

# The QTL-Shielding Test (QST): a Novel Technique for Genetic Network Discovery in a Microarray/Marker Dataset

There has been great interest in reverse engineering genetic networks from large genomic datasets in recent years. Particularly well-suited for genetic network analysis is the e-QTL or Expression-Quantitative Trait Loci dataset in which marker data and microarray data are collected for a set of segregating individuals from an experimental cross. Bayesian Networks is suited to the problem of genetic network inference since it can potentially infer causal relationships from the analysis of static data. However, there are several drawbacks to Bayesian Networks including high computational complexity, large sample size requirements, and strict assumptions that are often not satisfied in practice.

Here we present a technique for inferring genetic relationships from the analysis of expression-QTL (e-QTL) data called the QTL-Shielding Test (QST). The QST has its basis in Bayesian Networks but takes into account the specific structure of the genomic data being analyzed to simplify the problem of network inference. If we have two gene expression variables,  $Y_1$  and  $Y_2$ , and a QTL,  $X$ , with which  $Y_1$  and  $Y_2$  are significantly associated, the relationships among  $X$ ,  $Y_1$ , and  $Y_2$  can be described using one of the models shown in Figure 1. Model 2 is the “QTL-Shielding Model” in which  $Y_1$  shields  $Y_2$  from the QTL  $X$ , or in other words,  $Y_2$  is conditionally independent of  $X$  given  $Y_1$ . If this model can be inferred, then it is implied that  $Y_1$  causes or regulates  $Y_2$ .

In the QST, model selection is used to find gene pairs that satisfy the QTL-shielding model. One application of the QST is to find super-regulators, or genes that regulate the action of many other genes. Another application of the QST that is currently being investigated is the inference of larger networks of genes based on integrating the results of several QSTs. The QST has been applied to the analysis of a *Saccharomyces cerevisiae* dataset and a eucalyptus dataset. In the yeast dataset, key regulatory genes have been hypothesized in pathways related to leucine biosynthesis, daughter cell separation, pheromone response, and stress response. In the eucalyptus dataset, regulatory genes related to the lignin biosynthesis pathway have been predicted.

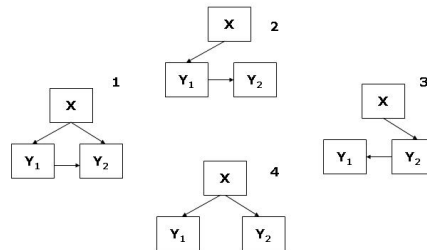


Figure 1: Graphical models describing the possible relationships among genes  $Y_1$  and  $Y_2$  and QTL  $X$ .