## Poster B-41

# Context-sensitive Evidence Integration for Predicting Biological Networks


**Authors:**
Chad L. Myers (*Lewis-Sigler Institute for Integrative Genomics, Department of Computer Science*)
Olga G. Troyanskaya (*Lewis-Sigler Institute for Integrative Genomics, Department of Computer Science*)

**Short Abstract:** We developed a general probabilistic system for context-sensitive discovery of biological networks from diverse genomic data. We measure significant variation in relevance of common experimental technologies across biological processes, and exploit this information to improve prediction accuracy. Using our approach, we predicted and experimentally confirmed novel network behavior for S. cerevisiae.

**Long Abstract:**
Understanding biological networks on a whole-genome scale is a key challenge in modern systems biology. For several organisms, broad availability of diverse functional genomic data including physical and genetic interactions, gene expression, localization, sequence, and regulation studies has made this problem amenable to computational approaches. As has been modeled by several recent computational approaches for integration and network prediction, these different experimental technologies are susceptible to different sources of biological and experimental noise. In addition to this variation, however, technologies can vary significantly in relevance and reliability across biological processes. In other words, the same technology can produce precise data for one process while producing noisy or unreliable data for another. This biological context-dependent variation is obvious to biologists who often choose a particular experimental assay to target biological processes of interest, but few computational approaches explicitly take advantage of this variation. We propose that utilization of information about biological context by network prediction approaches is critical to constructing models that are specific and accurate enough to drive laboratory experiments.
Exploiting process-dependent relevance variation requires methods that are able to recognize the biological context of the predictions being made. We have confirmed this variation in relevance for a large collection of published data including gene expression, physical and genetic interactions, upstream, downstream, and coding sequence data, and localization data, and show that accounting for biological context increases accuracy of biological network models. In addition to improving prediction accuracy, this notion of context-sensitivity is a natural paradigm for practical implementations of general data integration and prediction systems. Most biologists have a specific question or domain in mind when they consider using bioinformatics approaches for generating new hypotheses. This focus can be used to drive the prediction methodology and simultaneously define the biological context.
Based on the premise of context-sensitivity, we have developed a general methodology for robust integration of genomic evidence and prediction of biological networks. Our approach is composed of two major components: a Bayesian network that accomplishes

context-sensitive integration of the diverse evidence, and a probabilistic graph search algorithm for network prediction. Both Bayes net learning and inference are performed in real-time based on user input that specifies biological context. Given the result of the context-sensitive Bayesian integration, we developed a probabilistic graph search algorithm that extracts the specific portion of the predicted network relevant to the biological process present in the input query. We illustrate an extensive computational evaluation of our system and present several novel approaches for analysis of diverse sources of genomic evidence as a pre-processing step to the Bayesian data integration mentioned above.

We have used our context-sensitive integration and prediction framework to generate specific, testable hypotheses about previously uncharacterized proteins that play central roles in DNA replication and chromosome segregation, which we have confirmed experimentally. We have implemented this methodology in a public, web-accessible system for Saccharomyces cerevisiae (baker's yeast) and currently include genomic evidence from more than 6500 publications.