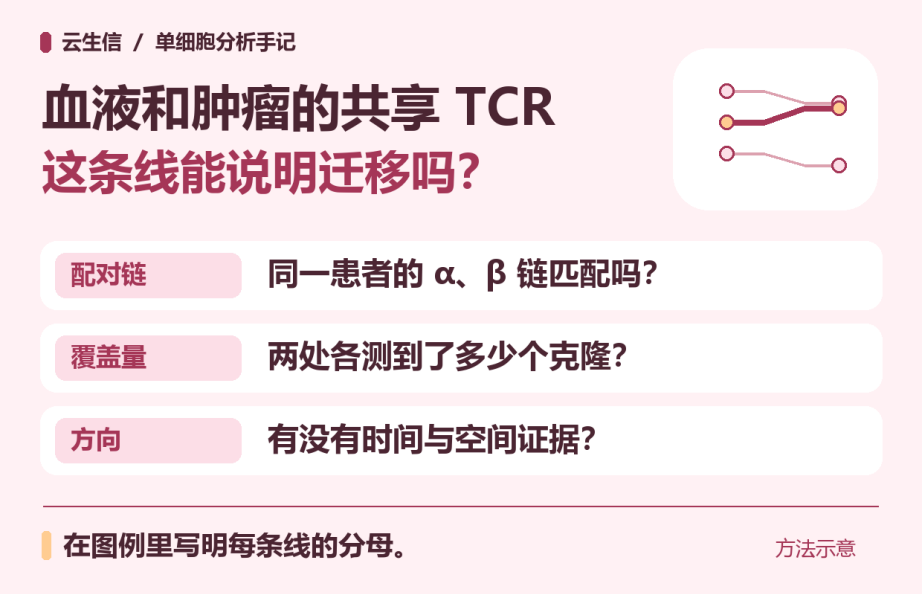
配对血液与肿瘤 TCR 克隆

从示例数据跑到第一张结果表

从经筛选的受体链和表达谱 T 细胞名单出发，先核对有效配对链的覆盖，再按患者比较两处克隆。即使文库复用条码，相同序列也只在同一位患者内部比较。



准备 R 环境

仅使用 R 自带的软件包。 示例已在 R 4.3.3 运行。

完整解压资料包，在 run.R 所在文件夹打开终端，运行以下命令。

```
Rscript run.R example results
```

若使用 RStudio，将工作目录设为同一文件夹，再在控制台输入下面这一行。

```
source("run.R", encoding="UTF-8")
```

打开 results 文件夹查看输出。expected 文件夹中有示例的参考结果，示例数字均来自教学模拟。

逐步复算示例

在示例已经跑完的同一个资料包目录中打开 R。下面两段代码可依次粘贴，也可以执行包内的“读表练习.R”。

读第一张表

```
z <- read.delim("results/cell_clonotypes.tsv")
paired <- subset(z, donor == "P1" & pair_status == "single_pair")
b <- unique(paired$clonotype[paired$tissue == "blood"])
t <- unique(paired$clonotype[paired$tissue == "tumor"])
c(blood = length(b), tumor = length(t),
  shared = length(intersect(b, t)))
```

P1 的血液有 6 个合格细胞、3 种克隆，肿瘤有 4 个合格细胞、2 种克隆。两处共享 1 种。用克隆种数算，血液侧比例为 1/3，肿瘤侧为 1/2；用并集作分母则为 1/4。

用第二项结果核对

```
length(intersect(b, t)) / length(union(b, t))
coverage <- read.delim("results/coverage.tsv")
subset(coverage, donor == "P1")
depth <- read.delim("results/equal_depth.tsv")
subset(depth, donor == "P1")
```

把同一患者两处抽到共同深度 4，固定种子重复 100 次，示例的 P1 共享交并比中位数为 0.25。该区间描述抽样敏感性。条形图和克隆表能说明两处检出同一克隆，方向与抗原特异性仍需各自的证据。

应能复现的三个观察

- P1 的合格链配对细胞为血液 6、肿瘤 4。
- P1 两处克隆种数为 3 和 2，交集 1、并集 4。
- P1 的 Jaccard 为 0.25，血液中属于共享克隆的细胞比例为 0.50。

缺失组织或没有有效配对链时，重叠统计记 NA。等深度抽样的分位范围也不能解读成患者群体置信区间。教学输入为人工短序列，仅用于验证匹配和分母，不用于受体生物学判定。

换成自己的数据

从每次 Cell Ranger 的 filtered_contig_annotations.csv 读取受体记录，并添加该文库的 library_id。表达谱里用同一文库编号与条码继续匹配，保留所有选定 T 细胞用于覆盖表。此练习要求高置信、有效、单 α 单 β，按两条 CDR3 核苷酸全等归并，不另要求 V/J 相同。

先对齐输入字段

cells.tsv	研究范围内的 T 细胞。含 library_id、barcode、donor、tissue；组织编码为 blood 和 tumor。
contigs.tsv	按文库追加 10x 的 barcode、chain、cdr3_nt、productive、high_confidence、is_cell。保留明确的 library_id。
settings.tsv	repetitions 控制等深度重复抽样次数，seed 控制随机种子，便于在同一软件版本复核。

请把路径和两个文库名换成自己的文件与标签。cells.tsv 需使用同样的 library_id，并保留文库内原始 barcode。多患者可按文库继续追加。

```
source("prepare.R", encoding="UTF-8")
b <- read_vdj_library("blood/filtered_contig_annotations.csv", "P1_B")
t <- read_vdj_library("tumor/filtered_contig_annotations.csv", "P1_T")
write.table(rbind(b,t), "contigs.tsv", sep="\t", quote=FALSE, row.names=FALSE)
```

将输入放入 my-data，补齐其余示例字段和设置，再运行 Rscript run.R my-data my-results。

输出文件怎样读

cell_clonotypes.tsv / coverage.tsv	逐条码的有效链、单链或多链状态，以及逐患者和组织的分母。
overlap.tsv / equal_depth.tsv	原始共享数量、各分母比例与等深度重复抽样分位数。
overlap-by-donor.pdf	每位具备两处有效样本的患者的 Jaccard 共享率图。

方法资料 / 10x 受体链字段

共享克隆至少要由几个细胞支持？亲手改三档门槛

一个共享克隆在两种组织各出现一次，与两侧都有多颗细胞支持，采样证据不同。限定在同一位患者内，分别为血液与肿瘤的克隆计数，试试双侧都要求至少一、二或三个细胞会留下什么。

```
Rscript advanced.R example results 1,2,3
```

末尾参数是组织内支持同一配对克隆的最低细胞数，多档门槛用英文逗号分开。纳入的细胞必须通过主流程的单条有效 TRA 与单条有效 TRB 判定，用配对 CDR3 核苷酸定义克隆。

P1 · 同一患者 · 双侧最低细胞数

门槛	血液克隆	肿瘤克隆	Jaccard
≥1	3	2	0.250
≥2	2	2	0.333
≥3	1	0	NA

在模拟患者 P1，门槛为 1 时血中有 3 个克隆，肿瘤中有 2 个，共享 1 个，Jaccard 为 1/4。门槛提高到 2 时两侧各保留 2 个，共享仍为 1 个，但并集缩为 3，于是指标升到 1/3。共享克隆的绝对个数并没有增加。

从结果表复算

```
x <- read.delim("results/clone-size-sensitivity.tsv")
subset(x, donor == "P1",
       c(min_cells, blood_kept, tumor_kept,
         jointly_kept, jaccard))
```

blood_kept、tumor_kept 记录每侧通过支持数门槛的不同克隆数。blood_cells_in_kept 和 tumor_cells_in_kept 记录这些克隆中的细胞总数。分清克隆数和细胞数才能看出分母变化。

门槛提到 3 时，血中还有一个克隆，肿瘤中已无可评估克隆。脚本把该行共享数和 Jaccard 标为 NA，提醒读者当前门槛下组织覆盖不足。回到 coverage.tsv 查看配对恢复率，再与主流程的等深抽样一起解释。