suppressPackageStartupMessages({

library(dplyr); library(stringr); library(tidyr)

library(readxl); library(clusterProfiler); library(enrichplot); library(ggplot2)

})

## 路径（按需修改）

kegg\_annotation\_path <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/kegg\_ko\_long.csv"

deg\_file\_path <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/各组差异基因/200vsCK.xlsx"

out\_prefix <- "/home/data/t210549/liuhuacheng/wrky注释/全基因组/200vsCK\_KEGG\_KO"

## 1) 读取 long 表，并按你的小写列名重命名

anno <- read.csv(kegg\_annotation\_path, header = TRUE, check.names = FALSE) %>%

rename(Gene\_ID = gene, KEGG\_ko = kegg\_ko) %>% # 关键：小写->标准名

filter(!is.na(KEGG\_ko), KEGG\_ko != "-") %>%

mutate(KEGG\_ko = str\_replace\_all(KEGG\_ko, "^ko:", ""), # 去掉 "ko:" 前缀

KEGG\_ko = str\_trim(KEGG\_ko)) %>%

filter(str\_detect(KEGG\_ko, "^K\\d+")) %>%

distinct(Gene\_ID, KEGG\_ko)

cat(sprintf("[INFO] Gene–KO 映射：%d 行，基因数=%d，KO数=%d\n",

nrow(anno), n\_distinct(anno$Gene\_ID), n\_distinct(anno$KEGG\_ko)))

## 2) 读取差异基因列表

deg\_ids <- read\_excel(deg\_file\_path, col\_names = FALSE)[[1]] %>%

as.character() %>% unique() %>% na.omit()

cat(sprintf("[INFO] 读入 DEG：%d 个\n", length(deg\_ids)))

## 3) DEG -> KO；背景 KO

deg\_ko <- anno %>% filter(Gene\_ID %in% deg\_ids) %>% pull(KEGG\_ko) %>% unique()

bg\_ko <- anno$KEGG\_ko %>% unique()

cat(sprintf("[INFO] DEG 映射到 KO：%d 个；背景 KO：%d 个\n", length(deg\_ko), length(bg\_ko)))

if (length(deg\_ko) < 10) stop("映射到 KO 的差异基因太少（<10）。")

## 4) KO 宇宙富集（organism='ko'）

ek <- enrichKEGG(

gene = deg\_ko,

organism = "ko",

keyType = "kegg",

universe = bg\_ko,

pvalueCutoff = 0.05,

qvalueCutoff = 0.20

)

res <- as.data.frame(ek)

if (nrow(res) == 0) stop("当前阈值下无显著通路，建议暂时把 pvalueCutoff 提到 0.2 观察趋势。")

## 检查是否已拿到通路名称

print(head(res[, c("ID","Description","pvalue","p.adjust","GeneRatio")]))

## 5) 画图与导出（dotplot 会自动用 Description 作为 y 轴）

p1 <- dotplot(ek, showCategory = 20, label\_format = 60) +

ggtitle("KEGG Enrichment: 200 mM vs CK")

ggsave(paste0(out\_prefix, "\_dotplot.pdf"), p1, width = 8, height = 9, dpi = 300)

ggsave(paste0(out\_prefix, "\_dotplot.png"), p1, width = 8, height = 9, dpi = 300)

write.csv(res, paste0(out\_prefix, "\_results.csv"), row.names = FALSE)

cat("[DONE] 已输出结果与图表；y 轴为通路名称（Description）。\n")