

## **494c A Bioinformatics Approach to the Analysis of Combinatorial Transcriptional Regulation**

*Sarita Nair, Praveen Chakravarthula, and Rajanikanth Vadigepalli*

Developing deeper understanding of the regulation at various levels in biological systems will enable us gain useful insights into the basis of differences between various cell types, cellular response to different environmental stimuli, and contribution of signaling pathways to various cellular processes. Eukaryotic transcriptional regulation involves coordination of multiple transcription factors and hence is combinatorial in nature. Combinatorial regulation addressed in this study is based on the concept of composite elements (CE). Composite regulatory elements contain two closely situated binding sites for two distinct transcription factors. Specific factor-DNA and factor-factor interactions contribute to the function of CEs. Coordinated action of transcription factors binding to the CEs results in highly specific patterns of transcription that cannot be individually produced by the involved factors. A database of known CEs is available (TRANSCOMPEL: Kel-Margoulis et al., 2002).

We have developed an approach to study the combinatorial aspect of gene regulation occurring through composite regulatory elements. The key aspect of our research centers on the statistical enrichment analysis of known composite elements in the upstream regulatory regions of co-expressed genes. The workflow includes automatic retrieval of promoter regions of clusters of co-expressed genes, computational identification of known composite elements, and generation of hypotheses on those elements, factors and genes that are playing a role in the biological process under study. The current implementation enables the identification of composite elements using CATCH, a web-based search tool associated with the TRANSCOMPEL database. The statistical significance of composite elements in clusters of co-expressed genes is conducted through Fisher's Exact Test based on the hypergeometric probability of observed number CEs compared to those in a reference list of genes. The reference typically is the microarray from which the gene expression clusters are obtained. Sensitivity analysis is performed by varying the binding site similarity score, distance between the binding sites in the CE and the number of allowed mismatches to the known binding sites. The individual parameter values range from 0.7-1.0 for the similarity score, 0-5 for the inter-site distance and 0-3 for allowed mismatches. Only those CEs that are significantly enriched for all the parameter values are considered in the subsequent generation of hypotheses.

We demonstrate our approach in two separate case studies. The first study involves a time series of differentially expressed genes from suprachiasmatic nucleus, which contains the master pacemaker controlling mammalian physiology and behaviour (Panda et al., 2002). The second study involves dynamic time profiles of differentially expressed genes from renal proximal tubule epithelial cells exposed to the toxin Staphylococcal enterotoxin B (Ionin et al., 2005). In each case, the hypotheses on CEs and corresponding factors are partially validated by their consistency with available literature and known biological function. The approach is integrated into the PAINT suite for transcriptional regulatory analysis (Vadigepalli et al., 2003).

### REFERENCES

Ionin B, Das R, Pontzer C, Jett M. Staphylococcal enterotoxin B induces cytoskeletal rearrangement and apoptosis in human kidney cells. In review.

Kel-Margoulis, O.V., Kel, A.V. et al. TRANSCompel: a database on composite regulatory elements in eukaryotic genes. *Nucleic Acids Research*. 2002, 30(1), 332-334.

Panda, S., Antoc, M.P. et al. Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock. *Cell*. 2002, 109, 307-320.

Vadigepalli, R., Chakravarthula, P., Zak, D.E., Schwaber, J.S., Gonye GE, PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. *Omics*. 2003, 7(3), 235-52.