

Book Reviews

Essentials of Medical Genomics

Stuart Brown with John Hay and Harry Ostrer
Wiley, Hoboken, NJ, USA; 2002;
ISBN 0-471-21003-X; 288 pp.;
£33.50; US\$50.00

This short book is intended primarily as an introduction to genomics for medical students, although it would also serve as a useful primer for anyone looking for a non-technical introduction to current work in genomics, with an emphasis on its impact on clinical practice. The first chapter gives an introduction to the Human Genome Project, and to basic genetic processes such as replication, translation, transcription and linkage. Subsequent background chapters address genomics technologies (restriction enzymes, polymerase chain reaction, sequencing); bioinformatics (sequence alignment, BLAST, hidden Markov models, phylogenetics); and genetic variation in humans, with an emphasis on single nucleotide polymorphisms (SNPs). Attention then turns to areas of more direct clinical relevance, with a discussion of genetic testing, gene therapy, pharmacogenomics and toxicogenomics. In addition, there are chapters on microarrays, particularly for the analysis of gene expression, on proteomics and on ethical considerations. Finally, a comprehensive and very useful glossary is given.

The book is clearly and informally written, with jargon avoided wherever possible. It is also easily accessible, requiring a minimum of background knowledge. The level throughout is introductory, with very little technical detail given: instead, an overview of these topics is attempted. This is not an easy undertaking, because of the breadth and cross-disciplinary nature of the field, and the speed with which it is moving, but I think this book succeeds reasonably well. It certainly gives impressively comprehensive and up-to-date coverage, with some of the material close to current research activity. Inevitably, this means that the book will go out of date quickly; it also means that specialists may disagree with the choice of topics or feel that their area was under-represented. I certainly felt that my areas of expertise — statistics and genetic epidemiology — deserved more emphasis. I felt dissatisfied, for instance,

at a description of hidden Markov models as a ‘pattern analysis technique’: the crucial role of such statistical methods underpinning genomic analyses should be acknowledged to give a full appreciation of the field. At present, the only genetic epidemiology material is in the chapter on genetic variation. In my view, this is inadequate, given the potential importance of these studies for medical practice. At a minimum, the basic principles and study designs of genetic epidemiology should have been discussed — my preference would have been for a separate chapter on the subject. I also felt that the commendable desire to avoid technical detail became problematic in places, with some important issues blurred. Examples include the failure to consistently distinguish between linkage and association in Chapter 5, and the description in the glossary of BLAST E-values as probabilities rather than expectations (an error which the text, however, is careful to avoid). Another criticism is that the book is rather under-referenced; given its introductory nature, suggestions for further reading on the topics discussed would have been very helpful.

Despite these caveats, I felt that this was a useful book. It gives a clearly written, commendably brief and up-to-date introduction to an increasingly important area, and I will be glad to have it on my bookshelf.

John Whittaker
Department of Epidemiology and Public Health
Imperial College London
London, UK

Proteomic and Genomic Analysis of Cardiovascular Diseases

Edited by Jennifer E. Van Eyk and Michael Dunn
Wiley-VCH, Darmstadt, Germany; 2003;
ISBN 3-527-30596-3; 396 pp.;
€119

This is a very good book, whose best virtue is that it reliably reflects the current status of genomics and proteomics in the cardiovascular area. The inclusion of both fields in the same book could be considered excessively ambitious, but the result is very positive. The reader is able to envisage the change of scenery in biology, and to see that proteomics and genomics are providing a new perspective on cardiovascular diseases. People working in

these fields are witnessing an unprecedented rapid development of techniques, methods and instruments, which are providing a huge amount of information that should lead to new insights into cellular functions in health and disease. The book is well balanced between technical instrumentation, background and discussion of the most recent discoveries in the genomics and proteomics of cardiovascular pathologies.

The book consists of three sections: the first focuses on genomics (ten chapters, 170 pages), the second on proteomics (nine chapters, 170 pages) and the third (two chapters, 50 pages) gives an overview of the relevance and potential of these two disciplines in drug and diagnostics discoveries. Many of the chapters, particularly in the first two sections, have a similar structure: an introduction giving background information and outlining the principles behind the technique, followed by its application to a cardiovascular pathology. The cluster of chapter introductions constitutes by itself a very clear description of the principles of DNA arrays (section I, genomics) and the proteomics techniques (two-dimensional electrophoresis, liquid chromatography and mass spectrometry; section II, proteomics). In this regard, the book could have been entitled 'Proteomic and Genomic Analysis'.

The genomics section contains a comprehensive and up-to-date collection of chapters on the many techniques associated with the identification of genes in the main cardiac pathologies in humans and in animal models. The first three chapters provide overviews of large-scale expression profiling in several chronic cardiac diseases ('Large scale expression profiling in cardiovascular disease using microarrays: Prospects and pitfalls'; 'Global genomic analysis of cardiovascular disease: A potential map or blind alley?'; 'Heart failure: A genomic approach'). A constant, almost obsessive, characteristic of these chapters is the description of the potential pitfalls that can be found in the construction, as well as in the interpretation, of the data generated by DNA arrays. In fact, Chapter 7 ('Pitfalls associated with cDNA microarrays — a cautionary tale') is a charming and candid story of every misfortune that can happen when working with cDNA arrays. Other chapters are dedicated to myocardial protection and the stress response of the pulmonary vascular system. Although the analysis of atherosclerosis using genomic techniques is still at a very early stage, the last chapter of section I presents the available data concerning differential gene expression patterns in atherosclerosis, including the upregulated

expression of periphilin specifically in ruptured plaques and a brief description of laser microdissection technology. Overall, the first section provides a thorough description of the advantages and limitations of genomics, emphasising the enormous quantity of new information that is accumulating by means of this approach.

The proteomics section follows a similar scheme to that of the previous section. After an initial chapter on the fundamentals of proteomics ('Proteomics, a step beyond genomics: Application to cardiovascular disease'), the next three chapters provide an excellent description of the main proteomics techniques ('Mass spectrometry — a powerful analytical tool'; 'Differential expression proteomic analysis using isotope coded affinity tags'; 'Protein chip technology in proteomic analysis'). They are written in a not-too-sophisticated manner for the novice, without being too simplistic for the most experienced worker in the field. There follow five chapters focusing on specific subproteomes, with a detailed study of particular groups of proteins within cardiac and vascular cells ('Recent applications of functional proteomics: Investigations in smooth muscle cell physiology'; 'Identification of targets of phosphorylation in heart mitochondria'; 'Proteomic characterization of protein kinase C signalling tasks'; 'Identification of secreted oxidative stress-induced factors (SOXF) and associated proteins: Proteomics in vascular biology'; and 'Myofilament proteomics: Understanding contractile dysfunction in cardiorespiratory disease'). These chapters are paradigms of what proteomic analysis can offer once a protein of interest has been identified, including the study of post-translational modifications, isoform determination etc. These chapters therefore include some more 'classical' biochemical approaches. Overall, in the proteomics section, I would have liked to see a more detailed description of the global analysis of some pathologies (eg dilated cardiomyopathy) from which a large number of new data have been generated using the proteomics approach. Furthermore, dilated cardiomyopathy is probably the best example of a pathology for which both genomic and proteomic data exist (this is also true for heart failure), and the inclusion of a comparative critical evaluation of the same pathology from both sides would have been very appropriate. The editors, who are also the authors, point out that this area will be better understood over the next few years.

This is a book which is suitable for graduate students and post-doctoral and clinical research fellows, as well as for researchers interested in the new scientific disciplines

as applied to the cardiovascular field. It will be a hard read, however, for clinical cardiologists who are not involved in research. The format is not well suited for a textbook, with an uneven quantity and quality of diagrams and illustrations (informative and well chosen in many chapters, but absent in others, eg in the last two chapters). Overall, this is a highly recommended book, written by specialists in their fields, offering a clear and concise overview of the basis for, and latest developments in, cardiovascular genomics and proteomics.

*Fernando Vivanco
Fundación Jiménez Díaz
Madrid, Spain*

Handbook of Comparative Genomics

Cecilia Saccone and Graziano Pesole
Wiley, Hoboken, NJ, USA; 2003;
ISBN 0-471-39128-X; 442 pp.;
£64.50; €90.90

The preface to this book states that it is an ambitious attempt to offer a tool for biologists, to enable them to tackle the wealth of data within the genomic era. Indeed, it is ambitious, taking an unusual inter-disciplinary look at genomic research, while always firmly rooted in the underlying biology. Its aim is to look at the effect of large-scale sequencing on our knowledge of the properties and functions of species through the structure and function of the genome. It views comparative evolutionary genomics as the key to further understanding, whereby its definition of genomics also includes expression and the regulatory mechanisms underlying biological processes.

The book is organised into three parts. The first part, which for me is the most successful, describes a broad range of genomic features for prokaryotes, eukaryotes and organelles. Topics such as genome shape and size, gene content, codon use, morphology, classification, replication and expression are discussed and contrasted between the different genomes. A snapshot of the current state of genomic sequencing is given, with comprehensive directions to data sources and a rich bibliography throughout. Furthermore, the seven species whose nuclear genomes have been sequenced to date have been given special consideration, with results from the study of those sequences. This part of the book gives

an illuminating account of the nature of genomes, and is full of fascinating facts. While it is useful for biologists, the clarity of the writing and the lack of technical complexity also make it an ideal introduction to genomes for students of bioinformatics.

The second part of the book describes molecular biological techniques, biological databases and computational methods for genome sequence analysis. Modern techniques such as DNA sequencing of whole genomes and expressed sequence tags (ESTs), DNA microarrays and proteomics are described, often citing software packages used in conjunction with the data. The biological databases described include many major DNA sequence and protein databases, together with the bioinformatics centres that disseminate them. An unusual but welcome part of the database chapter is the discussion of the need for well structured, interoperable data sources and good access to the data. The section that lacks most detail here concerns genomic databases. Only a few resources are covered, possibly because there are so many of them. I do, however, feel that an opportunity has been missed to describe major international resources that are structured on a comparative genomic framework (eg the US Gramene resource for comparative analysis of the grasses). This omission is probably due to the back focussing on sequence analysis and other resources centring on marker data, rather than on genomic sequence. The computational methods section states the need for high-quality algorithms and software to analyse the wealth of data becoming available. This chapter gives a good overview of popular areas such as sequence similarity for DNA and protein sequencing, database searching, pairwise and multiple alignment, sequence clustering and assembly, gene prediction, protein structure prediction and phylogenetic analysis. The algorithms are well described and are illustrated with many clear examples and figures, and pointers to appropriate software and further reading. In addition, many less well known areas such as linguistic analysis of sequences are described, again with the same care. One disappointment here is that tools (eg the Twinscan software for gene prediction) which aim to gain inferences from comparing homologous elements across genomes, are not mentioned. Emerging fields, such as the use of genomic elements other than sequence (eg gene content and order, repetitive elements) for phylogenetic analysis, are not described. Furthermore, resources such as the TIGR Gene Indices are mentioned, but not the EGO database that compares their elements between species.

In addition, resources are sometimes split confusingly across the three chapters. For example, the CATH and SCOP protein classification databases are described in the computational methods chapter rather than the biological databases chapter. On the whole, however, this part gives a good overview of techniques, databases and computational methods, with the major resources and software tools being described with clarity and with directions to an enormous range of software and literature.

The final part of the book focuses on molecular evolution and molecular phylogeny. These chapters outline ways in which to compare genomes. They touch upon topics such as genome size and base composition, tying in knowledge which the reader will have gained in the two previous parts of the book. Furthermore, the evolution of prokaryotes, eukaryotes and organelles is discussed, along with an in-depth look at the evolution of the nuclear genome. Other important topics, such as analysis of non-coding regions, gene family expansion and genome duplication are also discussed. The molecular phylogeny chapter gives an overview of the validity of the molecular clock and the need to distinguish between orthologous and paralogous elements. Furthermore, a discussion is given on the future use of whole genome sequences, rather than single genes, to gain

phylogenetic knowledge and the comparison of nuclear versus organellar genomes for this purpose.

The book concludes by noting that the emerging datasets and new, more complex analytical approaches will lead to consistent views of comparative genome evolution across many disparate datasets. As with the first part of the book, this made for a highly interesting read, although I did feel that it was a rather brief ending to the book, touching upon many uses of comparative genomics, but missing others.

In summary, this book gives an illuminating look at the study of genomes through their sequence. Although this is ultimately the most informative view, it is a shame that non-sequence comparative approaches have been largely overlooked. The book unites topics not usually seen together, and, to a large extent, does this successfully. Although it is an enjoyable read in its own right, it is perhaps most useful in introducing the reader to new areas and in guiding them to further study. All in all, this book will be a welcome addition to the library of any laboratory, and can be enjoyed by biologists and bioinformaticians alike.

*Jo Dicks
John Innes Centre
Norwich, UK*