

## BIOF3 组学数据分析

# 07 DNA 甲基化分析入门

导出日期：2026年5月12日

## 07 DNA 甲基化分析入门

DNA 甲基化（主要是 CpG 位点的 5-甲基胞嘧啶）是最稳定的表观修饰之一。和 CHIP/ATAC 不同，甲基化分析不做 peak calling，而是直接量化每个 CpG 位点的甲基化率（0~100%）。

### 实验类型

技术	覆盖度	成本	适用
WGBS	全基因组 ~28M CpG	高	全景图、发现新 DMR
RRBS	富集 CpG 岛附近 ~2M CpG	中	启动子甲基化
450K / EPIC 芯片	固定位点 450K~850K	低	大样本量、TCGA 数据

BioF3 这一章聚焦 WGBS/RRBS 的 bisulfite-seq 分析。芯片数据用 minfi 包处理，思路不同。

### 分析流程

```
FASTQ → Bismark (bisulfite-aware alignment) → methylation extraction
      → per-CpG methylation table → DMR calling (methylKit / DSS / dmrseq)
```

### Bismark 比对

```
# 建立 bisulfite 索引（一次性）
bismark_genome_preparation --bowtie2 reference/

# 比对
bismark --genome reference/ -1 sample_R1.fq.gz -2 sample_R2.fq.gz

# 去重
deduplicate_bismark sample_pe.bam

# 提取甲基化信息
bismark_methylation_extractor --paired-end --comprehensive --cytosine_report \
  --genome_folder reference/ sample_pe.deduplicated.bam
```

输出的 CpG\_context 文件每行一个 CpG 位点，包含染色体、位置、甲基化 reads 数、非甲基化 reads 数。

## R 里做差异甲基化

```
library(methylKit)

# 读入 Bismark 的 CpG 报告
file_list <- list("sample1.CpG_report.txt", "sample2.CpG_report.txt",
                 "sample3.CpG_report.txt", "sample4.CpG_report.txt")
obj <- methRead(file_list,
               sample.id = list("ctrl1", "ctrl2", "trt1", "trt2"),
               assembly = "hg38",
               treatment = c(0, 0, 1, 1),
               context = "CpG",
               mincov = 10)

# 合并所有样本的 CpG 位点
meth <- unite(obj, destrand = FALSE)

# 差异甲基化位点
diff <- calculateDiffMeth(meth)
diff_25 <- getMethylDiff(diff, difference = 25, qvalue = 0.01)

# 差异甲基化区域 (DMR)
# 用 tileMethylCounts 把基因组分成 1kb 窗口再做差异
tiles <- tileMethylCounts(obj, win.size = 1000, step.size = 1000)
meth_tiles <- unite(tiles)
diff_tiles <- calculateDiffMeth(meth_tiles)
```

## 关键概念

- 甲基化率：某个 CpG 位点被甲基化的 reads 占总 reads 的比例
- **DMC (Differentially Methylated Cytosine)**：单个 CpG 位点的差异
- **DMR (Differentially Methylated Region)**：连续多个 CpG 位点一起变化的区域
- 覆盖度过滤：覆盖度 < 10x 的位点噪声太大，通常过滤掉

## 和表达的关系

启动子区域的高甲基化通常和基因沉默相关；基因体内的甲基化和活跃转录正相关。把 DMR 和 RNA-seq 的差异基因做交叉，能找到“甲基化变化驱动表达变化”的候选基因。

## 参考资源

- [Bismark 手册](#)
- [methylKit 教程](#)
- [DSS 差异甲基化](#)
- [dmrseq](#)
- [minfi \(芯片数据\)](#)



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。



BioF3

SHENGXIN F3



BioF3

SHENGXIN F3



BioF3

SHENGXIN F3



BioF3

SHENGXIN F3



BioF3

SHENGXIN F3



BioF3

SHENGXIN F3



BioF3

SHENGXIN F3



BioF3

SHENGXIN F3