SEACR峰调用实操：bedGraph准备、对照归一化与阈值选择

从双端CUT&RUN/CUT&Tag比对结果制作片段覆盖bedGraph，再按是否有IgG对照选择norm/non及stringent/relaxed，最后检查输出和重复一致性。

SEACR适用于CUT&RUN及部分CUT&Tag等背景稀疏、零覆盖区域较多的信号，输入是片段覆盖的bedGraph，而不是普通read覆盖轨迹。结果会受片段定义、IgG对照、归一化和阈值模式影响。下面按SEACR 1.3给出一个双端样本的处理示例；请先确认实验方案和参考基因组，再运行峰调用。

从双端BAM构造片段覆盖轨迹

SEACR官方建议以read pair为单位生成信号：先把成对比对转为BEDPE，再提取一对read的5′和3′坐标生成片段区间，最后用bedtools genomecov输出非零bedGraph。下面演示同染色体、片段小于1000 bp的筛选条件；1000 bp是示例上限，不适用于所有文库。按实验设计调整，并对target与IgG使用相同定义、参考组装和过滤条件。

```
set -euo pipefail
bam=target.name_sorted.bam
genome=hg38.chrom.sizes
bedtools bamtobed -bedpe -i "$bam" > target.bedpe
awk '$1==$4 && $6-$2 < 1000 {print $0}' target.bedpe > target.clean.bedpe
cut -f1,2,6 target.clean.bedpe | sort -k1,1 -k2,2n -k3,3n > target.fragments.bed
bedtools genomecov -bg -i target.fragments.bed -g "$genome" > target.fragments.bedgraph
```

生成后核对文件非空、染色体名与chrom.sizes相符、区间按坐标排序、start小于end且信号值大于零。还要确认各行恰有4列、区间互不重叠，坐标遵循同一种0-based半开区间约定；target和controltarget和control必须来自相同组装、处理流程和信号定义。不要把输入的normalized bigWig简单改名为bedGraph；若信号已经按spike-in等方式严格归一化，记录归一化方法和计量单位。

选择对照与两个模式参数

场景	第二个参数和norm选项	解释
有配套IgG	control.bedgraph norm	SEACR按实验信号对control归一化，官方建议常规使用
已有严格spike-in归一化	control.bedgraph non	跳过SEACR内置control归一化，需能说明两轨迹可比
没有control	0.01 non	选择按总信号面积排序的前1%，不是FDR阈值

stringent按信号总量曲线峰值取阈值；relaxed使用曲线膝点与峰值之间的阈值。两者是调用策略，不代表统计学显著与否。可先在一个代表性样本上同时试跑，比较峰数、峰长和已知阳性区域，再按预先规定的实验问题统一选择。不要为了获得更多峰而在不同样本间切换模式。

运行并验收输出

```
SEACR=/opt/SEACR/SEACR_1.3.sh
bash "$SEACR" target.fragments.bedgraph control.fragments.bedgraph norm stringent S1
test -s S1.stringent.bed

# ■control■■■■■■■■■■■■
bash "$SEACR" target.fragments.bedgraph 0.01 non stringent S1_no_control
test -s S1_no_control.stringent.bed
```

每次运行使用唯一前缀并保存命令、SEACR版本、输入文件校验值和日志。SEACR 1.3输出BED通常包含染色体、起止坐标、区域总信号、最大信号和最大信号区间；检查列数、坐标有效性、染色体排序及空文件。若脚本报错，先查R、bedtools是否在PATH中、bedGraph是否含零信号行、输入参数顺序和文件路径。

再抽查几个已知阳性位点和邻近背景，在IGV或同类浏览器同时查看target、IgG和调用峰，确认坐标、信号方向及大片段异常。若峰集中在参考缺口、重复区域或单一染色体，先检查比对过滤和覆盖轨迹，再决定阈值是否合适。

重复和生物学结论要分开看

峰调用先按样本分别完成。评估重复一致性时，可比较共有区域、峰宽和信号强度，再按研究方案形成共识峰集；不要把多个生物学重复的bedGraph直接相加后，当作重复层面的统计检验。IgG差异、测序深度、片段筛选和目标蛋白富集都可能影响峰数。SEACR结果是候选富集区域，不自动证明直接结合或差异显著；条件比较还需在重复样本层面使用适当统计方法。

参考：SEACR官方README与输入bedGraph建议；SEACR 1.3参数和输出说明；bedtools genomecov官方文档。

参考资料

SEACR官方README：paired-end片段bedGraph准备、模式解释和输出 官方文档
SEACR 1.3官方脚本：命令参数与输出格式 官方文档
bedtools genomecov官方说明 官方文档