前　言

本书从多个侧面对组学数据生物信息学做了详尽的介绍。组学数据生物信息学是一个全新的研究领域，是分子生物学、应用信息学和统计学等多个学科的交叉和整合。近年来，生物信息学已成为以数据为驱动的组学研究领域的常规技术，也是组学应用研究人员必须掌握的重要技能。生命科学分析仪器的不断小型化，以及数据测定技术的快速发展，使我们能够同时处理并分析多种细胞组份及其状态，为解释特定生命现象提供了有力的工具。然而，如果不作恰当的处理和分析，组学数据对我们更好理解所研究的生命现象毫无裨益，就连对组学方法产生的海量原始数据的管理，在某种程度上也已经成为一种技术。

生物信息学通常被认为是一门技术科学。实际上，它是以计算机为工具进行生物学研究的交叉科学，即“计算生物学”，或“计算分子生物学”。组学的发展促使生命科学的研究模式发生了革命性的改变，传统的以问题为驱动的研究拓展为从数据出发的探索性研究。目前，组学恰恰处于上述两种研究模式彼此融合的发展阶段。从这一点上说，组学数据生物信息学不只关注数据的管理和分析，更关注生物问题的提出和假设的验证。目前，将数据分析策略整合在一起的生物信息工作流程已经出现，而这些流程本身的复杂性也反映了生物体和生命过程的复杂性。通过对大量调控方式和作用网络的综合分析，我们有可能从更高层次上理解细胞内的生命过程。不言而喻，组学生物信息学正逐步向计算系统生物学过渡，其最终目标是对高度动态的生命现象建立定量模型，用以预测由细胞内各种分子间相互作用产生的复杂生命活动的分子过程及功能。

毋庸置疑，组学数据生物信息学的研究对象十分复杂，它是引领现代分子生物学研究的重要分支学科。希望本书不仅能对组学研究起到具体指导的作用，也能为读者展现这一研究领域极具魅力的美好图景。

本书共分三篇。第一篇介绍核心分析策略、标准分析规范、数据管理指南，以及用于分析组学数据的基本统计方法。第二篇介绍用于基因组、转录组、蛋白质组、代谢组等各种不同组学数据的生物信息学分析方法，包括基本概念和实验背景，以及原始数据预处理和深入分析的基本方法。第三篇则介绍如何利用生物信息学进行组学数据分析的实例，包括人类疾病相关生物标记鉴定和靶标识别等具体例子。

衷心感谢参与本书各章编写的所有作者，感谢 John Walker 促成了编著本书的设想。书中若有纰漏之处，责任当由本人承担。希望大家阅读愉快。

奥地利　维也纳
贝恩德·迈尔

（罗静初　译）

Preface

This book discusses the multiple facets of "Bioinformatics for Omics Data," an area of research that intersects with and integrates diverse disciplines, including molecular biology, applied informatics, and statistics, among others. Bioinformatics has become a default technology for data-driven research in the Omics realm and a necessary skill set for the Omics practitioner. Progress in miniaturization, coupled with advancements in readout technologies, has enabled a multitude of cellular components and states to be assessed simultaneously, providing an unparalleled ability to characterize a given biological phenotype. However, without appropriate processing and analysis, Omics data add nothing to our understanding of the phenotype under study. Even managing the enormous amounts of raw data that these methods generate has become something of an art.

Viewed from one perspective, bioinformatics might be perceived as a purely technical discipline. However, as a research discipline, bioinformatics might more accurately be viewed as "[molecular] biology involving computation." Omics has triggered a paradigm shift in experimental study design, expanding beyond hypothesis-driven approaches to research that is basically explorative. At present, Omics is in the process of consolidating various intermediate forms between these two extremes. In this context, bioinformatics for Omics data serves both hypothesis generation and validation and is thus much more than mere data management and processing. Bioinformatics workflows with data interpretation strategies that reflect the complexity of biological organization have been designed. These approaches interrogate abundance profiles with regulatory elements, all expressed as interaction networks, thus allowing a one-step (descriptive) embodiment of wide-ranging cellular processes. Here, the seamless transition to computational Systems Biology becomes apparent, the ultimate goal of which is representing the dynamics of a phenotype in quantitative models capable of predicting the emergence of higher order molecular procedures and functions that arise from the interplay of basic molecular entities that constitute a living cell.

Bioinformatics for Omics data is certainly embedded in a highly complex technological and scientific environment, but it is also a component and driver of one of the most exciting developments in modern molecular biology. Thus, while this book seeks to provide practical guidelines, it hopefully also conveys a sense of fascination associated with this research field.

This volume is structured in three parts. Part I provides central analysis strategies, standardization, and data management guidelines, as well as fundamental statistics for analyzing Omics profiles. Part II addresses bioinformatics approaches for specific Omics tracks, spanning genome, transcriptome, proteome, and metabolome levels. For each track, the conceptual and experimental background is provided, together with specific guidelines for handling raw data, including preprocessing and analysis. Part III presents examples of integrated Omics bioinformatics applications, complemented by case studies on biomarker and target identification in the context of human disease.

I wish to express my gratitude to all authors for their dedication in providing excellent chapters, and to John Walker, who initiated this project. As for any omissions or errors, the responsibility is mine. In any case, enjoy reading.

Vienna, Austria *Bernd Mayer*