



# Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome

**Pamela Vernocchi<sup>1,2</sup>, Lucia Vannini<sup>1,3</sup>, Davide Gottardi<sup>1</sup>, Federica Del Chierico<sup>2</sup>, Diana I. Serrazanetti<sup>1</sup>, Maurice Ndagijimana<sup>4</sup> and Maria E. Guerzoni<sup>3\*</sup>**

<sup>1</sup> Interdepartmental Centre for Industrial Research-CIRI-AGRIFOOD, Alma Mater Studiorum, University of Bologna, Bologna, Italy

<sup>2</sup> Parasitology Unit, Department of Laboratories, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

<sup>3</sup> Department of Agricultural and Food Sciences, Alma Mater Studiorum, University of Bologna, Bologna, Italy

<sup>4</sup> Department of Agricultural Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada

## Edited by:

Lorenza Putignani, Children's Hospital and Research Institute Bambino Gesù, Italy

## Reviewed by:

Lorenza Putignani, Children's Hospital and Research Institute Bambino Gesù, Italy  
Andrea Petrucca, Azienda Ospedaliera Sant'Andrea, Italy

## \*Correspondence:

Maria E. Guerzoni, Professor Alma Mater, Department of Agricultural and Food Sciences, Alma Mater Studiorum, University of Bologna, Via Fanin, 46, 40127 Bologna, Italy. e-mail: [elisabetta.guerzoni@unibo.it](mailto:elisabetta.guerzoni@unibo.it)

Bacteria colonizing the human intestinal tract exhibit a high phylogenetic diversity that reflects their immense metabolic potentials. The catalytic activity of gut microbes has an important impact on gastrointestinal (GI) functions and host health. The microbial conversion of carbohydrates and other food components leads to the formation of a large number of compounds that affect the host metabolome and have beneficial or adverse effects on human health. Metabolomics is a metabolic-biology system approach focused on the metabolic responses understanding of living systems to physio-pathological stimuli by using multivariate statistical data on human body fluids obtained by different instrumental techniques. A metabolomic approach based on an analytical platform could be able to separate, detect, characterize and quantify a wide range of metabolites and its metabolic pathways. This approach has been recently applied to study the metabolic changes triggered in the gut microbiota by specific diet components and diet variations, specific diseases, probiotic and synbiotic food intake. This review describes the metabolomic data obtained by analyzing human fluids by using different techniques and particularly Gas Chromatography Mass Spectrometry Solid-phase Micro Extraction (GC-MS/SPME), Proton Nuclear Magnetic Resonance (<sup>1</sup>H-NMR) Spectroscopy and Fourier Transform Infrared (FTIR) Spectroscopy. This instrumental approach has a good potential in the identification and detection of specific food intake and diseases biomarkers.

**Keywords:** metabolites, instrumental methods, gut microbiota, diet, biomarkers

## INTRODUCTION

The human intestine is home of some 100 trillion of microorganisms of at least hundreds species. The density of bacterial cells in the colon has been estimated at  $10^{11}$ – $10^{12}$  per ml, which makes it one of the most densely populated known microbial habitats (Eckburg et al., 2005; Ley et al., 2006; Vitali et al., 2010). This microbial ecosystem serves numerous important functions for the human host, including protection against pathogens, nutrient processing, stimulation of angiogenesis, modulation of intestinal immune response, and regulation of host fat storage (Palmer et al., 2007). The composition of the adult gastrointestinal (GI) microbiota has been intensely studied, using both classical microbiology cultivation and, more recently, culture-independent, small subunit (SSU) ribosomal DNA (rDNA) sequence-based methods (Palmer et al., 2006, 2007; Huse et al., 2008; Vitali et al., 2010).

The prevention and treatment of gut-related diseases will strongly depend on understanding the mechanisms involved in the complex processes of digestion. However, the interactions between the diet, microbiota and host are largely unknown (Quigley, 2011; Del Chierico et al., 2012; Nicholson et al., 2012; Payne et al., 2012).

This review is addressed to evaluate the potential of different analytical approaches to identify and detect the molecules present in body fluids, such as blood, urine, feces, which can be considered originated by food digestion, diet/gut microbiota, and host/microbiota interactions as well as disease markers.

## DIET EFFECT ON HUMAN GUT METABOLOME

Foods contain thousands of compounds which, upon digestion and metabolism, give rise to complex physiological reactions and possible changes in the intestinal microbiota profile resulting in a plethora of metabolites present in body fluids such as blood, urine, feces, and saliva. The number and diversity of chemical compounds in foods is amazing. It has been estimated that an omnivore diet exposes humans to more than 15,000 components, 8000 out of which are non-nutrients such as dietary fiber, antioxidants, prebiotics, and probiotics (Wishart, 2008).

An individual food system such as milk contains more than 200 different oligosaccharides, and the edible plants metabolome consist of more than 10,000 detectable compounds and 800 non-nutrient phytochemicals (Wishart, 2008). These molecules and their accumulation in body fluids, and in particular in urine, can

be regarded as a repository of metabolites in which any nutrient or non-nutrient that is not needed or present in excess is found. Several food-specific biomarkers, e.g., related to wine, black tea, coffee, fruit, and vegetables, have been identified in urine and blood samples (Spencer et al., 2008; Wishart, 2008). Other compounds occurring in different body fluids, namely blood, can be regarded as biomarkers of physiological response to foods. In particular, they are products of lipid peroxidation such as 8-isoprostaglandin F<sub>2</sub>α, and/or oxidative stress markers such as malonaldehyde or glutathione.

Some of the diet components have a potentially direct quantitative and qualitative impact on the microbial species characterizing the gut microbiota. Their metabolic output can harbor physiologically active compounds for the human host which in turn provides a stable environment for proliferation. On the other hand the host has evolved and can use bacterial fermentation products as an energy source for the epithelial cells. Some of the fermentation products, e.g., short chain fatty acids (SCFAs), can positively affect the biochemical and physiological processes at colon level and support some important biological functions (Wong et al., 2006). However, the influence of the bacterial population composition on inter-individual differences in metabolites production and colon health status is poorly understood, except for the well-known relationships between some specific metabolites and metabolic diseases (Le Gall et al., 2011).

Metabolic profiling has a wide potential for understanding the complex interactions between components of the gut microbiota, and elucidate the cause/effects relationships associated with specific nutritional choices and the related shifts in the microbiota composition.

Particularly challenging is the metabolic signatures identification of many phenotypes and their linkage with nutritional choices. In fact, the whole set of metabolites which can be detected in body fluids, and which characterizes the metabolic phenotype, named “metabotype” (Waldram et al., 2009), of an individual, is affected by various intrinsic and extrinsic factors including environment, drugs, diets, lifestyle, and genetics. It should be also considered that the content of non-nutrients in diets is higher than that of nutrients. Such non-nutrients can exert a significant effect on metabolomic profiles thus resulting in several compounds that can be used as biomarkers to trace the food origin.

## HOST-MICROBIOME METABOLIC INTERACTIONS AND CELL-CELL COMMUNICATION

Mammalian–microbial symbiosis can play a role in the etiology and development of several diseases, e.g., insulin resistance, Crohn’s disease (Marchesi et al., 2007; Kinross et al., 2011), irritable bowel syndrome (IBS) (Martin et al., 2006; Sartor, 2008), food allergies, gastritis and peptic ulcers, obesity, cardiovascular disease, and GI cancers (Martin et al., 2011). Activities of gut microbiota can be highly specific, and it has been reported that the establishment of Bifidobacteria is important for the development of the immune system and management of gut functions (Ouweland, 2007). As the microbiome strongly interacts with the host to determine the metabotype,

which influences outcomes of drug interventions, the knowledge of these interactions can provide personalized healthcare solutions (Martin et al., 2011). Also the diet has a key role in the gut microbiota modulation and shaping and, as a consequence, the different foods or their ingredients play a crucial role in the microbes selection and in a metabolic signaling network construction.

The host and its gut microbiota coproduce a large array of small molecules during the conversion of food and xenobiotics (compounds of non-host origin that enter the gut with the diet or are produced by microbiota), many of which play critical roles in shuttling information between host cells and the host’s microbial symbionts. This chemical dialog includes signaling via low molecular weight metabolites, peptides, and proteins or may take place indirectly through immune-mediated pathways (Nicholson et al., 2012). The metabolites production by microbes may influence host health status and their detection in different body fluids can be considered as dysbiosis or disease biomarkers (Holmes et al., 2011; Nicholson et al., 2012) (Figure 1).

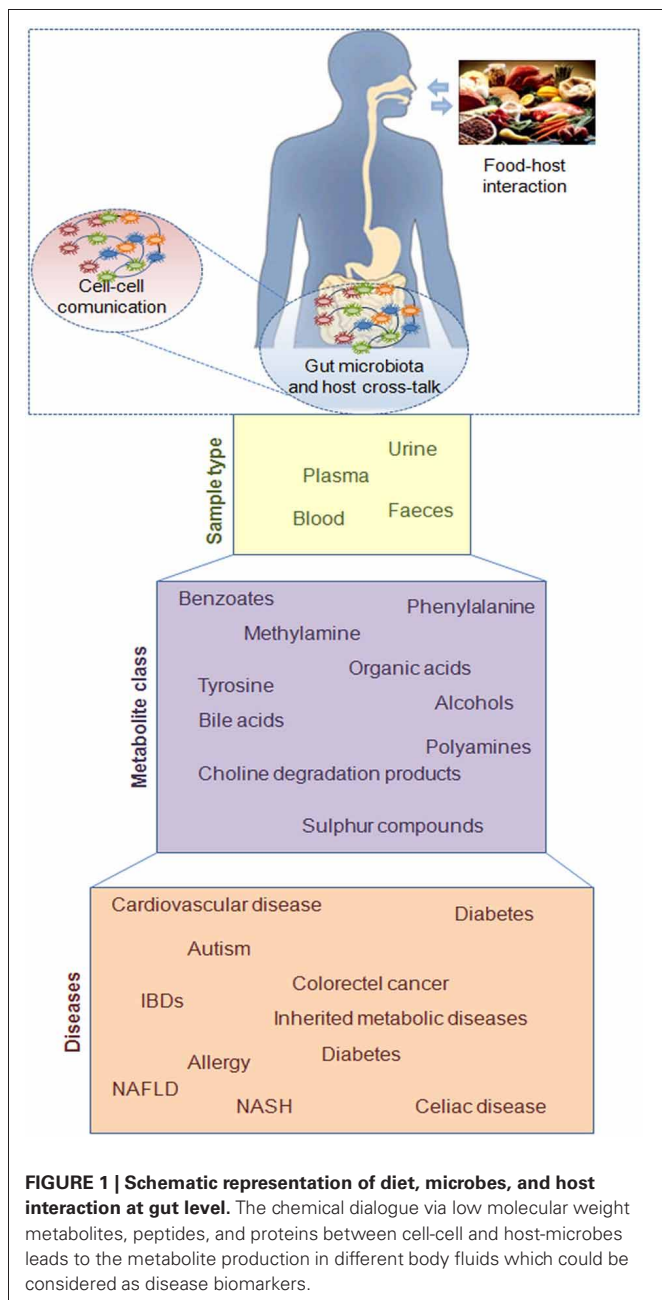
Different regions of the human GI tract vary in terms of composition of the indigenous microbiota (Gordon, 2012). For each compartment of the GI tract, a chemical dialog exists among different microbial species, e.g., direct substrate provision in the microbial food web, quorum sensing, contact-dependent signaling, and potentially gastro-transmitters (Ryan et al., 2009) as well as between microbial symbionts and host cells (Li et al., 2008).

Metabolites are the products and by-products of the many intricate biosynthetic and catabolic pathways existing in all living systems. Biological systems that exist at steady state maintain approximately constant concentrations of metabolic intermediates (Martin et al., 2011).

Fecal extracts are the most common material which can help to decipher the complex metabolic interplay between mammals and their intestinal ecosystems, and metabolic profiles-related to SCFAs, amino acids, organic acids, nucleosides and nucleotides, polyamines, phenolic compounds, bile acids, and lipid components can yield information on a range of gut diseases (Martin et al., 2011).

As described by Grider and Piliand (2007), SCFAs stimulate intestinal transit, and at physiological concentrations they induce an 8–10-fold increase in serotonin release in an *in vitro* colonic mucosal system. Moreover, Musso et al. (2011) showed that SCFAs are clearly one of the most important gut microbiota products and affect a range of host processes including energy utilization, host-microbe signaling, and control of colonic pH with consequent effects on microbiota composition, gut motility, and epithelial cells proliferation.

The monitoring of the fecal metabolome also may unravel diagnostic information for inflammatory bowel diseases (IBDs), including Crohn’s disease and ulcerative colitis (UC), Hirschsprung’s disease, celiac disease, allergy etc (Martin et al., 2011). Metabolomics has many potential applications. It supports functional genomics studies, systems biology, pharmacology, toxicology, and nutrigenomics. An approach called discovery metabolite profiling (DMP) is used to analyze all metabolites generated by a particular enzyme, providing a link between proteome and metabolome.



## ANALYTICAL TECHNIQUES TO DETECT METABOLITES

Metabonomics represents a well-recognized metabolic system approach that involves the study of multivariate metabolic responses of complex cellular organisms to different stimuli (Nicholson et al., 1999).

Ideally, it could be based on this sequence: samples collection (e.g., urine, blood, or some other body fluids), scanning them in a machine and finding a profile of tens or hundreds of chemicals that can predict whether an individual is on the road to a disease, or likely to experience side-effects from a particular drug (Pearson, 2007). However, this vision presents some drawbacks. The first complication

is that one person's profile of metabolites is likely to be dramatically different from another's, and each may fluctuate markedly depending on different aspects, for example the lifestyle, the diseases, the diet, the nutrition etc (Pearson, 2007).

The detection and identification of hundreds metabolites can offer deep insights on the influence of lifestyle and dietary factors in relation to specific diseases. The recent rapid development of a range of analytical platforms, including gas chromatography (GC), liquid chromatography (LC), high pressure LC (HPLC), ultra pressure LC (UPLC) coupled to mass spectrometry (MS), capillary electrophoresis (CE) coupled to MS, Fourier Transform Infrared (FTIR) spectroscopy, and nuclear magnetic resonance (NMR) [e.g., proton ( $^1\text{H}$ )-NMR] spectroscopy, can enable to separate, detect, characterize, and quantify such metabolites and related metabolic pathways (Zhang et al., 2011). Metabolomics focuses on the complex interactions of system components and highlights the whole system rather than the individual parts, providing a distinct perspective on cellular homeostasis (Liu et al., 2010). Even if nowadays NMR, GC-MS, LC-MS are the prevalent techniques used, none of them is a perfect technique that can meet the requirements of metabolomics for measuring all metabolites. MS-based metabolomics offers high selectivity and sensitivity for the identification and quantification of metabolites, and its combination with advanced and high-throughput separation techniques can reduce the complexity of metabolite separation (Zhang et al., 2011) (**Table 1**).

NMR-based metabolomics is able to provide a "holistic view" of the metabolites under certain conditions, and thus is well-suited and advantageous for metabolomic studies (Wu et al., 2010). In particular, it is becoming a useful tool in the study of body fluids and for a non-invasive detection of metabolites and into diagnosis of the diet effects on gut microbiota or significant public health problems (Martin et al., 2012).

Usually, the measurement of the gut microbial metabolism is confined to fecal samples, which is typically limited because of the elevated colonic absorption of bacterial metabolites. The Proton Nuclear Magnetic Resonance ( $^1\text{H}$ -NMR) monitoring of the human fecal metabolome also may unravel diagnostic information for IBD including UC (Marchesi et al., 2007). Saric et al. (2008) evaluated the similarity and dissimilarity across different mammals namely humans, mice, and rats by using  $^1\text{H}$ -NMR analysis. The authors reported how human fecal extracts showed greater inter-individual variation than in rodents, reflecting the natural genetic and environmental diversity in human population.

A major source of intestinal metabolites is both host and microbial processing of dietary nutrients. As reported by Martin et al. (2010) the  $^1\text{H}$ -NMR analysis of feces revealed that the supplementation with probiotics significantly affects the host microbiota interaction. Probiotic supplementation of "humanized" mice, inoculated with a model of human baby microbiota, was associated with metabolic changes in the protein metabolism of *Lactobacillus paracasei* in particular with specific aminoacidic pattern.

**Table 1 | Common analytical techniques used in metabolomics.**

Analytical method	Advantages	Disadvantages	Comments
NMR	<ul style="list-style-type: none"> <li>• Rapid analysis</li> <li>• High resolution</li> <li>• No derivatization method</li> <li>• Non-destructive</li> </ul>	<ul style="list-style-type: none"> <li>• Low sensitivity</li> <li>• Convoluted spectra</li> <li>• More than one peak per component</li> <li>• Libraries of limited use due to complex matrix</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Chemical consideration:</b> gives detailed structural information, particularly using 2-D-NMR of isolated metabolites</li> <li>• <b>Chemical bias:</b> these methods have little chemical bias and can be used directly on the sample</li> <li>• <b>Speed:</b> few minutes to hours. Depends on the strength of the magnet, sensitivity can be improved by magic angle spinning</li> </ul>
GC-MS	<ul style="list-style-type: none"> <li>• Sensitive</li> <li>• Robust</li> <li>• Large linear range</li> <li>• Large commercial and public libraries</li> </ul>	<ul style="list-style-type: none"> <li>• Slow</li> <li>• Often requires derivatization</li> <li>• Many analytes thermally-unstable or too large for analysis</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Chemical consideration:</b> on its own will not generally lead to metabolite identification. However, coupled with MS and NMR is very powerful for analyte identification</li> <li>• <b>Chemical bias:</b> solvent extraction bias: non-polar vs. polar analytes. Need for chemical derivatization</li> <li>• <b>Speed:</b> very useful for separation, but typically take 10–30 min</li> </ul>
LC-MS	<ul style="list-style-type: none"> <li>• No derivatization required (usually)</li> <li>• Many modes of separation available</li> <li>• Large sample capacity</li> </ul>	<ul style="list-style-type: none"> <li>• Slow</li> <li>• Limited commercial libraries</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Chemical consideration:</b> on its own will not generally lead to metabolite identification. However, coupled with MS and NMR is very powerful for analyte identification</li> <li>• <b>Chemical bias:</b> solvent bias means it is usually more applicable to polar compounds</li> <li>• <b>Speed:</b> very useful for separation, but typically take 10–30 min</li> </ul>
FT-IR	<ul style="list-style-type: none"> <li>• Rapid analysis</li> <li>• Complete fingerprint of sample chemical composition</li> <li>• No derivatization needed</li> </ul>	<ul style="list-style-type: none"> <li>• Extremely convoluted spectra</li> <li>• More than one peak per component</li> <li>• Metabolite identification nearly impossible</li> <li>• Requires samples drying</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Chemical consideration:</b> provide limited structural information, but useful for identification of functional groups</li> <li>• <b>Chemical bias:</b> these methods have little chemical bias and can be used directly on the sample</li> <li>• <b>Speed:</b> 10–60 s</li> </ul>

Dumas et al. (2006) have recently applied  $^1\text{H-NMR}$  technique to characterize the intergenome interactions in mice with synbiotic gut microbiota as well as to monitor the gut-microbial metabolite variation in rats and to study the intricate relationships between gut microbiota and host co-metabotype associated with dietary-induced changes. Ndagijimana et al. (2009) studied, by means of the  $^1\text{H-NMR}$  analysis of healthy human subjects faeces, the effect of the supplementation with a synbiotic food based on *Lactobacillus acidophilus*, *Bifidobacterium longum*, and Fructooligosaccharides. These authors reported that the number and the extent of metabolites in faecal slurries were strongly affected by the synbiotic food consumption and gave rise to characteristic metabolic signatures. Reproducibility of  $^1\text{H-NMR}$  metabolic profiles generated from water and methanol extracts

from human stools were assessed by Jacobs et al. (2008). On the other hand  $^1\text{H-NMR}$  is one of the preferred platforms also for urine and plasma analysis (Ala-Korpela, 2008). The first published study, in which a metabolomics approach on urine was described, used the  $^1\text{H-NMR}$  technique to monitor the effect of the inclusion of soy in the diet (Solanky et al., 2003). Urine samples have also been used to investigate responses to ingestion of chamomile tea, or other foods such as coffee, wine, and tea, evidencing that hippurate and glycine are important discriminatory metabolites (Ito et al., 2005; Wang et al., 2005).

The ability to predict the occurrence of exercise-induced ischemia in patients with suspected cardiovascular disease was investigated by  $^1\text{H-NMR}$  blood analysis. Barba et al. (2008) demonstrated that lactate, glucose, lipids, and long-CFAs are



the main metabolites involved. Xanthine and ascorbate were proposed as possible markers of plaque formation in an atherosclerotic mouse model (Leo and Darrow, 2009) and lipoprotein subclasses can now be analyzed by a commercial  $^1\text{H-NMR}$ -based protocol (Jeyarajah et al., 2006; Ala-Korpela, 2007).

Even if the information provided using  $^1\text{H-NMR}$  is highly valuable, it is still limited due to the low resolution and sensitivity which enables the annotation and quantification of only a limited number of low molecular weight molecules (Jansson et al., 2009).

GC-MS-based metabolomics requires a high-throughput technology to handle a large volume of samples and accurate peak identification through the standard retention times and mass spectra. GC-MS has been widely used for metabolomics and can provide efficient and reproducible analysis (Zhang et al., 2011). In fact, it is also possible to obtain, simultaneously profiles of several hundred compounds including organic acids, most amino acids, sugars, sugar alcohols, aromatic amines, and FAs. GC-MS Solid-Phase Microextraction (SPME) based analysis can be considered as a very effective method for rapidly qualitatively and quantitatively analyze faecal samples most of which have not been previously reported (Garner et al., 2007). In particular, GC-MS/SPME represents a novel method to study metabolic profiles of biological samples. This approach has been used to compare neonates and adult feces (De Lacy Costello et al., 2008) and to identify volatile markers of GI diseases (Garner et al., 2007). The main metabolites identified by GC-MS/SPME belong to sulphur compounds, nitrogen compounds such as pyridine, and its derivatives, pyrazines, indoles, aldehydes, ketones, esters, alcohols, phenols, organic acids, hydrocarbons, and stress molecules such as furans and furanones. Vitali et al. (2010) investigated the impact of a synbiotic food on a human gut microbial ecology and metabolic profiles by using this technique. While no significant changes in the structure of the gut microbiota of healthy subjects was observed, the synbiotic food intake generated significant changes in some gut metabolic activities. The Canonical discriminant Analysis of Principal coordinates (CAP) of the fecal metabolic profiles showed a separation of subjects depending on synbiotic food intake. Recently, the GC-MS/SPME analysis of human feces has been also proposed as a tool to evaluate the *in vitro* effect of prebiotics and probiotics on the human microbiota (Vitali et al., 2012).

Zheng et al. (2011) applied an untargeted GC-MS-based metabolomics approach to profile bacterial metabolites in normal Wistar rats administrated with a broad spectrum  $\beta$ -lactam antibiotic imipenem/cilastatin sodium. In depth, metabolic phenotyping allowed the identification of 202 urinary and 223 fecal metabolites, many of which not previously reported (e.g., oligopeptides and carbohydrates), significantly related to a functional metagenome (Zheng et al., 2011). Moreover, Maccaferri et al. (2010) investigated the impact of rifaximin administration on microbial metabolic profiles by using GC-MS/SPME. At the same time this technique was also used in combination with the  $^1\text{H-NMR}$  technique to investigate the metabolome of 19 celiac disease children under gluten-free diet (treated celiac disease, T-CD) and 15 healthy children (HC). The metabolome of

T-CD and HC children was studied using fecal and urine samples. With this approach the authors showed that the levels of volatile organic compounds and free amino acids in fecal and/or urine samples were markedly due to affected by CD (Di Cagno et al., 2011). Francavilla et al. (2012) used GC-MS/SPME to study the influence of lactose on the composition of the gut microbiota and metabolome of infants presenting with cow's milk allergy.

Moreover, MS and HPLC techniques are commonly used for compounds' characterization at structural level. In the field of metabolomics, both MS and HPLC are often combined to characterize unknown endogenous or exogenous from a complex biological matrix. LC is probably the most versatile separation method, as it allows separation of compounds of a wide range of polarity with little effort in sample preparation (compared to GC-MS) (Moco et al., 2007). Indeed, large-scale metabolomic technologies based on LC-MS are increasingly gaining attention for their use in human disease diagnosis (Courant et al., 2009).

In addition, GC-MS and LC-MS have been integrated to provide the comprehensive metabolic signature of the malnutrition in rat model and to discover differential metabolites (Wu et al., 2010).

Recently, the combination of UPLC with MS has covered a large number of polar metabolites enlarging the number of detectable analytes.

Like NMR, vibrational spectroscopies such as RAMAN and FTIR are comparable in sensibility, but the latter allows high-throughput screening and biological samples classification, and equally fits the "omics philosophy" of providing unbiased, whole-system measurements (Kell, 2004). In particular, the use of FTIR spectroscopy to monitor biochemical changes in living cells has gained considerable importance in the last decade. In fact, this technique presents the possibility to simultaneously identify various cellular biochemical targets, both *in vivo* and *in vitro* conditions, exploiting the differential infrared radiation absorption of each metabolites at specific wave number (Salman et al., 2004). Molecules and systems of biological relevance that can be detected by FTIR spectroscopy including molecules such as lipids and fatty acids, proteins, peptides, carbohydrates, nucleic acids as well as biomembranes, animal tissues, microbial cells plants, and clinical samples (Dole et al., 2011). This technique has been employed to characterize isolated biological molecules, particularly proteins, and lipids. However, more recently it has been used, with the aid of sophisticated sampling techniques such as infrared imaging, in the diagnosis of many diseases such as cervical cancer, Parkinson's disease, Alzheimer's disease, kidney stone and arthritis (Dole et al., 2011). In fact, FTIR spectroscopy can effectively provide chemical variation information about the structure and the composition of biological material at molecular level. Li et al. (2012) reported that colitis and colon cancer can be successfully identified and discern by analyzing colon biopsies through FTIR spectroscopy and chemometrics. Also, Argov et al. (2004) used FTIR microspectroscopy to distinguish IBDs from colitis-associated colon carcinomas, when pathological symptoms are similar. In particular, the study of differences in specific regions of the infrared spectra allowed the identification of the discriminating molecules, e.g., the phosphate

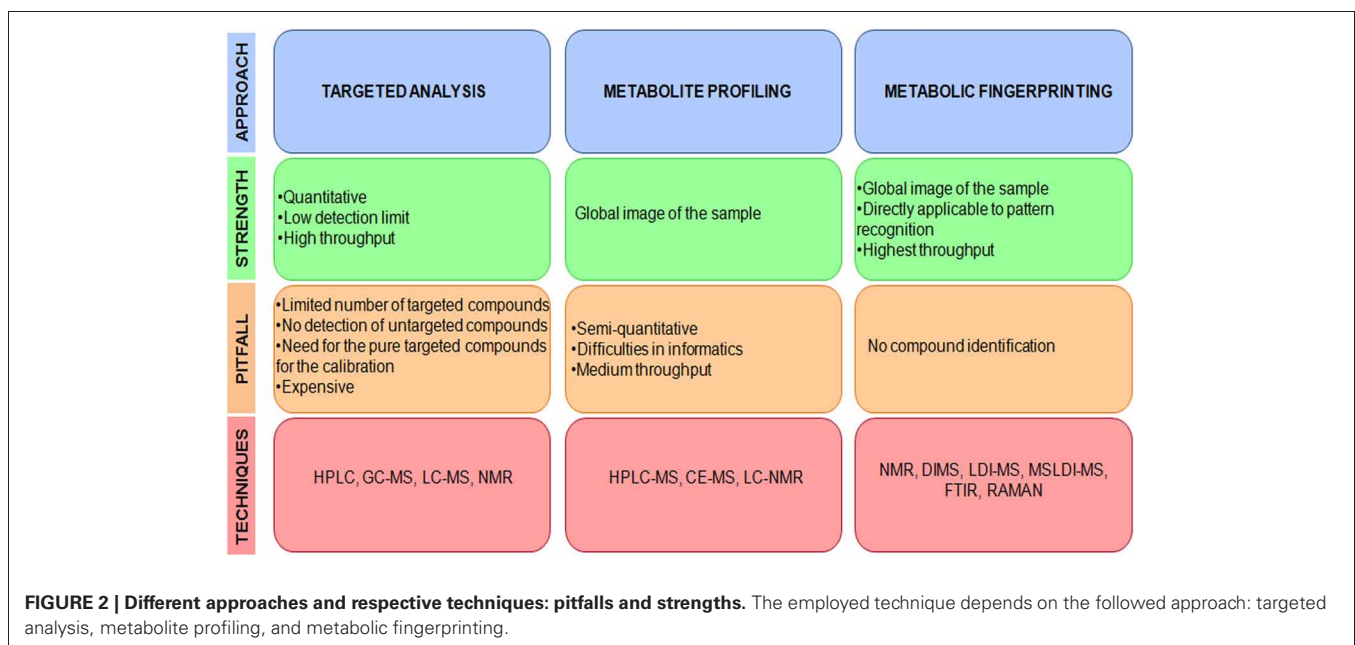
content and RNA/DNA ratio, as different in IBD, cancer, or normal tissues. Furthermore, FTIR spectroscopy has also been used by Bright et al. (2011) to determine the efficacy of plasma homocysteine levels on vitamin supplementation in diet in hyperlipidemic patients. Human plasma samples from healthy and chronic lymphocytic leukemia patients have been analyzed by FTIR spectroscopy by Erukhimovitch et al. (2006), leading to identify of specific spectral peaks as associated with biomarkers of the disease. Cluster analysis of the selected spectra provided excellent classifications correlated completely with clinical data, showing, not with standing. Although these results are preliminary, but promising, rapid, effective, and economic technique which can assist in the disease diagnosis.

However, the technique selection depends on the kind of followed approach: (1) HPLC, GC-MS, LC-MS are employed in metabolite target analysis, e.g., determination and quantification of known metabolites or products of particular biochemical pathways; (2) HPLC-MS, CE-MS, LC-NMR, are used in metabolite profiling studies, e.g., larger, defined set of compounds survey and lipid analysis; (3) NMR, Direct Infusion electrospray Ionization-MS (DIMS), Laser Desorption Ionization-MS (LDI-MS), Matrix Suppressed LDI-MS (MSLDI-MS), FTIR, and Raman spectroscopy are used in metabolic fingerprinting, e.g., generation and comparison of sample metabolic profiles to identify differences (Shulaev, 2006) (Figure 2).

Furthermore, the chromatography-MS systems are considered the most favorable for the detection of large numbers of metabolites in metabolomics coupling the chromatographic metabolite separation with the sensitivity of MS detection. GC-Time of Flight (TOF)-MS, two dimensional GC coupled to TOF MS (GCxGC-TOF-MS), HPLC-MS, the analytically superior UPLC-MS and CE-MS have all been employed in mammalian metabolomic studies, either in a metabolic profiling or targeted

analysis approach (Dunn et al., 2008). The GC-MS or GC-MS/SPME have been widely used in metabolomic and metabolomic studies, especially the second approach is useful because it does not require a difficult or strongly pre-analytical work flow. Garner et al. (2009) demonstrated the possibility to discriminate with GC-MS/SPME between the volatile organic compounds profile in fecal samples from preterm infants developing necrotizing enterocolitis (NEC) compared with non-NEC controls. Recent studies have suggested that the gut microbiota is involved in numerous important biochemical functions for the host, in healthy and pathological conditions (Del Chierico et al., 2012).

The extraction of valuable conclusions from the analysis of metabolomic data is important to perform the analytical measurements; in fact, there is a variety of methods that allow the instrument raw data transformation analyzed by the use of different software which provide a list of metabolites. Some parameters, such as biological variation present among individuals, sampling, sample preparation, and analytical measurement, influence the reproducibility of results, and these should be monitored as much as possible by measuring replicates, both analytical and biological. In principle, biological variance should surpass all analytical variance (Moco et al., 2007). Retention-time shifts are common in GC and more severely, LC, but only occasionally in NMR and FTIR spectra. In NMR spectra, non-reproducibility seems to be strictly related to sample preparation and hardly ever due to instrumental incoherence. Nevertheless, even in strictly controlled conditions, signal shifts may persist. For this reason, the use of signal-alignment software [e.g., MetAlign (De Vos et al., 2007), XCMS (Smith et al., 2006), and MZmine (Katajamaa et al., 2006)] has become a routine procedure for comparing chromatograms or spectra in MS applications, while HiRes is suitable in NMR (Zhao et al., 2006), transforming raw data into workable informative datasets.



## HOW TO MANAGE AND INTEGRATE ALL DATASETS?

It is difficult to correlate specific markers with risk for disease, nutritional state or diet choices. Instead, by examining the system as a whole and exploring multiple pathways simultaneously, a greater indication of a healthy or pathological state can be acquired. Hence, this new approach, characterized by a high-throughput determination of hundreds of metabolites, leads to a torrent of data, in fact, over the years, there have been constant upgrade in the hardware and the software of these technologies to meet the demands for robustness, practicality, applicability, and efficiency of the analyses. Therefore, powerful analytical strategies in combination with advanced multivariate statistical tools are required to have a look at this new “black box” and to extract maximum relevant information and knowledge about the complex biological system (Bictash et al., 2010).

The use of single or combined analytical techniques or different kind of samples inevitably leads to face the big problem of how to manage and represent the thousands of data collected.

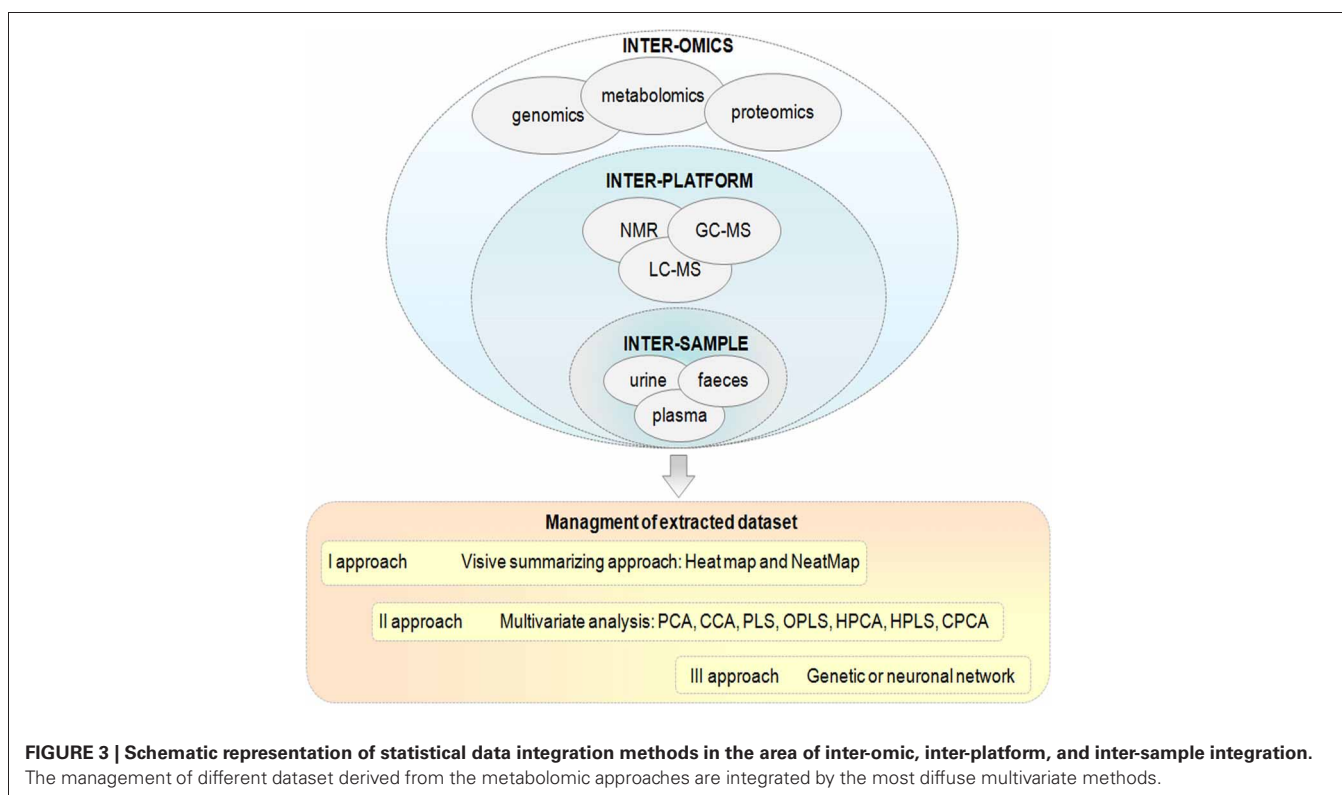
One of the expectations of the system biologist is that these datasets can be integrated to give a holistic picture of the state of the system, e.g., development, ageing, health, or disease, which provides insights enable a more biology fundamental understanding via unveiling network connections at molecular level (Richards et al., 2010).

Appropriate experimental design, sample numbers and statistical analyses are required to ensure generation of accurate and valid hypotheses and biological conclusions (Tseng and Wong, 2005). Powerful multivariate analyses, either supervised or unsupervised, are required to interrogate the data and define

structure-related to biological system similarities or differences (Thalamuthu et al., 2006).

The first step for simplification of all data, but not in an arbitrary way, could be the use of Heat maps, where the relative intensity value of one peak is replaced by a small colored block (Zhou et al., 2011). Initially, it has been developed for microarray studies, but it is already used in some metabolomic works (Spitale et al., 2012). Rajaram and Oono (2010) proposed a new analysis, the so-called NeatMap as an alternative to the traditional Heat map, this offers a variety of novel plots (in 2 and 3 dimensions) to be used in conjunction with principal component analysis (PCA) and multi-dimensional scaling (MDS). Although, the NeatMap has been used so far only in genomics, it could be exploited in metabolomics as well. The second step could be the use of multivariate statistical analysis methods that can interpret different datasets. According to Richards et al. (2010) there are three possible integration strategies: (1) inter-omic, or the integration of data obtained from different -omic platforms (metabonomics, genomics, proteomics, and transcriptomics); (2) inter-platform, or the integration of data from different spectroscopic platforms (NMR, GC-MS, LC-MS etc.); (3) inter-samples, or the integration of data obtained from different human samples (plasma, urine, tissue, and faeces) (**Figure 3**).

The most diffused multivariate methods are the PCA and the Partial Least Squares (PLS) (Holmes et al., 2011). In particular, PCA is widely used for multivariate of NMR or GC-MS analysis profiling data, it is based on modeling the natural variance within a dataset, which held to identify the underlying metabolite variables that contribute to that variance (Biais et al., 2009). PLS and



its derivative orthogonal PLS (OPLS) have been also used for NMR profiling data. In comparison with PLS, OPLS produces models which are more clear and therefore easier to interpret, leading to a better class-resolution in a discriminant problem (Stella et al., 2006). However, when PCA and PLS are used for the interpretation of several different, but potentially connected, datasets (called “blocks”), the loading plots are usually complex due to the co-variation in the spectrum, and therefore difficult to correlate to the corresponding score plot (Janné et al., 2001). For this reason a multiblock technique must be used. The common trend of the different blocks is revealed in the “super scores” plot, where the distribution of the samples of each individual block are shown in their respective “block scores,” and similarly to classical PCA, the contribution of variables to the trend shown in the blocks scores plot is shown in their “block loadings” plot (Biais et al., 2009).

The hierarchical multiblock segmentation techniques (HPCA, HPLS) are based on new variables created from the original data by blocking the spectra into sub-spectra, and then projecting the sub-spectra by PCA. These new variables are then used in the coming PCA or PLS calculations, reducing the random and non-wanted signals from e.g., light scatter, but still conserving all systematic information in the signals. This technique gives the greatest advantage of easier interpretation of the correlation between scores and loadings (Janné et al., 2001).

Another kind of multiblock PCA used is the Consensus PCA (CPCA), it can focus the data analysis on the relation between the specified metabolites and the remaining metabolites, searches for trends that explain as much as of the variation as possible (Biais

et al., 2009). CPCA was introduced as a method for comparing several blocks of descriptor variables measured on the same objects and its difference with respect to HPCA is in the data normalization (Westerhuis et al., 1998).

Also, the multivariate analysis, such as the Canonical Correlation Analysis (CCA) and the regularized CCA are widely used to integrate datasets obtained from different source materials (Yamamoto et al., 2008). However, when the “normal” variance is greater than that explained by the process of interest, the most sophisticated genetic (Johnson et al., 2001) or neural network (Ahmed et al., 2009) based algorithms, may identify the underlying variables of significance which are not associated with the greatest variance (the “normal” condition). In these cases, such algorithm-based models should not be overfit (Biais et al., 2009).

As the holistic approach to biological systems comprehension continually evolves, the techniques used to analyze and integrate the datasets must be steadily updated to manage the problems occurring during techniques application. Hence, system biologists have a key role in the interpretation of the meaning of the results obtained by the datasets management.

## ACKNOWLEDGMENTS

The authors declare absence of potential conflicts. The authors are especially grateful to Dr. Lorenza Putignani for critical revision and discussion of the manuscript. This work was supported by European Fund for Regional Development (Emilia-Romagna-Research Centre for Industrial Research-CIRIAGrifood) 2007–2013 (OP-ERDF) to Pamela Vernocchi.

## REFERENCES

- Ahmed, S. S., Santosh, W., Kumar, S., and Christlet, H. T. (2009). Metabolic profiling of Parkinson's disease: evidence of biomarker from gene expression analysis and rapid neural network detection. *J. Biomed. Sci.* 16, 63.
- Ala-Korpela, M. (2007). Potential role of body fluid  $^1\text{H}$  NMR metabolomics as a prognostic and diagnostic tool. *Expert Rev. Mol. Diagn.* 7, 761–773.
- Ala-Korpela, M. (2008). Critical evaluation of  $^1\text{H}$  NMR metabolomics of serum as a methodology for disease risk assessment and diagnostics. *Clin. Chem. Lab. Med.* 46, 27–42.
- Argov, S., Sahu, R. K., Bernshtain, E., Salman, A., Shohat, G., Zelig, U., et al. (2004). Inflammatory bowel diseases as an intermediate stage between normal and cancer: a FTIR-microspectroscopy approach. *Biopolymers* 75, 384–392.
- Barba, I., de León, G., Martín, E., Cuevas, A., Aguade, S., Candell-Riera, J., et al. (2008). Nuclear magnetic resonance-based metabolomics predicts exercise-induced ischemia in patients with suspected coronary artery disease. *Magn. Reson. Med.* 60, 27–32.
- Biais, B., Allwood, J. W., Deborde, C., Xu, Y., Maucourt, M., Beauvoit, B., et al. (2009).  $^1\text{H}$  NMR, GC-EI-TOFMS, and data set correlation for fruit metabolomics: application to spatial metabolite analysis in melon. *Anal. Chem.* 81, 2884–2894.
- Bictash, M., Ebbels, T. M., Chan, Q., Loo, R. L., Yap, I. K., Brown, I. J., et al. (2010). Opening up the “Black Box”: metabolic phenotyping and metabolome-wide association studies in epidemiology. *J. Clin. Epidemiol.* 63, 970–979.
- Bright, A., Renuga Devi, T. S., and Gunasekaran, S. (2011). Efficacy of vitamin supplementation on plasma homocysteine levels among hyperlipidemic patients – a spectral and clinical analysis. *Res. J. Pharm. Biol. Chem. Sci.* 2, 715–720.
- Courant, F., Pinel, G., Bichon, E., Monteau, F., Antignac, J. P., and Bizec, B. (2009). Development of a metabolomic approach based on liquid chromatography-high resolution mass spectrometry to screen for clenbuterol abuse in calves. *Analyst* 134, 1637–1646.
- De Lacy Costello, B., Ewen, R., Ewer, A. K., Garner, C. E., Probert, C. S., Ratcliffe, N. M., et al. (2008). An analysis of volatiles in the headspace of the faeces of neonates. *J. Breath Res.* 2, 037023.
- Del Chierico, F., Vernocchi, P., Bonizzi, L., Carsetti, R., Castellazzi, A. M., Dallapiccola, B., et al. (2012). Early-life gut microbiota under physiological and pathological conditions: the central role of combined meta-omics-based approaches. *J. Proteomics* 75, 4580–4587.
- De Vos, R. C., Moco, S., Lommen, A., Keurentjes, J. J., Bino, R. J., and Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2, 778–791.
- Di Cagno, R., De Angelis, M., De Pasquale, I., Ndagijimana, M., Vernocchi, P., Ricciuti, P., et al. (2011). Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization. *BMC Microbiol.* 11:219. doi: 10.1186/1471-2180-11-219
- Dole, M. N., Patel, P. A., Sawant, S. D., and Shedpure, P. S. (2011). Advance applications of Fourier transform Infrared Spectroscopy. *Int. J. Pharm. Sci. Rev. Res.* 7, 159–166.
- Dumas, M. E., Barton, R. H., Toye, A., Cloarec, O., Blancher, C., Rothwell, A., et al. (2006). Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12511–12516.
- Dunn, W. B., Broadhurst, D., Ellis, D. I., Brown, M., Halsall, A., O'Hagan, S., et al. (2008). A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. *Int. J. Epidemiol.* 37, i23–i30.
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638.
- Erukhimovitch, V., Talyshinsky, M., Souprun, Y., and Huleihel, M.



- (2006). FTIR spectroscopy examination of leukemia patients plasma. *Vib. Spectrosc.* 40, 40–46.
- Francavilla, R., Calasso, M., Calace, L., Siragusa, S., Ndagijimana, M., Vernocchi, P., et al. (2012). Effect of lactose on gut microbiota and metabolome of infants with cow's milk allergy. *Pediatr. Allergy Immunol.* 23, 420–427.
- Garner, C. E., Ewer, A. K., Elaso, K., Power, F., Greenwood, R., Ratcliffe, N. M., et al. (2009). Analysis of faecal volatile organic compounds in preterm infants who develop necrotising enterocolitis: a pilot study. *J. Pediatr. Gastroenterol. Nutr.* 49, 559–565.
- Garner, C. E., Smith, S., de Lacy Costello, B., White, P., Spencer, R., Probert, C. S., et al. (2007). Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB J.* 21, 1675–1688.
- Gordon, J. I. (2012). Honor thy gut symbionts redux. *Science* 336, 1251–1253.
- Grider, J. R., and Piland, B. E. (2007). The peristaltic reflex induced by short-chain fatty acids is mediated by sequential release of 5-HT and neuronal CGRP but not BDNF. *Am. J. Physiol. Gastrointest. Liver Physiol.* 292, 429–437.
- Holmes, E., Li, J. V., Athanasiou, T., Ashrafian, H., and Nicholson, J. K. (2011). Understanding the role of gut microbiome–host metabolic signal disruption in health and disease. *Trends Microbiol.* 19, 349–359.
- Huse, S. M., Dethlefsen, L., Huber, J. A., Mark Welch, D., Relman, D. A., and Sogin, M. L. (2008). Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* 4:e1000255. doi: 10.1371/journal.pgen.1000255. Erratum in: *PLoS Genet.* 4. doi: 10.1371/annotation/3d8a6578-ce56-45aa-bc71-05078355b851. Welch, David Mark [corrected to Mark Welch, David].
- Ito, H., Gonthier, M. P., Manach, C., Morand, C., Mennen, L., Révész, C., et al. (2005). Polyphenol levels in human urine after intake of six different polyphenol-rich beverages. *Br. J. Nutr.* 94, 500–509.
- Jacobs, D. M., Deltimple, N., van Velzen, E., van Dorsten, F. A., Bingham, M., Vaughan, E. E., et al. (2008). <sup>1</sup>H NMR metabolite profiling of feces as a tool to assess the impact of nutrition on the human microbiome. *NMR Biomed.* 21, 615–626.
- Janné, K., Pettersen, J., Lindberg, N. O., and Lundstedt, T. (2001). Hierarchical principal component analysis (PCA) and projection to latent structure (PLS) technique on spectroscopic data as a data pretreatment for calibration. *J. Chemom.* 15, 203–213.
- Jansson, J., Willing, B., Lucio, M., Fekete, A., Dicksved, J., Halfvarson, J., et al. (2009). Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* 4:e6386. doi: 10.1371/journal.pone.0006386
- Jeyarajah, E. J., Cromwell, W. C., and Otvos, J. D. (2006). Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. *Clin. Lab. Med.* 26, 847–870.
- Johnson, H. E., Gilbert, R. J., Winson, M. K., Goodacre, R., Smith, A. R., Rowland, J. J., et al. (2001). New food components and gastrointestinal health. *Proc. Nutr. Soc.* 60, 481–488.
- Katajamaa, M., Miettinen, J., and Oresic, M. (2006). MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22, 634–636.
- Kell, D. B. (2004). Metabolomics and system biology: making sense of the soup. *Curr. Opin. Microbiol.* 7, 296–307.
- Kinross, J. M., Darzi, A. W., and Nicholson, J. K. (2011). Gut microbiome–host interactions in health and disease. *Genome Med.* 3, 14.
- Le Gall, G., Noor, S. O., Ridgway, K., Scovell, L., Jamieson, C., Johnson, I. T., et al. (2011). Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. *J. Proteome Res.* 10, 4208–4218.
- Leo, G. C., and Darrow, A. L. (2009). NMR-based metabolomics of urine for the atherosclerotic mouse model using apolipoprotein-E deficient mice. *Magn. Reson. Chem.* 47, S20–S25.
- Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848.
- Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., et al. (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2117–2122.
- Li, X., Li, Q.-B., Zhang, G.-J., Xu, Y.-Z., Sun, X.-J., Shi, J.-S., et al. (2012). Identification of colitis and cancer in colon biopsies by Fourier transform infrared spectroscopy and chemometrics. *Sci. World J.* 2012, 936149.
- Liu, J. Y., Li, N., Yang, J., Li, N., Qiu, H., Ai, D., et al. (2010). Metabolic profiling of murine plasma reveals an unexpected biomarker in rofecoxib-mediated cardiovascular events. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17017–17022.
- Maccaferri, S., Vitali, B., Klinder, A., Kolida, S., Ndagijimana, M., Laghi, L., et al. (2010). Rifaximin modulates the colonic microbiota of patients with Crohn's disease: an *in vitro* approach using a continuous culture colonic model system. *J. Antimicrob. Chemother.* 65, 2556–2565.
- Marchesi, J. R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F., et al. (2007). Rapid and non-invasive metabolomic characterization of inflammatory bowel disease. *J. Proteome Res.* 6, 546–551.
- Martin, F. P., Collino, S., and Rezzi, S. (2011). <sup>1</sup>H NMR-based metabolomic applications to decipher gut microbial metabolic influence on mammalian health. *Magn. Reson. Chem.* 49, S47–S54.
- Martin, F. P., Collino, S., Rezzi, S., and Kochhar, S. (2012). Metabolomic applications to decipher gut microbial metabolic influence in health and disease. *Front. Physiol.* 3:113. doi: 10.3389/fphys.2012.00113
- Martin, F. P., Sprenger, N., Montoliu, I., Rezzi, S., Kochhar, S., and Nicholson, J. K. (2010). Dietary modulation of gut functional ecology studied by fecal metabolomics. *J. Proteome Res.* 9, 5284–5295.
- Martin, F. P., Verdu, E. F., Wang, Y., Dumas, M. E., Yap, I. K., Cloarec, O., et al. (2006). Transgenomic metabolic interactions in a mouse disease model: interactions of *Trichinella spiralis* infection with dietary *Lactobacillus paracasei* supplementation. *J. Proteome Res.* 5, 2185–2193.
- Moco, S., Bino, R. J., De Vos, R. C. H., and Vervoort, J. (2007). Metabolomics technologies and metabolite identification. *Trends Anal. Chem.* 26, 855–866.
- Musso, G., Gambino, R., and Cassader, M. (2011). Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annu. Rev. Med.* 62, 361–380.
- Ndagijimana, M., Laghi, L., Vitali, B., Placucci, G., Brigidi, P., and Guerzoni, M. E. (2009). Effect of a symbiotic food consumption on human gut metabolic profiles evaluated by <sup>1</sup>H Nuclear Magnetic Resonance spectroscopy. *Int. J. Food Microbiol.* 134, 147–153.
- Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., et al. (2012). Host-gut microbiota metabolic interactions. *Science* 336, 1262–1267.
- Nicholson, J. K., Lindon, J. C., and Holmes, E. (1999). 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29, 1181–1189.
- Ouweland, A. C. (2007). Antiallergic effects of probiotics. *J. Nutr.* 137, 794S–797S.
- Palmer, C., Bik, E. M., Digiulio, D. B., Relman, D. A., and Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol.* 5:e177. doi: 10.1371/journal.pbio.0050177
- Palmer, C., Bik, E. M., Eisen, M. B., Eckburg, P. B., Sana, T. R., Wolber, P. K., et al. (2006). Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res.* 3, e5.
- Payne, A. N., Chassard, C., and Lacroix, C. (2012). Gut microbial adaptation to dietary consumption of fructose, artificial sweeteners and sugar alcohols: implications for host-microbe interactions contributing to obesity. *Obes. Rev.* 13, 799–809.
- Pearson, H. (2007). Meet the human metabolome. *Nature* 446, 8.
- Quigley, E. M. (2011). Gut microbiota and the role of probiotics in therapy. *Curr. Opin. Pharmacol.* 11, 593–603.
- Rajaram, S., and Oono, Y. (2010). NeatMap – non-clustering heat map alternatives in R. *BMC Bioinformatics* 11:45. doi: 10.1186/1471-2105-11-45
- Richards, S. E., Dumas, M. E., Fonville, J. M., Ebbels, T. M. D., Holmes, E., and Nicholson, J. K. (2010). Intra- and inter-omic fusion of metabolic profiling data in a systems biology framework. *Chemometr. Intell. Lab.* 104, 121–131.
- Ryan, R. P., Romeo, T., De Keersmaecker, S. C., and Coulthurst, S. J. (2009). Nurturing scientific mutualism: a report from the 'Young Microbiologists Mini-Symposium on microbe signalling, organisation and pathogenesis'. *Mol. Microbiol.* 73, 760–774.
- Salman, A., Sahu, R. K., Bernshtain, E., Zelig, U., Goldstein, J., Walfisch, S., et al. (2004). Probing cell proliferation in the human colon using vibrational spectroscopy: a novel use of FTIR-microspectroscopy. *Vib. Spectrosc.* 34, 301–308.

- Saric, J., Wang, Y., Li, J., Coen, M., Utzinger, J., Marchesi, J. R., et al. (2008). Species variation in the fecal metabolome gives insight into differential gastrointestinal function. *J. Proteome Res.* 7, 352–360.
- Sartor, R. B. (2008). Microbial influences in inflammatory bowel diseases. *Gastroenterology* 134, 577–594.
- Shulaev, V. (2006). Metabolomics technology and bioinformatics. *Brief Bioinform.* 7, 128–139.
- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787.
- Solanky, K. S., Bailey, N. J., Beckwith-Hall, B. M., Davis, A., Bingham, S., Holmes, E., et al. (2003). Application of biofluid <sup>1</sup>H nuclear magnetic resonance-based metabolomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. *Anal. Biochem.* 32, 197–204.
- Spencer, J. P., Abd El Mohsen, M. M., Minihane, A. M., and Mathers, J. C. (2008). Biomarkers of the intake of dietary polyphenols: strengths, limitations and application in nutrition research. *Br. J. Nutr.* 99, 12–22.
- Spitale, R. C., Cheng, M. Y., Chun, K. A., Gorell, E. S., Munoz, C. A., Kern, D. G., et al. (2012). Differential effects of dietary supplements on metabolomic profile of smokers versus non-smokers. *Genome Med.* 4, 1–10.
- Stella, C., Beckwith-Hall, B., Cloarec, O., Holmes, E., Lindon, J. C., Powell, J., et al. (2006). Susceptibility of human metabolic phenotypes to dietary modulation. *J. Proteome Res.* 5, 2780–2788.
- Thalamuthu, A., Mukhopadhyay, I., Zheng, X., and Tseng, G. C. (2006). Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics* 22, 2405–2412.
- Tseng, G. C., and Wong, W. H. (2005). Tight clustering: a resampling-based approach for identifying stable and tight patterns in data. *Biometrics* 61, 10–16.
- Vitali, B., Ndagijimana, M., Cruciani, F., Carnevali, P., Candela, M., Guerzoni, M. E., et al. (2010). Impact of a symbiotic food on the gut microbial ecology and metabolic profiles. *BMC Microbiol.* 10:4. doi: 10.1186/1471-2180-10-4
- Vitali, B., Ndagijimana, M., Maccaferri, S., Biagi, E., Guerzoni, M. E., and Brigidi, P. (2012). An *in vitro* evaluation of the effect of probiotics and prebiotics on the metabolic profile of human microbiota. *Anaerobe* 18, 386–391.
- Waldram, A., Holmes, E., Wang, Y., Rantalainen, M., Wilson, I. D., Tuohy, K. M., et al. (2009). Top-down systems biology modeling of host metabotype-microbiome associations in obese rodents. *J. Proteome Res.* 8, 2361–2375.
- Wang, Y., Tang, H., Nicholson, J. K., Hylands, P. J., Sampson, J., and Holmes, E. (2005). A metabolomic strategy for the detection of the metabolic effects of chamomile (*Matricaria recutita* L.) ingestion. *J. Agric. Food Chem.* 53, 191–196.
- Westerhuis, J. A., Kourti, T., and Macgregor, J. E. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemom.* 12, 301–321.
- Wishart, D. S. (2008). Metabolomics: application to food science and nutrition research. *Trends Food Sci. Technol.* 19, 482–493.
- Wong, J. M., de Souza, R., Kendall, C. W., Emam, A., and Jenkins, D. J. (2006). Colonic health: fermentation and short chain fatty acids. *J. Clin. Gastroenterol.* 40, 235–243.
- Wu, J., An, Y., Yao, J., Wang, Y., and Tang, H. (2010). An optimised sample preparation method for NMR-based faecal metabolomic analysis. *Analyst* 135, 1023–1030.
- Yamamoto, H., Yamajib, H., Fukusakic, E., Ohnob, H., and Fukuda, H. (2008). Canonical correlation analysis for multivariate regression and its application to metabolic fingerprinting. *Biochem. Eng. J.* 40, 199–204.
- Zhang, A., Sun, H., Wang, P., Han, Y., and Wang, X. (2011). Modern analytical techniques in metabolomics analysis. *Analyst* 137, 293–300.
- Zhao, Q., Stoyanova, R., Du, S., Sajda, P., and Brown, T. R. (2006). HiRes—a tool for comprehensive assessment and interpretation of metabolomic data. *Bioinformatics* 22, 2562–2564.
- Zheng, X., Xie, G., Zhao, A., Zhao, L., Yao, C., Chiu, N. H., et al. (2011). The footprints of gut microbial-mammalian co-metabolism. *J. Proteome Res.* 10, 5512–5522.
- Zhou, J., Aa, J., Wang, G., Zhang, F., Gu, U. R., Wang, X., et al. (2011). Visualization of multivariate metabolomic data. *Chin. Herb. Med.* 3, 289–293.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 June 2012; accepted: 25 November 2012; published online: 11 December 2012.

Citation: Vernocchi P, Vannini L, Gottardi D, Del Chierico F, Serrazanetti DI, Ndagijimana M and Guerzoni ME (2012) Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome. *Front. Cell. Inf. Microbio.* 2:156. doi: 10.3389/fcimb.2012.00156  
Copyright © 2012 Vernocchi, Vannini, Gottardi, Del Chierico, Serrazanetti, Ndagijimana and Guerzoni. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.