Microarray analysis made simple

A Biologist's Guide to Analysis of DNA Microarray Data
By Steen Knudsen
John Wiley & Sons, $44.95 (cloth), ISBN: 0-471-22490-1, 2002

Reviewed by Douglas J. Gaasterland & Terry Gaasterland
Laboratory of Computational Genomics, The Rockefeller University, 1230 York Avenue, New York, New York, USA.

....

The field of DNA microarray data analysis is getting increasingly complicated as scientists develop new array technologies and analysis algorithms, yet accurate introductory information on the subject is hard to come by. Steen Knudsen's eminently readable (and skimable) *A Biologist's Guide to Analysis of DNA Microarray Data* aims to bridge this gap. The book offers an overview for anyone who wants to learn how to interpret the data from DNA microarray experiments without first earning a degree in statistics. It targets with great success the biologist who is confronted for the first time with large gene-expression data sets and the computer scientist who wants to know more about what gene expression is, what it aims to measure and its limitations. But the book's concise organization (it is just over 100 pages in length) and many references for further reading will also appeal to those who have already dealt with microarray data.
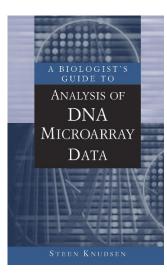
A DNA microarray is a small manufactured chip spotted with up to tens of thousands of different oligonucleotide probes, intended to test how well a test sample of RNA binds to any of the probes. The amount of raw data returned in a single microarray hybridization experiment is enormous; unfortunately, it cannot be directly compared to data from other hybridization experiments because there are many experimental parameters that can vary from lab to lab and technician to technician. Thus, proper analysis demands a lengthy procedural flowchart in which the data is scaled, possibly replicated, tested statistically to throw away outlying data points, clustered with multiple experiments and, finally, rendered usable to determine the significance of any expression hits.

The introductory chapters of Knudsen's book cover the basics of how DNA micro-arrays are made and used, including a survey of current technologies (Affymetrix, spotted glass slide arrays and serial analysis of gene expression (SAGE)), before launching into a primer on data analysis. In his introduction to data analysis, Knudsen keeps to simple formulas to describe the processes involved and refrains from delving into the more complicated, if more mathematically accurate, analysis tools. Each concept is discussed in a straightforward way that any scientist should be able to understand. Each chapter ends with a standard textbook summary of the main concepts and a thorough reading section listing primarily journal articles that the reader can look up for more detail.

The second half of the book is broader in scope. It includes several short chapters that discuss the conclusions that can be drawn from a series of microarray experiments and the medical and biological models that can be elucidated by multiple hybridization experiments. Topics covered in these chapters include how to design an informative experiment, how to pick which genes to spot on a custom array, what shortcomings there may be in microarray analysis and several real-world experiments that apply microarrays to cancer classification or use neural networks to classify genes on the basis of results from microarray experiments.

The final chapters discuss different types of software packages available for microarray analysis, from software provided by commercial vendors to that developed and released for free by researchers, and from software that is flexible but terribly manual to that which is automated but rigid. Again, there is little cohesive information in the real world to help understand the software that is available and how to use it, and so this survey is welcome. The text also includes some useful scripts for basic data transformation written in awk, a scripting language, and R, an open-source statistical package, both of which are easy to find and easy to learn.

Knudsen has put together a fine introduction to DNA microarray analysis using good, concrete examples, well reasoned descriptions of mathematical and statistical methods for data manipulation and comparison and much-needed examination of the tools, technologies and research resources available to microarray users. After reading this book, anyone new to microarray data should understand what can and cannot be measured using microarray gene expression techniques. The reader will be equipped with the knowledge to analyze the very large data sets generated by these types of experiments and be ready to use available bioinformatics tools and to consult efficiently with statisticians about follow-up analysis. To learn more about the details of statistical analysis, we recommend the book *Statistical Analysis of Gene Expression Microarray Data*, edited by Terry P. Speed. For a more detailed presentation of methods for microarray image processing, *DNA Array Image Analysis: Nuts & Bolts* by Gerda Kamberova and Shishir Shah is an excellent book.