# 加载包

library(dplyr)

library(readr)

library(purrr)

library(readxl) # 用于读取 Excel 文件

# ---------- 1. 读取数据 ----------

# 基因家族文件

gene\_family\_path <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/火山图/wrky分组.xlsx"

deseq\_results\_path <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/火山图/DESeq2\_all\_pairwise\_results.csv"

# 判断文件格式并读取

if (grepl("\\.xlsx$", gene\_family\_path, ignore.case = TRUE)) {

gene\_family <- read\_excel(gene\_family\_path)

} else {

gene\_family <- read\_csv(gene\_family\_path)

}

if (grepl("\\.xlsx$", deseq\_results\_path, ignore.case = TRUE)) {

deseq\_results <- read\_excel(deseq\_results\_path)

} else {

deseq\_results <- read\_csv(deseq\_results\_path)

}

# ---------- 2. 按 gene\_id 合并 Symbol ----------

merged\_results <- deseq\_results %>%

left\_join(gene\_family, by = "gene\_id")

# ---------- 3. 按 comparison 分组并输出 ----------

output\_dir <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/火山图/DESeq2\_by\_comparison"

dir.create(output\_dir, showWarnings = FALSE)

merged\_results %>%

group\_split(comparison) %>%

walk(function(df) {

comp <- unique(df$comparison) %>% as.character()

# 替换非法文件名字符（防止 comparison 中含有 "/", " " 等）

comp\_clean <- gsub("[^A-Za-z0-9.\_-]", "\_", comp)

filename <- file.path(output\_dir, paste0("DESeq2\_results\_", comp\_clean, ".csv"))

df %>%

select(-comparison) %>% # 去掉 comparison 列

write\_csv(filename)

})

message("✅ 处理完成！所有结果已保存到: ", output\_dir)

# ============================

# 🔥 Volcano Plot 绘图脚本（固定使用 pvalue + 同时保存 PNG 和 PDF）

# 作者: ChatGPT

# ============================

# 加载包

library(ggplot2)

library(readr)

library(dplyr)

library(ggrepel)

# ---------- 1. 输入文件 ----------

input\_file <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/火山图/DESeq2\_by\_comparison/DESeq2\_results\_CK200\_vs\_CKO.csv"

# ---------- 2. 读取数据 ----------

res <- read\_csv(input\_file)

colnames(res) <- tolower(colnames(res)) # 统一小写，避免大小写不匹配

# 检查是否存在 pvalue 列

if (!"pvalue" %in% colnames(res)) {

stop("❌ 未检测到 pvalue 列，请检查输入文件！")

}

message("✅ 使用显著性指标：pvalue")

# ---------- 3. 修正 pvalue = 0 的情况，避免 -log10(0) ----------

res <- res %>%

mutate(pvalue = ifelse(pvalue == 0, 1e-300, pvalue)) %>%

filter(!is.na(pvalue) & is.finite(pvalue))

# ---------- 4. 添加显著性分类 ----------

res <- res %>%

mutate(

significance = case\_when(

pvalue < 0.05 & log2foldchange > 1 ~ "Up",

pvalue < 0.05 & log2foldchange < -1 ~ "Down",

TRUE ~ "NotSig"

)

)

# ---------- 5. 选择 top10 显著基因用于标注 ----------

top\_labels <- res %>%

filter(significance != "NotSig", !is.na(symbol)) %>%

arrange(pvalue) %>%

slice\_head(n = 10)

message("🧩 标注的 top10 显著基因 Symbol：")

print(top\_labels$symbol)

# ---------- 6. 绘制火山图 ----------

p <- ggplot(res, aes(x = log2foldchange, y = -log10(pvalue), color = significance)) +

geom\_point(alpha = 0.6, size = 1.8) +

geom\_vline(xintercept = c(-1, 1), linetype = "dashed", color = "black") +

geom\_hline(yintercept = -log10(0.05), linetype = "dashed", color = "black") +

scale\_color\_manual(values = c("Down" = "#2DB2EB", "NotSig" = "#d8d8d8", "Up" = "#EB4232"

)) +

theme\_minimal(base\_size = 14) +

labs(

title = "Volcano Plot: 200mM vs CK",

x = "log2(Fold Change)",

y = "-log10(pvalue)",

color = "Significance"

) +

geom\_text\_repel(

data = top\_labels,

aes(label = symbol),

size = 4,

box.padding = 0.4,

point.padding = 0.3,

max.overlaps = 20,

segment.color = "grey30",

color = "black"

)

# ---------- 7. 自动命名输出文件 ----------

comparison\_name <- gsub("^.\*DESeq2\_results\_|\\.csv$", "", basename(input\_file))

output\_png <- file.path(dirname(input\_file), paste0("Volcano\_pvalue\_", comparison\_name, ".png"))

output\_pdf <- file.path(dirname(input\_file), paste0("Volcano\_pvalue\_", comparison\_name, ".pdf"))

# ---------- 8. 保存图像 ----------

ggsave(output\_png, plot = p, width = 8, height = 6, dpi = 300)

ggsave(output\_pdf, plot = p, width = 8, height = 6, dpi = 300)

message("🎨 火山图已保存为：")

message("🖼 PNG: ", output\_png)

message("📄 PDF: ", output\_pdf)

message("✅ 绘图数据行数：", nrow(res))

message("↑Up:", sum(res$significance == 'Up'),

" ↓Down:", sum(res$significance == 'Down'),

" ·NotSig:", sum(res$significance == 'NotSig'))