

基于转录组分析的冠状动脉疾病枢纽基因鉴定及诊断列线图构建

李晨宇¹, 孙雪纯^{2*}

¹山东大学齐鲁医院内分泌与代谢病科, 山东 济南

²山东大学公共卫生学院营养与食品卫生学系, 山东 济南

收稿日期: 2026年3月21日; 录用日期: 2026年4月16日; 发布日期: 2026年4月22日

摘要

背景: 冠状动脉疾病(CAD)是一类以冠状动脉粥样硬化为主要病理基础的慢性心血管疾病, 其分子机制尚未完全阐明。本研究旨在筛选CAD的诊断相关枢纽基因, 并探讨其在免疫病理过程中的潜在作用。**方法:** 从GEO数据库获取GSE20680数据集。采用limma筛选差异表达基因(DEGs), 并进行GO和KEGG富集分析; 采用WGCNA识别与CAD相关的模块; 利用CIBERSORT推断免疫细胞相对组成; 通过SVM-RFE和随机森林算法交集筛选枢纽基因; 进一步构建诊断列线图, 并采用校准曲线、DCA及ROC进行评价。**结果:** CAD相关差异基因主要涉及氧化磷酸化、核糖体、能量代谢及免疫炎症相关通路。最终筛选出3个枢纽基因: PSMA6、TOMM7和RPL9。免疫特征分析显示, CAD组与对照组在多种免疫细胞亚群相对丰度上存在差异。基于枢纽基因构建的列线图具有一定判别能力(AUC约为0.65)和校准性能。**结论:** PSMA6、TOMM7和RPL9可能作为CAD的潜在诊断生物标志物。基于枢纽基因构建的列线图可为CAD的早期识别及风险评估提供一定参考。

关键词

冠状动脉疾病, 诊断生物标志物, 免疫细胞组成, 转录组学, 枢纽基因, 列线图

Identification of Hub Genes and Construction of a Diagnostic Nomogram for Coronary Artery Disease Based on Transcriptomic Analysis

Chenyu Li¹, Xuechun Sun^{2*}

¹Department of Endocrinology and Metabolism, Qilu Hospital of Shandong University, Jinan Shandong

*通讯作者。

文章引用: 李晨宇, 孙雪纯. 基于转录组分析的冠状动脉疾病枢纽基因鉴定及诊断列线图构建[J]. 临床医学进展, 2026, 16(4): 4369-4382. DOI: 10.12677/acm.2026.1641706

Abstract

Background: Coronary artery disease (CAD) is a chronic cardiovascular disorder primarily based on coronary atherosclerosis, and its molecular mechanisms remain incompletely understood. This study aimed to identify diagnostic hub genes for CAD and to explore their potential roles in its immunopathological processes. **Methods:** The GSE20680 dataset was obtained from the GEO database. Differentially expressed genes (DEGs) were identified using limma, followed by GO and KEGG enrichment analyses. WGCNA was performed to identify CAD-related modules. CIBERSORT was used to infer the relative composition of immune cells. Hub genes were screened by intersecting the results of SVM-RFE and random forest. A diagnostic nomogram was then constructed and evaluated using calibration curves, decision curve analysis (DCA), and receiver operating characteristic (ROC) curves. **Results:** CAD-related DEGs were mainly involved in oxidative phosphorylation, ribosome, energy metabolism, and immune-inflammatory pathways. Three hub genes were ultimately identified: PSMA6, TOMM7, and RPL9. Immune profiling analysis showed differences in the relative abundance of multiple immune cell subsets between the CAD and control groups. The nomogram constructed based on the hub genes showed a certain discriminative ability (AUC approximately 0.65) and calibration performance. **Conclusion:** PSMA6, TOMM7, and RPL9 may serve as potential diagnostic biomarkers for CAD. The nomogram constructed based on these hub genes may provide a useful reference for the early identification and risk assessment of CAD.

Keywords

Coronary Artery Disease, Diagnostic Biomarkers, Immune Cell Composition, Transcriptomics, Hub Genes, Nomogram

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

冠状动脉疾病(coronary artery disease, CAD)是一类以冠状动脉粥样硬化为主要病理基础的慢性心血管疾病,其发生发展与脂质沉积、内皮功能障碍及炎症性重塑密切相关,具有较高的发病率和死亡率,仍是导致残疾和死亡的主要原因之一[1]。与一般人群相比,CAD患者心血管不良事件及死亡风险显著增加,预期寿命明显缩短。其沉重的疾病负担、较高的心肌梗死风险及长期并发症使其成为全球公共卫生领域的重大挑战。尽管近年来相关研究不断深入,CAD发生发展所涉及的分子机制仍未被完全阐明。因此,解析CAD的遗传学与分子基础,对于改进风险预测和实现精准预防具有重要意义[2]。

近年来,全转录组基因表达谱分析被广泛应用于CAD相关基因筛选、生物标志物发现及疾病相关分子通路解析[3]。既往研究表明,转录调控因子和应激反应基因,如NF- κ B通路相关基因和即刻早期基因,参与了动脉粥样硬化及CAD相关炎症反应;在CAD人群中,NF- κ B通路相关遗传因素也已得到直接评估[4]。此外,自噬相关竞争性内源RNA网络在CAD诊断中的应用潜力也受到关注。炎症介质如环氧合

酶-2 (cyclooxygenase-2, COX-2)和前列腺素 E2 (prostaglandin E2, PGE2)同样被认为与 CAD 血管炎症及斑块进展密切相关[5]。然而,目前仍缺乏稳定可靠的分子标志物,这在一定程度上限制了 CAD 的早期诊断和个体化治疗。筛选具有较高稳定性的基因特征有望提高临床风险预测和患者分层能力。

本研究旨在通过整合生物信息学方法,系统表征 CAD 中的基因表达改变并筛选潜在诊断生物标志物。我们分析了 GEO 数据库中的 GSE20680 芯片数据集,以识别 CAD 样本与对照样本之间的差异表达基因;随后采用加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)识别关键共表达模块,并结合支持向量机递归特征消除(support vector machine-recursive feature elimination, SVM-RFE)和随机森林(random forest, RF)方法筛选潜在枢纽基因,进一步构建诊断列线图模型,以期为 CAD 的早期识别、风险分层及相关分子机制研究提供新的理论依据。

2. 材料与方法

2.1. 数据处理

从 GEO 数据库获取微阵列数据集 GSE20680。本研究纳入 GSE20680 中的 CAD 样本和对照样本,采用 R 语言中的 limma 包进行差异表达分析[6]。由于本研究使用的公开表达谱数据集主要用于转录组分析,缺乏完整且可直接用于统一建模的常规临床危险因素和实验室生物标志物信息,因此本研究未能进一步构建包含血脂、血糖、血压、吸烟史或 hs-CRP、cTnI 等指标的联合比较模型。

2.2. 差异表达基因筛选

采用 R 语言中的 limma 包并结合经验贝叶斯方法对 CAD 组与对照组进行差异表达分析[7][8]。鉴于本研究所用样本来源于外周血表达谱, CAD 相关转录改变在血液中往往呈现中等效应幅度;同时,本研究后续还将差异表达结果与 WGCNA 模块归属及机器学习特征筛选进行交叉整合,因此初始 DEG 筛选更强调候选基因的敏感性保留,而非单一步骤中的严格收缩。基于此,以 $P < 0.05$ 且 $|\log_2 \text{fold change}| > 0.5$ 作为差异表达基因的初筛标准,以避免遗漏可能进入后续模块和特征选择流程的中等效应基因。火山图和热图分别采用 ggplot2 和 pheatmap 包绘制,用于展示差异表达基因的总体分布特征及其在各样本中的表达模式。

2.3. 富集分析

采用 clusterProfiler 包(结合 org.Hs.eg.db 注释库)对差异表达基因进行 Gene Ontology (GO)和 KEGG 功能富集分析[9]-[11]。以 $FDR < 0.05$ 作为显著富集标准。

2.4. 基因集富集分析

采用 MSigDB 数据库中的 c2.cp.kegg 基因集,在 symbol 水平进行基因集富集分析(GSEA) [12] [13]。根据 limma 分析得到的 t 统计量对基因进行排序;设定基因集大小为 10~200, $FDR < 0.05$ 作为显著标准,并以标准化富集分数(NES)表示富集方向。GSEA 中所用参考基因集为 MSigDB 的 c2.cp.kegg (symbol 水平),排序基因列表基于 limma 的 t 统计量构建。

2.5. 加权基因共表达网络分析

基于 GSE20680 表达矩阵进行 WGCNA 分析。根据标准 WGCNA 流程,采用 pickSoftThreshold 函数在一系列候选软阈值下计算无尺度拓扑拟合指数和平均连接度,并综合尽早达到近似无尺度网络与保留可接受网络连通性两方面信息选择软阈值功率(β)。本研究最终采用图 4(A)所示达到近似无尺度拓扑且未

导致平均连接度过度下降的最小 β 值构建加权网络; 随后构建无符号网络及拓扑重叠矩阵(TOM) [14]。

2.6. 支持向量机与随机森林模型构建

首先筛选差异表达基因与 CAD 相关 WGCNA 模块基因的交集, 作为候选特征基因。随后采用 2 种机器学习算法进行特征筛选: 支持向量机递归特征消除(support vector machine-recursive feature elimination, SVM-RFE) [15]和随机森林(random forest) [16] [17]。其中, SVM-RFE 用于筛选具有较优分类性能的特征基因集合; 随机森林用于评估候选基因的重要性, 并进一步筛选与 CAD 密切相关的特征基因。两种方法共同保留下来的基因定义为枢纽基因。

2.7. 免疫细胞反卷积分析(CIBERSORT)

基于标准化表达数据, 采用 CIBERSORT 算法和 LM22 特征矩阵(22 种白细胞亚群)推断样本中的免疫细胞组成[18]。对每个样本估算各类免疫细胞的相对比例, 包括活化 NK 细胞、M0/M1/M2 型巨噬细胞、记忆 B 细胞、浆细胞和 CD8⁺T 细胞等。仅保留 CIBERSORT 输出中 $P < 0.05$ 的样本结果用于后续分析。组间差异采用 Wilcoxon 检验比较, 基因表达与免疫细胞相对丰度之间的相关性采用 Spearman 相关分析。

2.8. 枢纽基因表达特征分析

对筛选得到的枢纽基因在 CAD 组与对照组中的表达情况进行可视化分析。采用热图展示 PSMA6、TOMM7 和 RPL9 在不同样本中的表达模式, 采用箱线图比较两组间枢纽基因的表达差异, 并构建基因关系网络图以展示枢纽基因之间的关联性。

2.9. 诊断模型构建与效能评价

采用 rms 包基于枢纽基因构建 CAD 诊断列线图模型(逻辑回归) [19]。采用 pROC 包绘制 ROC 曲线并计算曲线下面积(AUC)及其 95%置信区间, 以评价枢纽基因及列线图模型的诊断性能。采用校准曲线评估模型预测概率与实际观察结果的一致性。进一步采用临床影响曲线和决策曲线分析(decision curve analysis, DCA)评价模型的潜在临床应用价值[20]。

2.10. 统计学分析

所有分析均在 R 软件(v4.1.3)中完成。两组比较采用双侧 Wilcoxon 检验。差异表达基因筛选以 $P < 0.05$ 且 $|\log_2 \text{fold change}| > 0.5$ 为筛选标准; 其余涉及多重检验的分析采用 Benjamini-Hochberg 方法控制假发现率(FDR), 以 $FDR < 0.05$ 为差异有统计学意义。相关性分析采用 Spearman 秩相关。

3. 结果

3.1. 冠状动脉疾病组与健康对照组差异表达基因的鉴定

本研究利用微阵列数据集 GSE20680 进行差异表达分析。结果显示, 与对照组相比, CAD 组存在一系列显著失调的基因, 其总体分布可通过火山图和热图直观展示(图 1(A)、图 1(B))。火山图反映了基因表达变化的效应量及统计学显著性, 其中显著上调和下调的基因均被标出(图 1(A))。热图提示 CAD 组与对照组在差异表达基因层面存在一定的表达谱差异(图 1(B))。

3.2. 功能分析

GO 分析显示, 差异表达基因在分子功能、细胞组分和生物过程层面均显著富集, 主要涉及核糖体结

构成分、NAD(P)H 相关氧化还原酶活性及 NADH 脱氢酶活性等条目(图 2(A)), 提示 CAD 相关差异基因主要与核糖体结构及线粒体呼吸链功能改变有关。GO 条目 - 基因网络图进一步表明, 显著富集条目与 RPL/RPS 家族及 NDUF 家族基因密切相关(图 2(B))。

GSEA 结果显示, 差异基因主要富集于核糖体、基础转录因子、心肌收缩、肾素 - 血管紧张素系统、脂质与动脉粥样硬化、NOD 样受体信号通路、吞噬体及 IgA 生成相关肠道免疫网络等通路(图 3(A))。代表性负向和正向富集通路的运行曲线见图 3(B)、图 3(C), 提示 CAD 中存在代谢、炎症免疫及心血管相关通路的共同重塑。

3.3. CAD 相关模块基因的识别及其与差异表达基因的重叠分析

为探究 CAD 中的基因共表达模式, 本研究基于 GSE20680 表达矩阵进行了加权基因共表达网络分析(WGCNA)。根据无尺度拓扑拟合指数和平均连接度综合确定软阈值参数, 从而构建基因共表达网络

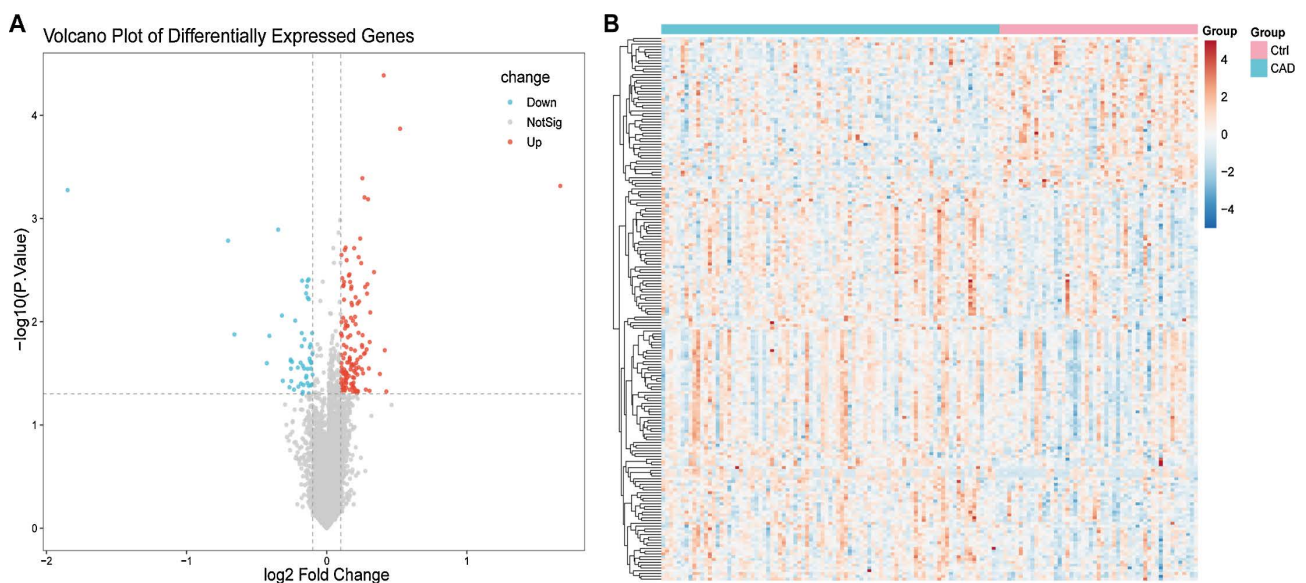


Figure 1. Differentially expressed gene analysis between the CAD group and the control group

图 1. CAD 组与对照组之间的差异表达基因分析结果

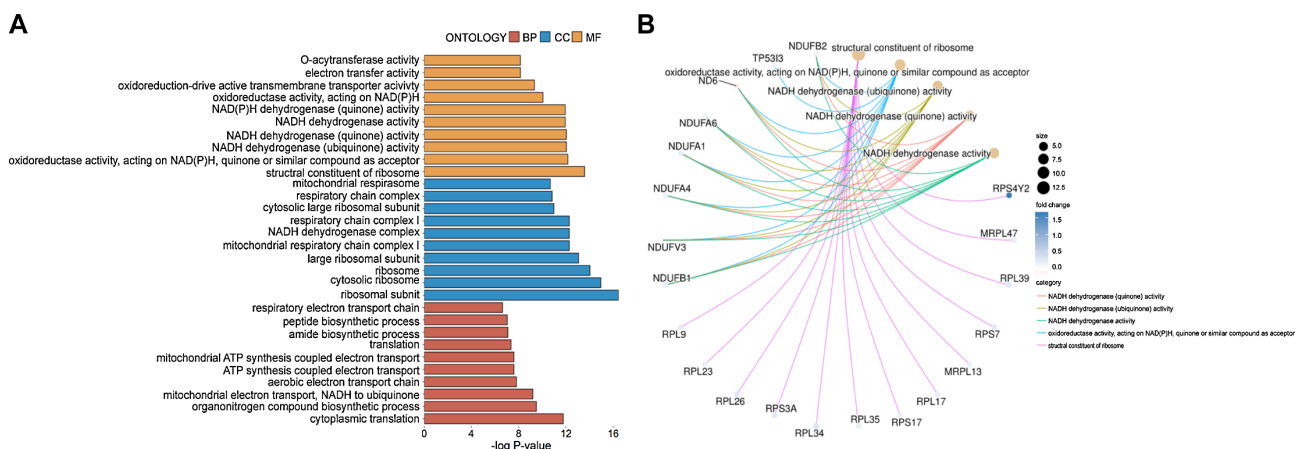


Figure 2. GO functional enrichment analysis of differentially expressed genes between the CAD group and the control group

图 2. CAD 组与对照组差异表达基因的 GO 功能富集分析

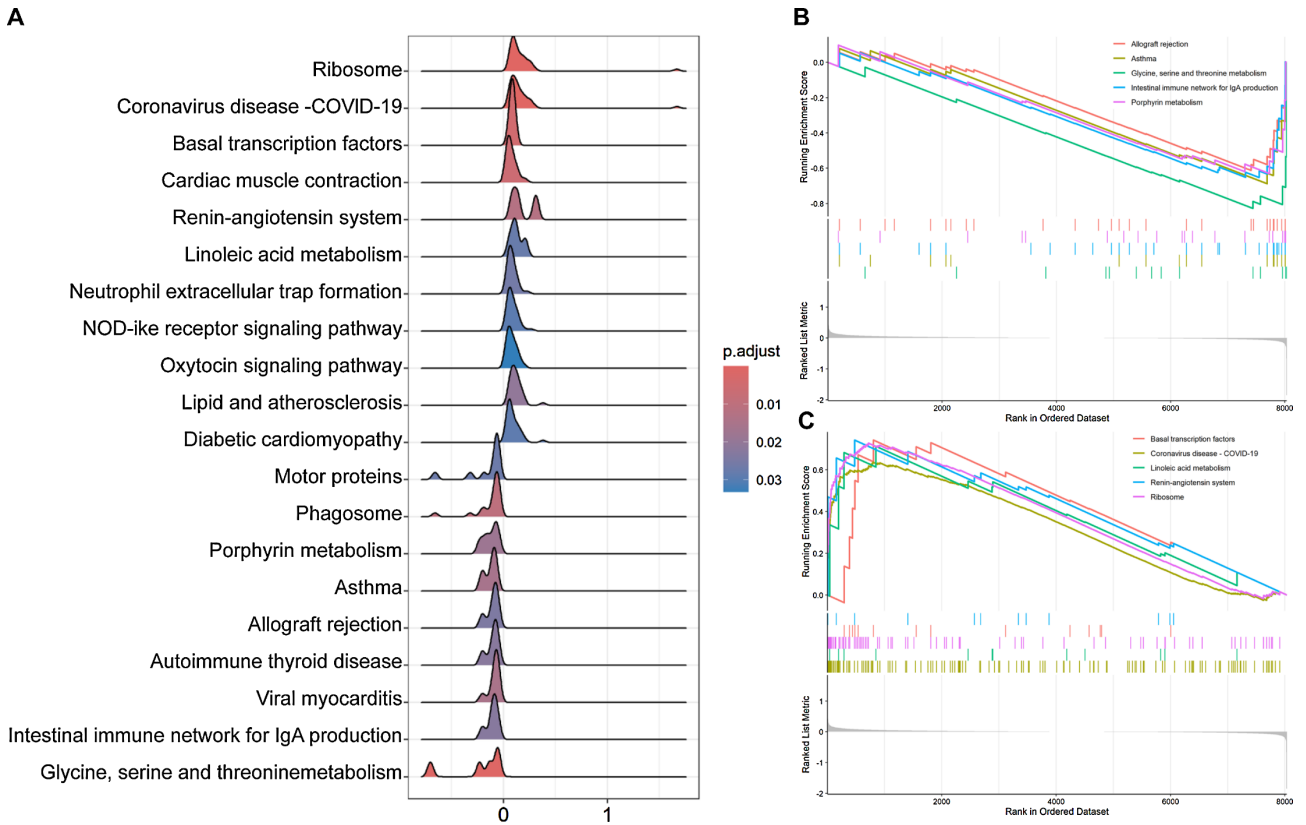
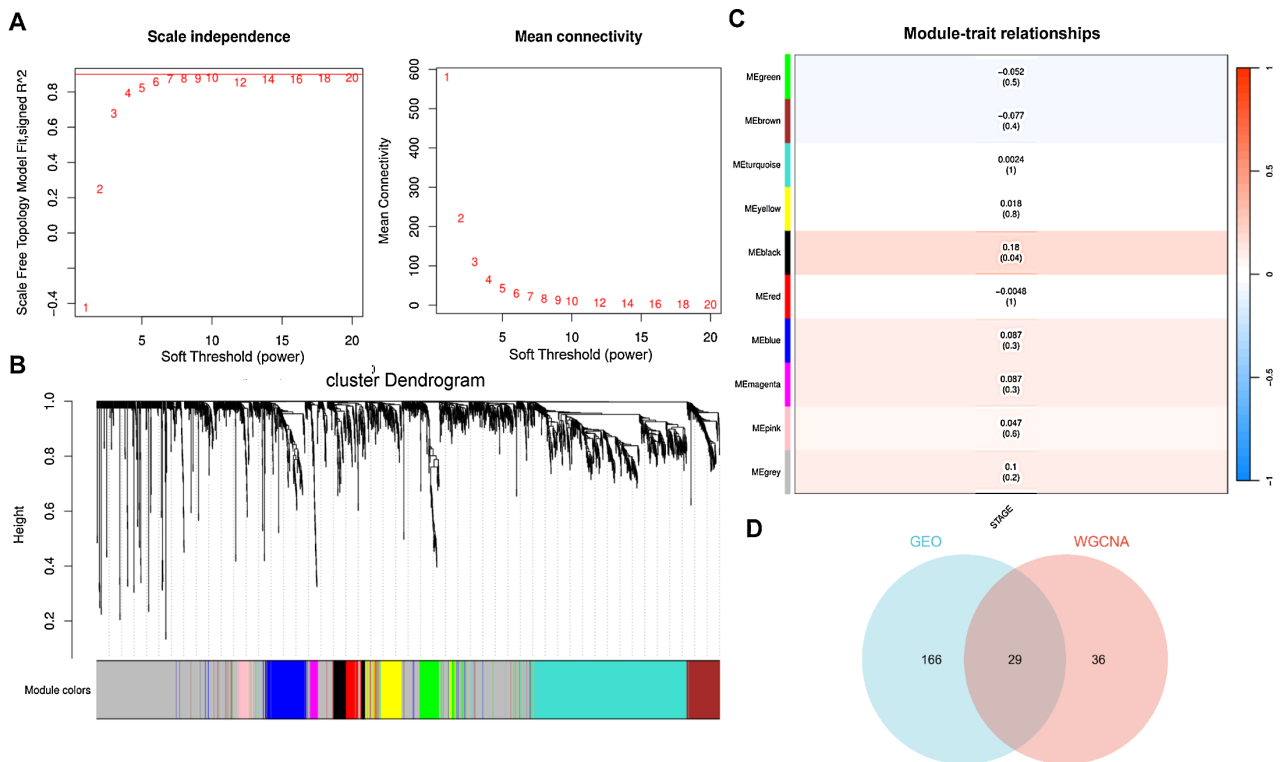


Figure 3. Gene set enrichment analysis (GSEA) results between the CAD group and the control group

图 3. CAD 组与对照组基因集富集分析(GSEA)结果



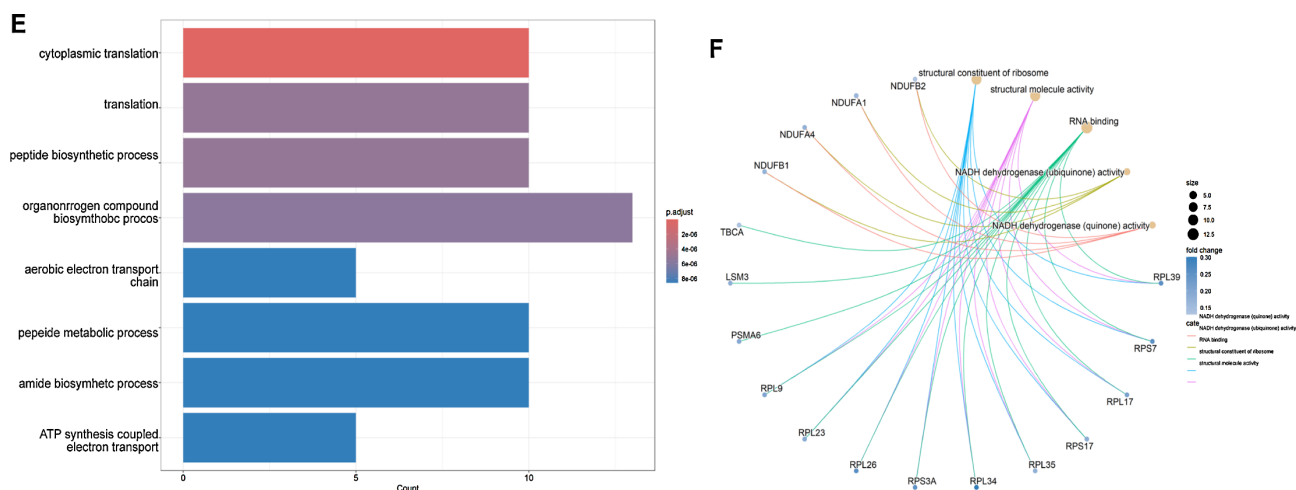


Figure 4. WGCNA analysis of CAD and functional annotation of the intersecting genes

图 4. CAD 的 WGCNA 分析及交集基因功能注释

(图 4(A))。基于层次聚类分析, 识别出多个基因共表达模块(图 4(B))。模块 - 表型相关性分析显示, 黑色模块与 CAD 表型的相关性相对较高, 因此将其作为后续分析的候选关键模块(图 4(C))。进一步将该模块基因与差异表达基因取交集, 共获得 29 个候选基因(图 4(D))。对交集基因进行功能富集分析显示, 这些基因主要涉及细胞质翻译、翻译、肽生物合成过程、有机含氮化合物生物合成过程及有氧电子传递链等生物过程(图 4(E))。基因 - 概念网络图进一步表明, 交集基因主要围绕核糖体结构组成、NADH 脱氢酶相关活性等功能条目形成关联网络(图 4(F))。上述结果为后续枢纽基因筛选提供了候选基因集。

3.4. 枢纽基因鉴定

为筛选与 CAD 相关的稳定基因特征, 本研究对候选基因采用 SVM-RFE 和随机森林两种机器学习算法进行特征筛选。如图 5(A)、图 5(B)所示, SVM-RFE 交叉验证结果显示, 当特征数为 17 时模型具有较优分类性能。随机森林分析结果见图 5(C), 随着决策树数量增加, 模型误差趋于稳定。进一步取 SVM-RFE 与随机森林筛选结果的交集, 最终获得 3 个与 CAD 密切相关的候选枢纽基因: PSMA6、TOMM7 和 RPL9 (图 5(D))。这些基因随后被用于后续诊断模型构建及免疫相关性分析。

3.5. 枢纽基因的基因集富集分析

为进一步探讨所筛选枢纽基因的潜在功能, 对 PSMA6、TOMM7 和 RPL9 分别进行了 GSEA 分析。结果显示, PSMA6 主要富集于病毒感染、免疫应答及蛋白稳态调控相关通路, 包括冠状病毒病、RIG-I 样受体信号通路、蛋白酶体及核糖体通路(图 6(A)); TOMM7 主要富集于氧化磷酸化、核糖体生物发生及代谢调控相关通路(图 6(B)); RPL9 则主要富集于 MAPK 信号通路、单纯疱疹病毒 1 型感染及核糖体相关通路(图 6(C))。这些结果提示, 上述枢纽基因可能参与 CAD 相关的免疫应答、能量代谢、蛋白质合成及转录后调控过程。

3.6. 枢纽基因与免疫细胞组成的相关性

本研究利用 CIBERSORT(LM22)分析了 CAD 样本与对照样本的免疫细胞组成特征。结果显示, 两组样本在多种免疫细胞亚群的相对丰度上存在差异(图 7(A)、图 7(B))。总体来看, B 细胞、NK 细胞、巨噬细胞及部分髓系细胞亚群在两组间呈现不同分布特征。相关性分析表明, PSMA6、TOMM7 和 RPL9 的

表达水平与多种免疫细胞亚群丰度相关, 尤其与 B 细胞、T 细胞、NK 细胞及巨噬细胞亚群关系较为密切(图 7(C)~(F))。这些结果提示, CAD 中可能存在免疫细胞组成失衡, 且枢纽基因可能参与相关免疫调控过程。

3.7. 枢纽基因表达特征分析

为进一步评估枢纽基因在 CAD 组与对照组中的表达特征, 对 PSMA6、TOMM7 和 RPL9 进行了可视化分析[19][20]。热图显示, 这 3 个枢纽基因在两组样本中存在不同的表达模式(图 8(A))。箱线图进一步表明, PSMA6、TOMM7 和 RPL9 在 CAD 组中的表达水平均高于对照组, 差异具有统计学意义(图 8(B))。此外, 基因关系网络图显示, 3 个枢纽基因之间存在一定关联性(图 8(C))。上述结果进一步支持 PSMA6、TOMM7 和 RPL9 可能作为 CAD 的潜在生物标志物。

3.8. 枢纽基因的诊断效能评价

进一步在 GSE20680 队列中评估 3 个枢纽基因(PSMA6、TOMM7 和 RPL9)及列线图模型的诊断性能。ROC 曲线分析显示, 列线图模型、PSMA6、TOMM7 和 RPL9 的 AUC 分别为 0.6541 (95% CI: 0.5618~0.7463)、0.6608 (95% CI: 0.5685~0.7530)、0.6771 (95% CI: 0.6445~0.6777) 和 0.5691 (95% CI: 0.4639~0.6743) (图 9(B))。校准曲线提示模型预测概率与实际观察结果具有一定一致性(图 9(C)), 临床影响曲线和决策曲线分析表明, 该模型在一定阈值概率范围内具有一定临床净获益(图 9(D)、图 9(E))。总体来看, PSMA6、TOMM7 和 RPL9 在当前数据集中具有一定探索性诊断信号, 基于枢纽基因构建的列线图模型可作为转录组层面的候选判别框架。但需指出, 在当前单队列条件下, 列线图模型的 AUC 并未明显优于最佳单基因 TOMM7, 提示多基因整合带来的增益仍较有限, 其增量预测价值仍需在具有完整临床信息的独立队列中进一步验证。

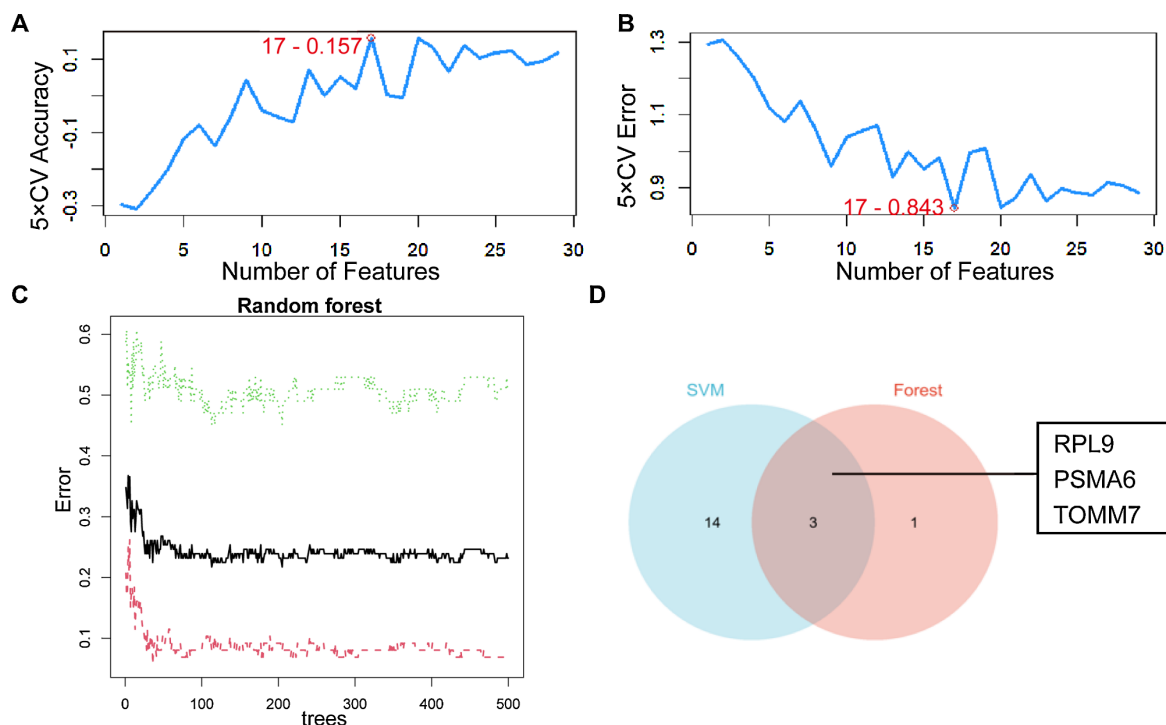


Figure 5. Screening of CAD-related hub genes

图 5. CAD 相关枢纽基因的筛选

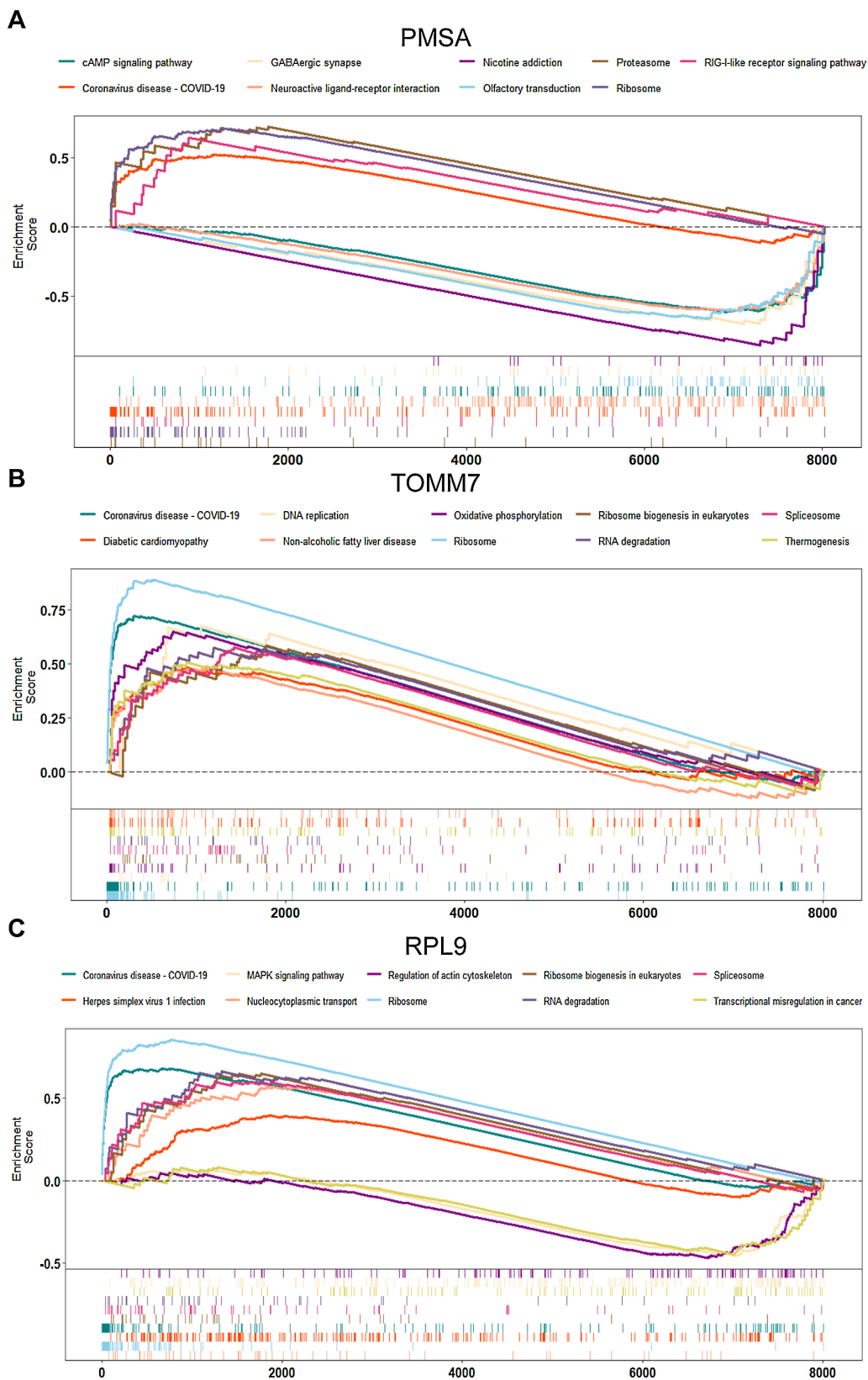


Figure 6. GSEA analysis of hub genes in CAD
图 6. CAD 中枢纽基因的 GSEA 分析

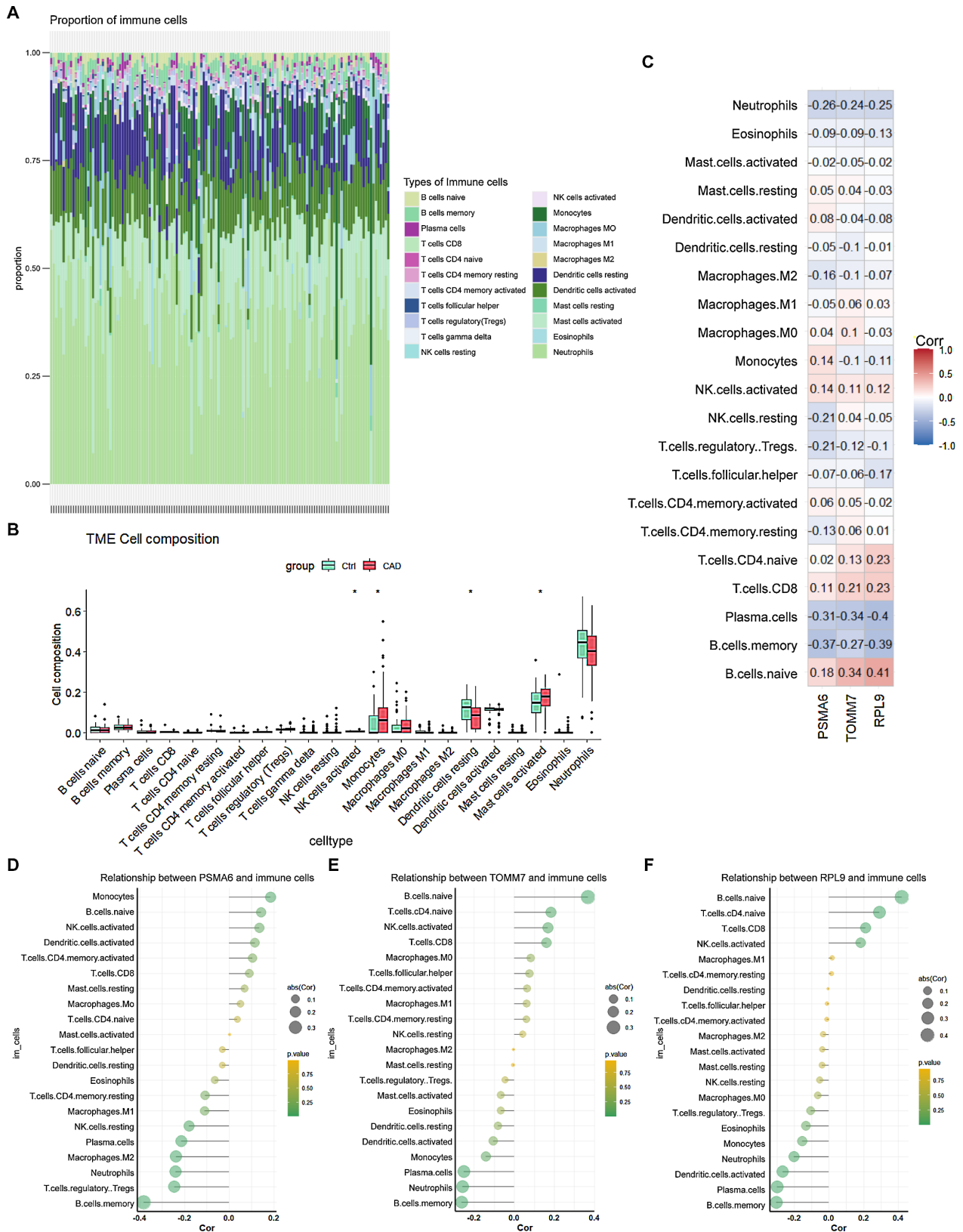


Figure 7. Analysis of immune cell composition in CAD based on CIBERSORT (LM22)
图 7. 基于 CIBERSORT(LM22)的 CAD 免疫细胞组成分析

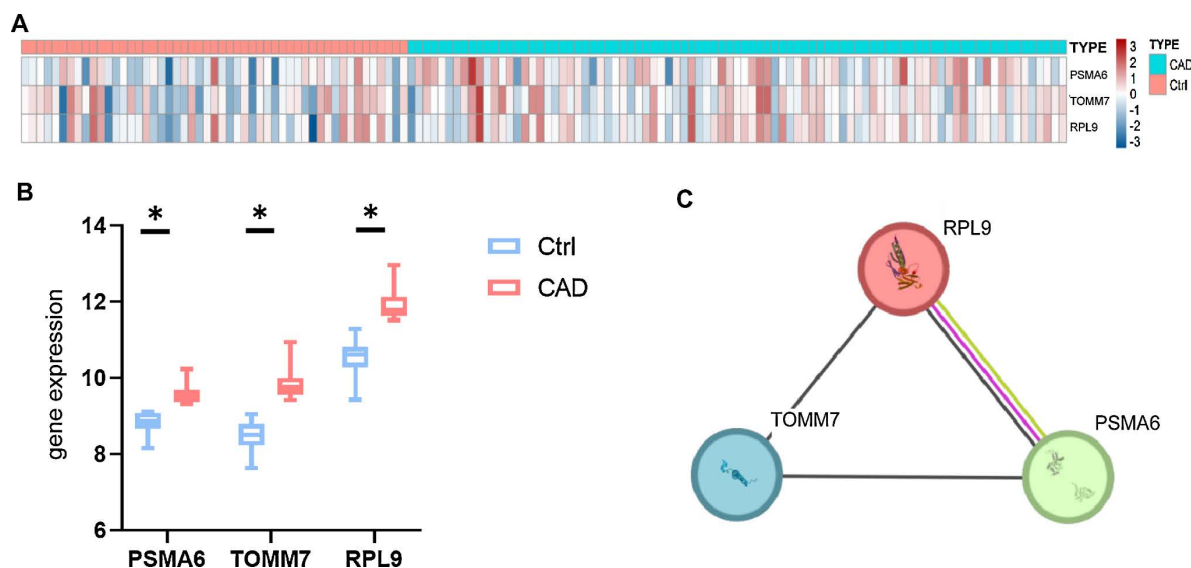


Figure 8. Expression profile analysis of hub genes
图 8. 枢纽基因的表达特征分析

4. 讨论

CAD 是一类以冠状动脉粥样硬化为主要病理基础的常见心血管疾病, 其发生发展与脂质沉积、内皮功能障碍及慢性血管炎症密切相关[5]。尽管相关研究已较多, 但其分子机制仍未完全阐明。临床上, CAD 常伴随心肌缺血及不良心血管事件, 其本质涉及脂质代谢紊乱、内皮损伤及免疫炎症失衡, 因此解析其关键通路和分子特征对于风险分层和精准防治具有重要意义[21]。

本研究显示, CAD 样本与对照样本在免疫细胞组成上存在差异, 且 PSMA6、TOMM7 和 RPL9 与 B 细胞、T 细胞、NK 细胞及巨噬细胞等亚群关系较为密切, 提示这些枢纽基因可能参与 CAD 相关免疫调控过程。GO 及 GSEA 分析进一步表明, CAD 相关差异基因主要涉及核糖体结构、线粒体呼吸链、能量代谢、蛋白质合成及部分免疫炎症相关过程, 提示 CAD 发生发展中可能存在生物能量代谢、蛋白质翻译与免疫调控通路的协同重塑。需要说明的是, 本研究在 DEG 初筛阶段未采用更严格的 $|\log_2FC| > 1$ 阈值, 主要考虑到外周血转录组中 CAD 相关变化多呈中等效应幅度, 且后续还结合 WGCNA 模块背景和机器学习算法进行交叉优先排序, 因此该阈值更适合作为探索性候选保留。

基于 SVM-RFE 和随机森林, 本研究筛选出 3 个 CAD 枢纽基因: PSMA6、TOMM7 和 RPL9。PSMA6 与蛋白稳态维持及免疫信号调控密切相关[22]; TOMM7 参与线粒体蛋白导入和生物能量调控[23]; RPL9 则反映细胞翻译活性水平[24]。这些功能与本研究富集结果相一致, 共同提示 CAD 中存在蛋白质周转、线粒体功能及蛋白质合成异常。进一步分析表明, 3 个枢纽基因在 CAD 组中的表达均高于对照组, 且彼此存在一定关联性, 提示其可能在 CAD 病理过程中协同发挥作用。

在此基础上, 本研究构建了基于枢纽基因的 CAD 风险预测列线图。在 GSE20680 队列中, 列线图模型、PSMA6、TOMM7 和 RPL9 的 AUC 分别为 0.6541、0.6608、0.6771 和 0.5691, 提示转录组层面存在一定但有限的诊断信号。总体而言, PSMA6、TOMM7 和 RPL9 及其构建的列线图可作为 CAD 风险评估的候选生物标志物, 但其临床转化仍需进一步验证。

本研究仍存在局限性。首先, 仅分析了单一公共队列(GSE20680)的微阵列表达数据, 仍需独立数据集进行外部验证。其次, 尚未开展体内或体外实验验证, 无法进一步证明相关基因及通路的因果作用。第三, 公开数据缺乏完整且可统一建模的临床危险因素和实验室生物标志物信息, 因此未能与现有临床

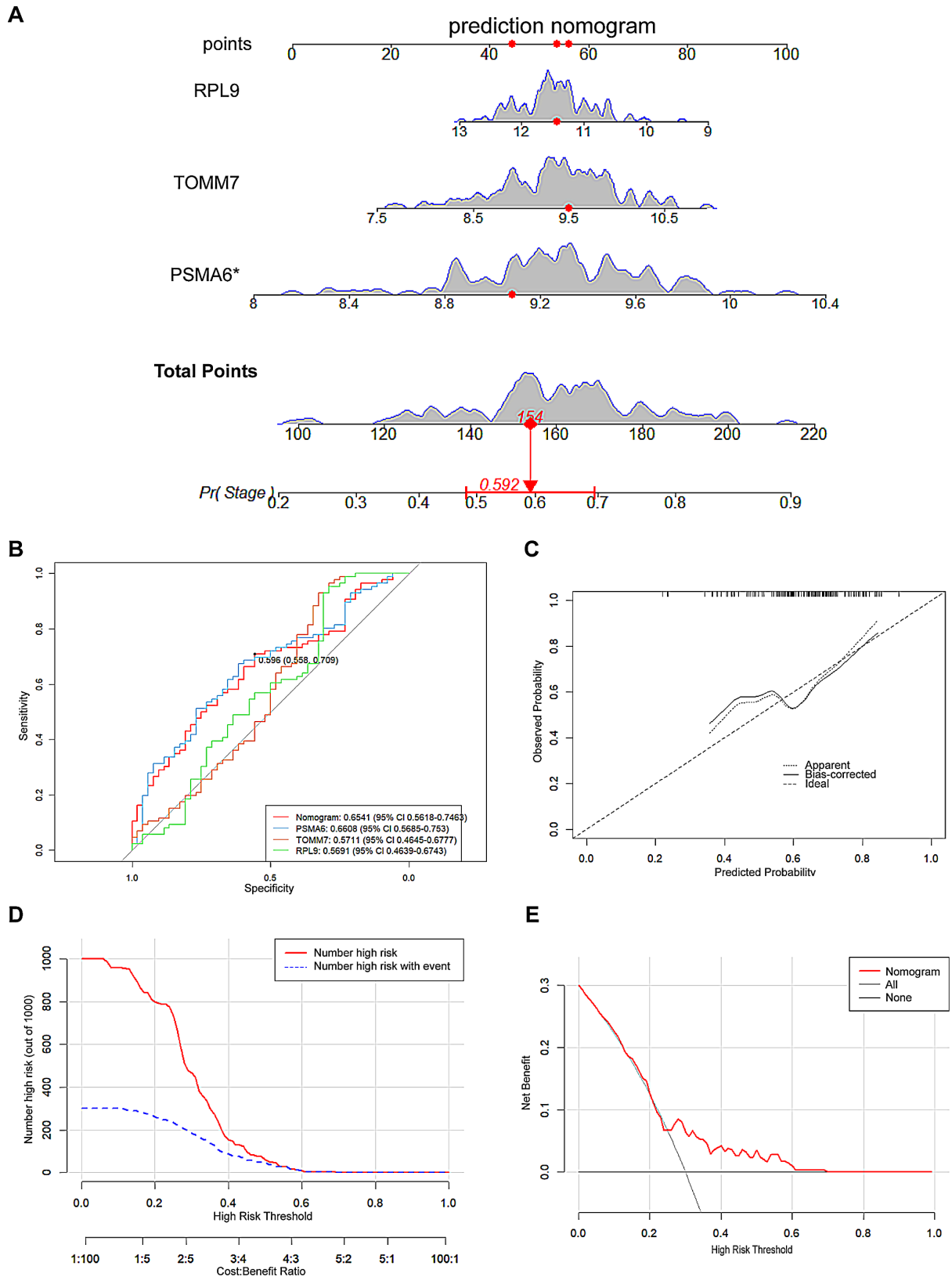


Figure 9. Validation of the CAD nomogram model based on hub genes
图 9. 基于枢纽基因的 CAD 列线图模型验证

指标模型进行直接比较, 也无法评估其增量价值。第四, 临床协变量和潜在混杂因素信息有限, 可能对血液表达谱产生影响。因此, 当前列线图更应被理解作为一种探索性的转录组候选模型, 而非成熟的临床诊断工具。未来仍需在具有完整临床变量和实验室指标的独立队列中, 系统评估其实际应用价值。

5. 结论

本研究筛选出与冠状动脉疾病相关的 3 个枢纽基因 PSMA6、TOMM7 和 RPL9, 并构建了列线图模型。该模型具有一定校准性能和决策分析价值, 但总体判别能力有限(AUC 约为 0.65)。上述基因可作为 CAD 分子分层和后续模型开发的候选生物标志物, 尚需独立队列进一步验证。

参考文献

- [1] Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., *et al.* (2020) Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019 Update from the GBD 2019 Study. *The Journal of the American College of Cardiology*, **76**, 2982-3021.
- [2] Kyu, H.H., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., *et al.* (2018) Global, Regional, and National Disability-Adjusted Life-Years (DALYs) for 359 Diseases and Injuries and Healthy Life Expectancy (HALE) for 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *The Lancet*, **392**, 1859-1922.
- [3] Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., *et al.* (2013) NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Research*, **41**, D991-D995. <https://doi.org/10.1093/nar/gks1193>
- [4] Lai, H.M., Li, X.M., Yang, Y.N., Ma, Y.T., *et al.* (2015) Genetic Variation in NFKB1 and NFKBIA and Susceptibility to Coronary Artery Disease in a Chinese Uyghur Population. *PLOS ONE*, **10**, e0129144. <https://doi.org/10.1371/journal.pone.0129144>
- [5] Libby, P. (2002) Inflammation in Atherosclerosis. *Nature*, **420**, 868-874. <https://doi.org/10.1038/nature01323>
- [6] Elashoff, M.R., Wingrove, J.A., Beineke, P., Daniels, S.E., Tingley, W.G., Rosenberg, S., *et al.* (2011) Development of a Blood-Based Gene Expression Algorithm for Assessment of Obstructive Coronary Artery Disease in Non-Diabetic Patients. *BMC Medical Genomics*, **4**, Article No. 26. <https://doi.org/10.1186/1755-8794-4-26>
- [7] 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>
- [8] Glickman, M.E., Rao, S.R. and Schultz, M.R. (2014) False Discovery Rate Control Is a Recommended Alternative to Bonferroni-Type Adjustments in Health Studies. *Journal of Clinical Epidemiology*, **67**, 850-857. <https://doi.org/10.1016/j.jclinepi.2014.03.012>
- [9] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., *et al.* (2000) Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, **25**, 25-29. <https://doi.org/10.1038/75556>
- [10] Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **28**, 27-30. <https://doi.org/10.1093/nar/28.1.27>
- [11] Yu, G., Wang, L., Han, Y. and He, Q. (2012) Clusterprofiler: An R Package for Comparing Biological Themes among Gene Clusters. *OMICS: A Journal of Integrative Biology*, **16**, 284-287. <https://doi.org/10.1089/omi.2011.0118>
- [12] Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., *et al.* (2005) Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proceedings of the National Academy of Sciences*, **102**, 15545-15550. <https://doi.org/10.1073/pnas.0506580102>
- [13] Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739-1740. <https://doi.org/10.1093/bioinformatics/btr260>
- [14] 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>
- [15] Huang, S., Cai, N., Pacheco, P.P., Narandes, S., *et al.* (2017) Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics*, **15**, 41-51.
- [16] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [17] Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate

- Descent. *Journal of Statistical Software*, **33**, 1-22.
- [18] 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. <https://doi.org/10.1038/nmeth.3337>
- [19] Harrell, F.E. (2015) Introduction. In: *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, Springer, 1-11.
- [20] Vickers, A.J. and Elkin, E.B. (2006) Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making*, **26**, 565-574. <https://doi.org/10.1177/0272989x06295361>
- [21] Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., *et al.* (2011) pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics*, **12**, Article No. 77. <https://doi.org/10.1186/1471-2105-12-77>
- [22] Bennett, D.A., Xu, P., Clarke, R., Zondervan, K., Parish, S., Palmer, A., *et al.* (2008) The Exon 1-8C/G SNP in the PSMA6 Gene Contributes Only a Small Amount to the Burden of Myocardial Infarction in 6946 Cases and 2720 Controls from a United Kingdom Population. *European Journal of Human Genetics*, **16**, 480-486. <https://doi.org/10.1038/sj.ejhg.5201948>
- [23] Shi, D., Qi, M., Zhou, L., Li, X., Ni, L., Li, C., *et al.* (2018) Endothelial Mitochondrial Preprotein Translocase Tomm7-Rac1 Signaling Axis Dominates Cerebrovascular Network Homeostasis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **38**, 2665-2677. <https://doi.org/10.1161/atvbaha.118.311538>
- [24] Lezzerini, M., Penzo, M., O'Donohue, M., Marques dos Santos Vieira, C., Saby, M., Elfrink, H.L., *et al.* (2020) Ribosomal Protein Gene RPL9 Variants Can Differentially Impair Ribosome Function and Cellular Metabolism. *Nucleic Acids Research*, **48**, 770-787. <https://doi.org/10.1093/nar/gkz1042>