Recent developments and application of metabolomics in cancer diseases

Afsaneh Arefi Oskouie*[,1], Salman Taheri[2]

[1]Faculty of Paramedical Sciences, Department of Basic Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Chemistry and Chemical Engineering Research Center of Iran, Tehran, Iran.

*Corresponding Author: email address: a.arefi@sbmu.ac.ir (A.Arefi Oskouie)

ABSTRACT

Metabolomics studies provide useful information about health and disease status. Metabolite based investigations on various cancers is a powerful approach to diagnosis, prognosis and therapy of cancer diseases. Recently by using advanced analytical techniques such as NMR and MS and its hyphenation methods, global metabolic profiling of diseases has been possible. It is predictable that international contributions and software developments in the future will lead to accurate instrumental analysis based on a large number of human samples that finally will improve validation methods and reach this field from the research phase to the clinical phase. In this review, we also discussed the latest developments in analytical methods, application of data analysis, investigation of useful databases and the curent application of metabolomics in cancer diseases that have led to the identification of related biomarkers. In continuation, we listed biomarkers involved in cancer diseases that have been published during recent years.

Keywords: Metabolomics; Biomarker; Cancer; Database.

Abbreviations: NMR: Nuclear Magnetic Resonance, MS: Mass Spectrometry, HMDB: Human Metabolome Database, RF: radio frequency ; EBC: Exhaled breath condensate ; CSF: cerebro-spinal fluid; AF: Amniotic fluid; JRES: J-resolved ; CPMG: Carr-Purcell-Meiboom-Gill ; TMS: Trimethylsilan ; TSP: 3-trimethyl silyl propionic acid; DSS: 2, 2-dimethyl-2-sila pentane-5-sulfonate sodium salt ; MAS-NMR: magic angle spinning Nuclear Magnetic Resonance; HR-MAS: High Resolution Magic Angle Spinning ; GC-MS: Gas chromatography-mass spectrometry; EI: electron impact; CI: chemical ionization ; MSTFA: N-methyl-trimethyl silyltriflouroacetamide ;GC-TOF-MS: Gas chromatography time of flight -mass spectrometry; LC-MS: Liquid chromatography mass spectrometry; RP-HPLC: reversed phase high performance liquid chromatography; ESI: electrospray ionization; MALDI: Matrix Assisted laser desorption/ionization ; DESI: Desorption electrospray ionization Mass Spectrometry; EESI: Extractive electrospray ionization ; UHPLC: Ultra high-performance liquid chromatography CE-MS: Capillary Electrophorese-mass spectrometry; MVA: Multivariate analysis; PCA: principle component analysis ; PCs: principal components; STOCSY : statistical correlation spectroscopy; CLASSY: cluster analysis statistical spectroscopy; PLS: partial least square ; DFA: discriminant function analysis ; PCR: Principal Component Regression ;PLS-DA: partial least square discriminant analysis; OPLS: Orthogonal partial least square; ANN: Artificial neural network ; ROC: receiver-operator characteristic ; OSCC: Oral squamous cell carcinoma ; HNSCC: Head and neck squamous cell carcinoma ; lysoPCs: Lysophosphatidylcholines .

INTRODUCTION

In the past few years, great changes have taken place in the biological sciences leading into the emergence field of systems biology. The metabolomics and metabonomics terms were used in the late nineteenth and early twentieth-century [1, 2]. Metabolomics is one of the important building blocks of systems biology, complementary of other "omic" sciences and the closest correlation to phenotype. On the other hand, metabolites are the reflection of physiological and disease status, then study of

metabolites facilitates understanding biochemical pathways. Metabolome is the collection of small molecules in a mass below approximately 1500 Dalton (metabolites) that produced by a cell as a result of response to environmental stresses[3] . Metabonomics were defined by Nicolson as "quantitative measurement of dynamic multi parametric metabolic response of living systems to pathophysiological stimuli or genetic modification"[4]. Metabolomics were defined as "study of the complete set of metabolites/low molecular intermediate, which are contexts dependent, varying according to the physiological, developmental or pathological state of cell, tissue and organ or organism" [3]. These two terminologies often use instead of each other.
Metabolites are divided into two categories; Endogenous and Exogenous. Endogenous metabolites, are the collection of chemicals that produce and consume by organism in the absence of extracellular materials and exogenous metabolites are foreign substances such as nutrients, drugs, xenobiotic and environmental change [5].
Analysis of metabolism has different ways; target analysis and integrated analysis. Target analysis deals with the measurement of specific metabolites or focuses on one or more metabolic pathways. [6]. Untargeted analysis is simultaneous measurement of metabolites without bias It must be noted that untargeted metabolomics somewhere refer metabolic profiling [7]. Untargeted analysis has two subsets: metabolic fingerprinting and metabolic footprinting. Metabolic Fingerprinting is a measurement of collection of metabolites without any previous assumption [8], although it is impossible to evaluate all present metabolites, but with recent advances in analytical instruments and Chemometric techniques can be closer to this goal. Metabolic Footprinting is defined as laboratory culture media that measurement of changes that caused by an organism. Another definition of Metabolic Footprinting is secreted metabolites that caused by an organism and often apply in biotechnology and microbiology [9].

General workflow of metabolic studies can be categorized into five steps: 1) Sample preparation 2) Analytical platforms based on NMR or MS spectroscopy, 3) pre-processing 4) data analysis (*via* Chemometrics methods) and 5) metabolite identification and data interpretation

Biological samples

Metabolites are generated by the processes of metabolism in cells, tissue or organ. The human body contains approximately 38000 (exactly 37,166, until now, according to a very recent report on the HMDB Version 3.6) detectable metabolites that are very diverse chemical compounds such as volatile, polar and more polar metabolites (Table 1)

Table 1. classes of chemical compounds

Volatile	**Polar**	**More polar**
Hydrocarbons, Alcohols	Eicosanoids, Steroids	Organic acids
Essential oils, Terpenoids	Esters, Amino Acids,	Organic amines
Aldehydes, Ketones	Hydrophilic Carbohydrate	Nucleosides
Hydrophobic lipids	Carotenoids, Alkaloids	Ionic species
Isocyanates, Isothiocyanates	Flavonoids/ Phenols	Nucleotides
Amines, Pyrols, Sulfides	Lipids, Fatty acids, Catecholamines	Polyamines

Because of difference in polarity of compounds, it is impossible to measure all of metabolites with a unique technics. Factors that affecting metabolomics analysis is gender [10], age [10], lifestyle [10], smoky [11]and physical activity [12]. One of the most important steps in metabolic approach is the selection of the sample and its preparation before instrumental analysis. The required analytical technique varies depending on types of sample. The common samples for metabolic analysis are blood (serum/plasma) [15], urine, , saliva, tissue, cell line [13], exhaled cerebro-spinal fluid(CSF) breath and amniotic fluids [14, 15]. Each of these samples has advantages and limitations of the analysis that summarized in Table 2.

Analytical Tools

A variety of analytical technologies have been applied to metabolomics studies. The most common analytical instrument that have been employed are nuclear magnetic resonance (NMR) spectroscopy and hyphenated mass spectrometry [16]. At first, a brief explanation is given and then their application in the diagnosis of cancer will be discussed.

Metabolomics based on NMR spectroscopy

NMR spectroscopy is application of strong magnetic fields and radio frequency (RF) pulses to the nuclei of atoms for atoms either an odd atomic number (e.g., ^{1}H) or odd mass number (e.g., 13C, 31P) or both of them (e.g. 14N) in the presence of a strong magnetic field, the nucleus will cause the possess spin (nuclear spin). Absorption of RF energy will then allow the nuclei to be promoted from low-energy to high-energy spin states, and the subsequent emission of radiation during the relaxation process is detected. The majority of applications employs 1H (proton) NMR for clinical studies and as the known metabolites contain various hydrogen atoms in their molecular structure [17]. Between used techniques; The NMR has attracted so much attention to itself because it is rapid, Quantitative, non-destructive and cost effective. It needs little or no sample preparation and produces repeatable and reproducible results. NMR is a powerful tool for assigning and identification in a large number of metabolites in complex biofluids because each metabolite has its chemical shifts and unique multiple patterns. [18].

This technique has some limitations, for example, low sensitivity (in comparison with MS spectroscopy) and signal overlapping. It must be noted that NMR sensitivity depends on some factors such as natural abundance of the atom that are studied (1H, 31P, 19F 100%; 13C1. 10%; 15N 0.37%).This limitation was improved by the construction of modern NMR instruments with higher magnetic fields and special microprobes for trace analysis. By using high magnetic fields (800-900 MHz) and cryogenic probe signal to noise were Improve dramatically [19].

Another approach is the use of microbes instead of conventional 5 mm probes, new probes have a diameter of 1-3 mm or micro scaled [20]. This also will reduce the sample amount that is needed for analysis (600µl to 60µl respectively).

Recently, two dimensional (2D) homonuclear 1H, J-resolved (JRES) nuclear magnetic resonance spectroscopy is used in metabolomics research. This NMR method contains many advantages such as disperses the overlapping resonances into a second dimension, reducing overlap, simplify spectral assignments and accurate quantification [21]. One of the challenges in NMR-based metabolomics is suppression of water in biofluids. This problem has been solved by applying pulse sequences of standard pulse sequence nuclear Overhauser effect spectroscopy (NOESY preset) [22]. This pulse program has been used especially for urine samples. Another challenge in NMR-based approach is signal's suppression of macromolecules that is possible by Carr-Purcell-Meiboom-Gill (CPMG) sequence [23, 24].

Table2. common biological samples in metabolic studies

Sample	Advantage	Limitations	No. of biomarker[a]
Blood Serum/plasma	-Are common biofluids in NMR metabolomics -Easy to obtain -Most of the relevant NMR detectable metabolites have been by (HMDB) -non-invasive	-Broad signals from proteins and lipids	4535
Urine	- non-invasive -Easy to collection and storage -Easy sample preparation -Good for toxicology or drug follow-up studies -Urine contains more metabolites than blood and CSF.	-More overlapping signals in an NMR spectrum -Need to maintain homeostasis results in it being one of the most complex Composition	3995

Tissue	- Tissue is even greater homeostasis regulation than plasma and urine therefore provide highly consistent metabolic measurements -Are suitable for the study of specific sites for body (toxicity) -Need HR-MAS	-More difficult to collect -Invasive - Highly sensitive analysis techniques are required -Expensive	439
Exhaled breath [25]	- Exhaled breath condensate (EBC) collection is an innovative method -Rich in volatile metabolites (such as aldehydes, ketones, alcohols, hydrocarbons and esters) -Inexpensive -Safe and noninvasive method of diagnostic and therapeutic monitoring	- Limited to the investigation of respiratory disease - Dilution is an issue that is a problem with all methods of sampling the airway and lungs including sputum collection and bronchoalveolar lavage - The small volume of sample collected - EBC is currently used only as a research tool, (due to the lack of appropriate standardization and the absence of reference values)	16
CSF [26, 27]	CSF indirectly reflects the biochemical processes occurring in the brain. Therefore provides information about states of normal or pathological metabolism of the brain. -The biochemical composition of CSF contains metabolites which are secreted by the central nervous system (CNS) -CSF has lower protein and lipid content - Signal overlaps in CSF is not as serious	-Samples are more difficult to obtain -Invasive	436
Saliva	Saliva encompasses the secretions of three major glands namely parotid gland, submandibular gland and sublingual gland and other minor salivary glands It reflects a large range of physiological needs and information -Most compounds found in blood Are also present in saliva, but usually in lower concentrations -non-invasive, - Reliable collection of sample is low-cost.	Highly sensitive analysis techniques is required because of low compound concentrations	1233
Cell line	Good for understanding of the in vitro and in vivo actions of drugs - and aid in their rapid incorporation into novel therapeutic settings -Cells can also be stored in a deep frozen state and doing so there is no alteration to their growth rate or genetic composition and they can be revived whenever needed -It is far more economical to use cell cultures instead of rearing animals and doing experiments with animals -There is no requirement for ethical approval	-Need for specialized equipment, -Their sensitivity to varying environmental conditions, e.g. power failure. -Is that cell cultures are very prone to infections -Very time consuming	49
Amniotic fluids	Amniotic fluid (AF) composition reflects the physiological status during fetal development and it may be used to detect potential pathological conditions excretions and placental tissues.	AF contains large amounts of proteins and metabolites produced by the amnion epithelial cells, fetal tissues	17

[a]Based on HMDB Version 3.6

Because of the low concentration of the Metabolites in biofluids (serum), proteins (such as albumin) with high concentration, has been masked the NMR signals of metabolites. For solving this major problem in NMR, researchers use this item that, relaxation time (T2) of Metabolites relatively long then produce narrow peak and in contrast, macromolecules such as proteins have short T2 and produce broad peaks. This difference in relaxation times causes a big background of protein in 1HNMR signals. Veenstra and co-workers [28] have designed WET–CPMG pulse sequence to remove these unwanted signals. In this pulse program, NMR peak intensities are decreased proportional to their T2 relaxation times. Therefore, CPMG used to suppress signals arising from macromolecules [29, 30]. Another great merit of NMR technique is ready to measure the concentration of metabolites (qNMR) because the intensity in a resonance line is directly proportional with the number of resonant nuclei (spins)[31].

It must be noted that in order to measure exact concentration of metabolite it is necessary to use an internal standard in NMR spectroscopy. Trimethylsilan (TMS) is useful for organic solvent but common reference for biological samples are (3-trimethyl silyl propionic acid-d4, TSP-d4) and 2, 2-dimethyl-2-sila pentane-5-sulfonate sodium salt (DSS) [32, 33].

One of the recent developments in NMR technique's automated NMR using Autosampler that is applying for NMR structure determination, epidemiological or population screening studies. Automated NMR can analysis 200-300 sample per day [34, 35]. High Resolution Magic Angle Spinning (HR-MAS) for solid samples is another powerful NMR technics in metabolomics. In this technique that invented for the 1950s use of magic angle spinning NMR (MAS-NMR) allows intact tissues and cells to be examined with little sample preparation (~20 mg) or no preparation [36, 37]. MAS-NMR significantly reduces the effect of magnetic-field inhomogeneity, because a sample in homogeneity leads differences in magnetic susceptibility and causes broad and distorted peak shapes. HR-MAS-NMR yield's remarkably high-resolution spectra from a range of intact tissue types, therefore, provides detailed analysis of complex matrices such as tissue. It has recently been applied in toxicology and oncology, such as the study of the toxins- and disease-induced changes in a variety of tissues, including tumors and kinds of cancers. MAS-NMR also provides a means to validate, and study of metabolism both in vitro and in vivo conditions. Another advantage of HR-MAS-NMR approaches is usefulness in MRS imaging studies [38]. Furthermore, HR-MAS-NMR also has been applied to help interpretation of proton MRS imaging of tumors [39].

Mass based metabolomics

Recently, Metabolomics based on mass spectrometry is being used increasingly as an alternative approach specially hyphenated methods coupled with separation methods such as GC-MS, LC-MS, CE-MS. These techniques are more sensitive, selective and have wide dynamic range compared to NMR. The number of compounds that measured in mass based techniques strongly will depend on the resolution of the chromatography system and the specificity of the detection technique. A mass spectrometer can function as highly specific chromatographies. In this article, we have explained recent research results briefly based on MS detection techniques.

Metabolomics based on GC-MS spectroscopy

Gas chromatography-mass spectrometry (GC-MS) is a hyphenated technique that has been used over many years. This technique is suitable for target-metabolomics In this approach inert carrier gas (i.e. N2, He) moves analytes between silica capillary. Separation occurs as a result of equilibrium formed between the solutes and the stationary phase. Useful detectors for GC-MS are electron impact/ chemical ionization (EI/CI) mass spectrometry. Electron Impact (EI) ion source leads to extensive fragmentation and is helpful for unambiguous identification of analytes while the CI ion source is mild and usually uses for determine accurate mass measurement of metabolites. (Because of little fragmentation)

GC-MS is highly sensitive, and it is suitable for volatile metabolites. One of the most advantages

of GC-MS is access to powerful libraries such as NIST, Wiley Library that allows users to search both chromatographies retention time and mass spectra of analytes. Derivatization for detection of some polar metabolites (for example: fatty acids, phenols, prostaglandins, Steroids, alkaloids, amino acids) is necessary before GC-MS analysis. Chemical derivatization is usually necessary due to decrease the polarities of functional groups to facilitate their separation by GC. Generally, there are three basic types of derivatization reactions: silylation, alkylation and acylation. Silylation is the most widespread derivatization method and is very suitable for non-volatile samples. In this respect, Silylating reagents react with active hydrogen and convert them to be silylated derivatives that are more volatile and thermally stable and creating narrow peaks. One of The most common silyating reagents is present N-methyl-Ntrimethylsilyltriflouro-acetamide (MSTFA); Acylating reagents react with highly polar functional groups such as amino acids or carbohydrates, and an acyl group is attached to an organic compound. In this method of derivatization, active hydrogens (e.g., -OH, -SH and -NH) converted into esters, thioesters and amides, respectively through acylation. One of Common reagent for the Acylation is Fluoracylimidazoles. Alkylating reagents protect active hydrogens and replace of active hydrogen by an aliphatic or aliphatic-aromatic (e.g., benzyl) group. Dialkylacetals are common derivatization reagents for the Alkylation reactions.

GC-MS has some limitations, for example. It is destructive, low reproducible, low quantitation, needs high sample amount and has limitations for the detection of more polar metabolites and thermally labile metabolites, one of most its disadvantage is incapability to identify unknown compounds after derivatization [40, 41].

Recently multidimensional GC, known GC×GC-MS, GC-TOF-MS improved limitations of GC-MS techniques such as resolution, sensitivity. GC×GC-MS has composed of two columns; the first column is longer than second and separates compounds based on volatility whereas the second column separates based on polarity [42].

LC-MS based metabolomics

Liquid chromatography hyphenated with mass spectrometry is the most important technique among analytical tools that has been used for metabolic studies. Commonly used column in LC is reversed phase column (RP-HPLC), and ion source is electrospray ionization (ESI) for detection ionic compounds and ion suppression. LC-MS separates, identify and quantify a very broad group of metabolites. It has advantages such as highly sensitive, no need for derivatization, able to analysis of thermo-labile metabolites, but this technique also has some limitations, for example, does not access to good libraries, thus limited structural information, has high matrix effects, expensive, time consuming, low reproducibility [43, 44].

In the few past years LC-MS-MS (tandem mass), Matrix Assisted laser desorption/ ionization (MALDI-MS) [45], Desorption electrospray ionization Mass Spectrometry (DESI) [46], Extractive electrospray ionization (EESI-MS) [47] methods have been used. These methods removed many problems of conventional technique (LC-MS) for example, LC-MS-MS is suitable to validate and identifying of unknown molecules because of the second mass analyzer [48]. MALDI-MS is highly sensitive, detect a wide range of molecules and is label free DESI and EESI-MS approaches are useful for high-throughput data [49].

Recently in LC-MS columns instead of conventional particulate have been used monolithic columns (for separation of peptides and proteins). Ultra high-performance liquid chromatography (UHPLC) has improved limitation of HPLC (chromatographic resolution, improved detection limit 3-5 fold) [50].

Recently, LC-NMR-MS has applied in metabolomics that it has advantages both techniques and has been noticed by scientists [16, 51].

CE-MS

Capillary Electrophorese-Mass spectroscopy is a powerful technique for the Analysis of small soluble compounds in biological fluids, but it has not been used widely for metabolic studies. CE-MS is useful for polar metabolites, complex

biological sample that leading to high resolution, but this approach suffers from insufficient sensitivity, low reproducibility, high cost and time consuming. Interface of CE with MS is so difficult and desalting step is necessary prior to mass spectrometry, it has Complex methodology and quantification is problematic [52-54].

Pre- reprocessing

Metabolomics data are highly informative and extraction of them has always been a challenging problem. The main objective of data preprocessing is to convert raw spectra to datasets and summarize them in a table (data matrix). Both NMR and mass spectra contain hundreds up to thousand signals. Analysis of such heavy datasets is a more challenging process. Before stepping of multivariate analysis, a number of data pretreatments are required. The data pretreatment in ^{1}HNMR and GC-MS are fundamentally different. It is for NMR data as baseline correction, alignment, binning, Normalization, scaling and transformations. Data pre-processing of mass peaks is one of the most challenging areas in the metabolomics field with regard to software development. Data preprocessing for MS based methods are noise Filtering, data binning, Peak detection, Time Alignment, Missing values, Normalization and Scaling. Preprocessing steps NMR and MS based methods are summarized in Table 3 and Table 4 respectively:

Table 3. preprocessing of NMR spectra

	Pre-processing	Explain	Solution
Preprocessing for NMR based approach[21]	Baseline correction	-Correction of distortions -Exclude outside 0.2-10.0ppm -Remove solvent water(in 4.7-5ppm) -Remove urea (in urine samples)	-Automated baseline correction -Iterative polynomial-fitting[55] -Asymmetric Least Squares[56] -B-splines[34]
	Alignment	Correction of peak shift caused by instrumental factors -Changes of pH, temperature - Changes of salt concentration and dilution	-Aligned with spectral reference[57] -Interval correction[58] shifting(icoshif)[59] -Warping correlation optimized warping (COW) [60] -Hierarchical cluster-based peak Alignment(CluPA)[61] -Time warping[62]
	Binning(bucketing)[63]	Used for quantitative, objective or scored integrals of specific spectral peak dimensional Reduce the data	-Equidistant binning of 0.04ppm -Non equidistant binning adaptive-intelligent binning (Al-binning) -Gaussian binning -Adaptive binning with wavelet transform -Dynamic Adaptive binning
	Normalization[64]	Make all samples comparable with each other -Dilution effects (especially for urine samples) Eliminate systematic errors Delete inter-sample variation	-Probabilistic normalization (PQ) -Histogram Normalization -Group aggregation normalization (GAN)
	Scaling[30]	-Scaling the most abundant metabolites before analysis	-Mean centering -Auto scaling -Pareto scaling -Range scaling -Vast scaling -Level scaling
	Transformation[65]	Remove hetero scedasticy Decrease the effect of non-normality	Log transform Box-Cox

Table 4. preprocessing of MS spectra (GC-MS) and (LC-MS)

	Pre-processing	Explain	Solution
Preprocessing for MS based approach[66]	Noise Filtering [67]	Mask the important components of the chromatographic data.	One of the common ways of removing noise is based on a so-called "moving window" filter (Antoniou 1993; Mitra 1998)
	Data binning [68]	Mass intensities are binned into Mass aligned vectors of uniform Length	Make slices (default 0.1m/z) Assign intensity to each slice For each scan. -Several methods available: - Bin (default) and Binlin, Binlinbase, Intlin
	Peak detection [68, 69]	-Identify all signals caused by true ions and avoid detection of false positives (i.e., noise, spikes) - Also aims to provide accurate quantitative information about ion concentrations	-Detection in two directions by finding peaks independently in both m/z and retention time direction - The other strategy is slicing data to extract ion chromatograms (XIC), with each one covering a narrow m/z range, therefore avoiding the problem of searching for peaks in m/z direction. (XCMS) - For feature extraction is model fitting against the original raw signal
	Retention Time Alignment[70]	Match peaks with similar retention times across Multiple samples, and use the groups of matched peaks for time alignment	-AMDIS automatically finds peaks and deconvolutes the mass spectra of co-eluting compounds - A simple linear retention time shift could align peaks at only one of the time points, but not both. -The nonlinear alignment algorithm in XCMS properly aligns peaks over all times
	Missing values[71]	Any peaks that failed to recognize	-Using information from peak detection about where peaks begin and end, and aligned retention times for each sample Then integrate the raw LC/MS data to fill in intensity values for each of the missing data points.
	Normalization[72]	It is necessary to remove the effect of the total amount from the analysis. Remove the unwanted systematic errors Bias in ion intensities between measurements, while Retaining the interesting biological variation	Normalization by unit norm Normalization by median intensities, Normalization by the maximum likelihood method Normalization by a single or multiple internal or external standard compounds based on empirical rules, such as specific regions of retention time
	Scaling[73]	Scaling the most abundant metabolites before analysis	Mean-centered, Pareto-scaled Range scaling Vast scaling

Data analysis

Chemometrics is one of an influential branch of chemistry and provides mathematical and statistical tools for data extraction and analysis of complex chemical and biological data. Multivariate analysis (MVA) is one of prominent subset of Chemometrics. This method of analysis is widely used in metabolomics studies that leads to reduces complexity and dimensionally of NMR/MS datasets and provides relevant

biological information from the input analytical data obtained from instrument. These initial data consists of many different variables in a matrix. This matrix has column and rows, that columns are results of observations (variables: metabolites) and rows are objects. Any variable can be considered as a different dimension. Reduction of dimensionality can be done in one of two ways; either unsupervised or supervised analysis[74, 75].

Unsupervised analysis

In unsupervised learning, input information (metabolite/peak data), are clustered into groups, revealing the overall structure on the data. One of the most applied methods is a principal component analysis (PCA). PCA is a method that reduces the dimension of data between variables in the dataset by application of covariance analysis. This technics is useful to identify hidden information within the data set by compressing of a data set and it is a linear combination of original data parameters. PCA utilizes a maximum of variables to the minimum of principal components (PCs). Each PC is orthogonal to the others. PCA can be visualized in two metrics, known Scores and Loadings. Score plot (T), is linear combinations of the original variables and reveals how the data is spread; each point expresses of a single spectrum, whereas loading plot/scatter plot (P) shows the effect (weight) of the individual X-variables in the model. Each point depicts a different spectral intensity. Detections of outliers are another application of PCA [62]. Other unsupervised methods which have been used in metabolomics are k-means, HCA(Hierarchical cluster analysis), STOCSY (statistical correlation spectroscopy), a hybrid of two methods of STOCSY and HCA that known as CLASSY (cluster analysis statistical spectroscopy)[76].

In table 5 briefly about each of them have been explained. Unsupervised analyzing methods have some limitations because of the character of unsupervised learning that cannot predict and evaluate model's parameters [62, 77].

Table 5. Other unsupervised methods

unsupervised methods	explain
Hierarchical cluster analysis (HCA)	-Make group based on similarity - Necessary choice of two input function
K-means	Clustering approach -Most apply in Transcripotomics
Statistical correlation spectroscopy (STOCSY)	Extract of information from 1D , 2D-NMR -Applied in drug metabolites
Cluster analysis, statistical spectroscopy (CLASSY)	-Better than STOCSY - Applied in hyphenated techniques -Is highly accurate

Supervised Analysis

In supervised techniques the input information (metabolite/peak date), is paired with defined outputs e.g. disease or control (training data) and a model is built to classify the data. Some of statistical strategies are used for prediction of models, include, partial least square (PLS) which is a linear regression method; discriminant function analysis (DFA), Principal Component Regression (PCR), partial least square discriminant analysis (PLS-DA) and OPLS method is modified form of the PLS method to help overcome the pitfalls. These techniques use a subset of the overall data (training data), to model the outcomes. The remaining data set is used to validate the predictive strength of the built model [78]. Validation is a very important step in evaluating a model for an independent Dataset. Several methods for validation of models exist. In a few years An Artificial neural network (ANN) method has been extensively used as a non-linear data modeling method in metabolic studies and pathway analysis [79]. A relevant criterion for assessing the quality of fit models And determine the correct prediction of the Performance is defined under the curve receiver operating characteristic (ROC) [80]. The most common methods are cross-validation, double cross-validation [81]. Many data in metabolic systems display non-linear dependencies because of biological

Complexity and environmental factors [81] . The linear methods are close to fail in such situations. Therefore Metabolites that are significant can be further analyzed using the receiver-operator Characteristic (ROC) curve. This is a plot of the degree of sensitivity (predicting a disease) versus specificity (predicting non-disease), and is used to identify the most discriminatory metabolites, based on a ROC score identify cutoff points for metabolites [82]. the Performance is defined under the curve receiver operating characteristic (ROC) [80]. The most common methods are cross-validation, double cross-validation [81]. Many data in metabolic systems display non-linear dependencies because of biological Complexity and environmental factors [81] . The linear methods are close to fail in such situations.

Therefore Metabolites that are significant can be further analyzed using the receiver-operator characteristic (ROC) curve. This is a plot of the degree of sensitivity (predicting a disease) versus specificity (predicting non-disease), and is used to identify the most discriminatory metabolites, based on a ROC score identify cutoff points for metabolites [82].

Common Database for Metabolomics

Metabolomics like other "Omics" sciences dealing with complex and high throughput data, then Researcher to identify metabolites, assignment of compounds, exact mass, GC-EI spectrum, MS/MS spectrum comparison, identity of pathways Require access to a comprehensive database. We listed here the major open-access resources available (Table6)

Application of Metabolomics

Recently, metabolomics has been used in many research fields, for example in the study of toxicology [83] responses to environmental stress [84], , nutrition [85], drug discovery [86] diagnosis of disease [87-94],identify biomarkers of disease and prognostic biomarkers [90] cancer [95-97], in natural product (plant metabolomics) [98], and in traditional medicine [99-101]. Metabolomics allow the accurate prediction of disease, etiology and mechanisms[102, 103] Cancer is a complex disease state that changes the normal of healthy cells into tumor cells. Due to the cancer is associated with metabolic changes, therefore metabolomics studies can help in early stage detection and diagnosis of cancer and in evaluation of clinical trial. In this section we briefly describe the biomarkers identified in cancer diseases. identifed biomarkers kinds of cancers listed in the Table7.

Table 6. Common database for metabolic study

Database	characteristics
METLIN[104]	The METLIN Database is a repository of metabolite information as well as tandem mass spectrometry data.
Human Metabolome Database(HMDB)[105]	The database contains three kinds of data: 1) chemical data, 2) clinical data, and 3) molecular biology/biochemistry data. Many data fields are hyperlinked to other databases (KEGG, PubChem, MetaCyc, ChEBI, PDB, Swiss-Prot, and GenBank) and a variety of structure and pathway viewing applets.
BiGG[106]	The BiGG database is a metabolic reconstruction of human metabolism designed for systems biology simulation and metabolic flux balance modeling. It is a comprehensive literature-based genome-scale metabolic
SetupX[107]	SetupX, developed by the Fiehn laboratory and is web-based. It displays GC-MS metabolomic data through its metabolic annotation database called BinBase. BinBase is a GC-TOF metabolomic database
SYSTOMONAS[108]	(SYSTems biology of pseudOMONAS) is a database for systems biology studies of Pseudomonas species. It contains transcriptomic, proteomic and metabolomic data as well as metabolic reconstructions of this pathogen.
MetaboLights[109]	MetaboLights is a database for metabolomics experiments. It is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations, concentrations and experimental data from metabolic experiments.

MassBank[110]	MassBank is a mass spectral database of experimentally acquired high resolution MS spectra of metabolites. Maintained and supported by the JST-BIRD project, it offers various query methods for standard spectra obtained from Keio University, RIKEN PSC, and other Japanese research institutions. It is officially sanctioned by the Mass Spectrometry Society of Japan. The database has very detailed MS data and excellent spectral/structure searching utilities. More than 13,000 spectra of 1900 different compounds are available.
Golm Metabolome Database[111]	The Golm Metabolome Database provides public access to GC/MS libraries which are stored as Mass Spectral (MS) and Retention Time Index (RI) Libraries (MSRI). These libraries of mass spectral and retention time indices can be used by the NIST/AMDIS software to identify metabolites according their spectral tags and RI's. The libraries are both searchable and downloadable and have been carefully collected under defined conditions on several types of GC/MS instruments (quadrupole and TOF).
Fiehn GC-MS Database	This database contains characteristics of compounds (name, structure, CAS ID, other links) and GC/MS data (spectra and retention indices)that have been collected by the Fiehn laboratory.
BML-NMR[112]	The Birmingham Metabolite Library Nuclear Magnetic Resonance database is a freely available resource .This database includes both 2-D 1H J-resolved spectra and 1-D 1H spectra, recorded at 500 MHz using various water suppression methods and acquisition parameters, for solutions at pH values of 6.6, 7.0 and 7.4.
MzCloud[113]	MzCloud features a searchable collection of high resolution/accurate mass spectral trees using a new third generation spectra correlation algorithm. MzCloud tries to address the identification bottleneck by considering all mass spectrometricaly relevant aspects, looking at number of experimental and computational details and in some cases, allowing identification of unknowns even if they are not present in library.
BMRB [114]	The BioMagResBank (BMRB) is the central repository for experimental NMR spectral data, primarily for macromolecules. The BMRB also contains a recently established subsection for metabolite data. The current metabolomics database contains structures, structure viewing applets, nomenclature data, extensive 1D and 2D spectra peak lists (from 1D, TOCSY, DEPT, HSQC experiments), raw spectra and FIDs for nearly 500 molecules.
MMCD [115]	The Madison Metabolomics Consortium Database (MMCD) is a database of chemicals. This database provides the chemical formula, names and synonyms, structure, physical and chemical properties, NMR and MS data on pure compounds under defined conditions, NMR chemical shifts determined by empirical and/or theoretical approaches, information on the presence of the metabolite in different biological species, and extensive links to images, references, and other public databases.

Table7. Identified biomarkers by metabolomics studies

Disease	**Sample**	**Instrument**	**Biomarker**	**Ref**
Ovarian borderline tumor	Tissue	GC-ToF-MS	51 metabolites were significantly different between borderline tumors and carcinomas Glycerol phosphate alpha, Lactic acid	[116]
Kidney cancer	Urine	GC-ToF-MS	High level quinolinate, α-ketoglutarate, and gentisate,	[117, 118]
Colorectal cancer	Serum and Urine	GC/MS	Taurine, lactate, choline, inositol, glycine, phosphocholine, proline, phenylalanine, alanine, threonine, valine and leucine { L-alanine, L-glutamine, glucoronoic lacaton}[119]	[120, 121]
Gastrointestinal cancer(early stage)	Serum	GC-MS	3-hydroxypropionic acid, pyruvic acid	[119]
Lung cancer	Urine	NMR	Hippurate, trigonelline, ß-hydroxyisovalerate, α-hydroxyisobutyrate, N-acetylglutamine and creatinine	[122]
Lung cancer	Plasma	NMR	High concentration(lactate, pyruvate) Low concentration (citrate, Formate, acetate, glucose,	[123]

			glutamine, alanine, tyrosine and Valine)	
Prostate cancer	Cell line	NMR	Myoinositol, glutathione, amino acids, methionine, phosphocholine, phosphocholine/glycerophosphocholine	[124, 125]
Human hepatocellular Carcinoma	Serum	NMR	Higher levels of acetate, N-acetylglycoproteins, Private, glutamine, a-ketoglutarate, glycerol, Tyrosine, 1-methylhistidine and phenylalanine, together with Lower levels of low-density lipoprotein, Isoleucine, Valine, acetoacetate, Creatine, choline and unsaturated lipids.	[126]
Human Brain cancer	Tumor cell	NMR	Phosphocholine/glycerophosphocholine	[127]
Human Brain cancer	Tissue	HR-MAS and solution NMR	Lactate, mobile lipids (correlate with the degree of tumor necrosis The ratio of (Phosphocholine/choline (correlates with malignancy of glioma)	[128]
Breast Cancer		HR-MAS	Coline, phosphocholine/glycerophosphocholine (degree of tumor) Glutamine, glutamate, glycine and taurine (in tumor type discrimination)	[129]
Breast Cancer	Breast tissue	HR-MAS	High level (taurine, choline containing metabolites)	[130]
Breast Cancer	Extract of tumor	NMR	In normal tissue (High level of glucose, Myo-inositol) In tumor tissue (phosphocholine, lactate, succinate)	[131]
Head and neck squamous cell carcinoma (HNSCC)	Tissue	HR-MAS	Higher levels of lactate, amino acids, choline and lower levels of triglyceride	[132, 133]
Oral squamous cell carcinoma (OSCC) early stage	Saliva	NMR LC-TOF-MS	14 potential salivary metabolites were identified. Eight biomarkers up-regulated in OSCC patients are compared with control and six down-regulated groups. Five salivary biomarkers (propionylcholine, N-Acetyl-L-phenylalanine, sphinganine, phytosphingosine, and S-carboxymethyl-L-cysteine	[80, 132, 134]

As seen in the table, levels of glucose, pyruvate, taurine , lactate have been increased and levels of phosphocholine, choline betaine and glycine have been decreased. Cancer cells request more energy that supplies it by high rate of glycolysis then going up lactic acid fermentation in the cytosol that it is known as the "Warburg" effect [135]. Proliferation of tumor in acidic environment elevates and in acidic media tumor cells invade to healthy cells.[136]**.** Therefore level of glucose has been decreased. [137, 138]. Glucose generates precursors of amino acids, lipids , and nucleotides [135]. Also Levels of fatty acids have been increased in early stage of cancer (specially breast cancer [139, 140]) and it is related to conversion of pyruvate to acetyl CoA. Acetyl CoA can be consumed for *de novo* fatty acid synthesis [141].

Reduced Level of Glycine has been observed in tumor cells because glycerin is a precursor for purine synthesis, and glycine has been consumed to produce sarcasm. Levels of urinary Sarcosion had been increased in prostate cancer. [142, 143]. Sarcosine is produced by methyl transfer from SAMe(*S*-adenosyl-L-methionine) to glycine.

The levels of Choline, phosphocholine, phosphatidylcholine, and glycerophospho-choline have been increased in brain, breast, prostate and liver cancers[96, 144] . Decreased levels of Lysophosphati-dylcholines (lysoPCs) have been observed in lung [145]and liver cancer [146]. while lysophosphatic acid has been reported as

increased in ovarian cancer[147] (lysoPCs) are lipid intermediates that can used to form PCs or are the products of phospholipases acting on PCs. ratios of (glycerophosphocholine and phosphocholine)/creatine, myo-inositol/ scyllo-inositol, choline/creatine and other ratios were demonstrated to correlate with the number of tumor cells, tumor cell proliferation[148].
Therefore in cancer cells, metabolites are associated with increased glycolysis, anaplerosis, and membrane choline metabolism.

CONCLUSION

Metabolomics use in a wide variety of disciplines, including toxicology, systems biology, pharmacology, and medicine, especially in prognoses, diagnoses and therapy. Actually, diseases are perturbation in metabolic orders. Therefore, measurement of metabolites in the healthy and disease state leads to a better understanding on the mechanism of the disease, finding metabolic pathways involved in disease. In this review, we briefly explained about research strategies and analytical tools that use and chemometric approaches (MVA) and the most common databases and finally application of metabolomics in diagnosis of cancer diseases.
It is noticeable that metabolic studies are being applied to better understand metabolism of cancer cells, for diagnosis and prognosis biomarkers. One of most limitations of metabolomics is that known biomakers are still experimental and in the research phase, there are some challenges include identification of all the variables (unknown peaks)by improvement in analytical tools and chemometrics, validation studies for high population, communion of clinical metabolomics data with the clinic, Detailed interpretation of clinical metabolomics data, and Following results of NMR and MS in the clinical. Due to metabolomics is the downstream approach and complimentary to the other omics approaches. The future of medical research within the field of cancer will use from information of systems biology particularly metabolomics to diagnosis biomarkers.
Systems biology approaches will combine together until acquire detailed information about cancer pathways, therefore, will facilitate understanding about the complexity of cancer.

References

1. Lindon, J.C., E. Holmes, and J.K. Nicholson, *Metabonomics techniques and applications to pharmaceutical research & development.* Pharmaceutical research, 2006. 23(6): p. 1075-1088.
2. Schnackenberg, L.K., J. Sun, and R.D. Beger, *Metabolomics in systems toxicology: Towards personalized medicine.* General, Applied and Systems Toxicology, 2009.
3. Oliver, S.G., et al., *Systematic functional analysis of the yeast genome.* Trends in biotechnology, 1998. 16(9): p. 373-378.
4. Nicholson, J.K., et al., *Metabonomics: a platform for studying drug toxicity and gene function.* Nature reviews Drug discovery, 2002. 1(2): p. 153-161.
5. Simón-Manso, Y., et al., *Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources.* Analytical chemistry, 2013. 85(24): p. 11725-11731.
6. Fiehn, O., *Metabolomics–the link between genotypes and phenotypes.* Plant molecular biology, 2002. 48(1-2): p. 155-171.
7. Cevallos-Cevallos, J.M., et al., *Metabolomic analysis in food science: a review.* Trends in Food Science & Technology, 2009. 20(11): p. 557-566.
8. Dettmer, K. and B.D. Hammock, *Metabolomics--a new exciting field within the" omics" sciences.* Environmental Health Perspectives, 2004. 112(7): p. A396.
9. Villas-Bôas, S.G., et al., *Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media.* Analytical biochemistry, 2006. 349(2): p. 297-305.
10. Lindon, J.C. and J.K. Nicholson, *Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery.* TrAC Trends in Analytical Chemistry, 2008. 27(3): p. 194-204.

11. Tikunov, Y.M., et al., *NON-SMOKY GLYCOSYLTRANSFERASE1 prevents the release of smoky aroma from tomato fruit.* The Plant Cell Online, 2013. 25(8): p. 3067-3078.
12. Walsh, M.C., et al., *Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans.* The American journal of clinical nutrition, 2006. 84(3): p. 531-539.
13. Čuperlović-Culf, M., et al., *Cell culture metabolomics: applications and future directions.* Drug discovery today, 2010. 15(15): p. 610-621.
14. Fanos, V., et al., *Metabolomics application in maternal-fetal medicine.* BioMed research international, 2013. 2013.
15. Graça, G., et al., *1H NMR based metabonomics of human amniotic fluid for the metabolic characterization of fetus malformations.* Journal of proteome research, 2009. 8(8): p. 4144-4150.
16. Dunn, W.B. and D.I. Ellis, *Metabolomics: current analytical platforms and methodologies.* TrAC Trends in Analytical Chemistry, 2005. 24(4): p. 285-294.
17. Shaw, D., *Fourier transform NMR spectroscopy*. *2.* 1984.
18. Coen, M. and P.W. Kuchel, *Metabonomics based on NMR spectroscopy.* Chemistry in Australia, 2004. 71(6): p. 13.
19. Beckonert, O., et al., *Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts.* Nature protocols, 2007. 2(11): p. 2692-2703.
20. Lane, A.N., *Principles of NMR for Applications in Metabolomics*, in *The Handbook of Metabolomics*. 2012, Springer. p. 127-197.
21. Parsons, H.M., et al., *Improved classification accuracy in 1-and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation.* BMC bioinformatics, 2007. 8(1): p. 234.
22. Lippens, G., C. Dhalluin, and J. Wieruszeski, *Use of a water flip-back pulse in the homonuclear NOESY experiment.* Journal of biomolecular NMR, 1995. 5(3): p. 327-331.
23. Song, Y.-Q., *Categories of coherence pathways for the CPMG sequence.* Journal of Magnetic Resonance, 2002. 157(1): p. 82-91.
24. Levitt, M.H. and R. Freeman, *Compensation for pulse imperfections in NMR spin-echo experiments.* Journal of Magnetic Resonance (1969), 1981. 43(1): p. 65-80.
25. Loukides, S., et al., *Biomarkers in the exhaled breath condensate of healthy adults: mapping the path towards reference values.* Current medicinal chemistry, 2008. 15(6): p. 620-630.
26. Zhang, A., et al., *Recent and potential developments of biofluid analyses in metabolomics.* Journal of proteomics, 2012. 75(4): p. 1079-1088.
27. Verwaest, K.A., et al., *1 H NMR based metabolomics of CSF and blood serum: A metabolic profile for a transgenic rat model of Huntington disease.* Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2011. 1812(11): p. 1371-1379.
28. Van, Q.N., G.N. Chmurny, and T.D. Veenstra, *The depletion of protein signals in metabonomics analysis with the WET–CPMG pulse sequence.* Biochemical and biophysical research communications, 2003. 301(4): p. 952-959.
29. Mittermaier, A. and L.E. Kay, *New tools provide new insights in NMR studies of protein dynamics.* Science, 2006. 312(5771): p. 224-228.
30. Weljie, A.M., et al., *Targeted profiling: quantitative analysis of 1H NMR metabolomics data.* Analytical Chemistry, 2006. 78(13): p. 4430-4442.
31. Barding Jr, G.A., R. Salditos, and C.K. Larive, *Quantitative NMR for bioanalysis and metabolomics.* Analytical and bioanalytical chemistry, 2012. 404(4): p. 1165-1179.
32. Torgrip, R., et al., *A note on normalization of biofluid 1D 1H-NMR data.* Metabolomics, 2008. 4(2): p. 114-121.
33. Sheedy, J.R., et al., *A sample preparation protocol for 1 H nuclear magnetic resonance studies of water-soluble metabolites in blood and urine.* Analytical biochemistry, 2010. 398(2): p. 263-265.
34. Mercier, P., et al., *Towards automatic metabolomic profiling of high-resolution one-*

dimensional proton NMR spectra. Journal of biomolecular NMR, 2011. 49(3-4): p. 307-323.
35. Zhang, F., et al., *Strategy for automated analysis of dynamic metabolic mixtures by NMR. Application to an insect venom.* Analytical chemistry, 2007. 79(20): p. 7748-7752.
36. Keun, H.C., *Metabonomic modeling of drug toxicity.* Pharmacology & therapeutics, 2006. 109(1): p. 92-106.
37. Lindon, J.C., et al., *Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis.* Biomarkers, 2004. 9(1): p. 1-31.
38. Li, M., et al., *An HR-MAS MR metabolomics study on breast tissues obtained with core needle biopsy.* PloS one, 2011. 6(10): p. e25563.
39. Bathen, T.F., et al., *Magnetic resonance metabolomics of intact tissue: a biotechnological tool in cancer diagnostics and treatment evaluation.* Cancer research, 2010. 70(17): p. 6692-6696.
40. Jonsson, P., et al., *A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS.* Analytical chemistry, 2004. 76(6): p. 1738-1745.
41. Kanani, H., P.K. Chrysanthopoulos, and M.I. Klapa, *Standardizing GC–MS metabolomics.* Journal of Chromatography B, 2008. 871(2): p. 191-201.
42. Culbertson, A.W., et al., *Inside the Personalized Medicine Toolbox: GCxGC-Mass Spectrometry for High-Throughput Profiling of the Human Plasma Metabolome.* LC GC North America, 2008. 26(6): p. 560.
43. Lu, W., B.D. Bennett, and J.D. Rabinowitz, *Analytical strategies for LC–MS-based targeted metabolomics.* Journal of Chromatography B, 2008. 871(2): p. 236-242.
44. Theodoridis, G., H.G. Gika, and I.D. Wilson, *LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics.* TrAC Trends in Analytical Chemistry, 2008. 27(3): p. 251-260.
45. Yukihira, D., et al., *MALDI− MS-based high-throughput metabolite analysis for intracellular metabolic dynamics.* Analytical chemistry, 2010. 82(10): p. 4278-4282.
46. Chen, H., et al., *Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation.* Rapid Communications in Mass Spectrometry, 2006. 20(10): p. 1577-1584.
47. Jackson, A.U., et al., *Targeted metabolomic analysis of Escherichia coli by desorption electrospray ionization and extractive electrospray ionization mass spectrometry.* Analytical biochemistry, 2008. 375(2): p. 272-281.
48. Xiao, J.F., B. Zhou, and H.W. Ressom, *Metabolite identification and quantitation in LC-MS/MS-based metabolomics.* TrAC Trends in Analytical Chemistry, 2012. 32: p. 1-14.
49. Zhang, A., et al., *Modern analytical techniques in metabolomics analysis.* Analyst, 2012. 137(2): p. 293-300.
50. Toh, D.-F., et al., *Ultra-high performance liquid chromatography/time-of-flight mass spectrometry (UHPLC/TOFMS) for time-dependent profiling of raw and steamed Panax notoginseng.* Journal of pharmaceutical and biomedical analysis, 2010. 52(1): p. 43-50.
51. Exarchou, V., et al., *LC–NMR coupling technology: recent advancements and applications in natural products analysis.* Magnetic Resonance in Chemistry, 2005. 43(9): p. 681-687.
52. Ramautar, R., G.W. Somsen, and G.J. de Jong, *CE-MS in metabolomics.* Electrophoresis, 2009. 30(1): p. 276-291.
53. Ramautar, R., et al., *CE-MS for metabolomics: Developments and applications in the period 2008–2010.* Electrophoresis, 2011. 32(1): p. 52-65.
54. Ramautar, R., G.W. Somsen, and G.J. de Jong, *CE-MS for metabolomics: Developments and applications in the period 2010–2012.* Electrophoresis, 2013. 34(1): p. 86-98.
55. Halouska, S. and R. Powers, *Negative impact of noise on the principal component analysis of NMR data.* Journal of Magnetic Resonance, 2006. 178(1): p. 88-95.
56. Blanchet, L., et al., *Fusion of metabolomics and proteomics data for biomarkers discovery: case study on the experimental autoimmune*

encephalomyelitis. BMC bioinformatics, 2011. 12(1): p. 254.
57. Savorani, F., G. Tomasi, and S.B. Engelsen, *icoshift: A versatile tool for the rapid alignment of 1D NMR spectra.* Journal of Magnetic Resonance, 2010. 202(2): p. 190-202.
58. Xi, Y. and D.M. Rocke, *Baseline correction for NMR spectroscopic metabolomics data analysis.* BMC bioinformatics, 2008. 9(1): p. 324.
59. Larsen, F.H., F. van den Berg, and S.B. Engelsen, *An exploratory chemometric study of 1H NMR spectra of table wines.* Journal of Chemometrics, 2006. 20(5): p. 198-208.
60. Veselkov, K.A., et al., *Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery.* Analytical Chemistry, 2008. 81(1): p. 56-66.
61. Smolinska, A., et al., *NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review.* Analytica chimica acta, 2012. 750: p. 82-97.
62. Goodacre, R., et al., *Proposed minimum reporting standards for data analysis in metabolomics.* Metabolomics, 2007. 3(3): p. 231-241.
63. De Meyer, T., et al., *NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm.* Analytical Chemistry, 2008. 80(10): p. 3783-3790.
64. Sysi-Aho, M., et al., *Normalization method for metabolomics data using optimal selection of multiple internal standards.* BMC bioinformatics, 2007. 8(1): p. 93.
65. Viant, M.R., *Improved methods for the acquisition and interpretation of NMR metabolomic data.* Biochemical and biophysical research communications, 2003. 310(3): p. 943-948.
66. Katajamaa, M. and M. Orešič, *Data processing for mass spectrometry-based metabolomics.* Journal of Chromatography A, 2007. 1158(1): p. 318-328.
67. Creek, D.J., et al., *IDEOM: an Excel interface for analysis of LC–MS-based metabolomics data.* Bioinformatics, 2012. 28(7): p. 1048-1049.
68. DeHaven, C.D., et al., *Organization of GC/MS and LC/MS metabolomics data into chemical libraries.* Journal of cheminformatics, 2010. 2(1): p. 9.
69. Trygg, J., et al., *Extraction and GC/MS analysis of the human blood plasma metabolome.* Analytical Chemistry, 2005. 77(24): p. 8086-8094.
70. Nordström, A., et al., *Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum.* Analytical chemistry, 2006. 78(10): p. 3289-3295.
71. Bijlsma, S., et al., *Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation.* Analytical chemistry, 2006. 78(2): p. 567-574.
72. Castillo, S., et al., *Algorithms and tools for the preprocessing of LC–MS metabolomics data.* Chemometrics and Intelligent Laboratory Systems, 2011. 108(1): p. 23-32.
73. van den Berg, R.A., et al., *Centering, scaling, and transformations: improving the biological information content of metabolomics data.* BMC genomics, 2006. 7(1): p. 142.
74. Xia, J., et al., *MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis.* Nucleic acids research, 2012. 40(W1): p. W127-W133.
75. Steuer, R., *Review: On the analysis and interpretation of correlations in metabolomic data.* Briefings in bioinformatics, 2006. 7(2): p. 151-158.
76. Smolinska, A., et al., *NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review.* Analytica Chimica Acta, 2012. 750(0): p. 82-97.
77. Weckwerth, W. and K. Morgenthal, *Metabolomics: from pattern recognition to biological interpretation.* Drug discovery today, 2005. 10(22): p. 1551-1558.
78. Catchpole, G.S., et al., *Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops.* Proceedings of the National academy of Sciences

of the United States of America, 2005. 102(40): p. 14458-14462.
79. Ott, K.-H., et al., *Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts.* Phytochemistry, 2003. 62(6): p. 971-985.
80. De Almeida, P.D.V., et al., *Saliva composition and functions: a comprehensive review.* The journal of contemporary dental practice, 2008(9): p. 72-80.
81. Madsen, R., T. Lundstedt, and J. Trygg, *Chemometrics in metabolomics—a review in human disease diagnosis.* Analytica Chimica Acta, 2010. 659(1): p. 23-33.
82. Seli, E., et al., *Receiver operating characteristic (ROC) analysis of day 5 morphology grading and metabolomic Viability Score on predicting implantation outcome.* Journal of assisted reproduction and genetics, 2011. 28(2): p. 137-144.
83. Robertson, D.G., P. Watkins, and M.D. Reily, *Metabolomics in toxicology: preclinical and clinical applications.* Toxicological Sciences, 2010: p. kfq358.
84. Bundy, J.G., M.P. Davey, and M.R. Viant, *Environmental metabolomics: a critical review and future perspectives.* Metabolomics, 2009. 5(1): p. 3-21.
85. Gibney, M.J., et al., *Metabolomics in human nutrition: opportunities and challenges.* The American journal of clinical nutrition, 2005. 82(3): p. 497-503.
86. Kell, D.B., *Systems biology, metabolic modelling and metabolomics in drug discovery and development.* Drug discovery today, 2006. 11(23): p. 1085-1092.
87. Fathi, F., et al., *Metabonomics based NMR in Crohn's disease applying PLS-DA.* Gastroenterology and Hepatology from bed to bench, 2013. 6(Suppl 1): p. S82.
88. Fathi, F., et al., *A Metabonomics Study on Celiac Disease by CART.* International Journal of Celiac Disease, 2014. 2(2): p. 44-46.
89. Fathi, F., et al., *NMR based metabonomics study on celiac disease in the blood serum.* Gastroenterology and Hepatology from bed to bench, 2013. 6(4): p. 190.
90. Nobakht M. Gh, B.F., et al., *The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis.* Biomarkers, 2014(0): p. 1-12.
91. Fathi, F., et al., *The differential diagnosis of Crohn's disease and celiac disease using nuclear magnetic resonance spectroscopy.* Applied Magnetic Resonance, 2014. 45(5): p. 451-459.
92. Fathi, F., et al., *1H NMR based metabolic profiling in Crohn's disease by random forest methodology.* Magnetic Resonance in Chemistry, 2014. 52(7): p. 370-376.
93. Mortazavi-Tabatabaei, S.A., et al., *Investigation of metabonomics technique by analyze of NMR data, which method is better? Mean center or auto scale?* Journal of Paramedical Sciences, 2012. 4(1).
94. Rezaei-Tavirani, M., et al., *Advantage of applying OSC to 1H NMR-based metabonomic data of celiac disease.* International Journal of Endocrinology and Metabolism, 2012. 10(3): p. 548.
95. Abate-Shen, C. and M.M. Shen, *Diagnostics: the prostate-cancer metabolome.* Nature, 2009. 457(7231): p. 799-800.
96. Tiziani, S., V. Lopes, and U.L. Günther, *Early stage diagnosis of oral cancer using 1 H NMR–based metabolomics.* Neoplasia, 2009. 11(3): p. 269-IN10.
97. Nishiumi, S., et al., *Serum metabolomics as a novel diagnostic approach for pancreatic cancer.* Metabolomics, 2010. 6(4): p. 518-528.
98. Sumner, L.W., P. Mendes, and R.A. Dixon, *Plant metabolomics: large-scale phytochemistry in the functional genomics era.* Phytochemistry, 2003. 62(6): p. 817-836.
99. Zhang, A., et al., *Metabolomics: towards understanding traditional Chinese medicine.* Planta medica, 2010. 76(17): p. 2026.
100. Verpoorte, R., et al., *Commentary:"A systems view on the future of medicine: inspiration from Chinese medicine?".* Journal of ethnopharmacology, 2009. 121(3): p. 479-481.
101. Jiang, Y., et al., *Recent analytical approaches in quality control of traditional Chinese medicines—a review.* Analytica chimica acta, 2010. 657(1): p. 9-18.

102. Jansson, J., et al., *Metabolomics reveals metabolic biomarkers of Crohn's disease.* PLoS One, 2009. 4(7): p. e6386.
103. Cakmak, A., et al., *Analyzing metabolomics data for automated prediction of underlying biological mechanisms.* BMC Bioinformatics (Submitted), 2009.
104. Tautenhahn, R., et al., *An accelerated workflow for untargeted metabolomics using the METLIN database.* Nature biotechnology, 2012. 30(9): p. 826-828.
105. Wishart, D.S., et al., *HMDB: the human metabolome database.* Nucleic acids research, 2007. 35(suppl 1): p. D521-D526.
106. Wishart, D.S., et al., *HMDB: a knowledgebase for the human metabolome.* Nucleic acids research, 2009. 37(suppl 1): p. D603-D610.
107. Scholz, M. and O. Fiehn. *SetupX--a public study design database for metabolomic projects*. in *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*. 2006.
108. Choi, C., et al., *Combination of a data warehouse concept with web services for the establishment of the Pseudomonas systems biology database SYSTOMONAS.* Journal of Integrative Bioinformatics, 2007. 4(1).
109. Haug, K., et al., *MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data.* Nucleic acids research, 2012: p. gks1004.
110. Horai, H., et al., *MassBank: a public repository for sharing mass spectral data for life sciences.* Journal of mass spectrometry, 2010. 45(7): p. 703-714.
111. Hummel, J., et al., *The Golm Metabolome Database: a database for GC-MS based metabolite profiling*, in *Metabolomics*. 2007, Springer. p. 75-95.
112. Ludwig, C., et al., *Birmingham Metabolite Library: a publicly accessible database of 1-D 1H and 2-D 1H J-resolved NMR spectra of authentic metabolite standards (BML-NMR).* Metabolomics, 2012. 8(1): p. 8-18.
113. Wang, J., et al., *A Platform to Identify Endogenous Metabolites Using a Novel High Performance Orbitrap MS and the mzCloud Library*. blood. 4: p. 2.
114. Markley, J.L., et al. *New bioinformatics resources for metabolomics*. in *Pacific Symposium on Biocomputing*. 2007.
115. Cui, Q., et al., *Metabolite identification via the madison metabolomics consortium database.* Nature biotechnology, 2008. 26(2): p. 162-164.
116. Denkert, C., et al., *Mass spectrometry–based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors.* Cancer research, 2006. 66(22): p. 10795-10804.
117. Kind, T., et al., *A comprehensive urinary metabolomic approach for identifying kidney cancer.* Analytical biochemistry, 2007. 363(2): p. 185-195.
118. Ganti, S. and R.H. Weiss. *Urine metabolomics for kidney cancer detection and biomarker discovery*. in *Urologic Oncology: Seminars and Original Investigations*. 2011. Elsevier.
119. Ikeda, A., et al., *Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer.* Biomedical Chromatography, 2012. 26(5): p. 548-558.
120. Nishiumi, S., et al., *A novel serum metabolomics-based diagnostic approach for colorectal cancer.* PLoS One, 2012. 7(7): p. e40459.
121. Bertini, I., et al., *Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer.* Cancer research, 2012. 72(1): p. 356-364.
122. Carrola, J., et al., *Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine.* Journal of proteome research, 2010. 10(1): p. 221-230.
123. Rocha, C.M., et al., *Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma.* Journal of proteome research, 2011. 10(9): p. 4314-4324.
124. Jordan, K.W. and L.L. Cheng, *NMR-based metabolomics approach to target biomarkers for human prostate cancer.* 2007.
125. Trock, B.J. *Application of metabolomics to prostate cancer*. in *Urologic Oncology: Seminars and Original Investigations*. 2011. Elsevier.
126. Chen, T., et al., *Serum and urine metabolite profiling reveals potential biomarkers of human*

hepatocellular carcinoma. Molecular & Cellular Proteomics, 2011. 10(7): p. M110. 004945.
127. Griffin, J.L. and R.A. Kauppinen, *A metabolomics perspective of human brain tumours.* FEBS journal, 2007. 274(5): p. 1132-1139.
128. Chan, E.C.Y., et al., *Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS).* Journal of proteome research, 2008. 8(1): p. 352-361.
129. Denkert, C., et al., *Metabolomics of human breast cancer: new approaches for tumor typing and biomarker discovery.* Genome Med, 2012. 4(4): p. 37.
130. Yang, C., et al. *Comparative metabolomics of breast cancer*. in *Pacific Symposium on Biocomputing*. 2007.
131. Weljie, A.M., et al., *1H NMR metabolomics identification of markers of hypoxia-induced metabolic shifts in a breast cancer model system.* Journal of biomolecular NMR, 2011. 49(3-4): p. 185-193.
132. Somashekar, B.S., et al., *Magic angle spinning NMR-based metabolic profiling of head and neck squamous cell carcinoma tissues.* Journal of proteome research, 2011. 10(11): p. 5232-5241.
133. Beger, R.D., *A review of applications of metabolomics in cancer.* Metabolites, 2013. 3(3): p. 552-574.
134. Wei, J., et al., *Salivary metabolite signatures of oral cancer and leukoplakia.* International Journal of Cancer, 2011. 129(9): p. 2207-2217.
135. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* science, 2009. 324(5930): p. 1029-1033.
136. Gatenby, R.A., et al., *Acid-mediated tumor invasion: a multidisciplinary study.* Cancer research, 2006. 66(10): p. 5216-5223.
137. Klement, R.J. and U. Kämmerer, *Is there a role for carbohydrate restriction in the treatment and prevention of cancer.* Nutr Metab (Lond), 2011. 8(75): p. 75.
138. Rotin, D., B. Robinson, and I.F. Tannock, *Influence of hypoxia and an acidic environment on the metabolism and viability of cultured cells: potential implications for cell death in tumors.* Cancer research, 1986. 46(6): p. 2821-2826.
139. Kuhajda, F.P., *Fatty-acid synthase and human cancer: new perspectives on its role in tumor biology.* Nutrition, 2000. 16(3): p. 202-208.
140. Pizer, E.S., et al., *Inhibition of fatty acid synthesis induces programmed cell death in human breast cancer cells.* Cancer research, 1996. 56(12): p. 2745-2747.
141. Mashima, T., H. Seimiya, and T. Tsuruo, *De novo fatty-acid synthesis and related pathways as molecular targets for cancer therapy.* British journal of cancer, 2009. 100(9): p. 1369-1372.
142. Sreekumar, A., et al., *Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression.* Nature, 2009. 457(7231): p. 910-914.
143. Jentzmik, F., et al., *Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours.* European urology, 2010. 58(1): p. 12-18.
144. Spratlin, J.L., N.J. Serkova, and S.G. Eckhardt, *Clinical applications of metabolomics in oncology: a review.* Clinical Cancer Research, 2009. 15(2): p. 431-440.
145. Hori, S., et al., *A metabolomic approach to lung cancer.* Lung Cancer, 2011. 74(2): p. 284-292.
146. Chen, J., et al., *Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations.* Analytica Chimica Acta, 2009. 650(1): p. 3-9.
147. Guan, W., et al., *Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines.* BMC bioinformatics, 2009. 10(1): p. 259.
148. Klein, M.S., et al., *NMR metabolomic analysis of dairy cows reveals milk glycerophosphocholine to phosphocholine ratio as prognostic biomarker for risk of ketosis.* Journal of proteome research, 2011. 11(2): p. 1373-1381.