

Chapter VII

Data Mining and Knowledge Discovery in Metabolomics

Christian Baumgartner

University for Health Sciences, Medical Informatics and Technology (UMIT), Austria

Armin Graber

BIOCRATES Life Sciences GmbH, Austria

ABSTRACT

This chapter provides an overview of the knowledge discovery process in metabolomics, a young discipline in the life sciences arena. It introduces two emerging bioanalytical concepts for generating biomolecular information, followed by various data mining and information retrieval procedures such as feature selection, classification, clustering, and biochemical interpretation of mined data, illustrated by real examples from preclinical and clinical studies. The authors trust that this chapter will provide an acceptable balance between bioanalytics background information, essential to understanding the complexity of data generation, and information on data mining principals, specific methods and processes, and biomedical applications. Thus, this chapter is anticipated to appeal to those with a metabolomics background as well as to basic researchers within the data mining community who are interested in novel life science applications.

INTRODUCTION

Metabolomics is an evolving discipline that studies unique chemical fingerprints reflecting metabolic changes related to disease onset and progression. Metabolite profiling, an area within

metabolomics, measures small molecules, or metabolites, contained in a human cell, tissue, or organ and involved in primary and intermediary metabolism. The biochemical information resulting from metabolite analysis reveals functional endpoints associated with physiological and

pathophysiological processes, influenced by both genetic predisposition and environmental factors such as nutrition exercise or medication (Daviss, 2005; Harrigan & Goodacre, 2003; Ryals, 2004; Schmidt, 2004).

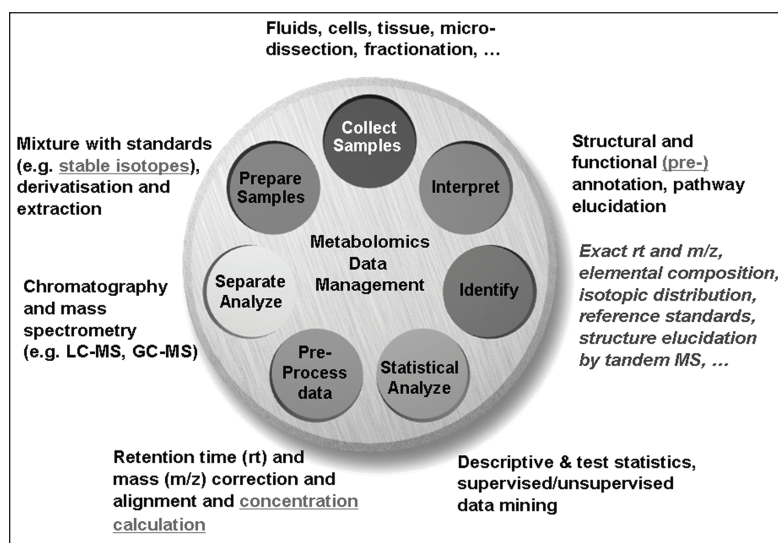
Recently, due to significant advances in high-throughput technologies, a wider set of the human metabolome—a thus far largely unexplored source of bioinformation—is now accessible (Beecher, 2003; Dunn, Bailey, & Johnson, 2005). Statistical comparison of metabolite profiles can expose multivariate patterns that have the potential to revolutionize the health care system by specifically capturing latent warning signs of upcoming diseases before any disease symptoms show up. Early disease screening and prevention, as opposed to late disease detection and expensive therapeutic interventions, is probably the primary health care coverage solution for the future.

By definition, these so-called biomarkers are “objectively measured indicators of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention, and ... are intended to substitute for a

clinical endpoint (predict benefit or harm) based on epidemiological, therapeutic, pathophysiological or other scientific evidence” (Biomarkers Definitions Working Group, 2001). Interest in the discovery of novel biomarkers originates from their broad range of potential applications and fundamental impact on pharmaceutical industry dynamics and current health care sector principles. Successful implementation of biomarkers in drug discovery can reduce the time and cost of drug development while the application to molecular diagnostics will improve patient compliance in clinical settings and reduce unnecessary costs resulting from false diagnosis in addition to late disease detection (McCandless, 2004; Morris & Watkins, 2005; Stoughton & Friend, 2005).

Qualitative and quantitative metabolite profiling technologies comprise a range of advanced analytical and data processing tools, with the objective of utilizing potential markers as a result of comparison of small molecule components of biological systems. Tandem mass spectrometry (MS/MS), for example, detects hundreds of metabolites simultaneously from microliter quanti-

Figure 1. Mass spectrometry (MS) based technologies used in metabolite profiling. Specific steps for qualitative nontargeted and quantitative targeted profiling are highlighted.



24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/data-mining-knowledge-discovery-metabolomics/29958

Related Content

A Survey on Overlapping Communities in Large-Scale Social Networks

S Rao Chintalapudi and H. M. Krishna Prasad M (2018). *Modern Technologies for Big Data Classification and Clustering* (pp. 198-215).

www.irma-international.org/chapter/a-survey-on-overlapping-communities-in-large-scale-social-networks/185984

Recent Developments on Security and Reliability in Large-Scale Data Processing with MapReduce

Christian Esposito and Massimo Ficco (2016). *International Journal of Data Warehousing and Mining* (pp. 49-68).

www.irma-international.org/article/recent-developments-on-security-and-reliability-in-large-scale-data-processing-with-mapreduce/143715

Mining Frequent Generalized Patterns for Web Personalization in the Presence of Taxonomies

Panagiotis Giannikopoulos, Iraklis Varlamis and Magdalini Eirinaki (2010). *International Journal of Data Warehousing and Mining* (pp. 58-76).

www.irma-international.org/article/mining-frequent-generalized-patterns-web/38954

Web Usage Mining for Ontology Management

Brigitte Trousse, Marie-Aude Aufaure, Bénédicte Le Grand, Yves Lechevallier and Florent Massegia (2008). *Data Mining with Ontologies: Implementations, Findings, and Frameworks* (pp. 37-64).

www.irma-international.org/chapter/web-usage-mining-ontology-management/7571

Sarcasm Detection Using RNN with Relation Vector

Satoshi Hiai and Kazutaka Shimada (2019). *International Journal of Data Warehousing and Mining* (pp. 66-78).

www.irma-international.org/article/sarcasm-detection-using-rnn-with-relation-vector/237138