Scientific Reports

https://doi.org/10.1038/s41598-026-41028-x

Article in Press

# Machine learning-driven discovery of *STAT1* and *TRIM22* as immune biomarkers for lupus nephritis: translational insights into diagnosis and pathogenesis

Received: 9 September 2025

Accepted: 17 February 2026

Published online: 20 February 2026

Cite this article as: Deng J., Zhang Z., Lai Y. *et al.* Machine learning-driven discovery of *STAT1* and *TRIM22* as immune biomarkers for lupus nephritis: translational insights into diagnosis and pathogenesis. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-41028-x

Jiayi Deng, Zimiao Zhang, Yueyuan Lai, Jinpeng Chen, Xiaomei Ma, Zhenkai Gao, Chao Lin, Xiaohong Li, Weihao Wu, Congjie Chen, Xiaohui Shangguan, Yanhong Huang, Haoran Qiu, Xiaoming Qiu & Longtian Chen

# Machine Learning-Driven Discovery of *STAT1* and *TRIM22* as Immune Biomarkers for Lupus Nephritis: Translational Insights into Diagnosis and Pathogenesis

Jiayi Deng [1,†], Zimiao Zhang[1,†], Yueyuan Lai[1], Jinpeng Chen[1], Xiaomei Ma[1], Zhenkai Gao[2], Chao Lin[1], Xiaohong Li[1], Weihao Wu[1], Congjie Chen[1], Xiaohui Shangguan[1], Yanhong Huang[1], Haoran Qiu[1], Xiaoming Qiu[1,*], Longtian Chen[1,*]

[1] Longyan First Affiliated Hospital of Fujian Medical University, Longyan, 364000, China

[2] The Third Hospital of Longyan, Longyan, 364000, China

† These authors contributed equally to this work.

* Corresponding authors

Correspondence:

**Primary contact:** Jiayi Deng, MD

Department of Pathology, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, 364000, China

Tel: (+86)05972958989 / Fax: (+86)05972292374

E-mail: 981164933@qq.com

**Co-correspondence may also be addressed to:**

Xiaoming Qiu, MD, E-mail: fjqiuxiaoming@163.com

Longtian Chen, MD, E-mail: ltchen@163.com

**Running title: Immune Biomarkers *STAT1* and *TRIM22* in Lupus Nephritis.**

## Abstract

**Background:** Lupus nephritis (LN) is a severe manifestation of systemic lupus erythematosus and a major cause of renal dysfunction, while reliable non-invasive biomarkers remain limited.

**Methods:** Transcriptomic data from three LN cohorts were analyzed to identify differentially expressed genes (DEGs). Immune-associated DEGs were selected using WGCNA and prioritized via multiple machine learning algorithms. Diagnostic performance was evaluated with ROC curves and nomogram modeling, accompanied by functional enrichment and immune infiltration analyses. Independent validation was performed by qRT-PCR in peripheral blood samples from 13 LN patients and 10 healthy controls.

**Results:** A total of 320 DEGs were identified, including 53 linked to immune processes. In the transcriptomic datasets, four candidate hub genes (*CD40LG, RETN, TRIM22, STAT1*) were initially identified. Furthermore, immune infiltration analysis suggested gene-specific

immune interaction patterns, particularly associating *TRIM22* with CD4⁺ T-cell–related signatures. qRT-PCR confirmed upregulation of *STAT1* and *TRIM22*, while *RETN* and *CD40LG* showed no significant elevation. Accordingly, a refined two-gene signature was constructed, showing consistent discriminatory trends in the training dataset and the clinical validation cohort (AUCs > 0.9).

**Conclusion:** *STAT1* and *TRIM22* were consistently upregulated in the peripheral blood of patients with lupus nephritis and may represent potential immune-related biomarkers.

**Keywords:** Lupus nephritis; *STAT1*; *TRIM22*; immune biomarkers; machine learning.



## Introduction

Systemic lupus erythematosus (SLE) is a complex autoimmune disease characterized by its multisystem involvement, posing significant challenges in clinical diagnosis and treatment[1]. The hallmark of SLE is an aberrant immune response targeting the body's own tissues, leading to damage in organs such as the skin, joints, kidneys, and heart[2]. Among these, lupus nephritis (LN) stands out as one of the most common and severe manifestations of SLE[3]. Approximately 40-60% of SLE patients develop LN, and the disease's prevalence and prognosis are closely linked, making LN a major

contributor to morbidity, mortality, and the progression to end-stage renal disease [4,5]. While immunosuppressants and glucocorticoids have shown some success in controlling disease activity, the long-term prognosis for LN remains unsatisfactory, necessitating further research to improve diagnostic and therapeutic strategies[6].

The pathogenesis of LN is multifactorial, involving immune complex deposition, autoantibody production, and immune dysregulation[7]. Dysregulated B and T cell responses, along with an imbalance in both innate and adaptive immunity, are key drivers of renal damage in LN[8]. Moreover, there is substantial immune cell infiltration in the kidneys of LN patients, particularly the accumulation of T cells, B cells, and macrophages, which exacerbate structural kidney damage by releasing pro-inflammatory cytokines and promoting tissue inflammation[9-11].



Currently, the diagnosis of LN heavily relies on renal biopsy, a method that is invasive and impractical for repeated sampling[12]. Although recent studies have explored non-invasive diagnostic approaches through the detection of serum and urinary biomarkers, none have yet been validated with sufficient sensitivity and specificity for clinical use[13,14]. Additionally, existing treatment regimens are limited by the severe side effects of long-term immunosuppressant and glucocorticoid use, with some patients

experiencing resistance or relapse[15]. Therefore, novel diagnostic and therapeutic strategies, particularly those based on molecular mechanisms of immune regulation, are urgently needed.

In recent years, advances in bioinformatics and high-throughput sequencing technologies have facilitated a deeper understanding of the molecular mechanisms underlying LN[16]. By integrating gene expression profiles and multi-omics data, researchers have identified differentially expressed genes (DEGs) closely associated with LN pathogenesis, particularly immune-related key genes[17]. These immune-related genes (IRGs) hold promise for early diagnosis and personalized therapy in LN[18].

The aim of this study is to use bioinformatics approaches to identify differentially expressed IRGs in LN patients by leveraging multi-omics data and to apply machine learning techniques to screen for potential key genes. By further analyzing the relationship between these genes and immune cell infiltration in LN, we hope to elucidate their potential applications in LN diagnosis and treatment.

## Materials and methods

### Data acquisition

The research flow diagram is presented in Figure 1. Three datasets related to LN, including clinical characteristics and gene

expression profiles, were obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The GSE72326 dataset consisted of 20 normal and 24 LN blood samples. The GSE99967 dataset included 29 LN and 17 normal blood samples[19]. These two datasets were combined after batch correction using the "sva" package to generate the training set[20]. The GSE81622 dataset, comprising 25 normal and 15 LN blood samples, was utilized for validation[21]. A total of 2,483 IRGs were retrieved from the ImmPort database (https://www.immport.org/home)[22] and included in Supplementary Table 1.

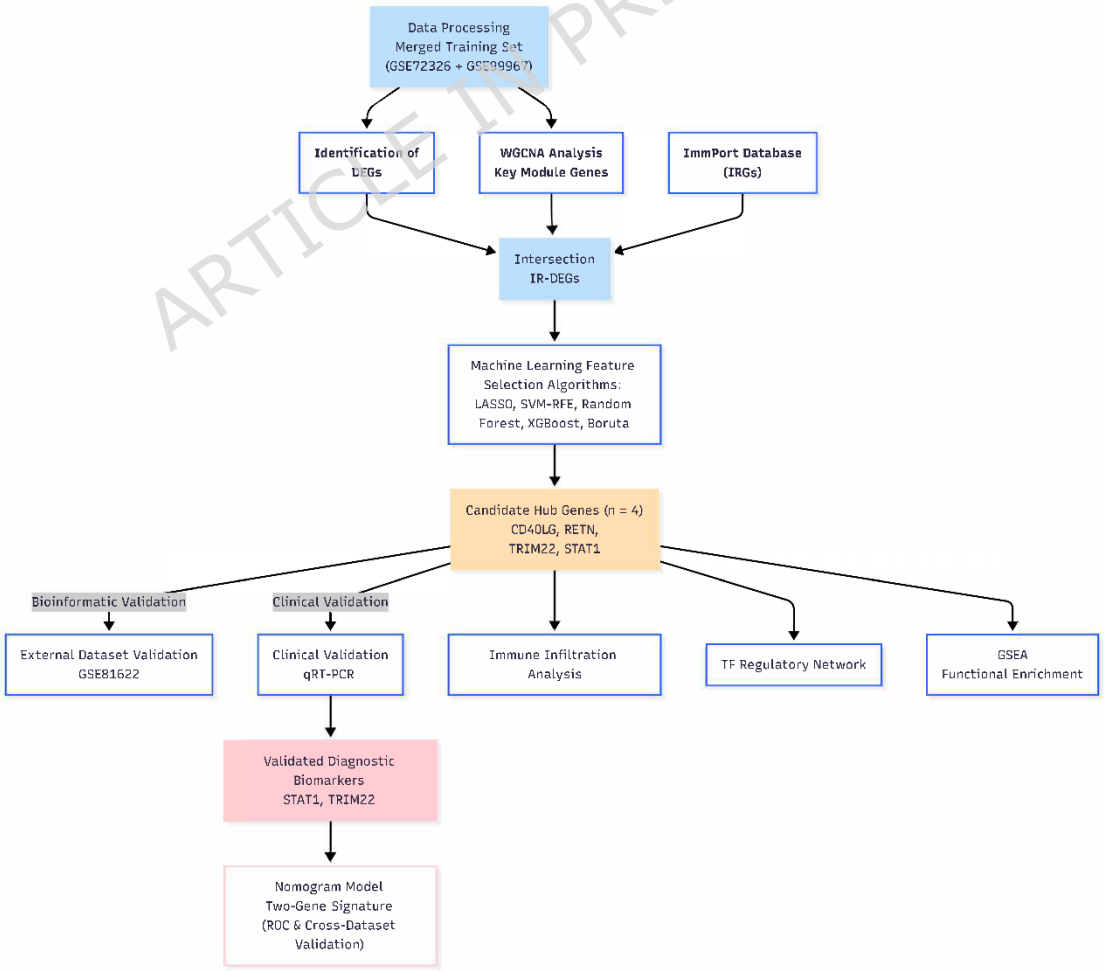


**Figure 1. The flow diagram of the research design.**

## Differential Expression and WGCNA

Differentially expressed genes (DEGs) between LN and control samples were identified using the “limma” R package with the criteria |Log2FC| > 0.585 and FDR < 0.05. Weighted gene co-expression network analysis (WGCNA) was performed using the “WGCNA” package[23]. Sample clustering was conducted via hierarchical clustering, excluding outliers with clustering height >70. The optimal soft-thresholding power was chosen to achieve scale-free topology. Modules were identified using the dynamic tree-cutting algorithm (minimum module size = 50) and merged if correlation >0.25. Modules significantly associated with LN (|correlation| >0.4, $p < 0.05$) were defined as key modules.

## Identification of Immune-Related DEGs

Immune-related DEGs (IR-DEGs) were defined as the intersection of DEGs, key module genes, and IRGs. Venn diagrams were used for visualization.

## Machine Learning for Biomarker Identification

Five machine learning algorithms—LASSO (glmnet)[24], Random Forest (randomForest)[25], Support Vector Machine (SVM, e1071)[26], Extreme Gradient Boosting (XGBoost, xgboost)[27], and Boruta (Boruta)[28]—were applied to identify signature genes. Genes

consistently selected across all methods were considered potential diagnostic biomarkers.

**Diagnostic Performance Evaluation**

Receiver operating characteristic (ROC) curves were generated using the “pROC” package to evaluate the diagnostic performance of candidate biomarkers[29]. Biomarkers with high AUC values in both the training and validation sets were considered key LN diagnostic genes. A diagnostic nomogram based on these biomarkers was constructed using the “rms” package and assessed by calibration plots and ROC curves [30].

To ensure the generalizability of the diagnostic tool and address potential overfitting, a leave-dataset-out validation was conducted as a sensitivity analysis. This step specifically focused on the simplified two-gene signature (*STAT1* and *TRIM22*), which was prioritized due to its superior consistency across multiple machine learning algorithms. The cross-validation involved training the model on one cohort and testing it on an independent cohort to verify the stability of the diagnostic signature across different datasets.

**Gene Set Enrichment and Immune Infiltration Analysis**

Gene set enrichment analysis (GSEA) was conducted with the “clusterProfiler” package to identify immune-related KEGG pathways associated with key biomarkers ($p$.adjust < 0.05). Principal

component analysis (PCA) was applied to visualize clustering of samples. Immune cell infiltration was evaluated using single-sample GSEA (ssGSEA) via the "GSVA" package based on a predefined immune cell marker gene set. Heatmaps and boxplots were generated with "pheatmap" and "ggpubr". Spearman correlation analyses between candidate biomarkers (*CD40LG, RETN, TRIM22, STAT1*) and immune cell types were performed using the "psych" package with FDR correction.

**Construction of TF-Gene Regulatory Network**

Transcription factors (TFs) associated with *CD40LG, RETN, TRIM22,* and *STAT1* were predicted using the NetworkAnalyst platform. Experimentally validated TF binding profiles from the JASPAR database were used to construct TF-gene interactions, which were visualized using Cytoscape (version 3.10.2).

**Clinical Characteristics and Sample Collection**

A total of 13 patients with LN and 10 age- and sex-matched healthy controls (HCs) were recruited from Longyan First Hospital (Fujian, China) under ethical approval (No. LYREC2024-k172-01) with written informed consent. LN diagnosis was established according to the International Society of Nephrology/Renal Pathology Society (ISN/RPS) criteria. Patients with malignancy, active infection, or other autoimmune diseases (e.g., rheumatoid

arthritis or systemic sclerosis) were excluded. Clinical and laboratory data, including age, sex, renal SLEDAI, serum creatinine, eGFR, 24-hour urine protein, hematuria, urinary casts, urine white blood cell count, anti-dsDNA, complement C3/C4 levels, ongoing treatment, and renal biopsy classification, were obtained from electronic medical records at the time of sampling.

Peripheral blood samples were collected from all participants into EDTA tubes, immediately mixed with TRIzol®-like reagent (CWBIO, CW0580S, China) at a 1:3 (v/v) ratio, flash-frozen in liquid nitrogen, and stored at −80°C until RNA extraction.

**RNA Extraction and qRT-PCR**

Total RNA was extracted from PBMCs using TRIzol-chloroform combined with UltraPure RNA column purification (CWBIO, CW0581M, China). RNA purity and integrity were assessed with a NanoPhotometer NP80 (Implen, Germany; OD260/280: 1.8–2.2) and 1% agarose gel electrophoresis. Genomic DNA was removed using 4×gDNA Wiper Mix (HiScript II, Vazyme, R223-01, China). cDNA synthesis was performed using HiScript II SuperMix in a 20 μL reaction at 50°C for 15 min, followed by heat inactivation at 85°C for 5 s. qRT-PCR was performed on a CFX Connect™ system (Bio-Rad, USA) using SYBR Master Mix (CWBIO, CW3360M, China) with the following cycling conditions: 95°C for 30 s, 40 cycles of 95°C for 10

s, 58°C for 30 s, 72°C for 30 s, and a melt curve from 65–95°C at 0.5°C increments. To ensure analytical specificity, particularly for the 317 bp STAT1 amplicon, PCR products were strictly verified by the presence of a single peak in melt curve analysis and visualized via agarose gel electrophoresis to rule out non-specific amplification. β-actin was used as an internal control, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method (primers in Table 1).

**Table 1. Primer sequences used for qRT-PCR.**

| Primer name | Sequence (5′–3′) | Product size (bp) | Annealing temperature (°C) |
|---|---|---|---|
| β-actin F | TGGCACCCAGCACAATGAA | 186 | 58.0 |
| β-actin R | CTAAGTCATAGTCCGCCTAGAAGCA | | |
| CD40LG F | GAGCAACAACTTGGTAACCCT | 134 | 58.0 |
| CD40LG R | GGCTGGCTATAAATGGAGCTTG | | |
| RETN F | CTGTTGGTGTCTAGCAAGACC | 104 | 58.0 |
| RETN R | CCAATGCTGCTTATTGCCCTAAA | | |
| TRIM22 F | CTGTCCTGTGTGTCAGACCAG | 102 | 58.0 |
| TRIM22 R | TGTGGGCTCATCTTGACCTCT | | |
| STAT1 F | GTTTCATTTGCCACCATCCG | 317 | 58.0 |
| STAT1 R | CACATAACCTTGTCCTTCACATTTC | | |



## Statistical Analyses

All statistical analyses were performed using R software (version 4.3.1) and GraphPad Prism 9.0 (GraphPad Software, USA). For bioinformatics data, differential expression was analyzed using

empirical Bayes statistics (moderated t-tests) within the 'limma' package. For clinical validation data, given the relatively small sample size, continuous variables were compared using the non-parametric Mann-Whitney U test (Wilcoxon rank-sum test) to avoid assumptions of normality. Categorical variables were compared using Fisher's exact test. Multiple group comparisons were assessed using the Kruskal–Wallis test. Associations between variables were evaluated using Spearman correlation analysis. ROC curves were generated to evaluate the diagnostic performance of candidate genes, and the AUC was calculated using logistic regression models. All statistical tests were two-tailed, and a $P$ value $< 0.05$ was considered statistically significant.



## Results

### Identification of differentially expressed genes in LN

According to PCA results, the merged data set (training set) has eliminated the batch effect (Supplementary Figure 1A, B). As shown in Figures 2A, B, we obtained the 320 DEGs between normal and LN groups, including 61 down-regulated and 259 up-regulated genes (Supplementary Table 2).

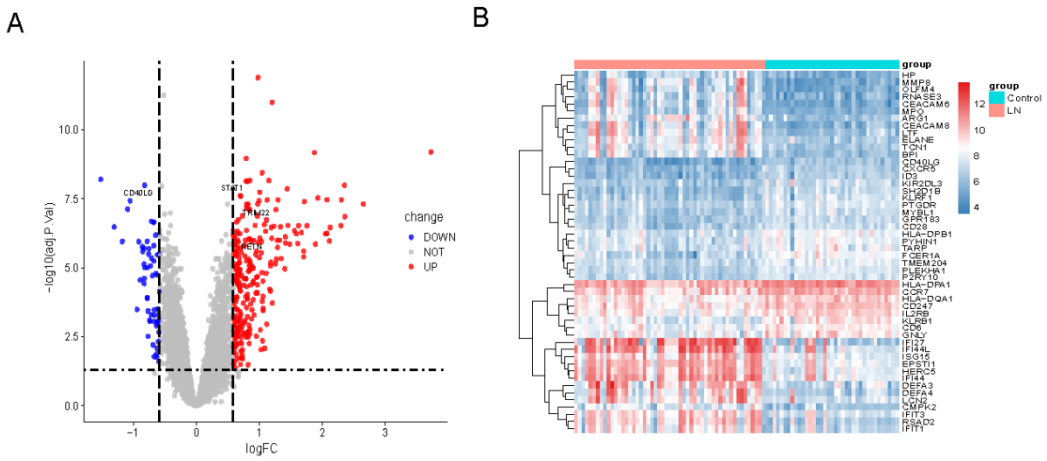



**Figure 2. Identification of differentially expressed genes in LN.**

(A) Volcano plot of DEGs between LN and normal controls. (B) Heatmap showing the TOP50 DEGs.

## Identification of Key Module Genes and IR-DEGs in LN

To identify genes associated with LN, WGCNA was performed. Hierarchical clustering of samples identified one outlier, which was removed before network construction (Supplementary Figure 2A–B). Several modules were significantly correlated with disease status (Figure 3A–C). Integrative analysis combining DEGs, key module genes from WGCNA, and IRG from the ImmPort database identified a set of 53 IR-DEGs (Figure 3D–E), which may represent promising candidates for further diagnostic and therapeutic investigation in LN.

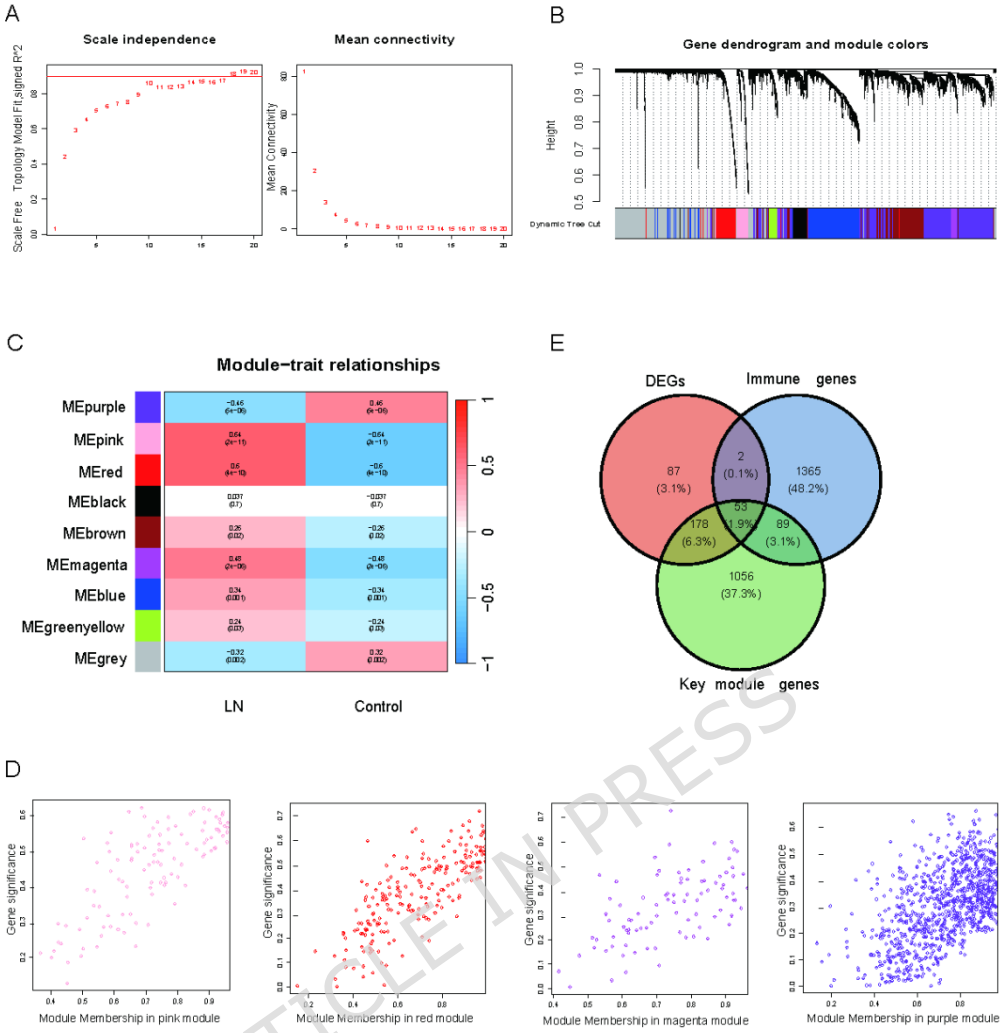



**Figure 3. Identification of key module genes and IR-DEGs in LN.**

(A) Scale-free topology fit and mean connectivity for different soft-thresholding powers. (B) Hierarchical clustering dendrogram of genes. (C) Module–trait relationships. (D) Scatterplots of gene significance versus module membership. (E) Venn diagram showing the intersection of DEGs, WGCNA module genes, and immune-related genes.

## Construction and Evaluation of the IR-DEGs Signature for LN

To further identify key immune-related genes in LN, multiple

feature selection methods were applied to the 53 IR-DEGs. First, the Boruta algorithm was used to assess feature importance, confirming several key predictors, including *CD40LG, RETN, TRIM22,* and *STAT1* (Figure 4A–B). Next, SVM-RFE analysis was performed to select an optimal subset of genes (Figure 4C), and LASSO regression provided a coefficient profile with cross-validation to identify candidate biomarkers associated with LN immune response (Figure 4D–E). Additionally, Random Forest and XGBoost analyses ranked genes based on importance (Figure 4F–G). By integrating results from all five methods, four characteristic hub genes—*CD40LG, RETN, TRIM22,* and *STAT1*—were identified as robust candidates for diagnostic and therapeutic exploration in LN (Figure 4H).

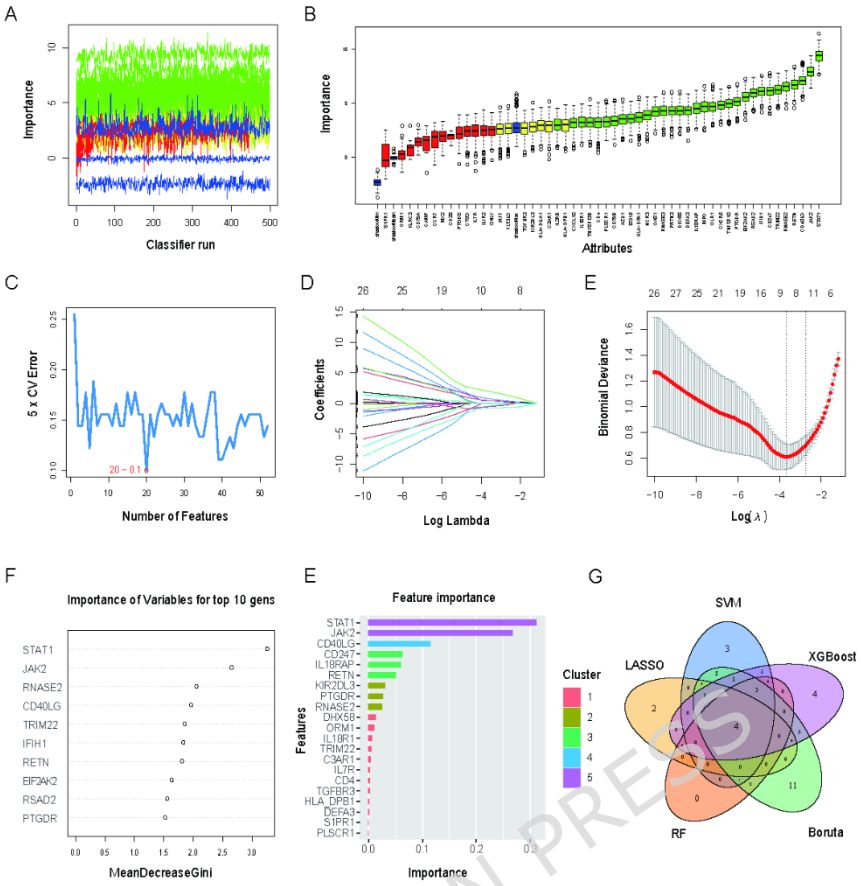



**Figure 4. Screening immune-related biomarkers of LN.**

(A) Z-score variation during Boruta feature selection. (B) Boruta feature selection plot. (C) SVM-RFE analysis. (D) LASSO regression coefficient profiles. (E) Cross-validation of the LASSO model using 10-fold cross-validation. The dotted vertical lines represent the optimal $\boldsymbol{\lambda}$ value (minimum criteria) and the 1-standard error criteria. (F) Random Forest feature importance ranking. (G) XGBoost feature importance ranking. (H) Venn diagram showing overlapping genes identified by multiple machine learning methods.

## Diagnostic Validation and Predictive Modeling of Hub IR-DEGs in LN

The diagnostic potential of the four hub IR-DEGs—*STAT1, RETN,*

*CD40LG,* and *TRIM22*—was first evaluated individually in the training cohort. Each gene demonstrated good discrimination between LN patients and healthy controls (Figure 5A), and these findings were confirmed in an independent validation cohort (Figure 5B). Comparative expression analysis showed consistent upregulation of these genes in LN across cohorts (Figures 5C–D).

Based on the training data, we initially constructed a preliminary four-gene nomogram (Supplementary Figure 3) which showed high theoretical discrimination. However, to ensure the robustness of the diagnostic model for clinical application, we prioritized experimental validation to determine the final diagnostic panel.

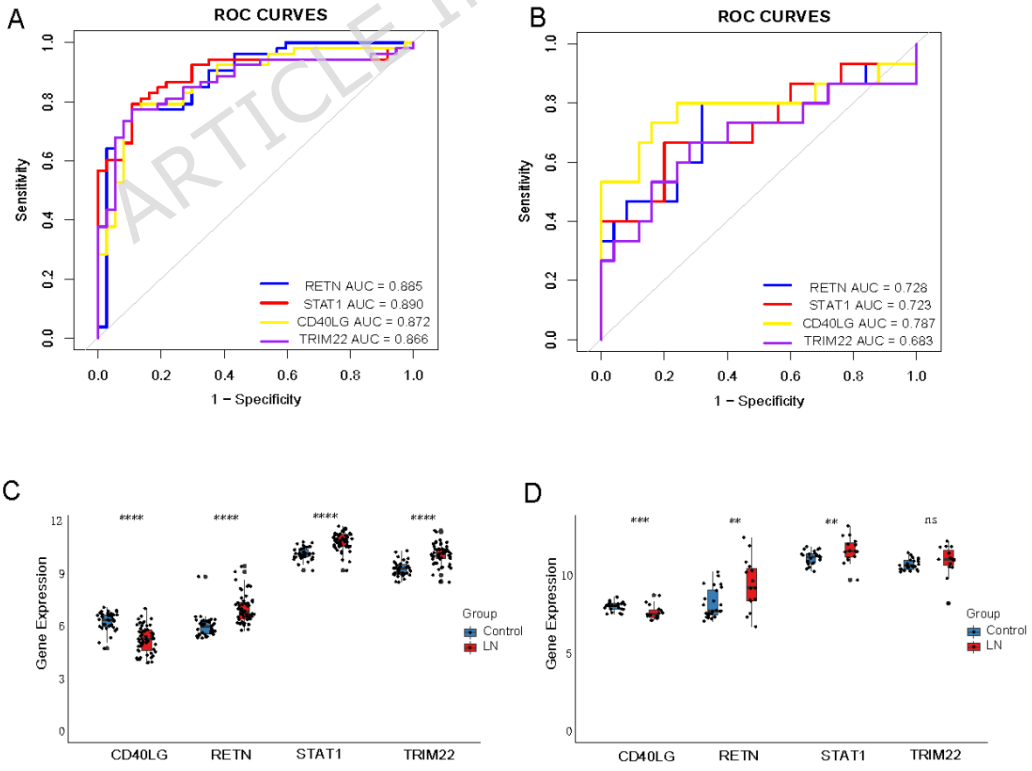



**Figure 5. Diagnostic performance and expression validation of hub IR-**

**DEGs in LN.**

(A, B) ROC curves of *CD40LG*, *RETN*, *TRIM22*, and *STAT1* in the training and independent validation cohorts. (C, D) Expression patterns of these four genes in the training and validation cohorts. Statistical comparisons were performed using the Mann-Whitney U test (Wilcoxon rank-sum test). ns, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

## Functional Enrichment Analysis of Biomarkers in LN

Single-gene GSEA indicated that the four hub biomarkers were significantly associated with immune-related and inflammatory pathways, including antigen presentation, complement and coagulation cascades, NK cell-mediated cytotoxicity, T cell receptor signaling, and graft-versus-host and autoimmune responses (Figure 6A–D). These results suggest that these genes play pivotal roles in immune regulation and inflammatory processes in LN.

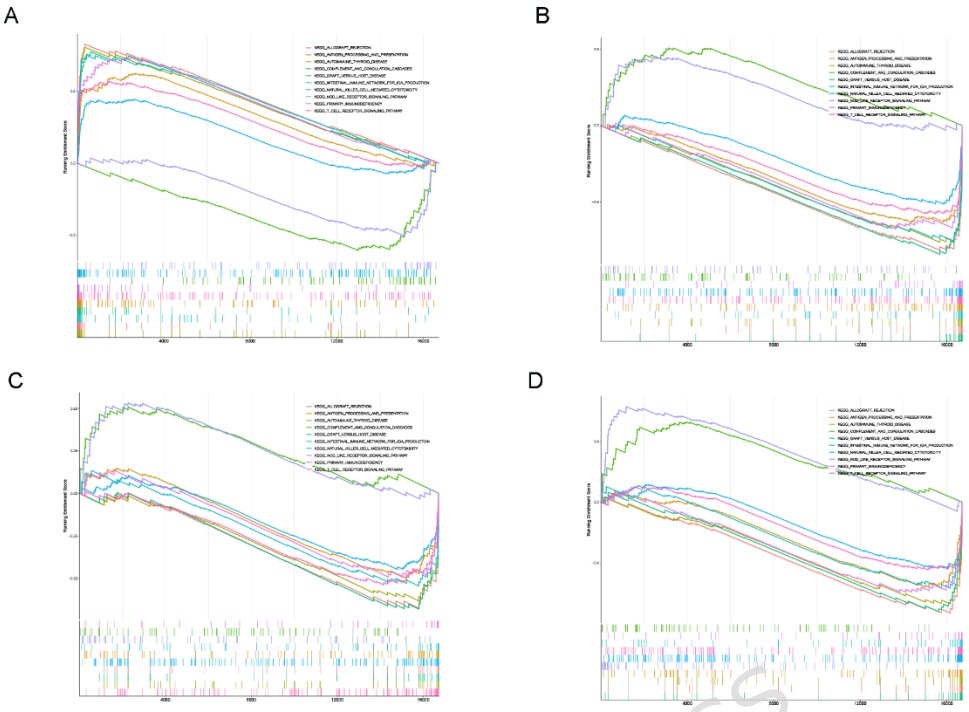



**Figure 6. Functional enrichment analysis of LN-associated biomarkers.**

## The Role of *CD40LG, RETN, TRIM22,* and *STAT1* in the LN Immune Microenvironment

PCA of *CD40LG, RETN, TRIM22,* and *STAT1* expression profiles revealed three distinct clusters (Cluster 1–3), accounting for 79.7% of total variance (Dim1: 56.2%, Dim2: 23.5%), suggesting notable heterogeneity in LN pathogenesis (Figure 7A). Immune deconvolution showed significant differences in the abundance of 17 out of 28 immune cell subsets across clusters (Figure 7B–C). Cluster 1 was enriched in innate immune signatures, including neutrophils and macrophages, whereas Clusters 2 and 3 displayed adaptive immune-dominant profiles, with higher levels of effector memory

CD8+ T cells and T follicular helper cells (Figure 7C).

Correlation analysis highlighted gene-specific immune interactions (Figure 7D–E). *CD40LG* positively correlated with T follicular helper cells and effector memory CD8+ T cells, but negatively with neutrophils and macrophages. *RETN* was associated with activated dendritic cells and mast cells, *TRIM22* with activated CD4+ T cells, and *STAT1* showed mild correlation with natural killer (NK) cells. A Spearman correlation heatmap further visualized the strength and direction of these associations (Figure7E).

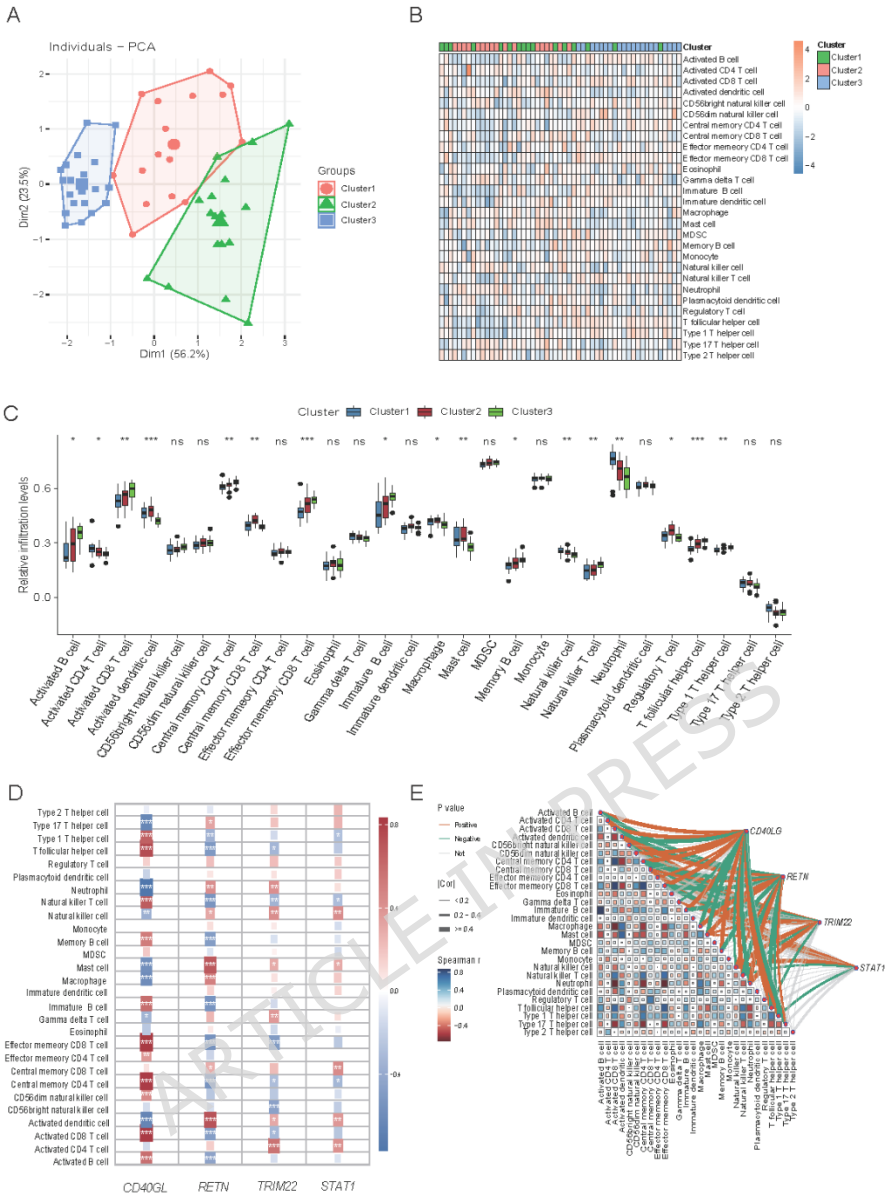



**Figure 7. Immune infiltration analysis of *CD40LG, RETN, TRIM22,* and *STAT1* in LN.**

(A) PCA plot showing three clusters based on gene expression profiles. (B) Relative proportions of immune cell infiltration in LN samples. (C) Abundance of 28 immune cell types across clusters. (D) Correlation analysis between the four biomarkers and immune cell types. (E) Spearman correlation heatmap visualizing associations between biomarkers and immune cells. (Kruskal-

Wallis test for panel C; Spearman correlation for panels D and E). ns, not significant; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

**Transcription Factor-Gene Regulatory Network Analysis in LN**

A transcription factor (TF)–gene regulatory network was constructed for the four hub genes (*CD40LG, STAT1, RETN,* and *TRIM22*) to explore potential upstream regulators (Figure 8). Several TFs were predicted to interact with each hub gene, forming distinct regulatory connections. These predicted interactions suggest potential transcriptional regulation of the hub genes in LN. All results are based on computational prediction and require experimental validation.



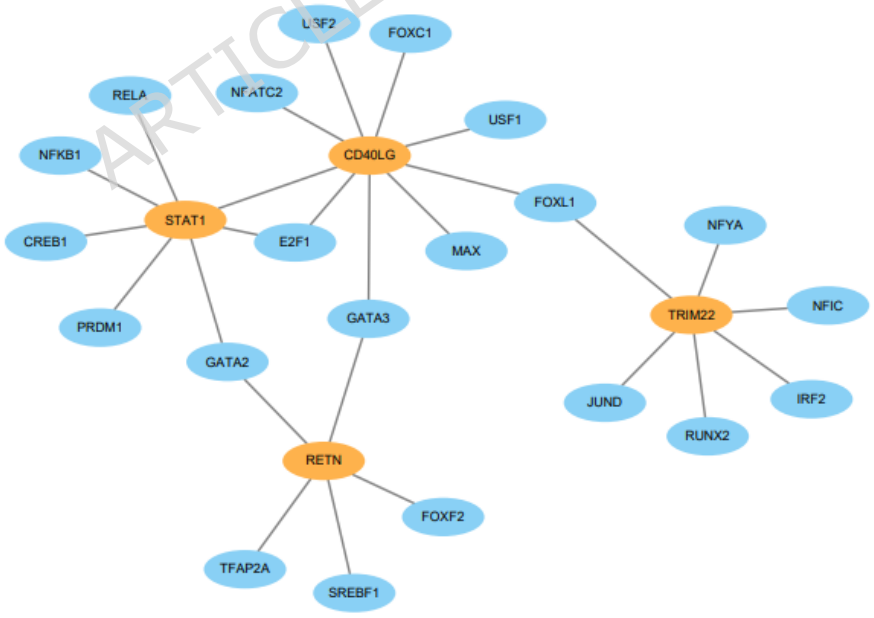



**Figure 8. TF-gene regulatory network of four immune-related genes in LN.**

**Clinical Characteristics**

The baseline characteristics of the study cohort are summarized in Table 2. The LN group included 12 females and 1 male (mean age 38.7 ± 12.2 years), and the HC group comprised 9 females and 1 male (mean age 34.6 ± 11.4 years). There were no statistically significant differences in age or sex distribution between the two groups ($P > 0.05$). In LN patients, the mean renal SLEDAI was 7.69 ± 5.76. Laboratory values showed a mean serum creatinine of 66.2 ± 20.8 μmol/L and eGFR of 104.8 ± 23.3 mL/min/1.73 $m^2$. Clinical features included proteinuria >0.5 g/24 h in 69.2% of patients, hematuria in 53.8%, urinary casts in 23.1%, and urine WBC >5/HPF in 38.5%. Anti-dsDNA antibodies were positive in 61.5% of patients, and mean complement levels were C3 0.90 ± 0.17 g/L and C4 0.178 ± 0.041 g/L.



Regarding treatment, all patients were receiving glucocorticoids (100%), and 53.8% were on immunosuppressants (primarily mycophenolate mofetil), while 23.1% were treated with biologics (belimumab). Renal biopsy revealed ISN/RPS classes III, IV, V, and mixed types.

**Table 2. Baseline characteristics of LN patients and healthy controls**

| Variable | LN (n = 13) | HC (n = 10) |
|---|---|---|
| **Demographics** | | |
| Age (years, mean ± SD) | 38.7 ± 12.2 | 34.6 ± 11.4 |
| Sex (female/male) | 12/1 (92.3% / 7.7%) | 9/1 (90% / 10%) |
| **Clinical and Laboratory Data** | | |
| Renal SLEDAI | 7.69 ± 5.76 | — |
| Serum creatinine (μmol/L) | 66.2 ± 20.8 | — |
| eGFR (mL/min/1.73 m²) | 104.8 ± 23.3 | — |
| 24h urine protein >0.5 g/24h (positive) | 9 (69.2%) | — |
| Hematuria (RBC > 5/HPF) | 7 (53.8%) | — |
| Urinary casts/cellular casts (positive) | 3 (23.1%) | — |
| Urine WBC > 5/HPF (positive) | 5 (38.5%) | — |
| Anti-dsDNA (positive) | 8 (61.5%) | — |
| C3 (g/L) | 0.90 ± 0.17 | — |
| C4 (g/L) | 0.178 ± 0.041 | — |
| **Main treatment at sampling** | | |
| **Glucocorticoids** | 13 (100.0%) | — |
| **Immunosuppressants** | | |
| Mycophenolate mofetil | 7 (53.8%) | — |
| Cyclophosphamide | 3 (23.1%) | — |
| Tacrolimus | 2 (15.4%) | — |
| **Biologics** | | |
| Belimumab | 3 (23.1%) | — |
| Renal biopsy (ISN/RPS classification) | III, IV, V, mixed types | — |

**Note:** Continuous variables are expressed as mean ± standard deviation (SD),

and categorical variables are expressed as frequency (percentage). There were no statistically significant differences in age (P = 0.456, Mann-Whitney U test) and sex distribution (P > 0.999, Fisher's exact test) between the LN and HC groups.

**Validation of hub gene expression levels in LN**

qRT-PCR analysis was performed on peripheral blood samples from 13 LN patients and 10 healthy controls to validate the expression of candidate hub genes. *STAT1* and *TRIM22* were significantly upregulated in LN patients compared to controls ($P$ = 0.0002 and $P$ = 0.0003, respectively), while *RETN* and *CD40LG* showed no significant differences ($P$ = 0.784 and $P$ = 0.784, respectively) (Figure 9A-D, Supplementary Table 3).

To investigate whether clinical variables influenced *STAT1* and *TRIM22* expression, univariable linear regression and Spearman correlation analyses were performed, including eGFR, serological markers (C3, C4, anti-dsDNA), and disease activity indices (renal SLEDAI, 24-hour proteinuria, log-transformed). No significant associations were observed (all $P > 0.05$, see Supplementary Table 4 and Supplementary Figure 4). This lack of correlation suggests that clinical severity scores did not independently drive gene expression in this cohort, potentially due to the stabilizing effects of concurrent medication. These results indicate that *STAT1* and

*TRIM22* serve as promising biomarkers for disease identification (diagnosis) even under treatment, whereas their utility for monitoring disease fluctuation appears limited in this context. Regarding *RETN* and *CD40LG*, their clinical relevance in peripheral blood warrants further investigation.

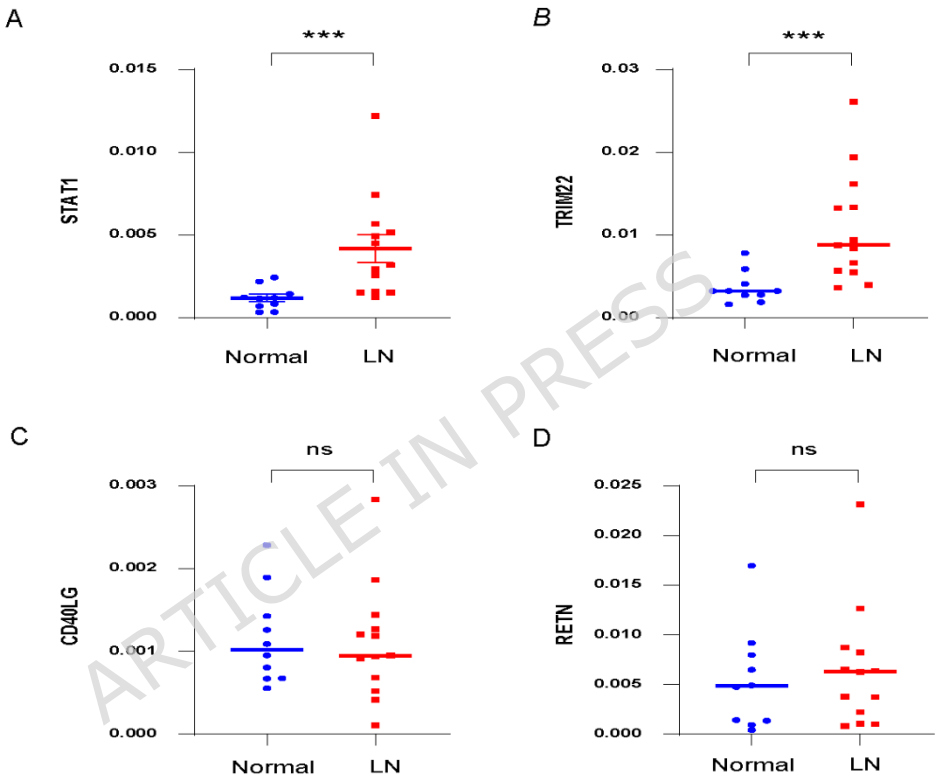



**Figure 9. Validation of hub gene expression levels in LN.**

(A–D) qRT-PCR analysis of *STAT1*, *TRIM22*, *RETN*, and *CD40LG* in peripheral blood from LN patients and healthy controls. Data are presented as median (IQR). Statistical comparisons between groups were performed using the Mann-Whitney U test. *** $P < 0.001$; ns, not significant.

**Construction and Clinical Validation of the Diagnostic**

**Nomogram and Two-Gene Model**

To translate the identified hub genes into a clinically applicable diagnostic tool, we prioritized genes validated by qRT-PCR experiments. Based on these validation results, only *STAT1* and *TRIM22* showed significant upregulation in our clinical samples; therefore, these two experimentally confirmed genes were selected to construct the diagnostic model.

Accordingly, we developed a diagnostic model based on the expression levels of *STAT1* and *TRIM22*. A nomogram was first established using the training set to visualize the predicted probability of LN risk (Figure 10A). ROC curve analysis in the training set demonstrated that the combined model based on *STAT1* and *TRIM22* exhibited high diagnostic accuracy (Figure 10B).

To verify the robustness and generalizability of the two-gene signature, we performed a "leave-dataset-out" cross-validation. The model was trained on one GEO cohort and tested on an independent GEO cohort, confirming stable diagnostic performance under cross-dataset conditions (Figure 10C-D).

Finally, the diagnostic efficacy of the signature was validated using the independent clinical qRT-PCR cohort. ROC curve analysis demonstrated strong diagnostic potential for *STAT1* and *TRIM22* individually, with AUCs of 0.923 and 0.915, respectively (Figure 10E).

Furthermore, a combined analysis integrating *STAT1* and *TRIM22* improved diagnostic performance to an AUC of 0.931 (Figure 10F). These results suggest that a multi-gene panel provides superior sensitivity and specificity for LN detection compared to single biomarkers.

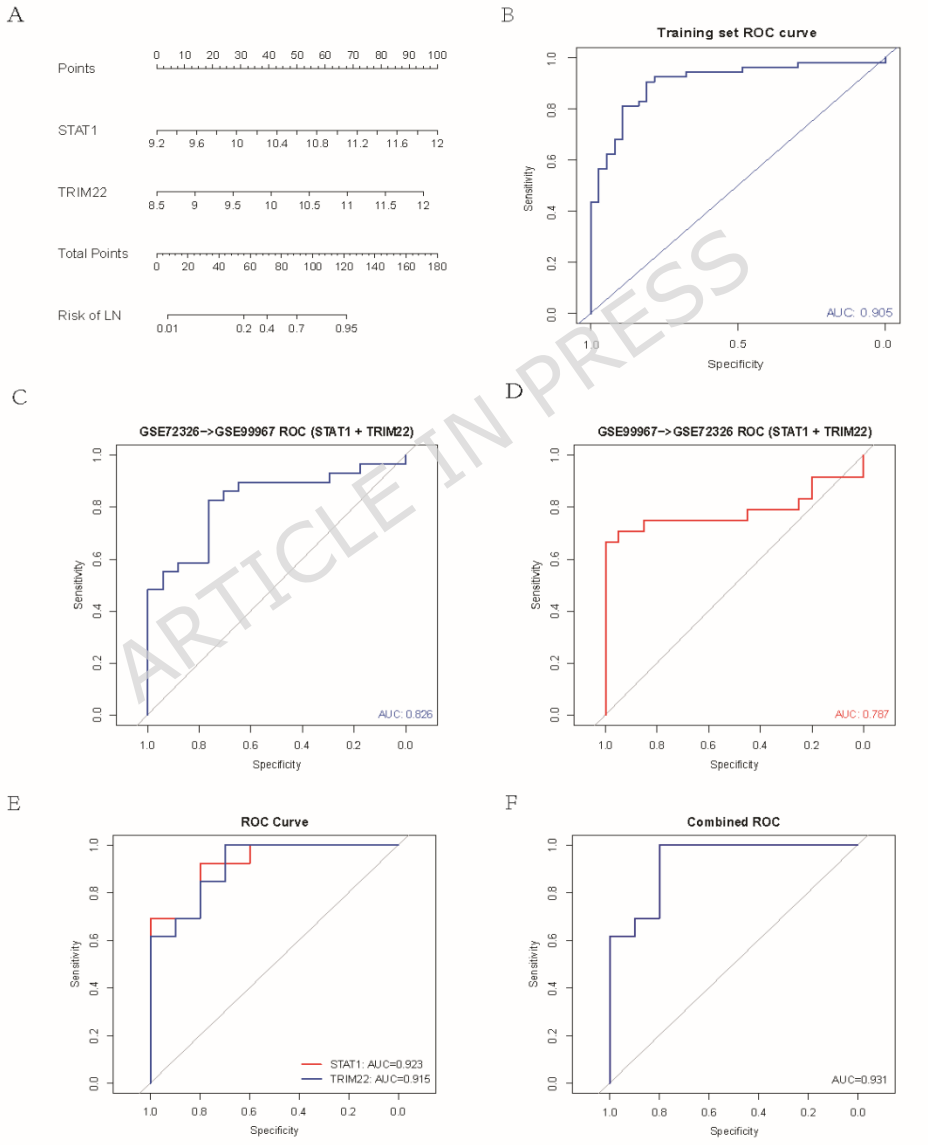



**Figure 10. Construction and validation of the diagnostic model for LN.**

(A) Diagnostic nomogram established in the training set based on *STAT1* and

*TRIM22*. (B) ROC curve analysis of the diagnostic performance of *STAT1* and *TRIM22* in the training set. (C–D) ROC curves of the two-gene model under leave-dataset-out cross-validation between independent GEO cohorts. (E) ROC curves of *STAT1* and *TRIM22* in the clinical validation cohort. (F) ROC curve of the combined model integrating *STAT1* and *TRIM22* in the clinical validation cohort. The nomogram was constructed using multivariate logistic regression analysis. Diagnostic performance was evaluated using ROC analysis and AUC calculations.

## Discussion

### Research Innovation and Key Findings

Integrating multi-cohort transcriptomics with machine learning, we identified *STAT1* and *TRIM22* as potential diagnostic biomarkers for LN based on peripheral blood mRNA expression. Upon validation in our clinical cohort, a notable observation was the divergence between their diagnostic accuracy and monitoring utility in peripheral blood: while both genes effectively distinguished LN patients from healthy controls, they did not correlate with disease activity indices (Fig.S4). Rather than a simple lack of relevance, we hypothesize that this discrepancy may reflect a 'therapeutic decoupling' of the interferon signature. As documented by Landolt-Marticorena et al., molecular interferon signatures—which are

fundamentally orchestrated by the *STAT1* pathway—often remain persistently elevated and relatively stable over time, even when clinical symptoms exhibit marked improvement under treatment[31-33]. This persistence suggests that *STAT1* and *TRIM22* represent a constitutive molecular imprint of LN that remains detectable despite the suppressive effects of glucocorticoids and immunosuppressants on clinical scores. Thus, these biomarkers may be more reliable for identifying the intrinsic autoimmune pathology rather than monitoring transient clinical fluctuations in treated populations.

Conversely, *CD40LG* and *RETN* were excluded from the final model as they did not reach statistical significance in our validation. This may be attributed to their biological characteristics and detection limitations. Specifically, *CD40LG* is transiently expressed in activated $CD4^+$ T cells, and its signal may be diluted in whole-blood samples by non-expressing cell populations[34-36]. Similarly, while *RETN* protein levels are elevated in LN serum, its peripheral mRNA abundance is predominantly macrophage-derived and may be limited by tissue-specific expression or post-transcriptional regulation[37]. While integrated multi-cohort analyses could capture these variable signals, they were likely below the detection threshold in our smaller, steroid-treated cohort**.**



### ***STAT1*** **in LN Pathogenesis**

*STAT1* is a central mediator of type I and II interferon signaling in the JAK–STAT pathway and plays a key role in autoimmune immune dysregulation[32]. Transcriptomic and single-cell studies have consistently identified *STAT1*-centered inflammatory signatures in SLE and related autoimmune diseases[33]. In LN renal biopsies, *STAT1* and its downstream chemokine *CXCL10* are markedly upregulated and pharmacologically modulated[38].

In this study, *STAT1* was significantly upregulated in LN peripheral blood and positively correlated with renal NK cell infiltration (Fig. 8D), suggesting interferon-driven amplification of renal inflammation. This finding aligns with animal studies showing sustained glomerular *STAT1* activation promotes renal injury[39,40]. Moreover, *STAT1* cooperates with *RELA* to enhance *CXCL10* transcription, facilitating monocyte recruitment and exacerbating renal inflammation[41-44]. Together, these data highlight the *STAT1*–*CXCL10* axis as a central pathway in LN pathogenesis.

**Immunoregulatory Role of *TRIM22* in LN**

*TRIM22* is an interferon-inducible E3 ubiquitin ligase involved in innate immune and inflammatory regulation[45]. We observed significant upregulation of *TRIM22* in LN peripheral blood, with positive correlations to $CD4^+$ T-cell activation (Fig. 8D). Experimental studies indicate that *TRIM22* enhances NF-κB

signaling and that its suppression alleviates renal inflammation in lupus-prone mice via inhibition of the NF-κB/NLRP3 axis[46].

Consistently, GSEA showed that high *STAT1* and *TRIM22* expression was associated with enrichment of NOD-like receptor and T-cell receptor signaling pathways, both of which converge on NF-κB activation [47-49]. NF-κB is a central regulator of pro-inflammatory cytokines, including TNF-α, IL-6, and MCP-1[50,51]. These findings suggest that *STAT1* and *TRIM22* may cooperatively amplify NF-κB-mediated inflammatory responses, promoting immune cell infiltration and renal injury in LN.

**Clinical Implications**

Current LN management relies on glucocorticoids and broad immunosuppression, with limited precision and substantial adverse effects[52]. While biologics such as belimumab improve outcomes in some patients, overall efficacy remains variable[53]. Traditional serological markers, including anti-dsDNA antibodies, show only moderate diagnostic performance for LN, with reported AUCs generally ranging from 0.6 to 0.8[54,55].

In contrast, the diagnostic model based on *STAT1* and *TRIM22* demonstrated favorable accuracy in distinguishing LN patients from healthy controls within our validation cohort. We acknowledge that distinguishing LN from healthy individuals does not fully reflect the

complex differential diagnosis faced by clinicians (e.g., distinguishing LN from other nephropathies). However, the robustness of this two-gene signature in treated patients suggests a specific utility: it may serve as a complementary, non-invasive tool to confirm the presence of an underlying interferon-driven autoimmune pathology when biopsy is contraindicated or inconclusive, rather than a standalone replacement for renal biopsy, pending further large-scale validation.

**Limitations**

Several limitations warrant mention. First, regarding the clinical cohort, the sample size was relatively small, and the majority of patients were receiving immunosuppressive therapy at the time of sampling. We acknowledge this as a potential confounding factor, as these medications may influence gene expression profiles. While this treatment background provides a potential explanation for the observed 'therapeutic decoupling' effect, it may limit the generalizability of our findings to treatment-naive populations. Furthermore, our study primarily compared LN patients with healthy controls. To fully address the clinical diagnostic dilemma, future large-scale studies including disease controls (e.g., patients with non-lupus glomerular diseases or SLE without nephritis) are necessary to rigorously validate the specificity of *STAT1* and *TRIM22*

in a broader differential diagnosis context. Second, regarding experimental verification, since samples were not stored in a manner that preserved protein content, it was not possible to assess the protein-level consistency of *STAT1* and *TRIM22*. We also note the methodological consideration regarding the 317 bp *STAT1* amplicon; however, specificity was verified via melt curve analysis and agarose gel electrophoresis. Finally, the proposed mechanistic links are inferred primarily from bioinformatic analyses; future studies incorporating *in vitro* or *in vivo* interventions are warranted to experimentally validate these pathways.



## Conclusion

In conclusion, we identified and characterized the immune-related hub genes *STAT1* and *TRIM22* in LN using transcriptomic data and machine learning. Both genes were significantly upregulated in peripheral blood and correlated with immune cell infiltration and inflammatory pathways. These biomarkers show promise as non-invasive molecular signatures for identifying underlying disease status. Building upon the insights and limitations of this exploratory study, further large-scale research is warranted to confirm their clinical utility and experimentally validate the proposed pathogenic mechanisms in LN.

# Abbreviations

| Abbreviation | Full Name |
|---|---|
| STAT1 | Signal transducer and activator of transcription 1 |
| TRIM22 | Tripartite motif containing 22 |
| CD40LG | CD40 ligand |
| RETN | Resistin |
| SLE | Systemic lupus erythematosus |
| LN | Lupus nephritis |
| DEGs | Differentially expressed genes |
| IRGs | Immune-related genes |
| WGCNA | Weighted gene co-expression network analysis |
| IR-DEGs | Immune-related differentially expressed genes |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| SVM | Support Vector Machine |
| XGBoost | Extreme Gradient Boosting |
| ROC | Receiver operating characteristic |
| GSEA | Gene Set Enrichment Analysis |
| PCA | Principal Component Analysis |
| ssGSEA | Single-sample Gene Set Enrichment Analysis |
| TFs | Transcription factors |
| qRT-PCR | Quantitative real-time PCR |
| HCs | Healthy controls |
| eGFR | Estimated glomerular filtration rate |
| SLEDAI | Systemic Lupus Erythematosus Disease Activity Index |
| AUC | Area Under the Curve |
| PCA | Principal component analysis |
| NK cells | Natural killer cells |

| Abbreviation | Full Name |
|---|---|
| ISGs | Interferon-stimulated genes |
| IFN | Interferon |

## Reference:


1 Rahman, A. & Isenberg, D. A. Systemic lupus erythematosus. *N Engl J Med* **358**, 929-939, doi:10.1056/NEJMra071297 (2008).

2 Tsokos, G. C. Systemic lupus erythematosus. *N Engl J Med* **365**, 2110-2121, doi:10.1056/NEJMra1100359 (2011).

3 Almaani, S., Meara, A. & Rovin, B. H. Update on Lupus Nephritis. *Clin J Am Soc Nephrol* **12**, 825-835, doi:10.2215/cjn.05780616 (2017).

4 Davidson, A. What is damaging the kidney in lupus nephritis? *Nat Rev Rheumatol* **12**, 143-153, doi:10.1038/nrrheum.2015.159 (2016).

5 Hocaoğlu, M. *et al.* Incidence, Prevalence, and Mortality of Lupus Nephritis: A Population-Based Study Over Four Decades Using the Lupus Midwest Network. *Arthritis Rheumatol* **75**, 567-573, doi:10.1002/art.42375 (2023).

6 Furie, R. *et al.* Two-Year, Randomized, Controlled Trial of Belimumab in Lupus Nephritis. *N Engl J Med* **383**, 1117-1128, doi:10.1056/NEJMoa2001180 (2020).

7 Lech, M. & Anders, H. J. The pathogenesis of lupus nephritis. *J Am Soc Nephrol* **24**, 1357-1366, doi:10.1681/asn.2013010026 (2013).

8 Banchereau, R. *et al.* Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell* **165**, 551-565, doi:10.1016/j.cell.2016.03.008 (2016).

9 Chang, A. *et al.* In situ B cell-mediated immune responses and tubulointerstitial inflammation in human lupus nephritis. *J Immunol* **186**, 1849-1860, doi:10.4049/jimmunol.1001983 (2011).

10 Kopetschke, K. *et al.* The cellular signature of urinary immune cells in Lupus nephritis: new insights into potential biomarkers. *Arthritis Res Ther* **17**, 94, doi:10.1186/s13075-015-0600-y (2015).

11 Sandersfeld, M. *et al.* Macrophage subpopulations in pediatric patients with lupus nephritis and other inflammatory diseases affecting the kidney. *Arthritis Res Ther* **26**, 46, doi:10.1186/s13075-024-03281-1 (2024).

12 Weening, J. J. *et al.* The classification of glomerulonephritis in systemic lupus erythematosus revisited. *Kidney Int* **65**, 521-530, doi:10.1111/j.1523-1755.2004.00443.x (2004).

13 Parikh, S. V., Almaani, S., Brodsky, S. & Rovin, B. H. Update on Lupus Nephritis: Core Curriculum 2020. *Am J Kidney Dis* **76**, 265-281, doi:10.1053/j.ajkd.2019.10.017 (2020).

14 Radin, M. *et al.* Prognostic and Diagnostic Values of Novel Serum and Urine Biomarkers in Lupus Nephritis: A Systematic Review. *Am J Nephrol* **52**, 559-571, doi:10.1159/000517852 (2021).

15 Morell, M., Pérez-Cózar, F. & Marañón, C. Immune-Related Urine Biomarkers for the Diagnosis of Lupus Nephritis. *Int J Mol Sci* **22**, doi:10.3390/ijms22137143 (2021).

16 Ha, E., Bae, S. C. & Kim, K. Recent advances in understanding the genetic basis of systemic lupus erythematosus. *Semin Immunopathol* **44**, 29-46,

doi:10.1007/s00281-021-00900-w (2022).

17 Toro-Domínguez, D., Carmona-Sáez, P. & Alarcón-Riquelme, M. E. Shared signatures between rheumatoid arthritis, systemic lupus erythematosus and Sjögren's syndrome uncovered through gene expression meta-analysis. *Arthritis Res Ther* **16**, 489, doi:10.1186/s13075-014-0489-x (2014).

18 Deng, Y. & Tsao, B. P. Advances in lupus genetics and epigenetics. *Curr Opin Rheumatol* **26**, 482-492, doi:10.1097/bor.0000000000000086 (2014).

19 Wither, J. E. *et al.* Identification of a neutrophil-related gene expression signature that is enriched in adult systemic lupus erythematosus patients with active nephritis: Clinical/pathologic associations and etiologic mechanisms. *PLoS One* **13**, e0196117, doi:10.1371/journal.pone.0196117 (2018).

20 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883, doi:10.1093/bioinformatics/bts034 (2012).

21 Zhu, H. *et al.* Whole-genome transcription and DNA methylation analysis of peripheral blood mononuclear cells identified aberrant gene regulation pathways in systemic lupus erythematosus. *Arthritis Res Ther* **18**, 162, doi:10.1186/s13075-016-1050-x (2016).

22 Bhattacharya, S. *et al.* ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data* **5**, 180015, doi:10.1038/sdata.2018.15 (2018).

23 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

24 Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267-288 (1996).

25 Rigatti, S. J. Random Forest. *J Insur Med* **47**, 31-39, doi:10.17849/insm-47-01-31-39.1 (2017).

26 Noble, W. S. What is a support vector machine? *Nature Biotechnology* **24**, 1565-1567, doi:10.1038/nbt1206-1565 (2006).

27 Chen, T. & Guestrin, C. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, San Francisco, California, USA, 2016).

28 Kursa, M. B. & Rudnicki, W. R. Feature Selection with the Boruta Package. *Journal of Statistical Software* **36**, 1 - 13, doi:10.18637/jss.v036.i11 (2010).

29 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77, doi:10.1186/1471-2105-12-77 (2011).

30 Iasonos, A., Schrag, D., Raj, G. V. & Panageas, K. S. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* **26**, 1364-1370, doi:10.1200/jco.2007.12.9791 (2008).

31 Landolt-Marticorena, C. *et al.* Lack of association between the interferon-

alpha signature and longitudinal changes in disease activity in systemic lupus erythematosus. *Annals of the rheumatic diseases* **68**, 1440-1446, doi:10.1136/ard.2008.093146 (2009).

32 Decker, T., Stockinger, S., Karaghiosoff, M., Müller, M. & Kovarik, P. IFNs and STATs in innate immunity to microorganisms. *The Journal of clinical investigation* **109**, 1271-1277, doi:10.1172/jci15770 (2002).

33 Cui, Y. *et al.* Exploring the shared molecular mechanisms between systemic lupus erythematosus and primary Sjögren's syndrome based on integrated bioinformatics and single-cell RNA-seq analysis. *Frontiers in immunology* **14**, 1212330, doi:10.3389/fimmu.2023.1212330 (2023).

34 Splawski, J. B., Nishioka, J., Nishioka, Y. & Lipsky, P. E. CD40 ligand is expressed and functional on activated neonatal T cells. *J Immunol* **156**, 119-127 (1996).

35 Ramanujam, M. *et al.* Phoenix from the flames: Rediscovering the role of the CD40-CD40L pathway in systemic lupus erythematosus and lupus nephritis. *Autoimmun Rev* **19**, 102668, doi:10.1016/j.autrev.2020.102668 (2020).

36 Koshy, M., Berger, D. & Crow, M. K. Increased expression of CD40 ligand on systemic lupus erythematosus lymphocytes. *J Clin Invest* **98**, 826-837, doi:10.1172/jci118855 (1996).

37 Hutcheson, J. *et al.* Resistin as a potential marker of renal disease in lupus nephritis. *Clin Exp Immunol* **179**, 435-443, doi:10.1111/cei.12473 (2015).

38 Shi, D. *et al.* Transcriptional expression of CXCL10 and STAT1 in lupus nephritis and the intervention effect of triptolide. *Clin Rheumatol* **42**, 539-548, doi:10.1007/s10067-022-06400-y (2023).

39 Dong, J., Wang, Q. X., Zhou, C. Y., Ma, X. F. & Zhang, Y. C. Activation of the STAT1 signalling pathway in lupus nephritis in MRL/lpr mice. *Lupus* **16**, 101-109, doi:10.1177/0961203306075383 (2007).

40 Zhang, L. H. *et al.* MiR-146b-5p targets IFI35 to inhibit inflammatory response and apoptosis via JAK1/STAT1 signalling in lipopolysaccharide-induced glomerular cells. *Autoimmunity* **54**, 430-438, doi:10.1080/08916934.2020.1864730 (2021).

41 Wang, X. *et al.* Reduced Renal CSE/CBS/H2S Contributes to the Progress of Lupus Nephritis. *Biology (Basel)* **12**, doi:10.3390/biology12020318 (2023).

42 Burke, S. J. *et al.* NF-κB and STAT1 control CXCL1 and CXCL2 gene transcription. *Am J Physiol Endocrinol Metab* **306**, E131-149, doi:10.1152/ajpendo.00347.2013 (2014).

43 Lee, E. Y., Lee, Z. H. & Song, Y. W. The interaction between CXCL10 and cytokines in chronic inflammatory arthritis. *Autoimmun Rev* **12**, 554-557, doi:10.1016/j.autrev.2012.10.001 (2013).

44 Rahmat-Zaie, R. *et al.* TNF-α/STAT1/CXCL10 mutual inflammatory axis that contributes to the pathogenesis of experimental models of multiple sclerosis: A promising signaling pathway for targeted therapies. *Cytokine* **168**, 156235, doi:10.1016/j.cyto.2023.156235 (2023).

45 Heo, H. *et al.* TRIM22 facilitates autophagosome-lysosome fusion by

mediating the association of GABARAPs and PLEKHM1. *Autophagy* **20**, 1098-1113, doi:10.1080/15548627.2023.2287925 (2024).

46 Chen, J. *et al.* Knockdown of TRIM22 regulates the expression of NF-κB/NLRP3 and alleviates inflammation and renal injury in mice with lupus nephritis. *Allergol Immunopathol (Madr)* **53**, 98-105, doi:10.15586/aei.v53i3.1313 (2025).

47 Barnabei, L., Laplantine, E., Mbongo, W., Rieux-Laucat, F. & Weil, R. NF-κB: At the Borders of Autoimmunity and Inflammation. *Front Immunol* **12**, 716469, doi:10.3389/fimmu.2021.716469 (2021).

48 Thaker, Y. R., Schneider, H. & Rudd, C. E. TCR and CD28 activate the transcription factor NF-κB in T-cells via distinct adaptor signaling complexes. *Immunol Lett* **163**, 113-119, doi:10.1016/j.imlet.2014.10.020 (2015).

49 Salek-Ardakani, S. & Croft, M. T cells need Nod too? *Nat Immunol* **10**, 1231-1233, doi:10.1038/ni1209-1231 (2009).

50 Liu, T., Zhang, L., Joo, D. & Sun, S. C. NF-κB signaling in inflammation. *Signal Transduct Target Ther* **2**, 17023-, doi:10.1038/sigtrans.2017.23 (2017).

51 Lawrence, T. The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harb Perspect Biol* **1**, a001651, doi:10.1101/cshperspect.a001651 (2009).

52 Shao, M. *et al.* [A multicenter study on the tolerance of intravenous low-dose cyclophosphamide in systemic lupus erythematosus]. *Beijing Da Xue Xue Bao Yi Xue Ban* **54**, 1112-1116, doi:10.19723/j.issn.1671-167X.2022.06.009 (2022).

53 Wang, X. *et al.* Advances in therapeutic targets-related study on systemic lupus erythematosus. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* **46**, 1267-1275, doi:10.11817/j.issn.1672-7347.2021.200056 (2021).

54 Vrabie, A., Obrișcă, B., Sorohan, B. M. & Ismail, G. Biomarkers in Lupus Nephritis: An Evidence-Based Comprehensive Review. *Life (Basel)* **15**, doi:10.3390/life15101497 (2025).

55 Ricchiuti, V. *et al.* Comparison of Five Assays for the Detection of Anti-dsDNA Antibodies and Their Correlation with Complement Consumption. *Diagnostics (Basel)* **15**, doi:10.3390/diagnostics15192430 (2025).



# Declarations

## Authors' contributions

JYD, ZMZ, and ZKG conceptualized and designed the study, with LTC and XMQ serving as guarantors. YYL and JPC conducted sample

collection, while CL and WHW coordinated all data acquisition. YHH and HRQ performed systematic analysis and interpretation of the data. JPC, CL, and YHL designed and plotted the figures and tables. JYD, ZMZ, XHSG, and XHL wrote the original manuscript draft. LTC, XMM, CJC, and WHW critically reviewed and edited the final version. This work was funded by LTC, XMM, and CJC. All authors (JYD, ZMZ, YYL, JPC, ZKG, CL, XHL, WHW, XMM, CJC, XHSG, YHH, HRQ, XMQ, LTC) read and approved the final manuscript.

**Availability of data and materials**

The datasets analyzed during the current study are publicly available in the Gene Expression Omnibus (GEO) database:

- GSE72326: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72326
- GSE99967: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99967
- GSE81622: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81622

A total of 2,483 IRGs were retrieved from the ImmPort database (https://www.immport.org/home) and are listed in Supplementary

Table 1. The list of 320 DEGs identified between LN and normal samples is available in Supplementary Table 2. The original qRT-PCR data used to validate the expression of selected hub genes are provided in Supplementary Table 3.

Results of the univariable linear regression analysis evaluating the associations between clinical parameters and *STAT1* and *TRIM22* expression levels in LN patients are shown in Supplementary Table 4. Detailed clinical characteristics of the LN patients, including demographic and disease-related information, are summarized in Supplementary Table 5. The baseline clinical and pathological characteristics of the integrated LN cohorts are provided in Supplementary Table 6.

All other relevant data generated or analyzed during this study are included in the main manuscript and the supplementary material.



**Ethics approval and consent to participate**

This study was approved by the Ethics Committee of Longyan First Hospital (Approval No. LYREC2024-k172-01). Whole blood samples from 13 lupus nephritis patients (diagnosed according to ISN/RPS criteria) and 10 healthy controls were collected after obtaining written informed consent from all participants. All procedures were conducted in accordance with the Declaration of Helsinki.

## Consent for publication

Not applicable.

## Competing interests

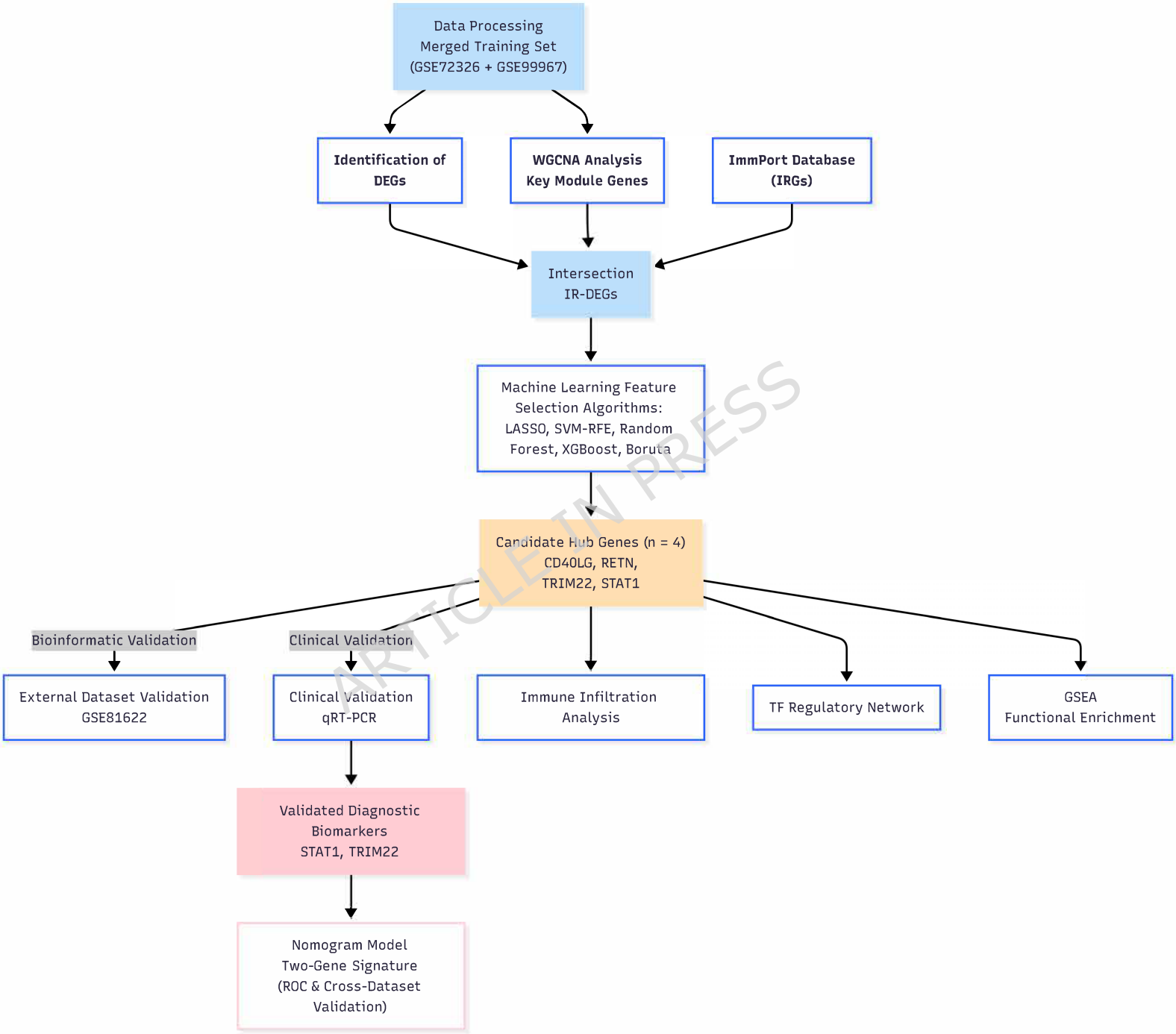
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding

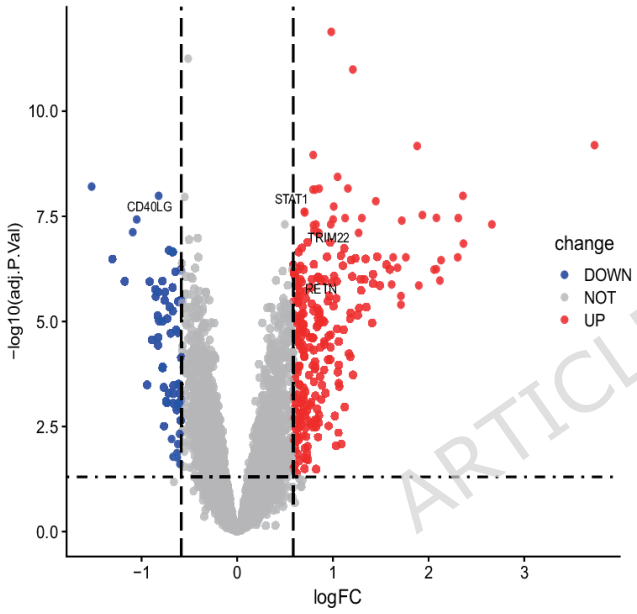
This work was supported by the Fujian Provincial Natural Science Foundation Program [grant number 2023J011894], the Longyan Municipal Science and Technology Planning Project [grant number 2022LYF17110], and the Fujian Provincial Health Technology Project [grant number 2022QNA102].

## Acknowledgements

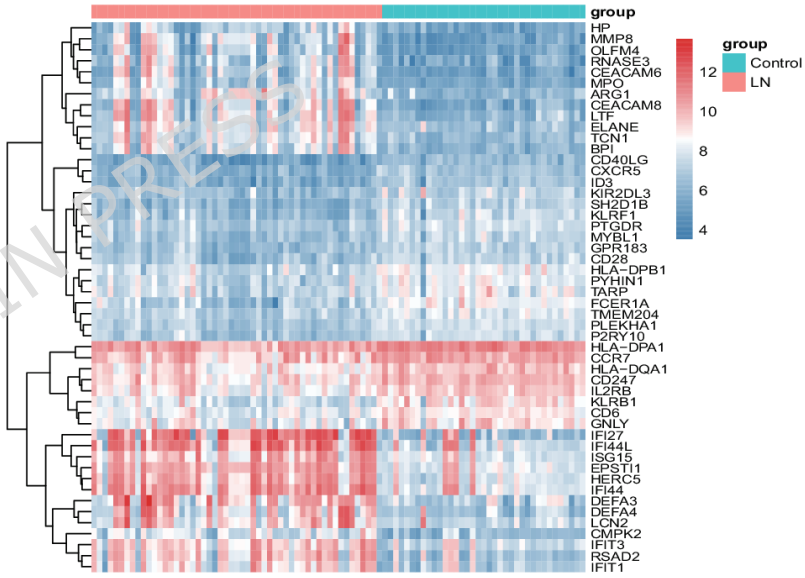
The authors thank all participants of the study. ChatGPT (OpenAI) was used solely to assist with language refinement and formatting during manuscript preparation. All intellectual content, data analysis, and interpretation were conducted independently by the authors.



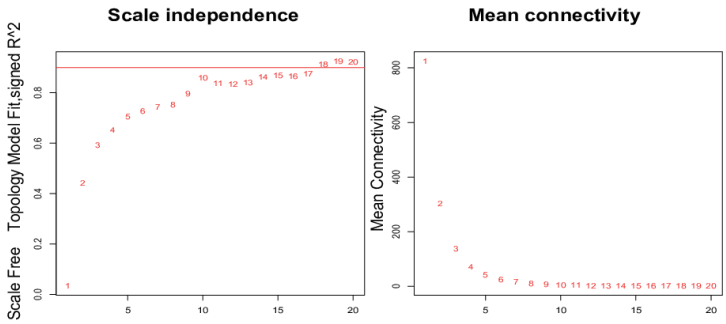
A



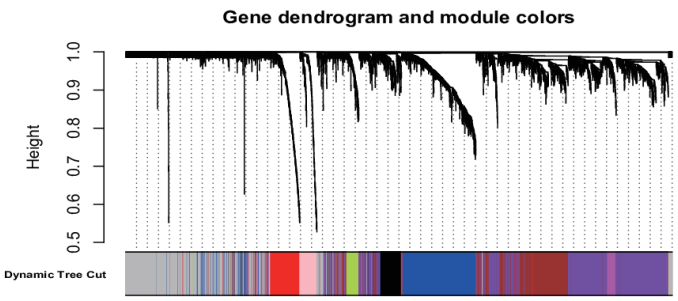

B



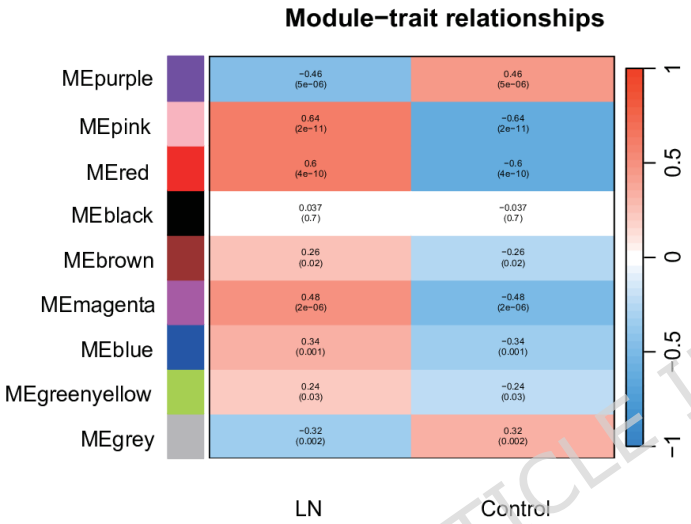
A



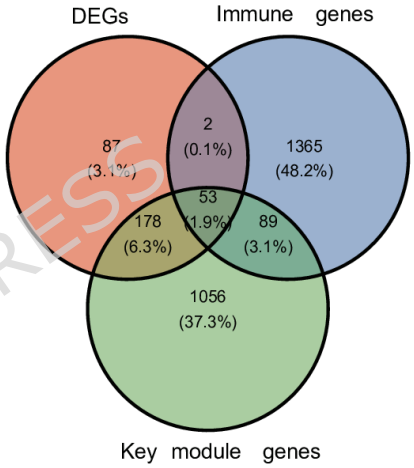

B



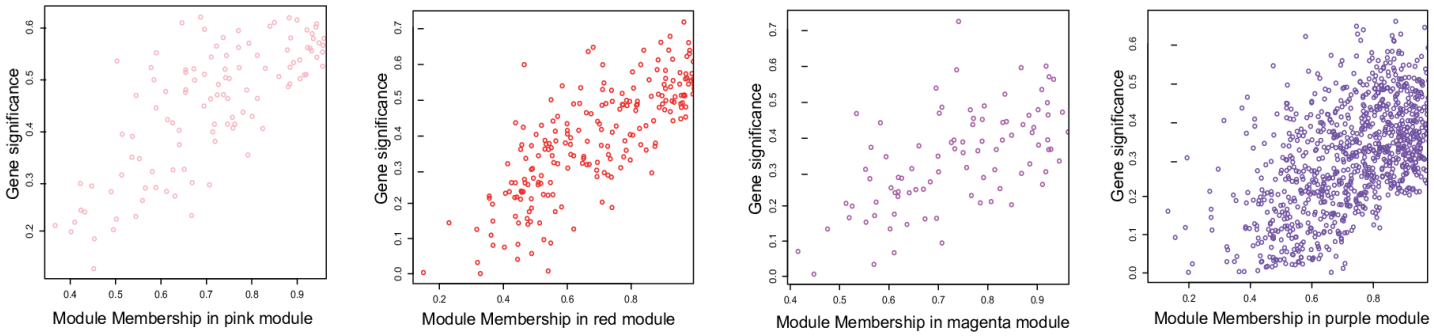

C

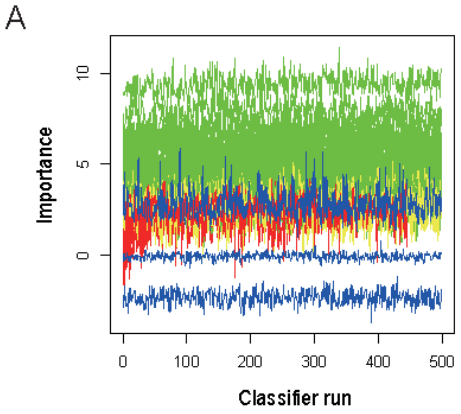


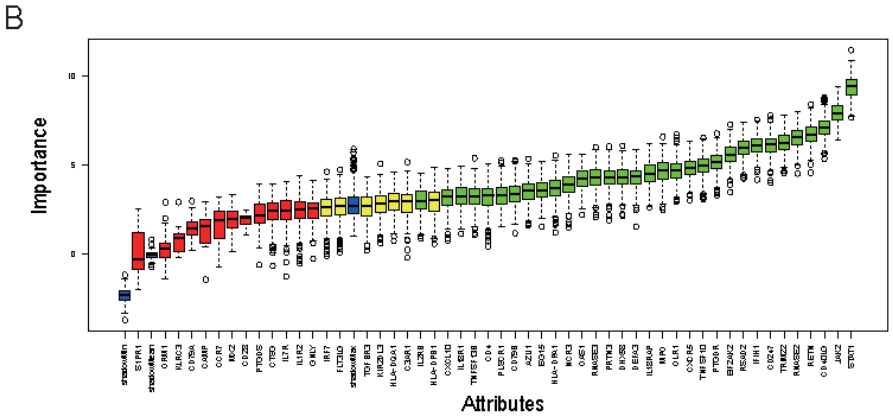

E

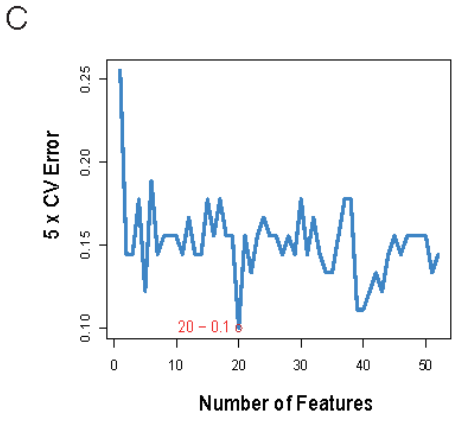


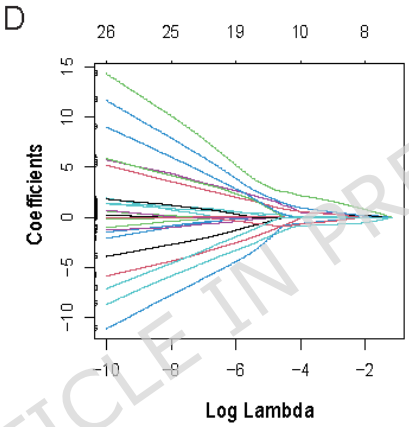

D

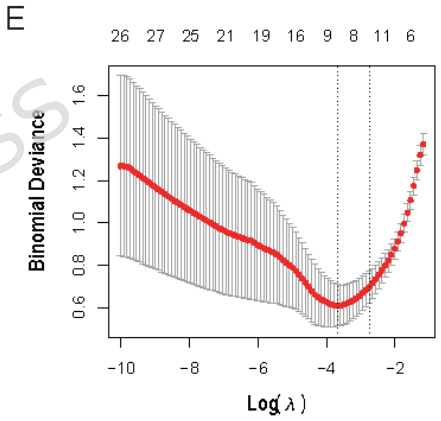


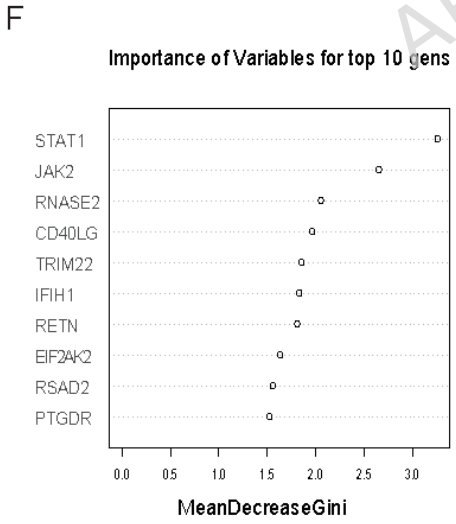


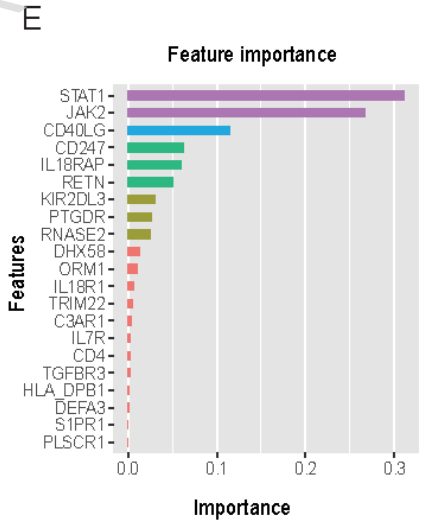



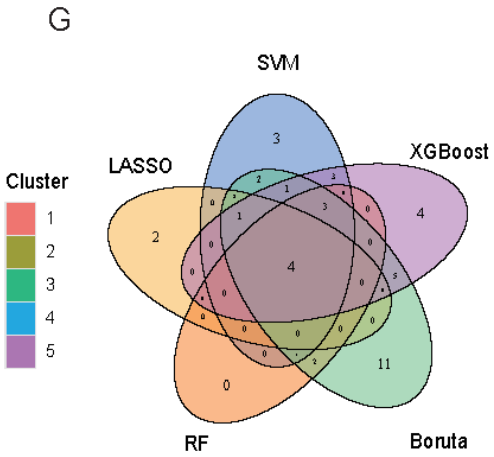



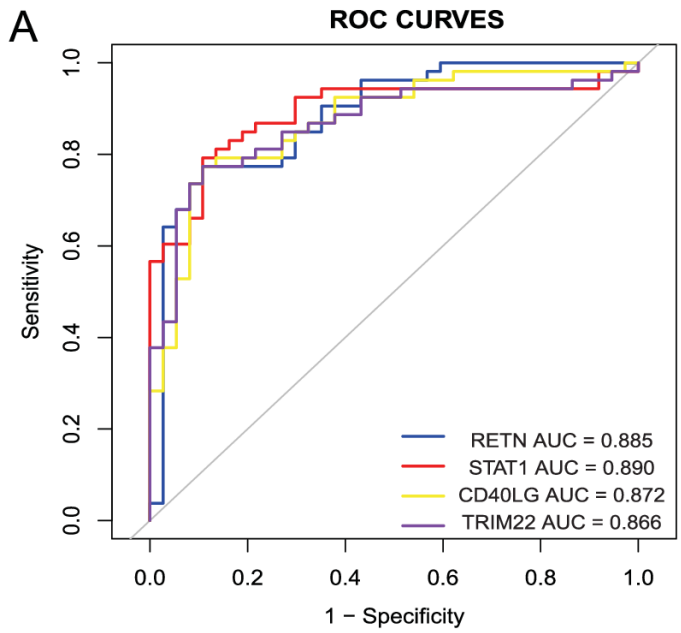



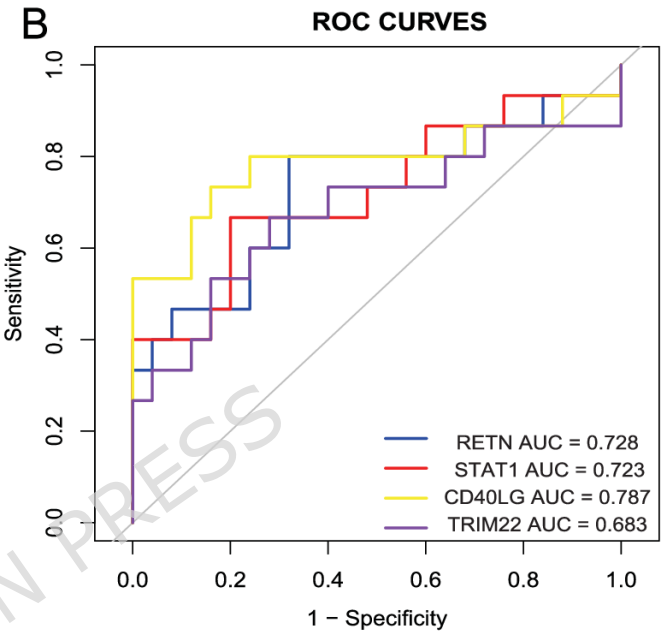



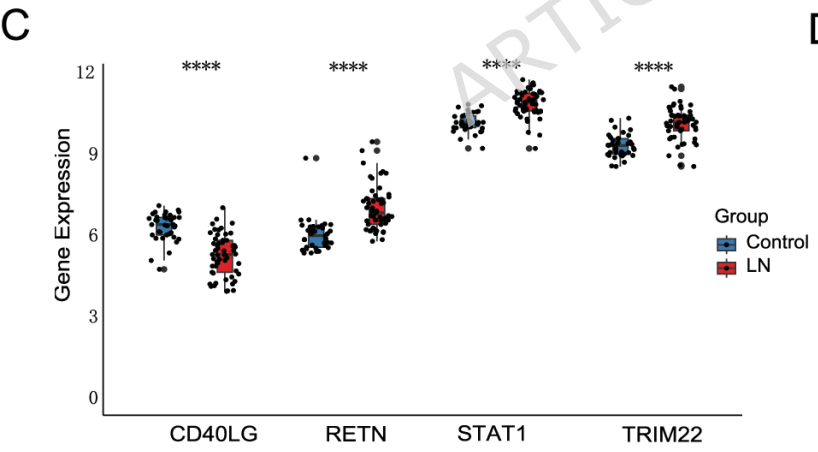



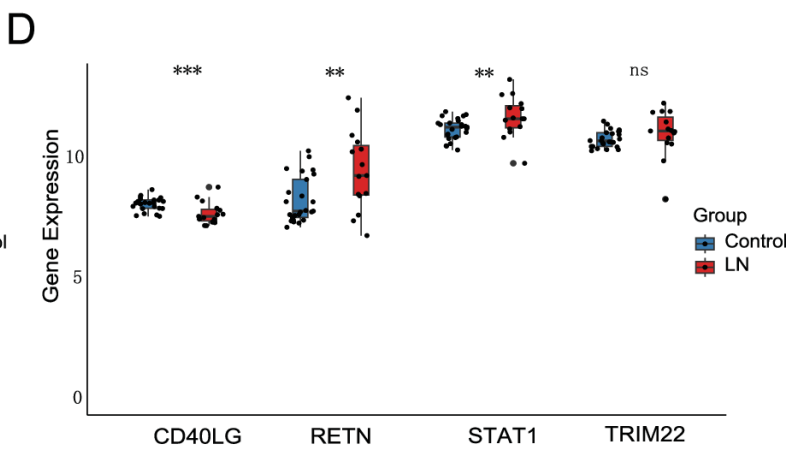



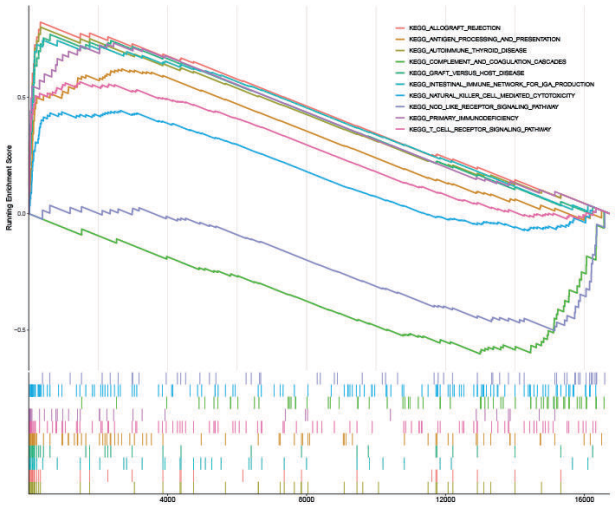


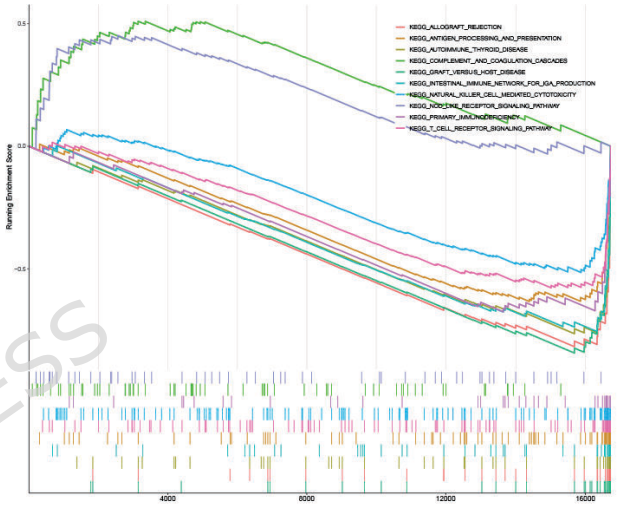



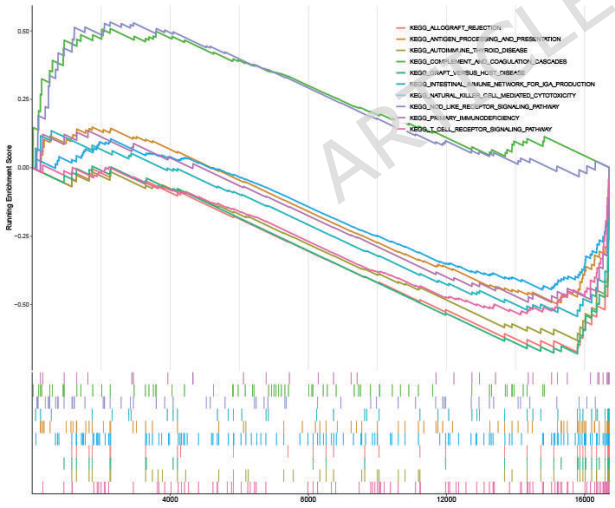
A



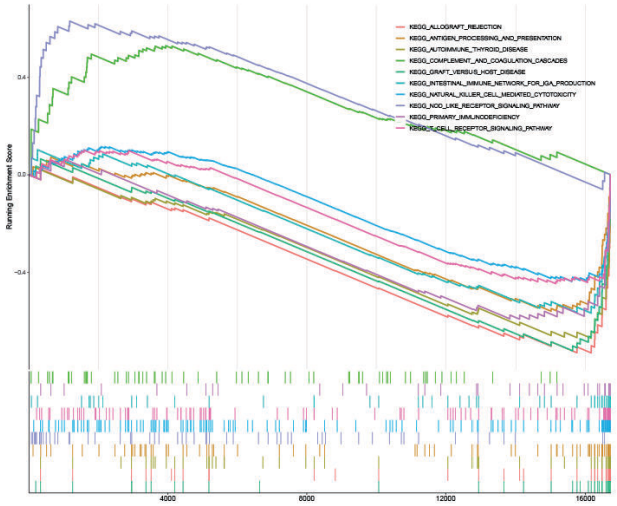

B

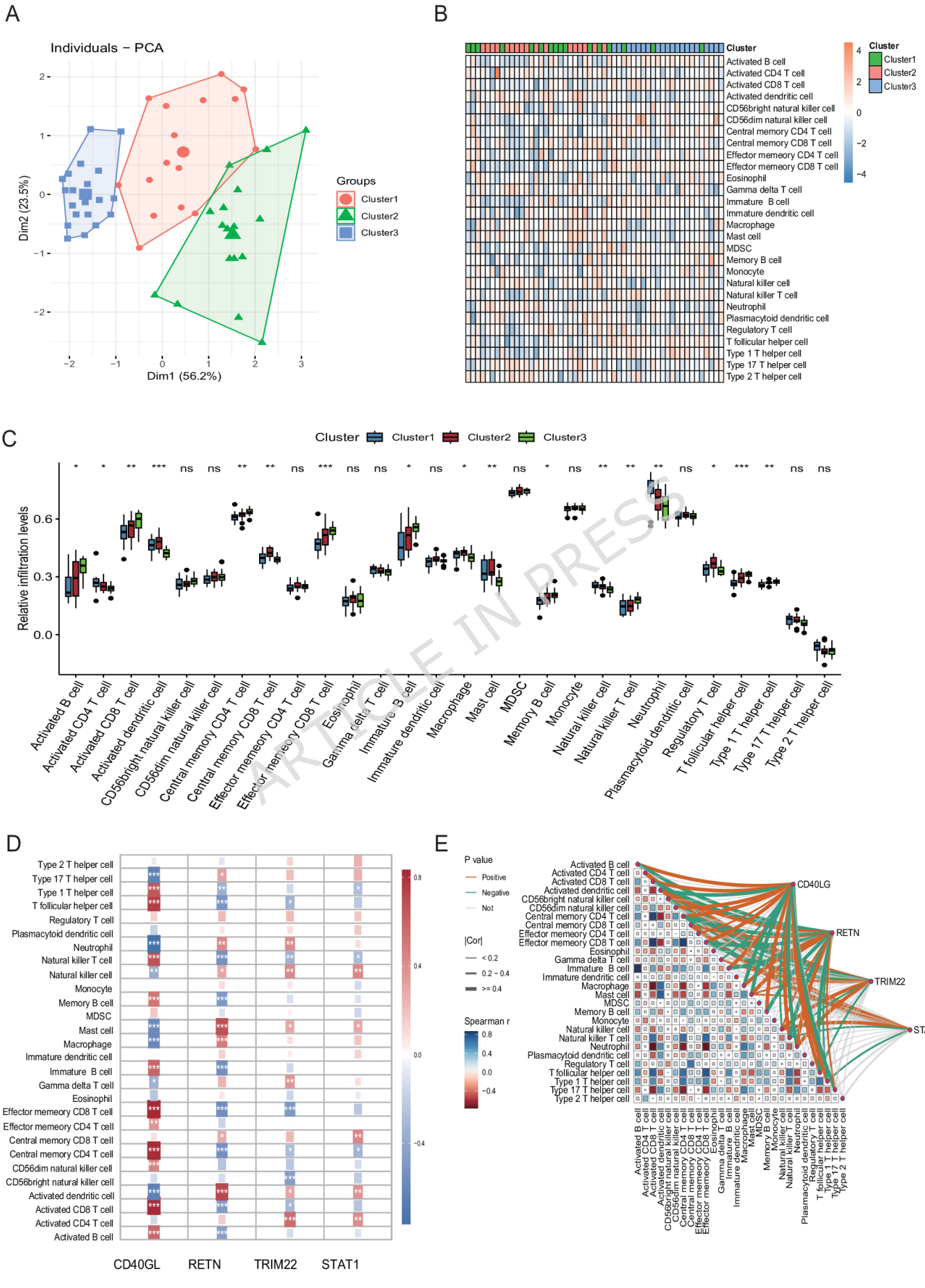


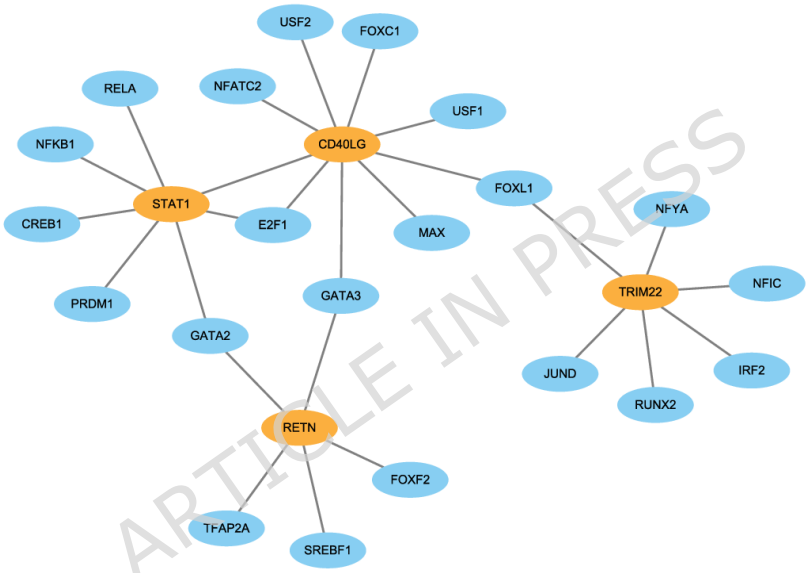

C

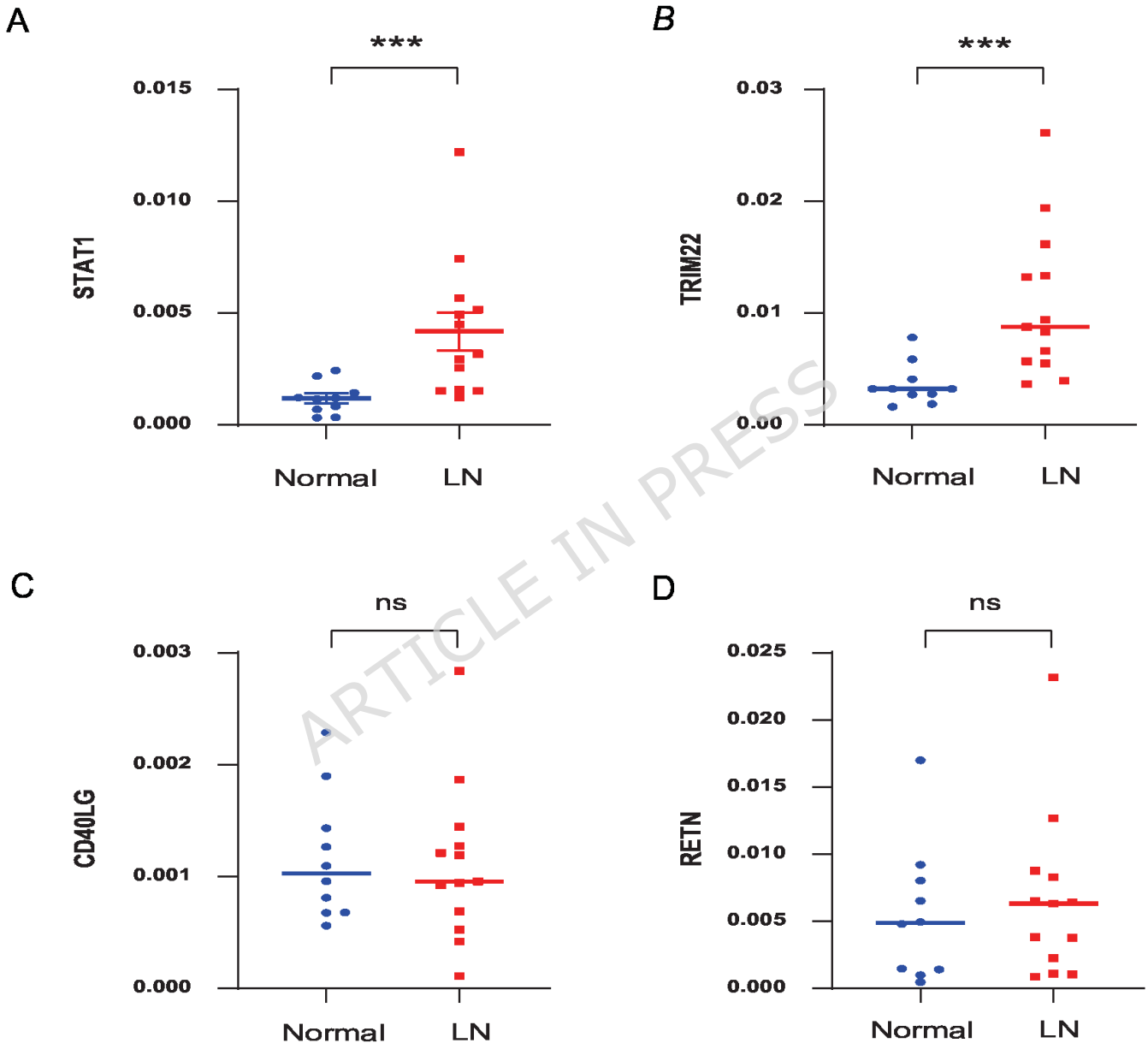



D

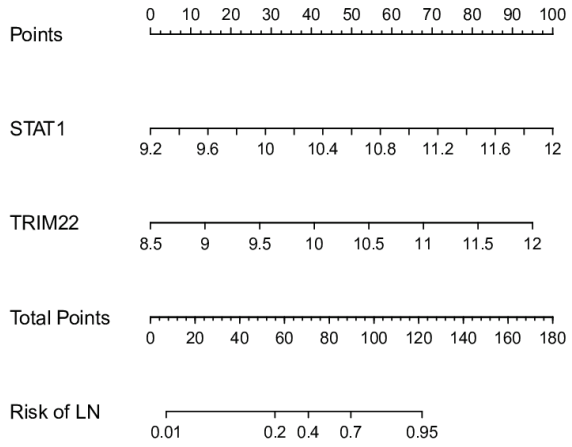




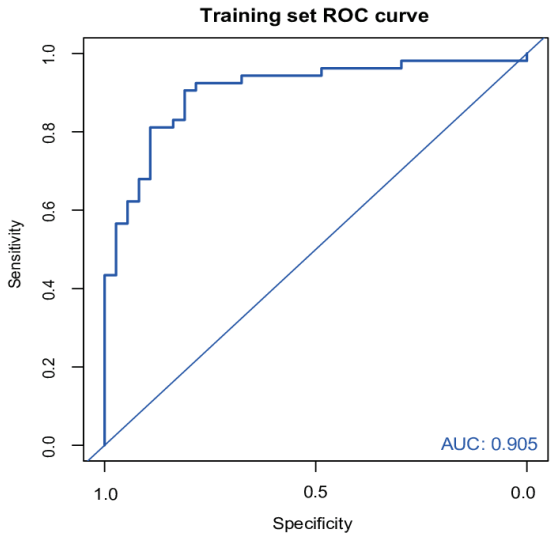




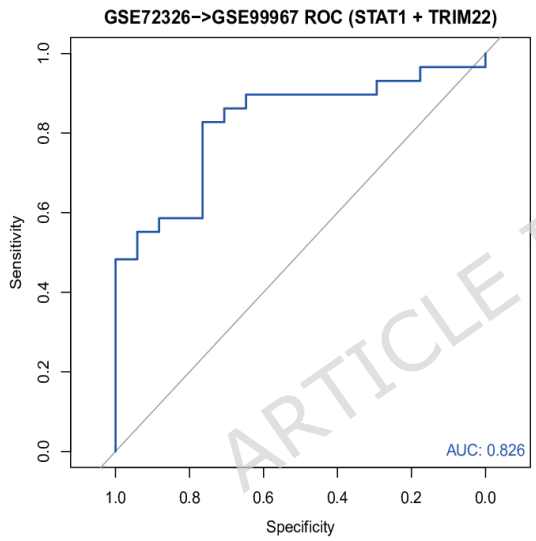
A



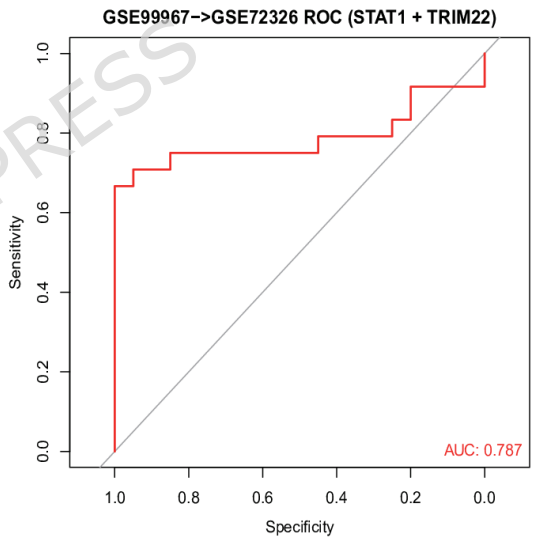

B



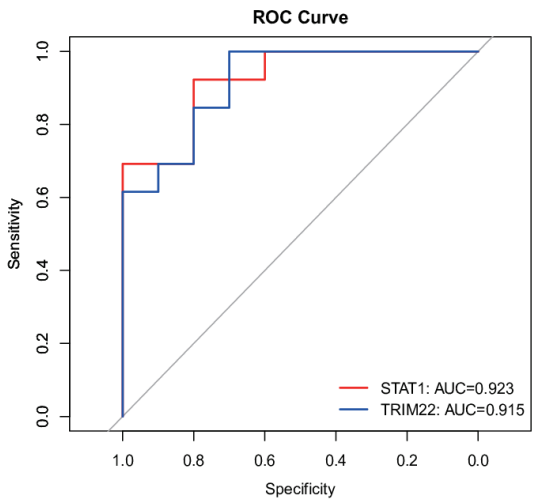

C




D




E




F