***Drug Development and Industrial Pharmacy*, 32:137, 2006**
Copyright © Taylor & Francis LLC
ISSN: 0363-9045 print / 1520-5762 online
DOI: 10.1080/03639040500390827

Book Review

Analyzing Microarray Gene Expression Data; G. L. McLachlan, K-A. Do, and C. Ambroise, Wiley-Interscience, Hoboken, New Jersey, 2004. Hardback (352 pages) ISBN: 0-471-22616-5

The use of microarrays to investigate gene expression in biological systems is revolutionizing biomedical research. These techniques allow for the screening of a large number of genes in a single experiment and hold great promise for developing novel therapeutic agents to combat genetic-based diseases. However, the quantity of data generated and the potential for misinterpretation are major obstacles that must be appropriately managed. The text, *Analyzing Microarray Gene Expression Data*, reviews methodologies for analysis of these data from a biostatistician's viewpoint. The authors begin with a basic overview of gene expression studies. Next, cleaning and normalization are discussed. The authors then cover cluster analysis methods, clustering of tissue samples, and clustering of genes. Other subjects discussed include discriminant analysis, supervised classification of tissue samples, and the linking of microarray data with survival analysis. In addition, an index provides a list of names of analytical approaches.

The text goes into a great deal of mathematical detail and appears to be written by mathematicians with little experience in microarray analysis. Several problems normally encountered in microarray analysis are not adequately addressed. Little attention is given to potentially devastating common problems in experimental design, microarray hybridization and scanning, and the different possibilities for clustering. The authors seem unaware of problems in slide-based arrays, such as spot-localized contaminating fluorescence, that can make global normalization impossible and actually can ruin good data. Quality control is mentioned without discussing specific ways to control processes. A combination of design, control, and randomization in the laboratory as well as fast approaches to determine the "goodness" of a set of slides or bootstrapping methods that can give confidence for each spot so the data can be trusted are not adequately covered. While the authors make sense in the statistical contents, a discussion of non-parametric approaches, which prove extremely useful, especially in clustering by experiment as is done with compendium experiments, is almost nonexistent. The text focuses on various data mining methods rather than the best approaches to solving microarray-specific problems so that the researchers know the data-mining is being done on good data.

While there is a need for a new text that discusses the analysis of microarray data, this text lacks a complete view of the issues and is unsuitable for biologists interested in microarray analysis because they need more discussion on the value of each approach or how to prioritize them; and, mathematicians need to realize that they have an investment in ensuring they get good data to analyze. Although the level of presentation is quite mathematical, it is incomplete in its scope. Someone trying to use this book, even to determine the best approaches for analysis, would have serious problems trying to get a clear or complete picture of the key issues.

Margaret Werner-Washburne, Ph.D.
Professor of Molecular Biology and Genetics
University of New Mexico