**Homework 3**

**Workshop of Genome-Wide Association Studies**

**Wuhan, July 4-9, 2016**

Professor: Zhiwu Zhang

Due on July 21, 2016, 5:00PM, Beijing time

**Objectives**: 1) Statistical power vs FDR and type I error; 2) mapping resolution; 3) multiple test correction; and 4) GWAS methods (Correlation, GLM, MLM, and FarmCPU).

**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 “GWAS2016HW3” to Dr. Xiaolei Liu ([xll19870827@hotmail.com](mailto:xll19870827@hotmail.com)). Name your files as following:

Homework3\_ firstname\_lastname.pdf and Homework3\_ firstname\_lastname.R

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

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. Perform GWAS with correlation method. Present Manhattan plot and label the cutoff of 1% after Bonferroni multiple test correction. For the first QTN with signal below the cutoff, perform permutation test and compare the signal with permutation cutoff (20 points).
2. Based on a resolution of 500 kb window, evaluate power vs. FDR and power vs. type I error on the association test in (1) on the simulated phenotypes before the permutation (20 points).
3. Redo (2) by excluding QTNs from the genotype before association tests (10 points).
4. Repeat (2) 100 times and evaluate the averages of power vs. FDR and power vs. type I error (20 points).
5. Compare (4) with methods of incorporating three PCs in GLM, MLM and FarmCPU (30 points).