Biological Significance of Gene Modules in an *Arabidopsis thaliana* Co-expression Network

Darryl Reeves[1]* (darryl.reeves@yale.edu), Gang Fang[2], **Mark Gerstein**[1,2,3]

[1]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; [2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; [3]Department of Computer Science, Yale University, New Haven, CT

Project Goals: Construction of a biologically meaningful co-expression network for *Arabidopsis thaliania* for use in the Systems Biology Knowledgebase (KBase)

Gene expression is subject to environmental and cell growth conditions. Gene co-expression modules which are associations between gene expression and experimental perturbations and/or cell phenotypes provide important clues for understanding gene biological function. Such modules can assist researchers in the design of additional experiments for identifying favorable genes and metabolic pathways for biofuel development. Thousands of gene expression datasets, mainly derived from microarray experiments, presenting a large range of conditions are publicly available now, and will be integrated into KBase. It is necessary to develop an online toolset to allow users to query, display and make *in silico* analysis of expression data. As one of the major packages for analyzing gene expression, WGCNA[1] (weighted correlation network analysis) will be used to build gene co-expression modules and to cluster cell phenotypes in KBase. In this poster, we present a work-flow showing the steps and options to perform this analysis.

Based on the test run of a number of *Arabidopsis thaliana* datasets, we showed how the choice of tree cutting algorithm affects co-expression module generation and, ultimately, functional annotation of the modules. The clustered co-expression network was grouped into modules using three tree-cutting methods provided by the WGCNA package: static, dynamic, and hybrid. We tested each method and optimized their module generation. The criteria for optimization were how enriched modules are for Gene Ontology (GO) terms and KEGG pathways. The goal is to help users determine biological significance of the modules generated by each set of parameters tested.

---

1 Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.