

Preface

Following the genomics revolution, proteomics, metabolomics, and lipidomics studies have emerged, among others, as a natural follow-up either in the investigation of human biology or, similarly, in animal or plant studies. The combined use of these omics fields may lead to a more comprehensive understanding of system biology. In medical research, it may generate novel biomarkers which may be used in early detection of disease and for the development of new screening programs. Examples are in practical clinical patient monitoring, and in the development of new patient disease management rules for complex diseases such as cancer and cardiovascular or inflammatory diseases, among others. Another key objective of such comprehensive system-level study is in the elucidation of the molecular biochemical process associated with these biomarkers. In plant biology, relevant applications may be in the prediction of desirable properties of novel foods or crops or in assisting genetic manipulation-based breeding programs for new varieties, for example.

Within the omics revolution, proteins play a key role in the study of living organisms as they provide the essential link with the genome on the one hand, while they are also key components of the physiological metabolic pathways of cells. Proteomics is therefore a fundamental research field which investigates the structure or function of protein complexes consisting of multiple proteins simultaneously. Metabolites, on the other hand, are small molecules formed from the breakdown products of larger molecules, such as proteins, after undergoing a metabolic process within an organism. They are involved in (cell) signaling processes through stimulatory and inhibitory effects on enzymes, among others. Comprehensive metabolic profiling or metabolomics can give an instantaneous snapshot of the physiology of the cell. Lipidomics is closely related to metabolomics but studies a specific set of non-water-soluble molecules consisting of glycerol and fatty acids. The collection of all lipids in an organism is referred to as *lipidome*, in analogy to *genome*, *proteome*, *metabolome*, and so on. Although all these molecules have differences in both structure and functions, they can all be studied experimentally using modern spectrometric technology—specifically mass spectrometry—to assess the required omics expression of interest.

Unlike measurement procedures and methodology in genomics research, which is reasonably standardized at the time of writing, mass spectrometry is itself a vast field with many forms and variants. Typical vehicles are time-of-flight (MALDI) mass spectrometry, liquid chromatography-mass spectrometry (LC-MS), and Fourier transform ion cyclotron resonance mass spectrometry (FTICRMS or FTMS), among others. The field is therefore still very much in flux, with many distinct mass spectrometric measurement technologies and hence also different study designs in use. The data types generated in these studies are also very different. They tend to have complex structures, while no consensus data analytical approaches have yet been agreed upon. For these reasons, expert knowledge gained with one specific measurement platform does not easily carry over to other mass spectrometric systems. Writing a comprehensive overview text on statistical data analytic methodology in the new mass spectrometry-based omics field would therefore not be realistic. Instead, we have chosen to bring together a group of established researchers to present their expert knowledge in their specific application area within this emerging field. With this book, we want to provide an overview of the current status of such mass spectrometry-based omics data analysis and give impetus to the emergence of a common view on the design and analysis of such data and experiments. In this way, the book could support the development of more standardized templates, research practices, and references for any data analyst such as statisticians, computer scientists, computational biologists, analytical chemists, and data scientists, both in the omics application fields we discuss and in related omics fields such as glycomics.

Materials presented in this book cover a broad range of topics. First, we discuss the preprocessing of mass spectrometry data such as data normalization, alignment, denoising, and peak detection including monoisotopic peaks. Second, it provides methods for identification of proteins from tandem mass spectrometry data. There is also a chapter on a software package for the analysis of such omics data. Additionally, it has chapters on downstream data analysis using Bayesian and frequentist statistical predictive modeling and classification techniques. Last but not the least, there are chapters on specific examples of biomarker detection using proteomics, metabolomics, and lipidomics data. We hope that this book will be suitable for the scientists, practitioners, software developers, as well as advanced students of statistics, computer science, computational biology, and biomedical sciences.

Gainesville, FL, USA
Leiden, The Netherlands

Susmita Datta
Bart J.A. Mertens

Statistical Analysis of Proteomics, Metabolomics, and
Lipidomics Data Using Mass Spectrometry

Datta, S.; Mertens, B.J.A. (Eds.)

2017, VIII, 295 p. 106 illus., 83 illus. in color., Hardcover

ISBN: 978-3-319-45807-6