

CIMIT Forum Summary – August 1, 2006

Timothy Gardner, PhD, Boston University, discussed “Shotgun Mapping of Transcription Regulation.” In his lab and in other labs, researchers are attempting to create microbial fuel cells, bacteria that anaerobically donate electrons to an electrode made out of ferric iron. These bacteria can run on carbon waste, but currently they do not produce much power (1.5 W/m<sup>2</sup>). To boost their power output, the bacteria must be modified, and this has traditionally been done by trial and error because the regulatory and metabolic networks of the bacteria are poorly understood. If these networks were better understood, it would be much easier to modify the bacteria in order to increase their power output.

Dr. Gardner’s team is seeking to map the regulatory networks in *E. coli*. Using data from 448 Affymetrix gene expression arrays run by other investigators, they tried a variety of machine-learning algorithms to create potential gene networks. First, they analyzed the data using a pairwise approach in which every pair of genes was tested for correlation, and pairs with high correlations were connected in an undirected network. The pairwise method is not sensitive to non-linear relationships, cannot differentiate between direct and indirect relationships, and struggles with genes that have multiple regulators. Dr. Gardner’s team also tried a multivariate (regression) model and the ARACNE method, which is similar to a pairwise approach except that it is sensitive to non-linear relationships. To determine which method worked best, they compared the results produced by each method to RegulonDB, a database that covers one quarter of the *E. coli* genome and details regulatory interactions. They found that the most sensitive and precise algorithm was context likelihood of regression (CLR), a pairwise approach that uses mutual information instead of correlation and that includes a background correction

step. Using the CLR method, they identified 429 regulatory interactions with 80 % precision. Of these interactions, 161 were true positives in RegulonDB, and as many as 228 may be novel regulatory interactions. They used chromatin immunoprecipitation followed by PCR to test some interactions and found 5 new interactions involving the genes in RegulonDB, and 8 new interactions outside of RegulonDB. They discovered a new pathway involving iron regulation, which affects virulence, biofilm formation, cell death, and stress protection in *E. coli*.

A.J. Marian Walhout, PhD, UMASS Medical School, discussed “Gene-centered Protein-DNA Interaction Networks in the Nematode *C. elegans*.” Researchers looking at gene-protein interactions have traditionally studied one gene or one protein at a time, but Dr. Walhout’s team attempted to create a network model encompassing as many interactions as possible. They chose to study the nematode *C. elegans* because it is a relatively simple eukaryote with precisely 959 somatic cells and because its development and its genome have been studied extensively. It is also transparent, enabling investigators to measure gene expression levels in real time via fluorescent proteins. To map a transcription regulatory network, Dr. Walhout’s team identified 934 putative *C. elegans* transcription factors from an existing database. They used yeast one-hybrid experiments (in which a given transcription factor is mixed with a promoter region attached to a reporter protein) to find the transcription factors and cDNA elements that bound to the promoter regions of the 167 genes expressed in the nematode digestive tract. Dr. Walhout’s team was able to identify 283 protein-DNA interactions. On average, each protein interacted with two promoters, and each promoter interacted with four proteins. The network that Dr.

Walhout's team mapped was highly connected and contained a few proteins that were connected to over a dozen promoters. Out of 117 proteins involved in the network, 107 were putative transcription factors from the list compiled at the beginning of the experiments. Eight out of the ten novel proteins were confirmed to have DNA-binding capability in chromatin immunoprecipitation (ChIP) experiments. These proteins may be novel transcription factors. The network mapped by Dr. Walhout's team was significantly enriched for transcription factors expressed in the digestive tract. Using quantitative RT-PCR, Dr. Walhout's team compared gene expression levels in wild-type nematodes to those in deletion strains missing certain transcription factors. In this way, they were able to validate some of the interactions (20 %) found via the yeast one-hybrid experiments. They used ChIP and knock-out experiments to study certain interactions and determined that a few transcription factors of previously unknown function were repressors.