```r
#### Author : Dina Gamaleldin Mansour Aly

#### Date : 4th October 2017

#### Program Usage: ./GWAS_plots.r ; this R script is used along with GWAS_pipeline for the

#### filtration, sorting and plotting of SNPTEST output files.

setwd("")

library("plyr")

library("qqman")

UNFILTERED_DATA= read.csv("GWAS_unfiltered.txt", sep=" ", header=T)

head(UNFILTERED_DATA)

MAF_HWE_Filtered_out= subset (UNFILTERED_DATA, UNFILTERED_DATA$all_maf < 0.05 &
UNFILTERED_DATA$controls_hwe < 0.0000057 )

write.csv (MAF_HWE_Filtered_out$rsid, "Filtered_snps.txt")

MAF_HWE_Filter_data= subset (UNFILTERED_DATA, UNFILTERED_DATA$all_maf > 0.05 &
UNFILTERED_DATA$controls_hwe > 0.0000057 )

head(MAF_HWE_Filter_data)

MAF_HWE_Filter_data <-
MAF_HWE_Filter_data[!(is.na(MAF_HWE_Filter_data$frequentist_add_pvalue)),]

CLEAN_data = as.data.frame (MAF_HWE_Filter_data)

SORTED_DATA = CLEAN_data[order(CLEAN_data$frequentist_add_pvalue),]

str(SORTED_DATA)

head(SORTED_DATA)

SORTED_DATA_1 = cbind(SORTED_DATA$rsid, SORTED_DATA$alternate_ids,
SORTED_DATA$position, SORTED_DATA$cases_total, SORTED_DATA$controls_total,
SORTED_DATA$all_maf, SORTED_DATA$all_OR,

        SORTED_DATA$all_OR_lower, SORTED_DATA$all_OR_upper,
SORTED_DATA$frequentist_add_pvalue)

SORTED_DATA_1 =as.data.frame(SORTED_DATA_1)

write.csv(SORTED_DATA_1, "Sorted_SNPTEST_output.csv")

PLOT_DATA =  cbind(CLEAN_data$rsid, CLEAN_data$alternate_ids, CLEAN_data$position,
CLEAN_data$frequentist_add_pvalue)

colnames(PLOT_DATA) = c("SNP","CHR", "BP","P")

PLOT_DATA = as.data.frame(PLOT_DATA)

pdf_file <- paste ("GWAS_manhattan_plot.pdf", sep = "")

pdf_qqplot <- paste ("GWAS_qqplot.pdf", sep = "")
```

```
pdf(pdf_file)

manhattan(PLOT_DATA, main = "Manhattan Plot of GWAS", ylim = c(0, 50), cex = 0.6, cex.axis = 0.9,
col = c("#56B4E9", "blue"), suggestiveline = -log10(1e-05), genomewideline = -log10(5e-08),
annotateTop = TRUE)

sink()

pdf(pdf_qqplot)

qq(PLOT_DATA$P, main = "Q-Q plot of GWAS", xlim = c(0, 7), ylim = c(0, 50), pch = 18, col = "blue",
cex = 1.5, las = 1)

sink()
```