**Homework 5**

**Statistical Genomics**

**CROPS 545, Spring 2020**

Professor: Zhiwu Zhang

Due on April 13, 2020, Monday, 3:10PM, PST

**Data files**: You can use either the same dataset you used in homework 2, or switch to a different dataset with same requirements.

**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](mailto: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 number of QTNs of your choice from the genetic markers used in homework2 and simulate QTN effects from a standard normal distribution. Assign genetic effects for each individual. 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. Count number of false positives for identifying half of your QTNs (20 points).
2. Repeat simulation in (1) 30 times and exam statistical power vs. FDR at mapping resolution of 100,000 base pairs (20 points).
3. Repeat simulation in (2) with additional two levels of heritability (0.25, and 0.5) (20 points).
4. Repeat simulation in (2) with additional two methods of your choice among MLM. CMLM, SUPER, FarmCPU, and BLINK (20 points).
5. Among above question (2-4), redo one of them by excluding the QTNs from the marker dataset before GWAS. Describe the impact of inclusion and exclusion of the QTNs (20 points.

**Extra credit**

1. Find another method and demonstrate that it has higher statistical power than BLINK (50 points).