



OPEN ACCESS

EDITED BY

Martina Cozzi,
San Giovanni Bosco Hospital, Italy

REVIEWED BY

Johannes Fessler,
Medical University of Graz, Austria
Georgia Damoraki,
National and Kapodistrian University of
Athens, Greece

*CORRESPONDENCE

Xiaoqing Chen
✉ chenxiaoqing202203@163.com

RECEIVED 08 January 2025

ACCEPTED 23 April 2025

PUBLISHED 14 May 2025

CITATION

Zhang S, Hu W, Tang Y, Lin H and
Chen X (2025) Identification of hub
immune-related genes and construction of
predictive models for systemic lupus
erythematosus by bioinformatics combined
with machine learning.
Front. Med. 12:1557307.
doi: 10.3389/fmed.2025.1557307

COPYRIGHT

© 2025 Zhang, Hu, Tang, Lin and Chen. 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.

Identification of hub immune-related genes and construction of predictive models for systemic lupus erythematosus by bioinformatics combined with machine learning

Su Zhang¹, Weitao Hu², Yuchao Tang², Hongjie Lin² and
Xiaoqing Chen^{1*}

¹Department of Rheumatology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China, ²Department of Gastroenterology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease that involves multiple systems. SLE is characterized by the production of autoantibodies and inflammatory tissue damage. This study further explored the role of immune-related genes in SLE. We downloaded the expression profiles of GSE50772 using the Gene Expression Omnibus (GEO) database for differentially expressed genes (DEGs) in SLE. The DEGs were also analyzed for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment. The gene modules most closely associated with SLE were then derived by Weighted Gene Co-expression Network Analysis (WGCNA). Differentially expressed immune-related genes (DE-IRGs) in SLE were obtained by DEGs, key gene modules and IRGs. The protein-protein interaction (PPI) network was constructed through the STRING database. Three machine learning algorithms were applied to DE-IRGs to screen for hub DE-IRGs. Then, we constructed a diagnostic model. The model was validated by external cohort GSE61635 and peripheral blood mononuclear cells (PBMC) from SLE patients. Immune cell abundance assessment was achieved by CIBERSORT. The hub DE-IRGs and miRNA networks were made accessible through the NetworkAnalyst database. We screened 945 DEGs, which are closely related to the type I interferon pathway and NOD-like receptor signaling pathway. Machine learning identified a total of five hub DE-IRGs (*CXCL2*, *CXCL8*, *FOS*, *NFKBIA*, *CXCR2*), and validated in GSE61635 and PBMC from SLE patients. Immune cell abundance analysis showed that the hub genes may be involved in the development of SLE by regulating immune cells (especially neutrophils). In this study, we identified five hub DE-IRGs in SLE and constructed an effective predictive model. These hub genes are closely associated with immune cell in SLE. These may provide new insights into the immune-related pathogenesis of SLE.

KEYWORDS

bioinformatics, hub genes, immune cell, machine learning, systemic lupus erythematosus

1 Introduction

Systemic lupus erythematosus (SLE) is an immune-mediated, complex, chronic systemic disease (1). Approximately 400,000 new cases of SLE are diagnosed globally each year, and it predominantly affects young women (2). Due to its complex pathogenesis and multi-organ involvement, SLE affects patients' quality of life and can endanger their lives, and may lead to psychological problems such as anxiety and depression (3). Genetics, hormones, and viral infections are all thought to contribute to the pathogenesis of SLE, but these factors ultimately result in immune dysregulation, which produces autoantibodies that lead to tissue damage (4). However, the pathogenesis of SLE is intricate and has not been thoroughly investigated. At this stage, SLE relies on drugs such as hydroxychloroquine and steroids to regulate immune function. Nevertheless, the toxic side effects of these drugs, such as infections, osteoporosis, and cardiovascular risks, should not be ignored (5, 6). Although the use of biologics offers hope for patients with refractory lupus, they are expensive for long-term use and new treatments are urgently needed.

The emergence of bioinformatics provides an effective way for people to process and analyze large datasets. It is capable of parsing data such as genomes and transcriptomes to identify specific biomarkers associated with certain diseases, thus aiding in early diagnosis and risk assessment (7). In recent years machine learning has become an increasingly promising tool for solving complex problems in the biomedical field. When combined with bioinformatics facilitates improved accuracy and reliability in exploring diseases (8).

In this study, a comprehensive bioinformatics analysis incorporating machine learning algorithms was performed to identify hub IRGs and pathways in SLE using the GEO and Immport databases. We then constructed a predictive model for SLE and validated the expression of the hub IRGs and the accuracy of the model using external datasets and RT-qPCR. Subsequently, we investigated a Pearson correlation analysis between hub genes and immune cells. Finally, we identified key miRNA molecules that interact with the hub genes. In summary, the study revealed hub IRGs in SLE, which will help to further elucidate the contribution of immune factors in SLE development and thus provide clues for exploring the complex etiology of SLE.

2 Materials and methods

2.1 Data collecting

The GEO database¹ (9) was searched for the keyword “systemic lupus erythematosus” to obtain the SLE-related dataset GSE50772 and GSE61635. GSE50772 was used as the training cohort, while GSE61635 served as the validation group. Both datasets are based on the GPL570 platform. The GSE50772 contains peripheral blood samples from 61 SLE patients and 20 normal controls (NC), while GSE61635 contains 109 blood samples from SLE and NC. In addition, we acquired datasets of primary Sjögren's Syndrome (pSS, GSE84844) and rheumatoid arthritis (RA, GSE17755). GSE84844 and GSE17755 were used for subsequent assessment of the diagnostic value of the hub genes for pSS and RA. Table 1 provides details of all the datasets.

2.2 Identification of DEGs and enrichment analysis

GSE50772 was normalized and filtered for DEGs using the “limma” package. The selection criteria for DEGs were set to $|\log_2 \text{FoldChange}| > 0.5$, corrected $p < 0.05$. DEGs were displayed using volcano and heatmaps. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) function analysis of DEGs was conducted with “clusterprofile” package to understand the biological processes and signaling pathways in which they are involved. A corrected $p < 0.05$ was considered to be statistical significance.

2.3 Construction of weighted gene co-expression network

To screen key gene modules from different modules that affect the SLE phenotype, we constructed a co-expression network using the “WGCNA” package of R (10). The best soft threshold was first established by pickSoftThreshold function. Then, the module merging threshold was set to 0.25 to obtain co-expression modules. Every module contains a minimum of 20 genes and non-significant genes were grouped into gray module. Finally, the correlation between every gene module and phenotype was computed. The correlation between gene modules and SLE patients was also assessed by the values of gene significance (GS) and module membership (MM).

2.4 Acquisition of common genes (CGs) and construction of PPI networks

The common genes (CGs) of DEGs and key gene modules were obtained by Venn diagram. The STRING database² (11) is commonly utilized to construct PPI networks. The minimum required reciprocal score was 0.4. Subsequent visualization was performed with Cytoscape software (version 3.9.1) (12). In addition, the PPI network nodes were scored utilizing Cytoscape's molecular complex detection (MCODE) plugin to filter out the most important modules and genes. The setup parameters for the MCODE plugin in this study were MCODE score > 5 , degree criticality = 2, node score criticality = 0.2, maximum depth = 100, k -score = 2.

2.5 Identification of DE-IRGs in SLE

There were 1793 IRGs were acquired from the ImmPort database³ (13). The Venn diagram showed that overlapping genes of IRGs and CGs are the DE-IRGs in SLE.

1 <https://www.ncbi.nlm.nih.gov/geo/>

2 <https://cn.string-db.org/>

3 <https://www.immport.org/shared>

TABLE 1 Details of the datasets included in this study.

Dataset	Platform	Species	Tissue	Number of cases and controls	Type of cohorts
GSE50772	GPL570	<i>Homo sapiens</i>	Peripheral blood	61 SLE/20NC	Training
GSE61635	GPL570			79 SLE/30NC	Validating
GSE84844	GPL570			30 pSS/30NC	Validating
GSE17755	GPL1291			112RA/53 NC	Validating

2.6 Screening of hub DE-IRGs in SLE

The least absolute shrinkage and selection operator (LASSO) regression is usually applied to select features for high-dimensional data, especially in gene expression data analysis (14). The basic principle is to perform variable selection by introducing L1 regularization terms, so as to efficiently screen out the important genes related to the target variables (15). Random forest (RF) is a machine learning method based on integrated learning, widely used in classification and regression problems, and can also be used to screen feature genes. In genomics and bioinformatics, random forests can help select the most relevant gene features to the target variable by assessing the importance of each gene to the prediction model (16). Support vector machine-recursive feature elimination (SVM-RFE) is a machine learning method commonly applied to screen signature genes. It is based on the principle of maximum interval of Support vector machine (SVM), through the model training samples, each feature score sorting, and then use the recursive feature elimination (RFE) algorithm step-by-step iterative way: remove the features with the smallest feature scores, and then use the remaining features to train the model again for the next iteration. The remaining characteristics are then utilized to train the model again for the next iteration, and finally the best combination of features is selected (17). In this study, we screened signature IRGs from 22 DE-IRGs using the three machine learning methods described above. The upset R diagram was subsequently utilized to obtain the intersecting genes of the three methods as the hub DE-IRGs of SLE.

2.7 Construction and validation of model

The accuracy of hub DE-IRGs selected by the machine learning methods was validated in another external SLE dataset. Subsequently, a model based on hub genes was constructed with an area under the curve (AUC) was greater than 0.8, indicating that the model has strong diagnostic value. Furthermore, we also assessed the diagnostic worth of the hub genes for pSS and RA by ROC curves.

2.8 Acquisition of peripheral blood mononuclear cells (PBMCs)

Peripheral blood was collected from 30 patients diagnosed with SLE from June 2024 to December 2024 in the Department of Rheumatology, the Second Affiliated Hospital of Fujian

Medical University. Peripheral blood was also collected from 22 normal controls who excluded hepatitis B, diabetes mellitus, pathogenic infection, malignant tumor and other types of autoimmune diseases, such as RA and pSS. The diagnosis of SLE was based on the European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) 2019 criteria. Furthermore, SLE disease activity was evaluated based on the SLE disease activity index 2000 (SLEDAI-2K) (18). We also collected gender, age, and relevant clinical and laboratory indicators for all participants (Table 2). Erythrocyte lysate (C3702) and lymphocyte isolate (C0025) were purchased from Beyotime (Shanghai, China). Mononuclear cells from peripheral blood were isolated according to the appropriate instructions (19).

TABLE 2 Clinical traits of SLE patients and normal controls.

Clinical traits*	SLE (n = 30)	Normal controls (n = 22)
Sex, male/female	2/28	2/20
Age (year)	35.77 ± 15.59	33.25 ± 9.99
Duration (year)	8.63 ± 7.69	
LN	12 (30)	
SLEDAI scores	11.5 ± 3.82	
ANA (positive)	29 (30)	
Anti-dsDNA antibody (positive)	26 (30)	
Lupus anticoagulant (positive)	12 (30)	
Leukocyte (10 ⁹ /L)	5.63 ± 1.79	
Platelets (10 ⁹ /L)	180.65 ± 94.51	
CRP (mg/L)	10.05 ± 13.24	
ESR (mm/h)	20.2 ± 15.69	
C3 (g/L)	0.73 ± 0.45	
C4 (g/L)	0.1 ± 0.05	
IgG (g/L)	15.14 ± 8.05	
IgA (g/L)	2.67 ± 1.8	
IgM (g/L)	0.8 ± 0.63	
Serum creatinine (μmol/L)	72.81 ± 58.24	
24 h urine protein (positive)	12 (30)	

*LN, lupus nephritis; SLEDAI, systemic lupus erythematosus disease activity index; ANA, antinuclear antibody; Anti-dsDNA antibody, anti-double stranded deoxyribonucleic acid antibody; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; C3/C4, complement 3/complement 4; Igs, immunoglobulins.

TABLE 3 The primers used in this study.

Gene names	Primers sequences (5' → 3')
B-actin-F	CATGTACGTTGCTATCCAGGC
B-actin-R	CTCCTTAATGTCACGCACGAT
CXCL2-F	GGCAGAAAGCTTGCTCAACCC
CXCL2-R	CTCCTTCAGGAACAGCCACCAA
CXCR2-F	TCCGTCACCTGATGTCTACCTGC
CXCR2-R	TCCTTCAGGAGTGAGACCACCT
CXCL8-F	GAGAGTGATTGAGAGTGGACCAC
CXCL8-R	CACAACCTCTGCACCCAGTTT
FOS-F	GCCTCTCTTACTACCACTCACC
FOS-R	AGATGGCAGTGACCGTGGGAAT
NFKBIA-F	TCCACTCCATCCTGAAGGCTAC
NFKBIA-R	CAAGGACACCAAAAGCTCCACG

2.9 RT-qPCR to validate hub genes expression

The RNA extraction kit was purchased from BioTeke (Beijing, China). See reference (19) for specific methodology. Reverse transcription reagents were purchased from Takara (Japan). The cDNA synthetic reaction was run at 37°C for 15 min and then heated at 85°C for 3 min to terminate. The cDNA was subsequently kept in liquid nitrogen. Finally, ABI PRISM 7500 PCR instrument (Applied Biosystems, United States) was used to amplify the target genes. The PCR cycle was performed as follows: 95°C for 15 min, 40 cycles of 95°C for 5 s, and 60°C for 30 s. *B-actin* was used as housekeeping gene to normalize target gene data. The primer sequences are shown in Table 3.

2.10 Evaluation of immune cell abundance

Since SLE is a classical autoimmune disease, immune cells play an important part in its development. CIBERSORT⁴ utilizes a gene expression matrix from a sample compared to a known set of genes that characterize the cell type using an inverse convolution algorithm to infer the relative level of each type of cell in the sample (20). We obtained the composition of 22 immune cells in SLE by Cibersort. We then compared the difference in the distribution of immune cells between the SLE and NC groups. Subsequent Pearson correlation analysis between hub genes and immune cells was calculated. For the above analysis, $p < 0.05$ represents statistical significance.

2.11 Construction of gene-miRNA networks

We uploaded hub genes to NetworkAnalyst website⁵ (21) to get miRNAs closely related to hub genes and constructed their interaction networks.

4 <https://cibersortx.stanford.edu/>

5 <https://www.networkanalyst.ca/NetworkAnalyst/>

2.12 Statistical analysis

The R software (version 4.4.2) was employed for all analyses. Pearson analysis was applied to investigate the correlation between hub genes and immune cells, and $p < 0.05$ was considered to be statistical significance. The specific flow chart of the study is summarized in Figure 1.

3 Results

3.1 Acquisition and enrichment analysis of DEGs

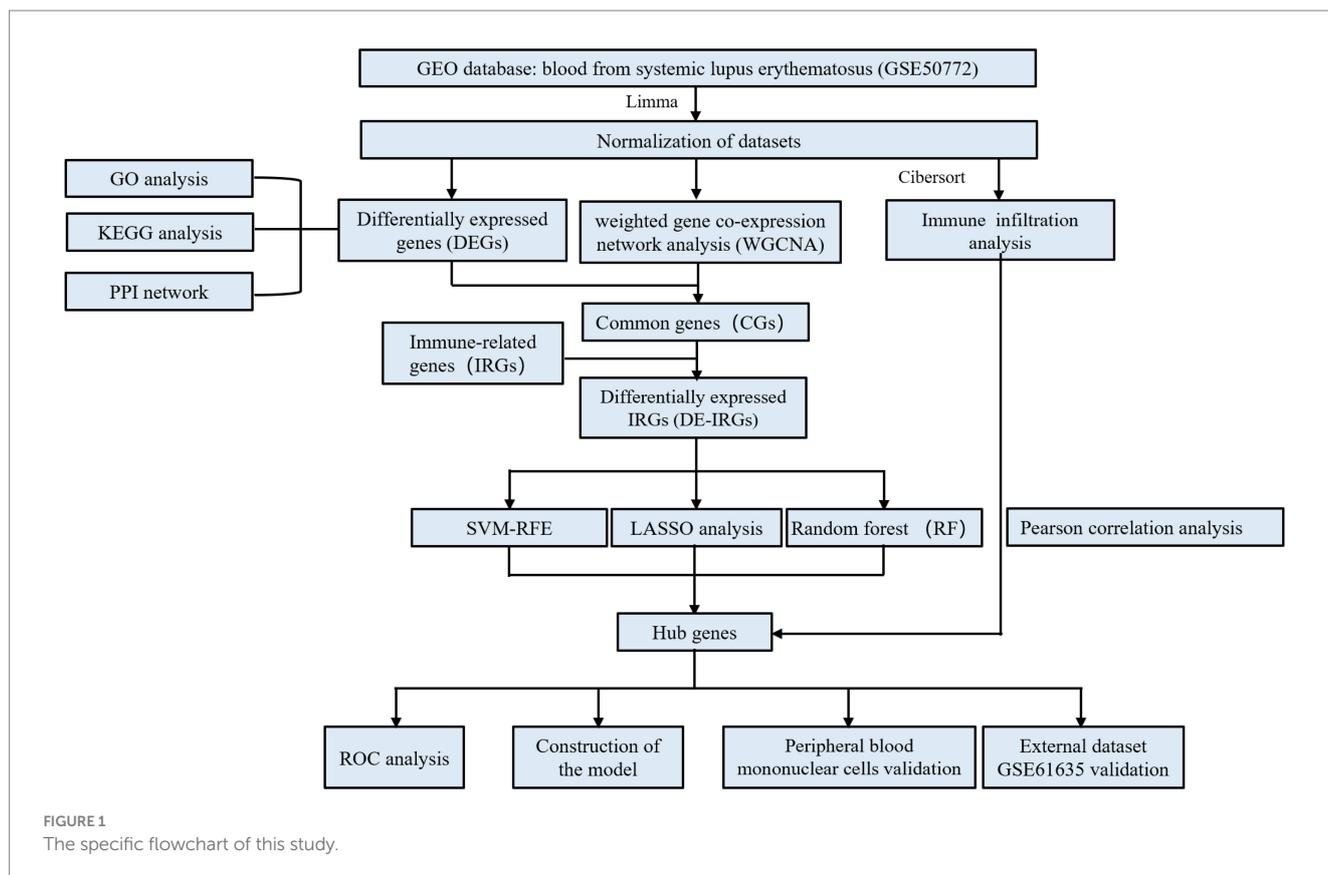
The median gene expression of single samples remained consistent after normalization to the training cohort, indicating that potential batch effects were rectified (Supplementary Figure 1). Based on the above selection criteria, we obtained 945 DEGs from GSE50772 (Figure 2A). The heatmap showed that they were expressed differently in NC and SLE groups (Figure 2B). The specific names of the DEGs were given in the Supplementary file. GO analysis showed that DEGs were mainly closely related to the type I interferon (Figure 2C). While KEGG enrichment showed that DEGs were primarily engaged in NOD-like receptor signaling pathway and TNF signaling pathway (Figure 2D). We chose the NOD-like receptor signaling pathway to demonstrate the distribution of DEGs in it (Supplementary Figure 2).

3.2 Identification of key module genes

Weighted Gene Co-expression Network Analysis showed that the mean connectivity is 0.9 when the soft threshold β is 5 (Figure 3A). A total of 17 gene modules were recognized (Figure 3B). We chose modules with a disease correlation coefficient greater than 0.7 as key modules. Green yellow and pink modules were found to fulfill our requirements and they satisfied $p < 0.05$ (Figures 3C–E). We take the intersection of the two modules' genes and DEGs to get their common genes (CGs). A total of 175 CGs were obtained (Figure 3F). We then uploaded the CGs to the STRING database and visualized the PPI network using Cytoscape. We got a PPI network consisting of 54 points and 247 edges (Supplementary Figure 3A). The most critical module is composed of 20 points and 147 edges (Supplementary Figure 3B). This suggested that CGs work together in the same biological process.

3.3 Identification of the hub DE-IRGs in SLE

The Venn plot identified 22 DE-IRGs in SLE (Figure 4A). Three machine learning methods was applied to screen signature genes. LASSO regression screened eight signature genes (Figures 4B,C). Random Forest ranked the 22 DE-IRGs for importance to get the top 10 genes with the highest scores (Figure 4D). A total of 22 signature genes were obtained from SVM-RFE (Figure 4E). The intersection of the signature genes



obtained from the three machine learning algorithms is taken to acquire the final five hub genes (*CXCL2*, *CXCL8*, *FOS*, *NFKBIA*, *CXCR2*) (Figure 4F). The hub genes are positively correlated with each other (Figure 4G). This implied that hub genes are synergistic in some functions.

3.4 Construction and validation of models

Interestingly, all of the hub genes are upregulated genes in the SLE training cohort ($p < 0.05$) (Figure 5A). We subsequently constructed a nomogram of SLE (Figure 5B). The AUC for ROC was found to be greater than 0.8, indicating that performed well in diagnosing SLE (Figure 5C). We then verified in the external cohort that the expression of hub genes was consistent with the training cohort (Figure 5D). And the accuracy of the model was verified again, and it was found that the AUCs were all greater than 0.8, which more strongly supported our results (Figure 5E). Further RT-qPCR results indicated that the expression of hub genes in the SLE group was obviously higher than that in the NC group (Figure 5F). Then, we detected that the expression of *CXCL8* in the pSS was significantly higher than that in the NC group, while *NFKBIA* was significantly lower than that in the NC group (Supplementary Figure 4A). Whereas in RA, the expression of *CXCL8*, *CXCR2* and *NFKBIA* was markedly higher than that in the NC group (Supplementary Figure 4B). Surprisingly, although the hub genes have some diagnostic value for pSS and RA, their diagnostic efficacy is not as good as that of

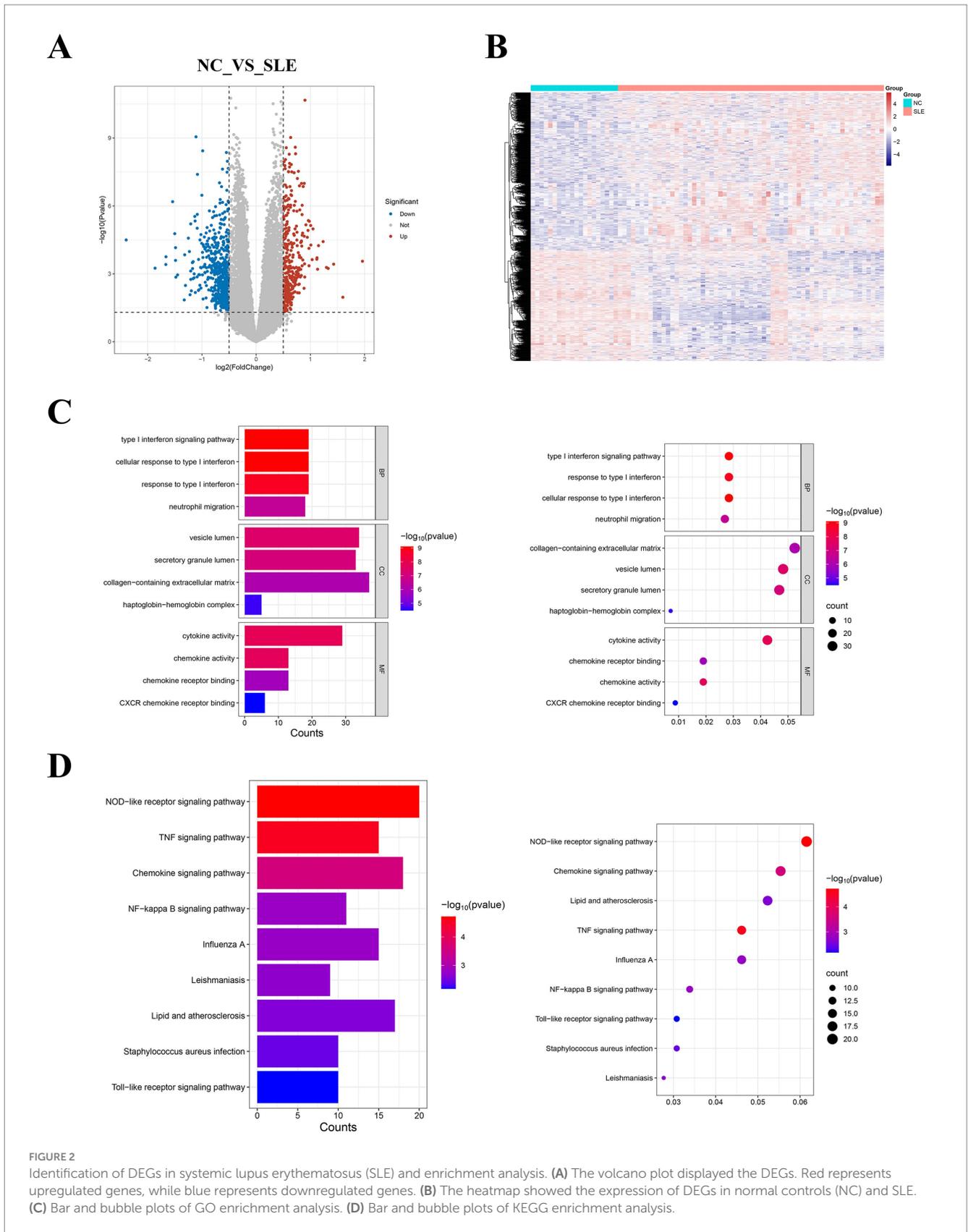
SLE (AUCs < 0.8) (Supplementary Figures 4C,D). This reinforces the specificity of the hub genes in the diagnosis of SLE.

3.5 Immune cell abundance analysis

Since SLE is a classical autoimmune disease, immune cells play an essential function in its pathogenesis. Our results showed that monocytes are the major immune component cells in SLE and NC groups (Figure 6A). The second is NK cells resting (Figure 6B). Meanwhile, our results revealed that T cells regulatory ($p < 0.05$), M2 macrophages ($p < 0.001$) and dendritic cells activated ($p < 0.001$), mast cells activated ($p < 0.01$) and neutrophils ($p < 0.0001$) were significantly higher in SLE (Figure 6C). In contrast, B cells naive ($p < 0.05$), T cells CD4 naive ($p < 0.05$), T cells CD4 memory resting ($p < 0.05$), NK cells resting ($p < 0.001$), mast cells resting ($p < 0.0001$) and eosinophils ($p < 0.05$) were significantly lower in SLE than in NC (Figure 6C). These results suggest that M2 macrophages and T cells are the major immune component cells in SLE patients. These cells may play an important role in the pathogenesis of SLE (22).

3.6 Correlation between the hub genes and immune cells

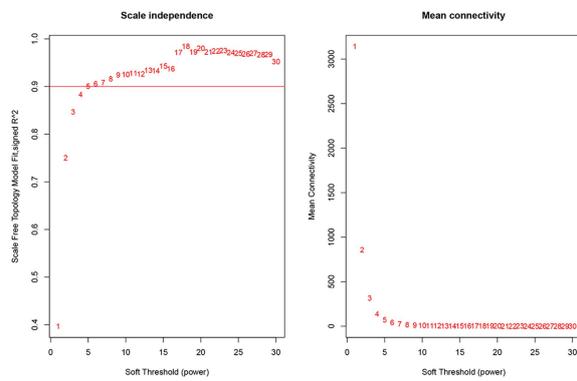
In order to further understand the relationship between hub genes and immune cells, we performed Pearson correlation analysis on them. The results of the analysis showed that all hub genes showed strong



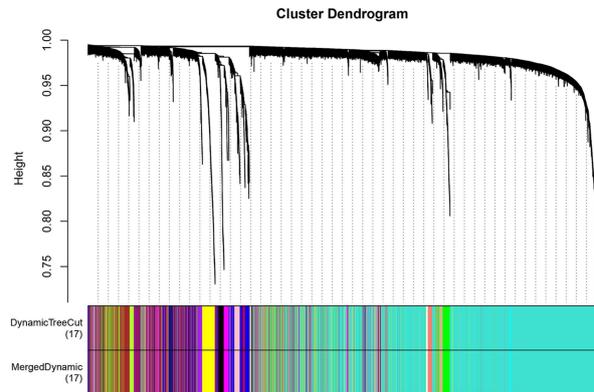
correlation with a variety of immune cells (Figure 7A). Specifically, *CXCL8* was positively correlated with neutrophils ($R = 0.64$) and negatively correlated with mast cells resting ($R = -0.59$) (Figure 7B).

FOS correlated with immune cells in the same way as *CXCL8*, also positively with neutrophils ($R = 0.56$) and negatively with mast cells resting ($R = -0.61$) (Figure 7C). *CXCL2* was positively related to

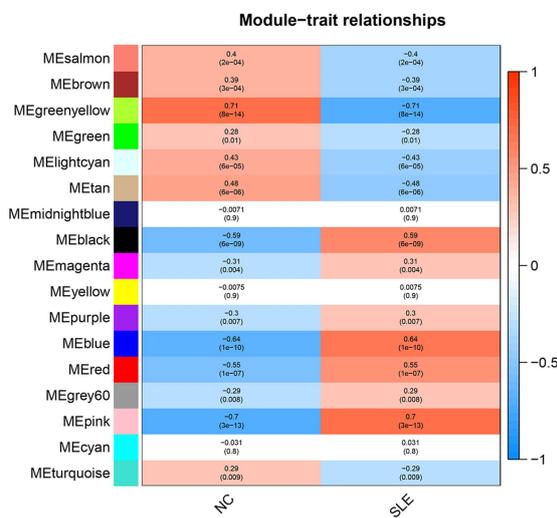
A



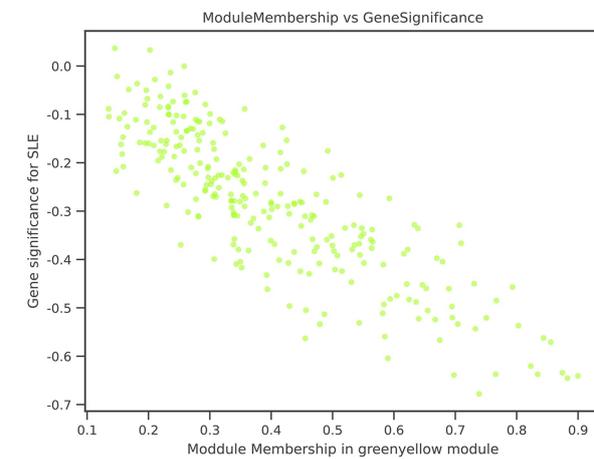
B



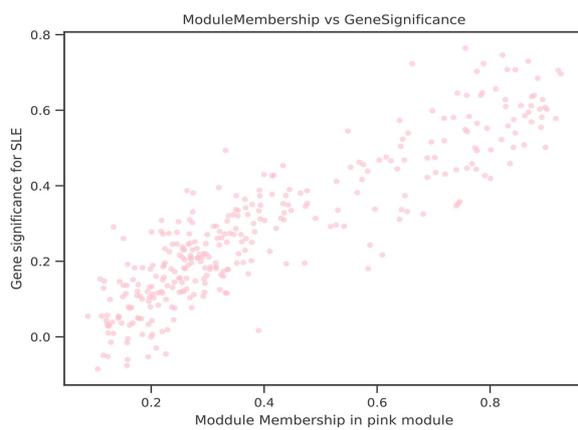
C



D



E



F

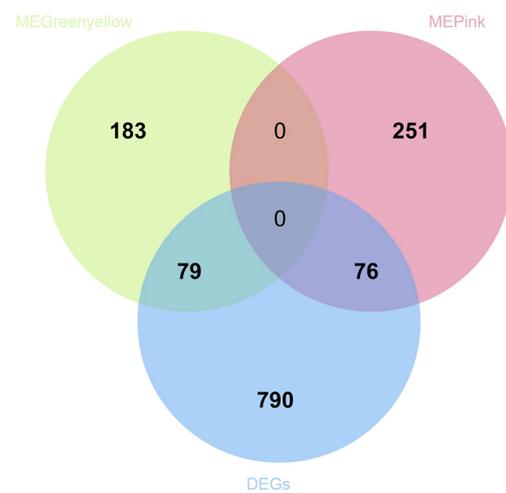
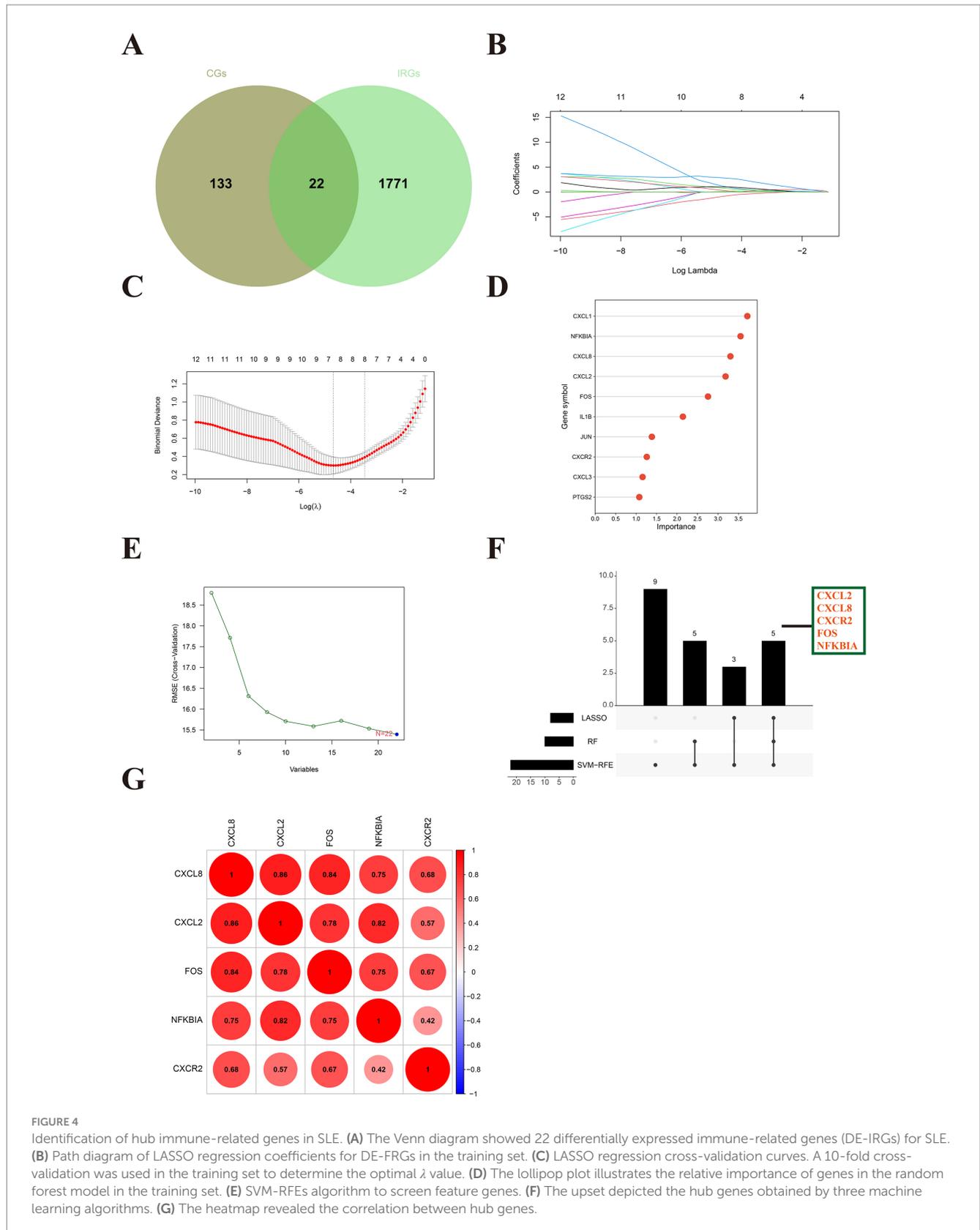


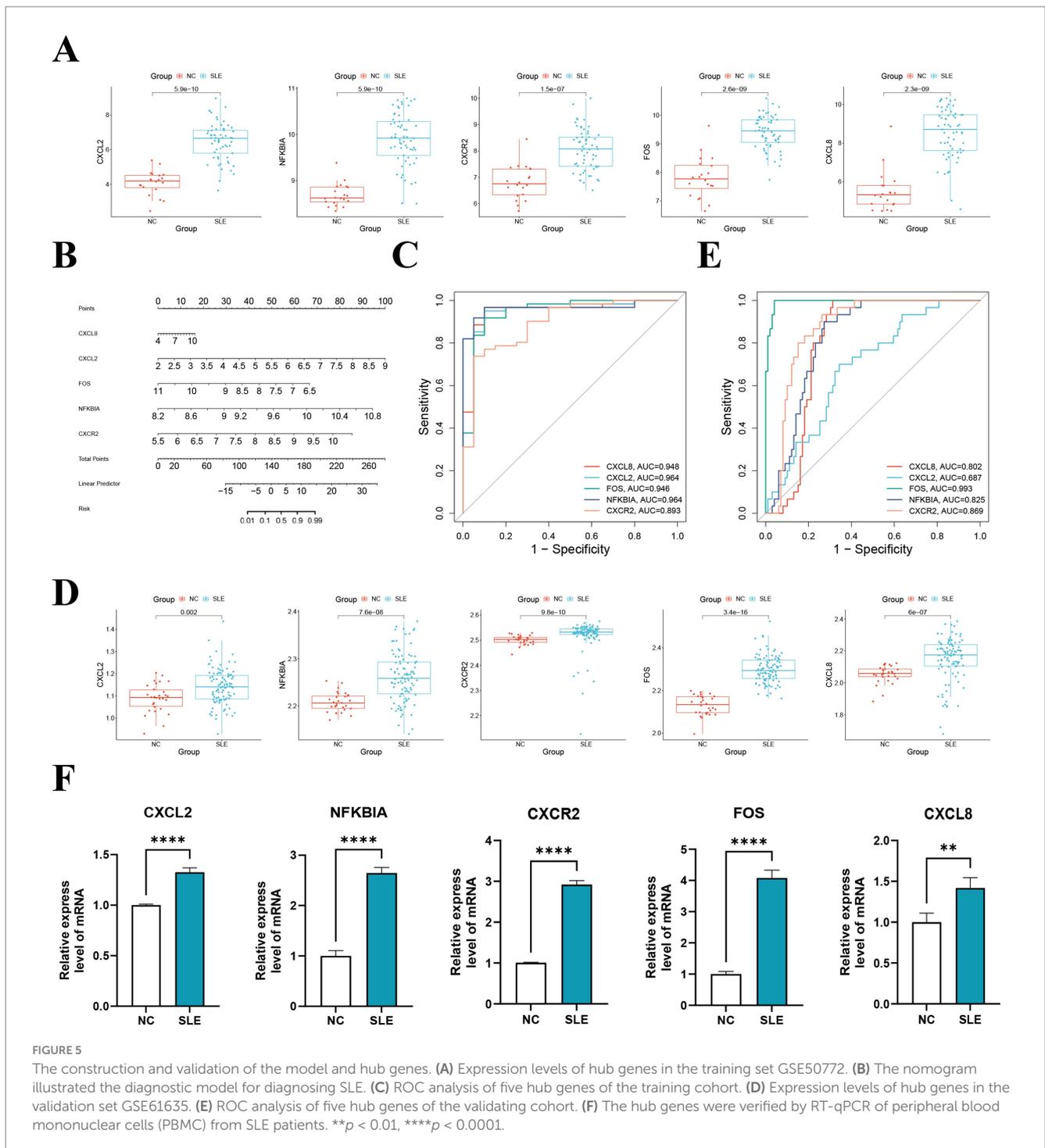
FIGURE 3

Weighted gene co-expression network analysis (WGCNA) for SLE. **(A)** The soft threshold and mean connectivity of the WGCNA network. **(B)** The clustering dendrogram of the WGCNA network. **(C)** The heatmap depicts the correlation of the different modules with clinical features, especially SLE and NC. **(D)** The scatter plot between the gene significance (GS) and module members (MM) in the green yellow module. **(E)** The scatter plot between the GS and MM in the pink module. **(F)** The Venn plot displayed the common genes (CGs) of yellow-green modular genes, pink modular genes, and DEGs.



multiple immune cells, which included M2 macrophages ($R = 0.47$), neutrophils ($R = 0.48$), and activated mast cells ($R = 0.5$), while it was negatively associated with mast cells resting ($R = -0.63$) (Figure 7D). *NFKBIA* showed a strong negative correlation with mast cells resting

($R = -0.73$) (Figure 7E). *CXCR2* possessed the strongest positive correlation with neutrophils ($R = 0.88$) and was negatively correlated with CD4⁺T cells naive ($R = -0.45$) (Figure 7F). These suggest that hub genes are strongly associated with immune cells in SLE.



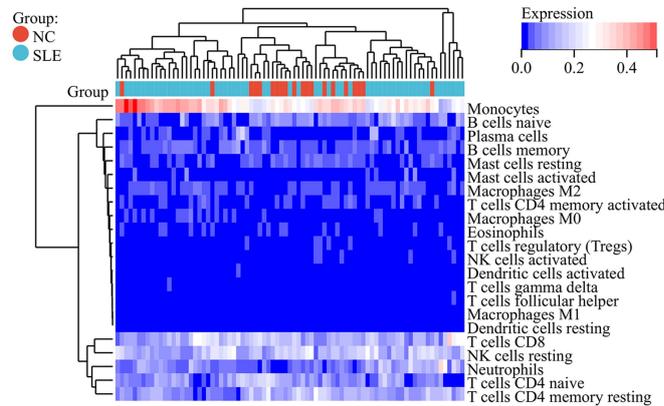
3.7 Construction of hub genes-miRNA network

Many studies have demonstrated that miRNAs perform their biological functions by participating in the regulation of their downstream gene translation. Therefore, we hope to find the key miRNAs that interact with these hub genes by constructing a hub genes-miRNAs network. The results showed that hsa-mir-335-5p is the molecule to which these hub genes are co-connected (Figure 8).

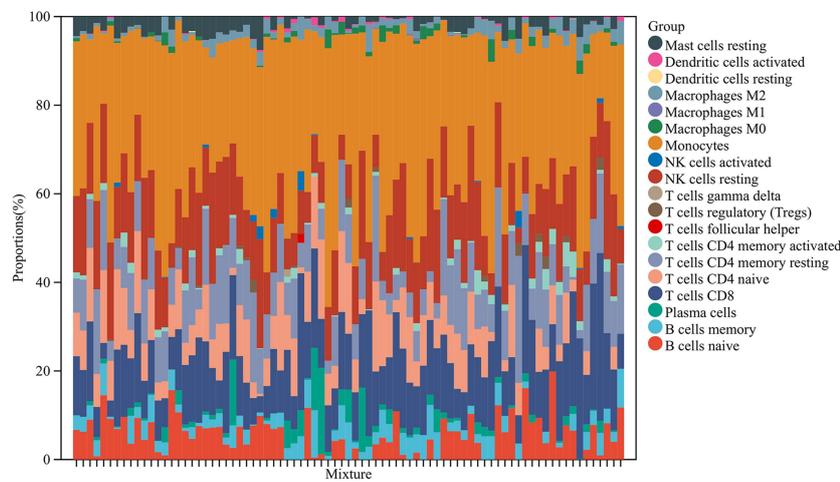
4 Discussion

Genome-wide studies have identified a number of susceptibility genes for SLE, but the IRGs for SLE remain largely unknown (23). In this study, we screened 945 DEGs for SLE. Further GO functional analysis of DEGs revealed that they were strongly associated with type I interferon and immune function. These results are consistent with previous studies (24–26). KEGG enrichment showed that DEGs were principally participated in NOD-like receptor signaling pathway and TNF signaling pathway. Inhibition of certain important molecules in

A



B



C

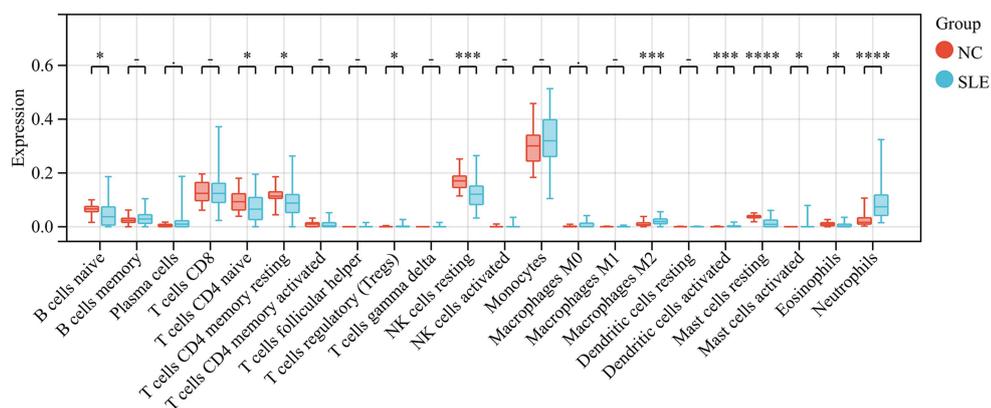


FIGURE 6 Analysis of immune cell abundance in SLE. **(A)** The heatmap showed the distribution of immune cells in SLE and NC. **(B)** Relative percentage of immune cell subpopulations in SLE and NC. **(C)** The box plot displayed the differences in the levels of immune cells in SLE and NC. * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$.

the NOD-like receptor signaling pathway and the TNF signaling pathway are effective strategies for the treatment of SLE (27–31). There are many studies on this type of research, which we will not elaborate here. In addition, WGCNA constructed modules that were closely related to SLE and selected the two modules with the strongest correlation to take the intersection with DEGs to get CGs. Then

we obtained DE-IRGs in SLE by CGs and IRGs. Three machine learning algorithms were selected for these DE-IRGs to obtain hub genes (*CXCL2*, *CXCL8*, *FOS*, *NFKBIA*, *CXCR2*). These hub genes showed high sensitivity and high specificity for the diagnosis of SLE ($AUC > 0.8$). We also validated the model by validation cohort and PBMC. Subsequently, we analyzed the large differences in immune

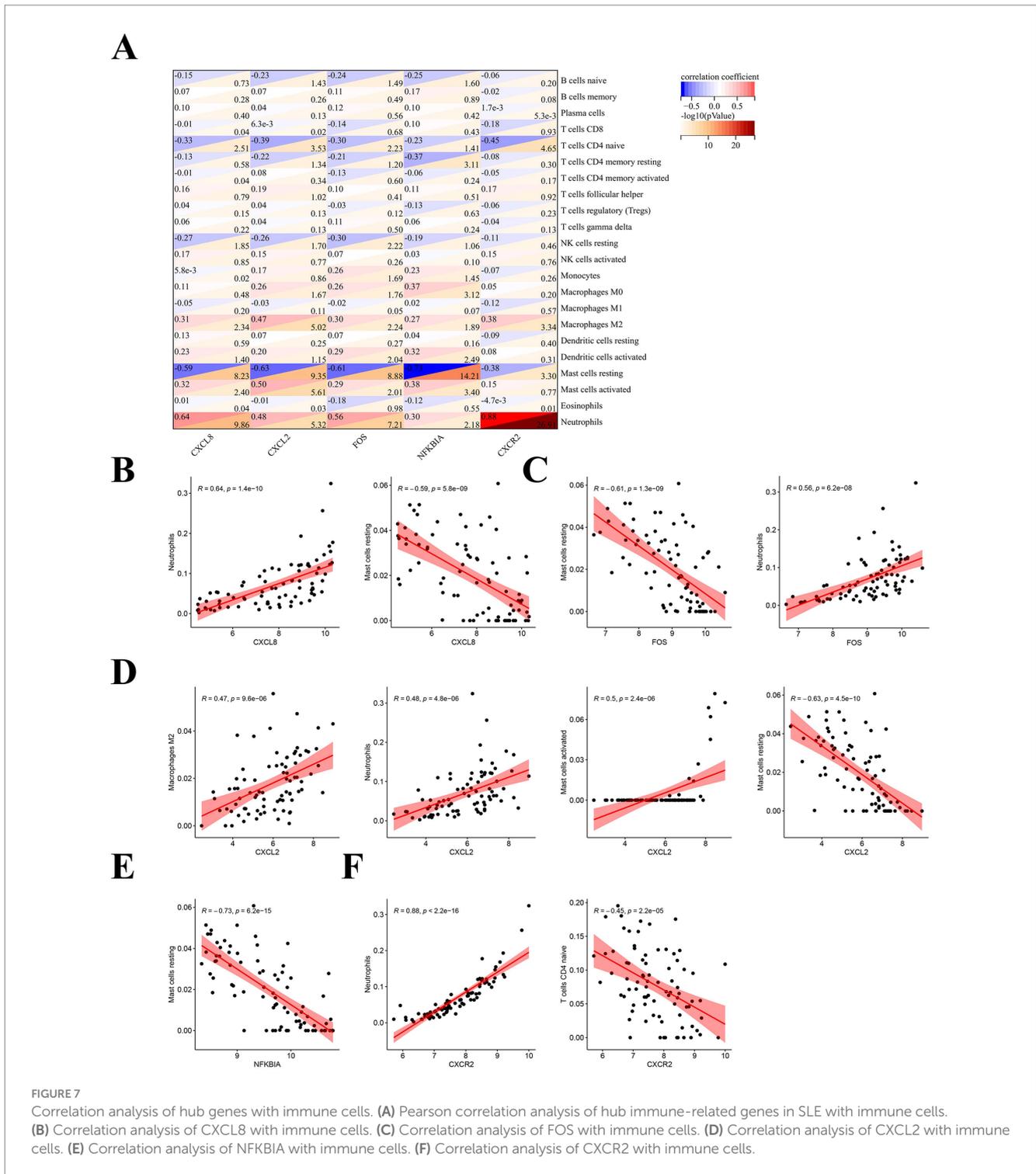


FIGURE 7
 Correlation analysis of hub genes with immune cells. **(A)** Pearson correlation analysis of hub immune-related genes in SLE with immune cells. **(B)** Correlation analysis of CXCL8 with immune cells. **(C)** Correlation analysis of FOS with immune cells. **(D)** Correlation analysis of CXCL2 with immune cells. **(E)** Correlation analysis of NFKBIA with immune cells. **(F)** Correlation analysis of CXCR2 with immune cells.

cells expression in SLE and NC as well as the possible influence of hub genes on the involvement of multiple immune cells in the pathogenesis of SLE. Therefore, we hypothesized that these five hub genes are important immune-related biomarkers for SLE. Finally, hsa-mir-335-5p was found to be tightly associated with the hub genes.

CXCL2 (C-X-C Motif Chemokine Ligand 2), CXCL8 (C-X-C Motif Chemokine Ligand 8) and CXCR2 (C-X-C Motif Chemokine Receptor 2) are all members of the chemokine family. CXCL2 and CXCR2 are a bunch of important chemokine ligands and receptors. CXCL2 – CXCR2

has been shown to play an important role in the development of a variety of tumors and is closely associated with neutrophil activation and migration (32–35). Neutrophils are the most abundant immune cells in the body and are inextricably linked to the development of SLE. Abnormal activation of neutrophils can exacerbate inflammatory responses and tissue damage. Neutrophils can also promote immune complex formation through the release of cytokines, the generation of neutrophil extracellular traps (NETs), and the production of oxidative stress, which can lead to exacerbation of the autoimmune response,

NFKB signaling pathway significantly reduced urinary protein and autoantibody levels in lupus mice, as well as reduced renal immune complex deposition. Therefore, inhibiting the over-activation of NFKB by enhancing the expression of NFKBIA may be an effective way to attenuate the inflammatory and immune responses in SLE.

Multiple studies have shown that miRNAs play an essential function in the development of SLE. For example, miR-590-3p ameliorated inflammation in lupus mice by inhibiting Th17 cell differentiation (53). miR-21 and miR-155 genetic variants were associated with susceptibility to SLE (54). Xu et al. (55) found that *IL-10* targeting E2F2-miR-17-5p inhibited autoantibody secretion in active SLE patients. The hsa-mir-335-5p is widely expressed in human and has been found to positively correlate with anti-CCP antibodies and C-reactive protein in rheumatoid arthritis (RA), and is a good biomarker for RA (56). It is also a valid marker for osteoarthritis (57). Inhibition of *FOS* expression by has-mir-335-3p regulates bone metabolic homeostasis in a stress mouse model. However, there is a lack of reports on the direct link between hub genes, microRNA and SLE.

Notably, the samples we chose for our dataset were all from peripheral blood, and we also validated this by peripheral blood from SLE patients, which may greatly support our results. Nevertheless, there are some shortcomings in our study. First, the dataset we analyze is an online public dataset, which is a secondary mining of the data. Second, the small sample size and the sample originating from one center in this study may be biased. Third, our immune cell analysis could not directly assess the limitations of tissue-resident immune cells, and it is hoped that future studies may combine tissue sampling with circulating cell analysis to gain a more comprehensive understanding. Fourth, due to the long duration of the disease in SLE patients and the wide range of medications used during treatment, the effect of medications on pivotal gene expression cannot be excluded. Finally, we lack more experiments to verify our results. Therefore, in the future, we will do further research in *in vivo* or *in vitro* experiments.

5 Conclusion

In summary, this is a study to screen for IRGs and metabolic pathways that are hubs in the peripheral blood of SLE. We identified five hub genes (*CXCL2*, *CXCL8*, *FOS*, *NFKBIA* and *CXCR2*), and constructed and validated a diagnostic model. We hope to provide new directions and evidence for the pathogenesis and treatment of SLE.

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 study was approved by the Ethics Committee of the Second Affiliated Hospital of Fujian Medical University, and all studies were conducted in accordance with relevant guidelines/regulations.

Informed consent was also obtained from the patients or their families, and the ethical approval number was [2024 (082)].

Author contributions

SZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. WH: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft. YT: Conceptualization, Investigation, Software, Writing – original draft. HL: Formal analysis, Resources, Visualization, Writing – original draft. XC: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Joint funds for the innovation of science and technology, Fujian province (Grant number: 2023Y9236) the Natural Science Foundation of Fujian Province (Grant number: 2024J01686).

Acknowledgments

We thank all those who participated in this study.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2025.1557307/full#supplementary-material>

References

- Hoi A, Igel T, Mok CC, Arnaud L. Systemic lupus erythematosus. *Lancet*. (2024) 403:2326–38. doi: 10.1016/S0140-6736(24)00398-2
- Tian J, Zhang D, Yao X, Huang Y, Lu Q. Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann Rheum Dis*. (2023) 82:351–6. doi: 10.1136/ard-2022-223035
- Bingham KS, Diaz Martinez J, Green R, Tartaglia MC, Ruttan L, Su J, et al. Longitudinal relationships between cognitive domains and depression and anxiety symptoms in systemic lupus erythematosus. *Semin Arthritis Rheum*. (2021) 51:1186–92. doi: 10.1016/j.semarthrit.2021.09.008
- Siegel CH, Sammaritano LR. Systemic lupus erythematosus: a review. *JAMA*. (2024) 331:1480–91. doi: 10.1001/jama.2024.2315
- Ugarte-Gil MF, Mak A, Leong J, Dharmadhikari B, Kow NY, Reátegui-Sokolova C, et al. Impact of glucocorticoids on the incidence of lupus-related major organ damage: a systematic literature review and meta-regression analysis of longitudinal observational studies. *Lupus Sci Med*. (2021) 8:e000590. doi: 10.1136/lupus-2021-000590
- Felten R, Arnaud L. Is it possible to stop glucocorticoids in systemic lupus? *Joint Bone Spine*. (2020) 87:528–30. doi: 10.1016/j.jbspin.2020.03.008
- O'Shea K, Misra BB. Software tools, databases and resources in metabolomics: updates from 2018 to 2019. *Metabolomics*. (2020) 16:36. doi: 10.1007/s11306-020-01657-3
- Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
- Clough E, Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res*. (2024) 52:D138–d144. doi: 10.1093/nar/gkad965
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. (2023) 51:D638–d646. doi: 10.1093/nar/gkac1000
- Doncheva NT, Morris JH, Holze H, Kirsch R, Nastou KC, Cuesta-Astroz Y, et al. Cytoscape string app 2.0: analysis and visualization of heterogeneous biological networks. *J Proteome Res*. (2023) 22:637–46. doi: 10.1021/acs.jproteome.2c00651
- Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, et al. Imm port, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific Data*. (2018) 5:180015. doi: 10.1038/sdata.2018.15
- Reichling C, Taieb J, Derangere V, Klopstein Q, Le Malicot K, Gornet JM, et al. Artificial intelligence-guided tissue analysis combined with immune infiltrate assessment predicts stage III colon cancer outcomes in PETACC08 study. *Gut*. (2020) 69:681–90. doi: 10.1136/gutjnl-2019-319292
- Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC Heart Failure*. (2020) 8:12–21. doi: 10.1016/j.jchf.2019.06.013
- Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform*. (2023) 24:bbad002. doi: 10.1093/bib/bbad002
- Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinform*. (2018) 19:432. doi: 10.1186/s12859-018-2451-4
- Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*. (2002) 29:288–91.
- Xing N, Dong Z, Wu Q, Kan P, Han Y, Cheng X, et al. Identification and validation of key molecules associated with humoral immune modulation in Parkinson's disease based on bioinformatics. *Front Immunol*. (2022) 13:948615. doi: 10.3389/fimmu.2022.948615
- Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. (2019) 37:773–82. doi: 10.1038/s41587-019-0114-2
- Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. Network analyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*. (2019) 47:W234–w241. doi: 10.1093/nar/gkz240
- Wang H, Shen M, Ma Y, Lan L, Jiang X, Cen X, et al. Novel mitophagy inducer alleviates lupus nephritis by reducing myeloid cell activation and autoantigen presentation. *Kidney Int*. (2024) 105:759–74. doi: 10.1016/j.kint.2023.12.017
- Wang YE, Zhang Y, Lin Z, Zhang H, Wang TY, Cao Y, et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat Commun*. (2021) 12:772. doi: 10.1038/s41467-021-21049-y
- Perez RK, Gordon MG, Subramaniam M, Kim MC, Hartoularos GC, Targ S, et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science*. (2022) 376:eabf1970. doi: 10.1126/science.abf1970
- Psarras A, Wittmann M, Vital EM. Emerging concepts of type I interferons in SLE pathogenesis and therapy. *Nat Rev Rheumatol*. (2022) 18:575–90. doi: 10.1038/s41584-022-00826-z
- Law C, Wacleche VS, Cao Y, Pillai A, Sowerby J, Hancock B, et al. Interferon subverts an AHR-JUN axis to promote CXCL13(+) T cells in lupus. *Nature*. (2024) 631:857–66. doi: 10.1038/s41586-024-07627-2
- Chen L, Cao SQ, Lin ZM, He SJ, Zuo JP. NOD-like receptors in autoimmune diseases. *Acta Pharmacol Sin*. (2021) 42:1742–56. doi: 10.1038/s41401-020-00603-2
- Wolf C, Lim EL, Mokhtari M, Kind B, Odainic A, Lara-Villacanas E, et al. UNC93B1 variants underlie TLR7-dependent autoimmunity. *Sci Immunol*. (2024) 9:eadi9769. doi: 10.1126/sciimmunol.adi9769
- Simmons DP, Nguyen HN, Gomez-Rivas E, Jeong Y, Jonsson AH, Chen AF, et al. SLAMF7 engagement superactivates macrophages in acute and chronic inflammation. *Sci Immunol*. (2022) 7:eabf2846. doi: 10.1126/sciimmunol.abf2846
- Zhao Z, Xu B, Wang S, Zhou M, Huang Y, Guo C, et al. Tfh cells with NLRP3 inflammasome activation are essential for high-affinity antibody generation, germinal center formation and autoimmunity. *Ann Rheum Dis*. (2022) 81:1006–12. doi: 10.1136/annrheumdis-2021-221985
- Singh MK, Rallabandi HR, Zhou XJ, Qi YY, Zhao ZZ, Gan T, et al. KLF2 enhancer variant rs4808485 increases lupus risk by modulating inflammasome machinery and cellular homeostasis. *Ann Rheum Dis*. (2024) 83:879–88. doi: 10.1136/ard-2023-224953
- Saito Y, Xiao Y, Yao J, Li Y, Liu W, Yuzhalin AE, et al. Targeting a chemo-induced adaptive signaling circuit confers therapeutic vulnerabilities in pancreatic cancer. *Cell Discovery*. (2024) 10:109. doi: 10.1038/s41421-024-00720-w
- Fieni C, Ciummo SL, Sorrentino C, Marchetti S, Vespa S, Lanuti P, et al. Prevention of prostate cancer metastasis by a CRISPR-delivering nanoplatfor for interleukin-30 genome editing. *Mol Ther*. (2024) 32:3932–54. doi: 10.1016/j.yimthe.2024.09.011
- Tan H, Jiang Y, Shen L, Nuerhashi G, Wen C, Gu L, et al. Cryoablation-induced neutrophil ca (2+) elevation and NET formation exacerbate immune escape in colorectal cancer liver metastasis. *J Exp Clin Cancer Res*. (2024) 43:319. doi: 10.1186/s13046-024-03244-z
- Dong X, Limjunyawong N, Sypek EI, Wang G, Ortines RV, Youn C, et al. Keratinocyte-derived defensins activate neutrophil-specific receptors Mrgpra2a/b to prevent skin dysbiosis and bacterial infection. *Immunity*. (2022) 55:1645–1662.e7. doi: 10.1016/j.immuni.2022.06.021
- Terui H, Yamasaki K, Wada-Irimada M, Onodera-Amagai M, Hatchome N, Mizuashi M, et al. *Staphylococcus aureus* skin colonization promotes SLE-like autoimmune inflammation via neutrophil activation and the IL-23/IL-17 axis. *Sci Immunol*. (2022) 7:eabm9811. doi: 10.1126/sciimmunol.abm9811
- Ambler WG, Kaplan MJ. Vascular damage in systemic lupus erythematosus. *Nat Rev Nephrol*. (2024) 20:251–65. doi: 10.1038/s41581-023-00797-8
- Antiochos B, Trejo-Zambrano D, Fenaroli P, Rosenberg A, Baer A, Garg A, et al. The DNA sensors AIM2 and IFI16 are SLE autoantigens that bind neutrophil extracellular traps. *eLife*. (2022) 11:11. doi: 10.7554/eLife.72103
- Li P, Jiang M, Li K, Li H, Zhou Y, Xiao X, et al. Glutathione peroxidase 4-regulated neutrophil ferroptosis induces systemic autoimmunity. *Nat Immunol*. (2021) 22:1107–17. doi: 10.1038/s41590-021-00993-3
- Guan Y, Peiffer B, Feng D, Parra MA, Wang Y, Fu Y, et al. IL-8+ neutrophils drive inexorable inflammation in severe alcohol-associated hepatitis. *J Clin Invest*. (2024) 134:8616. doi: 10.1172/JCI178616
- Haroun RA, Abdel Noor RA. Association of IL-8-251T/a (rs4073) gene polymorphism with systemic lupus erythematosus in a cohort of Egyptian patients. *Int Immunopharmacol*. (2023) 114:109528. doi: 10.1016/j.intimp.2022.109528
- Karakaslar EO, Katiyar N, Hasham M, Youn A, Sharma S, Chung CH, et al. Transcriptional activation of Jun and Fos members of the AP-1 complex is a conserved signature of immune aging that contributes to inflammaging. *Aging Cell*. (2023) 22:e13792. doi: 10.1111/acel.13792
- He J, Tsai LM, Leong YA, Hu X, Ma CS, Chevalier N, et al. Circulating precursor CCR7(lo)PD-1(hi) CXCR5+ CD4+ T cells indicate Th cell activity and promote antibody responses upon antigen reexposure. *Immunity*. (2013) 39:770–81. doi: 10.1016/j.immuni.2013.09.007
- Alaterre E, Ovejero S, Bret C, Dutrieux L, Sika D, Fernandez Perez R, et al. Integrative single-cell chromatin and transcriptome analysis of human plasma cell differentiation. *Blood*. (2024) 144:496–509. doi: 10.1182/blood.2023023237
- Freitas KA, Belk JA, Sotillo E, Quinn PJ, Ramello MC, Malipatolla M, et al. Enhanced T cell effector activity by targeting the mediator kinase module. *Science*. (2022) 378:eabn5647. doi: 10.1126/science.abn5647
- Luo S, Zhang H, Xie Y, Huang J, Luo D, Zhang Q. Decreased SUV39H1 at the promoter region leads to increased CREM α and accelerates autoimmune response in CD4(+) T cells from patients with systemic lupus erythematosus. *Clin Epigenetics*. (2022) 14:181. doi: 10.1186/s13148-022-01411-7
- Barnabei L, Laplantine E, Mbongo W, Rieux-Laucat F, Weil R. NF- κ B: at the Borders of autoimmunity and inflammation. *Front Immunol*. (2021) 12:716469. doi: 10.3389/fimmu.2021.716469
- Capece D, Verzella D, Flati I, Arboretto P, Cornice J, Franzoso G. NF- κ B: blending metabolism, immunity, and inflammation. *Trends Immunol*. (2022) 43:757–75. doi: 10.1016/j.it.2022.07.004

49. Caielli S, Balasubramanian P, Rodriguez-Alcazar J, Balaji U, Robinson L, Wan Z, et al. Type I IFN drives unconventional IL-1 β secretion in lupus monocytes. *Immunity*. (2024) 57:2497–2513.e12. doi: 10.1016/j.immuni.2024.09.004
50. Xu H, Zhang X, Wang X, Li B, Yu H, Quan Y, et al. Cellular spermine targets JAK signaling to restrain cytokine-mediated autoimmunity. *Immunity*. (2024) 57:1796–1811.e8. doi: 10.1016/j.immuni.2024.05.025
51. Ni H, Wang Y, Yao K, Wang L, Huang J, Xiao Y, et al. Cyclical palmitoylation regulates TLR9 signalling and systemic autoimmunity in mice. *Nat Commun*. (2024) 15:1. doi: 10.1038/s41467-023-43650-z
52. Yang L, Zhang T, Wang P, Chen W, Liu W, He X, et al. Imatinib and M351-0056 enhance the function of VISTA and ameliorate the development of SLE via IFN-I and noncanonical NF- κ B pathway. *Cell Biol Toxicol*. (2023) 39:3287–304. doi: 10.1007/s10565-023-09833-6
53. Huang J, Xu X, Wang X, Yang J, Xue M, Yang Y, et al. MicroRNA-590-3p inhibits T helper 17 cells and ameliorates inflammation in lupus mice. *Immunology*. (2022) 165:260–73. doi: 10.1111/imm.13434
54. Wang R, Wei A, Zhang Y, Xu G, Nong X, Liu C, et al. Association between genetic variants of microRNA-21 and microRNA-155 and systemic lupus erythematosus: a case-control study from a Chinese population. *J Clin Lab Anal*. (2022) 36:e24518. doi: 10.1002/jcla.24518
55. Xu L, Wang L, Shi Y, Deng Y, Oates JC, Kamen DL, et al. Up-regulated Interleukin-10 induced by E2F transcription factor 2-MicroRNA-17-5p circuitry in extrafollicular effector B cells contributes to autoantibody production in systemic lupus erythematosus. *Arthritis Rheumatol*. (2022) 74:496–507. doi: 10.1002/art.41987
56. Yu Y, Park S, Lee H, Kwon EJ, Park HR, Kim YH, et al. Exosomal hsa-miR-335-5p and hsa-miR-483-5p are novel biomarkers for rheumatoid arthritis: a development and validation study. *Int Immunopharmacol*. (2023) 120:110286. doi: 10.1016/j.intimp.2023.110286
57. Tornero-Esteban P, Rodríguez-Rodríguez L, Abásolo L, Tomé M, López-Romero P, Herranz E, et al. Signature of microRNA expression during osteogenic differentiation of bone marrow MSCs reveals a putative role of miR-335-5p in osteoarthritis. *BMC Musculoskelet Disord*. (2015) 16:182. doi: 10.1186/s12891-015-0652-9