Detailed Notes on Simulation Study for Paper

Andrew W. George

20 December, 2016

**Creation of Simulated Data sets for W S L A HS and HL scenarios**

\*.vcf.gz files located at /data/geo047/1000GenomeData/ on Bragg

Scripts need to convert these \*.vcf.gz snp genotype files into the different sized scenarios are in /home/geo047/MWAM/1000GenomeData/Scripts. The scripts contained in this directory produces the genotype data for the simulation study.

The phenotype data is created, from the genotype data, with the R command /home/geo047/MWAM/SimStudy/Timing/RScripts/gendata.R

Make sure you read the file

/home/geo047/MWAM/SimStudy/Timing/RScripts/README that has the instructions for the three steps to creating the phenotypic data with 100 reps in the flush1 directory.

**Running Simulation Study**

The slurm scripts for submitting jobs are in /home/geo047/MWAM/SimStudy/Timing/RunJobs/

Due to space requirements, I had to be careful what was running when. For A and HL, I could only run 30 replicates at a time. To run fast.slurm and fastALL.slurm, need to FIRST run /flush2/geo047/AM+/Timing/GenoData/PLINK.bash to create mydata files.

**Creating Results Files for Paper**

1. Timing and Memory data files
   1. Run extract\_mem.bash and extract\_time.bash from /home/geo047/MWAM/SimStudy/Timing/Results/
   2. Run form\_results.R from /home/geo047/MWAM/SimStudy/Timing/Results/ which creates memory.dat, cputimes.dat and abscputimes.dat files in this directory
   3. On my MAC, transfer files into ~/Papers/AM-Paper/AM+/Plots\_for\_Paper/
2. Results files for Power Calculation
   1. Core R file is in /home/geo047/MWAM/SimStudy/Timing/RScripts/ and