研究主题：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。

研究设计：

-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平台对来自额外乳腺癌组织的多个T细胞，进行配对的单细胞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算法等，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

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

题目：scRNA-seq探究三阴乳腺癌NAC治疗后肿瘤微环境

立项人：吉林大学 朱明远

项目意义：

-三阴乳腺癌的研究意义

因为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。其中，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。

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，其中新辅助化疗(neoadjuvantchemotherapy，NAC)更是得到了十分广泛的临床应用，发挥着举足轻重的作用。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。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受益及新辅助化疗的推广应用。现如今，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。

尽管约30％的TNBC对标准治疗化学疗法敏感，导致存活率提高，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（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，。虽然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。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。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了。

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。同样，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。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，显着增强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，在肿瘤亚型和患者中观察到免疫组成的显着异质性。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）分析的进一步研究已开始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。

内容：

1 患者和肿瘤样本

1.1 一般临床资料及NAC疗效相关因素分析

按照一定收入实验组的标准，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。记录年龄分布、BMI指数、恶性肿瘤家族史全部、是否存在腋窝淋巴结肿大、是否于术前发现远处转移。

通过对NAC疗效分析进行相关因素分析，应用x 2检验或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乳腺超声检查对NAC疗效的评估价值

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。

1.3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，按照制造商的说明使用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，并用与SMARTer Ultra Low RNA试剂盒从单个用到细胞中提取扩增RNA 。

1.4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I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。肿瘤的整个外显子组测序（WES）覆盖率为100×，配对的血液样本的覆盖率为50X。

2 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来构建测序文库，并使用HiSeq250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。为了评估阵列对照RNA加标的表达值，通过将三种对照RNA加标（ThermoFisher）与人基因组参考序列（hg19）和GENCODE 19注释合并来产生参考序列和相应的注释。然后使用STAR\_2.4.0b的2通模式（默认参数）55将RNA读数与参考序列比对，并使用RSEM v1.2.17（默认参数）将相对基因表达定量为百万转录（TPM）（默认参数）56。将每个基因的同种型表达水平相加以得到TPM值。使用RNA-SeQC 57进行对齐的单细胞RNA-seq读数的质量控制评估，并且结果总结在补充数据3中。为了去除具有低质量测序值的细胞，应用了四种过滤标准：（1）总读数; （2）映射率; （3）检测到的基因数量; （4）基因间区域的一部分。为了去除具有低表达值的基因，应用以下步骤。首先，TPM值<1被认为是不可靠的并且用零代替。其次，TPM值为log 2 - 在添加值1后转换。第三，除去在<10％的所有肿瘤组中表达的基因。总共有515个单细胞和17,779个基因通过了QC标准。过滤的17,779个基因也用于大量肿瘤分析。作为对单细胞RNA测序的灵敏度和重复性的估计，我们获得了RNA加标1（12,200个转录本）和RNA加标4（912个转录本; 补充图2a）的一致log 2（TPM + 1）比率。）。没有检测到具有估计的62个转录物输入的RNA加标7。对于转录组分析，表达数据的平均值是通过减去平均log 2来实现的（TPM + 1）每个基因的值，但以下情况除外：用于RNA加标分析; 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; 和自我归一化工具的应用，如基因集变异分析（GSVA）和TNBCtype。来自原

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，对来自8名从未接受过治疗的原发性乳腺癌患者的47,016个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。最终研究如同预期的那样，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。

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，研究人员通过10x Genomics单细胞平台，选取来自3个额外乳腺癌组织（BC9-11）的27,000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。最终表明TCR多样性并不是T细胞连续激活的唯一驱动因素，也存在

10x Genomcis 单细胞5'和VDJ测序的文库构建

将大约12,000个FACS分选的CD3+免疫细胞（90％-95％存活率），以700 cell/μl的浓度制成微滴，以期望能获得7,000个标记的单细胞mRNA信息（multiplet rate: 5.4%）。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。

从RNA-seq数据中复制数字推断

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log 2（TPM +） 1）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。第三，Z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分数。所有基因按染色体编号和起始位置分类。从150个基因的移动平均值作为窗口大小估计染色体表达模式，并将其调整为跨基因的中心值。

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。

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58中称为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相对途径激活水平。对于所有基因组，使用超几何测试进行过表达分析，并且使用P值<0.05的那些。

乳腺癌亚型特异性基因表达

我们采用将R包“修拉” 32来分析单细胞数据和所使用的LRT基于零膨胀数据和所述接收器操作特性测试，以确定亚型特异性标志物（其平均表达比双重和分类能力的情况下（曲线下面积）高于0.7，LRT P <0.05; 补充表2a-c）。为了显示每个亚型中的区室途径激活或鉴定与腔HER2 +肿瘤细胞相比的富含HER2的肿瘤细胞的特征基因，计算来自MSigDB v5.0的亚型相关途径或HER2 / HER3下游信号传导途径的GSVA富集评分。

三阴性乳腺癌亚型

使用TNBCtype（http://cbc.mc.vanderbilt.edu/tnbc/index.php）59将三阴性乳腺癌细胞分为六种亚型（图5c）。TNBCtype对于每个亚型具有六个质心，由2,188个亚型 - 特征基因和386个训练样品定义。通过将候选样本与六个质心进行比较，TNBCtype 为每个子类型提供Spearman相关系数和P值。一些细胞具有高相关系数（P.<0.05）具有多于一种亚型，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。然后上传输入文件而不居中以避免由于大多数TNBC单细胞中的ESR1表达为零而导致错误的ER +肿瘤细胞调用。

免疫细胞类型特异性基因表达谱

使用在11种非重叠免疫细胞类型中注释的412个基因，通过来自175个非肿瘤细胞的非阴性因子分解60鉴定了三个免疫细胞亚组37（补充表3）。为了表征三种免疫细胞簇，使用Seurat进行接收器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（LRT P <0.05; 补充表4a-c）。基因本体术语显着富集细胞类型特异性基因由DAVID 6.7注释（https://david.ncifcrf.gov/）使用默认选项。为了通过功能状态进一步表征T或B细胞，用文献中选择的基因组进行GSVA分析（补充表5a-c）。

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（1：200; MA5-12577，Thermo Fisher，Waltham，MA，USA）和抗MARK3（1：100; PA5-29328，Thermo Fisher）抗体对T淋巴细胞进行双染色。石蜡包埋（FFPE）载玻片。用抗CD20（1：200; MA5-13141，Thermo Fisher）和抗PRPSAP2（1:50; PA5-31237，Thermo Fisher）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0.125mm 2区域中的平均计数，具有最大阳性染色。

通过qPCR验证RNA-seq数据

使用来自6个大量和185个单细胞样品的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，并列于补充表6中。在比较qPCR和RNA-seq数据之前，将999（=未检测到）的Ct值替换为'NA'。Ct值被负转换，-20被设定为阈值。这些数据代表qPCR 的log 2表达水平，与RNA-seq的log 2（TPM + 1）相当。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。

方案进度：2018.7-2018.8，进行二细胞期样品的采集，送测序；

* + - 1. 样品的质控、对齐、表达矩阵的获取以及质控；

2018.10-2018.12,去干扰、差异表达、组成分分析、通路富集；

2019.1-2019.3；表观遗传相关基因、细胞周期相关基因以及新的转录本基因差异比对；

* 1. 结题报告的撰写。

# References