library(UpSetR)

library(tidyverse)

library(ggplot2)

# 读取数据

res <- read.csv("/home/data/t210549/liuhuacheng/wrky注释/全基因组/DESeq2\_all\_pairwise\_results.csv", header = TRUE)

res <- as.data.frame(res)

# 1. 获取所有可能的比较组（即使没有差异基因）

all\_comparisons <- unique(res$comparison)

# 2. 按比较组筛选差异基因（保留原始基因 ID）

sig\_res <- res %>%

filter(padj < 0.05) %>% # 筛选显著基因

select(gene\_id, comparison) # 只保留需要的列

# 3. 获取所有差异基因

all\_genes <- unique(sig\_res$gene\_id)

# 4. 构建逻辑矩阵（确保包含所有组别）

logic\_mat <- matrix(0,

nrow = length(all\_genes),

ncol = length(all\_comparisons),

dimnames = list(all\_genes, all\_comparisons))

# 5. 填充逻辑矩阵

for (comp in all\_comparisons) {

genes\_in\_comp <- sig\_res %>%

filter(comparison == comp) %>%

pull(gene\_id)

if (length(genes\_in\_comp) > 0) {

logic\_mat[genes\_in\_comp, comp] <- 1

}

# 如果该组别没有差异基因，该列会保持全0

}

# 6. 检查逻辑矩阵

cat("逻辑矩阵维度:", dim(logic\_mat), "\n")

cat("各比较组差异基因数量:\n")

print(colSums(logic\_mat))

# 7. 保存逻辑矩阵

write.csv(logic\_mat, "/home/data/t210549/liuhuacheng/wrky注释/全基因组/logic\_matrix\_for\_upset.csv")

# 8. 转换为数据框并绘制UpSet图

logic\_df <- as.data.frame(logic\_mat)

# 假设你的逻辑矩阵是 logic\_df

# 重命名列名（将原始名称改为更友好的名称）

new\_column\_names <- c(

"CK50 vs CKO" = "50 vs CK",

"CK100 vs CKO" = "100 vs CK",

"CK150 vs CKO" = "150 vs CK",

"CK200 vs CKO" = "200 vs CK",

"CK100 vs CK50" = "100 vs 50",

"CK150 vs CK50" = "150 vs 50",

"CK200 vs CK50" = "200 vs 50",

"CK150 vs CK100" = "150 vs 100",

"CK200 vs CK100" = "200 vs 100",

"CK200 vs CK150" = "200 vs 150"

)

# 应用新名称到逻辑矩阵

colnames(logic\_df) <- new\_column\_names[colnames(logic\_df)]

pdf(

file = "/home/data/t210549/liuhuacheng/wrky注释/全基因组/UpsetR.pdf",

width = 12, # 单位：英寸

height = 8, # 单位：英寸

onefile = FALSE,

family = "Helvetica"

)

# 绘制UpSet图（显示所有组别）

upset(

logic\_df,

nsets = length(all\_comparisons), # 显示所有组别

nintersects = 30, # 显示的交集数量

order.by = "freq", # 上方交集条按频率排序

keep.order = FALSE, # 保持集合顺序（左侧条形图不按大小排序）

main.bar.color = "black", # 上方条形图颜色

sets.bar.color = "skyblue", # 左侧条形图设为浅蓝色

text.scale = 1.5, # 调整文字大小

sets = colnames(logic\_df) # 显式指定要显示的组别

)

# 4. 关闭 PDF 设备

dev.off()