IDENTIFICATION OF THE GENE-GENE INTERACTIONS IN THE PRESENCE OF MISSING DATA : MULTIFACTORDIMENSIONALITY REDUCTION APPROACH

Y. Chung¹, S.Y. Lee², <u>T. Park^{$\dagger 1$}</u>

¹Seoul National University, Seoul, Korea ²Sejong University, Seoul, Korea

[†]E-mail: *tspark@stats.snu.ac.kr*

For common complex diseases such as diabetes and hypertensions, the effect of single genetic variation will be likely dependent on other genetic variations (gene-gene interaction) and environmental factors (gene-environment interaction). Thus, an identification and characterization of susceptibility genes is a challengeable task, in which To address this issue, the multifactor dimensionality reduction (MDR) has been proposed and implemented by Ritchie et al. (2001), Moore et al. (2002), Hahn et al. (2003) and Ritchie et al. (2003). The MDR approach effectively reduces the dimension of multilocus genotype predictors from n to one, which improves the identification of polymorphism combinations associated with disease risk. Unfortunately, the original MDR approach cannot handle appropriately the incomplete data with missing observations, in which missing observations are treated as an additional genotype category. However, this approach may suffer from a sparseness problem. When high-order interactions are considered, an additional missing category would make the contingency table cells become more sparse. We propose an extension of the MDR approach by considering missing data over all possible multifactor classes. We compare the proposed method with the original MDR method through simulation studies.