BIOINFORMATICS

Unraveling transcriptional networks

New bioinformatics approaches facilitate the analysis of microarray data to decipher regulatory networks.

Turn on! For a computer, push a button. For a car, turn a key. For a gene? Well, now things get more complicated. Transcription factors are proteins that bind to DNA, and they are important for turning on genes. But unlike a key, which is designed to start one specific car, one transcription factor can often regulate multiple genes. Furthermore, a single gene may be regulated by multiple transcription factors. Clearly, unraveling the transcriptional circuitry of an entire cell can be a daunting problem.

One approach to understand transcriptional networks relies on microarrays, small chips that can be used to measure the expression of thousands of genes. Microarray data from cells grown under different conditions can be used to draw inferences about transcriptional networks. Depending on the condition in which a cell is grown, certain genes in the cell either turn on or off; if a particular transcription factor and a particular gene appear to frequently turn on (or turn off) in response to the same conditions, it is likely that the transcription factor is controlling the gene.

But because microarrays simultaneously measure thousands of genes, it is difficult to spot such connections by eye. Therefore, computer algorithms have been developed to assist in explaining these regulatory connections. Unfortunately, it has been difficult to assess how well these algorithms perform on experimental data, and how well they are able to predict real interactions that occur in a cell. This dilemma caught the eye of Timothy Gardner, a bioengineer at Boston University, who explains, "In the past, algorithms were generally tested, with good results, on simulated regulatory network data or on smaller experimental networks. However, extensive tests on experimental networks and data were not performed because few organisms have a large set of gene expression profiles and a correspondingly large set of known regulatory interactions available for algorithm verification. Thus it was uncertain how such algorithms performed on a more global scale on experimental data sets."

Gardner, along with lead authors Jeremiah Faith and Boris Hayete, and collaborators Jim Collins and Simon Kasif, set out to more precisely benchmark the performance of these algorithms. They turned to the model organism *Escherichia coli*, in part because a large database of experimentally verified interactions between transcription factors and target genes already exists for *E. coli*. By analyzing the data from many microarray experiments in *E. coli* and then comparing the interactions predicted by each algorithm with the known interactions in the database, the scientists were able to evaluate how well the various algorithms performed.

But this was just the first step for Gardner's team. They realized there was room to improve upon the performance of existing algorithms, and they developed a new algorithm. Their new algorithm, context likelihood of relatedness, or CLR, is similar to existing algorithms, in that it identifies transcription factors and target genes with similar expression profiles and scores the similarity. Gardner explains, however, that CLR performs an additional level of analysis: "The algorithm asks the question 'How likely is it that this score represents a true interaction, given the scores for random interactions between this regulator and every possible target?'. Some regulators or targets have great similarity with many noninteracting genes; thus these genes must achieve much higher similarity scores to be considered real interactors."

Gardner and his team found that CLR was substantially better at predicting new regulatory interactions than existing algorithms. CLR is presently being used to make new discoveries in a variety of organisms; Gardner describes one such project: "We have applied CLR to *Arabidopsis thaliana* in collaboration with Dong Hee Lee at the Ewha Women's University in Korea. In this work, we analyzed a data set of expression profiles for 40 transgenic plants under cold stress. The work has identified most previously known cold regulatory factors as well as many new regulators. This work is aimed at finding way to enhance cold tolerance in Korean food crops."

Jesse Potash

RESEARCH PAPERS

Faith J.J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, published online 9 January 2007.