CHAPTER: 01

INTRODUCTION TO METABONOMICS

[1]Dr. SHATAKSHI SRIVASTAVA
[1]Apeejay Stya University, Sohna, Gurugram

[2]Prof. RAJA ROY
[2]Ex- Director, CBMR, SGPGIMS Campus

Ch.Id:-ASU/GRF/EB/IASINAM/2022/Ch-01
DOI: https://doi.org/10.52458/9789391842741.2022.eb.grf.asu.ch-01

"The Experimenter- Its an experience like no other experience I can describe, the best thing that can happen to a scientist, realizing that something that's happened in his or her mind exactly corresponds to something that happens in nature. It is startling every time it occurs. One is surprised that a construct of one's own mind can actually be realized in the honest-to-goodness world out there. A great shock, and a great, great joy." - Leo Kadanoff

1.1 METABOLISM, METABOLOME AND METABONOMICS

Metabolism is defined as a cascade of various complex cellular processes for maintaining and sustaining the homeostasis or internal environment of an organism. The metabolism is basically of two types: catabolism (breaks down products releasing energy) and anabolism (utilizes energy for constructing cellular infrastructure and other vital functions). These metabolic processes in human cells generate energy from three basic classes of nutrients: carbohydrates, lipids and amino acids which are taken through diet and mediate the synthesis of diverse essential molecules which cannot be taken in diet. It is well known that amino acids actively participate in protein biosyntheses which play an active role in cell signalling, immune responses, active transport across membranes, and cell cycle. Carbohydrates have an important role of storage and transport of energy in human body while lipids are not only a rich source of energy but also the integral part of cell membranes. Whereas these three molecules play an important role in energy metabolism and other important growth processes, other molecules like glutathione also exist which play an important role in maintaining the body homeostasis by intoxicating the poisonous substance or reactive oxygen species which get generated endogenously or are consumed exogenously. The first definition of metabolome was quipped by Oliver et al. in 1998 which defined "metabolome as the quantitative complement of all the low molecular weight molecules present in cells in a particular physiological or developmental state" (Oliver, Winson et al. 1998). Later on Beecher proposed that "metabolome consists only of those native small molecules (defined as non-polymeric compounds) that are participants in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell" (Beecher 2004). Thus the complete set of small molecules metabolites which are active in various metabolic processes and vary dynamically in a specific physiological condition is metabolome. It is the final down-stream product of gene expression and proteins. Any metabolic and physiologic regulatory derangements in response to a pathological condition will result in changes in dynamic metabolic processes and accumulation of metabolic end-products. This leads to loss of body homeostasis and will consequently lead to metabolic perturbations at the site of organ and in bio-fluids. The qualitative and quantitative measurement of these perturbations is metabonomics, as to put it in words of Prof. Lindon "the quantitative measurement of the multiparametric

metabolic response of living systems to pathophysiological stimuli or genetic modification". Past few decades have immensely contributed towards the identification of small molecule composition of different biological systems which consequently enables an unprejudiced investigation of changes in its metabolic status as a result of disease, toxic insult, genetic manipulation, environmental stress, etc. Consequently, metabonomics has become an emerging 'omics' discipline that measures the dynamic multi-parametric response of a living system's metabolome to genetic modifications or pathophysiological stimuli (Nicholson et al. 1999; Fiehn 2002). Several terms have been proposed for better understanding of cellular responses of the metabolome, which include metabolomics, metabonomics, metabolic and metabolite profiling, metabolite target analysis and metabolic fingerprinting (Fiehn 2002; Goodacre, Vaidyanathan et al. 2004). In contrast to classical biochemical approaches which usually focus on single metabolites or metabolic reactions (Kaddurah-Daouk et al. 2008), metabonomics allows comprehensive and simultaneous assessment of alterations occurring in metabolic pathways and gain an understanding of the changes in a biological system (Hood et al. 2004), since these fluctuations are the end-product of complex interaction between the genome, transcriptome, proteome, and the environment (Kaddurah-Daouk et al. 2008).

1.2 CONCEPTUAL BASIS FOR METABONOMICS

Any phenotypic representation of pathology is an after-effect of altered metabolic pathways. These alterations may arise from over- and/or under- activity of enzymes involved in the pathway and thus metabolic fluxes of affected pathway gets deferred. For example, in maple syrup urine disease (an inborn error of metabolism), defect in branched chain keto-acid decarboxylase activity causes alterations in metabolism of branched chain amino acids (BCAA): valine, leucine and isoleucine. The resultant variations were readily observable in metabolite profile of urine of such patients with BCAA as highly excreted products (Holmes, Foxall et al. 1997). Teusink et al. (1998) had seconded the above observation by proposing a model and suggested that the flux rate of a specific metabolic pathway is not a function of a single enzyme activity but it gets propagated throughout the metabolic network (Teusink, Walsh et al. 1998). Therefore, multivariate statistical analysis should be a part of metabolomics which indirectly measures the altered concentration coefficients of metabolites (Raamsdonk, Teusink et al. 2001). Metabolism consists of many small modules that are highly connected into a hierarchy where a few metabolites such as pyruvate and acetyl-CoA connect many such modules (Ravasz, Somera et al. 2002).

1.3 METABONOMICS APPROACH TO DISEASE DIAGNOSIS

During the initiation of Human Genome Project (HGP), it was envisioned that understanding of human genome will revolutionize the field of targeted medicine by providing an insight for the genetic disorders and predispositions before they really manifest themselves. But at its end it was realized that only genomics is not sufficient to serve the bigger purpose and therefore newer approaches are required for gaining the knowledge of gene-expression and phenotype correlation. This will provide a better understanding of functional consequences of an abiotic or biotic stimulus and will consequently assist the progress of medicine and health sciences. The quest for newer methods led to many 'omics' techniques like nutrigenomics, neuropharmacogenomics, toxiconomics etc. but only transcriptomics, proteomics, genomics and metabonomics contributed significantly, in terms of information, to the systems biology.

Genomics involve study of entire genome of an organism along with its gene-mapping, intra-genomic phenomena and gene interactions. The genome can be defined as the complete set of genes present in a cell. Thus, genomics is the study of genetic make-up of organisms. It includes three major sub-groups: (i) functional genomics (study the function of numerous genes), (Soininen, Kangas et al.) comparative genomics (to compare the genes among different organisms) and (iii) structural genomics (to generate the 3-D structure of one or more proteins from each protein family, thus offering clues to their function). However, after the discovery of mechanisms like RNA interference in switching off the specific gene, it is clear that genomics has its own limitations (Fire, Xu et al. 1998). Transcriptomics is the study of changes in gene-expression level by the measurement of messenger ribonucleic acid (mRNA) and non-coding RNA levels in a system under specific conditions but it cannot be considered as the only suitable approach due to its limitations in analytical precision and high cost. Proteomics involve the study of changes in proteins and their identification (Teusink, Walsh et al. 1998) but measurement of proteins is very labor-intensive due to their variable physical properties. Also, the enzyme concentrations may have a small impact on the metabolic fluxes which may not be statistically significant (Nicholson, Lindon et al. 1999). The exposure of environmental or chemical carcinogens causes numerous genetic mutations. However, all these mutations are not responsible for triggering cancer cascade of metabolic processes and may remain silent without showing any phenotypic effect. The measurement of effect of an external or internal stimulus on gene expression by other three 'omic' techniques shows low correlation between the gene-expression and phenotype. Thus, there is a need for defining the resultant phenotype from a genetic modification to understand how the metabolic processes work at

transcription or proteome level. In such cases, metabolomics based studies for biomarker identification have an upper-hand to the genomics- or transcriptomics- based markers because metabolites are the end-products of cellular processes unlike proteins and mRNA which are involved at intermediate levels and secondly, small metabolites are the closest link to phenotype. Therefore, newer modalities like mass spectrometry combined with gas chromatography (GC-MS) and NMR spectroscopy were introduced for simultaneous detection of numerous metabolites present in a tissue. These metabolites are part of normal cells as well but their concentration gets aberrant in pathological state. Nowadays, GC-MS and NMR spectroscopy are commonly used for metabonomics studies. However, GC-MS has more sensitivity in terms of concentration but also has longer sample pre-processing while NMR needs minimal sample pre-processing, non-destructive, reduced time for acquisition and accurate quantitation of metabolites is possible. In past two decades, nuclear magnetic spectroscopy along with its imaging variants has resulted in numerous in vitro, ex vivo and in vivo studies of associated cancer biomarkers for unravelling the complexity of metabolic information present in pathological conditions.

1.4 METABONOMICS AND CANCER RESEARCH

Cancer is a dreadful disease with high rate of mortality and poor morbidity. Even the most advance treatment modalities are not able to significantly increase the five-year survival rate of patients. This limitation is attributed to the poor prognosis of disease due to different responses of similar kind of tumors after the application of adjuvant therapy. Thus, clinical carcinomas are complex and present dilemmas in prognosis and diagnosis. Therefore, there has been an incessant quest for newer biomarkers in clinical medicine which can illuminate the following aspects: (i) differentiation among the malignant and non-malignant states of a tissue, (Soininen, Kangas et al.) if malignant, then identification of patients at minimal risk (low grade tumors) and at highest risk and (iii) study of response of the given adjuvant therapy to patients with high grade tumor. The criteria of a biomarker clearly indicates that only a single biomarker will not suffice and a broader spectrum of molecules associated with different biochemical aspects (tumor heterogeneity, different degree of cell differentiation and cell proliferation) of cancer will be required for improved prognosis, diagnosis and better survival rates.

Cancer cells exhibit a totally different phenotype as compared to the matched control cells as the initiation of carcinogenesis involves the accumulation of genetic mutations with course of time which results in overcoming the anti-cancer defence mechanisms of a cell leading to a growth advantage to the neoplastic cell. All such

phenotypes display common characteristics of tumorigenesis which involve inhibition of apoptosis (Costello and Franklin 2005), higher glycolytic rates, altered cellular bioenergetics, enhanced cell proliferation and related biochemical processes (high protein biosynthesis, nucleic acid biosynthesis etc.) which are represented by common metabolic markers like lactate, glucose, taurine, acetate, branched chain amino acids (valine, leucine, lysine and isoleucine), choline containing compounds and so on. However, at the same time, specific endogenous metabolites are markers of cancer of a specific tissue like citrate in prostate cancer (Dittrich, Kurth et al. 2012) and higher N-acetyl aspartate in brain cancer (Tanaka, Naruse et al. 1986; Usenius, Kauppinen et al. 1994; Warren 2004). These environmental factors will lead to the changing of tumor related genes and proteins, all of which will be reflected as changes in the metabolome (Hanahan and Weinberg 2000; Danial, Gramm et al. 2003; Cairns, Papandreou et al. 2006). Understanding of the metabolomic changes associated with uncontrolled proliferation of cells, metastasis, apoptosis and differentiation that characterize tumors remains an important undertaking that will advance early detection of the pathology.

Nuclear magnetic resonance (NMR) spectroscopy based metabonomics is an effective method to study the perturbations in metabolite composition of biological samples like tumor tissues and bio-fluids for early diagnosis of cancer and other diseases. The rapid development in field of NMR has made in vitro, ex vivo and in vivo analysis of main tumor and peri-tumoral region, more feasible, which has been proved by over ten thousand publications in past one decade. The metabolic derangements reflect the altered cellular activities in malignant state which could be easily probed by NMR spectroscopy. The in vitro and ex vivo NMR studies have been conducted on bio-fluids (blood serum/plasma, urine, cerebro-spinal fluid etc.), resected tumor biopsies, cultured cancer cell lines as well as in-vivo malignant tissues by using various advanced NMR techniques including recently developed high-resolution magic angle spinning (HR-MAS) and magnetic resonance spectroscopy and imaging (MRSI) methods. The NMR method of biomarker identification can be summarized as: metabolites are identified with the help of two-dimensional NMR spectroscopy, spiking method and NMR data banks available freely on internet (Screening method for biomarker). The metabolites thus identified are subjected to statistical tests for calculation of significance levels of alterations. These metabolic datasets can be readily mined using a range of pattern recognition techniques, including hierarchical cluster analysis, principal components analysis, partial least squares and neural networks, with the combined approach for fruitful information (validation for robustness of biomarker identified). The information can be used for differential diagnosis of a disease and its aggressiveness under clinical observations (biomarker translation).

1.5 GENERAL ASPECTS OF ^{1}H HR-MAS MR SPECTROSCOPY

^{1}H NMR spectroscopy plays a considerable role in various studies of toxic effects (Anthony, Gartland et al. 1994; Anthony, Sweatman et al. 1994; Anthony, Rose et al. 1995; Garrod, Humpher et al. 2001; Bundy, Lenz et al. 2002; Duarte, Stanley et al. 2005), cancer (Goodacre, Vaidyanathan et al. 2004; Warren 2004; Ma, Zhang et al. 2012) and inborn error of metabolism (Holmes, Foxall et al. 1997; Moolenaar, Engelke et al. 2003; Yap, Angley et al. 2010). The in vitro NMR analyses of bio-fluids and tissue extracts provide tissue biochemistry and also help in analyzing the cancer-related biomarkers present in tissues. In case of solid or semi-solid tissues, a nuclear spin experiences three major types of interactions: chemical shift anisotropy, dipolar interactions and quadrapolar interactions. In liquids, the chemical shift anisotropy and dipole-dipole couplings get averaged out to zero but in solids like tissues, both factors influence the spectrum largely resulting in broad resonances, therefore, tissue extraction processes came into existence. The perchloric acid method of extraction from tissues extracts all the water soluble metabolites like lactate, alanine, choline, glycerophosphocholine and other small metabolites, while the lipid extracts give an idea about the lipid composition of the tissues. However, tissue extraction procedures on cancer biopsies provided an insight of its biochemical processes, but at the cost of tissue destruction and modifications in its composition and also consume more time. These limitations were overcome by a technique HR-MAS NMR spectroscopy, introduced independently by Andrew et al. (Andrew, Bradbury et al. 1958; Andrew, Bradbury et al. 1959) and Lowe (Lowe 1959). In HR-MAS NMR spectroscopy, intact tissues are used for data acquisition in such a way that minimizes the effect of line broadening caused due to sample heterogeneity and residual anisotropic NMR parameters. Rapid spinning of the sample, speed (in kHz) depending upon the field strength, at an angle of 54.7° relative to the axis of magnetic field minimizes the line-broadening effects and so the loss of resolution and information contained in a tissue. Earlier applications of HR-MAS NMR were confined to the solid-state investigations (Stejskal and Memory 1994) but soon after the development of MAS probes with a deuterium channel for lock signal detection, the HR-MAS NMR based metabonomic studies came into realization. LL Cheng had first reported the use of magic angle spinning on malignant lymph nodes of rats with mammary adenocarcinoma (R13762) (Cheng, Lean et al. 1996). Since then a number of research articles and reviews were published regarding the metabolite identification and differentiation of benign and malignant human tissue specimens (Cheng, Chang et al. 1998; Tugnoli, Mucci et al. 2004) by ^{1}H NMR HR MAS spectroscopy. We, therefore, wish to define current status of cancer metabolomics with respect to the oral cancer, gall bladder cancer and urinary bladder cancer, and its role in identifying the metabolic

alterations in known pathways which occur during the oncological developments and progressions. The conditions of sample storage, role of additives and the experimental setup play an important role in defining the accurate and proper metabolic status of the bio-fluids, cell lines and tissue biopsies.

1.6 BIOLOGICAL SAMPLES

The human biofluids represent a complex and dynamic array of metabolites present in constant equilibrium for the proper homeostasis maintenance. In relation to the metabolic pathway being investigated during pathology, it is important to consider what type of biological sample would be appropriate for metabonomic study. Biological samples usually range from highly invasive and specific (e.g. cerebrospinal fluid, tissue biopsy, bile), to more general biological samples that are easy to collect and non-invasive for human patients (e.g. urine). The other biological samples include the bio-fluids like serum, urine, ascitic fluid and similar kinds of endogenous fluids of body, the cell lines and tissue biopsies. The storage conditions of these samples involving temperature, additives and pH critically affect the metabolite concentration and their stability and therefore, care must be taken to avoid the handling related alterations as these can mask the original cancer induced biochemical variations. The following section reviews different conditions of sample handling for the best results by retaining the sample stability and thus avoiding the analytical and experimental bias.

1.7 BLOOD- SERUM AND PLASMA

Blood is an omniscient bio-fluid of human body which contains a complex matrix of proteins, carbohydrates, hormones and blood cells. Its main functions involve transportation of gases, metabolic products, hormones and proteins from one organ to another. It is through the medium of the circulating blood that the constancy of the internal environment is maintained and disease processes or abnormalities anywhere in the body are reflected to various extents in altered blood composition. For disease biomarker studies, it becomes necessary to separate blood cells from the fluid matrix as the direct study of whole blood poses serious limitations to ^{1}H NMR studies. Therefore, two kinds of sub-fluids are separated from blood viz. serum and plasma. The process of separation either includes direct coagulation and centrifugation of blood, which results in a sub-fluid known as serum or by addition of an anti-coagulant (EDTA or heparin) in blood followed by centrifugation, which results into plasma. Thus, serum is a straw coloured sub-fluid of blood which lacks white blood corpuscles, red blood corpuscles and fibrinogen (clotting factor) while, plasma is similar to serum except the presence of clotting factor that makes it more viscous in nature. The constituents of serum include

proteins (glycoproteins, immunoglobulins, albumin and lipoproteins), electrolytes, antibodies, antigens, hormones, amino acids, fatty acids and any exogenous substances like drugs. In metabonomics, serum and plasma had been widely explored for three molecular windows in disease pathologies like nephrotoxicity, hepatotoxicity, inborn errors of metabolism, inflammatory diseases, cancer, rheumatoid arthritis, diabetis mellitus, coronary heart disease, atherosclerosis and bipolar disorder which are: lipids, lipoprotein sub-classes and water soluble low molecular weight small metabolites (Brindle, Antti et al. 2002; Brindle, Nicholson et al. 2003; Soedamah-Muthu, Colhoun et al. 2003; Beckonert, Keun et al. 2007; Soininen, Kangas et al. 2009; Sussulini, Prando et al. 2009). Moreover, the ^{1}H NMR spectra of serum and plasma looked similar, therefore, the protocols for sample handling and storage may be applicable equally on both bio-fluids. Long-term prospective or retrospective epidemiological studies with multiple withdrawals of blood in single subjects, however, require adequate plasma/serum sample storage. Thus, proper handling of these samples plays an important role in compositional analysis and its accuracy.

An important aspect of metabonomics is to assure that the biochemical variations arising in serum/plasma are due to the disease pathology only and not due to any pre-analytical error. The identification, removal or minimisation of these errors limits the analytical bias in a study and will provide a valid data which is a pre-requisite for the successful clinical diagnosis. In case of serum, several steps are involved from collection of blood to the final NMR analysis which may have significant impact on the final spectra and the concluding results. The factors involve (i) time and temperature of separation of serum/plasma from blood, (ii) time and temperature of storage of serum/plasma after separation (iii) duration of effective storage, (iv) number of freeze-thaw cycles, sample has undergone and (v) proper pH maintenance and optimum buffer solution addition for proper line widths in a spectrum. The presence of either of these factors during acquisition would cause erroneous interpretation of data and result in misleading biomarkers. The enzymatic activity and/or chemical reactions of a biological matrix may alter metabolic profile and no longer represent the actual state of physiology. Therefore, the identification and impact of these factors on biological fluids is needed. Numerous biochemical studies were performed on different aspects of serum lipid analysis (Howland 1990; Koukoulaki, O'Donovan et al. 2008; Bernini, Bertini et al. 2011) and small metabolites (Brindle, Antti et al. 2002; Bernini, Bertini et al. 2009; Sussulini, Prando et al. 2009; Carrola, Rocha et al. 2010; Srivastava, Roy et al. 2010). Few of recent NMR based studies regarding about sample stability under best conditions have been reviewed here.