Teaching genomics, proteomics and bioinformatics

Discovering genomics, proteomics and bioinformatics

By A. Malcolm Campbell and Laurie J. Heyer

Cold Spring Harbor Press • 2003

Paperback, £35.99/\$70.60

Sam Hanash

I was enthusiastic about reviewing a textbook with the title, *Discovering genomics, proteomics and bioinformatics*, out of curiosity to find out how the authors, one a biologist and the other a mathematician, would have masterminded covering all three fields. Covering just genomics and bioinformatics would seem reasonably straightforward. However adding proteomics could be particularly challenging. After all, rare are those individuals who have expertise, or even affinities, for both genomics and proteomics. At times, it can seem like the Great Wall of China has separated genome- and proteome-specialists. This can present problems when trying to write a book that covers both areas. However, this new book does represent an admirable and valiant effort at covering, in textbook fashion, all three topics.

Now that the major milestone of sequencing the human genome is behind us, sufficient wealth of knowledge is at hand to begin the synthesis of textbooks that cover both the field of genomics and the bioinformatics tools available to mine the genome. Given that the proteome is encoded in the genome, and given the emergence of proteomics as a 'hot field', it is understandable from a genomicists' point of view why the definition of genomics would be expanded to encompass proteomics. However, the tools used by the proteomicist are quite different from those used by the genomicist, hence the inherent difficulties in wanting to cover these disciplines in a seamless fashion.

Numerous features of the book are particularly attractive. The book is designed as an interactive resource, with real data that can be analysed alongside online databases to stimulate and teach the student. Useful features include "Discovery Questions" in each chapter to guide the student in problem solving. "Math minutes" is another feature that adequately addresses mathematical and statistical issues related to data analysis. A companion website also serves as a tool to enhance understanding of

In the 'hot' fields of genomics, proteomics and bioinformatics, there is an urgent need for training and education, and a textbook such as this is a good start.

genomics. Additionally, the "Methods" web pages clearly explain experimental procedures and show how they are applied to problem solving. Sequences are also provided for analysis using bioinformatics tools with links to relevant websites. A CD-ROM included within the textbook contains numerous colour figures that may be useful to display in a class setting.

The authors' writing style is easy to read and the book is built on stories and case studies. However, in some cases the authors have gone overboard. Do students really need to know that "Norm Dovichi has recently moved from the University of Alberta to the University of Washington"? Or what is the point of showing a picture of "tuberculosis-like lesions in the lung" that, even as a physician, left me baffled as to what is being pointed out? There are a very small number of glaring errors and inconsistencies; for example pRetinoblastoma is referred to as an oncogene in one place and as a tumor suppressor in another!

Clearly, proteomics is covered to a much lesser extent than the other two topics and much less than the title of the book would suggest. Even if this book is intended for undergraduates and it was somewhat deliberate to limit the proteomics content to one chapter, it is baffling why a currently esoteric topic such as single-cell proteomics is covered, whereas a major feature of proteins, such as post-translational modification, is

not adequately covered, if at all. Post-transcriptional regulation, in the form of translation initiation or targeted protein degradation, is also inadequately covered. These are important topics in the context of bioinformatics, from predicting sites of post-translational modifications to understanding post-transcriptional regulation. Given the limited coverage of proteomics, perhaps "proteomics" should not be in the title, as this could be misleading, even if the book contained a chapter devoted to proteomics.

So what about the coverage of genomics and bioinformatics? The book certainly does a much better job in teaching the basics of these subjects, although in some chapters one wonders if the topics covered are more in the field of genetics than genomics for some, and more in molecular biology for others. The book does teach the basics of how to capture, analyse and interpret genome data, formulate hypotheses and develop models. Much coverage is devoted to emerging technologies and fields such as DNA microarrays with a certain lack of critical evaluation of their merits and limitations. There is little discussion of the difficulties in mining microarray data and in assessing the functional relevance of RNA findings emerging from DNA microarray analysis.

Given the paucity of textbooks and the fact that the fields covered are still evolving, much of the book's shortcomings can be forgiven. Undoubtedly a second edition would have much more polish and greater information content. A case in point is that since publication of the text, the mouse genome has been sequenced and a lot of lessons will be learned (and taught) from careful comparisons of the human and mouse genomes. In the 'hot' fields of genomics, proteomics and bioinformatics, there is an urgent need for training and education, and a textbook such as this is a good start. □

Sam Hanash is in the Department of Pediatrics, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: shanash@med.umich.edu