

基于生物信息学与机器学习识别NAFLD与肌少症的共享候选诊断生物标志物

李思阅

山东大学齐鲁医院内分泌与代谢病科, 山东 济南

收稿日期: 2026年3月23日; 录用日期: 2026年4月18日; 发布日期: 2026年4月23日

摘要

目的: 探讨非酒精性脂肪性肝病(NAFLD)与肌少症共享的跨组织分子特征, 并筛选候选诊断生物标志物。方法: 基于GEO数据库中的NAFLD肝组织数据集GSE135251和肌少症骨骼肌数据集GSE111016, 采用差异表达分析、加权基因共表达网络分析(WGCNA)、逻辑回归和LASSO筛选共享候选分子, 并结合模块富集分析与单样本通路活性评分进行功能解释。结果: 共鉴定出11个在两种组织中一致下调的假基因注释转录本, 并构建了可区分病例与对照的共病评分。进一步优先确定4个跨组织核心特征基因: ATP5F1EP2、NME2P1、RPL37P2和RPL39P38。该四基因特征在肝组织和骨骼肌中均具有一定判别能力, AUC分别为0.983和0.802, 跨组织迁移分析AUC为0.715。富集分析提示, 氧化磷酸化、核糖体/翻译及核苷酸相关过程在两种组织中均发生协同扰动, 并伴随组织依赖性的自噬相关模式。结论: NAFLD与肌少症之间可能共享一种翻译-生物能量学应激程序, ATP5F1EP2、NME2P1、RPL37P2和RPL39P38可作为潜在候选诊断生物标志物, 为后续机制研究提供线索。

关键词

NAFLD, 肌少症, 肝-肌轴, 转录组学, WGCNA, 氧化磷酸化, 核糖体

Identification of Shared Candidate Diagnostic Biomarkers for NAFLD and Sarcopenia Based on Bioinformatics and Machine Learning

Siyue Li

Department of Endocrinology and Metabolism, Qilu Hospital of Shandong University, Jinan Shandong

Received: March 23, 2026; accepted: April 18, 2026; published: April 23, 2026

Abstract

Objective: To investigate the shared cross-tissue molecular features of nonalcoholic fatty liver disease (NAFLD) and sarcopenia, and to identify candidate diagnostic biomarkers. **Methods:** Based on the NAFLD liver tissue dataset GSE135251 and the sarcopenic skeletal muscle dataset GSE111016 from the GEO database, differential expression analysis, weighted gene co-expression network analysis (WGCNA), logistic regression, and LASSO were used to identify shared candidate molecules, combined with module enrichment analysis and single-sample pathway activity scoring for functional interpretation. **Results:** A total of 11 pseudogene-annotated transcripts consistently down-regulated in both tissues were identified, and a comorbidity score capable of distinguishing cases from controls was constructed. Four cross-tissue core feature genes were further prioritized: ATP5F1EP2, NME2P1, RPL37P2, and RPL39P38. This four-gene signature showed discriminative ability in both liver tissue and skeletal muscle, with AUCs of 0.983 and 0.802, respectively, while the cross-tissue transfer analysis yielded an AUC of 0.715. Enrichment analysis indicated coordinated perturbations of oxidative phosphorylation, ribosome/translation, and nucleotide-related processes in both tissues, accompanied by tissue-dependent autophagy-related patterns. **Conclusion:** NAFLD and sarcopenia may share a common translation-bioenergetic stress program. ATP5F1EP2, NME2P1, RPL37P2, and RPL39P38 may serve as potential candidate diagnostic biomarkers and provide clues for subsequent mechanistic studies.

Keywords

NAFLD, Sarcopenia, Liver-Muscle Axis, Transcriptomics, WGCNA, Oxidative Phosphorylation, Ribosome

Copyright © 2026 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/>



Open Access

1. 引言

非酒精性脂肪性肝病(NAFLD)是当前全球最常见的慢性肝病之一,其流行与肥胖和2型糖尿病患者率上升密切相关[1]-[3]。临床上,NAFLD涵盖从单纯性脂肪肝病变到脂肪性肝炎、进行性纤维化、肝硬化乃至肝细胞癌的一系列病理谱系,并与全身胰岛素抵抗及心脏代谢风险高度相关。

肌少症是以骨骼肌质量、肌力及身体功能进行性下降为特征的临床综合征,是虚弱、失能和死亡的重要促成因素[4][5]。除衰老外,肌少症还越来越被视为一种与肥胖、糖尿病及慢性低度炎症相互交织的代谢性并发症。骨骼肌作为机体最大的胰岛素敏感器官和底物氧化核心场所,其数量与质量下降可能放大全身代谢应激,并影响肝脏等肌外器官[6]。

越来越多的流行病学研究支持NAFLD与肌少症之间存在双向关联[7]。肌少症与NAFLD风险升高及更严重的肝纤维化相关,而NAFLD严重程度在多个队列中也与更高的肌少症患病率相关[8]。然而,二者的因果关系及混杂因素仍存在争议[9]。这些不确定性提示,研究NAFLD-肌少症共病需要跨组织分子视角,而非仅限于单器官层面的相关性分析。

从机制上看,肝-肌轴涉及代谢物、细胞因子、肝因子、肌因子及细胞外囊泡介导的内分泌和旁分泌通讯[10]。NAFLD与肌少症均被认为涉及多种共同过程,包括线粒体功能障碍及氧化磷酸化受损、翻译/蛋白稳态失衡、慢性炎症以及自噬-溶酶体通路异常[11]。值得注意的是,自噬等应激适应机制具有

明显的组织与情境依赖性：在肝脏中，自噬可能表现为保护性反应，也可能相对不足；而在骨骼肌中，自噬受损则可能妨碍细胞器更新并促进功能衰退[12]。因此，亟须开展以通路为中心的跨组织联合分析，以明确共享共病信号究竟反映共同的上游应激轴、组织特异性代偿反应，还是二者兼而有之[13]。

尽管上述认识不断深化，现有大多数转录组研究仍将 NAFLD 肝组织或肌少症肌肉组织分开分析，重点多集中于疾病特异性特征或单一通路。直接在人体肝脏与骨骼肌中量化共享共表达模块及其通路背景的整合性跨组织研究仍然有限。对于转化研究而言，至少存在两个关键问题：其一，是否可以用一个紧凑且便于解释的基因集合概括跨组织的共病相关生物学；其二，这一特征如何映射到翻译/核糖体活性、线粒体生物能量学、核苷酸相关代谢和自噬等连贯的通路级程序，而不仅仅是若干孤立的差异基因。

另一层值得关注的调控因素是假基因注释转录本[14][15]。假基因曾长期被认为没有功能，但如今已被证实在多种情境下具有转录活性，并可通过竞争性内源 RNA 或调控 RNA 稳定性等机制影响基因表达[16][17]。与此同时，假基因信号的生物学解释必须保持谨慎，因为注释歧义、与亲本基因的序列相似性以及情境特异性表达均可能增加推断难度[16][18]。因此，若能在不同组织中识别到可重复的假基因来源信号，并与一致的通路背景相结合，则可为系统层面的后续验证提供有价值且可检验的起点，而非直接作出因果性结论。

基于此，我们提出假设：NAFLD 与肌少症共享一种可由紧凑假基因来源特征概括的跨组织转录程序，该程序沿肝-肌轴反映翻译-生物能量学应激状态，并伴随具有组织依赖性的自噬相关反应。为检验这一假设，本研究分析了肝组织(GSE135251)和骨骼肌(GSE111016)的人类 RNA-seq 数据，整合差异表达分析与 WGCNA 以识别疾病相关模块及共享候选分子；随后利用惩罚回归构建简约预测特征，评估其在各队列内的判别能力及探索性跨组织迁移能力，并通过共表达富集、样本层面通路活性评分及网络互作分析解释核心特征的生物学意义。总体而言，本研究提出了一个具有可解释性的跨组织分析框架，可为 NAFLD-肌少症共病的后续验证与机制研究优先筛选共享候选分子和通路假说。

2. 材料与方法

2.1. 研究设计与数据来源

本研究的整体流程见图 1。我们从 Gene Expression Omnibus (GEO) 数据库中获取公开的人类 RNA-seq 转录组数据集，包括 NAFLD 数据集 GSE135251 和肌少症数据集 GSE111016。基因层面的表达矩阵采用 TPM 单位，样本表型信息则来自 GEO 系列矩阵文件及对应研究注释。最终纳入分析的样本包括 GSE135251 中 206 例 NAFLD 肝组织病例和 10 例健康肝组织对照，以及 GSE111016 中 20 例肌少症骨骼肌病例和 20 例健康肌肉对照。

2.2. 表达矩阵预处理

所有表达矩阵在后续分析前均进行 $\log_2(\text{TPM} + 1)$ 转换。样本标签与表达矩阵在统计检验和可视化分析前均进行了严格对应与校验。

2.3. 差异表达分析与功能富集分析

分别在 NAFLD (GSE135251) 和肌少症 (GSE111016) 队列中，基于 $\log_2(\text{TPM} + 1)$ 表达矩阵采用 limma 软件包进行差异表达分析[19]。差异表达基因 (DEGs) 筛选标准为名义 $P < 0.05$ 且 $|\log_2\text{FC}| \geq 0.263$ [20]。鉴于本研究更重视跨队列一致性及后续正交筛选，初始 DEG 步骤作为偏向灵敏度的探索性筛选，后续富集分析和建模阶段再实施多重校正。火山图和热图分别采用 ggplot2 和 pheatmap 绘制。

共享分子识别通过求取 NAFLD 肝组织与肌少症骨骼肌 DEGs 的交集，并保留变化方向一致性。

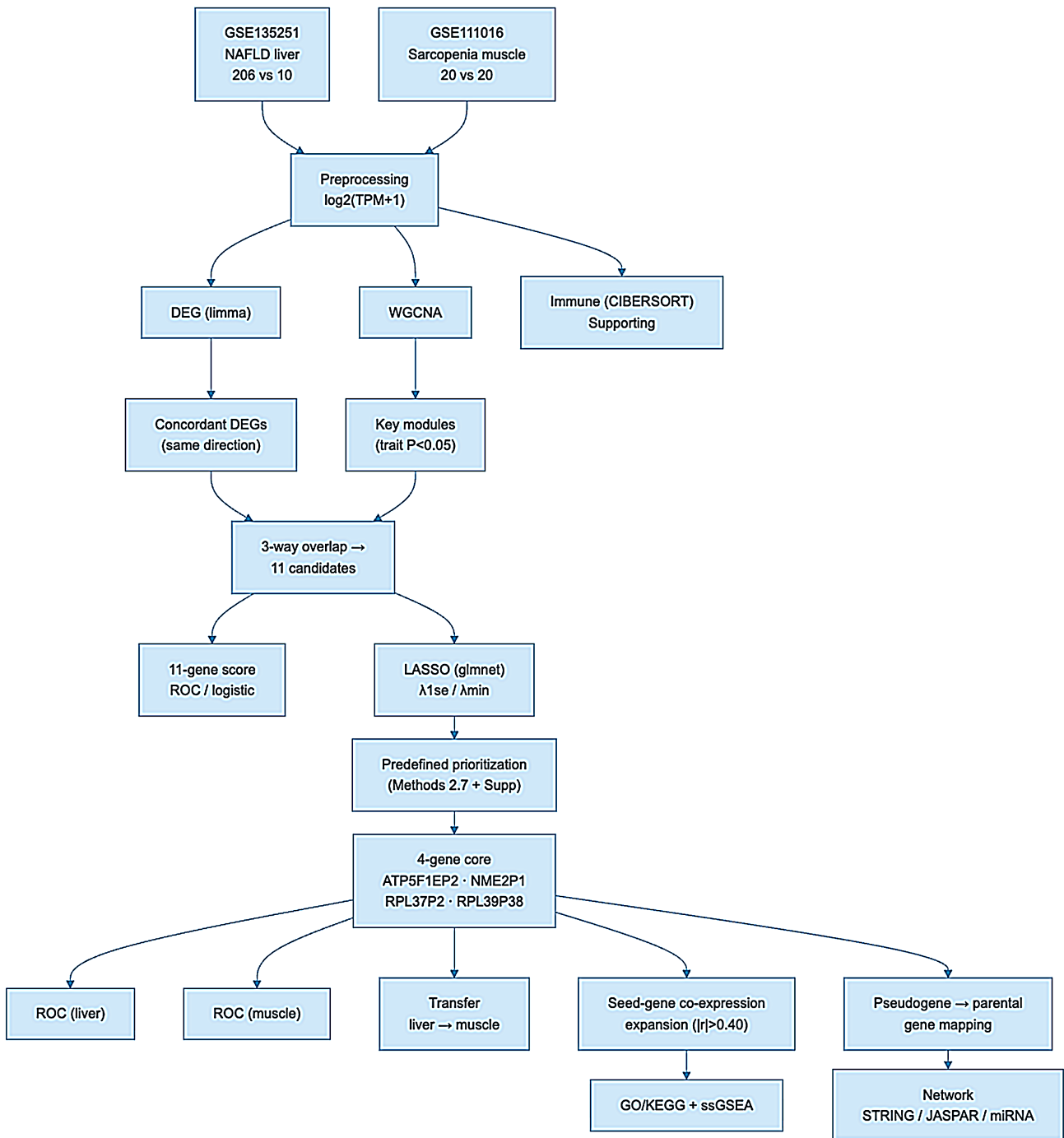


Figure 1. Study workflow for differential expression analysis and shared functional characterization of NAFLD and sarcopenia

图 1. NAFLD 与肌少症差异表达及共享功能特征的研究流程图

随后使用 clusterProfiler 对一致失调基因进行 GO 和 KEGG 富集分析，结果排序与可视化均基于 BH 校正后的 P 值(p.adjust)。

2.4. 加权基因共表达网络分析

采用 WGCNA 软件包构建加权基因共表达网络，以识别共表达模块并评估其与临床表型之间的关联

[21]。模块特征向量与表型之间采用 Pearson 相关分析，名义 $P < 0.05$ 的模块定义为性状相关模块，并纳入后续整合分析。之后，将关键模块基因与 DEGs 求交，以获得 NAFLD 与肌少症之间的共享候选基因，并结合方向一致性、FDR 控制的富集分析和惩罚建模进一步优先排序。

2.5. 核心基因表达统计与单基因 ROC 分析

从各队列中提取候选基因表达值($\log_2(\text{TPM} + 1)$)，组间差异主要采用 Wilcoxon 秩和检验评估，并以双侧 Studentt 检验作为敏感性分析。对 11 个共享共病候选基因，各队列内的 P 值均采用 BH 方法进行 FDR 校正[22]。单基因判别能力采用 pROC 软件包绘制 ROC 曲线，并分别在肝组织和肌肉中计算 AUC [23]。

2.6. 共病评分构建与逻辑回归分析

在每个队列内，先对每个基因进行 Z-score 标准化，再对各样本标准化值取均值，定义为队列特异性的共病评分。组间差异采用 Wilcoxon 秩和检验评估，诊断性能通过 pROC 进行 ROC 分析并报告 AUC [23]。同时采用二元逻辑回归评估共病评分与疾病状态的关联，并报告 $\text{OR} = \exp(\beta)$ 及其 Wald 95% 置信区间。

2.7. LASSO 特征筛选与跨组织迁移能力分析

采用 LASSO 逻辑回归(glmnet; $\alpha = 1$)，结合 10 折交叉验证确定 λ ，并根据 λ_{1se} 和/或 λ_{min} 下的非零系数构建简约模型[24] [25]。为评估跨组织可迁移性，将肝脏队列训练得到的系数应用于骨骼肌表达矩阵，计算迁移特征评分，并在肌少症队列中通过 ROC/AUC 评估其判别性能。

为从 11 个共享候选基因中定义跨组织核心特征，本研究采用预先设定的整合性优先排序策略，综合考虑：(i) 跨组织方向一致性；(ii) 单基因统计或诊断证据；(iii) LASSO 选择支持。需要强调的是，“跨组织核心特征”并非直接取自任一单组织 LASSO 模型，而是基于多项证据综合优先排序得到。某候选基因需满足以下多数原则：① 在肝脏与肌肉中方向一致；② 至少在一个组织中具有较强单基因证据，并在另一组织中保留基本证据；③ 获得组织特异性 LASSO 支持，双组织同时入选者优先级最高；④ 位于翻译、核糖体或线粒体生物能量学相关的共享模块或通路背景中。该策略旨在保留能够代表“肝脏主导信号”“肌肉主导信号”及“跨组织共同锚点”的最小特征集合，以提高跨组织解释性。

2.8. 种子基因共表达拓展与富集分析

在 \log_2 标准化表达矩阵中，将简约模型中的候选特征作为种子基因。分别计算每个种子基因与所有表达基因之间的 Pearson 相关性，并用 corPvalueStudent 计算相关 P 值[21]。当某基因与至少一个种子基因满足 $|r| > 0.40$ 且 $P < 0.05$ 时，即定义为共表达基因；不同种子基因的共表达集合随后合并，形成队列特异性的共表达基因集。之后以所有表达基因为背景，采用 clusterProfiler 将基因符号映射为 EntrezID，并对共表达基因开展 GOBiological Process 和 KEGG 富集分析[26]。富集阶段采用 BH 方法进行多重校正 [27]-[29]。

2.9. 基于 ssGSEA 的通路活性评分及其与特征评分的关联

基于肝脏和肌肉中共同富集的共表达 GO/KEGG 结果，筛选代表性通路并构建 ssGSEA 基因集。共构建 4 个自定义基因集：OXPHOS、RIBOSOME、AUTOPHAGY 和 NUCLEOTIDE [30] [31]。具体做法为：从以四基因为中心的共表达模块中提取对显著富集有贡献的基因，再取两种组织的交集作为最终 ssGSEA 基因集。随后利用 GSVA 软件包在 \log_2 转换表达矩阵上计算 ssGSEA 评分，并开展组间比较与相关性分析[32] [33]。

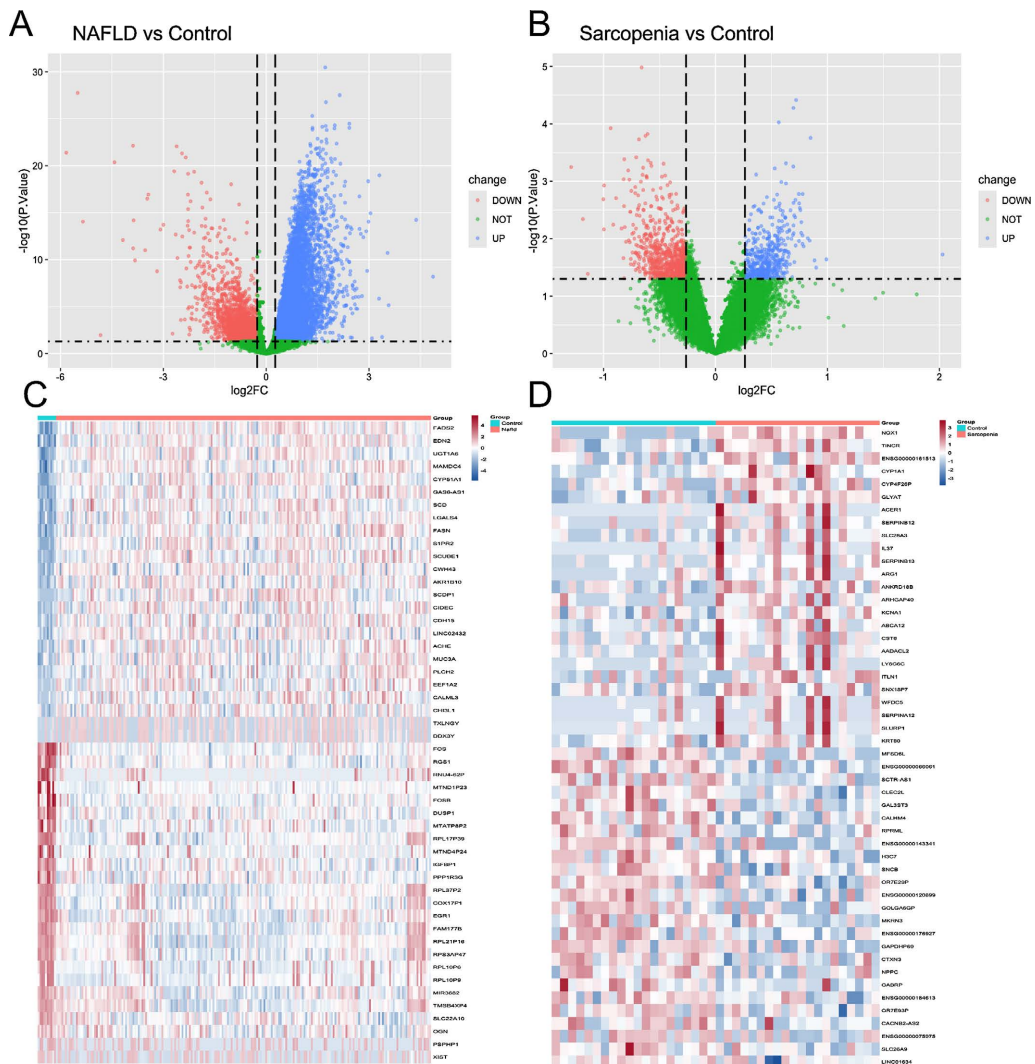
2.10. PPI 与调控网络分析

由于 STRING 和 NetworkAnalyst 对多数假基因标识符支持有限, 我们首先依据官方基因命名体系, 将四个枢纽假基因(ATP5F1EP2、NME2P1、RPL37P2 和 RPL39P38)映射至相应的蛋白编码亲本基因 ATP5F1E、NME2、RPL37 和 RPL39。随后在 STRING Web 界面开展 PPI 分析。TF-gene 调控关系基于 NetworkAnalyst 中的 JASPAR 数据库推断, 而 miRNA-mRNA 交互则基于 miRTarBase 的实验验证证据在 NetworkAnalyst 中获取。

3. 结果

3.1. 共享基因的差异表达及功能富集分析

两组队列均显示出广泛的转录组改变(图 2(A)~(D))。在 NAFLD 肝组织数据集(GSE135251; 206 例病例 vs 10 例对照)中, 共检测到 6936 个上调基因和 2511 个下调基因; 而在肌少症骨骼肌数据集(GSE111016; 20 例病例 vs 20 例对照)中, 共有 508 个上调基因和 747 个下调基因。按 logFC 排序后, 选取前 25 个上调和前 25 个下调基因绘制热图, 可见两种组织中均存在与分组高度一致的表达偏移模式(图 2(C), 图 2(D))。



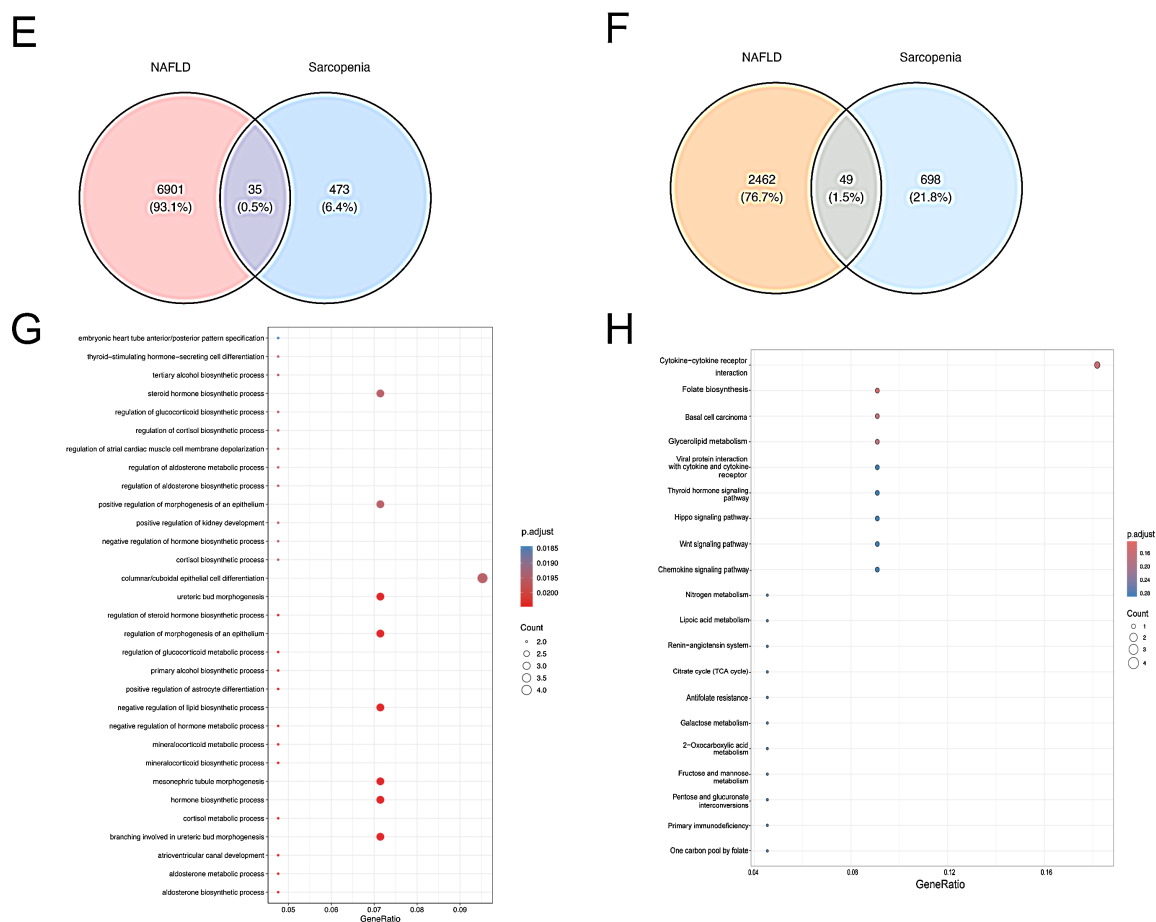


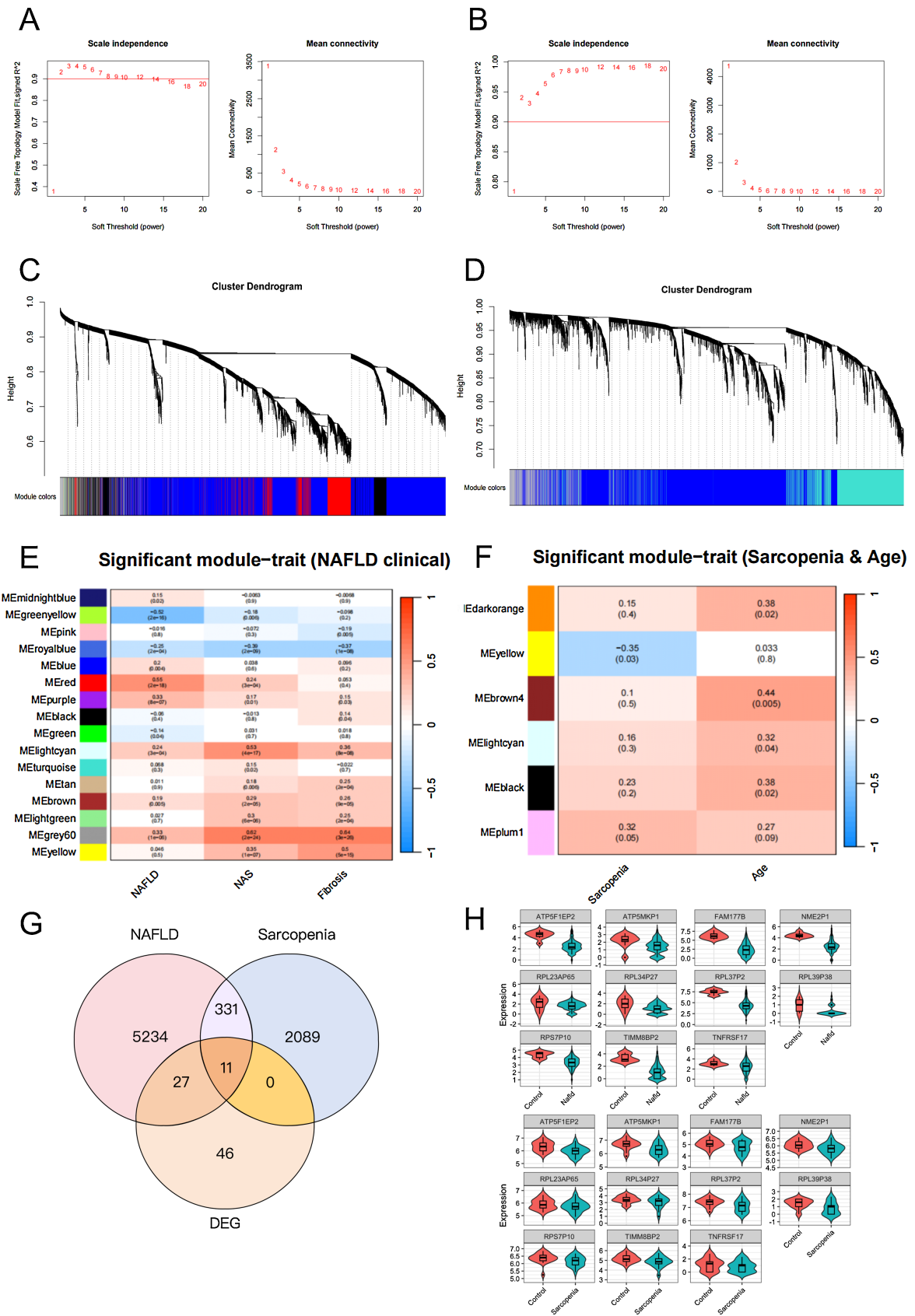
Figure 2. Differential expression patterns and shared functional features in NAFLD and sarcopenia
图 2. NAFLD 与肌少症中的差异表达及共享功能特征。

为识别共享分子信号，我们在保留变化方向的前提下对两队列 DEGs 求交，最终获得 35 个一致上调基因和 49 个一致下调基因，共计 84 个共享失调基因(图 2(E)，图 2(F))。对这 84 个基因开展 GO Biological Process 富集分析后发现，显著条目主要集中于类固醇/激素相关过程以及发育/形态发生相关过程(图 2(G))。KEGG 分析则提示其涉及多种代谢通路以及免疫炎症相关信号(图 2(H))。

3.2. WGCNA 识别疾病相关模块及共享的 11 基因共病特征

为解析与 NAFLD 和肌少症相关的协同转录程序，我们分别在肝组织(GSE135251)和骨骼肌(GSE111016)中构建加权基因共表达网络。基于 $\log_2(\text{TPM} + 1)$ 表达矩阵及无尺度拓扑分析结果，分别选择 $\beta = 3$ (NAFLD)和 $\beta = 4$ (肌少症)作为建网参数(图 3(A)，图 3(B))。最终，NAFLD 网络识别出 16 个共表达模块，肌少症网络识别出 6 个模块(图 3(C)，图 3(D))。模块 - 性状相关性分析显示，两种组织中均存在疾病及年龄相关的共表达程序(图 3(E)，图 3(F))：在 NAFLD 网络中，grey60 和 yellow 等模块与 NAFLD 诊断及疾病严重程度指标(NAS 和纤维化)呈正相关；在肌少症网络中，yellow 模块与肌少症状态呈负相关($r = -0.35$, $P = 0.03$)，同时多个模块与年龄呈正相关。

为获得稳健的跨组织共病特征，我们将名义 $P < 0.05$ 的模块定义为关键模块，提取各队列关键模块中的基因，并与肝脏和肌肉中方向一致的共享 DEGs 进行交集分析，最终得到 11 个共享共病基因：ATP5F1EP2、ATP5MKP1、FAM177B、NME2P1、RPL23AP65、RPL34P27、RPL37P2、RPL39P38、RPS7P10、



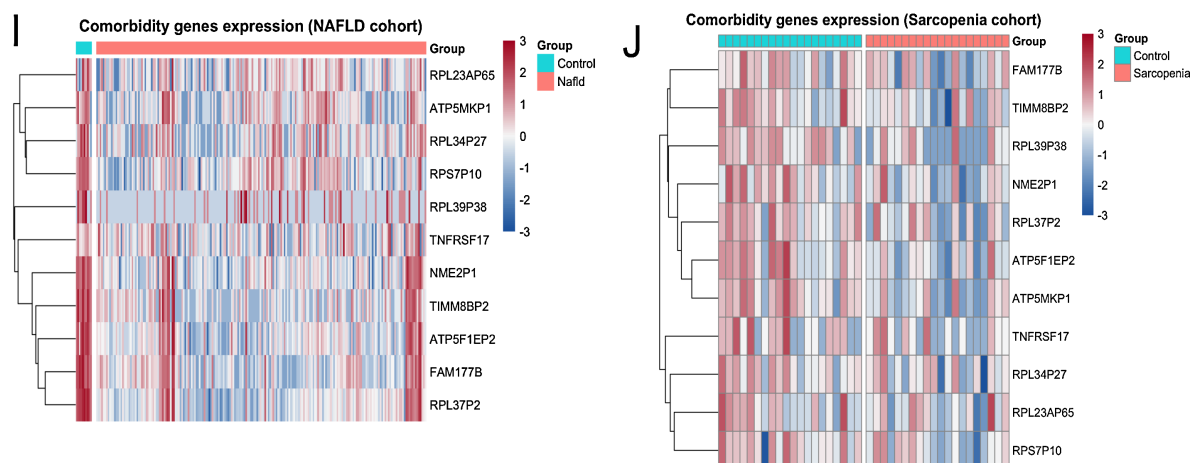


Figure 3. WGCNA identifies disease-related modules and shared comorbidity genes in NAFLD and sarcopenia
图 3. WGCNA 识别 NAFLD 与肌少症中的疾病相关模块和共享共病基因。

TIMM8BP2 和 TNFRSF17 (图 3(G))。

值得注意的是,该面板具有明显的假基因富集特征,其中 9 个为假基因,另有 2 个为蛋白编码基因,因此对这些假基因来源信号的功能解释需结合亲本基因映射及通路背景谨慎进行。11 个转录本在各队列内均显示出清晰且方向一致的疾病 - 对照差异(图 3(H)),热图亦显示出贯穿全部样本的协同共表达模式(图 3(I),图 3(J))。基于上述结果,这 11 个转录本被作为后续共病评分构建和预测建模的候选面板。

3.3. 11 个共享共病基因的表达变化及单基因诊断性能

在整合差异表达结果与 WGCNA 模块 - 性状关联后,我们获得了 11 个同时位于疾病相关模块中、且在肝脏与骨骼肌中方向一致失调的共享共病候选基因,即 ATP5F1EP2、ATP5MKP1、FAM177B、NME2P1、RPL23AP65、RPL34P27、RPL37P2、RPL39P38、RPS7P10、TIMM8BP2 和 TNFRSF17。这 11 个基因在 NAFLD 肝组织和肌少症骨骼肌中均表现为一致下调。NAFLD 队列中,这些基因的 \log_2FC 约介于 -0.57 至 -3.88,其中 9 个经多重校正后仍具有统计学意义($FDR < 0.05$);而在骨骼肌中,相应的 \log_2FC 幅度较小(约为 -0.16 至 -0.61),部分基因仅达到名义显著,可能与样本量较小有关。尽管统计效能存在差异,所有基因在两种组织中均保持一致下调,支持其作为共享共病信号的合理性。

单基因 ROC 分析进一步显示了该面板的判别潜力。在肝脏队列中,各基因 AUC 范围为 0.68~0.98,其中 FAM177B、RPL37P2、TIMM8BP2 和 NME2P1 的区分能力较强,AUC 均超过 0.95;在肌肉队列中,单基因 AUC 整体较低,但仍普遍高于随机水平(AUC 范围 0.59~0.75),其中 ATP5F1EP2、ATP5MKP1 和 RPL39P38 表现相对更优。总体来看,这 11 个共享共病基因不仅在肝脏和骨骼肌中呈现一致性下调,而且在单基因层面也提供了探索性的判别信号。基于此,我们进一步将其整合为复合共病评分,并采用 LASSO 逻辑回归推导简约多基因特征。

3.4. 11 基因共病评分的诊断性能

基于 WGCNA-DEG 交集得到的 11 个共享共病基因,我们对每个样本的 11 个基因表达进行逐基因 Z 标准化后取平均值,构建队列层面的共病评分。在两个队列中,该评分均能较清晰地地区分疾病组与对照组(图 4(A),图 4(B)): NAFLD 肝组织样本评分显著降低,肌少症骨骼肌样本亦呈一致性下降,提示该特征在不同组织中具有相同方向的变化。

ROC 分析显示,11 基因评分对 NAFLD 具有较高的表观判别能力($AUC = 0.968$;图 4(C)),对肌少症

则表现为中等判别能力(AUC = 0.747; 图 4(D))。以 NAFLD 状态为因变量的逻辑回归分析表明, 11 基因评分越高, 与 NAFLD 风险越低显著相关(OR = 0.03, 95% CI 0.01~0.14, P < 0.001; 图 4(E))。在肌少症队列中, 校正年龄后, 11 基因评分仍为独立预测因子(OR = 0.25, 95% CI 0.08~0.76, P = 0.015; 图 4(F))。

进一步分析显示, 11 个候选基因在两套数据集中的单基因 AUC 均高于随机水平(AUC > 0.5), 且肝脏中的单基因判别表现普遍优于骨骼肌(图 4(G))。综上, 11 基因共病评分捕捉到了一个连贯的跨组织信号, 与 NAFLD 显著相关, 并对肌少症状态具有一定判别能力。

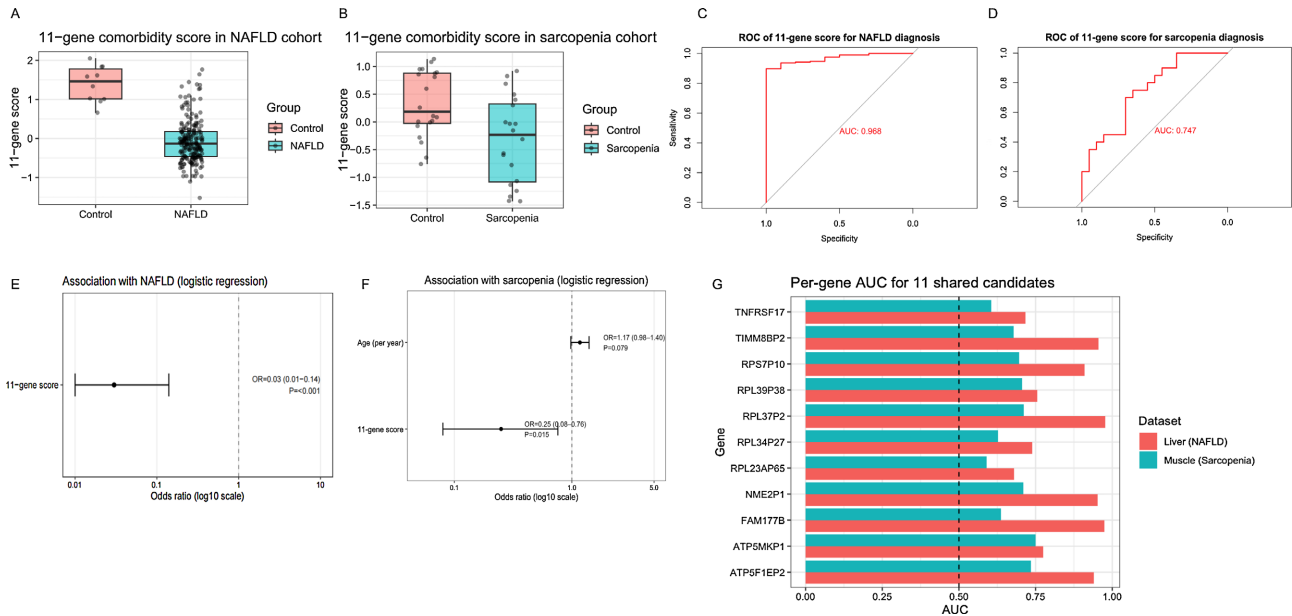


Figure 4. Diagnostic performance of the 11-gene comorbidity score and individual genes in NAFLD and sarcopenia
图 4. 11 基因共病评分及单基因在 NAFLD 与肌少症中的诊断表现

3.5. 跨组织四基因核心特征的识别

随后, 我们在每个队列中对 11 个共享候选基因应用 LASSO 逻辑回归, 以提取更简约的组织特异性稀疏特征(图 5(A)~(D))。在 NAFLD 肝脏队列中, LASSO 模型保留 4 个非零系数基因(FAM177B、RPL37P2、RPL39P38 和 TIMM8BP2; 图 5(A), 图 5(C)); 在肌少症骨骼肌队列中, 保留 4 个基因(ATP5F1EP2、ATP5MKP1、NME2P1 和 RPL39P38; 图 5(B), 图 5(D))。其中, RPL39P38 是唯一同时进入两种组织 LASSO 模型的基因, 可视为跨组织共同锚点。

需要强调的是, 本研究定义的四基因跨组织核心特征并非任何单一 LASSO 模型的直接输出, 也不是两组 LASSO 结果的简单拼接, 而是依据预先设定的整合性优先排序策略, 从两个组织的稀疏模型中进一步筛选得到。该策略综合考虑了跨组织方向一致性、各组织内最低限度的单基因证据、LASSO 支持以及共享模块/通路背景。基于此, 我们最终优先确定 ATP5F1EP2、NME2P1、RPL37P2 和 RPL39P38 作为最小跨组织核心特征集合: 其中 RPL39P38 代表双组织共同锚点, RPL37P2 更代表肝脏侧稳定信号, 而 ATP5F1EP2 和 NME2P1 更代表肌肉侧稳定信号。

在此基础上, 我们分别在各队列内重新拟合多变量逻辑回归模型, 并以线性预测子量化评分。该特征在 NAFLD 肝脏队列中表现出很高的表观判别力(AUC = 0.983; 图 5(F)), 在肌少症骨骼肌队列中也具有良好判别能力(AUC = 0.802; 图 5(G)); 在探索性跨组织迁移分析中, 将肝脏训练得到的四基因模型系数直接应用于肌肉表达矩阵后, 仍获得中等性能(AUC = 0.715; 图 5(E))。与此一致, 四基因评分在两种

组织中均表现为疾病组高于对照组(图 5(H), 图 5(I))。

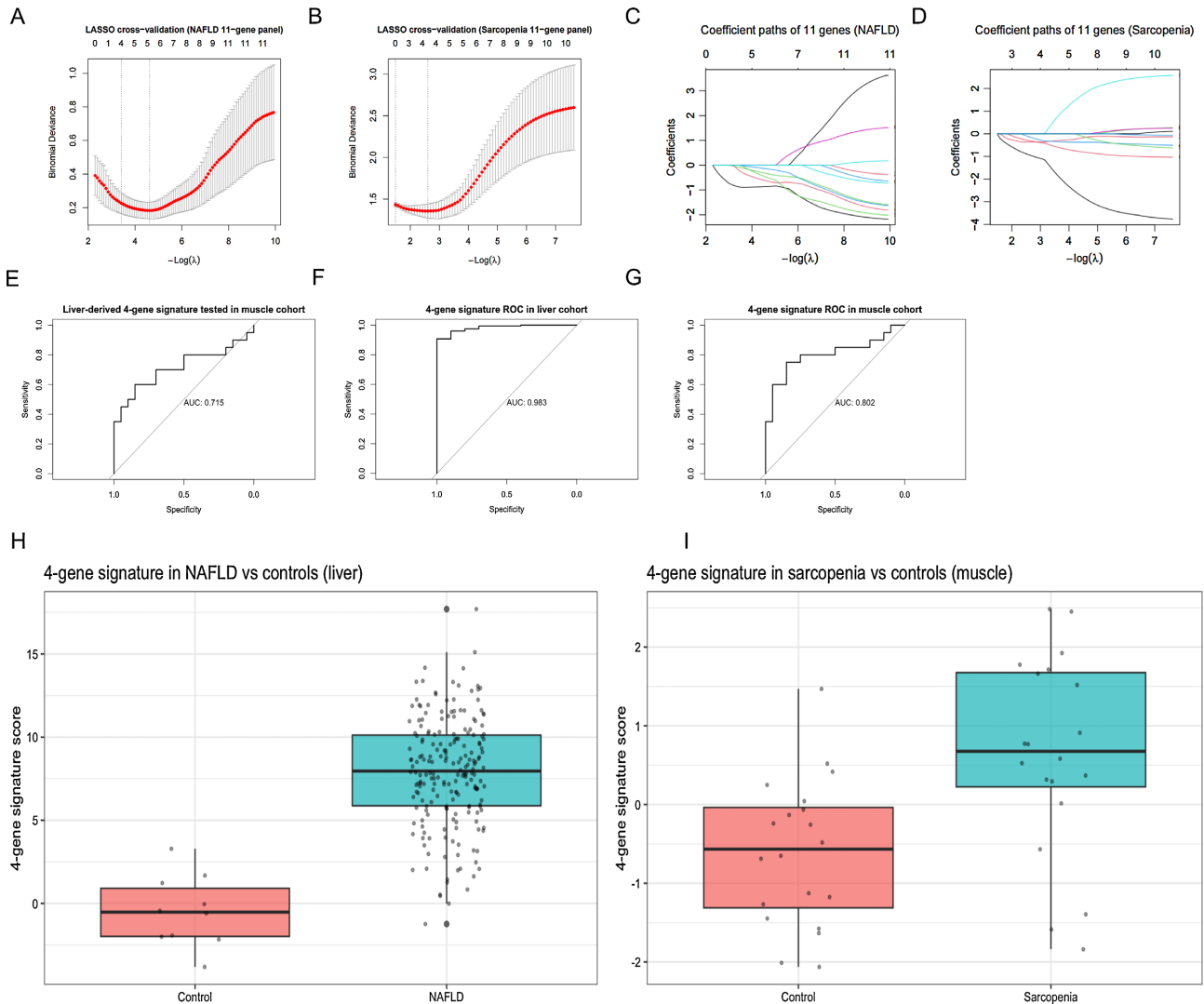


Figure 5. Construction and evaluation of the four-gene core signature in the NAFLD liver and sarcopenia muscle cohorts
图 5. NAFLD 肝组织与肌少症肌肉队列中四基因核心特征的构建与评估

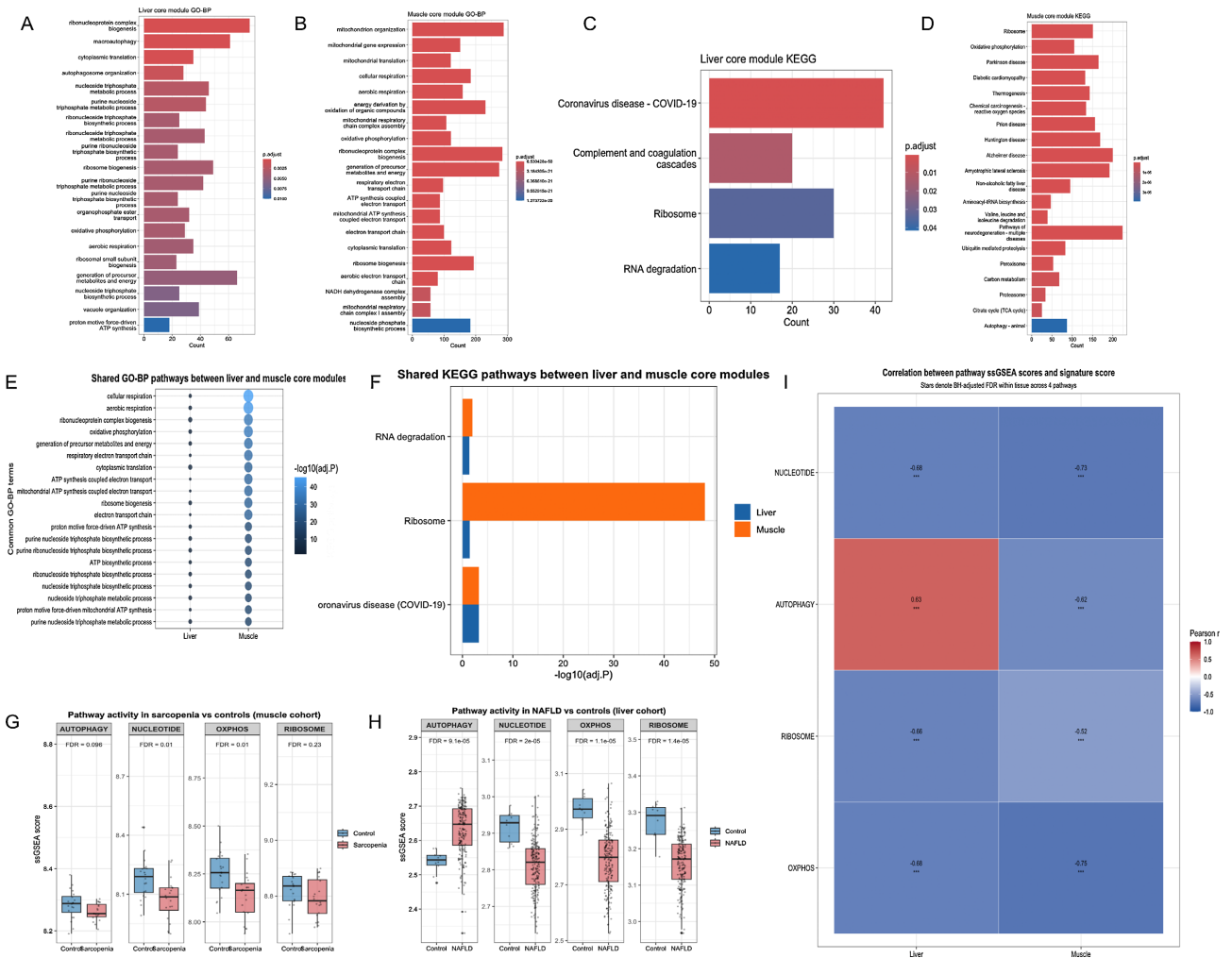
3.6. 基于共表达的核心模块富集及其通路背景

基于前述定义的肝脏和肌肉核心模块, 我们进一步开展 GO-BP 和 KEGG 富集分析。NAFLD 肝脏核心模块主要富集于核糖体生物发生、胞质翻译、核苷三磷酸代谢、线粒体氧化磷酸化与有氧呼吸及自噬相关过程(图 6(A)); 肌少症肌肉核心模块则更突出地富集于线粒体组织、呼吸电子传递、ATP 合成偶联电子传递及线粒体基因表达等过程(图 6(B))。KEGG 分析显示, 两种组织均显著富集于 Ribosome 和 RNA degradation, 而肌肉模块还进一步映射到 Oxidative phosphorylation、神经退行性疾病相关通路及 non-alcoholic fatty liver disease 通路(图 6(C)~(F))。这些结果提示, NAFLD 与肌少症之间可能共享翻译 - 生物能量学异常, 并伴随组织依赖性的自噬相关特征。

为在样本层面量化通路活性, 我们进一步构建了 4 个代表性基因集(OXPPOS、Ribosome、Autophagy 和 nucleoside triphosphate metabolism), 并计算 ssGSEA 评分。在肌少症骨骼肌队列中, OXPPOS 和

NUCLEOTIDE 评分在病例组较对照组下降(两者 FDR 均为 0.01), Autophagy 呈下降趋势(FDR = 0.096), 而 Ribosome 差异不显著(FDR = 0.23) (图 6(G))。在 NAFLD 肝脏队列中, OXPHOS、NUCLEOTIDE 和 Ribosome 评分均显著降低, 而 Autophagy 评分则高于对照(均为 FDR < 1 × 10⁻⁴; 图 6(H))。

进一步相关性分析显示, 在肝脏队列中, 较高的四基因特征评分与 OXPHOS、Ribosome 和 NUCLEOTIDE 活性呈负相关, 而与 Autophagy 呈正相关; 在肌肉队列中, 特征评分与四条通路均呈负相关(各组织内 4 条通路的 BH 校正 FDR 均<0.001)(图 6(I))。上述结果进一步支持 NAFLD 肝脏与肌少症肌肉共享翻译 - 生物能量学受损状态, 并提示自噬反应存在组织依赖性差异。



靶向网络框架(图 7(B))。基于 JASPAR 的 TF-gene 分析显示, FOXC1、GATA2 等高连接度转录因子可同时调控多个枢纽基因, 提示其可能受到共享上游调控因素影响(图 7(A))。此外, 基于 miRTarBase 的分析提取了 NME2 的 miRNA-mRNA 子网络(图 7(C))。总体而言, 这些网络结果从互动与调控层面为四基因核心特征提供了背景支持。

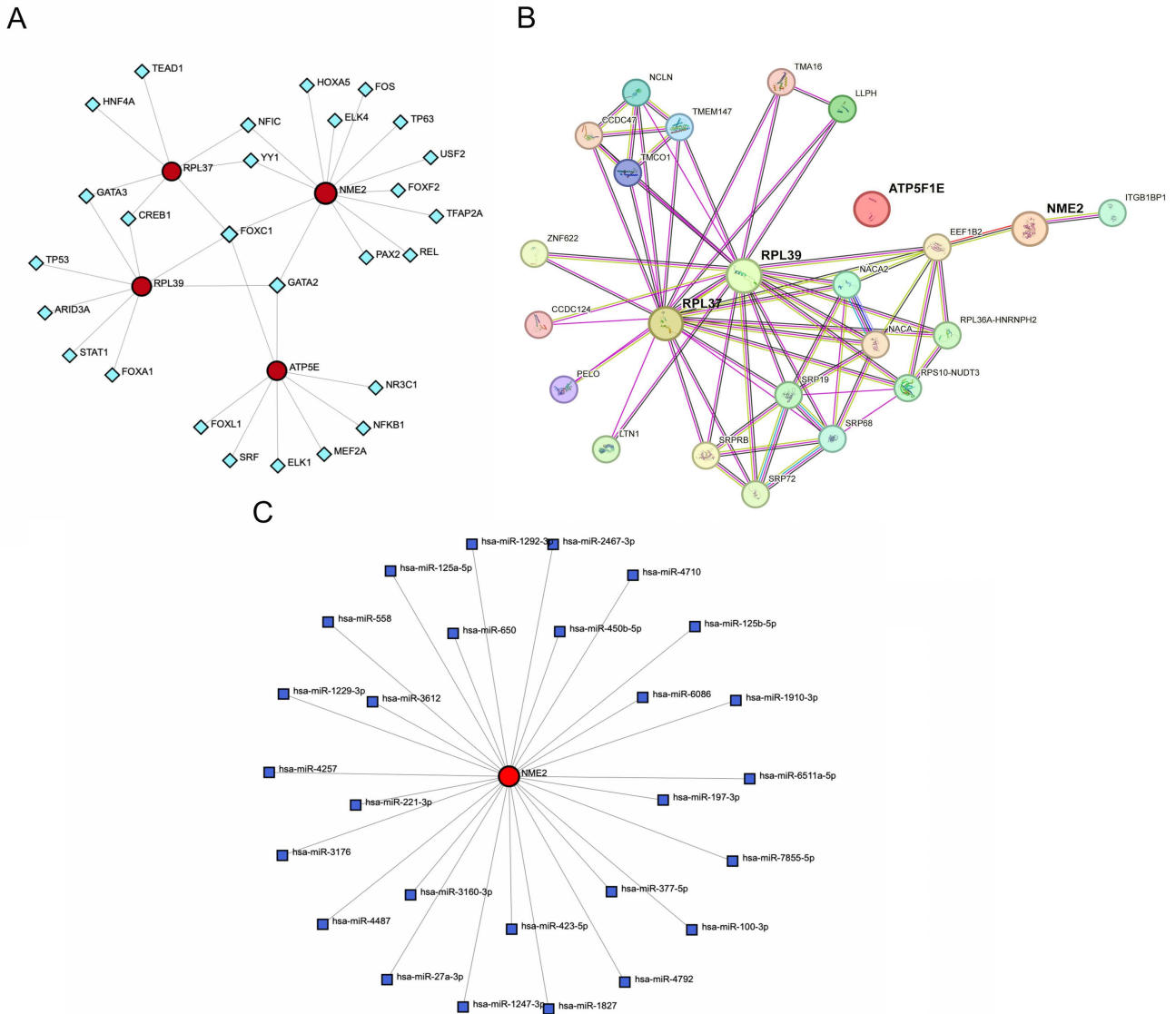


Figure 7. Network background analysis after mapping the four core pseudogenes to their parent genes
图 7. 四个核心假基因映射后的网络背景分析

4. 讨论

NAFLD 与肌少症在代谢性衰老背景下常常共存, 并越来越被视为一种多器官综合征[23] [24]。既往研究已分别提示两类疾病存在线粒体功能障碍、氧化磷酸化(OXPHOS)受损及蛋白稳态异常, 但这些程序是否在不同组织中呈现收敛仍不清楚[25] [26]。本研究通过整合肝脏与肌肉队列的 WGCNA 和 DEG 分析发现, 两种组织中存在一致下调的共享候选分子, 并在核糖体、翻译和线粒体能量代谢通路层面呈现稳定收敛, 提示 NAFLD 与肌少症之间可能共享一种跨组织的翻译 - 生物能量学应激程序。

本研究的重要发现之一，是识别出连接翻译相关过程与线粒体生物能量学的跨组织收敛模式。既往研究表明，NAFLD 常伴随肝脏线粒体呼吸能力下降和能量稳态紊乱，而肌少症则表现为肌肉线粒体功能减退和合成代谢能力下降[27]-[29]。本研究进一步提示，这些异常不仅表现为单个差异表达基因，还可在共表达模块和样本层面通路活性变化中得到体现。因此，这一共享信号更适合被理解为一种跨组织分子表型，即翻译能力与线粒体 ATP 生成程序共同下降的能量受限状态。值得注意的是，在肝脏中自噬相关评分升高，而在肌肉中则表现为较弱甚至下降趋势。结合既往研究，这种差异可能与肝脏中 AMPK-mTOR 轴介导的代偿性自噬激活，以及骨骼肌中 mTORC1-AMPK-ULK1 轴失衡导致的自噬诱导不足有关[30][31]。不过，由于本研究基于转录组推断通路活性，尚不能等同于真实自噬通量，仍需进一步实验验证。

通过疾病相关模块与方向一致 DEGs 的交集分析，我们鉴定了 11 个共享共病候选分子，其中多数为假基因转录本。假基因信号近年来在大规模转录组研究中受到越来越多关注，但其解释必须保持谨慎，因为其来源可能涉及亲本位点共调控、染色质状态或技术与注释依赖性[32]。为获得更紧凑且具有机制可解释性的特征，我们先在肝脏和肌肉中分别应用 LASSO 得到组织特异性稀疏候选集合，再依据预设的跨组织优先排序策略定义四基因核心特征。该四基因集合并非任何单一 LASSO 模型的直接输出，也不是事后任意拼接，而是综合跨组织方向一致性、单基因基本证据、LASSO 支持以及共享模块/通路背景后得到的最小代表集合。因此，四基因评分更适合作为翻译 - 生物能量学应激表型的低维读出指标，而非已完成独立外部验证的确定性诊断分类器[33]。

本研究仍存在局限。首先，两套队列样本量不平衡，尤其是 NAFLD 队列中对照样本较少，可能影响效应量稳定性。其次，整体组织转录组信号可能部分反映细胞组成差异或全局代谢状态，因此本研究结果更应被理解为系统层面的分子表型，而不宜直接归因于某一特定细胞类型。第三，跨组织迁移分析仅属探索性研究，不能替代独立外部验证。最后，假基因与亲本基因的映射仅为功能解释提供背景支持，并不能直接证明调控因果关系。尽管如此，两套队列在翻译与生物能量学相关过程上的一致收敛，仍支持 NAFLD 与肌少症之间可能存在共享的分子应激表型，并提示其值得进一步验证[34][35]。

5. 结论

本研究通过跨组织整合转录组分析，在肝脏和骨骼肌中识别出共享的 NAFLD-肌少症共病信号。该信号主要涉及翻译 - 生物能量学应激轴，可由四基因核心特征概括，并伴随组织依赖性的自噬相关差异。上述结果提示 NAFLD 与肌少症之间可能存在保守的跨组织分子表型，为后续机制研究和独立队列验证提供了依据。

参考文献

- [1] Younossi, Z.M. (2019) Non-Alcoholic Fatty Liver Disease—A Global Public Health Perspective. *Journal of Hepatology*, **70**, 531-544. <https://doi.org/10.1016/j.jhep.2018.10.033>
- [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. *Hepatology*, **77**, 1335-1347. <https://doi.org/10.1097/hep.0000000000000004>
- [3] Eslam, M., Newsome, P.N., Sarin, S.K., Anstee, Q.M., Targher, G., Romero-Gomez, M., *et al.* (2020) A New Definition for Metabolic Dysfunction-Associated Fatty Liver Disease: An International Expert Consensus Statement. *Journal of Hepatology*, **73**, 202-209. <https://doi.org/10.1016/j.jhep.2020.03.039>
- [4] Sayer, A.A. and Cruz-Jentoft, A. (2022) Sarcopenia Definition, Diagnosis and Treatment: Consensus Is Growing. *Age and Ageing*, **51**, afac220. <https://doi.org/10.1093/ageing/afac220>
- [5] Cruz-Jentoft, A.J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., *et al.* (2019) Sarcopenia: Revised European Consensus on Definition and Diagnosis. *Age and Ageing*, **48**, 16-31. <https://doi.org/10.1093/ageing/afy169>
- [6] Wong, R. and Yuan, L. (2024) Sarcopenia and Metabolic Dysfunction Associated Steatotic Liver Disease: Time to Address Both. *World Journal of Hepatology*, **16**, 871-877. <https://doi.org/10.4254/wjh.v16.i6.871>

- [7] Li, A.A., Kim, D. and Ahmed, A. (2020) Association of Sarcopenia and NAFLD: An Overview. *Clinical Liver Disease*, **16**, 73-76. <https://doi.org/10.1002/cld.900>
- [8] Joo, S.K. and Kim, W. (2023) Interaction between Sarcopenia and Nonalcoholic Fatty Liver Disease. *Clinical and Molecular Hepatology*, **29**, S68-S78. <https://doi.org/10.3350/cmh.2022.0358>
- [9] Yuan, J., Zhang, J., Luo, Q. and Peng, L. (2024) Effects of Nonalcoholic Fatty Liver Disease on Sarcopenia: Evidence from Genetic Methods. *Scientific Reports*, **14**, Article No. 2709. <https://doi.org/10.1038/s41598-024-53112-1>
- [10] Chakravarthy, M.V., Siddiqui, M.S., Forsgren, M.F. and Sanyal, A.J. (2020) Harnessing Muscle-Liver Crosstalk to Treat Nonalcoholic Steatohepatitis. *Frontiers in Endocrinology*, **11**, Article ID: 592373. <https://doi.org/10.3389/fendo.2020.592373>
- [11] Urbina-Varela, R., Castillo, N., Videla, L.A. and del Campo, A. (2020) Impact of Mitophagy and Mitochondrial Unfolded Protein Response as New Adaptive Mechanisms Underlying Old Pathologies: Sarcopenia and Non-Alcoholic Fatty Liver Disease. *International Journal of Molecular Sciences*, **21**, Article No. 7704. <https://doi.org/10.3390/ijms21207704>
- [12] Grefhorst, A., van de Peppel, I.P., Larsen, L.E., Jonker, J.W. and Holleboom, A.G. (2021) The Role of Lipophagy in the Development and Treatment of Non-Alcoholic Fatty Liver Disease. *Frontiers in Endocrinology*, **11**, Article ID: 601627. <https://doi.org/10.3389/fendo.2020.601627>
- [13] Mastoridou, E.M., Goussia, A.C., Kanavarnos, P. and Charchanti, A.V. (2023) Involvement of Lipophagy and Chaperone-Mediated Autophagy in the Pathogenesis of Non-Alcoholic Fatty Liver Disease by Regulation of Lipid Droplets. *International Journal of Molecular Sciences*, **24**, Article No. 15891. <https://doi.org/10.3390/ijms242115891>
- [14] An, Y., Furber, K.L. and Ji, S. (2017) Pseudogenes Regulate Parental Gene Expression via ceRNA Network. *Journal of Cellular and Molecular Medicine*, **21**, 185-192. <https://doi.org/10.1111/jcmm.12952>
- [15] Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L. and Francisco Carter, D.R. (2011) Pseudogenes: Pseudo-Functional or Key Regulators in Health and Disease? *RNA*, **17**, 792-798. <https://doi.org/10.1261/rna.2658311>
- [16] Chen, X., Wan, L., Wang, W., Xi, W., Yang, A. and Wang, T. (2020) Re-Recognition of Pseudogenes: From Molecular to Clinical Applications. *Theranostics*, **10**, 1479-1499. <https://doi.org/10.7150/thno.40659>
- [17] Ala, U. (2020) Competing Endogenous RNAs, Non-Coding RNAs and Diseases: An Intertwined Story. *Cells*, **9**, Article No. 1574. <https://doi.org/10.3390/cells9071574>
- [18] Karreth, F.A., Reschke, M., Ruocco, A., Ng, C., Chapuy, B., Léopold, V., *et al.* (2015) The BRAF Pseudogene Functions as a Competitive Endogenous RNA and Induces Lymphoma *in Vivo*. *Cell*, **161**, 319-332. <https://doi.org/10.1016/j.cell.2015.02.043>
- [19] Langfelder, P. and Horvath, S. (2008) WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics*, **9**, Article No. 559. <https://doi.org/10.1186/1471-2105-9-559>
- [20] Langfelder, P., Zhang, B. and Horvath, S. (2007) Defining Clusters from a Hierarchical Cluster Tree: The Dynamic Tree Cut Package for R. *Bioinformatics*, **24**, 719-720. <https://doi.org/10.1093/bioinformatics/btm563>
- [21] Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., *et al.* (2015) Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Research*, **43**, e47. <https://doi.org/10.1093/nar/gkv007>
- [22] Chen, Y. and Tian, Z. (2020) Roles of Hepatic Innate and Innate-Like Lymphocytes in Nonalcoholic Steatohepatitis. *Frontiers in Immunology*, **11**, Article No. 1500. <https://doi.org/10.3389/fimmu.2020.01500>
- [23] Li, X., Chen, W., Jia, Z., Xiao, Y., Shi, A. and Ma, X. (2025) Mitochondrial Dysfunction as a Pathogenesis and Therapeutic Strategy for Metabolic-Dysfunction-Associated Steatotic Liver Disease. *International Journal of Molecular Sciences*, **26**, Article No. 4256. <https://doi.org/10.3390/ijms26094256>
- [24] Arrese, M., Cabello-Verrugio, C., Arab, J.P., Barrera, F., Baudrand, R., Guarda, F.J., *et al.* (2022) Sarcopenia in the Setting of Nonalcoholic Fatty Liver. *Metabolism and Target Organ Damage*, **2**, 2. <https://doi.org/10.20517/mtod.2021.16>
- [25] Liu, Y., Zhang, M. and Wang, Y. (2025) Induction of Autophagy as a Therapeutic Breakthrough for NAFLD: Current Evidence and Perspectives. *Biology*, **14**, Article No. 989. <https://doi.org/10.3390/biology14080989>
- [26] Khambu, B., Yan, S., Huda, N., Liu, G. and Yin, X. (2018) Autophagy in Non-Alcoholic Fatty Liver Disease and Alcoholic Liver Disease. *Liver Research*, **2**, 112-119. <https://doi.org/10.1016/j.livres.2018.09.004>
- [27] Pérez-Carreras, M., Del Hoyo, P., Martín, M.A., Rubio, J.C., Martín, A., Castellano, G., *et al.* (2003) Defective Hepatic Mitochondrial Respiratory Chain in Patients with Nonalcoholic Steatohepatitis. *Hepatology*, **38**, 999-1007. <https://doi.org/10.1002/hep.1840380426>
- [28] Traussnigg, S., Kienbacher, C., Gajdošík, M., Valkovič, L., Halilbasic, E., Stift, J., *et al.* (2017) Ultra-High-Field Magnetic Resonance Spectroscopy in Non-Alcoholic Fatty Liver Disease: Novel Mechanistic and Diagnostic Insights of

- Energy Metabolism in Non-Alcoholic Steatohepatitis and Advanced Fibrosis. *Liver International*, **37**, 1544-1553. <https://doi.org/10.1111/liv.13451>
- [29] García-Ruiz, C. and Fernández-Checa, J.C. (2018) Mitochondrial Oxidative Stress and Antioxidants Balance in Fatty Liver Disease. *Hepatology Communications*, **2**, 1425-1439. <https://doi.org/10.1002/hep4.1271>
- [30] Maria Barbalho, S., Laurindo, L.F., Valenti, V.E., Méndez-Sánchez, N., Ramírez-Mejía, M.M. and Goulart, R.d.A. (2025) Organokine-Mediated Crosstalk: A Systems Biology Perspective on the Pathogenesis of MASLD—A Narrative Review. *International Journal of Molecular Sciences*, **26**, Article No. 11547. <https://doi.org/10.3390/ijms262311547>
- [31] Jin, M., Yu, H., Deng, Q., Wang, Z. and Liang, H. (2025) AMPK Affects the Development of Early-Stage NAFLD by Activating Autophagy and Fatty Acid Oxidation. *Scientific Reports*, **16**, Article No. 1425. <https://doi.org/10.1038/s41598-025-31181-0>
- [32] Polisenò, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J. and Pandolfi, P.P. (2010) A Coding-Independent Function of Gene and Pseudogene mRNAs Regulates Tumour Biology. *Nature*, **465**, 1033-1038. <https://doi.org/10.1038/nature09144>
- [33] Liang, R., Han, B., Li, Q., Yuan, Y., Li, J. and Sun, D. (2017) Using RNA Sequencing to Identify Putative Competing Endogenous RNAs (ceRNAs) Potentially Regulating Fat Metabolism in Bovine Liver. *Scientific Reports*, **7**, Article No. 6396. <https://doi.org/10.1038/s41598-017-06634-w>
- [34] Severinsen, M.C.K. and Pedersen, B.K. (2020) Muscle-Organ Crosstalk: The Emerging Roles of Myokines. *Endocrine Reviews*, **41**, 594-609. <https://doi.org/10.1210/endrev/bnaa016>
- [35] He, Y., Chen, Y., Qian, S., van Der Merwe, S., Dhar, D., Brenner, D.A., *et al.* (2025) Immunopathogenic Mechanisms and Immunoregulatory Therapies in MASLD. *Cellular & Molecular Immunology*, **22**, 1159-1177. <https://doi.org/10.1038/s41423-025-01307-5>