# 加载必要的包

library(DESeq2)

# 1. 读取 counts 矩阵（假设是 CSV 格式，第一列是基因名）

library(readxl)

# 读取Excel文件（自动检测编码）

counts\_data<-read\_xlsx("/home/data/t210549/liuhuacheng/wrky注释/全基因组/subset.xlsx",

sheet = 1, # 读取第一个sheet

na = "NA", # 缺失值标记

trim\_ws = TRUE) # 去除多余空格

# 1. 转换为 data.frame

counts\_data <- as.data.frame(counts\_data)

# 2. 设置行名为基因名

rownames(counts\_data) <- counts\_data[, 1]

# 3. 去掉基因名列

counts\_data <- counts\_data[, -1]

# 2. 定义样本分组信息（根据你的样本名修改）

sample\_names <- colnames(counts\_data)

condition <- factor(

c(rep("CKO", 3), rep("CK50",2),rep("CK100", 2), rep("CK150", 2), rep("CK200", 3)),

levels = c("CKO" ,"CK50","CK100", "CK150", "CK200")

)

# 构建样本信息表

coldata <- data.frame(

sample = sample\_names,

condition = condition,

row.names = sample\_names

)

# 3. 创建 DESeqDataSet 对象

dds <- DESeqDataSetFromMatrix(

countData = counts\_data,

colData = coldata,

design = ~ condition

)

# 4. 过滤低表达基因（至少在 3 个样本中 counts > 10）

keep <- rowSums(counts(dds) >= 10) >= 3

dds <- dds[keep, ]

# 5. 运行 DESeq2 分析

dds <- DESeq(dds)

# 1. 获取所有组名

groups <- levels(dds$condition)

cat("所有组名：", groups, "\n")

# 2. 生成所有两两组合（不重复，如 A vs B 和 B vs A 都会生成）

comparisons <- combn(groups, 2, simplify = FALSE)

cat("总共有", length(comparisons), "对两两比较\n")

# 3. 初始化一个空列表，用于存储所有比较结果

all\_results\_list <- list()

# 4. 循环提取每对比较的结果

for (comp in comparisons) {

group1 <- comp[1]

group2 <- comp[2]

# 提取比较结果

res <- results(dds, contrast = c("condition", group2, group1))

# 转换为数据框

res\_df <- as.data.frame(res)

# 添加比较组标签

res\_df$comparison <- paste(group2, "vs", group1)

# 添加到列表

all\_results\_list[[paste(group2, "vs", group1)]] <- res\_df

}

# 5. 合并所有结果

all\_res <- do.call(rbind, all\_results\_list)

# 6. 保存为 CSV 文件

output\_file <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/DESeq2\_all\_pairwise\_results.csv"

write.csv(all\_res, output\_file, row.names = TRUE)

cat("所有两两比较结果已保存至:", output\_file, "\n")