

基于单细胞 ATAC 测序数据的乳腺癌上皮细胞转录因子研究

张雅娜, 牟苑菁, 李苇祥, 赵俊焱, 郭志云*

(西南交通大学 生命科学与工程学院, 中国四川 成都 610031)

摘要: 研究表明, 乳腺肿瘤细胞染色质开放区域上的转录因子显著影响乳腺癌患者的临床表型和预后。然而, 在单细胞层面, 这些调控元件如何调控乳腺癌发生发展尚不明确。为此, 本研究从 GEO 数据库中下载了 45 216 个正常和肿瘤乳腺组织的单细胞染色质可及性测序(single-cell ATAC sequencing, scATAC-seq)数据, 并对其进行了分析。根据标记基因在细胞群的基因活性得分进行细胞类型注释后得到 7 种乳腺细胞类型。皮尔逊相关性分析表明, 肿瘤和正常乳腺样本的上皮细胞存在明显的染色质可及性差异, 且乳腺上皮细胞存在高度样本间异质性($PCC=-0.07$), 暗示其是乳腺肿瘤的主要恶性细胞类型。为了探究正常和恶性乳腺上皮细胞的差异可及性区域中转录因子结合基序的富集情况, 在提取上皮细胞群后对 4 个上皮细胞亚型的特征开放区域进行了转录因子结合基序富集分析。结果表明, 恶性乳腺上皮细胞有 194 个转录因子显著富集($P<0.001$), 它们可能涉及乳腺肿瘤发展和转移等调控过程。对上皮细胞亚型中富集程度较高的转录因子进行活性分数计算及转录组数据验证, 结果进一步显示这些差异调控元件可能与乳腺细胞恶性发展相关。此外, 对转录因子结合基序在不同乳腺癌亚型中的可及性差异进行了分析, 发现 *SNAI2* 在三阴性乳腺癌样本中具有显著高的可及性, 提示 *SNAI2* 在三阴性乳腺癌中具有潜在的特异性调控作用。

关键词: 乳腺癌; 单细胞; 单细胞染色质可及性测序(scATAC-seq); 转录因子

中图分类号: Q523+.1, Q811.4, R737.9

文献标志码: A

文章编号: 1007-7847(2024)02-0171-08

Analysis of Transcription Factors in Breast Cancer Epithelial Cells Based on Single-cell ATAC-seq Data

ZHANG Yana, MOU Yuanjing, LI Weixiang, ZHAO Junyan, GUO Zhiyun*

(School of Life Science and Engineering, Southwest Jiaotong University, Chengdu 610031, Sichuan, China)

Abstract: Studies have shown that transcription factors binding to open chromatin regions in breast tumor cells significantly influence the clinical phenotype and prognosis of breast cancer patients. However, it is unclear how these regulatory elements regulate the occurrence and development of breast cancer at the single-cell level. Herein, single-cell chromatin accessibility sequencing (single-cell ATAC sequencing, scATAC-seq) data of 45 216 normal and tumor breast tissues were downloaded from the GEO database, and seven breast cell types were obtained after cell type annotation based on gene activity scores of marker genes in cell populations. Pearson correlation analysis showed significant differences in chromatin accessibility between tumor and normal breast epithelial cells, and a high degree of inter-sample heterogeneity of breast epithelial cells ($PCC=-0.07$), implying that they are the main malignant cell type in breast tumors. To explore the enrichment of transcription factor motifs in the differentially accessible regions of normal and malignant mammary epithelial cells, motif enrichment analysis was performed on the characteristic open regions of the four epithelial cell subtypes after extraction of the epithelial cell populations. The results showed that malig-

收稿日期: 2023-05-21; 修回日期: 2023-07-27; 网络首发日期: 2023-10-17

基金项目: 四川省科技厅自然科学基金项目(2022NSFC0779); 传染病防治科技重大专项(2018ZX10101-003-001-008); 西南交通大学大学生创新创业训练计划(2023133); 第十七期个性化实验项目(GX202316027)

作者简介: 张雅娜(1997—), 女, 山西晋中人, 硕士研究生; *通信作者: 郭志云(1978—), 男, 山西大同人, 博士, 西南交通大学副教授, 主要从事肿瘤生物学研究, E-mail: zhiyunguo@swjtu.edu.cn

nant breast epithelial cells were significantly enriched for 194 transcription factors ($P < 0.001$), which may be involved in regulatory processes of breast tumor development and metastasis. Calculation of activity scores and validation of transcriptome data for transcription factors highly enriched in epithelial cell subtypes further showed that these differential regulatory elements may be associated with malignant development of breast cells. In addition, differences in the accessibility of transcription factor binding motifs in different breast cancer subtypes were analyzed, and *SNAI2* was found to have significantly high accessibility in triple-negative breast cancer samples, suggesting that *SNAI2* has a potential specific regulatory role in triple-negative breast cancer.

Key words: breast carcinoma; single nucleated cell; single-cell ATAC sequencing (scATAC-seq); transcription factor

(*Life Science Research*, 2024, 28(2): 171–178)

乳腺癌是严重危害全世界女性健康的第一大癌症, 其治疗目前仍然具有挑战性^[1]。染色质可及性影响转录因子与基因顺式调控元件的结合, 是基因表达调控的主要层面之一。研究表明, 乳腺肿瘤细胞染色质的开放可及区域富集关键转录因子, 这些元件调控乳腺癌发生和发展相关基因的表达, 显著影响着患者的临床表型和预后^[2-3]。通过染色质可及性测序技术(ATAC sequencing, ATAC-seq), 研究人员能够构建乳腺癌表观调控网络, 预测及识别关键基因的表达调控机制。迄今为止, 研究人员已使用 ATAC-seq 对乳腺癌瘤内表观调控状态进行了分析, 探究了乳腺癌进展等生物过程中的关键转录因子等调控元件^[4-6]。但目前针对乳腺癌的可及性研究大多是基于组织样本的, 仍缺少针对乳腺癌恶性转化和发展相关转录因子的单细胞精度研究。

为了从单细胞角度探索与乳腺癌发生发展相关的调控因子, 本研究从 GEO 数据库中下载了来自正常和肿瘤乳腺组织的单细胞染色质可及性测序(single-cell ATAC sequencing, scATAC-seq)数据, 并对其进行了分析。首先, 通过注释得到 7 种乳腺细胞类型; 相关性分析显示乳腺上皮细胞的样本间具有高度异质性, 提示乳腺上皮细胞是乳腺肿瘤恶性细胞的主要构成类型。然后, 对乳腺上皮细胞进一步注释, 获得 5 种细胞亚型; 通过亚型差异性转录因子富集分析及特征转录因子活性分析与转录组数据验证, 识别了可能与乳腺癌恶性发展相关的转录因子。最后, 对特征转录因子在不同乳腺癌亚型中的可及性差异进行分析, 结果提示 *SNAI2* 在三阴性乳腺癌(triple-negative breast cancer, TNBC)中具有潜在的特异性调控作用。综上所述, 本研究为理解乳腺癌的异质性和发病机

制提供了新的视角, 对临床乳腺癌早期筛查和治疗靶点开发的相关研究具有促进作用。

1 材料与方法

1.1 数据来源

人乳腺癌数据下载自公共基因芯片数据库 GEO (<https://www.ncbi.nlm.nih.gov/geo/>, 2022-06-10) 的 GSE198639, 检查数据质量后选择其中来自 14 名乳腺癌患者的 scATAC-seq 数据进行后续分析。人正常乳腺组织数据下载自 GSE184462, 包含来自两个供体的乳腺组织的 scATAC-seq 数据。

1.2 数据质控

使用 Cell Ranger ATAC v.2.1.0 对 26 223 个细胞的原始数据进行处理, 使用 R 包 ArchR v. 1.0.2 构建初始的单细胞特征矩阵, 并按照如下标准对数据进行严格的质量控制: 1) 转录起始位点富集分数 ≥ 8 ; 2) 每个细胞唯一的核片段 $\geq 1\ 700$; 3) 与 ENCODE (Encyclopedia of DNA Elements) 的 hg19 黑名单区域的读数比率 < 0.01 。随后, 使用 doubletCluster v.1.18.7 工具预测并去除可能产生的双细胞, 并使用 Harmony v.0.1 处理样本批次效应。

1.3 数据降维及细胞类型注释

为了减少样本初始特征矩阵的维度, 我们使用迭代 LSI 算法将稀疏矩阵的维度减少到 50, 然后采用统一流形近似与投影(uniform manifold approximation and projection, UMAP)技术实现数据二维可视化。为了对细胞群进行注释, 我们选择 ArchR v.1.0.2 的 addGeneScoreMatrix 函数计算基因的活性分数, 并从 PanglaoDB 数据库(<https://panglaodb.se/>)和相关文献中收集预期细胞类型的标记基因, 然后使用 plotEmbedding 函数将这些基因的活性得分显示在 UMAP 图中, 最后结合细

胞来源信息对细胞类型的分配进行人工检查。

1.4 细胞类型特异性开放峰的识别

使用 ArchR v.1.0.2 创建伪细胞群重复(pseudo-bulk replicates), 并对显著性分数进行标准化, 得到 500 bp 可及峰的集合。随后, 使用 getMarkerFeatures 函数进行差异可及性峰的识别, 选择 Benjamini-Hochberg 校正 P 值来确定多重检验后的差异可及性峰, $\log_2(\text{FC}) \geq 1.25$ (FC: fold change) 的峰被定义为细胞类型特异性峰。

1.5 转录因子结合基序的富集分析

使用 addMotifAnnotations 函数, 将 CIS-BP 数据库(<http://cisbp.cabr.utoronto.ca/>)中收录的转录因子结合基序(motif)注释添加到 ArchRProject 中, 设定显著富集的标准为 Benjamini-Hochberg 校正 P 值 < 0.001 。通过设置背景细胞, 我们识别了 4 个上皮细胞亚型中显著差异富集的转录因子。

1.6 特征转录因子在肿瘤和瘤旁组织中的差异表达分析

应用 GEPIA2 平台(<http://gepia2.cancer-pku.cn/>, 2022-07-20)的 limma 差异分析, 对特征转录因子在乳腺肿瘤组织和相应瘤旁组织之间的差异表达情况进行分析, $P < 0.05$ 为差异具有显著性。

1.7 转录因子的活性预测及共开放分析

为了预测基于单个细胞的、校正了 Tn5 转座酶结合偏好性的转录因子活性, 使用 R 包 chromVAR v.3.17 对特征转录因子的偏差校正标准分数(deviation z-score)进行计算。使用 Cicero v.1.18.0 计算基因组开放区域间的共可及性分数。

2 结果与讨论

2.1 乳腺上皮细胞具有高度异质性

在严格的数据质量控制后, 筛选得到可用于下游分析的 24 980 个单细胞数据。随后, 进行数据降维及聚类(图 1A), 共得到 23 个不同的细胞群。为了注释这些细胞群, 我们人工挑选了一组人乳腺组织细胞类型的标记基因^[7-8], 并以基因启动子和整个基因区域的染色质可及性作为依据对基因活性进行预测(图 1B)。根据标记基因的活性分数, 我们将这 23 个细胞群注释为 7 种细胞类型: 非恶性人乳腺上皮细胞(normal epithelial cell; C6~10: *EpCAM+*、*KRT+*)、乳腺肿瘤来源的上皮细胞(cancer epithelial cell; C1~2、C11~17: *EpCAM+*、*KRT+*、*KIT/CD44+*)、内皮细胞(endothelial cell; C3、C19: *KDR+*、*ENG+*、*PECAM1+*)、成纤维细胞(fibroblast;

C4、C18: *CXCL12+*、*THY1+*)、T 细胞(T cell; C5、C20、C22: *CD3D+*)、B 细胞(B cell; C23: *Pax5+*)和髓系细胞(myeloid cell; C21: *ITGAM+*、*ITGAX+*) (图 1A)。其中, 非恶性人乳腺上皮细胞类型存在两个明显的亚群, 可进一步注释为基质型上皮细胞(basal epithelial cell; C6~7: *KRT14+*)和管腔型上皮细胞(luminal epithelial cell; C8~10: *KRT18+*和 *KRT8+*)。

为了查看每种细胞类型的样本来源差异性, 我们统计了每个细胞群的样本构成(图 1D)。从图 1D 可知, 肿瘤来源的上皮细胞具有非常明显的样本间特异性, 该现象与肿瘤细胞调控的高度异质性有关。此外, 通过计算细胞群之间的皮尔逊相关系数(Pearson correlation coefficient, PCC), 我们系统地量化了正常和肿瘤乳腺组织中细胞类型间的相关性(图 1C)。结果显示, 乳腺上皮细胞群间的相关性相比其他细胞类型更低(PCC=-0.07), 并且肿瘤来源的上皮细胞与管腔型上皮细胞的相关性略高于基质型上皮细胞, 这说明即使同为上皮细胞, 不同亚型的细胞群在乳腺癌发展过程中的变异程度也存在差异。

2.2 乳腺上皮细胞不同亚群的差异染色质可及性分析

相关研究已经证明, 乳腺上皮细胞是乳腺癌恶性转化过程中的主要参与者, 其在乳腺组织发育及导致乳腺癌瘤内异质性中扮演重要角色^[9]。为了描述人乳腺癌样本与正常乳腺样本在表观遗传层面的异质性, 我们提取了注释为乳腺上皮细胞的 15 814 个细胞进一步聚类, 得到 17 个细胞群(图 2A)。其中, 细胞群 1~6 由正常乳腺样本来源的上皮细胞构成, 这些细胞聚集为 3 个亚群; 其余 11 个细胞群则来自肿瘤患者的样本, 并且大多是由单个患者来源的上皮细胞构成。上述结果与之前报道的单细胞 RNA 测序(single-cell RNA sequencing, scRNA-seq)数据的研究一致, 即: 癌细胞倾向于根据患者样本聚集在一起, 但非癌症来源的细胞则根据细胞类型聚集^[10-11]。随后, 我们基于标记基因^[7-8]的活性分数进行细胞类型注释, 得到 5 种上皮细胞亚型(图 2B、2C), 分别为基质型乳腺上皮细胞(basal epithelial cell; *KRT14+*、*KRT5+*、*ACTA2+*)、管腔乳腺上皮细胞 A 型(luminal-A epithelial cell; *KRT18+*、*KRT8+*、*Elf5+*、*Ltf+*、*Prom1+*)、管腔乳腺上皮细胞 B 型(luminal-B epithelial cell; *KRT18+*、*KRT8+*、*Prlr+*、*Pgr+*、*ANKRD30A+*)、癌症患者来源的恶性乳腺上皮细胞(cancer epithelial

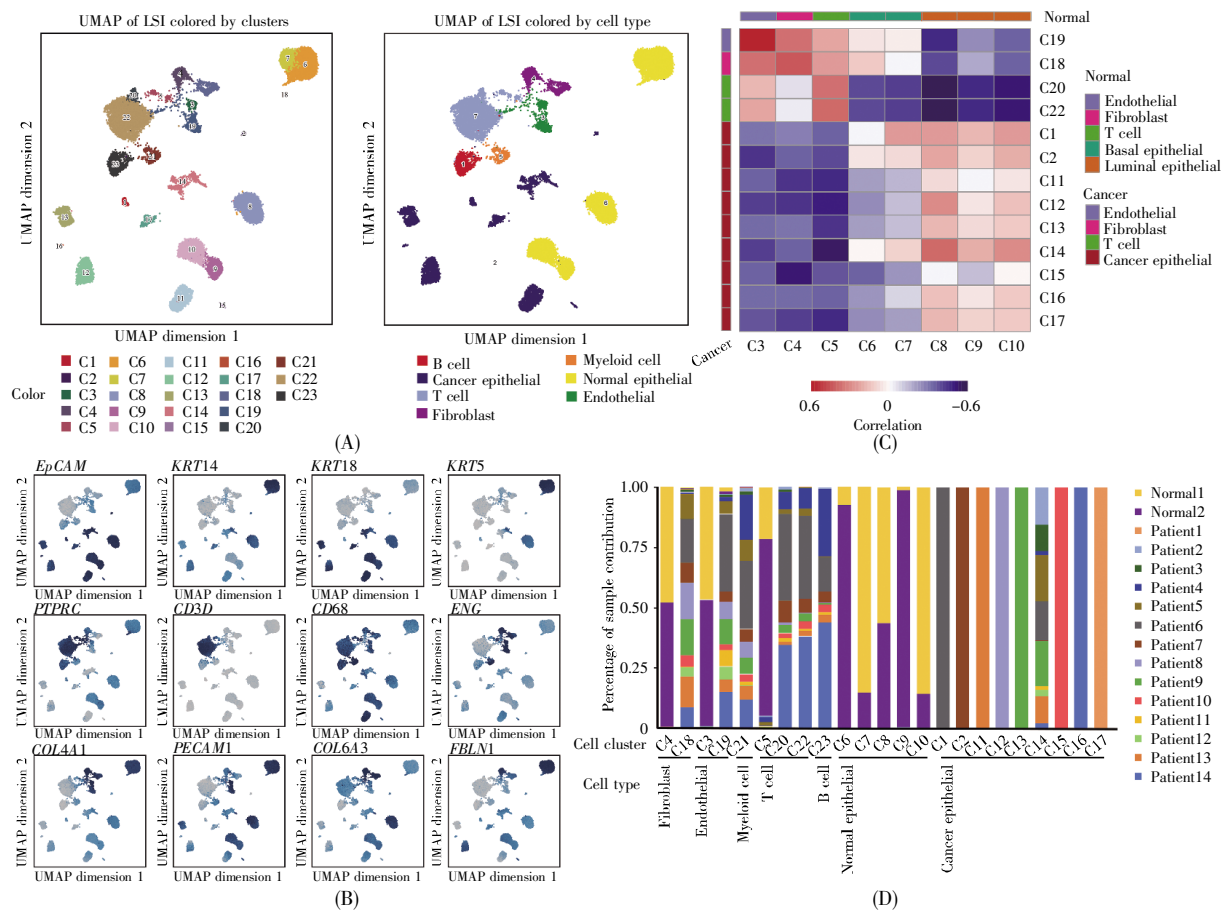


图1 人正常和肿瘤乳腺样本的单细胞染色质可及性

(A) 样本聚类结果的统一流形近似与投影图。左图和右图分别按照细胞群和注释的细胞类型着色；(B) 根据细胞类型标记基因的活性分数着色的二维投影结果图；(C) 细胞类型间的皮尔逊相关性热图。每行表示一个肿瘤样本细胞群，每列表示一个正常样本细胞群；(D) 各细胞类型的样本构成贡献。

Fig.1 Single-cell chromatin accessibility in human normal and tumor breast samples

(A) UMAP of clustering of sample cells, colored according to cell clusters (left) and cell types (right); (B) UMAP colored by z-scores of marker genes; (C) Heatmap showing Pearson correlation coefficient (PCC) between cell types. Each row represents a cluster in cancer sample, and each column represents a cluster in normal sample; (D) Distribution of cell types across human normal and tumor samples.

cell; *APOA1+*、*LEFTY1+*)以及癌症患者来源的非恶性乳腺上皮细胞(cancer-origin non-malignant epithelial cell, CN epithelial cell)。其中,管腔乳腺上皮细胞A型和B型分别对应了管腔细胞的两种亚型:能够产生乳汁的分泌性泡细胞(secretory alveolar cell, AV cell)以及能对内分泌信号做出反应的激素敏感细胞(hormone-sensing cell, HS cell)。

2.3 肿瘤和正常乳腺组织上皮细胞中转录因子的差异分析

先前研究表明,细胞间的表观遗传差异可能与转录因子的差异调控作用有关^[12-13]。为避免癌症患者来源的非恶性乳腺上皮细胞群(图2A C11~13)的畸变给分析带来干扰^[14],我们将其剔除,仅对其余4个上皮细胞亚型进行了下游分析。首先,对亚型差异可及性峰中富集的转录因子结合基序进

行识别,得到649个转录因子结合基序[Q -value < 0.05, $\log_2(FC) > 1$]。然后,使用R工具clusterProfiler v3.18对富集基序的结合转录因子参与的生物学过程进行分析(图3A)。结果显示,基序的富集与上皮细胞亚型的特征有密切联系:*ESR1*在恶性乳腺上皮细胞中富集,与基质上皮细胞发育相关的*TP63*在基质型乳腺上皮细胞中富集,而*SOX*和*FOX*家族转录因子分别在两种管腔乳腺上皮细胞中富集。这些结果也证实了我们基于基因活性评分得到的细胞类型分配结果的准确性。此外,基因本体论(Gene Ontology, GO)富集分析显示,癌症患者来源的恶性乳腺上皮细胞富集有与细胞内受体信号通路及细胞对类固醇激素刺激的反应相关的生物过程,提示其可能与乳腺癌进展中细胞响应类固醇激素(如雌激素和孕激素)有关。

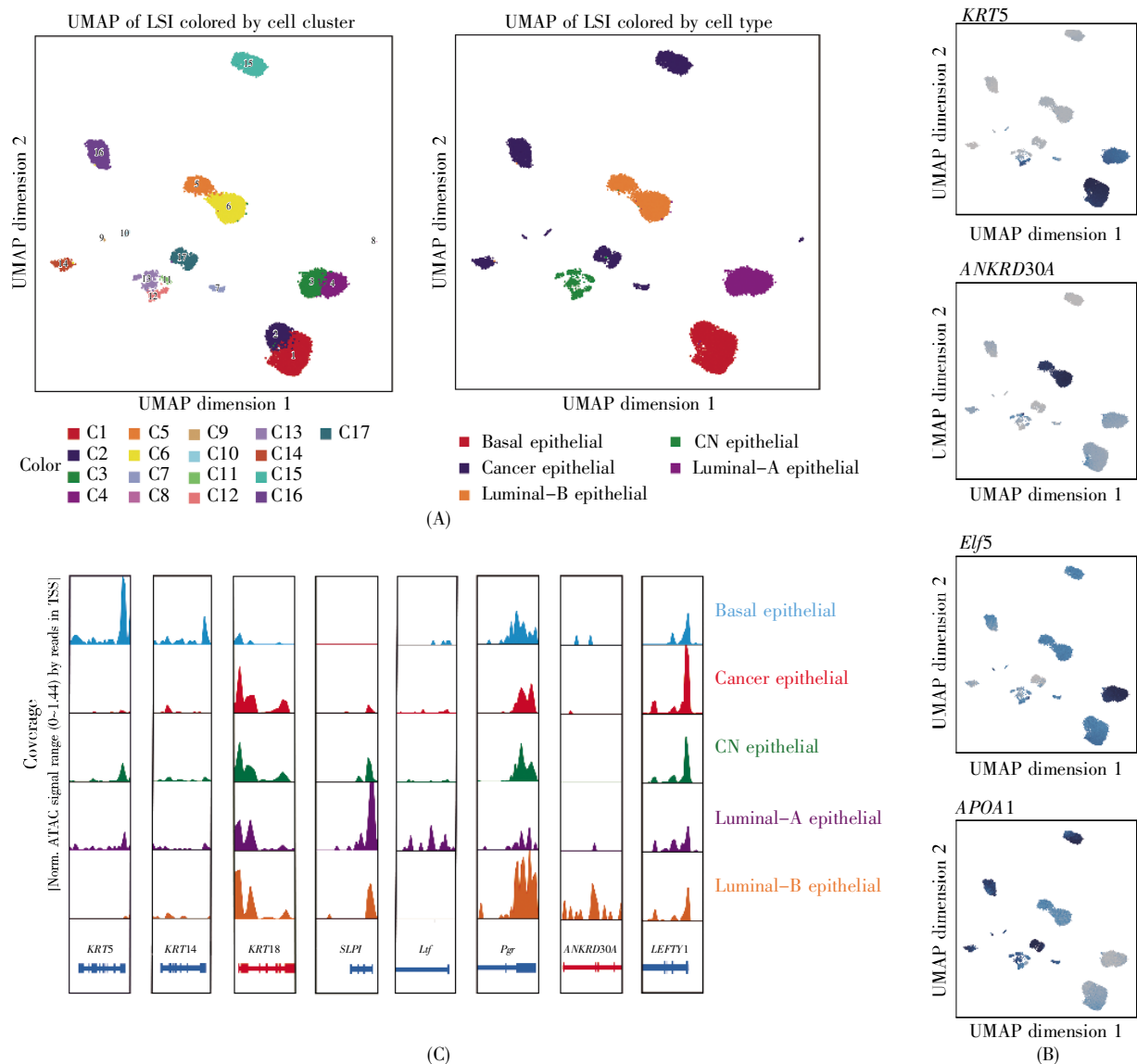


图 2 人乳腺上皮细胞的单细胞染色质可及性

(A) 样本聚类结果的统一流形近似与投影图。左图和右图分别按照细胞群和注释的细胞类型着色; (B) 根据细胞类型标记基因的活性分数着色的二维投影结果图; (C) 各上皮细胞亚型中标记基因的局部染色质开放状态轨迹图。

Fig.2 Single-cell chromatin accessibility in human mammary epithelial cells

(A) UMAP of clustering of sample cells, colored according to cell clusters (left) and cell types (right); (B) UMAP colored by z-scores of marker genes; (C) Trackplot of local chromatin accessibility at the marker genes on a per-cell type basis.

为了进一步探索在上皮细胞致癌转录调控中发挥关键作用的调控元件, 我们鉴定了恶性和非恶性乳腺上皮细胞亚型中显著差异富集的转录因子(P -adjust<0.001) (图 3B)。相关研究报道, 正常乳腺上皮细胞显著富集了 *TP63*^[15]、*TEAD4*^[16] 和 *NFIC*^[17] 等在上皮组织的发育、分化和形态发生上起重要作用的转录因子, 而我们在恶性乳腺上皮细胞中观察到与乳腺肿瘤发展和转移相关的转录因子, 如 *TFAP2C*^[18]、*ESR1*^[19] 和 *SNAI2*^[20] 等。

为了在总体可及性数据中预测这些特征转录因子的活跃程度, 我们计算了特征转录因子的活

性偏差校正标准分数(图 4A)。与预期一致, *TP63*、*NFIC* 和 *FOXA1* 分别在基质型乳腺上皮细胞、管腔乳腺上皮细胞 A 型和管腔乳腺上皮细胞 B 型中拥有最高的活性分数, 其中, 转录因子 *NFIC* 同样在基质型乳腺上皮细胞中具有较高的活性分数。对于恶性上皮细胞的特征转录因子而言, *ESR1*、*SNAI2* 在恶性乳腺上皮细胞亚型中的活跃程度明显更高, 但 *TFAP2C* 的活性差异不明显。

为了了解这些特征转录因子在临床乳腺癌患者转录组数据中的表达情况, 探究其与乳腺癌发生发展的关系, 我们接下来提取了来自公共数据

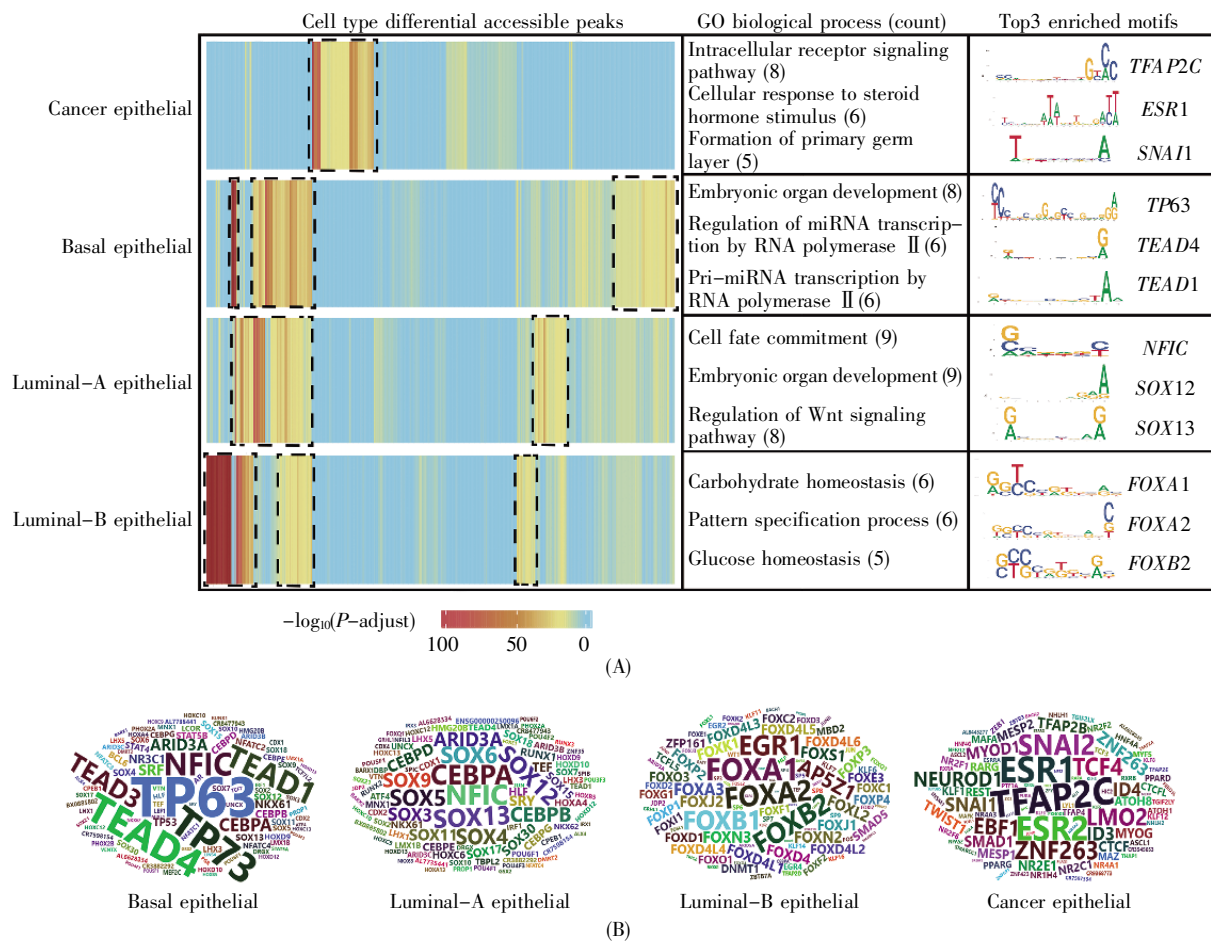


图 3 乳腺上皮细胞亚型特异性基因的调控图谱

(A) 上皮细胞亚型的转录因子结合基序标准化富集分数的热图, 以及相应 GO 功能富集分析结果。左图显示各上皮细胞亚型的特异性可及峰(differentially accessible peak, DAP), 中间图图是使用 GREAT 计算的 DAP 区域富集的 GO 项, 右图为各亚型 DAP 中排名前 3 的富集基序的结构图; (B) 各上皮细胞亚型富集基序情况。转录因子名称的标志越大, 其富集越显著。

Fig.3 Subtype-specific gene regulatory landscape of mammary epithelial cells

(A) Heatmap of normalized enrichment scores of transcription factor binding motifs in epithelial cell subsets and GO functional enrichment analysis results. Left: The differentially accessible peak (DAP) of each epithelial cell subtype. Middle: The enriched GO terms calculated by GREAT for DAPs. Right: The logos of the top 3 enriched motifs in each subtype DAP; (B) Enriched motifs for each epithelial cell subtype. The larger the logo of a transcription factor name is, the more significantly the transcription factor is enriched.

库 TCGA (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>) 的乳腺组织转录组测序(bulk RNA-seq)数据(2022 年 6 月 20 日), 对特征转录因子的基因表达情况进行了分析(图 4B)。结果显示, *TP63* 基因在瘤旁样本中的表达水平显著高于肿瘤样本, 与管腔上皮细胞分化密切相关的 *FOXA1* 基因在肿瘤样本中存在显著高表达, 恶性乳腺上皮细胞中的特征转录因子 *ESR1* 在肿瘤样本中的表达量显著高于瘤旁样本, 但 *NFIC*、*TFAP2C* 和 *SNAI2* 在两种类型样本中的表达量差异不显著, 推测转录因子 *NFIC*、*TFAP2C* 和 *SNAI2* 的差异可及性可能是上皮细胞谱系特异性的, 使用组织数据的分析无法体现这种差异性。

2.4 不同乳腺癌亚型中特征转录因子结合基序的可及性分析

乳腺癌的不同亚型都有着特殊的分子特征和生理学特性。因此, 在对恶性乳腺上皮细胞中特征富集的转录因子进行识别后, 我们期望了解这些特征转录因子结合基序在不同乳腺癌亚型中的可及性是否存在差异。通过对特征转录因子结合基序的局部染色质可及状态进行分析, 我们发现, *SNAI2* 仅在两个三阴性乳腺癌样本中具有高的可及性, 而 *TFAP2C*、*ESR1* 及 *LMO2* 无明显差异(图 5)。该结果提示, 转录因子 *SNAI2* 在三阴性乳腺癌亚型中具有调控作用。

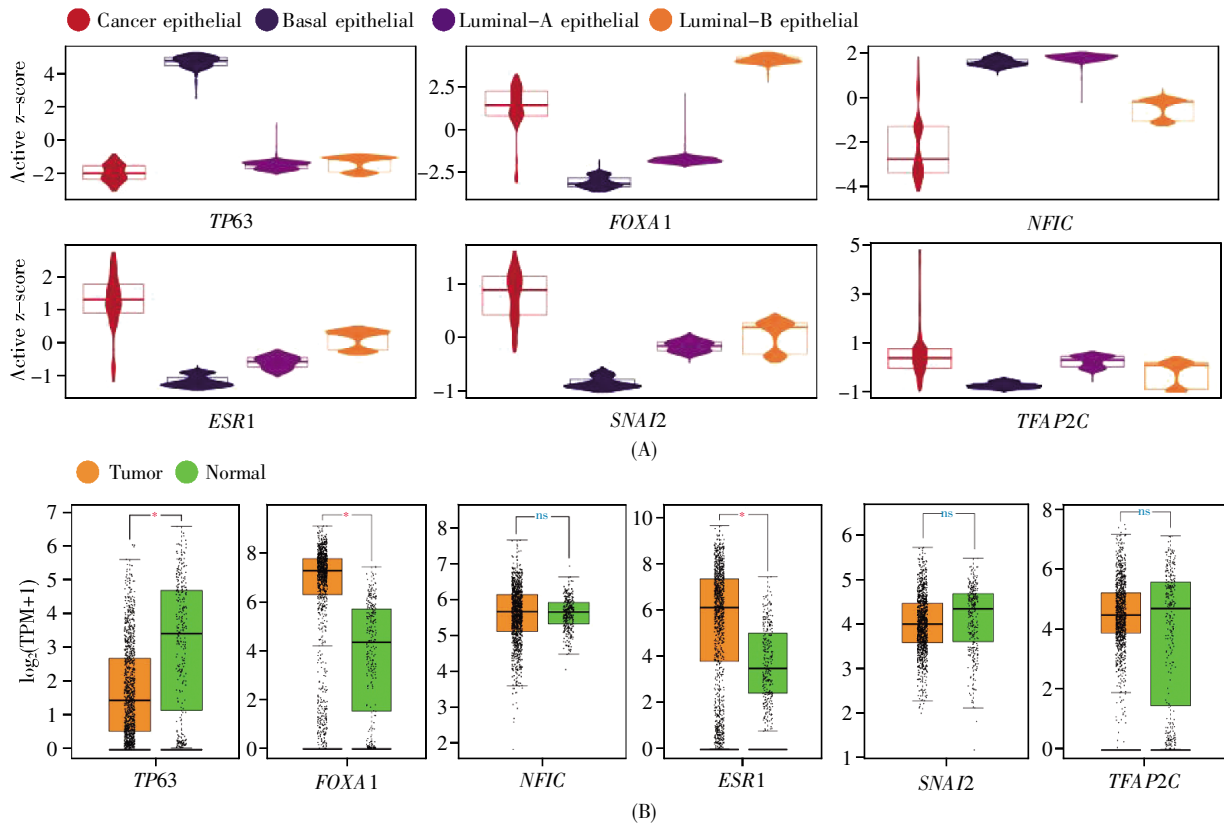


图 4 肿瘤和正常乳腺上皮细胞特征转录因子的活性差异

(A) 经 chromVAR 预测的特征转录因子活性分数的提琴图; (B) 特征转录因子在多细胞转录组测序结果中的表达分析。*表示 $P < 0.05$, ns 表示差异不显著。

Fig.4 Differences in the activity of characteristic transcription factors between tumor and normal mammary epithelial cells

(A) Violin plots of transcription factor activities predicted by chromVAR; (B) Analysis of characteristic transcription factor expression using multicellular transcriptome sequencing data. *: $P < 0.05$; ns: No significant difference.

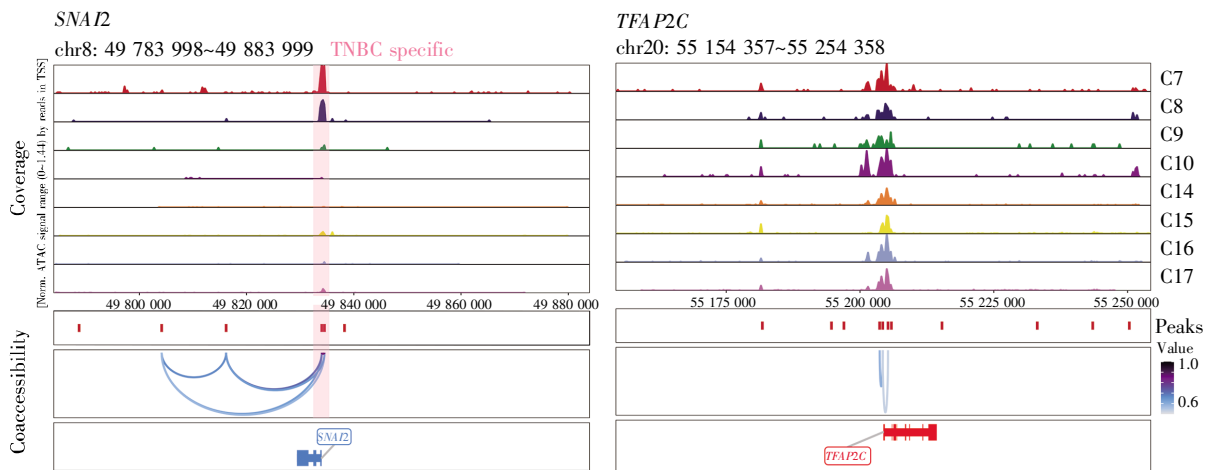


图 5 特征转录因子结合基序的局部染色质可及状态示意图

信号轨迹显示每个细胞群特征转录因子结合基序区域的染色质可及性峰。共可及关系在下方用弧线连接展示。正链基因用红色表示, 负链基因用蓝色表示。

Fig.5 Local chromatin accessibility track of the characteristic transcription factor binding motifs

The signal tracks show the chromatin accessibility peaks of the characteristic transcription factor binding motif region for each cell cluster. The coaccessibility is marked below with arc connections. Positive-strand genes are colored in red and negative-strand genes are colored in blue.

3 结论

本研究收集了公共数据库中的 14 名乳腺癌患者的乳腺肿瘤组织样本和两个正常乳腺组织样本的 scATAC-seq 数据, 使用 ArchR、Cicero 等生物信息学工具全面探索了乳腺癌组织与正常乳腺组织在染色质可及性层面的特征, 并重点聚焦于单细胞分辨率的上皮细胞类型特异性转录调控元件的识别和分析。研究发现, 来自肿瘤样本和正常乳腺样本的上皮细胞存在明显的染色质可及性差异, 提示其可能对乳腺癌内异质性以及患者间的治疗反应差异性具有重要影响。在针对 4 种上皮细胞亚型的分析中, 我们发现, 恶性乳腺上皮细胞显著富集了 *ESR1*、*SNAI2*、*LMO2* 等转录因子, 而 *TP63*、*NFIC* 以及 *SOX* 和 *FOX* 家族转录因子在正常的乳腺上皮细胞中显著富集, 表明这些调控元件在维持上皮细胞正常功能中具有重要作用。有趣的是, *SNAI2* 仅在两个三阴性乳腺癌样本中具有高的可及性, 这强调 *SNAI2* 是潜在的三阴性乳腺癌特异调节元件。综上所述, 本研究生成了基于单细胞 ATAC-seq 数据的正常及肿瘤乳腺组织的染色质可及性图谱, 为乳腺癌上皮细胞群的表观遗传特征提供了全新且更高分辨率的分析视角, 为破译正常乳腺组织和乳腺肿瘤组织的胞间异质性及临床差异性提供了新的参考角度。

参考文献(References):

- [1] NOLAN E, LINDEMAN G J, VISVADER J E. Deciphering breast cancer: from biology to the clinic[J]. *Cell*, 2023, 186(8): 1708–1728.
- [2] KARAAYVAZ M, CRISTEA S, GILLESPIE S M, *et al.* Unraveling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq[J]. *Nature Communications*, 2018, 9: 3588.
- [3] XU S W, LU Z X, SHAO W, *et al.* Integrative analysis of histopathological images and chromatin accessibility data for estrogen receptor-positive breast cancer[J]. *BMC Medical Genomics*, 2020, 13(Suppl. 11): 195.
- [4] LEE S, OSMANBEYOGLU H U. Chromatin accessibility landscape and active transcription factors in primary human invasive lobular and ductal breast carcinomas[J]. *Breast Cancer Research*, 2022, 24: 54.
- [5] CHEN D J, PARKER T M, BHAT-NAKSHATRI P, *et al.* Non-linear relationship between chromatin accessibility and estradiol-regulated gene expression[J]. *Oncogene*, 2021, 40(7): 1332–1346.
- [6] SANGHI A, GRUBER J J, METWALLY A, *et al.* Chromatin accessibility associates with protein-RNA correlation in human cancer[J]. *Nature Communications*, 2021, 12: 5732.
- [7] PAL B, CHEN Y S, VAILLANT F, *et al.* Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling[J]. *Nature Communications*, 2017, 8: 1627.
- [8] AZIZI E, CARR A J, PLITAS G, *et al.* Single-cell map of diverse immune phenotypes in the breast tumor microenvironment[J]. *Cell*, 2018, 174(5): 1293–1308.e36.
- [9] DIMRI G, BAND H, BAND V. Mammary epithelial cell transformation: insights from cell culture and mouse models[J]. *Breast Cancer Research*, 2005, 7(4): 171–179.
- [10] VENTEICHER A S, TIROSH I, HEBERT C, *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq[J]. *Science*, 2017, 355(6332): eaa18478.
- [11] FILBIN M G, TIROSH I, HOVESTADT V, *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq[J]. *Science*, 2018, 360(6386): 331–335.
- [12] YOSHIDA H, LAREAU C A, RAMIREZ R N, *et al.* The cis-regulatory atlas of the mouse immune system[J]. *Cell*, 2019, 176(4): 897–912.e20.
- [13] RUBIN A J, PARKER K R, SATPATHY A T, *et al.* Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks[J]. *Cell*, 2019, 176(1/2): 361–376.e17.
- [14] PETRIDIS C, BROOK M N, SHAH V, *et al.* Genetic predisposition to ductal carcinoma *in situ* of the breast[J]. *Breast Cancer Research*, 2016, 18: 22.
- [15] SU X H, CHAKRAVARTI D, FLORES E R. p63 steps into the limelight: crucial roles in the suppression of tumorigenesis and metastasis[J]. *Nature Reviews Cancer*, 2013, 13(2): 136–143.
- [16] LIU Y, WANG G, YANG Y, *et al.* Increased TEAD4 expression and nuclear localization in colorectal cancer promote epithelial-mesenchymal transition and metastasis in a YAP-independent manner[J]. *Oncogene*, 2016, 35(21): 2789–2800.
- [17] LEE H K, LEE D S, PARK J C. Nuclear factor I-C regulates E-cadherin via control of KLF4 in breast cancer[J]. *BMC Cancer*, 2015, 15: 113.
- [18] FRANKE C M, GU V W, GRIMM B G, *et al.* TFAP2C regulates carbonic anhydrase XII in human breast cancer[J]. *Oncogene*, 2020, 39(6): 1290–1301.
- [19] ROBINSON D R, WU Y M, VATS P, *et al.* Activating *ESR1* mutations in hormone-resistant metastatic breast cancer[J]. *Nature Genetics*, 2013, 45(12): 1446–1451.
- [20] BLANCO-GÓMEZ A, HONTECILLAS-PRÍETO L, CORCHADO-COBOS R, *et al.* Stromal SNAI2 is required for ERBB2 breast cancer progression[J]. *Cancer Research*, 2020, 80(23): 5216–5230.