Poster J-1

BiologicalNetworks: integrated environment for system level analysis and predictive modeling of molecular machines. Case study of eukaryotic early meiosis.

Authors:
Michael Baitaluk (*San Diego Supercomputer Center, UCSD*)
Mayya Sedova (*San Diego Supercomputer Center, UCSD*)
Biranchi Patra (*Keck Graduate Institute*)
Kirilynn Svay (*Keck Graduate Institute*)
Amarnath Gupta (*San Diego Supercomputer Center, UCSD*)
Animesh Ray (*Keck Graduate Institute*)

Short Abstract: We describe BiologicalNetworks: an information integration system for modeling of gene regulatory networks and show how system level models of genetic regulatory processes involved in development and differentiation can be derived from an integrated gene/protein network generated from: mRNA expression profiles and chromatin-immunoprecipitation experiments, protein-protein interaction, and genetic interaction data.

Long Abstract:
Although crucial aspects of genetic regulatory processes involved in development and differentiation have been uncovered through the study of individual genes and proteins, system level models for most developmental processes are lacking. An important impediment for systems level modeling of cell development is the absence of truly integrative computational modeling platforms. Here we describe BiologicalNetworks: an information integration system for automated reconstruction and modeling of gene regulatory networks and show how system level models can be derived from an integrated gene/protein network generated from three different types of functional relationships: integrated genome-wide time-series data obtained from mRNA expression profiles and chromatin-immunoprecipitation experiments, protein-protein interaction, and genetic interaction data. We apply this to the simple eukaryotic developmental systems—sporulation/meiosis of Saccharomyces cerevisiae.
The first step toward a system-level understanding of eukaryotic early meiosis is to provide 'first-draft' models both of the molecular assemblies involved and of the functional connections between them. These processes involve, among other factors, a transcriptional program of sporulation, patterns of chromatin modification, the transcriptional states of genes as defined by the binding of RNA polymerase II isoforms, and factors that control chromosome stability and dynamics. We have determined a dense (every 30 min) time series transcription profile of yeast during the first six hours of meiosis, and using genome-wide chromatin-immunoprecipitation (ChIP-CHIP and Double ChIP-CHIP) have measured binding profiles of RNA polymerase II, ubiquitylated histone H2B, and DNA double chain break sites during portions of this phase of meiosis.
Using whole genome microarrays, we identified almost 800 genes that show 8 representative expression patterns, which are active in the early-middle period of sporulation. About 600

genes are repressed during this same period.

Global correlations between transcriptome profiling and the interactome network were calculated. The PathSys (http://brak.sdsc.edu/pub/BiologicalNetworks/PathSys/) – database system, which integrates over 14 curated and publicly contributed data sources for 8 representative genomes, including datasets of transcription factor-DNA and protein-protein interactions, as well as low-throughput genetic and molecular data – was used to create high confidence yeast interactome network. Correlations between the interactome or the transcriptome data and the genetic datasets support the notion that these three types of data complement one another in predicting functional relationships, revealing modular organization of gene interaction networks and cross-talk between modules which govern the the process of sporulation and meiosis.

Through measuring genome-wide binding profiles by chromatin-immunoprecipitation we found ~400 'hot spot' genes for Spo11 binding that are the potential DNA double chain break sites. Analysis revealed correlation in regions of Spo11, RNA polymerase II and ubiquitylated histone H2B binding, suggesting that their patterns of association on the genome during portions of this phase in meiosis may be mechanistically or causally related.

Expression analysis of hot spot genes against all genes showing significant changes in mRNA levels during sporulation revealed functional modules representing major biological themes typical for portions of this phase of meiosis: 'double-strand break repair', 'DNA strand elongation', 'protein binding', 'ubiquitin dependent protein degradation', 'DNA repair' etc.

The final stage of our analysis is the predictive modeling of the dynamics of intracellular networks—how genes and gene modules interact and adapt according to cellular requirements.

To understand the dynamics of cellular function we are investigating automatic reconstruction of gene regulatory networks from genome-level information, and are examining their expression dynamics, using services provided by BiologicalNetworks and PathSys. The 'first-draft' networks obtained during the first phase of analysis are converted algorithmically to an equivalent mechanistic model, where each node has multiple states describing biological entities and transitions describing biological processes in the system. Implicit in the network are annotations of properties of nodes (genes or proteins) and edges (binding, positive or negative regulation, etc, and their values, if available). The transformation from the predicted graph to the mechanistic model is performed in a rule-based manner. The most basic rule is to convert every gene or gene product to a 'State' and every edge to a 'Process/Transition'. Using the mechanistic state-transition network as a guide, we construct a set of differential equations, the most widespread formalism to model dynamical systems.

San Diego Supercomputer Center computational capabilities are used to solve systems of differential equations describing network models. BilogicalNetworks is capable of sending tasks to the SDSC computational grid, using the OPAL web services (http://www.nbcr.net/services/). OPAL is a SOAP based web service, that uses strongly typed input and output parameters defined by XML Schemas and the Web Service Definition Language (WSDL).

The differential equation model of the particular network is then fitted to the experimentally observed values of mRNA expression of the network's member genes. Using Monte Carlo techniques, we generate regions in parameter space where the model output is consistent with the observed data. We also employ bifurcation theory to scan the parameter space to look for regions where interesting dynamics such as oscillatory or multistable behavior for some of the outputs may be found. We then investigate any overlap in the parameter spaces

using these two methods. Using the bounds of parameter space thus found we make predictions on its dynamics.
All described analysis are used and described in the context of BiologicalNetworks' analytical functionalities and modeling environment. BiologicalNetworks meets the critical needs and requirements of a systems level analysis of biological pathways: it stores, effectively retrieves, and performs analysis on single genes, gene families, patterns of molecular interactions, as well as on the global structure of the network. This software allows one to construct interaction networks by curation as well as computation (e.g., using algorithms that convert a time-series microarray data set into an influence network for example Pearson correlation networks). It enables the users to retrieve different interaction graphs through on-demand queries and construct new graphs by assembling them in a variety of ways. It also allows the incorporation of novel datasets locally, such as the user's own microarray expression data, and/or overlay these on biological networks to explore novel relationships among genes. Moreover it is capable of generating and simulating dynamical models from molecular interaction graphs.

This project is partly funded by the National Scientific Foundation (NSF) grants ITR EIA-0205061, and FIBR-0527023 to AR and AG, and the National Institutes of Health (NIH) grant (P41 RR08605) awarded to the National Biomedical Computation Resource (NBCR) (AG).
BiologicalNetworks is accessible at http://brak.sdsc.edu/pub/BiologicalNetworks

Contact: baitaluk@sdsc.edu