

## 单细胞转录组测序常见问题（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单细胞转录组测序？

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

- 3) 植物需要制备原生质体，原生质体直径小于 40um；
- 4) 组织尽量新鲜，冻存组织需要用冻存液，不可直接-80℃或液氮冷冻；
- 5) 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组织或血液样本？

答：组织样本 1cm<sup>3</sup> 左右，血液（全血，外周血，骨髓，脑脊液等）样本 5-10ml 左右。

#### 5. 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<sup>5</sup> 左右，建议浓度为 700-1200 cells/ul。

#### 6. 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。

#### 7. FACS 后的细胞可否进行单细胞转录组测序？

答：可以。很多文献也都会将 FACS 后的细胞做单细胞测序，但流式对细胞会产生一定程度的损伤，分选后的细胞需要尽快安排上机。

#### 8. 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。

#### 9. 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%)。

#### 10. 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。

#### 11. 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。

#### 12. 单细胞转录组测序能否检测表面蛋白?

答: 可以。目前 10x Genomics 平台和 BD 平台都能检测人和小鼠的细胞表面蛋白, 此外 BD 平台只支持检测小鼠 CD45 免疫细胞的表面蛋白。

#### 13. 如果能检测到表面蛋白, 是否能知道这些抗体对应的抗原是什么?

答: 目前还不能。

#### 14. 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。

#### 15. 病毒侵染类型的项目能否做单细胞转录组测序?

答: 可以。10x Genomics 捕获的是具有 poly (A) 尾巴的 RNA, 因此, 只要 RNA 具有 poly(A) 尾, 理论上是可以捕获得到的。目前已有病毒侵染类型的单细胞转录组文章。

## 二、数据分析问题

#### 16. 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 质控软件, 可以知道捕获到的细胞数, 每个细胞的平均 reads 数和每个细胞的中位基因数。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。

#### 17. 10x Genomics 除了 cellranger 质控外, 是否还进行其他质控?

答: 会的。除了 CellRanger 外, 还会对 UMI 数, 基因数和线粒体基因数进行质控, 从而过滤掉低质量的细胞。

#### 18. 单细胞测序的 reads 能否进行拼接?

答: 不能, reads1 只读取 barcode 和 UMI 信息, read2 仅仅用来比对到基因组上。

#### 19. 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 的划分是根据细胞基因表达模式相似性进行分析的，一般表达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。

#### 20. 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。

#### 21. 10x Genomics 单细胞转录组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。

#### 22. tSNE 降维聚类图是根据 marker 基因画出来的吗？

答：不是。数据分析时，首先根据细胞表达谱的相似性进行 tSNE 降维聚类，其次，筛选每个聚类中特异性高表达的基因，即 marker 基因。

#### 23. marker 基因的鉴定原则是什么，top10 是如何筛选的？

答：一般认为某个 cluster 的 marker 基因为在某特定 cluster 中高表达，在其他 cluster 中低表达或不表达的基因，可以根据 pct1 和 pct2 的值来评估，top10 的 marker 基因是按照 Gene Diff 由大到小来排列的。

Pct1: clusterA 中表达 a 基因的细胞数除以 clusterA 中细胞数的比值

Pct2: 除了 clusterA，其他 cluster 中表达 a 基因的细胞数除以其他 cluster 总细胞数的比值

$$\text{Gene Diff} = \frac{\text{pct1}}{\text{pct2}}$$

#### 24. 是否可以分析不同组在某一细胞类型中的差异基因？

答：可以分析。

#### 25. 差异基因的筛选标准是什么？

答：通常按照  $P < 0.05$ ,  $FC > 2$  来筛选差异基因，但由于某些样本差异基因比较少，可调整为  $FC > 1.2$  或  $1.5$ ,  $P < 0.05$  来筛选差异基因，也是比较认可的。

#### 26. 如何挑选 marker 基因进行验证，与差异基因挑选方式相同吗？

答：由于 Marker 的 P 值都很小，可以按照 Gene Diff 由大到小来筛选；差异基因筛选首先看 P 值，越小说明差异越显著，其次选 FC 较大的基因进行后续验证。

#### 27. 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。

#### 28. 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。

#### 29. 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。

#### 30. 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。

#### 31. 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。

#### 32. 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。

#### 33. 单细胞转录组测序可以进行差异基因及差异基因的 GO 和 KEGG 富集分析吗？

答：可以。目前常规转录组能分析的内容，单细胞转录组都能分析

#### 34. 单细胞测序数据可以与 bulk RNA-seq 联合分析吗，有何好处？

答：可以。做 bulk RNA-seq 的目的有两个：

- 1) 与单细胞转录组测序进行联合分析，计算 pearson 相关性指数，观察两组数据的相似性；
- 2) 由于细胞之间存在基因表达的异质性，因此 bulk RNA-seq 中某些可能很重要的基因的表达水平会被“掩盖”，通过比较单细胞测序的数据和 bulk RNA-seq 数据，可以找到这些被“掩盖”的基因。

#### 35. 关注的基因是融合基因（基因 A+B），为什么单细胞测序只检测到了其中一个基因？

答：单细胞转录组测序是利用 mRNA 3'末端捕获的 polyA 尾来捕获 mRNA，因此检测到的序列也是靠近 3'末端的序列，所以如果其中一个基因距离 3'末端较远，可能检测不到。

#### 36. 为什么基因 A 在 QPCR 结果中显示高表达，而在单细胞测序中表达很低或不表达？

答：QPCR 是基于 bulk 水平的结果，所以个别基因的 QPCR 结果与单细胞结果存在不一致，主要原因有两点

- 1) 单细胞测序只能检测中高丰度的基因，若基因 A 在单个细胞中表达很低，可能检测不到；
- 2) 基因 A 半衰期短，即转录之后马上翻译成蛋白，所以检测不到。

#### 37. 拟时序分析结果怎么看？

答：拟时序分析是通过软件模拟出来的细胞变化轨迹图，结果中分别提供了不同 cluster 及不同样本的细胞在轨迹中的位置，帮助我们判断细胞之间的变化轨迹（如非恶性细胞向肿瘤细胞转变，干细胞分化成不同的细胞类型等）。

### 38. 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。

### 39. 单细胞转录组测序结果是否需要通过 QPCR 验证？

答：一般不建议客户用 QPCR 结果验证单细胞转录组测序，主要因为个别基因半衰期短或在单个细胞中表达量较低，单细胞转录组测序可能检测不到该基因的表达，造成 QPCR 结果与转录组结果不符。

### 40. marker 基因的可视化图（feature plot 图）能否反映基因表达量的高低？

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的高低，建议通过小提琴图（violin）比较不同 cluster（或组别间）基因表达量的变化。

### 41. 10x Genomics 单细胞转录组测序是否可以分析可变剪切？

答：不可以。单细胞转录组测序只能捕获 mRNA 的 3'/5'端进行测序，因此无法分析可变剪切。

### 42. 10x Genomics 单细胞转录组测序是否可以分析 SNP 和 CNV？

答：目前单细胞转录组还不能分析 SNP， 可以进行 CNV 分析。