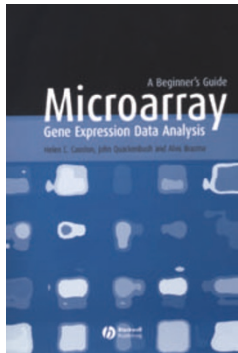


Microarray data analysis for beginners



Microarray Gene Expression Data Analysis: A Beginner's Guide

by Helen C Causton, John Quackenbush and Alvis Brazma

Blackwell Publishing, 2003.
£34.99/\$54.95

Terry Speed

Microarrays first appeared on the scene around 1995, and it was not long before their use became quite widespread. Early analyses were modelled on those of the pioneers. By around 2000, however, methods papers on microarray analyses started appearing in rapidly increasing numbers. Now the books are here. Quite a few recently published books discuss analysis of microarray gene expression data for beginners. *Microarray Gene Expression Data Analysis: A Beginner's Guide* by Helen Causton, John Quackenbush and Alvis Brazma is arguably the best of its kind in this regard.

So who is the book aimed at? The answer suggested by the back cover is “graduates and researchers in bioinformatics and the life sciences, and statisticians interested in the approaches currently used to study gene expression”. My guess is that although most bioinformaticians and statisticians are likely to want a more in-depth treatment of the subject, and will become impatient with the many explanations of the familiar, the book fulfils its goal of reaching its intended audience. The book is broad in scope, but remains reasonably sized; it is clear, precise and easy to read, and is thus well suited to the life science researcher who is only just beginning to analyse microarray data.

The authors come from complementary environments (a clinical sciences research centre, a genomics research institute and a bioinformatics institute), and jointly they draw upon a considerable amount of relevant experience. The reader can feel fairly confident that the authors have had first-hand experience of most, if not all, topics covered in the book.

The introductory chapter begins by explaining what microarrays are and why one would want use them, and ends with an overview of the entire book. The next two chapters deal, in turn, with experimental design and image processing, and normalization and data

*Terry Speed is in the Division of Genetics and Bioinformatics
The Walter & Eliza Hall Institute of Medical Research, 1G Royal Parade,
Parkville, Vic 3052, Australia
e-mail: terry@wehi.edu.au*

transformation. The fourth chapter describes the analysis of gene expression data matrices.

Much of this last chapter is devoted to the explanation of standard statistical techniques, but a few other topics, such as the gene ontology (GO), the KEGG metabolic pathway database and the identification of regulatory signals, are mentioned briefly.

Does the book cover all the topics that one might hope for? Almost, but not quite. One glaring omission is a discussion of the simple comparative experiment, where replicate arrays are used to compare gene expression in two mRNA samples from, say, the livers of knockout and wild-type mice. After assaying gene expression in these RNAs, either on two-colour microarrays or Affymetrix GeneChips, how should one analyse the resulting data to identify differentially expressed genes? Surprisingly, although this issue is mentioned at the start of the book, it is not discussed further. The appendix, however, does list some packages in which solutions can be found. In addition, it is a shame that so much space was given to explaining statistical ideas at the expense of important ideas central to the analysis of microarray data, such as creative use of the GO or KEGG databases and the use of tools with which to utilize them (for example, GenMAPP) to interpret gene expression patterns identified from microarray analysis. Furthermore, it would have been useful to reap the benefits of the authors' experience in solving the gene annotation problems that one encounters when attempting to follow up clones using the latest version of the relevant genome. One solution here might have been to refer to reliable statistics texts that cover topics of importance to microarray analysis, rather than present an extended discussion of statistical ideas.

The book is broad in scope, but remains reasonably sized; it is clear, precise and easy to read...

Minor irritations include four separate reference lists scattered throughout the book, two separate explanations of GeneChip PM and MM intensity measurements, the Minimal Information About a Microarray Experiment (MIAME) and loop designs in almost identical terms, and no mention of a very valuable website (www.bioconductor.org) providing free statistical tools for the analysis of microarray data. Perhaps most surprising of all, given that all three authors are microarray practitioners, rather than academics divorced from the action, there were no real case studies to motivate the reader and illustrate the material in the book — an unfortunate omission considering their potential instructive power.

In closing, let me endorse the authors' statement that “Expression data analysis methods are currently only in their infancy.” Although it is easy to carp, as I have done, it is no easy task to write a book for beginners under such conditions. Although not perfect, this is the best book for beginners currently available. The field will undoubtedly mature in coming years, and so will the introductory books. □