**Homework 5**

**Statistical Genomics**

**CROPS 545, Spring 2017**

Professor: Zhiwu Zhang

Due on April 12, 2017, Wednesday, 3:10PM, PST

**Data files**: mdp\_numeric.txt from GAPIT demo data. The data file can be download from <http://www.zzlab.net/GAPIT/GAPIT_Tutorial_Data.zip>. The data contains 281 individuals (row wise) and 3093 SNPs (column wise) coded as 0/1/2. The SNP ID, chromosome and position is indicated by a file named mdp\_SNP\_information.txt

**Hand in:** Email your report (PDF, limited to five page) and R source code (text file) with subject of “CROPS545 HW5” to Zhiwu.Zhang@WSU.edu. Name your files as following:

Homework5\_ firstname\_lastname.pdf and Homework5\_ firstname\_lastname.R

**Grade components**: 1) Hypothesis or statement; 2) Results; 3) Methods; 4 presentation; 5) R source code (clarity, simplicity and documenting comments)

**Objectives**: 1) Examine GLM package from previous homework; 2) Statistical power vs FDR and type I error; 3) mapping resolution; and 4) GWAS methods (GLM, MLM, SUPER, MLMM and FarmCPU).

1. Sample 10 QTNs from the genetic markers and simulate QTN effects from a standard normal distribution. Assign genetic effects for each of the 281 individuals. Simulate normal distributed residual effects with appropriate variance to have a heritability of 0.75. Add residual effects to the genetic effects to create phenotypes. Use the GLM GWAS package you developed in homework 4 to perform association analyses with three PCs included as covariates. Compare your package with GLM in GAPIT with same model (20 points).
2. Analyze above result of GLM in GAPIT and examine statistical power vs. FDR and Type I error at mapping resolution of one base pair (20 points).
3. Repeat (2) 100 times and make comparisons between mapping resolutions of one base pair and 100 KB (20 points).
4. At mapping resolution of 100 KB, compare statistical power by using GLM, MLM and SUPER in GAPIT with at least 30 replicates (20 points).
5. At mapping resolution of 100 KB, compare statistical power by using MLMM and FarmCPU with at least 30 replicates (20 points).

**Extra credit**

1. Find another method (BLINK excluded) and demonstrate that it has higher statistical power than FarmCPU (50 points. Report is limited to one extra page).