

# 单细胞 (核) 转录组 常见问题 (FAQ)

Single cell RNA-seq (scRNA-seq)



# 单细胞（核）转录组常见问题（FAQ）

单细胞（核）转录组常见问题（FAQ） .....	2
一、 前期常见问题 .....	7
（一） 技术选择.....	7
1. 10x Genomics、墨卓 MobiNova、BD Rhapsody 几个单细胞平台的区别是什么吗？如何选择？ .....	7
2. 什么样本都可以做单细胞吗？ .....	7
3. 什么情况下选择单细胞核转录组测序？ .....	7
4. FFPE 样本能做单细胞转录测序吗？ .....	7
5. 单细胞转录组测序能检测所有的基因（包含非编码）吗？ .....	8
6. 单细胞转录组测序需要生物学重复吗？ .....	8
7. 临床样本无法一次把所有样本取到，可以分批取样吗？ .....	8
8. 研究的物种没有参考基因组，可以做吗？ .....	8
9. 组织细胞太少，活细胞不够 2 万，能测吗？ .....	8
10. 单细胞转录组测序和 bulk 测序可以一起分析吗？ .....	8
11. 想要和其他组学联合，可以考虑哪些组学？ .....	8
（二） 样本准备.....	9
1. 单细胞悬液制备由客户自主完成还是公司完成？ .....	9

2.	需要多少组织量，血液多少？体液多少？ .....	9
3.	样本如何收集？ .....	9
4.	有多个组学联合分析时，如何取样？ .....	9
二、	关于 10x cellranger/ BD Rhapsody/墨卓 Mobivison 的问题 .....	10
1.	如何知道每个样本捕获到的细胞数？ .....	10
2.	单细胞测序的 reads 能否进行拼接？ .....	10
3.	Sequencing Saturation 需要达到多少才能继续分析？ .....	10
4.	10X Genomics cellranger 结果中 Reads Mapped Antisense to Gene 指标大于 10%，数据是否可用？ .....	10
5.	10X Genomics cellranger 结果中 Fraction Reads in Cells 指标低于 70%报错，对数据分析是否有影响？ .....	11
6.	10x cellranger 结果中的 filtered_feature_bc_matrix 与 raw_feature_bc_matrix 的区别是什么？ .....	11
7.	Cellranger/Mobivision 中的 Barcode Rank Plot 的曲线如何看？ .....	11
8.	什么情况下需要进行 force cell（提细胞数和降细胞数）？ .....	12
三、	关于报告和后续分析问题 .....	12
1.	单细胞转录组测序结果，除了 cellranger/ BD Rhapsody/Mobivison 软 件处理之外，是否还进行其他质控？ .....	12
2.	线粒体质控的标准是什么？ .....	13
3.	根据什么将细胞划分为不同的 cluster，可以调整 cluster 的数目吗？ .....	13

4.	批次处理的问题? .....	13
5.	常用的批次矫正方法.....	13
6.	为什么鉴定到的 cluster 数和细胞类型数不对应? .....	13
7.	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? 14	
8.	tSNE 降维聚类图是根据 marker 基因画出来的吗? .....	14
9.	marker 基因的鉴定原则是什么, top10 是如何筛选的? .....	14
10.	是否可以分析不同组在某一细胞类型中的差异基因? .....	14
11.	差异基因的筛选标准是什么? .....	14
12.	如何挑选 marker 基因进行验证, 与差异基因挑选方式相同吗? .....	15
13.	数据集鉴定的细胞类型一定是正确的吗? .....	15
14.	鉴定细胞类型的 marker 基因公司是否提供? .....	15
15.	为什么同一个 cluster 会鉴定到多种细胞类型? .....	15
16.	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? .....	15
17.	如果进行亚型分析, 能保证划分出很多 cluster 吗? .....	15
18.	细胞亚型分析对细胞数有要求吗? .....	15
19.	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? 16	
20.	单细胞测序数据可以与 bulk RNA-seq 联合分析吗, 有何好处? .....	16

21. 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？ .....	16
22. 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？ .....	16
23. 为什么送的是肿瘤组织却没有检测到肿瘤细胞？ .....	16
24. 单细胞转录组测序结果是否需要通过 QPCR 验证？ .....	16
25. marker 基因的可视化图（feature plot 图）能否反映基因表达量的高低？	17
26. 10X Genomics 单细胞转录组测序是否可以分析可变剪切？ .....	17
27. 10X Genomics 单细胞转录组测序是否可以分析 SNP 和 CNV？ .....	17
四、关于高级分析的内容 .....	17
（一）拟时序的相关问题 .....	17
1. 拟时序分析什么情况下适用？ .....	17
2. 拟时序分析需要认为判定方向吗？如何判断？ .....	17
3. 拟时序分析的轨迹如果预期不符，是否可以进行调整？ .....	18
（二）Scenic 的相关问题.....	18
1. 是否可以按照分组/cluster 进行 regulon 比较？ .....	18
2. 什么情况下按照分组分析 regulon？ .....	18
3. 如果 RAS 和 RSS 的数据不一致，怎么办？ .....	18

4.	活性高是否代表这个基因表达量也高? .....	18
5.	挑选的某个 cluster 里面活性高的 regulon, 是否下游靶基因也是上调的? .....	18
6.	确定了 regulon 之后还能做些什么? .....	18
(三) 细胞间通讯的相关问题.....		19
1.	cellphoneDB 和 cellchat 该如何选择? .....	19
2.	如何筛选关键的受配体基因? .....	19
3.	cellphoneDB 没有收录通路信息, 是否意味着不能关注通路? .....	19
(四) inferCNV 的相关问题 .....		19
1.	是否存在专门的数据库进行 CNV 预测? .....	19
2.	正常细胞要如何选择? .....	20
3.	CNV group 应该怎么用? .....	20

## 一、前期常见问题

### (一) 技术选择

#### 1. 10x Genomics、墨卓 MobiNova、BD Rhapsody 几个单细胞平台的区别是什么吗？

##### 如何选择？

答：几个平台的区别如下，可以对应选择

- 1) 单细胞分离捕获体系不同，前两个平台是微流控体系，BD 平台是微孔板；
- 2) 细胞活性要求不同，前两个要求活性 85%以上，BD 活性要求 65%以上；
- 3) BD 平台因无外界压力作用，对中性粒的捕获理论上更好；
- 4) 发文数量上来看，10x Genomics 平台更多，被客户了解的更多；
- 5) 墨卓 MobiNova 是国产化平台，价格更占优势，目前能达到 10x Genomics 平台的 V2.5 水平。

#### 2. 什么样本都可以做单细胞吗？

答：目前的几个平台对细胞的大小均有要求，一般细胞直径小于 40um 的细胞可以考虑进行单细胞转录组测序，如果细胞太大，建议考虑细胞核转录组或 FLEX 测序。

#### 3. 什么情况下选择单细胞核转录组测序？

答：单细胞核转录组测序，一般建议针对特殊情况采用，如下几种情况可以考虑：

- 1) 前期没考虑做单细胞，直接冻存且无法收集到新鲜样本的；
- 2) 关注的细胞比较大，直径超过 40um；
- 3) 多次解离不成功的样本，胰腺，肠道，脂肪等。

#### 4. FFPE 样本能做单细胞转录测序吗？

答：可以的。10x Genomic 平台推出 FLEX 测序技术可以解决 FFPE 样本做单细胞转录组测序的需求，可以通过制备细胞核悬液进行 FLEX 分析。

#### 5. 单细胞转录组测序能检测所有的基因（包含非编码）吗？

答：常规的单细胞转录组需要通过 polyA 捕获 mRNA，部分 lncRNA 具有 polyA 尾巴，可以被捕获到，其他非编码 RNA，暂时无法捕获。

#### 6. 单细胞转录组测序需要生物学重复吗？

答：建议要做生物学重复。对于小鼠的样本，背景比较单一，可以考虑每组 3 个样本；对于临床样本，个体差异比较大，建议每组 3-5 个甚至更多。

#### 7. 临床样本无法一次把所有样本取到，可以分批取样吗？

答：可以分批提供。单细胞后续分析过程中可以通过软件进行批次矫正。

#### 8. 研究的物种没有参考基因组，可以做吗？

答：不建议。单细胞测序主要在于后期判定细胞类型，如果没有参考基因组及注释文件，后续无法进行基因注释，和细胞类型判断，对分析有较大影响，很难达到老师的目的。

#### 9. 组织细胞太少，活细胞不够 2 万，能测吗？

答：上机前细胞，一般建议活性 85% 以上，细胞总数至少 2 万，是担心后续数据质量和捕获的细胞数太少不够分析。如果老师样本很珍贵，难取，可以根据情况考虑要不要上机进行油包水，但整个过程造成的任何风险，需要客户自行承担。

#### 10. 单细胞转录组测序和 bulk 测序可以一起分析吗？

答：可以的

#### 11. 想要和其他组学联合，可以考虑哪些组学？

答：组学联合可以考虑一下

- 1) DNA 层面：可以联合 scATAC-seq, WGS, WES
- 2) 转录层面：可以联合 bulkRNA-seq, ST-seq (空间转录组)
- 3) 蛋白层面：DIA、label-free, TMT 检测 (至少每组 3 重复，临床至少 5 对以上)



4) 代谢层面: GC-MS/LC-MS (可以通过通路等进行联合分析), 或可以联合空间代谢 (建议有目标细胞和目标代谢物的情况下)

## (二) 样本准备

### 1. 单细胞悬液制备由客户自主完成还是公司完成?

答: 解离可以有两种方法:

- 1) 客户自己完成解离, 公司上门进行油包水;
- 2) 样本寄到公司, 由公司解离并进行油包水。

### 2. 需要多少组织量, 血液多少? 体液多少?

答: 组织量一般 200mg 左右 (黄豆粒大小), 血液 3-5ml (不分选特定细胞情况, 分选特殊处理); 体液 (视情况 10-50ml 不等)。

### 3. 样本如何收集?

答: 样本收集存在几种情况

#### 1) 组织样本收集:

方法一: 保存液保存, 四度运输, 48h 内到公司;

方法二: 梯度冻存 (放梯度降温盒,  $-80^{\circ}\text{C}$  过夜或转液氮), 干冰寄送公司;

方法三: 直接液氮/ $-80^{\circ}\text{C}$  冻存, 干冰寄送公司。仅适用于 FLEX 或单细胞核转录组测序。

#### 2) 血液样本收集:

方法一: EDTA 抗凝管保存,  $4^{\circ}\text{C}$  快递到公司, 尽量当天到样;

方法二: 分选 PBMC/白细胞 (包含中性粒), 冻存后干冰寄送公司。

### 4. 有多个组学联合分析时, 如何取样?

答: 具体要看和哪些组学联合。建议不同组学样本按照各自的保存方法, 将组织分开后, 分别保存即可。

## 二、关于 10x cellranger/ BD Rhapsody/墨卓 Mobivison 的问题

### 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 质控软件，BD 平台有 scanner，墨卓的 Mobivison 结果，均可获得每个样本捕获到的细胞数，每个细胞的平均 reads 数和每个细胞的中位基因数。

### 2. 单细胞测序的 reads 能否进行拼接？

答：不能，由于测序之前打断的序列，除了 3' 端（3' 转录组测序），或 5' 端（5' 转录组测序）包含 barcode 和 UMI 信息之外，其余的序列并不包含这两个标签序列，拼接后的结果，并不能有效获得序列的来源，因此，单细胞转录组测序结果，并不会进行拼接。

### 3. Sequencing Saturation 需要达到多少才能继续分析？

答：目前 V3 版本的 cellranger，没有要求饱和度要达到多少指标，但是一般情况下，为了保证每个样本获得足够的数据进行下游分析，建议每个样本的测序饱和度尽量达到 40% 以上为最佳，如果低于 35%，建议加大测序数据量为最佳。

饱和度高一点的原因：

- 1) 差异基因会增多，marker 的筛选中部分基因的 PCT 值低；
- 2) 等丰度基因的检出会更高；

### 4. 10X Genomics cellranger 结果中 Reads Mapped Antisense to Gene 指标大于 10%，数据是否可用？

答：“Reads Mapped Antisense to Gene”是指比对到反义链上的 reads 的比例，这样的 reads 是不纳入后期分析的。通常情况下，对于单细胞转录组数据，这个指标基本不会超过 10%。但对于单细胞核转录组测序数据，这个指标一般在 25% 左右也是可接受的。

## 5. 10X Genomics cellranger 结果中 Fraction Reads in Cells 指标低于 70%报错，对数据分析是否有影响？

答：“Fraction Reads in Cells”是指比对到细胞中 reads 的比例。这个指标一般不需要太注意。这个指标主要和样本的背景，如细胞碎片，杂质过高，都会引起这个指标的偏低，根据我们的项目经验，如果是碎片引起的指标低，对于后续的分析，如果不低于 40%，几乎是没有任何影响的。

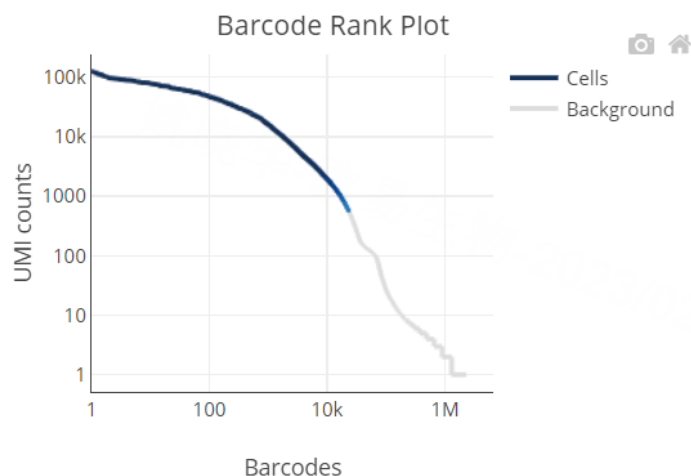
此外，针对做单细胞核测序的项目，由于解离过程造成的碎片比例普遍偏高，Fraction Reads in Cells 的指标部分会在 50%左右，对后续分析也几乎没有影响的，高于 40%，建议正常分析即可。

## 6. 10x cellranger 结果中的 filtered\_feature\_bc\_matrix 与 raw\_feature\_bc\_matrix 的区别是什么？

答：这两个 matrix 文件，都是 cellranger 输出的矩阵信息，包含 feature (Gene)，UMI 和 Barcode。其中“Raw\_feature\_bc\_matrix”文件是未过滤的所有的 barcode 信息（包含非细胞信息），而“filtered\_feature\_bc\_matrix”文件是针对 cellranger 捕获的细胞（即认为是完整的细胞）所包含的信息。

## 7. Cellranger/Mobivision 中的 Barcode Rank Plot 的曲线如何看？

## Cells ?



答：上图的曲线是每个细胞中 UMI counts 的分布曲线，横坐标按照 UMI 由高到低，对 barcode（细胞）进行排列，通常认为是细胞的，曲线为深蓝色，当细胞中 UMI 数量突然急剧下降到很低，说明此时的 barcode 对应的很可能不是细胞（通常所说的空包，或低质量细胞），曲线则用灰色表示。

### 8. 什么情况下需要进行 force cell（提细胞数和降细胞数）？

答： 以下情况下，可以考虑对 cellranger 或 Mobivison 结果进行 force cell，BD 平台暂不支持。

- 1) 各样本之间的细胞数差别太大，导致部分样本细胞丢失严重，计算比例时存在偏差
- 2) UMI 值低于 1000 甚至 500

## 三、关于报告和后续分析问题

### 1. 单细胞转录组测序结果,除了 cellranger/ BD Rhapsody/Mobivison 软件处理之外,是否还进行其他质控?

答：会的。还会通过 seurat 对每个细胞中 UMI（转录本）数，基因数和线粒体基因数甚至

血红蛋白进行质控，从而过滤掉样本中低质量的细胞。近年由于单细胞数据文章发表比较严格，部分审稿人建议在上述处理之后，还需要用部分双细胞处理的软件进行处理，因此会针对部分样本进行 Doubletfinder 处理双细胞。

## 2. 线粒体质控的标准是什么？

答：线粒体的质控，目前并没有严格的标准。目前主流的线粒体质控标准在 5%-30% 之间浮动，如遇胃、肝脏、肾脏、心脏等样本，因其样本的特殊性，可视情况调到 50% 左右。

## 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在同一个 cluster，cluster 分出来的多少，和参数（resolution）设置有关，必要情况下，可以针对 resolution 参数调整，从而实现 cluster 分群数目的调整。

## 4. 批次处理的问题？

答：如何判断是否存在批次效应：

- 1) 两个组完全分开，甚至是同一个 cluster 的不同分组细胞完全不重合
- 2) 同一个组的重复样本完全分开

如何判断过度矫正：

- 1) 一般总的细胞聚类做批次矫正，对于是否造成过度，并没有严格的标准作为参考。但是根据项目经验，可能出现的表现：1. 是 UMAP 会缩在一起，各 cluster 没有明显的边界；
2. 筛选不到明显的 marker 基因。
- 2) 亚群一般不进行批次矫正，原因在于亚群可能会校正太过，造成没有差异。

## 5. 常用的批次矫正方法

答：MNN, Harmony, CCA, RPCA, 不矫正批次为 PCA 分析

## 6. 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。

#### 7. 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。

#### 8. 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。

#### 9. 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

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

#### 10. 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。

#### 11. 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。

#### 12. 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。

#### 13. 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 (已发表文章中涉及的) 来辅助判断细胞类型。

#### 14. 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。

#### 15. 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, 只要数量不多, 可以忽略。

#### 16. 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, 如果客户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。

#### 17. 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。

#### 18. 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。

### 19. 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

### 20. 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。与 bulk RNA-seq 联合分析的目的有两个：

- 1) 与单细胞转录组测序进行联合分析,计算 person 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。

### 21. 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。

### 22. 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 水平检测的结果,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。

### 23. 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。

### 24. 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。

## 25. 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。

## 26. 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。

## 27. 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。

# 四、关于高级分析的内容

## (一) 拟时序的相关问题

### 1. 拟时序分析什么情况下适用?

答: 以下几种情况下可以考虑应用

- 1) 具有明确分化关系的细胞之间 (如干细胞分化、T 细胞活化等)
- 2) 针对某些组间差异很大的细胞亚型 (如处理之后出现的新亚型, 如何转化而来)

### 2. 拟时序分析需要认为判定方向吗? 如何判断?

答: 需要。软件自动预测的方向, 可能和实际生物学意义不符。判断方向的原则参考如下

- 1) 先验知识的积累 (常用)
- 2) 正常疾病的样本, 一般会指定正常组定位的细胞为起点 (常用)
- 3) 根据不同分支基因的功能/通路富集情况进行推测 (不常用)

4) 根据特定基因的表达, 如凋亡、初始相关基因定位的分支, 认为是起点 (不常用)

### 3. 拟时序分析的轨迹如果预期不符, 是否可以进行调整?

答: 是可以的, 必要情况下, 可以通过改变拟时序分析过程的高变基因筛选, 进行微调。

## (二) Scenic 的相关问题

### 1. 是否可以按照分组/cluster 进行 regulon 比较?

答: 可以的。

### 2. 什么情况下按照分组分析 regulon?

答: 理论上什么情况都可以,

1) 一般针对某种细胞的 subcluster 分析时, 如果不同的 subcluster 在组间差异比较大的情况, 建议按照 cluster 之间的差异 regulon 来分析

2) 如果针对很重要的细胞, 但是细胞数很少, 做 subcluster 效果不好的情况, 可以考虑直接做组间的差异 regulon 分析

3) 如果某些细胞类型在组间差异不大, subcluster 看不到想要的信息, 也可以考虑直接分析组间的 regulon

### 3. 如果 RAS 和 RSS 的数据不一致, 怎么办?

答: RAS 和 RSS 侧重点不一样, 通常情况下, RAS 高不代表 RSS 高, 因此如果数据不一致, 可以重点参考 RAS 指标。

### 4. 活性高是否代表这个基因表达量也高?

答: 不一定

### 5. 挑选的某个 cluster 里面活性高的 regulon, 是否下游靶基因也是上调的?

答: 不一定, 有些转录因子可以抑制靶基因的转录, 造成靶基因下调

### 6. 确定了 regulon 之后还能做些什么?

答：

- 1) 分析层面，可以针对 CSI 模块的 regulon 或重点关注的 regulon 及靶基因做富集分析，这样不仅可以关注到具体的基因，还能靶向相关的 pathway
- 2) 验证层面，可以对相关 TF 或靶基因进行免疫荧光或 KO 等实验验证

### (三) 细胞间通讯的相关问题

#### 1. cellphoneDB 和 cellchat 该如何选择？

答：两个方法都能用，但目前 cellchat 应用的广泛一些。两个方法主要有几点不同

- 1) cellphoneDB 收录的受配体对相对较多，cellchat 相对较少
- 2) cellchat 可视化的图更多样
- 3) cellchat 具有通路的信息

#### 2. 如何筛选关键的受配体基因？

答：如果没有关注的基因，可参考如下原则

- 1) 先锁定细胞类型
- 2) 筛选  $P < 0.05$  的关系对
- 3) 在 2) 的基础上，筛选 prob(通讯概率, cellchat) 或 express(L-R 的表达量, cellphoneDB) 由大到小

#### 3. cellphoneDB 没有收录通路信息，是否意味着不能关注通路？

答：cellphoneDB 本身没有通路信息，但如果老师想要关注一些细胞类型之间的受体配体信息，可以考虑针对  $P < 0.05$  的关系对，或特定关注的关系对，进行 GO/KEGG 富集。

### (四) inferCNV 的相关问题

#### 1. 是否存在专门的数据库进行 CNV 预测？

答：目前没有数据库，一般是基本项目中正常的细胞，作为对照进行 CNV 分析。

## 2. 正常细胞要如何选择?

答: 如果是上皮来源的肿瘤, 可以指定除了上皮以外的其他细胞做正常细胞; 如果是淋巴瘤细胞, 可以指定非淋巴细胞作为正常细胞。

## 3. CNV group 应该怎么用?

答: CNV group 是 inferCNV 软件通过判定肿瘤细胞的 CNV 水平进行的分组, 一般情况下, 如果想要探究是否肿瘤具有不同的恶性程度, 及不同恶性细胞之间的特征基因等信息, 可以考虑用 CNV group 进行分类。此外, 也可以通过原始聚类的 cluster 进行比较分析, 不用 CNV group, 也是可行的。