

## BIOF3 组学数据分析

# 06 ATAC-seq 分析要点

导出日期：2026年5月12日

## 06 ATAC-seq 分析要点

ATAC-seq 和 ChIP-seq 共享大部分分析流程（比对 → peak calling → 注释 → 差异），但有几个关键差异需要单独说明。

### 和 ChIP-seq 的区别

维度	ChIP-seq	ATAC-seq
测什么	特定蛋白结合位点	所有开放染色质区域
需要 input/control	是 (IgG 或 input DNA)	通常不需要
fragment size	单一分布	多模态 (nucleosome-free + mono/di/tri-nucleosome)
peak calling 参数	--nomodel 或默认	--nomodel --shift -100 --extsize 200
核心 QC	FRiP、IDR	TSS enrichment、fragment size 分布、FRiP

### Fragment size 分布

ATAC-seq 最重要的 QC 图是 fragment size 分布：

```

~100-150 bp → nucleosome-free fragments (信号最好的部分)
~200 bp     → mono-nucleosome
~400 bp     → di-nucleosome
~600 bp     → tri-nucleosome
  
```

好的 ATAC-seq 数据应该在 < 150bp 处有一个明显的 peak (nucleosome-free)，然后在 ~200bp 处有第二个 peak。如果第一个 peak 不明显，说明 Tn5 转座效率低或者细胞核裂解不充分。

### MACS2 参数

```

macs2 callpeak \
  -t sample.bam \
  -f BAMPE \
  --nomodel \
  --shift -100 --extsize 200 \
  -g hs \
  -n sample_atac \
  --keep-dup all \
  -q 0.05
  
```

关键参数：

- -f BAMPE：paired-end 模式，用实际 fragment 长度

- `--nomodel --shift -100 --extsize 200` : 不做 fragment size 建模, 直接用 Tn5 切割位点两侧 100bp
- `--keep-dup all` : ATAC-seq 的 PCR duplicate 已经在前面用 Picard 去过了

## 和单细胞 scATAC 的关系

[单细胞实践 10 scATAC-seq](#) 里用的 Signac 流程, 底层思路和 bulk ATAC 一样 (TF-IDF + LSI), 只是把"每个样本"换成了"每个细胞"。bulk ATAC 的 peak 可以直接作为 scATAC 的参考 peak set。

## 推荐流水线

如果不想手动跑每一步, 推荐用 nf-core/atacseq:

```
nextflow run nf-core/atacseq \  
  --input samplesheet.csv \  
  --genome GRCh38 \  
  --outdir results/
```

它会自动完成 trim → align → dedup → shift → peak call → QC report 全流程。

## 参考资源

- [ENCODE ATAC-seq pipeline](#)
- [nf-core/atacseq](#)
- [Buenrostro et al. 2013, ATAC-seq 原始论文](#)
- [Yan et al. 2020, ATAC-seq 分析最佳实践](#)



扫码关注微信公众号【生信F3】

获取文章完整内容, 分享生物信息学最新知识。