DNA-Protein Binding and Gene Expression Patterns

Hongyu Zhao, Baolin Wu and Ning Sun

Abstract

Although many clustering methods have been applied to analyze gene expression data, genes in the same cluster may have neither common functions nor common regulation. As a result, computational approaches have been developed to identify motifs in the regulatory regions of a cluster of genes or of genes with similar gene expression levels that are responsible for DNA-protein binding and similar gene expression levels. However, these motifs are neither sufficient nor necessary for a transcription factor to bind to the promoter region of a gene with these motif patterns. More recently, molecular methods have been developed to directly measure DNA-protein binding at the genomic level. In this article, we first evaluate the predictive power of computational approaches to predict DNA-protein binding from a study involving nine transcription factors in the cell cycle. We then compare how much variation in gene expression levels can be explained either by the observed DNA-protein binding or by the binding predicted through computational approaches. We find that current computational approaches may be limited both in predicting DNA-protein binding as well as in predicting gene expression levels. We also observe indirectly that the correspondence between gene expression levels and protein levels may be rather poor, which suggests that there may be difficulty in modeling genetic networks purely through gene expression data. To better understand gene expression patterns, an integrated approach to incorporating different kinds of information should be developed.

Keywords: gene expression; DNA-protein binding; motif; microarray

1 Introduction

With the completion of the Human Genome Project, large-scale gene expression experiments have become common practice in the scientific community. Such experiments normally have different objectives: (1) to identify differentially expressed genes, (2) to identify genes expressed in a coordinated manner across a set of conditions, (3) to identify gene expression patterns that distinguish different samples (*e.g.* normal versus tumor tissues), and (4) to define global biological pathways. Genomics research is different from traditional molecular biology in that traditional approaches focus on the study of individual genes considered in isolation, whereas functional genomics allows researchers to determine the principles underlying complex biological processes (*e.g.*