Screening biomarkers for systemic lupus erythematosus based on single-cell and bulk RNA sequencing

Meng Yang  
The Third Affiliated Hospital of Guangxi Medical University

Di Yang  
The Third Affiliated Hospital of Guangxi Medical University

Linquan Gan  
The Third Affiliated Hospital of Guangxi Medical University

Yongzhuo Wu  
The Second Affiliated Hospital of Kunming Medical University

Yanbin Pan  
The Third Affiliated Hospital of Guangxi Medical University

Guiyan Deng  
The Third Affiliated Hospital of Guangxi Medical University

Xiangui Cheng  
The Third Affiliated Hospital of Guangxi Medical University

Wenjun Zheng (✉ dermatology@tom.com)  
The First Affiliated Hospital of Guangxi Medical University

Research Article

Keywords: Systemic lupus erythematosus, Single-cell RNA sequencing, Bulk RNA sequencing, GEO

Posted Date: November 25th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3585730/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Abstract
Systemic lupus erythematosus (SLE) is an autoimmune disease. The pathogenesis of SLE remains unclear, and the aim of this study was to identify novel biomarkers of SLE. First, key modules and key cell clusters for the trait of sample grouping were screened by weighted gene coexpression network analysis (WGCNA). The differentially expressed genes (DEGs) between SLE and normal samples in GSE72326 were screened. The candidate genes were obtained by overlapping DEGs, key module genes, and the marker genes of key cell clusters. The random forest algorithm was executed based on candidate genes, and the top 5 genes were selected as the hub genes. In addition, gene set enrichment analysis (GSEA) of hub genes was performed. Finally, expression validation, methylation analysis, and immunoinfiltration analysis were completed. A total of 90 DEGs were obtained between SLE and control samples in the GSE72326 dataset. By random forest analysis, the hub genes (TNFSF13B, FCGR1A, TNFSF10, ISG15, LAP3) were obtained. GSEA revealed that TNFSF13B and FCGR1A were involved in primary immunodeficiency, cytosolic DNA sensing pathway, ribosome, and TNFSF10, ISG15, and LAP3 were related to pyruvate metabolism, complement and coagulation cascade. TNFSF13B, FCGR1A, TNFSF10, ISG15, and LAP3 were identified as hub genes of SLE, which provides a new perspective to study SLE.

Purpose: Systemic lupus erythematosus (SLE) is an autoimmune disease. The pathogenesis of SLE remains unclear, and the aim of this study was to identify novel biomarkers of SLE.

Patients and methods: First, key modules and key cell clusters for the trait of sample grouping were screened by weighted gene coexpression network analysis (WGCNA). The differentially expressed genes (DEGs) between SLE and normal samples in GSE72326 were screened. The candidate genes were obtained by overlapping DEGs, key module genes, and the marker genes of key cell clusters. The random forest algorithm was executed based on candidate genes, and the top 5 genes were selected as the hub genes. In addition, gene set enrichment analysis (GSEA) of hub genes was performed. Finally, expression validation, methylation analysis, and immunoinfiltration analysis were completed.

Results: A total of 90 DEGs were obtained between SLE and control samples in the GSE72326 dataset. By random forest analysis, the hub genes (TNFSF13B, FCGR1A, TNFSF10, ISG15, LAP3) were obtained. GSEA revealed that TNFSF13B and FCGR1A were involved in primary immunodeficiency, cytosolic DNA sensing pathway, ribosome, and TNFSF10, ISG15, and LAP3 were related to pyruvate metabolism, complement and coagulation cascade.

Conclusion: TNFSF13B, FCGR1A, TNFSF10, ISG15, and LAP3 were identified as hub genes of SLE, which provides a new perspective to study SLE.

Introduction
Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by the production of autoantibodies and the deposition of tissue-bound immune complexes, leading to systemic autoimmunity resulting in damage to various organs. The incidence and prevalence rates of SLE exhibit
significant variation across regions and even subpopulations globally. The global incidence over the last 5 years has been 1.5 ~ 11/100,000, and the global prevalence ranges from 13 ~ 7713.5/100,000. The pathophysiological mechanisms of SLE are complex, and both the innate immune system and the adaptive immune system play an important role in SLE. More than 100 genetic loci with polymorphisms that have been associated with SLE have been identified. These genes are associated with activation of the immune system in response to foreign antigens, production of autoantigens, and activation of the innate and adaptive immune systems. The genetic and phenotypic heterogeneity of SLE is a hindrance to diagnosis and treatment, and no single biomarker shows ideal sensitivity and specificity for SLE. In recent decades, SLE treatment has shifted from the use of hydroxychloroquine, systemic glucocorticoids, and conventional immunosuppressive drugs to novel targeted biologics. Biomarkers in SLE may play a role in predicting treatment response and monitoring the course of the disease, which is crucial to research and clinical practice. Although many SLE-related biomarkers have been identified, these are well validated in clinical practice. However, there are still many unmet needs in the diagnosis and management of diseases. Therefore, there is an urgent need to further explore the diagnostic markers and therapeutic targets of SLE to provide new methods for the diagnosis, treatment, and intervention of patients.

Bulk RNA-seq and scRNA-seq have been used in many studies to explore disease etiology and prognosis. This can not only confirm the accuracy of single-cell sequencing analysis results but also validate the sequencing results by using Bulk RNA-seq data evaluation of the same sample. This research method has achieved remarkable results in the study of breast cancer, lung cancer, lymphoma and other neoplastic diseases, revealing the heterogeneity of tumor cells and discovering new cellular immune subsets. In addition, it has also been used in a variety of immune diseases, such as rheumatoid arthritis and ulcerative colitis, but unfortunately, there has been no literature report in SLE.

In this study, by combining scRNA-seq and bulk RNA-seq data of SLE, weighted correlation network analysis (WGCNA) and machine learning methods were used to screen potential biomarkers of SLE. At the same time, the potential function and DNA methylation level of biomarkers were analyzed, and a gene regulatory network was constructed, which further provides new ideas for the diagnosis and treatment of SLE.

Materials and methods

2.1 Data source

The scRNA-seq dataset GSE135779, two bulk RNA sequencing datasets (GSE72326 and GSE81622), and the methylation dataset GSE96879 were downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/gds) database. Seven SLE samples and 5 normal samples from peripheral blood mononuclear cells (PBMCs) in GSE135779 were included in this study. GSE72326 was used as the training set and contained 157 SLE samples and 20 control samples from PBMCs. SE81622 was used as
the validation set and contained 30 SLE samples and 25 control samples from PBMCs. Fifty-seven SLE samples and 33 control samples from PBMCs in GSE96879 were included.

2.2 Data filtering and cell type annotation

The Seurat package\(^8\) in R was applied to the data filter. Cells that met the following quality control criteria were excluded: (i) cell counts < 3 and (ii) genes detected per cell < 200 or > 2500. (iii) Mitochondrial count ratio > 5%. The data were normalized and then subjected to principal component analysis (PCA). Annotation of cell clusters based on information from the Human Primary Cell Atlas Data database (https://www.humancellatlas.org/data-portal/) using the SingleR and celldex packages.\(^9\) The expression abundance of different cell clusters in the SLE and normal samples from the GSE72326 dataset was calculated using CIBERSORT.\(^10\)

2.3 WGCNA

The most relevant modules for the trait of sample grouping were screened by WGCNA in GSE72326 using the “WGCNA” package.\(^11\) The modules with the highest relevance to the sample grouping were defined as key modules. The correlation of modules with cell clusters was calculated, and the cell clusters with the highest relevance to the key module were defined as key cell clusters. Enrichment analysis of key module genes was performed by Metascape (http://Metascape.org/).

2.4 Differential Expression Analysis

Differential expression analysis was performed between SLE and normal samples in GSE72326 using the “limma” package\(^12\), and differentially expressed genes (DEGs) were screened by \(|\logFC| \geq 1\) and adj. \(P < 0.05\). Heatmaps for DEGs were drawn with the pheatmap package\(^13\) in R. The volcano map of DEGs was drawn with the ggplot package\(^14\) in R.

2.5 Acquisition and validation of hub genes

The candidate genes were obtained by overlapping DEGs, key module genes, and the marker genes of key cell clusters. The random forest algorithm was executed with the RandomForest package\(^15\) based on candidate genes, and the top 5 genes were selected as the hub genes for this study based on the Gini coefficient of the genes. Finally, Spearman correlation analysis was performed among hub genes. Receiver operating characteristic (ROC) curves were plotted by the “plotROC” R package\(^16\) for the hub genes in GSE72326 and GSE81622, and the validation of expression trends was completed. Moreover, gene set enrichment analysis (GSEA) of hub genes was performed by the “ClusterProfiler” R package.\(^17\)

2.6 Methylation analysis and the establishment of regulatory networks

The methylation sites corresponding to hub genes in samples from the methylation dataset GSE96879 were found. Whether the beta values of the methylation sites corresponding to each hub gene differed
between the normal and SLE samples was explored. The miRNAs and transcription factors (TFs) of hub genes were predicted using the Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining (TUST, https://www.grnpedia.org/trtrust/) database. The TF/miRNA-hub gene regulatory network was constructed.

2.7 Immune analysis

The abundance of 20 immune cells in the SLE and normal samples in GSE72326 was calculated using the CIBERSORT algorithm. Immune cells differentially expressed between SLE samples and normal samples were obtained, and correlations between differential immune cells and hub genes were calculated.

Results

3.1 Seventeen cell clusters were identified

Normalization was performed according to the previous quality control (Fig. 1A), and PCA was completed (Fig. 1B). By unbiased clustering based on UMAP, 17 cell clusters were identified (Fig. 2A). The annotated results for each cell group are shown in Table 1. The expression bubbles of marker genes in each cell cluster are shown in Fig. 2B. The results of the immune analysis showed that the abundances of monocytes (CD16+ and CD14+), NKT cells, T cells (CD8+), and monocyte-derived dendritic cells in the GSE72326 dataset were significantly different between the SLE and normal samples (P < 0.05, Fig. 2C).
Table 1: The annotated results for 17 cell groups

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0</td>
<td>CD4 + memory T-cell</td>
</tr>
<tr>
<td>C1</td>
<td>CD16 + and CD14 + monocytes</td>
</tr>
<tr>
<td>C2</td>
<td>Natural killer T (NKT) cell</td>
</tr>
<tr>
<td>C3</td>
<td>CD4 + T-cell</td>
</tr>
<tr>
<td>C4</td>
<td>CD4 + memory T-cell</td>
</tr>
<tr>
<td>C5</td>
<td>Naive T-cell</td>
</tr>
<tr>
<td>C6</td>
<td>Natural killer cell</td>
</tr>
<tr>
<td>C7</td>
<td>CD8 + T-cell</td>
</tr>
<tr>
<td>C8</td>
<td>Monocyte derived dendritic cell</td>
</tr>
<tr>
<td>C9</td>
<td>B-cell</td>
</tr>
<tr>
<td>C10</td>
<td>Macrophage</td>
</tr>
<tr>
<td>C11</td>
<td>Monocyte derived dendritic cell</td>
</tr>
<tr>
<td>C12</td>
<td>Natural killer cell</td>
</tr>
<tr>
<td>C13</td>
<td>Megakaryocyte progenitor cell</td>
</tr>
<tr>
<td>C14</td>
<td>B-cell</td>
</tr>
<tr>
<td>C15</td>
<td>CD4 + memory T-cell</td>
</tr>
<tr>
<td>C16</td>
<td>Naive CD8 + T-cell</td>
</tr>
</tbody>
</table>

**Abbreviations:** SLE, systemic lupus erythematosus; WGCNA, weighted gene coexpression network analysis; DEGs, differentially expressed genes; GSEA, gene set enrichment analysis; GEO, Gene Expression Omnibus; PBMC, peripheral blood mononuclear cell; PCA, principal component analysis; ROC, receiver operating characteristic; TF, transcription factor; TRRUST, transcriptional regulatory relationships unraveled by sentence-based text mining; BAFF, B-cell-activating factor; TRAIL, tumor necrosis factor-associated apoptosis-inducing ligand; TNFSF10, tumor necrosis factor superfamily member 10; ISG15, interferon stimulating gene 15; LAP3, leucine amino peptidase 3; CNVs, copy number variation; RLR, retinoic acid inducer gene-I receptor; GO, gene ontology.

### 3.2 Key module genes

The **MEgreen** module, which contained 246 genes, was the most relevant module for sample grouping (Cor = 0.52, P = 1e-13, Fig. 3A-E). Heatmaps of the correlation of each module with the sample groupings and with each cell cluster showed that C1 (CD4 + memory T cells) had the highest correlation (Cor = 0.43,
P = 2e-9) with the MEgreen module. C1 (CD4 + memory T cells), the key cell cluster in this study, contained 379 marker genes. The entries involved in the MEgreen module included acyl-CoA thioesterase 9, absent in melanoma 2, and basic transcription factor 3 (Fig. 3F), and the network of these entry interactions is shown in Fig. 3G.

3.3 Acquisition of five hub genes

A total of 90 DEGs were obtained between SLE and control samples in the GSE72326 dataset, including 83 upregulated and 7 downregulated genes (Fig. 4A-B). Nine candidate genes (FCGR1A, TNFSF10, TNFSF13B, LAP3, ISG15, IFI6, CCR1, SCO2, PLSCR1) were obtained by overlapping 90 DEGs, 246 key module genes and 379 marker genes for key cell clusters (Fig. 4C). To obtain the hub gene, machine learning algorithms were performed. By random forest analysis (Fig. 4D), the genes with Gini index top 5 were screened, which were TNFSF13B, FCGR1A, TNFSF10, ISG15, and LAP3. Positive correlations were found between all five hub genes (Fig. 4E).

3.4 GSEA and immune analysis

GSEA revealed that TNFSF13B and FCGR1A were involved in primary immunodeficiency, cytosolic DNA sensing pathway, ribosome, and TNFSF10, ISG15, and LAP3 were related to pyruvate metabolism, complement and coagulation cascade (Fig. 5A). There were eight immune cells that were significantly different between SLE and normal samples (Fig. 5B). There were negative correlations between naive B cells, resting memory CD4 T cells and hub genes and positive correlations between monocytes, activated dendritic cells and hub genes (Fig. 5C).

3.5 Methylation sites of hub genes

A total of 50 methylation sites for the hub gene were found in the GSE96879 dataset samples (Fig. 6). Of these, all 2 methylation sites of FCGR1A differed between the SLE and control samples, one of the 10 methylation sites for ISG15 differed between groups, seven of the 16 methylation sites for LAP3 differed between groups, two of the 7 methylation sites for TNFSF10 differed between groups, and seven of the 15 methylation sites for TNFSF13B differed between groups.

3.6 Establishment of the regulatory network

A total of 14 TFs and 91 miRNAs for 5 hub genes were obtained (Fig. 7). The lowest number of regulatory relationships was obtained for FCGR1A, at only 3. The highest number of regulatory relationships was
obtained for TNFSF10, with as many as 59. Some relationship pairs are present in regulatory networks, such as SPI1 and FCGR1A, hsa-mir-1-3p and ISG15, and hsa-mir-142-3p and LAP3.

Discussion

scRNA-seq is a breakthrough approach that can be used to study differences in gene expression and cell progression in different groups by aggregating cells. However, bulk RNA-seq is usually performed in a batch manner and can measure the average gene expression level of a large number of cells. The combination of scRNA-seq and bulk RNA-seq data enables a more comprehensive identification of disease biomarkers and novel molecular targets, which has not been applied to SLE, although it has some applications for rheumatoid arthritis and ulcerative colitis.

In this study, bioinformatics technology was used to search differentially expressed genes in SLE, and key genes were screened to explore the expression of their related functional pathways and methylation sites. Finally, five hub genes (TNFSF13B, FCGR1A, TNFSF10, ISG15, LAP3) were evaluated and clinically validated. TNFSF13B, FCGR1A, ISG15, and LAP3 expression in the training set and the validation set was consistent, there were differences between groups, and the prediction accuracy was good. In addition, we used qPCR to verify the expression of these five genes in 5 SLE patients and 5 healthy people. The results showed that FCGR1A and TNFSF10 were differentially expressed in the two groups of samples, which may be due to errors caused by the small sample size. We will further expand the sample size for verification in the following experiments.

The TNFSF13B gene is responsible for encoding B-cell-activating factor (BAFF), which is a key B-cell survival factor. BAFF plays a key role in the survival and proliferation of B cells, class switching recombination, and the selection of autoimmune B cells. BAFF is highly expressed in SLE patients, and there is a clear correlation between high serum BAFF levels and dsDNA titer and lupus activity. In the study of gene polymorphisms, it was found that the rs9514828C > T genotype could increase the expression of the TNFSF13B gene and lead to an increase in BAFF levels in the blood. BAFF antagonists, such as Belimumab, reduce soluble BAFF levels by specifically binding BAFF, antagonizing its biological activity, and promoting B-cell apoptosis and have significant efficacy in reducing the number of autoantibodies and disease activity in SLE patients. It became the first biological agent approved by the FDA for SLE in 2011. In addition, the methylation level of the TNFSF13B gene may lead to an increase in antinuclear antibodies, affect serum renin levels, and cause kidney damage and blood system damage, which plays an important role in the pathogenesis and prognosis of SLE.

FCGR1A is the only functional gene capable of producing the high-affinity FcγR CD64 and plays an important role in regulating immune responses. S Malekni et al. found that FCGR1A expression was increased in SLE patients using an integrated Bayesian network approach. Wu et al. found that different FCGR1A alleles have certain differences in functional expression. For example, rs1848781 (c. -
131C > G) gene substitution can affect FCGR1A promoter activity and gene functional expression, and the rs587598788 insertion/deletion genotype can lead to CD64 overexpression. However, the rs1050204 allele mutation affects CD64-mediated phagocytosis. In recent years, studies have found that CD64 can reflect the level of IFN-I in SLE patients. The expression of CD64 on circulating monocytes parallels the level of serum immune complexes and blood urea nitrogen, suggesting that CD64 may be related to disease activity and organ damage and can be used as a biomarker for detecting SLE activity. In an experimental lupus mouse model, MRL/lpr mice had higher IgG deposition and CD64 expression in skin lesions, and the inflammatory response was severe, while CD64-deficient mice had reduced skin inflammation. In addition, studies have shown that the methylation level of the FCGR1A gene is significantly higher in SLE patients than in healthy controls, which impairs the function of the autoimmune response, leading to increased inflammatory cells and increased SLE activity.

Tumor necrosis factor superfamily member 10 (TNFSF10) is a death receptor, also known as tumor necrosis factor-associated apoptosis-inducing ligand (TRAIL) or Apo-2 L. TNFSF10 causes apoptosis by attaching to its receptors, such as death receptor-4, and plays an important role in the development and apoptosis of immune cells. In the experimental lupus mouse model, MRL/lpr mice had higher IgG deposition and CD64 expression in skin lesions, and the inflammatory response was severe, while CD64-deficient mice had reduced skin inflammation.

ISG15 (Interferon Stimulating Gene 15) is a member of the ubiquitin-like protein family and a downstream effector gene that induces IFN production. In addition to its direct antiviral effect, IFN-I can also promote the proliferation and differentiation of immune cells and promote the production of antibodies by B lymphocytes. However, high levels of IFN-I can cause abnormal autoimmune reactions and play a key driving role in the pathogenesis of SLE. SLE patients with high IFN-I have higher disease activity and are more likely to develop lupus nephritis and other severe manifestations. The ISG15 mRNA expression level is significantly higher in active SLE patients. Studies have shown that conformational changes in myeloperoxidase induced by ubiquitin and NETs containing free ISG15 from systemic lupus erythematosus patients promote a proinflammatory cytokine response in CD4+ T cells. ISG15 has also been reported to be associated with lymphocytopenia in SLE. ISG15 gene methylation plays an important role in SLE and can affect the immune response and inflammatory response in SLE patients. Studies have shown that ISG15 gene methylation levels are significantly positively correlated with immunological markers (such as serum antinuclear antibodies) and clinical markers (such as SLEDAI) in patients with active SLE. In addition, the level of ISG15 gene methylation and the pathogenesis of SLE can be used as biomarkers of SLE patients. Abnormal methylation of the PI3K signaling pathway to the ISG15 gene can be used to explore the common disease pathway to RA and SLE. Therefore, controlling ISG15 gene methylation may be a feasible treatment method for SLE patients to evaluate the treatment of SLE patients more effectively. Previous SLE-related biogenic
analysis has been conducted to explore the shared genetic characteristics and molecular mechanisms between systemic lupus erythematosus and pulmonary hypertension, proving that a high IFN response in SLE patients may be a key susceptibility factor for PAH.\(^{(35)}\)

LAP3 (leucine amino peptidase 3), which is highly expressed in a variety of malignant tumors, has been proven to catalyze the hydrolysis of leucine residues and affect the formation of tumor blood vessels.\(^{(36)}\) LAP3 has not been reported to be associated with SLE. Studies have found that the upregulation of LAP3 gene expression in patients with Sjogren's syndrome is associated with the deletion of copy number variation (CNVs) on chromosome 4, and CNV is associated with SLE susceptibility,\(^{(37)}\) such as tumor necrosis factor (TNF) on chromosome 6 and small RNA-binding exonuclease protective factor La (SSB) on chromosome 2. This suggests that chromosome CNV deletion may affect the expression of the LAP3 gene.\(^{(38)}\)

The five hub genes obtained in this study are involved in the development of various signaling pathways in SLE, including Toll-like receptor signaling, ribosomal pathway, RIG-I receptor signaling, cytokine receptor pathway, and apoptosis pathway. The literature shows that Toll-like receptors can control the tolerance of B cells to autoantigens and affect the BCR signaling cascade. However, TLR7 and TLR9 may play opposite roles in B cells.\(^{(39)}\) TLR7 causes the loss of germ-center tolerance and drives B cells to produce anti-RNA autoantibodies,\(^{(40)}\) while the opposite TLR9 can maintain immune tolerance and protect against diseases.\(^{(41)}\) This is consistent with the results of our enrichment analysis that these 5 hub genes may affect the morbidity of patients by acting on the toll-like receptor pathway. Unfortunately, due to the limitations of biological information research, the specific type of TLR affected by the hub gene still needs to be explored.

Ribosome biogenesis is upregulated in lymphocytes of SLE patients,\(^{(42)}\) which is consistent with our enrichment analysis results that these five hub genes may affect the pathogenesis of patients by acting on the ribosomal pathway and upregulation.

Retinoic acid inducer gene-I receptor (RLR) is a cytoplasmic RNA innate sensor that induces IFN-I production and plays an antiviral role in innate immunity. Recent studies have found that it is also involved in apoptosis regulation and autoantibody production in SLE,\(^{(43)}\) which is consistent with the results of our enrichment analysis.

Programmed cell death contributes to the release of damage-related molecules, amplification of the inflammatory response, production of self-antigens, and tissue damage in SLE. The apoptosis of mononuclear macrophages is enhanced in SLE patients, but the clearance function of apoptotic cells is weakened, thus leading to the production of a large number of autoantibodies,\(^{(44)}\) which is consistent with our enrichment analysis results that the FCGR1A gene is involved in the autophagy pathway.

A large number of studies have shown that the activation of the classical complement pathway by the immune complex in SLE is a terminal event leading to multiorgan tissue injury,\(^{(45)}\) which is consistent with
our enrichment analysis results that FCGR1A, TNFSF13B, and TNFSF10 participate in the complement pathway.

Some scientists have used gene ontology (GO) functions and pathway enrichment analysis to screen for key candidate genes and signaling pathways in SLE. However, the advantage of our method is that we combine gene expression features at the single-cell level and tissue level, and the two complement each other to make the results more reliable.

**Conclusion**

In this study, we used the combination of scRNA-seq and bulk RNA-seq sequencing to screen biomarkers and new molecular targets for SLE for the first time and successfully screened 17 cell clusters and 5 central genes. WGCNA identified MEgreen, the key module with the strongest correlation with systemic lupus erythematosus. In addition, we performed gene enrichment analysis on Megreen and obtained 90 differentially expressed genes. We intersected key module genes, key cell group marker genes, and differential genes to obtain 9 candidate genes and used random forest screening to obtain 5 hub genes (TNFSF13B, FCGR1A, TNFSF10, ISG15, LAP3). We also performed differential expression analysis, methylation site analysis, and TF/miRNA prediction on hub genes. In the follow-up verification, we found that four genes with good effects (TNFSF13B, FCGR1A, ISG15, LAP3) may be biomarkers and new molecular targets for systemic lupus erythematosus. These results have important implications for our in-depth understanding of the pathogenesis of SLE and provide potential guidance for future treatment options. Although the findings suggest that 17 cell clusters and five central genes may serve as potential biomarkers and therapeutic targets, more clinical trials are needed to confirm these results. At the same time, we will continue to pay special attention to the specific role of these genes in the pathogenesis of SLE and their interaction with other factors.

**References**


Figures
Figure 1

Quality control of single-cell data. (A) The samples in the GSE135779 dataset were quality controlled by removing cells with cell counts less than 3, gene counts less than 200 or greater than 2500, and mitochondrial gene counts greater than 5%. (B) Principal component analysis (PCA) analysis of cells.
Figure 2

Identification of the cell subpopulation based on the single-cell sequencing data. (A) Cell clustering and subgroup annotation. (B) Expression of marker genes in each cell subpopulation. (C) The proportions of different cell subpopulations in systemic lupus erythematosus (SLE) and normal samples. ns, not significant; *p<0.05; ** p<0.01; ****p<0.0001.
Figure 3

Identification of key modules of SLE and functional enrichment analysis. (A) The samples in the GSE72326 dataset were clustered to remove outliers. (B) Selection of the optimal soft-thresholding (power). (C) Cluster tree of modules before merging. (D) Dynamic cut tree after module merging. (E) Heatmap of the correlation between gene modules and clinical traits (SLE, normal, and 17 cell
Identification of hub genes in the GSE72326 dataset. (A, B) The volcano map (A) and heatmap (B) of differentially expressed genes (DEGs) between SLE and normal samples. (C) Venn diagram of 9 subpopulation). (F) The results of functional enrichment for key module genes. (G) The interaction among enriched pathways.
Figure 5

Gene Set Enrichment Analysis (GSEA) and immune infiltration analysis. (A) The pathways enriched in TNFSF13B. (B) The proportion of immune cells in SLE and normal samples. (C) The relevance of hub genes and immune cells.
Figure 6

Analysis of methylation sites for hub genes.
Figure 7

The regulatory network of hub genes, transcription factors (TFs), and microRNAs (miRNAs). Yellow represents hub genes, green represents TFs, and blue represents miRNAs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GraphicalAbstract.tif