BST227 - Homework 2

Due:Wednesday, November 28

Introduction

You've recently discovered that there are in fact two types of cholesterol – both good (HDL) and bad (LDL). You are worried that you may have a genetic predisposition to having high levels of LDL cholesterol and decide to investigate what genes may be associated with LDL cholesterol. Here, we'll do a basic exploration of the data before investigating the specific genetic effects in subsequent lectures. We encourage you to use R markdown to include both your code and plots in the same document and then submit a pdf to Canvas.

Problem 1

Access the LDL summary statistics from the Global Lipids Genetics Consortium (GLGC) genome-wide association study. Visualize the summary statistics both as a Manhattan Plot and as a Q-Q plot. What chromosome(s) appear to have genome-wide significant hits?

Hint: The data can be found through LDHub: http://ldsc.broadinstitute.org/ldhub/. Click on "Get Started with LD Hub" > "Sign in with your Google account" > "Go GWAShare Center"and then click on the link for "LDL.sumstats_deGC.gz" in the leftmost column. This takes you to the GLGC website where you can read about the data and then download the file by clicking on the "LDL Cholesterol" link under "Joint Analysis of Metabochip and GWAS Data." Look back at Lab 2 for helpful R packages and functions for visualizing the data.

Problem 2

As part of the GLGC Consortium, the group analyzed data for a different SNP array, the Metabochip. Visualize the summary statistics both as a Manhattan Plot and as a Q-Q plot. What chromosome(s) appear to have genome-wide significant hits?

Hint: Download the data using the "LDL Cholesterol"" link found under the "Analysis of Metabochip Data" heading of the GLGC page from Problem 1.

Problem 3

Compute the measure of systematic inflation (λ_{GC}) associated with the summary statistics in Problem 1 and Problem 2. For which SNP array are the summary statistics more inflated?

Hint: $\lambda_{GC} = \text{median}(\chi^2)$ / `0.4549364` where the denominator is obtained from the median of a Chi-square distribution with 1 degree of freedom: `qchisq(0.5,1)`