Name:

Graded by: TA

**TOTAL POINTS: /100**

Problem 1: The package should contain at least three inputs: y, X, and C that are R numeric data frame objects. Their dimensions are n by 1, n by m, and n by t, corresponding to phenotype, genotype, and covariate data, where n is number of individuals, m is number of markers, and t is number of covariates. The function should return probability values with dimension of 1 by m for the association tests between phenotypes and markers. Markers are tested one at a time with covariates in C included as covariates (15 points).

* Has at least the three input R objects (3 points)
* Full Points
* Generate P values (6 points)
* Full Points
* Commenting (3 pts)
* Full Points
* Bug-free (3 pts)

Full Points

Comments:

POINTS: 15

Problem 2: The package should perform PCA and incorporate PCs as cofactors for GWAS. Your package should also automatically exclude the PCs that are in linear dependence with the covariates provided by users. (25 points).

* PCA (5 pts)

Full Points

* Automatically exclude PCs (10 pts)

Full Points

* Commenting (5 pts)
* Full Points
* Bug-free (5 pts)

Full Points

Comments:

POINTS: 25

Problem 3: Develop a user manual and tutorials. Name your package and create a logo. (20 points).

* Name package (1 pt)

Full Points

* Logo (1 pt)

Full Points

* Table of contents (1 pt)

Full Points

* Description of package functions (5 pts)

Full Points

* User guidance (5 pts)

Full Points

* Common questions and answers (5 pts)

Full Points

* Clarity (2 pts)

Full Points

Comments:

POINTS: 20

Problem 4: Perform GWAS on the data provided or your own data, which must contain cofactors (15 points).

* Manhattan plot (2 pts)

Full Points

* QQ plot (2 pts)

Full Points

* Genome-wide threshold (2 pts)

Full Points

* List of associated SNPs (2 pts)

Full Points

* Comments on MAF of associated SNPs (2 pts)

Full Points

* Commenting (2 pts)
* Full Points
* Bug-free (3 pts)

Full Points

Comments:

POINTS: 15

Problem 5: Demonstrate that your method is superior to the competing method (GWASbyCor) through simulation (25 points).

* Conduct the replicates correctly (5 pts)

Full Points

* Valid criteria for comparison (5 pts)

Full Points

* Make comparison conclusion based on statistical inference (5 pts)

Full Points

* Commenting (5 pts)
* Full Points
* Bug-free (5 pts)

Full Points

Comments:

POINTS: 25

Problem 6 (**Extra Credit**):

POINTS: 5

Problem 7 (**Extra Credit**):

Demonstrate that your package is better than BLINK C version (http://zzlab.net/blink) in either statistical power or speed (25 points).

You get either no credit or full credit if all of the following are satisfied:

1. Your package is either better on power or faster in speed.
2. The conclusion is based on at least 30 replicates.
3. At least three datasets are used for comparison.

Comments:

POINTS: 25