

BIOF3 组学数据分析

01 实验类型与数据格式

导出日期：2026年5月12日

01 实验类型与数据格式

表观组学实验类型多，但分析思路可以归为两大类：**开放区域检测**（ATAC-seq、DNase-seq）和**蛋白-DNA 结合检测**（ChIP-seq）。两者的分析流程几乎一样，区别在 peak calling 参数和质量指标。

三种主要实验

实验	测什么	peak 类型	典型 QC 指标
ATAC-seq	染色质开放区域	narrow	TSS enrichment、fragment size 分布
ChIP-seq	TF 结合位点 / 组蛋白修饰	narrow (TF) / broad (histone)	FRiP、IDR
WGBS/RRBS	DNA 甲基化	不做 peak calling	覆盖度、转化率

本专栏前 4 个模块聚焦 ATAC-seq 和 ChIP-seq（它们共享 peak-based 分析框架）。甲基化后续单独开。

从 FASTQ 到 peak 文件

不管是 ATAC 还是 ChIP，从原始数据到可分析的 peak 文件，标准流程是：

```
FASTQ → fastp (trim) → Bowtie2/BWA (align) → samtools (sort/filter)
→ Picard (dedup) → MACS2 (peak calling) → narrowPeak / broadPeak
```

BioF3 的表观组教程从 **peak 文件** 开始。如果你需要从 FASTQ 跑起，参考 [nf-core/chipseq](#) 或 [nf-core/atacseq](#) 流水线。

peak 文件格式

MACS2 输出的 `.narrowPeak` 是 BED6+4 格式：

```
chr1 9356548 9356648 peak_1 100 . 5.0 10.5 7.2 50
```

列	含义
1-3	染色体、起始、终止
4	peak 名
5	score
6	strand (通常 .)
7	fold enrichment
8	$-\log_{10}(\text{pvalue})$
9	$-\log_{10}(\text{qvalue})$
10	summit 相对于 start 的偏移

R 里用 `ChIPseeker::readPeakFile()` 或 `rtracklayer::import()` 读入, 得到 `GRanges` 对象。

下一步

- [02 Peak 注释与多样本比较](#)

参考资源

- [MACS2 文档](#)
- [nf-core/chipseq](#)
- [nf-core/atacseq](#)
- [ENCODE 实验标准](#)



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。