Omics-based integrated analysis identified IKZF2 as a biomarker associated with lupus nephritis

Mi Zhou¹,³, Yuening Kang¹,³, Jun Li¹,³, Rongxiu Li²* & Liangjing Lu¹*.

Lupus nephritis (LN) is a crucial complication of systemic lupus erythematosus (SLE). IKZF2 was identified as a lupus susceptibility locus, while its exact molecular function in LN is unknown. We aimed to explore the relationship between IKZF2 and LN based on multi-omics data. In our study, we carried out a meta-analysis of publicly available data, including not only tubulointerstitium, but also glomerulus tissue samples from LN patients and controls. Based on the common differentially expressed genes (co-DEGs) and previous researches, we selected IKZF2 for further analysis. Then, we analyzed potential molecular mechanisms of co-DEGs and IKZF2 in LN. To explore the possible targets of IKZF2, protein–protein interaction network (PPI) network and ceRNA network of IKZF2 were also constructed. Moreover, we performed immune infiltration analysis and evaluated clinical value of IKZF2. A total of 26 co-DEGs were observed in the integration of the above DEGs coming from the four sets of data, of which IKZF2 was selected for further analysis. Functional enrichment analysis from IKZF2 and related PPI network confirmed the tight relationship between IKZF2 and the immune reaction. Moreover, immune filtration analysis revealed the significant correlation between IKZF2 and naive B cell, NK cell activation, NK cell rest and other immune cells. Receiver operating characteristic (ROC) analysis showed that the areas under the ROC curves were 0.721, 0.80, 0.682, and 0.859 for IKZF2 in four datasets, which demonstrated the clinical value of IKZF2. Our study revealed that IKZF2 may play an essential role in the molecular function and development of LN, and might be a potential biomarker for distinguishing LN patients and healthy ones.

Abbreviations

LN  Lupus nephritis
co-DEGs  Common differentially expressed genes
ROC  Receiver operating characteristic
SLE  Systemic lupus erythematosus
ESKD  End-stage kidney disease
GWASs  Genome-wide association studies
PBCs  Peripheral blood cells
PPI  Protein–protein interaction network
GEO  Gene Expression Omnibus
TUB  Tubulointerstitium
GLOM  Glomerulus
DEGs  Differentially expressed genes
GO  Gene Ontology
KEGG  Kyoto Encyclopedia of Genes and Genomes
DO  Disease Ontology
MCODE  Molecular Complex Detection
DAVID  Database for Annotation, Visualization, and Integrate Discovery
GSEA  Gene Set Enrichment Analysis
ceRNA  Competing endogenous RNA
PCA  Principal components analysis

¹Department of Rheumatology, School of Medicine, Renji Hospital, Shanghai Jiaotong University, Shanghai 200001, China. ²State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China. ³These authors contributed equally: Mi Zhou, Yuening Kang and Jun Li. *email: rxli@sjtu.edu.cn; lu_liangjing@163.com
Systemic lupus erythematosus (SLE) is a multifactorial and multisystem autoimmune disease that commonly affects the kidneys\(^1\). LN is a major driver of mortality of SLE patients\(^2\), and up to 20% of SLE patients who have been afflicted by LN will finally develop end-stage kidney disease (ESKD) within the first decade of the disease course\(^3\). However, on the one hand, the clinical detection of early disease can be challenging because patients often lack overt signs of kidney disease\(^4\), and the conventional parameters such as proteinuria and serum creatinine are insensitive and non-specific\(^5\). On the other hand, current therapies are not sufficiently efficacious to control LN and not all patients show adequate treatment responses\(^6\). Thus, it is critical and necessary to search for new biomarkers.

Identifying the molecular mechanisms underlying the disease pathogenesis is an important step toward developing novel therapeutic targets for LN\(^7\). Genome-wide association studies (GWAS) in SLE have revealed more than 80 susceptibility loci\(^8-10\). Nevertheless, the identified risk variants could only explain a small portion of heritability of LN\(^11\). Besides, gene expression, as an intermediate phenotype between DNA and disease phenotypic variation, can provide us with a new perspective to our understanding of the molecular mechanisms leading to LN\(^12\). IKZF2, also known as HELIOS, was initially thought to be restricted to Treg cells\(^13,14\). In Treg cells, it was claimed to control IL2 expression by promoting binding of Foxp3 to the IL2 promoter\(^15\). It is expressed in other immune cells, epithelial tissues of the gut, the respiratory tract, and the tubules of the kidney\(^13\). Interestingly, it was also identified as a lupus susceptibility locus by GWAS\(^6,16\), and a recent research found it was enriched to regulating the related gene for T cells, B cells, and peripheral blood cells (PBCs) in LN\(^17\). However, its exact molecular function in LN is still unknown\(^18\).

In our study, we carried out a meta-analysis of publicly available data including not only tubulointerstitium but also glomerulus tissue samples from LN patients and pre-transplant healthy living donors (LD), and we identified common differentially expressed genes (co-DEGs) for each dataset. Among 26 co-DEGs, we selected IKZF2 for further analysis based on the GWAS\(^6-10\) and previous researches\(^9,13-18\). Then, we analyzed potential molecular mechanisms of co-DEGs and IKZF2 in LN. Protein–protein interaction network (PPI) network and ceRNA network of IKZF2 were also constructed to explore the possible targets of IKZF2. Moreover, we performed immune infiltration analysis and evaluated clinical value of IKZF2. Overall, this is the first study applying multiple datasets to understand the molecular mechanism of IKZF2 in LN in detail.

### Materials and methods

**Data collection.** We mined the Gene Expression Omnibus (GEO) database\(^19\) to find publicly available gene expression datasets for LN including both tubulointerstitium or glomerulus tissue samples. The datasets of GSE32591\(^20\) and GSE113342\(^21\) were selected and downloaded by the R package GEOquery\(^22\). Each dataset was divided into two groups of tubulointerstitium (TUB) or glomerulus (GLOM) for subsequent analysis. For details of the LN cases and LD included in the two gene expression profiles, see Table 1.

| ID        | GLOM (LN) | GLOM (LD) | TUB (LN) | TUB (LD) |
|-----------|-----------|-----------|----------|----------|
| GSE32591  | 32        | 14        | 32       | 15       |
| GSE113342 | 28        | 6         | 28       | 10       |

Table 1. Details of the LN cases and LD included in the two gene expression profiles.

**Common differentially expressed genes (co-DEGs).** Differentially expressed genes (DEGs) between LN cases and LD were identified by R package Limma\(^23\) in the two gene expression profiles for the tubulointerstitium and glomerulus tissue type, respectively. The criteria of DEGs were defined with adjust \(P\)-value < 0.05, and up/down-regulated DEGs was defined with adjust \(p\)-value < 0.05 and \(|\log FC| > 1\). Then, the co-DEGs was the intersection of the four gene sets. Thus, the DEGs of each dataset were also integrated with R package “RobustRankAggreg”\(^24\), one meta-analysis method, in order to make the result robust. In all, the results of DEGs were visualized by the R package ggplot2\(^25\), heatmap\(^26\), and VennDiagram\(^27\).

**Functional enrichment analysis of co-DEGs.** Gene Ontology (GO) analysis is commonly conducted for large-scale functional enrichment study\(^28\). Kyoto Encyclopedia of Genes and Genomes (KEGG, [www.kegg.jp/kegg/kegg1.html](https://www.kegg.jp/kegg/kegg1.html)) pathway is a database regarding diseases, genomes, biological pathways, chemical and drugs substances\(^29\). Disease Ontology (DO)\(^30\) annotates genes based on human diseases, helping to transform molecular discovery from high-throughput data into clinical relevance. We used the R package clusterProfiler\(^31\) to perform GO and KEGG analysis, and used the R package DOSE\(^32\) to conduct DO annotation. \(P\)-value < 0.05 was considered to be statistically significant for enrichment analysis.

**IKZF2-associated protein–protein interaction (PPI) network.** STRING\(^33\) is a public data source of known and predicted protein–protein interactions (PPI), from which we extracted the PPI of IKZF2-associated genes and used cytoscape\(^34\) to perform visualization. Then, we applied cytoscape plugin Molecular Complex...
Detection (MCODE)\cite{35} to identify highly interconnected clusters in the IKZF2-associated PPI network, and annotated GO terms of the specific cluster by ClueGO\cite{36} plug-in. Besides, the Database for Annotation, Visualization, and Integrate Discovery (DAVID)\cite{37} and Gene Set Enrichment Analysis (GSEA)\cite{38} were also used to understand biological significance underlying the IKZF2-associated PPI network.

**Construction of ceRNA network.** To further analyze the potential targets and biological function of IKZF2 in LN, we constructed a competing endogenous RNA (ceRNA)\cite{39} network related to IKZF2 including the mRNA, miRNA and lncRNA. The prediction of IKZF2-related miRNA targets was obtained from the starBase (http://starbase.sysu.edu.cn/index.php)\cite{40}, miRDB (http://www.mirdb.org)\cite{41}, Targetscan\cite{42} (http://www.targetscan.org/vert_72/), TarBase.v.8 (http://carolina.imis.athena-innovation.gr)\cite{43}, miRNA-related IncRNA targets were predicted by mirNet (https://www.mirnet.ca)\cite{44}. Based on miRNA-mRNA and miRNA-lncRNA, a ceRNA network was established and visualized using Cytoscape.

**Table 2.** Differentially expressed gene.

| Group       | DEGs  | Up-genes | Down-genes |
|-------------|-------|----------|------------|
| GSE32591- GLOM | 4442  | 101      | 250        |
| GSE32591- TUB  | 5823  | 25       | 104        |
| GSE113342- GLOM | 158   | 45       | 47         |
| GSE113342- TUB  | 199   | 33       | 42         |

**Figure 1.** Volcano dots of differentially expressed genes (DEGs) from GSE32591 and GSE113342. Comparing with GSE113342, more DEGs were obtained from GSE32591. There were 4442 DEGs in GSE32591_GLOM group (A) and 5823 DEGs in GSE32591_TUB group (B), as well as 158 DEGs in GSE113342_GLOM (C) and 199 DEGs in GSE113342_TUB (D).
Immune infiltration analysis. For the co-DEGs between LN cases and LD of the tubulointerstitium or glomerulus group in the two datasets, immune infiltration analysis was performed using CIBERSORT\(^4\), which was based linear support vector regression to deconvolute relative abundance of 22 infiltrated immune cells. Principal components analysis (PCA) was applied to illustrate the differences in immune infiltrating cells between LN cases and LD, and the stromal/immune scores of each sample were shown by the R package pheatmap\(^2\). Besides, the correlation between the IKZF2 expression and the types of 22 immune cell was determined by Spearman’s correlation test\(^4\).

Clinical characteristics correlation analysis and relative risk assessment. Spearman’s correlation test\(^4\) was adopted to explore the correlation between IKZF2 expression and clinical features. Receiver operating characteristic (ROC) analysis was applied to evaluate whether the expression of IKZF2 could distinguish LN tissues from normal tissues. R package pROC\(^7\) was used to calculate the area under the ROC curve and plot ROC curve.

Results
The identification of co-DEGs. To elucidate DEG between the sample from the LN cases and LD, the four samples from GSE32951 and GSE113342 datasets was analyzed using R package limma to identify the DEG and its related up/down-regulated genes according to the cutoff threshold (adjust \(P\)-value < 0.05, and |logFC|> 1 (Table 2 and Fig. 1A–D). A total of 26 co-DEGs were observed in the integration of the four sets of data (Fig. 2A). The color of heatmap represents each co-DEG expression in different sample (Fig. 3A–D). These results revealed that the 26 co-DEGs can easily distinguished LN samples from LD.

Based on the GWASs\(^8\) and previous researches\(^9\)–\(^10\), we selected IKZF2 as the focus of the follow-up research. Compared with normal controls, the expression value of IKZF2 was significantly downregulated in the LN samples (Fig. 2B). Besides, IKZF2 was also shown in the first 10 downregulated DEGs of all datasets determined by “RobustRankAgreg” (Fig. 2C).

Functional enrichment analysis of co-DEGs. In order to get further insight into the correlation between co-DEGs and biological processes, molecular functions, cellular components, biological pathways and specific diseases, functional enrichment was conducted. Firstly, GO analysis was conducted to reveal the biologi-
The characteristics of 26 co-DEGs. In the biological process category, these co-DEGs were mainly enriched in the mononuclear cell proliferation, leukocyte proliferation, regulation of inflammatory, etc. And from the point of molecular function, the hub mostly enriched GO terms included cytokine receptor binding, immune receptor activity, tumor necrosis factor and death receptor. Moreover, the cellular component category exhibited that co-DEGs mainly concentrated on the secretory granule membrane, membrane raft, membrane microdomain and membrane region (Fig. 4 A and Table 3).

The KEGG analysis upon the 26 co-DEGs showed that Epstein-Barr virus infection, complement and coagulation cascades and osteoclast differentiation, were mostly enriched terms of co-DEG (Fig. 4B and Table 4). The details of the significantly enriched pathways related to SLE were especially presented in Fig. 4E (https://www.kegg.jp/entry/map05322). The 26 co-DEGs were also mainly enriched in a number of signal pathways, which were highly related to cell proliferation and cell senescence including NFκB signaling pathway, MAPK signaling pathway, Th17 cell differentiation and Toll-like receptor signaling pathway (Fig. 4B-D). To explore the potential effects of co-DEGs upon disease, DO analysis was performed. The results depicted that co-DEGs were mostly enriched in kidney disease, urinary system disease, lymphoblastic leukemia, membranous glomerulonephritis and Human immunodeficiency virus infectious disease. (Fig. 4C and Table 5).
Figure 4. GO functional enrichment analysis, KEGG pathway enrichment analysis, and DO enrichment analysis. Three ways of enrichment analyses were performed upon the 26 co-DEGs. (A) GO analysis. The adjust p-value was calculated as x-axis and the GO terms was denoted as y-axis. Blue-purple strip, orange strip and range strip represented molecular function, cellular components and biological processes respectively. (B) The bubble diagram for the KEGG analysis results. Distinct size of dots represented the number of genes enriched in the corresponding pathway. (C) DO analysis. The x-axis represented 26 co-DEGs and the y-axis represented DO terms. (D) the co-DEGs in the enriched GO terms were particularly exhibited. (E) KEGG pathway associated with lupus nephritis.
Protein–protein interaction network (PPI) network construction. To identify the possible influence of IKZF2 on the development of LN, PPI network (Fig. 5A, B) was extracted from STRING database, comprising of 25 genes and 144 pairs of interaction. Enrichment analysis was carried out upon these IKZF2-associated genes by DAVID database. The results showed that those were mainly related with MAPK cascade, EGFR signal pathway, ERBB2 signal pathway, actively regulation of GTPase activity, Fc-ε receptor signal pathway, Phosphatidylinositol mediated signal transduction, Leukocyte migration, etc. (Fig. 5C and Table 6). Moreover, categorization by signal pathway revealed that Chronic myeloid leukemia, ErbB signal pathway, Glioma,
Figure 5. Protein–protein interaction network and its associated functional clusters. (A) STRING database was utilized to construct IKZF2 correlated PPI network. (B) Cytoscape was applied for visualization of the PPI network. Then, cytoscape plug-in MCODE was used to identify highly interconnected clusters and eventually the functional cluster 1 was obtained. (C) DAVID was adopted to annotate GO terms of special functional cluster from PPI network. (D) GO terms of the functional cluster 1 was annotated through ClueGO plug-in. (E), (F) GSEA was performed upon the genes involved in IKZF2 related PPI network. (G) pathway enrichment analysis on the genes of functional clusers from PPI network by the means of DAVID. Abbreviation, PPI protein–protein interaction network, GSEA gene set enrichment analysis, DAVID the Database for Annotation, Visualization, and Integrate Discovery.
Neurotrophin signaling pathway, PI3K-Akt signal pathway, Estrogen signal pathway and T cell receptor signal pathway were the hub enriched pathways (Fig. 5G and Table 7).

In order to further verify the latent functional implication of IKZF2 correlated genes in the PPI network, GESA analysis was conducted, and four mostly enriched GO terms were observed, including Cell intima binding organelle, Multicellular biological development, Multicellular biological process and Animal organ development (Fig. 5E). In addition, the genes in PPI network were mainly enriched in PI3K-Akt signal pathway, The cancer related pathway, JAK-STAT signal pathway and Ras signaling pathway (Fig. 5F). Next, we applied cytoscape plugin Molecular Complex Detection (MCODE) to identify highly interplayed clusters in the IKZF2 associated PPI network and finally obtained two functional clusters. And then, we utilized clueGO to annotate functional cluster 1 for its higher score (Fig. 5D). The results showed that multiple genes in functional cluster 1 were mainly enriched in two or three kinds of biological functions and pathways.

### Table 6. GO analysis of PPI.

| Go terms                                               | P value     |
|--------------------------------------------------------|-------------|
| MAPK cascade                                           | 1.24E-14    |
| Epidermal growth factor receptor signaling pathway     | 1.01E-10    |
| ERBB2 signaling pathway                                | 1.47E-09    |
| Positive regulation of GTPase activity                 | 2.78E-08    |
| Fc-epsilon receptor signaling pathway                  | 1.13E-07    |
| Phosphatidylinositol-mediated signaling                 | 2.8E-07     |
| Leukocyte migration                                    | 5.65E-07    |
| Regulation of phosphatidylinositol 3-kinase signaling   | 3.57E-06    |
| Protein phosphorylation                                | 2.64E-05    |
| Positive regulation of transcription from RNA polymerase II promoter | 2.94E-05 |
| Axon guidance                                           | 5.96E-05    |

### Table 7. KEGG analysis of PPI.

| Pathway                                      | P value     |
|----------------------------------------------|-------------|
| Chronic myeloid leukemia                      | 1.16E-12    |
| ErbB signaling pathway                        | 5.59E-12    |
| Glioma                                       | 5.17E-11    |
| Neurotrophin signaling pathway                | 7.77E-11    |
| PI3K-Akt signaling pathway                    | 5.42E-10    |
| Endometrial cancer                            | 1.05E-09    |
| Estrogen signaling pathway                    | 1.06E-09    |
| T cell receptor signaling pathway             | 1.14E-09    |
| Non-small cell lung cancer                    | 1.66E-09    |

Neurotrophin signaling pathway, PI3K-Akt signal pathway, Estrogen signal pathway and T cell receptor signal pathway were the hub enriched pathways (Fig. 5G and Table 7).

In order to further verify the latent functional implication of IKZF2 correlated genes in the PPI network, GESA analysis was conducted, and four mostly enriched GO terms were observed, including Cell intima binding organelle, Multicellular biological development, Multicellular biological process and Animal organ development (Fig. 5E). In addition, the genes in PPI network were mainly enriched in PI3K-Akt signal pathway, The cancer related pathway, JAK-STAT signal pathway and Ras signaling pathway (Fig. 5F). Next, we applied cytoscape plugin Molecular Complex Detection (MCODE) to identify highly interplayed clusters in the IKZF2 associated PPI network and finally obtained two functional clusters. And then, we utilized clueGO to annotate functional cluster 1 for its higher score (Fig. 5D). The results showed that multiple genes in functional cluster 1 were mainly enriched in two or three kinds of biological functions and pathways.

**Construction of the ceRNA regulatory network.** We next set out to better comprehend the potential regulation mechanism of IKZF2 in the progression of LN by establishing IKZF2 related ceRNA regulatory network. First, we utilized four databases to predict miRNA, which targeted the gene expression of IKZF2 and eventually obtained 22 unique miRNAs in all databases (Fig. 6A). Subsequently, 1178 LncRNA interplaying with the above 22 miRNAs were forecast based on miRNet database. And finally, cytoscape software was utilized for visualizing ceRNA regulatory network, which was composed of LncRNA-miRNA-mRNA interaction information. The cytohub plug-in in cytoscape was used to identify the key LncRNA-miRNA-mRNA regulatory mechanism in the ceRNA network (Fig. 7A and Supplementary1_miRNA.xls).
T cell, NK cell activated, neutrophils, and other immune cells (Figs. 6, 7). Between LN/LD samples, the results obviously showed the significant correlation between IKZF2 and kinds of γδT cells, T follicular helper, M1 Macrophage, Mast cell resting and Dendritic cell activated. In GLOM group GSE113342 was also analyzed in our study. Overall, in TUB group between LN/LD samples, the significant difference of abundance between LN group and normal one. (F) The heatmap of correlation between the expression level of IKZF2 and the relative abundance of 22 kinds of immune cells (G). (K) Immune infiltration abundance of co-DEGs and the correlation between 22 types of immune cells from TUB group of GSE32591 dataset.

Immune infiltration analysis. We further analyzed the relative abundance of immune infiltration of co-DEGs in 22 immune cells types from LN and normal tissues. The results exhibited that there was a significant distinction between LN samples and LD in the GLOM group from GSE32591 dataset (Fig. 6C). PCA analysis also confirmed this difference of immune cell infiltration between LN samples and LD (Fig. 6D). Subsequently, we conducted statistical analysis upon the 22 immune cells types to determine their difference in the two groups. Statistical analysis revealed that the significant difference of abundance in the 22 immune cells types were mainly focus on the following immune cells: neutrophils, monocytes, eosinophils and macrophages M0 (Fig. 6B). Moreover, Spearman correlation was calculated to reveal the link between 22 types of immune cells in LN tissues. The results indicated four immune cell subsets, including naïve T cells, monocytes, neutrophils, and γδT cells, were significantly correlated (Fig. 6E). The results also revealed the significant correlation between IKZF2 and naïve B cell, NK cell activation, NK cell rest and other immune cells. (Fig. 6F).

In addition, the immune infiltration between LN/LD samples from other group of datasets, including TUB group in the dataset GSE32591 (Fig. 6G–K), GLOM group (Fig. 7B–F) and TUB group (Fig. 7G–K) in the dataset GSE113342 was also analyzed in our study. Overall, in TUB group between LN/LD samples, the significant difference of abundance in the 22 immune cells types were mainly focus on the following immune cells (Cor > 0.5): γδT cells, T follicular helper, M1 Macrophage, Mast cell resting and Dendritic cell activated. In GLOM group between LN/LD samples, the results obviously showed the significant correlation between IKZF2 and kinds of T cell, NK cell activated, neutrophils, and other immune cells (Figs. 6, 7).

Clinical characteristics correlation analysis and relative risk assessment. In order to verify the clinical significance of IKZF2, we performed correlation analysis between the gene expression level of IKZF2 and the age of LN patients and related LN class. Although, no significant correlation was found between the IKZF2 expression level and the age of related patients (Fig. 8A/C), the results exhibited that the IKZF2 expression was related with the disease grade of LN.

In the GLOM group, higher LN class revealed higher expression of IKZF2. In the TUB group, on the contrary, higher disease grade of LN denoted lower expression of IKZF2. (Fig. 8B/D) After that, we next explored the correlation of IKZF2 with other clinicopathological features. Spearman’s correlation analysis were used to determine the association of immune cell infiltration and clinical characteristics. The results showed the significant correlation of IKZF2 with the disease grade of LN (Fig. 8B/D). The results also showed that the expression level of IKZF2 was negatively correlated with the disease activity index (Fig. 8C/E). Moreover, the expression level of IKZF2 was negatively correlated with the age of patients (Fig. 8C/F). The results of correlation analysis indicated that IKZF2 might be a potential biomarker for distinguishing LN patients and healthy ones. (Fig. 8E–H).

Validation with other GEO datasets. To verify our conclusion in a broader range, we analyzed the differentially expressed level of IKZF2 between LN and normal by datasets of GSE9842249 and GSE10494850 (Fig. 9A–C). The areas under the ROC curves were 0.721, 0.80, 0.682, and 0.859 for IKZF2 in GSE32591_GLOM, GSE32591_TUB,GSE113342_GLOM, and GSE113342_TUB datasets, respectively. The results showed that IKZF2 could distinguish LN samples and normal samples. Moreover, we valued IKZF2 expression between IgA Nephropathy and normal by datasets of GSE10494850 and GSE115857 (Fig. 9D/F), and IKZF2 didn’t exhibit significant transcriptional differences in the two datasets. We also examined the differentially expressed level of IKZF2 between ANCA associated vasculitis and normal by datasets of GSE10494850 and GSE108113 (Fig. 9E/H). IKZF2 was downregulated in ANCA associated vasculitis of GSE104948, but no significant difference was found in GSE108113.

Discussion

SLE is a chronic autoimmune disease characterized by loss of immune tolerance and various organ or tissue damage. About 25–50% patients diagnosed with SLE have symptoms of LN at onset, and approximately 60% of adult patients with SLE develop renal signs or symptoms during the disease course. LN remains ed a major cause of morbidity and death for SLE patients. Laboratory test has made great strides in diagnosis and therapy, but the incidence of renal flares ranged from 27 to 66%. However, SLE is a multifactorial process with potential involvement of different organs, and varied in its clinical severity, and the pathogenesis remains unclear. Hence, more accurate prediction and evaluation remain the great challenges for the improvement of LN outcomes.

The tissue injury or clinical manifestations are influenced by genetic, epigenetic, environmental, hormonal, and immunoregulatory factors. Many genes which encode related proteins have been found potential candidates associated with SLE, such as DNase I, DNasey, DNase III and TREX1 involved in DNA clearance, C1q and C4 od the complement pathway and other functions (AC5P, AGSS, FASL and STING) and Fc receptors. More than 80 genes have been identified associated with SLE by GWAS. Multiple signaling pathways were involved in LN, and immune infiltrate was exhibited in an ascending order from red to blue. The pink strip and the blue-green one represented LN group and normal control respectively. (D) The distinction of immunocyte infiltration between LN and normal group based on PCA algorithm. (E) Statistics on the distinction of 22 types of immune cell abundance between LN group and normal one. (F) The heatmap of correlation between the expression level of IKZF2 and the relative abundance of 22 kinds of immune cells (G). (K) Immune infiltration abundance of co-DEGs and the correlation between 22 types of immune cells from TUB group of GSE32591 dataset.
Figure 7. Immune infiltration analysis. (A) The cytohub plug-in in cytoscape was utilized for visualizing ceRNA regulatory network including lncRNA-miRNA-mRNA interaction information. (B)–(E) Immune infiltration abundance of co-DEGs and the correlation between 22 types of immune cells from GLOM group of GSE11342 dataset. (F)–(K) Immune infiltration abundance of co-DEGs and the correlation between 22 types of immune cells from TUB group of GSE11342 dataset.
Figure 8. Clinical characteristics correlation analysis and relative risk assessment. (A/C) Correlation analysis between the expression level of IKZF2 and the age of LN patients. The x-axis represented the age of LN patients, and y-axis denoted the expression of IKZF2 in the GLOM and TUB group from GSE11342 dataset. Different color represented distinct group of age. (B/D) Correlation analysis between the expression level of IKZF2 and the lupus nephritis class. The x-axis represented the expression of IKZF2 in the GLOM and TUB group from GSE11342 dataset, and y-axis represented the disease grade of lupus nephritis, which included class3, class3 + 5, class4 and class4 + 5. (E–H) ROC curve of IKZF2 expression level to distinguish LN samples and LD in the four sets of data. The results showed that IKZF2 revealed a good AUC value in GSE32591_GLOM, GSE32591_TUB and GSE113342_TUB datasets.
Figure 9. Validation with other GEO datasets. (A) ROC curve of IKZF family expression level to distinguish LN samples and LD in the dataset of GSE98422. (B) Comparison of IKZF2 expression value between LN/LD samples by datasets of GSE98422 and GSE104948. (C–F) Comparison of IKZF2 expression value between IgA nephropathy and normal by datasets of GSE104948 and GSE115857. (G–H) Comparison of IKZF2 expression value between ANCA associated vasculitis and normal by datasets of GSE104948 and GSE108113.
including type I interferon production (IRF5, IRF7, TLR7, TLR8, TLR9 and STAT4)\textsuperscript{34}, nuclear factor-kappa B (NF-kB) signaling pathway\textsuperscript{35} and effector and regulatory CD4(+) T cells especially Th 17 cells\textsuperscript{36}. People have focused on the possibility that viruses may trigger SLE for decades. It suggested that CD8 + T-cells are unable to control Epstein–Barr virus (EBV)-infected B cells in SLE patients, because of the higher viral load and faster seroconversion to infection and the molecular similarity between EBV nuclear antigen 1 and the common lupus autoantigen Ro\textsuperscript{37–39}.

In the current study, gene expression analysis was applied to identify co-DEGs and possible predictor factors between LN patients with healthy controls. A total of 26 co-DEGs, 717 miRNA and 10,125 lncRNA were identified and the PPI network was conducted. Most of the variable gene and pathways involved have been previously described in previous researches. Based on the analysis of co-DEGs expression, the IKZF2 was found to be down-regulated in the GLOM group and TUB group of the disease group. In order to explore the potential function of IKZF2 in LN patients, we analyzed the abundance of co-DEGs in 22 types of immune cells in LN tissues and normal tissues. The IKZF2 was correlated with immune cells including initial B cells and NK cell. The result showed the excellent value of IKZF2 for the AUC and we can substantially distinguish LN and healthy controls tissues.

The IKZF2 gene encodes a member of the Ikaros family of zinc-finger proteins including Ikaros, Aiolos and HELIOS, which are hematopoietic-specific transcription factors involved in the regulation of lymphocyte development\textsuperscript{60}. IKZF2, also known as HELIOS, a transcription factor expressing in 60–70% Foxp3+ Treg cells, has recently been demonstrated that it played a role in controlling certain aspects of Treg-suppressive function, differentiation, and survival\textsuperscript{61}. Treg cells play an important role in maintaining immune self-tolerance by suppressing effector T and B cells. And it’s believed that abnormalities of Treg cells may lead to T and B cell hyperactivity and contribute to the pathogenesis of SLE. It’s reported that Tregs expressing IKZF2 have stronger immunosuppressive capacity\textsuperscript{62}. In vitro experiments, it was found that IKZF2 deficient mice would acquire an auto-inflammatory phenotype similar to rheumatoid arthritis. And activated CD4 + and CD8 + T-cells, T-follicular helper cells and germinal center B-cells were noticed increased leading to production of autoantibody in circulation\textsuperscript{63}. In recent study, CD4 + CD25 − Foxp3 + T cells major expressing IKZF2 were indicated significantly increased in active SLE patients. Detailed cohort analysis revealed increased proportions of CD4 + CD25 − Foxp3 + T cells in SLE patients especially LN patients. And the same type T cell was detected in urine sediment samples of patients with active glomerulonephritis and correlated with the extent of proteinuria\textsuperscript{64}. Related data suggested that association between the percentage of FoxP3 + IKZF2 + Treg cells and the SLEDAI score is positive. There is a higher percentage of FoxP3 + IKZF2 + Treg cells in patients with more active disease compared to the patients with inactive disease or healthy donors\textsuperscript{65}. Thus, it’s assumed that CD4 + CD25 − Foxp3 + T cells might serve as an important tool to recognize and monitor SLE patients with renal involvement. However, IKZF2 is considered as a marker of T cell activation and proliferation\textsuperscript{66,67}. It’s possible that IKZF2 might become a new index for prediction and evaluation of LN.

Our study is the first study to explore the molecular mechanisms of IKZF2 in LN by multiple datasets and applying the immune infiltration analysis. Moreover, the AUC of IKZF2 expression had very high diagnostic value in each dataset based on ROC curves, revealing it might be a potential biomarker for distinguishing LN patients and healthy ones. However, due to the ethical and traumatic reasons, SLE patients without nephritis may not be punctured by the kidney. We cannot compare patients with nephritis against patients without nephritis. Nevertheless, we tried our best to obtain all available data and also analyzed clinical characteristics correlation and relative risk assessment. In addition, we valued IKZF2 expression in IgA Nephropathy and ANCA associated vasculitis, verifying our conclusion in a broader range. Besides, further experimental verifications are necessary to illuminate the biological functions of IKZF2 in LN. Subsequently, we would like to perform more careful examinations of the function of IKZF2 in LN by combining in vivo and in vitro experiments. Therefore, we consider that our study may help to investigate the progress of LN, and that IKZF2 may become biomarkers and therapeutic targets of LN in the future.

Conclusion

The absence of our patient cohort makes it difficult to verify the results. However, previous researches focused on the expression of IKZF2 in peripheral blood to evaluate the prognosis of the disease, and our study provided some evidence and new ideas about new sights for prediction and evaluation of LN. Our bioinformatics analysis identified the gene IKZF2 may be critical in the prediction and evaluation of disease activity. We hope this study may provide some potential evidence and ideas for the future treatment and monitoring of LN from new insights.

Received: 20 December 2021; Accepted: 23 May 2022
Published online: 10 June 2022

References

1. Parikh, S. V., Almaani, S., Brodsky, S. & Rovin, B. H. Update on Lupus Nephritis: Core Curriculum 2020. Am. J. Kidney Dis. 76(2), 265–281 (2020).
2. Tsokos, G. C. Systemic lupus erythematosus. N. Engl. J. Med. 365(22), 2110–2121 (2011).
3. Tektonidou, M. G., Dasgupta, A. & Ward, M. M. Risk of end-stage renal disease in patients with lupus Nephritis, 1971–2015: a systematic review and bayesian meta-analysis. Arthritis. Rheumatol. (Hoboken, NJ). 68(6), 1432–1441 (2016).
4. Cameron, J. J. Lupus nephritis. J. Am. Soc. Nephrol. 10(2), 413–424 (1999).
5. Yu, K. Y. C. et al. Serum syndecan-1, hyaluronan and thrombomodulin levels in patients with lupus nephritis. Rheumatol. (Oxf.) 60(2), 737–750 (2021).
6. Malvar, A. et al. Kidney biopsy-based management of maintenance immunosuppression is safe and may ameliorate flare rate in lupus nephritis. Kidney Int. 97(1), 156–162 (2020).
50. Grayson, P. C. et al. Metabolic pathways and immunometabolism in rare kidney diseases. *Ann. Rheum. Dis.* **77**(8), 1226–1233 (2018).
51. Anders, H. J. et al. Lupus nephritis. *Nat. Rev. Dis. Primers.* **6**(1), 7 (2020).
52. Sprangers, B., Monahan, M. & Appel, G. B. Diagnosis and treatment of lupus nephritis flares—an update. *Nat. Rev. Nephrol.* **8**(12), 709–717 (2012).
53. Brown, E. E., Edberg, J. C. & Kimberly, R. P. Fc receptor genes and the systemic lupus erythematosus diathesis. *Autoimmunity* **40**(4), 567–581 (2007).
54. Imgenberg-Kreuz, J. et al. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann. Rheum. Dis.* **77**(5), 736–743 (2018).
55. Cen, H. et al. Genetic interaction between genes involved in NF-kB signaling pathway in systemic lupus erythematosus. *Mol. Immunol.* **56**(4), 643–648 (2015).
56. Vincent, F. B., Northcott, M., Hoi, A., Mackay, F. & Morand, E. F. Clinical associations of serum interleukin-17 in systemic lupus erythematosus. *Arthritis Res. Ther.* **15**(4), R97 (2013).
57. Tsiokos, G. C., Magrath, I. T. & Balow, J. E. Epstein-Barr virus induces normal B cell responses but defective suppressor T cell responses in patients with systemic lupus erythematosus. *J. Immunol. (Baltim. Md: 1950)* **131**(4), 1797–1801 (1983).
58. Poole, B. D., Schofield, R. H., Harley, J. B. & James, J. A. Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus. *Autoimmunity* **39**(1), 63–70 (2006).
59. Draborg, A. H. et al. Impaired cytokine responses to Epstein-Barr virus antigens in systemic lupus erythematosus patients. *J. Immunol. Res.* **2016**, 6473204 (2016).
60. Odhams, C. A. et al. Mapping eQTLs with RNA-seq reveals novel susceptibility genes, non-coding RNAs and alternative-splicing events in systemic lupus erythematosus. *Hum. Mol. Genet.* **26**(5), 1003–1017 (2017).
61. Sebastian, M. et al. Helios Controls a Limited Subset of Regulatory T Cell Functions. *J. Immunol. (Baltim. Md: 1950)* **196**(1), 144–155 (2016).
62. Yin, Z. et al. Increased CD4(+)CD25(-)Foxp3(+) T cells in Chinese systemic lupus erythematosus: correlate with disease activity and organ involvement. *Lupus* **27**(13), 2057–2068 (2018).
63. Odhams, C. A., Cunninghame Graham, D. S. & Vyse, T. J. Profiling RNA-Seq at multiple resolutions markedly increases the number of causal eQTLs in autoimmune disease. *PLoS Genet.* **13**(10), e1007071 (2017).
64. Bonelli, M. et al. CD4(+)CD25(+)Foxp3(+) T cells: a marker for lupus nephritis?. *Arthritis Res. Ther.* **16**(2), R104 (2014).
65. Elkord, E. & Al-Ramadi, B. K. Helios expression in FoxP3(+) T regulatory cells. *Expert Opin. Biol. Ther.* **12**(11), 1423–1425 (2012).

**Author contributions**

LL and RL conceived and designed the project. MZ acquired and analyzed data. JL assisted data interpretation. MZ, JL and YK wrote the manuscript. All authors revised the manuscript and approved the final report.

**Funding**

This work was supported by the National Key Research and Development Program of China (2017YFC0909002), the National Natural Science Foundation of China (81974251), Shanghai Municipal Education Commission—Gaofeng Clinical Medicine Grant Support (20152214), Shanghai Hospital Development Center, joint research project (SHDC120181106), the pilot project construction of clinical collaboration between Chinese traditional medicine and Western medicine in Shanghai (ZY(2018–2020)-FWTX-1009), the pilot project construction of clinical collaboration between Chinese traditional medicine and Western medicine in Shanghai (ZY(2018–2020)-FWTX-1010).

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-13336-5.

**Correspondence** and requests for materials should be addressed to R.L. or L.L.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022