PREFACE

Omic studies aim at the abundance and (or) structural characterization of a wide variety of biological molecules in organisms under distinct scenarios in clinical, environmental or food science fields. Among all the different omics, those related to DNA (*genomics*, *epigenomics*), RNA transcripts (*transcriptomics*), proteins (*proteomics*) and metabolites (*metabolomics*) are of special interest due to their presence in the different layers of the biological information flux, from the genetic information to the observed phenotype. The comprehensive study of these molecules, which are interconnected and involved in key biological pathways, provides new insights underneath the biological process studied. The emergence of these omic fields has been possible due to the technological revolution of the analytical instrumentation. Nowadays, thousands of genes, proteins or metabolites can be simultaneously analysed by means of high-performance analytical technologies allowing the complete characterization of the biological state of a particular sample. The increasingly larger data sets generated by these technologies have motivated the use and development of advanced chemometric data analysis methods, in order to extract relevant chemical and biological information from them.

The primary goal of this volume is to become a reference for those researchers working in the omic field with interest in the data processing aspects. Data analysis methodologies developed in the chemometric field are covered in detail, from the initial experimental design, data preprocessing and traditional statistical approaches to the more advanced multivariate and megavariate integration data analysis methods. This volume complements other previously published books, such as Volumes 63 and 64, devoted to fundamentals and applications of advanced omic technologies in the *Comprehensive Analytical Chemistry* series related to the omic field. In addition, this volume is focused on the data analysis approaches used in transcriptomics and metabolomics, as they are currently used in a diverse type of applications.

Chapters 1 and 2 serve as a general introduction to the relevance of the omic sciences with particular attention to the experimental methodologies used in the different omic studies. The following three chapters are devoted to transcriptomic data analysis. While Chapter 3 is focused on microarray-generated data, Chapters 4 and 5 deal with high-throughput sequencing

techniques (RNA and DNA bisulphite, respectively). Next three chapters are related to metabolomic data preprocessing. The importance of quality assessment in metabolomic workflows and the description of some useful strategies are discussed in Chapter 6. Chapter 7 presents different methods for data normalization before their statistical analysis and their impact on the obtained results. Chapter 8 summarizes the required steps needed to preprocess NMR and MS metabolomic data. The following four chapters introduce some important chemometric methods commonly used in metabolomics. First, data exploration and data decomposition methods such as principal component analysis are depicted in Chapter 9. Later, Chapter 10 focuses on methods used for data classification and discrimination, with special attention to feature selection approaches derived from them. Advanced statistical multivariate data analysis methods are introduced in Chapter 11 where recent approaches such as ASCA (ANOVA simultaneous component analysis) are presented in detail. Chapter 12 describes the application of compression and resolution tools for nontargeted metabolomic MS data analysis with special emphasis in the recently proposed ROIMCR strategy. In Chapter 13, the most common data analysis methods used for the analysis of mass spectrometry images are introduced, highlighting that the spatial distribution of molecules adds valuable information to the studies of biological systems. Many of these different chemometric approaches are summarized in Chapter 14 where the variety of available approaches and tools is reviewed. Then, the different options for the annotation and identification of the detected metabolites are depicted in Chapter 15. Next three chapters present recent research topics about multiomics data integration analysis. These multiomics data integration approaches are discussed considering model-driven approaches (Chapter 16), the particular requirements of fusing information from different analytical platforms (Chapter 17) and the statistical techniques used when considering time series experiments (Chapter 18). The last four chapters of the volume demonstrate the usefulness of the previously presented approaches in different research fields, with examples in environmental metabolomics (Chapter 19) and risk assessment (Chapter 20) as well as in its application in biomedical (Chapter 21) and food (Chapter 22) analysis.

We would like to acknowledge all the authors for accepting our invitation and their efforts to prepare all these excellent chapters on omic data analysis in a timely manner. The proposal and preparation of this book have been possible due to the support from the European Research Council under the European Union's Seventh Framework

Programme (FP/2007–2013)/ERC Grant Agreement No. 320737, CHEMAGEB Project, particularly during the preparation of Chapters 1, 2, 12, 16, 19 and 20.

We also want to acknowledge Prof. Damià Barceló, editor of this series, for inviting us to prepare this multiauthor volume devoted to omics data analysis techniques. We appreciate the work done by the Elsevier editorial team and especially the efforts of the Editorial Project Managers Mr. Andrae Akeh and Ms. Teresa Pons-Ferrer.

We hope that the book final content will be useful for those researchers who want to deepen more in the omics data analysis.

Joaquim Jaumot, Carmen Bedia and Romà Tauler
Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Spain