

网络出版时间:2024-04-12 14:27:42 网络出版地址:https://link.cnki.net/urlid/34.1065.R.20240410.1009.009

糖尿病肾病核心基因的筛选与鉴定

吴晓俊^{1,2*},倪飞雪^{1,2*},徐玉雪^{1,2,3}

摘要 目的 基于生物信息学筛选糖尿病肾病(DKD)核心基因,探究 DKD 的治疗靶点,讨论其可能的调控机制。方法 提取 GEO 数据库(GSE30528,GSE47183)DKD 患者肾小球转录组表达数据矩阵,采用生物信息学方法对差异表达基因(DEGs)进行筛选,鉴定出核心差异基因,对核心差异基因进

行基因表达与富集分析(GSEA),预测有效靶点。结果 通过对 DKD mRNA 表达矩阵的筛选和鉴定,共筛选出 5 个核心基因,其中 Clorf21、NPHS1 表达显著下调,CD48、COL1A2、TGFB1 表达上调,研究得知,NPHS1、CD48 与免疫差异、细胞间交流及细胞表面相互作用等有显著相关性。通过受试者工作特征曲线(ROC)分析、GSEA 分析及药物靶点与 miRNA 预测显示,这些差异基因可能对 DKD 的治疗具有重要意义。结论 该研究筛选的核心基因与 DKD 具有显著相关性,有可能作为糖尿病治疗有效标志物,为 DKD 的治疗与鉴定提供理论依据。

关键词 糖尿病肾病;GEO 数据库;转录组;生物信息学;基因筛选;药物靶点

中图分类号 R 341.2

文献标志码 A 文章编号 1000-1492(2024)04-0610-09
doi:10.19405/j.cnki.issn1000-1492.2024.04.009

2024-02-29 接收

基金项目:国家自然科学基金(编号:31800985);山东省医药卫生科技发展计划项目(编号:2017WS694);烟台市科技创新发展计划基础研究类项目(编号:2023JCYJ072);滨州医学院科研启动基金(编号:BY2017KYQD09)

作者单位:¹滨州医学院药学院,烟台 264003

²滨州医学院山东省分子靶向智能诊疗技术创新中心,烟台 264003

³滨州医学院烟台市干细胞与再生医学重点实验室,烟台 264003

作者简介:吴晓俊,男,硕士研究生;

徐玉雪,女,副教授,硕士生导师,责任作者,E-mail:xuyuxue@bzmc.edu.cn

* 对本文具有同等贡献

糖尿病肾病(diabetic kidney disease, DKD)是糖尿病发展的常见并发症,主要特征是尿蛋白、肾小球过滤功能障碍,是终末期肾病的主要原因,严重

Rapamycin upregulates autophagy inhibits cell proliferation in human umbilical vein endothelial cells

Wang Yawen, Cheng Yanan, Yang Bin, Su Bihao, Xu Pu

(Dept of Stomatology, Haikou Hospital, Hainan Stomatological Center, Central South University Xiangya School of Medicine Affiliated Haikou Hospital, Haikou 570208)

Abstract Objective To investigate the effect of autophagy activation on cell proliferation in human umbilical vein endothelial cells (HUVECs). **Methods** HUVECs were treated with rapamycin (Rapa). Western blot assay was performed to examine the expression of protein of microtubule associated protein 1 light chain 3 (LC3), Beclin 1 and unc-51-like kinase 1 (ULK1). Autophagosomes were detected by transmission electron microscopy (TEM), and autophagy fluorescence was detected by monodansylcadaverine staining (MDC) assay. The effect of autophagy activation on cell proliferation was assessed by CCK-8 assay and EdU assay. Vascular formation experiments were used to detect vasculogenic ability. **Results** After Rapa treatment, LC3, Beclin1 and ULK1 expressions were enhanced, while the green autophagy fluorescence expression in the experimental group was stronger than that in the control group, and autophagosomes were visible by TEM; CCK-8 and EdU results showed that compared with the control group, the cell proliferation ability was weakened and tubes formation ability was reduced after the activation of autophagy in experimental cells. **Conclusion** Rapa upregulates autophagy activity in HUVECs to inhibit cell proliferation under certain time.

Key words rapamycin; HUVECs; autophagy; autophagic activity; autophagic protein; cell proliferation

增加了糖尿病患者的死亡率^[1]。肾小球的功能障碍在 DKD 的病程发展中起着重要的作用,高浓度的血糖可诱导氧化应激、细胞凋亡^[2]、免疫反应^[3]等。近年来,内皮素拮抗剂、肾素-血管紧张素-醛固酮系统阻滞剂、血管紧张素转换酶抑制剂等的研究进展,为 DKD 的治疗提供了方向^[4]。目前仍缺乏对 DKD 病理机制和药物靶点的探究,且临床治疗结果不理想^[5]。然而,生物信息学的发展为疾病研究提供了新的视角^[6]。基于对 DKD 肾小球基因表达矩阵进行的加权共表达网络分析(weighted gene co-expression network analysis, WGCNA)、最小绝对收缩和选择算法(least absolute shrinkage and selection operator, LASSO)、特征选择之支持向量机递归特征消除(support vector machine-recursive feature elimination, SVM-RFE)等分析方法,有助于对 DKD 的机制进行深入研究,筛选核心生物学标志物与治疗靶点,寻找与 DKD 相关的潜在治疗基因。

1 材料与方法

1.1 数据资料来源 使用基因表达数据库(GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), 搜索关键词“Diabetic kidney disease”和“Homo sapiens”, 选择“Expression profiling by array”以获得 DKD 的 mRNA 微阵列基因表达矩阵。筛选得到了 GSE47183 和 GSE30528 及对应数据平台文件(包括 30 个正常对照样本, 23 个 DKD 肾小球样本), 运用 R(4.2.1) 将基因表达数据整合并进行生物信息学分析。GEO 属于公共数据库, 数据库中涉及的患者已获得伦理批准, 因此本研究不涉及伦理问题及其他利益冲突。

1.2 生物信息学数据分析 使用 R 中的 limma 包对数据归一化处理, 通过 WGCNA 分别对 GSE47183 和 GSE30528 数据聚类分析, 筛选与疾病具有显著相关性的核心基因集, 并通过韦恩图(Venn diagram, Venn)对基因集进行重叠, 对共有的显著核心基因使用 LASSO 和 SVM-RFE 算法进一步分析, 筛选目标核心基因。同时, 为了得到与目标核心基因相关的药物靶点, 运用基因药物分析网络分析(<https://dgidb.genome.wustl.edu/>) 预测基因药物作用靶点, 并且使用基因集合富集分析(gene set enrichment analysis, GSEA) 筛选目标核心基因可能参与调控的信号通路。

1.3 RNA 提取及实时荧光定量 PCR 使用南京诺唯赞 RNA 抽提试剂盒(L/N 7E632K2) 提取高脂饮食(high-fat diet, HFD) 喂养糖尿病小鼠与同龄健

康 C57 小鼠肾脏组织总 mRNA, 使用 Takara 试剂盒将 RNA 逆转录为模板 cDNA, 检测样本量为 1 μ g, 依照 SYBR Ex Tap II 试剂说明书进行扩增, β -actin 为内参, 验证核心基因足细胞标记蛋白(NPHS1)、I 型胶原蛋白 α 2 链(collagen type I alpha 2 chain, COL1A2)、转化生长因子 β 诱导蛋白(transforming growth factor beta induced, TGFBI)、分化集群 48(cluster of differentiation 48, CD48)、1 号染色体开放阅读框 21(chromosome 1 open reading frame 21, C1orf21) 在肾脏中的基因表达水平。

1.4 统计学处理 差异表达水平使用 R(4.2.1) 及 GraphPad Prime 8.0 统计分析, 使用 *t* 检验, 以标准误差(SEM) 表示平均值的误差, $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 差异基因 mRNA 的筛选和分析 提取 GEO 数据库 GSE30528 和 GSE47183 中 DKD 肾小球的 mRNA 数据, 作火山图分析。结果表明, DKD 数据 GSE30528 共有 317 个具有显著性差异的基因($P_{adj} < 0.05$, $|\log_2 FC| > 1$), 235 个基因表达下调(绿色), 82 个基因表达上调(红色)(图 1A)。GSE47183 数据共有 209 个具有显著表达的差异基因($P_{adj} < 0.05$, $|\log_2 FC| > 1$), 100 个基因表达下调(绿色), 109 个基因表达上调(红色)(图 1B)。使用 Venn 图重叠差异基因, 62 个差异基因在两个数据库均表达(图 1C)。对此 62 个差异基因进行 KEGG 分析, 结果显示这些差异基因与细胞外基质(extracellular matrix, ECM) 受体相互作用, PI3K-Akt 信号通路具有显著相关性。GO 结果表明, 差异基因主要与细胞因子的正调节具有相关性, 主要在细胞外间质结构成分中起作用(图 2B-2D)。

2.2 WGCNA 筛选疾病核心基因 将 DKD 表达矩阵 GSE30528 和 GSE47183 分别用 R(limma 包) 对数据归一化, 皮尔逊相关系数对样品聚类, 进行 WGCNA, 筛选与疾病关系密切相关的核心基因。软阈值分别为 16 和 12 时, 满足 $R^2 = 0.9$ 的无标度的网络拓扑关系(图 3A、3B)。将数据集的邻接矩阵转换为 TOM 矩阵, 对样本聚类, 构建 WGCNA 网络(图 3C), 筛选出不同颜色的模块基因集。值得关注的模块分别是 GSE30528 的 ME turquoise($r = 0.93$, $P = 5e-10$) 和 GSE47183 的 ME blue($r = 0.72$, $P = 5e-6$), 其与疾病状态最具有关联性(图 3D)。同时, 证实了两个模块(turquoise 模块, blue 模块) 中基因

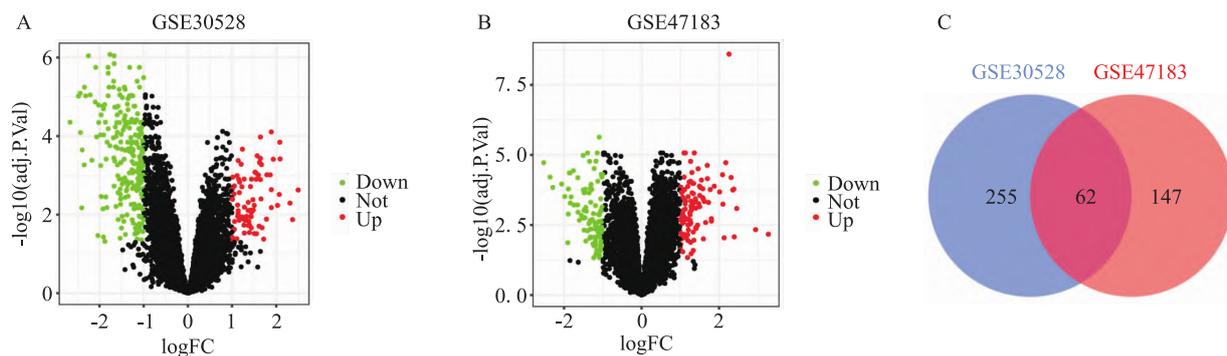


图1 差异基因的提取

A:GSE30528 差异基因火山图;B:GSE47183 差异基因火山图;C:Veen 图

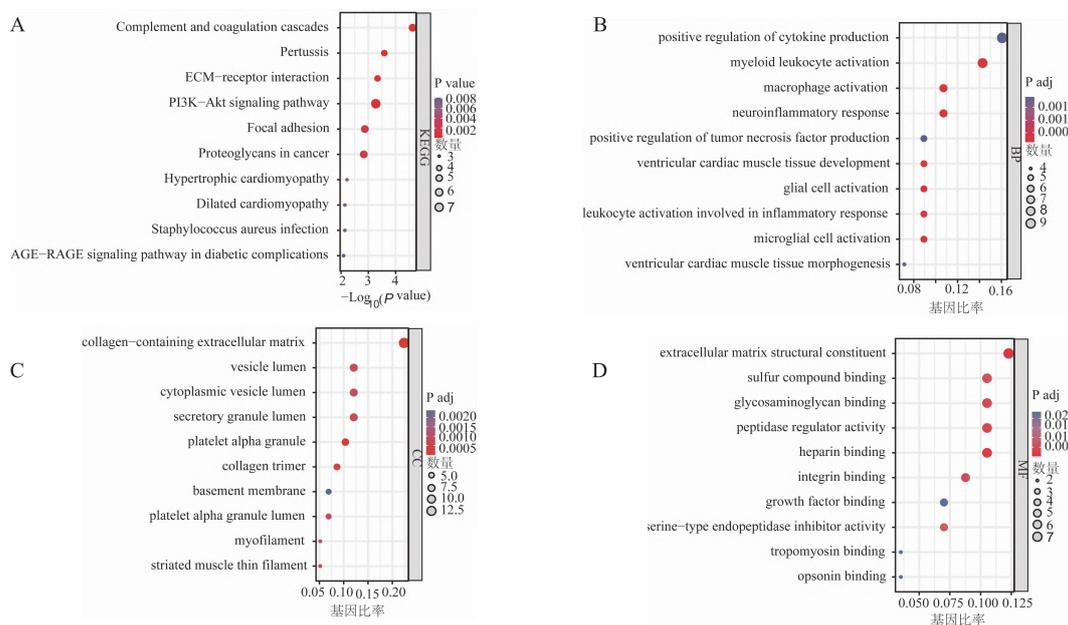


图2 差异基因 KEGG 和 GO 分析

A:KEGG 分析;B:生物进程 (BP); C:细胞成分 (CC);D:分子功能 (MF)

集的相关性和显著性, turquoise 模块: $Cor = 0.92, P = 3.8e-172$ (GSE30528), blue 模块: $Cor = 0.63, P = 2.7e-21$ (GES47183), 见图 3E、3F。分别将 GSE30528、GSE47183 的显著性差异基因与 WGCNA 筛选的核心基因使用 Veen 图重叠筛选, 富集出 37 个核心基因(图 4A), 对两组数据进行合并, 热图展现每个基因的表达水平, 共有 10 个基因表达下调(蓝色), 27 个基因表达上调(红色)(图 4B)。

2.3 机器学习算法筛选特征基因 对 37 个核心基因使用 LASSO 模型算法, 绘制 LASSO 回归图形和交叉验证图形, 筛选出 6 个特征基因(图 5A、5B)。同时, 使用 SVM-RFE 算法对 37 个核心基因进行过滤, 筛选特征基因的最优组合, 10 个特征基因被显著富集(图 5C、5D)。Veen 图对两种算法的核心基因取交集, 共同筛选出 5 个目标核心基因, 这 5 个重

叠交叉的核心基因将是进一步研究的重点(图 5E)。

2.4 目标核心基因的表达水平及相关性 将 GEO 数据库中的 GSE30528 和 GSE47183 数据合并, 共有 30 个正常肾小球样本, 23 个 DKD 肾小球样本, 分析 5 个核心基因在数据中的表达情况。结果表明, CD48、COL1A2、TGFB1 的表达水平显著升高, NPSH1、C1orf21 的表达水平显著降低 ($P. adj < 0.05, |log_2FC| > 1$)(图 6A-6E)。皮尔森相关性分析确定 5 个基因间的相关性(图 6F)。基于差异基因集构建的逻辑回归模型, 绘制的受试者工作特征曲线(receiver operating characteristic curve, ROC)显示, 5 个核心基因的 ROC 曲线下面积(area under the receiver operating characteristic curve, AUC)均大于 0.8(图 7A-7E), 进一步表明这些核心基因可能与 DKD 疾病具有密切关系。

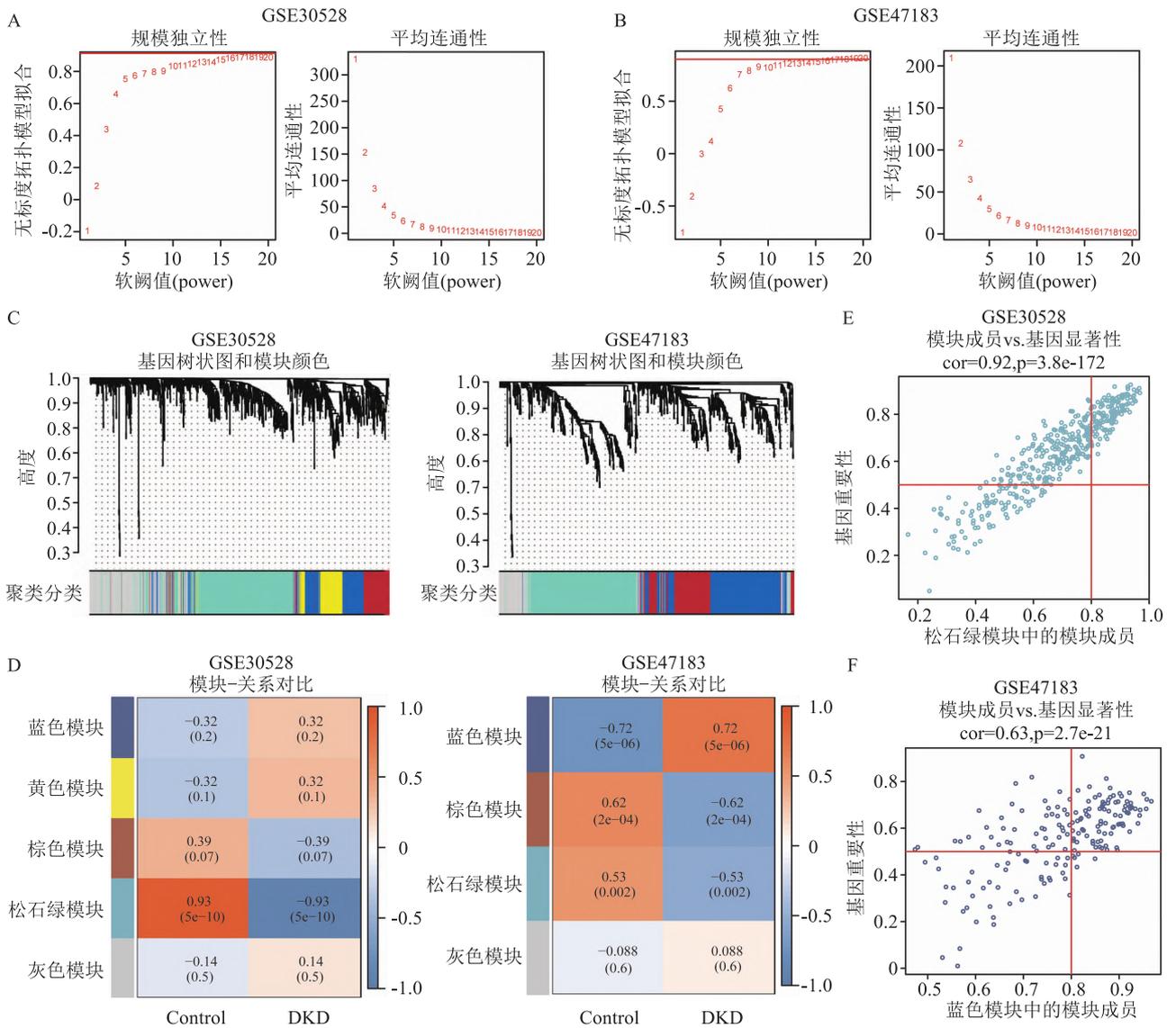


图3 WGCNA 筛选疾病核心基因

A、B: 样品软阈值的拓扑关系及连通关系; C: 基因聚类树状图; D: 不同颜色代表的基因模块; E、F: turquoise 模块和 blue 模块的基因显著性

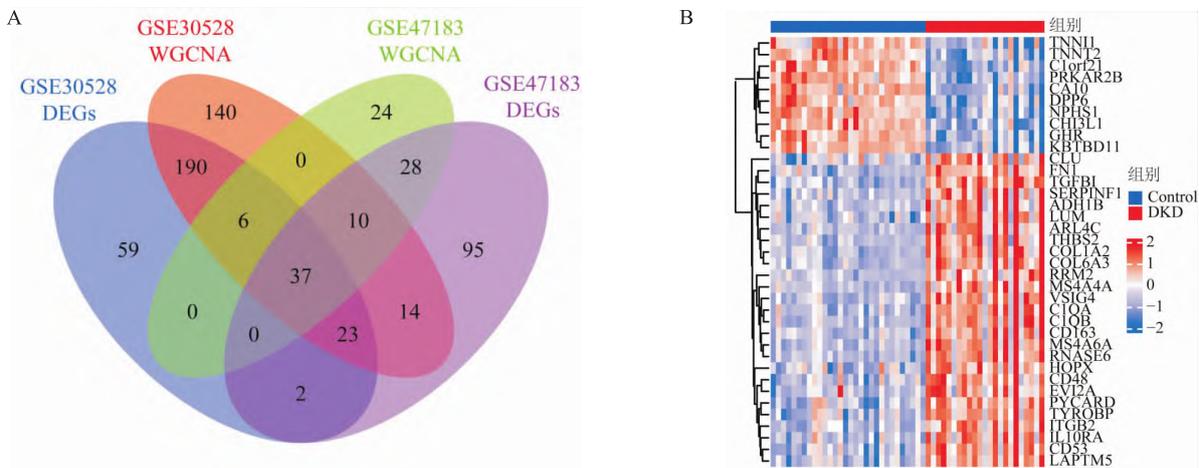


图4 核心基因的筛选和展示

A: Venn 图筛选 GSE30528 和 GSE47183 共有的基因; B: 37 个重叠基因热图

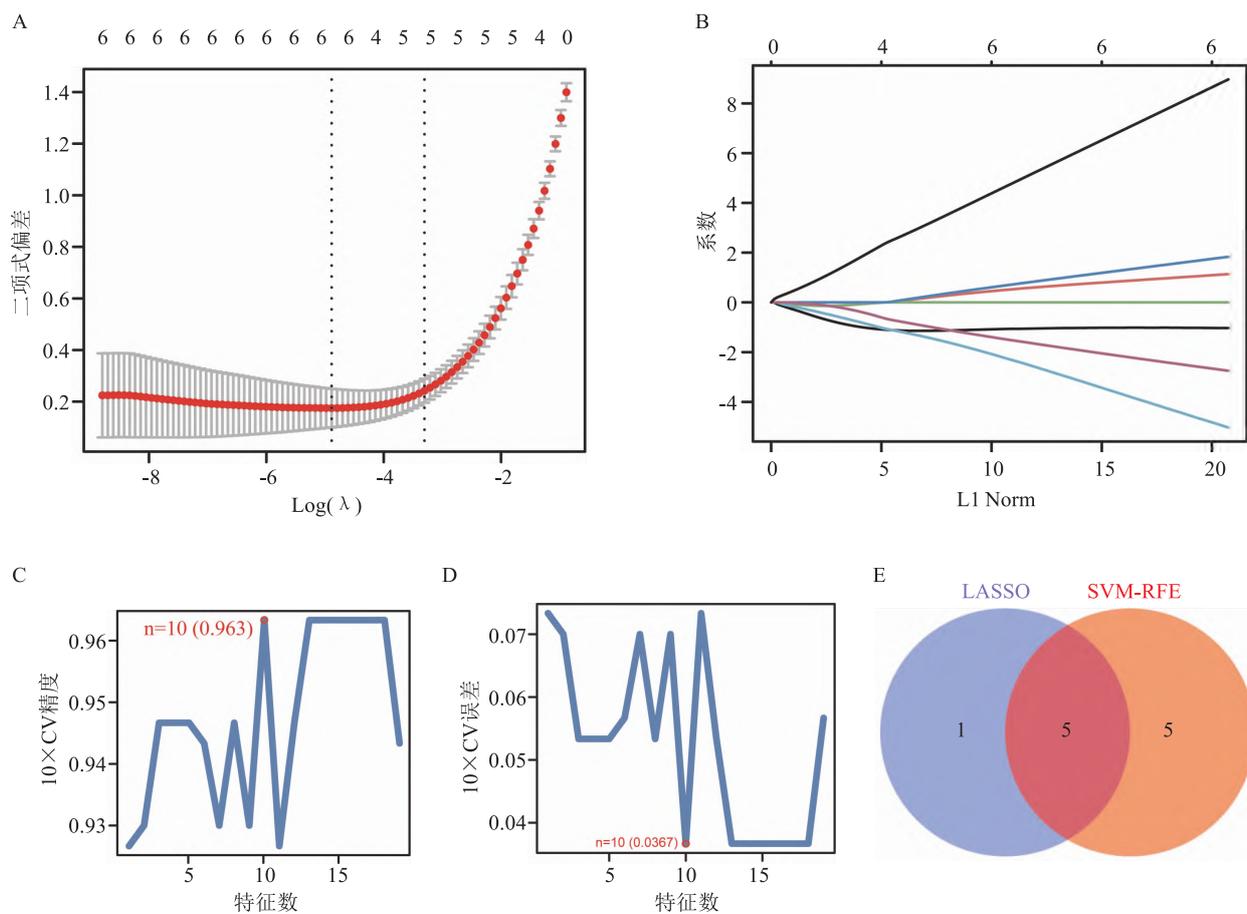


图5 目标核心基因的筛选

A, B: LASSO 算法筛选核心基因; C, D: SVM-REF 算法筛选核心基因; E: Veen 图

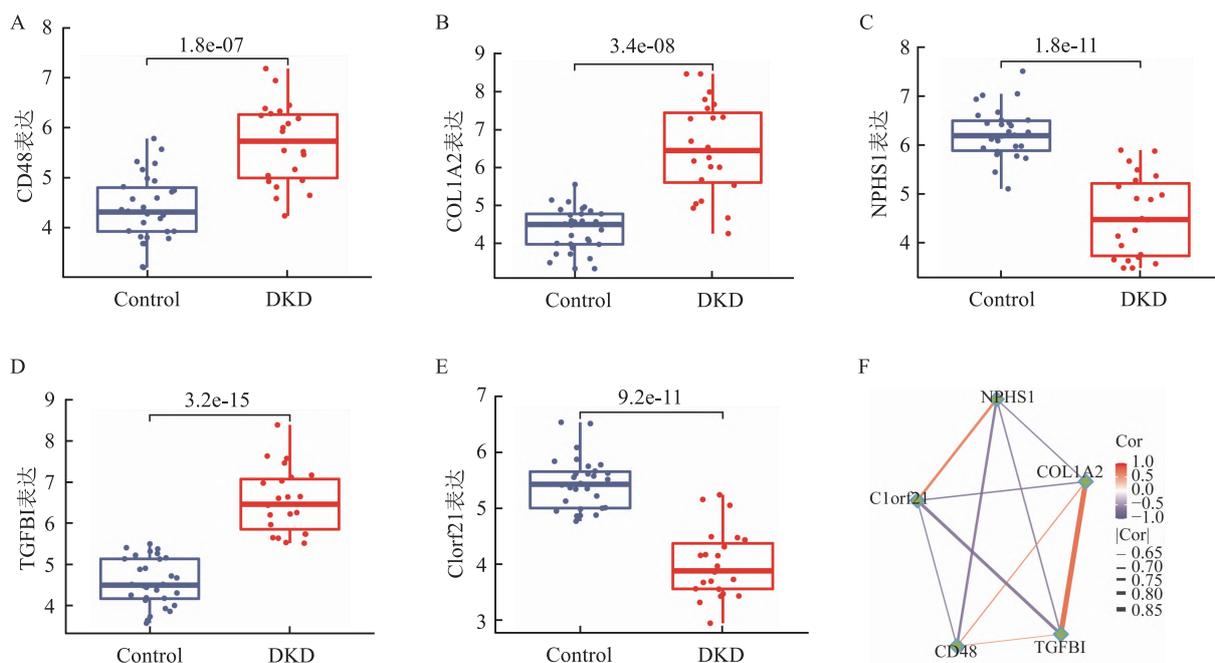


图6 目标核心基因的表达水平及皮尔森相关性分析

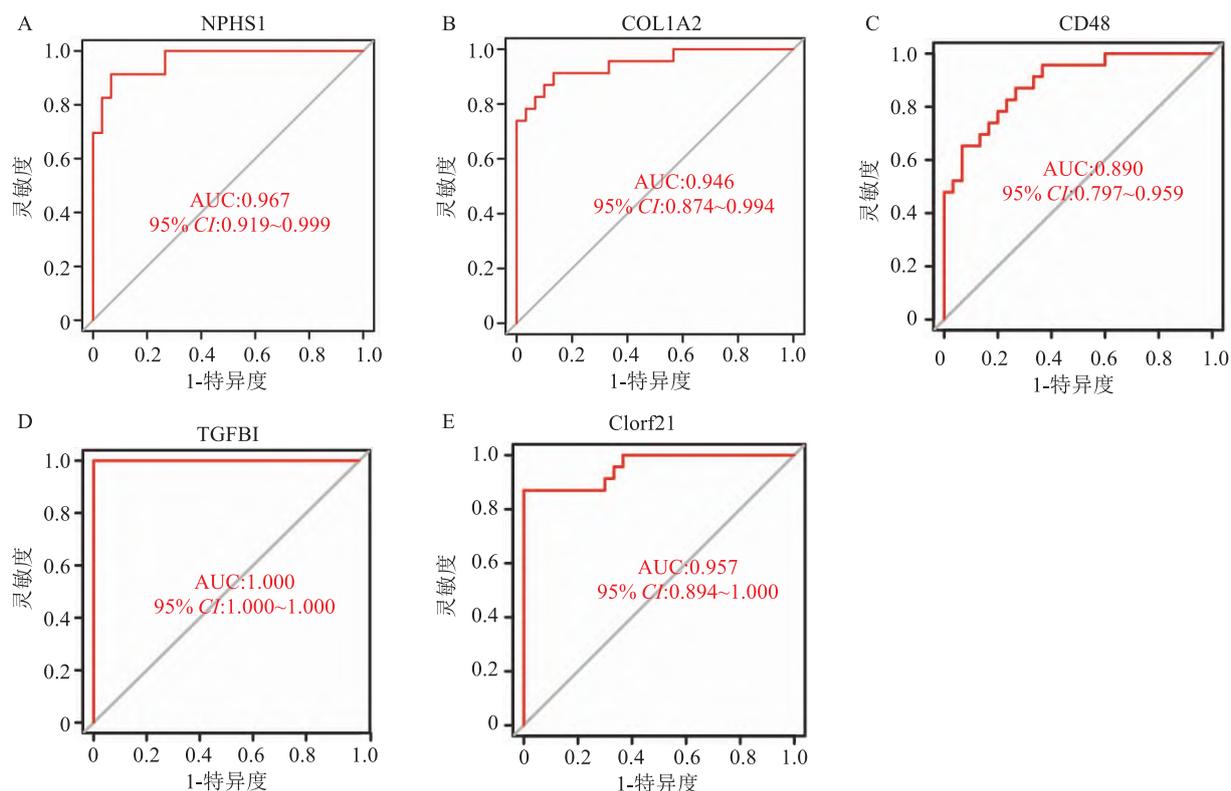


图7 目标核心基因的 ROC 曲线

2.5 核心基因的生物功能通路分析 使用 GSEA 进行数据分析 (c2. cp. all 基因集作富集使用), 结果显示 NPFS1 在细胞外机制及细胞交流信号通路显著富集 (图 8A), COL1A2 主要与整合素细胞表面互作相关 (图 8B), CD48 在自然杀伤细胞毒性和细胞表面相互作用中起作用 (图 8C), 而 TGFBI 与 ECM 糖蛋白和整合素途径有关 (图 8D)。然而 Clorf21 的研究目前匮乏, 功能尚不明确, 无法进行 GSEA 富集。

2.6 目标核心基因生物学验证 为进一步研究目标基因在 DKD 的表达, 使用 HFD (加小剂量链脲佐菌素) 诱导的糖尿病小鼠肾脏组织, 检测 5 个核心基因 (NPFS1、CD48、COL1A2、TGFBI、Clorf21) 表达水平, 如图 9 所示, NPFS1 表达降低 (图 9A), CD48 表达升高 (图 9B), 而 COL1A2、TGFBI 和 Clorf21 的表达则无显著变化 (图 9C-9E)。结果表明, NPFS1 和 CD48 符合上述的生物信息学分析结果, 它们可能与 DKD 的发生和发展具有密切关系。

2.7 核心基因的药物靶点预测 为了研究核心基因对 DKD 的有效治疗, 进行了目标基因的基因-药物的靶点预测, 以及指导 mRNA 合成的有效 miRNA, 通过基因药物相互作用网站预测基因靶点。仅预测出 LOSARTAN 以 NPFS1 为靶点发挥治疗作

用 (图 10A)。使用 3 种软件 (miRanda、miRDB、TargetScan) 分析 NPFS1 合成相关的 miRNA, 结果表明, 多种 miRNA 直接指导 NPFS1 的合成 (图 10B)。

3 讨论

DKD 是导致肾衰竭的主要原因。尽管目前对于 DKD 疾病已有较清晰的了解, 但是 DKD 的治疗方法十分有限, 具体病理机制仍不甚清楚^[7]。近年来, 基于 mRNA 微阵列分析技术的生物信息学迅速发展, 为探究疾病发展的分子机制起到指导作用。在本研究中, 通过 GEO 公共数据采用 WGCNA 对数据进行分类, 筛选出 37 个与 DKD 具有密切相关性的显著性差异基因, 包括 10 个下调基因和 27 个上调基因。对此 37 个差异基因采用 LASSO、SVM-RFE 机器算法及皮尔森相关性分析进一步筛选出与 DKD 具有密切关联性的 5 个核心基因 (NPFS1、COL1A2、CD48、TGFBI、Clorf21), 并通过 HFD 造模糖尿病小鼠肾脏组织的 RNA 检测, 进一步验证了核心基因的表达情况。

对核心基因的 GSEA 分析结果发现, NPFS1 与 NEPH1 信号通路及细胞交流呈负相关。有研究^[8]表明, NEPH1 是免疫球蛋白超家族中的一员, 在结构上与肾素相关, 诱导细胞黏附。NPFS1 和 NEPH1

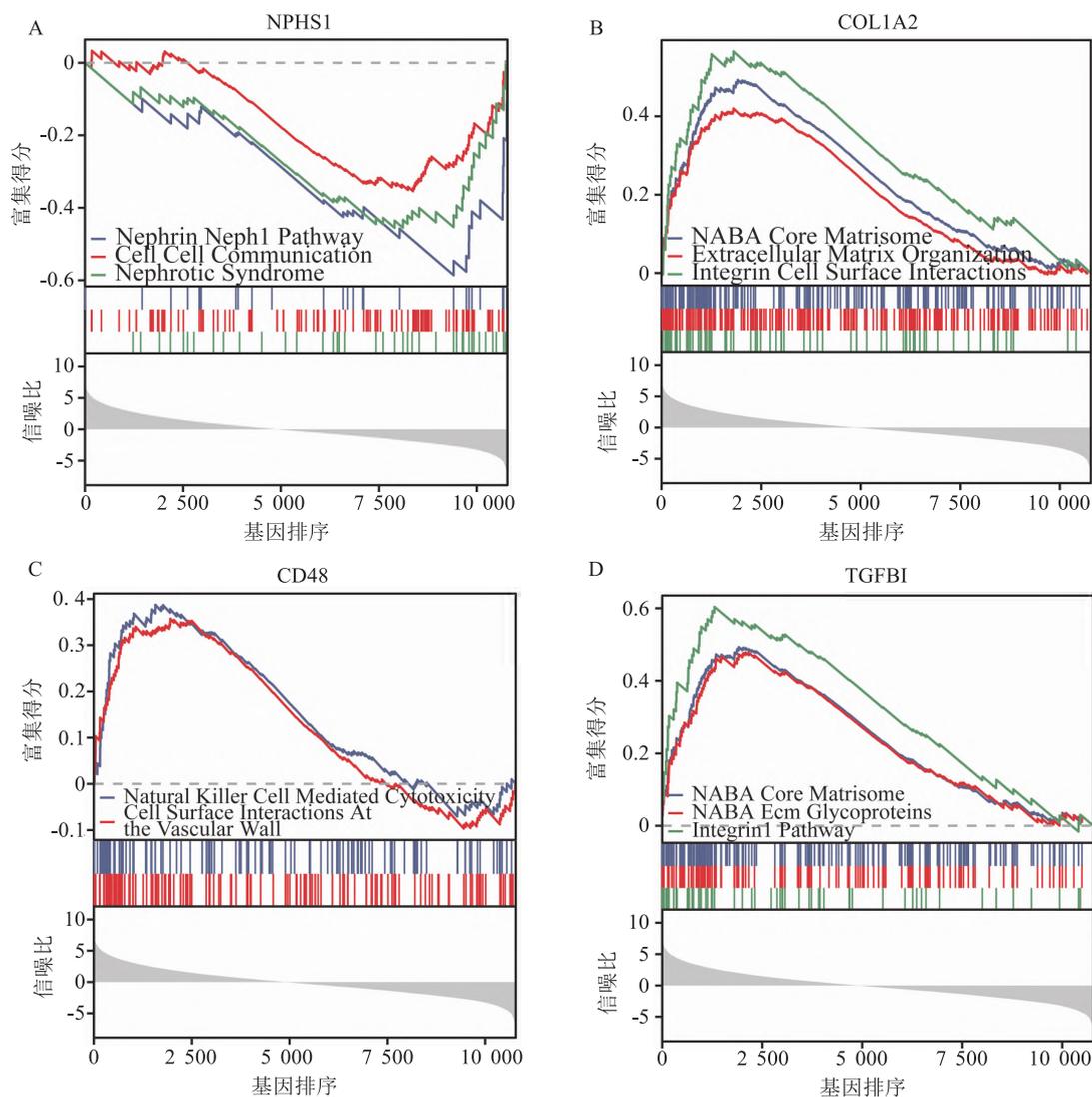


图8 目标基因的 GSEA 分析

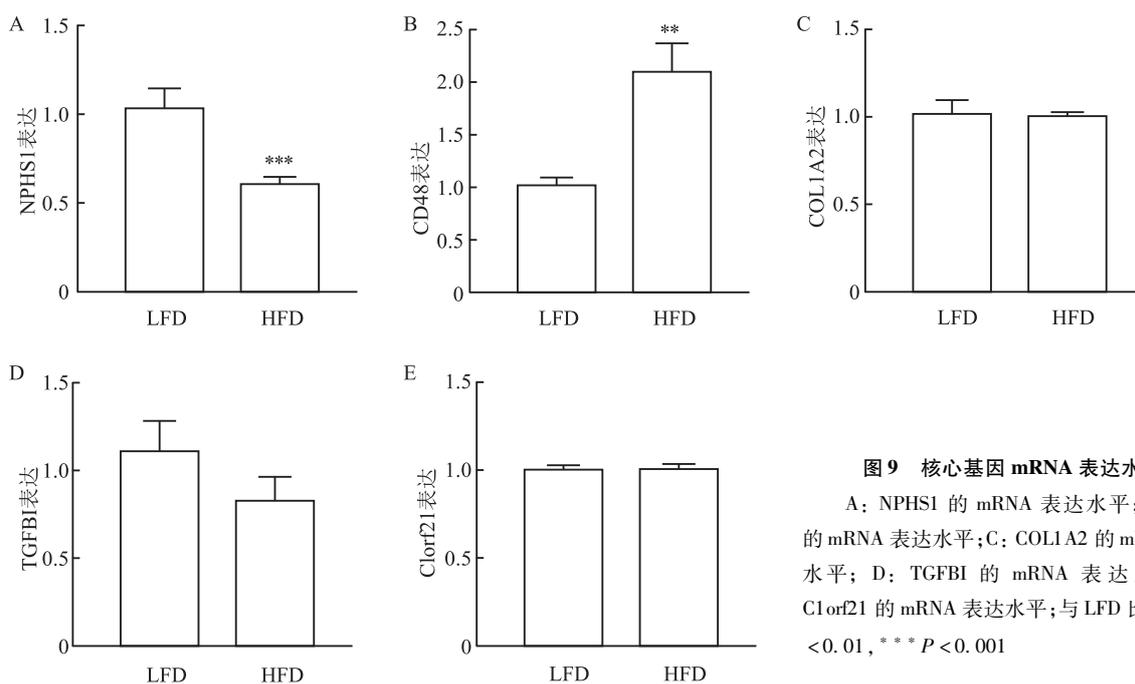


图9 核心基因 mRNA 表达水平

A: NPHS1 的 mRNA 表达水平; B: CD48 的 mRNA 表达水平; C: COL1A2 的 mRNA 表达水平; D: TGFBI 的 mRNA 表达水平; E: C1orf21 的 mRNA 表达水平; 与 LFD 比较: ** $P < 0.01$, *** $P < 0.001$

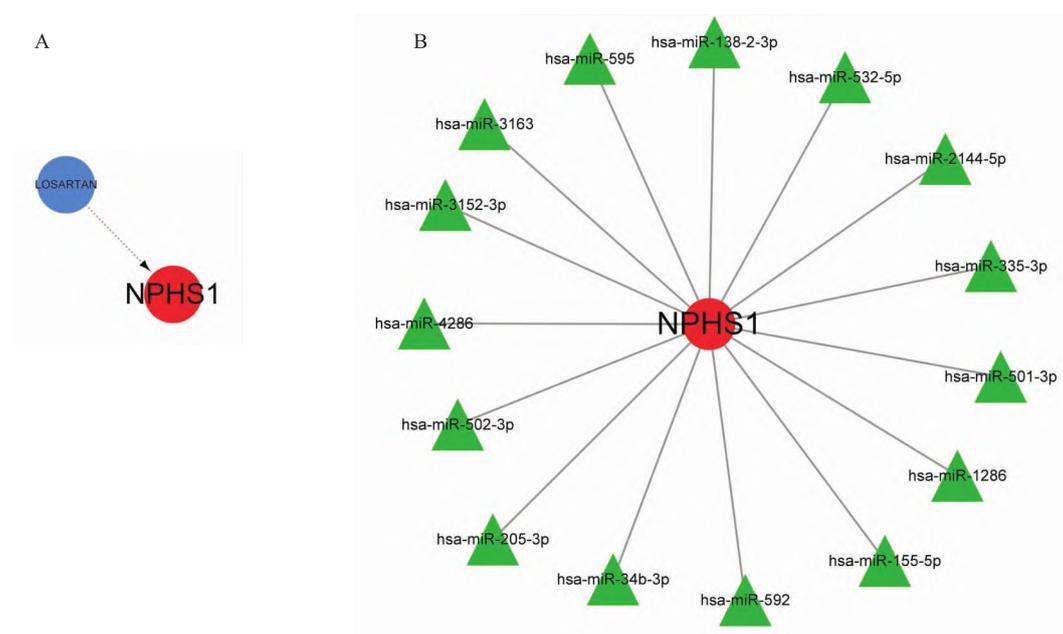


图 10 核心基因药物及靶点预测
A: 基因 NPHS1 的基因 - 药物预测; B: NPHS1 的 miRNA 预测

的基因产物 Nephrin 和 NEPH1 是 Ig 超家族的足细胞膜蛋白, NPHS1 的缺乏会严重导致尿蛋白症, 使疾病迅速恶化^[9]。在 DKD 中, 分析结果表明, COL1A2 的表达与 DKD 细胞外基质及细胞表面整合素相互作用显著相关, COL1A2 是 I 型胶原的成员, 属纤维化基因, 因此 COL1A2 的高表达与 DKD 中的肾纤维化有关^[10]。有证据指出 COL1A2 的表达与 DKD 中由高糖诱导的炎症反应具有显著相关性^[11], 针对 COL1A2 的治疗, 是 DKD 管理的潜在干预策略, 然而, 在糖尿病小鼠肾脏组织中, COL1A2 却没有显著变化。与此同时, 越来越多的研究^[12]发现, DKD 的发展中存在着免疫机制的参与, T 淋巴细胞、B 淋巴细胞、巨噬细胞等与 DKD 密切相关, CD48 主要参与自然杀伤细胞介导的细胞毒性及反应体细胞表面相互作用(图 8C)。CD48 是信号淋巴细胞激活分子家族的成员, 参与免疫细胞的黏附和激活^[13], 在 DKD 中的高表达激活了自然杀伤细胞的活力, 引起了免疫细胞的激活。同样的, TGFBI 是转化因子诱导蛋白, 参与细胞的生长、分化以及细胞黏附等。在肿瘤研究中, TGFBI 在很大程度上促进了肿瘤细胞的进展^[14], 在 DKD 中却没有报道。然而, TGFBI 能够调控 ECM 糖蛋白, 而 ECM 蛋白与脂肪组织相关, 并调节炎症、纤维化、血管生成和代谢恶化^[15]。因此推测 TGFBI 的表达调节 DKD 中的肾纤维化和免疫炎症具有密切关系, 但本研究显示,

TGFBI 在 HFD 小鼠肾脏组织中并没有发生显著变化。除此之外, 发现 C1orf21 同样是被显著富集出的基因, C1orf21 是一个目前还未功能注释的非特征性蛋白质编码基因。针对 C1orf21 的研究十分匮乏, 其表达差异意味着 1 号染色体的功能性障碍, 这有待于进一步研究确认。综上所述, NPHS1、CD48、COL1A2、TGFBI、C1orf21 基因可能与 DKD 具有密切相关性, 通过参与细胞表面相互作用, 影响免疫细胞的激活、细胞黏附以及促进肾纤维化影响疾病的恶化。而基因靶点药物和有效 miRNA 的预测为研究 DKD 的治疗提供了重要指导。本研究表明, COL1A2、CD48 与 DKD 的发展具有显著相关性, 而针对 NPHS1 的基因靶点药物治疗及 miRNA 的预测在 DKD 中有可能作为重要的突破点。

参考文献

- [1] Wei R, Qiao J, Cui D, et al. Screening and identification of Hub genes in the development of early diabetic kidney disease based on weighted gene co-expression network analysis[J]. *Front Endocrinol (Lausanne)*, 2022, 13: 883658.
- [2] Sifuentes-Franco S, Padilla-Tejeda D E, Carrillo-Ibarra S, et al. Oxidative stress, apoptosis, and mitochondrial function in diabetic nephropathy[J]. *Int J Endocrinol*, 2018, 2018: 1875870.
- [3] Donate-Correa J, Luis-Rodriguez D, Martin-Nunez E, et al. Inflammatory targets in diabetic nephropathy [J]. *J Clin Med*, 2020, 9(2):458.
- [4] Sawaf H, Thomas G, Talierecio J J, et al. Therapeutic advances in

- diabetic nephropathy[J]. *J Clin Med*, 2022, 11(2):378.
- [5] Kopel J, Pena-Hernandez C, Nugent K. Evolving spectrum of diabetic nephropathy[J]. *World J Diabetes*, 2019, 10(5): 269 – 79.
- [6] Milano M, Cannataro M. Network models in bioinformatics; modeling and analysis for complex diseases [J]. *Brief Bioinform*, 2023, 24(2):bbad016.
- [7] Budge K, Dellepiane S, Yu S M, et al. Complement, a therapeutic target in diabetic kidney disease[J]. *Front Med (Lausanne)*, 2020, 7: 599236.
- [8] Heikkila E, Ristola M, Havana M, et al. Trans-interaction of nephrin and Neph1/Neph3 induces cell adhesion that associates with decreased tyrosine phosphorylation of nephrin [J]. *Biochem J*, 2011, 435(3): 619 – 28.
- [9] Gerke P, Huber T B, Sellin L, et al. Homodimerization and heterodimerization of the glomerular podocyte proteins nephrin and NPH1 [J]. *J Am Soc Nephrol*, 2003, 14(4): 918 – 26.
- [10] Chen J, Luo S F, Yuan X, et al. Diabetic kidney disease-predisposing proinflammatory and profibrotic genes identified by weighted gene co-expression network analysis (WGCNA) [J]. *J Cell Biochem*, 2022, 123(2): 481 – 92.
- [11] Wang S, Sun K, Hu H, et al. MiR-1297 attenuates high glucose-induced injury in HK-2 cells *via* targeting COL1A2 [J]. *Nephrology (Carlton)*, 2021, 26(7):623 – 31.
- [12] Chang T T, Li S Y, Lin L Y, et al. Macrophage inflammatory protein-1beta as a novel therapeutic target for renal protection in diabetic kidney disease [J]. *Biomed Pharmacother*, 2023, 161: 114450.
- [13] Mcardel S L, Terhorst C, Sharpe A H. Roles of CD48 in regulating immunity and tolerance [J]. *Clin Immunol*, 2016, 164: 10 – 20.
- [14] Corona A, Blobe G C. The role of the extracellular matrix protein TGFBI in cancer [J]. *Cell Signal*, 2021, 84: 110028.
- [15] Lee S G, Chae J, Woo S M, et al. TGFBI remodels adipose metabolism by regulating the Notch-1 signaling pathway [J]. *Exp Mol Med*, 2023, 55(3): 520 – 31.

Screening and identification of core genes for diabetic nephropathy

Wu Xiaojun^{1,2}, Ni Feixue^{1,2}, Xu Yuxue^{1,2,3}

(¹*School of Pharmacology, Binzhou Medical University, Yantai 264003*; ²*Shandong Technology Innovation Center of Molecular Targeting and Intelligent Diagnosis and Treatment, Binzhou Medical University, Yantai 264003*; ³*Yantai Stem Cell and Regenerative Medicine Key Laboratory, Binzhou Medical University, Yantai 264003*)

Abstract Objective To screen the core genes of diabetic kidney disease (DKD) based on bioinformatics, explore the therapeutic targets of DKD, and discuss its possible regulatory mechanism. **Methods** The expression data matrix of glomerular transcriptome in patients with DKD in GEO database (GSE30528, GSE47183) was extracted, and the differentially expressed genes (DEGs) were screened by bioinformatics methods to identify the core differential genes, and then gene expression and enrichment analysis (GSEA) were conducted to predict effective targets. **Results** By screening and identifying the mRNA expression matrix of DKD, five core genes were screened out. Among them, C1orf21 and NPHS1 were significantly down regulated, and CD48, COL1A2, and TGFBI were up regulated. NPHS1 and CD48 were significantly related to immune differences, intercellular communication, and cell surface interaction. Through receiver operating characteristic curve (ROC) analysis and GSEA analysis and drug target prediction, it might be of great significance for the treatment of DKD. **Conclusion** The core genes screened in this study have significant correlation with DKD, which may be used as effective markers for the treatment of diabetes. And then, this study provides a theoretical basis for the treatment and identification of DKD.

Key words diabetic kidney disease; GEO database; transcriptome; bioinformatics; genetic screening; drug target