同一份组织分两管建库，怎么判断两管的结果是否一致?

把同一组织分管建库的重复性拆成两项检查，先看各谱系有多少细胞被捕获，再在两管共同覆盖的谱系内汇总原始计数，检查相关图中的偏离基因。



运行教学示例

Python 3.10 或以上，需 numpy、scipy、matplotlib。练习从原始 10x MEX 和细胞标签开始。

```
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/、cells.tsv 原始 RNA 稀疏矩阵，条码的取材标本、分管编号和共同口径的谱系标签。

replicates.tsv 一对真实同标本的分管、各群纳入的最低细胞数、双管合计的最低基因 UMI。

核对输出，再换自己的数据

composition.tsv、coverage_correlation.tsv 比较两管各类细胞数、占比、百分点差，以及共同群的覆盖和表达相关性。

gene_deviations.tsv、agreement.png 列出每群聚合计数、归一化表达、偏离方向，并绘制两管散点图。

用示例验算

1 A 管 20 颗、B 管 18 颗，T 比例为 60% 和约 33.3%，相差约 −26.7 个百分点。

2 NK 在 A 有 2 颗、在 B 为 0，不输出虚构的表达相关。

3 T 群 FOS 的分管聚合计数为 9、47，归一化后的 B 减 A 约为 +3.289。

换用真实数据

采用同一参考、计数口径与注释层级，给每颗条码加对应管号和标本号。先检查全体组成，再比较共同谱系，并以多个独立标本重复同样的核查。

单个标本分成两管不能提供独立生物学重复。跨基因相关性描述谱系内轮廓，不是等效性检验；局部基因偏差和稀有谱系覆盖需要另读。

方法资料

Bioconductor 的样本和细胞类型聚合示例

Seurat 的逐样本 pseudobulk 示例