

Chapter 2

Data Analysis and Interpretation in Metabolomics

Jose M. Garcia-Manteiga

Protein Transport and Secretion Unit, San Raffaele – DIBIT, Italy

ABSTRACT

Metabolomics represents the new ‘omics’ approach of the functional genomics era. It consists in the identification and quantification of all small molecules, namely metabolites, in a given biological system. While metabolomics refers to the analysis of any possible biological system, metabonomics is specifically applied to disease and physiopathological situations. The data collected within these approaches is highly integrative of the other higher levels and is hence amenable to be explored with a top-down systems biology point of view. The aim of this chapter is to give a global view of the state of the art in metabolomics describing the two analytical techniques usually used to give rise to this kind of data, nuclear magnetic resonance, NMR, and mass spectrometry. In addition, the author will focus on the different data analysis tools that can be applied to such studies to extract information with special interest at the attempts to integrate metabolomics with other ‘omics’ approaches and its relevance in systems biology modeling.

1. INTRODUCTION

Metabolomics is a relatively new ‘omics’ strategy aimed at the identification and quantification of potentially all small metabolites in a given biological system. As for genomics, transcriptomics,

and proteomics before, metabolomics represents a picture, a downstream snapshot of the molecular biology dogma flow of information, from genes to proteins and beyond. Small metabolites of almost all kinds, amino acids, lipids, sugars, nucleotides, organic acids, amines etc., define the metabolic composition of a biological system and its changes are governed by chemistry, physiology, biochem-

DOI: 10.4018/978-1-61350-435-2.ch002

istry, molecular biology, and genetics in the end. It is therefore one of the most integrative ‘omics’ approach available today since very small changes in the upstream levels, the transcriptome and the proteome, which could be hardly detectable with the current technologies, could give rise to detectable changes in the levels of metabolites. Hence, metabolomics pursues the highest level of integration within the functional genomics field and is therefore, alone or in combination with other techniques, perfectly suitable to systems biology approaches for their use at different levels such as ecology, physiology, biochemistry, and cell biology studies amongst others.

The metabolomics field puts together a relevant amount of scientific skills, from analytical chemistry, cell biology, physiology, and biochemistry, passing through bioinformatics, statistics, and mathematics. Beginning with good study design and sampling procedures, not only is mandatory to obtain good quality data but to have good expertise in analytical chemistry, such as nuclear magnetic resonance or mass spectrometry skills, which will be required after samples are obtained. Once spectra are acquired, a good statistical and bioinformatics approach will distinguish a success from a failure. Moreover, we must define the goal of our metabolomics study in an early state since there are plenty of bioinformatics and chemometric approaches to apply to the huge amount of data that the analytical high-throughput techniques are currently able to supply with. Each of these approaches is best fitted for different outputs of the study. For example, in clinical metabolomics, PCA (Principal Component Analysis), which will be discussed later in this chapter, is currently the most frequently used tool for extracting metabolic information combined with PLS-DA (Partial Least Square Discriminant Analysis) in a first approach to the data. If big data matrices are available, however, other statistical tools, such as SVM (Support Vector Machines) or genetic algorithms are best suited for the discovery of biomarkers, which is very often the final goal in clinical metabolomics

approaches. One of the goals of this chapter is to bring about a discussion of the state of the art data analysis and interpretation in metabolomics studies, focusing on the attempts to integrate the metabolome information with the proteome and its application at the systems biology level.

In this chapter I will try to give an overall prospective of metabolomics studies, beginning with some definitions and historical approaches to the field as well as its several applications. I will describe the process from the study design and sampling procedures for the different types of approaches covering also the two main analytical techniques, NMR and Mass Spectrometry. I will show their advantages and disadvantages, discussing their technical challenges for the future of the field, such as single-cell metabolomics applications. I will then highlight the most recent aim of combining NMR and MS techniques to get as close to the global metabolomics goal as possible. In the final section I will describe the databases and mining software solutions for data interpretation with special focus on the modeling uses of metabolomics data for systems biology and the integrative approaches with other ‘omics’ techniques such as proteomics.

1.1 Background: History and Applications

Although the term metabolomics is widely accepted in the research community as a scientific area focused on the analysis and interpretation of metabolite levels in biological samples, the different approaches to such a goal have received different names, giving rise to a vast terminology during the last 20 years, i.e. metabolic profiling, fingerprinting, footprinting etc. I will define a few of the most used approaches in the field but for a good review on their historical hints you can refer to Oldiges (Oldiges et al., 2007). According to Fiehn (Fiehn, 2002), metabolomics technologies are subdivided into *target analysis*, aimed at quantitative analysis of substrate and/or

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/data-analysis-interpretation-metabolomics/60827

Related Content

MicroRNA Precursor Prediction Using SVM with RNA Pairing Continuity Feature

Huan Yang, Yan Wang, Trupti Joshi, Dong Xu, Shoupeng Yu and Yanchun Liang (2011). *Interdisciplinary Research and Applications in Bioinformatics, Computational Biology, and Environmental Sciences* (pp. 73-82).

www.irma-international.org/chapter/microrna-precursor-prediction-using-svm/48366

Molecular Dynamics Simulation of Interlayer Structure and Hydration Properties of Glycine Intercalated Layered Double Hydroxides

Guo-Xiang Pan, Feng Cao, Pei-Song Tang, Hai-Feng Chen, Zhe-Ming Ni, Jin-Tian Yang, Li-Geng Wang and Min-Hong Xu (2011). *Interdisciplinary Research and Applications in Bioinformatics, Computational Biology, and Environmental Sciences* (pp. 242-250).

www.irma-international.org/chapter/molecular-dynamics-simulation-interlayer-structure/48381

Genomatrices and the Genetic Octet Yin-Yang-Algebras

Sergey Petoukhov and Matthew He (2010). *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications* (pp. 132-167).

www.irma-international.org/chapter/genomatrices-genetic-octet-yin-yang/37900

Pharmacy Data Integrity for Optimal Analytics

C. David Butler (2014). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 21-45).

www.irma-international.org/article/pharmacy-data-integrity-for-optimal-analytics/147302

Genetic Code: Emergence of Symmetrical Pattern, Beginnings of Matrix Genetics

Sergey Petoukhov and Matthew He (2010). *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications* (pp. 1-30).

www.irma-international.org/chapter/genetic-code-emergence-symmetrical-pattern/37894