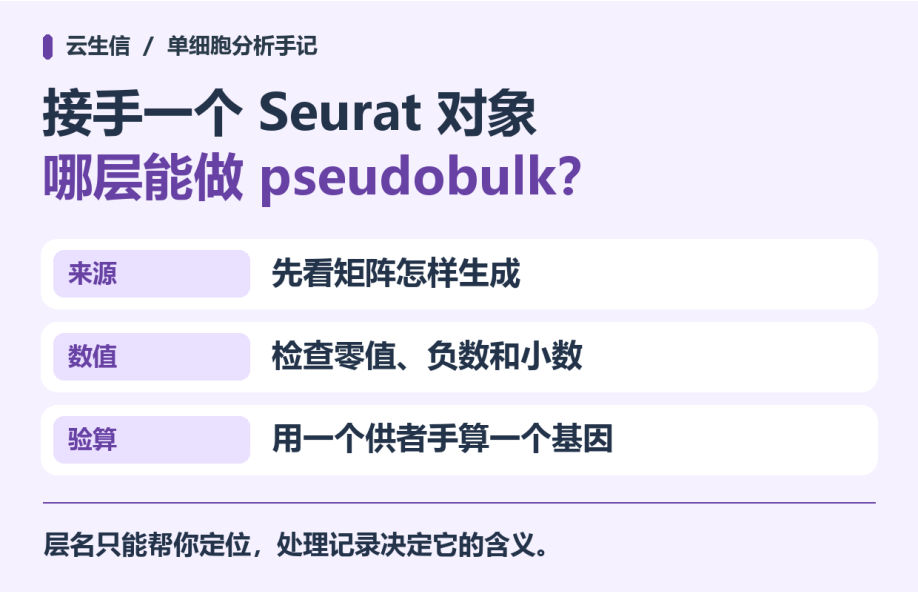
接手一个 Seurat 对象，怎样知道哪一层还能用来做 pseudobulk?

从 Seurat 对象逐层导出基因与条码，核对来源、整数计数和基本单位，用一个供者的一格原始计数验算 pseudobulk 汇总。



运行教学示例

Python 3.10 以上，含 numpy、scipy；安装命令见 requirements.txt。已有 Seurat v5 的读者可用包内 R 导出脚本。

```
python run.py --input example --out results
```

在解压后的资料包根目录运行。example 是教学输入，results 保存本次输出，expected 保存可供核对的参考结果。

输入字段

layers.tsv 每行一层，记录名称、来源类型、来源说明和四个对应输入位置中的矩阵及基因、条码文件。

.mtx / genes.tsv / cells.tsv Matrix Market 格式的基因 × 细胞稀疏矩阵。表头 gene_id 或 cell_id。

metadata.tsv 每个 cell_id 一个 sample_id 和 cell_type，细胞键须全局唯一。

核对输出，再换自己的数据

layer_audit.tsv 每层的维度、非零数、小数、负数、来源与处置。

pseudobulk.tsv 选中的原始计数按生物样本与细胞类型求和。

manual_check.tsv 第一项汇总的逐条码原始值以及组内合计，用来验算。

用示例验算

1 两个有原始来源记录且条码不重叠的计数层进入汇总。

2 故意叫 counts_backup 的整数层因来源未知而待核实。

3 P01 的 T 细胞中 G1 为 1 + 2 = 3，P02 对应值为 4。

换用真实数据

用 export_seurat.R 从 Seurat v5 的 RNA assay 导出原始层名和条码。对照建库输出和上游脚本填写 provenance、source_note，再把 meta 中的样本与细胞类型映射到 metadata.tsv。真实数据只为目标 RNA 层运行，聚合时另存参考基因清单。

矩阵数值检查不能证明计数来源。此示例面向普通 UMI 整数计数，去污染、期望计数及其他实验读出需要按下游模型重新判定。

方法资料

Seurat v5 基本操作

SeuratObject 的 Layers 与 LayerData