

## BIOF3 组学数据分析

# 05 Motif 富集与 HOMER

导出日期：2026年5月12日

## 05 Motif 富集与 HOMER

Peak 本身只是"这里有信号"，motif 分析回答的是"什么转录因子可能在这里结合"。思路是：从 peak 序列里统计过表达的短序列模式 (motif)，和已知 TF motif 数据库比对。

### 常用工具

工具	特点
HOMER	一条命令出完整报告 (已知 + de novo motif)，最常用
MEME Suite	学术标准，de novo 发现能力强
motifmatchr + JASPAR	R 里做 motif scanning，适合和 DiffBind 结果联动

### HOMER 典型用法

```
# 安装 HOMER (一次性)
# http://homer.ucsd.edu/homer/introduction/install.html

# 对 narrowPeak 做 motif 富集
findMotifsGenome.pl peaks.narrowPeak hg38 motif_output/ \
  -size 200 -mask -p 8
```

-size 200 表示取 peak summit 两侧各 100bp 做分析。输出目录里会有：

- knownResults.html — 已知 motif 的富集排名
- homerResults.html — de novo 发现的 motif
- 每个 motif 的 logo 图

## R 里做 motif scanning

```
library(motifmatchr)
library(TFBSTools)
library(JASPAR2020)

# 获取 JASPAR 的人类 TF motif
pfm_list <- getMatrixSet(JASPAR2020, opts = list(species = "Homo sapiens"))

# 在 peak 序列里扫描 motif
library(BSgenome.Hsapiens.UCSC.hg38)
motif_hits <- matchMotifs(pfm_list, peaks_gr, genome = BSgenome.Hsapiens.UCSC.hg38)

# 统计每个 motif 在 DE peaks vs non-DE peaks 里的富集
```

这种方式的好处是能和 DiffBind 的差异结合结果直接联动：只看“在耐药细胞里 gained 的 peak 富集了什么 motif”。

### 解读要点

- **已知 motif 富集**：如果你做的是 AR ChIP-seq，排第一的应该是 ARE（androgen response element）。如果不是，说明实验或分析有问题。
- **de novo motif**：HOMER 会尝试从头发现新 motif。如果 de novo 结果和已知 motif 高度相似，说明信号很强。
- **背景选择**：默认用随机基因组区域做背景。如果你的 peak 集中在启动子，用“所有启动子”做背景会更严格。

### 参考资源

- [HOMER 官方文档](#)
- [MEME Suite](#)
- [JASPAR 数据库](#)
- [motifmatchr Bioconductor](#)



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。