) 11(4):e0153252. doi: 10.1371/journal.pone.0153252
46. Bernales I, Fullaondo A, Marín-Vidal MJ, Ucar E, Martínez-Taboada V, López-Hoyos M, et al. Innate Immune Response Gene Expression Profiles Characterize Primary Antiphospholipid Syndrome. *Genes Immun* (2008) 9(1):38–46. doi: 10.1038/sj.gene.6364443
47. O'Hanlon TP, Rider LG, Gan L, Fannin R, Paules RS, Umbach DM, et al. Gene Expression Profiles From Discordant Monozygotic Twins Suggest That Molecular Pathways Are Shared Among Multiple Systemic Autoimmune Diseases. *Arthritis Res Ther* (2011) 13(2):R69. doi: 10.1186/ar3330
48. Deng Y, Zheng Y, Li D, Hong Q, Zhang M, Li Q, et al. Expression Characteristics of Interferon-Stimulated Genes and Possible Regulatory Mechanisms in Lupus Patients Using Transcriptomics Analyses. *EBioMedicine* (2021) 70:103477. doi: 10.1016/j.ebiom.2021.103477

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhong, Zhang, Hong, Zeng, Chen, Liao, Cai, Xu, Wang, Liu, Tang and Dai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.