doi: 10.3978/j.issn.2095-6959.2021.12.005
View this article at: https://dx.doi.org/10.3978/j.issn.2095-6959.2021.12.005

非酒精性脂肪性肝炎关键基因的鉴定及通路分析

陈懿[1]，严丹[2]，李雪[1]，张友才[1]

(1. 温州医科大学附属第一医院感染内科，浙江 温州 325000；2. 金华市中心医院呼吸内科，浙江 金华 321000)

[摘　要]　**目的：**非酒精性脂肪性肝病(nonalcoholic fatty liver disease，NAFLD)是最常见的慢性肝病之一，包括非酒精性单纯性脂肪肝(nonalcoholic fatty liver，NAFL)、非酒精性脂肪性肝炎(nonalcoholic steatohepatitis，NASH)、肝硬化，最终可能发展为肝细胞癌。本研究旨在通过加权基因共表达网络分析(weighted gene co-expression network analysis，WGCNA)构建一个基因共表达网络，以探索NASH发生发展的潜在关键通路和基因。**方法：**数据来自基因表达综合(Gene Expression Omnibus，GEO)数据库中GSE48452和GSE89632两个数据集。使用WGCNA确定与NASH相关的模块。通过功能富集分析，探讨其潜在的生物学功能，并对这些基因进行蛋白质-蛋白质相互作用分析。**结果：**GSE48452数据集中的黑色模块、GSE89632数据集中的灰色模块与NASH相关最显著。基因本体(Gene Ontology，GO)富集分析显示目标基因的生物学功能分别集中在脂质降解、核染色体及有机酸的结合。京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes，KEGG)通路显示基因在过氧化物酶体增殖物激活受体(peroxisome proliferator-activated receptor，PPAR)通路富集尤为显著。**结论：**通过WGCNA分析，确定了与NASH相关的潜在基因[脂蛋白脂肪酶(lipoprotein lipase，*LPL*)、胸苷酸合成酶(thymidylate synthetase，*TYMS*)、细胞色素P450家族26亚家族A成员1(cytochrome P450 family 26 subfamily A member 1，*CYP26A1*)、脂肪酸结合蛋白4(fatty acid-binding protein 4，*FABP4*)、脂肪酸结合蛋白5(fatty acid-binding protein 5，*FABP5*)、微染色体维持蛋白2(minichromosome maintenance protein 2，*MCM2*)、微染色体维持蛋白5(minichromosome maintenance protein 5，*MCM5*)、GINS复合物亚单位2(GINS complex subunit 2，*GINS2*)]及关键通路。这些基因可能参与NAFL到NASH的转化过程，具有一定的诊断和治疗价值。

[关键词]　非酒精性单纯性脂肪肝；非酒精性脂肪性肝炎；共表达网络；基因

Identification and pathway analysis of key genes in nonalcoholic steatohepatitis

CHEN Yi[1], YAN Dan[2], LI Xue[1], ZHANG Youcai[1]

(1. Department of Infection, First Affiliated Hospital of Wenzhou Medical University, Wenzhou Zhejiang 325000; 2. Department of Respiratory, Jinhua Municipal Central Hospital, Jinhua Zhejiang 321000, China)

Abstract　**Objective:** Nonalcoholic fatty liver disease (NAFLD) is one of the most common chronic liver diseases, including

收稿日期 (Date of reception)：2020–09–25

通信作者 (Corresponding author)：张友才，Email: zhangyc0815@163.com

基金项目 (Foundation item)：温州市科技计划项目 (Y20150016)。This work was supported by Wenzhou Science and Technology Project, China (Y20150016).

a series of pathological changes, from benign nonalcoholic fatty liver (NAFL), nonalcoholic steatohepatitis (NASH), liver cirrhosis, and finally to hepatocellular carcinoma. The purpose of this study was to construct a gene co-expression network by using weighted gene co-expression network analysis (WGCNA) to explore the potential key pathways and genes in the genesis and development of NASH. **Methods:** Data were collected from two data sets GSE48452 and GSE89632 in the Gene Expression Omnibus (GEO) database. WGCNA was used to identify NASH-related modules. The potential biological functions of NASH were discussed by functional enrichment analysis. These genes were analyzed with protein-protein interactions. **Results:** The black module in GSE48452 data set and gray module in GSE89632 data set had the most significant correlation with NASH. The enrichment analysis of Gene Ontology (GO) showed that the biological functions of the target gene focused on lipid degradation, nuclear chromosome, and organic acid binding. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway showed that genes were enriched in peroxisome proliferator-activated receptor (PPAR) pathway. **Conclusion:** Through WGCNA analysis, the potential genes [lipoprotein lipase (*LPL*), thymidylate synthetase (*TYMS*), cytochrome P450 family 26 subfamily A member 1 (*CYP26A1*), fatty acid-binding protein 4 (*FABP4*), fatty acid-binding protein 5 (*FABP5*), minichromosome maintenance protein 2 (*MCM2*), minichromosome maintenance protein 5 (*MCM5*), GINS complex subunit 2 (*GINS2*)] and key pathways related to NASH have been identified. These genes may be involved in the transformation process from NAFL to NASH and have certain diagnostic and therapeutic value.

Keywords nonalcoholic fatty liver; nonalcoholic steatohepatitis; co-expression network; gene

随着肥胖及其相关代谢综合征的流行，非酒精性脂肪性肝病(nonalcoholic fatty liver disease，NAFLD)的发病率逐渐升高，成为世界范围内慢性肝病的常见病因之一[1]。NAFLD是指在缺乏其他肝病病因，如慢性病毒性肝炎，或使用胺碘酮等引起脂肪变性的药物，以及排除其他慢性肝病，如自身免疫性肝炎、血色素沉着症、威尔逊病等的情况下，肝脏存在脂肪变性(≥5%)[2]。NAFLD包括3种具有不同预后的组织病理学疾病：非酒精性单纯性脂肪肝(nonalcoholic fatty liver，NAFL)、非酒精性脂肪性肝炎(nonalcoholic steatohepatitis，NASH)和与NAFL相关的纤维化/肝硬化[3-4]。NASH是一种侵袭性的NAFLD，可发展为肝硬化和肝细胞癌，并迅速成为终末期肝病或肝移植的主要病因之一[5]。

NAFLD的发生发展受到多种因素的影响，目前为大多数人所接受的是“多重打击假说”，认为遗传易感性、环境因素和饮食习惯等多个因素均可以通过不同途径影响肥胖、代谢综合征和胰岛素抵抗的发展，从而导致肝损伤[6]。目前，生活方式干预，如饮食热量限制和运动，是NASH/NAFLD治疗的基石，但很难实现和维持，这突出了药物治疗的迫切需要[7]。加权基因共表达网络分析(weighted gene co-expression network，WGCNA)作为一种常用的基因模块化分析技术，已经被广泛应用于鉴定和筛选复杂疾病的分子标志物或药物靶点[8]。

本研究利用WGCNA筛选出NAFL进展为NASH的相关基因模块，并探究其分子机制研究，为临床治疗提供可能作用的靶点。

1 资料与方法

1.1 数据收集与预处理

本研究中所有数据来自基因表达综合(Gene Expression Omnibus，GEO)数据库(http://www.ncbi.nlm.nih.gov/)。选择GSE48452和GSE89632两个数据集。GSE48452基于GPL11532平台，包含14例NAFL患者及18例NASH患者的肝脏基因表达数据[9]。GSE89632基于平台GPL14951，包含20例NAFL患者及19例NASH患者[10]。采用Robust Multi-array Average (RMA)方法对两个数据集进行归一化处理。

1.2 共表达网络的构建

使用R中的WGCNA包进行共表达网络分析[11]。WGCNA的目的是鉴定基因模块和不同性状(NAFL，NASH)的相关性。基因共表达的相关矩阵由两个基因间的相关系数组成。通过基因间的Pearson系数构建层次聚类树，其中不同分支表示

不同的基因模块，不同颜色代表不同模块，将表达模式类似的基因置于同一模块中。基因模块与疾病性状相关的结果以热图的形式显示。选取与性状相关度最高的模块进行后续分析。

1.3 模块共同基因的筛选

对两个共表达网络中筛选出的模块基因取交集，重叠基因为目标基因，用R语言进行可视化输出。

1.4 目标基因的功能富集分析

基因本体(Gene Ontology，GO)富集分析为功能基因组学提供了一个全面的来源。GO注释可分为3类，包括分子功能(molecular function，MF)、生物过程(biological process，BP)和细胞成分(cellular component，CC)[12]。京都基因和基因组百科全书(Kyoto Encyclopedia of Genes and Genomes，KEGG)是一个集基因组、化学和系统功能信息于一体的综合性数据库，其中KEGG通路被专门用来存储不同物种之间的基因途径信息[13]。对目标基因进行GO和KEGG富集分析，以阐明目标基因的生物学功能。

1.5 PPI网络的构建

使用用于预测蛋白质-蛋白质相互作用(protein-protein interaction，PPI)的在线工具STRING (Search Tool for the Retrieval of Interaction Genes)来构建一个基因的PPI网络[14]。选择互相作用得分≥0.4的基因，用Cytoscape (v3.7.2)构建可视化的网络模型。

1.6 统计学处理

用R (3.6.0)软件进行统计分析及图像生成，$P<0.05$为差异有统计学意义。

2 结果

2.1 构建加权基因共表达网络

根据WGCNA的相应步骤建立基于层次聚类树的基因网络。设置模块最小基因数为50，在GSE48452中筛选出7个模块，GSE89632中筛选出6个模块，绘制基因聚类树状图(图1A、1B)。根据每个模块与两种临床性状(NAFL和NASH)之间的关联，绘制模块-特征关系热图(图2A、2B)。分析结果表明：黑色模块和灰色模块是与NASH显著相关的2个基因模块，其中黑色模块共包含65个基因，相关系数0.53 ($P<0.01$)；灰色模块包含1 247个基因，相关系数0.5 ($P<0.01$)。

2.2 目标基因的筛选

选取GSE48452数据集中的黑色模块与GSE89632数据集中的灰色模块，选取重叠的16个基因作为目标基因[凝血因子XIII A1多肽(coagulation factor XIII, A1 polypeptide，*F13A1*)、脂蛋白脂肪酶(lipoprotein lipase，*LPL*)、细胞色素P450家族26亚家族A成员1(cytochrome P450 family 26 subfamily A member 1，*CYP26A1*)、小脑肽3前体蛋白(cerebellin 3 precursor protein，*CBLN3*)、胸苷酸合成酶(thymidylate synthetase，*TYMS*)、脂肪酸结合蛋白4(fatty acid-binding protein 4，*FABP4*)、脂肪酸结合蛋白5(fatty acid-binding protein 5，*FABP5*)、X-脯氨酰氨肽酶2(X-prolyl aminopeptidase 2，*XPNPEP2*)、微染色体维持蛋白2(minichromosome maintenance protein 2，*MCM2*)、Rho家族GTP酶2(Rho family GTPase 2，*RND2*)、微染色体维持蛋白5(minichromosome maintenance protein 5，*MCM5*)、AAA ATP酶Midasin 1(AAA-ATPase midasin 1，*MDN1*)、GINS复合物亚单位2(GINS complex subunit 2，*GINS2*)、G蛋白偶联受体88(G-protein coupled receptor 88，*GPR88*)、白细胞介素20受体亚单位beta(interleukin-20 receptor subunit beta，*IL20RB*)、序列相似家族111成员B(family with sequence similarity 111 member B，*FAM111B*)]进行后续分析(图3)。

2.3 目标基因的功能富集

对16个目标基因进行GO及KEGG富集分析。GO显示目标基因的生物学功能主要集中在脂质降解、核染色体及有机酸的结合(图4A)。KEGG通路显示基因在过氧化物酶体增殖物激活受体(peroxisome proliferator-activated receptor，PPAR)通路富集尤为显著(图4B)。

2.4 PPI网络的构建

用String数据库构建目标基因的PPI网络，其中8个基因(*LPL*、*TYMS*、*CYP26A1*、*FABP4*、*FABP5*、*MCM2*、*MCM5*、*GINS2*)存在互相作用，网络中包含8个节点和10条边(图5)。

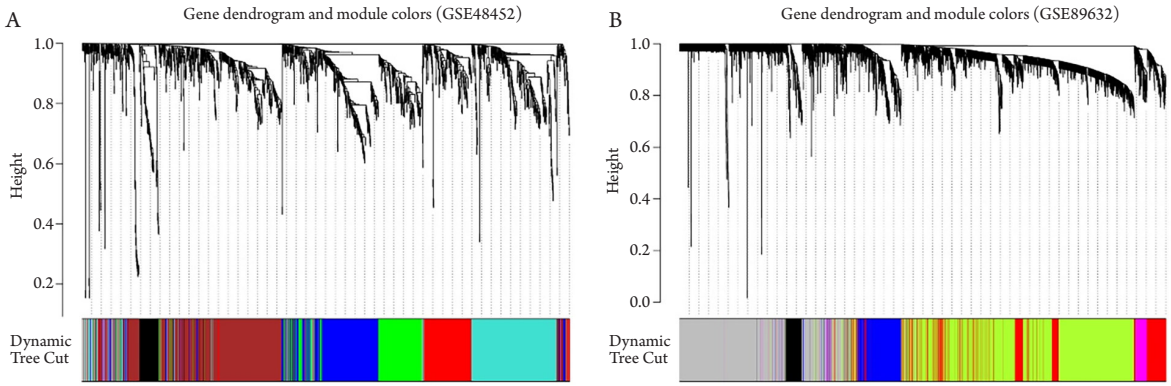

图1 聚类树状图显示功能相似的基因按照不同的度量划分为相同的模块

Figure 1 Clustering dendrogram shows that genes with similar functions are divided into the same modules according to different metrics

(A)GSE48452数据集；(B)GSE89632数据集。

(A) GSE48452 data set; (B) GSE89632 data set.

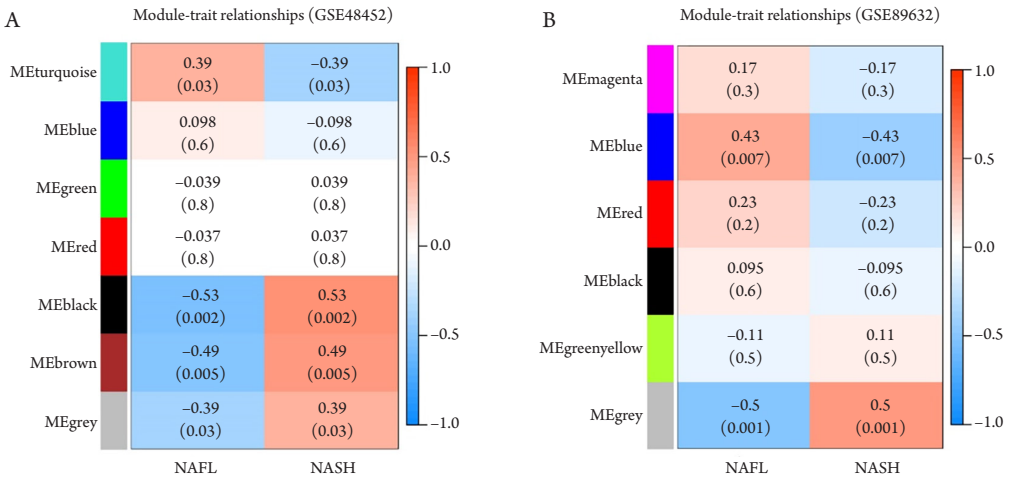

图2 基因模块和疾病相关性热图

Figure 2 Heat maps of gene modules correlated with disease

(A)GSE48452数据集；(B)GSE89632数据集。

(A) GSE48452 data set; (B) GSE89632 data set.

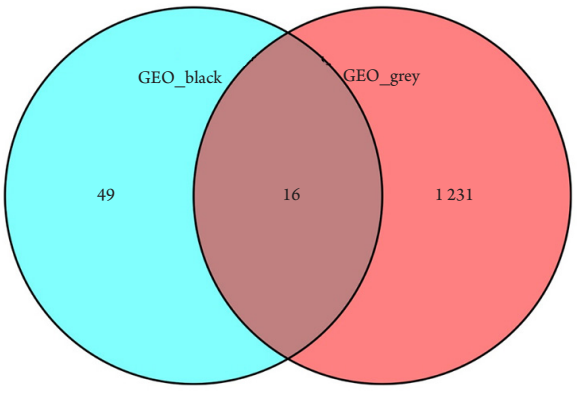

图3 共表达模块重叠基因的Venn图

Figure 3 Venn diagram of overlapping genes in co-expression modules

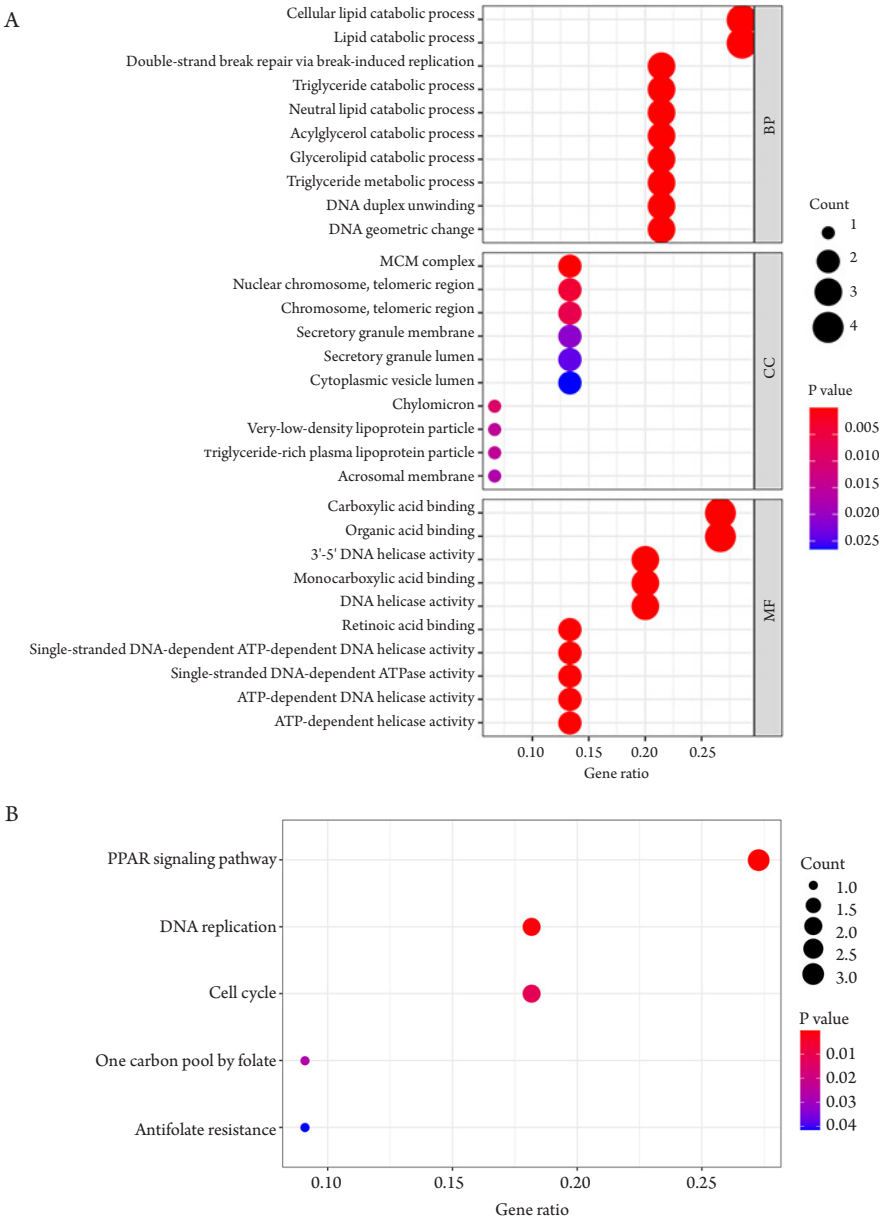

图4 目的基因的GO和KEGG富集分析

Figure 4 GO and KEGG enrichment analysis of the target gene

(A)GO富集分析；(B)KEGG富集分析。

(A) GO enrichment analysis; (B) KEGG enrichment analysis.

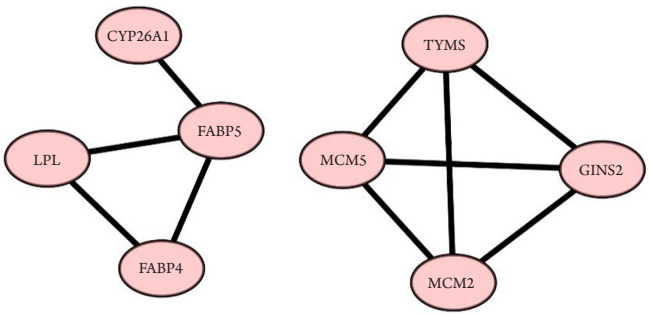

图5 PPI网络和枢纽基因的可视化

Figure 5 Visualization of PPI network and hub genes

3 讨论

在全球范围内，NAFLD是最常见的慢性肝病，病理过程为从良性肝脂肪变性到NASH、肝硬化，最终可能发展为肝细胞癌[15-16]。单纯肝脂肪变性向NASH转变，是重症肝病发展中最重要的一步[5]。NASH患者肝纤维化和终末期肝病发生风险较单纯性脂肪性肝病患者明显升高。因此，鉴定NASH的易感基因对研究NASH的发病原因和寻找潜在的治疗方法具有重要意义。

近年研究[17]表明：在探索性研究的框架内构建基因共表达网络是非常必要的，因为这种努力可以帮助识别与疾病相关的重要模块和基因。本研究采用WGCNA技术对NAFL和NASH的肝组织基因表达数据进行分析，发现黑色模块和灰色模块与NASH显著相关。对两个模块中的共同基因进行富集分析，发现NASH与脂质降解、核染色体、有机酸的结合及PPAR通路等生物学途径显著相关。有研究[18]报道了番茄红素可减轻肝脏过度脂质积聚，增强脂解，逆转NASH中的胰岛素抵抗，以及肝脏炎症和纤维化。Veskovi?等[19]报道称：补充甜菜碱可增加细胞核的均匀性和染色质的复杂性，通过减少脂肪积累和抑制肝细胞增殖来减轻小鼠的NAFLD。过量的脂肪酸摄取引起的氧化应激导致肝损伤和炎性免疫细胞的连续聚集，从而促进单纯脂肪变性向NASH的进展[18-20]。PPAR调节碳水化合物和脂质代谢，改善胰岛素敏感性、三酰甘油水平、炎症和氧化应激，已经成为整个生物体和细胞代谢功能的重要调节因子[21-23]。

PPI网络包含*LPL*、*TYMS*、*CYP26A1*、*FABP4*、*FABP5*、*MCM2*、*MCM5*、*GINS2* 8个基因。泽泻醇B醋酸酯(Alisol B 23-acetate，AB23A)通过诱导PPAR、CPT1、LPL等增加脂质代谢，对小鼠产生抗NASH的保护作用[24]。TYMS是叶酸代谢的关键限速酶，有研究[25-26]证明叶酸可改善肝脏脂肪积累及肝脏炎症。CYP26A1的高表达表明全反式维甲酸(all-trans retinoic acid，ATRA)在NAFLD中的降解增强，可能是NAFLD发展为包括肝硬化在内的慢性肝病的可能机制[27]。Milner等[28]和Coilly等[29]的研究报道了脂肪酸结合蛋白(fatty acid binding protein，FABP)水平可区分脂肪性肝炎和单纯脂肪变性[28-29]。MCM与DNA复制过程密切相关，线粒体复制和转录功能障碍在NASH发生发展中起重要作用[30-31]。近年来的研究[32]指出：GINS2的表达可以作为酒精所致神经炎症的生物标志物，并有助于识别具有炎症能力的小胶质细胞。

本研究在NASH方面提供了新的见解，但其潜在的关键途径和基因都是基于生物信息学方法确定的。因此，目前还不清楚这些基因是因果关系，还是仅仅是NASH的标记，需要在分子实验中加以验证。

参考文献

1. Vernon G, Baranova A, Younossi ZM. Systematic review: The epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults[J]. Aliment Pharmacol Ther, 2011, 34(3): 274-285.
2. Younossi ZM, Koenig AB, Abdelatif D, et al. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes[J]. Hepatology, 2016, 64(1): 73-84.
3. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of non-alcoholic fatty liver disease: Practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association[J]. Hepatology, 2012, 55(6): 2005-2023.
4. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease[J]. J Hepatol, 2016, 64(6): 1388-1402.
5. Goldberg D, Ditah IC, Saeian K, et al. Changes in the prevalence of hepatitis C virus infection, nonalcoholic steatohepatitis, and alcoholic liver disease among patients with cirrhosis or liver failure on the waitlist for liver transplantation[J]. Gastroenterology, 2017, 152(5): 1090-1099.e1.
6. Buzzetti E, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD)[J]. Metabolism, 2016, 65(8): 1038-1048.
7. Sumida Y, Yoneda M. Current and future pharmacological therapies for NAFLD/NASH[J]. J Gastroenterol, 2018, 53(3): 362-376.
8. Jia R, Zhao H, Jia M. Identification of co-expression modules and potential biomarkers of breast cancer by WGCNA[J]. Gene, 2020, 750: 144757.
9. Morrison MC, Kleemann R, Van Koppen A, et al. Key inflammatory processes in human NASH are reflected in Ldlr(-/-).Leiden mice: A translational gene profiling study[J]. Front Physiol, 2018, 9: 132.
10. Arendt BM, Comelli EM, Ma DW, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids[J]. Hepatology, 2015, 61(5): 1565-1578.
11. Langfelder P, Horvath S. WGCNA: An R package for weighted

correlation network analysis[J]. BMC Bioinformatics, 2008, 9: 559.
12. Gene Ontology Consortium. Gene Ontology Consortium: going forward[J]. Nucleic Acids Res, 2015, 43(Database issue): D1049-D1056.
13. Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation[J]. Nucleic Acids Res, 2016, 44(D1): D457-462.
14. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets[J]. Nucleic Acids Res, 2019, 47(D1): D607-D613.
15. Ratziu V, Goodman Z, Sanyal A. Current efforts and trends in the treatment of NASH[J]. J Hepatol, 2015, 62(1 Suppl): S65-S75.
16. Abd El-Kader SM, El-Den Ashmawy EM. Non-alcoholic fatty liver disease: The diagnosis and management[J]. World J Hepatol, 2015, 7(6): 846-858.
17. Zhao W, Langfelder P, Fuller T, et al. Weighted gene coexpression network analysis: State of the art[J]. J Biopharm Stat, 2010, 20(2): 281-300.
18. Ni Y, Zhuge F, Nagashimada M, et al. Lycopene prevents the progression of lipotoxicity-induced nonalcoholic steatohepatitis by decreasing oxidative stress in mice[J]. Free Radic Biol Med, 2020, 152: 571-582.
19. Vesković M, Labudović-Borović M, Zaletel I, et al. The effects of betaine on the nuclear fractal dimension, chromatin texture, and proliferative activity in hepatocytes in mouse model of nonalcoholic fatty liver disease[J]. Microsc Microanal, 2018, 24(2): 132-138.
20. Dasarathy S, Yang Y, Mccullough AJ, et al. Elevated hepatic fatty acid oxidation, high plasma fibroblast growth factor 21, and fasting bile acids in nonalcoholic steatohepatitis[J]. Eur J Gastroenterol Hepatol, 2011, 23(5): 382-388.
21. Gutiérrez-Cuevas J, Sandoval-Rodríguez A, Monroy-Ramírez HC, et al. Prolonged-release pirfenidone prevents obesity-induced cardiac steatosis and fibrosis in a mouse NASH model[J]. Cardiovasc Drugs Ther, 2021, 35(5): 927-938.
22. Ndisang JF. Cross-talk between heme oxygenase and peroxisome proliferator-activated receptors in the regulation of physiological functions[J]. Front Biosci (Landmark Ed), 2014, 19: 916-935.
23. Han L, Shen WJ, Bittner S, et al. PPARs: Regulators of metabolism and as therapeutic targets in cardiovascular disease. Part I: PPAR-α[J]. Future Cardiol, 2017, 13(3): 259-278.
24. Meng Q, Duan XP, Wang CY, et al. Alisol B 23-acetate protects against non-alcoholic steatohepatitis in mice via farnesoid X receptor activation[J]. Acta Pharmacol Sin, 2017, 38(1): 69-79.
25. Wang X, Sun X, Du X, et al. Thymidylate synthase gene polymorphisms as important contributors affecting hepatocellular carcinoma prognosis[J]. Clin Res Hepatol Gastroenterol, 2017, 41(3): 319-326.
26. Sid V, Shang Y, Siow YL, et al. Folic acid supplementation attenuates chronic hepatic inflammation in high-fat diet fed mice[J]. Lipids, 2018, 53(7): 709-716.
27. Ashla AA, Hoshikawa Y, Tsuchiya H, et al. Genetic analysis of expression profile involved in retinoid metabolism in non-alcoholic fatty liver disease[J]. Hepatol Res, 2010, 40(6): 594-604.
28. Milner KL, Van Der Poorten D, Xu A, et al. Adipocyte fatty acid binding protein levels relate to inflammation and fibrosis in nonalcoholic fatty liver disease[J]. Hepatology, 2009, 49(6): 1926-1934.
29. Coilly A, Desterke C, Guettier C, et al. FABP4 and MMP9 levels identified as predictive factors for poor prognosis in patients with nonalcoholic fatty liver using data mining approaches and gene expression analysis[J]. Sci Rep, 2019, 9(1): 19785.
30. Li Z, Xu X. Post-translational modifications of the mini-chromosome maintenance proteins in DNA replication[J]. Genes (Basel), 2019, 10(5): 331.
31. Hasturk B, Yilmaz Y, Eren F. Potential clinical variants detected in mitochondrial DNA D-loop hypervariable region I of patients with non-alcoholic steatohepatitis[J]. Hormones (Athens), 2019, 18(4): 463-475.
32. Liu C, Wang R, Zhang Y. GINS complex subunit 2 (GINS2) plays a protective role in alcohol-induced brain injury[J]. Artif Cells Nanomed Biotechnol, 2019, 47(1): 1-9.

本文引用： 陈懿, 严丹, 李雪, 张友才. 非酒精性脂肪性肝炎关键基因的鉴定及通路分析[J]. 临床与病理杂志, 2021, 41(12): 2778-2784. doi: 10.3978/j.issn.2095-6959.2021.12.005

Cite this article as: CHEN Yi, YAN Dan, LI Xue, ZHANG Youcai. Identification and pathway analysis of key genes in nonalcoholic steatohepatitis[J]. Journal of Clinical and Pathological Research, 2021, 41(12): 2778-2784. doi: 10.3978/j.issn.2095-6959.2021.12.005