Exploratory CART analysis of BMI:
Mining gene-gene interactions and generating hypotheses

Tae-Hwi An

Mentors:
Ingrid Borecki, Ph.D.
Michael Province, Ph.D.

Abstract

Obesity is an important public health problem that is highly prevalent in developed part of the world. It is a complex trait that is mediated by gene-gene and gene-environmental interactions. There are many mathematical methods to model epistatic interactions among/between genes and covariates but are often computationally expensive. Classification and Regression Tree (CART) was used in exploratory analysis to generate hypotheses of inter and intra gene-gene interactions of chromosome 7 and 13 regions for Body-mass index (BMI) phenotype. A case-control sample of Family Heart Study (FHS) (n=861) was used to generate regression trees and the randomly selected Framingham Family Heart study (n=6284) sample was used for validation of the proposed models. Imputation was done to impute SNPs present on FHS sample but not on Framingham sample. Using CART, initially overfitted regression trees were obtained then pruning was done to increase predictive power of the trees. By using SNPs with marginal additive association to BMI phenotype (p<0.01), the search regions on both chromosomes were reduced. The three proposed regression trees explained about 2% of BMI phenotypic variance in the FHS sample but failed to duplicate in Framingham sample. Bootstrapping was done to stabilize sampling variation and to get a more accurate test statistic. Bootstrapping resulted in chr7 (p=0.455), chr13 (p=0.295), and chr7&13 (p=0.325). These findings do not support the validity of the proposed regression trees, but sample ascertainment, imputation, and sampling variation between iterations of bootstrapping may have influenced the finding.