

Preface

The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work.

John von Neumann (1903–1957)

These ambiguities, redundancies, and deficiencies recall those attributed by Dr. Franz Kuhn to a certain Chinese encyclopaedia entitled Celestial Emporium of Benevolent Knowledge. On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance.

Jorge Luis Borges (1899–1986)

The analytical language of John Wilkins. In *Other Inquisitions* (1937–1952). University of Texas Press, 1984.

One of the central goals in biological sciences is to develop predictive models for the analysis and visualization of information. However, the analysis and visualization of biological data patterns have traditionally been approached as independent problems. Until now, biological data analysis has emphasized the automation aspects of tools and relatively little attention has been given to the integration and visualization of information and models.

One fundamental question for the development of a systems biology approach is how to build prediction models able to identify and combine multiple, relevant information resources in order to provide scientists with more meaningful results.

Unsatisfactory answers exist in part because scientists deal with incomplete, inaccurate data and in part because we have not fully exploited the advantages of integrating data analysis and visualization models. Moreover, given the vast amounts of data generated by high-throughput technologies, there is a risk of identifying spurious associations between genes and functional properties owing to a lack of an adequate understanding of these data and analysis tools.

This book aims to provide scientists and students with the basis for the development and application of integrative computational methods to analyse and understand biological data on a systemic scale. We have adopted a fairly broad definition for the areas of *genomics* and *proteomics*, which also comprises a wider spectrum of 'omic' approaches required for the understanding of the functions of genes and their products. This book will also be of interest to advanced undergraduate or graduate students and researchers in the area of bioinformatics and life sciences with a fairly limited background in data mining, statistics or machine learning. Similarly, it will be useful for computer scientists interested in supporting the development of applications for systems biology.

This book places emphasis on the processing of multiple data and knowledge resources, and the combination of different models and systems. Our goal is to address existing limitations, new requirements and solutions, by providing a comprehensive description of some of the most relevant and recent techniques and applications.

Above all, we have made a significant effort in selecting the content of these contributions, which has allowed us to achieve a unity and continuity of concepts and topics relevant to information analysis, visualization and integration. But clearly, a single book cannot do justice to all aspects, problems and applications of data analysis and visualization approaches to systems biology. However, this book covers fundamental design, application and evaluation principles, which may be adapted to related systems biology problems. Furthermore, these contributions reflect significant advances and emerging solutions for integrative data analysis and visualization. We hope that this book will demonstrate the advantages and opportunities offered by integrative bioinformatic approaches.

We are proud to present chapters from internationally recognized scientists working in prestigious research teams in the areas of biological sciences, bioinformatics and computer science. We thank them for their contributions and continuous motivation to support this project.

The European Science Foundation Programme on Integrated Approaches for Functional Genomics deserves acknowledgement for supporting workshops and research visits that led to many discussions and collaboration relevant to the production of this book.

We are grateful to our Publishing Editor, Joan Marsh, for her continuing encouragement and guidance during the proposal and production phases. We thank her Publishing Assistant, Andrea Baier, for diligently supporting the production process.

Francisco Azuaje and Joaquin Dopazo

Jordanstown and Madrid

October 2004