一组测到的 T 细胞多一倍，如何比较哪组的克隆更多样？

对真实肿瘤和癌旁 TCR 统一配对链定义，在每名患者每种组织内重复不放回抽样，同时比较克隆丰富度和对扩增更敏感的有效克隆数。



运行教学示例

Python 3.10 或以上与 requirements.txt；从原始 H5MU 重新提取另需 h5py 和 awkward。

```
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

paired_chains.tsv 14 位患者肿瘤与癌旁的真实配对 TCR 氨基酸序列、条码与组织标签。

extract_wu.py、prepare_clones.py 从公开教学对象重建同一链定义和入组细胞。

核对输出，再换自己的数据

sample_diversity.tsv、draws.tsv 原始丰富度、每次抽样的三种有效克隆数及条件分位数。

paired_donors.tsv、rarefied_pairs.png 可比较的患者、配对差值及低覆盖缺失。

用示例验算

1 深度 30、每份 300 次，不足深度的样本标记缺失。28 份中 18 份可抽，6 位患者两侧都足够。

2 Lung3 原始癌旁与肿瘤分别有 116、55 个克隆，稀释后中位数 30 与 29。

3 Renal2 同深度中位数为癌旁 28、肿瘤 18；肿瘤的二阶有效克隆数中位数约 9.18。

换用真实数据

统一链定义，再在同一患者、同一组织内不放回抽样。报告留存患者、三项深度敏感性和扩增主导程度。

相同细胞数不保证同等覆盖度；缺失低深度组织会改变入组人群，重复抽样不能充当生物学重复。

方法资料

Wu 等的 TCR 原始研究

Chao 与 Jost 的覆盖度方法