

Chapter 1

Unsupervised Data Analysis Methods used in Qualitative and Quantitative Metabolomics and Metabonomics

Miroslava Cuperlovic-Culf

Institute for Information Technology, National Research Council, Canada

ABSTRACT

Metabolomics or metabonomics is one of the major high throughput analysis methods that endeavors holistic measurement of metabolic profiles of biological systems. Data analysis approaches in metabolomics can broadly be divided into qualitative – analysis of spectral data and quantitative – analysis of individual metabolite concentrations. In this work, the author will demonstrate the benefits and limitations of different unsupervised analysis tools currently utilized in qualitative and quantitative metabolomics data analysis. Following a detailed literature review outlining different applications of unsupervised methods in metabolomics, the author shows examples of an application of the major previously utilized unsupervised analysis methods. The testing of these methods was performed using qualitative as well as corresponding quantitative metabolite data derived to represent a large set of 2,000 objects. Spectra of mixtures were obtained from different combinations of experimental NMR measurements of 13 prevalent metabolites at five different groups of concentrations representing different phenotypes. The analysis shows advantages and disadvantages of standard tools when applied specifically to metabolomics.

DOI: 10.4018/978-1-61350-435-2.ch001

INTRODUCTION

Over the last decade many new technologies have been developed and utilized in order to understand and describe the complexity of biological systems. High throughput methods involving the parallel measurement of biological molecule concentrations have been applied to many different systems in different environments (Schena, 1995; Fiehn, 2000; Nicholson, 2002). Large amount of data created by these methods have made data analysis a major part of most biological explorations nowadays. In many examples of high throughput methodology a strong emphasis in the first level of statistical analysis is on unsupervised approaches. Unsupervised data analysis is employed for obtaining connections between samples and/or molecular features without biasing the results by the introduction of prior knowledge. Thus, the development of methods and tools for unsupervised data analysis as well as the exploration of the most optimal methods and their most appropriate uses in the analysis of various high throughput, *i.e.* omics, biological data is an active research area.

One of the major, fast developing high throughput analysis methods, attempts a holistic measurement of metabolic profiles, *i.e.* metabolom or metabonom, of biological systems. Metabolomics, initially defined as the global analysis of all metabolites in a sample (Fiehn, 2001; Oliver, 1998) and metabonomics perceived as the analysis of metabolic responses to drugs or diseases (Nicholson, 1999) are nowadays often interchangeable terms broadly referring to the multi-component analysis of metabolites in a biological system (Beckoners, 2007). Metabolomics provides to systems biology a functional readout of changes determined by genetic code, regulation, protein abundance and modifications as well as environmental influences. Metabolic differences in biological fluids, cells and tissues provide the closest link to the various phenotypical responses showing the actual effects of various phenotype changes, drug effects, toxicological responses or disease states. The other functional

genomics technologies, such as transcriptomics and proteomics indicate the potential causes for phenotype response. Metabolomics can be viewed as a “re-invention” or extension of the approaches of analytical biochemistry of the 1960s. However, there are some major differences between modern metabolomics and analytical biochemistry of the past. First, is the introduction of highly advanced and reliable instrumentation such as nuclear magnetic resonance (NMR) and mass spectrometers (MS) for parallel, quantitative analysis of complex biological samples. Second, is the introduction of a novel data driven approach aimed at observing all measurable metabolites without any preconceptions or preselection. Third, is the introduction of various data analysis procedures, computational tools and methodologies that are able to quantitatively and accurately analyze large amounts of data. These tools must handle, store, pre-process and analyze complex and often large datasets (Cuperlovic-Culf, 2010). This new approach to metabolic phenotyping has emerged as a powerful new way to augment genomic, transcriptomic and proteomic methods for the capture of molecular information already applied to a range of biological systems (Bictash, 2009). Similarly to other types of omics methodologies large amounts of data produced in metabolomics experiments require application of appropriate and often complex data analysis tools.

Depending on the type of metabolic data available the data analysis approach in metabolomics can broadly be divided into qualitative and quantitative with the major steps and applications of these two approaches outlined in Figure 1. The type of analysis employed defines the necessary pre-processing steps. In qualitative metabolic analysis complete metabolomics spectra or spectral regions are used. In contrast, quantitative metabolomics initially performs compound identification and quantification. Once metabolite concentration information is obtained these data can be used for various applications including the development of systems biology models or biomarker discovery.

26 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/unsupervised-data-analysis-methods-used/60826

Related Content

Blockchain Technology Changing Traditional Methods of Applied Research in Bioinformatics

Natalia Rasskazova and Liudmila Ratushnaia (2019). *International Journal of Applied Research in Bioinformatics* (pp. 66-74).

www.irma-international.org/article/blockchain-technology-changing-traditional-methods-of-applied-research-in-bioinformatics/231591

Identification of Distinguishing Motifs

Wangsen Feng and Lusheng Wang (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 53-67).

www.irma-international.org/article/identification-distinguishing-motifs/47096

A Web Database IR-PDB for Sequence Repeats of Proteins in the Protein Data Bank

Selvaraj Samuel and Mary Rajathej (2017). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 1-10).

www.irma-international.org/article/a-web-database-ir-pdb-for-sequence-repeats-of-proteins-in-the-protein-data-bank/190790

Electromyography-Based Functional Electrical Stimulation (FES) in Rehabilitation

Poulami Ghosh, Ankita Mazumder, Anwesha Banerjee and D.N. Tibarewala (2016). *Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes* (pp. 337-355).

www.irma-international.org/chapter/electromyography-based-functional-electrical-stimulation-fes-in-rehabilitation/140499

Interaction of Nucleic Acids: Hidden Order of Interaction

Gennadiy Vladimirovich Zhizhin (2021). *International Journal of Applied Research in Bioinformatics* (pp. 1-8).

www.irma-international.org/article/interaction-of-nucleic-acids/278747