Learning gene regulatory networks in *Arabidopsis thaliana*

Chris Needham[1], James Bradford[2], Andrew Bulpitt[1],
Phil Gilmartin[3], Iain Manfield[3] and David Westhead[2]

[1] School of Computing, University of Leeds
[2] Institute of Molecular and Cellular Biology, University of Leeds
[3] Institute of Integrative and Comparative Biology, University of Leeds

Gene regulatory networks govern the functional development and biological processes of cells in all organisms. Genes regulate each other as part of a complex system, of which it is vitally important to gain an understanding. For example, discovery of the complete gene regulatory networks in humans would allow the identification of genes which cause disease, and could be used for drug discovery to identify genes interacting with compounds of interest. Similarly in plants knowledge of the gene regulatory networks would allow the development of stress (drought/salt/temperature) resistant crops.

Learning large gene regulatory networks with thousands of genes with any certainty from microarray data is extremely challenging. This research aims to build around known networks from the literature on gene regulation, and assesses which other genes are likely to play a regulatory role or be in the same regulatory pathways. The gene regulatory networks are modelled with a Bayesian network. The gene expression levels are quantised and a greedy hill climbing search method is used within a network structure learning algorithm.

Large sets of microarray experiments are used in this analysis, specifically 2466 NASC *Arabidopsis thaliana* microarrays containing gene expression levels of over twenty thousand genes in a number of experimental conditions. Initial investigation of this data is very promising. We have learned gene transcription sub-networks (see Figure 1) regulated by the plant's circadian clock. The network shown was generated from microarray data without the use of any prior information, and yet the method managed to identify the strong causal relationships between clock components (TOC1, LHY, ELF3, ELF4, CCA1) and to link these to further key regulators of important processes (e.g. ZAT, myb and GATA transcription factors).

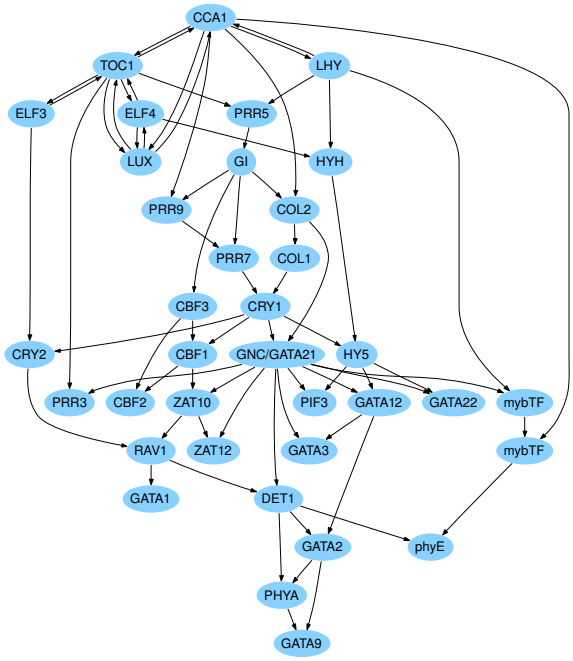

Figure 1: Learned gene transcription network