

## 10x 单细胞多组学 (ATAC + 基因表达) 常见问题 (FAQ)

### 一、前期常见问题

#### 1. 组织样本如何制备单细胞悬液?

答：目前制备单细胞悬液的方法有两种

- 1) 客户用酶自己解离，公司会提供样本准备指南供老师参考，但不同的组织解离难度不同，需要客户自己调整实验条件；
- 2) 欧易生物提供样本解离服务，如果选择公司解离，我们会给客户寄组织保存液，需要在规定时间 (30h 内) 将组织寄回公司。

#### 2. 是否所有细胞都符合 10x 单细胞多组学 (ATAC + 基因表达) 测序?

答：不是。满足以下条件的细胞才能做单细胞转录组测序

- 1) 细胞直径小于 40um，大于 40um 的单核细胞，可考虑用细胞核来做；
- 2) 细胞活性大于 85% (10x 平台)，或 65% (BD 平台)。

#### 3. 组织或植物做单细胞测序需要注意哪些因素?

答：需要注意三点

- 1) 植物需要制备原生质体，原生质体直径小于 40um；
- 2) 组织尽量新鲜，冻存组织需要用冻存液，不可直接-80°C或液氮冷冻；
- 3) 组织含有红细胞，需要裂红，以免红细胞混杂降低目标细胞数量。

#### 4. 单细胞测序需要多少细胞数?

答：单细胞测序需要细胞悬液，总细胞数为  $10^5$  左右，建议浓度为 700-1200 cells/ul。



#### 5. 如果细胞量小于 10,000，能否进行实验？

答：一般不建议。如果因为客户原因要做单细胞转录组测序，需要客户承担风险。

#### 6. 如果重点关注某一细胞类型，能否保证检测到这种细胞？

答：不一定。通常情况，建议客户做预实验，如果关注的细胞在样本中占比大于 10%，一般可以检测到，如果细胞在样本中占比偏低（小于 10%），可能检测不到，或检测到很少，不利于后续分析，建议客户用磁珠或流式分选后上机测序。

#### 7. 直接拿组织做单细胞测序和 FACS 后做单细胞测序，有何区别？

答：通常做 FACS 的原因有两种：

- 1) 老师只关注组织中的某些细胞类型；
- 2) 老师关注组织中的所有细胞，但某些细胞类型在其中占比很低（小于 10%）。

#### 8. 单细胞转录组测序需要做重复吗？

答：没有严格要求。但由于样本间存在差异（尤其是临床样本）以及目前文章的趋势来看，建议老师至少 3-5 个重复。

#### 9. 单细胞测序是否会存在双细胞（或多细胞）的情况？

答：会。根据 10x Genomics 官方提供的数据显示，当上样细胞数量为 9,000 左右的时候，多细胞概率仅 3.9 %。BD 平台的 Scanner 也能实时检测双细胞率。此外，在后续的数据分析中，会有细胞基因定量质控分析，去除离域细胞，从而保证结果的准确性。



## 10. 单一的细胞株能否做单细胞测序？

答：可以，但要注意以下几种情况

- 1) 细胞间是否具有异质性，若细胞株太纯，结果可能没差异，所以不建议；
- 2) 癌症细胞异质性通常很高，可以尝试；
- 3) 若前期发现培养过的细胞会分化成不同功能的细胞，可以尝试。

## 二、数据分析问题

### 1. 如何知道每个样本捕获到的细胞数？

答：10x Genomics 平台有官方推出的 CellRanger ARC 质控软件，可以知道捕获到的细胞数。BD 平台有 Scanner，可以监测细胞的捕获过程及捕获到的细胞数目。

### 2. 10x 单细胞多组学 (ATAC + 基因表达) 除了 CellRanger ARC 质控外，是否还进行其他质控？

答：会的。除了 CellRanger ARC 外，还会对 UMI 数，基因数、线粒体基因数和 Peak 中片段数量所占比例等进行质控，从而过滤掉低质量的细胞。

### 3. 根据什么将细胞划分为不同的 cluster，可以调整 cluster 的数目吗？

答：cluster 的划分是根据特征模式相似性进行分析的，一般特征模式相似性高的细胞会聚为 1 个 cluster。此外，Cluster 的数目可以根据需要进行调整，目前主要有两种聚类方式：K-means 和 Graph-based。K-means 可认为指定 cluster，而 Graph-based 可根据样本具体情况计算最适的 cluster。

### 4. 为什么鉴定到的 cluster 数和细胞类型数不对应？

答：通常情况，同一种细胞类型也会存在基因表达异质性，即细胞亚型，这种



基因表达的异质性会反映在 cluster 上, 因此, 如果客户比较关注某一细胞类型, 可以将这种细胞单独进行亚型分析。

### 5. 10x 单细胞多组学 (ATAC + 基因表达) 中每个细胞能捕获到的基因数和转录本数是多少?

答: 根据不同的细胞类型, 捕获到的基因数会有区别, 有些细胞类型在 3000 左右, 而像 pmBC 这种基因表达水平比较低的大概在 1000 左右, 具体视细胞类型而定。每个细胞捕获到的转录本也会有区别, 一般在一万左右。

### 6. 降维聚类图是根据 Marker 基因或 Peak 画出来的吗?

答: 不是。数据分析时, 首先根据细胞特征模式的相似性进行降维聚类分析, 其次, 筛选每个聚类中特异性高表达或、高开放的特征, 即 Marker 基因或 Peak。

### 7. Marker 的鉴定原则是什么, top10 是如何筛选的?

答: 一般认为某个 cluster 的 Marker 基因或 Peak 为在某特定 cluster 中高表达或高开放的, 在其他 cluster 中低/不表达或开放的特征, 可以根据 pct1 和 pct2 的值来评估, top10 的 marker 基因是按照 GeneDiff 由大到小来排列的。

pct1: clusterA 中表达或开放特征 a 的细胞数除以 clusterA 中细胞数的比值

pct2: 除了 clusterA, 其他 cluster 中表达或开放特征 a 的细胞数除以其他 cluster 总细胞数的比值

$\text{GeneDiff} = \text{pct1} / \text{pct2}$

### 8. 数据集鉴定的细胞类型一定是正确的吗?

答: 不一定。目前人和小鼠的样本可以通过数据集来鉴定细胞类型, 但数据集提供的只是参考, 还要通过 marker 基因来辅助判断细胞类型。



#### 9. 鉴定细胞类型的 Marker 基因公司是否提供？

答：关于人和小鼠，有已知的 Marker 基因数据库，公司或客户都可以查找。  
除人和小鼠外或者鉴定细胞亚型（包括人和小鼠）需要客户提供 Marker 基因。

#### 10. 为什么同一个 cluster 会鉴定到多种细胞类型？

答：一般认为相同的细胞类型，细胞具有相似的表达谱，但通常情况下同一个 cluster 也会混有其他细胞类型，只要不多，可以忽略。

#### 11. 是否可以通过提高 cluster 数目将所有细胞类型的亚型划分出来？

答：有可能，但一般不建议直接这样做，直接将 cluster 数目提高可能导致某些细胞聚类效果不好，如果老师关注某特定细胞类型，可将其单独进行降维聚类，即亚型分析。

#### 12. 如果进行亚型分析，能保证划分出很多 cluster 吗？

答：cluster 的多少取决于细胞的异质程度，如果细胞类型本身差异很小的话，cluster 可能不会很多。

#### 13. 细胞亚型分析对细胞数有要求吗？

答：有的。细胞数太少，可能差异很小，很难划分，一般至少 50 个细胞以上才能进行亚型的划分。

#### 14. 为什么送的是肿瘤组织却没有检测到肿瘤细胞？

答：通常肿瘤细胞是由一些正常细胞转变而成，在细胞类型鉴定时只能鉴定到属于哪种正常细胞（或基质细胞），如果想进一步分析哪些细胞属于肿瘤细胞，可通过 CNV 分析来判断。



### 15. Marker 的可视化图 (feature plot 图) 能否反映基因表达量或 Peak 开放度的高低?

答: feature plot 图是经过归一化的图, 只能反映基因在哪些 cluster 中有表达, 并不能反映基因表达量或 Peak 开放度的高低, 建议通过小提琴图 (violin) 比较不同 cluster (或组别间) 的变化。

### 16. 染色质开放区与染色质可及性该如何解释?

答: 当基因处于转录状态时, 染色质由紧密压缩状态转变为松散状态, 为转录因子的结合提供机会。此时, 松散的染色质区域称为染色质开放区, 转录因子能够与染色质开放区结合的性质则称为染色质可及性。目前对染色质可及性的研究方法为 ATAC, 主要着眼于获得全基因组染色质开放区, 比较细胞分化、环境响应、疾病发生发展相关表观遗传改变的影响。

### 17. 为什么要用 peak 邻近基因对 peak 进行注释, 是否存在调控关系?

答: 进行 peak 分析时会以 peak 在基因组上的位置对其进行命名 (如: chr1-10089-10275), 但大多数客户无法对其进行理解, 故我们将 peak 对应至最邻近的基因以便对 peak 位置进行理解, peak 与其最邻近注释基因间并不一定存在调控关系。

### 18. 10x 单细胞多组学 (ATAC + 基因表达) 如何进行整合分析?

答: 加权最近邻(WNN)是一种联合多模态数据重新定义细胞状态的分析框架, 该方法基于无监督策略学习细胞特异的模式“权重”, 每种模式同时包含了多模态的信息, 并确定了其在下游分析中的相对重要性。WNN 整合分析大大提高了



我们在多种生物背景和数据类型中定义细胞状态的能力。

