

MAR05-2004-006032

Abstract for an Invited Paper
for the MAR05 Meeting of
the American Physical Society

Mining gene-chip data

MORTEN KLOSTER, NEC Laboratories America

DNA microarray (“gene chip”) technology has enabled a rapid accumulation of gene-expression data for model organisms such as *S. cerevisiae* and *C. elegans*, as well as for *H. sapiens*, raising the issue of how best to extract information about the gene regulatory networks of these organisms from this data. While basic clustering algorithms have been successful at finding genes that are coregulated for a small, specific set of experimental conditions, these algorithms are less effective when applied to large, varied data sets. One of the major challenges in analyzing the data is the diversity in both size and signal strength of the various transcriptional modules, *i.e.* sets of coregulated genes along with the sets of conditions for which the genes are strongly coregulated. One method that has proven successful at identifying large and/or strong modules is the Iterative Signature Algorithm (ISA) [1]. A modified version of the ISA algorithm, the Progressive Iterative Signature Algorithm (PISA), is also able to identify smaller, weaker modules by sequentially eliminating transcriptional modules as they are identified. Applying these algorithms to a large set of yeast gene expression data illustrates the strengths and weaknesses of each approach. [1] Bergmann, S., Ihmels, J., and Barkai, N., Phys. Rev. E **67**, 031902 (2002).