单细胞数据太稀疏，怎样用 metacell 找共表达模块，又保住供者差异?

把八位供者的真实单核细胞在供者和刺激条件内聚合，用 hdWGCNA 找到实际共表达模块，随后回到原条码核查成员复用、供者变化及测序深度。



运行教学示例

常规复核用 Python 3.10 或以上与 requirements.txt。重新建网用 R、hdWGCNA 0.4.12 和 Seurat，运行命令见数据来源。

```
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

cells.tsv、counts.tsv、prespecified_markers.txt 真实细胞元数据、257 个基因的原始计数面板和预先声明的干扰素基因。

fit_hdwgcna.R、example/fit/ 从细胞 PCA、分层聚合和计数归一化到有符号网络、模块与原条码映射。

核对输出，再换自己的数据

barcode_reuse.tsv、membership_totals.tsv、module_inventory.tsv 审计条码归属、复用量、模块基因数及预设基因覆盖。

paired_donors.tsv、depth_correlations.tsv、donors_and_depth.png 按供者成对计算条件差，并与原始测序深度对照。

用示例验算

1 1,152 个原条码进入步骤，148 个 metacell 每个采用 12 颗，共覆盖 997 个不同条码。

2 所选 94 基因模块包含 22 个预设干扰素标记，八位供者刺激条件中位数均高于对照。

3 模块均分与 metacell 原始整库 UMI 总和的秩相关约 0.409，分条件仍有相关。

换用真实数据

在同一细胞类型、供者和实验条件内聚合，并单独报告成员复用与未覆盖条码。将网络模块回到每位供者看方向，再用新的样本检验模块。

小面板有意保留干扰素标记，适用于解释过程。分组中位数和相关是描述量，不能把 metacell 数当成独立供者数。

方法资料

hdWGCNA 作者教程

Kang 等原始实验