

论著·基础研究

基于GEO数据库筛选狼疮性肾炎的关键基因和信号通路

刘 音¹, 杨 涛¹, 谢裕赛², 王玉柱¹

1. 北京市海淀区医院肾内科, 北京 100080; 2. 中国医科大学基础医学院病理学教研室, 沈阳 110122

[摘要] **目的**·利用生物信息学分析方法筛选狼疮性肾炎相关差异表达基因及相关信号通路。**方法**·从GEO公共数据库中下载GSE32591数据集矩阵数据,应用R软件limma包进行标准化以及筛选差异表达基因,应用ggpubr和pheatmap包对差异基因绘制火山图及热图。应用DAVID在线数据库对差异表达基因进行GO(Gene Ontology)分析和KEGG(Kyoto Encyclopedia of Genes and Genomes)通路分析,使用R语言ggplot包绘制柱状图及气泡图。运用STRING和Cytoscape软件构建差异表达基因的蛋白质-蛋白质相互作用网络,应用MCODE及cytohubba插件筛选出参与狼疮性肾炎的最显著模块及枢纽基因。采用GSE99339数据集验证枢纽基因的差异表达。**结果**·通过limma包分析GSE32591数据集,获得了367个差异表达基因,包括253个上调基因及114个下调基因。GO分析和KEGG通路富集分析表明,差异表达基因在病毒防御反应、质膜外侧、甲型流感、结核病、EB病毒感染、补体途径等方面显著富集。应用STRING和Cytoscape构建了差异表达基因的蛋白互作网络,并鉴定出与狼疮性肾炎相关的10个枢纽基因,在验证数据集GSE99339中有显著的差异表达。**结论**·获得的367个差异表达基因和10个枢纽基因可能是狼疮性肾炎潜在的生物标志物。

[关键词] GEO数据库; 狼疮性肾炎; 系统性红斑狼疮; 生物信息学; 差异表达基因

[DOI] 10.3969/j.issn.1674-8115.2021.06.007 **[中图分类号]** R593.24; Q786 **[文献标志码]** A

Screening of key genes and pathways involved in lupus nephritis based on GEO database

LIU Yin¹, YANG Tao¹, XIE Yu-sai², WANG Yu-zhu¹

1. Department of Nephrology, Beijing Haidian Hospital, Beijing 100080, China; 2. Department of Pathology, College of Basic Medical Sciences, China Medical University, Shenyang 110112, China

[Abstract] **Objective**·To identify the differentially expressed genes and pathways involved in lupus nephritis (LN) using bioinformatics analysis. **Methods**·The matrix data of GSE32591 dataset was downloaded from the GEO database, and the limma package of R software was applied to standardize and screen the differentially expressed genes. The volcano map and heatmap of the differentially expressed genes were drawn by ggpubr and pheatmap packages. Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of differentially expressed genes were performed using DAVID online database, and bar plot and bubble chart were drawn using R software ggplot package. STRING database and Cytoscape software were used to construct protein interaction networks of differentially expressed genes, and MCODE and cytohubba plug-ins were used to screen out the most significant modules and key genes involved in lupus nephritis. GSE99339 dataset was used to verify the differential expression of hub genes. **Results**·The GSE32591 data set was analyzed through the limma package, and 367 differentially expressed genes were obtained, including 253 up-regulated genes and 114 down-regulated genes. GO analysis and KEGG pathway enrichment analysis showed that differentially expressed genes were significantly enriched in virus defense response, cytoplasmic matrix, influenza A, tuberculosis, EB virus infection, complement pathway, etc. A protein interaction network of differentially expressed genes was constructed by STRING and Cytoscape, and 10 pivotal genes related to LN were identified. The hub genes are significantly differentially expressed in the GSE99339 validation dataset. **Conclusion**·The 367 differentially expressed genes and 10 hub genes are potential biomarkers of LN.

[Key words] GEO database; lupus nephritis; systemic lupus erythematosus; bioinformatics; differentially expressed genes

系统性红斑狼疮(systemic lupus erythematosus, SLE)是一种慢性系统性自身免疫性疾病,其特征是自身抗体产生、补体激活和免疫复合物沉积。每年全球范围内SLE的发生率在0.003%~0.232%^[1]。狼疮性肾炎(lupus nephritis, LN)是SLE患者中最常见、最严重的器官表现之一,其特征通常是血尿、蛋白尿和肾小球滤过

率受损^[2]。约50%的SLE患者会发展为临床特征明显的肾脏疾病,其中11%的患者在病程5年时会发展为终末期肾脏疾病(end-stage renal disease, ESRD)^[3-4]。LN是导致SLE患者发展至ESRD和死亡的重要原因。LN的初始和后续治疗主要由免疫抑制剂和糖皮质激素组成,这意味着几乎没有有效且特异性的疗法。因此,研究LN涉及

[作者简介] 刘 音(1980—),女,主治医师,学士;电子信箱:790144859@qq.com。

[通信作者] 杨 涛,电子信箱:yangtaodoctor@163.com。

[Corresponding Author] YANG Tao, E-mail: yangtaodoctor@163.com.

[网络首发] <https://kns.cnki.net/kcms/detail/31.2045.R.20210524.0952.002.html> (2021-05-26 10:02:57)。



的分子机制对于LN的临床治疗具有重要意义。

LN的病理生理机制复杂。在LN患者的不同肾小球区室中,免疫复合物的形成、先天性免疫信号途径的激活、免疫细胞的渗透和促炎症介质可通过多种途径损害肾小球细胞^[5]。尽管许多研究已经确定了LN的某些病理机制,但其发病机制仍不清楚。

在基因组水平上追踪LN的生物学变化是一种值得关注的策略。近年来,许多学者已经进行了基因测序技术与生物信息学分析相结合的研究,以鉴定可能与预后生物标志物有关的疾病相关基因,并在未来将其开发为治疗靶标。生物信息学分析可以在极短的时间内处理大量的样品,并提供有关疾病的有价值的信息,而且与SLE密切相关的几个基因已经确定,可以驱动研究的创新^[6-7]。但是,很少有研究利用生物信息学分析来鉴定和分析LN背景下的肾脏组织。

尽管LN会影响肾脏的所有组成部分,但肾小球是最适合研究的组织,并且与疾病的发病机制和治疗密切相关。本研究选择GEO数据库中的肾小球组织数据集GSE32591研究LN。应用R语言limma包获取GSE32591数据集中的差异表达基因(differentially expressed genes, DEGs)。使用DAVID数据库来分析这些DEGs的GO(Gene Ontology)注释信息,包括生物过程、细胞组分、分子功能和KEGG(Kyoto Encyclopedia of Genes and Genomes)通路。构建了蛋白质-蛋白质相互作用(protein-protein interaction, PPI)网络,并应用Cytoscape中MCODE及Cytohubba插件分析DEGs,鉴定并验证了与LN相关的10个枢纽基因(hub genes)。最终应用GSE99339数据集验证了枢纽基因的显著差异表达,以为LN的诊断和治疗提供新的思路。

1 资料与方法

1.1 数据来源及筛选

在GEO公共数据库(<http://www.ncbi.nlm.nih.gov/geo>)^[8]中以“lupus nephritis”及“glomeruli”作为关键词检索,选择GSE32591数据集作为分析数据集,GSE99339数据集作为枢纽基因验证数据集。下载GSE32591和GSE99339数据集矩阵数据及平台注释信息。GSE32591数据集基于GPL14663平台,其中包含32例LN患者和14例正常供体对照的肾小球组织基因表达信息;GSE99339数据集基于GPL19184平台,含有30例LN患者和8例肾脏肿瘤切除对照的肾小球组织基因

表达信息。

1.2 筛选DEGs

应用R软件(4.0.2版)limma包标准化数据及筛选差异基因^[9],以 $|\log FC| > 1$ 和校正后 $P < 0.05$ 为阈值筛选差异基因。分别运用R语言ggpubr和pheatmap包对差异基因绘制火山图及热图。

1.3 DEGs的功能富集分析

使用GO分析确定枢纽基因的生物过程、细胞成分和分子功能属性^[10],通过KEGG数据库确定其通路功能^[11]。DAVID(<https://david.ncifcrf.gov/>)是进行基因富集分析的常用数据库^[12]。应用DAVID在线工具对筛选的差异表达枢纽基因进行GO和KEGG分析, $P < 0.05$ 为差异具有统计学意义,使用R语言ggplot包绘制柱状图及气泡图。

1.4 构建PPI网络及筛选枢纽基因

STRING数据库常用来构建PPI网络^[13]。将DEGs输入STRING在线工具中,筛选combined score > 0.9 的互作蛋白。将得到的PPI结果导入Cytoscape软件中,通过分子复合物检测(molecular complex detection, MCODE),以degree cutoff = 2、node score cutoff = 0.2、k-core = 2和max.depth = 100为标准,筛选出PPI中最显著模块^[14]。在最显著模块中应用Cytohubba插件,按degree算法计算出评分位于前10位的枢纽基因^[15]。

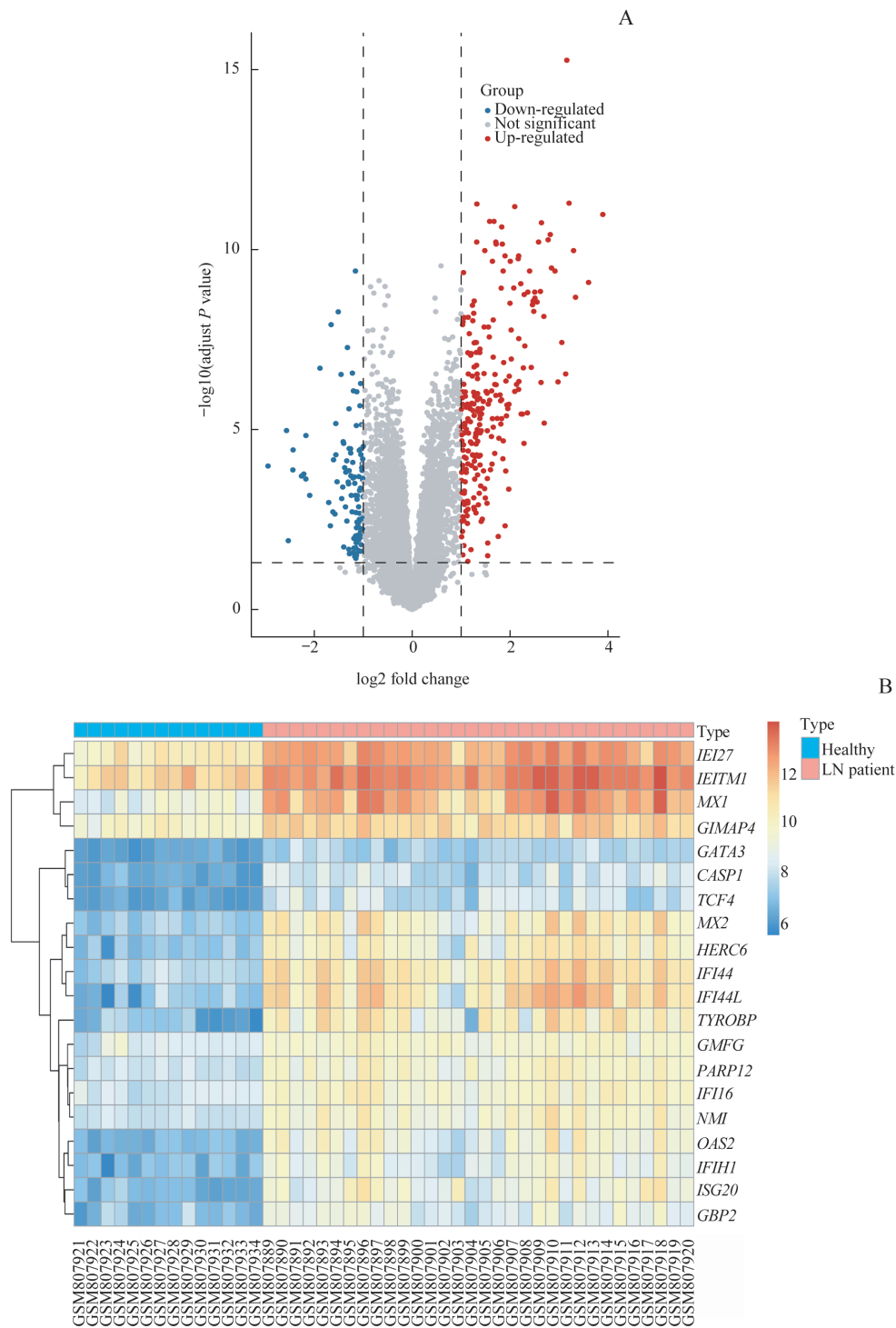
1.5 枢纽基因的验证

为了进一步验证正常组织和LN组织中上述枢纽基因的表达差异,应用GEO数据集GSE99339的数据,分析枢纽基因的表达水平。

2 结果

2.1 DEGs的筛选

应用limma包对GSE32591数据集进行数据矫正及差异分析,以 $|\log FC| > 1$ 和校正后 $P < 0.05$ 为阈值,获得了367个DEGs,包括253个上调基因及114个下调基因。使用ggpubr包绘制DEGs的火山图,按照校正后 P 值对显著DEGs进行排序;应用pheatmap包绘制出前20个DEGs的表达热图。结果显示,IFI27、MX1、ISG20等基因在LN中显著高表达(图1)。



Note: A. Volcano plot. B. Heatmap of the top 20 express genes.

图1 DEGs的鉴定

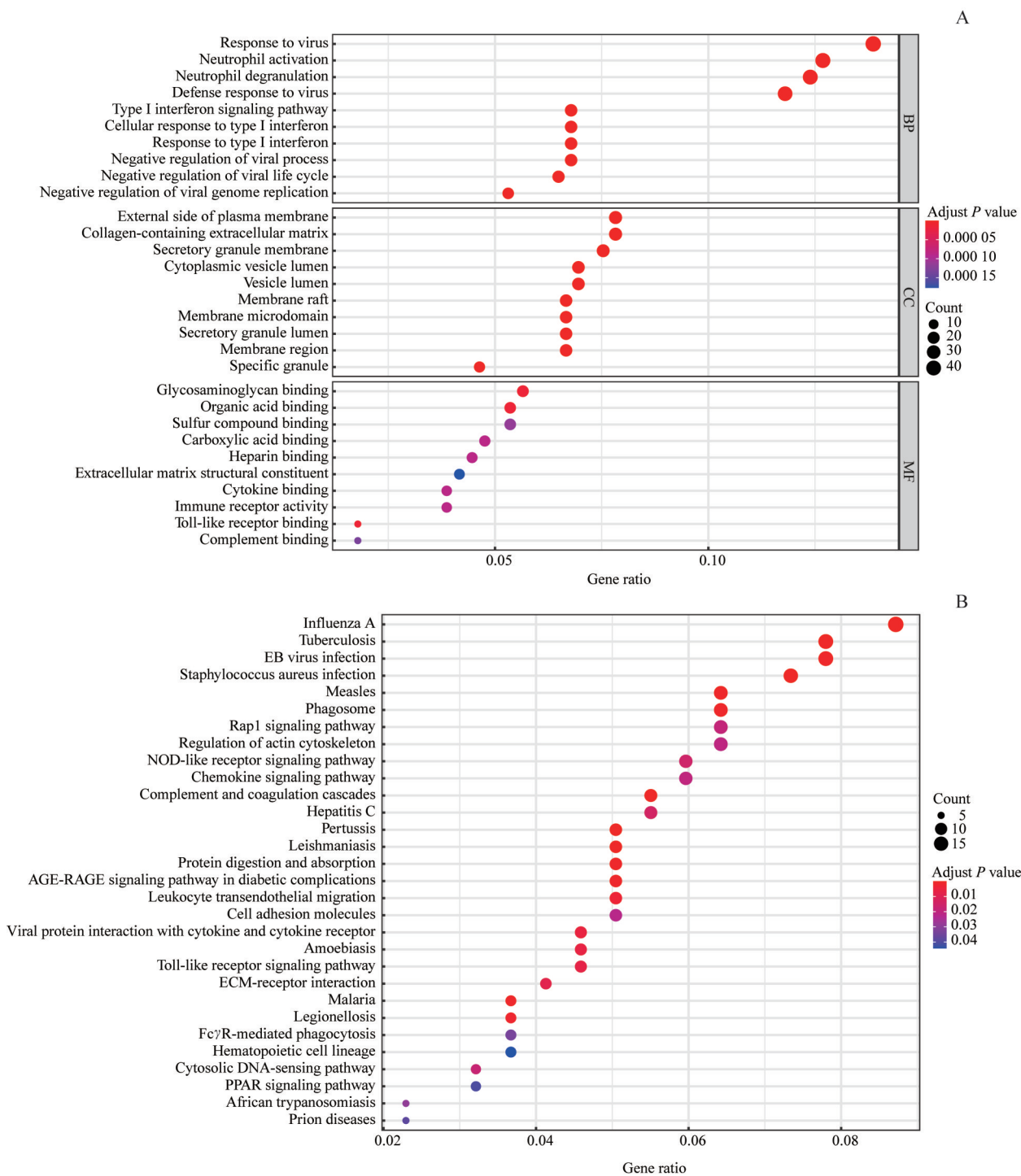
Fig 1 Identification of DEGs

2.2 DEGs的GO分析和KEGG信号通路分析

应用DAVID软件对DEGs进行GO富集分析和KEGG信号通路富集分析。GO富集分析显示这些DEGs主要参与病毒反应应答、对病毒的防御反应，并且参与I型干扰素信号通路、活化中性粒细胞以及负调控病毒生命周期等生物过程；主要细胞成分位于质膜外侧、囊腔及胶

原纤维（含细胞外基质）；主要分子功能与Toll样受体结合、有机酸结合、补体结合及免疫受体活性等相关（图2A）。KEGG通路分析显示，这些差异基因主要参与甲型流感、结核病、EB病毒感染、金黄色葡萄球菌感染和补体途径等信号通路（图2B）。





Note: A. The bar plot of GO analysis. B. The bubble chart of KEGG pathway analysis.

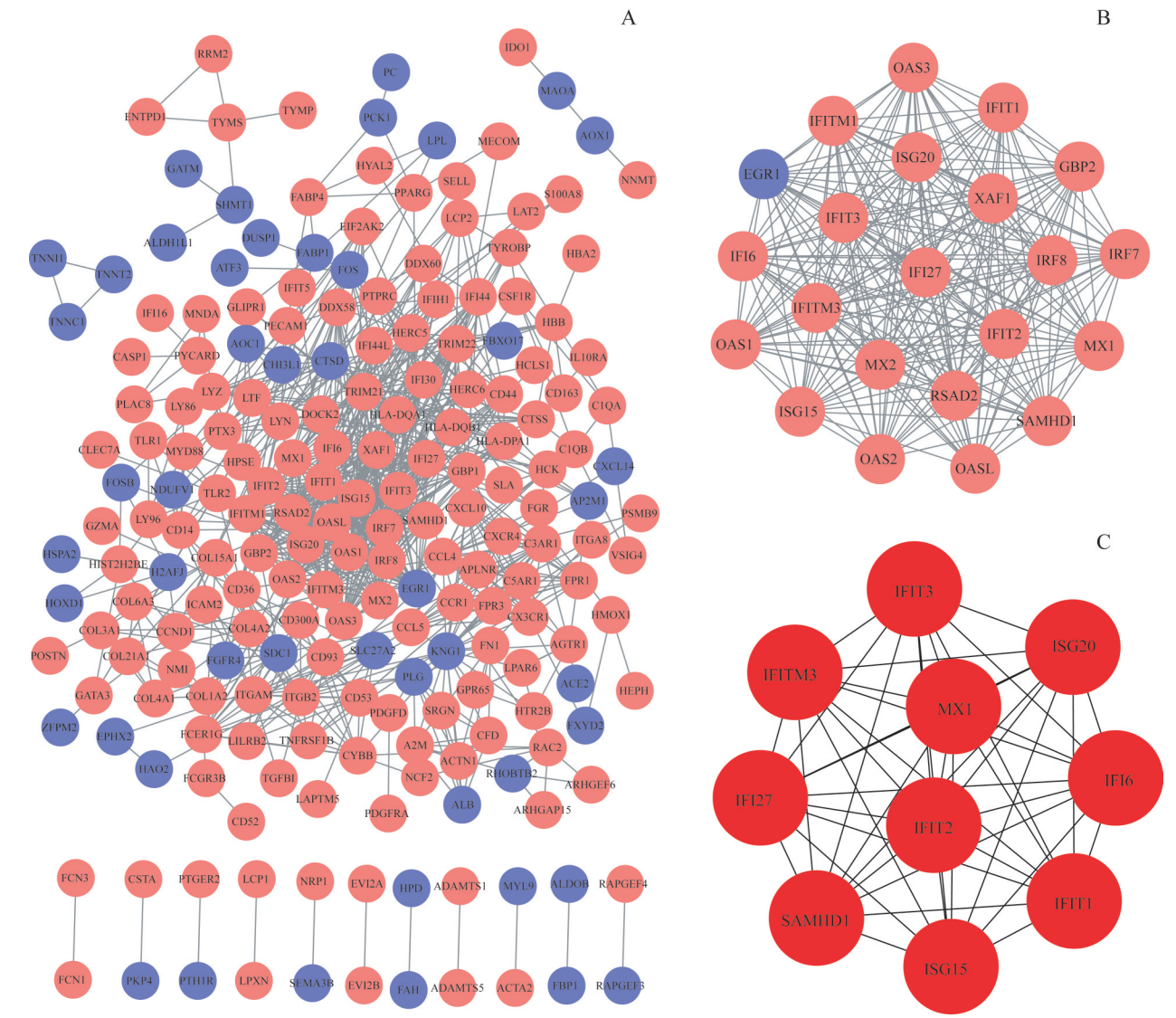
图2 DEGs的GO分析和KEGG通路分析

Fig 2 GO analysis and KEGG pathway analysis of DEGs

2.3 DEGs编码蛋白的PPI网络及枢纽基因的筛选

通过 STRING 在线工具和 Cytoscape 软件对 367 个 DEGs 进行 PPI 网络分析, 筛选 combined score > 0.9 的相互作用蛋白对。在 Cytoscape 中得到 206 个节点, 831 条相互作用的网络图 (图 3A)。在 PPI 网络中, 利用 Cytoscape 内的 MCODE 插件, 以 degree cutoff = 2、node score

cutoff = 0.2、k-core = 2 和 max.depth = 100 为标准, 筛选出最显著模块。最显著模块包括 22 个重要节点及 227 条相互作用 (图 3B)。对最显著模块运用 cytohubba 插件, 以 degree 算法计算出总分排名前 10 位的枢纽基因, 分别是 *IFI6*、*IFI27*、*IFIT1*、*IFIT2*、*IFIT3*、*IFITM3*、*ISG15*、*ISG20*、*MX1*、*SAMHD1* (图 3C, 表 1)。



Note: A. A PPI network was constructed; red nodes represent up-regulated genes, and blue nodes represent down-regulated genes. B. The most significant module was selected from the PPI network. C. The network of hub genes with the top 10 was screened with cytohubba.

图3通过STRING和MCODE方法筛选枢纽基因

Fig 3 Identification of hub genes from the DEGs by STRING and MCODE

表 1 10个枢纽基因的功能

Tab 1 Functions of 10 hub genes

No.	Gene symbol	Full name	Function
1	<i>IFI6</i>	Interferon α inducible protein 6	May play a critical role in cell apoptosis, negatively regulate intracellular apoptosis signaling pathways and TNFSF10-induced cell apoptosis
2	<i>IFI27</i>	Interferon α inducible protein 27	May play a role in type I interferon-induced apoptosis and TNFSF10-induced apoptosis, and has antiviral activity against hepatitis C virus (HCV) in the innate immune response
3	<i>IFIT1</i>	Interferon induced protein with tetratricopeptide repeats 1	Inhibits expression of viral messenger RNAs
4	<i>IFIT2</i>	Interferon induced protein with tetratricopeptide repeats 2	Inhibits expression of viral messenger RNAs and promotes apoptosis
5	<i>IFIT3</i>	Interferon induced protein with tetratricopeptide repeats 3	Inhibits cellular as well as viral processes, cell migration, proliferation, signaling, and viral. Negatively regulates the apoptotic effects of IFIT2
6	<i>IFITM3</i>	Interferon induced transmembrane protein 3	Interferon-induced antiviral protein inhibits virus from entering the cytoplasm of host cells
7	<i>ISG15</i>	ISG15 ubiquitin-like modifier	Mediates innate immune response to viral and mycobacterial infections



Continued Tab

No.	Gene symbol	Full name	Function
8	<i>ISG20</i>	Interferon stimulated exonuclease gene 20	Interferon-induced antiviral exo-ribonuclease acts on single-stranded RNA and has antiviral activity against RNA viruses such as HCV
9	<i>MX1</i>	MX dynamin like GTPase 1	Interferon-induced dynamin-like GTPase with antiviral activity against a wide range of RNA viruses and some DNA viruses
10	<i>SAMHD1</i>	SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1	May play a role in the regulation of innate immune response

2.4 枢纽基因的验证

在 GEO 数据库检索其他 LN 数据集。最终选择 GSE99339 数据集，分析 LN 与正常对照之间这些枢纽基

因的差异表达水平（图4）。在 GSE99339 数据集中，除了缺失 *IFITM3* 的探针信息，上述枢纽基因均明显上调。说明了枢纽基因筛选结果的可靠性。

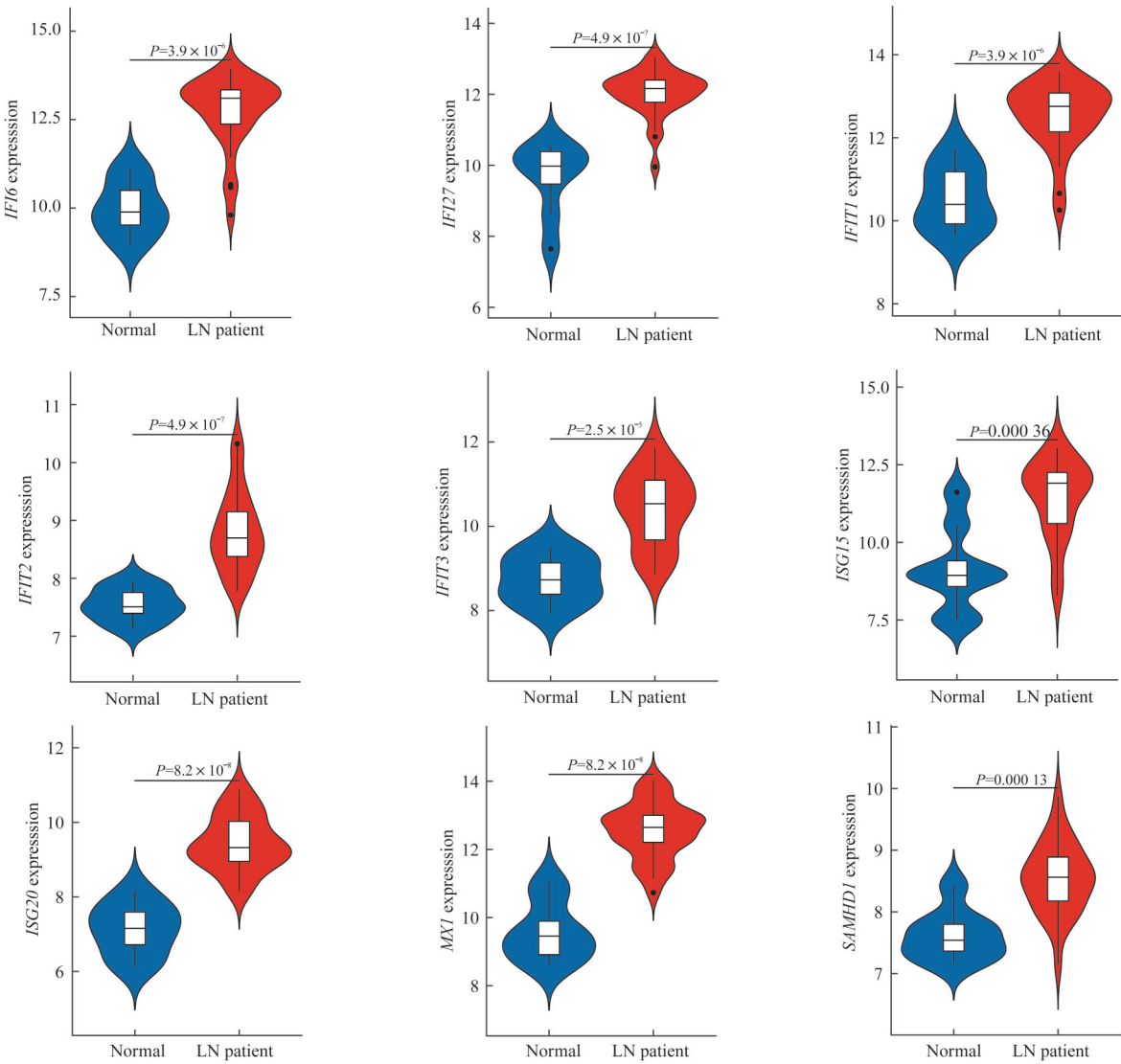


图4 在 GSE99339 数据集中验证枢纽基因在 LN 与正常组织之间的表达差异
Fig 4 Differentially expressed levels of the hub genes between LN and normal in GSE99339 datasets

3 讨论

随着生物信息学的发展，人们对寻找各种疾病枢纽

基因的关注日益增加，收集到的疾病枢纽基因信息可为治疗疾病提供新的手段。我们利用基于 GPL14663 平台的数据集 GSE32591，对 32 例 LN 患者的肾小球组织基因表达数据进行分析，筛选 DEGs。最终筛选出 367 个 DEGs，

其中包括 253 个上调基因和 114 个下调基因;再运用 ggpubr 和 pheatmap 包绘制火山图及差异表达显著的前 20 个 DEGs 的表达热图。热图结果显示, *IFI27*、*MX1*、*ISG20* 等基因在 LN 中显著高表达,提示这些基因可能参与 LN 的疾病进展。

对 DEGs 的功能富集分析显示,免疫应答、感染和干扰素相关基因参与了 LN 的发病过程。GO 分析和 KEGG 途径富集分析表明,DEGs 在病毒防御反应、质膜外侧、甲型流感、结核病、EB 病毒感染和补体途径等方面显著富集,提示了 LN 疾病进展中免疫活动的增强。本研究应用 Cytoscape 中 MCODE 及 Cytohubba 插件分析这些 DEGs,鉴定了与 LN 相关的最显著模块及 10 个枢纽基因(*IFI16*、*IFI27*、*IFIT1*、*IFIT2*、*IFIT3*、*IFITM3*、*ISG15*、*ISG20*、*MX1* 和 *SAMHD1*)。最终,应用其他 GEO 数据集(GSE99339)的数据信息,验证了上述枢纽基因显著表达的可靠性。

在枢纽基因中,干扰素家族蛋白的表达显著上调,这也验证了之前的研究结果^[16-17]。干扰素是一类信号蛋白,分为 I 型、II 型和 III 型;其中 I 型干扰素是其最大

的家族^[18]。GO 分析结果也表明,I 型干扰素信号通路可能与 SLE 的发病机制密切相关。SLE 的大多数枢纽基因都是 I 型干扰素诱导型基因。SLE 患者显示出较高水平的 I 型干扰素及 I 型干扰素诱导型基因,这些基因在 SLE 发病机制中起主要作用^[19-20]。干扰素诱导的四肽重复序列 1 (IFIT1) 蛋白质也属于 I 型干扰素家族,在大多数 SLE 人群中 *IFIT1* mRNA 的表达明显上调^[21]。*IFIT1* 是首个鉴定为主要由干扰素- α/β 诱导的细胞质和线粒体干扰素刺激基因^[22]。*IFIT1* 可调节病毒复制以及翻译、增殖、凋亡和信号转导等细胞过程^[23],可能与 Rho/Rac 鸟嘌呤核苷酸交换因子发生相互作用,从而参与 SLE 免疫反应。

在本研究中,我们使用系统的生物信息学方法来分析 LN 中肾小球组织表达信息。研究存在一些局限性,如 GEO 数据库仅提供了少量 LN 数据集。在后续研究中,我们将进一步收集临床数据,以作更深入的探索。

总之,本研究使用 GEO 数据库筛选 LN 肾小球组织的 367 个 DEGs 及 10 个枢纽基因,其可能是 LN 潜在的生物标志物。生物标志物的揭示,有助于提高临床诊断水平和开发个性化治疗方法。

参 · 考 · 文 · 献

- [1] Rees F, Doherty M, Grainge MJ, et al. The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies[J]. Rheumatology (Oxford), 2017, 56(11): 1945-1961.
- [2] Yu F, Haas M, Glasscock R, et al. Redefining lupus nephritis: clinical implications of pathophysiologic subtypes[J]. Nat Rev Nephrol, 2017, 13(8): 483-495.
- [3] Almaani S, Meara A, Rovin BH. Update on lupus nephritis[J]. Clin J Am Soc Nephrol, 2017, 12(5): 825-835.
- [4] Tektonidou MG, Dasgupta A, Ward MM. Risk of end-stage renal disease in patients with lupus nephritis, 1971-2015: a systematic review and Bayesian meta-analysis[J]. Arthritis Rheumatol, 2016, 68(6): 1432-1441.
- [5] Devarapu SK, Lorenz G, Kulkarni OP, et al. Cellular and molecular mechanisms of autoimmunity and lupus nephritis[J]. Int Rev Cell Mol Biol, 2017, 332: 43-154.
- [6] Bing PF, Xia W, Wang L, et al. Common marker genes identified from various sample types for systemic lupus erythematosus[J]. PLoS One, 2016, 11(6): e0156234.
- [7] Coit P, Jeffries M, Altork N, et al. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naïve CD4⁺T cells from lupus patients[J]. J Autoimmun, 2013, 43: 78-84.
- [8] Clough E, Barrett T. The gene expression omnibus database[J]. Methods Mol Biol Clifton N J, 2016, 1418: 93-110.
- [9] Piao J, Sun J, Yang Y, et al. Target gene screening and evaluation of prognostic values in non-small cell lung cancers by bioinformatics analysis[J]. Gene, 2018, 647: 306-311.
- [10] Gene Ontology Consortium. Gene Ontology Consortium: going forward[J]. Nucleic Acids Res, 2015, 43(database issue): D1049-D1056.
- [11] Du JL, Yuan ZF, Ma ZW, et al. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model [J]. Mol Biosyst, 2014, 10(9): 2441-2447.
- [12] Dennis G, JR., Sherman BT, Hosack DA, et al. DAVID: Database for annotation, visualization, and integrated discovery[J]. Genome Biol, 2003, 4(5): P3.
- [13] Kumar MS, Adki KM. Marine natural products for multi-targeted cancer treatment: a future insight[J]. Biomed Pharmacother, 2018, 105: 233-245.
- [14] Wang J, Zhong J, Chen G, et al. ClusterViz: a cytoscape APP for cluster analysis of biological network[J]. IEEE/ACM Trans Comput Biol Bioinform, 2015, 12(4): 815-822.
- [15] Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. BMC Syst Biol, 2014, 8(Suppl 4): S11.
- [16] Yang H, Li H. *CD36* identified by weighted gene co-expression network analysis as a hub candidate gene in lupus nephritis[J]. PeerJ, 2019, 7: e7722.
- [17] Shu B, Fang Y, He W, et al. Identification of macrophage-related candidate genes in lupus nephritis using bioinformatics analysis[J]. Cell Signal, 2018, 46: 43-51.
- [18] Reder AT, Feng X. Aberrant type I interferon regulation in autoimmunity: opposite directions in MS and SLE, shaped by evolution and body ecology[J]. Front Immunol, 2013, 4: 281.
- [19] Bezalel S, Guri KM, Elbirt D, et al. Type I interferon signature in systemic lupus erythematosus[J]. Isr Med Assoc J, 2014, 16(4): 246-249.
- [20] Chen JY, Wang CM, Chen TD, et al. Interferon- λ 3/4 genetic variants and interferon- λ 3 serum levels are biomarkers of lupus nephritis and disease activity in Taiwanese [J]. Arthritis Res Ther, 2018, 20(1): 193.
- [21] Becker AM, Dao KH, Han BK, et al. SLE peripheral blood B cell, T cell and myeloid cell transcriptomes display unique profiles and each subset contributes to the interferon signature[J]. PLoS One, 2013, 8(6): e67003.
- [22] Reynaud JM, Kim DY, Atasheva S, et al. IFIT1 differentially interferes with translation and replication of alphavirus genomes and promotes induction of type I interferon[J]. PLoS Pathog, 2015, 11(4): e1004863.
- [23] Zhang L, Wang B, Li L, et al. Antiviral effects of IFIT1 in human cytomegalovirus-infected fetal astrocytes[J]. J Med Virol, 2017, 89(4): 672-684.

[收稿日期] 2020-09-08

[本文编辑] 吴 洋

