

BIOF3 FIGCODE · SCI 绘图代码集

# Top 基因 (Top Gene Counts)

导出日期: 2026年6月27日



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。

分类: 分布可视化 | 依赖包: ggplot2

## 解决的生物学问题

哪些基因占据了最多的测序资源? 是否存在单个基因主导文库的情况?

## 应用场景

- 检测 rRNA 或线粒体基因污染
- 识别高表达管家基因
- 评估文库复杂度
- 判断是否需要过滤特定基因

## 输入数据格式

DESeqDataSet 或归一化计数矩阵

## 关键参数

- ntop 展示基因数 (10-20)
- normalized 是否使用归一化值
- intgroup 分组变量
- transform 是否 log 变换
- geom 展示方式 (bar / point)

## 快速开始

### 1. 直接在线运行

打开 [FigCode 在线绘图](#)，点击本工具卡片的"在线绘图"按钮即可用内置示例数据出图，零环境配置。

## 2. 本地复现

下载脚本和示例数据，本地 RStudio 运行：

```
# 下载
curl -O https://<your-site>/figcode/scripts/top-genes-count.R
curl -O https://<your-site>/figcode/data/top-genes-count.csv
```

## 3. 安装依赖

```
# CRAN 包
install.packages(c("ggplot2"))

# Bioconductor 包 (如需)
# BiocManager::install(c())
```

## 完整代码

```
library(ggplot2)

# --- Demo data ---
set.seed(42)
genes <- c("ACTB", "GAPDH", "MT-CO1", "MT-ND4", "EEF1A1", "RPL13",
           "RPS27A", "MALAT1", "FTL", "B2M", "TMSB4X", "TPT1",
           "MT-CO3", "MT-CYB", "RPS4X")
counts_df <- data.frame(
  gene = rep(genes, each = 8),
  sample = rep(paste0("S", 1:8), times = 15),
  counts = rpois(120, lambda = rep(seq(50000, 5000, length.out = 15), each = 8))
)

# --- Plot ---
ggplot(counts_df, aes(x = reorder(gene, counts, FUN = median), y = counts / 1000)) +
  geom_boxplot(fill = "#e0f2fe", color = "#0369a1", outlier.size = 0.5) +
  coord_flip() +
  labs(x = NULL, y = "Normalized counts (x1000)", title = "Top Expressed Genes") +
  theme_classic()
```

## 替换为自己的数据

脚本中以 `# --- Demo data ---` 标注的段落是示例数据生成代码。替换为自己的数据时，**保持列名一致**即可：

- 输入格式：DESeqDataSet 或归一化计数矩阵
- 使用 `read.csv()` / `readRDS()` 读取本地文件

## 延伸阅读

- 相关教程：[Top 基因 完整流程](#)
- 出现 bug？欢迎在 [FigCode 页面](#) 点击对应卡片，在评论区留言。