doi:10.3971/j.issn.1000-8578.2026.25.0694

• 基础研究 •

基于单细胞与机器学习的三阴性乳腺癌上皮细胞亚群解析与预后模型构建

吴金鹏[1]，郭雪[1,2*]，刘恩谷[2]，林峰[1]，李鸿涛[1]

Single-Cell and Machine Learning-Based Identification of Epithelial Subsets and Prognostic Modeling in Triple-Negative Breast Cancer

WU Jinpeng[1], GUO Xue[1,2*], LIU Engu[2], LIN Feng[1], LI Hongtao[1] (*: Contributed Equally as the First Author)

1. Department of Breast and Thyroid Surgery, Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi 830000, China; 2. School of Clinical Medicine, Xinjiang Medical University, Urumqi 830017, China

Corresponding Author: LI Hongtao, E-mail: lht4656@163.com

Abstract: Objective To investigate the heterogeneity and key molecular features of epithelial cells in triple-negative breast cancer (TNBC), identify prognostic biomarkers, and develop a robust survival prediction model. **Methods** Using TNBC single-cell transcriptomic data, epithelial cells were extracted, normalized, and subclustered to characterize their molecular signatures and functional differences. High-dimensional weighted gene co-expression network analysis (hdWGCNA) was applied to establish co-expression modules in epithelial cells. Multiple machine learning algorithms were integrated to select key prognostic genes and develop a risk-score model, whose performance was evaluated using receiver operating characteristic (ROC) curves and Kaplan-Meier (K-M) survival analysis. In addition, the immune microenvironment features and potential drug-response differences between the high- and low-risk groups were systematically assessed. Finally, PCR was performed to validate the expression differences of the key genes between tumor and normal tissues. **Results** We characterized the composition and molecular features of TNBC epithelial subpopulations and identified a TNBC-associated epithelial subset. By integrating hdWGCNA with machine learning approaches, 10 key genes were selected to construct a prognostic model, which effectively stratified patients into distinct survival-risk groups and demonstrated favorable predictive performance in ROC and K-M analyses. Immune profiling revealed the differences in the infiltration levels of seven immune cell types and immune function-related features between the high- and low-risk groups. Drug-sensitivity analysis suggested potential differential responses to eight agents across the risk groups. PCR validation further confirmed the differential expression of the ten signature genes between tumor and normal tissues. **Conclusion** This study reveals epithelial heterogeneity in TNBC at single-cell resolution and establishes a 10-gene prognostic model, which may facilitate the stratification of TNBC risk and the evaluation of immune characteristics and potential therapeutic strategies.

Key words: Triple-negative breast cancer; Epithelial cell; Machine learning; Prognostic model; scRNA-seq; hdWGCNA

Funding: "Tianshan Talents Cultivation Program" Technology and Innovation Leading Talents Project (No. 2023TSYCLJ0039)

Competing interests: The authors declare that they have no competing interests.

摘 要：目的 探讨三阴性乳腺癌（TNBC）上皮细胞的异质性与关键分子特征，筛选预后相关标志物并构建稳定的生存预测模型。**方法** 基于TNBC单细胞转录组数据，对上皮细胞进行提取、标准化及聚类分型，并解析其分子特征与功能差异；采用hdWGCNA构建上皮细胞共表达模块，结合多种机器学习算法筛选关键预后基因并建立风险评分模型；通过ROC曲线与Kaplan-Meier（K-M）生存分析验证模型性能。进一步评估高、低风险组的免疫微环境特征及潜在药物反应差异，并采用PCR验证关

收稿日期：2025-09-05；**修回日期**：2025-11-14

基金项目：“天山英才”培养计划科技创新领军人才项目（No. 2023TSYCLJ0039）

作者单位：1. 830000 乌鲁木齐，新疆医科大学附属肿瘤医院乳腺甲状腺外科；2. 830017 乌鲁木齐，新疆医科大学临床医学院

通信作者：李鸿涛，男，博士，主任医师，教授，主要从事乳腺癌相关临床与教学、科研工作，E-mail: lht4656@163.com，ORCID: 0000-0003-2003-0015

作者简介：吴金鹏，男，硕士在读，主要从事乳腺癌的相关研究，ORCID: 0009-0004-6275-7452；郭雪，女，硕士在读，主要从事乳腺癌的相关研究，ORCID: 0009-0008-6999-2999（*：并列第一作者）

键基因在肿瘤与正常组织中的表达差异。**结果**　本研究解析了TNBC上皮细胞的亚群组成及其分子特征，识别出与TNBC相关的上皮细胞亚群。基于hdWGCNA与机器学习联合筛得10个关键基因，构建的预后模型能够有效区分不同生存风险人群，并在ROC与K-M分析中表现出良好的预测能力。免疫分析显示，高、低风险组在7类免疫细胞浸润水平及免疫功能相关特征上存在差异；药物敏感性评估提示8种药物在不同风险组间存在潜在响应差异。PCR进一步验证了10个关键基因在肿瘤与正常组织中的表达差异。**结论**　本研究在单细胞分辨率下揭示了TNBC上皮细胞异质性，并建立了基于10基因的预后模型，可用于TNBC患者风险分层及免疫特征与潜在治疗策略评估。

关键词：三阴性乳腺癌；上皮细胞；机器学习；预后模型；单细胞RNA测序；hdWGCNA

中图分类号：R737.9

0　引言

三阴性乳腺癌（Triple-negative breast cancer, TNBC）因缺乏ER、PR、HER2表达，对内分泌与HER2靶向治疗不敏感，较其他亚型更易转移和复发，预后也更差[1-3]。流行病学显示其约占乳腺癌的10%~20%，在年轻女性中更常见，非裔美国人与西班牙裔的发病率更高[4-6]。临床诊断通常结合影像学、病理（含免疫组织化学）与分子标志物，治疗以紫杉醇、蒽环类新辅助化疗为基础，并辅以个体化与免疫治疗[7]。对BRCA1/2突变患者，PARP抑制剂正逐步纳入治疗版图，联用化疗或免疫治疗有望提升疗效并缓解耐药[8-9]。尽管相关认识不断进展，TNBC的诊疗仍面临挑战，亟需可靠的生物标志物与预后模型支撑临床决策。

单细胞转录组测序可在细胞分辨率层面解析肿瘤细胞谱系组成及其与微环境的相互作用，为挖掘更稳健的预后相关分子模块提供新的研究路径。因此，本研究聚焦TNBC相关上皮细胞亚群，筛选其特征基因并提取中心基因构建生存预测模型，进一步系统评估不同风险组的免疫浸润特征、功能状态、免疫逃逸倾向及潜在治疗靶点，并对关键预后基因进行验证，为TNBC的机制研究与临床风险评估提供依据。

1　材料与方法

1.1　细胞与培养条件

人乳腺上皮细胞MCF-10A与三阴性乳腺癌细胞MDA-MB-231、MDA-MB-468细胞均购自ATCC细胞库（美国弗吉尼亚州马纳萨斯），并按ATCC公布的培养方案进行培养（https://www.atcc.org/）。

1.2　数据下载

从GEO数据库下载数据GSE161529[10-11]，用以获得三阴性乳腺癌的单细胞（scRNA-seq）数据，包括13个正常样本和8个三阴性乳腺癌样本（共126 297个细胞）。

从TCGA下载乳腺癌（BRCA）数据后，筛得117例TNBC与113例正常乳腺样本，用于预后建模和差异分析[12-13]。从GEO获取GSE58812数据集，包含107例TNBC的常规转录组及完整生存信息[14]。

1.3　伦理学声明

本研究已获新疆医科大学附属肿瘤医院伦理委员会批准（批准号：K-2021076）。

1.4　单细胞测序数据的初步处理

采用Seuratv4.3.0软件包进行单细胞分析[15]。质控：保留线粒体基因比例<25%、检测基因数200~7 000的细胞。去除细胞周期影响后，使用SCTransform进行标准化和缩放并提取高变基因[16]。随后以PCA降维并聚类，采用Harmony方法校正批次效应。用FindAllMarkers鉴定各簇特异基因并完成细胞注释[17]。提取上皮细胞后，重复上述流程（标准化/缩放、降维聚类与批次校正）。

1.5　细胞-细胞通讯分析

采用CellChatv1.5.0[18]按推荐流程（默认参数）构建细胞间通信网络，并筛得与其他细胞关联最强的上皮细胞亚群。

1.6　富集分析和拟时序分析

使用“FindAllMarkers”筛选差异基因，采用clusterProfilerv4.7.1进行细胞亚群间基因集富集分析[19-20]。提取7个肿瘤样本的上皮细胞，使用Monocle[21]开展拟时序分析，以推断其发育轨迹。

1.7　高维WGCNA分析

高维WGCNA（hdWGCNA）可在基因组学数据中更稳健地识别模块并解析其与临床特征的关系[22]。本研究按标准流程使用hdWGCNA包[23-24]：先以MetacellsByGroups为每个样本与细胞簇构建元细胞（每个含50个细胞），提取上皮细胞重建Seurat对象，依次运行TestSoftPowers、ConstructNetworks、ModuleEigengenes、ModuleConnectivity、RunModuleUMAP，最终用HubGeneNetworkPlot进行枢纽基因网络/交互分析[25]。

1.8　预后基因筛选

以韦恩图求取0簇上皮细胞标志基因与TNBC相关上皮细胞模块基因的交集[26]；随后使用survival包对交集基因进行单因素Cox回归，筛选与生存相关基因（$HR\neq1$，$P<0.05$）[27]。

1.9　机器学习结合预后模型

基于TCGA的TNBC与正常样本筛选差异基因。为评估hdWGCNA获得的TNBC相关上皮细胞基因的预后价值，将其与差异基因求交后用于建模。采用随机森林[28]与SVM-RFE[29]双重筛选：随机森林采用R randomForest包，set.seed=123。先以ntree=500获得OOB误差曲线，按OOB最小确定最终树数（ntree=71）；按MeanDecreaseGini由高到低取前30位基因。随后以SVM-RFE分析采用外层5折交叉验证，RFE内部k=5，并设halve.above=100；对不同特征子集扫描，以CV错误率最低确定最终特征数。最终以RF Top30 ∩ SVM-RFE Top30用于后续建模。以该交集基因为特征，进行LASSO与多因素Cox回归[30]构建预后模型。用ROC曲线以及Kaplan-Meier（K-M）生存曲线值观察模型可靠性[31]。

1.10 免疫微环境分析

在TCGA与GSE58812整合的TNBC队列中，计算22种免疫细胞浸润评分，并以Wilcoxon检验比较两风险组差异（P<0.05）。同时基于已发表的13组免疫相关功能基因集[32]，采用GSVA[33]包的ssGSEA算法计算功能评分并比较组间差异（P<0.05）。另对两组的免疫表型评分与代谢通路评分进行对比。此外，还研究了免疫评分、基质评分和TumorPurity评分的差异[34]。

1.11 预后基因的表达验证

在GSE161529单细胞数据中判定这些基因在TNBC各细胞类型的表达；在TCGA中以Wilcoxon检验比较肿瘤与正常样本的预后基因表达差异，并用K-M生存分析评估其对生存的影响；同时利用HPA在线数据库观察这些预后基因的蛋白水平表达差异[35]。

1.12 实时荧光定量PCR（qRT-PCR）

使用RNAsimple Total RNA Kit（Solarbio, 中国）提取总RNA，并以NanoDrop测定纯度与浓度（A260/280≈1.8~2.1）。取1 μg RNA，用PrimeScript RT试剂（TaKaRa Bio, 中国）反转录为cDNA（20 μl体系）。qPCR采用TB Green体系（TaKaRa Bio, 中国），在QuantStudio 6 Flex（Applied Biosystems）上进行；每孔20 μl：TB Green Mix 10 μl、正/反向引物各0.4 μl（10 μmol/L）、cDNA 2 μl、无核酸酶水补足。循环条件：95℃ 30 s；随后40个循环（95℃ 5 s，60℃ 30 s）；末端进行熔解曲线以验证特异性。内参基因使用GAPDH；目标基因包括TCF7L2、CBR1、CHI3L1、KRT15、PDAP1、ISCU、IMPDH2、DDIT4、BTG3、RPN1。每个条件n=3生物学重复，每孔技术重复3次。相对表达按$2^{-\Delta\Delta Ct}$方法计算。

1.13 统计学方法

使用GraphPadPrism9.0统计与R软件（版本4.3.2）对数据进行处理和分析[36]。所有统计检验均为双侧检验。对于定量数据，使用双尾、非配对Student's t检验或单因素方差分析（ANOVA）结合Tukey多重比较检验来比较两个或多个组之间的均值。使用Wilcoxon方法检验高、低风险组间参数的差异[37]。P<0.05认为差异有统计学意义。

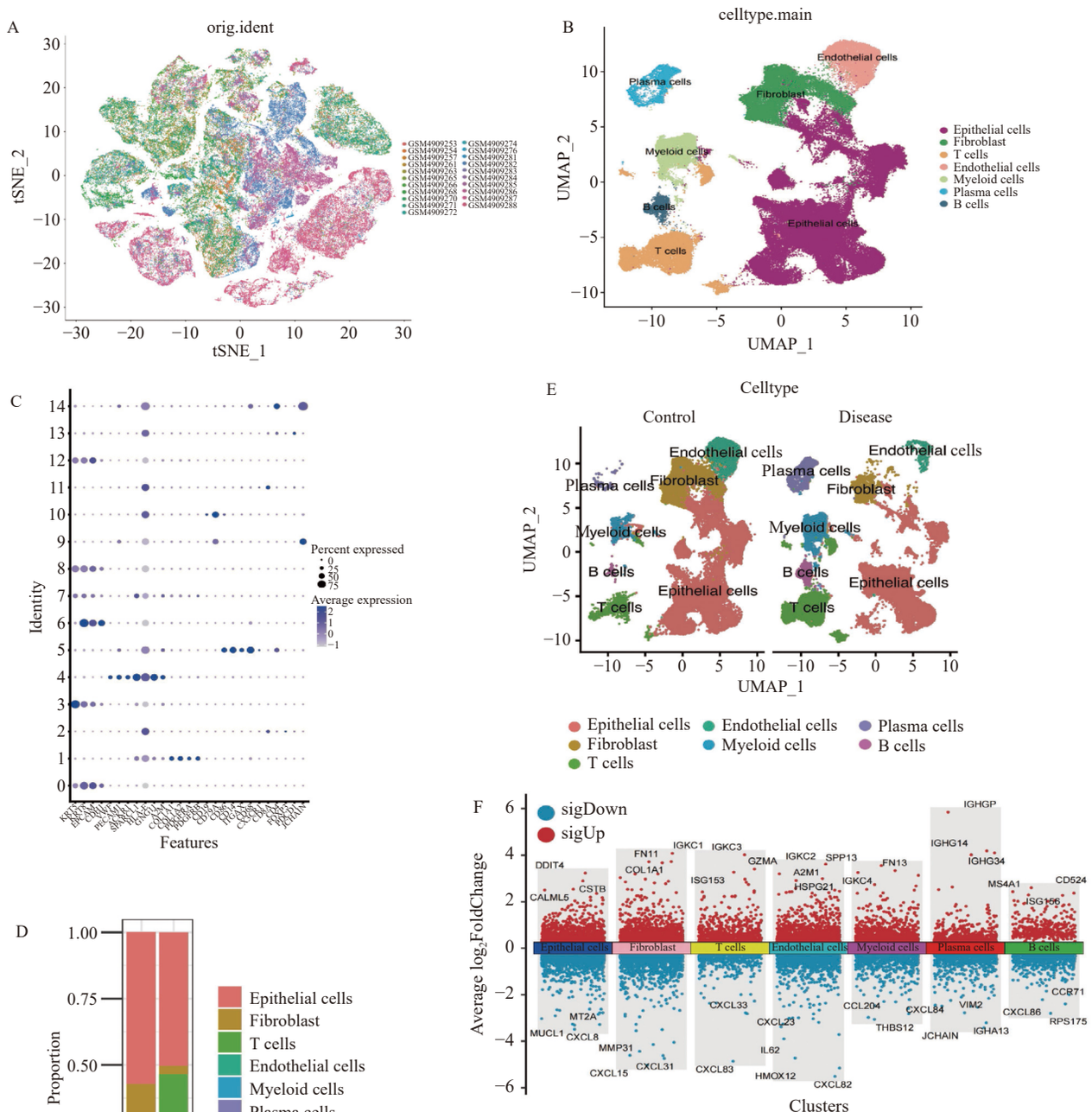
2 结果

2.1 单细胞测序分析揭示三阴性乳腺癌组织中细胞组成的变化

为探讨TNBC组织的细胞组成变化，我们整合正常与TNBC的单细胞测序数据。经质控与批次校正后，获得118 739个合格细胞（每细胞表达25 951个基因）。经无监督降维与聚类并注释，鉴定7类细胞并以UMAP展示（图1A、1B）：上皮细胞（KRT5、KRT8、EPCAM、CDH1）、内皮细胞（VWF、PECAM1、ACKR1、SPARCL1、HLA-E、GNG11、A2M）、成纤维细胞（COL1A1、COL1A2、PDGFRA、PDGFRB）、T细胞（CD8A、CD4、FOXP3、PDCD1）、B细胞（CD19、CD79A）、髓系细胞（CD86、CD14、ITGAX、CD68、CX3CR1）及浆细胞（JCHAIN）（图1C）。与正常组织相比，TNBC组织中T细胞、髓系细胞、B细胞与浆细胞显著增多，成纤维细胞明显减少，上皮细胞相对减少（图1D、E）。差异表达分析显示，TNBC上皮细胞上调最多的3个基因为DDIT4、CSTB、CALML5，下调最多的3个基因为MT2A、MUCL1、CXCL8（图1F）。

2.2 单细胞分析显示三阴性乳腺癌组织中上皮细胞的异质性

为进一步解析上皮细胞的特异性变化，我们提取上皮细胞后重新标准化、降维聚类并校正批次效应，得到10个亚簇（图2A）。与正常组织相比，TNBC中1、2、4簇显著减少，而0、3簇显著增加，其中3簇被鉴定为TNBC相关上皮细胞（图2B）。对10个簇的前十标志基因绘制热图（图2C）。随后聚焦0簇，进行差异表达分析并绘制火山图（图2F），开展富集分析：GSEA显示阿尔茨海默病、氧化磷酸化、核糖体等通路上调；趋化因子信号、细胞因子—受体相互作用、ECM—受体相互作用、焦点黏附及NOD样受体信号通路下调（图2D、E）。



A: t-SNE map showing the cellular distribution in normal and triple-negative breast cancer (TNBC) tissues; B: unsupervised clustering categorized cells into seven clusters, color-coded and labeled by cell identity; C: marker-gene expression depicted as UMAP feature maps and a bubble plot; D, E: proportional comparison of the seven cell types between the normal and tumor groups; F: for each cell type, visualization of the top three differentially expressed genes between normal and TNBC ($|\log_2FC|>0.5$, FDR<0.05).

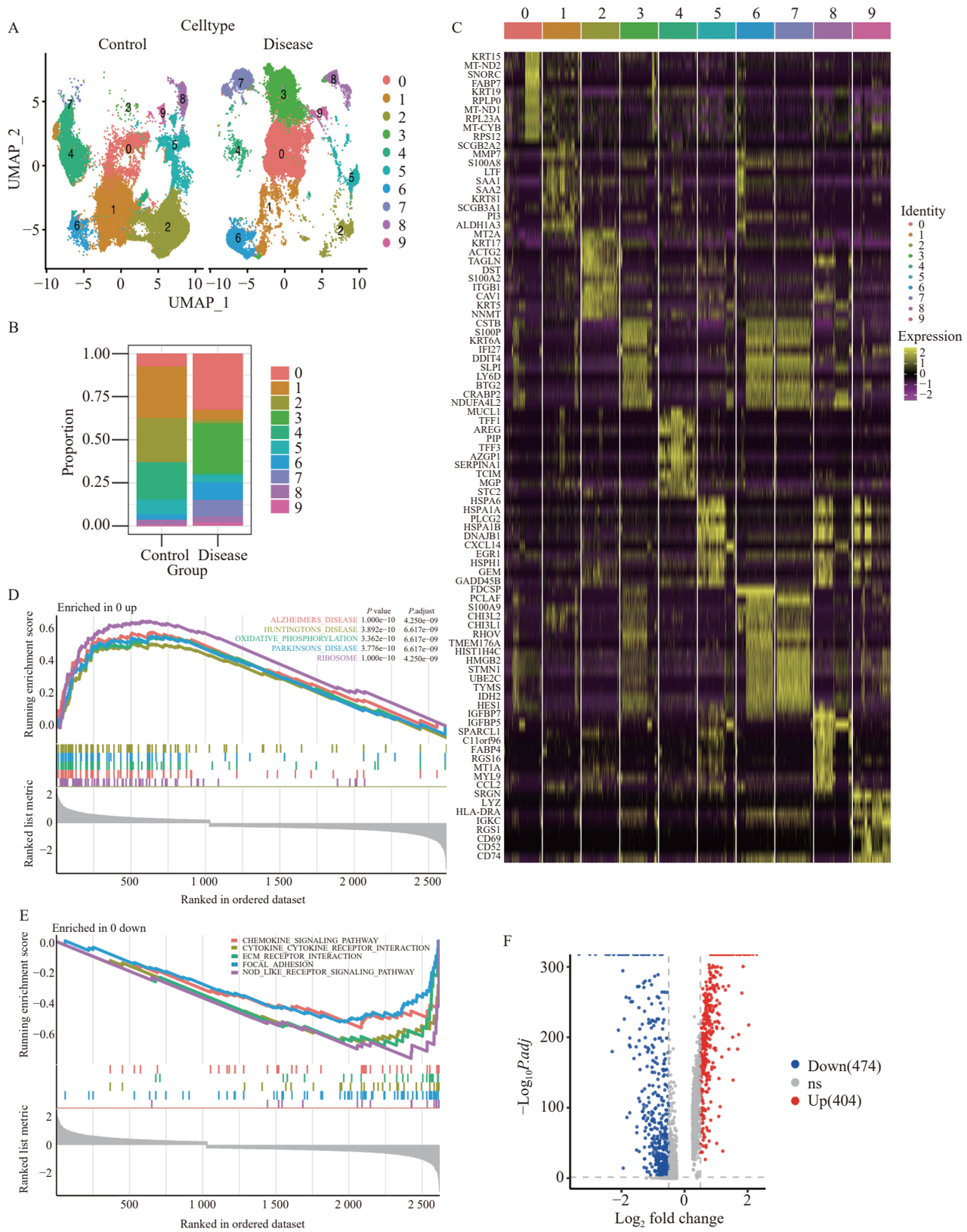
图 1　正常组织与三阴性乳腺癌组织的单细胞图谱与差异特征

Figure 1　Single-cell atlas and differential signatures in normal versus TNBC tissues

2.3　三阴性乳腺癌相关上皮细胞的细胞通讯分析和拟时序分析

结合细胞通讯与拟时序分析，我们系统解析了上皮细胞在TNBC中的作用。CellChat[38]显示3簇上皮细胞与多类细胞交互最为密集（图3A、B），与内皮、髓系和T细胞的通讯尤为显著（图3C）。其强交互主要经APP–CD74、CD99–CD99、MIF–（CD74+CXCR4/CD44）、HLA-C–KIR2DL3等配体—受体轴介导（图3C）。在MIF、CD99、MHC-Ⅰ、APP四类信号中，3簇参与APP信号比例最高；其在CD99、MHC-I通路中以influencer为主，在APP、MIF通路兼具sender与influencer（图3E）。上述通路涉及黏附、迁移、增殖与存活，影响转移、纤维化、免疫反应及NK细胞杀伤，提示其在TNBC发病中可能很关键。

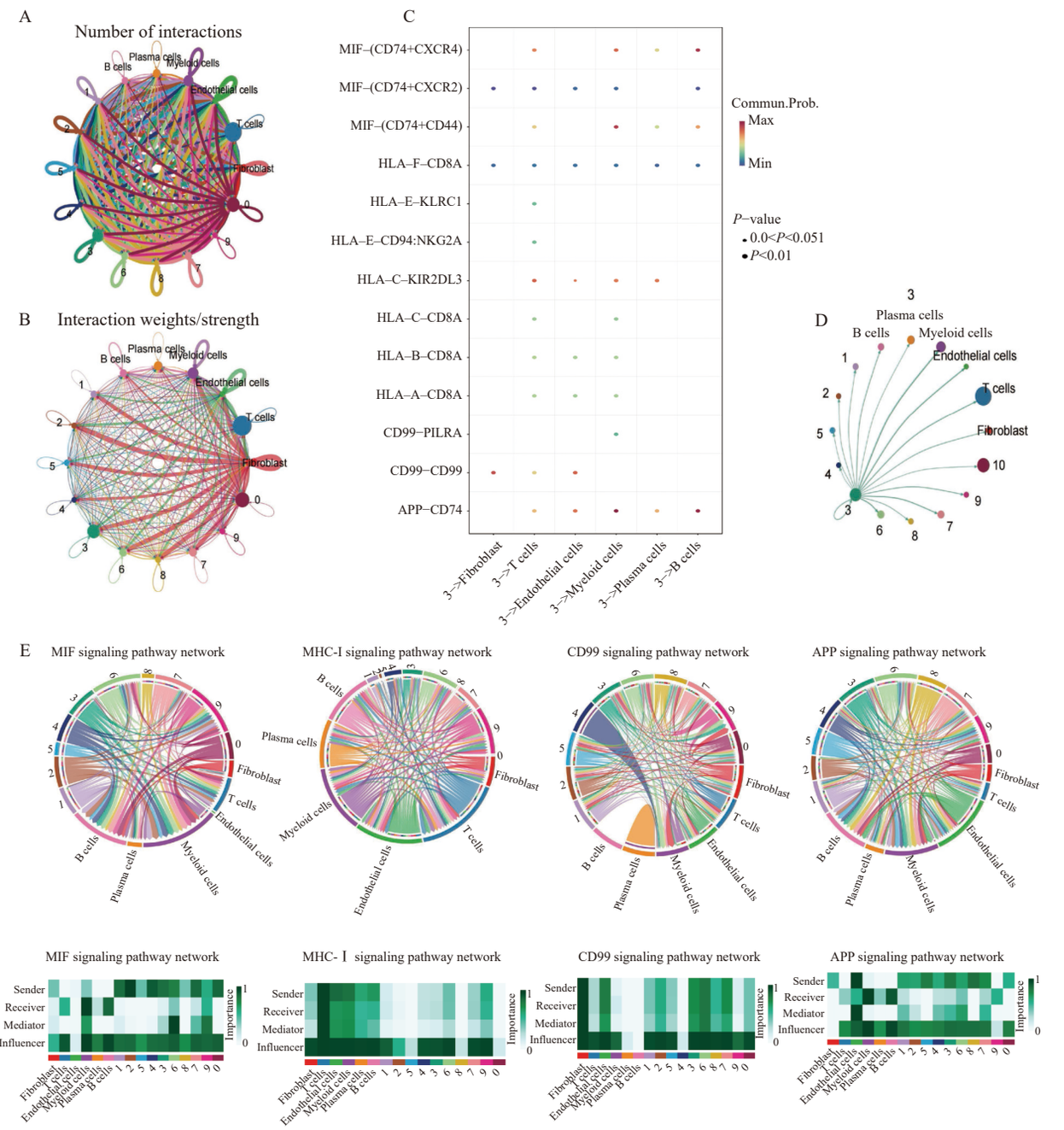
拟时序结果显示，4簇位于轨迹起始，3簇出现



A: UMAP visualization showing the distribution of ten epithelial subtypes across normal and TNBC tissues; B: proportional composition of epithelial cells by subtype in the normal versus tumor groups; C: heatmap of the top 10 marker genes for each epithelial subtype identified by Seurat's FindAllMarkers function; D, E: GSEA results for DEGs in cluster 0 epithelial cells, highlighting the top enriched pathways/GO terms (e.g., NES and FDR q-values as indicated); F: volcano plot of differentially expressed genes in cluster 0 epithelial cells comparing TNBC with normal tissues.

图 2　上皮细胞亚群分析

Figure 2　Analysis of epithelial subpopulations



A, B: global visualization of communication patterns among all cell types, including the number and strength of inferred interactions; C: bubble plot showing ligand-receptor-mediated communication probabilities between the three epithelial cluster sand the six other cell classes; D: network-style visualization of communications between the three epithelial clusters and other cell types; E: donut charts and heatmaps summarizing cell-type contributions and pathway activity for four signaling pathways: MIF, CD99, MHC-Ⅰ, and APP.

图 3　上皮细胞亚群与肿瘤微环境的配体–受体通讯特征

Figure 3　Ligand-receptor communication between epithelial subsets and the tumor microenvironment

在分支1/2，指示其特定演化过程（图4A~C）。

2.4　hdWGCNA揭示三阴性乳腺癌相关上皮细胞中的模块中枢基因

为揭示TNBC上皮细胞亚型的潜在功能，采用hdWGCNA以功率10构建无标度网络，获得11个模块（图5A）。相关性分析显示，模块3/4/6/7/8/10/11与上皮细胞关联最强，模块2与T细胞最相关，模块5、9与内皮细胞最相关，模块1与成纤维细胞最相关（图5C）。雷达图提示上述7个模块与3簇上皮细胞相关性均≥50%（图5D），据此将3簇界定为TNBC关键上皮细胞，并以这些模块基因作为其中心基因。随后对各模块前25个中心基因绘制PPI网络（图5E）。模块间相关性显示，模块5与1显著正相关，模块7与2显著负相关（图6A）。

2.5　机器学习算法揭示三阴性乳腺癌相关上皮细胞的特征基因

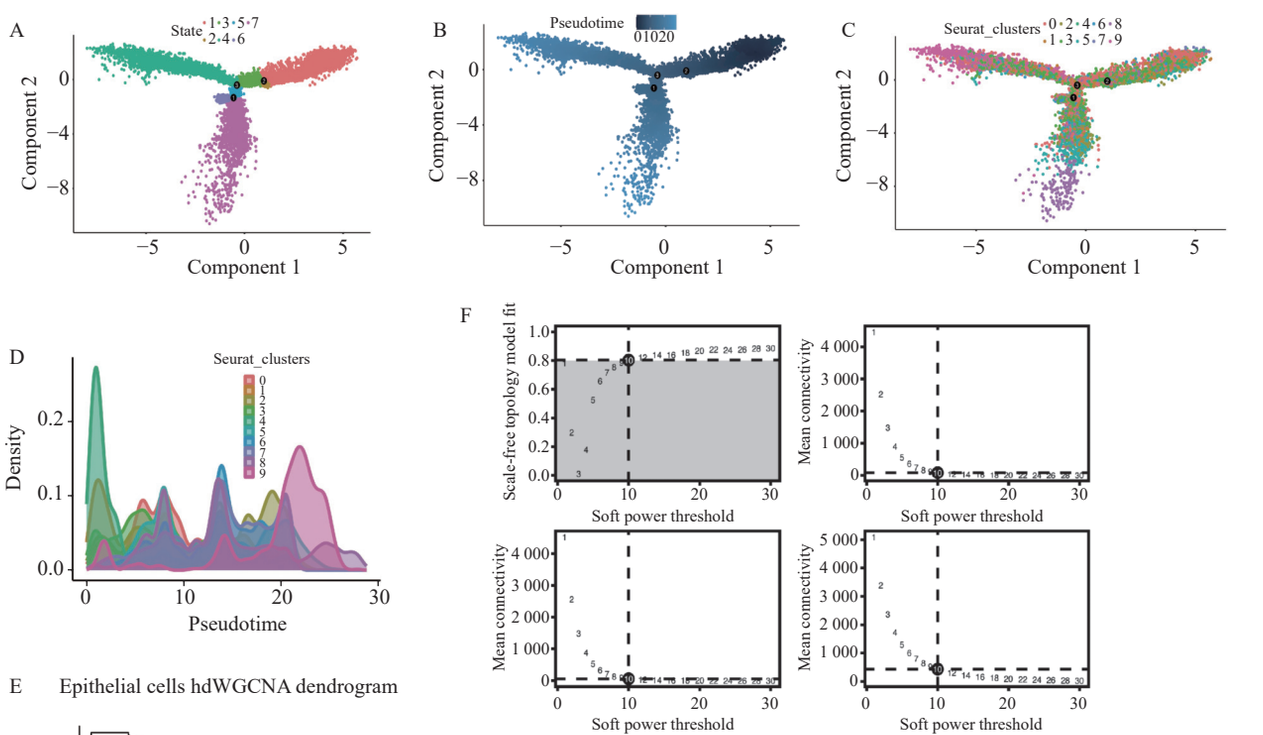
为评估3簇上皮细胞特征基因的预后价值，我们将其与TNBC的差异基因求交建模。差异分析获7 423个基因（图6B）；与3簇特征基因取交得987个（图6C）。整合TCGA与GSE58812共224例，对交集基因做单因素Cox，得到69个预后相关基因，其中前20个预后相关基因见图6E。随后以随机森林筛前30个（图6F），SVM-RFE（五折交叉验证）筛前30个（图6H、6I），二者交集得18个关键基因（图7A）。

以这18个基因为特征，LASSO缩减至14个（图7B~D），再经多因素Cox构建10基因预后模型（图7E）。时间依赖ROC显示模型稳健：总体AUC为0.839（图7F）；在TCGA的1/3/5年AUC分别为0.792、0.843、0.875（图7G，P<0.05）。基于surv_cutpoint函数确定最佳阈值（图7H），K-M曲线表明低风险组生存显著更优（图7J，P<0.01）。

2.6 不同组间存在不同的免疫微环境

三阴性乳腺癌患者的免疫细胞浸润情况分析显示，低风险组的ESTIMATE、Stromal、Immune评分更高，而高风险组肿瘤纯度更高（图8A，均P<0.05）。两组在7类免疫细胞上显著不同（Bcellsnaive、TcellsCD4memoryactivated、Tfollicularhelper、Monocytes、MacrophagesM0/M1/M2，图8D）。低风险组的MHC_IPS、EC_IPS、AZ_IPS及总体IPS更高，提示抗原呈递与免疫效应更强；高风险组的CP_IPS更高，提示对免疫检查点抑制可能更敏感（图8E）。高风险组多条糖代谢通路活性更高（己糖胺、果糖/甘露糖、淀粉/蔗糖等，图8F），并呈更强免疫逃逸趋势：8个检查点基因下调（CD244、CTLA4、ICOS、IDO1、LAG3、PDCD1LG2、TNFRSF14、TNFRSF18，图8G）表明高危组肿瘤免疫逃避的可能性更高，免疫微环境抑制更强。相关性分析结果显示RPN1与M0型巨噬细胞呈正相关，与未激活的B细胞呈负相关，且RPN1与M0型巨噬细胞的相关性最强（图8J）。

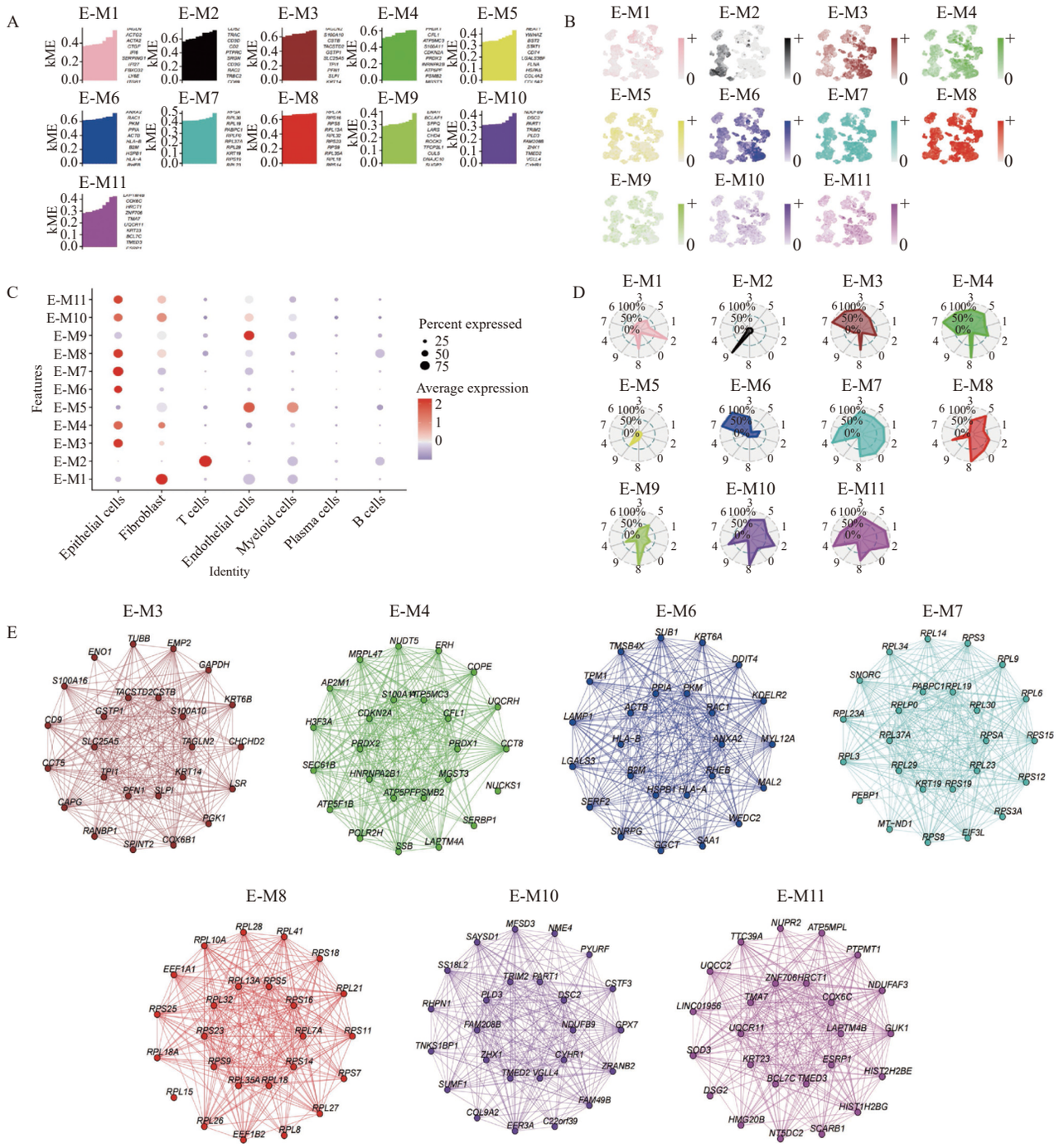
我们利用oncoPredict软件包将整合的两组患者对不同药物的敏感性进行分析，发现了10种药物的半数抑制浓度（IC_{50}）在两组间不同（图8K），其中XAV939在低风险组的患者中更敏感，而KU-55933、NU7441、RO-3306、MK-2206、GSK269-962A等药物则在高风险组患者中更敏感。



A–D: pseudotime analysis delineating epithelial differentiation in TNBC, highlighting branch points and lineage directions; E: hdWGCNA dendrogram with dynamic tree cut; colored modules are retained as biologically meaningful, whereas the grey module contains genes not assigned to any module; F: scale-free topology analysis showing the fit index and mean connectivity across soft-thresholding powers (the optimal power is 10).

图 4 TNBC上皮细胞拟时序轨迹与hdWGCNA模块筛选

Figure 4 Pseudotime trajectories and hdWGCNA-based module selection in TNBC epithelium



A: kME values (module membership) for the 11 gene modules; the top 10 genes in each module are listed by kME rank; B: UMAP visualization showing module eigengene expression of the 11 modules across annotated cell types; C: hdWGCNA-based estimates of module activity across seven cell clusters; D: radar chart depicting correlations between the 11 gene modules and 10 epithelial subtypes; E: protein–protein interaction (PPI) networks for the top 25 hub genes of seven representative modules.

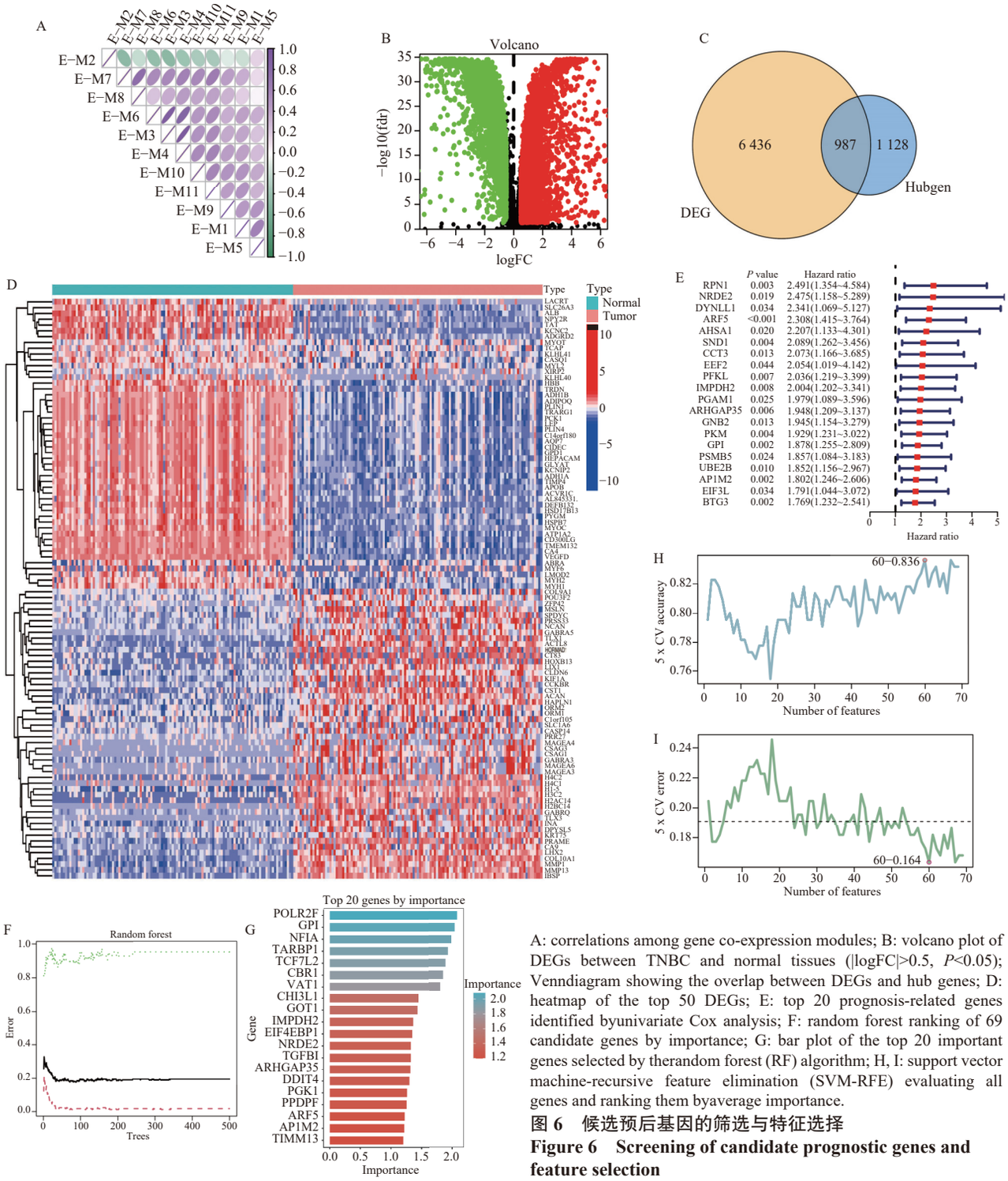
图 5　基于hdWGCNA的模块活性、细胞类型特异性与枢纽基因网络

Figure 5　hdWGCNA-derived module activity, cell-type specificity, and hub-gene networks

2.7　预后基因在三阴性乳腺癌和正常样本中有不同的表达

在220例TNBC样本中，热图显示TCF7L2、CBR1、CHI3L1、KRT15、PDAP1、ISCU在低风险组更高，而IMPDH2、DDIT4、BTG3、RPN1在高风险组更高（图9A）。染色体定位：IMPDH2、RPN1位于2号；TCF7L2、DDIT4位于10号；CBR1、BTG3位于21号；CHI3L1、PDAP1、ISCU、KRT15分别位于1、7、12、17号（图9B）。单细胞数据表明大量上皮细胞表达上述10个基因（图9C、9D），并呈拟时序动态（图9E）。

在TCGA中，10个基因在肿瘤与正常组间均存



A: correlations among gene co-expression modules; B: volcano plot of DEGs between TNBC and normal tissues (|logFC|>0.5, P<0.05); Venndiagram showing the overlap between DEGs and hub genes; D: heatmap of the top 50 DEGs; E: top 20 prognosis-related genes identified byunivariate Cox analysis; F: random forest ranking of 69 candidate genes by importance; G: bar plot of the top 20 important genes selected by therandom forest (RF) algorithm; H, I: support vector machine-recursive feature elimination (SVM-RFE) evaluating all genes and ranking them byaverage importance.

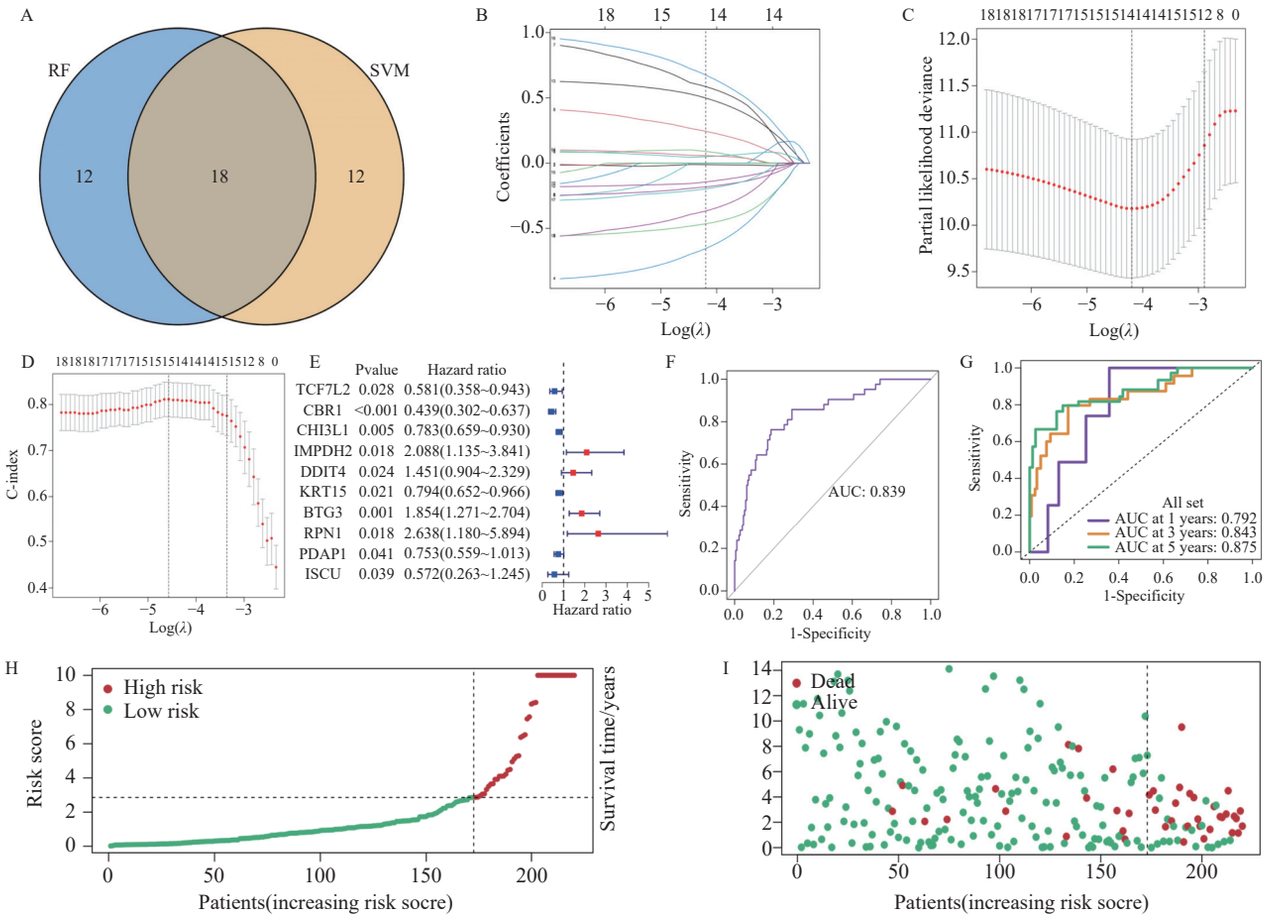
图 6 候选预后基因的筛选与特征选择

Figure 6 Screening of candidate prognostic genes and feature selection

在差异：CBR1、CHI3L1、DDIT4、BTG3、RPN1、PDAP1上调，TCF7L2、IMPDH2、KRT15、ISCU下调（图9F）。对上调的6个基因做K-M分析，DDIT4、BTG3、RPN1高表达预后更差，CBR1、CHI3L1、PDAP1高表达预后更好（图10A）；因CHI3L1、PDAP1生存差异不显著，后续聚焦BTG3、RPN1、DDIT4、CBR1。HPA显示这4个基因在乳腺癌组织染色更深（图10B）；单基因预测AUC分别为0.632（BTG3）、0.643（RPN1）、0.618（DDIT4）、0.608（CBR1）（图10C）。qRT-PCR验证发现：MDA-MB-231/468中CBR1、CHI3L1、DDIT4、BTG3、RPN1、PDAP1更高；MCF-10A中TCF7L2、IMPDH2、KRT15、ISCU更高（图10D）。

3 讨论

三阴性乳腺癌对女性健康危害显著，尽管近年来TNBC在系统治疗及预后分层方面取得了一定进展，但不同患者间治疗反应和临床转归仍存在明显



A: Venn diagram showing the intersection of genes selected by random forest (RF) and SVM–RFE; B–D: LASSO regression applied to the RF/SVM–RFE intersection, ultimately identifying 14 key genes; E: multivariable Cox proportional hazards analysis to screen independent prognostic genes and construct the risk-score model; F: ROC curves evaluating the accuracy of the model's risk score for survival prediction; G: time-dependent ROC (tdROC) curves demonstrating predictive performance in the integrated TCGA + GSE26193 dataset (P<0.05); H: risk-score distribution with the optimal cutoff; I: survival status plot at the optimal cutoff; J: Kaplan–Meier survival curves comparing outcomes between the two risk groups in 220 patients with TNBC.

图 7　整合机器学习与生存分析构建TNBC预后模型

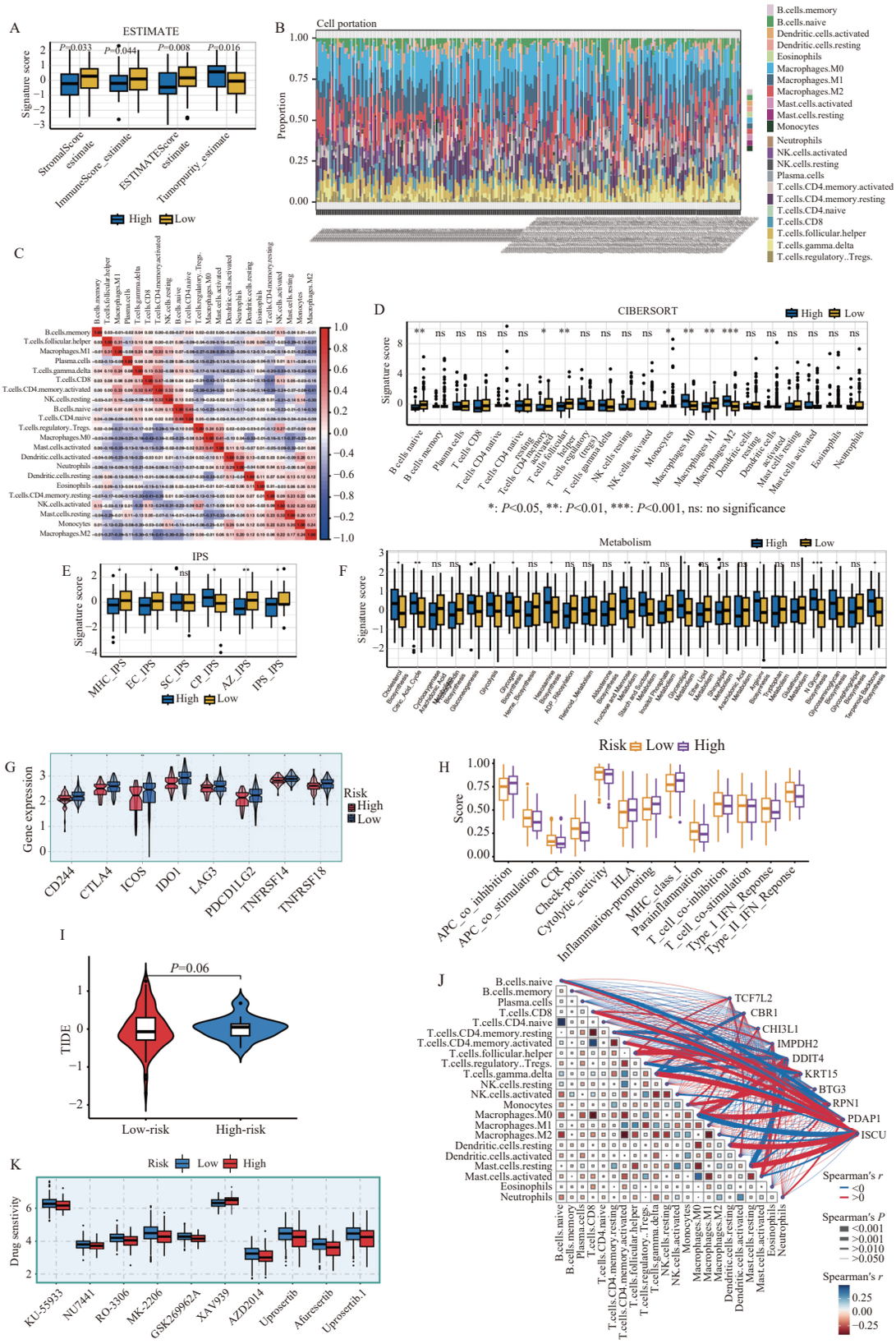
Figure 7　Building a TNBC prognostic model by integrated machine learning and survival analysis

分化[39-40]；与此同时，空间与分子层面的显著异质性进一步提示，仍需持续发掘更稳定、可转化的预后标志物和潜在治疗靶点[41]。本研究发现TNBC肿瘤组织的细胞组成较正常组织显著改变，在上皮细胞精细分型中，鉴定到3簇在TNBC中富集。细胞通讯与拟时序结果表明3簇具独特发育轨迹并与多细胞类型强交互。hdWGCNA显示其共表达模块与整体上皮细胞相近，故将3簇界定为TNBC相关上皮细胞，并据此构建预后模型，为TNBC风险分层及潜在靶点挖掘提供依据。

在0簇上皮细胞中，GSEA富集分析提示阿尔茨海默病、氧化磷酸化、核糖体等通路上调，指示高应激/增殖/代谢状态[42-43]；而趋化因子、细胞因子–受体、ECM–受体、焦点黏附、NOD样受体等下调，多与免疫逃逸、转移潜能增强和微环境重塑相关，例如趋化因子信号下调可削弱免疫细胞募集[44]、焦点黏附下调可降低黏附并增强迁移[45]。

细胞互作方面，TNBC相关上皮细胞与内皮、髓系及T细胞通讯更强，关键配体–受体轴包括MIF–（CD74+CXCR4/CD44）、CD99–CD99、APP–CD74、HLA-C–KIR2DL3等。MIF为多功能细胞因子，通过CD74/CD44及CXCR2/4/7等受体和胞内互作，促增殖、迁移、生存及代谢重编程[46-47]；CD99在多种肿瘤（如Ewing肉瘤）与转移相关，CD99–CD99轴参与免疫调节与迁移[48]。这些通路可能驱动TNBC进展与免疫调控。

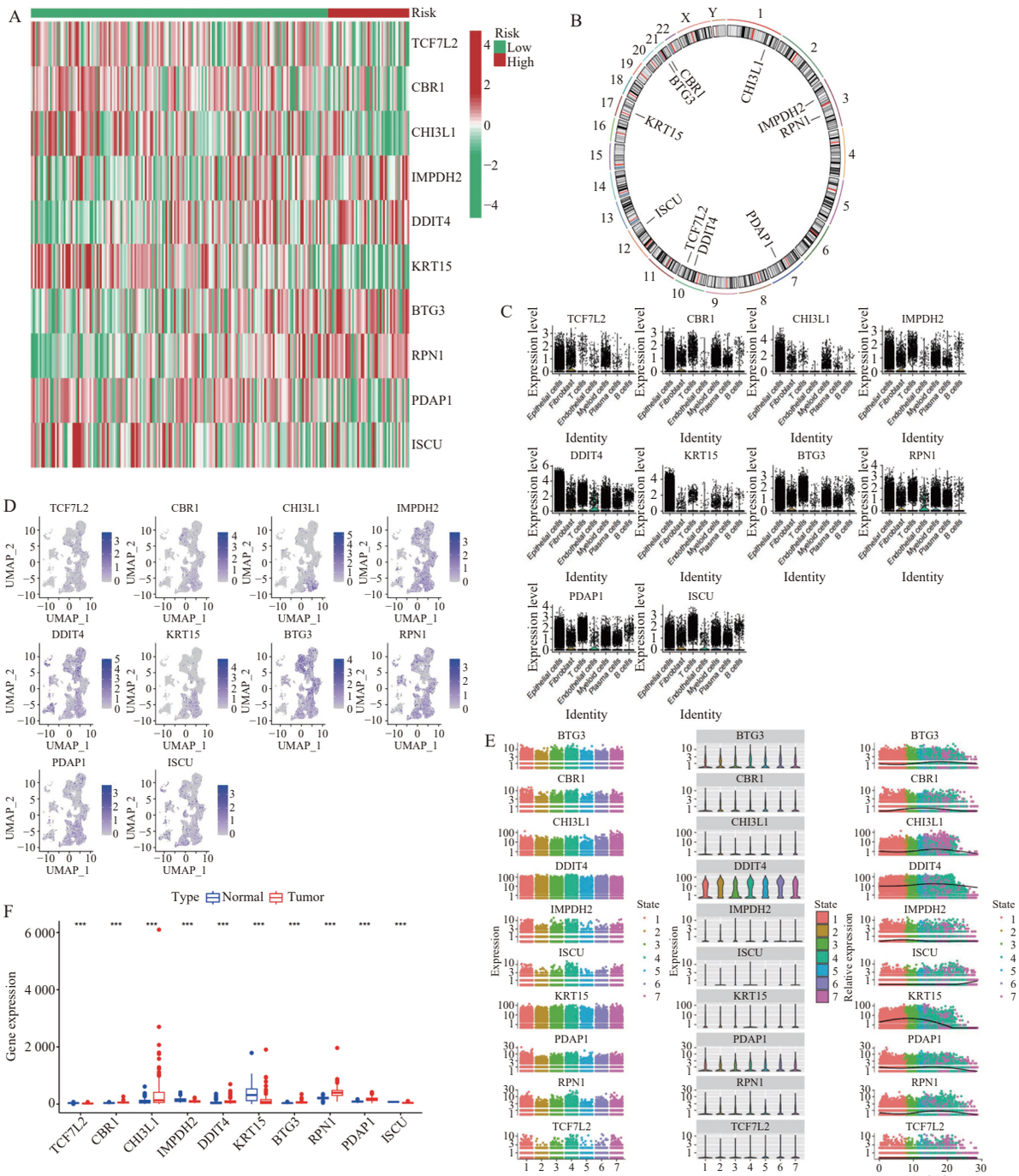
本研究通过hdWGCNA构建并分析三阴性乳腺癌单细胞数据中的共表达网络，获得三阴性乳腺癌



A: ESTIMATE-derived stromal, immune, and composite scores in the high- and low-risk groups; B: stacked bar plot of immune-cell composition across 220 patients with TNBC; C: correlation matrix among inferred immune-cell fractions; D: boxplots showing CIBERSORT-inferred differences in immune-cell infiltration between risk groups; E: comparison of immunophenoscore (IPS) between the two groups; F: comparison of metabolic pathway activity scores between the two groups; G: expression levels of immune checkpoint genes in the two groups; H: ssGSEA-based differences in immune functions between the two groups; I: TIDE scores in the high- versus low-risk groups; J: butterfly plot depicting correlations between the 10 prognostic genes and immune cell fractions; K: predicted drug sensitivity (IC_{50}) differences between the two groups.

图 8　TNBC风险分层下的免疫微环境与治疗反应图谱

Figure 8　Immune microenvironment and therapy–response landscape across risk strata in TNBC

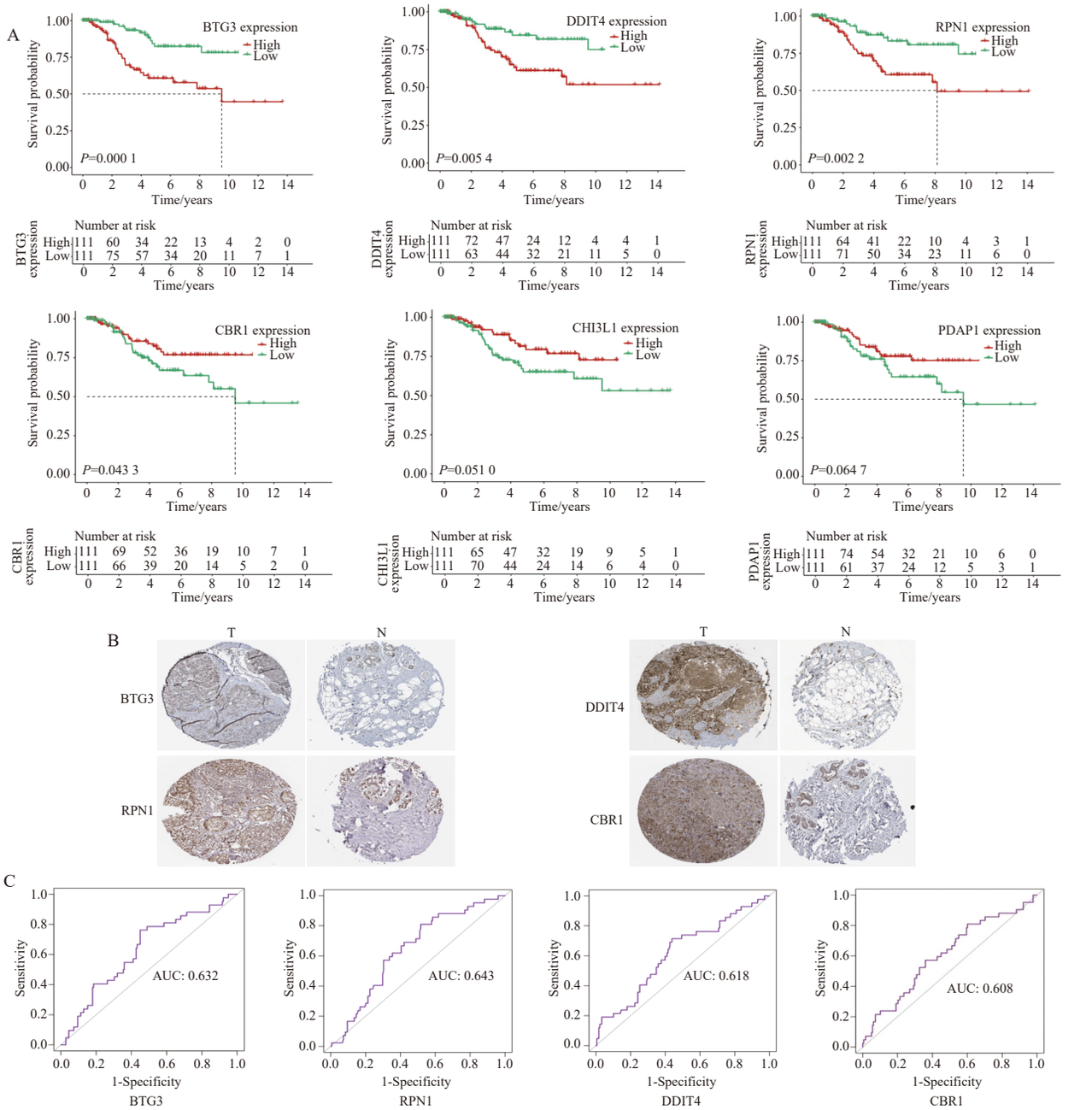


A: heatmap showing expression levels of the 10 prognostic genes; B: chromosomal localization of the prognostic genes; C, D: violin plots and UMAP projections depicting expression patterns of the prognostic genes across cells from TNBC tissues; E: dynamic expression of the 10 prognostic genes along the inferred pseudotime trajectory; F: differential expression analysis of the prognostic genes between the indicated groups. ***: P<0.001.

图 9　预后基因的表达分析

Figure 9　Expression analysis of prognostic genes

相关性上皮细胞的共表达基因模块。使用多种机器学习算法结合LASSO回归分析和多因素COX回归来提高模型的可靠性，最终筛选CBR1、CHI3L1等10个关键基因。随后采取ROC曲线对该模型进行验证，确定了其可靠性。我们根据模型的高低评分进行分组，系统性的分析了不同组间存在的免疫微环境及药敏差异，尽管本研究在图8G展示的8个检查点面板未包含PD-L1，但在TNBC的治疗决策中，PD-L1仍是最具实践意义的免疫标志物之一：在转移性TNBC中，KEYNOTE-355显示CPS≥10（22C3抗体）的人群从帕博利珠单抗联合化疗中获益显著，已成为一线治疗的重要依据；而在早期TNBC，KEYNOTE-522证实围手术期帕博利珠单抗带来的获益不依赖PD-L1状态，提示其在不同疾病阶段的



A: Kaplan-Meier survival analyses for six prognostic genes; B: immuno-his-to-chemistry (IHC) of four genes in representative tissues (T: tumor; N: normal tissues); C: receiver operating characteristic (ROC) curves depicting the predictive accuracy of each gene, individually, for survival in patientswith TNBC; D: qRT-PCR showing differential expression of the 10 prognostic genes in TNBC versus normal breast cell lines.

图 10 TNBC预后基因的多维度评估：Kaplan-Meier生存、IHC、ROC与qRT-PCR

Figure 10 Multimodal assessment of TNBC prognostic genes: Kaplan-Meier survival, IHC, ROC, and qRT-PCR

指征差异[49]。结合我们的结果，高风险组呈现糖代谢重编程增强与免疫抑制/逃逸倾向。需要强调的是，PD-L1的蛋白稳态受糖基化等翻译后修饰强烈调控：十六碳胺途径（HBP）生成的UDP-GlcNAc供体可促进PD-L1 N-糖基化并稳定其膜表达，这意味着即使CD274转录本差异不显著，蛋白层面的“功能性PD-L1”仍可能上调，与我们观察到的糖代谢增强具有生物学一致性[50]。

经验证发现，模型基因在正常组与肿瘤组中的表达确实存在差异。研究选取了在肿瘤组织中表达增加的6个基因继续分析，K-M曲线生存分析发现，DDIT4、BTG3、RPN1基因高表达时与预后差相关，而CBR1、CHI3L1、PDAP1低表达时与预后差相关，但CHI3L1与PDAP1基因的K-M曲线*P*值大

于0.05，差异并无统计学意义。利用HPA在线数据库分析发现，DDIT4、BTG3、RPN1以及CBR1基因在肿瘤组织的蛋白表达更高。DNA损伤可导转录物4（DDIT4）是一种肿瘤相关蛋白，也是代谢和免疫相关蛋白，在化疗、缺氧和DNA损伤等压力条件下高度表达。研究表明，DDIT4与卵巢癌[51]、肺癌[52]、膀胱上皮癌[53]等多种癌症的进展有关。Foltyn等研究发现DDIT4通过抑制mTORC1促进胶质母细胞的生存和抗肿瘤治疗抗性[54]。Jiang等发现lncRNA DDIT4-AS1介导的自噬激活促进了TNBC的进展和化学抵抗，DDIT4-AS1的靶向可以作为提高TNBC化疗疗效的新方法[55]。B细胞易位基因3（BTG3）是BTG/Tob家族中的一个成员，这个家族的基因普遍具有抗增殖作用[56]。Zhang等研究发现，TRIM65可以通过降解BTG3来调节细胞周期进而促进肾细胞癌的细胞增殖[57]。RPN1（Ribophorin Ⅰ）是一个编码膜蛋白的基因，研究表明，该基因的表达与肿瘤微环境、免疫反应因子和抗肿瘤治疗反应相关，Luo等发现，该基因可以促进三阴性乳腺癌、胶质母细胞瘤和膀胱癌的肿瘤细胞增殖和迁移[58]。大麻素受体1（CBR1）是人体内的一种重要的G蛋白偶联受体（GPCR），主要参与调节神经系统的多种功能。研究表明，CBR1参与减少动物模型中肿瘤的进展，并在体内具有抗肿瘤作用[59]。有证据表明，CBR1改变了雌激素受体（ER）和孕酮受体（PR）的迁移和入侵，是乳腺癌的潜在治疗靶点[60]。

总之，本研究为TNBC的预后评估与治疗策略提供了有价值的证据。所构建的10基因预后模型可作为临床辅助风险分层与决策支持工具：于初诊或围手术期量化复发/死亡风险，并与既有临床病理指标（分期、肿瘤大小、分级、Ki-67、BRCA 状态）及治疗反应指标（如 pCR）联合，用于讨论治疗强度的强化/去强化；在免疫治疗中，可与PD-L1（CPS）、TILs、IFN-γ相关签名、IPS等联合判读，识别“高风险且免疫活性高”的患者以优先考虑免疫联合策略，或“高风险且PD-L1低、代谢活跃”的患者以优先考虑临床试验/联合方案；在随访管理中，依据风险等级调整复诊频次与影像学评估，并将高风险患者优先纳入ctDNA/分子残留病灶（MRD）等动态监测；模型建议以FFPE兼容的qRT-PCR/定向表达面板固化实现，并固定阈值形成可解释的低/中/高三档分层。需强调，该模型用于补充而非替代现行指征与指南；在多中心外部验证并形成标准化报告与SOP后，可纳入MDT综合决策。

本研究也有一定的局限性。在确定预后相关基因作为可行的生物标志物之前，未进行体外实验进一步筛选和鉴定生物标志物，尽管我们利用qRT-PCR证实了这些预后基因在正常细胞系与TNBC细胞系之间的表达差异，未来仍需开展相关细胞与动物实验验证上述结果，并进一步深入研究上皮细胞新亚群以及这些预后基因在疾病进展中的调控机制。另外，三阴性乳腺癌的发病机制非常复杂，目前的研究只集中在上皮细胞的分析上，未来可开展工作对其他细胞类型进行深入研究。

总的来说，本研究为三阴性乳腺癌的发病机制提供了新的见解，也为预后评估和治疗策略提供了有价值的科学证据。

利益冲突声明：

所有作者均声明不存在利益冲突。

参考文献：

[1] Garrido-Castro AC, Lin NU, Polyak K. Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment[J]. Cancer Discov, 2019, 9(2): 176-198.

[2] Bai X, Ni J, Beretov J, *et al*. Triple-negative breast cancer therapeutic resistance: where is the Achilles' heel?[J]. Cancer Lett, 2021, 497: 100-111.

[3] Loibl S, André F, Bachelot T, *et al*. Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up[J]. Ann Oncol, 2024, 35(2): 159-182.

[4] Bao B, Prasad AS. Targeting CSC in a Most Aggressive Subtype of Breast Cancer TNBC[J]. Adv Exp Med Biol, 2019, 1152: 311-334.

[5] So JY, Ohm J, Lipkowitz S, *et al*. Triple negative breast cancer (TNBC): non-genetic tumor heterogeneity and immune microenvironment: emerging treatment options[J]. Pharmacol Ther, 2022, 237: 108253.

[6] Hua Z, White J, Zhou J. Cancer stem cells in TNBC[J]. Semin Cancer Biol, 2022, 82: 26-34.

[7] Spring LM, Fell G, Arfe A, *et al*. Pathologic complete response after neoadjuvant chemotherapy and impact on breast cancer recurrence and survival: a comprehensive meta-analysis[J]. Clin Cancer Res, 2020, 26(12): 2838-2848.

[8] Zhu H, Wei M, Xu J, *et al*. PARP inhibitors in pancreatic cancer: molecular mechanisms and clinical applications[J]. Mol Cancer, 2020, 19(1): 49.

[9] Morganti S, Marra A, De Angelis C, *et al*. PARP Inhibitors for Breast Cancer Treatment: A Review[J]. JAMA Oncol, 2024, 10(5): 658-670.

[10] Ye Z, Yuan J, Hong D, *et al*. Multimodal diagnostic models and subtype analysis for neoadjuvant therapy in breast cancer[J]. Front Immunol, 2025, 16: 1559200.

[11] Wang H, Zhu YN, Zhang S, *et al*. Transcriptome-wide analysis

reveals potential roles of CFD and ANGPTL4 in fibroblasts regulating B cell lineage for extracellular matrix-driven clustering and novel avenues for immunotherapy in breast cancer[J]. Mol Med, 2025, 31(1): 179.

[12] Zhong L, Zhu J, Chen J, *et al*. MGAT4EP promotes tumor progression and serves as a prognostic marker for breast cancer[J]. Cancer Biol Ther, 2025, 26(1): 2475604.

[13] Zou Y, Xie J, Zheng S, *et al*. Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery[J]. Int J Surg, 2022, 107: 106936.

[14] Chen C, Qi J, Lin W, *et al*. Single-cell ligand-receptor profiling reveals an immunotherapy-responsive subtype and prognostic signature in triple-negative breast cancer[J]. Front Immunol, 2025, 16: 1590951.

[15] Stuart T, Butler A, Hoffman P, *et al*. Comprehensive integration of single-cell data[J]. Cell, 2019, 177(7): 1888-1902. e21.

[16] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression[J]. Genome Biol, 2019, 20(1): 296.

[17] Tan Z, Chen X, Zuo J, *et al*. Comprehensive analysis of scRNA-Seq and bulk RNA-Seq reveals dynamic changes in the tumor immune microenvironment of bladder cancer and establishes a prognostic model[J]. J Transl Med, 2023, 21(1): 223.

[18] Jin S, Guerrero-Juarez CF, Zhang L, *et al*. Inference and analysis of cell-cell communication using CellChat[J]. Nat Commun, 2021, 12(1): 1088.

[19] Yu G, Wang LG, Han Y, *et al*. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. OMICS, 2012, 16(5): 284-287.

[20] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data[J]. BMC Bioinformatics, 2013, 14: 7.

[21] Trapnell C, Cacchiarelli D, Grimsby J, *et al*. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells[J]. Nat Biotechnol, 2014, 32(4): 381-386.

[22] Sun Y, Wu J, Zhang Q, *et al*. Single-cell hdWGCNA reveals metastatic protective macrophages and development of deep learning model in uveal melanoma[J]. J Transl Med, 2024, 22(1): 695.

[23] Wang X, Jin Y, Xu L, *et al*. Integrating Single-Cell RNA-Seq and Bulk RNA-Seq to Construct a Novel γδT Cell-Related Prognostic Signature for Human Papillomavirus-Infected Cervical Cancer[J]. Cancer Control, 2024, 31: 10732748241274228.

[24] Morabito S, Reese F, Rahimzadeh N, *et al*. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data[J]. Cell Rep Methods, 2023, 3(6): 100498.

[25] Tian Y, Xiao H, Yang Y, *et al*. Crosstalk between 5-methylcytosine and N6-methyladenosine machinery defines disease progression, therapeutic response and pharmacogenomic landscape in hepatocellular carcinoma[J]. Mol Cancer, 2023, 22(1): 5.

[26] Jia A, Xu L, Wang Y. Venn diagrams in bioinformatics[J]. Brief Bioinform, 2021, 22(5): bbab108.

[27] Ishwaran H, Kogalur UB. Consistency of random survival forests[J]. Stat Probab Lett, 2010, 80(13-14): 1056-1064.

[28] Sanz H, Valim C, Vegas E, *et al*. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels[J]. BMC Bioinformatics, 2018, 19(1): 432.

[29] Dai P, Chang W, Xin Z, *et al*. Retrospective study on the influencing factors and prediction of hospitalization expenses for chronic renal failure in China based on random forest and LASSO regression[J]. Front Public Health, 2021, 9: 678276.

[30] Shen C, Chai W, Han J, *et al*. Identification and validation of a dysregulated TME-related gene signature for predicting prognosis, and immunological properties in bladder cancer[J]. Front Immunol, 2023, 14: 1213947.

[31] Lätti S, Niinivehmas S, Pentikäinen OT. Rocker: open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization[J]. J Cheminform, 2016, 8(1): 45.

[32] Wang Y, Xu J, Fang Y, *et al*. Comprehensive analysis of a novel signature incorporating lipid metabolism and immune-related genes for assessing prognosis and immune landscape in lung adenocarcinoma[J]. Front Immunol, 2022, 13: 950001.

[33] Zhao Z, Ding Y, Tran LJ, *et al*. Innovative breakthroughs facilitated by single-cell multi-omics: manipulating natural killer cell functionality correlates with a novel subcategory of melanoma cells[J]. Front Immunol, 2023, 14: 1196892.

[34] Gu X, Chen B, Zhang S, *et al*. The expression of CCL17 and potential prognostic value on tumor immunity in thyroid carcinoma based on bioinformatics analysis[J]. Sci Rep, 2024, 14(1): 31580.

[35] Zhao P, Zhen H, Zhao H, *et al*. Identification of hub genes and potential molecular mechanisms related to radiotherapy sensitivity in rectal cancer based on multiple datasets[J]. J Transl Med, 2023, 21(1): 176.

[36] Zhang X, Wu D, Zhang L, *et al*. Predicting the potential mechanism of radix chimonanthi pracecocis in treating osteoarthritis by network pharmacology analysis combined with experimental validation[J]. J Ethnopharmacol, 2024, 331: 118231.

[37] Karran JC, Moodie EE, Wallace MP. Statistical method use in public health research[J]. Scand J Public Health, 2015, 43(7): 776-782.

[38] Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics[J]. Nat Protoc, 2025, 20(1): 180-219.

[39] Schmid P, Cortes J, Dent R, *et al*. Overall survival with pembrolizumab in early-stage triple-negative breast cancer[J]. N Engl J Med, 2024, 391(21): 1981-1991.

[40] Leon-Ferre RA, Jonas SF, Salgado R, *et al*. Tumor-infiltrating lymphocytes in triple-negative breast cancer[J]. JAMA, 2024, 331(13): 1135-1144.

[41] Wang X, Venet D, Lifrange F, *et al*. Spatial transcriptomics reveals substantial heterog eneity in triple-negative breast cancer with potential clinical implications[J]. Nat Commun, 2024, 15(1): 10232.

[42] Li C, Ji H, Zhuang S, *et al*. Update on the correlation between

mitochondrial function and osteonecrosis of the femoral head osteocytes[J]. Redox Rep, 2025, 30(1): 2491846.

[43] Chaudhary S, Siddiqui JA, Pothuraju R, *et al*. Ribosome biogenesis, altered metabolism and ribotoxic stress response in pancreatic ductal adenocarcinoma tumor microenvironment[J]. Cancer Lett, 2025, 612: 217484.

[44] Lai Y, Zhong Z, Li R, *et al*. Unveiling the crucial role of CD8+ T cell and endothelial cell interaction in rheumatoid arthritis through the integrated analysis of spatial transcriptomics and bulk/single-cell RNA-seq[J]. J Transl Autoimmun, 2025, 10: 100285.

[45] Kamal KY, Trombetta-Lima M. Mechanotransduction and skeletal muscle atrophy: the interplay between focal adhesions and oxidative stress[J]. Int J Mol Sci, 2025, 26(6): 2802.

[46] Sumaiya K, Langford D, Natarajaseenivasan K, *et al*. Macrophage migration inhibitory factor (MIF): a multifaceted cytokine regulated by genetic and physiological strategies[J]. Pharmacol Ther, 2022, 233: 108024.

[47] Jankauskas SS, Wong DWL, Bucala R, *et al*. Evolving complexity of MIF signaling[J]. Cell Signal, 2019, 57: 76-88.

[48] Grünewald TGP, Cidre-Aranaz F, Surdez D, *et al*. Ewing sarcoma[J]. Nat Rev Dis Primers, 2018, 4(1): 5.

[49] Cortes J, Rugo HS, Cescon DW, *et al*. Pembrolizumab plus chemotherapy in advanced triple-negative breast cancer[J]. N Engl J Med, 2022, 387(3): 217-226.

[50] Ruegenberg S, Horn M, Pichlo C, *et al*. Loss of GFAT-1 feedback regulation activates the hexosamine pathway that modulates protein homeostasis[J]. Nat Commun, 2020, 11(1): 687.

[51] Chang B, Liu G, Yang G, *et al*. REDD1 is required for RAS-mediated transformation of human ovarian epithelial cells[J]. Cell Cycle, 2009, 8(5): 780-786.

[52] Jin HO, Hong SE, Kim JY, *et al*. Induction of HSP27 and HSP70 by constitutive overexpression of Redd1 confers resistance of lung cancer cells to ionizing radiation[J]. Oncol Rep, 2019, 41(5): 3119-3126.

[53] Zeng Q, Liu J, Cao P, *et al*. Inhibition of REDD1 sensitizes bladder urothelial carcinoma to paclitaxel by inhibiting autophagy[J]. Clin Cancer Res, 2018, 24(2): 445-459.

[54] Foltyn M, Luger AL, Lorenz NI, *et al*. The physiological mTOR complex 1 inhibitor DDIT4 mediates therapy resistance in glioblastoma[J]. Br J Cancer, 2019, 120(5): 481-487.

[55] Jiang T, Zhu J, Jiang S, *et al*. Targeting lncRNA DDIT4-AS1 sensitizes triple negative breast cancer to chemotherapy via suppressing of autophagy[J]. Adv Sci (Weinh), 2023, 10(17): e2207257.

[56] Zheng HC, Xue H, Zhang CY, *et al*. The clinicopathological significances and related signal pathways of BTG3 mRNA expression in cancers: a bioinformatics analysis[J]. Front Genet, 2022, 13: 1006582.

[57] Zhang Q, Li Y, Zhu Q, *et al*. TRIM65 promotes renal cell carcinoma through ubiquitination and degradation of BTG3[J]. Cell Death Dis, 2024, 15(5): 355.

[58] Luo P, Li Z, He H, *et al*. Exploring the impact of RPN1 on tumorigenesis and immune response in cancer[J]. FASEB J, 2025, 39(6): e70345.

[59] Morin-Buote J, Ennour-Idrissi K, Poirier É, *et al*. Association of breast tumour expression of cannabinoid receptors CBR1 and CBR2 with prognostic factors and survival in breast cancer patients[J]. J Pers Med, 2021, 11(9): 852.

[60] Caffarel MM, Andradas C, Pérez-Gómez E, *et al*. Cannabinoids: a new hope for breast cancer therapy?[J]. Cancer Treat Rev, 2012, 38(7): 911-918.

[**编辑**：邱颖慧；**校对**：尤婷婷]

作者贡献：

吴金鹏：论文撰写，数据处理与分析，图表绘制

郭　雪：完成部分实验及相关数据整理

刘恩谷：完成部分实验

林　峰：参与实验与结果分析

李鸿涛：研究设计、指导与修改，实验监督与数据分析