给 PBMC 手工注释，第一张 DotPlot 应该放哪些基因?

从一位健康供者的公开 PBMC3k 原计数画一张有分母的标记点图，按条码查看成套标记的共检出，再用公开参考标签复核手工判断。



运行教学示例

Python 3.10 或以上及 requirements.txt。离线检查不需下载全矩阵。重取原计数时另需 h5py 与 scipy。

```
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

panel_counts.tsv、panels.json 2,638 个已处理条码的 17 项真实原计数、全库总量、不带名字的簇号与可编辑标记组。

reference_labels.tsv、rebuild_public.py 公开原注释留作判读后对照，附从两个校验过的 H5AD 抽取数据的脚本。

核对输出，再换自己的数据

dot_values.tsv、pbmc_dotplot.png 每个点的分子、分母、平均标准化计数、用作颜色的数值和点图。

barcode_evidence.tsv、cluster_panel_evidence.tsv、T_NK_overlap.tsv 同一条码内的多基因共检出、逐簇的证据及 T 与杀伤面板的交叉统计。

用示例验算

1 C2 的 MS4A1 是 294/342，CD79A 是 320/342，三种 B 标记中至少两种为 337/342。

2 C3 的 316 颗中有 89 颗、C4 的 154 颗中有 22 颗同时满足 T 与杀伤面板。

3 小簇 C7 的 15 颗全部同时检出 PPBP 和 PF4。

换用真实数据

将真实注释对照表留到最后才打开，先凭成组表达、邻近谱系与同一条码共表达给出大类和待查范围，多供者应做供者分面。

单个健康供者的已有标签仅供对照。检出至少两种是教学核对规则；原表达无法单独确认所有双细胞或稀有细胞身份。

方法资料

Seurat PBMC 注释示例

10x PBMC3k 公开数据