ABSTRACT: Algorithms for Inferring *Cis*-Regulatory Structures and Protein Interaction Networks

Richard M. Karp
International Computer Science Institute and University of California at Berkeley

A major focus of functional genomics today is the discovery of the interactions between genes and proteins that regulate the transcription of genes and the responses of cells to external signals. The speaker will describe his recent efforts with several coworkers to solve pieces of this puzzle. The work divides into several parts:

- A new approach to the recognition of transcription-factor binding sites, based on the principle that transcription factors divide naturally into families such as the leucine zippers and the zinc fingers, and that the binding site motifs for transcription factors within the same family have common features. These features may be obscure at the sequence level, but can be characterized at a higher level of description. By discovering and modeling such *meta-sequence features* one can improve the sensitivity and specificity with which binding sites can be determined for transcription factors within a family. [5],[6]
- An algorithm and an associated web-based tool for finding recurrent *cis*-regulatory modules in the promoter regions of human genes. Each such module consists of a set of transcription factors that often bind to the same promoter regions and collectively enhance or inhibit the transcription of the corresponding genes. [4]
- An algorithm for minimizing the number of gene perturbation experiments required to reconstruct signal transduction pathways whose regulatory structures can be described within the mathematical framework of *chain functions*. [1]
- Algorithms for discovering protein complexes and regulatory pathways that are conserved in evolution, using protein sequence data and protein-protein interaction data for two or more organisms. [2], [3]

[1] I. Gat-Viks, R.M. Karp, R. Shamir and R. Sharan, "Reconstructing Chain Functions in Genetic Networks," In *Proceedings of the Ninth Pacific Symposium on Biocomputing (PSB '04)*, 2004.

[2] B. Kelley, R. Sharan, R. Karp, T. Sittler, D. Root, B. Stockwell and T. Ideker, "Conserved Pathways within Bacteria and Yeast as Revealed by Global Protein Network Alignment," *PNAS* 100, pp. 11394-11399, 2003.

[3] R. Sharan, T. Ideker, B. Kelley, R. Shamir, R. Karp, "Identification of Protein Complexes by Comparative Analysis of Yeast and Bacterial Protein Interaction Data," *Proceedings of RECOMB 2004*.

[4] R. Sharan, I. Ovcharenko, A. Ben-Hur and R.M. Karp, "CREME: A Framework for Identifying Cis-Regulatory Modules in Human-Mouse Conserved Segments," *Bioinformatics* 19, Supplement 1, pages 1283-1291, 2003.

[5] E.P. Xing and R.M. Karp, "MotifPrototyper: A Profile Bayesian Model for Motif Families," manuscript, 2004.

[6] E.P. Xing, R.M. Karp, W. Wu and M. Jordan, "A Modular Bayesian Model for *de novo* Motif Detection, *Proc. IEEE Computer Society Bioinformatics Conference* (2003).