

基于生物信息学分析 *MX1*、*IFI44* 和 *STAT1* 在狼疮性肾炎中的作用

崔道林¹⁾, 陈春丽¹⁾, 周正宏²⁾, 龚蕾¹⁾

(1) 曲靖医学高等专科学校基础医学院; 2) 曲靖医学高等专科学校医学技术学院, 云南 曲靖 655100)

[摘要] **目的** 旨在筛选与 LN 相关的潜在生物标志物, 以期用于早期诊断、病情监测和更精准的治疗方案制定。**方法** 从基因表达谱数据库 (gene expression omnibus, GEO) 下载了 GSE22221、GSE112943、GSE99967 和 GSE32591 的基因表达数据。通过加权基因共表达网络分析 (weighted gene co-expression network analysis, WGCNA) 和微阵列数据的线性模型 (linear models for microarray data, LIMMA), 获得了交集基因。随后, 利用基因本体论 (gene ontology, GO) 和京都基因与基因组百科全书 (kyoto encyclopedia of genes and genomes, KEGG) 对这些交集基因进行了生物功能和信号通路分析。接着, 通过蛋白质-蛋白质相互作用 (protein-protein interaction, PPI) 网络分析、CytoHubba 算法、支持向量机 (support vector machine, SVM) 和随机森林 (random forest, RF) 方法, 筛选出了与 LN 高度相关的枢纽基因, 进行了受试者工作特征曲线 (receiver operating characteristic, ROC) 分析, 并利用 GSE72798 数据集对 3 个潜在的生物标志物进行了验证。**结果** WGCNA 分析获得绿黄色模块 ($P = 7.4e-40$) 和青色模块 ($P = 1.5e-14$); 利用 LIMMA 法筛选到 193 个差异表达基因; 共鉴定出 113 个 LN 相关的交集基因, GO 和 KEGG 分析显示这些基因主要富集在病毒或细菌的防御、I 型干扰素信号途径、中性粒细胞介导的免疫和 Toll 样受体信号等方面。通过 CytoHubba、SVM 和 RF 3 种方法筛选出 *MX1*、*IFI44* 和 *STAT1*, 其曲线下的面积 (area under the curve, AUC) 分别为 0.874、0.879 和 0.833。验证数据集显示, *MX1*、*IFI44* 和 *STAT1* 在 LN 患者中的表达显著高于健康人群 ($P < 0.001$)。**结论** *MX1*、*IFI44* 和 *STAT1* 在 LN 的发病机制中起到了关键作用, 可能成为 LN 的重要生物标志物和未来的潜在治疗靶点。

[关键词] 狼疮肾炎; 系统性红斑狼疮; 生物标志物; 枢纽基因; I 型干扰素

[中图分类号] R392.9 **[文献标志码]** A **[文章编号]** 2095-610X(2024)12-0105-10

Bioinformatics-Based Analysis of the Roles of *MX1*, *IFI44*, and *STAT1* in Lupus Nephritis

CUI Daolin¹⁾, CHEN Chunli¹⁾, ZHOU Zhenghong²⁾, GONG Lei¹⁾

(1) School of Basic Medical Sciences; 2) School of Medical Technology, Qujing Medical College, Qujing Yunnan 655100, China)

[Abstract] **Objective** To identify potential biomarkers associated with LN, with the goal of improving early diagnosis, disease monitoring, and the development of more precise treatment strategies. **Methods** Gene expression data were downloaded from the Gene Expression Omnibus (GEO) database for datasets GSE22221, GSE112943, GSE99967, and GSE32591. Intersecting genes were obtained through the application of weighted gene co-expression network analysis (WGCNA) and linear models for microarray data (LIMMA). Subsequently, biological function and pathway analyses were conducted on these intersecting genes using Gene Ontology (GO) and

[收稿日期] 2024-06-15

[基金项目] 云南省教育厅科学研究基金资助项目(2023J1760); 曲靖医学高等专科学校大学生科技创新资助项目(2024DZ004)

[作者简介] 崔道林(1979~), 男, 山东成武人, 理学硕士, 副教授, 主要从事狼疮肾炎研究工作。

[通信作者] 龚蕾, E-mail: yngonglei@163.com

the Kyoto Encyclopedia of Genes and Genomes (KEGG). Next, protein-protein interaction (PPI) network analysis was performed, and hub genes highly associated with LN were identified using the CytoHubba algorithm, support vector machine (SVM), and random forest (RF) methods. Receiver operating characteristic (ROC) analysis was performed, and three potential biomarkers were validated using the GSE72798 dataset. **Results** The green-yellow module ($P = 7.4e-40$) and the cyan module ($P = 1.5e-14$) were identified through WGCNA analysis. A total of 193 differentially expressed genes were identified using LIMMA, with 113 intersecting genes related to LN being identified. GO and KEGG analyses indicated that these genes were mainly enriched in viral or bacterial defense, type I interferon signaling pathway, neutrophil-mediated immunity, and Toll-like receptor signaling. MX1, IFI44, and STAT1 were identified as hub genes using CytoHubba, SVM, and RF methods, with AUC values of 0.874, 0.879, and 0.833, respectively. Validation using the GSE72798 dataset demonstrated that the expression of MX1, IFI44, and STAT1 was significantly higher in LN patients compared to healthy individuals ($P < 0.001$ for all). **Conclusion** MX1, IFI44, and STAT1 play crucial roles in the pathogenesis of LN and may serve as important biomarkers and potential therapeutic targets for LN.

[**Key words**] Lupus nephritis; Systemic lupus erythematosus; Biomarker; Hub genes; Type I interferon

系统性红斑狼疮(systemic lupus erythematosus, SLE)是1种由多种因素诱发的自身免疫性疾病,这些因素包括雌激素、紫外线、药物、污染、维生素D缺乏和病毒等^[1]。该病可影响多个器官,包括关节、肺、皮肤和肾脏。在其并发症中,狼疮性肾炎(lupus nephritis, LN)是最严重的,其病理特征为肾内淋巴组织形成、产生自身抗体和免疫复合物沉积。全球LN的发病率因地区和种族的不同而有所差异,亚洲的发病率和严重性均高于西方国家^[2]。约60%的SLE患者在10 a内发展为LN,其中10%在5 a内进展为终末期肾病(end stage renal disease, ESRD),需要透析或移植^[3]。LN是导致患者残疾和死亡的主要原因。

2021年1月, voclosporin的批准上市,结束了LN治疗数10 a无新药的局面^[4]。在初始治疗阶段,低剂量环磷酰胺或霉酚酸酯(mycophenolate mofetil, MMF)被用作高剂量环磷酰胺的替代。尽管这些治疗减少了相关副作用,但它们并不比高剂量环磷酰胺更为有效^[5]。

研究显示, LN诊断时如出现严重器官损害、高血压、白蛋白减少和肾功能下降, ESRD风险增加,这强调了早期诊断的重要性^[3]。与无LN患者相比,伴有LN的狼疮患者死亡率显著增加,并且在增生性LN患者中,5 a内约有5%至25%的死亡直接归因于肾病。此外,10%至30%的LN患者会进展至需要肾脏替代治疗的肾衰竭。患有增生性LN(III、IV或III/IV + V类)的人群面临更高的肾脏替代治疗需求风险^[6]。因此,早期诊断和治疗显著提高生存率。狼疮性肾炎的组织学分类指导治疗策略,旨在减少蛋白尿并维护肾功能。尽管肾活检是LN确诊的金标准,但由于其

侵入性,使用受到限制。关于肾损伤与LN患者肾活检实际价值的争议仍然存在^[7]。

随着表达谱芯片数据的兴起和生物信息学的进步,研究人员开始在基因水平上探讨疾病机制^[8]。开发基于基因或蛋白的生物标志物以预测和监测LN肾损害对早期诊断和治疗至关重要。WGCNA是1种描述基因关联模式的系统方法,能够筛查与疾病相关的基因模块^[9]。SVM是1种主要用于分类和回归的有监督学习模型,是深度学习之前的关键算法^[10]。RF是1种由多决策树组成的分类器集合,通过所有树的投票决定输出类别,并在属性选择中引入随机性^[11]。这些方法广泛应用于疾病基因研究。然而,利用这些方法筛选LN核心基因的研究仍然很少,大多数研究所选用的芯片数量有限且来源单一,导致筛选基因存在差异。

在本研究中,笔者选择了人类的肾和血液组织样本,采用了4个表达谱芯片并进行了批次校正。首先,笔者使用“limma”包确定了LN的差异表达基因(DEGs),并使用WGCNA构建基因共表达网络,找出与LN相关的基因模块。然后,将DEGs与基因模块交叉,并进行GO和KEGG分析。接着,结合Cytohubba插件、SVM和RF方法分析交集基因,并取这些方法得到基因的交集,确定目标核心基因。最后,利用ROC曲线和GSE72798数据集对核心基因进行了验证,并分析了它们与免疫细胞的相关性及在ceRNA调控网络中的作用,见图1。这些核心基因可能与LN的病理机制和进展紧密相关,对LN的早期诊断和治疗具有关键意义。

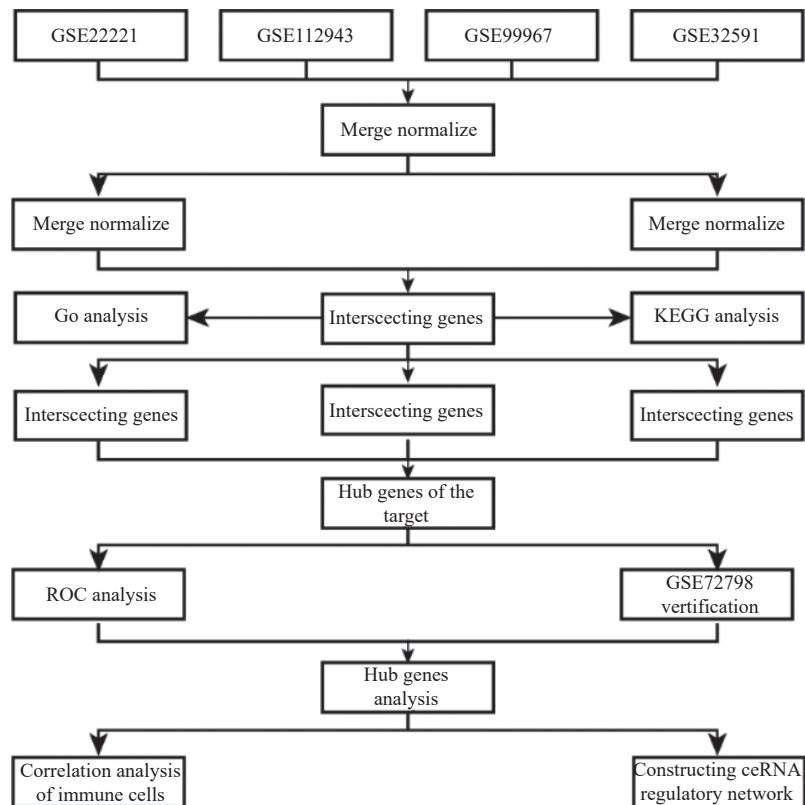


图 1 整体研究设计

Fig. 1 Overall research design

1 资料与方法

1.1 数据来源

从 GEO 数据库 (<https://www.ncbi.nlm.nih.gov/geo/>) 提取以下基因芯片数据: GSE22221 中, 15 份红斑狼疮患者血样 (GSM2159869–GSM2159883) 与 25 份健康者血样 (GSM2159829–GSM2159853); GSE112943 中, 14 份红斑狼疮患者肾脏组织样 (GSM3091782、GSM3091791–GSM3091795、GSM3091783–GSM3091790) 与 7 份健康者肾脏组织样 (GSM3091775–GSM3091781); GSE99967 中, 29 份红斑狼疮患者血样 (如 GSM2666720、GSM2666722–GSM2666725 等) 与 17 份健康者血样 (GSM2666758–GSM2666774); GSE32591 中, 64 份红斑狼疮患者肾脏组织样 (GSM807842–GSM807873、GSM8078

89–GSM807920) 与 29 份健康者肾脏组织样 (GSM807874–GSM807888、GSM807921–GSM807934); GSE72798 中, 30 份红斑狼疮患者血样 (GSM1869396–GSM1869425) 与 10 份健康者血样 (GSM1869520–GSM1869529)。注释平台包括 GPL10558、GPL21970、GPL14663 和 GPL570。训练组为 GSE22221、GSE112943、GSE99967、GSE32591, 验证组为 GSE72798, 见表 1。

1.2 数据预处理

利用 “strawberry-perl-5.32.0.1” 将表达矩阵的探针信息映射至基因名。笔者用 R-4.2.3 对原始表达矩阵进行细致处理, 通过 “limma” 包完成数据的规范化、标准化和归一化, 并使用 “sva” 包的 ComBat 函数对训练组数据执行批次修正, 同时对修正前后数据进行可视化。

表 1 实验数据集信息

Tab. 1 Experimental dataset information

基因芯片	注释平台	LN患者(n)	正常人(n)	分组信息/组	数据来源
GSE22221	GPL10558	15	25	训练	Morris A C, et al ^[12]
GSE112943	GPL10558	14	7	训练	Ko W C, et al ^[13]
GSE99967	GPL21970	29	17	训练	Wither J E, et al ^[14]
GSE32591	GPL14663	64	29	训练	Berthier C C, et al ^[15]
GSE72798	GPL570	30	10	验证	Ducreux J, et al ^[16]

1.3 加权共表达网络构建与分析

WGCNA 方法分析基因表达数据, 该方法在系统生物学中被广泛采用。用 R 语言的 WGCNA 包创建共表达网络, 通过 pickSoftThreshold 函数确定最佳的软阈值参数。接着, 建立加权邻接矩阵, 并基于拓扑重叠矩阵(TOM)的分层聚类创建相关基因模块。

1.4 交集基因的筛选和功能富集分析

用 R 语言中的“limma”包, 设置条件为 $|\log_2 \text{FC}| > 0.585$ 和 $P < 0.05$, 从 LN 和正常肾组织样本中挑选差异表达基因。通过“VennDiagram”包得到并可视化交集基因, 这些交集基因是 WGCNA 中筛选模块与差异表达基因的共同致病相关基因。借助“clusterProfiler”、“limma”、“org.Hs.eg.db”、“enrichplot”及“pathview”包, 对差异基因进行 GO 功能和 KEGG 信号通路富集分析。

1.5 Cytohubba 筛选核心基因

将筛选出的交集基因导入 STRING 数据库构建 PPI 网络, 关键参数设置: Multiple proteins; Homo sapiens; highconfidence (0.700); hide disconnected nodes in thenetwork。使用 Cytoscape3.9.1 的插件“Cytohubba”, 获取 MCC、DMNC、MNC、Degree、EcCentricity、BottleNeck、Closeness、Radiality、Betweenness 和 Stress 打分最高的前 50 个基因。再从这些基因中找出交集, 得到 Cytohubba 的核心基因。

1.6 支持向量机和随机森林筛选核心基因

RF 和 SVM 算法在多学科交叉和机器学习的推动下, 已成为基因表达谱数据分析的重要工具。利用 R 中的“glmnet”和“ggplot2”包, 笔者对交集基因进行 SVM 分析并进行了可视化。准确性最高的点和误差最低的点所对应的基因数, 均为 SVM 筛选 LN 特征基因的数量。同样地, 使用“randomForest”和“ggplot2”包对交集基因进行了 RF 分析, 并进行了可视化。决策树数量是决定模型表现的关键因素。通过交叉验证, 笔者在误差曲线上找到了最小的交叉验证误差, 并据此确定了最优决策树数量。此外, 笔者还计算了基因的重要性得分。

1.7 获得目标核心基因

通过 Cytohubba、SVM 和 RF 方法分别筛选得到核心基因, 利用 R 语言中的‘VennDiagram’包对这些核心基因进行交集分析和可视化, 最终获得目标核心基因。

1.8 验证目标核心基因

本研究使用 GSE72798 数据集作为外部验证数

据集, 利用 R 语言的“limma”、“ggpubr”和“ph-teamap”包比较目标核心基因在 LN 患者和健康人组织之间的表达差异分析。利用 R 语言的“pROC”包用以 ROC 分析并计算了 AUC, 以评价目标核心基因的预测效力。

1.9 统计学处理

所有的统计分析都使用 R 4.2.3 软件, 采用 Wilcoxon 检验比较 2 组间差异, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 联合芯片批次矫正

鉴于 LN 的复杂性和多样性, 本研究致力于探索和识别其相关的生物标记物。由于相同组织不同批次的芯片以及不同组织来源的芯片在基因表达上都存在较大的差异, 这种差异可能影响生物标记物的准确鉴定。因此, 本研究的核心目标是通过多芯片批次矫正技术, 消除这些差异, 以获得更为稳定和准确的基因表达数据, 见图 2。在没有进行多芯片批次矫正的情况下, 4 个选定的表达谱芯片之间的基因表达差异明显。为了提高数据的稳定性和一致性, 采用了多芯片批次矫正技术。矫正后, 这 4 个芯片在基因表达量上的差异显著减少, 数据呈现出 1 种归一化的分布特征, 说明矫正技术有效地消除了不同批次和不同组织来源的芯片间的差异。通过这种方法, 笔者期望在后续的生物信息学分析中获得更为准确和有力的支持, 从而更有效地识别 LN 的潜在生物标记物。

2.2 共表达网络构建及重要模块的识别

为了更深入地探索 LN 的生物学特性及其潜在生物标志物, 笔者计划应用 WGCNA 技术, 分析 LN 患者和正常对照组的芯片数据, 识别与 LN 密切相关的基因模块。基于 200 个样本(包括 122 个 LN 患者样本和 78 个正常对照样本)的联合芯片数据集, 构建了 WGCNA 网络。选择软阈值是关键步骤, 因此, 根据无尺度拟合指数 R^2 首次达到 0.9, 确定最佳软阈值为 $\beta = 7$, 见图 3A。在这一软阈值下, 网络的平均连接程度较高, 确保在分析中捕获足够的信息量。进一步采用了模块合并策略, 将模块相似性阈值(MEDissThres)设置为 0.25, 以确保合并距离较近且具有相似表达模式的模块。同时, 设定每个模块至少包含 50 个基因, 并设置剪切高度为 0.25。经过这些步骤, 共得到

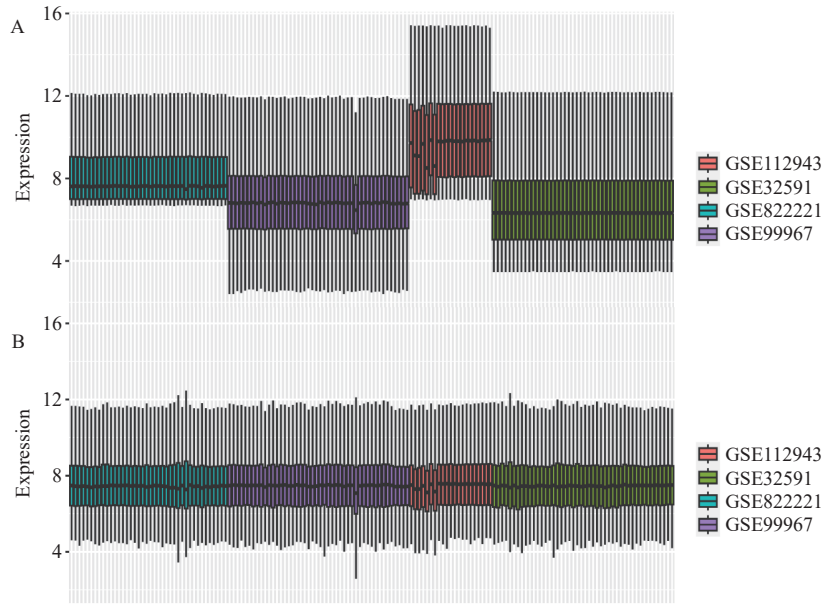


图 2 多基因芯片联合分析

Fig. 2 Multi-gene chip combined analysis

A: 矫正前分析; B: 矫正后分析。

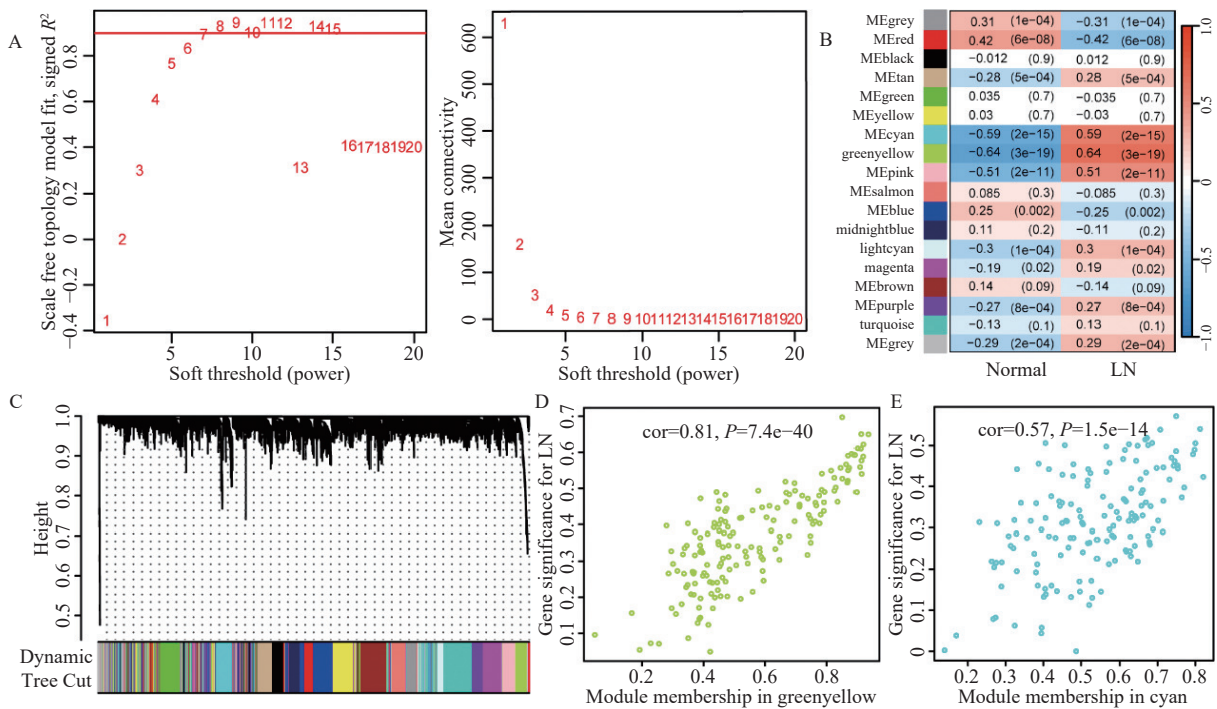


图 3 构建 WGCNA 网络

Fig. 3 Construction of the WGCNA network

A: LN 共表达模块中 power 值筛选; B: 模块特征基因与样本特征向量之间的皮尔逊相关系数; C: 筛选 LN 的共同表达模块; D: 黄绿模块和基因重要性之间的相关性, cor 表示 基因显著性(gene significance, GS)和模块成员(module membership, MM)之间的绝对相关系数; E: 蓝绿模块和基因重要性之间的相关性, cor 表示 GS 和 MM 之间的绝对相关系数。

了 18 个与 LN 相关的基因模块, 见图 3B, 3C。在随后的模块-特征关系分析中, 热图显示了各个模块与 LN 的相关性。值得注意的是, 绿黄色模块和青色模块与 LN 的相关性最为显著。具体而言, 绿黄色模块的相关性为 $cor = 0.81 (P =$

$7.4e-40)$; 青色模块的相关性为 $cor = 0.57 (P = 1.5e-14)$, 见图 3D。

2.3 LN 差异表达基因的筛选

为了深入了解 LN 的分子机制, 计划鉴定与 LN 相关的差异表达基因。在对联合基因芯片数据

进行标准化处理后, 成功鉴定出 193 个与 LN 相关的差异表达基因, 其中 183 个基因呈现上调表

达(占比为94.8%), 而 10 个基因呈现下调表达(占比为5.2%), 见表 2。

表 2 正常组与 LN 患者差异基因

Tab. 2 Differential genes between normal group and LN patients

差异基因	
上调基因	<i>IFI27, IFI44L, IFI44, MX1, LTF, RSAD2, CD163, HERC5, PLSCR1, OAS3, IFIT3, OAS1, OAS2, IFI6, IFIT1, LY96, MX2, HERC6, CREG1, SIGLEC1, EIF2AK2, CEACAM8, CIQB, IFITM1, IFIH1, IFIT2, DDX60, DYSF, ELANE, DEFA4, ABCA1, XAF1, IFITM3, TRIM22, ISG15, LCN2, RNASE6, OLFM4, APOBEC3A, IFI35, RNASE1, VSIG4, MMP9, DHX58, RNASE2, TOP2A, TLR2, LYN, HCK, HBD, MMP8, IRF7, ISG20, ZCCHC2, OLR1, CDC20, GBP2, MS4A4A, DDX58, CCL2, RRM2, SPATS2L, LHFPL2, GBP1, CTSG, CEACAM1, STAT1, CD38, ENTPD1, UBE2J1, TCN2, CXCL10, RTP4, FCER1G, TYMS, B4GALT5, CCNA2, RP2, C3AR1, SAMD9, YIPF6, CARHSP1, TAP1, IFI30, LGALS3BP, TRIM21, FAR2, STIL, PARP12, MFGE8, RNASE3, H2AFZ, CEACAM6, LAP3, YWHAH, GMPR, RIN2, SCARB2, LMNB1, MPO, LCP1, BPGM, TLR4, PRC1, LY6E, CFD, MS4A6A, TPX2, XK, LRRC32, SH3GLB1, TGM2, CTSS, OASL, COL1A2, GINS2, ANP32B, ADAR, RAPGEF5, LYZ, MARCKS, HMGB2, ATP6V1D, SPI100, FAM46C, CAMP, COL4A1, TMEM140, ANXA3, FCN1, NETO2, MYOF, EVI2A, FUT4, DNAJC13, CD36, IFI16, POU2AF1, SAMSNI, TCN1, MRI, MNDA, GIMAP4, PRCP, NFE2, CA1, STAT2, PPP1R3D, TACC3, SLPI, CIQA, S100A8, NMI, TLR1, NUSAP1, HLX, ARG1, TMOD1, REXO2, PLA2G4A, PTEN, TNFSF10, NPL, PLEK, GINS3, NOD2, LXN, DNAJC15, HIST1H2AI, MFSD1, DEGS1, FAM46A, CENPN, ITGB2, SCO2, PSMC2, COL15A1, CDC123, DHRS9, HK3, SERPINB1, NOL7, ITGAM</i>
下调基因	<i>CLIC5, ASB1, PTGDS, EPHX2, CDKN1C, TNNC2, KIR2DL3, FCER1A, TGFB3, FOSB</i>

2.4 交集基因的筛选及富集分析

为了鉴定在 LN 中与差异表达基因(differentially expressed genes, DEGs)相关的关键模块基因, 选取了绿黄色模块和青色模块的基因进行分析。与 DEGs 的交集分析鉴定出 113 个可能的 LN 相关的关键基因, 见图 4。利用 R 4.2.3 版本进行的功能富集分析表明: 在生物过程(biological process, BP)中, 这些基因主要与病毒的防御、I 型干扰素信号、细菌的防御、细胞因子信号、模式识别受体信号、RIG-I 信号、中性粒细胞介导的免疫、Toll 样受体信号以及调节性 T 细胞的免疫等过程有关; 在细胞组分(cellular component, CC)中, 显著富集于特异性颗粒、分泌颗粒管腔、细胞质囊泡腔体、初级溶酶体和吞噬细胞囊泡; 而在分子功能(molecular function, MF)中, 它们主要涉及双

链 RNA 的结合、DNA 的结合、核酸酶活性、趋化因子受体的结合以及激酶的调节活性, 见表 3。此外, KEGG 路径分析揭示了这些基因在甲型流感、COVID-19、NOD 受体信号、RIG-I 信号、金黄色葡萄球菌感染、SLE、IL-17 信号、细胞膜 DNA 感应、补体途径、程序性坏死、Toll 样受体信号以及中性粒细胞外陷阱形成等路径中可能发挥的重要作用, 见表 4。

2.5 枢纽基因的筛选

在无尺度网络中, 大多数节点只有少量的连接, 而少数节点(称为“枢纽节点”)有大量的连接。生物网络中的小部分基因(枢纽基因)可能在许多生物过程中起到非常重要的作用。为了深入探讨 LN 高度相关的枢纽基因, 笔者从交集中挑选了 113 个基因并以此构建了 PPI 网络。该网络涉及 76 个节点和 615 个互作关系。通过使用 Cytoscape 软件的 cytoHubba 插件, 成功鉴定了以下 15 个网络中的核心基因: *DDX60, IFI44, HERC5, HERC6, RTP4, DHX58, GBP1, IFI44L, IFIT1, IRF7, MX1, RSAD2, XAF1, STAT1* 和 *CXCL10*。利用 SVM 方法, 进一步从交集基因中确定了 39 个具有显著特征的基因: *GINS3, LY6E, CDC20, MARCKS, ELANE, PRC1, PTGDS, MX1, ARG1, RSAD2, FAR2, TAP1, TPX2, LGALS3BP, TCN2, RNASE1, ADAR, SPI100IFI6,*

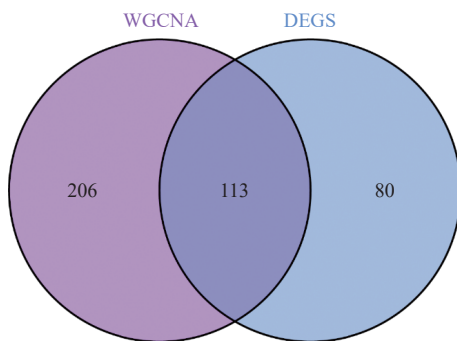


图 4 WGCNA 和 DEGs 的交集基因

Fig. 4 Intersection genes of WGCNA and DEGs

表 3 交集基因 GO 富集分析
Tab. 3 GO enrichment analysis of intersection genes

ONTOLOGY	ID	描述	P
BP	GO:0051607	defense response to virus	0.000
	GO:0060337	type I interferon signaling pathway	0.000
	GO:0042742	defense response to bacterium	0.000
	GO:0019221	cytokine-mediated signaling pathway	0.000
	GO:0002221	pattern recognition receptor signaling pathway	0.000
	GO:0039529	RIG-I signaling pathway	0.000
	GO:0002446	neutrophil mediated immunity	0.000
	GO:0002709	toll-like receptor signaling pathway	0.000
	GO:0002709	regulation of T cell mediated immunity	0.002
CC	GO:0042581	specific granule	0.000
	GO:0034774	secretory granule lumen	0.000
	GO:0060205	cytoplasmic vesicle lumen	0.000
	GO:0005766	primary lysosome	0.000
	GO:0045335	phagocytic vesicle	0.001
MF	GO:0003725	double-stranded RNA binding	0.000
	GO:0008301	DNA binding	0.006
	GO:0004518	nuclease activity	0.008
	GO:0042379	chemokine receptor binding	0.009
	GO:0019207	kinase regulator activity	0.015

表 4 交集基因 KEGG 信号通路分析

Tab. 4 KEGG pathway analysis of intersection genes

ID	描述	P
hsa05164	Influenza A	0.000
hsa05171	COVID-19	0.000
hsa04621	NOD-like receptor signaling pathway	0.000
hsa04622	RIG-I-like receptor signaling pathway	0.000
hsa05150	Staphylococcus aureus infection	0.001
hsa05322	Systemic lupus erythematosus	0.002
hsa04657	IL-17 signaling pathway	0.004
hsa04623	Cytosolic DNA-sensing pathway	0.016
hsa04610	Complement signaling pathway	0.022
hsa04217	Necroptosis	0.025
hsa04620	Toll-like receptor signaling pathway	0.037
hsa04613	Neutrophil extracellular trap formation	0.044

SAMD9、*ZCCHC2*、*CEACAM1*、*RTP4*、*MYOF*、*XAF1*、*STAT1*、*OAS2*、*IFITM3*、*IFI44*、*EIF2AK2*、*ANXA3*、*FUT4*、*IFIT2*、*TYMS*、*RAPGEF5*、*RNASE2*、*CTSG*、*SLPI*和 *LAP3*。此外, 通过 RF 算法, 鉴定了 56 个基因具有重要的诊断价值, 包括: *IFIH1*、*MX1*、*CDC20*、*RNASE1*、*LTF*、*EIF2AK2*、*RNASE2*、*TRIM22*、*IFI44*、*HMGB2*、*HERC5*、*OAS2*、*PLSCR1*、*TPX2*、*MPO*、*IFIT3*、

ELANE、*GINS2*、*OAS3*、*IFI6*、*DEFA4*、*OAS1*、*IFI27*、*RRM2*、*HERC6*、*STIL*、*LXN*、*GBP1*、*GINS3*、*SCARB2*、*FAR2*、*DDX58*、*PRC1*、*TOP2A*、*RAPGEF5*、*ARG1*、*CEACAM1*、*TCN2*、*IFI44L*、*TGFBR3*、*MX2*、*IFITM1*、*NUSAP1*、*CDKN1C*、*OLR1*、*MMP8*、*MS4A4A*、*CTSG*、*RIN2*、*SP100*、*FAM46A*、*SLPI*、*STAT1*、*IRF7*、*SAMD9*和 *TRIM21*。

综合以上 3 种方法的结果, 并通过韦恩图分析, 确定了 3 个具有重要诊断潜力的 LN 相关生物标记物, 分别是: *IFI44*、*MX1* 和 *STAT1*, 见图 5。

2.6 枢纽基因的验证分析

为了验证生物信息学方法所预测的 3 个关键基因 *IFI44*、*MX1* 和 *STAT1* 的准确性, 首先通过接收者操作特性曲线(ROC)的分析, 发现 *IFI44*、*MX1* 和 *STAT1* 在 LN 的诊断和预后上均表现出良好的鉴别能力, 其 AUC 值分别为 0.874、0.879 和 0.833, 这表明这 3 个基因在 LN 中具有较高的诊断准确性。为进一步验证这些初步结果, 采用了外部数据集 GSE72798。结果显示, 在 LN 患者与健康对照组之间, *IFI44*、*MX1* 和 *STAT1* 在血液组织中的相对表达量有统计学意义, 其中 LN 患者的表达显著高于健康人群($P < 0.001$), 见图 6。

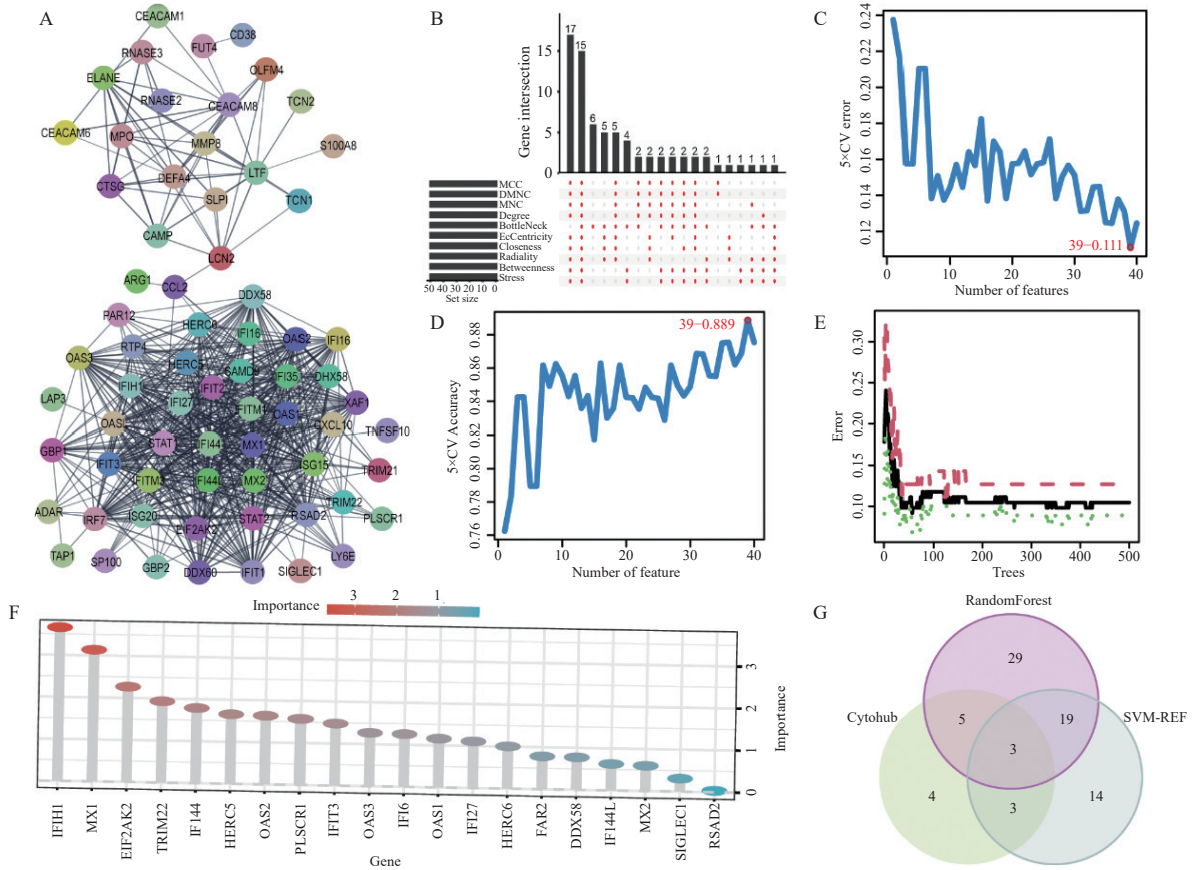


图 5 筛选枢纽基因

Fig. 5 Screening of hub genes

A: 交集基因的蛋白质互作网络; B: 由蛋白质互作网络分析确定的 10 个拓扑算法的 UpSet 图; C ~ D: 使用 SVM-RFE 算法筛选 SVM 基因; E ~ F: 从交集基因里筛选 RandomForest 基因及其重要性分析; G: 韦恩图显示了 3 种方法获得的中心基因的交集。

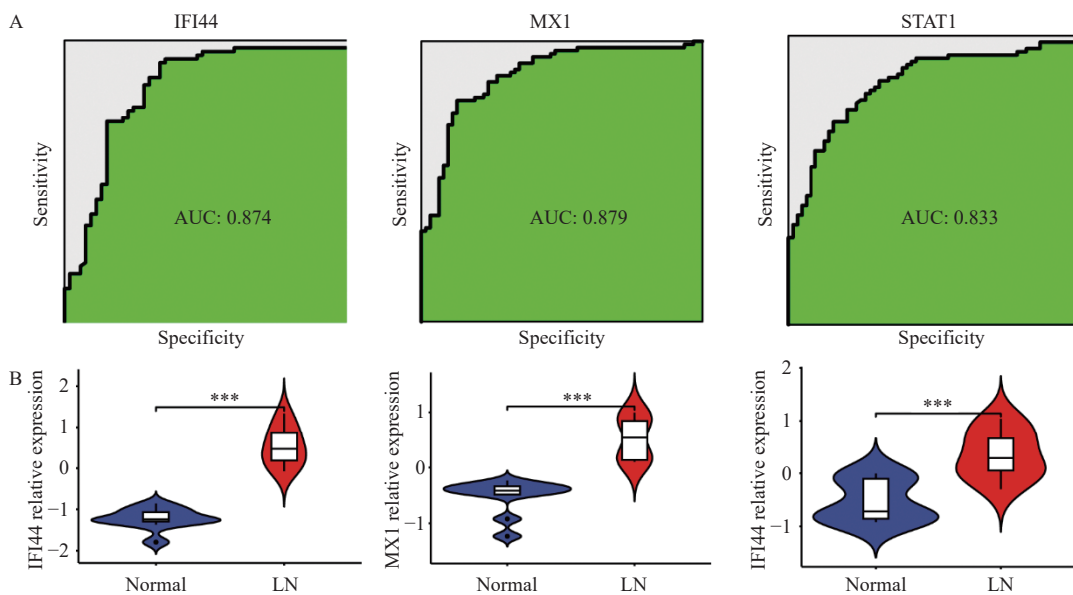


图 6 验证枢纽基因

Fig. 6 Validation of hub genes

A: *IFI44*, *MX1* 和 *STAT1* 表达水平的 ROC 曲线; B: *IFI44*, *MX1* 和 *STAT1* 在验证集中的相对表达量。*** $P < 0.001$ 。

3 讨论

3.1 I 型干扰素信号通路在 LN 发病中的关键作用

LN 是 SLE 的主要并发症, 严重影响患者的生活质量和预期寿命。本研究结合多基因芯片分析、WGCNA 与 LIMMA 技术, 筛选出与 LN 相关的模块基因及差异表达基因, 并进行了 GO 和 KEGG 功能富集分析。结果表明, 相关通路主要涉及病毒感染、I 型干扰素信号通路、细胞因子介导的信号传导及免疫反应等。研究表明, LN 的发生可能与免疫复合物、病毒感染及干扰素介导的免疫反应密切相关。I 型干扰素, 特别是 IFN- α 和 IFN- β , 在 LN 的发病中起关键作用, 模式识别受体感知病毒并激活 IFN-I 的产生, 进一步加剧肾脏损伤。因此, 干扰素信号通路在 LN 的病理机制中占据核心地位, 验证了其作为潜在治疗靶点的价值。

3.2 核心基因 *MX1*、*IFI44* 和 *STAT1* 的作用机制

通过多种生物信息学方法, 本研究鉴定了 3 个与 LN 密切相关的核心基因: *MX1*、*IFI44* 和 *STAT1*, 这些基因在 I 型干扰素信号通路中发挥重要作用。

3.2.1 *MX1* 的作用 *MX1* 是由干扰素激活的先天免疫应答蛋白, 主要通过抑制 RNA 病毒的复制发挥作用^[17]。研究显示, *MX1* 在 LN 的免疫反应和肾脏纤维化过程中起核心作用, 且与干扰素信号途径密切相关^[18-19]。抑制 *MX1* 可显著减轻肾纤维化^[18], 表明其在 LN 中的免疫调节作用。

3.2.2 *IFI44* 的作用 *IFI44* 是干扰素刺激基因, 在免疫反应中具有重要作用^[20]。研究发现, LN 患者中 *IFI44* 的表达显著增高, 可能通过调节先天免疫反应来参与 LN 的发病过程^[21], 表明 *IFI44* 可能成为 LN 的关键生物标志物。

3.2.3 *STAT1* 的作用 *STAT1* 在干扰素信号通路中扮演关键角色, 并与免疫应答密切相关^[22]。研究表明, *STAT1* 的活化与 LN 的病理变化及疾病活跃度密切相关, 尤其在巨噬细胞的聚集和肾功能损害中具有重要作用^[23-25]。尽管 *STAT1* 在 LN 中的作用已得到初步验证, 但其具体机制仍需进一步探索。

MX1、*IFI44* 和 *STAT1* 作为干扰素响应基因, 其上调表达表明 I 型干扰素信号通路的激活在 LN 中的关键作用。干扰素通过促进免疫细胞迁移、激活炎症因子及诱导细胞凋亡, 可能导致肾脏的

持续损伤和纤维化^[26-27]。针对干扰素信号通路的干预可能为 LN 的治疗提供新的思路。

3.3 研究创新点及展望

本研究首次综合应用 WGCNA、LIMMA、PPI、CytoHubba、SVM 和 RF 等多种分析方法, 成功鉴定出 *MX1*、*IFI44* 和 *STAT1* 这 3 个与 LN 密切相关的核心基因, 并通过 GO 和 KEGG 功能富集分析验证了它们在 I 型干扰素信号通路中的关键作用。这些基因不仅为 LN 的诊断和预后提供了潜在的生物标志物, 还为 LN 的靶向治疗提供了新的研究方向。未来的研究可进一步探索这些基因在 LN 中的具体机制, 并验证其在临床应用中的可行性。

[参考文献]

- [1] Bastidas Goyes A R, Mora C, Arsanios D M, et al. Systemic lupus erythematosus, a disease conditioned by the environment[J]. *Revista Colombiana de Reumatología*, 2021, 28(6): 12-20.
- [2] Barber M R W, Drenkard C, Falasinnu T, et al. Global epidemiology of systemic lupus erythematosus[J]. *Nat Rev Rheumatol*, 2021, 17(9): 515-532.
- [3] Gasparotto M, Gatto M, Binda V, et al. Lupus nephritis: Clinical presentations and outcomes in the 21st century[J]. *Rheumatology (Oxford)*, 2020, 59(Suppl5): v39-v51.
- [4] Heo YA. Voclosporin: First approval[J]. *Drugs*, 2021, 81(5): 605-610.
- [5] Mejia-Vilet J M, Malvar A, Arazi A, et al. The lupus nephritis management renaissance[J]. *Kidney Int*, 2022, 101(2): 242-255.
- [6] Parikh S V, Almaani S, Brodsky S, et al. Update on lupus nephritis: Core curriculum 2020[J]. *American Journal of Kidney Diseases*, 2020, 76(2): 265-281.
- [7] Bernard L, Wang A R, Menez S, et al. Kidney biopsy utility: Patient and clinician perspectives from the kidney precision medicine project[J]. *Kidney Med*, 2023, 5(10): 100707-100715.
- [8] Satam H, Joshi K, Mangrolia U, et al. Next-generation sequencing technology: Current trends and advancements[J]. *Biology (Basel)*, 2023, 12(7): 997-1021.
- [9] Langfelder P, Horvath S. Wgcna: An r package for

- weighted correlation network analysis[J]. *BMC Bioinformatics*, 2008, 9(12): 559–571.
- [10] Nedaie A, Najafi A A. Support vector machine with dirichlet feature mapping[J]. *Neural Netw*, 2017, 98(2): 87–101.
- [11] Ubels J, Schaefer T, Punt C, et al. Rainforest: A random forest approach to predict treatment benefit in data from (failed) clinical drug trials[J]. *Bioinformatics*, 2020, 36(Suppl_2): i601–i609.
- [12] Morris A C, Forbes–Osborne M A, Pillai L S, et al. Microarray analysis of xops–mcfp zebrafish retina identifies genes associated with rod photoreceptor degeneration and regeneration[J]. *Invest Ophthalmol Vis Sci*, 2011, 52(5): 2255–2266.
- [13] Ko WC C, Li L, Young T R, et al. Gene expression profiling in the skin reveals strong similarities between subacute and chronic cutaneous lupus that are distinct from lupus nephritis[J]. *J Invest Dermatol*, 2021, 141(12): 2808–2819.
- [14] Wither J E, Prokopec S D, Noamani B, 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[J]. *PLoS One*, 2018, 13(5) : e0196117.
- [15] Berthier C C, Bethunaickan R, Gonzalez–Rivera T, et al. Cross–species transcriptional network analysis defines shared inflammatory responses in murine and human lupus nephritis[J]. *J Immunol*, 2012, 189(2): 988–1001.
- [16] Ducreux J, Houssiau F A, Vandepapeli è re P, et al. Interferon α kinoid induces neutralizing anti–interferon α antibodies that decrease the expression of interferon–induced and b cell activation associated transcripts: Analysis of extended follow–up data from the interferon α kinoid phase i/ii study[J]. *Rheumatology (Oxford)*, 2016, 55(10): 1901–1905.
- [17] McKellar J, Moncorg é O, Goujon C. Mx1 proteins and influenza a virus under the microscope[J]. *Virologie (Montrouge)*, 2023, 27(1): 31–34.
- [18] Tao S, Tan X, Chai W, et al. Knockdown of klf5 ameliorates renal fibrosis in mrl/lpr mice via inhibition of mx1 transcription[J]. *Immun Inflamm Dis*, 2023, 11(7): e937.
- [19] Feng X, Wu H, Grossman J M, et al. Association of increased interferon–inducible gene expression with disease activity and lupus nephritis in patients with systemic lupus eryt hematosus[J]. *Arthritis Rheum*, 2006, 54(9): 2951–2962.
- [20] Busse D C, Habgood–Coote D, Clare S, et al. Interferon–induced protein 44 and interferon–induced protein 44–like restrict replication of respiratory syncytial virus[J]. *J Virol*, 2020, 94(18): e00297.
- [21] DeDiego M L, Nogales A, Martinez–Sobrido L, et al. Interferon–induced protein 44 interacts with cellular fk506–binding protein 5, negatively regulates host antiviral responses, and supports virus replication[J]. *mBio*, 2019, 10(4): e01839.
- [22] Philips R L, Wang Y, Cheon H, et al. The jak–stat pathway at 30: Much learned, much more to do[J]. *Cell*, 2022, 185(21): 3857–3876.
- [23] Xu W, Chen Z, Shen X, et al. Reno–protective effect of realgar nanoparticles on lupus nephritis of mrl/lpr mice through stat1 [J]. *Iran J Immunol*, 2019, 16(2): 170–181.
- [24] Fu Y, Xiang Y, Wang Y, et al. The stat1/hmgb1/nf– κ b pathway in chronic inflammation and kidney injury after cisplatin exposure[J]. *Theranostics*, 2023, 13(9): 2757–2773.
- [25] Kemmner S, Bachmann Q, Steiger S, et al. Stat1 regulates macrophage number and phenotype and prevents renal fibrosis after ischemia–reperfusion injury[J]. *American Journal of Physiology–Renal Physiology*, 2019, 316(2): F277–F291.
- [26] Ding X, Ren Y, He X. Ifn–i mediates lupus nephritis from the beginning to renal fibrosis[J]. *Front Immunol*, 2021, 12(4): 676082–676099.
- [27] Arazi A, Rao D A, Berthier C C, et al. The immune cell landscape in kidneys of patients with lupus nephritis[J]. *Nature immunology*, 2019, 20(7): 902–914.