

## Seurat分析单细胞转录组数据

### 1. 环境准备

1.1 读取R包: `library()`

### 2. 数据加载与创建Seurat对象

2.1 读取10X 数据: `Read10X()`  
2.2 创建 Seurat 对象: `CreateSeuratObject()`

### 3. 质量控制与过滤

3.1 计算线粒体基因百分比: `PercentageFeatureSet()`  
3.2 可视化QC指标: `VlnPlot()`、`FeatureScatter()`  
3.3 根据可视化结果设定过滤阈值并过滤细胞: `subset()`

### 4. 数据标准化

4.1 用`NormalizeData`函数可对单细胞counts矩阵数据归一化处理:  
`NormalizeData(MEs)`

### 5. 特征选择

5.1 识别在细胞间表达变化大的基因: `FindVariableFeatures()`  
5.2 可视化高变基因: `VariableFeaturePlot()`、`LabelPoints()`、`plot`

### 6. 数据缩放与线性降维

6.1 对所有基因缩放: `ScaleData()`  
6.2 主成分分析 (PCA): `RunPCA()`、`enrichGO()`  
6.3 可视化 PCA 结果:  
`VizDimLoadings()`\`DimPlot()`\`DimHeatmap()`\`ElbowPlot()`  
6.4 根据肘点图确定使用前 10 个 PCs

### 7. 细胞聚类

7.1 构建K近邻图(KNN)和共享最近邻图(SNN): `FindNeighbors()`  
7.2 在SNN图上应用聚类算法(Leiden): `FindClusters()`  
7.3 查看前几个细胞的聚类ID: `head(Idents(scRNA))`  
7.4 查看各簇细胞数量: `table(scRNA@meta.data$seurat_clusters)`

### 8. 非线性降维与可视化

8.1 Umap降维: `RunUMAP()`  
8.2 t-SNE降维: `RunTSNE()`  
8.3 可视化聚类结果: `DimPlot()`

### 9. 寻找差异表达Marker基因

9.1 为每个簇寻找相对于所有其他簇的 marker 基因: `FindAllMarkers()`  
9.2 可视化特定 marker 基因的表达: `FeaturePlot()`、`VlnPlot()`

### 10. 细胞类型注释 (手动注释)

10.1 查阅文献或数据库 (CellMarker) 确定每个簇最可能对应的细胞类型: `levels()`、`RenameIdents()`  
10.2 可视化带注释标签的 UMAP: `DimPlot()`

## 常见问题解析

### 1. 如何设置合理的质控阈值？

①阈值设置过松导致低质量细胞污染下游分析，阈值设置过严导致损失过多真实细胞类型，影响生物学发现；②绘制VlnPlot或FeatureScatter查看各指标的分布及相互关系（`nFeature_RNA` vs `percent.mt`），并寻找离群点，如极高`percent.mt`的死细胞、极低`nFeature_RNA/nCount_RNA`的空液滴或碎片、极高`nFeature_RNA`的双细胞；③可参考类似组织或实验条件的已发表论文使用的阈值；④如果整合多个样本，可能需要为每个样本单独评估和设置阈值，尝试不同阈值组合，观察过滤后细胞数量变化和后续聚类/标记基因的合理性。

### 2. 线粒体基因比例高一定是死细胞吗？

①某些具有高代谢活性的细胞类型（如心肌细胞、肝细胞）天然具有较高的线粒体基因表达；②`percent.mt`阈值设置不合理可能过滤掉这些真实的生物学细胞类型；③结合其他指标（如`nFeature_RNA`、`nCount_RNA`）和细胞类型特异性标记基因进行综合判断，避免对所有细胞类型使用统一的、过于严格的`percent.mt`阈值。

### 3. 如何设置合适的分辨率参数（`resolution`）？

①分辨率过低导致生物学上不同的细胞类型被聚在一起，分辨率过高导致单个细胞类型被拆分成多个亚群（可能是技术噪音或过度解读），Seurat的聚类算法基于共享最近邻图和模块化优化算法（如Louvain、Leiden），`resolution`参数直接控制聚类的粒度，值越大，得到的簇越多、越小。②最佳分辨率高度依赖于数据的生物学复杂度和分析目标（想看清大类还是细分亚型），可尝试一系列分辨率（如0.2、0.4、0.6、0.8、1.0）进行不断调整；③可对每个分辨率下的聚类结果绘制UMAP图，观察细胞群的分裂/合并情况。④使用已知的细胞类型标记基因（FeaturePlot、VlnPlot、DotPlot）检查每个簇的身份，确定在某个分辨率下，期望的细胞类型是否被清晰地分离出来，并且没有明显的不合理分裂。