Poster J-6

Dynamical Modeling of the Gene Regulatory Network of Halobacterium from Diverse Global Systems Biology Datasets

Authors:
David J Reiss (*Institute for Systems Biology*)
Vesteinn Thorsson (*Institute for Systems Biology*)
Paul Shannon (*Institute for Systems Biology*)
Lee Hood (*Institute for Systems Biology*)
Nitin S Baliga (*Institute for Systems Biology*)
Richard Bonneau (*Depts. of Biology and Computer Science, New York University*)

Short Abstract: We have developed a computational strategy for inferring a global model of the regulatory network of the archaeon Halobacterium NRC-1, using genome-wide high-throughput data, de novo. The procedure integrates mRNA and protein levels with functional associations, physical interactions, and regulatory sequence to generate a dynamic, predictive regulatory model that is biologically and statistically parsimonious.

Long Abstract:
We have developed a computational strategy for inferring a model of the regulatory network of the archaeon Halobacterium NRC-1 on a global scale, using genome-wide high-throughput data, de novo. The procedure is composed of three components, which iteratively: (1) groups genes into putatively co-regulated clusters that are co-expressed over subsets of observed experiments, contain the same (predicted) cis-regulatory motifs, and are highly-connected in various functional and physical association networks, (2) generates a dynamic, predictive model of the expression levels of the putative regulons by statistically associating them with regulators and environmental influences (and interactions between them), across both time series and equlibrium experiments, and (3) uses the inferred influences to make dynamical predictions for use in planning future experiments. The resulting regulatory model contains predictions of binding sites for regulators including combinatorial control of genes via regulator interactions.

The first step of the procedure fits a joint statistical model for gene biclusters that integrates a model for co-expression across a subset of observed experiments, with the simultaneous detection of multiple position-specific-scoring-matrix motif models in the genes' upstream sequences and a hypergeometric model for connectivity in various functional and physical association networks (operons, metabolic pathways, phylogenetic profiles, protein-DNA and protein-protein interactions). It uses a simulated annealing—like procedure to iteratively (a) optimize the joint bicluster model as a function of the genes/experiments the bicluster contains and then (b) update the bicluster gene/experiment membership as a function of the model. The second step of the network inference procedure uses linear or logistic regression and variable selection (statistical techniques for the selection of a parsimonious subset from a number of potential predictors) to identify the most likely transcriptional influences on each bicluster, based upon the integration of genome annotation and expression data. The network model can be fit using both steady-state measurements and heterogeneous time

series simultaneously, and incorporates the learning of binary logic interactions between regulators that requires no discretization of data. The resulting network of inferred regulatory influences includes numerical estimates of the relative strength of each regulatory influence, and instances of predicted combinatorial control. The third step in the procedure uses the learned linear or log-linear responses to quantitatively predict the downstream effects of environmental perturbations or regulator knockouts or over-expressions on the gene expression levels. The learned regulatory model is capable of predicting global expression changes under novel perturbations (not part of the original training set) with predictive power similar to that seen over the training set.

We have used the model to predict the regulatory roles for many uncharacterized transcription factors, such as control of phosphate homeostasis by the transcription regulator Trh3. Specific hypotheses learned with this method were supported by further experiments: identification of the Bat-binding motif (UAS) and its coordinated regulation of phototrophy with anaerobic respiration; the activation of a Cu-efflux ATPase by VNG1179C; the regulation of Mn-uptake genes by SirR; and the regulation ribosomal genes by TfbF.