单细胞 ATAC 里片段数很高、TSS 富集却低，这些细胞该留吗?

从公开的人 PBMC 片段文件取 160 颗细胞，逐条重算片段数、TSS 富集和核小体比例，并回看 MS4A1 与 LYZ 的覆盖。



运行教学示例

Python 3.11 或以上，pip install -r requirements.txt。直接使用包内片段可离线重算。

```
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、fragments.tsv.gz 160 个通过原流程判定的真实条码、完整片段记录与其原始深度档。

tss.tsv.gz、file_hashes.txt 主染色体的 hg38 TSS 坐标以及三个源文件 SHA-256。

核对输出，再换自己的数据

cell_qc.tsv、filter_groups.tsv 每个条码的中心与侧翼切口、片段长度以及不同阈值组的汇总。

locus_coverage.tsv、quality_and_locus.png MS4A1、LYZ 的固定窗口覆盖、深度与 TSS 联合图。

用示例验算

1 两项均过线 155 颗，高深度低 TSS 为 3 颗。

2 后者的片段中位数 154,814，核小体比例中位数约 5.32。

3 其余两颗深度不足，其中一颗 TSS 超过 4。

换用真实数据

每个独立文库单独编制条码前缀和深度图，保持组装、注释和计算窗口一致，再对改变过滤线的条码复核。

四档取样让读者能看到边缘情况。该取样和教学阈值不能估计原标本的不合格率。

方法资料

Signac 的 PBMC 样本和原片段

UCSC hg38 RefSeq Curated 注释