Brad Olinger

Quant Bio

Week 3

9/24/21

1. Visualize genetic relatedness between the strains by performing principal component analysis and plotting the first two components.

Note: In Python notebook.

Command: module load plink

Creating a bed file and plink files

Command: plink --vcf genotypes.vcf --make-bed

#make proper files from vcf file

plink --vcf genotypes.vcf --make-bed

#generata PCa objects (ped, map files)

Command: vcftools --vcf genotypes.vcf --out my\_data –plink

Run Plink

##### Command: plink --file my\_data

Generate eigenval and eigenvectors for PCA

plink --vcf genotypes.vcf --pca 10

Note: Imported eigenvect to Python, plotted PC1/PC2

1. Visualize the allele frequency spectrum by plotting a histogram of allele frequencies.

##### Command: plink --file my\_data --freq

Above line generated plink.fq with allele frequencies

Python contains code for histogram.

1. Using plink, perform quantitative association testing for each phenotype. Use the top 10 principal components (eigenvectors) as covariates in your analysis, to adjust for non-independence due to relatedness.
   * Be sure to use the --allow-no-sex option
   * You may find this portion of the [plink documentation](https://zzz.bwh.harvard.edu/plink/anal.shtml) helpful for performing association testing on each of the phenotypes

Command that doesn’t work: plink --vcf genotypes.vcf --allow-no-sex --pheno CB1908\_IC50.txt --assoc --covar plink.eigenvec --out results

#create qassoc file with p values per snip and phenotype

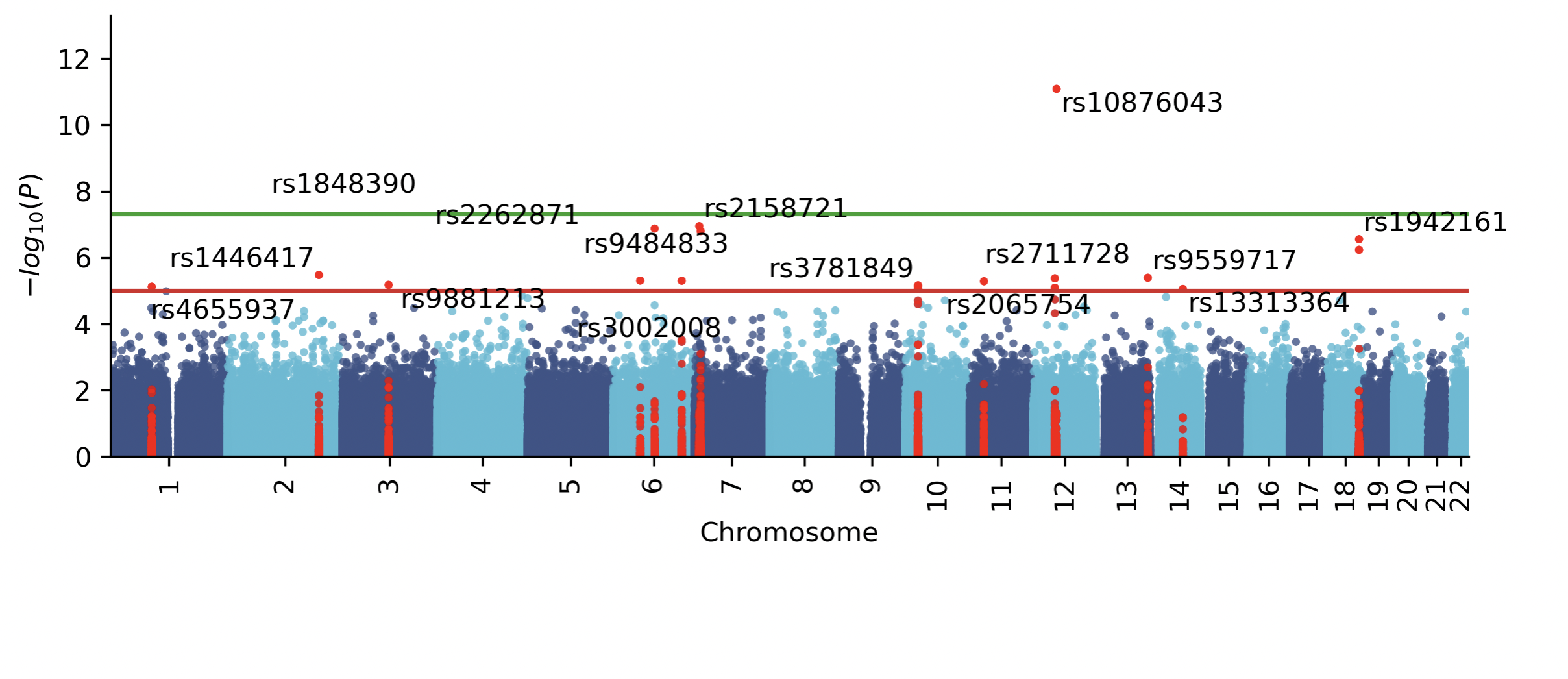
Command: plink --vcf genotypes.vcf --allow-no-sex --pheno GS451\_IC50.txt --assoc --covar plink.eigenvec --out resultsGS

Command #2: plink --vcf genotypes.vcf --allow-no-sex --pheno CB1908\_IC50.txt --assoc --covar plink.eigenvec --out resultsCB

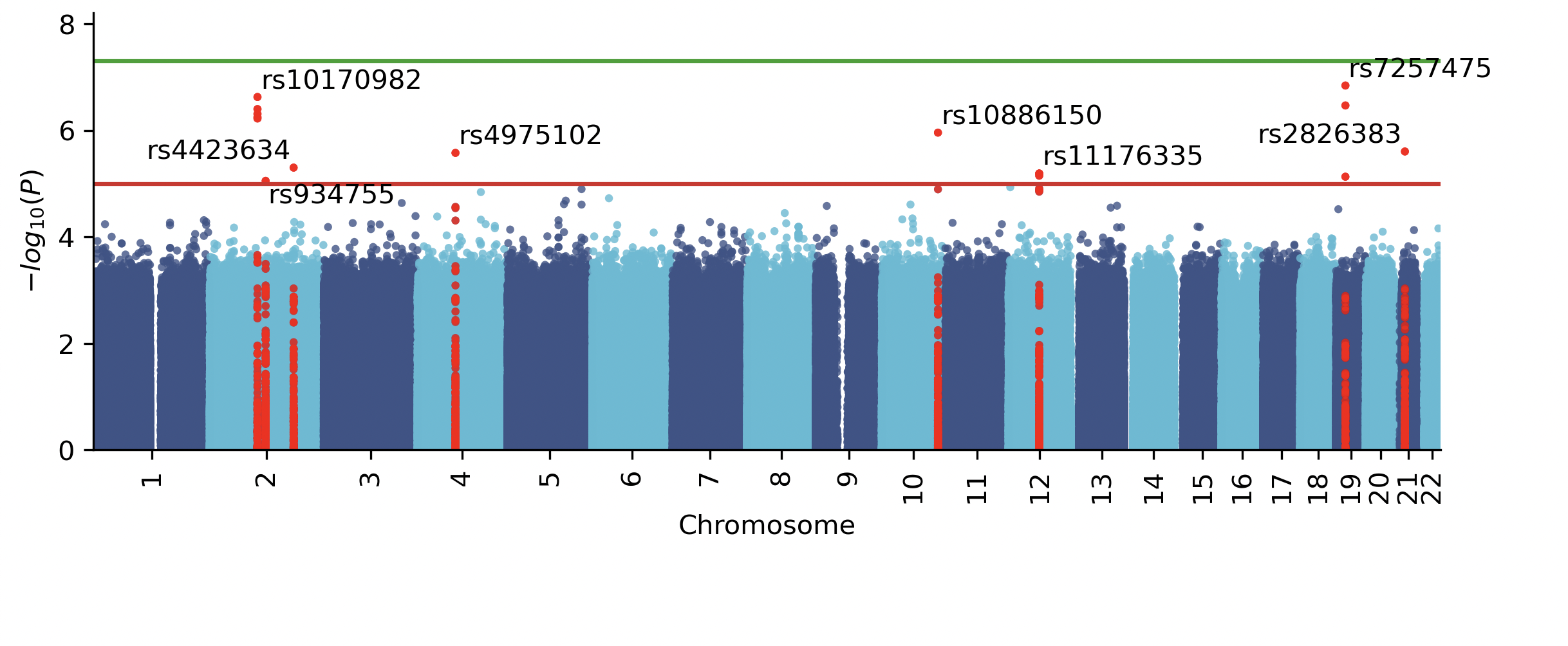
1. For each phenotype, produce a QQ plot and Manhattan plot. For the Manhattan plot, highlight SNPs with p-values less than 10-5 in a different color.

Note: code and QQplots are in Python; Manhattan plots are below:

CB drug:



GS drug:



1. Choose one of the traits for which you performed GWAS. For the top associated SNP, visualize the effect size by creating a boxplot of the phenotype stratified by genotype.

Note: this is in python