We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Computational Methods to Interpret and Integrate Metabolomic Data

Feng Li[1], Jiangxin Wang[2], Lei Nie[3] and Weiwen Zhang[2]
[1]Department of Mathematics and Statistics, University of Maryland,
Baltimore County, Baltimore, MD,
[2]School of Chemical Engineering and Technology, Tianjin University, Tianjin,
[3]Division of Biostatistics, Department of Epidemiology and Preventive Medicine,
University of Maryland Baltimore, Baltimore, MD,
[2]P.R. China
[1,3]U.S.A.

1. Introduction

Revolutionary improvements in high-throughput DNA sequencing technologies have made it possible to measure gene, mRNA, proteins and metabolites, as well as their interaction at global level. In the past decades, significant efforts in improving analytical technologies pertaining to measuring mRNA, proteins and metabolites have been made. These efforts have led to the generation of several new 'omics' research fields: transcriptomics, proteomics, metabolomics, interactomics and so on (Singh & Nagaraj, 2006; Fiehn 2007; Lin & Qian, 2007; Kandpal et al., 2009; Ishii & Tomita, 2009). Among them, metabolomics is an approach to obtain a comprehensive evaluation of metabolites in cells. Compared with transcriptomics and proteomics approaches, metabolomics can achieve large-scale quantitative and qualitative measurements of cellular metabolites, which can thus generate a high-resolution biochemical and functional information of an organism.

Due to the chemical complexity of cellular metabolites, it is generally accepted that no single analytical technique can provide a comprehensive visualization of all metabolites, so multiple technologies are generally employed (Dunn & Ellis, 2005; Villas-Boas Silas et al., 2005; Hollywood et al., 2006; Dettmer et al., 2007; Lenz & Wilson, 2007; Seger & Sturm, 2007). The selection of the most suitable technology is typically a compromise between speed, chemical selectivity, and instrumental sensitivity. Tools such as nuclear magnetic resonance spectroscopy (NMR) are rapid, highly selective and non-destructive, but have relatively lower sensitivities. Other tools such as capillary electrophoresis (CE) coupled to laser-induced fluorescence detection are highly sensitive, but have limited chemical selectivity (Ramautar et al., 2006). So far mass spectrometry (MS) measurement following chromatographic separation offers the best combination of sensitivity and selectivity (Dunn & Ellis, 2005; Bedair & Sumner, 2008). Mass-selective detection provides highly specific chemical information including molecular mass and/or characteristic fragment-ion that is directly related to chemical structure of molecules. This information can be utilized for compound identification through spectral matching with data compiled in libraries for

authentic compounds or used for *de novo* structural elucidation. Further, chemically selective MS information can be obtained from extremely small quantities of metabolites in the *p*mole and *f*mole level for many primary and secondary metabolites. Different technologies, either individual or integrated, could be employed for different study aims, based on metabolite identification, detection speed, high throughput and sensitivity. In the chapter, we will first review current MS technologies that have been incorporated into many metabolomics research programs as well as some of the emerging MS technologies that hold additional promise for the future advancement of metabolomics.

In contrast to classical biochemical approaches that typically focus on a single metabolite, single metabolic reaction or their kinetic properties, metabolomics involves collection of large amount of quantitative data on a broad series of metabolites in an attempt to gain an overall understanding of metabolism and/or metabolic dynamics associated with conditions of interest, such as disease or drug exposure. Generally speaking, metabolomics data share a great deal of similarity with transcriptomics data: both types of data matrices are large, feature rich, and challenged with issues of dealing with a limited sample size and a high-dimensional feature space. Thus in many cases the robust data processing algorithms originally developed for transcriptomic analysis have been adapted directly for metabolomic analysis. However, the challenges with metabolomic data can be unique, and may require new methodologies supported with a detailed knowledge of cheminformatics, bioinformatics, optimization, dynamic system theory and statistics. In recent years, many computational methods have been developed specifically for metabolomic data, varying from metabolic network analysis oriented or feature selection/data mining oriented. In the chapter, we will introduce each of these methods and their relevant applications, and will also discuss all the computational challenges associated.

2. Major analytical technologies of metabolomics

2.1 Mass spectrometry (MS)-based metabolomics

Several MS-based metabolomics technologies have been developed in the past decades, a brief introduction was presented here. Much detailed information regarding these technologies can be found from several recent excellent reviews (Dunn & Ellis, 2005; Hollywood et al., 2006; Dettmer et al., 2007; Bedair & Sumner, 2008).

Direct infusion mass spectrometry (DIMS) is a method for direct analysis of complex metabolic extracts without extraction or separation *via* electrospray ionization (ESI) MS, which provides a sensitive, high-throughput method to make it possible for several hundred samples per day. In terms of disadvantages, DIMS analysis is susceptible to ionization suppression due to competitive ionization with other components in the matrix (*e.g.*, salts and other ionic compounds, organic acids/bases, and hydrophobic compounds), although ionization suppression effects could be reduced by nano electrospray (nano-DIMS) with increased ionization efficiency (Southam et al., 2007). In addition, typical DIMS is not able to discriminate isomeric compounds; however, when coupled with tandem MS (DIMSMS) or Fourier transform ion cyclotron resonance (FT-ICR) spectrometers can trap and accumulate fragment ions that often enable to determine different isomeric structures (Aharoni et al., 2002). A comparison between direct infusion negative-ESI iontrap MS and GC-quadrupole MS analysis for the metabolic fingerprinting of five yeast mutants was reported recently

(Mas et al., 2007). Negative ESI LCQ ion-trap MS was reported as an effective method for the characterization of plant extracts with well-defined clusters in comparison to positive-ion ESI and 1H-NMR profiling (Mattoli et al., 2006).

Gas chromatography-mass spectrometry (GC-MS) has been a very useful technology for volatile and thermally stable polar and nonpolar metabolites (Tanaka et al., 1980). Metabolite identification or confirmation is performed by retention time or index comparisons with pure compounds and mass spectral interpretation or comparison using retention index/mass spectral library databases (Wagner et al., 2003). Metabolites can be classified into two classes: volatile metabolites not requiring chemical derivatisation (Yassaa et al., 2001; Mallouchos et al., 2002; Deng et al., 2004) and non-volatile metabolites requiring chemical derivatisation (Roessner et al., 2000). GC-MS based metabolic profiling has been used to compare four *Arabidopsis* genotypes and showed each genotype exhibited a different metabolite profile (Birkemeyer et al., 2003), and to compare transgenic tomato plants over expressing hexokinase (Roessner-Tunali et al., 2003). Silent phenotypes of potatoes have been distinguished from their parental background by employing metabolic profiling (Weckwerth et al., 2004). The same approach has recently been employed in microbial metabolomics to study the effect of different growth conditions on *Corynebacterium glutamicum* (Strelkov et al., 2004).

The application of Liquid Chromatography coupled to Mass Spectrometry (LC-MS) in metabolomics has been growing over the past few years (Wittmann et al., 2004). As a universal separation technique that can be tailored for the targeted analysis of specific metabolite groups or utilized in a broader non-targeted manner, LC offers additional benefits of analyte recovery by fraction collection and/or concentration, which has been difficult for GC separation. In addition, LC-MS operates at lower analysis temperatures than GC-MS, which enables the analysis of heatlabile metabolites. LC-MS analysis does not involve sample derivatization, which simplifies the sample-preparation and improves the identification of metabolites. The major disadvantage of LC-MS relative to GC-MS is the lack of transferable LC-MS libraries for metabolite identifications, although some efforts have been initiated to construct in-house LC-MS or LC-MS-MS libraries for automated metabolite identifications (Noteborn et al., 2000). Two-dimensional LC has also been utilized to increase the peak separation capacity (Aharoni et al., 2002). Recent LC-MS metabolite-profiling examples include the identification of flavonoids and isoflavonoids in *Medicago truncatula* (Daykin et al., 2002), the revelation of novel pathways by studying the differential and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures (Farag et al., 2008), and the investigation of small polar-metabolite responses to salt stress in *Arabidopsis thaliana* (Lindon et al., 2000). LC-MS has also been used in the non-targeted analysis of endogenous metabolites in an unbiased manner (Rashed et al., 1997; De Vos et al., 2007).

Capillary electrophoresis mass spectrometry (CE-MS) is a powerful separation technique for charged metabolites (Ramautar et al., 2006; Monton et al., 2007). CE has superior separation efficiencies compared to LC due to the plug-flow profile generated by the electroosmotic flow (EOF) as compared to the parabolic flow in LC. Capillary zone electrophoresis (CZE) has been the major CE mode used for CE-MS analysis of metabolites, due to the simplicity of the running buffer. Simultaneous separation of charged and neutral metabolites can be achieved using other CE modes (*e.g.*, micellar electrokinetic chromatography (MEKC) or

capillary electrochromatography (CEC)). Cationic and anionic CE-MS analysis of *Bacillus subtilis* extracts detected 1692 metabolite features of which 150 were identified (Soga et al., 2003). The same analytical procedure was recently used to study the alteration of metabolic pathways in transgenic rice lines that over-express dihydroflavonol-4-reductase (Sato et al., 2004). Non-aqueous CE-ESI ion-trap MSn was utilized for quantitative and qualitative profiling of isoquinoline alkaloids in single-plant tubers of four central European Corydalis species (Sturm et al., 2007).

2.2 Other emerging mass spectrometry technologies

Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) is a popular analytical technique for biopolymer analysis. It has a high throughput capacity and a higher tolerance for salts than ESI. In metabolomics, MALDI has largely been confined to the targeted analysis of high-molecular-weight metabolites due to the substantial chemical-background signals generated by the matrix in the low-molecular-weight region (<1,000 m/z) (e.g., the analysis of phospholipids in mammalian tissues (Jones et al., 2006), plant carotenoids (Fraser et al., 2007), and plant cell wall xyloglycans (Lerouxel et al., 2002). MALDI has also been used for imaging MS (IMS) of proteins and small molecules in tissues (Reyzer & Caprioli, 2007). Whole organisms or selected tissue sections are analyzed through an array of spots in which MS spectra are acquired at spatial intervals that define the image resolution. The m/z intensities of the acquired spectra are then plotted in the x and y coordinates to form a 2D image of the m/z values, which represents the spatial distribution of that metabolite/ion in the tissue (Rubakhin et al., 2005). MALDI-TOF IMS has been applied successfully for the study of drug and metabolite distributions in rat-brain tissues (Hsieh et al., 2006) and whole rat body (Khatib-Shahidi et al., 2006).

Desorption electrospray ionization (DESI) is a new, ambient, soft-ionization technique that combines features from both ESI and desorption ionization (DI) methods (Takats et al, 2004; Cooks et al., 2006). In DESI, an electrospray emitter is used to generate a spray of charged micro droplets that is directed towards an ambient sample surface. There is virtually no sample preparation required for DESI, thus allowing the direct analysis of animal and plant tissues. The application of DESI in metabolomics is relatively new, but its ambient DI properties as well as its high-throughput capacity make it an attractive tool for metabolomics. One promising area for DESI is *in vivo* metabolomics, which was demonstrated through the direct profiling of alkaloids from plant tissues of Conium maculatum without sample preparation while still identifying all of its previously reported alkaloids using tandem MS (Talaty et al., 2005). Although this technique can be incorporated into an IMS configuration, the spatial resolution of the DESI source (0.5–1.0 mm) is currently less than that of MALDI ion imaging (50–100 µm) (Wiseman et al., 2005).

Extractive electrospray ionization (EESI) is another new ESI technique that uses two separate sprayers. Although the exact sample-ionization mechanism is still unclear, the ionization process depends on liquid–liquid extraction between the colliding micro-droplets of the sample spray and the charged reagent-solvent spray (Chen et al., 2006). The advantage of EESI is its ability to analyze complex biological samples, such as urine and serum, directly with minimum or no sample preparation for an extended period of time. EESI-MS along with ^{1}H NMR was recently used to monitor the effect of diet on the metabolites founds in rate urine (Gu et al., 2007).

2.3 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is a technique that exploits the magnetic properties of certain atomic nuclei to determine physical and chemical properties of atoms or the molecules in which they are contained. It relies on the phenomenon of nuclear magnetic resonance and can provide detailed information about the structure, dynamics, reaction state, and chemical environment of molecules. Most frequently, NMR spectroscopy is used to investigate the properties of organic molecules, though it is applicable to any nucleus possessing spin. This can range from small compounds analyzed with 1-dimensional proton or carbon-13 NMR to large proteins or nucleic acids using 3 or 4-dimensional techniques (Grivet & Delort, 2009). NMR presents an unbiased technique for metabolite fingerprinting that is quantitative even in complex mixtures. Nicholson and co-workers have pioneered the application of NMR for metabolite fingerprinting (Lindon et al., 2003). The analysis of six yeast knock-out strains proved to classify and relate the genotypes by multivariate statistics, which potentially can be applied for functional genomics (Raamsdonk et al., 2001). NMR is non-destructive and therefore *in vivo* analysis is also possible (Gmati et al., 2005). Thus, NMR provides a powerful method for accessing metabolite complement (metabolome) and metabolic fluxes (fluxome) at a fine scale (metabolite identification) and a global scale (metabolomics).

2.4 Vibrational spectroscopy

Vibrational spectroscopy is one of the oldest spectroscopic methods. The vibrational states of a molecule can be probed in a variety of ways. The most direct way is through infrared spectroscopy (IRS), as vibrational transitions typically require an amount of energy that corresponds to the infrared region of the spectrum. Raman spectroscopy (RS), which typically uses visible light, can also be used to measure vibration frequencies directly. RS has been used for the identification of microorganisms of medical relevance (Dunn et al., 2005); however, its application for complex biological systems outside the area of microbiology is still in its infancy, although the potential of using 1064 nm excitation has been demonstrated in studies of the biochemical analysis of honey (de Oliveira et al., 2002) and in the analysis of plant pigments and essential oils (Schrader et al., 1999). In contrast, IRS has been applied for diagnostics, characterisation of microorganism and plant, adulteration and quality assurance, biomarker discovery and biochemical responses (Dunn et al., 2005).

2.5 Single-cell metabolomics

Single cell analysis is the new frontier in "OMICS" (Wang & Bodoritz, 2010). Most current metabolomic technologies only collect data averaged over thousands or millions of cells. However, cellular heterogeneity within a cell population is a widespread event (Irish et al., 2006; Graf & Stadtfeld, 2008). Since the metabolome provides biological processes occurring in the cells, it will be imperative to establish a reliable metabolomic method to measure at a single-cell level.

The challenges for single-cell metabolomics include: *i)* tiny quantities of metabolites from a single cell. A type single cell is about 1-500 *f*L in volume (from *E. coli* to large mammalian),

and with metabolites as low as *a*mole to *f*mole (Schmid A et al., 2010), 10^6 times lower than a typical population-based metabolomics. While amplification of DNA/RNA and highly sensitive fluorescence measurements could be employed in single cell genomics, transcriptomics and proteomics, no similar technique is available for single-cell metabolomics. *ii*) Sample processing for a single cell is extremely challenging. Even though detection limits for metabolites using MS can be as low as *f*moles to *a*mole range (Amantonico et al., 2008); however, transferring of a cell or cell content to mass spectrometer, conserving the original metabolome, and separating metabolites from cell debris, proteins and salts, would be critical.

In recent years, several approaches have been established for MS-based single-cell metabolomics (Figure 1) (see review by Heinemann & Zenobi, 2011). *i*) Sampling the cell contents with a micropipette, followed by injection into a mass spectrometer using a nano-electrospray ionization (nano-ESI) source (Masujima 2009). This approach, probably only suitable for very large cells, can only measure a few cells per hour; *ii*) Sample preparation on a microfluidic chip, followed by deposition on a sample plate for (matrix-assisted) laser desorption/ionization (MALDI or LDI) mass spectrometry (Lu et al., 2006; Mellors et al., 2008; Amantonico et al., 2008, 2010; Holmes et al., 2009). Once a complete setup is realized, it has the potential to generate high throughput data in an automated way; *iii*) Cell arraying, single cells are deposited on a sample plate for LDI or MALDI covered by a solvent-repelling, application of a MALDI matrix in an organic solvent will then lyse the cells and extract the compounds of interest for analysis by MALDI. This approach is a true high-throughput operation because the sample arraying can be automated, and thus the speed of MS instrument is the only limited factor (1000s of cells/hour) (Urban et al., 2010); *iv*) Imaging mass spectrometry, many modern mass spectrometers have imaging capabilities, with a spatial resolution of typically ~50 µm (MALDI or LDI), and ~1 µm (secondary ion mass spectrometry, SIMS), at relatively fast acquisition speed (Fletcher, 2009). With SIMS, the distribution of ions such as Na^+, K^+, Ca^{2+}, as well as cationized cholesterol, lipids present at cell surfaces can be imaged (Fletcher 2009). However, so far the data generated through this approach has been less quantitative (Heinemann & Zenobi, 2011).
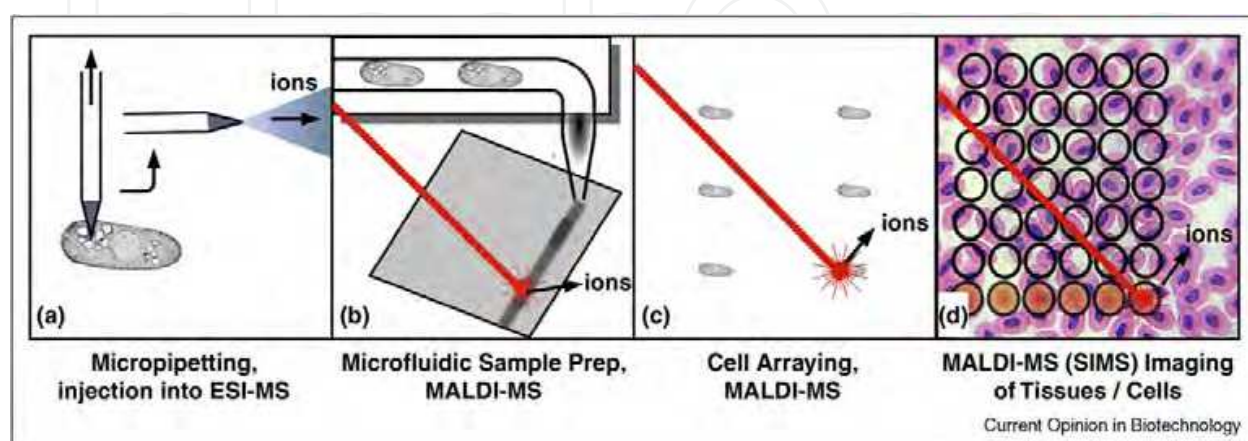

Fig. 1. Schematics of the four MS-based approaches for single cell metabolomics. (Heinemann & Zenobi, 2011)

Only very few metabolites can be analyzed directly in single cells by autofluorescence (Amantonico et al., 2010); however, by incorporating fluorescent tags or probes, researchers have been able to detect more metabolites. For example, Fehr et al. (2002) developed a protein based nanosensor for detection of maltose uptake by living yeast cells. In another study, a genetically encoded fluorescent sensor was expressed in living cells for detecting adenylate nucleotides (Berg et al., 2009). However, it is still arguable whether the formation of these foreign complexes between sensor and metabolites in cells will cause damages and lead to alteration of physiological status of cells.

The high sensitivity of electrochemical detectors to electroactive species makes them suitable for targeted studies of metabolites in single-cell analysis. A range of microscale electrochemical methods have been introduced to monitor various physiological processes (Huang & Kennedy, 1995). For example, release of metabolites, such as catecholamines and oxygen, can be readily measured electrochemically (Cannon et al., 2000). Specific methods targeting particular metabolic pathways in single cells have been in use for a long time, including autoradiography of cells preincubated with radioisotopically labeled compounds (Fliermans & Schmidt, 1975).

Other methods used in single cells including single-cell spectroscopy in conjunction with image analysis for glycogen metabolism in yeast cells (Cahill et al., 2000), enzyme-catalyzed luminescence method for dopamine release from a mammalian nerve cell (Shinohara & Wang, 2007), synchrotron Fourier transform infrared spectromicroscopy for ethanol formation in single living cells of unicellular algae (Goff et al., 2009), raman spectroscopy (Schuster et al., 2000; Buckmaster et al., 2009; Hermelink et al., 2009) for detecting nucleic bases and amino acids in single cells, and nuclear magnetic resonance (NMR) for structural characterization of organic compounds, including metabolites (Beckonert et al., 2007; Motta et al., 2010). Brief summary and comparison for various metabolomics techniques discussed above are listed in Table 1.

3. Computational methods for analyzing metabolomic data

Several computational methods have been developed in recent years to analyze metabolomic data. The overview of the metabolomic data processing is shown in Figure 2. These computational methods can be divided into two major categories: methods for data pre-processing (low-level, such as noise reduction) or methods for interpretation (high-level, such as feature selection). Pre-processing methods concern the improvement and the enhancement of raw signals, which typically include noise reduction, peak detection, baseline correction, peak alignment and normalization. Pre-processing methods for metabolomic data has been reviewed in details by researchers from different perspectives (Jewett et al., 2007; Enot et al., 2011). In this chapter, we will focus on the most widely used or recently developed high-level methods for interpreting metabolomic data. Some computational methods may also depend on the platform or instruments used. The reader should refer Section 2 of the chapter for detailed explanation on popular platforms such as GC-MS, LC-MS and NMR etc. Some software packages have been developed for interpreting metabolomic data in recent years, although review of the software tools are beyond the scope of this chapter, most of the software utilize the statistical methods discussed here.

Techniques	Advantages	Disadvatanges
DIMS	high-throughput, simple sample preparation	nondiscrimination of isomeric compounds, susceptilble to ionization
GC-MS	high sensitivity, ideal to complex samples, versatile	derivation needed for non-volatile metabolites
LC-MS	high sensitivity, average/high resolution	limited structural information, matrix effects
CE-MS	high sensitivity, quantification, label free, small volume	no robust, destructive
MALDI-MS	high sensitivity, detection of a wide range of molecule, label free	no quantification, destructive
DESI	direct analysis, high-throughput, ambient depsorption ionization	optimization required for each sample, low resolution
EESI	high-throughput, simple sample preparation	destructive
NMR	structural information, qualitative and quantitative, versatile	expensive, time consuming, difficult interpration
IR	versatile, easy to identify functional groups	poor structural information, difficult sample preparation, destructive
Raman	label free	low selectivity, poor sturctural information
FM	high sensitivity, imaging capabilities, dynamic	no structural information, targeted analysis only, difficult labelling
EC	high sensitivity, quantification, label free	not comprehensive, no structural information, vulnerable to interferences

FM: fluorescence microscopy
EC: electrochemical

Table 1. Summary for metabolomics techniques: characteristic of the main techniques considered for application in metabolomics
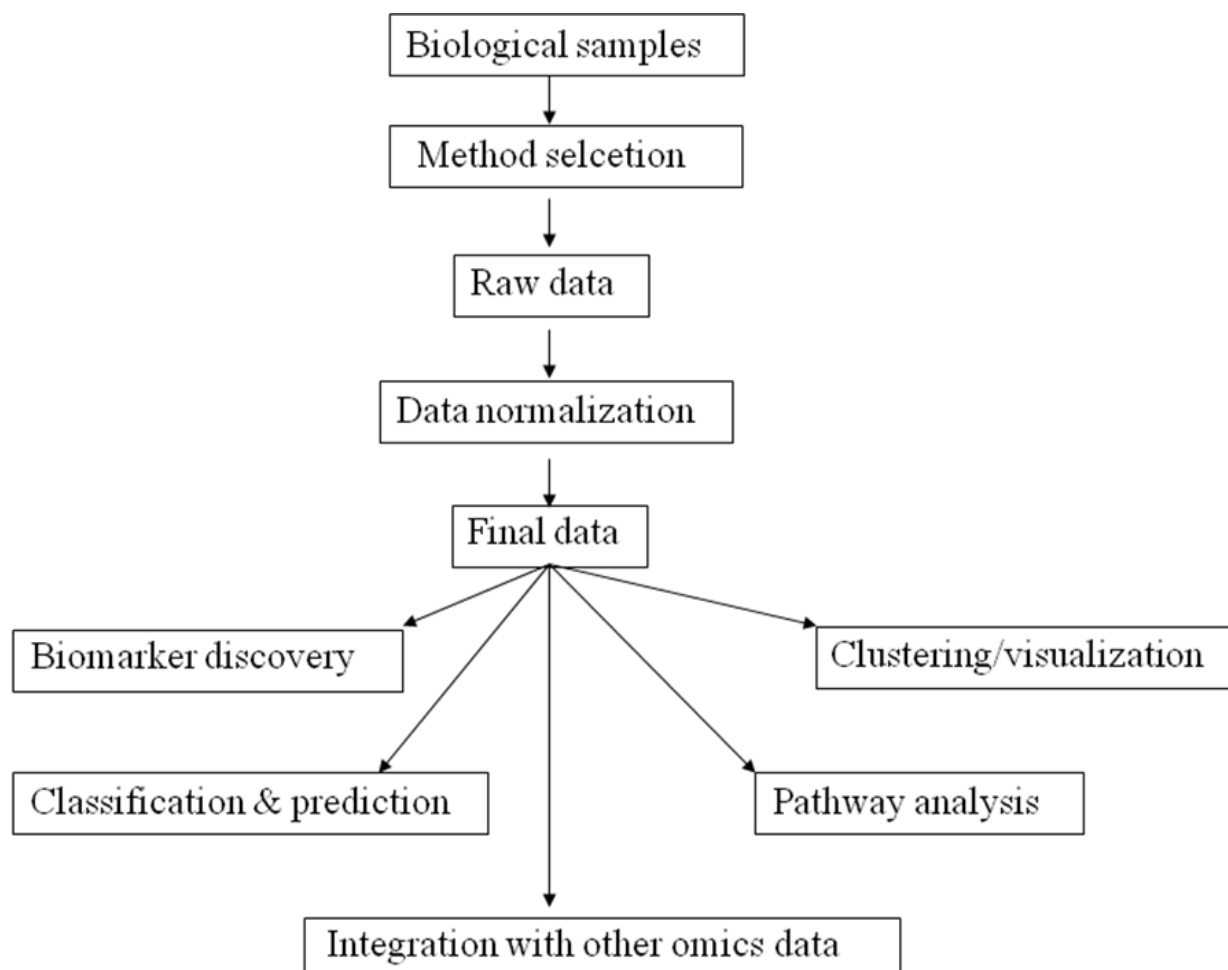
Fig. 2. An overview of metabolomic processing pipeline.

3.1 Computational methods for interpreting metabolomics

The computational methods to interpreting metabolomics data should be selected according to the aim of a study. If the aim is for sample classification and *prior* information about the sample identity is unknown, unsupervised methods such as hierarchical clustering analysis (HCA), principal component analysis (PCA), or self-organizing map (SOM) are typically used. On the other hand, in the case that sample identity is known and the aim of a study is to discover characteristic biomarkers (*e.g.*, search for biomarkers of a disease comparing samples from healthy and diseased subjects), supervised methods such as linear discriminant analysis, artificial neural networks (ANN) or support vector machine (SVM) can be used. The supervised methods use *prior* information about sample class and perform better in biomarker discovery. However, if the aim of the study is a mere biomarker discovery in samples of known classes, common statistical methods such as ANOVA with multiplicity control can also be used (Jonsson et al., 2005). Below, we categorize and discuss different computational methods according to their usage in practice. We will not specifically differentiate them by the mathematical forms such as univariate or multivariate models. It should be noted that most of the methods are multivariate and are often used in combination in practical application.

Biomarker discovery: Like other "omics" studies, the primary objective of many metabolomics studies is to find biomarkers that are discriminative between matched "case" and "control" samples, *i.e.,* which metabolites are apparently altered under different physiological conditions. In pharmaceutical research, metabolomics study has been used for biomarker discovery for different diseases, safety markers, or drug mechanism research. However, given the large number of metabolites studied simultaneously with usually small sample size, it is very common to find metabolites that appear persuasive but in fact spurious. Thus, it is of crucial importance to control the rate of false positive (Broadhurst & Kell, 2006). To tackle this problem, many statistical methods have been developed under the term of large scale hypothesis testing (Benjamini & Hochberg, 1995; Storey 2002; Efron 2003, 2004a, 2004b, 2007a, 2007b, 2008; Storey & Tibshirani, 2003; Reiner et al., 2003; Xie et al., 2005). In classical hypothesis testing, the fundamental problem is to control type I error, the probability that a non-trivial finding is declared while it actually happens by chance. Type I error increases with the number of independent hypothesis considered simultaneously. The well-known and widely used strategy to control the overall type I error rate is Bonferroni-correction, in which the critical value for individual hypothesis testing is obtained by dividing the significance level by the number of hypothesis considered. For example, in metabolomics, if the search for discriminating biomarkers is performed using 500 metabolites and an acceptable chance to reject one true hypothesis is 0.05, and then the Bonferroni-corrected critical value for rejecting an individual null hypothesis for a metabolite should be 0.05/500 = 0.0001). Berferroni-correction is conservative in the sense that it excludes type I errors at the cost of increasing the potential for type II errors (false negatives) (Broadhurst & Kell, 2006).

A widely accepted error measure in microarray literature for large-scale hypothesis testing is the false discovery rate (FDR), the proportion of false positives among all the discoveries. The procedure controlling false discovery rate proposed by Benjamini and Hochberg (1995) has been recognized as a breakthrough and widely applied or adapted by statistical researchers (Efron, 2004a). Most of the literatures assume that the theoretical null hypothesis is known in advance. However, Efron argued that in large scale hypothesis testing, like in 'omics' studies, the theoretical null often fails for reasons like correlations among proteins or genes, unknown confounding factors, or systematic bias (Efron, 2004b, 2007a, 2008). Thus, it will be appropriate to estimate the distribution of the null statistics from the data in order to have a more meaningful discovery. Translating to the setting of a metabolomics study, Efron's concept aims to find a subset of metabolites that behave very differently from the majority of the metabolites. Efron's creative idea has received significant attention from the research field. Applications to metabolite biomarker finding has not been reported but certainly expected.

Metabolomics-based biomarker discovery have been reported. In one example for invasive ovarian carcinomas and ovarian borderline tumors, a differential analysis of 291 detected metabolites in sixty-six invasive ovarian carcinomas and nine borderline tumors of the ovary revealed 51 metabolites that were significantly different between borderline tumors and carcinoma with a FDR controlled at 7.8% (Denkert et al., 2006). For *Onchocera volvulus*, analysis of an African sample set of 73 serum and plasm samples based on LC-MS revealed a set of 14 biomarkers that showed excellent discrimination between *Onchocera volvulus*-positive and negative individuals (Denery et al., 2010). Controlling FDR at 54% using

Storey's *q*-value approach (Storey, 2002) resulted in 194 features selected from a total of 2350 mass features. Among the 194 features, the authors selected the top 14 feature for investigation.

Data clustering and visualization: Clustering or unsupervised modeling is useful for class discovery and provides information on data similarity: metabolomic samples clustered or grouped together can be objectively considered to be similar.

Principal components analysis (PCA) is probably the most widely used unsupervised approach to data mining or visualization. PCA is a multivariate technique that transforms the data into a coordinate system where each new projection (also called principal components (PC)) is a linear combination of the original variables. PCs are orthogonal so that each dimension is related to different data characteristics and source of variability in a mathematical sense (Enot et al., 2008). As a dimension reduction tool, PCA is very useful for metabolomics data visualization and further data clustering. However, PCA may not work well for metabolomics data where the differences between groups are minor and obscured by other covariates.

Self-organizing map (SOM) is another visualization tool for high-dimensional data (Kohonen, 1998, 2001). The SOM describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the closest weight vector to the vector taken from data space. Once the closest node is located, it is assigned the values from the vector taken from the data space. The SOM places similar input data in adjacent nodes. Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar apart. One disadvantage of SOM is that the final map solution is dependent on the order of the presentation of the training data. The batching-learning version of the algorithm (Kohonen, 2001) overcame this problem and yields reproducible maps. SOM has been applied to metabolic profiling for clustering blood plasma (Kaartinen et al., 1998), and NMR spectra of breast cancer tissues (Beckonet et al., 2003b). More recently, Kouskoumvekaki et al. (2008) applied SOM to identify similarities among the metabolic profiles of different filamentous fungi. Meinicke et al. (2008) proposed one-dimensional SOM for metabolite-based clustering and visualization of marker candidates. In a case study on the wound response of *Arabidopsis thaliana*, they showed how the clustering and visualization capabilities of SOM can be utilized to identify relevant groups of biomarkers.

As a popular unsupervised learning method, Hierarchical cluster analysis (HCA) clusters the data to form a tree diagram or dendrogram which shows the relationships between samples (Ebbels, 2007). The algorithm begins by computing the distances between all pairs of samples. Initially each cluster consists of a single sample. The algorithm proceeds iteratively until all samples are members of a single cluster. The final structure of the resulting clusters depends on the choice of distance function or "linkage" between two clusters as well as a similarity cut-off. The most popular choices of linkage are centroidal, average, single (nearest neighbor) and complete (farthest neighbor) linkages. The centroidal linkage defines the inter-cluster distance as the distance between the centroids. To determine the cluster membership, one must decide on a similarity cut-off which breaks the dendrogram into a number of separate clusters. As an example, Beckobert et al. (2003) used the HCA method to explore a set of toxicology studies. HCA allowed interpretation of the

data in terms of the magnitude and site of toxicological effect, and helped to explain misclassifications by other methods.

The *k*-means clustering is a method of cluster analysis that aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. In *k*-means clustering, the Euclidean distance is used as a distance metric and variance is used as a measure of cluster scatter. The number of clusters *k* is an input parameter. When performing *k*-means, it is important to run diagnostic checks for determining the number of clusters in the data set. Thus, *k*-means is often used in conjunction with other clustering and visualization methods (Ebbels, 2007).

Fuzzy *k*-means (also called Fuzzy c-means) is an extension of *k*-means clustering technique based on Fuzzy logic. While *k*-means discovers hard clusters (a point belong to only one cluster), Fuzzy *k*-means is a more statistically formalized method and discovers soft clusters where a particular point can belong to more than one cluster with certain probability. In Fuzzy *k*-means algorithm, one sample can be assigned to more than one class instead of only one. The membership of each sample is calculated and then represented by a membership value between 0 and 1, instead of just 0 and 1 in the hard clustering. Cuperlovic-Culf et al. (2009) presented the application of fuzzy *k*-means clustering method for the classification of samples based on metabolomics 1D ^{1}H-NMR fingerprints. The sample classification was performed on NMR spectra of cancer cell line extracts and of urine samples of type 2 diabetes patients and animal models. The fuzzy *k*-means clustering method allowed more accurate sample classification in both datasets relative to the other tested methods including PCA, HCA and *k*-means clustering. Li et al. (2009) applied fuzzy *k*-means to cluster three gene types of *Escherichia coli* on the basis of their metabolic profiles and delivered better results than PCA. On the basis of the optimized parameters, the fuzzy *k*-means was able to reveal main phenotype changes and individual characters of three gene types of *E. coli*, while PCA failed to model the metabolite data.

Clustering of metabolomics data can be hampered by noise originating from biological variation, physical sampling error and analytical error. Bootstrap aggregating (bagging) is a resampling technique that can deal with noise and improves accuracy. Hageman et al. (2006) demonstrated the application of bagged clustering to metabolomics data. It was argued that the bagged *k*-means should be favored against ordinary *k*-means clustering when dealing with noisy metabolomics data.

In practice, it is common to combine dimensionality reduction and clustering methods. For example, first, a sample-based principal component analysis (PCA) is performed to compute a subset of principal components. Then the metabolite-specific PCA loadings of these components are used for metabolite-based clustering using *k*-means or hierarchical methods (Pohjanen et al., 2006).

Classification and prediction: While the purpose of clustering is to group similar data together, classification aims at finding a rule to discriminate the classes in an optimal way as well as selecting the subset of features that are most discriminative or predicative. In contrast to clustering applications, the class label and the number of classes are known for a subset of data (training samples) in *a priori* in classification problem. Once the rule or classifies are determined using training dataset, it can be used to predict the class label (such as diseased or not) of a test sample.

The *k*-Nearest Neighbour (*k*NN) rule for classification may be the simplest of all supervised classification approaches (Ebbels, 2007). Different from other supervised learning methods, the training phase of *k*NN consists of only storing the training samples and the corresponding class labels. In the classification (prediction) phase, a test sample is classified by assigning the label which is most frequent among the *k* nearest training samples. The method requires only the choice of *k*, the number of neighbors to be considered when making the classification. Greater values of *k* reduce the effect of noise on the classification, but make boundaries between classes less distinct. A limitation of this classification method is that the classes with the more frequent examples tend to dominate the prediction of a test sample. Usually, *k* is chosen through a cross validation procedure. *k*NN has been often used as a comparator for other methods in literature (Beckonert et al., 2003; Baumgartner et al., 2004).

Partial least square for discriminant analysis (PLS-DA) is a regression extension of PCA that takes advantage of class label information to maximize separation between groups of observations. PLS-DA models the relationship between the class affiliation matrix (Y) and feature matrix (X), which is a generalization of multiple linear regressions. It determines a set of latent variables explaining as much as possible of the covariance between the two matrices. PLS-DA can deal with uncompleted dataset and multicollinearity problem. The output of PLS-DA is the score matrix that can be plotted similarly as in PCA and the predictor matrix containing estimated class affiliation (Ciosek et al., 2005; Trygg & Lundstedt, 2007). The ortogonal-PLS (OPLS) method (Cloarec et al., 2005; Trygg & Lundstedt, 2007) is a recent modification of the PLS method. The main idea of OPLS is to separate the systematic variation in X into two parts, one that is linearly related to Y and one that is orthogonal to Y. The OPLS method provides a prediction similar to that of PLS. However, the interpretation of the models is improved because the structured noise is modeled separately from the variation common in X and Y. Analogous to PLS, when the Y matrix is class affiliation, the corresponding analysis is named OPLS-DA. Cloarec et al. (2005) illustrated the applicability of the method in combination with statistical total correlation spectroscopy to ^{1}H NMR spectra of urine from a metabonomic study of a model of insulin resistance based on the administration of a carbohydrate diet to three different mice strains. Tapp and Kemsley (2009) recently discussed similarities and differences between PLS-DA and OPLS-DA with a focus on the usage of OPLS in the analytical chemistry literature. They concluded that the two methods are very similar and no one outperforms the other, and the reported discrepancies in the literature must be due to differences in the implementation details, or some otherwise ''unfair'' comparison between the methods

An Artificial neural network (ANN) is a widely used non-linear data modeling tool (Bishop, 1995). An ANN is a computational model that is inspired by the structure and functional aspects of biological neural network. An ANN consists of a layered network of nodes with simple linear or sigmoid activation function. The most widely used type of ANN is the multilayer perceptron (MLP), which has at least three layers including the input layer, one or more hidden layers and output layer. The most attractive feature of a MLP is its capacity to approximate any continuous function in arbitrary precision given enough number of nodes in the hidden layers with sigmoid-type activation functions. ANN has been applied to classification of tumor cells by different researchers (Maxwell et al., 1998; Ott et al., 2003).

One disadvantage of ANN is that non-linear optimization algorithm is needed to train a neural network. It is easy for the optimization procedure to be stuck in local minima while finding the optimal parameters. Thus, global optimal solution is not guaranteed. Moreover, it is difficult to interpret the connection weights to gain some biological insight for feature ranking in applications.

Support vector machines (SVMs) is a supervised learning technique for classification and regression (Cristianini & Shawe-Taylor, 2000). A support vector machine constructs an optimal hyperplane for classification. The hyperplane is constructed in such a way that it has the largest distance to the nearest training data points of any class. In contrast to ANN, SVMs is trained by using convex programming such as quadratic programming. Thus, it can find the global optimal solution efficiently. The computation complexity of SVMs depends on the number of samples instead of the dimension of each sample. Thus there is no curse of dimensionality problem. When cases are not linearly separable, appropriate kernel functions can be chosen to transform the original data into high dimensional feature space. SVMs is well known for its good generalization capacity. For nearly a decade, SVMs has been used in the field of bioinformatics for classifying and evaluating gene expression microarray data (Furey et al., 2000; Guyon et al., 2002). Mahadevan et al. (2008) compared the performance of PLS-DA multivariate analysis to SVMs and showed that SVMs were superior to PLS-DA in terms of predictive accuracy with the least number of features. With fewer features, SVMs were able to give better predictive model when compared to that of PLS-DA.

Computation methods for identification of metabolites: Metabolomics studies are targeted at identifying and quantifying all metabolites in a given biological context. One central problem is the translation of the measured mass into molecular formulae. Given the observed mass and atoms, the first problem is to find all the possible elemental compositions by solving the Diophantine equation:

$$\sum_i n_i \cdot m_i = M \ ,$$

where M is the observed mass and m_i is the mass of the ith atom. Diophantine equation is the basis for much of the mass spectrometry software to obtain compositions. Usually, there are many integer solutions to it mathematically. Among all the mathematical solutions, we then seek all of the integers n_i that are chemically feasible considering some chemical contextual information such as the valence rules, double-bond equivalents or exact mass (Meija, 2006). Even with very high mass accuracy (<1 ppm), many chemically possible formulae can be obtained in higher mass regions. To further reduce the number of potential elemental compositions, it is necessary to utilize isotope abundance pattern (Kind & Fiehn, 2006).

The identification of small metabolites has been seen as one of the bottlenecks in interpreting metabolomics data. Neumann & Bocker (2010) provided a review focusing on the computational methods for electronspary ionization (ESI) mass spectrometry. One of the most common methods for the identification of compounds using mass spectrometry is the comparison with spectra of authentic standards. The Metabolomics Standards Initiative

(MSI) has defined several confidence levels for the identification of non-novel chemical compounds, ranging from level 1 for a rigorous identification to unidentified signals at level 4 (Sumner et al., 2007). The difference between level 1 and 2 is that the former requires the comparison with authentic standards based on in-house data measured under identical analytical conditions, whereas the latter allows one to use literature or external databases. Level 1 or level 2 identifications are based on a comparison of "exact mass and isotope pattern". Even with the most exact mass and isotope pattern the identification will be limited to the elemental composition. Many compounds share the same sum formula for known metabolites in databases such as KEGG or PubChem. For all other MSI levels, the "identification" usually reduces to an annotation with lower levels of confidence (Neumann & Bocker, 2010).

When a reference spectrum is used, a similarity or distance function is needed for selecting database entries. The most basic similarity functions are those based on counting the number of matching peaks between a query spectrum and each of the database spectra. For this, both spectra can be considered as binary vectors with 0's and 1's for "peak absent" and "peak present", respectively (Neumann & Bocker, 2010). Common distance functions on binary vectors are the Hamming distance (counting any difference) or the Jaccard coefficient (the fraction of matching peaks). Besides counting matches, other measures also consider their actual mass and intensity, such as the Euclidean distance, the probability-based matching (PBM), the normalized dot product (NDP), and a modified cosine distance for the database search of EI spectra (Stein, 1994).

Oberacher et al. (2009) proposed and optimized a search function for tandem mass spectrometry (MS/MS-spectra) based on a combination of relative and absolute match probabilities, which combines the principle of peak counting and summed intensities of matching peaks. The X-Rank algorithm (Mylonas et al., 2009) for MS/MS-spectra match is based on probability calculations. It sorts peak intensities of a spectrum and then establishes a correlation between two sorted spectra. X-Rank computes the probability that a rank from an experimental spectrum matches a rank from a reference library spectrum. The solution requires training on a representative dataset. In a training step, characteristic parameter values are generated for a given data set. Identification of small compounds is still challenging, especially for compounds that have not been recorded in any library or structure database. Methods for these tasks are highly sought.

3.2 Computational methods for metabolic pathway analysis using metabolomic data

Metabolomics data provides a series of snapshots of cellular metabolism, which can be combined with metabolic flux data for further analysis. Metabolic pathways are the true functional units of metabolic systems (Schilling et al., 2000b). Finding biochemically plausible pathways between metabolites in a metabolic network is a central problem in computational metabolic modeling. Mathematical modeling approaches to metabolic regulation analysis involve different levels of details and complexities ranging from detailed kinetic models, stoichiometric analysis, structural kinetic models, to large scale topological network analysis (Steuer, 2007).

Detailed kinetic models of metabolic pathways, based on explicit enzymekinetic rate equations, is a bottom-up approach towards more comprehensive large-scale dynamic

models. It allows for the most detailed quantitative evaluation of the dynamics of metabolic systems, which is very important for improving the understanding of metabolic regulation and control. The metabolic control analysis (MCA) is the culminating mathematical theory from kinetic models, which describes the control and regulatory properties of metabolic systems (Heinrich & Schuster, 1996; Fell, 1997).

A metabolic network is a collection of enzyme-catalyzed reactions and transport processes that serve to dissipate substrate metabolites and generate final metabolites. The dynamics of a metabolic system can be described by a set of ordinary differential equations:

$$\frac{dX_i}{dt} = \sum_j S_{ij} v_j(X,k) = S \cdot v(X,k), \tag{1}$$

where X_i represents the concentration of the metabolite and S_{ij} stands for the stoichiometric coefficient for the reactant i in the jth reaction. $v_j(X,k)$ corresponds to the flux through the jth reaction. The vector $v(X,k)$ consists of nonlinear enzyme-kinetic rate functions, which depends on the concentration X and a set of kinetic parameters k. Given the $v(X,k)$ function form, the set of kinetic parameters k and an initial state $X(0)$, the differential equations can be solved numerically to obtain the time-dependent behavior of all metabolites under consideration. The stoichiometric matrix S is an m by n matrix where m corresponds to the number of metabolites and n is the total number of metabolites and n is the total number of fluxes taking place in the network.

The stoichiometric analysis approach takes advantage of the structure nature of metabolic system. Knowledge of the stoichiometry puts constraints on the feasible flux distributions, which can be utilized to model the functional capabilities of metabolic networks (Varma & Palsson 1994; Edwards & Palsson, 2000; Stelling et al., 2002; Famili et al., 2003; Price et al., 2003).

The pathway structure should be an invariant property of the network along with stoichiometry. Under steady-state, the set of ordinary equations reduce to linear equations:

$$0 = S \cdot v(X^0,k) = S \cdot v^0, \tag{2}$$

which is typically underdetermined as the number of reactions exceeds the number of participating metabolites. The set of all solutions v^0 to the steady state equation (2) is called null space. A set of basis vectors corresponding to a steady-state biochemical pathway can be selected to describe the null space. Additional constraints from biological insights are necessary to determine the system completely (Schilling et al., 2000b), which generally results in a set of linear equalities and inequalities. The set of feasible solutions under bio-chemical constraints form a convex steady-state flux cone. The convex analysis of biochemical networks was founded by Clarke (Clarke, 1980, 1988). The study of convex flux cones utilizes methods and concepts rooting in linear algebra and optimization and forms the underlying mathematical structure for metabolic pathway analysis.

Flux balance analysis (FBA) is a computational approach to reduction of the admissible flux space (Schilling et al., 1999; Edwards & Palsson, 2000). FBA optimizes an objective function such as maximal biomass yield or maximal energy production through the steady-state flux space has resulted in many applications (Papin et al., 2004; Almaas et al., 2004; Stephanopoulos et al., 2004). Using FBA, *in silico* studies on the systemic properties of the *Haemophilus influenzae* and *E. coli* (Edwards & Palsson, 2000) metabolic networks have been completed. Under various substrate conditions, Schilling et al. (2000a) explored the metabolic capabilities and predicted functions of a sub-system of the *E. coli* using FBA.

A metabolic network can be decomposed into distinct pathways, termed elementary flux modes (EFM). An EFM is the minimal set of reactions capable of working together in a steady state, which is unique for a given metabolic network. Another closely related concept is extreme pathways, which are a subset of elementary modes (Klamt & Stelling, 2003). All feasible flux vectors can be described as linear combinations of EFMs. The concept of EFM has resulted in a vast number of applications for metabolic network analysis (Stelling et al., 2002; Schuster et al., 2002; Klamt & Schuster, 2002; Klamt & Gilles, 2004; Klamt et al., 2006). For medium-sized metabolic networks, software packages have been developed for the computation of elementary flux modes (Hoops et al., 2006; Klamt et al., 2007). Owing to a combinatorial explosion of the number of elementary vectors, this approach becomes computationally intractable for genome scale networks. To develop an analysis approach computationally feasible even for genome scale networks, Urbanczik & Wagner (2005) proposed to focus on conversion cone, the projection of the flux cone, which describes the interaction of the metabolism with its external chemical environment. The method for calculating the elementary vectors of this cone was applied to study the metabolism of *Saccharomyces cerevisiae*.

Stoichiometric analysis does not incorporate dynamic properties into the description of the system. Steuer (2006, 2007) proposed a structure kinetic modeling approach to augmenting the stoichiometric analysis with kinetic properties. The idea of the proposed approach is to use a local linear approximation to explicit kinetic model to capture the dynamic response to perturbations, the stability of a metabolic state, as well as the transition to oscillatory behavior. The local linear approximation is obtained from a Taylor series expansion of the metabolic system. The linear term of the expansion is the derivative of the kinetic rate equations with respect to the metabolic concentration X at a given state, which usually requires knowledge of the enzyme-kinetic rate equations. Even in the absence of enzyme-kinetic information, it is still possible to specify the structure of the linear term. Structure kinetic modeling approach allows quantitative conclusions about the possible dynamics of the system, based on only a minimal amount of additional information.

The extension of the detailed kinetic models to whole cell models is faced with some fundamental difficulties including the absence of comprehensive measured kinetic parameter values, and the observed inconsistency in the available kinetic data, and the computational complexity of such models. Traditionally, kinetic models are constructed using rate equations derived to describe conditions in vitro and thus rely on the use of in vitro measured kinetic parameters. However, the conditions at which *in vitro* experiments are performed are often very different from those inside the cell. Thus, *in vitro* kinetic rates and *in vitro* kinetic parameters describe enzymatic behaviors that may not truly represent the observed physiological kinetic behavior in the cell. Several methods have been proposed

to address this issue by incorporating *in vivo* measurements in constructing kinetic models (Visser & Heijnen, 2003) or estimating kinetic parameters in biochemical networks using measured variables (Lei & Jorgensen, 2001; Moles et al., 2003; Segre et al., 2003). These methods require considerable mathematical efforts or utilize nonlinear optimization techniques. (Famili et al., 2003) proposed an approach for incorporating steady-state *in vivo* data with constraint based modeling approach to determine all candidate numerical values of kinetic constants. The kinetic solution space, termed *k*-cone, contains all the allowable numerical values of the kinetic constants. The *k*-cone is obtained by approximating the nonlinear kinetic rate laws as a linear or bilinear function of the kinetic constants. The *k*-cone approach can be used to determine consistency between *in vitro* measured kinetic values and *in vivo* concentration and flux measurements when used in a network-scale kinetic model. To calculate the relationship between kinetic parameters measured in vitro and the *k*-cone, optimization methods were essential. It was successful in determining whether *in vitro* measured kinetic values used in the reconstruction of a kinetic-based model of *Saccharomyces cerevisiae* central metabolism could reproduce *in vivo* measurements.

Despite the number of research published, information gained from theoretical or experimental metabolic network has not fully enabled probing biochemical pathway structure with the aim at detecting novel metabolic routes (Fiehn, 2007).

4. Computational methods to integrate metabolomic data with other "omics" datasets

It is also becoming clear that any single "omics" approach may not be sufficient to characterize the complexity of biological systems and an integrated "omics" approach may be a key to decipher complex biological systems (Gygi et al., 1999; Zhang et al., 2010). In general, integrated analysis of metabolomics dataset with other types of "omics" datasets can increase both dimension of information sources and statistical power in order to generate a conclusion with high confidence. In recent years, several computational methods have been applied to integrate metabolomics with other "omics" datasets, and the results demonstrated that better pattern recognition and association identification can be achieved when proper mathematical, statistical and other computational tools were applied.

The simple correlation analysis, *Pearson* or *Spearman*'s correlation has been used to assess degree of association between metabolomic and transcriptomic data. For example, it has been used to distinguish different potato tuber systems, and to determine the relationship between genes and their paired metabolites (Urbanczyk-Wochniak et al., 2003, 2007), through integrating metabolomic and transcriptomic datasets.

Because of the high dimensionality involved in the metabolomic data and other "omics" data, the dimension reduction tool such as PCA can demonstrate its power in the integrated analysis. It has been used in two different but related scenarios: *i*) PCA applies to each "omics" data to reduce the dimensionality of each omics data so that the "omics" data are ready for integration; *ii*) PCA applies to the integrated metabolomics and other "omics" data directly to identify a particular pattern. In scenario 1, metabolomic data or other "omics" data are not directly used in any integrated analysis. Rather, PCA was used to reduce the dimension of each data, so that researchers can focus on the most important components of metabolomic data or other "omics" data (Urbanczyk-Wochniak et al., 2003; Rubingh et al.,

2009; van den Berg et al., 2009). In scenario 2, PCA was directly used to identify pattern of the integrated metabolomic data with other "omics" data. For example, it was used to a integrated metabolomic data and a proteomic data to reveal clustering of the two genotypes (Weckwerth et al., 2004). While PCA can also be used for exploring polynomial relationships and for multivariate outlier detection, this method is restricted to linear relationships.

In both correlation and PCA analysis, the roles of all variables are the same and they are interchangeable. They are used to explore associations between factors. In some other analyses, some factors (independent variables X) are used to explain or predict the variable of main interest (dependent variable Y). For example, PLS is a statistical method that models Y over X through a linear relationship. Rather than considering all dependent variables as regressors in a multivariate regression analysis, PLS regresses Y over principal components resulted from a principal components analysis (Garthwaite 1994). The method was in fact previously used to model metabolomic variables as a function of the transcriptome profiles (Pir et al., 2006). The analysis allowed the discrimination between the effects that the growth media, dilution rate and deletion of specific genes on the transcriptomic and metabolomic profiles (Pir et al. 2006). The method was also used to relate quantifiable phenotypes of interest such as protease activity or productivity, to concentrations of each of the metabolites determined (Braaksma et al., 2011). The analysis revealed various sugar derivatives correlated with glucoamylase activity.

As an extension of PLS, Le Cao and colleagues proposed a sparse PLS approach to combine integration and simultaneous variable (*e.g.*, gene) selection in one step (Le Cao et al., 2008, 2009). In the approach, the PLS was penalized by the sum of the absolute values of the coefficients through least absolute shrinkage and selection operator (LASSO) (Tibshirani 1996), therefore automatically eliminating variables (e.g., genes) with negligible effects. The model selection approach, together with the smoothly clipped absolute deviation approach (Fan & Li, 2001) is effective in analyzing data with sparsity (*e.g.*, only a few genes have significant effects).

The methods previously discussed in this section, including *Perason* correlation, PCA, and PLS, are all methods to explore linear relationship. On the other hand, the kinetic model and artificial network could be more sensible when nonlinearity occurs. In an analysis to integrate metabolomics and pharmacokinetics (or nutrikinetics), Van Velzen et al. (2009) presented a one-compartment nutrikinetic model with first-order excretion, a lag time, and a baseline function was fitted to the time courses of these selected biomarkers based on metabolomic data. A kietic model was also used to model the relationship between enzyme kinetics and intracellular metabolites through a two-substrate Michaelis–Menten equation with competitive substrate inhibition or competitive product inhibition (Schroer et al., 2009). The kinetic constants were estimated by nonlinear regression of initial rate measurements. Martense & Vanrolleghem (2010) summarized a few other modeling approaches. Compared to data driving unsupervised analysis, the mathematical modeling may provide a meaning relationship for a better understanding. However, because the modeling is generally based on approximation under some restricted assumptions, the simple model may not precisely describe the complex biology system.

Artificial network is another method to cope with the nonlinearity. A batch-learning self-organizing network was utilized to classify the metabolomes and the transcriptomes

according to their time-dependent pattern of changes (Kanaya et al., 2001); the results showed that the metabolomes and transcriptomes regulated by the same mechanism tended to be clustered together (Hirai et al., 2004, 2005). The A batch-learning self-organizing network is artificial neural network that is trained using unsupervised learning to produce a low-dimensional representation of the input space of the training samples. A Network-embedded thermodynamic analysis (NET analysis) is presented as a framework for mechanistic and model-based analysis of metabolite data. By coupling the data to an operating metabolic network via the second law of thermodynamics and the metabolites' Gibbs energies of formation, NET analysis allows inferring functional principles from quantitative metabolite data; for example it identifies reactions that are subject to active allosteric or genetic regulation as exemplified with quantitative metabolomic data from *E. coli* and *S. cerevisiae* (Kummel et al., 2006). The network typically creates a graphic representing the global relationship. In a review article, Feist et al. (2006) classified studies using network into three categories: studies that use a reconstruction to examine topological network properties, studies that use a reconstruction in constraint-based modeling for quantitative or qualitative analyses, and studies that are purely data driven. Some of the networks' mathematical frameworks are graph theory. It provides a visual presentation of the complex biology system. However, when it involves more features, the network approach often become too complicated to provide a clear clue.

Bayesian graphical modeling approaches infer biological regulatory networks by integrating expression levels of different types. Specific sequence/structure information will be incorporated into the *prior* probability models (Webb-Robertson et al., 2009) presented a Bayesian approach to integration that uses posterior probabilities to assign class memberships to samples using individual and multiple data sources; these probabilities are based on lower-level likelihood functions derived from standard statistical learning algorithms. The approach was demonstrated by integrating two proteomic datasets and one metabolic dataset from microbial infections of mice; the results showed that integration of the different datasets improved classification accuracy to ~89% from the best individual dataset at ~83 %.

Integrative interpretations of data from different "omics" including metabolomics, are still in it early development stage. More thoughtful interpretation methods that are capable to reveal biology at a system level are yet to come. The collaboration between mathematicians, statisticians, bioinformaticians and experimental biologists will be the key to the success of these efforts.

5. Final remarks

Although comprehensive coverage of metabolome in cells is not yet possible, significant advancements in the large-scale profiling of metabolites have been achieved in recent years and these analyses have offered unique insight into the metabolic and regulatory networks of cells. In this chapter, we first reviewed some of the widely used and emerging technologies for metabolomics analysis, and then focus on recent progress in developing computational methodologies to improve biological interpretation of high throughput metabolomic data. In addition, we present some mathematical, statistical and bioinformatics methods that have been utilized for the integration of metabolomics data with other type of "omics" datasets and how this integrative analysis has improved our interpretation of biological systems.

6. Acknowledgements

7. Disclaimer

Feng Li is currently working for Office of Biostatistics, Food and Drug Administration. Views expressed in this paper are the author's professional opinions and do not necessarily represent the official positions of the U.S. Food and Drug Administration

8. References

Aharoni, A., C. H. R. De Vos, H. A. Verhoeven, C. A. Maliepaard, G. Kruppa, R. Bino & D. B. Goodenowe (2002) Nontargeted metabolome analysis by use of fourier transform ion cyclotron mass spectrometry. *OMICS A Journal of Integrative Biology, 6*, 217-234.

Almaas, E., B. Kovacs, T. Vicsek, Z. N. Oltvai & A. L. Barabasi (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature, 427*, 839-843.

Amantonico, A., P. L. Urban, J. Y. Oh & R. Zenobi (2009) Interfacing microfluidics and laser desorption/Ionization mass spectrometry by continuous deposition for application in single cell analysis. *Chimia, 63*, 185-188.

Amantonico, A., J. Y. Oh, J. Sobek, M. Heinemann & R. Zenobi (2008) Mass spectrometric method for analyzing metabolites in yeast with single cell sensitivity. *Angewandte Chemie-International Edition, 47*, 5382-5385.

Amantonico, A., P. L. Urban & R. Zenobi (2010) Analytical techniques for single-cell metabolomics: state of the art and trends. *Analytical and Bioanalytical Chemistry, 398*, 2493-2504.

Anderle, M., S. Roy, H. Lin, C. Becker & K. Joho (2004) Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatography-mass spectrometry of human serum. *Bioinformatics, 20*, 3575-3582.

Baumgartner, C., C. Bohm, D. Baumgartner, G. Marini, K. Weinberger, B. Olgemoller, B. Liebl & A. A. Roscher. 2004. Supervised machine learning techniques for the classification of metabolic disorders in newborns. In *Bioinformatics*, 2985-96. England.

Beckonert, O., M. E. Bollard, T. M. D. Ebbels, H. C. Keun, H. Antti, E. Holmes, J. C. Lindon & J. K. Nicholson (2003) NMR-based metabonomic toxicity classification: hierarchical cluster analysis and k-nearest-neighbour approaches. *Analytica Chimica Acta, 490*, 3-15.

Beckonert, O., H. C. Keun, T. M. D. Ebbels, J. G. Bundy, E. Holmes, J. C. Lindon & J. K. Nicholson (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols, 2*, 2692-2703.

Bedair, M. & L. W. Sumner (2008) Current and emerging mass-spectrometry technologies for metabolomics. *Trac-Trends in Analytical Chemistry, 27*, 238-250.

Benjamini, Y. & Y. Hochberg (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*, 289-300.

Berg, J., Y. P. Hung & G. Yellen (2009) A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nature Methods, 6*, 161-166.

Birkemeyer, C., A. Kolasa & J. Kopka (2003) Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. *Journal of Chromatography A, 993*, 89-102.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford, UK: Clarendon Press.

Braaksma, M., S. Bijlsma, L. Coulier, P. J. Punt & M. J. van der Werf (2011) Metabolomics as a tool for target identification in strain improvement: the influence of phenotype definition. *Microbiology, 157*, 147-59.

Breitling, R., A. R. Pitt & M. P. Barrett (2006) Precision mapping of the metabolome. *Trends in Biotechnology, 24*, 543-548.

Broadhurst, D. & D. Kell (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics, 2*, 171-196.

Buckmaster, R., F. Asphahani, M. Thein, J. Xu & M. Zhang (2009) Detection of drug-induced cellular changes using confocal Raman spectroscopy on patterned single-cell biosensors. *Analyst, 134*, 1440-1446.

Cahill, G., P. K. Walsh & D. Donnelly (2000) Determination of yeast glycogen content by individual cell spectroscopy using image analysis. *Biotechnology and Bioengineering, 69*, 312-322.

Cannon, D. M., N. Winograd & A. G. Ewing (2000) Quantitative chemical analysis of single cells. *Annual Review of Biophysics and Biomolecular Structure, 29*, 239-263.

Chen, H. W., A. Venter & R. G. Cooks (2006) Extractive electrospray ionization for direct analysis of undiluted urine, milk and other complex mixtures without sample preparation. *Chemical Communications*, 2042-2044.

Ciosek, P., Z. Brzozka, W. Wroblewski, E. Martinelli, C. Di Natale & A. D'Amico. 2005. Direct and two-stage data analysis procedures based on PCA, PLS-DA and ANN for ISE-based electronic tongue-Effect of supervised feature extraction. In *Talanta*, 590-6. England.

Clarke, B. L. 1980. Stability of Complex Reaction Networks. In *Advances in Chemical Physics*, 1-215. John Wiley & Sons, Inc.

--- (1988) Stoichiometric network analysis. *Cell Biophysics, 12*, 237-53.

Cloarec, O., M. E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes & J. Nicholson (2005) Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Analytical Chemistry, 77*, 1282-9.

Cooks, R. G., Z. Ouyang, Z. Takats & J. M. Wiseman (2006) Ambient mass spectrometry. *Science, 311*, 1566-1570.

Cristianini, N. & J. Shawe-Taylor. 2000. *n Introduction to Support Vector Machines*. Cambridge University Press.

Cuperlovic-Culf, M., N. Belacel, A. S. Culf, I. C. Chute, R. J. Ouellette, I. W. Burton, T. K. Karakach & J. A. Walter (2009) NMR metabolic analysis of samples using fuzzy K-means clustering. *Magnetic Resonance in Chemistry, 47* Suppl 1, S96-104.

Daykin, C. A., P. J. D. Foxall, S. C. Connor, J. C. Lindon & J. K. Nicholson (2002) The comparison of plasma deproteinization methods for the detection of low-

molecular-weight metabolites by H-1 nuclear magnetic resonance spectroscopy. *Analytical Biochemistry,* 304, 220-230.

Denery, J., A. Nunes, M. Hixon, T. Dickerson, K. Janda (2010) Metabololics-based discovery of diagnostic biomarkers for onchoceriasis. PLoS Neglected Tropical Diseases, 4(10), e384.

Deng, C. H., X. M. Zhang & N. Li (2004) Investigation of volatile biomarkers in lung cancer blood using solid-phase microextraction and capillary gas chromatography-mass spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 808, 269-277.

Denkert, C., J. Budczies, T. Kind, W. Weichert, P. Tablack, J. Sehouli, S. Niesporek, D., Konsgen, M. Dietel, and O. Fiehn (2006) Mass Spectrometry–Based Metabolic Profiling Reveals Different Metabolite Patterns in Invasive Ovarian Carcinomas and Ovarian Borderline TumorsSpec. Cancer Res, 66: (22).

de Oliveira, L. F. C., R. Colombara & H. G. M. Edwards (2002) Fourier transform Raman spectroscopy of honey. *Applied Spectroscopy,* 56, 306-311.

Dettmer, K., P. A. Aronov & B. D. Hammock (2007) Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews,* 26, 51-78.

De Vos, R. C. H., S. Moco, A. Lommen, J. J. B. Keurentjes, R. J. Bino & R. D. Hall (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols,* 2, 778-791.

Dieterle, F., A. Ross, G. Schlotterbeck & H. Senn (2006) Metabolite projection analysis for fast identification of metabolites in metabonomics. Application in an amiodarone study. *Analytical Chemistry,* 78, 3551-61.

Dunn, W. B., N. J. C. Bailey & H. E. Johnson (2005) Measuring the metabolome: current analytical technologies. *Analyst,* 130, 606-625.

Dunn, W. B. & D. I. Ellis (2005) Metabolomics: current analytical platforms and methodologies. *Trac-Trends in Analytical Chemistry,* 24, 285-294.

Ebbels, T. M. D. 2007. Non-linear Methods for the Analysis of Metabolic Profiles. In *The Handbook of Metabonomics and Metabolomics,* eds. J. C. Lindon, J. K. Nicholson & E. Holmes, 201--226. London: Elsevier.

Edwards, J. S. & B. O. Palsson. 2000. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. In *Proceedings of the National Academy of Sciences of the United States of America*, 5528-33. United States.

Efron, B. (2003) Robbins, Empirical Bayes and Microarrays. *The Annals of Statistics,* 31, 366-378.

--- (2004a) Large-Scale Simultaneous Hypothesis Testing. *Journal of the American Statistical Association,* 99, 96-104.

--- (2004b) Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association,* 99, 96-104.

--- (2007a) Correlation and Large-Scale Simultaneous Significance Testing. *Journal of the American Statistical Association,* 102, 93-103.

--- (2007b) Size, power and false discovery rates. *The Annals of Statistics,* 35, 1351-1377.

--- (2008) Microarrays, Empirical Bayes and the Two-Groups Model. *Statistical Science,* 23, 1-22.

Enot, D. P., B. Haas & K. M. Weinberger (2011) Bioinformatics for mass spectrometry-based metabolomics. *Methods in Molecular Biology,* 719, 351-75.

Enot, D. P., W. Lin, M. Beckmann, D. Parker, D. P. Overy & J. Draper (2008) Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nature Protocols,* 3, 446-70.

Famili, I., J. Forster, J. Nielsen & B. O. Palsson. 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. In *Proceedings of the National Academy of Sciences of the United States of America*, 13134-9. United States.

Fan, J. Q. & R. Z. Li (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association,* 96, 1348-1360.

Farag, M. A., D. V. Huhman, R. A. Dixon & L. W. Sumner (2008) Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in Medicago truncatula cell cultures. *Plant Physiology,* 146, 387-402.

Fehr, M., W. B. Frommer & S. Lalonde (2002) Visualization of maltose uptake in living yeast cells by fluorescent nanosensors. *Proceedings of the National Academy of Sciences of the United States of America,* 99, 9846-9851.

Feist, A. M., J. C. Scholten, B. O. Palsson, F. J. Brockman & T. Ideker (2006) Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Molecular Systems Biology,* 2, 2006 0004.

Fell, D. 1997. *Understanding the Control of Metabolism*. Potland Press.

Fiehn, O. (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics,* 2, 155-168.

Fiehn, O. 2007. Cellular Metabolomics: The Quest for Pathway Structure. In *The Handbook of Metabonomics and Metabolomics,* eds. J. C. Lindon, J. K. Nicholson & E. Holmes, 35-54. Elsevier.

Fletcher, J. S. (2009) Cellular imaging with secondary ion mass spectrometry. *Analyst,* 134, 2204-2215.

Fliermans, C. B. & E. L. Schmidt (1975) Authoradiography and immunofluorescence combined for autecological study of single cell activity with nitrobacter as a model system. *Applied Microbiology,* 30, 676-684.

Fraser, P. D., E. M. A. Enfissi, M. Goodfellow, T. Eguchi & P. M. Bramley (2007) Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Plant Journal,* 49, 552-564.

Furey, T. S., N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer & D. Haussler (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics,* 16, 906-914.

Garthwaite, P. H. (1994) An Interpretation of Partial Least-Squares. *Journal of the American Statistical Association,* 89, 122-127.

Gmati, D., J. K. Chen & M. Jolicoeur (2005) Development of a small-scale bioreactor: Application to *in vivo* NMR measurement. *Biotechnology and Bioengineering,* 89, 138-147.

Goff, K. L., L. Quaroni & K. E. Wilson (2009) Measurement of metabolite formation in single living cells of *Chlamydomonas reinhardtii* using synchrotron Fourier-Transform Infrared spectromicroscopy. *Analyst,* 134, 2216-2219.

Graf, T. & M. Stadtfeld (2008) Heterogeneity of Embryonic and Adult Stem Cells. *Cell Stem Cell,* 3, 480-483.

Grivet, J.-P. & A.-M. Delort (2009) NMR for microbiology: In vivo and in situ applications. *Progress in Nuclear Magnetic Resonance Spectroscopy,* 54, 1-53.

Gu, H., H. Chen, Z. Pan, A. U. Jackson, N. Talaty, B. Xi, C. Kissinger, C. Duda, D. Mann, D. Raftery & R. G. Cooks (2007) Monitoring diet effects via biofluids and their implications for metabolomics studies. *Analytical Chemistry,* 79, 89-97.

Guyon, I., J. Weston, S. Barnhill & V. Vapnik (2002) Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning,* 46, 389-422.

Gygi, S. P., Y. Rochon, B. R. Franza & R. Aebersold (1999) Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology,* 19, 1720-30.

Hageman, J. A., R. A. van den Berg, J. A. Westerhuis, H. C. J. Hoefsloot & A. K. Smilde (2006) Bagged K-Means Clustering of Metabolome Data. *Critical Reviews in Analytical Chemistry,* 36, 211-220.

Heinemann, M. & R. Zenobi (2011) Single cell metabolomics. *Current Opinion in Biotechnology.* 22, 26-31.

Heinrich, R. & S. Schuster. 1996. *The Regulation of Cellular Systems.* Chapman and Hall.

Hermelink, A., A. Brauer, P. Lasch & D. Naumann (2009) Phenotypic heterogeneity within microbial populations at the single-cell level investigated by confocal Raman microspectroscopy. *Analyst,* 134, 1149-1153.

Hirai, M. Y., M. Klein, Y. Fujikawa, M. Yano, D. B. Goodenowe, Y. Yamazaki, S. Kanaya, Y. Nakamura, M. Kitayama, H. Suzuki, N. Sakurai, D. Shibata, J. Tokuhisa, M. Reichelt, J. Gershenzon, J. Papenbrock & K. Saito (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. *Journal of Biological Chemistry,* 280, 25590-5.

Hirai, M. Y., M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara & K. Saito (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 10205-10.

Hollywood, K., D. R. Brison & R. Goodacre (2006) Metabolomics: Current technologies and future trends. *Proteomics,* 6, 4716-4723.

Holmes, D., D. Pettigrew, C. H. Reccius, J. D. Gwyer, C. van Berkel, J. Holloway, D. E. Davies & H. Morgan (2009) Leukocyte analysis and differentiation using high speed microfluidic single cell impedance cytometry. *Lab on a Chip,* 9, 2881-2889.

Hoops, S., S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes & U. Kummer. 2006. COPASI--a COmplex PAthway SImulator. In *Bioinformatics*, 3067-74. England.

Hsieh, Y., R. Casale, E. Fukuda, J. W. Chen, I. Knemeyer, J. Wingate, R. Morrison & W. Korfmacher (2006) Matrix-assisted laser desorption/ionization imaging mass spectrometry for direct measurement of clozapine in rat brain tissue. *Rapid Communications in Mass Spectrometry,* 20, 965-972.

Huang, L. & R. T. Kennedy (1995) Exploring single-cell dynamics using chemically-modified microelectrodes. *Trends in Analytical Chemistry,* 14, 158-164.

Irish, J. M., N. Kotecha & G. P. Nolan (2006) Innovation - Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nature Reviews Cancer,* 6, 146-155.

Ishii, N. & M. Tomita. 2009. *Multi-Omics Data-Driven Systems Biology of E. coli.*

Jones, J. J., S. Borgmann, C. L. Wilkins & R. M. O'Brien (2006) Characterizing the phospholipid profiles in mammalian tissues by MALDI FTMS. *Analytical Chemistry,* 78, 3062-3071.

Jewett, M., M. Hansen & J. Nielsen. 2007a. Data acquisition, analysis, and mining: Integrative tools for discerning. In *Metabolomics: A Powerful Tool in Systems Biology,* eds. J. Nielsen & M. C. Jewett, 159-187. Springer.

---. 2007b. Data acquisition, analysis, and mining: Integrative tools for discerning. In *Metabolomics, A Powerful Tool in Systems Biology,* eds. J. Nielsen & M. C. Jewett, 159-187. Springer.

Jonsson, P., S. J. Bruce, T. Moritz, J. Trygg, M. Sjostrom, R. Plumb, J. Granger, E. Maibaum, J. K. Nicholson, E. Holmes & H. Antti (2005) Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst,* 130, 701-7.

Kaartinen, J., Y. Hiltunen, P. T. Kovanen & M. Ala-Korpela (1998) Application of self-organizing maps for the detection and classification of human blood plasma lipoprotein lipid profiles on the basis of 1H NMR spectroscopy data. *NMR in Biomedicine,* 11, 168-176.

Kandpal, R. P., B. Saviola & J. Felton (2009) The era of 'omics unlimited. *Biotechniques,* 46, 351.

Kanaya, S., M. Kinouchi, T. Abe, Y. Kudo, Y. Yamada, T. Nishi, H. Mori & T. Ikemura (2001) Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene,* 276, 89-99.

Khatib-Shahidi, S., M. Andersson, J. L. Herman, T. A. Gillespie & R. M. Caprioli (2006) Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. *Analytical Chemistry,* 78, 6448-6456.

Kind, T. & O. Fiehn. 2006. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. In *BMC Bioinformatics*, 234 7:234. England.

Klamt, S. & E. D. Gilles (2004) Minimal cut sets in biochemical reaction networks. *Bioinformatics,* 20, 226-34.

Klamt, S., J. Saez-Rodriguez & E. D. Gilles (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology,* 1, 2.

Klamt, S., J. Saez-Rodriguez, J. A. Lindquist, L. Simeoni & E. D. Gilles (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics, 7,* 56.

Klamt, S. & S. Schuster (2002) Calculating as many fluxes as possible in underdetermined metabolic networks. *Mol Biol Rep,* 29, 243-8.

Klamt, S. & J. Stelling (2003) Two approaches for metabolic pathway analysis? *Trends in Biotechnology,* 21, 64-9.

Kohonen, T. (1998) The self-organizing map. *Neurocomputing,* 21, 1-6.

---. 2001. *Self-Organizing Maps*. Springer.

Kouskoumvekaki, I., Z. Yang, S. O. Jonsdottir, L. Olsson & G. Panagiotou. 2008. Identification of biomarkers for genotyping Aspergilli using non-linear methods for clustering and classification. In *BMC Bioinformatics*, 59. England.

Kummel, A., S. Panke & M. Heinemann (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Molecular Systems Biology,* 2, 2006 0034.

Le Cao, K. A., P. G. Martin, C. Robert-Granie & P. Besse (2009) Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics,* 10, 34.

Le Cao, K. A., D. Rossouw, C. Robert-Granie & P. Besse (2008) A sparse PLS for variable selection when integrating omics data. *Statistical and Applied Genetics and Molecular Biology,* 7, Article 35.

Lenz, E. M. & I. D. Wilson (2007) Analytical strategies in metabonomics. *Journal of Proteome Research,* 6, 443-458.

Lei, F. & S. B. Jorgensen. 2001. Estimation of kinetic parameters in a structured yeast model using regularisation. In *Journal of Biotechnology*, 223-37. Netherlands.

Lerouxel, O., T. S. Choo, M. Seveno, B. Usadel, L. Faye, P. Lerouge & M. Pauly (2002) Rapid structural phenotyping of plant cell wall mutants by enzymatic oligosaccharide fingerprinting. *Plant Physiology,* 130, 1754-1763.

Li, X., X. Lu, J. Tian, P. Gao, H. Kong & G. Xu (2009) Application of fuzzy c-means clustering in data analysis of metabolomics. *Analytical Chemistry,* 81, 4468-75.

Lin, J. & J. Qian (2007) Systems biology approach to integrative comparative genomics. *Expert Review of Proteomics,* 4, 107-119.

Lindon, J. C., E. Holmes & J. K. Nicholson (2003) So whats the deal with metabonomics? Metabonomics measures the fingerprint of biochemical perturbations caused by disease, drugs, and toxins. *Analytical Chemistry,* 75, 384A-391A.

Lindon, J. C., J. K. Nicholson & I. D. Wilson (2000) Directly coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. *Journal of Chromatography B,* 748, 233-258.

Listgarten, J. & A. Emili. 2005. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. In *Molecular and Cellular Proteomics*, 419-34. United States.

Lu, K.-Y., A. M. Wo, Y.-J. Lo, K.-C. Chen, C.-M. Lin & C.-R. Yang (2006) Three dimensional electrode array for cell lysis via electroporation. *Biosensors & Bioelectronics,* 22, 568-574.

Mahadevan, S., S. L. Shah, T. J. Marrie & C. M. Slupsky (2008) Analysis of metabolomic data using support vector machines. *Analytical Chemistry,* 80, 7562-70.

Mallouchos, A., M. Komaitis, A. Koutinas & M. Kanellaki (2002) Investigation of volatiles evolution during the alcoholic fermentation of grape must using free and immobilized cells with the help of solid phase microextraction (SPME) headspace sampling. *Journal of Agricultural and Food Chemistry,* 50, 3840-3848.

Mas, S., S. G. Villas-Boas, M. E. Hansen, M. Akesson & J. Nielsen (2007) A comparison of direct infusion MS and GC-MS for metabolic footprinting of yeast mutants. *Biotechnology and Bioengineering,* 96, 1014-1022.

Masujima, T. (2009) Live single-cell mass spectrometry. *Analytical Sciences,* 25, 953-960.

Mattoli, L., F. Cangi, A. Maidecchi, C. Ghiara, E. Ragazzi, M. Tubaro, L. Stella, F. Tisato & P. Traldi (2006) Metabolomic fingerprinting of plant extracts. *Journal of Mass Spectrometry,* 41, 1534-1545.

Maxwell, R. J., I. Martinez-Perez, S. Cerdan, M. E. Cabanas, C. Arus, A. Moreno, A. Capdevila, E. Ferrer, F. Bartomeus, A. Aparicio, G. Conesa, J. M. Roda, F. Carceller, J. M. Pascual, S. L. Howells, R. Mazucco & J. R. Griffiths (1998) Pattern recognition analysis of 1H NMR spectra from perchloric acid extracts of human brain tumor biopsies. *Magnetic Resonance in Medicine,* 39, 869-77.

Mellors, J. S., V. Gorbounov, R. S. Ramsey & J. M. Ramsey (2008) Fully integrated glass microfluidic device for performing high-efficiency capillary electrophoresis and electrospray ionization mass spectrometry. *Analytical Chemistry,* 80, 6881-6887.

Meija, J. (2006) Mathematical tools in analytical mass spectrometry. *Analytical and Bioanalytical Chemistry*, 486-499.

Meinicke, P., T. Lingner, A. Kaever, K. Feussner, C. Gobel, I. Feussner, P. Karlovsky & B. Morgenstern. 2008. Metabolite-based clustering and visualization of mass spectrometry data using one-dimensional self-organizing maps. In *Algorithms Mol Biol*, 9. England.

Moles, C. G., P. Mendes & J. R. Banga (2003) Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. *Genome Research,* 13, 2467-2474.

Monton, M. R. N. & T. Soga (2007) Metabolome analysis by capillary electrophoresis-mass spectrometry. *Journal of Chromatography A,* 1168, 237-246.

Motta, A., D. Paris & D. Melck (2010) Monitoring real-time metabolism of living cells by fast two-dimensional NMR spectroscopy. *Analytical Chemistry,* 82, 2405-2411.

Mylonas, R., Y. Mauron, A. Masselot, P. A. Binz, N. Budin, M. Fathi, V. Viette, D. F. Hochstrasser & F. Lisacek (2009) X-Rank: a robust algorithm for small molecule identification using tandem mass spectrometry. *Analytical Chemistry,* 81, 7604-10.

Neumann, S. & S. Bocker (2010) Computational mass spectrometry for metabolomics: Identification of metabolites and small molecules. *Analytical and Bioanalytical Chemistry*, 2779-2788.

Noteborn, H., A. Lommen, R. C. van der Jagt & J. M. Weseman (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *Journal of Biotechnology,* 77, 103-114.

Oberacher, H., M. Pavlic, K. Libiseller, B. Schubert, M. Sulyok, R. Schuhmacher, E. Csaszar & H. C. Kofeler (2009) On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library: 2. Optimization and characterization of the search algorithm. *Journal of Mass Spectrometry,* 44, 494-502.

Ott, K. H., N. Aranibar, B. Singh & G. W. Stockton. 2003. Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts. In *Phytochemistry*, 971-85. United States.

Papin, J. A., J. Stelling, N. D. Price, S. Klamt, S. Schuster & B. O. Palsson. 2004. Comparison of network-based pathway analysis methods. In *Trends in Biotechnology*, 400-5. England.

Pir, P., B. Kirdar, A. Hayes, Z. Y. Onsan, K. O. Ulgen & S. G. Oliver (2006) Integrative investigation of metabolic and transcriptomic data. *BMC Bioinformatics,* 7, 203.

Podwojski, K., A. Fritsch, D. C. Chamrad, W. Paul, B. Sitek, P. Mutzel, C. Stephan, H. E. Meyer, W. Urfer, K. Ickstadt & J. Rahnenführer (2009) Retention Time Alignment Algorithms for LC/MS Data must consider Nonlinear Shifts. *Bioinformatics*. 25: 758-64.

Pohjanen, E., E. Thysell, J. Lindberg, I. Schuppe-Koistinen, T. Moritz, P. Jonsson & H. Antti (2006) Statistical multivariate metabolite profiling for aiding biomarker pattern detection and mechanistic interpretations in GC/MS based metabolomics. *Metabolomics,* 2, 257-268-268.

Price, N. D., J. A. Papin, C. H. Schilling & B. O. Palsson (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends in Biotechnology,* 21, 162-9.

Raamsdonk, L. M., B. Teusink, D. Broadhurst, N. S. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam & S. G. Oliver (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology,* 19, 45-50.

Ramautar, R., A. Demirci & G. J. de Jong (2006) Capillary electrophoresis in metabolomics. *Trac-Trends in Analytical Chemistry,* 25, 455-466.

Rashed, M. S., M. P. Bucknall, D. Little, A. Awad, M. Jacob, M. Alamoudi, M. Alwattar & P. T. Ozand (1997) Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles. *Clinical Chemistry,* 43, 1129-1141.

Reiner, A., D. Yekutieli & Y. Benjamini (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics,* 19, 368-75.

Reyzer, M. L. & R. M. Caprioli (2007) MALDI-MS-based imaging of small molecules and proteins in tissues. *Current Opinion in Chemical Biology,* 11, 29-35.

Roessner, U., C. Wagner, J. Kopka, R. N. Trethewey & L. Willmitzer (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant Journal,* 23, 131-142.

Roessner-Tunali, U., B. Hegemann, A. Lytovchenko, F. Carrari, C. Bruedigam, D. Granot & A. R. Fernie (2003) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. *Plant Physiology,* 133, 84-99.

Rubakhin, S. S., J. C. Jurchen, E. B. Monroe & J. V. Sweedler (2005) Imaging mass spectrometry: fundamentals and applications to drug discovery. *Drug Discovery Today,* 10, 823-837.

Rubingh, C. M., S. Bijlsma, R. H. Jellema, K. M. Overkamp, M. J. van der Werf & A. K. Smilde (2009) Analyzing longitudinal microbial metabolomics data. *Journal of Proteome Research,* 8, 4319-27.

Sato, S., T. Soga, T. Nishioka & M. Tomita (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *Plant Journal,* 40, 151-163.

Schmid, A., H. Kortmann, P. S. Dittrich & L. M. Blank (2010) Chemical and biological single cell analysis. *Current Opinion in Biotechnology,* 21, 12-20.

Schilling, C. H., J. S. Edwards, D. Letscher & B. Palsson (2000a) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology & Bioengineering,* 71, 286-306.

Schilling, C. H., D. Letscher & B. O. Palsson (2000b) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology,* 203, 229-48.

Schilling, C. H., S. Schuster, B. O. Palsson & R. Heinrich (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress,* 15, 296-303.

Schrader, B., H. H. Klump, K. Schenzel & H. Schulz (1999) Non-destructive NIR FT Raman analysis of plants. *Journal of Molecular Structure,* 509, 201-212.

Schroer, K., B. Zelic, M. Oldiges & S. Lutz (2009) Metabolomics for biotransformations: Intracellular redox cofactor analysis and enzyme kinetics offer insight into whole cell processes. *Biotechnology & Bioengineering,* 104, 251-60.

Schuster, K. C., I. Reese, E. Urlaub, J. R. Gapes & B. Lendl (2000) Multidimensional information on the chemical composition of single bacterial cells by confocal Raman microspectroscopy. *Analytical Chemistry,* 72, 5529-5534.

Schuster, S., T. Pfeiffer, F. Moldenhauer, I. Koch & T. Dandekar (2002) Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae. *Bioinformatics,* 18, 351-361.

Seger, C. & S. Sturm (2007) Analytical aspects of plant metabolite profiling platforms: Current standings and future aims. *Journal of Proteome Research,* 6, 480-497.

Segre, D., J. Zucker, J. Katz, X. Lin, P. D'Haeseleer, W. P. Rindone, P. Kharchenko, D. H. Nguyen, M. A. Wright & G. M. Church (2003) From annotated genomes to metabolic flux models and kinetic parameter fitting. *OMICS,* 7, 301-16.

Shinohara, H. & F. Wang (2007) Real-time detection of dopamine released from a nerve model cell by an enzyme-catalyzed luminescence method and its application to drug assessment. *Analytical Sciences,* 23, 81-84.

Singh, O. V. & N. S. Nagaraj (2006) Transcriptomics, proteomics and interactomics: unique approaches to track the insights of bioremediation. *Briefings in Functional Genomics & Proteomics,* 4, 355-362.

Soga, T., Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita & T. Nishioka (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research,* 2, 488-494.

Southam, A. D., T. G. Payne, H. J. Cooper, T. N. Arvanitis & M. R. Viant (2007) Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Analytical Chemistry,* 79, 4595-4602.

Stein, S. E. (1994) Estimating probabilities of correct identification from results of mass spectral library searches. *Journal of the American Society for Mass Spectrometry,* 5, 316-323.

Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster & E. D. Gilles (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature,* 420, 190-193.

Stephanopoulos, G., H. Alper & J. Moxley (2004) Exploiting biological complexity for strain improvement through systems biology. *Nature Biotechnology,* 22, 1261-1267.

Steuer, R. 2006. Review: on the analysis and interpretation of correlations in metabolomic data. In *Briefing in Bioinformatics*, 151-8. England.

---. 2007. Computational approaches to the topology, stability and dynamics of metabolic networks. In *Phytochemistry*, 2139-51. United States.

Storey, J. D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 64, 479-498.

Storey, J. D. & R. Tibshirani. 2003. Statistical significance for genomewide studies. In *Proceedings of the National Academy of Sciences of the United States of America*, 9440-5. United States.

Strelkov, S., M. von Elstermann & D. Schomburg (2004) Comprehensive analysis of metabolites in *Corynebacterium glutamicum* by gas chromatography/mass spectrometry. *Biological Chemistry,* 385, 853-861.

Sturm, S., C. Seger & H. Stuppner (2007) Analysis of Central European Corydalis species by nonaqueous capillary electrophoresis-electrospray ion trap mass spectrometry. *Journal of Chromatography A,* 1159, 42-50.

Suits, F., J. Lepre, P. Du, R. Bischoff & P. Horvatovich (2008) Two-dimensional method for time aligning liquid chromatography-mass spectrometry data. *Analytical Chemistry,* 80, 3095-104.

Sumner, L., A. Amberg, D. Barrett, M. Beale, R. Beger, C. Daykin, T. Fan, O. Fiehn, R. Goodacre, J. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. Lane, J. Lindon, P. Marriott, A. Nicholls, M. Reily, J. Thaden & M. Viant (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 211-221.

Sysi-Aho, M., M. Katajamaa, L. Yetukuri & M. Oresic. 2007. Normalization method for metabolomics data using optimal selection of multiple internal standards. In *BMC Bioinformatics*, 8: 93. England.

Takats, Z., J. M. Wiseman, B. Gologan & R. G. Cooks (2004) Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science,* 306, 471-473.

Talaty, N., Z. Takats & R. G. Cooks (2005) Rapid in situ detection of alkaloids in plant tissue under ambient conditions using desorption electrospray ionization. *Analyst,* 130, 1624-1633.

Tanaka, K., A. Westdull, D. G. Hine, T. B. Lynn & T. Lowe (1980) Gas-chromategraphic method of analysis for urinary organic-acids. 2. Description of the procedure, and its application to diagnosis of patients with organic acidurias. *Clinical Chemistry,* 26, 1847-1853.

Tapp, H. S. & E. K. Kemsley (2009) Notes on the practical utility of OPLS. *TrAC Trends in Analytical Chemistry,* 28, 1322-1327.

Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological,* 58, 267-288.

Trygg, J. & T. Lundstedt. 2007. Chemometrics Techniques for Metabonomics. In *The Handbook of Metabonomics and Metabolomics,* eds. J. C. Lindon, J. K. Nicholson & E. Holmes. London: Elseivier.

Urban, P. L., K. Jefimovs, A. Amantonico, S. R. Fagerer, T. Schmid, S. Maedler, J. Puigmarti-Luis, N. Goedecke & R. Zenobi (2010) High-density micro-arrays for mass spectrometry. *Lab on a Chip,* 10, 3206-3209.

Urbanczik, R. & C. Wagner (2005) An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics,* 21, 1203-10.

Urbanczyk-Wochniak, E., A. Luedemann, J. Kopka, J. Selbig, U. Roessner-Tunali, L. Willmitzer & A. R. Fernie (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Report,* 4, 989-93.

Urbanczyk-Wochniak, E., L. Willmitzer & A. R. Fernie (2007) Integrating profiling data: using linear correlation to reveal coregulation of transcript and metabolites. *Methods in Molecular Biology,* 358, 77-85.

van den Berg, R. A., C. M. Rubingh, J. A. Westerhuis, M. J. van der Werf & A. K. Smilde (2009) Metabolomics data exploration guided by prior knowledge. *Anal Chim Acta,* 651, 173-81.

van Nederkassel, A. M., M. Daszykowski, P. H. Eilers & Y. V. Heyden. 2006. A comparison of three algorithms for chromatograms alignment. In *J Chromatogr A*, 199-210. Netherlands.

van Velzen, E. J., J. A. Westerhuis, J. P. van Duynhoven, F. A. van Dorsten, C. H. Grun, D. M. Jacobs, G. S. Duchateau, D. J. Vis & A. K. Smilde (2009) Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. *Journal of Proteome Research,* 8, 3317-30.

Vandenbogaert, M., S. Li-Thiao-Te, H. M. Kaltenbach, R. Zhang, T. Aittokallio & B. Schwikowski (2008) Alignment of LC-MS images, with applications to biomarker discovery and protein identification. *Proteomics*, 8, 650-72.

Varma, A. & B. O. Palsson (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nature Biotechnology*, 12, 994-998.

Villas-Boas, S. G., S. Mas, M. Akesson, J. Smedsgaard & J. Nielsen (2005) Mass spectrometry in metabolome analysis. *Mass Spectrometry Reviews*, 24, 613-646.

Visser, D. & J. J. Heijnen. 2003. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. In *Metabolic Engineering*, 164-76. United States.

Wagner, C., M. Sefkow & J. Kopka (2003) Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry*, 62, 887-900.

Wang, D. & S. Bodovitz (2010) Single cell analysis: the new frontier in 'omics'. *Trends in Biotechnology*, 28, 281-290.

Warrack, B. M., S. Hnatyshyn, K. H. Ott, M. D. Reily, M. Sanders, H. Zhang & D. M. Drexler. 2009. Normalization strategies for metabonomic analysis of urine samples. In *J Chromatogr B Analyt Technol Biomed Life Sci*, 547-52. Netherlands.

Webb-Robertson, B. J., L. A. McCue, N. Beagley, J. E. McDermott, D. S. Wunschel, S. M. Varnum, J. Z. Hu, N. G. Isern, G. W. Buchko, K. McAteer, J. G. Pounds, S. J. Skerrett, D. Liggitt & C. W. Frevert (2009) A Bayesian integration model of high-throughput proteomics and metabolomics data for improved early detection of microbial infections. *Pac Symp Biocomput*, 451-63.

Weckwerth, W., M. E. Loureiro, K. Wenzel & O. Fiehn (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 7809-7814.

Weckwerth, W., K. Wenzel & O. Fiehn (2004) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics*, 4, 78-83.

Wiseman, J. M., S. M. Puolitaival, Z. Takats, R. G. Cooks & R. M. Caprioli (2005) Mass spectrometric profiling of intact biological tissue by using desorption electrospray ionization. *Angewandte Chemie-International Edition*, 44, 7094-7097.

Wittmann, C., J. O. Kromer, P. Kiefer, T. Binz & E. Heinzle (2004) Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Analytical Biochemistry*, 327, 135-139.

Xie, Y., W. Pan & A. B. Khodursky (2005) A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics*, 21, 4280-4288.

Yassaa, N., E. Brancaleoni, M. Frattoni & P. Ciccioli (2001) Trace level determination of enantiomeric monoterpenes in terrestrial plant emission and in the atmosphere using a beta-cyclodextrin capillary column coupled with thermal desorption and mass spectrometry. *Journal of Chromatography A*, 915, 185-197.

Zhang, J., E. Gonzalez, T. Hestilow, W. Haskins & Y. Huang (2009) Review of peak detection algorithms in liquid-chromatography-mass spectrometry. *Current Genomics*, 10, 388-401.

Zhang, W., F. Li & L. Nie (2010) Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology*, 156, 287-301.
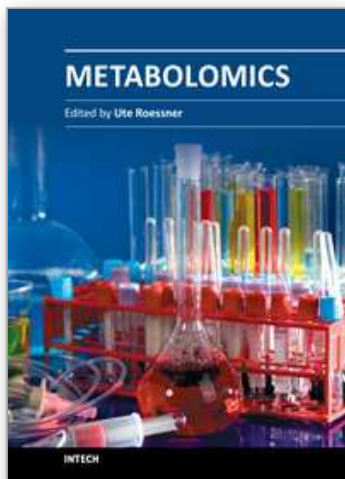
Metabolomics

Edited by Dr Ute Roessner

Metabolomics is a rapidly emerging field in life sciences, which aims to identify and quantify metabolites in a biological system. Analytical chemistry is combined with sophisticated informatics and statistics tools to determine and understand metabolic changes upon genetic or environmental perturbations. Together with other 'omics analyses, such as genomics and proteomics, metabolomics plays an important role in functional genomics and systems biology studies in any biological science. This book will provide the reader with summaries of the state-of-the-art of technologies and methodologies, especially in the data analysis and interpretation approaches, as well as give insights into exciting applications of metabolomics in human health studies, safety assessments, and plant and microbial research.

INTECH

open science | open minds