基于生物信息学和机器学习算法的非酒精性 脂肪性肝炎细胞衰老关键基因的分析与验证

朱 月,毋中明

山东第一医科大学附属省立医院内分泌科,山东 济南

收稿日期: 2025年5月11日; 录用日期: 2025年6月5日; 发布日期: 2025年6月13日

摘要

目的:非酒精性脂肪性肝炎(Non-alcoholic Steatohepatitis, NASH)是发病机制涉及多重病理生理过程, 其中细胞衰老对NASH的发生发展起着关键作用。基于此,针对细胞衰老相关基因及其调控网络的深入研 究,可能为NASH的靶向治疗提供新的干预策略。方法:从GEO数据库获取NASH相关转录组测序数据集 GSE89632和GSE37031。采用WGCNA筛选出与NASH高度相关的基因模块,对其进行富集分析,以阐明 其潜在生物学功能。将差异表达基因(Differentially expressed genes, DEGs)与CellAge数据库中的细胞 衰老基因集取交集,筛选出细胞衰老DEGs。对细胞衰老DEGs进行功能富集分析,系统揭示其在NASH中 的分子调控网络。随后,整合机器学习算法筛选Hub基因。最后,构建ROC曲线并计算AUC评估Hub基因 的诊断效能。结果:WGCNA分析表明,共有10个基因模块被识别,其中黑色模块基因与NASH呈现最显 著相关。富集分析显示,关键模块的基因与细胞衰老和免疫炎症密切相关。对训练基因集进行差异分析, 共获得细胞衰老相关DEGs共51个。进一步GO和KEGG分析显示,细胞衰老DEGs主要与衰老和免疫炎症 相关。对细胞衰老DEGs进行三种机器学习算法,综合对比后,FOS、MYC和PIM1被确定为关键基因。通 过绘制ROC曲线检验在训练集和测试集中的诊断效能。关键基因的ROC曲线下面积均大于0.7且这三个基 因较对照组均显著低表达。结论:本研究通过整合生物信息学分析和机器学习算法,最终鉴定出FOS、 MYC和PIM1三个关键调控基因,均在NASH患者中显著低表达,且与免疫炎症反应密切相关。

关键词

非酒精性脂肪性肝炎,细胞衰老,生物信息学,机器学习算法

Analysis and Validation of Key Cellular Senescence Genes in Non-Alcoholic Steatohepatitis Based on Bioinformatics and Machine Learning Algorithm

Yue Zhu, Zhongming Wu

Endocrinology Department, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan Shandong

Received: May 11th, 2025; accepted: Jun. 5th, 2025; published: Jun. 13th, 2025

Abstract

Objective: Non-alcoholic steatohepatitis (NASH) is a pathogenesis involving multiple pathophysiological processes, in which cellular senescence plays a key role in the development of NASH. Based on this, in-depth studies of cellular senescence-related genes and their regulatory networks may provide new intervention strategies for targeted therapy of NASH. Methods: NASH-related transcriptome sequencing datasets GSE89632 and GSE37031 were obtained from the GEO database, and WGCNA was used to screen out the gene modules that were highly related to NASH, and they were enriched and analyzed to elucidate their potential biological functions. Differentially expressed genes (DEGs) were intersected with the cellular senescence gene sets in the CellAge database to screen the cellular senescence DEGs. Functional enrichment analysis was performed on the cellular senescence DEGs to systematically reveal their molecular regulatory networks in NASH. Subsequently, machine learning algorithms were integrated to screen Hub genes. Finally, ROC curves were constructed and AUC was calculated to evaluate the diagnostic efficacy of Hub genes. Results: WGCNA analysis showed that a total of 10 gene modules were identified, with the black module genes showing the most significant association with NASH. Enrichment analysis showed that the key module genes were closely associated with cellular senescence and immune inflammation. Differential analysis of the training gene set yielded a total of 51 cellular senescence-related DEGs. Further GO and KEGG analysis showed that cellular senescence DEGs were mainly associated with senescence and immune inflammation. Three machine learning algorithms were performed on the cellular senescence DEGs, and FOS, MYC and PIM1 were identified as key genes after comprehensive comparison. Diagnostic efficacy in the training and test sets was examined by plotting ROC curves. The area under the ROC curve for the key genes was greater than 0.7 and all three genes were significantly under expressed compared to the control. Conclusion: By integrating bioinformatics analysis and machine learning algorithms, the present study finally identified three key regulatory genes, FOS, MYC and PIM1, which were significantly under expressed in NASH patients and closely associated with immune-inflammatory responses.

Keywords

Non-Alcoholic Steatohepatitis, Cellular Senescence, Bioinformatics, Machine Learning Algorithm

Copyright © 2025 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

1. 引言

非酒精性脂肪性肝病(Non-alcoholic Fatty Liver Disease, NAFLD)是以肝内脂肪异常堆积和肝细胞脂肪 变性为特征的一系列疾病谱[1]。非酒精性脂肪性肝炎(Non-alcoholic Steatohepatitis, NASH)以肝细胞脂肪 变性、肝小叶炎症和进行性纤维化为特征,是非酒精性脂肪肝向肝硬化甚至肝细胞癌(Hepatocellular Carcinoma, HCC)转变的关键环节[2]。此外, NASH 还是多系统疾病的重要危险因素,这进一步加剧了 NASH 对全球健康负担的影响[3] [4]。NASH 的发病机制涉及多种病理生理过程,包括脂质代谢异常、氧化应激、

免疫炎症改变、肠道菌群失衡和遗传变异等[5]。目前 NASH 治疗手段仍十分有限,生活方式的改变仍是 NASH 治疗的基石[6]。

目前认为,NASH 患者肝细胞中普遍存在衰老表型。NASH 中的脂肪酸代谢紊乱和线粒体功能障碍 可导致活性氧过量产生与积累,从而加重氧化应激损伤并引起肝细胞衰老的发生[7],而肝脏衰老细胞的 积累又促进肝脏内脂质沉积和肝脂肪变性[8],形成了恶性循环。此外,肝细胞衰老通常伴随着衰老相关 分泌表型分泌,一方面,这些因子可以引发特定区域的炎症应答反应、介导免疫效应细胞的定向迁移与 聚集,实现对衰老细胞的识别与清除;另一方面还能通过细胞间旁分泌信号传导机制促使周边正常细胞 衰老,从而使组织微环境发生重构[9][10]。值得关注的是,细胞衰老还能够调节免疫微环境,从而影响 多种疾病病理机制的进展[11]。例如,肝脏衰老会导致巨噬细胞等固有免疫细胞的功能失调,进而驱动肝 脏慢性炎症、纤维化和 HCC 的进展[12]。因此靶向免疫细胞的调控机制,可能为治疗衰老相关肝脏疾病 提供新的策略。

生物信息学作为一种新兴且有效的工具,能够整合各种疾病的多组学数据,系统揭示疾病相关的分子网络和信号通路,提高了疾病诊断、预防和治疗的能力[13][14]。机器学习算法则可通过特征选择和模式识别,从海量数据中提取关键生物标志物[15]。这些技术的应用为 NASH 研究提供了新的思路和方法。

基于以上背景,本研究旨在整合生物信息学和机器学习算法识别 NASH 诊断的潜在基因,为 NASH 的精准治疗提供新的理论依据和潜在靶点。

2. 材料与方法

2.1. 数据的获取和预处理

以"Nonalcoholic steatohepatitis"为关键词,从GEO数据库中数据集,符合纳入标准的数据集包括: (1)实验分组包括 NASH 患者和健康对照组; (2)样本量至少为 10 人,每组至少 5 人。最终筛选出 2 个符合以上条件的数据集 GSE89632 (24 个健康对照组,19 个 NASH)和 GSE37031 (7 个健康对照组,8 个 NASH)。然后,从 CellAge 数据库下载了 279 个细胞衰老相关基因。将从 GEO 数据库下载的原始数据用 R 软件的"limma"包进行背景校正和归一化处理以去除样本间批次效应。

2.2. 加权基因共表达网络分析

加权基因共表达网络分析(Weighted Gene Co-Expression Network Analysis, WGCNA)是一种基于高通 量转录组数据构建基因关联网络的生物学方法[16],可以识别功能相关的基因模块,从而揭示潜在的生物 学机制[17]。根据无标度拓扑准则,我们使用 R 软件中的"WGCNA"包在 GSE89632 队列中构建了共表 达网络,使用 Pearson 算法计算基因模块和性状之间的相关系数,相关系数最高的模块用于下一步研究。

2.3. 关键基因模块的功能富集分析

基因本体论(Gene Ontology, GO)是用于生物信息学研究者描述基因及其产物功能的标准分类系统,涵盖生物学过程(Biological Process, BP)、细胞成分(Cellular Component, CC)和分子功能(Molecular Function, MF)三方面[18]。京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)通路分析则用于揭示这些基因在代谢和信号传导通路中的作用,提供更广泛的基因组背景信息[19]。为了深入探索关键基因模块的功能和通路,本研究使用 R 的 "clusterProfiler"包对该模块进行 GO 和 KEGG 富集分析。

2.4. 细胞衰老差异表达基因的筛选

使用 R 的"limma"包筛选训练集中 NASH 组和健康对照组之间的差异表达基因(DEGs),筛选标准

为: *P* < 0.05, |log2 fold change (FC)| > 0.5。利用韦恩图将 DEGs 与从 CellAge 数据库中下载的细胞衰老 基因取交集。为了将这些 DEGs 可视化,使用 R 软件的 "ggplot2"包生成火山图,使用 "pheatmap"包 生成热图。

2.5. 细胞衰老差异表达基因的功能富集分析

为了深入探索细胞衰老差异表达基因的功能和通路,本研究使用 R 的 "clusterProfiler"包对细胞衰 老 DEGs 进行 GO 和 KEGG 通路富集分析,并以 P < 0.05 作为统计显著性标准,最终进行可视化分析。

2.6. 机器学习算法筛选 NASH 中细胞衰老差异表达基因的关键基因

分别应用 R 软件中的 "glmnet"、 "randomForest"和 "e1071"包执行最小绝对收缩和选择算子 (Least Absolute Shrinkage and Selection Operator, LASSO)逻辑回归、随机森林(Random Forest, RF)算法及 支持向量机递归特征消除(Support Vector Machine-Recursive Feature Elimination, SVM-RFE)这三种算法, 来筛选细胞衰老 DEGs 中差异有统计学意义的基因。使用 "Venn"包对三种机器学习分析结果进行可视 化,取三组结果的交集,视为 NASH 细胞衰老 DEGs 的 Hub 基因。

2.7. Hub 基因的诊断价值评估

使用 R 软件中的 "ggplot2" 包绘制 Hub 基因的箱线图,评估训练集和测试集中 NASH 和健康对照 样本的 Hub 的基因表达水平, *P*<0.05 表示有显著差异。使用 "pROC" 包绘制受试者工作特征(Receiver operating characteristic, ROC)曲线,并计算曲线下面积(Area under the curve, AUC)和 95%可信区间来量化 其诊断效能,AUC 数值越接近于 1,表明 Hub 具有更高的准确性和敏感性。因此,差异最显著且 AUC> 0.7 的 Hub 基因为理想诊断基因。

2.8. NASH 的免疫浸润景观分析及其与关键基因的相关性

CIBERSORT 算法是一种基于生物信息学原理的计算工具,能够精确估算特定细胞亚群在复杂组织 微环境中的相对丰度分布[20]。本研究采用 CIBERSORT 算法评估免疫细胞的浸润情况,并计算浸润免疫 细胞之间的相关性。使用 R 软件的 "vioplot"包绘制条形图和小提琴图,对 NASH 组和对照组之间的免 疫细胞浸润差异进行可视化;相关热图用 R 软件的 "corrplot"包绘制。

3. 结果

3.1. 加权基因共表达网络分析及关键模块的识别

利用 R 软件中的 "pickSoftThreshold"函数来计算软阈值,当无尺度拓扑拟合指数 R²为 0.9 时, GSE89632数据集的最佳软阈值 β 为 13 (图 1(a))。根据确定的最佳软阈值,进行层次聚类建树,去除离群 样本(图 1(b)),然后通过动态切割树法切割基因模块,合并高度类似的模块,最终得到 10 个基因模块, 每种颜色代表一个基因模块(图 1(c))。最后,将 10 个基因模块与 NASH 进行 Pearson 相关性分析(图 1(d))。 结果表明,黑色基因模块(MEblack)与 NASH 之间的相关性最高(r=-0.94, P=2e-21),这表明其内所包含 的基因可能与 NASH 的发病机制显著相关。

3.2. NASH 关键基因模块的富集分析

GO 分析显示,黑色模块中的基因所参与的 BP 主要有:脂肪细胞分化、DNA 结合转录因子活性的 调节和细胞对白细胞介素-1 的反应等(图 2(a));MF 主要富集在:DNA 结合转录因子的结合、RNA 聚合酶 II 特异性 DNA 结合转录因子结合等(图 2(b));所参与的 CC 主要为:转录调节因子复合物、转录抑制



Figure 1. Weighted gene co-expression network analysis and identification of key modules 图 1. 加权基因共表达网络分析及关键模块的识别

因子复合物等(图 2(c))。KEGG 分析显示,富集程度最高的通路包括癌症中的转录失调、TNF 信号通路、 IL-17 信号通路以及 p53 信号通路(图 2(d))。此外,在 KEGG 分析可视化的前十个结果中,我们还发现了 细胞衰老、糖尿病并发症中的 AGE-RAGE 信号通路等与细胞衰老和炎症密切相关的通路。

2.3. 细胞衰老差异表达基因的筛选

使用 R 软件中的"limma"包分析训练集,得到 NASH 组与正常对照组之间的差异表达基因共 2752 个。随后,将这些差异表达基因与 CellAge 数据库中收录的 279 个细胞衰老相关基因通过韦恩图相交后,确定了 51 个细胞衰老 DEGs (图 3(a))。其中上调基因 25 个,下调基因 26 个(图 3(b))。使用"pheatmap" 包将筛选到的细胞衰老 DEGs 中的前 50 个基因制作聚类热图(图 3(c))。此外,我们还使用 Spearman 相关 性检验评估了这些细胞衰老 DEGs 之间的相关性(图 3(d))。

2.4. 细胞衰老差异表达基因的富集分析

将上述 51 个细胞衰老 DEGs 作为一个目标基因集,使用 R 的"clusterProfiler"包进行 GO 和 KEGG 富集分析,以确定涉及他们的的生物学功能及相关的信号通路,以 P < 0.05 作为筛选标准,可视化富集 分析的结果。GO 分析表明,细胞衰老 DEGs 主要在有丝分裂细胞周期的调节、细胞周期的负调控和有丝



 Figure 2. Enrichment analysis of key gene modules

 图 2. 关键基因模块的富集分析



Figure 3. Screening of differentially expressed genes in cellular senescence 图 3. 细胞衰老差异表达基因的筛选

分裂细胞周期相变等 BP 中显著富集(图 4(a)); MF 主要富集在:蛋白质丝氨酸/苏氨酸/酪氨酸激酶活性、 蛋白质丝氨酸激酶活性、蛋白质丝氨酸/苏氨酸激酶活性、甲基化组蛋白结合和甲基化依赖性蛋白结合等 (图 4(a)); CC 主要富集在:转录调节因子复合物、有丝分裂纺锤体、核染色质、Sin3 复合物和 Sin3 型复 合物等(图 4(a))。

KEGG 分析显示(图 4(b)),这些细胞衰老 DEGs 主要富集到细胞周期、人 T 细胞白血病病毒 1 感染、 p53 信号通路和细胞衰老等通路。此外,我们还发现了这些差异基因在免疫炎症相关通路如 IL-17 信号通路和 TNF 信号通路等富集。



Figure 4. Enrichment analysis of differentially expressed genes in cellular senescence 图 4. 细胞衰老差异表达基因的富集分析

3.5. 机器学习算法识别 Hub 基因

使用"glmnet"包执行 LASSO 回归,共筛选出 7 个候选基因(图 5(a))。使用"randomForest"包执行随机森林算法,最终筛选出 30 个候选基因(图 5(b)、图 5(c))。使用"e1071"包执行 SVM-RFE 算法筛选 出 8 个候选基因(图 5(d)、图 5(e))。将三者的筛选结果通过韦恩图取交集并进行可视化,分别获得 MYC、PIM1、MAD2L1、FOS 和 ZFP36⑤个基因。最终,我们综合对比了三种算法的结果,发现 FOS、MYC 和 PIM1 这三个基因在所有算法中均位于前五名,因此被确定为关键基因(图 5(f))。



Figure 5. Machine learning algorithms identify Hub genes 图 5. 机器学习算法识别 Hub 基因

3.6. 关键基因的表达水平及诊断价值评估

使用 R 软件中的 "ggplot2"包绘制箱线图评估机器学习算法分析得到的三个关键基因在 GSE89632 和 GSE37031 的表达水平。结果显示,在训练数据集 GSE89632 中(图 6(a)~(c)),相较于健康对照组,NASH 患者中 FOS、MYC 和 PIM1 的基因表达水平均显著下降(P<0.05)。在测试数据集 GSE37031 中,我们同 样观察到 NASH 患者中 FOS 和 MYC 的表达水平显著低于健康对照组(P<0.05,图 6(g)、图 6(h)),虽然 健康对照组和非酒精性脂肪性肝炎样本中 PIM1 的表达无显著差异(P = 0.12,图 6(i)),但与对照组相比, NASH 患者中 PIM1 的表达仍有下降趋势。为了验证这 3 个 Hub 基因对 NASH 的诊断价值,我们构建了 ROC 曲线并计算 AUC,AUC 值大于 0.7 的 Hub 基因被认为具有较好的诊断价值。结果显示,训练集 GSE89632 中三个关键基因的 AUC 值为 1.000 (图 6(d),图 6(f)),表明每个 Hub 基因对 NASH 的预测均 具有良好的诊断价值。此外,我们还使用测试集 GSE37031 进行了外部验证。通过分析该数据集,我们 观察到三个关键基因的 AUC 值在 0.714 至 1.000 之间变化(图 6(j)~(1)),进一步证实了这些基因出色的诊断性能。



Figure 6. The expression level and diagnostic value of key genes were evaluated 图 6. 关键基因的表达水平及诊断价值评估

3.7. NASH 中的免疫细胞浸润及其与关键基因的相关性

条形图展示了各种免疫细胞在 NASH 和健康对照组中的分布情况(图 7(a))。该图直观的展示出 NASH 组和健康对照组的免疫细胞存在差异。小提琴图(图 7(b))显示, NASH 和健康对照组之间有 9 种免疫细胞 浸润丰度存在显著差异(P < 0.05)。与健康对照组相比, NASH 中 CD8 T 细胞、γδT 细胞、M2 巨噬细胞、静息肥大细胞的比例显著增高, 而初始 B 细胞、单核细胞、活化树突状细胞、活化肥大细胞和中性粒细 胞的比例显著降低。

朱月, 毋中明



Figure 7. Immune cell infiltration in NASH 图 7. NASH 中的免疫细胞浸润

4. 讨论

NASH 是全球最常见的慢性肝病,其发病率随着肥胖和代谢综合征的流行而逐年升高,给全球医疗体系带来了沉重负担[21][22]。NASH 的病理生理机制涉及多重因素复杂交互作用,其精准的分子调控网络尚未完全阐明,且 NASH 的治疗方法也较为局限,生活方式的改变及外科手术仍是 NASH 治疗的主要手段[23]。在此背景下,探索 NASH 早期诊断的生物标志物及创新治疗靶点具有重大的临床意义。

首先,我们基于 GEO 数据库系统筛选出 GSE89632 及 GSE37031 两个转录组数据集。通过 WGCNA 分析筛选出 NASH 的关键基因模块,随后,对关键基因模块进行富集分析,以揭示 NASH 潜在的发病机制。GO 富集分析显示,关键基因模块主要富集在转录调控、炎症反应以及脂肪细胞分化等过程。这与先前的研究结果一致,提示关键基因模块通过代谢失调、炎症激活和表观遗传调控等途径参与 NASH 的发病机制[24][25]。KEGG 分析表明,这些基因通过免疫炎症改变、细胞周期失调和细胞衰老共同驱动 NASH 的进展。Zhang 等[26]证实,高脂饮食(HFD)喂养的 NAFLD 大鼠模型呈现出衰老标志物 p21、p16 的 mRNA 表达水平上调。临床研究数据同样支持细胞衰老与 NASH/NAFLD 的密切关联。Laish 等[27]发现,NAFLD 患者外周血淋巴细胞的端粒显著缩短,同时伴有端粒酶逆转录酶 mRNA 的表达上调,这一发现为细胞衰 老在 NAFLD 中的作用提供了有力证据。

尽管细胞衰老与 NASH 的关联已得到初步探索,但针对二者的特征基因的研究仍较为有限。基于此,本研究通过细胞衰老 DEGs 并对其进行富集分析,探究细胞衰老 DEGs 在 NASH 发病机制中的潜在作用。 GO 结果表明细胞衰老可能通过细胞周期调控、转录因子的信号传导、代谢调节等方面促进 NASH 的发 生发展[28]。KEGG 分析显示,除细胞周期、p53 信号通路和细胞衰老外,DEGs 还显著富集在 IL-17 信 号通路和 TNF 信号通路。研究表明,IL-17 和 TNF 促进炎症反应和免疫微环境重塑[29] [30],提示细胞 衰老可能通过调控代谢、免疫和炎症反应参与 NASH 的病理进程。这项发现为我们后续的免疫浸润分析 提供了理论依据。

随后,我们通过 LASSO 回归、随机森林和 SVM-RFE 三种机器学习算法联合筛选,确定了三个关键 基因: FOS、MYC 和 PIM1。这些基因在数据集 GSE89632 和外部数据集 GSE37031 构建 ROC 曲线验证 三个关键基因具有良好的诊断价值。

FOS 是转录因子家族成员,由 c-Fos、FosB、Fra-1/Fosl1 和 Fra-2/Fosl2 组成,在细胞增殖、分化和转 化中发挥关键调控作用。研究表明,其表达水平与炎症和肿瘤的发生发展密切相关[31]-[33]。在 NASH 研 究领域,动物模型实验证实,AP-1 是代谢、转运、信号转导和氧化还原相关的基因变化的关键转录调节 因子[34]。NASH 患者中 NF-κB 与 AP-1 的 DNA 结合活性显著增强,这一现象促进氧化应激和胰岛素抵 抗等病理过程[35]。此外,FOS 家族与衰老过程密切相关,Fra-1 通过激活 p21 和 p16 信号通路在衰老过 程中发挥着关键作用,提示干预 Fra-1 可能成为预防衰老相关疾病的潜在策略[36]。MYC 作为重要的原 癌基因,在调节新陈代谢和 RNA 剪切等过程发挥关键调控作用[37]。然而,目前关于该基因在肝脏代谢 疾病中的研究相对有限。现有证据表明,MYC 表达不足可能导致电子传递链功能异常和/或对脂肪酸 β 氧化的依赖增加,从而促进 ROS 产生,加速衰老进程[38]。考虑到 ROS 积累和细胞衰老是 NASH 发病 的重要环节,本研究推测 MYC 可能通过调控细胞衰老相关过程参与 NASH 的病理机制,这一发现为探 索 NASH 的分子机制提供新思路,但具体途径仍需深入探讨。

在 NASH 的病理过程中,慢性炎症反应以及免疫系统的异常激活发挥着重要作用[39]。为了深入探 讨免疫细胞在疾病中的角色,本研究对免疫细胞浸润进行了全面分析。结果显示,M2 巨噬细胞、静止肥 大细胞、CD8 T 细胞和 yðT 细胞是 NASH 中浸润程度最高的四种免疫细胞。巨噬细胞是免疫系统的重要 组分,其在 NASH 发病机制中的作用己得到广泛证实,这主要与其独特的分化特征及在疾病组织病理学 中的显著地位有关[40]。Rensen 等研究发现,M2 巨噬细胞标志物在 NASH 中呈现高表达特征,提示这类 巨噬细胞可能通过促进组织修复而参与肝细胞损伤后的肝脏重塑,但同时也可能增加肝纤维化的风险[41]。此外,CD8 T 细胞在 NASH 的发展过程中被特异性激活,活化的 CD8 T 细胞与自然杀伤 T 细胞直接与肝 细胞相互作用,激活核因子 NF-kB 信号通路,促进脂肪变性,并推动疾病向 HCC 恶性转化[42]。这些研 究证据充分支持了本研究中免疫细胞浸润分析结果的可靠性。这些发现为未来 NASH 免疫治疗提供了新 的思路,也为深入探索疾病机制奠定了重要基础。

5. 结论

本研究基于生物信息学和机器学习算法筛选出 FOS、MYC 和 PIM1 基因。内部及外部验证表明, FOS、MYC 和 PIM1 在 NASH 组中的表达水平显著降低,表明三者在 NASH 的病理过程中可能发挥重 要调控作用。进一步分析发现,M2 巨噬细胞、静止肥大细胞、yoT 细胞和 CD8 T 细胞与 NASH 的病理 过程呈现显著相关性,提示免疫微环境在 NASH 进展中的关键作用。这些研究结果为进一步探索 NASH 的致病机理奠定了理论依据,同时也为开发 NASH 的靶向治疗提供了潜在的分子靶点。

然而,本研究也有一定的局限性: (1) 本研究的转录组数据均来自公共数据库,无法完全排除患者群体和临床特征的差异性对本研究的潜在影响。(2) 用于分析和验证的样本量相对有限,可能影响结果的准确性,未来的研究需要扩大样本量。(3) 仅依靠基因表达水平的变化来推测基因的功能和作用机制不够严谨,需要进一步的实验验证来确认,且关键基因的作用机制缺乏深入的探讨。(4) 本研究仅分析了免疫细胞的浸润情况,未来仍需对免疫细胞功能进行深入探究。

参考文献

[1] Wang, T., Wang, R., Bu, Z., Targher, G., Byrne, C.D., Sun, D., et al. (2022) Association of Metabolic Dysfunction-

Associated Fatty Liver Disease with Kidney Disease. *Nature Reviews Nephrology*, **18**, 259-268. <u>https://doi.org/10.1038/s41581-021-00519-y</u>

- [2] Younossi, Z.M., Golabi, P., Paik, J.M., Henry, A., Van Dongen, C. and Henry, L. (2023) The Global Epidemiology of Nonalcoholic Fatty Liver Disease (NAFLD) and Nonalcoholic Steatohepatitis (NASH): A Systematic Review. *Hepatol*ogy, 77, 1335-1347. <u>https://doi.org/10.1097/hep.00000000000004</u>
- [3] Deprince, A., Haas, J.T. and Staels, B. (2020) Dysregulated Lipid Metabolism Links NAFLD to Cardiovascular Disease. *Molecular Metabolism*, **42**, Article ID: 101092.
- [4] Younossi, Z., Anstee, Q.M., Marietti, M., *et al.* (2018) Global Burden of NAFLD and NASH: Trends, Predictions, Risk Factors and Prevention. *Nature Reviews Gastroenterology & Hepatology*, **15**, 11-20.
- [5] 姜珊. 基于转录组和代谢组探讨非酒精性脂肪肝炎动态发病机制及转化熊胆粉干预机制研究[D]: [博士学位论 文]. 北京: 中国中医科学院, 2024.
- [6] Zeng, J., Fan, J. and Francque, S.M. (2024) Therapeutic Management of Metabolic Dysfunction Associated Steatotic Liver Disease. United European Gastroenterology Journal, 12, 177-186. <u>https://doi.org/10.1002/ueg2.12525</u>
- [7] Minamino, T., Orimo, M., Shimizu, I., Kunieda, T., Yokoyama, M., Ito, T., et al. (2009) A Crucial Role for Adipose Tissue P53 in the Regulation of Insulin Resistance. *Nature Medicine*, 15, 1082-1087. <u>https://doi.org/10.1038/nm.2014</u>
- [8] Aravinthan, A., Scarpini, C., Tachtatzis, P., Verma, S., Penrhyn-Lowe, S., Harvey, R., et al. (2013) Hepatocyte Senescence Predicts Progression in Non-Alcohol-Related Fatty Liver Disease. Journal of Hepatology, 58, 549-556. <u>https://doi.org/10.1016/j.jhep.2012.10.031</u>
- [9] Bonnet, L., Alexandersson, I., Baboota, R.K., Kroon, T., Oscarsson, J., Smith, U., et al. (2022) Cellular Senescence in Hepatocytes Contributes to Metabolic Disturbances in Nash. Frontiers in Endocrinology, 13, Article ID: 957616. <u>https://doi.org/10.3389/fendo.2022.957616</u>
- [10] Gorgoulis, V., Adams, P.D., Alimonti, A., Bennett, D.C., Bischof, O., Bischop, C., et al. (2019) Cellular Senescence: Defining a Path Forward. Cell, 179, 813-827. <u>https://doi.org/10.1016/j.cell.2019.10.005</u>
- [11] Takasugi, M., Yoshida, Y. and Ohtani, N. (2022) Cellular Senescence and the Tumour Microenvironment. *Molecular Oncology*, 16, 3333-3351. <u>https://doi.org/10.1002/1878-0261.13268</u>
- [12] Yao, J., Li, Y. and Wang, H. (2023) The Roles of Myeloid Cells in Aging-Related Liver Diseases. *International Journal of Biological Sciences*, **19**, 1564-1578.
- [13] 陈铭. 大数据时代的整合生物信息学[J]. 生物信息学, 2022, 20(2): 75-83.
- [14] Xue, G., Hua, L., Zhou, N., et al. (2021) Characteristics of Immune Cell Infiltration and Associated Diagnostic Biomarkers in Ulcerative Colitis: Results from Bioinformatics Analysis. Bioengineered, 12, 252-265.
- [15] Yang, Y., Cao, Y., Han, X., Ma, X., Li, R., Wang, R., et al. (2023) Revealing EXPH5 as a Potential Diagnostic Gene Biomarker of the Late Stage of COPD Based on Machine Learning Analysis. Computers in Biology and Medicine, 154, Article ID: 106621. <u>https://doi.org/10.1016/j.compbiomed.2023.106621</u>
- [16] 王建茹,李彬,彭广操,等.基于加权基因共表达网络分析挖掘 ECMO 治疗后心源性休克结局的核心枢纽基因 [J].临床心血管病杂志,2021,37(5):433-440.
- [17] Meng, Q., Li, X. and Xiong, X. (2022) Identification of Hub Genes Associated with Non-Alcoholic Steatohepatitis Using Integrated Bioinformatics Analysis. *Frontiers in Genetics*, 13, Article ID: 872518. https://doi.org/10.3389/fgene.2022.872518
- [18] The Gene Ontology Consortium, Aleksander, S.A., Balhoff, J., et al. (2023) The Gene Ontology Knowledgebase in 2023. Genetics, 224, iyad031.
- [19] Kanehisa, M. (2002) The KEGG Database. Novartis Foundation Symposium, Vol. 247, 91-101.
- [20] Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., et al. (2015) Robust Enumeration of Cell Subsets from Tissue Expression Profiles. Nature Methods, 12, 453-457. <u>https://doi.org/10.1038/nmeth.3337</u>
- [21] Hardy, T., Oakley, F., Anstee, Q.M., et al. (2016) Nonalcoholic Fatty Liver Disease: Pathogenesis and Disease Spectrum. Annual Review of Pathology, 11, 451-496.
- [22] Vilar-Gomez, E., Vuppalanchi, R., Mladenovic, A., et al. (2023) Prevalence of High-Risk Nonalcoholic Steatohepatitis (NASH) in the United States: Results from NHANES 2017-2018. Clinical Gastroenterology and Hepatology, 21, 115-124e7.
- [23] 马美娜, 刘巾玮. 治疗伴有肝纤维化的非酒精性脂肪性肝炎新药 Resmetirom 的临床评价[J]. 中国处方药, 2025, 23(1): 107-110.
- [24] Loft, A., Alfaro, A.J., Schmidt, S.F., Pedersen, F.B., Terkelsen, M.K., Puglia, M., et al. (2021) Liver-Fibrosis-Activated Transcriptional Networks Govern Hepatocyte Reprogramming and Intra-Hepatic Communication. Cell Metabolism, 33, 1685-1700.e9. <u>https://doi.org/10.1016/j.cmet.2021.06.005</u>

- [25] Xiao, Y., Batmanov, K., Hu, W., Zhu, K., Tom, A.Y., Guan, D., et al. (2023) Hepatocytes Demarcated by EphB2 Contribute to the Progression of Nonalcoholic Steatohepatitis. Science Translational Medicine, 15, eadc9653. https://doi.org/10.1126/scitranslmed.adc9653
- [26] Zhang, X., Zhou, D., Strakovsky, R., et al. (2012) Hepatic Cellular Senescence Pathway Genes Are Induced through Histone Modifications in a Diet-Induced Obese Rat Model. American Journal of Physiology Gastrointestinal and Liver Physiology, 302, G558-G564.
- [27] Laish, I., Mannasse-Green, B., Hadary, R., et al. (2016) Telomere Dysfunction in Nonalcoholic Fatty Liver Disease and Cryptogenic Cirrhosis. Cytogenetic and Genome Research, 150, 93-99.
- [28] Campisi, J. and d'Adda di Fagagna, F. (2007) Cellular Senescence: When Bad Things Happen to Good Cells. Nature Reviews Molecular Cell Biology, 8, 729-740. <u>https://doi.org/10.1038/nrm2233</u>
- [29] Mills, K.H.G. (2022) IL-17 and Il-17-Producing Cells in Protection versus Pathology. Nature Reviews Immunology, 23, 38-54. <u>https://doi.org/10.1038/s41577-022-00746-9</u>
- [30] Van Loo, G. and Bertrand, M.J.M. (2023) Death by TNF: A Road to Inflammation. *Nature Reviews Immunology*, 23, 289-303.
- [31] Vallejo, A., Valencia, K. and Vicent, S. (2017) All for One and FOSL1 for All: FOSL1 at the Crossroads of Lung and Pancreatic Cancer Driven by Mutant KRAS. *Molecular & Cellular Oncology*, **4**, e1314239.
- [32] Jiang, X., Xie, H., Dou, Y., Yuan, J., Zeng, D. and Xiao, S. (2019) Expression and Function of FRA1 Protein in Tumors. Molecular Biology Reports, 47, 737-752. <u>https://doi.org/10.1007/s11033-019-05123-9</u>
- [33] Tulchinsky, E. (2000) Fos Family Members: Regulation, Structure and Role in Oncogenic Transformation. *Histology and Histopathology*, 15, 921-928.
- [34] Dorn, C., Engelmann, J.C., Saugspier, M., Koch, A., Hartmann, A., Müller, M., et al. (2014) Increased Expression of C-Jun in Nonalcoholic Fatty Liver Disease. Laboratory Investigation, 94, 394-408. <u>https://doi.org/10.1038/labinvest.2014.3</u>
- [35] Videla, L.A., Tapia, G., Rodrigo, R., et al. (2009) Liver NF-kappaB and AP-1 DNA Binding in Obese Patients. Obesity (Silver Spring), 17, 973-979.
- [36] Yang, D., Xiao, C., Long, F., et al. (2019) Fra-1 Plays a Critical Role in Angiotensin II-Induced Vascular Senescence. The FASEB Journal, 33, 7603-7614.
- [37] Koh, C.M., Bezzi, M., Low, D.H.P., Ang, W.X., Teo, S.X., Gay, F.P.H., et al. (2015) MYC Regulates the Core PremRNA Splicing Machinery as an Essential Step in Lymphomagenesis. *Nature*, 523, 96-100. <u>https://doi.org/10.1038/nature14351</u>
- [38] Wang, H., Lu, J., Stevens, T., Roberts, A., Mandel, J., Avula, R., et al. (2023) Premature Aging and Reduced Cancer Incidence Associated with Near-Complete Body-Wide Myc Inactivation. Cell Reports, 42, Article ID: 112830. https://doi.org/10.1016/j.celrep.2023.112830
- [39] Schuster, S., Cabrera, D., Arrese, M. and Feldstein, A.E. (2018) Triggering and Resolution of Inflammation in NASH. *Nature Reviews Gastroenterology & Hepatology*, 15, 349-364. <u>https://doi.org/10.1038/s41575-018-0009-6</u>
- [40] Kazankov, K., Jørgensen, S.M.D., Thomsen, K.L., Møller, H.J., Vilstrup, H., George, J., et al. (2018) The Role of Macrophages in Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. Nature Reviews Gastroenterology & Hepatology, 16, 145-159. <u>https://doi.org/10.1038/s41575-018-0082-x</u>
- [41] Rensen, S.S., Slaats, Y., Nijhuis, J., Jans, A., Bieghs, V., Driessen, A., et al. (2009) Increased Hepatic Myeloperoxidase Activity in Obese Subjects with Nonalcoholic Steatohepatitis. The American Journal of Pathology, 175, 1473-1482. https://doi.org/10.2353/ajpath.2009.080999
- [42] Wolf, M.J., Adili, A., Piotrowitz, K., Abdullah, Z., Boege, Y., Stemmer, K., et al. (2014) Metabolic Activation of Intrahepatic CD8+ T Cells and NKT Cells Causes Nonalcoholic Steatohepatitis and Liver Cancer via Cross-Talk with Hepatocytes. Cancer Cell, 26, 549-564. <u>https://doi.org/10.1016/j.ccell.2014.09.003</u>