

— Keynote Address —

## Revealing the Structure and Dynamics of *Cis*-Regulation Using Heterogeneous, Genome-Wide, Multi-Species Data

Ron Shamir

rshamir@post.tau.ac.il

School of Computer Science, Tel Aviv University, Tel Aviv, 69978 Israel

**Keywords:** gene networks

Gene regulation is based on interactions between transcription factors and their DNA binding sites. We report on three studies on the structure and dynamics of *cis*-regulation.

(1) By studying the rate of changes of motifs in promoters of four yeast genomes, we provide a first global view of the selection forces acting in the evolution of binding sites. Our analysis [2] suggests new models for binding site activity, identifies families of related binding sites, and discovers optimized selective pressures operating on these families. In several cases, it reveals that a single transcription factor has multiple functional modes.

(2) We have developed a method called SAMBA (Statistical-Algorithmic Method for Bicluster Analysis) for finding subsets of genes that manifest a significant co-expression within particular subsets of the conditions [3]. The method is graph-theoretic and based on a statistical model of the data generation. We demonstrate the utility of SAMBA in extracting biological knowledge from large and highly heterogeneous genome-wide yeast datasets. SAMBA dissected the yeast system into *modules*, each comprising a set of genes that share common features over diverse data sources. Using these modules, we can predict the function of over 800 unknown genes, and have validated some predictions experimentally. We also obtain broad perspectives on the interaction of transcription factors and modules, and on the hierarchical organization of modules in yeast. SAMBA's analysis results are available for use interactively [4] and the software is also available as part of the EXPANDER software [1].

(3) We explore the evolution of *cis*-regulatory programs associated with conserved gene modules (detected by SAMBA), by integrating expression profiles for two yeast species with sequence data of 11 other fungal genomes. While some conserved modules have highly conserved regulatory mechanisms, the regulation of other conserved modules is remarkably diverged. We infer the evolutionary history of these changes, showing how the infiltration of new binding sites, the establishment of redundant regulatory mechanisms, and the loss of other sites allow the emergence of different mechanisms that perform similar regulatory roles.

Our studies show that deciphering the evolutionary forces shaping regulatory systems is crucial for understanding their organization and function.

Joint work, in parts, with **Amos Tanay**, **Irit Gat-Viks** (School of Computer Science, Sackler Faculty of Exact Sciences, Tel Aviv University), **Martin Kupiec** (Faculty of Life Sciences, Tel Aviv University), **Roded Sharan** (International Computer Science Institute, Berkeley, California, and **Aviv Regev** (Bauer Center for Genomics Research, Harvard University).

## References

- [1] Sharan, R., Maron-Katz, A., and Shamir, R., Click and expander: A system for clustering and visualizing gene expression data, *Bioinformatics*, 19(14):1787–1799, 2003.
- [2] Tanay, A., Gat-Viks, I., and Shamir, R., A global view of the selection forces in the evolution of yeast *cis*-regulation, *Genome Res.*, 14:829–834, 2004.
- [3] Tanay, A., Sharan, R., Kupiec, M., and Shamir, R., Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genome-wide data, *Proc. Nat. Acad. Sci. USA*, 101(9):2981–2986, 2004.
- [4] <http://www.cs.tau.ac.il/~rshamir/samba/>