

BIOF3 FIGCODE · SCI 绘图代码集

Python 火山图 (PyDESeq2 Volcano)

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: 差异分析 | 依赖包: numpy、pandas、matplotlib

解决的生物学问题

哪些基因在两组之间差异最显著? Python 生态下如何做差异分析可视化?

应用场景

- Python 用户的差异分析可视化
- 与 R DESeq2 结果对比
- 纯 Python 流程展示

输入数据格式

数据框: gene, log2FoldChange, padj

关键参数

- padj 阈值 (0.05)
- |log2FC| 阈值 (1)
- 标注基因数
- 配色方案

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/py-volcano-pydeseq2.R
curl -O https://<your-site>/figcode/data/py-volcano-pydeseq2.csv
```

3. 安装依赖

```
# CRAN 包
install.packages(c("numpy", "pandas", "matplotlib"))

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

完整代码

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# --- Demo: simulate DEG results ---
np.random.seed(42)
n = 2000
genes = [f"Gene_{i}" for i in range(n)]
log2fc = np.random.normal(0, 1.5, n)
padj = np.random.uniform(0, 1, n)
# Make some genes significant
sig_idx = np.random.choice(n, 150, replace=False)
padj[sig_idx] = np.random.uniform(1e-10, 0.01, 150)

df = pd.DataFrame({"gene": genes, "log2FC": log2fc, "padj": padj})
df["sig"] = "NS"
df.loc[(df.padj < 0.05) & (df.log2FC > 1), "sig"] = "Up"
df.loc[(df.padj < 0.05) & (df.log2FC < -1), "sig"] = "Down"

# --- Plot ---
colors = {"Up": "#dc2626", "Down": "#2563eb", "NS": "#9ca3af"}
fig, ax = plt.subplots(figsize=(7, 5))

for sig, color in colors.items():
    mask = df["sig"] == sig
    ax.scatter(df.loc[mask, "log2FC"], -np.log10(df.loc[mask, "padj"]),
```

```
c=color, s=8, alpha=0.6, label=f"{sig} ({mask.sum()})")

ax.axhline(-np.log10(0.05), ls="--", color="#9ca3af", lw=0.8)
ax.axvline(-1, ls="--", color="#9ca3af", lw=0.8)
ax.axvline(1, ls="--", color="#9ca3af", lw=0.8)

# Label top genes
top = df.nsmallest(10, "padj")
for _, row in top.iterrows():
    ax.annotate(row["gene"], (row["log2FC"], -np.log10(row["padj"])),
                fontsize=7, ha="center", va="bottom")

ax.set_xlabel("log2 Fold Change")
ax.set_ylabel("-log10(adjusted p-value)")
ax.set_title("Volcano Plot (Python)")
ax.legend(loc="upper right", frameon=False)
ax.spines[["top", "right"]].set_visible(False)
plt.tight_layout()
plt.savefig("plot_001.png", dpi=150, bbox_inches="tight")
print(f"Up: {(df.sig=='Up').sum()}, Down: {(df.sig=='Down').sum()}")
```

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, log2FoldChange, padj
- 使用 `read.csv()` / `readRDS()` 读取本地文件



BioF3
SHENGXIN F3

延伸阅读

- 后续将补充配套教程
- 出现 bug? 欢迎在 [FigCode 页面](#) 点击对应卡片，在评论区留言。



BioF3
SHENGXIN F3



BioF3
SHENGXIN F3