BST 227 Final Project
Fall 2018

Assignment:
The goal of this assignment is to provide a taste for what analyzing genome-wide association data may look like in a research context. For each group, we have simulated genetic data from two differential ancestral groups with variable disease status for a particular chromosome. We've assigned you into groups of 4-5 and hope that you'll work together and collaborate in solving the data analysis challenge while creating a narrative associated with your work. In particular, we expect you to consider the following elements in working on this project:

1. Provide an introduction to your disease and discuss why it may (or may not) have been a good candidate for GWAS. Do heritability estimates exist for this disease? If so, relate this value to other complex disorders and interpret it in the context of performing a genetic association study.
2. Examine the provided data set and provide relevant QC metrics for its characterization.
3. Perform per-variant associations to examine the relationship between a variant and disease status. Consider appropriate covariates to include in your model and what impact their exclusion may have on your results.
4. Interpret the top association markers from your analysis. In particular, show plots of the summary statistics both globally (i.e. a Manhattan plot) and locally (around the top associations using, for example, http://locuszoom.org/). Does the effect size vary between the ethnic groups, sex, etc?
5. Discuss the possible biological activity of the associated variant with respect to interesting genetic features near the variant of interest. Have these genetic loci been implicated in other disorders?

Though not required, consider some of the following ways you can further characterize your disease and association results.

1. Consider what other complex phenotypes may be related to your trait by examining its genetic correlation with results from http://ldsc.broadinstitute.org/. Discuss the effect of ancestry on these analyses.
2. Examine your top association markers for potential functional enrichment here: http://www.regulomedb.org/.
3. Variants can also be associated with variable epigenetic markings. Though less accessible, do any associated variants from your study also associate with differential epigenetic signal found here: http://eqtl.uchicago.edu/Home.html? Consider the cell types used for these analyses and whether those would be particularly relevant in your disease of interest.

Assessment:
All groups will present in class on **Wednesday, December 19**. The presentation should be ~7-10 minutes with 3-4 minutes of Q/A. *We encourage questions from other students for other groups*.

Additionally, we ask that you provide a write up of your project and presentation, including key figures. Please submit pdf versions of the report and presentation via Canvas (one per group) before the end of day, **Wednesday, December 19**.