线粒体比例统一卡在 10%，会从你的组织里删掉哪类细胞?

直接从稀疏原始 RNA 计数求出每颗细胞的线粒体比例，按样本和已有谱系比较三档过滤线，逐个找回受影响细胞并核查标记与复杂度。



运行教学示例

Python 3.10 或以上，需 numpy、scipy、matplotlib。先用 python -m pip install -r requirements.txt 安装。

```
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

counts/、mt_features.txt 原始 Gene Expression MEX 和按实际参考确定的线粒体基因 ID 清单。

cells.tsv、settings.json 按唯一条码给出样本与已有谱系，设置百分比上限、检出基因下限以及关注标记。

核对输出，再换自己的数据

cells_qc.tsv、joint-qc.png 每颗细胞的分子、线粒体比例、检出基因和标记，及按样本绘制的联合分布。

thresholds.tsv、above_cutoff.tsv 各谱系每档条件命中、重叠、保留人数，以及超线条码。

用示例验算

1 S1_M1 有 100 个总 UMI，其中 20 个为线粒体分子，比例为 20%。

2 S1 三个模拟肌细胞在 10% 上限下有两个越线，三者都通过教学基因数下限。

3 教学零总量条码 S2_Z0 的比例为缺失，不进入阈值可评估分母。

换用真实数据

用实际参考的线粒体染色体注释导出特征 ID，保留每个样本过滤前的 RNA 原始计数。把初步谱系、代表性标记和样本列附上，先扫若干候选上限，再查受影响细胞的复杂度和标记。

上限的命中说明细胞值得复核，不能单独判定细胞死亡或确认某个谱系。模拟矩阵仅保留少数基因，教学检出基因数下限不能用于完整转录组。

方法资料

Bioconductor 单细胞质量控制

scuttle 单细胞质控工具