

单细胞转录组测序探究三阴乳腺癌新辅助化疗后肿瘤微环境

吉林大学 朱明远

研究现状、成果

乳腺癌是威胁全世界女性健康的恶性肿瘤之一，乳腺癌在西方国家居女性恶性肿瘤发病率的首位，近年来，中国妇女尤其是城市妇女的乳腺癌发病率在逐年上升，其患病率跃居中国女性恶性肿瘤的第一位。而且中国乳腺癌发病高峰呈现年青化趋势，发病高峰已由45岁提前为35岁，乳腺癌已成为严重危害中国女性健康的杀手之一。

随着的诊疗理念的不断进步和科学技术的飞速发展，乳腺癌的治疗效果已较前显著提高，但其仍是我国女性的一大死亡原因。如今，乳腺癌的治疗以根治性手术为核心，放疗、化疗、靶向治疗相结合。由于肿瘤间以及肿瘤内的异质性，针对乳腺癌的控制是十分困难的，尤其是三阴乳腺癌（TNBC）由于缺少特异性的靶点，在多个水平上表现出异质性，给诊断和治疗提出了极大的挑战。尽管约30%的TNBC对标准化学治疗疗法敏感，使其存活率提高，但有些患者缺乏雌激素和孕激素受体以及人表皮生长因子2（neoadjuvant chemotherapy, HER2）受体，使得他们不适合激素或者抗HER2治疗，因此新辅助化疗（NAC）成为了许多TNBC患者的护理标准。除针对局部晚期乳腺癌患者外，NAC还可检测化疗方案敏感性、杀伤微转移灶、争取降期保乳机会等，得到了十分广泛的临床应用。相关研究表明，若乳腺癌患者接受NAC后，得以达到病理完全缓解，则其无病生存期(disease free survival, DFS)及总生存期(overall survival, OS)较未达到者更长。但NAC患者能达到乳腺癌化疗病理学缓解（pCR）者只有20%左右，更有少数患者在NAC期间出现了肿瘤进展。目前前对于新辅助化疗的适应症、具体化疗方案、化疗时程、疗效评价等尚未完全达成共识，一定程度上影响了病人的治疗及新辅助化疗的推广应用。乳腺癌新辅助化疗已经过了三十余年的发展，但还存在诸多问题待进一步研究，如新辅助化疗对象的选择，预测和监测及评估疗效。因而，研究NAC疗效的预测因子，对乳腺癌个体化综合治疗的开展具有很大的临床意义和实用价值。

NAC治疗对一些TNBC患者有效，但大约一般患者会产生耐药性，导致总体生存率低下。另一方面，肿瘤异质性对治疗效果以及患者的治愈后影响仍然不明确。以前其他癌症的基因组学研究经验表明，在不同的癌症类型中，会产生两种以上不同的抗性，即先前已有的亚克隆的选择和扩增（即适应性抵抗，adaptive resistance），及由于化疗导致的新突变而产生的（即获得性抵抗，acquired resistance）。由于目前尚不清楚化疗抗性的产生，对精细敏感肿瘤依赖性的缺乏机械性理解和缺乏预测因素阻碍了NAC向乳腺癌精准医学范式的转化。我们目的之一通过对肿瘤组织进行测序，以试图了解是否化疗后新突变会自发产生，或者是否在治疗前突变就已经以很低的水平存在。

尽管TNBC微环境（基质）在肿瘤进展中可能起关键作用，但其异质性尚未很好的被大家所认知。免疫系统细胞在非淋巴正常组织和肿瘤中起着重要的辅助功能，幼稚、效应和记忆T淋巴细胞和慢性刺激的功能失调的T淋巴细胞被认为是主要的T细胞分化状态。共刺激受体，如CD28, ICOS, OX40, CD40L和CD137, 显著增强T细胞受体（TCR）依赖性T细胞活化，而增加水平的共抑制受体CTLA-4, PD-1, TIGIT, LAG3, TIM-3和CD160是进行性T细胞功能障碍和自我更新潜力丧失（衰竭）的特征。这种专门的抑制细胞谱系被认为在癌症进展中起着重要作用，在乳腺癌中，在肿瘤亚型患者中观察到免疫组成具有显著异质性。这些观察提出了关于免疫细胞状态在不同异质性的肿瘤组织中是否不同以及它们是否代表有限数量的离散分化或活化中间体的问题。

了解肿瘤微环境中的免疫细胞表型对于理解癌症进展和免疫治疗反应的机制是必不可少的。随着下一代测序的迅速出现，采用单细胞RNA-seq（scRNA-seq）分析的进一步研究、探索肿瘤中免疫细胞亚群的更精细定义，但其规模有限。我们使用单细胞RNA-seq分析了来自NAC治疗前后乳腺癌的若干个免疫细胞及匹配乳腺组织、血液和淋巴结，探究三阴乳腺癌肿瘤异质性表型多样性的组合影响，建立一个动态模型。我们的结果对于表征肿瘤浸润性免疫细胞具有重要意义，最终将有助于确定乳腺癌中NAC耐药反应的机制基础。

津白2系(Tientsin Albino 2 strain,TA2)小鼠是开始经过了多年的筛选培育建立起来的自发性乳腺癌高发模型。TA2小鼠源于昆明种小鼠，在无任何外来化学药物等诱癌刺激下TA2小鼠的具有很高的自发性乳腺癌发生率，前人研究表明TA2小鼠乳腺癌的发生与是否妊娠和妊娠次数密切相关，见瘤时鼠龄平均为330天，且易发生肺转移、肝转移和脾脏转移。TA2小鼠自发乳腺癌病理形态类似人浸润性癌，免疫组织化学染色显示ER、PR和HER2阴性。TA2小鼠

自发乳腺癌生物学行为特征与人三阴乳腺癌极为相近，因此TA2小鼠是研究人三阴乳腺癌的一个良好动物模型，可用于研究三阴乳腺癌发生发展和转移的分子机制，筛选诊断标志物和治疗靶点及检测新药物和新治疗方案的有效性。

研究目的、方法

目的

采用单细胞RNA-seq (scRNA-seq) 分析，研究三阴乳腺癌 (TNBC) 新辅助化疗 (NAC) 治疗后肿瘤异质性及更精细免疫细胞亚群，通过建立动态模型探究乳腺癌中NAC耐药反应的机制基础。

研究方法：

- 招募患者，NAC治疗前后患者肿瘤样本及配对的正常乳腺组织、血液和淋巴结样本，收集一般临床资料，对NAC疗效相关因素分析
- 分离单细胞、RNA测序和生物信息学分析
- 全外显子组测序和数据处理
- 10x Genomics Chromium平台对来自乳腺癌组织的多个T细胞，进行配对单细胞RNA测序和TCR测序，揭示单个T细胞克隆型活化状态的范围
- SEQC流程、贝叶斯聚类、Biscuit算法对这些组织中的免疫细胞进行聚类和鉴定。
- 乳腺癌亚型特异性基因表达及免疫细胞类型特异性基因表达谱
- 通过qPCR验证RNA-seq数据
- 小鼠模型验证

1 患者招募与肿瘤样本

一般临床资料及NAC疗效相关因素分析

按照一定实验组纳入的标准，重点收集2010年1月至2018年12月于医院乳腺外科收治的行新辅助化疗的乳腺癌患者病例。比较三阴乳腺癌患者与非三阴乳腺癌患者临床病理学指标(年龄、肿瘤大小、腋淋巴结转移、治疗方式、局部复发及远处转移)之间是否存在差异，运用Kaplan-Meier生存分析比较两组患者的生存愈后情况。

通过对NAC疗效分析进行相关因素分析，应用卡方、Fisher精确检验单因素分析患者临床特征、乳腺癌分子分型、组织学分级和免疫组化指标等与病理缓解情况的关系。通过分析超声所示肿瘤缩小程度与病理缓解情况的一致性考察NAC疗效的价值。

本研究需要得到医院和医疗中心的机构审查委员会的批准，所有患者都提供了签署知情同意书，以收集标本并对衍生遗传物质进行详细分析。收集20个原发性肿瘤标本和转移性淋巴结并进行单细胞RNA测序，样本进行两轮单细胞RNA测序并组合用于下游分析。使用R包genefu对肿瘤的分子亚型进行确定。

2 分离单细胞和cDNA扩增

通过机械解离和酶消化获得乳腺癌组织或淋巴结转移的单细胞悬液，通过Ficoll-Paque PLUS分离除去死细胞，并将剩余细胞加载到C1 Single-Cell Auto中的单个集成流体回路mRNA测序芯片上。用显微镜检查加载的芯片以验证单细胞已经载入。对于细胞裂解、cDNA合成和扩增，按说明使用SMARTer Ultra Low RNA试剂盒。将来自ArrayControl RNA Spikes的spike-ins 1和4加入裂解混合物中。使用Qubit 2.0荧光计和2100 Bioanalyzer对扩增后的cDNA进行定量和鉴定。使用RNeasy Plus Micro试剂盒制备肿瘤组织分离物或肿瘤组织，并用SMARTer Ultra Low RNA试剂盒从单个用到细胞中提取扩增RNA。

3 全外显子组测序和数据处理

使用SureSelect XT Human All Exon V5 kit试剂盒捕获外显子，采用100-bp配对末端模式测序配套的TruSeq Rapid PE Cluster试剂盒和TruSeq Rapid SBS试剂盒构建HiSeq 2500系统测序文库。使用BWA将外显子组测序读长对其到参考基因组hg38。采用Picard标记潜在的重复突变 (<https://sourceforge.net/projects/picard/files/picard-tools/1.119/>)。通过应用GATK的模块重新校准可能包含小插入或缺失的位点，该模块具有千人基因组计划第1阶段确定的已知变异位点 (<http://www.1000genomes.org/>) 和dbSNP-137 (<http://www.ncbi.nlm.nih.gov/SNP/>) 的数据。我们使用MuTect来检测单核苷酸变异 (SNV)，并使用Control-FREEC软件包来检测CNV。肿瘤的外显子组测序 (WES) 覆盖率为100×，配对的血液样本的覆盖率为50X。

4 RNA测序和生物信息学分析

使用Nextera XT DNA样品制备试剂盒扩增cDNA构建测序文库，并使用HiSeq2500测序仪与TruSeq Rapid PE Cluster试剂盒和TruSeq Rapid SBS的100bp配对末端模式中测序。评估阵列对照spike-in的表达值，通过将三种对照spike-in (ThermoFisher) 与人基因组参考序列 (hg38) 和GENCODE 38注释合并来产生参考序列和相应的注释。然后使用STAR将RNA读数与参考序列比对，并使用RSEM将相对基因表达定量为百万转录数 (TPM)。将每个基因的同种型表达水平相加以得到TPM值。使用RNA-SeQC 2.0对对齐的scRNA-seq读长进行质量控制评估。为了去除具有低质量测序值的细胞，应用了总读数、映射率、总的基因数量等标准。去除具有低表达值的基因，首先TPM值 < 1被认为是不可靠的并用零代替。其次，TPM值lg转换。第三，除去在 < 10% 的所有肿瘤组中表达的基因。通过QC标准确定过滤后的细胞以及基因，过滤的基因将用于肿瘤分析。为了获得单细胞RNA测序的灵敏度和重复性的估计，使用加入外源RNA进行评估。对于转录组分析，RNA-seq和定量PCR (qPCR) 结果之间的比较; 检测染色体表达模式; 测量肿瘤内相关性; 与免疫荧光染色结果比较; 免疫标志物表达的比较; 自我归一化工具的应用 (如基因集变异分析 (GSVA) 和TNBCtype)。

5 发性乳腺癌患者免疫细胞的单细胞转录组测序

本研究通过整合Illumina的HiSeq 2500测序仪、Droplet Genomics的芯片，以及定制的inDrop单细胞RNA-Seq微流体平台，对接受过NAC治疗的原发性乳腺癌患者的免疫细胞进行转录组分析。研究对象包括表达雌激素受体 (ER) 的肿瘤、表达孕激素受体 (PR) 的肿瘤、表达人表皮生长因子受体2 (HER2) 的肿瘤，以及ER、PR和HER2都不表达的三阴肿瘤。预期在结果中看到了各种免疫细胞簇，包括单核细胞、巨噬细胞、肥大细胞、T细胞、B细胞、树突状细胞和嗜中性粒细胞。

6 10x Chromium平台配对的单细胞RNA-seq和TCR-seq揭示单个T细胞克隆型活化状态的范围

在后续实验中，为了更深入的了解TCR库的多样性是否有助于观察T细胞活化谱和整体表型的多样性，我们通过10x Genomics单细胞平台，选取来自3个额外乳腺癌组织多个T细胞，进行配对的单细胞RNA测序和TCR测序。由此方法获得的数据能通过相同的单个细胞，将基因表达直接映射到TCR应用中。结果将会观察到TCR多样性与T细胞连续激活的驱动因素的联系。

7 10x Genomics 单细胞5'和VDJ测序的文库构建

将FACS分选的CD3+免疫细胞，以700 cell/μl的浓度制成微滴，以期能获得标记的单细胞mRNA信息。液滴破碎，mRNA逆转录后，产生带有Barcode和UMI的cDNA，barcoded-cDNA纯化，PCR扩增，构建5'基因表达文库和富集的T细胞受体文库，上机测序。

本次试验观察的结果，以及基于10x Chromium平台的免疫细胞RNA-seq和TCR-seq数据集，加上全面的分析手段，将有助于我们更好地了解免疫细胞促进和延缓肿瘤生长背后的潜在分子机制。

8 从RNA-seq数据中推测重复序列

为了鉴定与推定的非癌细胞相比的癌细胞的不同染色体基因表达模式，来自GTEx门户 (<http://www.gtexportal.org/>) 的正常乳腺组织的表达谱首先转化为与我们的数据集相当的值。其次，计算在我们过滤的单细胞数据中检测到的基因的平均基因表达及其对所有正常乳腺组织的变化。通过将我们的单细胞数据归一化为正常乳腺组织的平均表达谱来计算每个基因的数值。所有基因按染色体编号和起始位置分类。从150个bp作为滑动窗口估计染色体表达模式，并将其调整为跨基因的中心值。

9 基因组CNV与推断的CNV之间的相关性

对体细胞CNV和大规模片段 (> 10,000bp) 进行CNV的比较。为了并行比较WES和RNA-seq之间的CNV，将两个CNV分成10Mb窗口大小，并推断CNV在单个细胞中取平均值。使用R函数通过Pearson相关分析计算相关系数。

10 途径分析

为了评估基因表达特征和途径激活，使用具有RNA-seq模式的R包GSVA软件的非参数和无监督软件算法。我们在GSVA中包括GTEx正常乳腺组织样品，以评估对照途径激活水平。对于所有基因组，使用超几何测试进行过表达分析，并且分离出P值<0.05的那些下游基因。

11 乳腺癌亚型特异性基因表达

我们采用R包“scater”来分析单细胞数据和所使用的LRT基于零膨胀数据特性测试，以确定亚型特异性标志物。为了显示每个亚型中的区室途径激活或鉴定与肿瘤细胞相比的富含HER2的肿瘤细胞的特征基因，计算亚型相关途径或HER2 / HER3下游信号传导途径的GSVA富集评分。

12 三阴性乳腺癌亚型

使用TNBctype (<http://cbc.mc.vanderbilt.edu/tnbc/index.php>) 将三阴性乳腺癌细胞分为六种亚型。TNBctype对于每个亚型具有六个类型，由亚型 - 特征基因和训练集组成。通过将候选样本与六个类型进行比较，TNBctype为每个子类型提供Spearman相关系数和P值。一些具有多于一种亚型的细胞具有较高的相关系数 (P<0.05)，因此被分类为多种亚型。在应用TNBctype之前，除去在任何单个TNBC细胞中不表达的基因。

13 免疫细胞类型特异性基因表达谱

使用在非重叠免疫细胞类型中注释的基因，通过来自非肿瘤细胞的非阴性因子分解，鉴定免疫细胞亚组。为了表征三种免疫细胞簇，使用Seurat进行接收操作特性测试和基于零膨胀数据的LRT。然后，获得倍数> 2且曲线下面积> 0.7的基因作为细胞类型特异性基因。基因本体论富集细胞类型特异性基因由DAVID注释 (<https://david.ncifcrf.gov/>)。为了通过功能及状态进一步表征T或B细胞，用选择的基因组进行GSVA分析。

14 免疫荧光染色

进行免疫荧光染色以评估肿瘤组织中肿瘤浸润性T或B细胞的存在。用福尔马林固定的抗CD3和抗MARK3抗体对T淋巴细胞进行双染色。载玻片石蜡包埋 (FFPE) 用抗CD20和抗PRPSAP2抗体对B淋巴细胞进行双染色。Alexa488标记的抗小鼠和Alexa568标记的抗兔抗体 (1:50; Invitrogen) 用于具有4,6-二脒基-2-苯基衍生物的双重免疫荧光。将CD3 +或CD20 +细胞的数量评估为三个0.125mm²区域中的平均计数，统计具有最大阳性染色。

15 通过qPCR验证RNA-seq数据

使用来自混杂样本以及单细胞样品的cDNA，用DELTA基因测定法 (PN100-3035, Fluidigm) 进行qPCR。使用D3软件 (Fluidigm) 设计引物序列。在比较qPCR和RNA-seq数据之前，未检测到值替换为'NA'，进行对数转换以及设置阈值。通过Pearson的相关性，Spearman的秩次相关和线性回归分析来评估相互关系。

项目计划

- 2019-2020年，完成项目立项、评估答辩、患者招募、肿瘤样本的收集及测序
- 2020-2021年，数据分析基础上，持续跟踪患者的病理状况，并收集相关临床资料；研究论文的撰写
- 2021-2022年，项目答辩，数据的补充以及上游分析的下游验证，小鼠用药模型的建立

Reference

- [1]Elham Azizi;Ambrose J. Carr;George Plitas;Andrew E. Cornish;Catherine Konopacki;Sandhya Prabhakaran;Juozas Nainys;Kenmin Wu;Vaidotas Kisieliovas;Manu Setty;Kristy Choi;Rachel M. Fromme;Phuong Dao;Peter T. McKenney;Ruby C. Wasti;Krishna Kadaveru;Linaz Mazutis;Alexander Y. Rudensky;Dana Pe'er.Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment[J].Cell,2018,
- [2]Woosung Chung;Hye Hyeon Eum;Hae-Ock Lee;Kyung-Min Lee;Han-Byoel Lee;Kyu-Tae Kim;Han Suk Ryu;Sangmin Kim;Jeong Eon Lee;Yeon Hee Park;Zhengyan Kan;Wonshik Han;Woong-Yang Park.Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer[J].Nature Communications,2017,Vol.8: 15081
- [3]Sun H;Miao Z;Zhang X;Chan UI;Su SM;Guo S;Wong CKH;Xu X;Deng CX..Single-cell RNA-Seq reveals cell heterogeneity and hierarchy within mouse mammary epithelia.[J].J Biol Chem,2018,
- [4]Peter Savas;Balaji Virassamy;Chengzhong Ye;Agus Salim;Christopher P. Mintoff;Franco Caramia;Roberto Salgado;David J. Byrne;Zhi L. Teo;Sathana Dushyanthen;Ann Byrne;Lironne Wein;Stephen J. Luen;Catherine Poliness;Sophie S. Nightingale;Anita S. Skandarajah;David E. Gyorki;Chantel M. Thornton;Paul A. Beavis;Stephen B. Fox;Kathleen Cuninghame Foundation Consortium for Research into Familial Breast Cancer (kConFab), Phillip K. Darcy;Terence P. Speed;Laura K. Mackay;Paul J. Neeson;Sherene Loi.Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis[J].Nature Medicine,2018,
- [5]Tanioka Maki;Fan Cheng;Parker Joel S;Hoadley Katherine A;Hu Zhiyuan;Li Yan;Hyslop Terry M;Pitcher Brandelyn N;Soloway Matthew G;Spears Patricia A;Henry N Lynn;Tolaney Sara;Dang Chau T;Krop Ian;Harris Lyndsay N;Berry Donald A;Mardis Elaine R;Winer Eric P;Hudis Clifford A;Carey Lisa A;Perou Charles M.Integrated analysis of RNA and DNA from a phase III trial of trastuzumab-based neoadjuvant chemotherapy identifies response predictors in HER2-positive breast cancer.[J].Clinical cancer research : an official journal of the American Association for Cancer Research,2018,
- [6]Bjørklund, Sunniva Stordal;Kristensen, Vessela N.;Seiler, Michael;Kumar, Surendra;Grenaker Alnæs, Grethe I.;Yao Ming;Kerrigan, John;Naume, Bjørn;Sachidanandam, Ravi;Bhanot, Gyan;Børresen-Dale, Anne-Lise;Ganesan, Shridhar.Expression of an estrogen-regulated variant transcript of the peroxisomal branched chain fatty acid oxidase ACOX2 in breast carcinomas.[J].BMC Cancer,2015,Vol.15(1): 1-13
- [7]Gupta SC 1 ;Tripathi YN 1 ..Potential of Long Non-coding RNAs in Cancer Patients: From Bio-markers to Therapeutic Targets.[J].Int J Cancer,2017,Vol.140(9): 1955-1967
- [8]Harrod A;Fulton J;Nguyen VT;Periyasamy M;Ramos-Garcia L;Lai CF;Metodieva G;de Giorgio A;Williams RL;Santos DB;Gomez PJ;Lin ML;Metodiev MV;Stebbing J;Castellano L;Magnani L;Coombes RC;Buluwela L;Ali S.Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer.[J].Oncogene,2016,
- [9]Freeman BT 1, 2, 3 , Jung JP 1, 2 , Ogle BM 1, 2, 3, 4, 5, 6 ..Single-cell RNA-seq reveals activation of unique gene groups as a consequence of stem cell-parenchymal cell fusion.[J].Scientific Reports,2016,Vol.6

[10]Athreya AP;Kalari KR;Cairns J;Gaglio AJ;Wills QF;Niu N;Weinshilboum R;Iyer RK;Wang L..Model-based unsupervised learning informs metformin-induced cell-migration inhibition through an AMPK-independent mechanism in breast cancer.[J].Oncotarget,2017,Vol.8(16): 27199-27215