

From Genome to Candidate Cis-Regulatory Networks: A Bioinformatics Approach

Vadigepalli, Rajanikanth^{*1}, Chakravarthula, Praveen¹, Zak, Daniel², Schwaber, James¹, Gonye, Gregory¹
¹Daniel Baugh Institute for Functional Genomics and Computational Biology, Thomas Jefferson University, Philadelphia, PA, USA; ²Department of Chemical Engineering, University of Delaware, Newark, DE, USA

Present technological enhancements have resulted in public databases containing data sets of various types: gene expression, protein-DNA interaction and transcription factor (TF) activity data, protein-protein interactions, and genomic sequence and ontology information. The analysis of these large volumes of information holds the promise of uncovering the complex dynamic function of the biochemical regulatory networks. Various attempts at reverse engineering the gene regulatory networks from microarray data alone have met with limited success. In contrast, the analysis of promoters for the genes of interest for known transcriptional regulatory elements (TREs) will directly provide a good candidate set of network interactions. To this end, we have developed PAINT: Promoter Analysis and Interaction Network Tool. PAINT is available at www.dbi.tju.edu/dbi/tools/paint.

PAINT currently consists of (1) a database of predicted promoter sequences of known or predicted genes in the Ensembl annotated mouse genome database, and (2) various modules that can retrieve and process the upstream sequences for known TREs. PAINT utilizes TRANSFAC database and the associated TRE identification tool MATCH. PAINT computes statistical significance of the frequency of appearance of TREs in genes of interest paving way for further experimental study on a limited set of highly significant TFs. The candidate interactions generated by PAINT can be incorporated into Bayesian networks or linear dynamic model identification as constraints to render the algorithms tractable for large-scale systems. The PAINT output can be loaded into clustering software such as Cluster and TreeView, as well as into network visualization tools such as Cytoscape and Pajek for further analysis. A network layout diagram is also produced using GraphViz to aid in visual analysis. In addition to command-line and web-based interfaces, PAINT also has an Open Agent Architecture module allowing it to function as an 'agent' for BioSPICE, the simulation and analysis software platform for systems biology being developed under DARPA's BioCOMP program.

Application of PAINT is demonstrated in two separate case studies; the first one involving 100 genes in neuroblastoma cell differentiation, and the second study involving differentially expressed genes in neuroblastoma cells upon activation of the AT1 receptor by angiotensin II. In both examples, the candidate interactions generated by PAINT reduce the number of computed parameters by six to seven fold, thus improving the accuracy and computational tractability of the network identification methods. Preliminary analysis of PAINT output indicates the combinatorial aspects of transcriptional regulation and the need for combining TF activity data with gene expression information for improved accuracy of regulatory network identification.

This work is supported by the DARPA BioCOMP Initiative under the Project "Multi-Timescale Complex Adaptation," contract number F30602-01-2-0578, James Schwaber, PI.