

BIOF3 FIGCODE · SCI 绘图代码集

CNV-表达散点 (CNV-Expression Scatter)

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、ggrepel

解决的生物学问题

哪些基因的拷贝数变化真的影响了表达? 哪些是"沉默扩增"或"补偿性下调"?

应用场景

- CNV-RNA dosage effect 验证
- 驱动基因筛选 (CNV + 表达双高/双低)
- iCluster 多组学整合质控
- TCGA 多组学论文 Figure 3 经典图

输入数据格式

数据框: gene, chr, cnv_log2, expression_log2

关键参数

- method = "lm" / "loess" (拟合方法)
- color_by = "chromosome"
- highlight 高亮特定基因

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/cnv-expression.R
curl -O https://<your-site>/figcode/data/cnv-expression.csv
```

3. 安装依赖

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

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

完整代码

```
library(ggplot2)
library(ggrepel)

# --- Demo data ---
set.seed(42)
n <- 1500
df <- data.frame(
  gene = paste0("Gene", 1:n),
  chr = sample(paste0("chr", c(1:22, "X")), n, replace = TRUE),
  cnv_log2 = rnorm(n, 0, 0.6),
  noise = rnorm(n, 0, 1)
)
# Real dosage effect: expression follows CNV with noise
df$expr_log2 <- 0.7 * df$cnv_log2 + df$noise

# Highlight top |CNV × expr|
df$score <- df$cnv_log2 * df$expr_log2
top <- df[order(-abs(df$score)), ][1:15, ]

# --- Plot ---
ggplot(df, aes(cnv_log2, expr_log2)) +
  geom_point(aes(color = chr), alpha = 0.4, size = 0.9) +
  geom_smooth(method = "lm", color = "#0f172a", se = TRUE, size = 0.8) +
  geom_text_repel(data = top, aes(label = gene), size = 3,
    max.overlaps = 15) +
  geom_vline(xintercept = c(-0.3, 0.3), linetype = "dashed", color = "grey60") +
```

```
geom_hline(yintercept = 0, linetype = "dashed", color = "grey60") +
scale_color_manual(values = rep(c("#dc2626", "#2563eb", "#059669", "#d97706",
                                "#7c3aed", "#0891b2"), length.out = 23)) +
labs(title = "CNV vs Expression (per gene)",
     x = "CNV log2 ratio", y = "Expression log2 (z-score)") +
theme_classic() +
theme(legend.position = "none")
```

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：数据框：gene, chr, cnv_log2, expression_log2
- 使用 `read.csv()` / `readRDS()` 读取本地文件

延伸阅读

- 相关教程：[CNV-表达散点 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。

