Lecture 7 Exercises:

ALL of the code and scripts are inside of the file: Lecture_7_code.txt. Because the code for doing some of these tasks is quite involved, please copy and paste from that file, but do make efforts to understand what the code is generally doing.

Exercise 1. As in lecture 4, using PLINK, extract the SNPs from the transferrin data set that are within the 60kb of the *TF* gene and recode them in additive fashion (i.e. as 0, 1, 2 for the number of minor alleles).

Recall: From figure 2 of Benyamin et al. (2009), the coordinates for the region of interest is chr 3 from 134840K to 135052K.

Hang onto this as we will use it again in a moment.

Exercise 2. Within plink, test for GxE interaction between the SNPs and the dichotomous covariate in covD.dat. Please do this genome wide and then also using only the SNPs within 6 0kb of the *TF* gene.

Exercise 3. Now try to use the same code to do GxE analysis with the continuous covariate in covC.dat. What happens?

Exercise 4. Now let's open up R. Read in the SNPs inside of the TF gene (within 60kb) that you exported earlier. It's not imperative, but it would be good to clean up missing values again.

a. Test for interaction between each of the SNPs and the continuous covariate of interest using standard 1-df test.
b. Now, let's test for GxE using 2-df test. What happens? How do you interpret the results?

Exercise 5. Gene-gene interaction testing. Tr_dich.pheno is a dichotomized version of the transferrin levels. Let's do GxG analysis using this alternative phenotype in plink.

a. Let's first try looking across all possible pairs of SNPs. (you can try to stop it using control-C)
b. Instead of looking at all SNPs, let's only focus in on the SNPs in the TF gene.
c. Try doing case-only analysis with the SNPs in the TF gene. Does it run? Do you think it should run? Try tricking it into running.