

分子标志物的挖掘与鉴定对糖尿病健康管理的研究

马 林*, 巩建敏

首都医科大学附属北京康复医院劳模健康管理中心, 北京

收稿日期: 2025年9月27日; 录用日期: 2025年10月21日; 发布日期: 2025年10月27日

摘 要

背景: 糖尿病视网膜病变(DR)是糖尿病最主要的致盲性并发症, 其早期诊断困难且缺乏有效的干预靶点, 构成了全球性的健康管理挑战。前体mRNA加工因子4B (PRPF4B)在其他糖尿病并发症中扮演关键角色, 但其在DR中的作用尚不明确。方法: 本研究采用生物信息学分析方法, 整合GEO数据库中的GSE185011和GSE146615两个DR数据集。通过批次效应校正、差异表达基因(DEGs)筛选、加权基因共表达网络分析(WGCNA)、蛋白质-蛋白质相互作用(PPI)网络构建以及功能富集分析, 筛选DR的核心枢纽基因。并利用CTD数据库和TargetScan平台分析核心基因的疾病关联及上游调控机制。结果: 共鉴定出1598个DEGs。WGCNA与PPI网络分析联合筛选出9个核心枢纽基因, 其中PRPF4B在DR组织中显著高表达。功能分析表明, PRPF4B与炎症反应、纤维化等DR核心病理过程密切相关。CTD分析进一步证实其与肾脏疾病、纤维化等并发症相关。TargetScan预测其表达可能受hsa-miR-139-5p调控。结论: PRPF4B是DR中一个关键的高表达基因, 可能通过调控剪接过程介导炎症和纤维化, 从而推动疾病进展。PRPF4B作为一个潜在的生物标志物, 其在DR诊断、预后和治疗中的潜力值得进一步研究。

关键词

生物信息学分析, 蛋白质相互作用网络, 早期诊断, 精准健康管理

Mining and Identification of Molecular Markers for Diabetes Health Management

Lin Ma*, Jianmin Gong

Labor Model Health Management Center, Beijing Rehabilitation Hospital Affiliated to Capital Medical University, Beijing

Received: September 27, 2025; accepted: October 21, 2025; published: October 27, 2025

*通讯作者。

文章引用: 马林, 巩建敏. 分子标志物的挖掘与鉴定对糖尿病健康管理的研究[J]. 临床医学进展, 2025, 15(10): 2378-2395. DOI: 10.12677/acm.2025.15103024

Abstract

Background: Diabetic retinopathy (DR) is the leading cause of blindness in patients with diabetes mellitus. The difficulty in early diagnosis and the lack of effective intervention targets pose a global health challenge. Precursor mRNA processing factor 4B (PRPF4B) plays a key role in other diabetic complications, but its role in DR remains unclear. **Methods:** Two DR Datasets (GSE185011 and GSE146615) from GEO database were integrated using bioinformatics analysis. The core hub genes of DR were screened by batch effect correction, differentially expressed genes (DEGs) screening, weighted gene co-expression network analysis (WGCNA), protein-protein interaction (PPI) network construction and functional enrichment analysis. The CTD database and TargetScan platform were used to analyze the disease association and upstream regulatory mechanism of core genes. **Results:** A total of 1598 DEGs were identified. WGCNA combined with PPI network analysis screened out 9 core hub genes, among which PRPF4B was significantly highly expressed in DR Tissues. Functional analysis showed that PRPF4B was closely related to inflammatory response, fibrosis and other core pathological processes of DR. CTD analysis further confirmed its association with complications such as kidney disease and fibrosis. TargetScan predicted that its expression may be regulated by hsa-miR-139-5p. **Conclusions:** PRPF4B is a key highly expressed gene in DR, which may mediate inflammation and fibrosis by regulating the splicing process, thereby promoting the progression of DR. As a potential biomarker, PRPF4B deserves further investigation for its potential in the diagnosis, prognosis and treatment of DR.

Keywords

Bioinformatics Analysis, Protein Interaction Network, Early Diagnosis, Precision Health Management

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/>



Open Access

1. 引言

糖尿病(DM)作为一种以持续性高血糖为特征的慢性代谢性疾病,已构成全球性公共卫生危机。其中,2型糖尿病(T2DM)约占病例总数的90%以上,其核心病理机制涉及胰岛素抵抗或分泌缺陷[1][2]。2024年流行病学数据显示,全球成年糖尿病患者数量已达5.3亿(530 million),预计至2045年将攀升至7亿[3]。该疾病具有隐匿进展特征,早期常无显著临床症状,当出现多尿、视力模糊等典型症状时,多已伴发不可逆性器官损伤[4][5]。

糖尿病视网膜病变(DR)作为最严重的微血管并发症之一,是劳动年龄人群首位致盲性疾病[6]。约30%~50%的糖尿病患者在确诊后10~15年内发生DR,早期非增殖期(NPDR)以视网膜微血管瘤为病理标志;若进展至增殖期(PDR)或糖尿病性黄斑水肿(DME),则可因病理性新生血管形成及纤维增生导致不可逆性视力丧失[7]。当前DR健康管理策略面临双重挑战:其一,传统筛查手段难以识别亚临床期视网膜损伤,导致约40%的早期患者漏诊[8];其二,病情评估多依赖主观指标,临床干预仅能在晚期实施对症治疗,无法逆转早期结构性损伤[9]。

糖尿病病因尚未完全阐明,现有证据表明其为遗传易感性与环境因素共同作用的结果。T2DM遗传度高达70%~80%,全基因组研究已鉴定超过100个风险位点[10],而染色体异常及表观遗传修饰等因素

亦可增加并发症发生风险[11]。深入解析 DR 分子机制并筛选关键调控基因, 对推动糖尿病管理范式从“被动血糖控制”向“主动健康管理”转型具有关键意义。

生物信息学技术为复杂疾病研究提供了突破性工具, 可整合基因表达汇编(GEO)等公共数据库的大规模组学数据, 通过批次效应校正、加权基因共表达网络分析(WGCNA)、蛋白质互作(PPI)网络构建等技术筛选核心分子[12]。尽管已有研究筛选出 DR 差异表达基因谱, 其向健康管理工具的转化仍存在显著缺口。

前体 mRNA 加工因子 4B (PRPF4B)编码剪接体核心组分, 参与调控 pre-mRNA 剪接过程。其功能异常与多种疾病相关, 研究表明该基因高表达与糖尿病心肌病、糖尿病肾病中的炎症反应及纤维化进程密切相关[13][14]。然而, PRPF4B 在 DR 中的调控作用及其健康管理价值尚未明确。

本研究拟采用生物信息学分析的策略, 系统阐明 PRPF4B 在 DR 发病机制中的作用: 整合 GSE185011 与 GSE146615 数据集筛选核心基因; 通过基因本体(GO)及京都基因与基因组百科全书(KEGG)通路富集分析解析其生物学功能; 验证 PRPF4B 在 DR 中的差异表达模式; 利用细胞模型验证其对 DR 关键病理表型的调控效应。

2. 实验材料与方法

2.1. 糖尿病视网膜病变数据集

本研究从基因表达综合数据库(Gene Expression Omnibus, GEO; <http://www.ncbi.nlm.nih.gov/geo/>)下载了基于 GPL24676 和 GPL10558 平台生成的糖尿病视网膜病变相关数据集 GSE185011 和 GSE146615。GSE185011 数据集包含 5 例糖尿病视网膜病变样本及 5 例正常对照样本, GSE146615 数据集包含 44 例糖尿病视网膜病变样本及 20 例正常对照样本。这些数据集用于识别糖尿病视网膜病变相关的差异表达基因(Differentially Expressed Genes, DEGs)。

2.2. 批次效应校正

为整合 GSE185011 与 GSE146615 数据集并进行批次效应校正, 首先利用 R 语言环境中的 inSilicoMerging 软件包(<https://doi.org/10.1186/1471-2105-13-335>)对数据进行合并, 生成合并表达矩阵。随后, 应用 limma 软件包(版本 3.42.2)中的“removeBatchEffect”函数对合并矩阵进行批次效应校正, 最终获得校正后的表达矩阵, 并将其用于后续分析。

2.3. 差异表达基因筛选

采用 R 包“limma”对合并校正后的表达矩阵进行探针注释整合及背景校正。使用 Benjamini-Hochberg 方法校正原始 p 值。差异倍数(Fold Change, FC)的计算基于错误发现率(False Discovery Rate, FDR)。设定 DEGs 的筛选阈值为校正后 p 值(adj. p value) < 0.05 且 $|\log_2 FC| > \log_2(1.5)$ 。筛选结果通过火山图进行可视化展示。

2.4. 加权基因共表达网络分析(WGCNA)

首先, 基于基因表达谱计算各基因的中位绝对偏差(Median Absolute Deviation, MAD), 并剔除 MAD 值最低的 50%基因。随后, 利用 WGCNA R 包中的“goodSamplesGenes”函数移除离群基因及样本。进而, 构建无标度共表达网络(scale-free co-expression network)。具体流程如下: 计算所有基因对的皮尔逊相关系数矩阵, 并采用平均连锁法进行层次聚类。利用幂函数法则构建加权邻接矩阵: $A_{mn} = |C_{mn}|^{\beta}$, 其中 C_{mn} 代表基因 m 与基因 n 之间的皮尔逊相关系数, A_{mn} 代表基因 m 与基因 n 之间的邻接强度。软阈值参数 β 用于强化强相关关系并弱化弱相

关关系。选定 β 值为 12 后, 将邻接矩阵转换为拓扑重叠矩阵(Topological Overlap Matrix, TOM), 其值度量基因的网络连通性(定义为该基因与网络中所有其他基因邻接强度的总和)。进而计算拓扑重叠相异度(1-TOM)。为将表达模式相似的基因聚类至同一模块, 基于 TOM 相异度进行平均连锁层次聚类, 设定基因树状图的最小模块大小(基因数量)为 30。设置合并敏感度参数(deepSplit)为 3。为优化模块划分, 计算模块特征基因(Module Eigengene, ME)的相异度, 据此确定模块树状图的切割高度, 合并部分模块。此外, 将相异度小于 0.25 的模块进行合并。需注意, “grey” 模块包含未能被归类至任何特定模块的基因。

2.5. 功能富集分析

采用基因本体论(Gene Ontology, GO)和京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)通路分析评估基因功能及生物学通路。本研究首先通过 KEGG REST API (<https://www.kegg.jp/kegg/rest/keggapi.html>)获取最新的 KEGG 通路基因注释信息作为背景集。将筛选得到的差异基因列表映射至该背景集后, 利用 clusterProfiler R 包(版本 3.14.3)进行富集分析, 以获得基因集富集结果。同时, 使用 org.Hs.eg.db R 包(版本 3.1.0)提供的基因 GO 注释作为背景集进行 GO 富集分析。设定最小基因集大小为 5, 最大基因集大小为 5000。富集显著性判定标准为 p 值 < 0.05 且 $FDR < 0.25$ 。

此外, 利用 Metascape 数据库(<http://metascape.org/gp/index.html>)对上述差异基因列表进行功能富集分析, 该数据库提供全面的基因注释、分析资源及可视化功能, 结果可导出。

2.6. 基因集富集分析(GSEA)

进行基因集富集分析(Gene Set Enrichment Analysis, GSEA)时, 从 GSEA 官网 (<https://doi.org/10.1073/pnas.0506580102>, <http://software.broadinstitute.org/gsea/index.jsp>)获取 GSEA 软件(版本 3.0)。依据样本表型(糖尿病视网膜病变 vs. 正常对照)将样本分为两组。从分子特征数据库(Molecular Signatures Database (MSigDB; <https://doi.org/10.1093/bioinformatics/btr260>, <http://www.gsea-msigdb.org/gsea/downloads.jsp>)下载基因集文件 c2.cp.kegg.v7.4.symbols.gmt。基于基因表达谱及表型分组信息, 利用 GSEA 软件评估相关通路及分子机制。分析参数设定如下: 最小基因集大小 5, 最大基因集大小 5000, 置换次数(permutation)为 1000 次。显著性判定标准为 p 值 < 0.05 且 $FDR < 0.25$ 。同时, 通过 GSEA 对全基因组进行了 GO 与 KEGG 富集分析。

2.7. 蛋白质 - 蛋白质相互作用(PPI)网络的构建与分析

STRING 数据库(<http://string-db.org/>)旨在收集、评估并整合所有公开可用的蛋白质 - 蛋白质相互作用信息, 并通过计算预测进行补充。本研究将差异基因列表输入 STRING 数据库, 构建置信度阈值 > 0.4 的预测性 PPI 网络, 用于识别核心基因。

Cytoscape 软件为生物网络分析及二维(2D)可视化提供了强大工具。本研究利用 Cytoscape 对 STRING 生成的 PPI 网络进行可视化及核心基因预测。具体步骤包括: 将 PPI 网络导入 Cytoscape; 使用 MCODE 插件识别网络中最紧密连接的模块; 分别应用 MCC (Maximal Clique Centrality)和 MNC (Maximum Neighborhood Component)两种算法计算网络中连通性最高的基因, 并取其并集作为核心基因候选集; 最终可视化并导出核心基因列表。

2.8. 基因表达热图

利用 R 包 “heatmap”, 对通过上述两种算法(MCC, MNC)筛选出的核心基因, 在 GSE185011 与 GSE146615 合并并经批次效应校正后的表达矩阵中的表达水平进行可视化, 生成热图, 以展示核心基因在糖尿病视网膜病变组与正常对照组间的表达差异。

2.9. CTD 数据库分析

比较毒物基因组学数据库(Comparative Toxicogenomics Database, CTD)整合了大量化学物质、基因、功能表型及疾病间的相互作用数据,为研究疾病相关环境暴露因素及潜在药物作用机制提供了重要资源。本研究将核心基因列表输入 CTD 数据库(<http://ctdbase.org/>),检索与各核心基因最显著相关的疾病,并利用 Excel 软件绘制各核心基因表达差异的雷达图。

2.10. miRNA 靶基因预测

TargetScan (www.targetscan.org)是一个在线数据库,用于预测 miRNA 与靶基因的调控关系。本研究利用 TargetScan 筛选可能调控核心差异表达基因(中枢 DEGs)的 miRNA。

3. 结果

3.1. DEGs 的筛选

在本研究中,根据 GSE185011 和 GSE146615 的去批次化合并矩阵中鉴定 DEG,共鉴定出 1598 个 DEGs (图 1)。

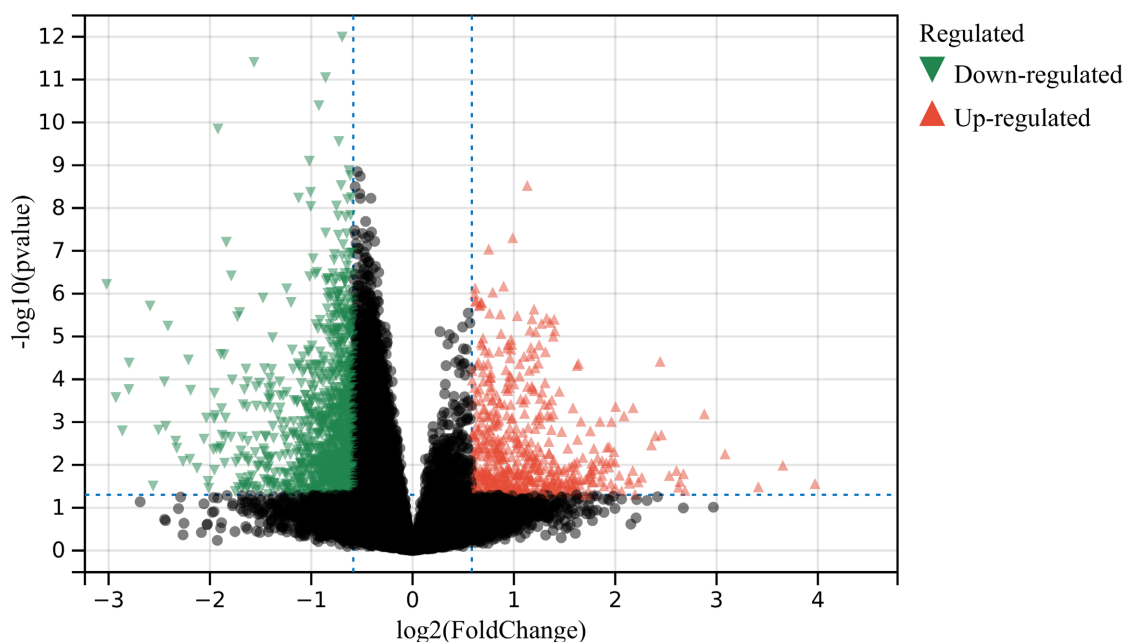


Figure 1. Screening of DEGs. A total of 1598 DEGs were identified

图 1. DEGs 的筛选。共鉴定出 1598 个 DEGs

3.2. 功能富集分析

3.2.1. DEGs 功能富集分析

我们对这些差异表达基因进行 GO 和 KEGG 分析,根据 GObp 分析,它们主要富集在分解代谢过程、细胞分解代谢过程、T 细胞活化、建立对细胞器的蛋白质定位(图 2(A))。根据 GOcc 分析,它们主要富集在质膜部分、细胞投影、质膜结合的细胞投射、全膜(图 2(C))。根据 GOMf 分析,它们主要富集在催化活性、阳离子结合、金属离子结合、氧化还原酶活性(图 2(E))。KEGG 分析结果显示,靶基因主要富集于类风湿性关节炎、炎症性肠病、I 型糖尿病、同种异体移植排斥反应(图 2(G))。

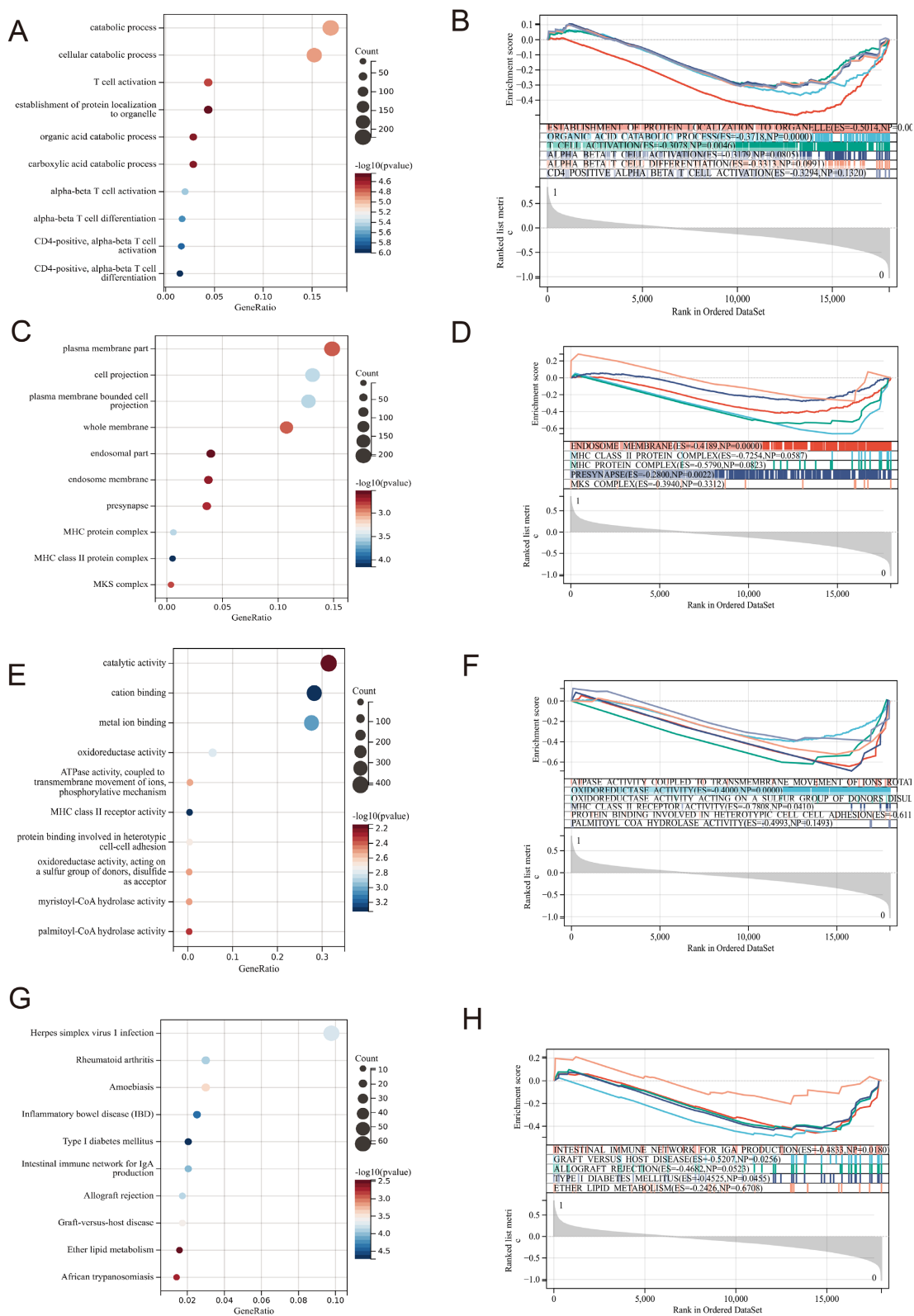


Figure 2. Functional enrichment analysis. (A) GOBP analysis. (C) GOCC analysis. (E) GOMF analysis. (G) KEGG analysis. (B), (D), (F), (H) GSEA analysis

图 2. 功能富集分析。(A) GOBP 分析。(C) GOCC 分析。(E) GOMF 分析。(G) KEGG 分析。(B)、(D)、(F)、(H) GSEA 分析

3.2.2. GSEA 分析

另外, 我们对全基因组进行 GSEA 富集分析, 旨在寻找非差异表达基因中可能存在的富集项目, 并验证差异表达基因的结果。富集项与差异表达基因的 GO KEGG 富集项的交集如图所示, 主要富集在同种异体移植排斥反应、I 型糖尿病、乙醚脂代谢(图 2(B)、图 2(D)、图 2(F)、图 2(H))。

3.2.3. Metascape 富集分析

在 Metascape 的富集项目中, GO 有细胞粘附的调节、循环系统过程、正向调节 T 细胞介导的免疫、有机磷的分解代谢过程(图 3(A)), 同时我们还输出了以富集项着色和 p 值着色的富集网络(图 3(B)、图 3(C)), 可视化表示各富集项目的关联和置信度。

3.3. WGCNA 分析

软阈值功率的选择是 WGCNA 分析中的重要一步。进行网络拓扑分析以确定软阈值功率。WGCNA 分析中的软阈值功率设置为 12 (图 4(A), 图 4(B))。构建了所有基因的层次聚类树, 并生成了重要模块(图 4(D))。然后分析这些模块之间的交互(图 4(C)), 并生成了模块与表型相关性热图(图 5(A))和相关核心基因的 GS 与 MM 相关性散点图(图 5(B)~(E)、图 6(A)~(D))。

我们合并了距离小于 0.25 的模块, 最终获得了 20 个共表达模块, 计算了模块特征向量与基因的表达相关性以获得 MM, 基于截止标准(|MM| > 0.8), 754 个在临床显著模块中具有高连通性的基因被鉴定为枢纽基因。通过 WGCNA 与 DEGs 筛选出的差异基因绘制了韦恩图并取交集, 用于蛋白质相互作用网络的创建和分析(图 6(E))。

3.4. 蛋白质 - 蛋白质相互作用(PPI)网络的构建与分析

DEGs 的 PPI 网络是由 STRING 在线数据库构建并由 Cytoscape 软件分析(图 7(A))采用两种不同的算法识别中枢基因(图 7(B)、图 7(C)), 绘制了韦恩图并取交集(图 7(D)), 获得了 9 个核心基因(RPL36A, SF3A1, EIF5B, TMEM147, PHF5A, MFAP1, RPL22, PRPF4B)。

3.5. 基因表达量热图

我们得到了核心基因在糖尿病视网膜病变和正常样本间的可视化表达差异热图, 我们发现 4 个核心基因(RPL36A, SF3A1, EIF5B, TMEM147)在糖尿病视网膜病变样本中低表达, 在正常样本中高表达; PRPF4B 基因在糖尿病视网膜病变样本中高表达, 在正常样本中低表达; RPL7A、RPL22、PHF5A、MFAP1 基因在糖尿病视网膜病变样本中高表达, 推测其可能对于糖尿病视网膜病变具有调控作用(图 8)。

3.6. CTD 分析

在这项研究中, 我们将核心基因列表输入到 CTD 网站中寻找与核心基因有关的疾病, 提高了对基因与疾病关联的理解。发现 9 个基因(RPL36A, SF3A1, EIF5B, TMEM147, PHF5A, MFAP1, RPL22, PRPF4B)和炎症、坏死、肝肿大、纤维化、肾脏疾病有关(图 9)。

3.7. 与核心基因相关的 miRNA 预测与功能注释

在这项研究中, 我们将核心基因列表输入到 targetsacan 中寻找相关的 miRNA, 提高对基因表达调控的理解(表 1)。我们发现 RPL36A 基因的相关 mirna 是 hsa-miR-325-3p; SF3A1 基因的相关 mirna 是 hsa-miR-3139、hsa-miR-28-5p、hsa-miR-708-5p; EIF5B 基因的相关 mirna 是 hsa-miR-199b-5p、hsa-miR-199a-5p; TMEM147 基因的相关 mirna 是 hsa-miR-452-5p、hsa-miR-4676-3p、hsa-miR-892c-3p; PHF5A 基因的相关 mirna 是 hsa-miR-107、hsa-miR-103a-3p; MFAP1 基因的相关 mirna 是 hsa-miR-302c-3p、hsa-miR-520f-3p;

RPL22 基因的相关 mirna 是 hsa-miR-153-3p; PRPF4B 基因的相关 mirna 是 hsa-miR-139-5p。

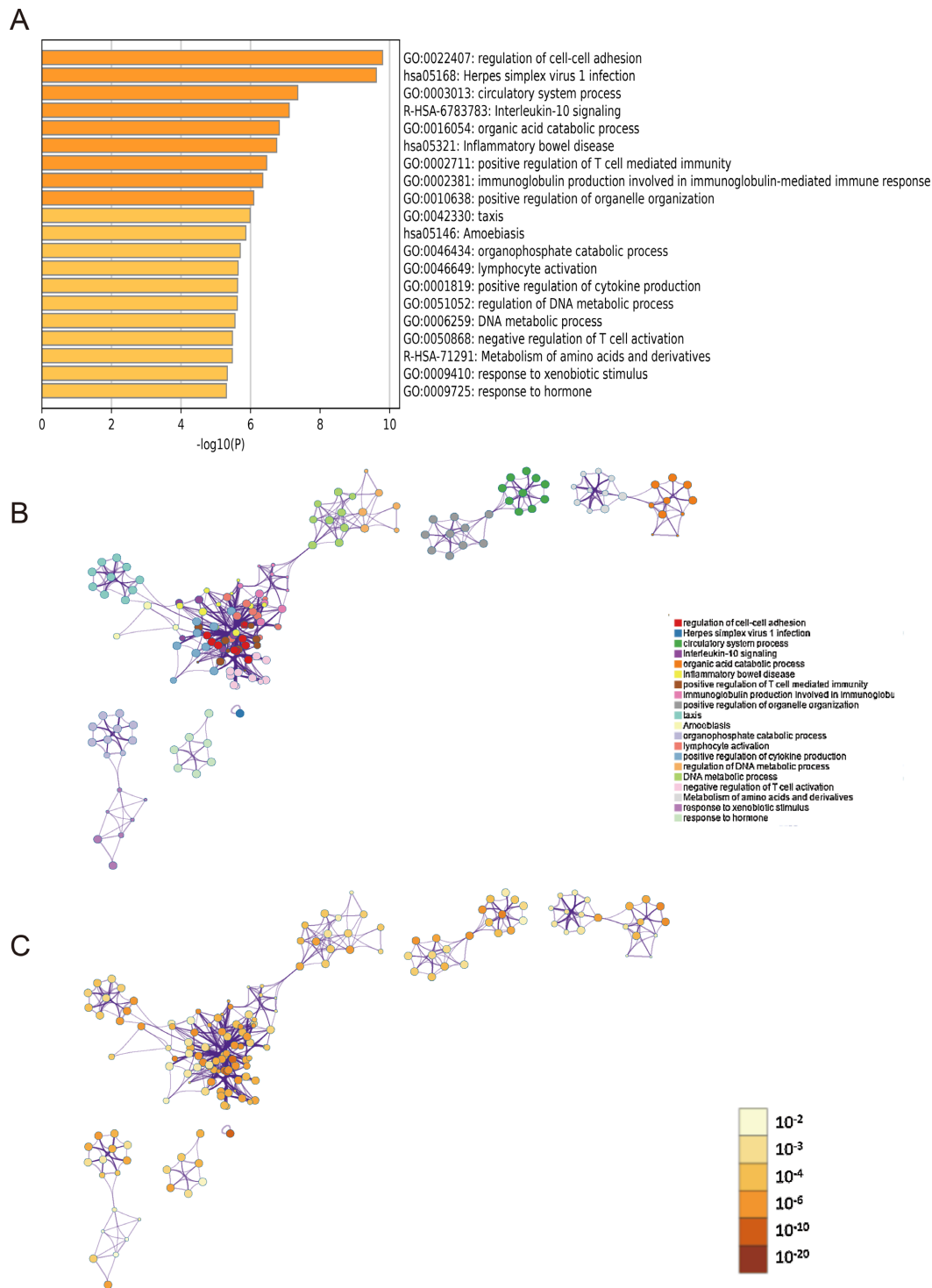


Figure 3. Enrichment analysis by Metascape. (A) In the enrichment project of metascape, GO had the regulation of cell adhesion, circulatory system process, positive regulation of T cell-mediated immunity, and the catabolic process of organophosphorus. (B) Enrichment networks colored by enrichment terms. (C) Enrichment network colored by p-value

图 3. Metascape 富集分析。(A) 在 metascape 的富集项目中, GO 有细胞粘附的调节、循环系统过程、正向调节 T 细胞介导的免疫、有机磷的分解代谢过程。(B) 以富集项着色的富集网络。(C) p 值着色的富集网络

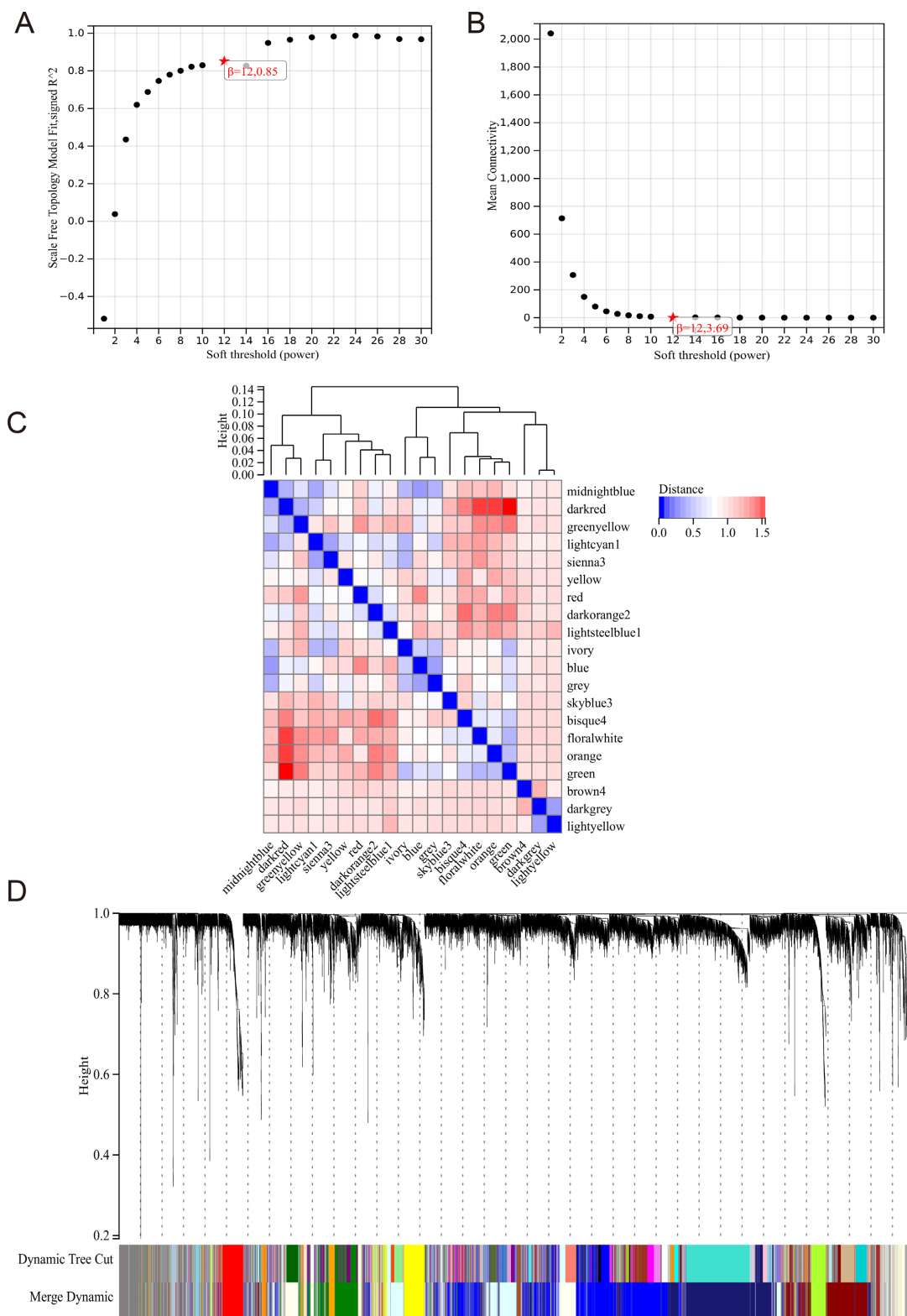


Figure 4. WGCNA analysis. (A) (B) The soft threshold power in WGCNA analysis was set to 12. (D) Hierarchical clustering trees of all genes were constructed and significant modules were generated. (C) Interaction between modules

图 4. WGCNA 分析。(A) (B) WGCNA 分析中的软阈值功率设置为 12。(D) 构建了所有基因的层次聚类树，并生成了重要模块。(C) 模块之间的交互

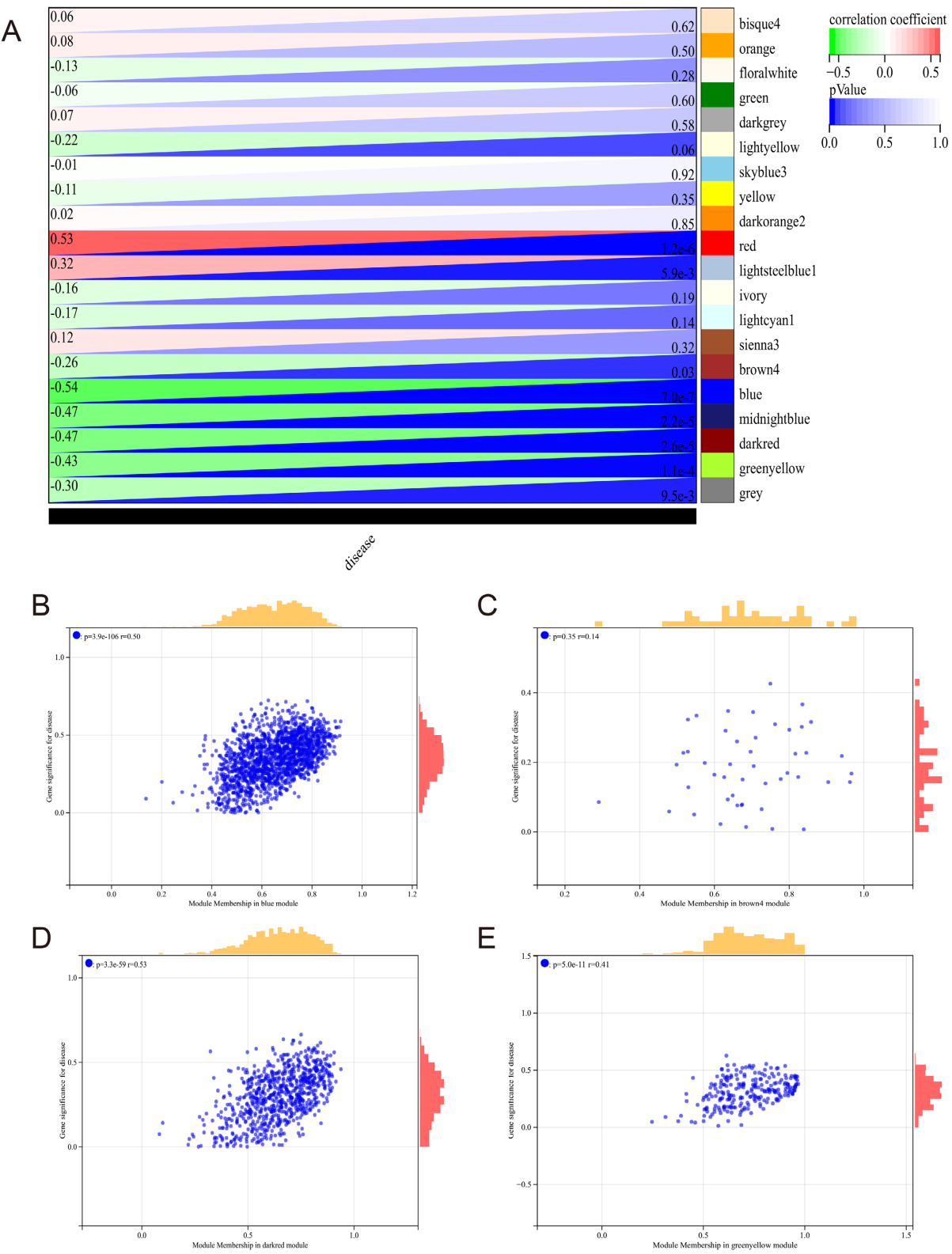


Figure 5. (A) Heatmap of module and phenotype correlation. (B)~(E) Scatter plot of correlation between GS and MM for related core genes

图 5. (A) 模块与表型相关性热图。 (B)~(E) 相关核心基因的 GS 与 MM 相关性散点图

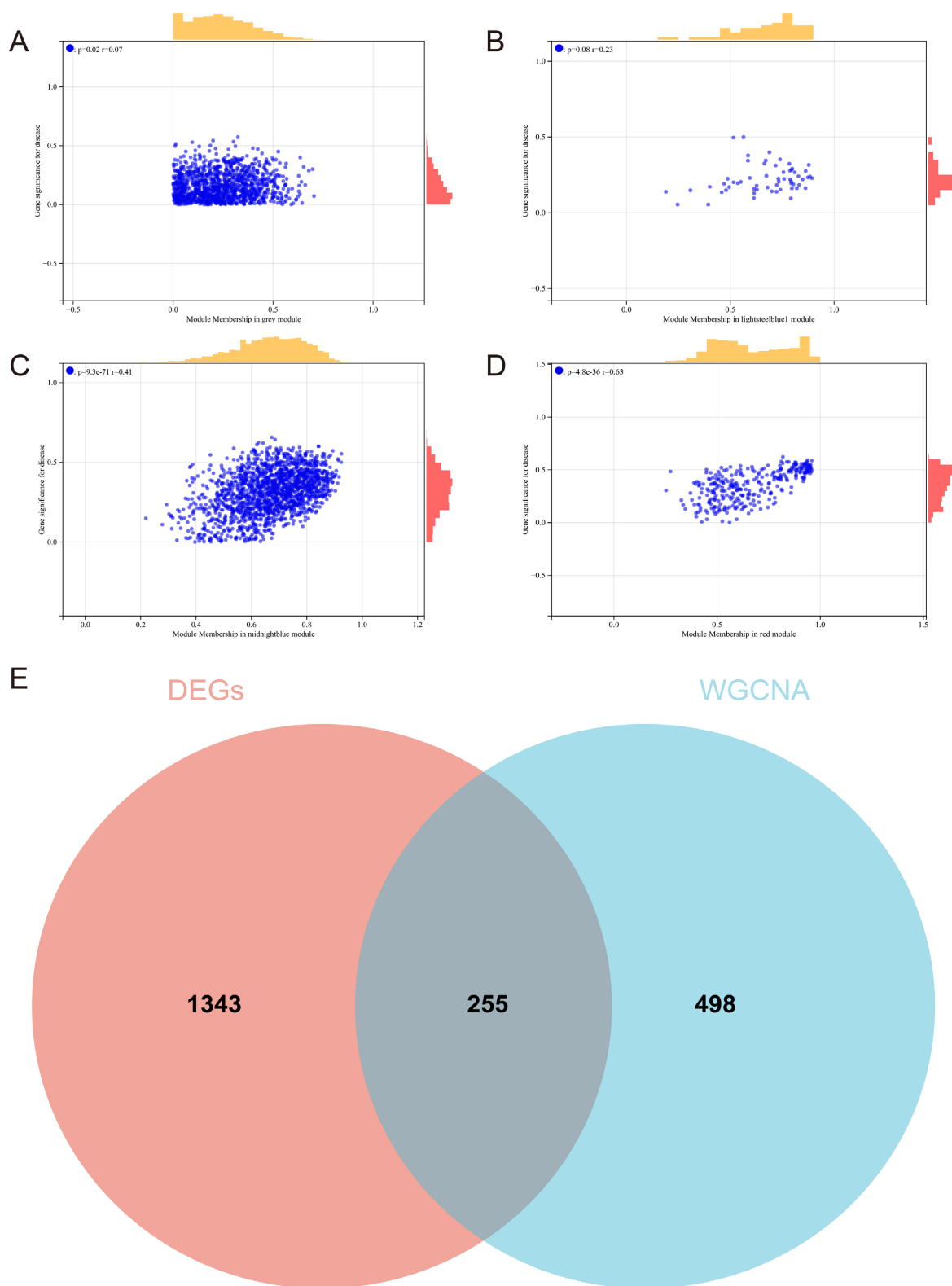


Figure 6. (A)~(D) Scatter plot of correlation between GS and MM of related core genes. (E) Venn diagram was drawn by WGCNA and differentially expressed genes screened by DEGs and the intersection was taken

图 6. (A)~(D) 相关核心基因的 GS 与 MM 相关性散点图。(E) 通过 WGCNA 与 DEGs 筛选出的差异基因绘制了韦恩图并取交集

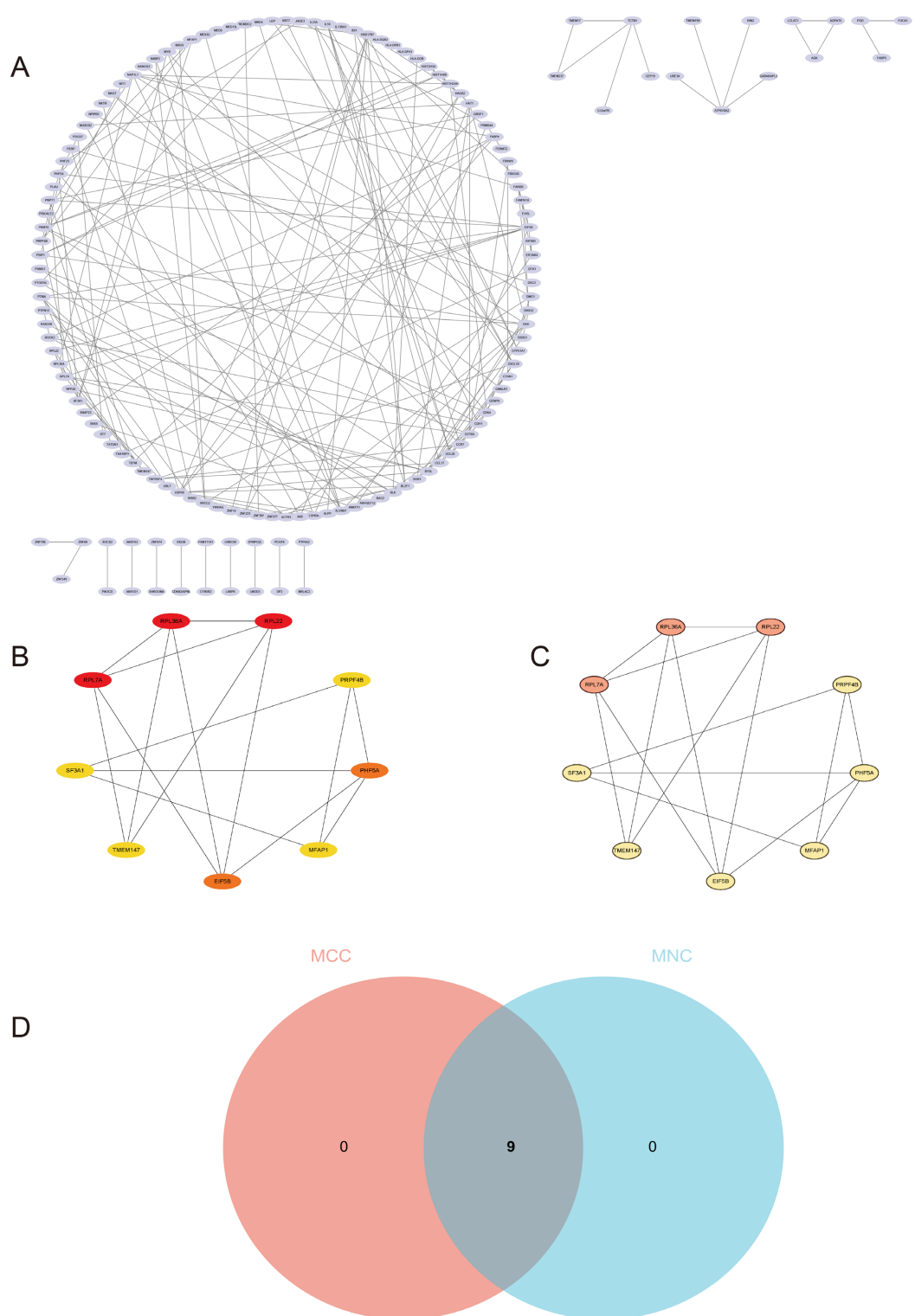


Figure 7. Construction and analysis of protein-protein interaction (PPI) networks. (A) PPI network of DEGs. (B) (C) Two different algorithms were used to identify the hub genes. (D) Venn diagram was drawn and intersection was taken, and 9 core genes (RPL36A, SF3A1, EIF5B, TMEM147, PHF5A, MFAP1, RPL22, PRPF4B) were obtained

图 7. 蛋白质 - 蛋白质相互作用(PPI)网络的构建与分析。(A) DEGs 的 PPI 网络。(B) (C) 采用两种不同的算法识别中枢基因。(D) 绘制了韦恩图并取交集, 获得了 9 个核心基因(RPL36A, SF3A1, EIF5B, TMEM147, PHF5A, MFAP1, RPL22, PRPF4B)

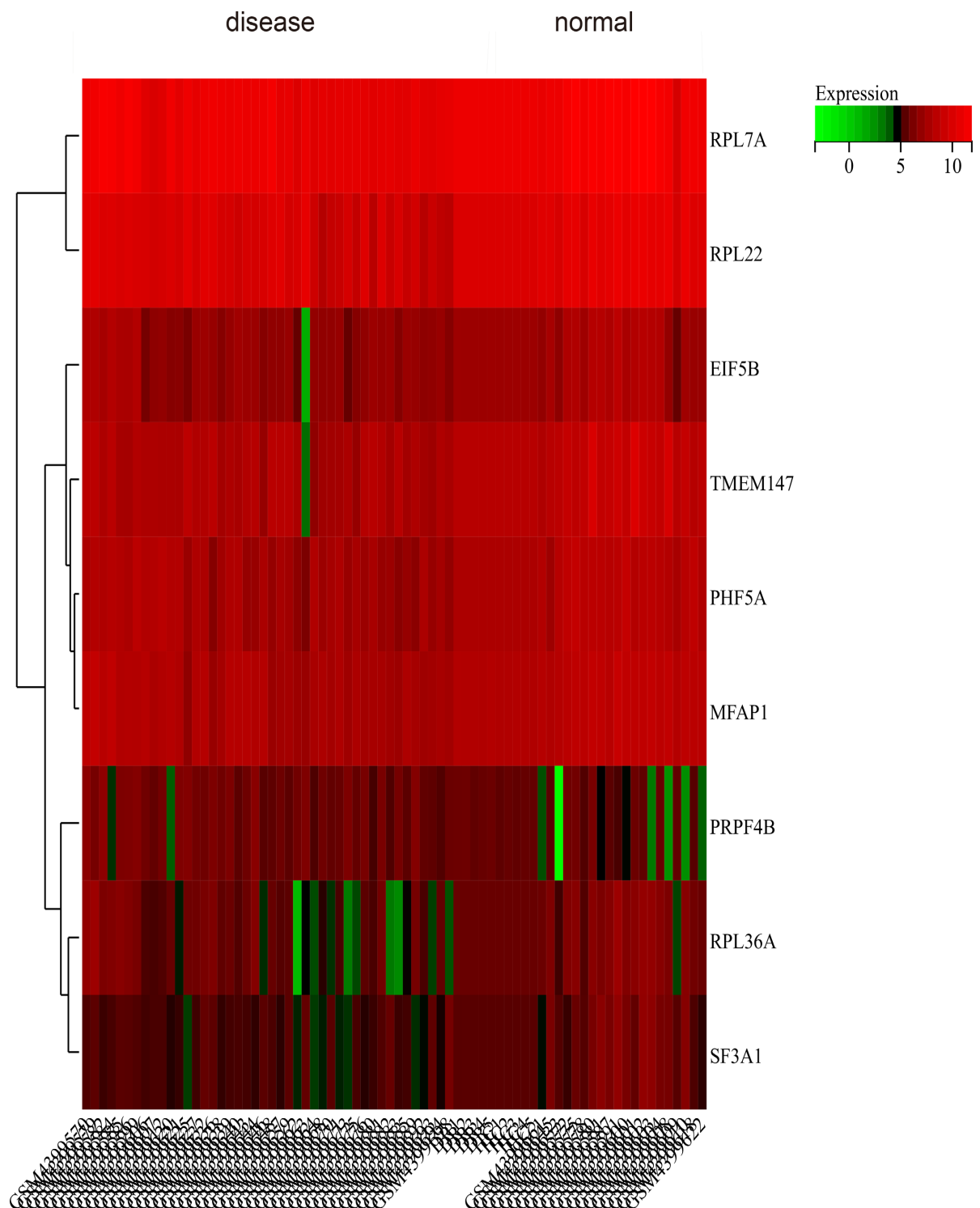


Figure 8. Heat map of gene expression. Four core genes (RPL36A, SF3A1, EIF5B and TMEM147) were lowly expressed in diabetic retinopathy samples but highly expressed in normal samples. PRPF4B was highly expressed in diabetic retinopathy samples and low expressed in normal samples. RPL7A, RPL22, PHF5A and MFAP1 genes were highly expressed in diabetic retinopathy samples

图 8. 基因表达量热图。4 个核心基因(RPL36A, SF3A1, EIF5B and TMEM147)在糖尿病视网膜病变样本中低表达, 在正常样本中高表达; PRPF4B 基因在糖尿病视网膜病变样本中高表达, 在正常样本中低表达; RPL7A、RPL22、PHF5A、MFAP1 基因在糖尿病视网膜病变样本中高表达

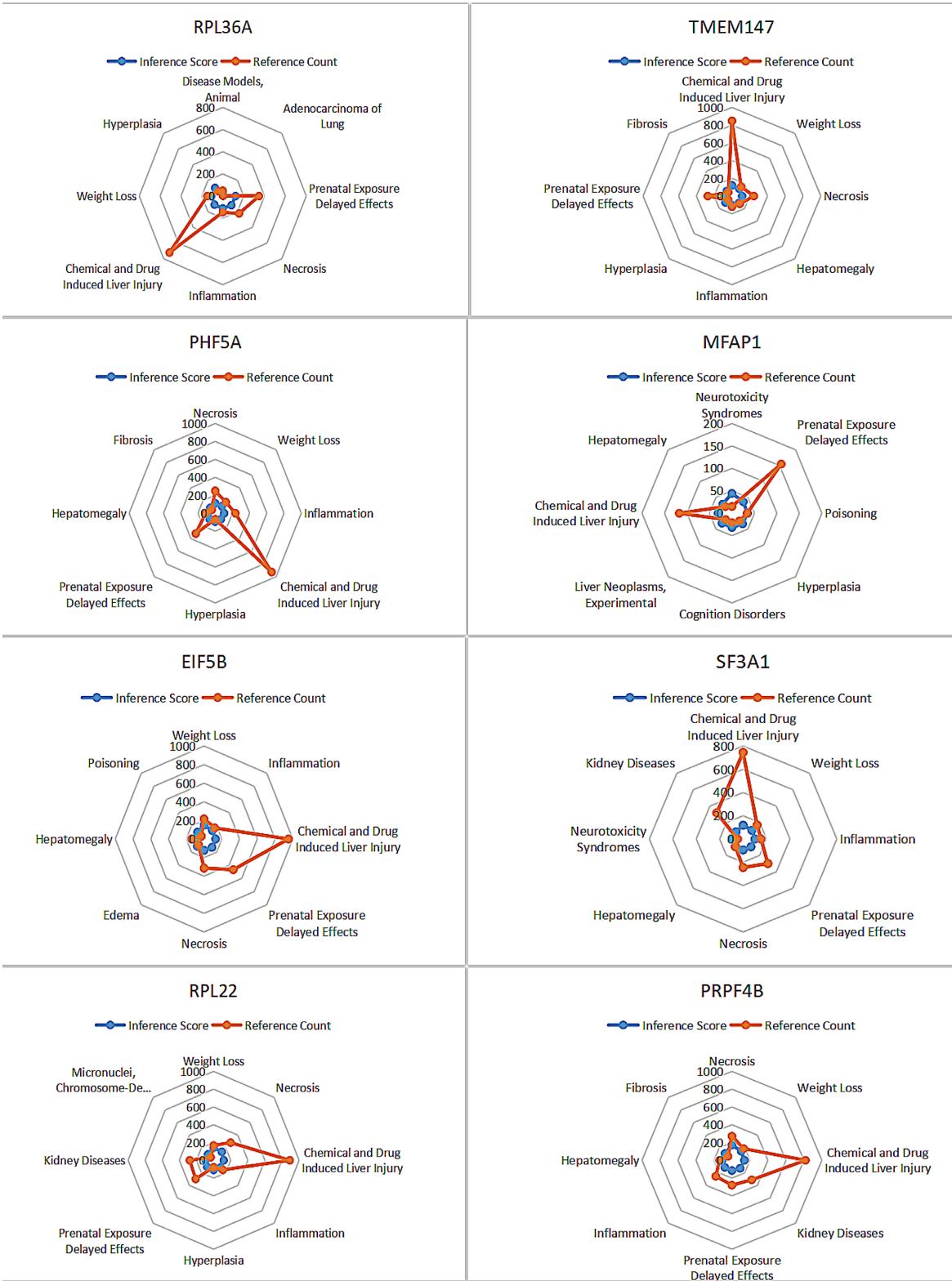


Figure 9. CTD analysis. Nine genes (RPL36A, SF3A1, EIF5B, TMEM147, PHF5A, MFAP1, RPL22, PRPF4B) were associated with inflammation, necrosis, hepatomegaly, fibrosis, and kidney disease

图 9. CTD 分析。9 个基因(RPL36A, SF3A1, EIF5B, TMEM147, PHF5A, MFAP1, RPL22, PRPF4B)和炎症、坏死、肝肿大、纤维化、肾脏疾病有关

Table 1. A summary of miRNAs that regulate hub genes
表 1. 调控关键基因的 miRNAs 概述

	Gene	MIRNA		
1	RPL36A	hsa-miR-325-3p		
2	SF3A1	hsa-miR-3139	hsa-miR-28-5p	hsa-miR-708-5p
3	EIF5B	hsa-miR-199b-5p	hsa-miR-199a-5p	
4	TMEM147	hsa-miR-452-5p	hsa-miR-4676-3p	hsa-miR-892c-3p
5	PHF5A	hsa-miR-107	hsa-miR-103a-3p	
6	MFAP1	hsa-miR-302c-3p.2	hsa-miR-520f-3p	
7	RPL22	hsa-miR-153-3p		
8	PRPF4B	hsa-miR-139-5p		

4. 讨论

糖尿病是全球高发代谢病，2024 年全球成年患者超 5.3 亿，30%~50%并发糖尿病视网膜病变(DR) [15]。DR 是主要致盲并发症，病程隐匿。早期常无症状，进展至增殖期或黄斑水肿时视网膜损伤多不可逆，严重降低生活质量并加重医疗负担[16]。当前 DR 管理面临挑战：早期筛查缺乏敏感标志物，影像学易漏诊；评估依赖主观指标，难以个体化分层；晚期手段无法逆转早期损伤[17] [18]。探索 DR 分子机制、筛选兼具早期预警和干预潜力的分子，对构建全周期管理体系至关重要。

本研究整合 GSE185011 与 GSE146615 两个 DR 相关 GEO 数据集，经批次效应校正、差异表达基因 (DEGs)筛选、加权基因共表达网络分析(WGCNA)及蛋白质 - 蛋白质相互作用(PPI)网络构建，鉴定出 9 个 DR 核心枢纽基因，其中前体 mRNA 加工因子 4B (PRPF4B)的特征尤为突出。结果显示，PRPF4B 在 DR 患者视网膜样本中显著高表达，正常对照中低表达；WGCNA 证实其为 DR 分子网络核心枢纽基因，CTD 数据库分析显示其与 DR 核心病理(炎症、坏死、纤维化)高度相关；TargetScan 预测其可被 hsa-miR-139-5p 调控。参考 PRPF4B 在癌症中“高表达关联不良预后”的特征，推测其在 DR 中也具预后预测价值，表达水平越高，视网膜炎症越活跃、血管损伤越严重，未来发生黄斑水肿、牵拉性视网膜脱离的风险越高，有望成为 DR 健康管理中“预后分层”的关键指标。

PRPF4B 是 U4/U6.U5 三小核核糖核蛋白(tri-snRNP)复合体的核心组分，核心功能为调控 pre-mRNA 剪接，这一转录后调控过程直接决定基因产物的多样性与功能性，对维持细胞生理功能至关重要[19] [20]。已有研究证实，PRPF4B 与剪接体组分协同调控靶基因剪接，其异常表达会导致大量功能基因剪接失调，诱发疾病[21] [22]。近年来，国外研究已将 PRPF4B 与糖尿病并发症关联：在糖尿病心肌病(DCM)中，Zhang 等[23]通过 db/db 小鼠模型发现，心肌组织 PRPF4B 高表达与心肌细胞凋亡率、间质纤维化程度正相关，其通过异常剪接 ERK 通路基因(如 MAP2K1)激活 MAPK/ERK 通路，促进心肌损伤，敲低 PRPF4B 可改善 DCM 小鼠心功能；在糖尿病肾病(DN)中，Kishore 等[13] [24]发现 DN 患者肾小球系膜细胞 PRPF4B 高表达，与尿微量白蛋白/肌酐比值(UACR)、肾小球硬化指数正相关，高糖可诱导其表达上调，通过调控胶原基因 COL4A1 剪接促进细胞外基质(ECM)积聚，这与本研究 PRPF4B 在 DR 中的纤维化关联特征高度一致。

此外，PRPF4B 与炎症反应的调控关联已被证实，而慢性低度炎症被认为是糖尿病视网膜病变(DR)发病的核心驱动因素之一[25]。在糖尿病视网膜病变中，活化的视网膜 Müller 细胞和小胶质细胞释放大 量炎症因子，如肿瘤坏死因子- α (TNF- α)和白细胞介素-6 (IL-6)，这些因子是导致血视网膜屏障(BRB)破坏

和血管渗漏的关键诱因[26][27]。值得注意的是, NF- κ B 信号通路是介导这一炎症反应的核心枢纽, 其持续激活会加剧视网膜的病理损伤[28]。另一方面, BRB 的完整性对于维持视网膜内环境稳定至关重要, 其破坏是 DR 早期血管病变的核心特征, 也是糖尿病性黄斑水肿发生的重要基础[29]。本研究及既往研究表明, PRPF4B 作为关键的剪接调控因子, 其功能异常很可能通过影响炎症通路相关基因(如 NF- κ B 通路组分)及血管完整性相关基因(如 VE-cadherin)的剪接过程, 参与“剪接失调-炎症激活-血管损伤-纤维化”的病理链条, 从而驱动糖尿病并发症的进展。这一机制与本研究 PRPF4B 作为 DR 核心枢纽基因的定位高度一致, 进一步支持了其在 DR 病理进程中扮演的重要角色。

值得注意的是, PRPF4B 调控网络与 DR 健康管理需求高度契合: 其一, 其“DR 高表达、正常低表达”的疾病特异性, 符合“早期筛查标志物”要求; 其二, 其表达水平与 DR 病理进程(炎症、纤维化)的关联, 可用于“病情分期评估”, 如将高表达患者列为“高风险人群”, 制定 3 个月密集随访计划, 低表达患者延长至 6~12 个月随访, 优化健康管理资源[30]; 其三, TargetScan 预测其受 hsa-miR-139-5p 调控(该 miRNA 在 DR 中低表达[31]), 若后续验证 hsa-miR-139-5p 可通过抑制 PRPF4B 改善高糖诱导的视网膜细胞损伤, 可开发眼用制剂, 填补 DR 早期干预空白。

然而, 本研究仍存在若干局限。首先, 分析基于生物信息学数据, 缺乏临床样本中蛋白水平的验证及在血液、房水等易获取样本中的表达检测, 而这是标志物转化应用的前提[32], 这可能导致对 PRPF4B 及相关分子标志物临床适用性的判断存在一定不确定性, 降低了直接外推至临床应用的可靠性。其次, 未开展功能实验, 无法明确 PRPF4B 在视网膜细胞中的具体作用及调控机制, 从而可能影响对其生物学意义及潜在干预价值的准确评估。最后, PRPF4B 与 hsa-miR-139-5p 的调控关系及其下游剪接靶基因尚未验证, 限制了其作为干预靶点的精准开发[33], 进而对结论的稳健性带来一定影响。

综上所述, DR 作为糖尿病致盲的首要原因, 健康管理核心需求是“早识别、早干预、精准预后”。本研究通过生物信息学分析, 首次将 PRPF4B 与 DR 全周期健康管理关联, 证实其高表达且为核心枢纽基因, 与炎症、纤维化密切相关, 兼具“早期筛查标志物-预后分层指标-潜在干预靶点”三重特征。结合国外研究, PRPF4B 在糖尿病心肌病、肾病中的调控机制进一步支持其在 DR 中的驱动作用。尽管缺乏实验验证与临床转化数据, 但 PRPF4B 的发现为 DR 健康管理提供新视角。若后续验证其可靠性与有效性, 有望纳入 DR “早筛-评估-干预-预后”全周期管理体系, 通过“PRPF4B 检测+个体化随访+靶向干预”, 提升早期识别率、降低致盲风险, 改善糖尿病患者健康质量。

5. 结论

本研究发现 PRPF4B 基因在糖尿病视网膜病变中显著高表达, 其通过干扰胰岛素信号、介导细胞应激及促进纤维化等多重机制推动疾病进展。这一发现为糖尿病并发症的精准健康管理提供了新思路: 监测 PRPF4B 表达水平可作为早期风险预警的生物标志物, 实现高危个体的精准识别; 而针对 PRPF4B 的靶向干预则有望成为延缓并发症进展的新型治疗策略, 推动糖尿病管理从被动血糖控制向主动分子分层防控转变。

参考文献

- [1] American Diabetes Association Professional Practice Committee (2021) 2. Classification and Diagnosis of Diabetes: standards of Medical Care in Diabetes—2022. *Diabetes Care*, **45**, S17-S38. <https://doi.org/10.2337/dc22-s002>
- [2] Williams, R., Karuranga, S., Malanda, B., Saedi, P., Basit, A., Besançon, S., *et al.* (2020) Global and Regional Estimates and Projections of Diabetes-Related Health Expenditure: Results from the International Diabetes Federation Diabetes Atlas, 9th Edition. *Diabetes Research and Clinical Practice*, **162**, Article ID: 108072. <https://doi.org/10.1016/j.diabres.2020.108072>
- [3] Collum, L.M.T. (2000) Immune-Mediated Corneal Melting Disease. *Acta Ophthalmologica Scandinavica*, **78**, 78-80.

- <https://doi.org/10.1034/j.1600-0420.2000.078s230078.x>
- [4] Chatterjee, S., Khunti, K. and Davies, M.J. (2017) Type 2 Diabetes. *The Lancet*, **389**, 2239-2251. [https://doi.org/10.1016/s0140-6736\(17\)30058-2](https://doi.org/10.1016/s0140-6736(17)30058-2)
- [5] Forbes, J.M. and Cooper, M.E. (2013) Mechanisms of Diabetic Complications. *Physiological Reviews*, **93**, 137-188. <https://doi.org/10.1152/physrev.00045.2011>
- [6] Flaxel, C.J., Adelman, R.A., Bailey, S.T., Fawzi, A., Lim, J.I., Vemulakonda, G.A., *et al.* (2020) Diabetic Retinopathy Preferred Practice Pattern®. *Ophthalmology*, **127**, P66-P145. <https://doi.org/10.1016/j.ophtha.2019.09.025>
- [7] Pop-Busui, R., Boulton, A.J.M., Feldman, E.L., Bril, V., Freeman, R., Malik, R.A., *et al.* (2016) Diabetic Neuropathy: A Position Statement by the American Diabetes Association. *Diabetes Care*, **40**, 136-154. <https://doi.org/10.2337/dc16-2042>
- [8] Scanlon, P.H., Malhotra, R., Thomas, G., Foy, C., Kirkpatrick, J.N., Lewis-Barned, N., *et al.* (2003) The Effectiveness of Screening for Diabetic Retinopathy by Digital Imaging Photography and Technician Ophthalmoscopy. *Diabetic Medicine*, **20**, 467-474. <https://doi.org/10.1046/j.1464-5491.2003.00954.x>
- [9] Wong, T.Y., Cheung, C.M.G., Larsen, M., Sharma, S. and Simó, R. (2016) Diabetic Retinopathy. *Nature Reviews Disease Primers*, **2**, Article No. 16012. <https://doi.org/10.1038/nrdp.2016.12>
- [10] Diabetes Genetics Replication and Meta-Analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, *et al.* (2014) Genome-Wide Trans-Ancestry Meta-Analysis Provides Insight into the Genetic Architecture of Type 2 Diabetes Susceptibility. *Nature Genetics*, **46**, 234-244.
- [11] Sandbach, D. (1982) The Long-Term Implications of the Short Report—The Time to Redesign Medical Careers. *Health Services Manpower Review*, **8**, 3-6.
- [12] Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., *et al.* (2012) NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Research*, **41**, D991-D995. <https://doi.org/10.1093/nar/gks1193>
- [13] Du, J., Zhu, G., Cai, J., Wang, B., Luo, Y., Chen, C., *et al.* (2021) Splicing Factors: Insights into Their Regulatory Network in Alternative Splicing in Cancer. *Cancer Letters*, **501**, 83-104. <https://doi.org/10.1016/j.canlet.2020.11.043>
- [14] Liu, Y., Liu, W., Zhang, Z., Hu, Y., Zhang, X., Sun, Y., *et al.* (2021) Yishen Capsule Promotes Podocyte Autophagy through Regulating SIRT1/NF- κ B Signaling Pathway to Improve Diabetic Nephropathy. *Renal Failure*, **43**, 128-140. <https://doi.org/10.1080/0886022x.2020.1869043>
- [15] Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B.B., *et al.* (2022) IDF Diabetes Atlas: Global, Regional and Country-Level Diabetes Prevalence Estimates for 2021 and Projections for 2045. *Diabetes Research and Clinical Practice*, **183**, Article ID: 109119. <https://doi.org/10.1016/j.diabres.2021.109119>
- [16] Teo, Z.L., Tham, Y., Yu, M., Chee, M.L., Rim, T.H., Cheung, N., *et al.* (2021) Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-Analysis. *Ophthalmology*, **128**, 1580-1591. <https://doi.org/10.1016/j.ophtha.2021.04.027>
- [17] Scanlon, P.H. (2008) Article Commentary: The English National Screening Programme for Sight-Threatening Diabetic Retinopathy. *Journal of Medical Screening*, **15**, 1-4. <https://doi.org/10.1258/jms.2008.008015>
- [18] Das, A., McGuire, P.G. and Rangasamy, S. (2015) Diabetic Macular Edema: Pathophysiology and Novel Therapeutic Targets. *Ophthalmology*, **122**, 1375-1394. <https://doi.org/10.1016/j.ophtha.2015.03.024>
- [19] Wahl, M.C., Will, C.L. and Lührmann, R. (2009) The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, **136**, 701-718. <https://doi.org/10.1016/j.cell.2009.02.009>
- [20] Will, C.L. and Lührmann, R. (2001) Spliceosomal UsnRNP Biogenesis, Structure and Function. *Current Opinion in Cell Biology*, **13**, 290-301. [https://doi.org/10.1016/s0955-0674\(00\)00211-8](https://doi.org/10.1016/s0955-0674(00)00211-8)
- [21] Scott, L.M. and Rebel, V.I. (2013) Acquired Mutations That Affect Pre-mRNA Splicing in Hematologic Malignancies and Solid Tumors. *JNCI Journal of the National Cancer Institute*, **105**, 1540-1549. <https://doi.org/10.1093/jnci/djt257>
- [22] Lee, S.C. and Abdel-Wahab, O. (2016) Therapeutic Targeting of Splicing in Cancer. *Nature Medicine*, **22**, 976-986. <https://doi.org/10.1038/nm.4165>
- [23] Lv, Y., Zhang, W., Zhao, J., Sun, B., Qi, Y., Ji, H., *et al.* (2021) SRSF1 Inhibits Autophagy through Regulating BCL-X Splicing and Interacting with PIK3C3 in Lung Cancer. *Signal Transduction and Targeted Therapy*, **6**, Article No. 108. <https://doi.org/10.1038/s41392-021-00495-6>
- [24] Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., *et al.* (2019) Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, **176**, 535-548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>
- [25] Tang, J. and Kern, T.S. (2011) Inflammation in Diabetic Retinopathy. *Progress in Retinal and Eye Research*, **30**, 343-358. <https://doi.org/10.1016/j.preteyeres.2011.05.002>

-
- [26] Zeng, H.Y., Green, W.R. and Tso, M.O. (2008) Microglial Activation in Human Diabetic Retinopathy. *Archives of Ophthalmology*, **126**, 227-232. <https://doi.org/10.1001/archophthalmol.2007.65>
- [27] Joussen, A.M., Poulaki, V., Le, M.L., Koizumi, K., Esser, C., Janicki, H., *et al.* (2004) A Central Role for Inflammation in the Pathogenesis of Diabetic Retinopathy. *The FASEB Journal*, **18**, 1450-1452. <https://doi.org/10.1096/fj.03-1476fje>
- [28] Zhang, T., Ma, C., Zhang, Z., Zhang, H. and Hu, H. (2021) NF- κ B Signaling in Inflammation and Cancer. *MedComm*, **2**, 618-653. <https://doi.org/10.1002/mco2.104>
- [29] Antonetti, D.A., Klein, R. and Gardner, T.W. (2012) Diabetic Retinopathy. *New England Journal of Medicine*, **366**, 1227-1239. <https://doi.org/10.1056/nejmra1005073>
- [30] García-Fiñana, M., Hughes, D.M., Cheyne, C.P., Broadbent, D.M., Wang, A., Komárek, A., *et al.* (2018) Personalized Risk-Based Screening for Diabetic Retinopathy: A Multivariate Approach versus the Use of Stratification Rules. *Diabetes, Obesity and Metabolism*, **21**, 560-568. <https://doi.org/10.1111/dom.13552>
- [31] Chamberlain, S., Koenig, M., Richter, A., Palau, F. and Pandolfo, M. (1993) Molecular Analysis of the Friedreich's Ataxia Locus. *Advances in Neurology*, **61**, 193-204
- [32] Kori, M., Gov, E., Arga, K.Y. and Sinha, R. (2024) Biomarkers from Discovery to Clinical Application: *In Silico* Pre-Clinical Validation Approach in the Face of Lung Cancer. *Biomarker Insights*, **19**, 1-9. <https://doi.org/10.1177/11772719241287400>
- [33] Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., *et al.* (2019) Therapeutic Target Database 2020: Enriched Resource for Facilitating Research and Early Development of Targeted Therapeutics. *Nucleic Acids Research*, **48**, D1031-D1041. <https://doi.org/10.1093/nar/gkz981>