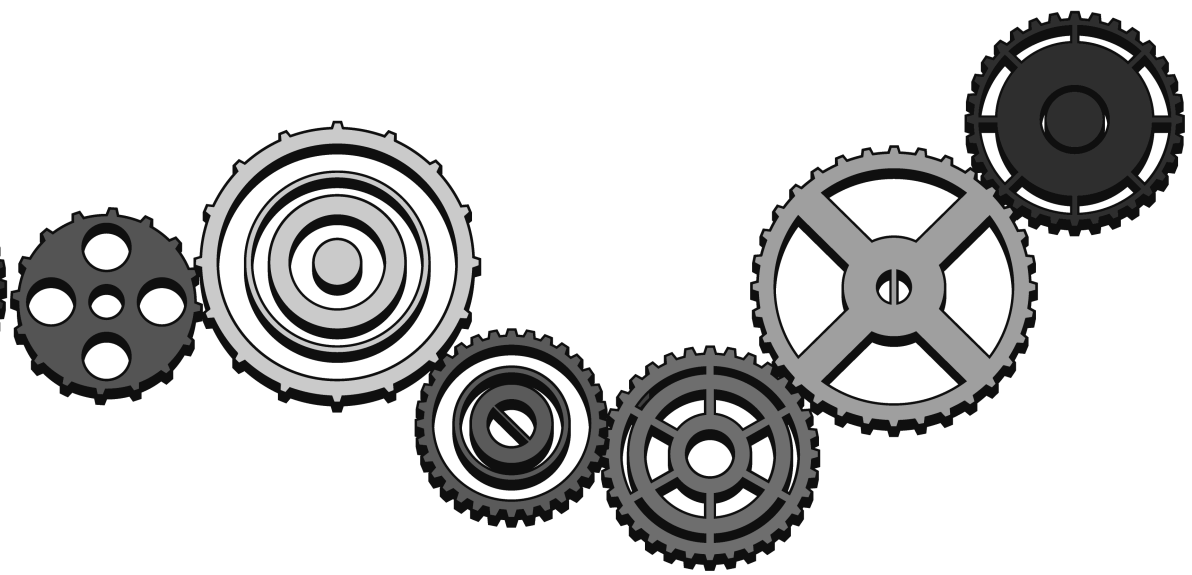
Introduction



CLINICAL METABOLOMICS: AN OVERVIEW

Metabolomics is an example of a modern multidisciplinary research area. It has its roots in the technological revolutions of the sixties, the impressive breakthrough of the computer industry of the seventies and the re-birth of the holistic views in biology triggered by such a major undertaking as the human genome project. Metabolomics is a post-genomic discipline aiming at the study of metabolites: the end points and the intermediate products of the metabolism. Historically, the understanding of intracellular metabolic pathways and fluxes in model (mostly unicellular) organisms has been a dominant trend in the field [1]. However, clinical metabolomics, which has emerged as a spin-off of the 'mainstream application', focuses on the research of the metabolic regulation of the entire human organism and aims at the development of novel diagnostic/prognostic tools, prediction of patients responses to the treatment, exploration of the human metabolic individuality *et cetera.* Metabolomics of model systems and clinical metabolomics use the same analytical strategies but they differ with regard to experimental design and data treatment/interpretation.

Clinical metabolomics relies on the analysis of body fluids as the main source of information. The metabolic composition of the various body fluids provides the most accurate description of the individual physiological phenotype and as such is an essential element in the development of the concept of personalized medicine, the prediction of the susceptibility of an individual to a disease, the prediction of a state of disease or/and the outcome of a treatment. In a way, the most immediate goal of the clinical metabolomics is to reach the level of the technical confidence for being accepted as an independent measure in epidemiological studies.

Indeed, classical epidemiology operates within a limited space of clinical measures/readouts while metabolomics-based 'molecular epidemiology' is contributing by bringing in a comprehensive description of the chemical phenotype. Thus, instead of linking pre-existing knowledge to a certain level of the biological regulation (genes, gene expression protein expression), metabolomics offers a transverse approach to the problem evaluating the global outcome without taking into account the effect of any single variable.

At the time when metabolomics was just an emerging area, many studies were logically focused on revealing the potential of various analytical techniques and of the data processing. At the current state-of-art, metabolomics studies do not only provide class separation (e.g. diseased versus controls, response to a therapy and/or disease progression) but also a critical identification of those metabolites involved in the separation and their role in a meaningful biological context which can be used for further validation.

The metabolomics approach has been extensively used in the profiling of cancer phenotypes with the main intention to provide a valuable tool for highly needed alternative, earlier and molecule-based detection methods. A big portion of the present studies is dedicated to rapid diagnosis and monitoring of therapy. Probably, breast cancer represents the best example of metabolomics applied to diagnostic purposes: this tumor type showed at tissue level an elevated phosphocholine level, low glycerophosphocholine and low glucose compared with healthy tissue or benign tumors which allowed differentiation with a sensitivity of the 83% and a specificity of 100% [2]. Furthermore, it was calculated that magnetic resonance spectroscopic imaging analysis performed on choline-positive tissues could have prevented a biopsy in 66% of the cases [3]. Metabolite profiling has also contributed to the research of early diagnostic markers in lung cancer which, according to the World Health Organization, kills more than any cancer – a situation that is estimated

to continue till 2030. For this pathological condition, standard screening methods such as chest radiography and sputum histology have not increased the survival rate [4]. However, it has been proved that high risk cytochrome p450 genotypes may accelerate the catabolism of naturally occurring volatile organic compounds (VOCs) in the breath of lung cancer patients. Metabolite-based detection techniques have pointed at alkanes, alkenes and benzene derivatives in the classification of lung cancer patients' breath [5-7]. The additional investigation of body fluids showed an increased concentration of two aldehydes –hexanal and heptanal– in blood [8].

Among all the clinical conditions, infectious diseases deserve special attention: the metabolome of an infected organism, in fact, represents not only the metabolic phenotype of the host and/or the pathogen but it also depicts their cross-talk. Furthermore, for a comprehensive view of the metabolic pattern, multiple factors have to be taken into account such as the risk of exposure and the host's individual susceptibility to a given infection. Moreover, the word 'pathogen' encompasses a variety of infectious agents with very diverse invasion properties which are able to activate different pathways of the host immune system. The bulk of the literature dedicated to the metabolomics of infectious diseases concerns animal studies. For example, explorative studies of parasitic infections have been dedicated to the understanding of the pathogen metabolism and/or the host response after inducing the infection. Explicative examples are the most widespread parasite infections, namely malaria and schistosomiasis. For malaria, it has been shown that the dominant change in the urinary metabolic profile may provide a sensitive diagnostic tool for *Plasmodium* infection and for the evaluation of the malaria progression in mice [9]. A particular relevant finding is the specific conversion of arginine to ornithine induced by the parasite [10]. Tissue material has also provided interesting data for the understanding of the metabolic pathways. In fact, most tissues obtained from *S. mansoni*-infected mice were characterized by high levels of amino acids (leucine, isoleucine, lysine, glutamine and asparagine) and high levels of membrane phospholipid metabolites (glycerophosphoryl choline and phosphoryl choline). Additionally, low levels of energy-related metabolites (lipids, glucose and glycogen) were observed in ileum, spleen and liver samples of infected mice. These findings show that a *S. mansoni* infection causes a clear disruption of the metabolism in some tissues, which are in accordance with the previously reported indications in a biofluid (urine) [11].

METABOLOMICS OF THE URINARY TRACT INFECTION (UTI)

Despite the encouraging results achieved in animal model and *in vitro* studies, real clinical metabolomics studies in the field of infectious diseases are still lacking. Although large population studies are nowadays a matter of fact, the step from an ideal and controlled situation represented by a 'synthetic' study to human situation still appears tedious. In this respect, urinary tract infection (UTI) represents a real life situation in a clinical context and the comprehensive study of the UTI-induced changes in the metabolic pattern may offer the possibility to extensively investigate such a complex infection in man. This clinically well-defined disease represents a good model for the development of an analytically based scoring system of disease severity or/and evaluation of the effectiveness of a treatment. Urinary tract infection (UTI) is an inflammatory response to the presence of single or multiple species of pathogens in the urinary tract, with a high incidence of *Escherichia coli*, *Pseudonomas aeruginosa*, *Klebsialla pneumoniae*, *Enterobacter sp.*, *Acinetobacter sp.*, and *Proteus mirabilis*.

Despite the efficacy of the medical care and antibiotic treatment, UTI still contributes significantly to the annual morbidity and mortality rate in Western countries.

In a first approximation, UTI can be grouped into acute uncomplicated UTI (non-pregnant young women with cystitis but no evidence of urological abnormalities) and complicated UTI (UTI in males, all febrile UTI syndromes and those with urological abnormalities). In this respect, fever reflects the presence of a tissue invasive disease such as pyelonephritis, prostatitis or the urosepsis syndrome. The clinical management of UTI is relatively straightforward, but obtaining guidance for the duration of the treatment or for scaling of the morbidity inflicted to, for example, prostate or kidney is not such a simple task.

The identity of the causal pathogen is confirmed by a urine culture test. This typically needs 24-48 hours before the results are available. An alternative diagnostic method routinely used in clinical laboratories is a nitrite dipstick test: a fast and cheap method but with a high incidence of false-negative results, especially in the case of a Gram-positive bacteria pathogen.

An early and accurate diagnostic method would be valuable for a correct and prompt medical treatment, especially in view of the fact that epidemiological data suggest that a poorly diagnosed UTI or a recurrent UTI due to antibiotic-resistant pathogens is associated with a higher mortality rate [12, 13]. To this end, the scientific literature presents a growing range of NMR- and MS-based methods for fast identification of the pathogen [14]. It has been shown that uropathogens present a specific metabolite pattern which under controlled conditions can be used for accurate identification within six hours. For instance, *in vitro* studies have revealed specific metabolic 'signatures' for the different uropathogens: *E. coli* produces lactose from the lactate metabolism, *K. pneumoniae* produces 1,3 propanediol from the glycerol metabolism, *P. aeruginosa* produces 6-hydroxy nicotinic acid from the nicotinic acid metabolism and *P. mirabilis* produces 4-methylthio-2- oxobutyric acid from the methionine metabolism.

It has become evident that a molecule-based approach has practical value for the development of novel diagnostic tools. Furthermore, metabolomics creates new possibilities for the investigation of the host-pathogen cross-talk. This new aspect might help in more accurately scaling the progression of morbidity, a subject which has been largely ignored till this moment, and can help in the evaluation of the risk of the recurrent UTI. In fact, from a clinical point of view, recurrent and complicated UTI are more challenging problems since they can lead to more serious and life-threatening problems till organ failures.

The modeling of the metabolic phenotype has matured and improved over the years and is, nowadays, influencing translational medicine through the screening of a large cohorts of individuals with the ambition to deliver a classification in many different scenarios. An obvious application is the investigation of metabolic diseases often caused by a dysregulation in the metabolism of small molecules. It has been shown how a multiplatform approach to the metabolome-wide analyses of diabetes leads to a more sophisticated classification of the disease and the evaluation of the disease progression. Key observations could be linked to a series of metabolic plasma perturbations linked to kidney dysfunctions, lipid metabolism and gut microflora [15]. One of the first examples of a metabolome-wide association study has analysed risk factors (diet and blood pressure) in relation to exploratory spectroscopic data allowing the identification of good candidates for the metabolic signature of disease risk in different populations [16].

On the basis of the findings for metabolic diseases, a multiplatform explorative study could be expected to reveal the uropathogen expression and, even further, might uncover the response of the host and the complex signaling between the host and the (uro)pathogen. The so defined molecular pattern might lead to a deeper knowledge of the metabolome of UTI patients and thus to a more accurate disease classification and better understanding of the severity of the morbidity.

ANALYTICAL TECHNOLOGIES IN METABOLOMICS

The last decades of 20th century are often called the 'years of biology', a time when biomedical sciences made an enormous progress. Analytical and technical solutions, enabling large scale analysis of the 'biopolymers' (DNA, RNA and proteins), have played an important part in this phenomenon. The term 'biopolymer' implies the biological origin of a molecule and the limited number of monomeric units it consists of. A restricted number of the building blocks confines the chemical diversity of the oligomers (the analysis of which constitutes the basis of the analytical workflows) and at the same time one of the major analytical challenge, for example, in protein analysis remains the dynamic range. For metabolomics, dynamic range is just one of the challenges. The diversity of the physico-chemical properties of the metabolites is another one. Indeed, unlike the terms polynucleotides and polypeptides, the term 'metabolites' does not refer to a particular chemical family; even more, the most logical definition of metabolites is not a chemical but a functional one – the end products and intermediates of metabolism. Thus, metabolites belong to different chemical families and have a diverse size and polarity; their 'apparent simplicity' hides a concrete obstacle for the standardization of the analysis methods. Moreover, none of the existing analytical techniques provides a coverage of the entire metabolic space and therefore the analysis of the metabolome demands a multiplatform approach.

For metabolite profiling purposes, Nuclear Magnetic Resonance spectroscopy (NMR) and Mass Spectrometry (MS) are the main techniques applied.

Since the early 80's, NMR has been the first choice in metabolomics studies aiming at the analysis of body fluids [17, 18]. NMR is a highly robust and reproducible technology based on the detection of the resonance frequency of the atoms' nuclei when these are exposed to a magnetic field. Due to the high abundance of the ^{1}H nucleus, ^{1}H-NMR has been widely employed for metabolome profiling: any hydrogen-containing molecule in body fluids and tissues is potentially detectable. NMR is a non-destructive technique, can be largely automated and requires minimal sample preparation, all perfect characteristics for implementation in large scale clinical studies. In addition, according to its detection modality high-resolution NMR offers simultaneously structural elucidation and absolute quantification (at sub-micromolar range).

Despite the incomparable high throughput and the power of structural analysis, NMR only (un)covers a fraction of the metabolome, and thus mass spectrometry which may involve multiple ionization techniques, mass analyzers and hyphenation solutions represents a strong complementary technology. In mass spectrometry, molecules are detected on the basis of their mass over charge ratio (*m/z*) after an ionization process. Over the last decade, MS has become one of the key analytical methods in the biomedical sciences, primarily due to its potential to measure hundreds of metabolites over a wide dynamic range with a sensitivity that is several orders of magnitude higher than NMR. As a drawback, it is a

destructive technique, the metabolite profile is more susceptible to variability and results are platform-dependent. Recent improvements in achievable mass resolution and mass accuracy have enormously facilitated the task of structural elucidation of metabolites, with a mass accuracy often lower than a few ppm.

MS techniques show their best potential when used in the classical set-up of hyphenation with a separation technique which brings both benefits and complications. An MS analysis is more time consuming than NMR because of a number of reasons, such as the need of a more extensive sample preparation, a longer analysis-time and a more difficult interpretation of the resultant spectra. However, the hyphenation also provides extra information on the compounds' polarity, size or charge and reduces the complexity of the metabolic profile. The most commonly used separation techniques in combination with MS encountered in literature are liquid chromatography (LC-MS), gas chromatography (GC-MS) and capillary electrophoresis (CE-MS).

Of these, probably, the most applied hyphenation for profiling purposes is LC-MS due to its high versatility and the capability to analyze a wide variety of compounds. In addition there is a number of different columns, organic modifiers and technical solutions in order to maximize the chromatographic separation and MS ionization/detection with regard to a single compound, a certain matrix, or a specific compound family. For instance, the 'straight forward approach' encompasses the use of reversed phase columns that poorly retain polar compounds; to overcome this issue, hydrophilic interaction chromatography offers a solution. The advent of ultra high pressure LC systems (UPLC) in combination with MS analyzers has dramatically improved the chromatographic resolution, the peak capacity, and has reduced the limits of detection down to the picograms level while it also allows a better high-throughput capability than classic HPLC systems. Moreover, this type of platform allows coverage of a greater number of polar metabolites which increases its applicability and makes LC-MS an even more powerful technique for biomolecular research. Apart from improvement at the LC level, electrospray ionization (ESI) has significantly contributed to the success of LC-MS in its role of the perfect hyphenated technique for the analysis of the body fluids. For example, urine can be directly injected into LC-MS system and a plasma sample can be analyzed after minimal sample processing (protein precipitation).

GC is one of the oldest separation techniques and the first of the three separation techniques mentioned above which was used in a metabolite profiling study [19]. Despite its 'age', it still remains the only chromatographic technique able to separate, with high reproducibility, hundreds of compounds in one single run. GC-MS allows the simultaneous detection of a wide variety of analytes, although a derivatization step is often required which may be time-consuming and costly. The traditional sources for GC-MS are equipped with a vacuum stage interface – namely electronical impact (EI) and chemical ionization (CI). Under these conditions, EI fragmentation patterns have shown to be reproducible and as a result a number of open source and commercial spectral libraries are available which can be used for structural assignment. Nevertheless, the occurrence of extensive fragmentation can make the identification of unknown compounds in complex samples a challenging task. CI provides a softer ionization but the resultant spectra are strongly dependent on the experimental conditions (gas reagent, gas pressure) which makes the data less suitable for library search. The structural assignment of the compounds is, by its nature, the slowest element in an exploratory 'omics' workflow and there is always a place for new technology. Atmospheric Pressure Chemical Ionization (APCI) appears as a natural choice for a GC

platform. GC-APCI has recently been reintroduced for profiling aims after decades of oblivion. Reviving this technique has required a general optimization and definition of the analytical parameters as is shown in chapter 4 of this thesis. Moreover, one of the factors which seriously prevents GC-APCI from becoming a routine platform for explorative studies is the lack of a spectral such as extensively available for the more traditional sources. The first public spectra library for the APCI source is now available (http://metams.lumc.nl) (chapter 5) including a wide number of compounds (primary and secondary metabolites) most commonly encountered in the analysis of body fluids.

As indicated above, any single analytical detection modality brings its specific pros and cons, and thus the use of a multiplatform strategy to undertake an explorative metabolomics study of a clinical disease seems the most logical approach. UTI is almost an ideal test to prove how a metabolomics approach can offer a more comprehensive understanding of this complex clinical entity. The various analytical techniques have each worked as a magnifying lens providing a view on the metabolic signature and allowing a different perspective on the same pathological problem. The measurement of different panels of metabolites is the consequence of the complex status of the infection which represents a complex system where the host and the invader 'fight' their battle releasing specific chemical signals which may or may not be detected.

REFERENCES

1] Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, Walsh MC, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nature Biotechnology. 2001;19:45-50.

[2] Bathen TF, Jensen LR, Sitter B, Fjoesne HE, Halgunset J, Axelson DE, et al. MR-determined metabolic phenotype of breast cancer in prediction of lymphatic spread, grade, and hormone status. Breast Cancer Research and Treatment. 2007;104:181-9.

[3] Jacobs MA, Barker PB, Bottomley PA, Bhujwalla Z, Bluemke DA. Proton magnetic resonance spectroscopic imaging of human breast cancer: A preliminary study. J Magn Reson Imaging. 2004;19:68-75.

[4] Humphrey LL, Teutsch S, Johnson M. Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: An update for the US Preventive Services Task Force. Annals of Internal Medicine. 2004;140:740-53.

[5] Phillips M, Altorki N, Austin JHM, Cameron RB, Cataneo RN, Greenberg J, et al. Prediction of lung cancer using volatile biomarkers in breath. Cancer biomarkers : section A of Disease markers. 2007;3:95-109.

[6] Phillips M, Gleeson K, Hughes JMB, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. Lancet. 1999;353:1930-3.

[7] Poli D, Carbognani P, Corradi M, Goldoni M, Acampa O, Balbi B, et al. Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. Respiratory Research. 2005;6.

[8] Deng CH, Li N, Zhang XM. Development of headspace solid-phase microextraction with on-fiber derivatization for determination of hexanal and heptanal in human blood. J Chromatogr B. 2004;813:47-52.

[9] Li JV, Wang Y, Saric J, Nicholson JK, Dirnhofer S, Singer BH, et al. Global metabolic responses of NMRI mice to an experimental Plasmodium berghei infection. Journal of Proteome Research. 2008;7:3948-56.

[10] Olszewski KL, Morrisey JM, Wilinski D, Burns JM, Vaidya AB, Rabinowitz JD, et al. Host-Parasite Interactions Revealed by Plasmodium falciparum Metabolomics. Cell Host & Microbe. 2009;5:191-9.

[11] Li JV, Holmes E, Saric J, Keiser J, Dirnhofer S, Utzinger J, et al. Metabolic profiling of a Schistosoma

mansoni infection in mouse tissues using magic angle spinning-nuclear magnetic resonance spectroscopy. International Journal for Parasitology. 2009;39:547-58.

[12] Foxman B. The epidemiology of urinary tract infection. Nature Reviews Urology. 2010;7:653-60.

[13] Dielubanza EJ, Schaeffer AJ. Urinary Tract Infections in Women. Medical Clinics of North America. 2011;95:27-+.

[14] Gupta A, Bansal N, Houston B. Metabolomics of urinary tract infection: a new uroscope in town. Expert Review of Molecular Diagnostics. 2012;12:361-9.

[15] Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PloS one. 2010;5:e13953-e.

[16] Holmes E, Loo RL, Stamler J, Bictash M, Yap IKS, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008;453:396-U50.

[17] Nicholson JK, Oflynn MP, Sadler PJ, Macleod AF, Juul SM, Sonksen PH. PROTON-NUCLEAR-MAGNETIC-RESONANCE STUDIES OF SERUM, PLASMA AND URINE FROM FASTING NORMAL AND DIABETIC SUBJECTS. Biochemical Journal. 1984;217:365-75.

[18] Bell JD, Brown JC, Sadler PJ. NMR studies of body fluids. NMR in biomedicine. 1989;2:246-56.

[19] Pauling L, Robinson AB, Teranish.R, Cary P. QUANTITATIVE ANALYSIS OF URINE VAPOR AND BREATH BY GAS-LIQUID PARTITION CHROMATOGRAPHY. Proceedings of the National Academy of Sciences of the United States of America. 1971;68:2374-&.