Final Project for MDSC 403

**Genotype-phenotype association mapping**

Please download the genotype file (genotype.csv) and the phenotype file (FT10.txt) from the D2L. (You have analyzed the genotype file in the Linux session). Try to find out the association between the genotype and phenotype. The associations may be detected by any statistical approaches as long as you can justify them. Feel free to use PLINK or R packages or anything you’d prefer.

Please (1) identify your research goal, (2) design your pipeline of the study, (3) implement the pipeline, (4) conduct the analysis, and (5) interpret the results based on the gene models (GFF file in D2L) and online information regarding the annotation of the genes. Write a report on the outcome of your research project. Please provide sufficient details to ensure that your outcome is replicable by others who may want to follow your work.

Grading rubric of the final report:

* (20%) Provide the rationale for the proposed analytic pipeline. Hypothesis and Objectives are clearly stated.
  + What is the goal of this work?
  + Why you think this project is feasible?
* (40%) Methods and pipelines are clearly presented so that the readers can replicate the whole work based on the same data and your report. This includes:
  + A general illustration of the pipeline should be concise and to the point
  + Detailed codes are attached, clearly linked with the general illustration.
  + If existing tools are used, please specify the parameters used.
* (40%) The main findings (if any) are clearly stated and well interpreted. This includes:
  + Figures are polished and figure captions are informative.
  + Statistical justifications are sound.
  + The text justifying the findings should integrate figures and statistics.
* (Bonus 20%) In the event you have tried multiple methods, provide a discussion of the trade-off between alternative analytic methods. What lessons are learnt during your exploration?