

BIOF3 组学数据分析

02 Peak 注释与多样本比较

导出日期: 2026年5月12日

02 Peak 注释与多样本比较

拿到 peak 文件之后, 第一个问题是"这些 peak 落在基因组的什么位置"。ChIPseeker 把 peak 注释到最近的基因、标注它在启动子 / 内含子 / 基因间等区域, 一步出图。

本章用 ChIPseeker 自带的 AR (雄激素受体) ChIP-seq 数据演示: 3 个剂量 (0M / 1nM / 100nM) 的 peak 文件, 看剂量增加时 peak 数量、位置分布和关联基因如何变化。

核心流程

```
library(ChIPseeker)
library(TxDb.Hsapiens.UCSC.hg19.knownGene)

peak <- readPeakFile("peaks.narrowPeak")
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene

anno <- annotatePeak(peak, TxDb = txdb, tssRegion = c(-3000, 3000))
plotAnnoPie(anno)
plotDistToTSS(anno)
```

tssRegion = c(-3000, 3000) 定义"启动子"的范围: TSS 上下游 3kb 以内的 peak 算启动子区域。

真实示例

配套脚本 [epi02_chipseeker_sci.R](#) 在 ChIPseeker 内置的 AR ChIP-seq 数据上跑完整流程:

```
Rscript scripts/epigenomics/epi02_chipseeker_sci.R
```

每张图看什么

Peak annotation: AR ChIP-seq 100nM

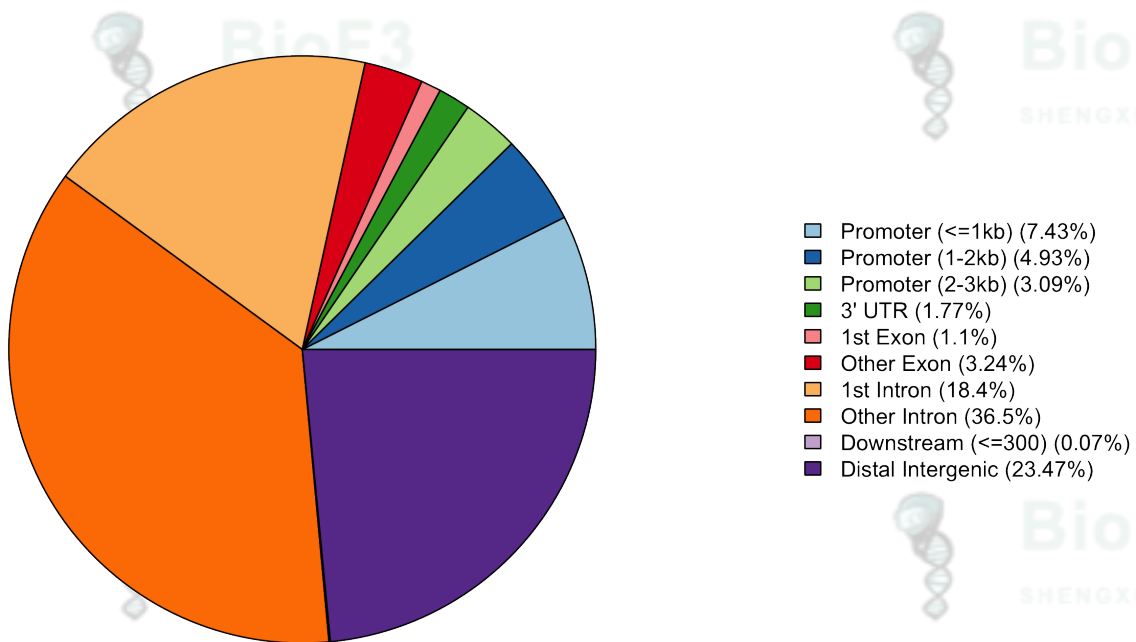


图 1: AR 100nM 的 peak 落在哪些基因组区域。启动子占比越高, 说明这个 TF 越倾向于结合在基因起始位点附近。AR 是经典的启动子 + 增强子结合 TF, 所以启动子和远端基因间区域都有不少。

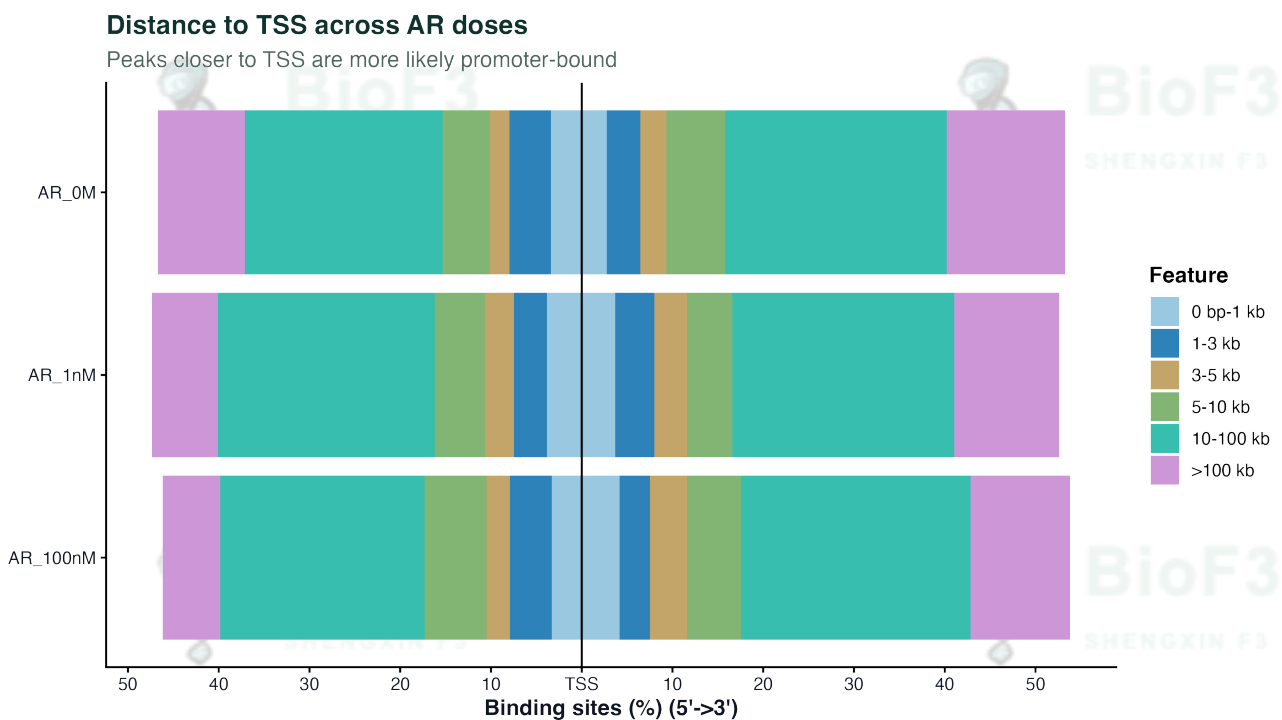


图 2: 三个剂量的 peak 到最近 TSS 的距离分布。剂量越高, 靠近 TSS 的 peak 比例越大 —— 说明高剂量下 AR 更多地占据启动子。

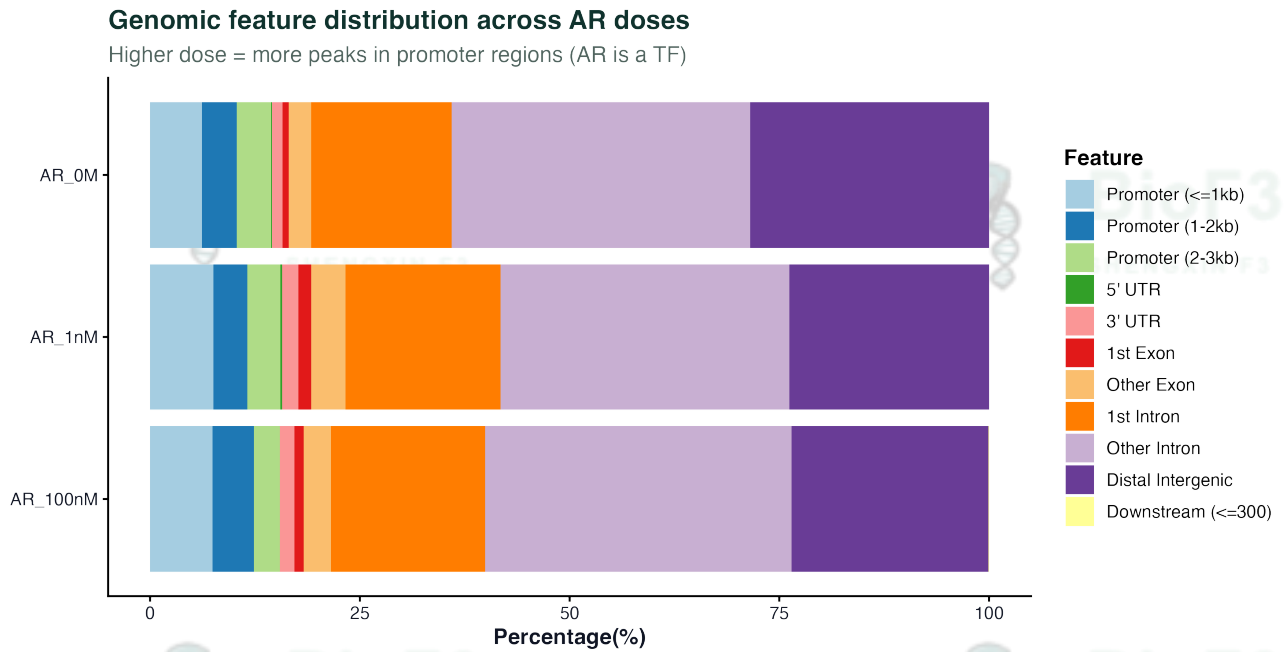


图 3：三个剂量的基因组区域分布对比。堆叠条形图一眼看出"启动子占比随剂量增加"。

Peak-associated gene overlap across AR doses

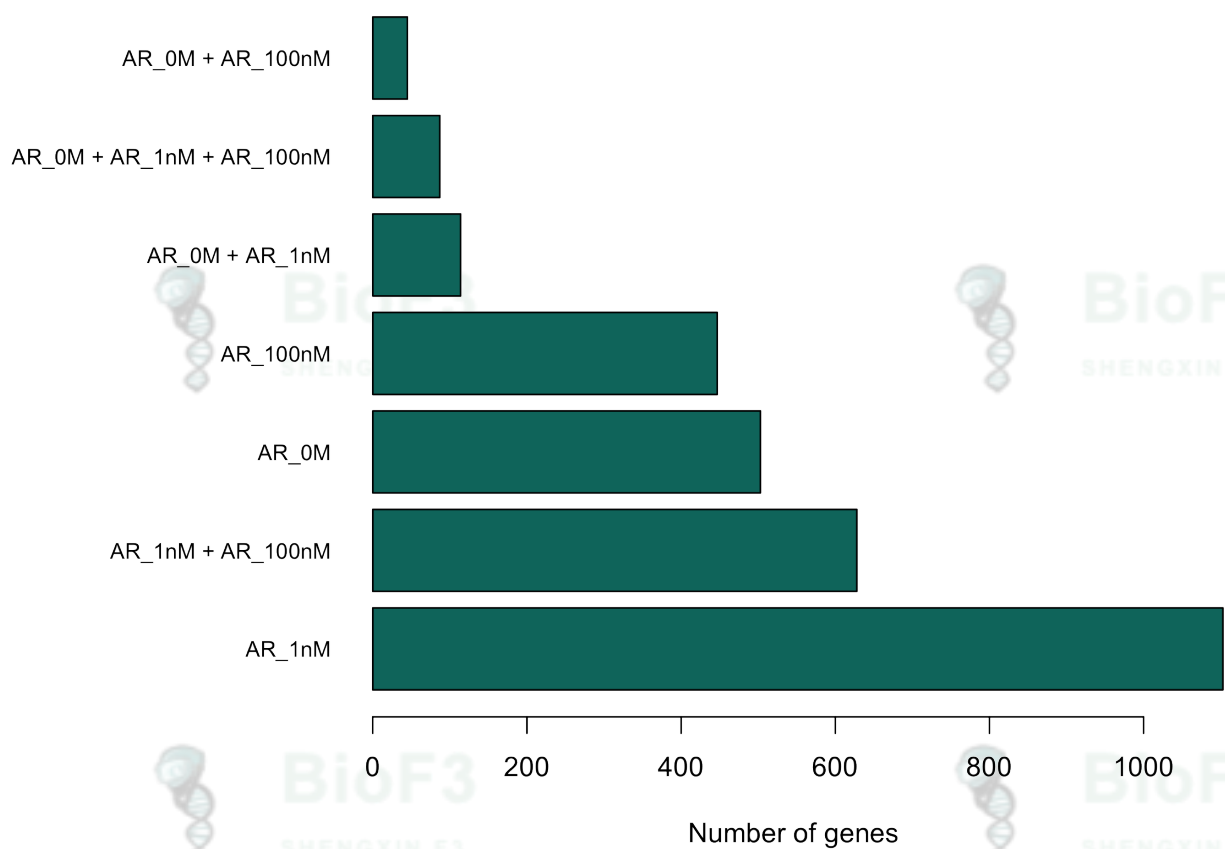


图 4：三个剂量的 peak 关联基因重叠。"三个剂量都有"的基因是 AR 的核心靶基因；"只在 100nM 出现"的是高剂量特有的。

GO BP enrichment: AR 100nM peak-associated genes

Genes nearest to AR binding peaks at highest dose

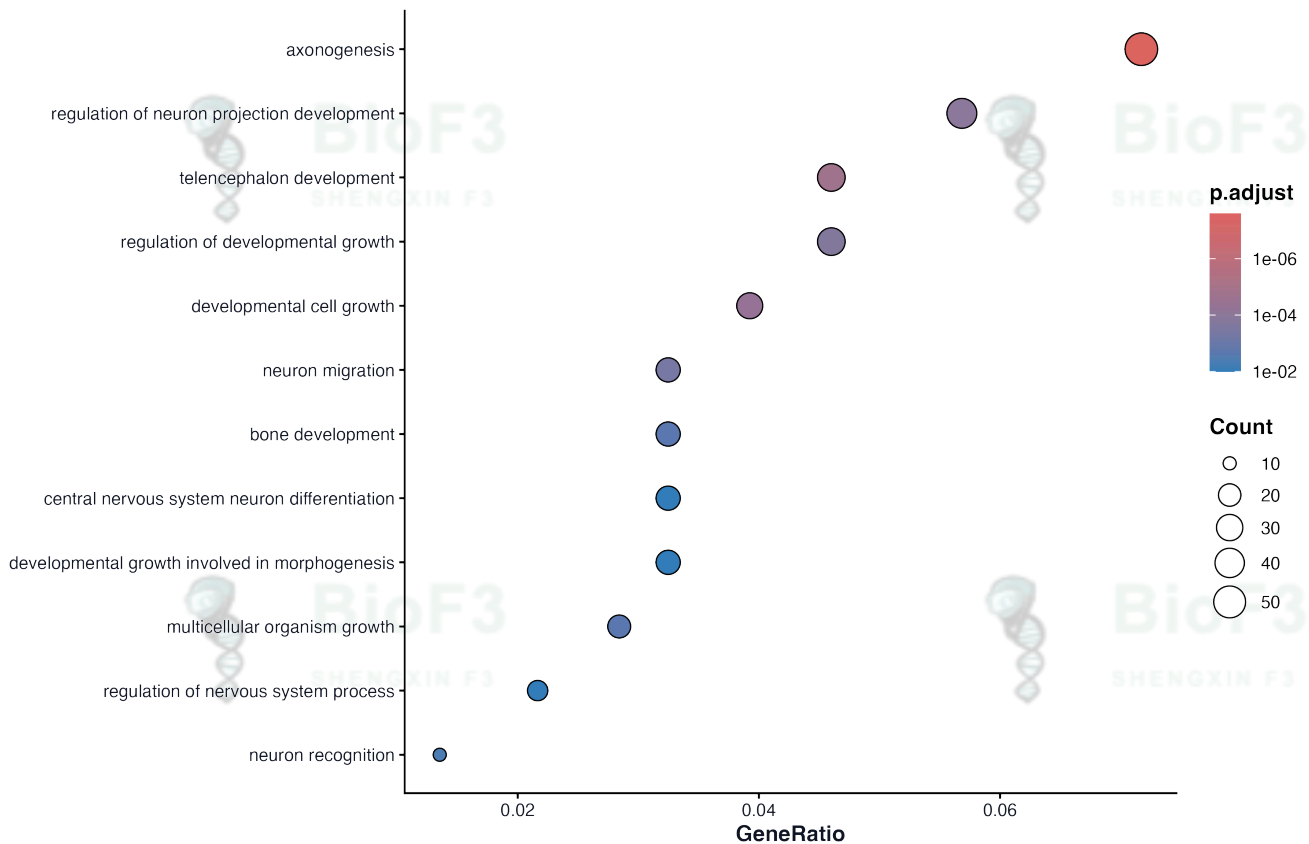


图 5: AR 100nM peak 关联基因的 GO BP 富集。应该能看到雄激素响应、细胞增殖调控等通路。

Peak count per chromosome across AR doses

Higher dose = more peaks genome-wide; distribution roughly proportional to chr size

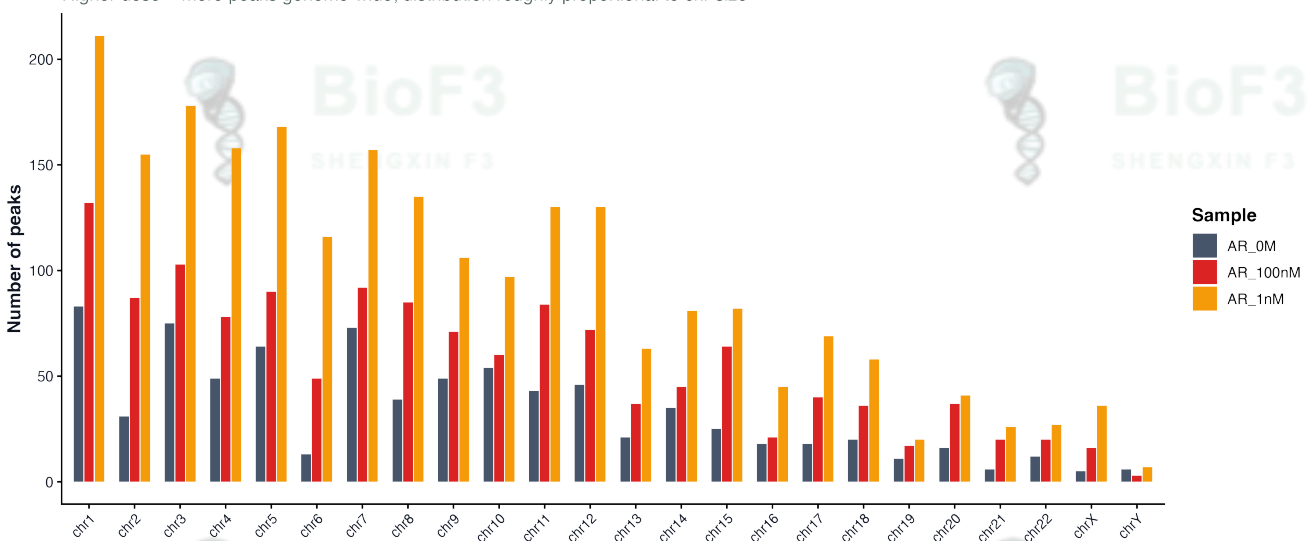


图 6: 每条染色体上的 peak 数量。分布大致和染色体大小成正比, 但某些染色体可能因为基因密度高而偏多。

套到自己数据上

把 `getSampleFiles()` 换成自己的 `.narrowPeak` 文件路径即可。注意:

- 参考基因组版本要和 peak 文件一致 (hg19 / hg38 / mm10)
- 对应的 TxDb 包: `TxDb.Hsapiens.UCSC.hg38.knownGene` (hg38) 或 `TxDb.Mmusculus.UCSC.mm10.knownGene` (小鼠)

- 如果是 broadPeak (组蛋白修饰), `annotatePeak` 的参数不用改, 但解读时"启动子"的含义要注意

下载资源

epi02_chipseeker_sci.R
8 KB

[下载 ChIPseeker peak 注释完整脚本 ↗](#)

参考资源

- [ChIPseeker Bioconductor 文档](#)
- [Yu et al. 2015, ChIPseeker 论文](#)
- [TxDb 包列表](#)



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