Matrix methods for gene expression analysis

Rachel Brem

Abstract

Matrix methods have been used to find dominant temporal patterns in gene expression data.

Significance and context

A major new branch of genomics involves expression arrays, in which the levels of RNA produced by many genes are measured simultaneously. Technical advances have made array experiments fairly easy to do, but tools for analysis of data produced have lagged behind. Here, Holter *et al*. pioneer the use of matrix methods on time-course expression arrays. From this analysis, the authors find that most genes in these systems undergo only one or a few simple patterns of expression over time. The technique is important because it may be more rigorous than standard clustering approaches; the paradigm of Holter *et al*. may find expression patterns that would not be detected using other methods.

Key results

Holter *et al*. apply a standard technique called singular value decomposition (SVD) to array data: two sets from time-course expression arrays from yeast and one from human fibroblasts. SVD finds eigenvectors, or fundamental patterns of expression with time, of the array matrices (see Methodological innovations for details of the methods used). Holter *et al*. find that in the yeast and fibroblast data only a few eigenvectors dominate, but the same is not true in a control calculation on random data. The dominant eigenvectors usually involve just one or two increases and decreases in expression levels. This suggests that at a gross level, most time-dependent expression patterns are very simple. The authors also find that many related genes - for example, those activated at a certain stage in yeast sporulation - can be found by multiplying the dominant eigenvectors by the same constant coefficients. This implies that related proteins are expressed with the same time course. Thus, data from SVD agree with previous knowledge of expression patterns.

Methodological innovations

One can understand the method of Holter *et al*. as follows. First, picture many graphs of a variable *y* versus time, each following a different curve. Here, *y* stands for the expression of one gene over time. Stacked vertically, these graphs represent the matrix of array data. Now picture another set of graphs of *y* versus time. These have the special property that, when added together, multiplied by different coefficients, the sum reproduces each experimental graph. The latter special graphs are the eigenvectors of the array matrix, and Holter *et al*. find them using a standard technique called SVD. At maximum, there are as many eigenvectors as there are genes in the array. But sometimes a few eigenvectors dominate the system: sums of these do a reasonable job of reporting back the experimental data.

Conclusions

The authors conclude that most genes obey one of only a few patterns of expression over time, and that their method performs well on positive controls with known protein classes.

Reporter's comments

This paper provides an excellent new idea for finding global trends in expression data. The next step is to use it in a clinical setting: do cancer cells express genes on a different time course from healthy cells, for example, and can the cause be identified?

Table of links

Proceedings of the National Academy of Sciences of the United States of America

References

1. Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR, Fedoroff NV: Fundamental patterns underlying gene expression profiles: simplicity from complexity. Proc Natl Acad Sci U S A. 2000, 97: 8409-8414. 0027-8424

This PDF file was created after publication.