

# BST227 - Homework 1

*Due: Monday, October 30*

## 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 bad cholesterol and decide to investigate what genes may be associated with bad cholesterol. Here, we'll do a basic exploration of the data before investigating the specific genetic effects in subsequent lectures.

## Problem 1

**Access the LDL summary statistics from the GLGC Consortium's 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: use the **qqman** package in **R** to visualize the data, which can be accessed through **LDHub**.*

## 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: use the Metabochip summary statistics can be found on the same page as the GWAS summary statistics*

### Problem 3

Compute the measure of systematic inflation ( $\lambda_{GC}$ ) 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 last number comes from `qchisq(0.5,1)`