

MAR05-2004-000680

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

### **Combining gene-chip data and bioinformatics to define transcription networks**

HARMEN J. BUSSEMAKER, Columbia University

The development of DNA microarray technology has made it possible to simultaneously monitor the mRNA abundance of all genes (“transcriptome”) for a variety of cellular conditions. In addition, microarrays have been used to map protein-DNA interactions by measuring occupancy profiles along the chromosome for an increasing number of transcription factors (TFs), especially in the yeast *S. cerevisiae*. With this data and the complete genome sequence on hand, it is becoming possible to quantitatively model the molecular computation performed near the transcription start site of the gene. This computation has as input the nuclear concentrations of the active form of various regulatory proteins (“regulome”) and as output a transcription rate, which together with the half-life of the transcript determines the mRNA abundance. Our laboratory has pioneered the use of multivariate regression methods to link mRNA expression data with genome sequence data and TF occupancy data. This allows us to: (i) discover cis-regulatory elements in non-coding regulatory regions; (ii) infer the condition-dependent regulatory activities of transcription factors as “hidden variables”; and (iii) accurately determine which genes are controlled by which transcription factors. Together, our results show that model-based analysis of functional genomics data provides a versatile and extensible conceptual and practical framework for the elucidation of regulatory circuitry, and a powerful alternative to the currently popular methods based on clustering and “modules”.