

# HR+/HER2-乳腺癌高复发风险与DNA修复异常有关

## ——基于METABRIC数据库分析

黄永, 魏一睿, 李朝阳, 李一凤, 程巧\*

重庆医科大学附属第一医院, 乳腺甲状腺外科, 重庆

收稿日期: 2026年2月13日; 录用日期: 2026年3月6日; 发布日期: 2026年3月17日

### 摘要

目的: 乳腺癌是全球最常见的恶性肿瘤之一, 其中激素受体阳性(HR+)/人表皮生长因子受体2阴性(HER2-)亚型通常预后较好。然而, 约5%~10%的HR+/HER2-患者在2.5年内发生转移, 这凸显了优化风险分层和实施靶向干预的迫切性。本研究旨在鉴定该高危亚群的遗传特征并探索潜在的治疗靶点。方法: 利用METABRIC数据库数据, 根据临床预后将患者分为“高危”和“低危”两组。通过多组学差异分析, 系统表征高危患者的分子图谱。利用TCGA乳腺癌队列对研究结果的稳健性进行外部验证。结果: 多组学分析揭示了高危HR+/HER2-乳腺癌患者独特的分子特征。通路富集分析显示关键生物学过程显著受损, 特别是DNA双链断裂修复机制。外部验证显示, 较高的同源重组缺陷(HRD)评分与较差的无进展生存期相关。结论: 这些发现加深了对高危HR+/HER2-乳腺癌遗传和分子驱动因素的理解, 并为开发靶向治疗策略提供了关键见解, 有助于推动精准肿瘤学的发展。

### 关键词

乳腺癌, HR+/HER2-, 遗传特征, DNA修复, 精准肿瘤学

# High Recurrence Risk in HR+/HER2- Breast Cancer Is Associated with DNA Repair Defects

## —An Analysis Based on the METABRIC Database

Yong Huang, Yirui Wei, Zhaoyang Li, Yifeng Li, Qiao Cheng\*

\*通讯作者。

文章引用: 黄永, 魏一睿, 李朝阳, 李一凤, 程巧. HR+/HER2-乳腺癌高复发风险与DNA修复异常有关[J]. 临床医学进展, 2026, 16(3): 2964-2978. DOI: 10.12677/acm.2026.1631100

Department of Breast and Thyroid Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing

Received: February 13, 2026; accepted: March 6, 2026; published: March 17, 2026

## Abstract

**Objective:** Breast cancer remains one of the most prevalent malignancies worldwide, with the hormone receptor-positive (HR+)/human epidermal growth factor receptor 2-negative (HER2-) subtype generally associated with favorable outcomes. However, approximately 5%~10% of HR+/HER2- patients develop metastases within 2.5 years, highlighting the need for improved risk stratification and targeted interventions. This study aims to identify the genetic characteristics of this high-risk subgroup and explore potential therapeutic targets. **Methods:** Using data from the METABRIC database, patients were classified into distinct “high-risk” and “low-risk” categories based on clinical outcomes. Multi-omics differential analysis was performed to systematically characterize the molecular landscape of high-risk patients. The robustness of the findings was externally validated using the TCGA breast cancer cohort. **Results:** Multi-omics analysis revealed distinct molecular features in high-risk HR+/HER2- breast cancer patients. Pathway enrichment analysis demonstrated significant impairments in key biological processes, particularly in DNA double-strand break repair mechanisms. In external validation, higher homologous recombination deficiency (HRD) scores were associated with poorer progression-free survival. **Conclusion:** These findings enhance the understanding of the genetic and molecular drivers of high-risk HR+/HER2- breast cancer and provide critical insights for developing targeted therapeutic strategies, contributing to advances in precision oncology.

## Keywords

Breast Cancer, HR+/HER2-, Genetic Characteristics, DNA Repair, Precision Oncology

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

乳腺癌已成为全球女性最常见的恶性肿瘤[1]。虽然在西方发达国家它仍是女性死亡的主要原因之一，但在中国，其发病率正迅速上升，并伴有发病年龄年轻化的趋势。临床上，根据激素受体(HR)状态——包括雌激素受体(ER)和孕激素受体(PR)——以及人表皮生长因子受体 2 (HER2)的表达，将乳腺癌分为四种主要亚型。其中，HR 阳性/HER2 阴性(HR+/HER2-)乳腺癌约占所有病例的三分之二，因其相对惰性的生物学行为和对内分泌治疗的敏感性，传统上被视为“预后良好”的亚型[2]-[5]。

然而，HR+/HER2-乳腺癌在生物学上并非单一、同质的疾病，其显著的内在异质性导致不同患者治疗反应及复发模式上存在巨大差异[6]。尽管接受了标准治疗，仍有 15%的患者在 5 年内复发，5%~10%的患者在 2.5 年内发生转移[7]-[10]，表明该亚群面临更高的复发风险和更差的生存预后。这种临床变异性强调了进行更深层次分子特征分析的必要性，以识别高危亚群并指导靶向治疗策略。

分子谱分析的进步进一步凸显了 HR+/HER2-乳腺癌内部的巨大异质性。基于 PAM50 的基因谱分析

显示,免疫组化(IHC)与分子分型之间存在 37.8%的不一致率,其中 27.0%的 IHC 定义的 Luminal A 型实际上被归类为 Luminal B 型,10.2%归类为 HER2 富集型[4] [11]。在基因组水平上,Luminal A 型肿瘤表现出独特的突变谱,其特征是 TP53 突变率较低(12% vs Luminal B 型的 29%),但 PIK3CA (45% vs 29%)和 MAP3K1 (13% vs 5%)的突变频率较高[12]。这些分子差异导致了临床观察中预后和治疗反应的异质性。

虽然目前的风险分层工具(如 21 基因和 70 基因检测)改善了 HR+/HER2-患者的化疗决策,但它们在准确识别具有生物学侵袭性的亚群方面仍有局限[13]-[18]。最近的多组学分析为这种异质性提供了更深入的见解。复旦大学附属肿瘤医院 2023 年的一项研究利用整合基因组、转录组和代谢组谱分析,将 HR+/HER2-乳腺癌分为四个不同的分子亚群,强调了治疗策略的潜在差异[19]。CDK4/6 抑制剂的开发和临床成功(如 MonarchE 和 NATALEE 试验所示)进一步强调了精确识别高危人群的必要性。

然而,驱动 HR+/HER2-乳腺癌极差预后(如早期复发和转移)的关键分子机制仍知之甚少。特别是,驱动高危疾病发生和发展的遗传及分子特征尚未得到全面系统的阐明[20]。这一知识空白限制了开发靶向治疗策略的能力,使得部分高危患者缺乏最佳治疗选择。在此背景下,本研究利用 METABRIC 数据库调查与 HR+/HER2-乳腺癌极差预后相关的多组学特征。通过全面的差异和富集分析,我们旨在阐明驱动高危疾病的分子机制并确定潜在的治疗靶点,为推进精准肿瘤学提供有价值的见解。

## 2. 材料与方法

### 2.1. 数据获取与整合

多组学数据,包括临床信息、基因组改变(SNV/INDEL)、拷贝数变异(CNV)、转录组谱(芯片)和 DNA 甲基化数据(RRBS),均通过 cBioPortal for Cancer Genomics 从 METABRIC 数据库获取。经过严格的质量控制和数据过滤,共有 965 例具有完整临床和多组学资料的 HR+/HER2-乳腺癌样本纳入分析。

### 2.2. 多组学差异分析

为了鉴定与 HR+/HER2-乳腺癌预后相关的关键分子特征,使用 R 包“MOVICS”(v0.99.17)在预后不良和预后良好患者组之间进行了全面的多组学差异分析。所有分析均在 R(v4.2.2)中进行,并由 CancerSubtypes 包(v1.24.0)提供支持。本研究特别关注药物敏感性和免疫检查点治疗反应的差异,旨在为更精准的治疗策略提供见解。

#### 2.2.1. 突变分析

使用 MOVICS 中的 compMut 函数比较两组间的突变频率。应用卡方检验识别差异突变基因,频率截断值(freq.cutoff)设定为 freq.cutoff = 0.05,确保仅包含在至少 5%的样本中发生突变的基因。统计显著性由校正后  $P < 0.05$  确定。

#### 2.2.2. 转录组分析

对于转录组分析,使用 na.omit 函数去除了含有缺失值的基因。差异表达分析使用针对芯片数据优化的 limma 方法通过 runDEA 函数进行。使用 runMarker 函数鉴定显著差异表达基因,显著性定义为  $P < 0.05$  和校正后  $P < 0.05$ 。

#### 2.2.3. 表观基因组分析

RRBS 的甲基化数据采用  $\beta$  值方法进行分析。缺失值使用 impute 包中的 impute.knn 函数进行填补。差异甲基化分析结合 ChAMP 包中的 DMP 方法与 limma 进行,显著性阈值设定为  $P < 0.05$  且校正后  $P < 0.05$ 。

#### 2.2.4. 药物敏感性分析

我们使用 `compDrugsen` 函数基于转录组信息预测不同亚型对药物的响应。该分析通过计算基因表达与药物反应指标(IC50 或 EC50)之间的相关性,评估亚型特异性的药物敏感性。参数设置如下: `tissueType` 设定为“breast”,显著性检验采用非参数检验。该分析旨在扩大筛选范围,挖掘亚型特异性的潜在治疗靶点。

### 2.3. 基因集富集分析(GSEA)

为了阐明预后不良组基因表达变化的生物学意义,我们进行了 GSEA 分析。这是一种用于评估预定义的基因集是否在排序基因列表的顶部或底部显著富集的计算方法。该方法能够检测生物学通路和过程的潜在改变。GSEA 通过 R 中的 `MOVICS` 包内的 `GSEA` 函数实现,允许对全局基因表达谱进行综合分析,并检测基因集中细微但协调的变化。

### 2.4. 基因集变异分析(GSVA)和单样本基因集富集分析(ssGSEA)

鉴于通过 GSEA 发现高危乳腺癌样本中 DNA 损伤修复通路显著富集,进一步采用 GSVA 和 ssGSEA 对这些通路进行详细探索。

GSVA 是一种非参数、无监督的方法,可将基因水平的表达数据转换为每个样本的通路水平富集评分,从而能够详细评估各患者组的代谢通路活性。ssGSEA 是 GSEA 的扩展,计算特定基因集在每个样本上的富集评分,提供个体水平通路激活的精细视图。

本分析从 `cBioPortal` 获取了 475 例 HR+/HER2-乳腺癌样本的 FPKM 标准化基因表达矩阵。根据既往研究[21]得到的一份包含 276 个与 DNA 损伤修复通路相关基因的列表进行分析。使用 R 中的 `GSVA` 包,应用 `GSVA` 和 `ssGSEA` 函数计算通路富集评分, `kcdf` 参数设置为“Gaussian”以适应连续表达数据。结果使用 `ggplot2` 和 `ggpubr` 包进行可视化,并使用 Student's t 检验评估组间差异。

### 2.5. 预后独立影响因素的 Cox 比例风险回归分析

为了评估关键分子特征对患者预后的独立预测价值,我们构建了多变量 Cox 比例风险回归模型。为消除治疗异质性的干扰,该分析仅纳入接受了标准内分泌治疗的患者队列。模型校正的协变量包括:年龄、化疗史、组织学分级、阳性淋巴结数量、TP53 突变状态(Mutation vs. Wild-type)以及 MMEJ 通路评分(连续变量)。通过计算风险比(Hazard Ratio, HR)及其 95%置信区间(CI),确定各因素对无复发生存期(RFS)的独立影响。

### 2.6. 使用 TCGA 数据库进行外部验证

为检验结论的普适性,我们引入 TCGA-BRCA 数据作为验证集。通过 UCSC Xena 平台获取临床及生存数据,筛选出 484 例 ER+/HER2-病例。考虑到 PFI (无进展间隔)与 DFS (无病生存期)的高度相关性( $r = 1$ ),我们选取 PFI 作为预后终点。

同源重组缺陷(Homologous Recombination Deficiency, HRD)是指细胞无法通过高保真的同源重组通路有效修复 DNA 双链断裂,从而导致基因组高度不稳定及特定基因组“疤痕”积累的状态。它是反映肿瘤 DNA 修复能力受损程度的关键生物标志物。我们获取的 HRD 评分引自既往文献[22],经匹配后保留 435 例有效数据。为规避生存偏倚,我们剔除了 PFI 与 OS 值相同且无事件记录的“无效随访”病例,最终锁定 43 例具可评估数据的患者。

受限于样本量,我们调整了分组策略,以 PFI 是否超过 2.5 年为界进行二分(PFI\_gt2.5y vs PFI\_le2.5y),并运用 Wilcoxon 秩和检验、Welch's t 检验及 Cohen's d 效应量,多维度评估 HRD 评分的组间差异。

### 3. 结果

#### 3.1. 入组患者的基线特征

高危 HR+/HER2-乳腺癌的常规治疗方案通常涵盖半年的化疗及后续的长期内分泌治疗。若患者在 2 年内复发,即被视为原发性内分泌耐药,此类人群经过目前优化治疗后总生存期(Overall Survival after Recurrence, OS)约为 4 年。鉴于部分患者未接受标准内分泌治疗,本研究剔除了这部分样本,以确保研究群体的同质性。

基于此,我们确立了严苛的高危(CS1 组, 54 例)纳入标准:无复发生存期(RFS) < 2.5 年且复发后总生存期(OS) ≤ 4 年,同时剔除初始肿瘤 > 5 cm 或 III 期患者以排除局部晚期干扰。相对地,低危组(CS2 组, 134 例)则限定为 RFS > 5 年且排除临床 I 期患者。

我们总结了两组样本预后情况,其中,CS1 组 RFS 中位值 20.41 个月显著低于 CS2 组中位值 140.14 个月;CS1 组 OS 中位值 34.87 个月亦显著低于 CS2 组中位值 142 个月,CS1 组预后更差。尽管在年龄、肿瘤大小、分级及手术化疗方式上无显著差异,但阳性淋巴结数量( $P < 0.001$ )的差异具有统计学意义(表 1)。

**Table 1.** Clinical characteristics and survival outcomes of patients with CS1 and CS2 signatures

**表 1.** CS1 和 CS2 特征患者的临床特征及生存结局

特征	CS1 组 (N = 54)	CS2 组 (N = 134)	P 值
<b>年龄</b>			0.529
<40 岁	1	2	
40~44 岁	0	5	
45~49 岁	3	9	
≥50 岁	50	118	
<b>阳性淋巴结数量</b>			<0.001
0	12	0	
1~3	22	104	
≥4	20	30	
<b>肿瘤直径</b>			0.403
≤20 mm	12	22	
>20 mm	42	112	
<b>组织学分级</b>			0.291
I (良好)	2	10	
II (中等)	24	70	
III (差)	27	52	
缺失	1	2	
<b>手术方式</b>			0.860
保乳手术	16	38	
乳房切除术	38	96	
<b>化疗</b>			1.000
无	44	110	
有	10	24	

续表

<b>肿瘤分期</b>			1.000
1 期和 2 期	35	90	
3 期	4	9	
缺失	15	35	
<b>生存结局</b>			
中位 RFS (月)	20.41	140.14	
中位 OS (月)	34.87	142	

### 3.2. 两个预后组之间的多组学差异

为了从分子层面解码预后差异，我们对基因组、表观组及转录组进行了深度挖掘，结果显示：

**基因组分析：**预后不良的 CS1 组中观察到显著更高的 *TP53* 突变频率(表 2)。总体而言，*TP53* 的肿瘤突变负荷(TMB)为 20%。卡方检验得出了极显著的 *P* 值( $2.24 \times 10^{-3}$ )和校正后 *P* 值( $5.38 \times 10^{-2}$ )，证实了 CS1 和 CS2 之间 *TP53* 突变存在实质性差异。

**转录组分析：**在转录组水平上，大量基因表现出显著的差异表达。在 CS1 中，157 个基因显著上调( $P < 0.05$ )(图 1(A))，而 346 个基因显著下调(图 1(B))。CS1 中前 10 个上调和下调的基因汇总于(表 3)。

**表观基因组分析：**表观基因组分析鉴定出 109 个基因在 CS1 的转录起始位点(TSS)区域显著高甲基化(校正后  $P < 0.05$ )。前 10 个高甲基化基因列于(表 4)。值得注意的是，在 TSS 区域未检测到差异低甲基化基因。

**Table 2.** Chi-square test for the independence between molecular subtype and *TP53* mutation

**表 2.** 分子亚型与 *TP53* 突变独立性的卡方检验

突变基因	总体(TMB)	CS1	CS2	<i>P</i> 值	校正后 <i>P</i> 值
<i>TP53</i>	38 (20%)	19 (35.2%)	19 (14.2%)	0.00224	0.0538

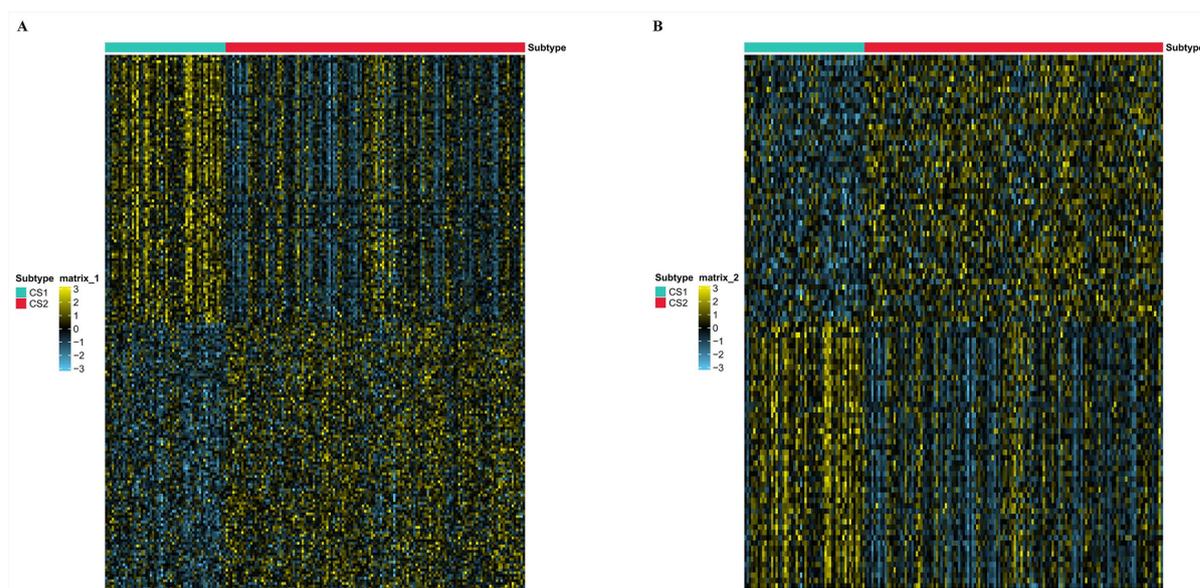
**Table 3.** Top 10 upregulated genes in CS1 identified by transcriptomic differential analysis (left); Top 10 downregulated genes in CS1 identified by transcriptomic differential analysis (right)

**表 3.** 转录组差异分析鉴定的 CS1 中前 10 个上调基因(左)及下调基因(右)

序号	基因	<i>P</i> 值	校正后 <i>P</i> 值	基因	<i>P</i> 值	校正后 <i>P</i> 值
1	KIF20A	$1.89 \times 10^{-10}$	$3.85 \times 10^{-6}$	TRIM4	$5.41 \times 10^{-7}$	$2.74 \times 10^{-4}$
2	BIRC5	$1.59 \times 10^{-9}$	$1.45 \times 10^{-5}$	GASK1B	$6.38 \times 10^{-7}$	$2.89 \times 10^{-4}$
3	GTSE1	$2.14 \times 10^{-9}$	$1.45 \times 10^{-5}$	CHRNB1	$1.39 \times 10^{-6}$	$5.55 \times 10^{-4}$
4	LMNB2	$3.87 \times 10^{-9}$	$1.93 \times 10^{-5}$	ABAT	$2.94 \times 10^{-6}$	$8.94 \times 10^{-4}$
5	SAPCD2	$4.73 \times 10^{-9}$	$1.93 \times 10^{-5}$	TMEM26	$3.11 \times 10^{-6}$	$9.32 \times 10^{-4}$
6	PLK1	$6.37 \times 10^{-9}$	$2.16 \times 10^{-5}$	PNPLA4	$4.11 \times 10^{-6}$	$1.12 \times 10^{-3}$
7	TROAP	$8.74 \times 10^{-9}$	$2.54 \times 10^{-5}$	FAM241A	$4.24 \times 10^{-6}$	$1.14 \times 10^{-3}$
8	PTTG1	$1.00 \times 10^{-8}$	$2.55 \times 10^{-5}$	MAPT	$4.66 \times 10^{-6}$	$1.23 \times 10^{-3}$
9	CDC20	$3.14 \times 10^{-8}$	$6.63 \times 10^{-5}$	PIGV	$5.15 \times 10^{-6}$	$1.32 \times 10^{-3}$
10	BLM	$3.25 \times 10^{-8}$	$6.63 \times 10^{-5}$	SMARCA2	$5.16 \times 10^{-6}$	$1.32 \times 10^{-3}$

**Table 4.** Top 10 hypermethylated TSSs in CS1 identified by epigenomic differential analysis**表 4.** 表观基因组差异分析鉴定的 CS1 中前 10 个高甲基化转录起始位点(TSS)

序号	基因	<i>P</i> 值	校正后 <i>P</i> 值
1	ADCYAP1	$6.14 \times 10^{-8}$	$8.10 \times 10^{-4}$
2	LINC00577	$2.92 \times 10^{-6}$	0.0126
3	SMIM17	$3.61 \times 10^{-6}$	0.0126
4	CYP7B1	$4.28 \times 10^{-6}$	0.0126
5	MEG3	$4.79 \times 10^{-6}$	0.0126
6	MTNR1A	$6.29 \times 10^{-6}$	0.0138
7	ACAN	$7.55 \times 10^{-6}$	0.0142
8	SLC6A3	$1.00 \times 10^{-5}$	0.0162
9	LOC283683	$1.11 \times 10^{-5}$	0.0162
10	PRDM12	$1.29 \times 10^{-5}$	0.017

**Figure 1.** (A) Heatmap of upregulated biomarkers in CS1. (B) Heatmap of downregulated biomarkers in CS1**图 1.** (A) CS1 组中上调生物标志物的热图。(B) CS1 组中下调生物标志物的热图

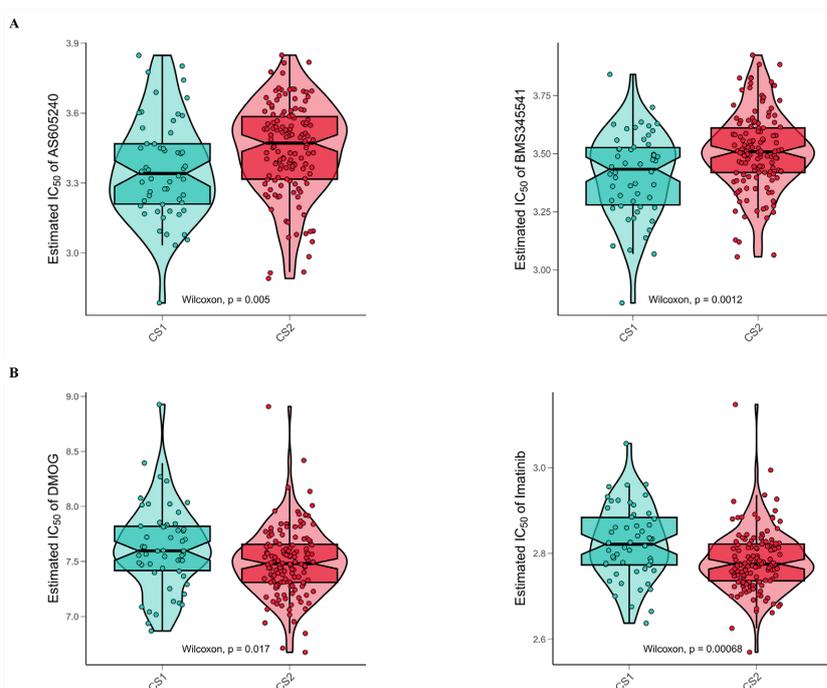
### 3.3. CS1 组药物敏感性降低及耐药性增加

对两个患者亚组间药物敏感性和耐药性的比较分析显示,对多种化合物的反应存在显著差异。CS1 组对 BMS345541 (一种 I $\kappa$ B 激酶抑制剂)和 AS605240 (一种 PI3K 抑制剂)表现出增强的敏感性(图 2(A))。相反, CS1 组对 Imatinib (一种酪氨酸激酶抑制剂)和 DMOG (一种 HIF-PH 抑制剂)的敏感性降低(图 2(B))。这些结果突显了根据分子谱定制治疗策略的机会,并强调了发现生物标志物的必要性,以便对这一高危、预后不良的人群进行更精准的治疗靶向。

### 3.4. GSEA 揭示 DNA 损伤修复通路显著上调

鉴于表观组仅发现高甲基化基因,而转录组差异广泛,我们进一步通过 GSEA 寻找通路层面的线索。结果显示,两组间多个 KEGG 通路中表现出显著差异(图 3)。我们观察到在高危患者群体中,上调通路主要集中于 DNA 损伤修复相关领域(图 3(A)),包括 KEGG\_CELL\_CYCLE (细胞周期通路)、

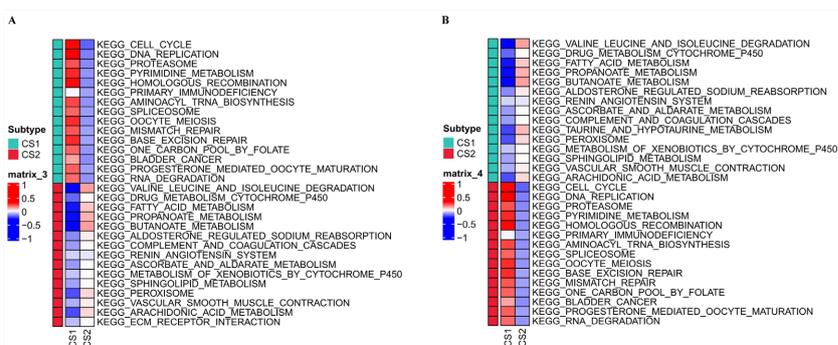
KEGG\_DNA\_REPLICATION (DNA 复制通路)、KEGG\_HOMOLOGOUS\_RECOMBINATION (同源重组修复通路)、KEGG\_MISMATCH\_REPAIR (错配修复通路)、KEGG\_SPLICEOSOME (剪接体通路)和 KEGG\_BASE\_EXCISION\_REPAIR (碱基切除修复通路)等多个环节。这个结果揭示了高危患者生物学行为可能与高增殖、高 DNA 修复活性密切相关。相反,免疫炎症反应及生物代谢相关通路则在高危患者中被抑制(图 3(B)),如:KEGG\_COMPLEMENT\_AND\_COAGULATION\_CASCADES (补体和凝血级联通路)等。这表明高危亚组中可能存在免疫反应抑制和代谢改变。



(A) BMS345541 和 AS605240 在 CS1 (蓝绿色)和 CS2 (红色)亚组中的敏感性。两种药物在 CS1 中的 IC<sub>50</sub>值均低于 CS2,表明 CS1 组敏感性增加,且具有统计学显著性。(B) Imatinib 和 DMOG 在两个亚组中的敏感性。两种药物在 CS1 中的 IC<sub>50</sub>值均高于 CS2,表明 CS2 组敏感性增加,且具有统计学显著性。

Figure 2. Results of the drug sensitivity analysis

图 2. 药物敏感性分析结果



(A) 上调通路和 (B) 下调通路。每一行对应一个 KEGG 通路,每一列代表一个亚型(CS1, CS2)。颜色表示标准化富集评分(NES):红色表示富集程度较高,蓝色表示富集程度较低。细胞周期、DNA 复制和同源重组等通路在不同亚型间表现出独特的富集模式。

Figure 3. Results of GSEA enrichment analysis

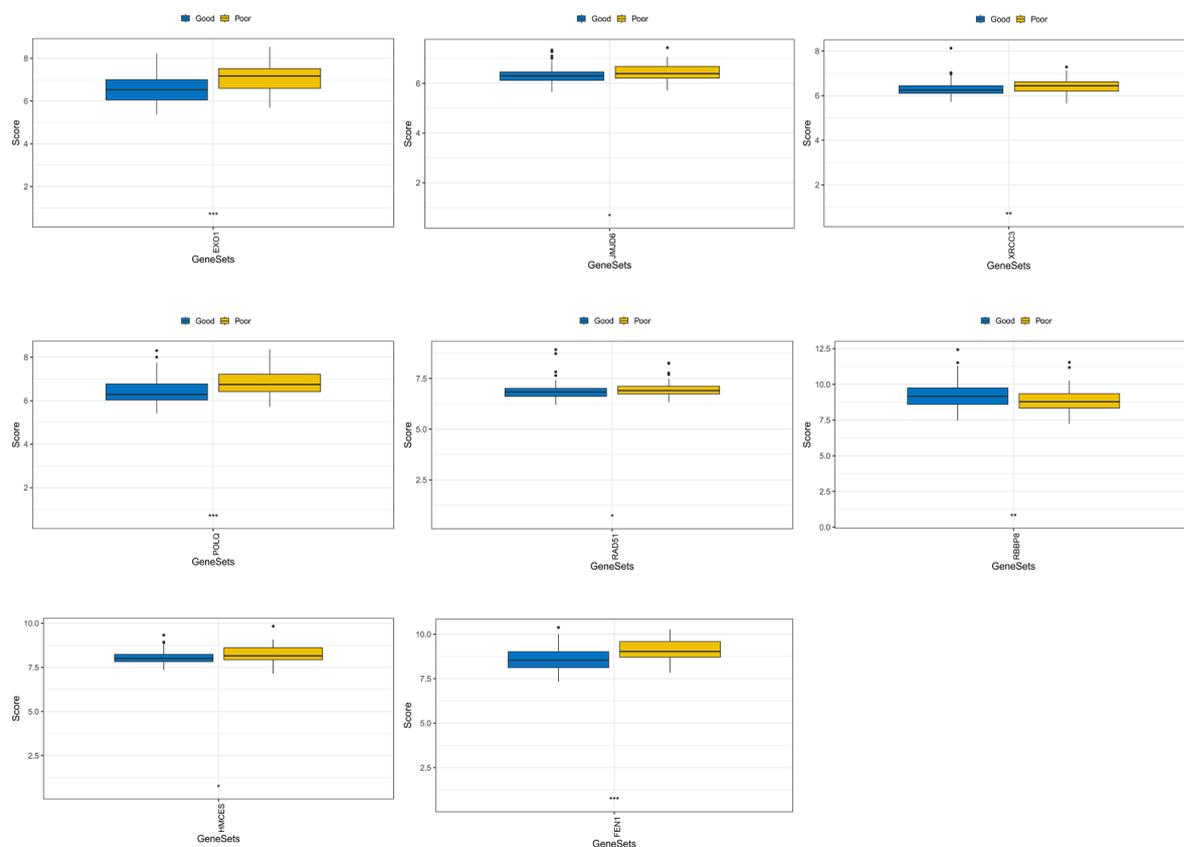
图 3. 基因集富集分析(GSEA)的结果

### 3.5. DNA 损伤修复通路的 GSVA 分析

GSVA 分析进一步细化了我们对 DNA 修复异常的认识。8 个关键基因在两个预后组之间显示出统计学显著差异(图 4)。RBBP8 表达在高危组中显著较低, 而 *EXO1*、*FEN1*、*HMCES*、*JMJD6*、*POLQ*、*RAD51* 和 *XRCC3* 则显著上调(图 4)。尤为关键的是, GSVA(图 5(A))与 ssGSEA(图 5(B))均指向同一个结论: 高致癌性的微同源介导末端连接(MMEJ)通路(一种 DNA 双链断裂(DSB)修复的替代机制)在高危患者中显著富集。然而, 两组之间在同源重组(HR)和非同源末端连接(NHEJ)(经典的 DSB 修复通路)方面未观察到显著差异。

### 3.6. TP53 突变与 MMEJ 评分的独立预后价值验证

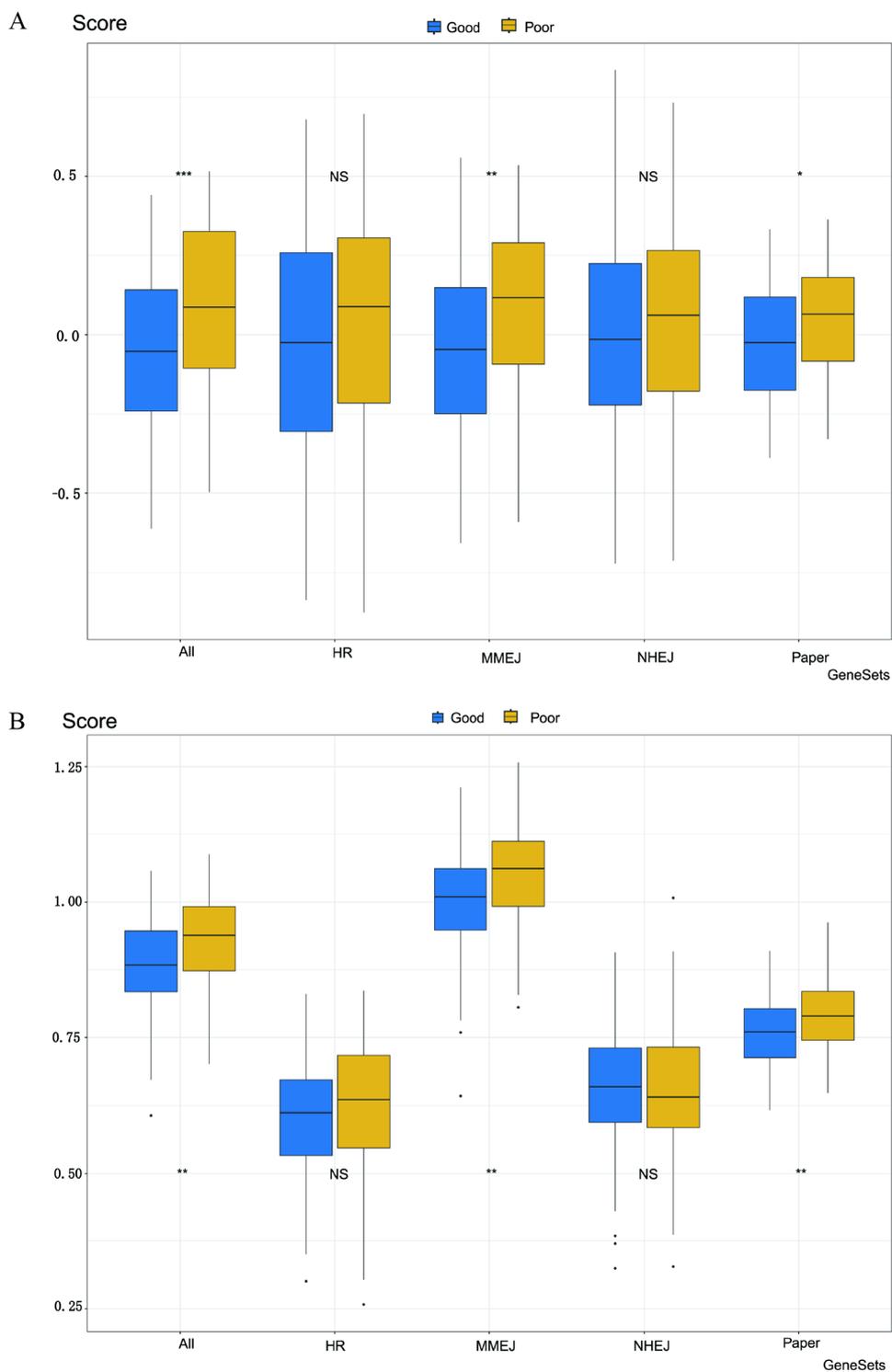
多因素 Cox 回归模型进一步验证了 *TP53* 突变与 MMEJ 通路在预后评估中的独立性(图 6)。在纳入年龄、淋巴结状态、分级及化疗等临床协变量后, *TP53* 突变(HR = 2.12,  $P = 0.013$ )和 MMEJ 评分(HR = 3.78,  $P = 0.013$ )仍显著关联于更差的 RFS。值得注意的是, MMEJ 评分作为连续变量显示出较高的风险比, 提示随着易错修复通路的激活程度增加, 复发风险呈显著上升趋势。该结果巩固了二者作为 HR+/HER2-乳腺癌不良预后核心分子驱动因子的地位。



箱线图描绘了各组评分的分布, 水平线表示中位数, 箱体边界代表四分位距(IQR)。触须延伸至  $1.5 \times IQR$ , 异常值显示为单独的点。X 轴标签代表关键的 DNA 损伤修复因子, Y 轴表示基因表达值。星号表示两组之间的统计学显著差异: \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; NS, 不显著。

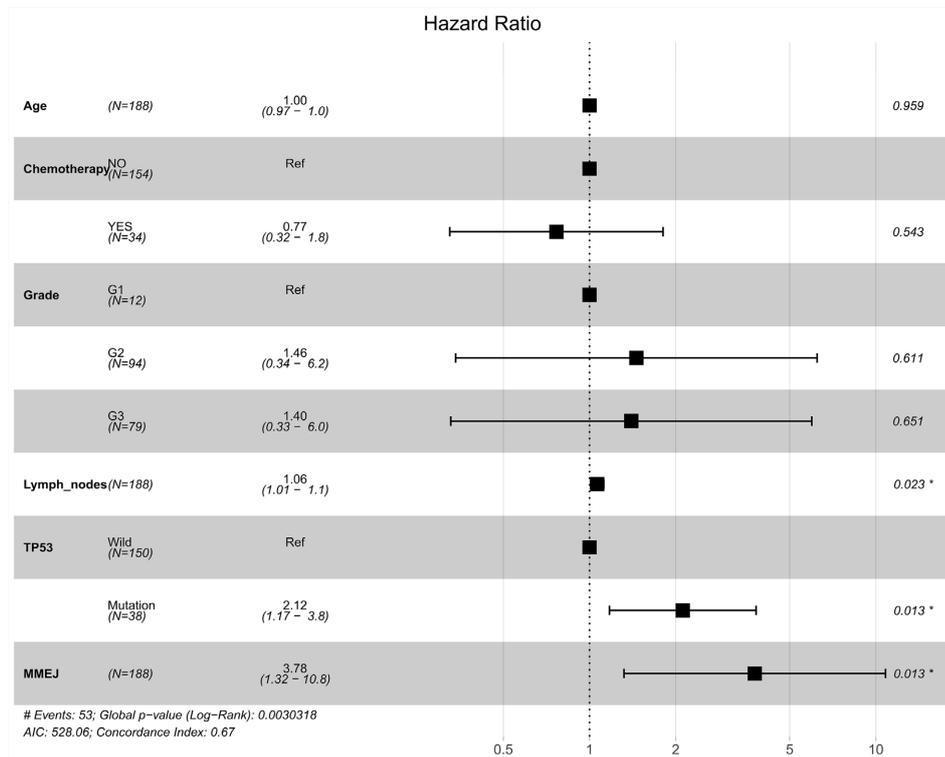
**Figure 4.** Comparison of gene set scores for DNA damage repair pathway factors between the CS1 (yellow) and CS2 (blue) groups

**图 4.** CS1 (黄色)和 CS2 (蓝色)组之间 DNA 损伤修复通路因子基因集评分的比较



箱线图展示了各组评分的分布，X轴标签表示不同的基因集。在GSVA分析(A)中，Y轴(评分)代表基因表达值，而在ssGSEA分析(B)中，Y轴(评分)表示基因集富集评分。在所有基因及MMEJ基因集中，GSVA(A)和ssGSEA(B)分析均显示两组之间存在显著差异。其中，“Paper”指代基于文献报道的276个DNA损伤修复相关基因组成的综合基因集，用于评估样本的总体DNA修复通路活性

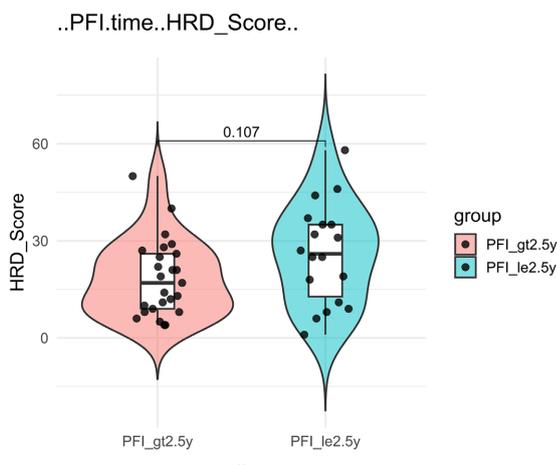
**Figure 5.** Comparison of gene set scores between the CS1 (yellow) and CS2 (blue) groups  
**图 5.** CS1 (黄色)和CS2 (蓝色)组之间基因集评分的比较



该模型纳入了年龄、化疗史、组织学分级、阳性淋巴结数量、TP53 突变状态及 MMEJ 通路评分等变量。图中展示了各变量的风险比(Hazard Ratio, HR)及其 95%置信区间(CI)。MMEJ 评分和淋巴结数量作为连续变量分析,其余为分类变量。结果显示,在调整其他临床因素后,TP53 突变(HR = 2.12, P = 0.013)和高 MMEJ 评分(HR = 3.78, P = 0.013)均为显著的独立预后危险因素

**Figure 6.** Forest plot of multivariate Cox proportional hazards regression analysis for Relapse-Free Survival in HR+/HER2- patients receiving endocrine therapy

**图 6.** 接受内分泌治疗的 HR+/HER2-乳腺癌患者无复发生存期的多因素 Cox 回归分析森林图



本图展示了根据无进展间隔(PFI)分层的两组患者的 HRD 评分分布情况: PFI\_gt2.5y 组(>2.5 年, 粉色)和 PFI\_le2.5y 组(≤2.5 年, 蓝色)。小提琴图描绘了各亚组中 HRD 评分的密度分布,其中嵌入的箱线图指示了中位数和四分位距。图中的黑点代表单个样本。组间的统计比较结果以 P 值(P = 0.107)显示,用于评估不同 PFI 类别之间 HRD 评分差异的显著性

**Figure 7.** Distribution of HRD scores between PFI-based subgroups in the TCGA-BRCA cohort

**图 7.** TCGA-BRCA 队列中基于 PFI 分组的 HRD 评分分布

### 3.7. 使用 TCGA 数据库进行外部验证

在 TCGA-BRCA 验证集中, 共有 43 例患者被纳入研究, 其中 25 例被分配到 PFI\_gt2.5y 组(PFI 大于 2.5 年), 18 例被分配到 PFI\_le2.5y 组(PFI 小于 2.5 年)。

同源重组缺陷(HRD)评分的比较表明, 短 PFI 组有 HRD 水平升高的趋势。尽管受限于样本量, Wilcoxon 检验( $P = 0.1066$ )与 Welch's t 检验( $P = 0.094$ )未达传统显著性水平, 但 Cohen's d 值为 $-0.56$ , 提示中等效应量。这一结果与我们在 METABRIC 数据库分析得到的结果一致, 佐证了 HRD 水平升高与不良预后之间的潜在联系(图 7)。

## 4. 讨论

乳腺癌是全球女性中最常见的恶性肿瘤[1], 在癌症相关死亡率中排名第五。激素受体阳性(HR+)/HER2-乳腺癌约占所有病例的 70%, 通常预后良好。然而, 由于其显著的异质性, 约 15%的 HR+/HER2-患者尽管接受了标准治疗, 仍会在 5 年内复发。与其他乳腺癌亚型不同, HR+/HER2-疾病的复发风险无限期持续, 可延长至确诊后 20~30 年[23]。

临床上, 某些 HR+/HER2-患者尽管肿瘤较小, 但在诊断时即有淋巴结受累或远处转移, 表现出侵袭性疾病特征。另一些患者经历早期复发——无论是在手术后不久、辅助化疗期间或一年内, 还是辅助内分泌治疗两年内——表明对化疗和内分泌治疗均存在原发性耐药[24]。这种内在的异质性显著复杂化了治疗策略的制定。尽管进行了广泛研究, 目前仍缺乏用于精确分型和预后的确切分子标志物。因此, 阐明驱动该亚型异质性的分子机制对于改善患者预后至关重要, 这也是本研究的核心临床依据。

基于这一临床痛点, 本研究利用多组学手段对 HR+/HER2-乳腺癌进行了深度剖析。我们的核心目标有二: 一方面是通过研究各组学层面的差异, 揭示 HR+/HER2-高危乳腺癌内在的遗传及生物学特征, 推动对其的进一步理解; 另一方面是发现高危 HR+/HER2-乳腺癌亚型潜在的治疗靶点, 以期转化为新的治疗策略或药物, 实现对该类型人群早期、精准和强化治疗, 最终改善患者的生活质量和生存率。

我们首先定义了一种极端的高危患者人群, RFS < 2.5 年并 OS < 4 年, 同时, 为了排除非生物学因素(如治疗延迟)造成的不良预后, 排除了初诊局部晚期的患者。而低危患者人群, 我们不仅定义其 RFS > 5 年, 并且同时排除因初诊 I 期而掩盖的不良预后。通过设定严苛的筛选标准, 我们构建了极具代表性的高危与低危对照组, 从而捕捉到了以往研究中可能被稀释的分子差异。基于 METABRIC 数据库的多组学分析, 我们揭示了基因组、转录组和表观基因组水平的显著差异。基因组分析首先指出 *TP53* 突变在高危组中的富集。*TP53* 突变是人类癌症中最普遍的遗传改变之一[25]。作为“基因组卫士”, *TP53* 的功能缺失往往意味着细胞对 DNA 损伤监控能力的丧失。在 HR+/HER2-乳腺癌中, *TP53* 突变常与同源重组修复(HRR)相关基因突变(如 *BRCA1/2*)共发生, 可能加剧基因组不稳定性并加速癌症进展[26]。

我们也观察到了表观基因组差异, 特别是涉及先前与乳腺癌进展相关的基因。例如, *CYP7B1* 基因表达上调, 该基因编码的酶负责催化胆固醇氧化产物如 27-羟基胆固醇(27-HC)7 $\alpha$ -羟基化。在 HR+乳腺癌中, 27-HC 作为选择性雌激素受体调节剂, 促进肿瘤生长[27]。然而, *CYP7B1* 通常在该乳腺癌亚型中下调, 导致 27-HC 在肿瘤微环境中的积累[28]。

在转录组水平上, 观察到了最显著的差异, 表明基因表达改变可能驱动高危 HR+/HER2-乳腺癌的不良预后。诸如 *UBE2C* 和 *CBX2* 等已知增强肿瘤细胞增殖的基因[29] [30]在高危组中显著上调。相反, 增强化疗敏感性并抑制肿瘤进展的 *PIP* 基因[31]则下调。

然而, 转录组中最令人担忧的差异是 DNA 损伤修复通路的显著差异。DNA 修复缺陷是肿瘤发生和发展的关键驱动因素[32]。同源重组(HR)DNA 修复通路中的致病突变, 如 *BRCA1/2* 突变, 是乳腺癌中最常见的遗传改变之一[33]。这些突变损害了细胞修复双链 DNA 断裂的能力, 从而增加了基因组不稳定性。

通路富集分析进一步揭示, 高危肿瘤表现出 DNA 损伤修复通路的显著上调, 特别是微同源介导的末端连接(MMEJ)通路——这是一种极易出错的机制, 与基因组不稳定性及耐药性相关[34]。相比之下, 经典的同源重组(HR)和非同源末端连接(NHEJ)通路未显示显著差异。7 个关键的 DNA 修复基因——*EXO1*、*FEN1*、*HMCES*、*JMJD6*、*POLQ*、*RAD51* 和 *XRCC3*——在 高危组中显著上调。随后的 GSEA 和 ssGSEA 分析证实了高危肿瘤中 MMEJ 通路的过度激活。鉴于其与复杂基因组重排和治疗耐药的强相关性, MMEJ 过度激活可能促成了这些肿瘤的侵袭性表型。

与此同时, 我们使用预后相关因素的 Cox 比例风险模型进一步评估 *TP53* 突变状态和 MMEJ 通路评分对患者预后的独立影响。得到了两者均为独立不良预后因素的结论。

为了验证这一发现, 我们在 TCGA 队列中进行了外部验证。结果显示, 尽管可评估病例数量有限( $n=43$ )降低了统计效能, 但 PFI 较短的患者表现出较高的 HRD 评分, 且具有中等效应量, 表明 HRD 升高与更短的无进展生存期之间存在潜在关联。这一趋势与我们主队列的结果一致, 并加强了 HRD 作为 ER+/HER2-乳腺癌潜在预后生物标志物的地位。

然而, 本研究仍存在若干局限性。主要局限包括: 依赖单一公共数据集且外部验证有限; 风险组定义具有主观性; 高危亚组样本量相对较小; 芯片和 RRBS 数据的固有局限性。此外, 缺乏实验验证也限制了对 MMEJ 激活的功能解释。

## 5. 结论

本研究利用 METABRIC 数据库, 首次证明了在预后极差的 HR+/HER2-乳腺癌中存在 MMEJ 介导的 DNA 修复过度激活。这些发现为该亚组肿瘤侵袭性和耐药性背后的分子机制提供了有价值的见解。然而, 样本量相对较小和外部验证有限等局限性仍需进一步调查。未来的研究应侧重于在更大的队列中验证这些结果, 并探索靶向 MMEJ 及相关通路在 高危 HR+/HER2-乳腺癌中的治疗潜力。

## 致 谢

感谢公共数据库 METABRIC 数据库和 TCGA 数据库提供的数据支持, 感谢重庆市自然科学基金创新发展联合基金对本研究的资助, 感谢所有对本文提供帮助的机构和个人。

## 声 明

本研究获得重庆医科大学附属第一医院临床科研伦理委员会批准(审批号: 2024 年科研伦理(2024-044-01)号), 本研究数据来源为公共数据库。

## 基金项目

重庆市自然科学基金创新发展联合基金(万州区)资助(CSTB2023NSCQ-LMX0018)。

## 参考文献

- [1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **71**, 209-249. <https://doi.org/10.3322/caac.21660>
- [2] Goldhirsch, A., Winer, E.P., Coates, A.S., Gelber, R.D., Piccart-Gebhart, M., Thürlimann, B., *et al.* (2013) Personalizing the Treatment of Women with Early Breast Cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology*, **24**, 2206-2223. <https://doi.org/10.1093/annonc/mdt303>
- [3] Huppert, L.A., Gumusay, O., Idossa, D. and Rugo, H.S. (2023) Systemic Therapy for Hormone Receptor-positive/human Epidermal Growth Factor Receptor 2-negative Early Stage and Metastatic Breast Cancer. *CA: A Cancer Journal for*

- Clinicians*, **73**, 480-515. <https://doi.org/10.3322/caac.21777>
- [4] Prat, A., Pineda, E., Adamo, B., Galván, P., Fernández, A., Gaba, L., *et al.* (2015) Clinical Implications of the Intrinsic Molecular Subtypes of Breast Cancer. *The Breast*, **24**, S26-S35. <https://doi.org/10.1016/j.breast.2015.07.008>
- [5] Pagani, O., Walley, B.A., Fleming, G.F., Colleoni, M., Láng, I., Gomez, H.L., *et al.* (2023) Adjuvant Exemestane with Ovarian Suppression in Premenopausal Breast Cancer: Long-Term Follow-Up of the Combined TEXT and SOFT Trials. *Journal of Clinical Oncology*, **41**, 1376-1382. <https://doi.org/10.1200/jco.22.01064>
- [6] Zhu, X., Zhang, H., Zhou, Y., Chen, Y., Fu, T., Jin, M., *et al.* (2025) Subtyping-Directed Precision Treatment Refines Traditional One-Size-Fits-All Therapy for HR+/HER2-Breast Cancer. *Cancer Research*, **85**, 3983-3998. <https://doi.org/10.1158/0008-5472.can-24-5002>
- [7] Ma, C.X., Reinert, T., Chmielewska, I. and Ellis, M.J. (2015) Mechanisms of Aromatase Inhibitor Resistance. *Nature Reviews Cancer*, **15**, 261-275. <https://doi.org/10.1038/nrc3920>
- [8] Dowsett, M., Cuzick, J., Ingle, J., Coates, A., Forbes, J., Bliss, J., *et al.* (2010) Meta-Analysis of Breast Cancer Outcomes in Adjuvant Trials of Aromatase Inhibitors versus Tamoxifen. *Journal of Clinical Oncology*, **28**, 509-518. <https://doi.org/10.1200/jco.2009.23.1274>
- [9] Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2011) Relevance of Breast Cancer Hormone Receptors and Other Factors to the Efficacy of Adjuvant Tamoxifen: Patient-Level Meta-Analysis of Randomised Trials. *The Lancet*, **378**, 771-784. [https://doi.org/10.1016/s0140-6736\(11\)60993-8](https://doi.org/10.1016/s0140-6736(11)60993-8)
- [10] Francis, P.A., Pagani, O., Fleming, G.F., Walley, B.A., Colleoni, M., Láng, I., *et al.* (2018) Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer. *New England Journal of Medicine*, **379**, 122-137. <https://doi.org/10.1056/nejmoa1803164>
- [11] Filipits, M., Nielsen, T.O., Rudas, M., Greil, R., Stöger, H., Jakesz, R., *et al.* (2014) The PAM50 Risk-of-Recurrence Score Predicts Risk for Late Distant Recurrence after Endocrine Therapy in Postmenopausal Women with Endocrine-Responsive Early Breast Cancer. *Clinical Cancer Research*, **20**, 1298-1305. <https://doi.org/10.1158/1078-0432.ccr-13-1845>
- [12] (2012) Comprehensive Molecular Portraits of Human Breast Tumours. *Nature*, **490**, 61-70. <https://doi.org/10.1038/nature11412>
- [13] Puppe, J., Seifert, T., Eichler, C., Pilch, H., Mallmann, P. and Malter, W. (2020) Genomic Signatures in Luminal Breast Cancer. *Breast Care*, **15**, 355-365. <https://doi.org/10.1159/000509846>
- [14] Zhang, S., Fitzsimmons, K.C. and Hurvitz, S.A. (2022) Oncotype DX Recurrence Score in Premenopausal Women. *Therapeutic Advances in Medical Oncology*, **14**, 17588359221081077. <https://doi.org/10.1177/17588359221081077>
- [15] Sparano, J.A., Gray, R.J., Makower, D.F., Albain, K.S., Saphner, T.J., Badve, S.S., *et al.* (2020) Clinical Outcomes in Early Breast Cancer with a High 21-Gene Recurrence Score of 26 to 100 Assigned to Adjuvant Chemotherapy plus Endocrine Therapy: A Secondary Analysis of the TAILORx Randomized Clinical Trial. *JAMA Oncology*, **6**, 367-374. <https://doi.org/10.1001/jamaoncol.2019.4794>
- [16] Cardoso, F., van't Veer, L.J., Bogaerts, J., Slaets, L., Viale, G., Delaloge, S., *et al.* (2016) 70-gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *New England Journal of Medicine*, **375**, 717-729. <https://doi.org/10.1056/nejmoa1602253>
- [17] van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., *et al.* (2002) Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. *Nature*, **415**, 530-536. <https://doi.org/10.1038/415530a>
- [18] Piccart, M., van 't Veer, L.J., Poncet, C., Lopes Cardoso, J.M.N., Delaloge, S., Pierga, J., *et al.* (2021) 70-Gene Signature as an Aid for Treatment Decisions in Early Breast Cancer: Updated Results of the Phase 3 Randomised MINDACT Trial with an Exploratory Analysis by Age. *The Lancet Oncology*, **22**, 476-488. [https://doi.org/10.1016/s1470-2045\(21\)00007-3](https://doi.org/10.1016/s1470-2045(21)00007-3)
- [19] Jin, X., Zhou, Y., Ma, D., Zhao, S., Lin, C., Xiao, Y., *et al.* (2023) Molecular Classification of Hormone Receptor-Positive Her2-Negative Breast Cancer. *Nature Genetics*, **55**, 1696-1708. <https://doi.org/10.1038/s41588-023-01507-7>
- [20] Garutti, M., Griguolo, G., Botticelli, A., Buzzatti, G., De Angelis, C., Gerratana, L., *et al.* (2022) Definition of High-Risk Early Hormone-Positive Her2-Negative Breast Cancer: A Consensus Review. *Cancers*, **14**, Article No. 1898. <https://doi.org/10.3390/cancers14081898>
- [21] Knijnenburg, T.A., *et al.* (2018) Genomic and Molecular Landscape of DNA Damage Repair Deficiency across the Cancer Genome Atlas. *Cell Reports*, **23**, 239-254.e6.
- [22] Ballot, E., Galland, L., Mananet, H., Boidot, R., Arnould, L., Desmoulins, I., *et al.* (2022) Molecular Intrinsic Subtypes, Genomic, and Immune Landscapes of BRCA-Proficient but HRD-High ER-Positive/HER2-Negative Early Breast Cancers. *Breast Cancer Research*, **24**, Article No. 80. <https://doi.org/10.1186/s13058-022-01572-6>
- [23] Pan, H., Gray, R., Braybrooke, J., Davies, C., Taylor, C., McGale, P., *et al.* (2017) 20-Year Risks of Breast-Cancer

- Recurrence after Stopping Endocrine Therapy at 5 Years. *New England Journal of Medicine*, **377**, 1836-1846. <https://doi.org/10.1056/nejmoa1701830>
- [24] Johnston, S.R.D., Toi, M., O'Shaughnessy, J., Rastogi, P., Campone, M., Neven, P., *et al.* (2023) Abemaciclib plus Endocrine Therapy for Hormone Receptor-Positive, Her2-Negative, Node-Positive, High-Risk Early Breast Cancer (monarchE): Results from a Preplanned Interim Analysis of a Randomised, Open-Label, Phase 3 Trial. *The Lancet Oncology*, **24**, 77-90. [https://doi.org/10.1016/s1470-2045\(22\)00694-5](https://doi.org/10.1016/s1470-2045(22)00694-5)
- [25] Olivier, M., Hollstein, M. and Hainaut, P. (2009) TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspectives in Biology*, **2**, a001008. <https://doi.org/10.1101/cshperspect.a001008>
- [26] Kim, J., Jeong, K., Jun, H., Kim, K., Bae, J.M., Song, M.G., *et al.* (2023) Mutations of TP53 and Genes Related to Homologous Recombination Repair in Breast Cancer with Germline BRCA1/2 Mutations. *Human Genomics*, **17**, Article No. 2. <https://doi.org/10.1186/s40246-022-00447-3>
- [27] Avena, P., Casaburi, I., Zavaglia, L., Nocito, M.C., La Padula, D., Rago, V., *et al.* (2022) 27-Hydroxycholesterol Binds GPER and Induces Progression of Estrogen Receptor-Negative Breast Cancer. *Cancers*, **14**, Article No. 1521. <https://doi.org/10.3390/cancers14061521>
- [28] Shi, S.Z., *et al.* (2019) Recruitment of Monocytes and Epigenetic Silencing of Intratumoral CYP7B1 Primarily Contribute to the Accumulation of 27-Hydroxycholesterol in Breast Cancer. *American Journal of Cancer Research*, **9**, 2194-2208.
- [29] Lu, Z.N., Song, J., Sun, T. and Sun, G. (2021) UBE2C Affects Breast Cancer Proliferation through the AKT/mTOR Signaling Pathway. *Chinese Medical Journal*, **134**, 2465-2474. <https://doi.org/10.1097/cm9.0000000000001708>
- [30] Zheng, S., *et al.* (2019) Overexpression of CBX2 in Breast Cancer Promotes Tumor Progression through the PI3K/AKT Signaling Pathway. *American Journal of Translational Research*, **11**, 1668-1682.
- [31] Urbaniak, A., Jablonska, K., Suchanski, J., Partynska, A., Szymczak-Kulus, K., Matkowski, R., *et al.* (2023) Prolactin-Induced Protein (PIP) Increases the Sensitivity of Breast Cancer Cells to Drug-Induced Apoptosis. *Scientific Reports*, **13**, Article No. 6574. <https://doi.org/10.1038/s41598-023-33707-w>
- [32] Willis, R. (2016) Targeted Cancer Therapy: Vital Oncogenes and a New Molecular Genetic Paradigm for Cancer Initiation Progression and Treatment. *International Journal of Molecular Sciences*, **17**, Article No. 1552. <https://doi.org/10.3390/ijms17091552>
- [33] Heeke, A.L., Xiu, J., Elliott, A., Korn, W.M., Lynce, F., Pohlmann, P.R., *et al.* (2020) Actionable Co-Alterations in Breast Tumors with Pathogenic Mutations in the Homologous Recombination DNA Damage Repair Pathway. *Breast Cancer Research and Treatment*, **184**, 265-275. <https://doi.org/10.1007/s10549-020-05849-2>
- [34] Patterson-Fortin, J. and D'Andrea, A.D. (2020) Exploiting the Microhomology-Mediated End-Joining Pathway in Cancer Therapy. *Cancer Research*, **80**, 4593-4600. <https://doi.org/10.1158/0008-5472.can-20-1672>