



Statistics Seminar

Statistical and Bioinformatic Analysis of Breast Cancer Co-Methylation Pattern

Shuying Sun

Friday, March 26

12:00 - 1:00 pm

Abstract: Breast cancer is one of the most commonly diagnosed cancers. It is associated with DNA methylation, which occurs when a methyl group is covalently added to a cytosine base in a DNA sequence. DNA methylation can have different effects on gene functions and is therefore associated with breast cancer. The methylation patterns of different genes or sites in a genome may be correlated in certain ways. This correlation pattern is known as co-methylation. However, it is still not clear how different genes co-methylate with each other in the whole genome. Two previous studies are conducted based on relatively smaller datasets, Illumina methylation 27K data. That is, their findings are based on methylation signals of 27,000 probes or sites for each sample. In this study, we will analyze much larger publicly available datasets, Illumina methylation 450K (i.e., with 450,000 sites) array data of 53 breast cancer samples. We will show our analysis results on overall co-methylation patterns with relation to physical distance, sign (i.e., positive or negative correlation), and number in normal and tumor datasets. We will show specific sites whose co-methylation patterns change significantly from normal to tumor samples. We will report relationships of genes related to these sites. We will also show the super-connector sites/genes that have high correlations with many other genes and report pathway analysis results. Due to the large data size, our analysis is computationally challenging. For example, because of the limit of R, we cannot even create a 450,000 by 450,000 matrix to store our correlation analysis results. Our ability of analyzing datasets of this size can provide researchers with a new and improved understanding of co-methylation patterns in breast cancer. Our new findings will furthermore allow researchers to help establish relationships and associations between different genes in the future.

Bio: Dr. Shuying Sun received her Ph.D. in statistics from the University of Toronto in Canada and is currently an associate professor in the Department of Mathematics, Texas State University. Dr. Sun has interests in statistical genetics and bioinformatics and has published 30 peer-reviewed research articles in these fields. Dr. Sun's research focuses on addressing challenging genetic and epigenetic questions using statistical and computational methods. She has also been developing statistical methodologies and software packages for genomic problems using Bayesian methods, hidden Markov models, Markov Chain Monte Carlo algorithms, and linear models.