手上只有普通基因表达矩阵，还缺哪些输入才能做 RNA velocity?

用公开小鼠胰腺 3,696 个条码比对两层原始计数，修复刻意打乱的行顺序，画 Cpe 和 Neurog3 的原始相图，判断哪些基因有足够双层观察。



运行教学示例

Python 3.10 或以上及 requirements.txt，默认无需联网；提取脚本可从公开 H5AD 重建。

```
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

spliced.tsv、unspliced.tsv、cell_metadata.tsv 五个基因的两层原始计数，公开细胞大类与统一条码。

extract_pancreas.py、source.json 源 H5AD 的哈希、原始维度、提取代码和教学打乱说明。

核对输出，再换自己的数据

alignment.tsv、gene_layer_support.tsv 对齐前后同位置条码、每个基因的各层非零与双层共检出。

barcode_phase.tsv、support_by_type.tsv、raw_phase_portraits.png 逐条码原始读数、分类型覆盖以及 Cpe 与 Neurog3 的对照相图。

用示例验算

1 两表原顺序匹配 0 行，按键对齐后恢复 3,696 颗。

2 Cpe 已剪接非零 2,130、未剪接非零 1,454，两者同颗非零 1,376。

3 Neurog3 已剪接非零 1,569、未剪接非零 7，其中两者同颗非零 6。

换用真实数据

先查对象的两层、参考注释、条码和基因标识；如需从比对文件重计，请按实验化学与工具要求检查带条码和 UMI 的比对及参考版本。

本文做输入核对与原始相图，未拟合全基因组速度或迁移方向。五基因子集不够替代完整矩阵、邻域及模型诊断。

方法资料

RNA velocity 原始论文

scVelo 胰腺官方示例

velocyto 比对文件与计数命令