“National research program for genomic medicine: advanced bioinformatics core for genomic statistics for complex diseases”
PI: Chun-Houh Chen (Academia Sinica)
Role: Co-principal investigator
National Science Council of Taiwan: 05/01/2005-04/30/2008

Novel statistical methodology can enhance understanding of the interactions between multiple genes and environmental factors, as well as therapeutic or preventive effects, on a complex disease. The massive amount of high-throughput microarray, SNPs and other biological data bring a great challenge of developing advanced statistical and computational data mining tools. In the Genomics Statistics Team for Complex Disease, we continuously make efforts to develop effective statistical design/analysis methods via a comprehensive data management/analysis platform and provide statistical consulting services for the genomic research based on genotypes and clinical phenotypes by adjusting for the physiological and pathogenetic variations among study subject.

We have developed a user friendly statistical platform Gene Expression Study Design and Analysis Suite (GESDAS) for microarray gene expression studies, and will expand the platform GESDAS to the Gene-Environment Analysis Refining System (GEARS), by enlarging its capacities for SNPs data analysis, linkage analysis and association studies, information mining and visualization in genotype-to-phenotype mapping, diagnostic accuracy assessment, dose-response study, and age-at-onset/survival analysis.

We are devoted to routine service, technological research & development, and collaborative research with the following specific aims:

1. Continue to provide and expand statistical consulting services, based on our established software suite GESDAS for design, management and analysis of gene expression experiments, using one-channel oligonucleotide arrays or two-channel cDNA arrays, towards the quality standards of the MIAME (Minimum Information About a Microarray Experiment) Guidelines. More new methodologies with user friendly analysis interfaces will be updated and enhanced.
2. Provide statistical practice of data mining on classification or clustering of disease heterogeneity with genotypic, endophenotypic, phenotypic and clinical profiles for designated human diseases from database repository in an integrated platform. In particular, state-of-the-art information mining and visualization

methods, including the continuous and categorical versions of the Generalized Association Plot, statistical learning methods and nonlinear dimension reduction techniques, latent class models and their improvements, will be advocated.

3. Elaborate new statistical methods and software packages for multipoint linkage and association mapping analysis, and for age-at-onset analysis to investigate the gene-environment interactions on phenotypes of complex diseases in case-control or familial studies. Robust methods without artificial assumptions for flexible data types will be implemented.
4. Establish clinical information systems for important disease categories and ultimately link them to functional genomics databases by developing standard operation procedures and collaborating with informatics technology team concerning metadata structures with web-browser interfaces.
5. Build a prototype system to integrate information between human-disease-specific animal models and high-throughput genomic researches, and to establish useful statistical tools for the design and analysis of animal experiments as well as comparative genomic studies. The constructed animal models are essential for biomedical researchers to validate genetic findings of human diseases or to study system biology in vivo.

By active and efficient cooperation with teams of information technology, functional and comparative bioinformatics in the Advanced Bioinformatics Core, this team will collaborate with researchers in genomic medicine to promote diagnoses and prognoses of major disease areas.