Preface

Since their discovery, microarrays have emerged as one of the most important biotechnological tools for studying the behavior of thousands of genes and genomic features on a genome-wide scale. The development of accurate and affordable technologies for effective and quick detection becomes paramount when considering the cost and time allowed to conduct a biological experiment. This has influenced the way in which transcriptional, proteomic and cell developmental studies are being carried out, providing a systematic way to study gene expression, elucidation of the proteome, discovering binding sites and methylation studies, all of these in a synchronized way. Transcriptional studies are one of the most important applications of microarrays, and for this, the development of gene expression arrays has helped enormously. Measuring the transcriptome or the entire repertoire of transcripts in an organism or cell lines provides a wealth of data about the relationships among DNA, transcription, translation, and phenotype. Advances in the microarray field have also allowed accurate detection that can uncover specific genetic mutations that may result in a number of human disorders. In this regard, microarrays, whether on full-length genes or short oligonucleotide assays, constitute an efficient tool for parallel and systematic analysis.

Although different shapes, resolutions, and features of microarrays may vary substantially across platforms, techniques and applications, they can be defined in general as a small glass slide with a collection of spots or elements in which biochemical reactions take place in order to reveal quantitative information about gene or protein expression, or other biological interaction of the molecules, under specific conditions. A microarray can be conceived as a small device (Greek word "mikro" for small), in which biomolecules are arranged in a systematic manner (French word "arayer" for array). In practice, a microarray is a device that has a rectangular shape and measures a few cm long – it is typically ordered, microscopic, planar, and specific.

Microarrays have been widely used by many researchers in diverse studies in biology. To mention a few facts, there have been more than 50,000 publications pertaining to microarrays, if PubMed is used as the source of information. Extending the scope to publications on "array" would list more than 120,000 abstracts in PubMed. Various databases and resources contain a wealth of microarray data for studies of gene expression, proteomics, tissue analysis, DNA methylation, genome-wide association studies, copy number variations and aberrations, and single nucleotide polymorphism, among others. The Gene Expression Omnibus database of the National Center for Integrative Biomedical Informatics of the United States contains a wealth of gene expression data, which are publicly accessible via the Internet. These databases contain millions of samples, mostly based on microarray studies.

Since the introduction of microarrays, huge amounts of data have been generated, and hence it has been imperative to create and improve specialized tools and

algorithms for the acquisition, analysis, and management and interpretation of the data. In every aspect, microarray data analysis is invariably an essential component of microarray studies. Roughly speaking, data analysis tools for microarrays are derived from a variety of bioinformatics methods and techniques that bridge the gap among molecular biology, computer science, statistics, and mathematics. Microarray data analysis takes on the images that are obtained from scanning and up to the biological and functional analysis, involving a sequence of well-established steps aiming at identifying locations of spots or probe cells, background, and noise, and to subsequently quantify and normalize the signals. Numerical data resulting from the pre-processing stages are analyzed by means of different techniques for clustering, data mining, and classification, followed by visualization, and network and functional biological analysis. Data management involves storage, databases, resources, and compression, which are important steps that complement the whole process of analysis. Microarray image and data analysis is essential, since the aim of the whole experimental procedure is to obtain meaningful biological conclusions, which depends on the accuracy of the analytical steps, mainly those at the beginning of the pre-processing stage.

This book provides a comprehensive review of the main, up-to-date methods, tools, and techniques for microarray image and data analysis. Internationally recognized experts address specific research topics and challenges in their areas of expertise, providing valuable knowledge about the state-of-the-art methods in the field, and covering the main steps for image processing, gridding, segmentation, noise treatment, and normalization. Biological aspects and applications of microarrays extend to a wide range of biological studies in cancer research, and different microarray types and platforms such as DNA, various types of oligonucleotide and tissue microarrays, as well as the most recent tools for microarray data analysis in the Bioconductor suite. Machine learning methods for microarray data analysis cover the main aspects of clustering, biclustering, multi-dimensional microarray data analysis, and reconstruction of regulatory networks.

The book is intended for students as a graduate textbook and as a reference for the design of graduate courses at an advanced level, while providing sufficient level of detail for general readers as well. In addition, one of the goals of this book is to provide a useful reference for researchers and practitioners in academia and industry who develop tools and carry out critical analysis, management, and interpretation of microarray data. By covering state-of-the-art methods in microarray image and data analysis, the book provides a handy reference for professors, researchers, and graduate and advanced-level undergraduate students in multi-disciplinary fields including bioinformatics, computer science, statistics, biology, biochemistry, genomics, and biomedical engineering. While the readers will also learn about the current state of the microarray technology and how data generated by the microarray experiments are analyzed, they will understand the current landscape regarding the applications, availability, and affordability of the research utilizing this technology.

This book has been made possible through the help of many people who participated at different stages of the production process. I would like to thank all of them

for their valuable help and support to make this happen. I am indebted to the book series editor, Rastislav Lukac, who specially invited me to edit this book and guided me through the main steps. My special thanks are for Adnan Ali for adding a self-contained biological insight to the book, which provides an excellent complement to the computing and statistical methods being presented. Special thanks are also for Iman Rezaeian for his tireless hours of work put into the edition and format conversions, as well as his contribution to a chapter. I would also like to thank Alioune Ngom for his contribution and help in reviewing, and for his constant logistic support, and Yifeng Li for his additional help in the review process.

Finally, I would like to thank the expert contributors for putting their efforts in carefully writing up-to-date and high quality chapters about their specific topic, while keeping the whole volume well synchronized, and for their help in the reviewing process. I would like to acknowledge the external reviewers, Michael Crawford, Meng Li, A. Sri Nagesh, Moysés Nascimento, Tuan Pham, Armando Pinho, Alexander Pozhitkov, Khedidja Seridi, Guifang Shao, Ruisheng Wang, Xiongwu Wu, Hong Yan, and Jie Zheng. With their solid expertise in the specific topics, they added the valuable feedback needed to establish these contributions as leading in the field.