Structural Analysis of Microarray Data Using Singular Value Decomposition

Scott A. Rifkin, Kevin Atteson and Junhyong Kim
Department of Ecology and Evolutionary Biology, Yale University
<scott.rifkin, junhyong.kim>@yale.edu, atteson@peaplant.biology.yale.edu

Previous analyses of genomic expression data have used various clustering methods to group genes with similar expression patterns. While these approaches can identify groups of genes with similar regulation and perhaps function, they do not take advantage of the power of microarray and related technologies to explore genomes as systems of interacting components. We develop a structural approach to analyzing genomic expression data which enables us to investigate both the contributions of particular genes to organismal responses to experimental manipulations and the organismal responses themselves explicitly in the context of the expression patterns of other genes.

A microarray experiment gives a snapshot of the state of an organism in terms of the relative abundances of its mRNA transcripts. Consequently, each experiment locates the organism at a point in a high dimensional state space where each axis represents the relative expression level of a single gene. Multiple experiments generate a cloud of points in this gene expression space. We investigate the variational properties of such a cloud using a *Saccharomyces cerevisiae* microarray dataset generated by the Brown and Botstein labs at Stanford University and summarized in Eisen *et. al.* (1998). In particular, we perform a dimension reduction analysis using singular value decomposition to find the linear least-squares approximation to the data and thereby extract the magnitude and directions of the major variance.

We find that (1) there exist major directions of variation in the data, (2) these directions differ between groups of experiments, and (3) the singular vectors project differentially onto the basis axes of the gene expression space. These findings suggest that the yeast cell state is restricted to small subspaces in gene expression space and that the responses to different experimental manipulations occupy different regions of the space. We use the magnitude of the projections onto the expression axes of particular genes both to identify genes involved in responses to specific manipulations and those involved in maintenance and cell cycle activities and to illustrate pleiotropy and epistasis in the yeast genome.

Eisen, MB, PT Spellman, PO Brown, and D Botstein. (1998). *Proc. Natl. Acad. Sci. USA* **95**, 14863-14868.