A Bayesian Network Approach to Building Gene Regulatory Networks

Jeremy Burrell, Chao-Jen Wong and Joan Chen

ACSESS proceedings

2008

CSRC
gateway to the future
an interdisciplinary

SAN DIEGO STATE UNIVERSITY

A Bayesian Network Approach to Building Gene Regulatory Networks

Jeremy Burrell, Chao-Jen Wong, Joan Chen

Abstract

Continuous abiotic stresses (such as extreme temperatures, high winds and edaphic conditions) can have adverse effects on plant life. This can become a major constraint in crop production. In order to alleviate these problems, it is important to understand the cold stress mechanisms at the molecular and cellular levels, regarding the signaling pathways from cold perception to activation of gene expression. We took a system biology approach, aiming to build the gene regulatory network using the cold stress microarray data that have been obtained, to gain a better understanding of plant responses to cold stress at the molecular level. The results generated from this approach will be utilized to generate hypothesis, which can be tested by other experimental approaches.

To discover the gene regulatory network in response to cold stress on plants, we have employed a new proposed reverse engineering method implemented in the "GeneNet" R package (Opgen-Rhein and Strimmer 2007). This method is a heuristic algorithm applying an extended graphical Gaussian model and dynamical correlation shrinkage estimators to the inference of partial directed causal networks from high-dimensional time series expression data.

1   Introduction

Background:

Plants respond to various environmental stresses such as cold stress by activating the expression of a large number of genes through a series of signal transduction cascades. These signal transduction cascades, often times interact with each other, compose the signaling networks, which eventually govern the transcriptional regulation of expression of genes which are involved in protecting plants from cold damage. Thus understanding the molecular mechanisms of the signal transduction network will greatly help in plant's stress tolerance, which in turn, will have a big economic influence on agricultural business.

Previous research has suggested that plant cold responsive pathways involve a large number of genes which interact with each other. Some of the smaller signaling networks have already provided valuable information about how genes interact with each other. To gain a more complete hierarchical view of the networks in plant cold stress response, we employed a large dataset which contains the microarray data from more than 20 ATH1 chips. This data was obtained from plants which were cold treated for various periods of time, thus we aim to identify genes which might play central regulatory roles in cold response.

The GeneChip® microarray data was downloaded from publicly available database (http://www.arabidopsis.org/info/expression/ATGenExpress.jsp). Plant total RNA were extracted from leaves of 18-day old plants treated with cold (4^0C) for the indicated time periods, according to the standard protocol. cRNA labeling, hybridization and scanning was described in Hannah et al (2006). The Arabidopsis ATH1 whole genome microarray from Affymetrix was used for this set of cold stress experiments.

All the data analysis was performed in R. The raw data file (.cel) files from chip hybridization were read directly into R. Robust Multi-Chip Analysis (RMA) algorithm was used for chip background correction, normalization and to obtain gene expression estimates.

2 Local Pooled Error (LPE) Test

Using the Local-Pooled Error (LPE) test we estimate the significance of each genes differential expression. LPE test is based on pooling errors within genes and between replicate arrays for genes in which expression values are similar (Jain et. al. 2003).
LPE is most useful when dealing with a low number of replicates (i.e. 2-3). This is one reason for using the LPE test rather than the traditional 2-sample t-statistics. Also, within-gene estimates of variance do not provide a reliable hypothesis-testing framework.

We applied the LPE test to genes at each time point (30 min, 1 hr, 3 hr, 6 hr, 12 hr, 24 hr) in relation to the initial time point (0 minutes) and computed the corresponding p-values for each gene. Genes with a significant p-value of $p \leq 0.01$ were chosen to input into the GeneNet algorithm.

3 GeneNet

GeneNet implements a statistical learning algorithm proceeding in two steps: (i) transform correlation network into a partial correlation network, which is an undirected graph displaying the linear associations, and (ii) convert the undirected graph into a partially directed graph by estimating pairwise log-ratios of standardized partial variances.

Consider a linear regression with Y as response and the set of vectors $\{X_1,...,X_k,...,X_K\}$ as covariates. The regression coefficient estimator of Y is defined by

$$\beta_k^{\gamma} = \underbrace{\tilde{\rho}_{\gamma k}}_{A} \underbrace{\sqrt{\frac{SPV_{\gamma}}{SPV_k}}}_{B} \underbrace{\sqrt{\frac{\sigma_{\gamma}^2}{\sigma_k^2}}}_{C} \ ,$$

where $\tilde{\rho}_{\gamma k}$ is the partial correlation between Y and X_k, and $SPV_{\gamma} = \tilde{\sigma}_{\gamma}^2 / \sigma_{\gamma}^2$ is called the standardized partial variance.

A: Establish edges between nodes. An edge is drawn by the multiple testing $A \neq 0$.
B: Determine the directionality of edges of a causal network. If $\log B \neq 0$ evaluated by the multiple testing, edges are directed in a fashion such that the direction of an arrow points from the node with larger standardized partial variance to the node with smaller standardized partial variance.

4   Conclusion

Using the union of up-regulated genes (figure 1), the union of down-regulated genes as well as the union of both up and down-regulated genes together, we have built a gene regulatory network in plant cold stress response, based on the time point microarray data. We have identified a number of master regulatory genes which might play important roles in transducing cold stress signals to other genes. The network built in this work sets up a framework for bench scientists to design experiments to further test the importance of those genes, which might be important in cold stress response.
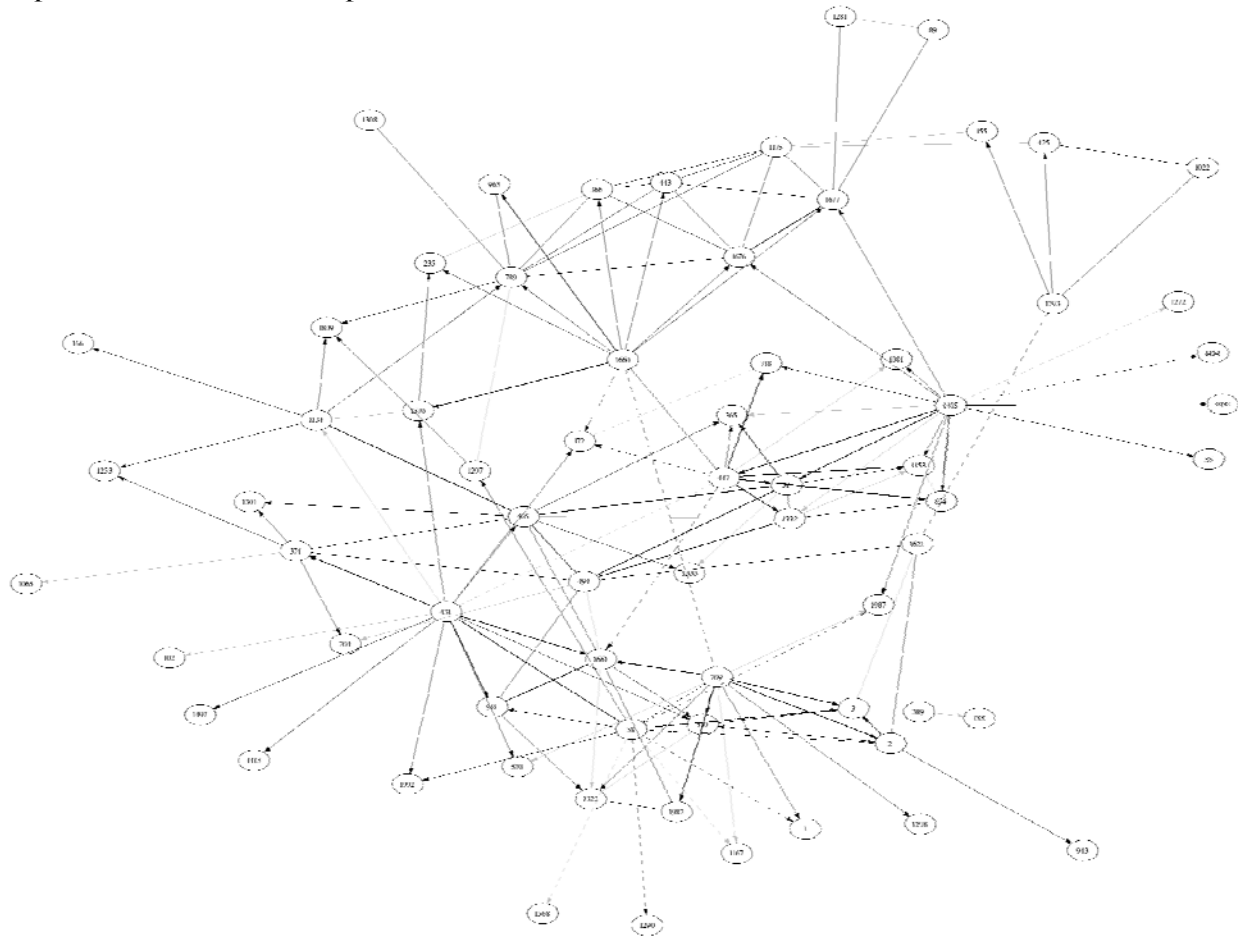
Figure 1: Up Regulated Network

References

[1] Efron, B. (2004), "Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis", *J Amer Statist Assoc*, 99:96-104.

[2] Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, Hincha DK. "Natural Genetic Variation of Freezing Tolerance in Arabidopsis", *Plant Physiol*. 2006 Sep;142(1):98-112. Epub 2006 Jul 14.

[3] Jain, N. et. al. (2003), "Local-Pooled-Error Test for Identifying Differentially Expressed Genes With a Small Number of Replicated Microarrys", *Bioinformatics*, 19, 15:1945-1951.

[4] Opgen-Rhein, R. and Strimmer, K. (2007), "From Correlation to Causation Networks: A Simple Approximate Learning Algorithm and its Application to High-Dimensional Plant Gene Expression Data", *BMC Systems Biology*, 1:37.