

BIOF3 FIGCODE · SCI 绘图代码集

PCA 散点图 (PCA 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: 降维可视化 | 依赖包: ggplot2、ggrepel

解决的生物学问题

样本的主要变异来源是什么? 实验分组是否是最大的变异因素?

应用场景

- RNA-seq 样本质控与分组验证
- 检测批次效应
- 多因素实验设计的变异分解
- 处理前后样本变化趋势

输入数据格式

DESeqDataSet 或归一化矩阵

关键参数

- ntop 用于 PCA 的高变基因数 (500)
- 归一化方法 (VST / rlog)
- 展示的 PC 组合 (PC1 vs PC2)
- intgroup 分组变量
- 是否添加置信椭圆

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/pca.R
curl -O https://<your-site>/figcode/data/pca.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)
samples <- paste0("S", 1:10)
condition <- rep(c("Treatment", "Control"), each = 5)
# Simulate PCA coordinates
pc1 <- c(rnorm(5, 10, 2), rnorm(5, -10, 2))
pc2 <- rnorm(10, 0, 3)
pca_df <- data.frame(PC1 = pc1, PC2 = pc2, condition = condition, name = samples)
pct <- c(45, 18) # variance explained

# --- Plot ---
ggplot(pca_df, aes(PC1, PC2, color = condition)) +
  geom_point(size = 4) +
  geom_text_repel(aes(label = name), size = 3) +
  stat_ellipse(level = 0.95, linetype = "dashed") +
  labs(x = paste0("PC1 (", pct[1], "%)"), y = paste0("PC2 (", pct[2], "%)"),
       title = "PCA - Sample Clustering") +
  scale_color_manual(values = c(Treatment="#dc2626", Control="#2563eb")) +
  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 或归一化矩阵
- 使用 `read.csv()` / `readRDS()` 读取本地文件

延伸阅读

- 相关教程：[PCA 散点图 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。