DOI: 10.14188/j.1671-8852.2024.0721 **文章编号** 1671-8852(2025)11-1429-07

基于免疫相关基因的HER2阳性乳腺癌新辅助化疗联合靶向治疗疗效预测模型

陈梦婷 陈芳芳 张京伟
武汉大学中南医院甲状腺乳腺外科 湖北 武汉 430071

摘要 目的：构建基于免疫相关基因的预测模型，评估HER2阳性乳腺癌患者新辅助化疗联合靶向治疗的疗效，并探讨与免疫微环境的关系。方法：从GEO数据库获取新辅助化疗联合靶向治疗HER2阳性乳腺癌样本数据，结合差异基因、WGCNA挖掘的免疫相关基因模块、ImmPort数据库的免疫基因集，通过LASSO回归构建预测模型，评估不同风险组的免疫微环境和免疫治疗潜力。结果：筛选出APOD、CCL18、CCR7、CXCL13、GDF15、IGHM、S100A8、SERPINA3八个关键基因构建的预测模型在训练集和验证集上表现出良好的预测效能，低风险组患者免疫活性较高，显示更大的免疫治疗潜力。结论：该预测模型能够有效预测HER2阳性乳腺癌新辅助治疗效果，且免疫微环境在治疗中起重要作用。

关键词 HER2阳性乳腺癌；新辅助治疗；靶向治疗；免疫浸润；预测模型
中图分类号 R737.9
文献标识码 A **开放科学（资源服务）标识码（OSID）：**

Predictive model for neoadjuvant chemotherapy combined with targeted therapy in HER2-positive breast cancer based on immune-related genes

CHEN Mengting, CHEN Fangfang, ZHANG Jingwei
Department of Thyroid and Breast Surgery, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei, China

Abstract

Objective: To construct a predictive model based on immune-related genes to assess the efficacy of neoadjuvant chemotherapy combined with targeted therapy in HER2-positive breast cancer patients and explore its association with the immune microenvironment. **Methods:** The data of HER2-positive breast cancer samples receiving neoadjuvant chemotherapy combined with targeted therapy were obtained from the GEO database. Differentially expressed genes, immune-related gene modules identified by WGCNA, and immune gene sets from the ImmPort database were integrated, and a predictive model was constructed using LASSO regression. The immune microenvironment and immunotherapy potential were evaluated in different risk groups. **Results:** Eight key genes (APOD, CCL18, CCR7, CXCL13, GDF15, IGHM, S100A8, SERPINA3) were selected to construct the predictive model, which showed good predictive performance in both the training and validation sets. The low-risk group exhibited higher immune activity and greater immune therapy potential. **Conclusion:** This predictive model can effectively predict the efficacy of neoadjuvant therapy in HER2-positive breast

作者简介：陈梦婷，女，1998—，医学硕士生，主要从事乳腺癌诊疗相关研究，E-mail：chenmt@whu.edu.cn
通信作者：张京伟，男，1968—，医学博士，教授，博士生导师，主要从事乳腺癌的机制及预后等相关研究，E-mail：zn001213@whu.edu.cn

cancer, and the immune microenvironment plays an important role in the treatment.

Key Words HER2-Positive Breast Cancer; Neoadjuvant Therapy; Targeted Therapy; Immune Infiltration; Predictive Model

世界卫生组织国际癌症研究机构最新报告显示,2022年全球乳腺癌的发病率仅次于肺癌,位居第二,并且在女性中乳腺癌的发病率和致死率均为最高[1]。其中人类表皮生长因子受体2(HER2)阳性的乳腺癌约占15%~20%,且与高复发率和低生存率相关[2]。对于局部晚期的乳腺癌患者,新辅助治疗可缩小肿瘤范围,增加可手术切除的机会,甚至实现保乳和保留腋窝。目前,帕妥珠单抗和曲妥珠单抗双重靶向治疗联合化疗成为了HER2阳性乳腺癌的国际标准一线治疗,病理完全缓解率(pCR)可达约60%[3],但仍有部分患者无法获益。

目前临床主要依据经典的临床和病理信息(如肿瘤大小、淋巴结或远处转移状态)辅助决策HER2阳性乳腺癌患者的治疗,实际应用的生物标志物仅限于HER2的过表达或扩增。因此,迫切需要探索更多有潜在应用价值的生物标志物来预测治疗效果。研究表明,免疫浸润较高的患者更可能从HER2靶向治疗中受益。已有研究[4]发现,肿瘤浸润淋巴细胞(TIL)水平与接受含曲妥珠单抗、帕妥珠单抗新辅助治疗方案的HER2阳性乳腺癌病人的pCR率正相关。此外,抗体依赖性的细胞介导的细胞毒作用(ADCC)是曲妥珠单抗的主要作用机制之一[5],且其还可通过作用于巨噬细胞上的Fcγ受体促进抗体依赖性细胞吞噬作用(ADCP)[6]。因此,免疫特征是预测HER2阳性乳腺癌治疗功效的重要预后指标。本研究旨在探索能有效预测HER2阳性乳腺癌病人新辅助治疗效果的免疫相关基因,并构建预测模型。

1 材料与方法

1.1 数据收集与整理 从GEO数据库中筛选出GSE22358、GSE37946、GSE42822、GSE66399四个数据集,包含接受新辅助化疗联合靶向治疗方案的HER2阳性乳腺癌患者信息及治疗前乳腺癌穿刺样本的转录组数据。以GSE22358、GSE37946、GSE4282作为训练集,下载并合并转录组表达矩阵,使用ComBat方法校正批次效应,共101个样本,其中53个达到pCR,48个为病灶残余(RD)。

1.2 免疫浸润分析 使用R包"MCP-counter"量化9种免疫相关细胞,利用R包"estimate"计算了每个样本的基质评分(Stromal Score)、免疫评分(Immune Score)及综合评分(ESTIMATE Score),比较pCR组与RD组的免疫浸润水平。

1.3 筛选关键基因及构建预测模型 在训练集中,使用R包"limma"进行差异基因(DEGs)分析(调整后的$P<0.05$,$|\log_2 \text{fold change(FC)}|>0.5$)。取方差排名前5 000的基因使用R包"WGCNA"进行加权基因共表达网络分析(WGCNA)来构建基因共表达网络,以筛选出与MCP-counter、Estimate评分显著相关的模块。从ImmPort数据库下载得到1 793个免疫相关的基因。用R包"VennDiagram"将差异基因、WGCNA基因模块和ImmPort免疫相关基因集取交集。使用LASSO回归分析,利用R包"glmnet"构建基因模型,风险评分公式如下:

$$\text{风险评分}=\sum_{i=1}^{n}(Exp_i \times Coef_i)$$

其中,Exp_i表示基因的表达量,$Coef_i$为相应基因的回归系数。根据评分中位数将样本分为高风险组和低风险组。使用R包"survivalROC"中的受试者工作特征(ROC)曲线评估基因模型的性能。此外,我们还使用GSE66399队列(含23个样本)作为独立数据集验证了模型的性能。

1.4 关键基因与免疫细胞浸润的相关性分析 TIMER肿瘤免疫浸润评估资源数据库通过TIMER算法系统地分析各种癌症类型中的浸润免疫细胞[7]。我们分析了所筛选的关键基因与免疫细胞浸润的相关性。

1.5 信号通路富集分析 使用基因集富集分析(GSEA),分析不同风险组中的显著富集通路。选取KEGG和Hallmark基因集进行分析,采用1 000次随机抽样,若$P<0.05$,错误发现率(FDR)<0.25,则结果具有统计学意义。

1.6 免疫微环境和免疫治疗潜力评估

1.6.1 IPS评分 IPS评分综合了免疫细胞活性、免疫调节因子和MHC分子等信息,反映了肿瘤微环境的免疫活跃程度。通过Cancer Immunome Atlas(TCIA)在线工具计算GEO样本的IPS评分,比较高风险组与低风险组的差异[8]。

1.6.2　TIDE 评分　TIDE 评分通过分析肿瘤中的 T 细胞功能障碍和免疫排斥状态来评估免疫治疗的潜在效果，能够预测免疫治疗对不同肿瘤的响应性。使用 TIDE 官网工具计算 GEO 样本的 TIDE 评分，比较高风险组与低风险组的差异[9]。

1.6.3　免疫检查点表达水平分析　选择了多个免疫检查点基因（如 PD-1、CTLA-4、CD80 等），并使用 R 语言中的 limma 包对 GEO 数据库的高风险组和低风险组样本的基因表达数据进行差异分析。

1.6.4　肿瘤突变负荷（TMB）分析　TMB 水平高的肿瘤能被免疫系统识别的肿瘤新抗原数量可能越多，更容易对免疫检查点抑制剂药物产生反应，激活免疫系统，更好地识别、杀死癌细胞[10]。在 TCGA 队列中，使用 R 包“maftools”可视化 TMB 水平，比较高风险组与低风险组之间的差异。

1.7　统计学方法　所有数据均采用 R 4.3.3 软件统计分析，两组间比较采用 Wilcoxon 符号秩检验，相关性分析采用 Spearman 方法。$P<0.05$ 则差异具有统计学意义。

2　结果

2.1　pCR 组和 RD 组免疫浸润的差别　pCR 组的免疫细胞浸润程度普遍高于 RD 组（图 1A），细胞毒性 T 细胞（图 1B）、B 细胞（图 1C）、NK 细胞（图 1D）、单核细胞（图 1E）在 pCR 组均表现出显著较高的浸润。pCR 组的基质评分（Stromal Score）（图 1F）、免疫评分（Immune Score）（图 1G）、ESTIMATE 评分（图 1H）也显著高于 RD 组，提示 pCR 组免疫活性更强。

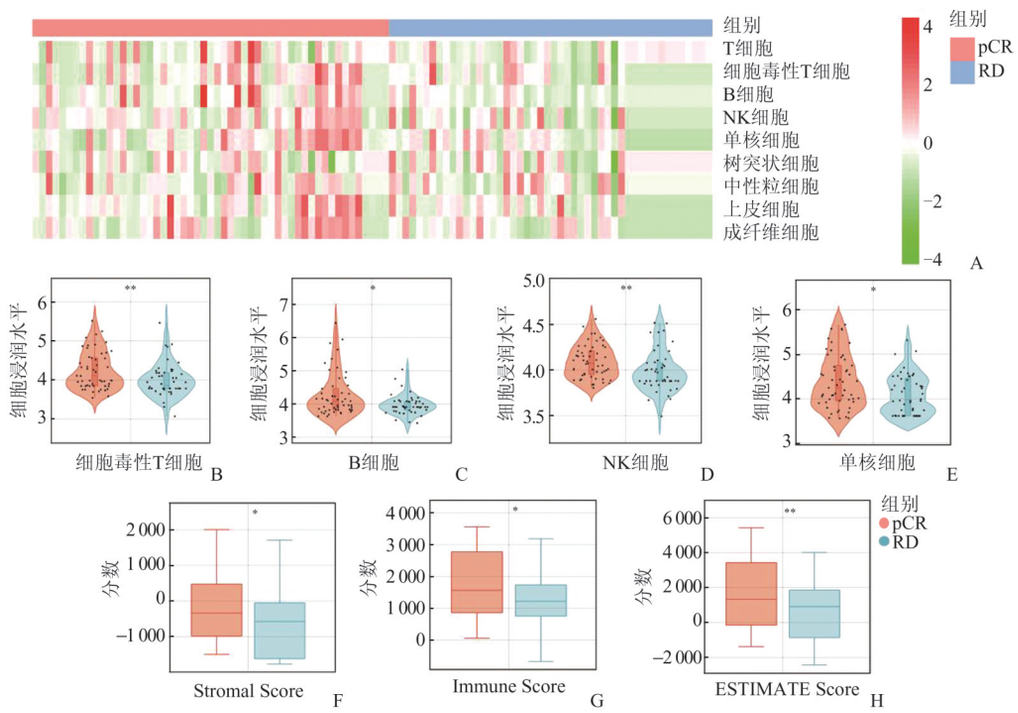


图 1　pCR 组和 RD 组的免疫浸润分析和比较

A. MCP-counter 算法分析 pCR 组和 RD 组免疫细胞浸润水平的热图；B～E. pCR 组和 RD 组细胞毒性 T 细胞、B 细胞、NK 细胞、单核细胞浸润水平的比较；F～H. ESTIMATE 算法比较 pCR 组和 RD 组的基质评分、免疫评分、ESTIMATE 评分。$^{*}P<0.05$；$^{**}P<0.01$。

2.2　基因模型的构建与验证　在训练集中，筛选出 137 个差异基因，其中 106 个在 pCR 组上调，31 个下调（图 2A）。WGCNA 分析得到 7 个基因模块（图 2B），其中棕色、绿色、黄色模块与多数免疫细胞的浸润程度、ESTIMATE 评分显著正相关，蓝绿色模块则显著负相关，共得到 3 080 个与免疫浸润相关的基因。从 ImmPort 数据库获取 1 793 个免疫相关基因。结合差异基因、WGCNA 模块和 ImmPort 数据库的免疫基因，共筛选出 39 个关键基因（图 2C）。

随后，LASSO 回归分析筛选出 8 个关键基因（图 2D，图 2E），并构建风险评分模型，公式如下：风险评分＝（−0.083 81）×APOD＋（−0.262 86）×CCL18＋（−0.293 14）×CCR7＋（−0.034 09）×CXCL13＋0.133 91×GDF15＋（−0.069 72）×IGHM＋（−0.015 92）×S100A8＋0.102 23×

SERPINA3。根据中位数风险评分将样本分为高低风险组。在训练集和验证集中均显示该模型具有良好的预测性能(训练集AUC=0.772,验证集AUC=0.863)(图2F,2G)。

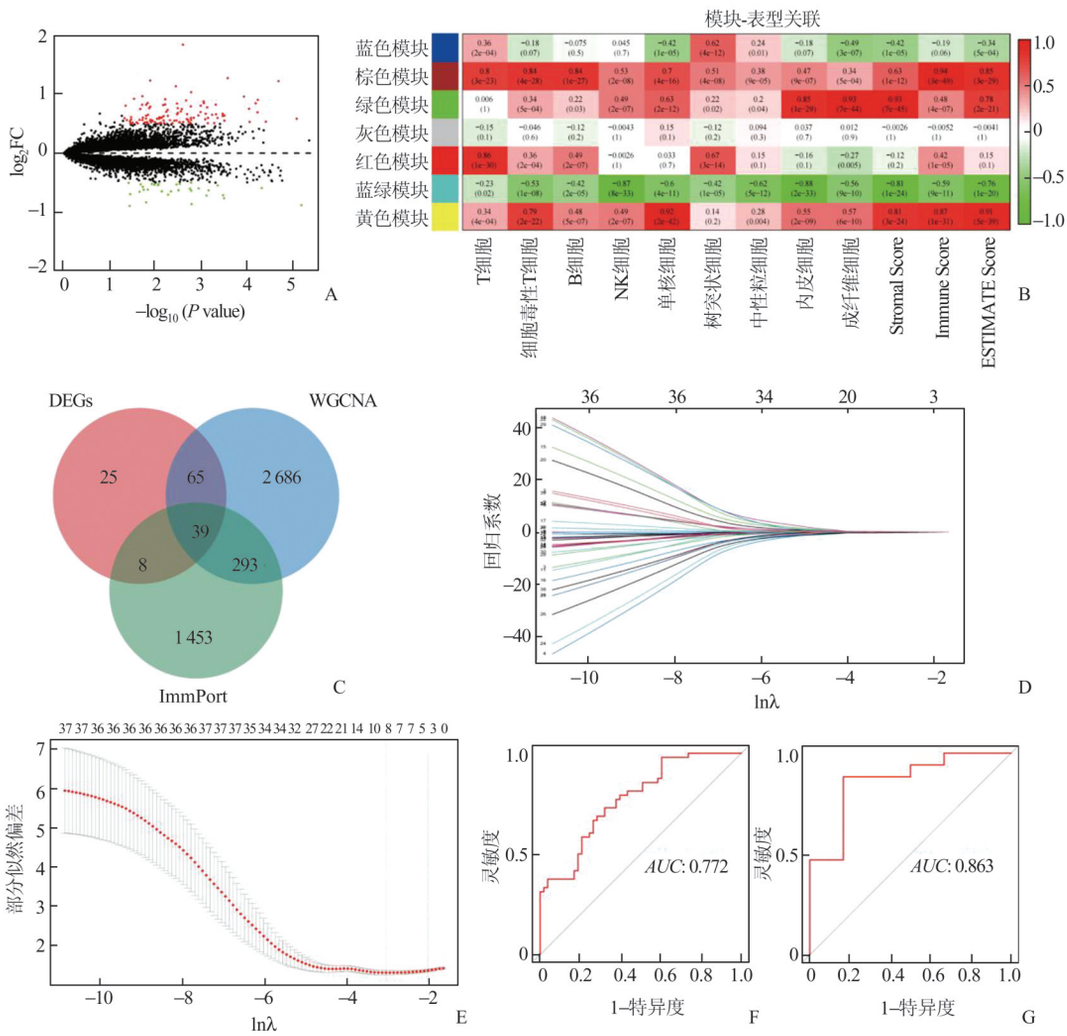


图2 HER2阳性乳腺癌新辅助化疗联合靶向治疗效果基因预测模型的构建

A. pCR组和RD组差异表达基因的火山图;B. WGCNA分析所得基因模块与MCP-counter、Estimate评分之间的相关性热图;C. 差异基因、WGCNA免疫高度相关基因模块、ImmPort数据库来源免疫相关基因集的韦恩图;D. 惩罚线性回归模型的参数选择;E. 正则化路径分析;F. 训练集的ROC曲线;G. 验证集的ROC曲线。

2.3 关键基因与免疫细胞浸润的相关性分析 APOD、CCL18、CCR7、CXCL13、IGHM和S100A8与风险评分负相关,GDF15和SERPINA3与风险评分正相关,提示前者与治疗敏感性正相关,后者与耐药性相关。TIMER分析显示,APOD与$CD8^+$ T细胞浸润正相关(图3A),CCL18、CCR7、CXCL13与$CD8^+$ T细胞、$CD4^+$ T细胞、NK细胞和巨噬细胞的浸润均正相关(图3B~3M)。GDF15与$CD8^+$ T细胞、$CD4^+$ T细胞、NK细胞和巨噬细胞的浸润负相关(图3N~3Q),S100A8与NK细胞和巨噬细胞的浸润正相关(图3R, 3S),SERPINA3与免疫细胞浸润负相关(图3T, 3U)。IGHM编码mu重链的C区,它定义了IgM同种型,在B细胞活化中占重要地位[11]。

2.4 高风险组和低风险组的GSEA分析 根据KEGG基因集进行的GSEA分析,结果显示抗原加工和呈递、细胞黏附分子、趋化因子、细胞因子相互作用、白细胞跨内皮迁移、自然杀伤细胞介导的细胞毒性、TOLL样受体相关的信号通路在低风险组富集,而细胞周期、DNA复制、核苷酸切除修复相关通路在高风险组富集(图4A)。根据Hallmark基因集进行的GSEA分析结果显示补体、IL-6/JAK/STAT3、炎症反应、干扰素α应答、干扰素γ应答相关的通路在低风险组富集,而E2F靶基因、雌激素

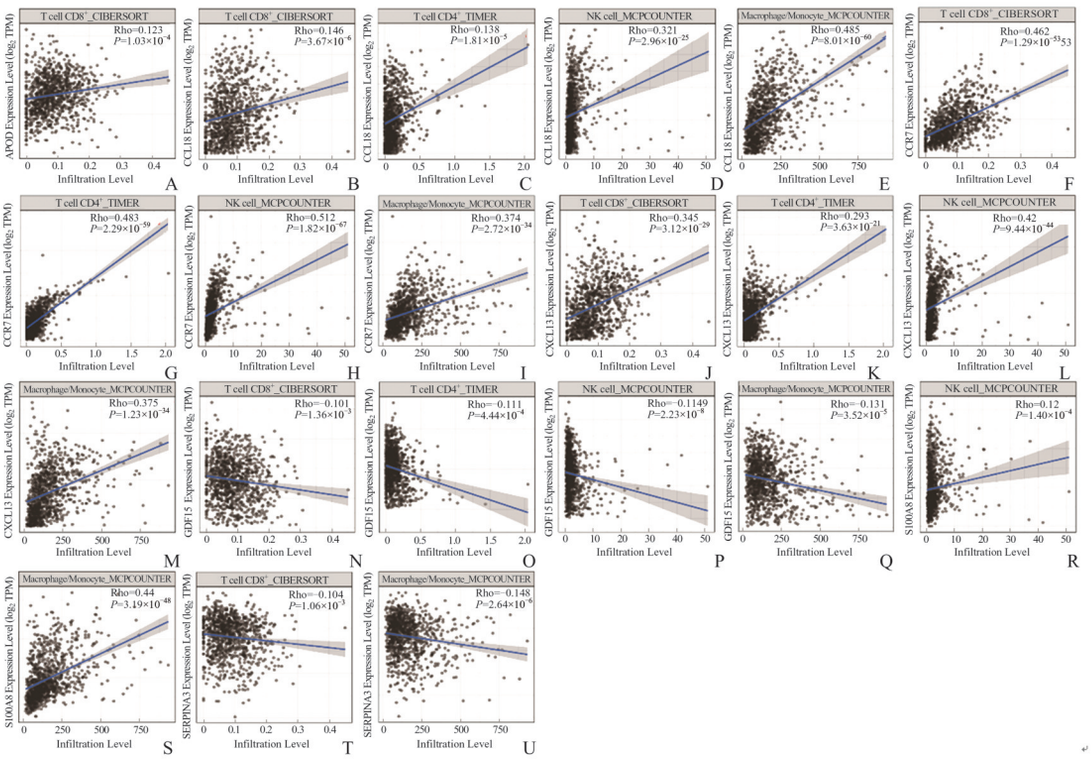


图 3　关键基因的表达水平与免疫细胞浸润水平的相关性

A. APOD 与 $CD8^+$ T 细胞 TIMER 数据库分析结果散点图；B～E. CCL18 与 $CD8^+$ T 细胞、$CD4^+$ T 细胞、NK 细胞、巨噬细胞 TIMER 数据库分析结果散点图；F～I. CCR7 与 $CD8^+$ T 细胞、$CD4^+$ T 细胞、NK 细胞、巨噬细胞 TIMER 数据库分析结果散点图；J～M. CXCL13 与 $CD8^+$ T 细胞、$CD4^+$ T 细胞、NK 细胞、巨噬细胞 TIMER 数据库分析结果散点图；N～Q. GDF15 与 $CD8^+$ T 细胞、$CD4^+$ T 细胞、NK 细胞、巨噬细胞 TIMER 数据库分析结果散点图；R～S. S100A8 与 NK 细胞、巨噬细胞 TIMER 数据库分析结果散点图；T～U. SERPINA3 与 $CD8^+$ T 细胞、巨噬细胞 TIMER 数据库分析结果散点图。

反应、G_2/M 检查点、MYC 靶基因相关通路在高风险组富集（图 4B）。

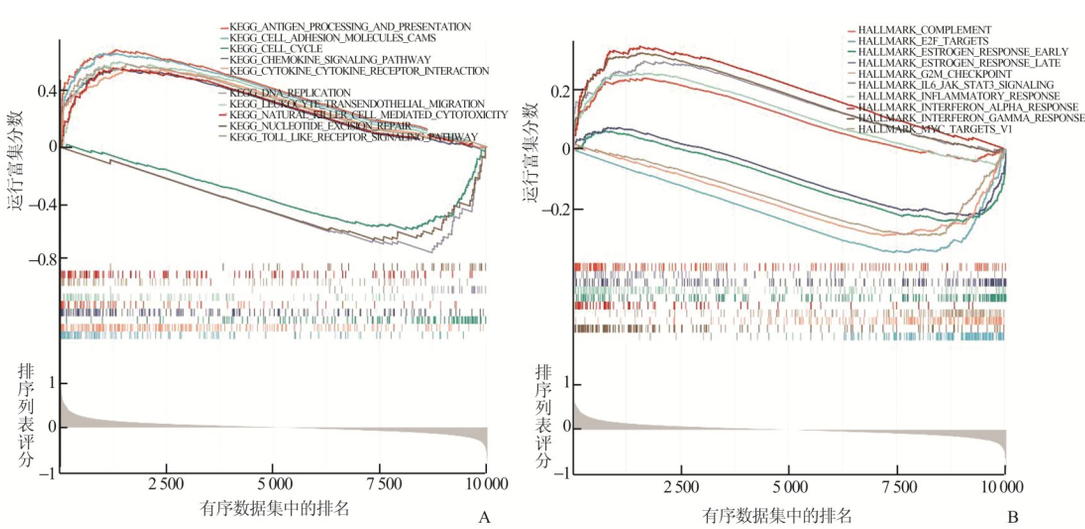


图 4　高风险组和低风险组的 GSEA 分析

A. 基于 KEGG 基因集的 GSEA 通路富集图；B. 基于 Hallmark 基因集的 GSEA 通路富集图。

2.5　免疫微环境和免疫治疗潜力的评估

2.5.1　IPS 评分和 TIDE 评分　低风险组的 IPS 评分显著高于高风险组（$P<0.01$），表明低风险组免疫反应更活跃（图 5A）；低风险组 TIDE 评分显著低于高风险组（$P<0.000\,1$），提示其免疫抑制状态较轻，免疫逃逸机制较弱，可能更易响应免疫治疗（图 5B）。

2.5.2 免疫检查点基因表达 低风险组的免疫检查点基因(如CD160、CTLA-4、PDCD1等)表达普遍高于高风险组($P<0.05$)(图5C),表明低风险组可能存在更多免疫治疗靶点。

2.5.3 TMB分析 TCGA队列的体细胞突变数据显示,高低风险组在前20个突变基因谱上高度重叠,但高风险组特有的突变基因(MAP2K4、ZFHX4、TBX3、DMD、ARID1A)与低风险组的特异性基因(MUC17、SPTA1、RYR3、APOB、CSMD3)有所不同(图5D,5E)。低风险组TMB水平显著高于高风险组($P<0.05$)(图F),提示低风险组对免疫治疗可能具有更高响应度。

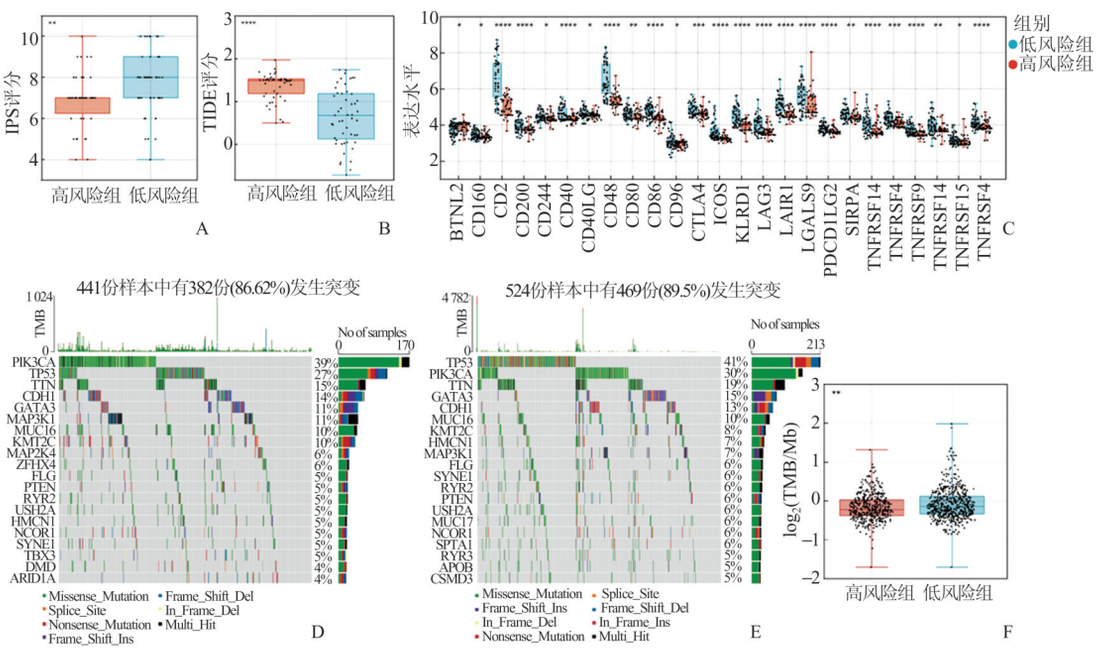


图5 高风险组和低风险组的TMB分析和比较

A. 高风险组和低风险组IPS评分的比较;B. 高风险组和低风险组TIDE评分的比较;C. 高风险组和低风险组的免疫检查点表达水平;D. TCGA队列中高风险组的TMB瀑布图;E. TCGA队列中低风险组的TMB瀑布图;F. 高风险组和低风险组TMB水平的比较。$^{*}P<0.05$;$^{**}P<0.01$;$^{***}P<0.001$;$^{****}P<0.0001$。

3 讨论

尽管HER2阳性乳腺癌患者在新辅助化疗联合靶向治疗中取得了一定疗效,但仍有一部分患者无法获益。因此,需要寻找新的生物标志物来预测治疗效果并指导个体化治疗。先前有研究[4-6]提出HER2阳性乳腺癌患者的免疫特征与含靶向药物的新辅助治疗效果相关。本研究构建了基于免疫相关基因的疗效预测模型,深入分析免疫微环境并探讨其对治疗反应的潜在机制。结果表明,免疫微环境对治疗反应起着至关重要的作用,筛选出的关键基因能够有效预测疗效,并为临床治疗提供新方向。

我们发现,pCR组患者的免疫细胞浸润程度显著高于RD组,提示免疫活跃的肿瘤微环境可能是影响新辅助治疗疗效的关键因素。随后,我们筛选出8个关键基因,TIMER分析显示与治疗敏感性正相关的基因(APOD、CCL18、CCR7、CXCL13、IGHM、S100A8)与多种免疫细胞(如$CD8^+$ T细胞、$CD4^+$ T细胞、NK细胞和巨噬细胞)的浸润正相关。既往研究[12]发现,CCR7是细胞迁移至次级淋巴器官的关键趋化因子受体,能够促进免疫应答;CXCL13在淋巴器官发育、淋巴细胞的趋化和分化中起着关键作用,CXCL13的高表达与乳腺癌预后改善相关[13];IGHM编码mu重链的C区,在B细胞活化中占重要地位,与免疫应答及细胞吞噬作用密切相关,与乳腺癌良好的预后相关[11];S100A8可促进白细胞的黏附、聚集、迁移等活动,从而引起免疫反应,在癌症免疫调节和靶向治疗中有着潜在的作用[14]。相反,与治疗敏感性负相关的基因(GDF15、SERPINA3)与多种免疫细胞(如$CD8^+$ T细胞、$CD4^+$ T细胞、NK细胞和巨噬细胞)的浸润水平负相关。既往研究[15]发现,GDF15抑制树突状细胞成熟和功能,进一步抑制T淋巴细胞活化,从而促进

肿瘤细胞免疫逃逸；而SERPINA3通过调控免疫逃逸相关通路，可能进一步促进肿瘤的免疫抑制状态[16]。这些基因的表达特点进一步揭示了治疗敏感性与肿瘤免疫微环境之间的复杂关系。

低风险组患者表现出更为活跃的免疫微环境，其IPS评分显著高于高风险组，提示其免疫细胞的功能更强。其免疫检查点基因表达水平普遍更高，可能增强抗肿瘤免疫反应。其TIDE评分低、TMB高，表明免疫逃逸机制较弱，肿瘤免疫原性较高，更可能受益于免疫治疗。与此相反，高风险组表现出更显著的免疫抑制状态。此外，高风险组的肿瘤富集了与DNA修复和细胞周期相关的通路，提示其肿瘤可能通过增强增殖信号和遗传稳定性来抵抗治疗。因此，对于高风险组患者，可尝试结合PARP抑制剂靶向DNA修复通路，或联合其他免疫治疗手段克服其免疫抑制状态。

本研究验证了基于免疫相关基因的预测模型在HER2阳性乳腺癌中的有效性，仍存在一些局限性。首先，研究基于公开数据库，缺乏真实世界样本验证，未来需要通过前瞻性临床研究验证模型的适用性。其次，关键基因在免疫调控中的机制尚不完全明确，需通过体外和体内实验进一步探索。总之，本研究揭示了免疫微环境在HER2阳性乳腺癌新辅助治疗中的关键作用，并为个体化治疗策略的优化提供了理论依据。进一步的临床研究和功能实验将帮助验证这些发现，并推动其应用。

参考文献

[1] BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA A Cancer J Clinicians, 2024, 74(3): 229-263.

[2] CHOONG G M, CULLEN G D, O′SULLIVAN C C. Evolving standards of care and new challenges in the management of HER2-positive breast cancer[J]. CA A Cancer J Clin, 2020, 70(5): 355-374.

[3] TAKADA M, TOI M. Neoadjuvant treatment for HER2-positive breast cancer[J]. Chin Clin Oncol, 2020, 9(3): 32.

[4] HWANG H W, JUNG H, HYEON J, et al. A nomogram to predict pathologic complete response (pCR) and the value of tumor-infiltrating lymphocytes (TILs) for prediction of response to neoadjuvant chemotherapy (NAC) in breast cancer patients[J]. Breast Cancer Res Treat, 2019, 173(2): 255-266.

[5] SCHLAM I, TARANTINO P, TOLANEY S M. Overcoming resistance to HER2-directed therapies in breast cancer[J]. Cancers, 2022, 14(16): 3996.

[6] SWAIN S M, SHASTRY M, HAMILTON E. Targeting HER2-positive breast cancer: Advances and future directions[J]. Nat Rev Drug Discov, 2023, 22(2): 101-126.

[7] LI T, FU J, ZENG Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells[J]. Nucleic Acids Res, 2020, 48(W1): W509-W514.

[8] GUO J N, CHEN D, DENG S H, et al. Identification and quantification of immune infiltration landscape on therapy and prognosis in left- and right-sided colon cancer[J]. Cancer Immunol Immunother, 2022, 71(6): 1313-1330.

[9] ZHENG K, HAI Y, CHEN H, et al. Tumor immune dysfunction and exclusion subtypes in bladder cancer and pan-cancer: A novel molecular subtyping strategy and immunotherapeutic prediction model[J]. J Transl Med, 2024, 22(1): 365.

[10] CAO D, XU H, XU X, et al. High tumor mutation burden predicts better efficacy of immunotherapy: A pooled analysis of 103078 cancer patients[J]. OncoImmunology, 2019, 8(9): e1629258.

[11] LEE H, KWON M J, KOO B M, et al. A novel immune prognostic index for stratification of high-risk patients with early breast cancer[J]. Sci Rep, 2021, 11(1): 128.

[12] BILL C A, ALLEN C M, VINES C M. C-C chemokine receptor 7 in cancer[J]. Cells, 2022, 11(4): 656.

[13] CRISCITIELLO C, BAYAR M A, CURIGLIANO G, et al. A gene signature to predict high tumor-infiltrating lymphocytes after neoadjuvant chemotherapy and outcome in patients with triple-negative breast cancer[J]. Ann Oncol, 2018, 29(1): 162-169.

[14] SONG R, STRUHL K. S100A8/S100A9 cytokine acts as a transcriptional coactivator during breast cellular transformation[J]. Sci Adv, 2021, 7(1): eabe5357.

[15] HAAKE M, HAACK B, SCHÄFER T, et al. Tumor-derived GDF-15 blocks LFA-1 dependent T cell recruitment and suppresses responses to anti-PD-1 treatment[J]. Nat Commun, 2023, 14(1): 4253.

[16] YUAN Q, WANG S, ZHANG G, et al. Highly expressed of SERPINA3 indicated poor prognosis and involved in immune suppression in glioma[J]. Immun Inflamm Dis, 2021, 9(4): 1618-1630.

（2024-09-21 收稿）
编辑　孙孝云