Statistical Science for Genetics and Genomics

Chair: Karen Kopciuk (Cancer Epidemiology and Prevention Research, AHS/University of Calgary)

SHOFIQUL ISLAM, McMaster University

Comparing the Performance of Linear and Nonlinear Principal Components in the Context of High Dimensional Genomic Data Integration

We integrate phenotype, gene expression and miRNA expression data to classify tumor or death of Lung Cancer patients. We apply linear and kernel principal component analysis (PCA) to reduce the dimension and subsequently integrate three domains based on several nonlinear approaches including logistic regression. We also conducted an extensive simulation to compare the performances. We observe that a reduced set of kernel principal component carries more information from a genetic process but failed to show better predictive ability in this occasion. Reducing dimension with linear PCA and a logistic regression model for classification seems to be sufficient for this purpose.

GUN HO JANG, Ontario Institute for Cancer Research

A Copy-Number Recalibration Method using Object Integration for Next Generation Sequencing Data

Both point mutations and copy number alterations are widespread characteristics of tumour. Whole genome sequencing is one of the most popular technologies to detect both of them. Many methods have been developed for calling copy-number alteration using reads counts and non-reference allele frequencies. The differences in normalization method and presumed statistical models result in different outcomes. An object integrated copy-number recalibration method is proposed to evaluate and improve copy-number calls. The performance of the proposed method is presented and compared with several copy-number callers including HMMcopy, TitanCNA and Control-FREEC.

LI LI, Dalhousie University

A New Approach to Detect Genetic Recombination in Phylogeny

Genetic Recombination is a process where parts of different genes are combined to form a new gene. Recombination has large effects on the phylogenetic analysis of DNA sequences, and needs to be identified correctly. We propose a new algorithm for recombination detection based on the site log-likelihood matrix. The method is computationally efficient and shows great ability to detect recombination events. Performance of the method is evaluated on both simulated and real data examples. Results are compared with other methods. abstract

KUN LIANG, University of Waterloo

A Hidden Markov Tree Model for Multiple Hypotheses Testing of Gene Ontology Gene Sets

Testing predefined gene categories has become a common practice for scientists analyzing high throughput transcriptome data. A systematic way of testing gene categories leads to testing hundreds of null hypotheses that correspond to nodes in a directed acyclic graph. The relationships among gene categories induce logical restrictions among the corresponding null hypotheses. Under a hidden Markov tree model, we develop a computationally efficient method to incorporate the dependence information among the null hypotheses. Our method provides more powerful results than existing methods that honor the logical restrictions.

FARHAD SHOKOOHI, McGill University

DNA Methylation Analysis; A Thorough Comparison of Available Analytic Tools

DNA methylation studies have increased in number recently thanks to recent advances in next-generation sequencing and microarray technology, providing many data sets at high resolution, enabling researchers to understand methylation patterns and their regulatory roles in biological processes and diseases. Diverse methods and software create ample opportunities for researchers to do quantitative analysis, however, this also makes it difficult for practitioners to choose one that is suitable and efficient in analyzing the data. The goal of this talk is to compare the available analytic tools using real and simulated data sets to give insights into their validity, efficiency and utility.

THIERRY CHEKOUO TEKOUGANG, UT MD Anderson Cancer Center
A Bayesian Predictive Model for Imaging Genetics with Application to Schizophrenia

In this talk, I'll introduce an integrative Bayesian risk prediction model that allows to discriminate between schizophrenic patients and healthy controls, based on a sparse set of discriminatory ROIs (region of interests) and SNPs (single-nucleotide polymorphism). Inference on a regulatory network between SNPs and ROI intensities (ROI-SNP network) is used in a single modeling framework to inform the selection of the discriminatory ROIs and SNPs. Using simulation studies, we assess the performance of our method, and we apply it to a schizophrenia dataset. We confirm that some biomarkers involved in the ROI-SNP network are more likely to be discriminating.