|  |
| --- |
| R library(data.table) library(TwoSampleMR) library(MRcML) library(dplyr) suppressMessages(library(tidyverse)) library(MRPRESSO)  # 设置结果所在环境 setwd("~/GWAS/GM\_data") # 创建新的文件夹 new\_folder <- "GM\_or\_data" dir.create(new\_folder)  #设置工作环境为GM数据所在 setwd("~/GWAS/GM") # 读入数据集 GM <- fread("MBG.allHits.p1e4.txt", header = TRUE)  # 设置结果所在环境 setwd("~/GWAS/GM\_data")  # 将数据集按照ID分为不同的子集 data\_list <- split(GM, GM$bac)  # 对每个子集进行导出为csv文件操作，并保存到新文件夹 lapply(names(data\_list), function(x) {  filepath <- file.path(new\_folder, paste0(x, ".csv"))  write.csv(data\_list[[x]], filepath, row.names = FALSE) })  ## 创建新文件夹 dir.create("GM\_Pclean")  ## 读取文件夹中所有csv文件 file\_names <- list.files(path = "~/GWAS/GM\_data/GM\_or\_data",  pattern = "\\.csv$",   full.names = TRUE)  # 定义一个函数对单个文件进行操作 p\_clean\_file <- function(file\_path) {  dat <- read.csv(file\_path)  dat <- subset(dat,P.weightedSumZ < 1e-5)  new\_file\_path <- file.path("GM\_Pclean",   paste0("subset\_",   basename(file\_path)))  write.csv(dat, file = new\_file\_path, row.names = FALSE) }  # 对所有文件执行操作，并保存子集到新文件夹 lapply(file\_names, p\_clean\_file)  #读取文件夹中所有csv文件 file\_names <- list.files("~/GWAS/GM\_data/GM\_Pclean", pattern = "\\.csv$", full.names = TRUE)   #新建一个文件夹 dir.create("GM\_clumped")  clumped\_file <- function(file\_paths) {    #遍历所有文件  for (file\_path in file\_paths) {  # 从文件路径中提取文件名，作为输出文件的前缀  prefix <- basename(file\_path) %>% tools::file\_path\_sans\_ext()    # 运行您的代码  data\_clumped <- read\_exposure\_data(file\_path, sep = ",",    snp\_col = "rsID",    phenotype\_col = "bac",    beta\_col = "beta",     se\_col = "SE",     effect\_allele\_col = "eff.allele",     other\_allele\_col = "ref.allele",    clump = TRUE)    # 将处理结果保存为 CSV 文件  new\_file\_path <- file.path("GM\_clumped", paste0("clumped\_", basename(file\_path)))  write.csv(data\_clumped, new\_file\_path, row.names = FALSE)  } }  #运行函数 # 对所有文件执行操作，并保存子集到新文件夹 大概六七分钟 lapply(file\_names, clumped\_file)   #读取文件夹中所有csv文件 file\_names <- list.files("~/GWAS/GM\_data/GM\_clumped",   pattern = "\\.csv$",   full.names = TRUE)       ##################新建一个文件夹 dir.create("MR\_GM\_ieu-a-989")  #设工作环境 setwd("~/GWAS/MR\_GM\_ieu-a-989") #读取文件夹中所有csv文件 # 获取文件夹中的文件名 exposure\_files <- list.files("~/GWAS/GM\_data/GM\_clumped",   pattern = "\\.csv$", full.names = TRUE)  #新建一个文件夹         dir.create("GM\_harmonise")  # 对每个exposure文件和outcome文件进行操作 for(i in seq\_along(exposure\_files)){    # 读取exposure文件  exposure\_dat <- fread(exposure\_files[i], header = T)    # 读取outcome文件  outcome\_dat1 <- extract\_outcome\_data(snps = exposure\_dat$SNP, outcomes = 'ieu-a-989')  outcome\_dat <- outcome\_dat1[!duplicated(outcome\_dat1$SNP),]  required\_cols <- c("SNP", "id.outcome", "outcome", "beta.outcome", "se.outcome", "effect\_allele.outcome", "other\_allele.outcome")  missing\_cols <- setdiff(required\_cols, colnames(outcome\_dat))    if (length(missing\_cols) > 0) {  no\_req\_cols\_file <- paste("Outcome file", basename(exposure\_files[i]), "does not contain all required columns:", paste(missing\_cols, collapse = ", "), sep = " ")  cat(no\_req\_cols\_file, file = "no\_outcome\_files.txt", append = TRUE, sep = "\n")  next  }    # 合并数据  dat <- harmonise\_data(exposure\_dat = exposure\_dat, outcome\_dat = outcome\_dat)   dat$`R2` <- 2\*dat$beta.exposure\*dat$beta.exposure\*dat$eaf.outcome\*(1- dat$eaf.outcome)    dat$`F` <- 18338 \* dat$R2/(1- dat$R2)    dat <- subset(dat, F > 10)    new\_file\_path <- file.path("GM\_harmonise", paste0("harmonise\_", basename(exposure\_files[i])))  # 导出数据为csv文件  write.csv(dat, file = new\_file\_path, row.names = FALSE) }    #新建一个文件夹 dir.create("GM\_mr")  ###MR files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_harmonise",   pattern = "\\.csv$", full.names = TRUE)     for(i in seq\_along(files\_name)){    # 读取exposure文件  dat <- read.csv(files\_name[i], header = T)  data <- mr(dat)     new\_file\_path <- file.path("GM\_mr", paste0("mr\_", basename(files\_name[i])))  write.csv(data, file = new\_file\_path, row.names = FALSE) }  ###MR files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_mr", pattern = "\\.csv$", full.names = TRUE)  ##将所有结果整合到一个表格里  df\_list <- lapply(files\_name, fread) mr\_all\_data <- rbindlist(df\_list, fill = TRUE)  # 将数据框保存到新的 csv 文件 write.csv(mr\_all\_data,"mr\_all\_data.csv", row.names = FALSE)   #新建一个文件夹 dir.create("GM\_mr\_or")  ###MR files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_mr",  pattern = "\\.csv$", full.names = TRUE)  for(i in seq\_along(files\_name)){    # 读取exposure文件  dat <- fread(files\_name[i])    data <- generate\_odds\_ratios(mr\_res = dat)     new\_file\_path <- file.path("GM\_mr\_or", paste0("mr\_or\_", basename(files\_name[i])))  write.csv(data, file = new\_file\_path, row.names = FALSE) } ###MR\_or的结果 files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_mr", pattern = "\\.csv$", full.names = TRUE)  ##将所有结果整合到一个表格里  df\_list <- lapply(files\_name, read.csv) mr\_or\_all\_data <- do.call(rbind, df\_list)  class\_rows <- startsWith(mr\_or\_all\_data$exposure, "class") & mr\_or\_all\_data$method == "Inverse variance weighted" fam\_rows <- startsWith(mr\_or\_all\_data$exposure, "fam") & mr\_or\_all\_data$method == "Inverse variance weighted" phy\_rows <- startsWith(mr\_or\_all\_data$exposure, "phy") & mr\_or\_all\_data$method == "Inverse variance weighted" ord\_rows <- startsWith(mr\_or\_all\_data$exposure, "ord") & mr\_or\_all\_data$method == "Inverse variance weighted" gen\_rows <- startsWith(mr\_or\_all\_data$exposure, "gen") & mr\_or\_all\_data$method == "Inverse variance weighted"  p\_adj\_class <- p.adjust(mr\_or\_all\_data$pval[class\_rows], method = "fdr") p\_adj\_fam <- p.adjust(mr\_or\_all\_data$pval[fam\_rows], method = "fdr") p\_adj\_phy <- p.adjust(mr\_or\_all\_data$pval[phy\_rows], method = "fdr") p\_adj\_ord <- p.adjust(mr\_or\_all\_data$pval[ord\_rows], method = "fdr") p\_adj\_gen <- p.adjust(mr\_or\_all\_data$pval[gen\_rows], method = "fdr")  p\_adj <- rep(NA, nrow(mr\_or\_all\_data)) p\_adj[class\_rows] <- p\_adj\_class p\_adj[fam\_rows] <- p\_adj\_fam p\_adj[phy\_rows] <- p\_adj\_phy p\_adj[ord\_rows] <- p\_adj\_ord p\_adj[gen\_rows] <- p\_adj\_gen  mr\_or\_all\_data$p\_adj <- p\_adj  # 将数据框保存到新的 csv 文件 write.csv(mr\_or\_all\_data, "mr\_or\_all\_data.csv", row.names = FALSE)  #####新建一个文件夹 dir.create("GM\_mr\_heterogeneity")  ###MR files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_harmonise", pattern = "\\.csv$",full.names = TRUE)  for(i in seq\_along(files\_name)){    # 读取exposure文件  dat <- fread(files\_name[i])    df <- mr\_heterogeneity(dat)     new\_file\_path <- file.path("GM\_mr\_heterogeneity", paste0("mr\_heterogeneity\_",   basename(files\_name[i])))  write.csv(df, file = new\_file\_path, row.names = FALSE) } ################################################################################### # 读取所有mr\_heterogeneity结果文件 files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_mr\_heterogeneity",   pattern = "\\.csv$", full.names = TRUE) df\_list <- lapply(files\_name, fread)  # 添加文件名作为新的一列 for (i in seq\_along(df\_list)) {  df\_list[[i]]$file\_name <- basename(files\_name[i]) }  # 合并所有数据框 mr\_heterogeneity\_all\_data <- rbindlist(df\_list, fill = TRUE)  # 将数据框保存到新的 csv 文件 write.csv(mr\_heterogeneity\_all\_data, "mr\_heterogeneity\_all\_data.csv", row.names = FALSE)   ###############################################################################  #新建一个文件夹 dir.create("GM\_mr\_pleiotropy\_test")  ###MR files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_harmonise",   pattern = "\\.csv$", full.names = TRUE)  for(i in seq\_along(files\_name)){    # 读取exposure文件  dat <- fread(files\_name[i])    df <- mr\_pleiotropy\_test(dat)     new\_file\_path <- file.path("GM\_mr\_pleiotropy\_test",   paste0("mr\_pleiotropy\_test\_",   basename(files\_name[i])))  write.csv(df, file = new\_file\_path, row.names = FALSE) }  ############################################################################### # 读取所有mr\_pleiotropy\_test结果文件 files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_mr\_pleiotropy\_test", pattern = "\\.csv$", full.names = TRUE) df\_list <- lapply(files\_name, read.csv)  # 添加文件名作为新的一列 for (i in seq\_along(df\_list)) {  df\_list[[i]]$file\_name <- basename(files\_name[i]) }  # 合并所有数据框 mr\_pleiotropy\_test\_all\_data <- rbindlist(df\_list, fill = TRUE)  # 将数据框保存到新的 csv 文件 write.csv(mr\_pleiotropy\_test\_all\_data, "mr\_pleiotropy\_test\_all\_data.csv",   row.names = FALSE)  #新建一个文件夹 dir.create("GM\_mr\_presso") file.create("selected\_files.txt")  ###MR  files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_harmonise",   pattern = "\\.csv$", full.names = TRUE)  for(i in seq\_along(files\_name)){    # 读取exposure文件  SummaryStats <- fread(files\_name[i], header = T)    SummaryStats = as.data.frame(SummaryStats)    # 判断行数是否小于等于3  if(nrow(SummaryStats) <= 3) {  # 将文件名写入文件  cat(files\_name[i], "\n", file = "selected\_files.txt", append = TRUE)  next # 直接进入下一次循环  }    df <- mr\_presso(BetaOutcome = "beta.outcome",  BetaExposure = "beta.exposure",   SdOutcome = "se.outcome",   SdExposure = "se.exposure",   OUTLIERtest = F,   DISTORTIONtest = F,   data = SummaryStats,   NbDistribution = 1000,   SignifThreshold = 0.05)     new\_file\_path <- file.path("GM\_mr\_presso", paste0("mr\_\_presso\_",   basename(files\_name[i])))    write.csv(df, file = new\_file\_path, row.names = FALSE) }  # 读取所有mr\_presso结果文件 files\_name <- list.files("~/GWAS/MR\_GM\_ieu-a-989/GM\_mr\_presso",   pattern = "\\.csv$", full.names = TRUE) df\_list <- lapply(files\_name, read.csv)  # 添加文件名作为新的一列 for (i in seq\_along(df\_list)) {  df\_list[[i]]$file\_name <- basename(files\_name[i]) }  # 合并所有数据框 mr\_presso\_all\_data <- do.call(rbind, df\_list)  # 将数据框保存到新的 csv 文件 write.csv(mr\_presso\_all\_data, "mr\_\_presso\_all\_data.csv", row.names = FALSE) |