Discovery of Novel Genes Linked to Breast Cancer Progression Using a Bayesian Framework

Khatri, Shania

With over 19,000 genes in the human genome, an effective method to link genes to the progression of cancer is needed. The identification of even one novel significant gene in the progression of breast cancer can lead to improvement in breast cancer monitoring and treatment. In this study, I extracted and used a large public gene expression data set for breast cancer progression to learn gene-gene and gene-stage interactions. I created a set of known genes from a literature review on breast cancer progression and combined them with genes chosen by feature selection to create my data set. I built Bayesian networks to understand the relationships between gene expression and breast cancer stage, with parsimony and biological accuracy being the main goal in model selection. The genes selected through literature review were organized into a Bayesian structure which was used to validate associations in my models. In my model, all five genes that are direct parents of stage have been biologically validated as related to breast cancer progression. My study demonstrated that knowledge-based Bayesian models can be used on public data to yield parsimonious and accurate networks. My computational modeling finds genes relating to the progression of breast cancer, including a novel gene CLDN11, which has since been shown to be a biomarker for human breast cancer. My results validate the use of this framework as a novel gene discovery tool and suggest its use for accurately finding drug targets and biomarkers for disease.

Awards Won: