

本文引文格式:卢甜甜,房钰奇,邱博渊,等.利用单细胞RNA测序与拷贝数变异分析挖掘雌激素受体阳性乳腺癌标志物[J].右江民族医学院学报,2026,48(2):164-173.

【论著与临床报道】

## 利用单细胞 RNA 测序与拷贝数变异分析 挖掘雌激素受体阳性乳腺癌标志物

卢甜甜<sup>1</sup>,房钰奇<sup>1</sup>,邱博渊<sup>2</sup>,蒋丹璐<sup>1</sup>,申琳<sup>2</sup>,赵劲民<sup>1,2</sup>,朱博<sup>1,2</sup>

(1. 广西医科大学再生医学与医用生物资源开发应用省部共建协同创新中心,广西南宁 530021;  
2. 广西医科大学生命科学研究院,广西南宁 530021)

**摘要:**目的 本研究旨在整合单细胞 RNA 测序(single-cell RNA-seq, scRNA-seq)与拷贝数变异(copy number variation, CNV)分析,系统筛选雌激素受体阳性(estrogen receptor-positive, ER+)乳腺癌的潜在分子标志物。方法 利用 ER+乳腺癌 scRNA-seq 数据结合 CNV 评分识别癌细胞与正常上皮细胞;通过差异表达分析、富集分析、共识聚类及蛋白质相互作用网络分析筛选候选基因;基于 scRNA-seq 数据评估各基因对 ER+乳腺癌细胞的标记效能以及实验验证,最终确定最具潜力的标志物。结果 在 scRNA-seq 数据分析中,成功区分癌细胞与正常上皮细胞,并筛选出若干 ER+乳腺癌潜在标志物。验证结果表明本筛选流程准确性良好,所获标志物具有较高特异性;与仅依赖 RNA-seq 数据的分析相比,scRNA-seq 结合 CNV 分析能识别出更多潜在的癌细胞标志物。最终,基于该整合分析筛选出的钾电压门控通道亚家族 E 调节亚基 4(potassium voltage-gated channel subfamily E regulatory subunit 4, KCNE4)被确定为 ER+乳腺癌的潜在新型标志物与治疗靶点;进一步通过 qRT-PCR 和 Western Blot 实验证实, KCNE4 在 ER+乳腺癌中表达显著上调。结论 本研究筛选出的 KCNE4 是 ER+乳腺癌中极具潜力的分子标志物,为针对癌症的研究提供了新思路。

**关键词:**单细胞基因表达分析;DNA 拷贝数变异;雌激素受体阳性乳腺肿瘤;癌症靶点基因;钾电压门控通道亚家族 E 调节亚基 4

中图分类号:R737.9 文献标识码:A 文章编号:1001-5817(2026)02-0164-10  
doi:10.3969/j.issn.1001-5817.2026.02.003

### Identification of biomarkers for estrogen receptor-positive breast cancer using single-cell RNA sequencing and copy number variation analysis

LU Tiantian<sup>1</sup>, FANG Yuqi<sup>1</sup>, QIU Boyuan<sup>2</sup>, JIANG Danlu<sup>1</sup>,  
SHEN Lin<sup>2</sup>, ZHAO Jinmin<sup>1,2</sup>, ZHU Bo<sup>1,2</sup>

(1. Collaborative Innovation Centre of Regenerative Medicine and Medical Bioresource Development and Application Co-constructed by the Province and Ministry, Guangxi Medical University, Nanning 530021, Guangxi, China; 2. Life Sciences Institute, Guangxi Medical University, Nanning, Guangxi 530021)

**Abstract:** **Objective** To comprehensively screen potential molecular biomarkers for estrogen receptor-positive (ER+) breast cancer by integrating single-cell RNA sequencing (scRNA-seq) with copy number variation (CNV) analysis. **Methods** scRNA-seq data of ER+ breast cancer were utilized to distinguish cancer cells from normal epithelial cells in combination with CNV scoring. Differential expression analysis, enrichment analysis, consensus clustering, and protein-protein interaction network analysis were employed to screen

基金项目:国家自然科学基金项目(82560600);广西自然科学基金项目(2023GXNSFAA026002)

第一作者:卢甜甜,在读硕士研究生,研究方向:再生医学与干细胞及其肿瘤靶向分子机制研究, E-mail: t1095734174@163.com

通讯作者:朱博,博士,副研究员,硕士研究生导师,研究方向:再生医学与干细胞及其肿瘤靶向分子机制研究, E-mail: gxzyx-yzhuo@163.com

candidate genes. The marker efficacy of each gene for ER+ breast cancer cells was evaluated based on scRNA-seq data, together with experimental validation, to ultimately identify the most promising biomarkers. **Results** In the scRNA-seq data analysis, cancer cells were successfully distinguished from normal epithelial cells, and several potential biomarkers for ER+ breast cancer were identified. The validation results demonstrated good accuracy of the screening process, with the identified biomarkers showing high specificity. Compared with analysis relying solely on RNA-seq data, scRNA-seq combined with CNV analysis identified more potential cancer cell biomarkers. Ultimately, potassium voltage-gated channel subfamily E regulatory subunit 4 (KCNE4) was identified as a potential novel biomarker and therapeutic target for ER+ breast cancer through this integrated analysis. Further qRT-PCR and Western Blot experiments confirmed that KCNE4 expression was significantly upregulated in ER+ breast cancer. **Conclusion** KCNE4, screened in this study, is a highly promising molecular biomarker for ER+ breast cancer and provides new insights for cancer-targeted research.

**Key words:** single-cell gene expression analysis; DNA copy number variation; estrogen receptor-positive breast neoplasms; cancer target genes; potassium voltage-gated channel subfamily E regulatory subunit 4

乳腺癌是全球女性中发病率最高的恶性肿瘤之一,在女性各类癌症中占比约 30%;其中,雌激素受体阳性(estrogen receptor-positive, ER+)亚型最为常见<sup>[1]</sup>。随着癌基因研究的深入与测序技术的日益成熟,ER+乳腺癌的相关标志物与治疗靶点(如 CDK4/6)已被陆续识别<sup>[2-3]</sup>。目前,基于差异分析的方法在挖掘该类癌症标志物方面发挥了重要作用,助力乳腺癌研究取得了一系列进展<sup>[4-6]</sup>。尽管众多研究成果及相应靶向药物的应用已显著降低患者的死亡率,但耐药问题也逐渐显现,尤其在癌症转移阶段更为突出<sup>[7]</sup>。因此,利用更先进的技术深入挖掘潜在的 ER+乳腺癌标志物,显得尤为迫切。

单细胞 RNA 测序(scRNA-seq)是一种强大的技术,能够精准识别样本组织中多种细胞类型<sup>[8]</sup>,已在多项乳腺癌研究中得到应用<sup>[9]</sup>。同时,关于细胞拷贝数变异(copy number variation, CNV)的研究也在逐步深入<sup>[10]</sup>。目前也有多种 CNV 评分工具被开发用于识别 RNA-seq 数据中发生染色体非整倍体变异的细胞,并已证实此类细胞通常具有恶性特征<sup>[11-12]</sup>。其中,拷贝数核型分析(copy number karyotyping, CopyKAT)是一种专门针对 scRNA-seq 数据设计的 CNV 分析工具<sup>[13]</sup>,并且对高通量 scRNA-seq 方法更具兼容性<sup>[14]</sup>。结合 CopyKAT 的分析,scRNA-seq 差异基因分析有望更准确地挖掘潜在的癌症标志物<sup>[15-16]</sup>。然而,当前标志物的挖掘仍主要依靠常规 RNA-seq 的差异分析方法,基于 scRNA-seq 工具的系统分析思路相对较少。

因此,本研究旨在通过 scRNA-seq 数据并结合 CNV 分析等方法来挖掘新的 ER+乳腺癌标志物,为未来开发更多替代性治疗靶点提供理论依据。

## 1 材料与方法

1.1 单细胞 RNA 测序原始数据 从 GEO 数据库(<https://www.ncbi.nlm.nih.gov/geo/>)获得 ER+

乳腺癌样本原始 scRNA-seq 数据(GSE161529),并用于进一步的生物信息学分析。

1.2 数据质量控制 使用 R 包 Seurat(版本 4.3.0)处理单细胞分子标签计数矩阵<sup>[17]</sup>。通过 Seurat 中的 NormalizeData 函数进行数据标准化。全局缩放标准化方法“lognormalize”将每个细胞的基因表达测量值通过总表达量,乘以一个缩放因子(缺省为 10000),最后将结果对数转换后,将其标准化。通过过滤表达基因数少于 200 个或线粒体基因占比超过总表达基因 5%的细胞后,总计获得 8 576 个单细胞。

1.3 UMAP 可视化 使用 ScaleData 函数选择所有基因作为主成分分析的特征<sup>[18]</sup>。随后,对每个样本提取前 20 个主成分,并基于 Seurat 工具包中的 FindNeighbors 与 FindClusters 函数执行细胞聚类,进一步通过 RunUMAP 实现结果的可视化。二次降维中通过 R 包 Harmony(版本 0.1.1)去除批次效应。

1.4 注释细胞类型 使用 R 包 SingleR<sup>[19]</sup>和参考转录组数据集“Human Primary Cell Atlas”<sup>[20]</sup>鉴定每个单细胞的起源并识别细胞类型。使用 Seurat 包的 FindMarkers 函数识别差异表达基因(differentially expressed genes, DEGs)。同时在 CellMarker(<http://xteam.xbio.top/CellMarker>)比对 DEGs 以识别细胞类型。

1.5 上皮细胞拟时序分析 使用 R 包 monocle(版本 2.24.0)进行拟时序分析。

1.6 CopyKAT-CNV 分析 使用 R 包 CopyKAT(版本 1.1.0)对通过 UMAP 降维的上皮细胞样本进行 CNV 分析。

1.7 差异分析 结合 dplyr 和 tidyverse 包的数据处理功能,使用 R 包 limma(版本 3.54.2)提供的差异分析方法。设置了  $P < 0.05$  和  $|\log_2\text{foldchange}| > 0.5$  为过滤标准。

1.8 临床数据和共识聚类 通过 GEO 数据库和癌

症基因组图谱(the cancer genome atlas, TCGA)数据库分别获得 ER+ 乳腺癌样本 RNA-seq 数据集(GSE159956)和 1 085 份带有临床信息的女性乳腺癌 RNA-seq 数据,两个数据集使用 ConsensusCluster-Plus 包(版本 1.62.0)进行共识聚类。

1.9 肿瘤分型与免疫分析 通过 R 包 survminer(版本 0.4.9)和 R 包 survival(版本 3.5-5)对两个数据集的分型结果进行了分析。使用 R 包 limma 和 R 包 estimate(版本 1.0.13)分析了 TCGA 样本的肿瘤微环境评分、亚型免疫细胞差异以及亚型免疫检查点的差异。

1.10 富集分析 在 DAVID(<https://david.ncifcrf.gov/home.jsp>)中进行 GO 和 KEGG 富集分析。

1.11 蛋白质相互作用网络 通过 GeneMANIA(<https://genemania.org>)构建蛋白质相互作用(protein-protein interaction, PPI)网络。

1.12 细胞系与培养 人正常乳腺上皮细胞 MCF10A(Procell, 中国)在 DMEM/F12(Biosharp, 中国)中培养,该培养基添加 5% 马血清、20 ng/mL 表皮生长因子(EGF)、0.5  $\mu$ g/mL 氢化可的松、10  $\mu$ g/mL 胰岛素、1% 非必需氨基酸溶液(NEAA)以及 1% 青霉素/链霉素。MCF7 和 T47D(Procell, 中国)细胞系的培养方法参照之前的方法<sup>[21]</sup>所述操作。

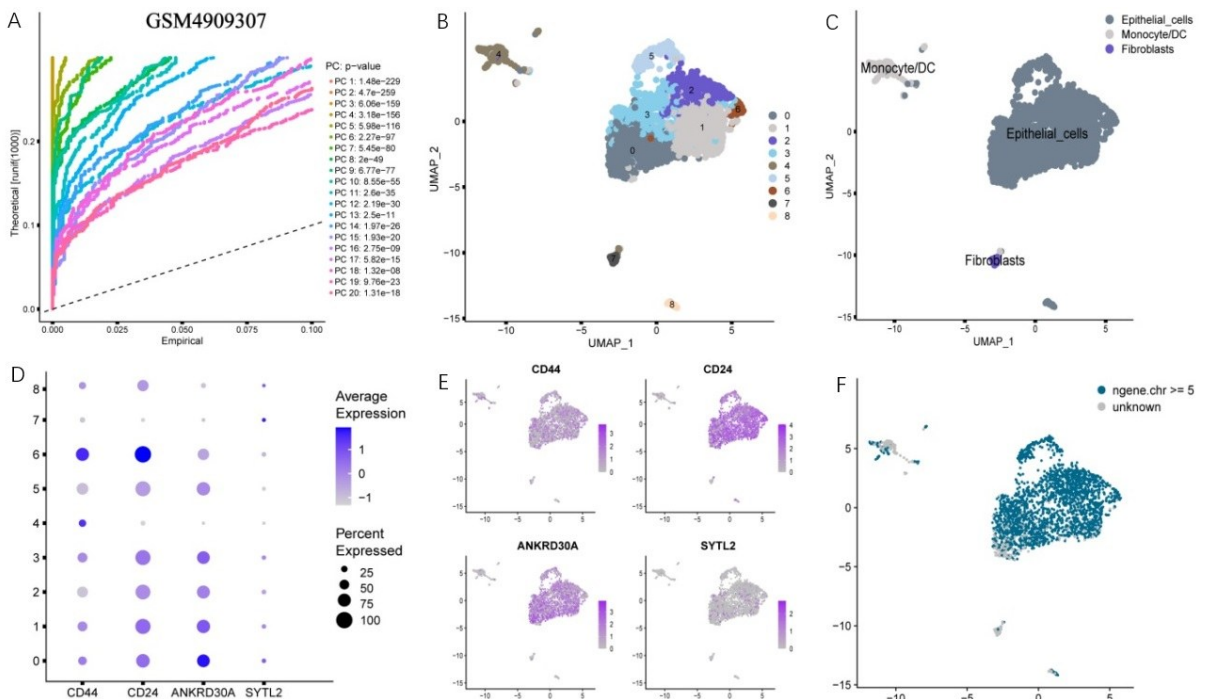
1.13 qRT-PCR 检测 KCNE4 的 mRNA 表达 用 TRIzol 提取细胞总 RNA,反转录成 cDNA 后,采用 iQ5 Real-time PCR 检测系统进行 PCR 扩增。PCR 反应条件:95  $^{\circ}$ C 预变性 5 min;95  $^{\circ}$ C 变性 10 s,60  $^{\circ}$ C 退

火 30 s,72  $^{\circ}$ C 延伸 1 min,共 35 个循环;最后 72  $^{\circ}$ C 延伸 5 min。以 GAPDH 为内参,采用  $2^{-\Delta\Delta CT}$  法计算 KCNE4 mRNA 相对表达量。GAPDH 引物序列:上游引物:5'-AGCCGAAGCAGGAACACC-3',下游引物:5'-TCAGCCTCGTGGTTGCAG-3';KCNE4 引物序列:上游引物:5'-AAACTGTCAGCCTTGCCCT-3',下游引物:5'-AGTACTCGTTGCCATTGCCG-3',实验重复 3 次。

1.14 Western Blot 分析 KCNE4 蛋白表达 在细胞中加入 RIPA(Beyotime, 中国)裂解液,冰上孵育 30 min,离心取上清,BCA 法测定蛋白浓度,取 50  $\mu$ g 蛋白进行 SDS-PAGE 电泳,转移至 PVDF 膜,5% 脱脂牛奶封闭 1 h,加入 KCNE4 抗体(1:1000, Proteintech, 中国),4  $^{\circ}$ C 孵育过夜,洗膜 3 次后加入二抗(1:1000, Beyotime, 中国),室温孵育 1 h 后 ECL 显色,最后使用 ImageJ 分析 KCNE4 蛋白表达量。

## 2 结果

2.1 利用单细胞 RNA 测序从 ER+ 乳腺癌样本中鉴定出上皮细胞 首先进行主成分分析并获得样本的前 20 个主成分用于降维(见图 1A),并通过 UMAP 方法对样本进行聚类后(见图 1B),根据 CellMarker 官网的 4 种上皮细胞标志物(CD44、CD24、ANKRD30A 和 SYTL2)与细胞簇间差异基因对细胞进行注释(见图 1C~图 1E)。随后使用 CopyKAT 方法中要求的每条染色体至少包含 5 个基因来计算 DNA 拷贝数的阈值条件( $\text{ngene. chr} \geq 5$ ),筛选出 CopyKAT 技术可识别的上皮细胞(见图 1F)。

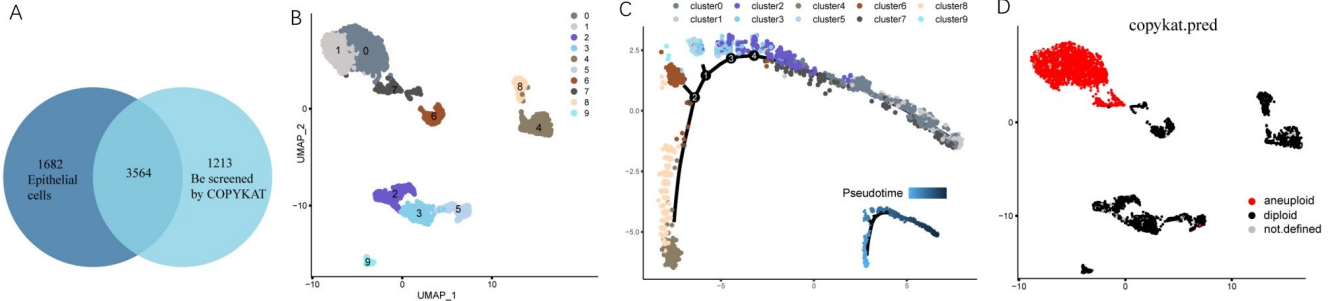


注:A 为样本通过主成分分析,获得前 20 个主成分;B 为样本的 UMAP 降维结果,分为 9 个簇;C 为细胞簇注释结果;D 为 4 个常见上皮细胞标志物的表达气泡图;E 为 4 个常见上皮细胞标志物的表达分布;F 为满足过滤标准  $\text{ngene. chr} \geq 5$  的细胞分布。

图 1 从单细胞样本中筛选上皮细胞

2.2 利用 CNV 分析区分癌细胞与正常上皮细胞 通过注释得到 5 146 个上皮细胞,其中 CopyKAT 识别出 3 564 个上皮细胞(见图 2A)。对这些细胞进行 UMAP 二次降维聚类,共得到 10 个上皮细胞亚群(见图 2B)。拟时序分析中的良性细胞和恶性细胞倾向于

聚集在不同分支中<sup>[22]</sup>。于是本研究结合拟时序分析结果与 CopyKAT 的 CNV 评分,发现 0/1/7 簇具有染色体非整倍体变异且倾向于聚集在拟时序轨迹的同一端(见图 2C、图 2D),提示 0/1/7 簇为癌细胞簇。



注:A为韦恩图显示注释上皮细胞与 CopyKAT 识别上皮细胞的交集;B为 3 564 个上皮细胞的 UMAP 二次降维聚类结果;C为 3 564 个上皮细胞的拟时序分析结果;D为 CopyKAT-CNV 评分可视化。

图 2 上皮细胞的二次降维聚类

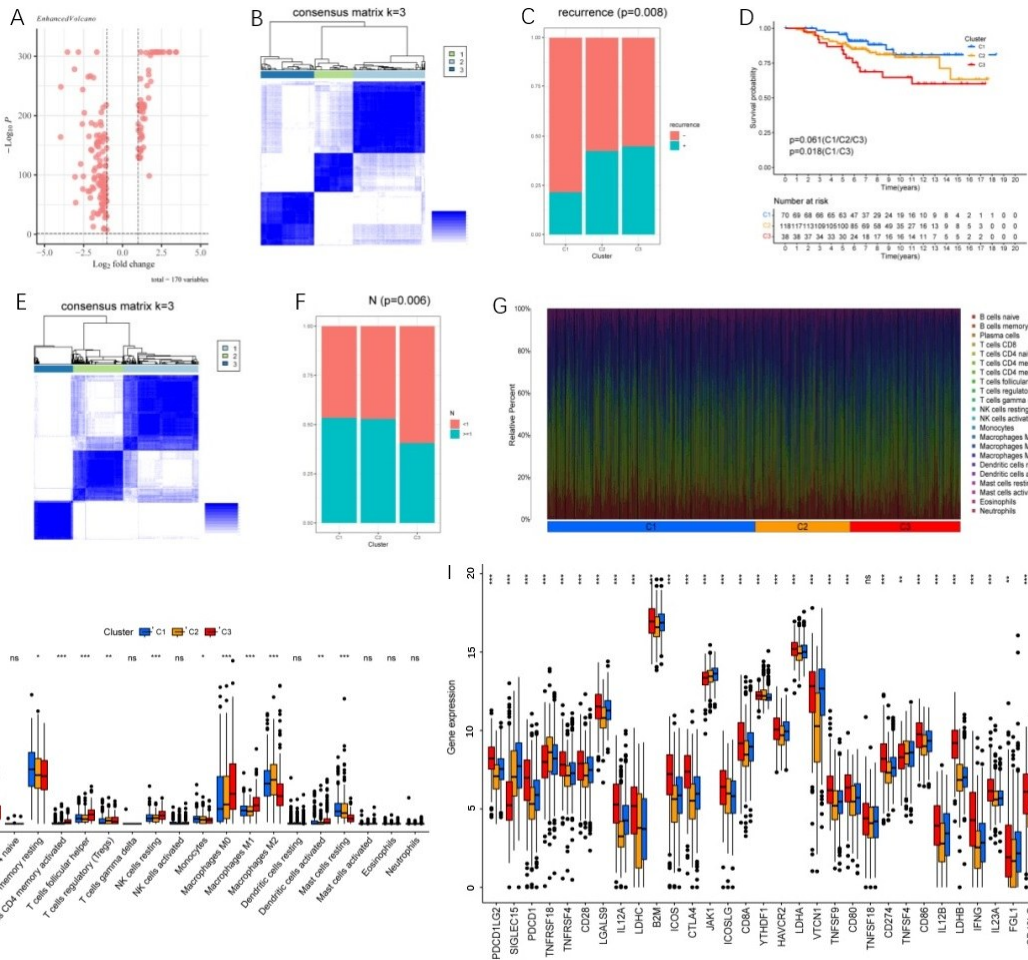
2.3 通过共识聚类验证差异基因的可靠性 通过分析 CopyKAT 区分的正常上皮细胞与 ER+乳腺癌细胞之间的差异基因表达水平,获得了 170 个显著差异表达基因(见图 3A)。随后,收集数据集(GSE159956)与 TCGA 数据库的乳腺癌 RNA-seq 样本,并通过共识聚类观察这 170 个基因对乳腺癌进展的影响。对于 GEO 数据库样本,共识聚类将样本分为 3 个亚型(见图 3B),发现肿瘤亚型在疾病复发方面存在显著差异(见图 3C),C1 亚型相对较低的复发率( $P < 0.05$ )也体现在其相对较好的生存分析结果中(见图 3D)。对于 TCGA 数据库样本,共识聚类将样本分为 3 个亚型(见图 3E),发现 C3 亚型肿瘤 N 分期存在显著差异(见图 3F),并且肿瘤免疫微环境分析(见图 3G、图 3H)和肿瘤免疫检查点分析(见图 3I)结果显示 C3 亚型中多种免疫细胞浸润程度以及免疫检查点基因表达水平与其他两亚型存在差异。这些结果提示 170 个差异表达基因可能对肿瘤生长和转移具有重要影响。

2.4 通过差异表达变化筛选癌细胞标志物 随后,从上皮细胞二次降维聚类结果中获取癌细胞与上皮细胞间差异表达变化最大的基因,同时展示它们在图 2B 上皮细胞簇中的表达情况。包括癌细胞中表达最高的 8 个基因: *SLITRK6*、*KCNE4*、*ECM1*、*CXCL14*、*KCNJ3*、*MRPS30*、*MRPS30-DT*、*LINC01238*(见图 4A),以及癌细胞中表达最低的 8 个基因: *TSC22D1*、*IGDBP5*、*AREG*、*CRIP2*、*MARCKS*、*PERP*、*MUCL1* 和 *GSN*(见图 4B)。

2.5 使用 RNA-seq 分析 16 个差异基因 使用 scRNA-seq 能够挖掘出 RNA-seq 难以发现的结果。上述 16 个基因在 TCGA RNA-seq 中的差异分析结果

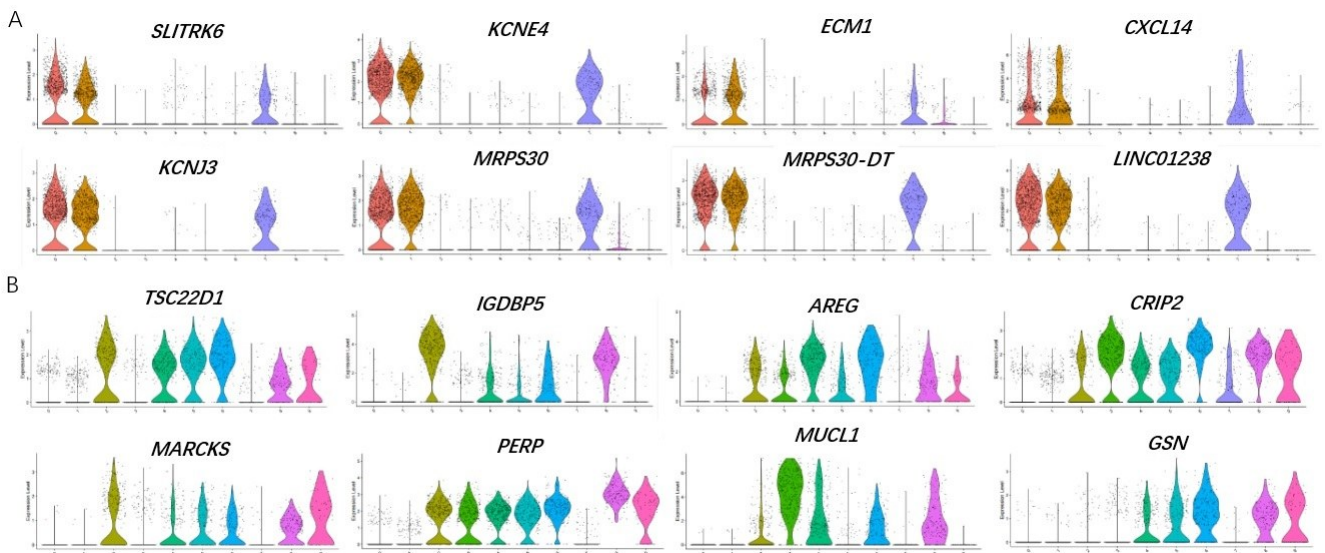
提示多个基因的表达差异并不显著(见图 5A、图 5B)。生存分析显示,16 个基因中只有 *MRPS30-DT* 和 *MARCKS* 的结果显著( $P < 0.05$ )(见图 5C)。这一现象表明 scRNA-seq 与 RNA-seq 结果会有差异,也暗示使用 scRNA-seq 数据挖掘的癌细胞标志物会更准确。

2.6 利用 scRNA-seq 验证 ER+乳腺癌高水平 mRNA 的表达 在 scRNA-seq 数据中标志物的选择以表达量增强的 mRNA 为主。从 GEO 数据库中选取 ER+乳腺癌样本(GSM4909296、GSM4909297、GSM4909298、GSM4909299、GSM4909300、GSM4909301)来展示 6 个现有标志物(*SOX2*、*EPCAM*、*MKI67*、*ESR1*、*ERBB2*、*PGR*)和剩余 7 个高表达 mRNA(*ECM1*、*KCNJ3*、*MRPS30*、*MRPS30-DT*、*SLITRK6*、*KCNE4*、*CXCL14*)的表达。第一个样本显示,*KCNE4* 与现有标志物表达相似(见图 6A)。第二个样本显示,*ECM1* 和 *KCNE4* 与现有标志物表达相似,而 *KCNJ3* 与现有标志物效果不一致(见图 6B)。第三个样本显示,*MRPS30*、*MRPS30-DT*、*SLITRK6* 和 *KCNE4* 与现有标志物表达相似,而 *KCNJ3* 与 *CXCL14* 现有标志物效果不一致(见图 6C)。第四个样本显示,*KCNJ3* 和 *SLITRK6* 与现有标志物表达相似,而 *ECM1* 和 *CXCL14* 在部分非癌细胞中也具有较高表达(见图 6D)。第五个样本显示,*KCNJ3*、*SLITRK6* 和 *KCNE4* 与现有标志物表达相似,而 *ECM1*、*MRPS30*、*MRPS30-DT* 和 *CXCL14* 与现有标志物效果不一致(见图 6E)。结果表明,*SLITRK6* 与 *KCNE4* 的标记效果与现有标志物更相似,且两者仅在 ER+乳腺癌细胞中高表达。因此,*SLITRK6* 和 *KCNE4* 可能更具有成为标志物的潜力。



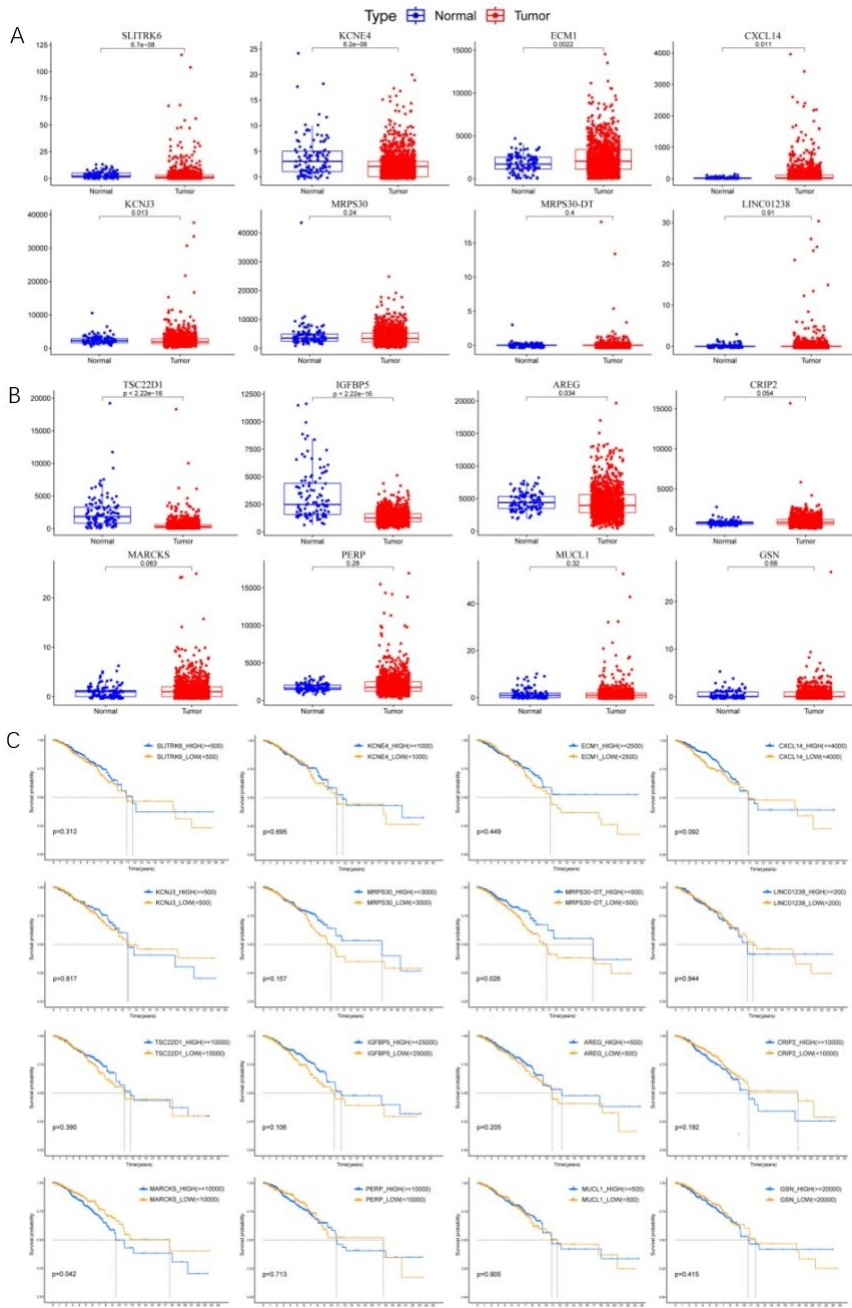
注:A为火山图显示了170个差异基因的表达;B为 GEO 数据库样本的共识聚类结果;C为 GEO 肿瘤分子亚型与癌症复发的关联结果;D为 GEO 肿瘤分子亚型的生存分析;E为 TCGA 数据库样本的共识聚类结果;F为 TCGA 肿瘤分子亚型与癌症N期的相关性;G为 TCGA 肿瘤分子分类的免疫细胞浸润情况;H为 TCGA 肿瘤分子亚型中免疫细胞浸润的差异分析;I为 TCGA 肿瘤亚型的免疫检查点分析。  
 ns:  $P > 0.05$ , \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ 。

图 3 乳腺癌细胞簇差异分析与共识聚类分析



注:A为癌细胞中8个表达最高基因;B为癌细胞中8个表达最低基因。

图 4 差异表达基因进一步筛选

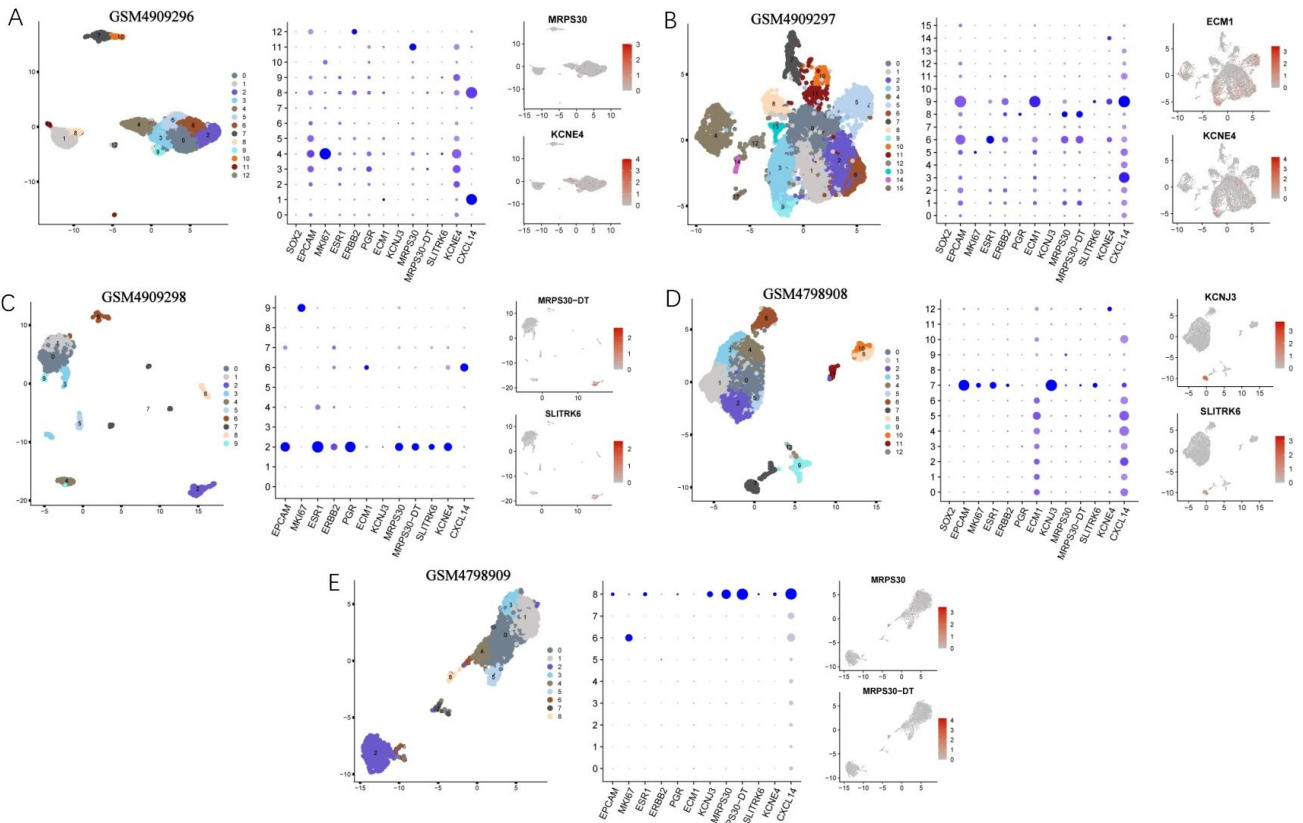


注:A为TCGA RNA-seq显示癌细胞中8个表达最高基因的表达情况;B为TCGA RNA-seq显示癌细胞中8个表达最低基因的表达情况;C为TCGA临床数据库BRCA样本中16个差异最显著基因的生存分析。

图5 16个差异基因在RNA-seq数据集中的表达分析

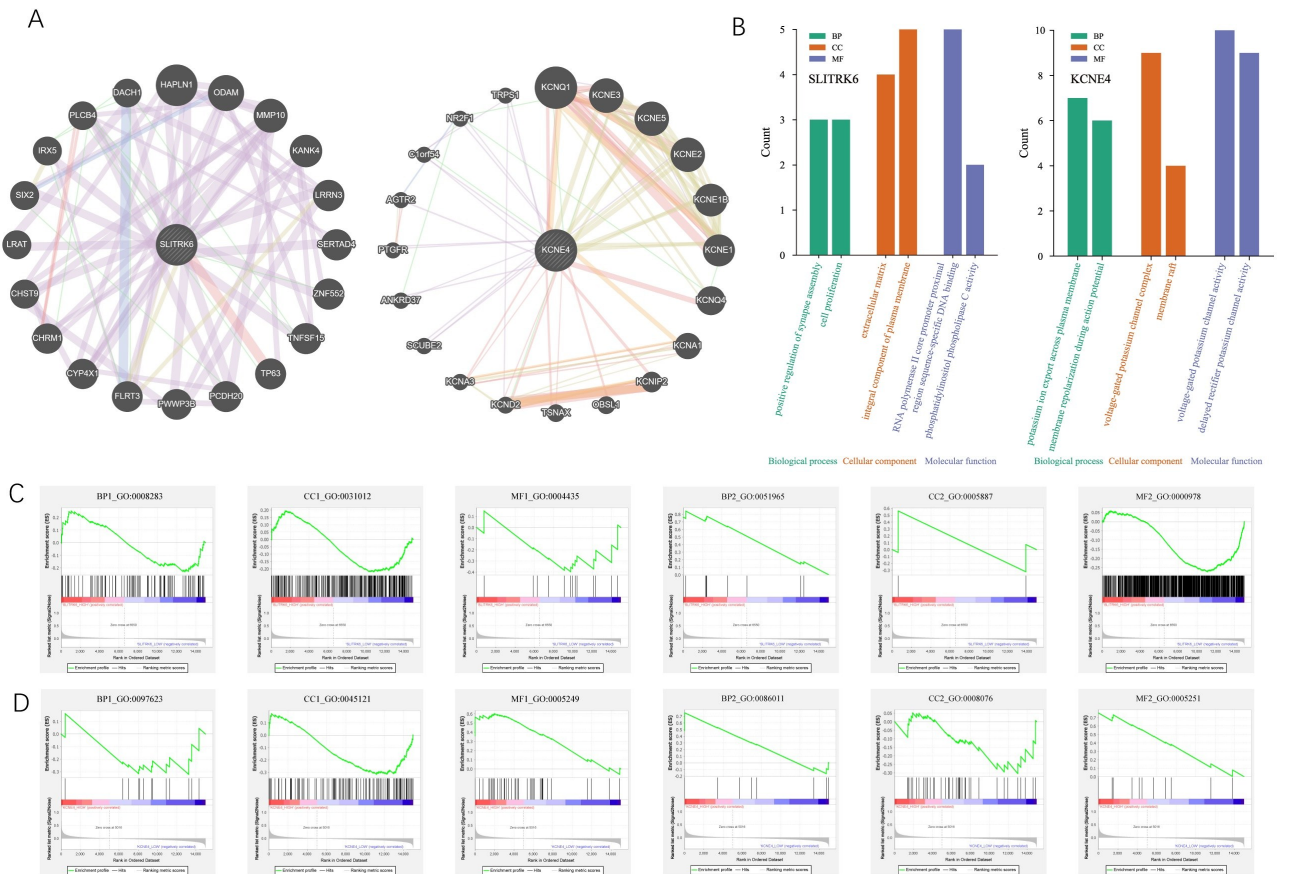
2.7 GSEA分析筛选癌标志物 由于癌细胞标志物往往与癌症进展联系密切,因此,为了进一步在SLITRK6和KCNE4中鉴定出哪个基因对癌症进展影响更重要,分别构建了两者的PPI网络(见图7A),并将与两者相互作用的基因分别进行GO富集分析。根据P值和富集基因数量,显示了前两个BP通路、前两个CC通路和前两个MF通路(见图7B)。随后,对这些通路进行GSEA富集分析(见图7C、图7D),获得每个通路的富集分数(ES)、FDR值和P值,并将这些

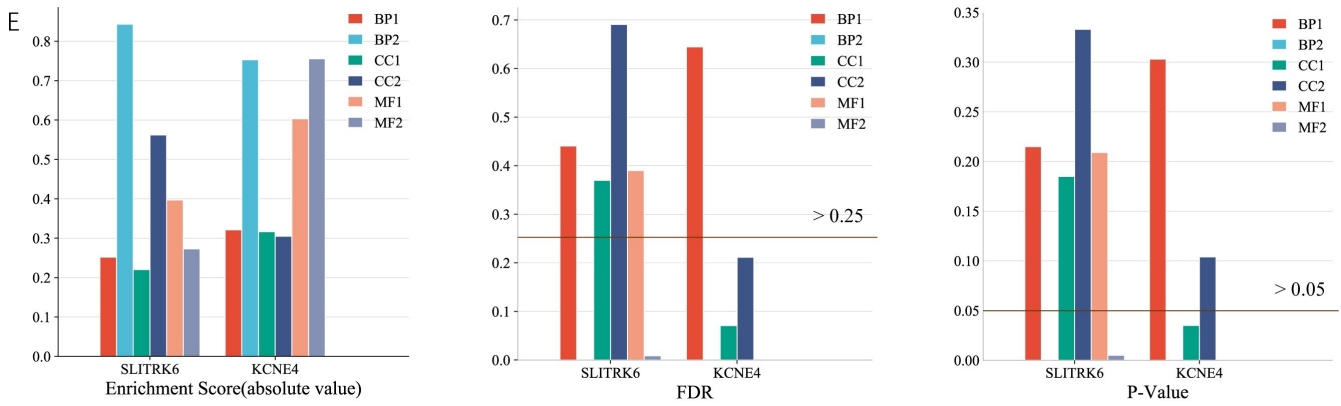
结果绘制为柱状图(见图7E)。统计结果显示,SLITRK6和KCNE4的总ES相似,然而,KCNE4显著影响到4条通路( $FDR < 0.25, P < 0.05$ ),而SLITRK6仅对2条通路影响显著。这表明KCNE4相关基因集在更多的生物学通路中显示出显著富集,提示KCNE4可能关联着一个更为广泛或核心的调控网络,这或许与其作为离子通道调节亚基的多功能特性有关。根据上述分析,KCNE4可能通过调控更广泛的信号网络影响癌症进展。



注:A为样本 GSM4909296 降维结果与 MRPS30 和 KCNE4 的表达谱;B为样本 GSM4909297 降维结果与 ECM1 和 KCNE4 的表达谱;C为样本 GSM4909298 降维结果与 MRPS30-DT 和 SLITRK6 的表达谱;D为样本 GSM4798908 降维结果与 KCNJ3 和 SLITRK6 的表达谱;E为样本 GSM4798909 降维结果与 MRPS30 和 MRPS30-DT 的表达谱。

图 6 ER+乳腺癌高表达 mRNA 在 scRNA-seq 样本中的验证



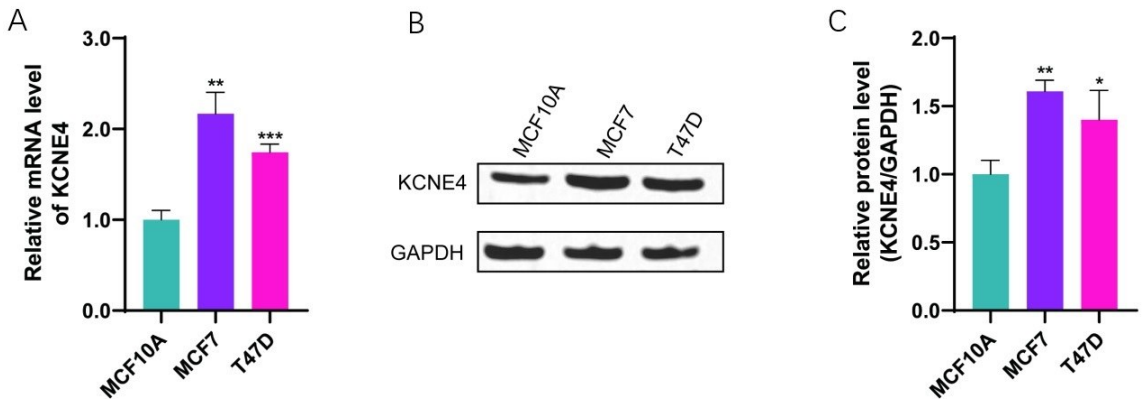


注:A为构建了SLITRK6和KCNE4的PPI网络;B为图A中SLITRK6、KCNE4的PPI基因的GO富集结果;C为SLITRK6的GSEA结果;D为KCNE4的GSEA结果;E为柱状图汇总了SLITRK6和KCNE4的GSEA结果(ES、FDR、P值)。

图7 SLITRK6和KCNE4的PPI基因在GO和GSEA中的富集结果

2.8 qRT-PCR和Western Blot验证KCNE4在ER+乳腺癌中呈上调表达 为探讨KCNE4表达是否与ER+乳腺癌细胞相关,对正常乳腺上皮细胞和ER+乳腺癌细胞进行qRT-PCR和Western Blot分析,结

果显示,与人类正常乳腺上皮细胞(MCF10A)相比,ER+乳腺癌细胞系(MCF7和T47D)的KCNE4表达在mRNA和蛋白质水平上均显著升高(见图8A~图8C)。



注:A为采用qRT-PCR测定正常乳腺上皮细胞和ER+乳腺癌细胞中KCNE4的相对mRNA表达水平; B为采用Western Blot测定正常乳腺上皮细胞和ER+乳腺癌细胞中KCNE4蛋白的表达水平; C为蛋白质印迹结果的定量分析。ns:  $P > 0.05$ , \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ 。

图8 KCNE4基因在ER+乳腺癌中qRT-PCR和Western Blot表达结果

### 3 讨论

ER+乳腺癌作为乳腺癌中最常见的亚型,其靶向治疗常因耐药问题而效果受限,因此挖掘新的ER+乳腺癌治疗靶点成为当前研究的重要挑战。癌症转录组差异分析是寻找有效治疗靶点的基础<sup>[23]</sup>,然而传统RNA-seq方法在分析肿瘤相关差异基因的方法存在一定局限,特别是由于基质细胞、免疫细胞与肿瘤细胞混合,导致对癌症表型的评估不准确。近年来,高通量scRNA-seq已成为解析细胞基因表达特征的强有力工具<sup>[24-25]</sup>。ER+乳腺癌细胞与正常上皮细胞具有一定同源性,利用scRNA-seq可深入分析ER+乳腺癌细胞的独特基因表型,并识别更可靠的癌细胞标志物。

在本研究中,对公共数据库中scRNA-seq数据进

行了UMAP降维,并基于CopyKAT提供的CNV评分方法筛选出癌细胞与正常上皮细胞,进而鉴定出170个在两类细胞间差异表达的基因。随后,采用能够分析癌症分子亚型的共识聚类<sup>[26]</sup>,验证了这些基因所定义的肿瘤亚型与临床特征之间的关联。进一步通过在上皮细胞中评估这些基因的表达水平,本研究筛选出16个候选基因,并在scRNA-seq样本中重点分析了其中7个癌细胞高表达的mRNA。结合GSEA富集结果,本研究发现KCNE4具备作为ER+乳腺癌标志物的潜力。KCNE4是电压门控钾通道的辅助亚基,参与多种生物学功能<sup>[27]</sup>。已有研究显示,人类KCNE4基因的突变与各种病理生理状态相关,尤其是心律失常<sup>[28]</sup>。同时,通过qRT-PCR和Western Blot

验证, KCNE4 在 ER+乳腺癌细胞高表达。本研究结果提示, KCNE4 也可能在 ER+乳腺癌进展中发挥重要作用, 是一个值得深入探索的重要癌标志物。本研究的核心贡献在于, 通过整合单细胞测序与多组学分析, 首次系统性地将 KCNE4 鉴定为一个在 ER+乳腺癌中特异性高表达且与不良预后显著相关的潜在新型生物标志物。本研究发现基于跨平台数据(单细胞及批量转录组)的一致性, 并初步在蛋白水平得到验证, 这为将其从候选基因推向临床应用提供了关键的第一步证据。然而, 必须指出的是, 本研究仍存在一定的局限性。首先, 当前的研究主要确立了 KCNE4 表达与 ER+乳腺癌之间的强相关性及其预后价值, 但尚未在细胞或动物模型中通过功能获得或功能缺失实验, 证实 KCNE4 对肿瘤恶性表型的“必要性”和“充分性”, 其具体的促癌分子机制也有待阐明。其次, 尽管本研究在现有样本中进行了验证, 但 KCNE4 作为独立预后标志物的临床效用, 仍需在将来通过大规模、多中心的前瞻性队列研究进行更严格的评估。未来的研究工作将聚焦于探索 KCNE4 影响 ER+乳腺癌细胞增殖、侵袭及药物敏感性等具体功能, 并深入解析其上下游调控网络, 以评估其作为治疗靶点的潜力。

综上所述, 本研究整合 scRNA-seq 数据与 CNV 分析方法, 系统筛选了 ER+乳腺癌细胞与正常上皮细胞之间的差异表达基因, 最终挖掘出 KCNE4 作为 ER+乳腺癌的潜在标志物。该发现为在单细胞水平上注释 ER+乳腺癌细胞提供了新的依据, 也为后续靶点开发奠定了理论基础。

#### 参考文献:

- [1] OCEGUERA-BASURTO P E, FIGUEROA-OCHOA E B, ANGUIANO-SEVILLA L A, et al. Evaluation of a polymeric topical formulation of Endoxifen in an estrogen receptor positive breast cancer murine model[J]. *Int J Pharm*, 2023, 642: 123175.
- [2] LIU YB, PARK S, LI Y. Breaking cancer's momentum: CDK4/6 inhibitors and the promise of combination therapy[J]. *Cancers*, 2025, 17(12): 1941.
- [3] ZAPATERO-SOLANA E, DING Y, PULLIAM N, et al. Models of early resistance to CDK4/6 inhibitors unveil potential therapeutic treatment sequencing[J]. *Int J Mol Sci*, 2025, 26(6): 2643.
- [4] HUSAIN A, KHAN H, FATIMA J E, et al. Recent advancements in breast cancer therapies and biomarkers: mechanisms and clinical significance[J]. *Curr Pharm Biotechnol*, 2025, 26(15): 2381-2402.
- [5] SHI X, WANG X, JIN L, et al. Bayesian identification of differentially expressed isoforms using a novel joint model of RNA-seq data[J]. *PLoS Comput Biol*, 2025, 21(1): e1012750.
- [6] HUANG GY, YU Y, SU H, et al. Integrating RNA-seq and scRNA-seq to explore the prognostic features and immune landscape of exosome-related genes in breast cancer metastasis[J]. *Ann Med*, 2025, 57(1): 2447917.
- [7] TOSKA E. Epigenetic mechanisms of cancer progression and therapy resistance in estrogen-receptor (ER+) breast cancer[J]. *Biochim Biophys Acta Rev Cancer*, 2024, 1879(3): 189097.
- [8] NADUKKANDY A S, KALAISELVAN S, LIN L, et al. Clinical application of single-cell RNA sequencing in disease and therapy[J]. *Clin Transl Med*, 2025, 15(11): e70512.
- [9] HU L, SU L M, CHENG H N, et al. Single-cell RNA sequencing reveals the cellular origin and evolution of breast cancer in BRCA1 Mutation carriers[J]. *Cancer Res*, 2021, 81(10): 2600-2611.
- [10] BREWER T, YEHIA L, BAZELEY P, et al. Integrating somatic CNV and gene expression in breast cancers from women with PTEN hamartoma tumor syndrome[J]. *NPJ Genom Med*, 2023, 8(1): 14.
- [11] LI Q Q, WANG R, YANG Z L, et al. Molecular profiling of human non-small cell lung cancer by single-cell RNA-seq[J]. *Genome Med*, 2022, 14(1): 87.
- [12] FAN J, LEE H O, LEE S, et al. Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data[J]. *Genome Res*, 2018, 28(8): 1217-1227.
- [13] SONG M F, MA S, WANG G, et al. Benchmarking copy number aberrations inference tools using single-cell multi-omics datasets[J]. *Brief Bioinform*, 2025, 26(2): bbaf076.
- [14] ZHANG X W, PENG M J, ZHU J H, et al. Benchmarking metabolic RNA labeling techniques for high-throughput single-cell RNA sequencing[J]. *Nat Commun*, 2025, 16(1): 5952.
- [15] GOVINDARAJAN M, WOHLMUTH C, WAAS M, et al. High-throughput approaches for precision medicine in high-grade serous ovarian cancer[J]. *J Hematol Oncol*, 2020, 13(1): 134.
- [16] YANG J Q, ZHANG J, NA S, et al. Integration of single-cell RNA sequencing and bulk RNA sequencing to reveal an immunogenic cell death-related 5-gene panel as a prognostic model for osteosarcoma[J]. *Front Immunol*, 2022, 13: 994034.
- [17] MONCADA R, BARKLEY D, WAGNER F, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas [J]. *Nat Biotechnol*, 2020, 38(3): 333-342.
- [18] TSUYUZAKI K, SATO H, SATO K, et al. Benchmark-

- ing principal component analysis for large-scale single-cell RNA-sequencing[J]. *Genome Biol*, 2020, 21(1):9.
- [19] ARAN D, LOONEY A P, LIU L Q, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage[J]. *Nat Immunol*, 2019, 20(2):163-172.
- [20] MABBOTT N A, BAILLIE J K, BROWN H, et al. An expression atlas of human primary cells; inference of gene function from coexpression networks[J]. *BMC Genomics*, 2013, 14:632.
- [21] BHASIN S, DUSEK C, PEACOCK J W, et al. Dependency of tamoxifen sensitive and resistant ER(+) breast cancer cells on semaphorin 3C (SEMA3C) for growth[J]. *Cells*, 2023, 12(13):1715.
- [22] PENG J Y, SUN B F, CHEN C Y, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma[J]. *Cell Res*, 2019, 29(9):725-738.
- [23] LOIBL S, POORTMANS P, MORROW M, et al. Breast cancer[J]. *Lancet*, 2021, 397(10286):1750-1769.
- [24] XIAO M L, XIE L, CAO GS, ET AL. CD4<sup>+</sup> T-cell epitope-based heterologous prime-boost vaccination potentiates anti-tumor immunity and PD-1/PD-L1 immunotherapy[J]. *J Immunother Cancer*, 2022, 10(5):e004022.
- [25] GALA DE PABLO J, LINDLEY M, HIRAMATSU K, et al. High-throughput Raman flow cytometry and beyond[J]. *Acc Chem Res*, 2021, 54(9):2132-2143.
- [26] RAKHA E A, PAREJA F G. New advances in molecular breast cancer pathology[J]. *Semin Cancer Biol*, 2021, 72:102-113.
- [27] XIAO J S, CHENG X, HUANG D, et al. Functions of the KCNE gene family in ion channels[J]. *Biochem Genet*, 2025. DOI:10.1007/s10528-025-11202-3.
- [28] KELEMEN K, GREENER I D, WAN X P, et al. Heterogeneous repolarization creates ventricular tachycardia circuits in healed myocardial infarction scar[J]. *Nat Commun*, 2022, 13(1):830.

收稿日期:2026-01-06;修回日期:2026-01-11

(本文编辑 覃黎黎)

(上接第 156 页)

- [19] FU J N, WU H. Structural mechanisms of NLRP3 inflammasome assembly and activation[J]. *Annu Rev Immunol*, 2023, 41:301-316.
- [20] LI X D, ZHANG Z Y, HAN Y, et al. NLRP3 inflammasome and pyroptosis; implications in inflammation and multisystem disorders[J]. *PeerJ*, 2025, 13:e19887.
- [21] TOLDO S, MEZZAROMA E, BUCKLEY L F, et al. Targeting the NLRP3 inflammasome in cardiovascular diseases[J]. *Pharmacol Ther*, 2022, 236:108053.
- [22] LV D M, CAO X L, ZHONG L, et al. Targeting phenylpyruvate restrains excessive NLRP3 inflammasome activation and pathological inflammation in diabetic wound healing[J]. *Cell Rep Med*, 2023, 4(8):101129.
- [23] LAABEI J, VEGLIANTE G, STROGULSKI N R, et al. The NOX2-ROS-NLRP3 inflammasome axis in traumatic brain injury[J]. *J Neuroinflammation*, 2025, 22(1):242.
- [24] DOĞAN MF, KAYA K, DEMIREL H H, et al. The effect of vitamin C supplementation on favipiravir-induced oxidative stress and proinflammatory damage in livers and kidneys of rats[J]. *Immunopharmacol Immunotoxicol*, 2023, 45(5):521-526.
- [25] BABAEENEZHAD E, FARAHMANDIAN N, SOTOUDEHEIAN M, et al. Resveratrol relieves hepatic steatosis and enhances the effects of atorvastatin in a mouse model of NAFLD by regulating the renin-angiotensin system, oxidative stress, and inflammation[J]. *Food Sci Nutr*, 2025, 13(3):e70073.
- [26] PAIS M. The effect of HPA-axis dysregulation and glucocorticoid administration on strength and generalization of extinction memory[J]. *Psychoneuroendocrinology*, 2024, 160:106740.
- [27] 门钟兰, 黄竹芸, 章彦文, 等. 基于 AMPK/p38MAPK/NF-κB 通路探究湿润烧伤膏改善咪喹莫特诱导小鼠银屑病样的机制研究[J]. *中国药物评价*, 2025(1):21-27.

收稿日期:2025-11-14;修回日期:2026-01-08

(本文编辑 覃黎黎)