

Session 3: Introduction to the PLINK software for GWAS

Performing GWAS for Transferrin and Height with PLINK

1. Using R, investigate the Transferrin.bim and Transferrin.fam files. How many individuals are included in the study. How many SNPs are available? (Hint: Use the R-script that is available on the course website)
2. Also investigate the adjusted transferrin phenotype values in the file “Tr.pheno” with R. How many individuals in the study have transferrin measurements? What is the distribution of the transferrin phenotype? (Hint: Plot a histogram of the transferrin.) Similarly investigate the adjusted height phenotypes in the file “Ht.pheno.”
3. Using PLINK, perform a GWAS of transferrin using the phenotype file “Tr.pheno” and the Transferrin.bed, Transferrin.bim, and Transferrin.fam files. For the association analysis, use the following quality control threshold filters: minor allele frequency > 0.05 , at least a 99% genotyping call rate (less than 1% missing), and HWE p-values greater than 0.001. (Hint: Use a similar command to what is given in the PLINK intro slides).
4. Now perform a GWAS of height (using the phenotype file “Ht.pheno”) with PLINK. Use the same QC criteria given in question 3.
5. Using R, read in your association results for transferrin and height from PLINK. Make Manhattan plots and Q-Q plots of the association results. (Hint: Use the R-script available on the course website). Compare your Manhattan plot for Transferrin to the Manhattan plot given in Benjamin et al. (2009, AJHG) that is posted on the module website.
6. Using R, obtain the top ten most significant SNPs for transferrin. Are the SNPs in LD? (Use the commands in the R script). Create a text file containing the SNP names of the top ten SNPs, then use PLINK to obtain the R^2 measures of LD for the SNPs. (Hint: Use a similar command to what is given in the PLINK intro slides).