

## **Modeling and Identifying Regulatory Modules in (*Glycine max*) Soybean Time Series Gene Expression Data using Bayesian Networks**

Delasa Aghamirzaie, Dhruv Batra, Lenwood Heath, Eva Collakova, Ruth Grene

Much cellular activity is organized as networks of interacting modules. Oftentimes cellular responses can take the form of a network of proteins interacting together under different internal and external conditions. Recently, there has been much interest in reverse engineering these networks using gene expression data. Bayesian approaches have frequently been used to reverse engineer regulatory networks based on transcriptomics data. Gene expression data by itself suffers from high noise and lack of statistical power for inferring causal relationship between genes. Incorporating prior biological knowledge has the potential to improve performance, allowing the production of biologically meaningful networks. It has been shown that incorporating multiple sources of prior knowledge including co-expression and protein-protein interaction data into Bayesian networks creates more predictive and biologically meaningful networks. In this project, a Bayesian network was generated using selected genes from time-series RNA-sequencing-based transcriptomics of developing soybean embryos. A group of genes whose expression changed significantly ( $p\text{-value} < 0.05$ ) in at least one of the time points were identified for Bayesian network learning. We take advantage of multiple data types, including co-expression and co-localization between Arabidopsis homologs of soybean genes from GeneMania ([www.genemania.org](http://www.genemania.org)), and incorporate these data as prior knowledge in order to assess the likelihood of functional linkage between gene pairs. The final Bayesian network yields more biologically meaningful relationships among genes than is obtained from one set of gene expression data only. This is a fresh approach to inferring regulatory network in plant biology.