

基于加权基因共表达网络分析鉴定 2 型糖尿病枢纽基因

谭铨广¹ 张东然² 黄淑兰^{3*}

1. 罗定市人民医院 罗定 527200

2. 普宁市人民医院 普宁 515300

3. 中国人民解放军南部战区总医院内分泌科 广州 510010

摘要: 目的: 利用生物信息学技术筛选 2 型糖尿病 (T2DM) 枢纽基因。方法: 从 GEO 数据库下载 T2DM 相关芯片 GSE20966, 筛选差异表达基因 (DEGs), 通过加权基因共表达网络 (WGCNA) 分析基因与临床性状的关系, 获取关键 DEGs 并进行功能富集。采用 LASSO 回归筛选枢纽基因, 并绘制 ROC 曲线评估诊断效能。结果: 共鉴定 209 个 DEGs, 其中 157 个上调、52 个下调。WGCNA 显示黄色模块与 T2DM 最相关 ($r=0.69$, $P=0.001$)。交集获得 59 个关键 DEGs, GO 和 KEGG 富集分析显示其主要参与转运、细胞外区域、胰腺分泌等过程。LASSO 回归筛选出 7 个枢纽基因 (MDFIC、EFHD2、LMBRD2、GALNT14、IGFBP3、KCNG3、GPR116), ROC 曲线 AUC=0.96, 显示良好诊断效能。结论: 这 7 个基因有望作为 T2DM 的新型诊断标志物和治疗靶点。

关键词: 2 型糖尿病; 生物信息学; 枢纽基因

2 型糖尿病 (T2DM) 是一种因胰岛素分泌或作用缺陷导致的慢性代谢性疾病。国际糖尿病联盟报告显示, 全球糖尿病患者约 5.37 亿, 其中 T2DM 占 90% 以上, 预计 2045 年将增至 7.82 亿^[1]。我国成人糖尿病患病率达 12.8%^[2]。T2DM 是多基因遗传病, 已发现约 400 个相关基因变异^[3], 但其发病机制尚未完全明确。利用生物信息学筛选关键基因对揭示 T2DM 机制及开发新靶点具有重要意义^[4]。

1 材料与方法

1.1 材料

通过从 GEO 数据库 (<https://www.ncbi.nlm.nih.gov/geo/>) 下载与 T2DM 相关的基因芯片数据集 GSE20966, 所使用的平台为 GPL10295。芯片数据是由 10 例 T2DM 患者胰腺组织样本及 10 例正常人胰腺组织样本组成。

1.2 方法

(1) 差异表达基因筛选: 运用 R 语言的 “limma” 包来筛选差异表达基因 (DEGs), 并绘制 DEGs 火山图。其筛选的统一标准是 P 值 < 0.05 并且 $|\log_2$ 差异倍数 (fold change, FC) $| > 2$ 。

(2) WGCNA 模块构建及可视化: WGCNA 分析作为一种常见的模块化分析技术常被用于疾病的生物标志物或者药物靶点的识别与筛选以及高度协同变化的基因表达矩阵

的鉴定, 为疾病的诊治提供新方向^[5]。

(3) 功能富集分析: 将 WGCNA 分析获得的关键模块基因与 DEGs 进行交集运算, 并筛选出重叠的基因作为关键 DEGs。对于关键 DEGs, 充分利用 R 语言中的 “clusterProfiler” 包进行基因本体论功能 (Gene Ontology, GO)^[6] 和基因组百科全书通路途径 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 富集分析^[7]。经校正后, 具有统计学意义的结果被定义为 P 值 < 0.05 。

(4) LASSO 回归分析筛选枢纽基因: LASSO 回归分析实现了变量的选择与正则化的同步进行, 从而提高生成的统计模型预测的准确性与可解释性^[8, 9]。使用 R 语言中的 “pROC” 包绘制枢纽基因的诊断 ROC 曲线并算出 AUC 值。

2 结果

2.1 差异基因的筛选结果

对数据集 GSE20966 进行独立归一化和标准化后, 使用 R 语言中的 “limma” 包进行差异基因分析。结果发现, 与正常对照组相比, T2DM 组有 157 个基因显著上调, 52 个基因表达显著下降 (图 1)。

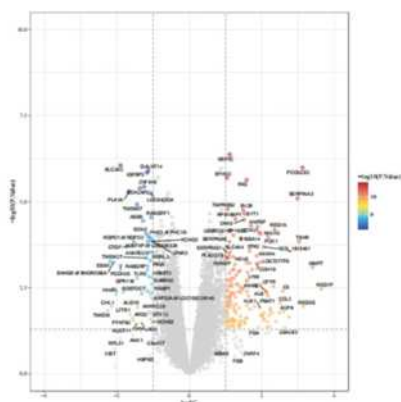
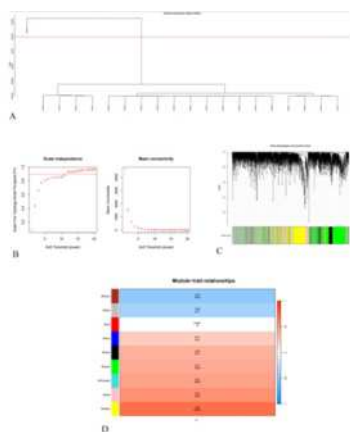


图 1 T2DM 差异表达基因火山图

2.2 WGCNA 网络的构建

数据集 GSE20966 去除批次效应后, R 语言排除异常离群样本 GSM524155 (图 2A)。通过表达矩阵与表型数据的共同载入, 确定软阈值为 14, 构建表达网络 (图 2B)。灵活运用 WGCNA 分析使得基因共被分为 9 个模块 (图 2C)。利用 Pearson 法分析对基因模块与临床特性的相关性进行了分析, 得到模块—临床特性的相关性并绘制关联图。结果发现黄色模块的 1287 个基因与 T2DM 的相关性最高, 其相关性系数为 0.69, P 值为 0.001, 因此为关键基因模块 (图 2D)。



注 :A.GEO 样本聚类树; B. 尺度独立性和平均连接阈值图; C. 基因模块聚类树; D. 模块—特征相关性热图。

图 2 WGCNA 分析样本聚类图及软阈值图

2.3 关键 DEGs 的筛选及富集分析

将黄色模块中的 1287 个基因与 209 个差异表达基因取交集, 得到 59 个关键 DEGs, 如图 3。对这 59 个关键 DEGs 展开 GO 和 KEGG 富集分析。GO-BP 分析结果显示, 关键 DEGs 主要富集在转运、定位的建立、转运的调节过程、

跨膜转运、跨膜转运的调节等生物学过程 (图 4A)。对于 GO-CC 分析而言, 前 3 个主要富集的术语是细胞外区域、囊泡、胞外分泌体 (图 4B)。根据 GO-MF 分析结果显示, 前 3 个显著富集的通路包括 L-氨基酸肽酶活性、丝氨酸内肽酶活性和丝氨酸肽酶活性 (图 4C)。KEGG 信号途径显示关键 DEGs 主要富集在胰腺的分泌、神经活性配体—受体相互作用以及蛋白质的消化与吸收等通路 (图 5)。

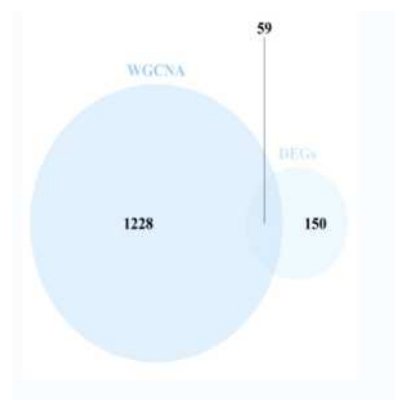
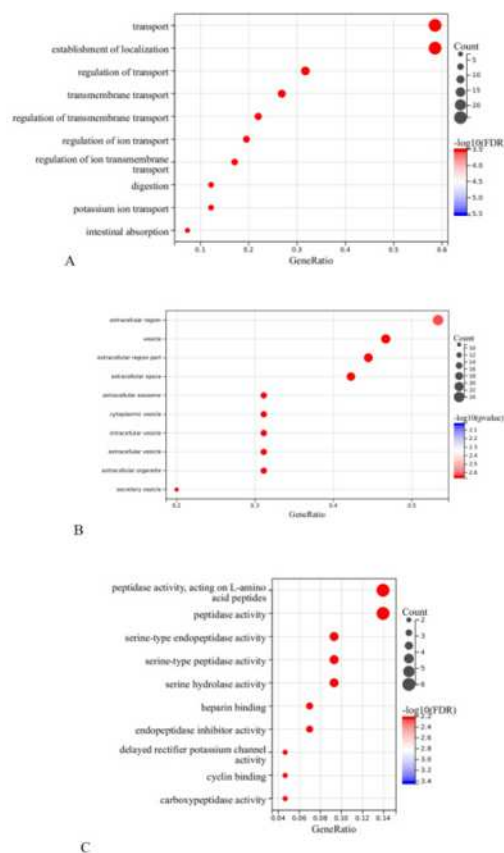


图 3 DEGs 与模块核心基因韦恩图



注 :A. 生物学工程; B. 细胞组分; C. 分子功能。

图 4 关键 DEGs GO 富集分析

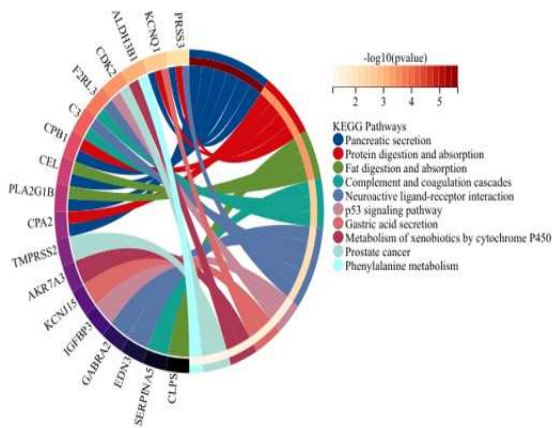
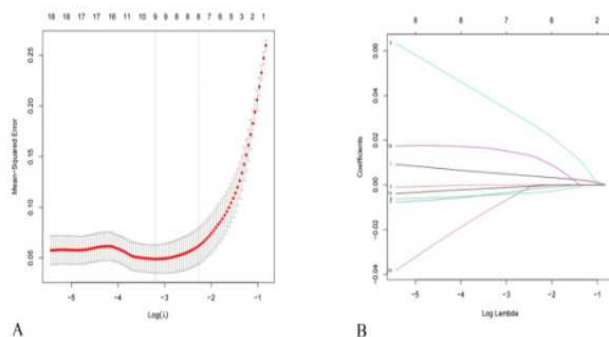


图 5 关键 DEGs KEGG 富集分析

2.4 Lasso 回归筛选枢纽基因

通过 R 语言中的“glmnet”包对关键 DEGs 进行 Lasso 回归分析 (图 6A), 并选择 10 折交叉验证误差最小值作为模型的最优值 (图 6B), 此时筛选枢纽基因有 7 个, 其中上调基因有 3 个, 分别是 MyoD 家族抑制域含蛋白 (MyoD family inhibitor domain containing protein, MDFIC)、EF 手性域家族成员 D2 (EF-hand domain family member D2, EFHD2)、含 LMBR1 结构域 2 (LMBR1 domain containing 2, LMBRD2)。下调基因有 4 个, 分别是多肽 N-乙酰半乳糖胺基转移酶 14 (Polypeptide N-acetylgalactosaminotransferase 14, GALNT14)、胰岛素样生长因子结合蛋白 (insulin like growth factor binding protein 3, IGFBP3)、电压门控钾离子通道蛋白家族成员 3 (voltage-gated channel subfamily G member 3, KCNG3)、孤儿粘附 G 蛋白偶联受体 116 (adhesion G protein-coupled receptor F5, GPR116)。通过 ROC 曲线验证枢纽基因, 其 AUC 值为 0.96 (图 7)。表明这些基因具有较高的诊断准确性。



注: A: Lasso 回归系数筛选; B: Lasso 回归轨迹图

图 6 Lasso 回归分析结果

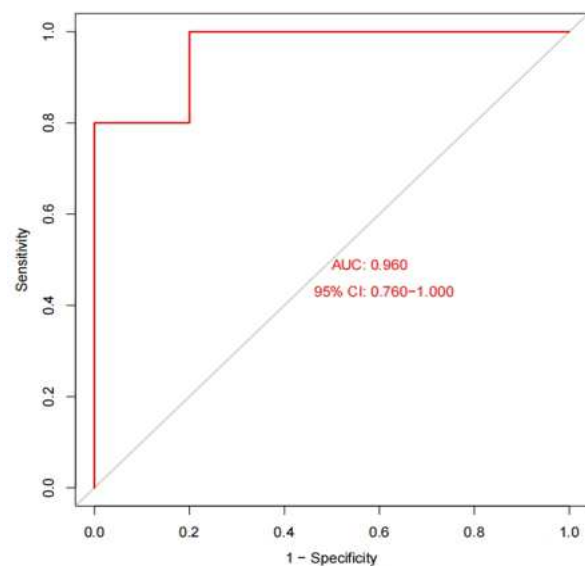


图 7 枢纽基因在训练集中的 ROC 曲线

3 讨论

本研究基于 GSE20966 数据集, 在 T2DM 患者与正常人胰腺组织之间鉴定出 209 个差异表达基因 (DEGs), 其中上调 157 个、下调 52 个。通过 WGCNA 构建共表达网络, 共识别出 9 个基因模块, 其中黄色模块与 T2DM 相关性最高 ($r=0.69$, $P=0.001$), 提示其在疾病发生中起关键作用。将该模块 1287 个基因与 DEGs 取交集, 获得 59 个关键 DEGs。GO 富集分析显示, 这些基因主要参与转运、细胞外区域、囊泡及丝氨酸内肽酶活性等生物过程; KEGG 分析表明它们显著富集于胰腺分泌、神经活性配体-受体相互作用和蛋白质消化吸收等通路, 这些通路已被报道与 T2DM 发病及进展密切相关^[10,11]。进一步经 LASSO 回归筛选出 7 个枢纽基因, 包括 3 个上调基因 (MDFIC、EFHD2、LMBRD2) 和 4 个下调基因 (GALNT14、IGFBP3、KCNG3、GPR116)。ROC 曲线显示其 AUC 值为 0.96, 表明这些基因具有优异的诊断性能。

EFHD2 作为 EF-hand 钙结合蛋白家族成员, 通过蛋白互作参与多种生物学过程^[12]。研究表明其过表达可促进糖尿病肾病早期足细胞线粒体依赖性凋亡, 提示其在糖尿病肾病进程中起重要作用^[13,14]。IGFBP3 通过结合胰岛素样生长因子及其受体, 参与代谢与生长调节^[15]; 其在 2 型糖尿病骨质疏松 (T2DOP) 中表达下调, 可能通过抑制骨细胞活性及矿化促进 T2DOP 发生^[16]。GPR116 是白色脂肪组织中可溶性纤维连接蛋白 III 型结构域包含 4 (fibronectin type III

domain containing 4, sFNDC4) 的受体^[17], 在机体建立了控制系统葡萄糖稳态的内分泌 FNDC4-GPR116 轴。而该内分泌轴在糖尿病患者中受损, 注射重组 Fc sFNDC4 可以起到治疗糖尿病前期小鼠高血糖症的作用, 可作为临床治疗糖尿病前期的靶点^[18]。

目前 MDFIC、LMBRD2、GALNT14 和 KCNG3 与 T2DM 的关系尚不明确。MDFIC 可结合并调控机械敏感离子通道 PIEZO 活性, 可能通过影响肥胖相关机制间接参与 T2DM^[19-21]。LMBRD2 参与 G 蛋白信号转导与肾上腺素受体调节, 推测其可能通过影响骨骼肌葡萄糖摄取参与降糖过程^[22,23]。GALNT14 作为 O-糖基化起始酶, 参与细胞凋亡和铁死亡调控; 其表达下调可能通过促进脂质过氧化影响 T2DM 进程^[24,25]。KCNG3 是一类编码电压门控钾离子通道亚单位的重要蛋白, 该电压门控钾离子通道在调节神经递质释放、胰岛素释放、平滑肌收缩以及细胞的形态等方面发挥着重要的作用^[26]。关于 KCNG3 基因参与调控 T2DM 的具体机制需要实验验证。本研究提示 7 个枢纽基因 (MDFIC、EFHD2、LMBRD2、GALNT14、IGFBP3、KCNG3、GPR116) 具有良好的诊断价值, 有望成为 T2DM 潜在生物标志物和治疗靶点。但结果基于生物信息学分析, 仍需后续实验与临床验证。

作者贡献: 谭铨广: 论文总体构思, 撰写初稿; 张东然: 论文构思; 黄淑兰: 论文总体构思, 审读, 修订;

利益冲突: 所有作者声明无利益冲突。

参考文献:

- [1] 中华医学会糖尿病学分会, 朱大龙, 郭立新. 中国糖尿病防治指南(2024 版)[J]. 中华糖尿病杂志, 2025, 17(1): 16-139.
- [2] 刘月, 管敏鑫, 陈烨, 等. 线粒体基因突变相关的糖尿病研究进展[J]. 基因组学与应用生物学, 2017, 36(11): 4562-4569.
- [3] Ahmad E, Lim S, Lamprey R, et al. Type 2 diabetes[J]. Lancet, 2022, 400(10365):1803-1820.
- [4] Mulvey LPA, May MR, Brown JM, Höhna S, Wright AM, Warnock RCM. Assessing the Adequacy of Morphological Models Using Posterior Predictive Simulations. Syst Biol. 2025;74(1):34-52
- [5] Li J, Yan N, Li X, He S, Yu X. Identification and analysis of hub genes of hypoxia-immunity in type 2 diabetes mellitus. Front Genet. 2023;14:1154839. Published 2023 Apr 21.

doi:10.3389/fgene.2023.1154839

- [6] Zhang YD, Chen YR, Zhang W, Tang BQ. Assessing prospective molecular biomarkers and functional pathways in severe asthma based on a machine learning method and bioinformatics analyses. J Asthma. 2025;62(3):465-480.
- [7] Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587-D592.
- [8] Zhang X, Jiang W, Zhao H (2024) Integration of expression QTLs with fine mapping via SuSiE. PLOS Genetics 20(1): e1010929.
- [9] Alhamzawi R, Ali H. The Bayesian adaptive lasso regression[J]. Math Biosci, 2018, 303:75-82.
- [10] Li J, Yan N, Li X, et al. Identification and analysis of hub genes of hypoxia-immunity in type 2 diabetes mellitus[J]. Frontiers in Genetics, 2023, 14.
- [11] Ding L, Fan L, Xu X, et al. Identification of core genes and pathways in type 2 diabetes mellitus by bioinformatics analysis[J]. Molecular medicine reports, 2019, 20(3):2597-2608.
- [12] Fu JT, Liu J, Wu WB, et al. Targeting EFHD2 inhibits interferon- γ signaling and ameliorates non-alcoholic steatohepatitis. J Hepatol. 2024;81(3):389-403.
- [13] Wang Z B, Zhang S, Li Y, et al. LY333531, a PKC β inhibitor, attenuates glomerular endothelial cell apoptosis in the early stage of mouse diabetic nephropathy via down-regulating swiprosin-1[J]. Acta Pharmacol Sin, 2017, 38(7):1009-1023.
- [14] Wang R M, Wang Z B, Wang Y, et al. Swiprosin-1 Promotes Mitochondria-Dependent Apoptosis of Glomerular Podocytes via P38 MAPK Pathway in Early-Stage Diabetic Nephropathy[J]. Cell Physiol Biochem, 2018, 45(3):899-916.
- [15] Cakir B, Hellström W, Tomita Y, et al. IGF1, serum glucose, and retinopathy of prematurity in extremely preterm infants. JCI Insight. 2020;5(19):e140363. Published 2020 Oct 2. doi:10.1172/jci.insight.140363
- [16] Yang J, He Q, Wang Y, et al. Gegen Qinlian Decoction ameliorates type 2 diabetes osteoporosis via IGFBP3/MAPK/NFATc1 signaling pathway based on cytokine antibody array[J].

Phytomedicine, 2022,94:153810.

[17]Kang H, Fichna J, Matlawska-Wasowska K, et al. The Expression Pattern of Adhesion G Protein-Coupled Receptor F5 Is Related to Cell Adhesion and Metastatic Pathways in Colorectal Cancer—Comprehensive Study Based on In Silico Analysis[J]. Cells, 2022,11(23):3876.

[18]Georgiadi A, Lopez-Salazar V, Merahbi R E, et al. Orphan GPR116 mediates the insulin sensitizing effects of the hepatokine FNDC4 in adipose tissue[J]. Nature Communications, 2021,12(1).

[19]Oakley R H, Busillo J M, Cidlowski J A. Cross-talk between the glucocorticoid receptor and MyoD family inhibitor domain-containing protein provides a new mechanism for generating tissue-specific responses to glucocorticoids[J]. The Journal of biological chemistry, 2017,292(14):5825–5844.

[20]Zhou Z, Ma X, Lin Y, et al. MyoD-family inhibitor proteins act as auxiliary subunits of Piezo channels[J]. Science, 2023,381(6659):799–804.

[21]Aberle J, Lautenbach A, Meyhofer S, et al. Obesity and Diabetes[J]. Exp Clin Endocrinol Diabetes, 2021,129(S 01):S44–S51.

[22]Malhotra A, Ziegler A, Shu L, et al. De novo missense

variants inLMBRD2 are associated with developmental and motor delays, brain structure abnormalities and dysmorphic features[J]. Journal of Medical Genetics, 2021,58(10):712–716.

[23]Ekman Ryding A. A new approach to treat type 2 diabetes – Targeting a non-insulin dependent pathway[J]. Drug Discovery Today, 2021,26(11):2487–2488.

[24] 吕依倡, 张红飞, 张伟. GALNT14 与肿瘤 [J]. 中国生物化学与分子生物学报, 2019,35(01):7–12.

[25]Li H W, Liu M B, Jiang X, et al. GALNT14 regulates ferroptosis and apoptosis of ovarian cancer through theEGFR/mTOR pathway[J]. Future Oncol, 2022,18(2):149–161.

[26]O Donnell A M, Nakamura H, Tomuschat C, et al. Altered expression of KCNG3 and KCNG4 in Hirschsprung’ s disease[J]. Pediatric Surgery International, 2019,35(2):193–197.

作者简介: 谭铨广 (1996), 男, 汉族, 广东省罗定市人, 研究生学历, 罗定市人民医院, 呼吸内科住院医师, 研究方向: 呼吸系统疾病

通讯作者: 黄淑兰 (1996), 女, 汉族, 广东省揭阳市人, 研究生学历, 中国人民解放军南部战区总医院, 内分泌科医师, 研究方向: 糖尿病、甲状腺疾病等内分泌疾病相关的机制研究