GWAS Implementation

Benny Xie (A17465776) Yifei Ding(A17007213) Marina Xu (A16235362)

The goal of this project is to build a tool in python that can perform GWAS provided with genotypes and phenotypes of a population. It will produce a Manhattan and a QQ plot of the results. We will compare our tool to plink in lab 3.

Our program will accept the same input as plink –vcf <vcf> –phenon <phen>, calculate p-value and generate the corresponding Manhattan and QQ plot with a Minor Allele Frequency filter and a count filter.

First, to ensure the correctness of our tool, we will compare the ranking of the SNP p-values and the plots from our tool and from plink. Second, we will compare the amount of time each program used to perform GWAS.

We will use the genotype (lab3_gwas.vcf.gz) and phenotype (lab3_gwas.phen) data from lab3 for benchmarking. Another dataset we could use is the UK Bio Bank (https://www.ukbiobank.ac.uk/enable-your-research), We have submitted an application that's currently pending approval. In addition, we will generate random and non-random phenotype data as additional test cases.