Homework 4

The homework uses data from the FAMuSS study (Functional SNPs Associated with Muscle Size and Strength). The data are available as tab-delimited text file. Background information can be found in Thompson et al. 2004.

Q1: Look up what is the required input data format for Haploview. Obtain a subset of data containing as trait pre.BMI>25 and as genotypes the SNPs within the resistin gene, in this format.

Q2: Use Haploview to study the LD and haplotype block structure within the resistin gene. Also assess association of haplotypes within this gene with the trait pre.BMI>25.

Q3: Construct and prune a regression tree predicting NRDM.CH from the SNPs within the gene, treating the SNPs as three-level categorical variables. Interpret your findings. Which is the most predictive variable? Repeat your analysis treating the SNPs as ordinal variables. Discuss how the results differ from the previous ones.

Q4: Construct and prune a regression tree predicting NRDM.CH from the SNPs within the actn3 and resistin genes. Repeat your analysis including Race as a potential predictor. Discuss your findings

Q5: Grow a random forest predicting NRDM.CH from the SNPs within the akt1 gene. Restrict your analysis to Caucasians. Assess missing genotype rates for the different SNPs. Apply single imputation before growing the forest. Interpret your results.

Q6: Repeat your analysis for the complete set of individuals. It is known that genotype frequencies can vary substantially in different racial and ethnic strata. Carefully consider how you should acknowledge this in your analysis.