**Machine learning-based network-centric analysis of transcriptomic data**

Chuang Ma and Xiangfeng Wang

School of Plant Sciences, University of Arizona

Systematically integrating and analyzing transcriptome data to unravel novel biological knowledge remains a challenge in plant systems biology with great needs in introducing novel systematic methodologies. I’d like to demonstrate our success in applying machine learning strategy to decipher stress-responsive regulatory networks from gene expression data in plants. Machine learning is a form of intelligent data mining technique widely used in informatics, engineering and computer sciences, with the central idea of making the best use of prior “knowledge” to intelligently analyze and interpret a new set of “data”. In developing this systems biology framework for understanding abiotic stress in plants, knowledgebase was first compiled from TAIR, Gene Ontology and DRASTIC databases, including 895 (salt), 433 (cold), 394 (drought), 357 (wound), 46 (heat) and 42 (genotoxic) known stress-related genes in Arabidopsis. Then, we used the expression data of 22,591 Arabidopsis genes profiled at six time-series points with and without the corresponding stress treatments to identify the characteristics of known stress-related genes in the stress networks and non-stress networks constructed from gene expression data. These network-scale characteristics of stress-related genes, such as degree, connectivity, centrality, edge length, closeness, and eigenvector etc., will be constructed as feature matrix that will be used as training datasets to conduct a series of machine learning-based network-scale analyses on new or existing datasets, including: a) improved network construction of stress networks; b) prediction of novel stress genes from differential network comparison; c) prediction of stress-responsive pathways by network decomposition; d) correlational analyses of stress pathways; e) cross-talk analyses of different stress; f) cross-species prediction of stress-responsive genes in non-model plants.