



OPEN ACCESS

EDITED BY

Xiaolin Sun,
Peking University People's Hospital, China

REVIEWED BY

Qiang Shu,
Shandong University, China
Hua Ye,
Peking University People's Hospital, China
Chun Li,
Peking University People's Hospital, China

*CORRESPONDENCE

Xinwang Duan
✉ dxw_efyfsmyk@163.com

SPECIALTY SECTION

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

RECEIVED 25 January 2023

ACCEPTED 04 April 2023

PUBLISHED 18 April 2023

CITATION

Li H, Zhang X, Shang J, Feng X, Yu L, Fan J,
Ren J, Zhang R and Duan X (2023)
Identification of NETs-related
biomarkers and molecular clusters in
systemic lupus erythematosus.
Front. Immunol. 14:1150828.
doi: 10.3389/fimmu.2023.1150828

COPYRIGHT

© 2023 Li, Zhang, Shang, Feng, Yu, Fan, Ren,
Zhang and Duan. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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 NETs-related biomarkers and molecular clusters in systemic lupus erythematosus

Haoguang Li, Xiuling Zhang, Jingjing Shang, Xueqin Feng,
Le Yu, Jie Fan, Jie Ren, Rongwei Zhang and Xinwang Duan*

Department of Rheumatology and Immunology, The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China

Neutrophil extracellular traps (NETs) is an important process involved in the pathogenesis of systemic lupus erythematosus (SLE), but the potential mechanisms of NETs contributing to SLE at the genetic level have not been clearly investigated. This investigation aimed to explore the molecular characteristics of NETs-related genes (NRGs) in SLE based on bioinformatics analysis, and identify associated reliable biomarkers and molecular clusters. Dataset GSE45291 was acquired from the Gene Expression Omnibus repository and used as a training set for subsequent analysis. A total of 1006 differentially expressed genes (DEGs) were obtained, most of which were associated with multiple viral infections. The interaction of DEGs with NRGs revealed 8 differentially expressed NRGs (DE-NRGs). The correlation and protein-protein interaction analyses of these DE-NRGs were performed. Among them, HMGB1, ITGB2, and CREB5 were selected as hub genes by random forest, support vector machine, and least absolute shrinkage and selection operator algorithms. The significant diagnostic value for SLE was confirmed in the training set and three validation sets (GSE81622, GSE61635, and GSE122459). Additionally, three NETs-related sub-clusters were identified based on the hub genes' expression profiles analyzed by unsupervised consensus cluster assessment. Functional enrichment was performed among the three NETs subgroups, and the data revealed that cluster 1 highly expressed DEGs were prevalent in innate immune response pathways while that of cluster 3 were enriched in adaptive immune response pathways. Moreover, immune infiltration analysis also revealed that innate immune cells were markedly infiltrated in cluster 1 while the adaptive immune cells were upregulated in cluster 3. As per our knowledge, this investigation is the first to explore the molecular characteristics of NRGs in SLE, identify three potential biomarkers (HMGB1, ITGB2, and CREB5), and three distinct clusters based on these hub biomarkers.

KEYWORDS

systemic lupus erythematosus (SLE), neutrophil extracellular traps (NETs), biomarker, bioinformatics, machine learning, clusters

Introduction

Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease characterized by the appearance of a variety of autoantibodies, resulting in a wide range of tissue and organs damaged (1). It has a higher frequency in females than males, with a prevalence of 9: 1 (2). Global SLE prevalence in adults is estimated to range from 30-150/100,000 and an incidence of about 2.2-23.1/100,000 annually, that is more than five million people suffering this disease worldwide (3). Despite the increase in the SLE survival rate over the last several decades (4, 5), it is still linked with early mortality (6), indicating substantial medical challenges especially linked with diagnosis and treatment. The clinical presentation of SLE is highly heterogeneous and there is an increasing number of atypical cases, which poses challenges in diagnosing the disease. Moreover, SLE treatment usually involves broad-spectrum immunosuppressive therapy, the efficacy of which can vary considerably between individuals and may have significant adverse effects, particularly when taken for prolonged periods. Different etiology has resulted in distinct clinical manifestations and cellular and molecular bases for SLE. Therefore, it is of immense clinical importance to uncover the underlying pathogenesis of SLE for its diagnosis and management.

The pathogenesis of SLE is complex and associated with enhanced dysregulation of innate and adaptive immunity (7). Neutrophils, the most abundant innate immune cells in humans, have long been considered as non-specific at the front line of defense against infections, however, the elucidation of neutrophil extracellular traps (NETs) has drastically revolutionized our current knowledge about neutrophil's activity and their significance during the immune response. NETs are extracellular networks composed of chromatin associated with cytosolic and granular proteins that are released by neutrophils in response to various stimuli, a process called NETosis (8–10). Recent research has highlighted that NETs are crucial for the stimulation and advancement of systemic autoimmune diseases and perform complex inflammatory responses that cause organ injury (11–14). For example, aberrant NETs production and/or its decreased clearance, as well as NETosis-associated molecule alterations, in both animals and humans suffering from lupus. The NETs secrete extruded nuclear antigens, these are the source of autoantigens, which also participate in the breakdown of self-tolerance during lupus. Increased NETs can also stimulate the release of pro-inflammatory cytokine interferon- α , produce a direct cytotoxic effect on different renal cells, and induce capillary necrosis and podocyte loss. Additionally, NETs can induce endothelial-mesenchymal trans-differentiation (EndMT), thereby, stimulating activated myofibroblasts and causing the production of extracellular matrix. Thus, NETs have also been suggested as a critical contributor to the development of SLE, and targeting NETs has also been shown to have therapeutic effects. However, a detailed and integrated investigation of NETs-associated genes in SLE remains to be conducted.

Because of significant advancements in gene microarray technology, researchers can now rapidly assess the expression levels of thousands of genes, which has contributed to increased

genetic knowledge of the disease etiology. Therefore, this investigation employs bioinformatics tools to explore the molecular characteristics of NETs-related genes (NRGs) in SLE and identify reliable biomarkers and molecular clusters for better diagnosis and effective management.

Materials and methods

Data source

The GEOquery R package (15) was used to download microarray datasets relevant to SLE from the Gene Expression Omnibus (GEO) online database (16). The following filtering criteria were used: (1) they were all from SLE patients and healthy controls; (2) the sample size was relatively large, with more than 20 sample; (3) the test specimens were from humans; and (4) the tissues used for sequencing were whole blood (WB) or peripheral blood mononuclear cell (PBMC). Additionally, the data were available for free download from the GEO database. Based on the above criteria, four datasets (GSE45291, GSE81622, GSE61635, and GSE122459) relevant to SLE were finally included in this study. There were 312 samples in the GSE45291 dataset (GPL13158 platform), of which 292 were of SLE and 20 were of normal WB. This dataset was utilized as a training set for analysis. The GSE81622 dataset (GPL10558 platform) included 30 SLE and 25 normal PBMC samples, the GSE61635 dataset (GPL570 platform) had 99 SLE and 30 normal WB samples, and the GSE122459 dataset (GPL16791 and GPL18573 platforms) comprised 20 SLE and 6 normal PBMC samples. These three datasets were selected for validation analysis. Additionally, a total of 69 NRGs were obtained from a previous study (17) (Supplementary Table S1).

All raw data in our investigation was subjected to normalization and adjustment for background, and we also cross-referenced all probe names with their respective gene symbols. Additionally, we also used the “ComBat” algorithm (18) to address the batch effects to improve the efficacy of the subsequent analysis.

Identification of differentially expressed genes

The adjust p-value < 0.05 and fold changes (FC) > 1.5 were used set as the cut-off criteria to evaluate DEGs between SLE and healthy controls *via* the R-program limma package (19). The data acquired was presented in volcano plot and heatmap with the help of the R-packages ggplot2 (20) and pheatmap (21), respectively.

Functional annotation and pathway enrichment of DEGs

To determine DEGs' role in SLE, clusterProfiler of the R-package (22) was used for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. The Benjamini-

Hochberg multiple correction approach was used to compensate for FDR-adjusted ($P < 0.05$) data. The three categories; cellular component (CC), molecular function (MF), and biological process (BP) of the GO analysis were crucial for examining physiological functions (23). KEGG analysis was performed to investigate potential routes (24). The ggplot (20) and GOplot (25) R programs were used to plot the findings of GO and KEGG, respectively.

Gene expression patterns and protein-protein interaction analyses of differentially expressed NRGs

Next, the expression patterns of the DE-NRGs were analyzed. First, the DE-NRGs were obtained through the interaction of DEGs with NRGs, and their chromosomal locations were revealed and visualized using Package RCircos (26). Thereafter, using the corrplot function (R package corrplot) (27) the correlations among the expression levels of DE-NRGs were confirmed. Furthermore, a PPI enrichment analysis was performed using the online software GeneMANIA (<https://genemania.org/>) (28), and any combined score > 0.4 interaction was deemed statistically significant. Gene interactions are represented as a network, where nodes stand in for individual genes and linkages for whole networks. To further elucidate the biological features and mechanism of interacting proteins of the DE-NRGs, functional annotation analysis of inter was performed using another web tool, Metascape (<http://metascape.org/>) (29).

Identification of NETs-related hub gene based on machine learning algorithm

Following three machine learning algorithms were adopted, support vector machine (SVM), least absolute shrinkage and selection operator (LASSO), and random forest (RF) to screen hub genes from DE-NRGs, using R packages of glmnet (30), e1071 (31) and caret (32), and RF (33), respectively. LASSO logistic regression identified variables by looking for those with the lowest likelihood of classification mistakes (34). The SVM technique creates a hyperplane in the characteristic space with the maximum margin for isolating negative from positive instances (35). RF is a group of independent decision trees used in an ensemble machine-learning technique to predict regression or clustering (36). The optimal lambda for LASSO regression was selected based on 10 resampling iterations of 10-fold cross-validation. Furthermore, the performances of SVM and RF were also evaluated based on 10-fold cross-validations. The selection of hub SLE genes was based on the overlapping genes produced from the three algorithms.

Construction and validation of a diagnostic model for SLE

With the R package of rms, multivariate logistic regression analysis was performed to establish a nomogram model based on the NETs-related hub genes (37) for the diagnosis of SLE. The “total points” reflect

the total of the points assigned to the aforementioned predictors, whereas each predictor has a corresponding point. The receiver operating characteristic (ROC) curves and area under the curve (AUC) were obtained from both training and validation sets to confirm nomograms’ diagnostic value. ROC curves were assessed *via* the ROC R package (38). Moreover, the decision curve analysis (DCA) and calibration curve were applied to evaluate the accuracy and practical applicability of the diagnostic model with the rms R package (37).

Consensus cluster analysis

With the help of the ConsensusClusterPlus R package (39), an unsupervised hierarchical clustering (50 iterations, resample rate of 80%) was performed on the 292 SLE samples from the training set based on the expression of the NETs-related hub genes. This clustering was based on: 1) the gradual and smooth increase of the cumulative distribution function (CDF) curve. 2) no small sample size in any group. 3) delta area should have the largest decrease. 4) after clustering, the intra-group correlation should increase, while the inter-group correlation should decrease.

Functional distinctions among the sub-clusters

Afterward, the NETs-related sub-clusters were compared pairwise, and the highly expressed DEGs were determined according to statistical criteria ($FC > 1.5$ and $\text{adj.p} < 0.05$) in each cluster. Followed by that GO and KEGG enrichment analyses were carried out to depict their physiological activities and visualized with the circle (40) and GOplot (25) R packages, respectively.

Evaluating the immune infiltration of the sub-clusters

To evaluate sub-clusters’ immune infiltration, the R packages of GSVA (41) and GSEAbase (42) were utilized to quantify the relative abundance of immune cells and immune-related pathways and then compared enrichment scores across the sub-clusters, the results of which were visualized using the ggpubr R package (43).

Statistical analysis

Unless otherwise stated, all investigations and visualization were carried out *via* R software (version.4.2.1). $P < 0.05$ and $FC > 1.5$ were set as the criteria for statistical importance.

Results

Identification of DEGs

Figure 1 is a detailed flowchart of the research procedure. Based on significance criteria, a total of 1006 (393 upregulated and 613

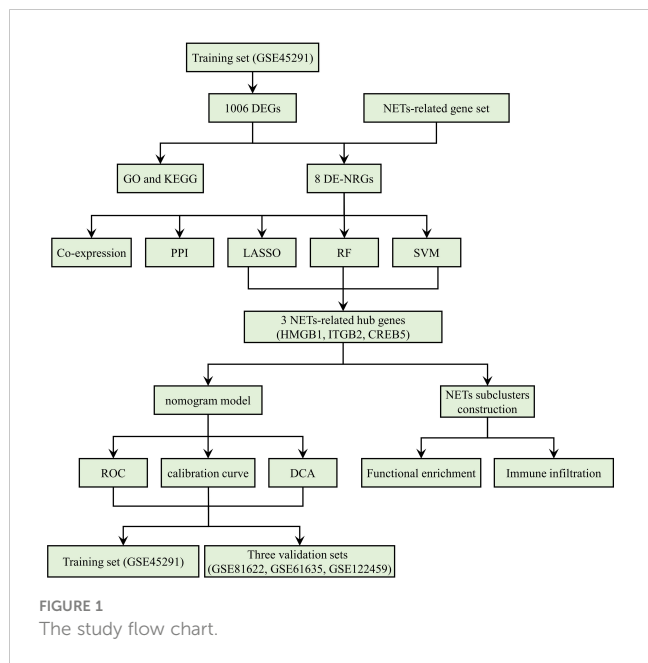


FIGURE 1
The study flow chart.

downregulated) SLE-associated DEGs (Supplementary Table S2). Figures 2A, B depict the volcano plot and heatmap of DEGs.

Functional annotation and pathway enrichment of DEGs

According to GO-BP analysis, the DEGs were considerably increased in viral processes, viral life cycle, the response against virus and type I interferon, regulation of biotic stimulus and innate immune responses, negative viral process regulation, defense response to virus and symbiont, and cellular response to type I interferon (Figure 3A). The findings of the GO-CC analysis were primarily lytic vacuole membrane, lysosomal membrane, ficolin-1-rich granule lumen, early endosome, endocytic vesicle, coated vesicle membrane, and clathrin-coated vesicle, vesicle membrane, and endocytic vesicle (Figure 3B).

The activity of ubiquitin-protein ligase, ubiquitin-binding, ubiquitin-protein transferase, ubiquitin-like protein transferase, ubiquitin-like protein ligase, MHC protein complex binding, immune receptor, double-stranded RNA binding, and DNA-binding transcription factor binding were the key outcomes of enrichment analysis in GO-MF (Figure 3C). The KEGG pathway analysis primarily indicated that these DEGs were associated with virus-related diseases including Hepatitis C, Natural killer cell mediated cytotoxicity, Influenza A, Tuberculosis, Th17 cell differentiation, Leishmaniasis, Human T-cell leukemia virus 1 infection, Viral life cycle – HIV – 1, Graft-versus-host disease, and NOD-like receptor signaling pathway (Figure 3D).

Gene expression patterns and PPI network of DE-NRGs

The interaction of DEGs with NRGs resulted in 8 DE-NRGs (Figure 4A), and the location of these DE-NRGs on chromosomes was: ALPL (chr1), CD93 (chr20), CREB5 (chr7), FCGR3B (chr1), HMGB1 (chr13), ITGB2 (chr21), SLC22A4 (chr5), and VNN3 (chr6) (Figure 4B). The gene relationship network diagram indicated the correlation among these DE-NRGs (Figure 4C). Additionally, a DE-NRGs network was constructed, comprising 20 other genes including AGER, FCGR2A, IRAK2, SPON2, JAM3, F10, CD1C, GP1BA, PROC, TLR2, HCK, TLR4, TIRAP, ITGAL, ITGAD, MS4A7, ICAM3, TBXAS1, TYROBP, and ICAM4 (Figure 4D), which were mainly involved in PID_INTEGRIN2_PATHWAY (Figure 4E).

Identification of NETs-related hub genes for SLE

The LASSO, SVM, and RF used DE-NRGs expression levels to discriminate SLE from healthy controls. In the RF classifier, the optimal number of trees selected was 44 as it has the lowest error rate and stability (Figure 5A), to obtain the dimensional importance of the DE-NRGs. Furthermore, the top 5 genes (HMGB1, ITGB2, CREB5,

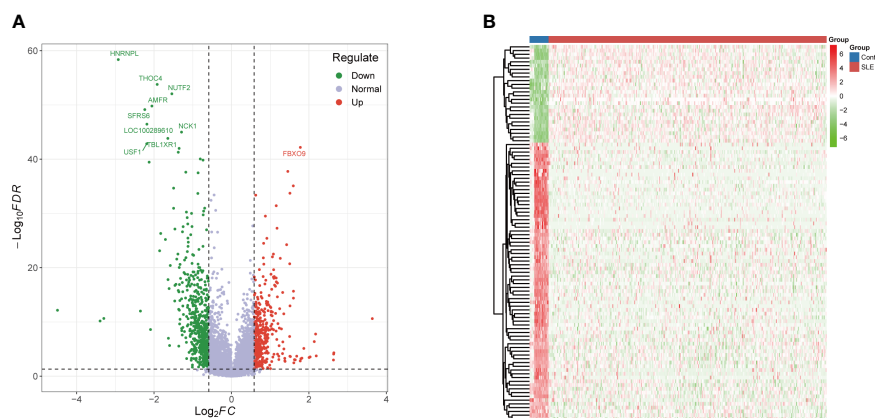


FIGURE 2
Identification of DEGs. (A) Volcano plot: The volcano plot was constructed using the fold change values >1.5 and adjusted P-values < 0.05. (B) The heatmap of the DEGs and different colors represent the trend of gene expression in different PBMCs. The top 100 genes ranked according to adjusted p-values as shown in this figure. DEGs differentially expressed genes.

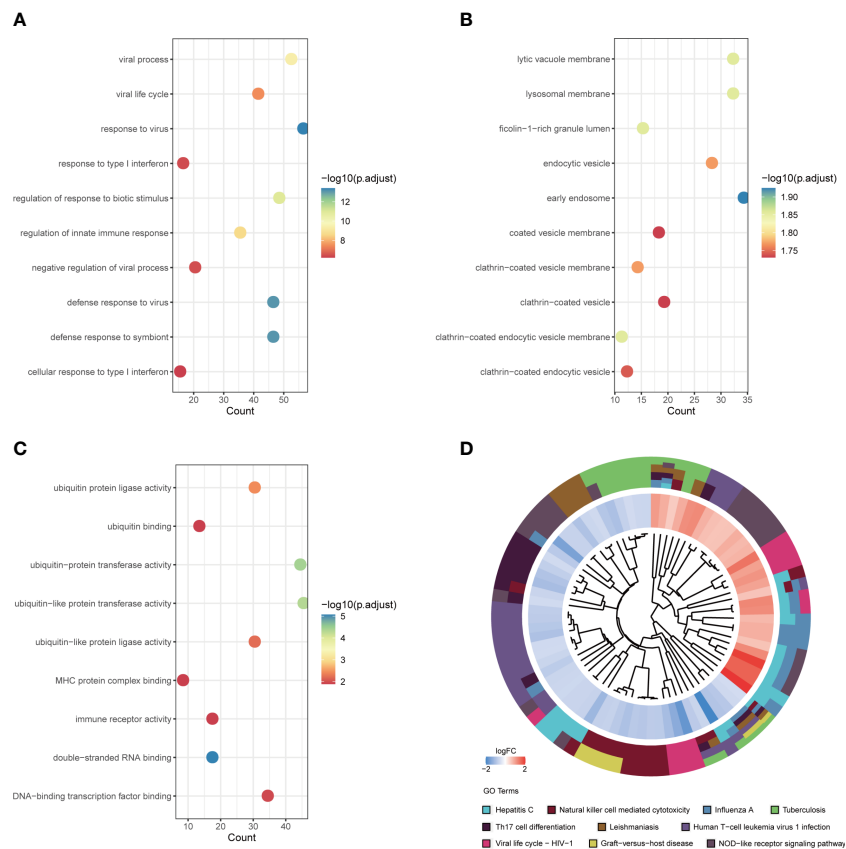


FIGURE 3 Functional annotation and pathway enrichment of DEGs. **(A)** Top 10 GO BP pathway; **(B)** Top 10 GO CC pathway; **(C)** Top 10 GO MF pathway; **(D)** Top 10 KEGG pathway. DEGs, differentially expressed genes; GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto encyclopedia of genes and genomes.

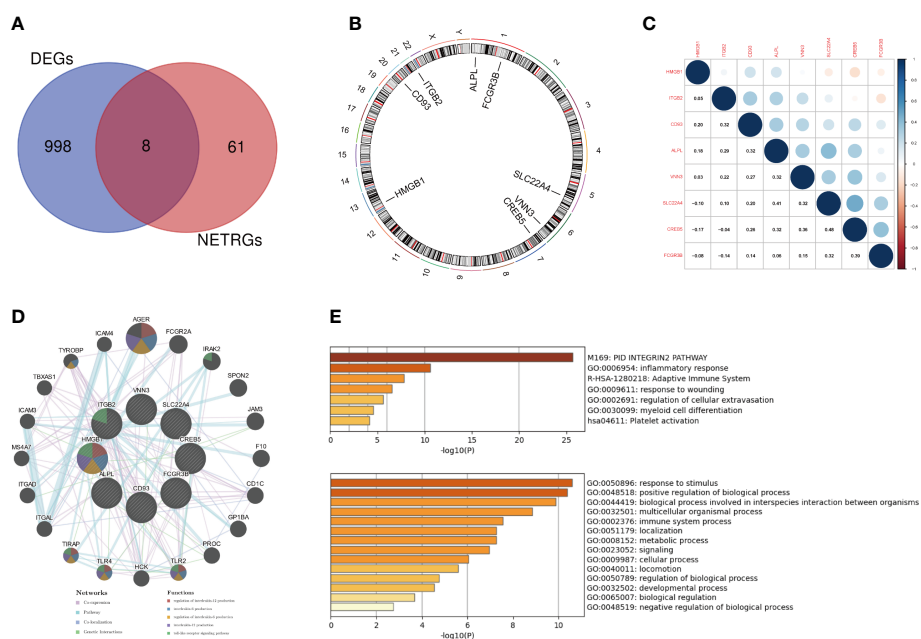


FIGURE 4 Gene expression patterns and PPI network of DE-NRGs. **(A)** The overlap of genes between DEGs and NRGs. **(B)** The location of the DE-NRGs on chromosomes. **(C)** Correlation matrix of DE-NRGs. **(D)** PPI network of DE-NRGs and its interacting proteins. **(E)** GO enrichment analysis of interacting proteins of DE-NRGs. PPI, protein-protein interaction; DEGs, differentially expressed genes; NRGs, NETs-related genes; GO, gene ontology.

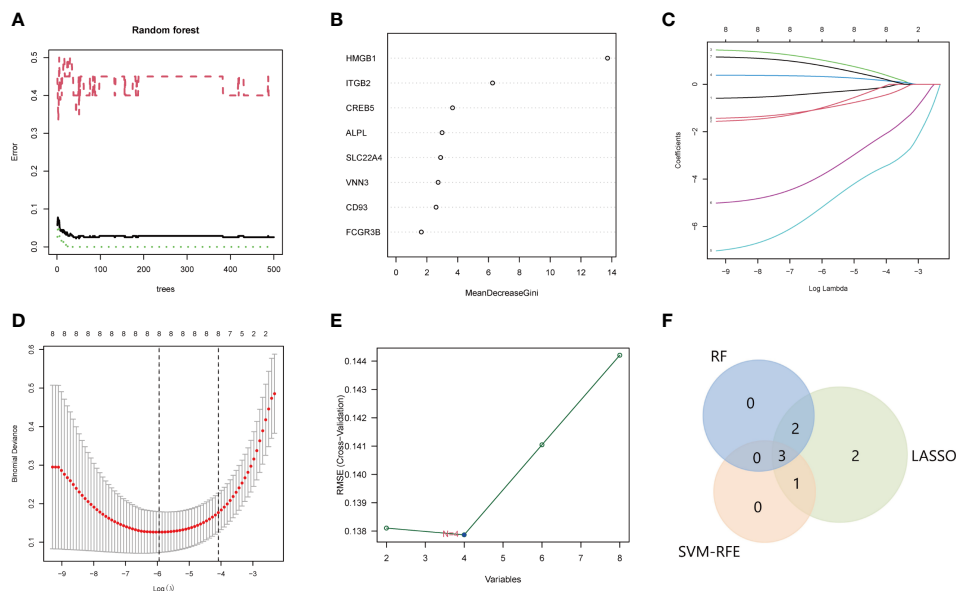


FIGURE 5

Identification of the hub genes for SLE. A-B. RF to screen candidate hub genes. (A) The influence of the number of decision trees on the error rate. The x-axis is the number of decision trees and the y-axis is the error rate. Green represents the SLE samples, red represents the non-SLE samples, and black represents the overall samples. (B) Gini importance measure. The horizontal axis represents the mean decrease Gini, and the vertical axis represents characteristic NRGs. C-D. LASSO logistic regression to screen candidate hub genes. (C) LASSO coefficient spectrum of eight genes enrolled and generate a coefficient distribution map for a logarithmic sequence. (D) The 10-fold cross-validation process was repeated 10 times to select the optimal penalization coefficient lambda. The value of lambda yielded the minimum average binomial deviance that was used to select features. (E) The RMSE of candidate hub genes combination of the SVM algorithm. The minimum classification error (minimal RMSE = 0.1377) in the condition of four candidate genes. (F) The Venn diagram shows the overlap of candidate genes among the above three algorithms. SLE, systemic lupus erythematosus; RF, random forest; NRGs, NETs-related genes; LASSO, least absolute shrinkage, and selection operator; RMSE, root-mean-square error; SVM, support vector machine.

ALPL, and SLC22A4) based on their importance in MeanDecreaseGini result (Figure 3B), were selected as candidate hub genes. In LASSO logistic regression, all the DE-NRGs were recognized as candidate hub genes according to the optimal lambda = 0.002619094 (Figures 5C, D). The SVM model has the minimum classification error [minimal root-mean-square error (RMSE) = 0.1377] in the case of four candidate genes; HMGB1, ITGB2, CREB5, and VNN3 (Figure 5E). Finally, HMGB1, ITGB2, and CREB5, overlapping genes by the three algorithms, were selected as hub genes (Figure 5F).

Performance of NETs-related hub genes

Based on the three NETs-related hub Genes, a nomogram model was constructed to calculate the odds of developing SLE and to further evaluate their predictive effectiveness (Figure 6A). The calibration curve (Figure 6B) and DCA (Figure 6C) confirmed the nomogram's performance. A combined ROC analysis of the three hub genes was carried out with an AUC as high as 0.983 (Figure 6D), indicating outstanding diagnostic efficiency. Furthermore, AUCs of 0.788, 0.930, and 0.901 for CREB5, HMGB1, and ITGB2 respectively, showed their potency as useful diagnostic biomarkers (Figures 6E-G). In addition, the three validation datasets, GSE81622 (Figures 7A-C), GSE61635 (Figures 7D-F), and GSE122459 (Figures 7G-I), were utilized to

independently assess this model's performance by ROC curves, calibration curve, and DCA, and found that it performed well in the validation sets with AUCs of 0.904, 0.945, and 0.883, respectively.

Identification of NETs-related sub-clusters in SLE

The CREB5, HMGB1, and ITGB2 expression profiles were used to group the 292 SLE samples *via* a consensus clustering algorithm. The k value of three (k = 3) gave the most stable cluster numbers, based on the above criteria (Figures 8A-C). As the PCA plot indicates, gene expression patterns were specific across the sub-clusters (Figure 8D).

Functional distinctions among the sub-clusters

A total of 107, 25, and 38 DEGs, highly expressed in cluster1 (Supplementary Table S3), cluster2 (Supplementary Table S4), and cluster3 (Supplementary Table S5) respectively were identified. As GO analysis indicates, the cluster 1 highly expressed DEGs were enriched in cell killing (GO:0001906), neutrophil activation (GO:0042119), neutrophil-mediated immunity (GO:0002446), cellular response to type I interferon (GO:0071357), and response

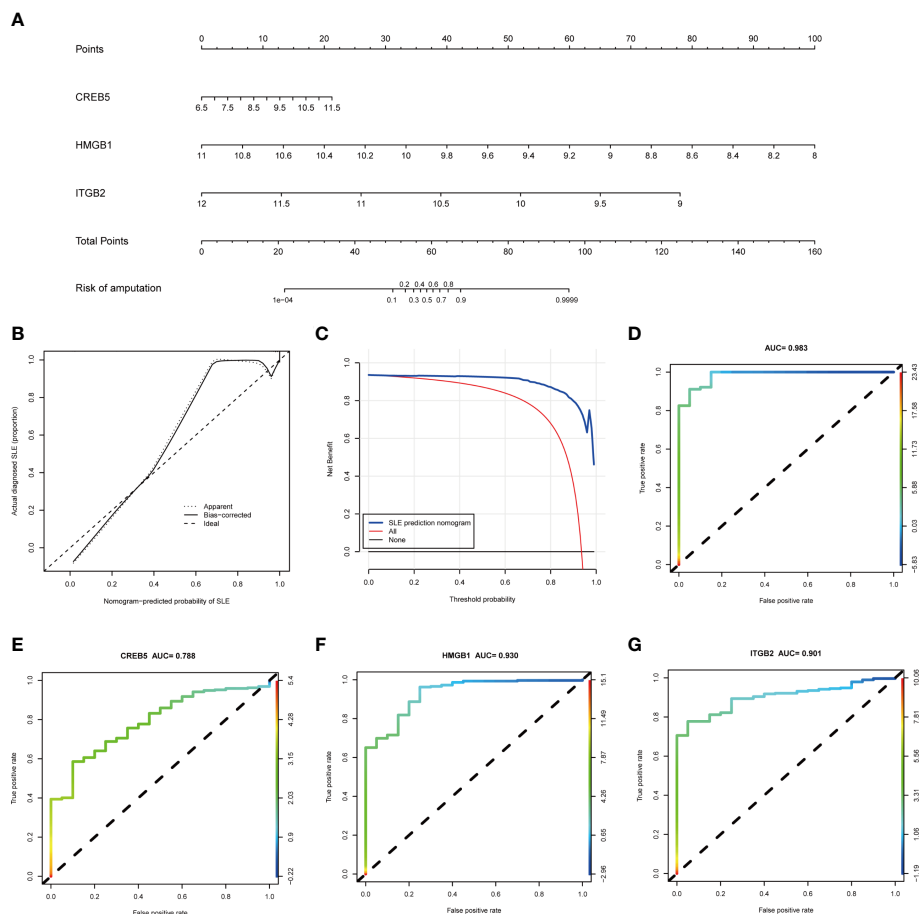


FIGURE 6

The performance of hub genes in the training set. **(A)** Construction of a nomogram for predicting the occurrence of SLE based on hub genes. **(B)** Calibration curve. The y-axis is the actual rate of SLE diagnosis and the x-axis is the predicted risk of SLE. The diagonal dotted line represents a perfect prediction by an ideal model. The solid line represents the bias-corrected performance of the nomogram, where a closer fit to the diagonal dotted line represents a better prediction. **(C)** Decision curve analysis. The blue represents the net benefit of the nomogram in the prediction of SLE occurrence. The red line represents the assumption that all people have SLE. The black line represents the assumption that all people do not have SLE. **(D)** ROC curve of the three hub genes combined diagnosis of SLE. **(E)** ROC curve of CREB5 gene diagnosis of SLE; **(F)** ROC curve of HMGB1 gene diagnosis of SLE; **(G)** ROC curve of ITGB2 gene diagnosis of SLE. SLE, systemic lupus erythematosus; ROC, receiver operating characteristic.

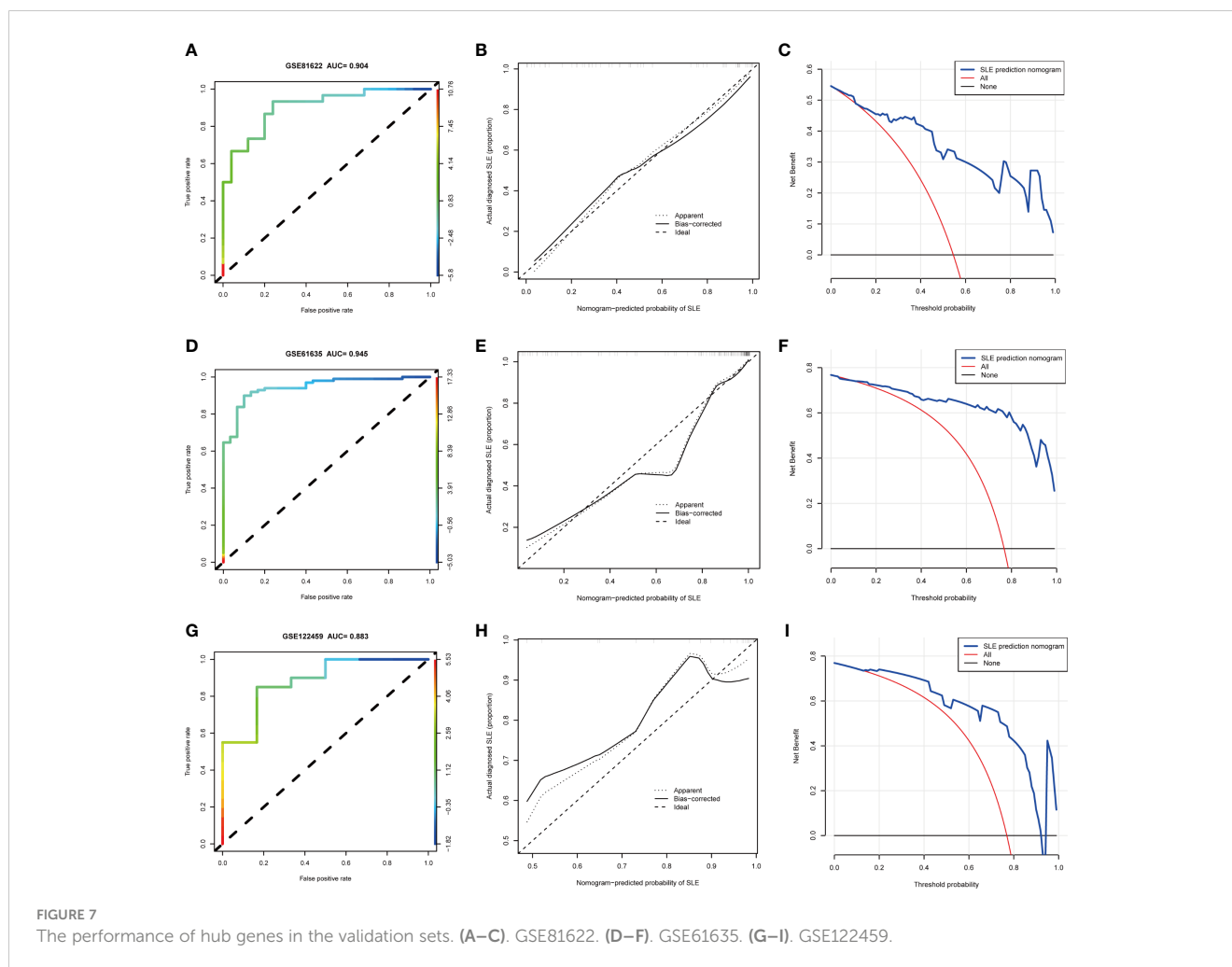
to type I interferon (GO:0034340) (Figure 9A), and the cluster 2 highly expressed DEGs were enriched in response to type I interferon (GO:0034340), response to the virus (GO:0009615), defense response to symbiont (GO:0140546), negative regulation of viral genome replication (GO:0045071), and defense response to the virus (GO:0051607) (Figure 9B), and the DEGs highly expressed in cluster 3 were enriched in CD4-positive, alpha-beta T cell cytokine production (GO:0035743), immune response-activating signal transduction (GO:0002757), immune response-activating cell surface receptor signaling pathway (GO:0002429), humoral immune response (GO:0006959), and immune response-regulating signaling pathway (GO:0002764) (Figure 9C).

The KEGG enrichment analysis revealed that cluster 1 highly expressed DEGs were mostly prevalent in Hepatitis C, Measles, Hematopoietic cell lineage, Influenza A, NOD-like receptor signaling pathway, Epstein-Barr virus infection, Prostate cancer, Amoebiasis, Osteoclast differentiation, and Insulin resistance (Figure 9D), and that of cluster 2 were mostly enriched in Hepatitis C, Influenza A, Necroptosis, Human T-cell leukemia virus 1 infection,

Phosphatidylinositol signaling system, Epstein-Barr virus infection, Measles, cGMP-PKG signaling pathway, Neutrophil extracellular trap formation, and Viral carcinogenesis (Figure 9E). Cluster 3 highly expressed DEGs were mostly associated with Staphylococcus aureus infection, Hematopoietic cell lineage, production of intestinal immune network for IgA, Starch and sucrose metabolism, Human T-cell leukemia virus 1 infection, Inflammatory bowel disease, Renin-angiotensin system, Fluid shear stress and atherosclerosis, Asthma, and Longevity regulating pathway (Figure 9F).

Differentiation of immune infiltration characteristics among the sub-clusters

The immune infiltration analysis indicated the presence of an altered immune microenvironment between the sub-clusters (Figure 9G). Cluster1 exhibited the highest proportions of macrophages and neutrophils, whereas pDCs, aDCs, and CD8+ T, B, Th1, T helper, Th2, and TIL cells were abundant in Cluster3.



Discussion

SLE is an autoimmune disorder comprising a complex network of immune-inflammatory pathological mechanisms. The NETs functions have been associated with the development and progression of SLE, however, its potential mechanisms that contribute to SLE at the genetic level have not been investigated. This investigation explores the molecular characteristics of NRGs in SLE based on bioinformatics analyses and machine learning, and identify three reliable biomarkers and three molecular clusters of SLE.

Here GEO database was utilized to elucidate the gene expression levels of healthy controls and SLE patients, and 1006 DEGs, including 393 upregulated and 613 downregulated genes were identified. The DEGs were then annotated using the research results on function-related enrichment. As results indicate, these genes were most prevalent in different viral infections and their associated with the immune response pathways. Previous research has revealed that pathogens, especially viruses, are important environmental triggers of SLE (44), which reinforce our results. Previous studies have highlighted the significance of NETs in the pathophysiology of SLE (8–10). In order to investigate the molecular properties of NRGs in SLE, the intersection of DEGs and NRGs was carried out, and 8 DE-NRGs were subsequently

obtained. Further analysis of these 8 DE-NRGs revealed that most of them had significant co-expression tendencies, which indicated that these genes may work together in the pathogenesis of SLE. Notably, the PPI analysis revealed that the DE-NRGs interacting genes were mainly involved in the PID_INTEGRIN2_PATHWAY. The Beta2-integrins, comprising a common Beta2 (CD18) subunit complexed with rare subunits (CD11a-d), including LFA-1 (leukocyte function antigen 1, CD11a/CD18), Mac-1 (CD11b/CD18, aMb2 or complement receptor 3), alphaXbeta2 (CD11c/CD18) and alphaDbeta2 (CD11d/CD18), have significant activity in immigration by endothelial and chemotaxis, leukocyte adhesion, and immune and inflammatory reactions (45). Among them, LFA-1 and Mac-1 are the most studied and have been reported to be linked with SLE pathogenesis (46, 47). For instance, loss of LFA-1 activity protects against the development of lupus in mouse models whereas its overexpression causes lupus-like syndrome (48). Furthermore, in a spontaneous SLE model, the MRL/lpr mice, Mac-1 reduction causes severe glomerulonephritis that is related to an increase in neutrophil infiltration in the kidneys (49). However, how these genes contribute to SLE and its clear mechanisms need to be discussed with functional studies in the future.

With the rapid advancements of artificial intelligence (AI), machine learning algorithm, as an important branch, can better

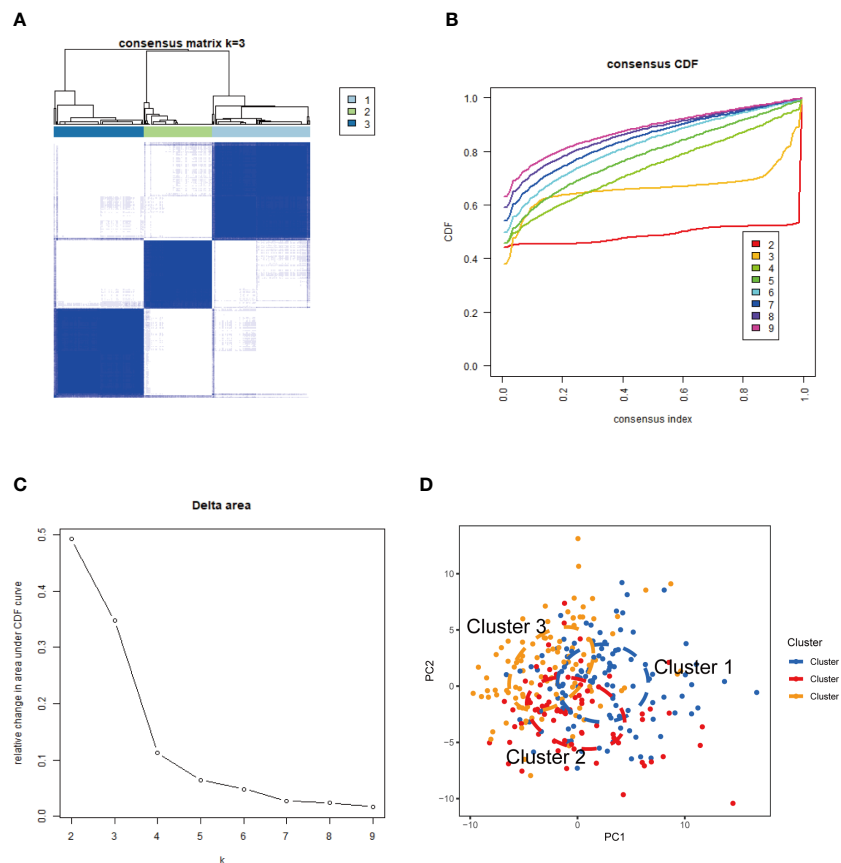


FIGURE 8

NETs-related molecular clusters identification in SLE. (A) Consensus clustering matrix when $k = 3$; (B) Consensus clustering CDF for $k = 2-9$. (C) Relative change in area under CDF curve for $k = 2-9$. (D) PCA was used to verify the three distinct subgroups divided by consensus clustering. SLE, systemic lupus erythematosus; CDF, cumulative distribution function; PCA, principal component analysis.

discriminate and have higher-dimensional feature data, and has been widely used for hub gene identification and screening (50). In this investigation, the predictive performance of three selected machine learning classifiers (RF, SVM, and LASSO) was integrated based on DE-NRGs expression profiles. Finally, CREB5, HMGB1, and ITGB2 were selected as hub genes, whose diagnostic values for SLE have been validated in the training and three validation sets.

CREB5 (cAMP Responsive Element Binding Protein 5), belongs to the CRE (cAMP response element)-binding protein family and encodes a transcription activator in eukaryotic cells (51). CREB5-associated pathways include PI3K-Akt and Toll-like receptor signaling pathways (52). The PI3K-AKT pathway regulates various physiological responses and is critical for lymphocyte development and optimal immune responses (53, 54). In the lupus MRL/lpr mouse model, it was found that excessive PI3K/Akt pathway activation leads to the severity of inflammation and fibrosis in the kidney while its inhibition reduces glomerulonephritis and prolonged life span (55-57). The toll-like receptors (TLRs) respond to molecular patterns associated with microbes such as bacteria and viruses, stimulating innate and acquired immunity as well as an inflammatory response and cytokines upregulation (58). Several studies suggest that the signaling pathway has essential activity in the pathogenesis of

autoimmune diseases including SLE (59, 60). Therefore, we speculate that CREB5 may participate in the pathogenesis of SLE with the involvement of PI3K-Akt and Toll-like receptor signaling pathways and serve as a potential therapeutic target. However, the relevance of CREB5 in SLE has not been documented and further research is required. HMGB1 (High Mobility Group Box 1), encodes a protein that belongs to the High Mobility Group-box superfamily and exerts pro-inflammatory effects in SLE through both innate and adaptive immunity (61, 62). SLE individuals have aberrant apoptotic debris removal, and apoptotic or necrotic cells release HMGB1 as a damage-associated molecular pattern (DAMP), causing tissue and organ damage and dysfunction (61). HMGB1 has been associated with multiple disease phenotypes in SLE, including lupus nephritis, neuropsychiatric lupus, and skin lesions (61, 63). ITGB2 (Integrin Subunit Beta 2) gene encodes an integrin beta chain (CD18) and combines with multiple different alpha chains to form different Beta2-integrins (45), as mentioned earlier. Thus, the NETs hub genes have potential roles in promoting the occurrence and development of SLE, and some specific pathways can act as therapeutic targets for controlling SLE.

In addition, unsupervised cluster analysis illustrated different NETs regulation patterns in SLE patients based on hub gene expression, and three distinct clusters were identified. Interestingly,

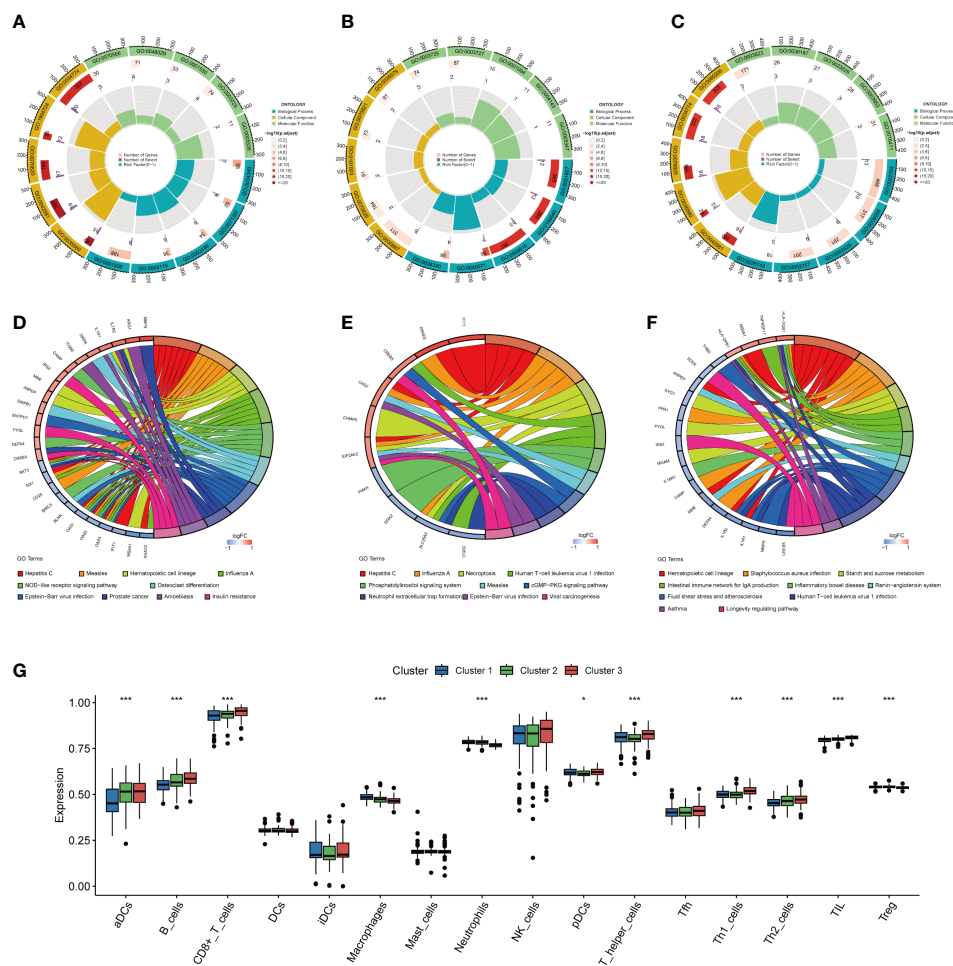


FIGURE 9 Functional enrichment and immune cell infiltration analyses among three NETs subgroups. (A–C). GO enrichment analysis of DEGs highly expressed in cluster 1 (A), cluster 2 (B), and cluster 3 (C). (D–F). KEGG enrichment analysis of DEGs highly expressed in cluster 1 (D), in cluster 2 (E), and in cluster 3 (F). (G) Box plot showing the relative abundances of immune cells among three clusters, * $P < 0.05$, *** $P < 0.001$. NETs, neutrophil extracellular traps; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes.

functional enrichment was performed among three NETs subgroups, and the data showed that the DEGs highly expressed in cluster 1 were enriched in innate immune response pathways, such as neutrophil activation, neutrophil-mediated immunity, response to type I interferon, and so on, whereas that of cluster 3 were enriched in adaptive immune response pathways. Moreover, immune infiltration analysis also revealed that innate immune cells including macrophages and neutrophils were significantly upregulated in cluster 1, in contrast, cluster 3, which was dominated by upregulated T and B cells of the adaptive immunity. It is well known that SLE is characterized by its clinical and therapy heterogeneity, with a wide range of clinical manifestations and drug responses reflecting its complex etiopathogenesis, which means that stratified management planned accordingly may provide better outcomes. Consequently, the data of this investigation may provide potential biological insights into the different clinical phenotypes and a rationale for the stratified therapy of patients.

This investigation has several limitations that should be taken into consideration. Firstly, the study only relies on bioinformatics analyses, and further experimental validation is needed to confirm the results. Additionally, the findings were obtained from a relatively small sample size of SLE patients, and larger cohorts are necessary to provide more robust and reliable results. Moreover, the SLE diagnostic model developed in this study requires further evaluation and external validation before its clinical application can be considered. Finally, this investigation only focused on gene expression data, and future studies should also explore epigenetic, proteomic, and metabolomic alterations in SLE pathogenesis.

In conclusion, this research is the first as per our knowledge to explore the molecular characteristics of NRGs in SLE, identify three potential biomarkers, HMGB1, ITGB2, and CREB5, and three distinct clusters based on these hub biomarkers. These findings may assist in diagnosing and managing SLE with the ultimate objective of improving patient outcomes.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

Author contributions

XD and HL designed the study. HL, XZ, JS, XF, LY, JF, JR, and RZ performed data analysis. HL drafted the manuscript. XD revised the manuscript. All authors read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grants from the National Key Research and Development Program of China (2021YFC2501304), the Science and Technology Program of Department of Health of Jiangxi Province (20204254), and the Key Research and Development Program of Jiangxi municipal Science and Technology Department (20192BBGL70024). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Dörner T, Furie R. Novel paradigms in systemic lupus erythematosus. *Lancet* (2019) 393(10188):2344–58. doi: 10.1016/s0140-6736(19)30546-x
- Pons-Estel GJ, Ugarte-Gil MF, Alarcón GS. Epidemiology of systemic lupus erythematosus. *Expert Rev Clin Immunol* (2017) 13(8):799–814. doi: 10.1080/1744666x.2017.1327352
- Durcan L, O'Dwyer T, Petri M. Management strategies and future directions for systemic lupus erythematosus in adults. *Lancet* (2019) 393(10188):2332–43. doi: 10.1016/s0140-6736(19)30237-5
- Tektonidou MG, Lewandowski LB, Hu J, Dasgupta A, Ward MM. Survival in adults and children with systemic lupus erythematosus: a systematic review and Bayesian meta-analysis of studies from 1950 to 2016. *Ann Rheum Dis* (2017) 76(12):2009–16. doi: 10.1136/annrheumdis-2017-211663
- Yen EY, Shaheen M, Woo JMP, Mercer N, Li N, McCurdy DK, et al. 46-year trends in systemic lupus erythematosus mortality in the united states, 1968 to 2013: a nationwide population-based study. *Ann Intern Med* (2017) 167(11):777–85. doi: 10.7326/m17-0102
- Lee YH, Choi SJ, Ji JD, Song GG. Overall and cause-specific mortality in systemic lupus erythematosus: an updated meta-analysis. *Lupus* (2016) 25(7):727–34. doi: 10.1177/0961203315627202
- 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
- Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. *Science* (2004) 303(5663):1532–5. doi: 10.1126/science.1092385
- Brinkmann V, Zychlinsky A. Beneficial suicide: why neutrophils die to make NETs. *Nat Rev Microbiol* (2007) 5(8):577–82. doi: 10.1038/nrmicro1710
- Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol* (2018) 18(2):134–47. doi: 10.1038/nri.2017.105
- Wang M, Ishikawa T, Lai Y, Nallapothula D, Singh RR. Diverse roles of NETosis in the pathogenesis of lupus. *Front Immunol* (2022) 13:895216. doi: 10.3389/fimmu.2022.895216

Acknowledgments

We thank Editideas (www.editideas.cn) for its linguistic assistance during the preparation of this manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1150828/full#supplementary-material>

- Wigerblad G, Kaplan MJ. Neutrophil extracellular traps in systemic autoimmune and autoinflammatory diseases. *Nat Rev Immunol* (2022) 18:1–15. doi: 10.1038/s41577-022-00787-0
- Jariwala MP, Laxer RM. NETosis in rheumatic diseases. *Curr Rheumatol Rep* (2021) 23(2):9. doi: 10.1007/s11926-020-00977-6
- Mutua V, Gershwin LJ. A review of neutrophil extracellular traps (NETs) in disease: potential anti-NETs therapeutics. *Clin Rev Allergy Immunol* (2021) 61(2):194–211. doi: 10.1007/s12016-020-08804-7
- Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics* (2007) 23(14):1846–7. doi: 10.1093/bioinformatics/btm254
- Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol* (2016) 1418:93–110. doi: 10.1007/978-1-4939-3578-9_5
- Zhang Y, Guo L, Dai Q, Shang B, Xiao T, Di X, et al. A signature for pan-cancer prognosis based on neutrophil extracellular traps. *J Immunother Cancer* (2022) 10(6):e004210. doi: 10.1136/jitc-2021-004210
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics* (2007) 8(1):118–27. doi: 10.1093/biostatistics/kxj037
- Liu J, Zhou S, Li S, Jiang Y, Wan Y, Ma X, et al. Eleven genes associated with progression and prognosis of endometrial cancer (EC) identified by comprehensive bioinformatics analysis. *Cancer Cell Int* (2019) 19:136. doi: 10.1186/s12935-019-0859-1
- Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer Publishing Company, Incorporated. (2009).
- Kolde R. Pheatmap: pretty heatmaps. (2015).
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2(3):100141. doi: 10.1016/j.xinn.2021.100141
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Gene Ontology Consortium. Nat Genet* (2000) 25(1):25–9. doi: 10.1038/75556

24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27
25. Walter W, Sánchez-Cabo F, Ricote M. GPlot: an R package for visually combining expression data with functional analysis. *Bioinformatics* (2015) 31(17):2912–4. doi: 10.1093/bioinformatics/btv300
26. Zhang H, Meltzer P, Davis S. Rcircos: an R package for circos 2D track plots. *BMC Bioinf* (2013) 14:244. doi: 10.1186/1471-2105-14-244
27. Wei T, Simko V. R package “corrplot”: visualization of a correlation matrix (Version 0.84) (2017). Available at: <https://github.com/taiyun/corrplot>.
28. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* (2010) 38(Web Server issue):W214–20. doi: 10.1093/nar/gkq537
29. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6
30. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01
31. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang CC, et al. e1071: misc functions of the department of statistics, probability theory group (Formerly: E1071), TU wient. (2015).
32. Kuhn M. *Caret package*. (2008).
33. Liaw A, Wiener M. Classification and regression by randomForest. *R News* (2002) 23(23).
34. Tibshirani R. Regression shrinkage and selection Via the lasso. *J R Stat Society: Ser B (Methodological)* (1996) 58(1):267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
35. Gold C, Sollich P. Model selection for support vector machine classification. *Neurocomputing* (2003) 55(1):221–49. doi: 10.1016/S0925-2312(03)00375-8
36. Wang H, Yang F, Luo Z. An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics* (2016) 17:60. doi: 10.1186/s12859-016-0900-5
37. Harrell FE, Harrell F, Harrell FE. Rms: regression modeling strategies. In: *R package version 4.0-0* (2013).
38. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf* (2011) 12:77. doi: 10.1186/1471-2105-12-77
39. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26(12):1572–3. doi: 10.1093/bioinformatics/btq170
40. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics* (2014) 30(19):2811–2. doi: 10.1093/bioinformatics/btu393
41. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
42. Bauer S, Robinson PN, Gagneur J. Model-based gene set analysis for bioconductor. *Bioinformatics* (2011) 27(13):1882–3. doi: 10.1093/bioinformatics/btr296
43. Ggpubr: ‘ggplot2’ based publication ready plots (2017).
44. Illescas-Montes R, Corona-Castro CC, Melguizo-Rodríguez L, Ruiz C, Costela-Ruiz VJ. Infectious processes and systemic lupus erythematosus. *Immunology* (2019) 158(3):153–60. doi: 10.1111/imm.13103
45. Tan SM. The leucocyte $\beta 2$ (CD18) integrins: the structure, functional regulation and signalling properties. *Biosci Rep* (2012) 32(3):241–69. doi: 10.1042/bsr20110101
46. Fagerholm SC, MacPherson M, James MJ, Sevier-Guy C, Lau CS. The CD11b-integrin (ITGAM) and systemic lupus erythematosus. *Lupus* (2013) 22(7):657–63. doi: 10.1177/0961203313491851
47. Rosetti F, Mayadas TN. The many faces of mac-1 in autoimmune disease. *Immunol Rev* (2016) 269(1):175–93. doi: 10.1111/imr.12373
48. Kevil CG, Hicks MJ, He X, Zhang J, Ballantyne CM, Raman C, et al. Loss of LFA-1, but not mac-1, protects MRL/MpJ-fas(lpr) mice from autoimmune disease. *Am J Pathol* (2004) 165(2):609–16. doi: 10.1016/s0002-9440(10)63325-1
49. Rosetti F, Tsuboi N, Chen K, Nishi H, Hernandez T, Sethi S, et al. Human lupus serum induces neutrophil-mediated organ damage in mice that is enabled by mac-1 deficiency. *J Immunol* (2012) 189(7):3714–23. doi: 10.4049/jimmunol.1201594
50. Stafford IS, Kellermann M, Mossotto E, Beattie RM, MacArthur BD, Ennis S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ Digit Med* (2020) 3:30. doi: 10.1038/s41746-020-0229-3
51. Nomura N, Zu YL, Maekawa T, Tabata S, Akiyama T, Ishii S. Isolation and characterization of a novel member of the gene family encoding the cAMP response element-binding protein CRE-BP1. *J Biol Chem* (1993) 268(6):4259–66. doi: 10.1016/S0021-9258(18)53604-8
52. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinf* (2016) 54:30. doi: 10.1002/cpbi.5
53. Hemmings BA, Restuccia DF. PI3K-PKB/Akt pathway. *Cold Spring Harb Perspect Biol* (2012) 4(9):a011189. doi: 10.1101/cshperspect.a011189
54. Hemmings BA, Restuccia DF. The PI3K-PKB/Akt pathway. *Cold Spring Harb Perspect Biol* (2015) 7(4):a011189. doi: 10.1101/cshperspect.a026609
55. Zhao C, Gu Y, Chen L, Su X. Upregulation of FoxO3a expression through PI3K/Akt pathway attenuates the progression of lupus nephritis in MRL/lpr mice. *Int Immunopharmacol* (2020) 89(Pt A):107027. doi: 10.1016/j.intimp.2020.107027
56. Ge F, Wang F, Yan X, Li Z, Wang X. Association of BAFF with PI3K/Akt/mTOR signaling in lupus nephritis. *Mol Med Rep* (2017) 16(5):5793–98. doi: 10.3892/mmr.2017.7367
57. Stylianou K, Petrakis I, Mavroei V, Stratakis S, Vardaki E, Perakis K, et al. The PI3K/Akt/mTOR pathway is activated in murine lupus nephritis and downregulated by rapamycin. *Nephrol Dial Transplant* (2011) 26(2):498–508. doi: 10.1093/ndt/gfq496
58. Kawasaki T, Kawai T. Toll-like receptor signaling pathways. *Front Immunol* (2014) 5:461. doi: 10.3389/fimmu.2014.00461
59. Chen JQ, Szodoray P, Zeher M. Toll-like receptor pathways in autoimmune diseases. *Clin Rev Allergy Immunol* (2016) 50(1):1–17. doi: 10.1007/s12016-015-8473-z
60. Baek WY, Choi YS, Lee SW, Son IO, Jeon KW, Choi SD, et al. Toll-like receptor signaling inhibitory peptide improves inflammation in animal model and human systemic lupus erythematosus. *Int J Mol Sci* (2021) 22(23):12764. doi: 10.3390/ijms222312764
61. Liu T, Son M, Diamond B. HMGB1 in systemic lupus erythematosus. *Front Immunol* (2020) 11:1057. doi: 10.3389/fimmu.2020.01057
62. Schaper F, Westra J, Bijl M. Recent developments in the role of high-mobility group box 1 in systemic lupus erythematosus. *Mol Med* (2014) 20(1):72–9. doi: 10.2119/molmed.2014.00019
63. Abdulahad DA, Westra J, Limburg PC, Kallenberg CG, Bijl M. HMGB1 in systemic lupus erythematosus: its role in cutaneous lesions development. *Autoimmun Rev* (2010) 9(10):661–5. doi: 10.1016/j.autrev.2010.05.015