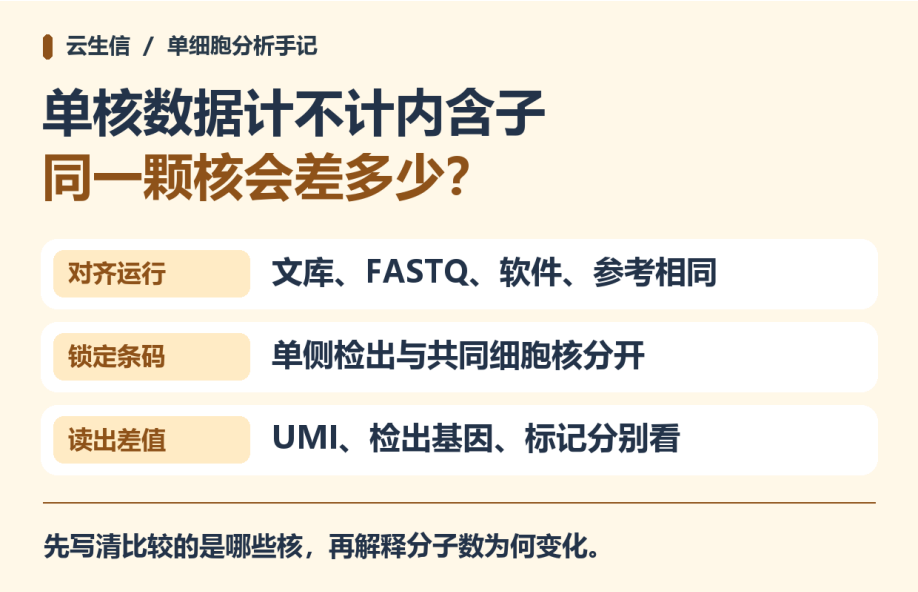
同一份单核 RNA 数据，计不计内含子会让哪些结果变化？

将同一单核文库的两套计数按特征和条码重新对齐，分别检查呼叫名单变化、共同细胞核内的分子增量，以及同一个标记基因是否获得额外支持。



运行教学示例

Python 3.10 或以上，需 numpy 和 scipy。运行前可用 python -m pip install -r requirements.txt 安装依赖。

```
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

exonic/、inclusive/ 同一文库两套标准 10x filtered MEX 矩阵，支持 gzip 压缩。

runs.json、markers.txt 两次运行的元数据与内含子开关。标记文件每行一个参考中的 RNA 特征 ID。

核对输出，再换自己的数据

calls.tsv 两份上游细胞识别名单的并集及两侧是否出现。

paired_nuclei.tsv 每个共同核的 UMI、检出基因数与配对差值。

paired_markers.tsv、summary.json 指定基因在每颗共同核的计数变化，以及名单覆盖和差值中位数。

用示例验算

1 两套名单各 4 颗核，交集 3 颗，并集 5 颗。

2 共有条码 N1-1 的总 UMI 从 5 增为 10，检出基因从 2 增为 3。

3 在同一颗核中，教学标记 DEMO_A 的计数从 3 增至 6。

换用真实数据

用相同 FASTQ、相同参考及软件版本生成两套计数，将 Gene Expression 矩阵放入对应目录。用原始特征 ID 填标记文件，保留原始条码和各自的细胞识别输出，先比较一份文库，再逐文库汇总。

配对比较仅覆盖两侧共同通过识别的核。分子计数差包含比对归属及 UMI 折叠规则的影响，不能直接解释为每个基因真实产生的内含子分子数。

方法资料

Cell Ranger 基因表达计数算法

Cell Ranger 版本与内含子默认设置