

“Statistical methods for analyzing high-throughput genomic data”

PI: Guan-Hua Huang (National Chiao Tung University)

Role: Principal investigator

National Science Council of Taiwan: 08/01/2009-07/31/2011

Novel statistical methodology can enhance understanding of the interactions between multiple genes and environmental factors on a complex disease. The massive amount of high-throughput genomic data brings a great challenge of developing advanced statistical and computational data mining tools. In this project, we make efforts to develop effective statistical methods for analyzing these high-throughput data. The project focuses on two types of high-throughput data: gene expression microarray and single nucleotide polymorphism (SNP) markers. In gene expression microarray analysis, we will evaluate combinations of the most popular preprocessing and differential expression methods in terms of validity and reliability. In the candidate-gene approach of SNP marker analysis, we will consider various SNP tagging criteria, haplotype block definitions and association tests, and estimate the power and type I error of these combinations. In the genome-wide approach of SNP marker analysis, we will develop and compare methods to search for a set of marker loci in different genes and to analyze these loci jointly.

Keywords : Candidate-Gene Association Study, Gene Expression Microarray, Genome-Wide Association Study, High-Throughput Genomic Data, Single Nucleotide Polymorphism Markers