

## Book Reviews

---

*Am. J. Hum. Genet.* 71:1483–1484, 2002

*A Biologist's Guide to Analysis of DNA Microarray Data.*  
By Steen Knudsen. New York: John Wiley & Sons, 2002.  
Pp. 125. \$44.95.

With the dramatic increase in the use of microarrays in research over the past five years has come a concomitant increase in the number of biologists who wish they had paid more attention in their undergraduate statistics courses. The widespread availability of microarrays and the astonishing range of questions to which they can be applied mean that more and more researchers are turning to this methodology. The resulting vast bulk of numerical data that have been generated is truly astonishing. In many laboratories, this glut of numbers now represents a bottleneck of data that need to be subjected to rigorous analysis and data mining...if only one knew how.

This is why a book that calls itself a “biologist’s guide to analysis of microarray data” and professes to take over where most image analysis software takes its leave will appear as a light in the darkness for many. If it were to deliver on these promises, this slim volume would, indeed, be a guide that no scientist who is even contemplating microarray experimentation should be without.

The basic position of the book is that “mathematical stringency is sacrificed for intuitive and visual introduction of concepts” (p. *xi*). Although this is desirable from the point of view of anyone to whom long mathematical equations are a foreign language, there is a fine balance between making complex mathematical arguments understandable and simplifying them to the point where they lose meaning. For the most part, this book manages to fulfill its objectives admirably, although there are sections where a more in-depth treatment of the subject matter might have helped.

The book is divided into a series of very short chapters, each devoted to a particular facet of analysis and arranged in roughly the same order as the issues one might encounter during a real experiment. It begins with a very brief overview of hybridization, which nicely summarizes the microarray technology and highlights the current limitations of the most commonly used methods. Affymetrix and cDNA-based chips are treated separately, here and throughout the book. Although this is useful for users of each technology, the authors seem to place a somewhat disproportionate emphasis on the analysis of data derived from the Affymetrix system. However, for the

most part, readers can apply the concepts to data generated from cDNA arrays with little difficulty.

The book continues with the shortest, and perhaps most important, chapter of all. Chapter 2 is simply a flow diagram, outlining the stages at which users might apply different analysis tools. The visual presentation is very simple and is highly effective. It provides a framework for the future chapters and allows the reader to grasp the range of possibilities for analysis of data. It also brings home the critical idea that there is no cookbook approach for analysis of microarray data and that experimental design and variables play an integral role in decisions about analysis tools.

Chapter 3 provides some of the basics, such as the ever-popular topic of scaling and how to assess the significance of any fold changes in expression. The first half of this chapter is heavily devoted to the use of Affymetrix and could have benefited from a more even treatment of both types of chips. However, the important points made about scaling and measuring within the linear range of a scanner are equally applicable to the gathering of cDNA array data, and readers should have no trouble extrapolating the examples to their own data. There would be some benefit from a more in-depth treatment of nonlinear scaling options, although there are a fairly large number of references to which readers can turn for additional help. The section in this chapter on fold change and how to assess its significance is particularly helpful, because it is illustrated with a very simple example, which shows the reader the impact of using different statistical tests to assess data. The reader can follow this same example through the book, as it appears in different chapters to illustrate the use of other statistical tests. This continuity is very helpful.

Chapter 4 is a brief treatment of principal components analysis, whereas chapter 5 expands to include  $\kappa$ -means clustering, hierarchical clustering, and self-organizing maps. Each of these treatments is quite short, and not many details are provided, although enough information is given to allow the reader to understand the major differences between the approaches. Again, the chapters are well referenced for those who need to know more.

The next three chapters move beyond the realm of organizing data into groups of genes and move toward trying to make biological sense of the whole thing. Chapter 6 is devoted to strategies for the dissection of information about promoter regions as a way to infer the function of groups of genes. Chapter 7 covers some very interesting examples of the use of expression level changes to help to define regulatory networks

of genes, and chapter 8 deals with the construction of molecular classifiers for the identification of subtypes of samples, such as tumors. Each of these chapters contains interesting ideas and applications, as well as small examples that use real data to illustrate major concepts. Chapter 8, in particular, is quite short and contains only passing references to some of the current hot topics, such as neural nets or vector support machines. It does, however, suggest how these types of classification schemes can be used and what should already have happened to the data prior to their use, which will be useful information for most researchers.

The remaining chapters cover a variety of topics. Chapter 9 deals with some of the issues encountered when selecting genes for printing on a chip. Chapter 10 is an extremely short passage outlining some caveats about the limitations of data derived from microarray experiments. Chapter 11 is a passing mention of genotyping chips but also has another very brief overview of neural networks and how they can be used in data analysis. Chapters 12 and 13 deal with available software for tackling various issues raised in the book, with chapter 12 containing scripts for performing various types of analyses for those enterprising researchers who would prefer to work in a Unix or Linux environment. There are also some instructions for programming in R for many of the operations dealt with earlier in the book, again using the familiar example that has been used throughout. For those of us with little time or inclination to learn to program in Awl, Perl, or R, chapter 13 is a listing of some commercial software packages that are available to help with microarray analyses. Readers should be aware, however, that this list is necessarily incomplete, and there is no discussion of the relative merits of any of the programs.

Overall, the major shortcoming of the book could be viewed as the extremely brief treatment of most of the topics, which are sometimes given no more than a passing mention. However, this might equally be regarded as a strength, since the reader is never bogged down in a heavy discussion of mathematical arguments. In addition, the brevity is countered by extensive reading lists, which accompany every chapter, and by references to a variety of Web sites and other resources to which readers may go to find more-detailed help. True to its word, the book remains accessible throughout to the biologist who may have no formal training in the statistical arts. What it does manage to do quite nicely is to introduce the fascinating world of microarray analysis beyond scanning and to provide a framework for the kinds of questions that can be asked and the tools available to help answer them. In that respect, this book has fulfilled its promise as a biologist's guide.

HARRIET FEILOTTER

*Department of Pathology  
Queen's University  
Kingston  
Canada*

*Am. J. Hum. Genet.* 71:1484–1485, 2002

*Mapping Human History: Discovering the Past through Our Genes.* By Steve Olson. Boston: Houghton Mifflin, 2002. Pp. 278. \$25.00.

As human genetics becomes increasingly a part of everyday experience, there is a need for literature that accurately explains our science to the general public. Steve Olson, an experienced science writer, has done an excellent job of balancing scientific content with journalistic spice in this new book about the genetic history of our species.

Olson writes clearly and gracefully, and his enthusiasm for the subject is readily apparent. He defines a number of key concepts, including speciation, mutation, natural selection, genetic drift, and even the concept of coalescence. The definitions are usually quite approachable, as in the following description of recombination: “the chromosome pairs delicately intertwine and exchange pieces in a process known as recombination. The result is two hybrid chromosomes, as if a husband and wife had exchanged arms and legs” (p. 15).

As the book progresses, we are presented with genetic evidence bearing on major events in our history. Considerable attention is devoted to the competing “out of Africa” and “multiregional” hypotheses. Olson agrees with most geneticists in finding the former more convincing than the latter, although I think he concludes too readily that archaic humans made no contribution at all to our genetic heritage. A chapter is devoted to the origins of agriculture, and there are chapters on the genetic histories of Africa, Asia, Europe, the Americas, and the Pacific Islands. Although the focus is on genetic studies, some attention is also devoted to inferences made from archaeological and linguistic evidence. Each chapter is laced with vivid descriptions of places visited by the author in the course of writing the book, including the grasslands of Botswana, the city of Jericho, Skhul Cave, the Stonehenge monument, and—not surprisingly—the beaches of Hawaii. During these travels, the author interviewed many of our colleagues, and we read excerpts of interviews with Becky Cann, Luca Cavalli-Sforza, Li Jin, Himla Soodyall, and Doug Wallace, among others. These personal touches enliven the book and will doubtless make it more inviting to lay readers.

A book about human genetic variation must inevitably address the delicate issue of “race.” In fact, considerable discussion is devoted to this subject, and I found it to be, for the most part, accurate and sensitive. Olson does not make the mistake of simply declaring that we are “all the same.” He rightly emphasizes the low degree of nucleotide diversity in humans, relative to many other species, but he also recognizes that “real physical differences exist between the average Nigerian, the average Norwegian, and the average Filipino. Most of the members of these groups share a common biological history, which is reflected in their DNA” (p. 49). He achieves an appropriate balance by highlighting the massive exchanges of genetic material that have occurred during our species' history, such that “in terms of our DNA, all humans overlap” (p. 50). Because of this overlap, we observe a geographic continuum of genetic variation, and individual humans (and even populations) do not always fit neatly into the categories some-