肿瘤与血液的 TCR 要算作同一克隆，需要哪些序列信息一致?

用真实血液与肿瘤 TCR 链表，在同一患者内逐步放宽 V/J、核苷酸、氨基酸和配对链规则，找出新出现的共享克隆，并检查它们到底差在哪条链。



运行教学示例

基础复核用 Python 3.10 或以上及 requirements.txt。重新导出完整 H5MU 另需 awkward 和 h5py。

```
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、chains.tsv 公开三千条码的组织与患者信息、7,544 行受体注释和真实核苷酸及氨基酸。

extract_wu.py、source.json 完整教学对象、固定校验值与 AIRR 字段的导出说明。

核对输出，再换自己的数据

rules_by_donor.tsv、rule_totals.tsv 每种定义每个患者的分母、共享键及落入共享键的细胞数。

shared_sequences.tsv、cell_keys.tsv、shared_by_definition.png 共享键、两种组织的条码和每颗细胞在每种规则中的去向。

用示例验算

1 仅四位患者的血与肿瘤两处都有数据；严格可判定的血液条码 103 个、肿瘤条码 159 个。

2 四种规则患者内共享键合计依次为 10、10、11、18 个。

3 Renal1 的 CASSEYWAKTQYF 在一颗血细胞及两颗肿瘤细胞中出现，但两处 α 链不同。

换用真实数据

先确认相同患者与样本组织，用一致的 AIRR 字段和链过滤定义，先固定条码再放宽键，最后才纳入新增有缺链或多链的条码。

数据为含 TCR 的教学子集。序列共享不确定祖先关系、抗原特异性或组织间迁移方向。

方法资料

Wu 等原始研究

Scirpy 肿瘤免疫受体教程