Gulishana Adilijiang (adiliji2)

STAT 530 Bioinformatics: Homework 2

Problem1:

1) Samples:

After filtering, 0 of 89 individuals are removed for low genotyping (MIND > 0.06).

2) SNPs:

2 markers to be excluded based on HWE test (p <= 0.0001), 1 in cases, 1 in controls.

859 SNPs failed missingness test (GENO > 0.1).

24114 SNPs failed frequency test (MAF < 0.05).

After frequency and genotyping pruning, there are 58737 SNPs left.

Problem2:

Genomic inflation factor (based on median chi-squared) is 1.16769.

The OR of the SNP rs2222162 is 0.2221.

OR < 1, which means having more minor allele of this SNP associated with lower risk.

Problem3:

Top 5 lines of qcd.pca are shown as below:

3
1.6190
1.1200
1.1120
 0.1068 -0.0439 -0.0735

Problem4:

Previously, the genomic inflation factor (based on median chi-squared) is 1.16769 > 1.05, which means it needs to adjust for population structure. Now the genomic inflation factor (based on median chi-squared) is 1.02424 < 1.05, which means we have adequately controlled for population stratification.

The OR of the SNP rs2222162 is 0.01695.

OR < 1, which means having more minor allele of this SNP associated with lower risk.

Problem5:

None of the SNPs reach genome-wide significance.

The Bonferroni threshold adjusting for 58737 SNPs tested is $0.05/58737 = 8.51*10^{-7}$

None of the SNPs pass this threshold.

Problem6:

R Notebook

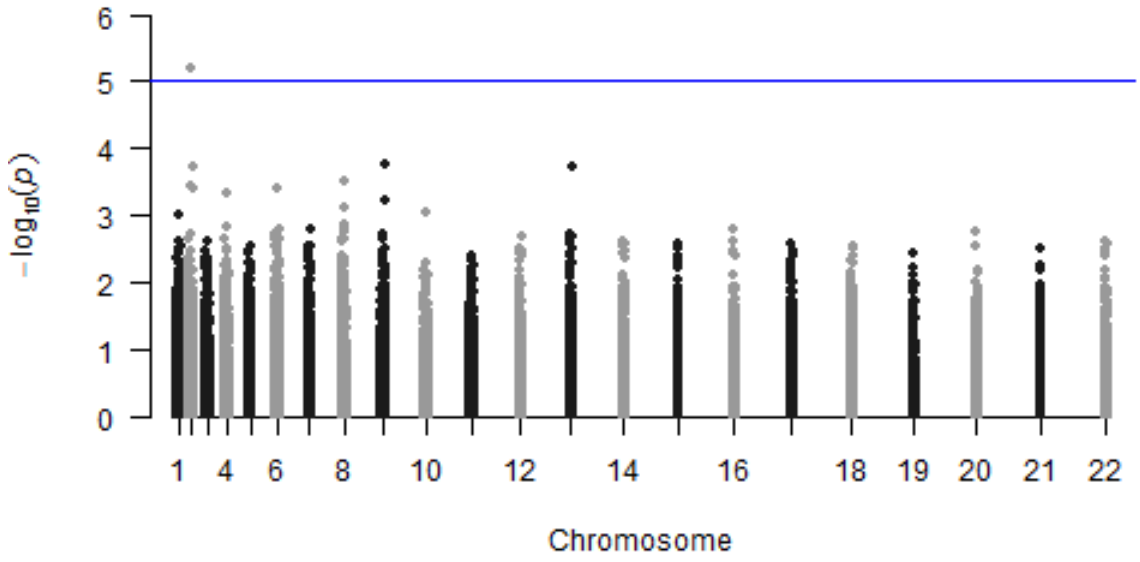
This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
setwd("D:/VirtualBox_Ubuntu/Ubuntu_LTS_shared folders/STAT530/hapmap1")
pc1 <- read.table("pc1.assoc.logistic",header=TRUE)
pc1 <- pc1[pc1$TEST=="ADD",]
library(qqman)

##

## For example usage please run: vignette('qqman')

##

## Citation appreciated but not required:

## Turner, S.D. qqman: an R package for visualizing GWAS results using Q-Q an
d manhattan plots. biorXiv DOI: 10.1101/005165 (2014).

##

png(file="Manhattan.png",width=500,heigh=300)
manhattan(pc1[,c("SNP","CHR","BP","P")])
dev.off()

## png
##   2
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).



The SNP rs2222162 exceeds the 10^{-5} significance threshold.

The RefSeq gene closet to SNP rs2222162 is GYPC.