In this issue: Recent Developments in the Clinical Application of Mass Spectrometry

The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine

eJIFCC

Progress in metabolomics standardisation and its significance in future clinical laboratory medicine

Daniel A. Dias[1], Therese Koal[2]

[1] *School of Health and Biomedical Sciences, Discipline of Laboratory Medicine, RMIT University, Victoria, Australia*
[2] *Biocrates Life Sciences AG, Eduard-Bodem-Gasse 8, 6020 Innsbruck, Austria*

ARTICLE INFO

Corresponding author:
Dr. Daniel Anthony Dias
School of Health and Biomedical Sciences
Discipline of Laboratory Medicine
RMIT University
PO Box 71, Bundoora, 3083, Victoria
Australia
Phone: +61 3 9925 7071
Email: daniel.dias@rmit.edu.au

Key words:
metabolomics, standardization, laboratory medicine, clinical chemistry, biomarker discovery

ABSTRACT

Today, the technology of 'targeted' based metabolomics is pivotal in the clinical analysis workflow as it provides information of metabolic phenotyping (metabotypes) by enhancing our understanding of metabolism of complex diseases, biomarker discovery for disease development, progression, treatment, and drug function and assessment. This review is focused on surveying and providing a gap analysis on metabolic phenotyping with a focus on targeted based metabolomics from an instrumental, technical *point-of-view* discussing the *state-of-the-art* instrumentation, pre- to post- analytical aspects as well as an overall future necessity for biomarker discovery and future (pre-) clinical routine application.

1. INTRODUCTION

The well-established field of metabolomics aims to comprehensively identify qualitatively and/or quantitatively detectable, endogenous metabolites in biological systems. It is the study of the complete biochemical phenotype of a cell, tissue, or whole organism mainly using analytical platforms such as: nuclear magnetic resonance spectroscopy (NMR), liquid chromatography–mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS). Metabolomics interrogates biological systems since it is an unbiased, data-driven approach that may ultimately lead to hypotheses providing new biological knowledge. The term 'metabolomics' typically describing the state of an organisms' metabolism, was first coined by one its pioneers, Jeremy Nicholson, Chair in Biological Chemistry at Imperial College, London. United Kingdom [1]. The term 'metabolomics' was concerned with measuring the metabolites in one sample and may be derived from only one cell type. The term 'metabonomics' (common term used almost a decade ago) was defined as the global study of the systems that regulate metabolism, including variations over time, nevertheless, metabolomics is generally the accepted term used to date [1]. This can be confirmed in the field by new developments of metabolomics core units and research centers in academia with an increased interest in the pharmaceutical and biotechnology industries. These developments and the progressive increase in the number of publications in the field of metabolomics, justifies the growing interest of metabolomics in biomarker discovery and its applications in complex (long-term developing) diseases with open diagnostic, predictive, patient stratification, treatment response, relapse questions, in metabolic disorders (e.g. metabolic syndrome, hypertension, diabetes [2]), neurological disorders (Alzheimer´s disease [3-5]), cardiovascular diseases (heart failure [6, 7], inflammatory diseases (rheumatoid arthritis [8]), oncology [9-13] as well as in toxicology and drug assessment [14]. It is evident that metabolomics is a promising tool to aid in providing insight to answer a biological question, especially in biomarker discovery, however one must assume that it is unlikely there is a single *'golden'* endogenous biomarker that predicts or diagnoses disease. The consensus in the research community is that most studies in the scientific literature document that *'metabolic signatures'* will be the answer. By definition, a metabolic signature contains a panel and/or combination of affected endogenous metabolites and not just an individual metabolite, which is plausible due to the relevance of affected metabolic pathways. It also appears that metabolic signatures can provide an improvement in statistical outcomes and robustness of candidates in biomarker discovery. As a consequence, the instrumental technologies need to possess cross-platform capabilities established by MS and/or NMR and need to be reliable and robust for high-throughput routine analyses. This is the case for *state-of-the-art* LC-MS/MS instrumentation (as detailed below), which are routinely used in the clinical environment for toxicology screening, therapeutic drug monitoring, vitamin and hormone quantitative analysis [15, 16].

Metabolites may have a concentration range in the pico-millimolar range, a mass range of the order of ~1000 amu and polarity of molecules ranging from highly hydrophilic to hydrophobic. There is no single methodology able to separate, detect, and quantify the range of a chemically diverse range of metabolites [17], therefore multiple analytical techniques and sample preparation strategies are necessary to capture most of the metabolome [18]. A typical metabolomics workflow is comprised of sample harvesting and metabolic quenching; metabolite extraction, data acquisition, interrogation and bioinformatic analysis. For sample harvesting

and metabolic quenching, many methods have been published in the literature for various biological systems [19-22]. Individual differences of varied sample types (e.g. WT (Wild-type)/ Control vs KO (Knock-out)/salinity stress), will determine the techniques used at each of these steps. The aim of a typical metabolomics experiment is to analyze as many metabolites as possible. To date, there are several established analytical platforms which details the semi-quantitative detection (relative intensities) of metabolites, however the field of metabolomics is now spearheading towards the absolute quantitation of metabolites in biological systems. As a significant number of metabolites are present in an organism, the data acquired is substantially large requiring interrogation and needs to be processed and treated (eg. normalized and statistically transformed) to obtain a meaningful biological interpretation. Multivariate data analyses such as Principal Component Analysis (PCA) and Partial Least Squares-Discriminant Analysis (PLS-DA), Hierarchical Clustering Analysis (HCA), Heatmaps, Volcano plots and pairwise *t*-test, are routinely employed to extract information from large metabolomics data sets [23]. Once the statistically significant variables (metabolites) are identified, correlations between metabolites and responses, groupings and all the experimental data can be used, permitting to constructing a hypothesis or explain observations [23]. The differences between the samples and the identified metabolites connected with them will provide a holistic vision about the interrogated biological system.

Since its establishment in the late 90s [24, 25], metabolomics has proven to be a valuable tool in the analysis of biological systems where it has been used in an ever increasing number of diverse applications such as identifying key genes for important traits, to clarify events of physiological mechanisms and to reveal unknown metabolic pathways in crops [26]; response of salinity in chickpea cultivars [27] and salinity research [28]. It has been applied to assess meat quality traits in pig, cattle and chicken [29]; produce, preserve, and distribute high-quality foods for health promotion [30]; beer [31]; and natural products [32, 33]. By performing global metabolite profiling, also known as *"untargeted"* metabolomics, new discoveries linking cellular pathways to biological mechanisms are being revealed and are shaping our understanding of cell biology, physiology and medicine. These pathways can potentially be targeted to diagnose and treat patients with immune-mediated diseases [34]; understanding the physiological changes occurring in *"normal"* aging and the molecular multi-mechanistic processes involved during senescence [35]; human related studies, that is, genomics, epigenomics, microbiomics, transcriptomics, proteomics and metabolomics (systems biology) [36]; forensic science [37]; response to high intensity exercise [38] and to aid in precision medicine for patients with multiple sclerosis [39].

In this review, the choice and characteristics of all major metabolomics technologies will be addressed together with a discussion on current trends and requirements of biomarker discovery in a clinical environment as well as future (pre-) clinical routine capabilities. Finally, the current state of knowledge with respect to metabolomics standardization and a gap analysis, which needs to be addressed to bring metabolic signatures to clinical routine applications, will be presented.

2. GC-MS and LC-MS

2.1 Gas Chromatography –Mass Spectrometry (GC-MS)

It is generally assumed that GC-MS is only amenable for the analysis of volatile compounds or those classes of lipophilic compounds extracted from apolar, organic solvents. This is not

the case for the well-established 'polar GC-MS metabolomics', hydrophilic compounds can be made volatile due to chemical derivatisation which selectively alters known functional groups making them amenable for GC-MS analysis. Nevertheless, GC-MS, though limited to the analysis of compounds smaller < 1,000 Da, can unambiguously, comprehensively resolve >400 compounds including sugars, sugar alcohols, sugar phosphates, amino and organic acids, amines, sterols and fatty acids in one acquisition and is typically suited to the analysis of primary metabolites - those involved in fundamental biological processes (e.g. glycolysis, TCA cycle and amino acid synthesis) of the growth and development of a cell. To increase the number of identified metabolites, authentic standards are required to match mass spectra and retention time with metabolites in the sample [40]. In addition to the highly reproducible electron impact (electron impact ionization, EI) mass spectra, some commercially available libraries provide retention times or retention time indices under standardized conditions for each metabolite [41] increasing confidence in metabolite identification. One of the greatest advantages of GC-MS is that the ionization mode used in this technique is highly standardized and reproducible across GC-MS systems from different vendors worldwide (based on 70 eV ionization) which allows for the establishment of comprehensive GC-MS mass spectral libraries such as the NIST (http://www.nist.gov/srd/nist1a.htm), Agilent's FienLab (http://fiehnlab.ucdavis.edu/db) [42], or publically available [Golm Metabolome Database (GMD, http://gmd.mpimp-golm.mpg.de) [41], which contain TMS (tri-methylsilylated)-derivatized metabolites. Undoubtedly, for more than a decade and a half, GC-MS has been accepted as the *"work-horse"* platform due to its notable separation, reproducibility, robustness, ease of establishment and operation and its relatively low costs and standardization worldwide.

Recently, there has been a focus within the metabolomics community to obtain quantitative data for biological studies since they describe accurately the actual concentration of the metabolites of interest. In the current literature, >90% of published metabolomics studies are semi-quantitative with <10% of published metabolomics studies using absolute quantification. In one well-known example, a quantitative database was curated by Psychogios *et al.,* who systematically characterized the human metabolome through the Human Metabolome Project [43]. In another study, Schwarz and colleagues identified and quantified 476 metabolites in cerebrospinal fluid as part of an integrative metabolome-proteome CSF database towards biomedical research [44]. Boutara *et. al.,* utilized a number of analytical multi-platform (NMR, GC-MS, DFI/LC-MS/MS, ICP-MS and HPLC) analyses which led to the identification of 445 and quantification of 378 unique urine metabolites or metabolite species. An online database containing 2651 confirmed human urine metabolite species, 3079 in total, concentrations, related literature references and links to their known disease associations are freely available at http://www.urinemetabolome.ca [45].

2.2 Liquid Chromatography –Mass Spectrometry (LC-MS)

Liquid chromatography-mass spectrometry (LC-MS) is a complimentary analytical platform used to identify metabolites that generally do not require chemical derivatisation, typically suited to the analysis of higher molecular weights metabolites in their 'intact' form. There are also derivatization reagents available which provide greater selectivity and sensitivity for the analysis of 'targeted' metabolites classes [46-48]. Depending on the metabolite(s) or metabolite classes of interest requiring LC-MS analysis, the choice of extraction solvent is crucial in separating polar and apolar metabolites, achieved

through either a monophasic or bi-phasic solvent extraction (e.g. chloroform, methanol and water) covering the polarity range.

Once a sample has been extracted prior to carrying out a LC-MS analysis the chromatographic separation of eluting metabolites requires subsequent optimization. A number of stationary phases [(e.g. ion exchange, reversed phase (C18), hydrophilic interaction (HILIC) chromatography and aqueous normal phase separation)] with varying solvent systems in either isocratic or gradient elution are used to separate and attempt to capture the metabolome. The most commonly used ionization modes include, electrospray ionization (ESI) and atmospheric chemical pressure ionization (APCI). Depending on the metabolite classes to be targeted, a number of LC-MS based platforms and modes can be used. For example, the lipophilic extract can be profiled in an untargeted manner using Liquid Chromatography–Quadrupole Time of Flight–Mass Spectrometry (LC–*Q*TOF–MS) which allows for the identification of intact lipid species with its corresponding high resolution mass spectra. Quantification of each lipid species can be achieved using Multiple Reaction Monitoring (MRM) using authentic standards on a Liquid Chromatography–Triple Quadrupole–Mass Spectrometry (LC–*QQQ*–MS), a rapidly important field referred to as lipidomics. The LIPID MAPS Structure Database (LMSD, www.lipidmaps.org/data/structure) encompasses structures and annotations of biologically relevant lipids.

Subsequently, the polar extract which may contain secondary metabolites or higher molecular weight metabolites and those that are not GC-MS amenable can be separated by reversed phase (C18) chromatography. The lipophilic or non-polar extract can be separated using hydrophilic interaction (HILIC). Further to these separation methodologies, positive and negative, soft ionization modalities need to be applied to cover both positively $[M+H]^+$ and negatively $[M-H]^-$ charged metabolites present. Each mode can result in 500 - 3000 mass features detectable represented with their accurate mass, isotopic pattern, and retention time. Accurate mass (as obtained with a *Q*TOF or FT-type instrument, and isotopic pattern as well as additional secondary and tertiary MS^n can assist with the structural elucidation of the metabolite of interest. Nevertheless, the unambiguous identification is not always possible as a number of metabolites can have exactly the same molecular formula and mass ultimately requiring the isolation and complete elucidation by NMR by first principles.

For LC-MS based untargeted metabolomics a number of databases currently exist. As mentioned, the HMDB Version 3.6 (www.hmdb.ca/) is a comprehensive, web-accessible electronic database containing information on metabolites found in the human body. For food compounds, FoodDB (http://foodb.ca/), however its mass spectral information is also duplicated in the HMDB. METLIN is a metabolite repository is a web and freely accessible electronic database (http://metlin.scripps.edu/) to facilitate metabolite annotation through MS analysis. MassBank is an open-community mass spectra repository designed for public sharing of reference mass spectra from authentic chemical standards for metabolite annotation (www.massbank.jp). mzCloud is an open community of academic and industrial partners who provides MS/MS and MS^n spectral trees that can be freely searched (https://www.mzcloud.org/). For further details please *see* [49].

2.3 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY (NMR)

For decades NMR has proven to be the summit in the structure elucidation of organic compounds but the last decade and a half has seen its integration in metabolomics given that it is *non*-destructive, providing highly accurate quantification of

metabolites and unambiguous structure elucidation of compounds. Furthermore, NMR is regarded the *'gold standard method'* of choice for the structural elucidation of unknown compounds if accurate mass measurements and mass fragmentation pattern analysis has not resulted in sufficient information. Though, ^{1}H NMR is highly reproducible and signal intensity is directly related to the molar concentration, a drawback is that it has low sensitivity and resolution compared to MS-based methods [50]. For literature describing NMR metabolomics *see* [51-54].

3.0 NEED AND GAP ANALYSIS IN STANDARDIZATION/HARMONIZATION IN TARGETED METABOLOMICS

Today there is a motivation in the research community that the standardization and harmonization in metabolomics are mandatory to deliver comparable, reliable, high-quality and precise quantitative data [15, 55-59]. These requirements need to establish a set of standards to guarantee longitudinal, robust *lab-to-lab* and *inter-instrument* comparability and bring emerging applications of metabolomics into future routine analyses. Biomarker signatures identified using untargeted metabolomics profiling needs to be translated to targeted-based metabolomics later in the biomarker validation workflow to allow for the absolute quantitation and improved comparability of studies. The following gap analysis summarized in Table 1 below, presents the requirements for targeted metabolomics studies, related metabolic signature validations as well as a first outlook for the requirements in routine clinical applications.

3.1 Metabolomics study design

Published biomarker (signatures) discovery case/control studies today are often quite limited in sample cohort sizes and lacking in appropriate validation studies for confirmation. It is critical to invest in validation studies using a greater cohort of samples including the development of valuable biomarker/metabolite associations, which are powerful for future clinical applications. Research focused on population based cohorts is able to fulfil this requirement. Standardized targeted metabolomics analytics (e.g. kits) acquired in the same data formats is certainly advantageous, due to ease in comparing data and translating results, increasing statistical power. Secondly, studies are often not well designed to address biomarker specificity based on appropriate control cohort selection. Frequently, only healthy controls are included. Consideration of all relevant disease control cohorts as delimitation controls are mandatory to test and approve specificity of biomarker candidates resulting in increased value for future clinical applications (e.g. consideration of alternate cancers in women, benign breast diseases, endometriosis, polycystic ovary syndrome (PCOS) for breast cancer studies). In addition case/control studies needs to be translated to longitudinal studies, both retrospective as well as prospective, to investigate the biomarker performance for predictive value and to improve knowledge about the follow-up of individuals. Additional factors which can contribute to a cohesive study design are the inclusions of typically used clinical (less invasive) biofluids as potential target matrices to allow for parallel data translation/interpretation from bodily compartments to systemic biofluids (e.g. plasma, serum, dried blood spots), which are appealing in (pre)clinical applications.

3.2 Pre-analytics

The consideration of pre-analytical aspects in biomarker discovery and the validation is mandatory for the robust and reliable performance of detecting biomarkers. Avoiding "artificial" interferences and to investigate how biomarker candidates are affected and sensitive to sample generation (e.g. venepuncture), transport and storage (short-term

Table 1 Requirements for improved metabolomics biomarker studies and for future clinical applications

Study design	a) Number of samples in a cohort are often limited b) Validation studies are missing c) Gap of disease specificity in biomarker studies: case/control studies require inclusion of more disease and gender related delimitated controls not only healthy controls to prove and deliver specificity d) Translate case/control studies to longitudinal studies (population based, retrospective followed by prospective) e) Inclusion of typically used clinical (less invasive) biofluids as matrices into study protocols to ensure data translation/interpretation from body compartment to systemic biofluids
Pre-analytics	a) Pre-analytical quality markers based on endogenous metabolites (stability markers for sample generation and storage), appropriate database is required also to prove biomarker candidates robustness, standardized pre-analytics
Analytics	a) Quantitative metabolomic data b) Standardization (e.g. kits) (from sample to results including sample preparation, analysis, technical validated analytical results to deliver *lab-to-lab* comparability, *inter-instrument* comparability, long-term comparability c) Gap of reference materials and reference laboratories, round-robin/ring trial tests d) Gap of standard materials (external, internal standards) e) Established QMS system in the analytical laboratory (ISO 9001, ISO 17025, GLP etc.)
Post-analytics	a) Gap in standardized data pre-processing for statistical data analysis: • identification of pre-analytical affected samples in the study • normalization, batch correction, data cleaning (e.g. LOD imputation), confounder adjustment and multivariate outlier detection b) Standardized data formats
Needs for future clinical applications (bench to bed side)	a) High performing biomarker signatures in defined/standardized biological matrices for clinical question b) Translate disease/metabolite association to causality c) Reference methods/kits (medical device regulatory, FDA, CE/IVD), reference laboratories (e.g. CLIA) d) Traceability and commutability of standards and reagents (e.g. calibrators) e) Standardized sample/sampling device f) External quality assurance programs (proficiency tests, ring trials) g) Certified reference materials (for metabolic signatures /metabolite panels)

mainly at room temperature, mid/long-term (biobanking) in accordance to sample collection time and storage) are important. The value of targeted based metabolomics has the potential to monitor pre-analytical sample treatment based on the analysis of endogenous metabolites concentrations and signatures (e.g. sums, ratios) demonstrate the feasibility for targeted metabolomics [60-64]. The best approach is to test a panel of markers in combination and in parallel, to identify affected samples with high confidence when all quality markers (combinations) are identified. In our opinion, an appropriate database is required to collate (non)pre-analytical affected endogenous metabolites, allowing to evaluate the value and quality of biomarker candidates from a pre-analytical perspective.

3.3 Analytics

The automation, standardization and harmonization of providing quantitative metabolite data in metabolomics is required to bring together *lab-to-lab*, inter-instrument and long-term robust analytical analyses into biomarker discovery and development, critical in further developing biomarker signatures into the clinical environment. Commercially available kits are available for targeted metabolomics (e.g. Biocrates Life Sciences), which can *fill-the-gap*, by delivering standardized data formats. Biomarker discovery need to apply the rules which are standard practice in routine clinical analyses today to bring new biomarker signatures into the (pre-) clinical routine [15, 56, 59]. A second aspect to consider is the current gap of broad target analyte coverage of available standards (external and internal standards) and (certified) reference materials (covering metabolite classes) as well as appropriate reference laboratories. This gap is obligatory, not only for quantitative data but also to enable quantitation with approved accuracy and traceability. One of the initial activities of the National Institute of Standards and Technology (NIST) in collaboration with NIH (National Institute of Health) confirms the awareness for the need of standard reference materials (e.g. SRM 1950) [15, 65, 66]. External and internal standards are essential to define quantitative linear range, sensitivity, selectivity, and correction of matrix effects of the metabolites in the assay resulting in improved analytical accuracy and precision. Finally, an established quality system including consideration, for example, sample entrance and storage control, instrument and quality performance tests within the laboratory organization will and should assist in common laboratory quality rules.

As previously discussed, it is evident that GC-MS based metabolomics is very-well standardized with commercial and publically MS-databases and methods available. However, with new, sensitive instrumentation, enhancements in LODs, mass accuracy and resolution, dynamic ranges and the push for 'quantitative' data, existing libraries may need to be revisited because of the differences in mass accuracy and the relative intensities of fragment ions that can affect mass spectra similarity scores. As previously mentioned, there is numerous metabolomics applications and technologies and there has been discussions and consortiums formed to attempt to standardize aspects in metabolomics.

The bottleneck lies predominantly with LC-MS based profiling typically due to the complexity of variables including: extraction, chromatographic method or solvent system to comprehensively profile the whole metabolome in a single analytical run. Establishing unbiased analytical methods are not trivial due to a number of factors such as: issues in combining data from different MS analyses which hinder correlation of data obtained in different instruments/laboratories which include large scale studies acquired overt time, instrument drift and maintenance; analyte quantification is a problem compared to conventional targeted

methods; data acquired in untargeted LC-MS profiling from same/different laboratories acquired on different instruments cannot be easily compared. Standardized protocols from the phase of study design, sample collection and handling, up to the phase of chemical and statistical analysis remains an issue to be resolved [67]. The standardization of metabolomics data formats (more easier in targeted metabolomics compared to profiling) is absolutely necessary to simplify data comparability. A logical consequence is also the need and consideration of standardization in data pre-processing including data filtering (e.g. agreement of common rules for limit of detection (LOD) and lower limit of quantitation (LLOQ) imputation), sample batch correction, and apply pre-analytical quality marker signatures over the sample data set) before any downstream analysis can begin.

There have been a number of worldwide initiatives which have attempted to standardize LC-MS methodologies and protocols including column stationary phases, elution solvents and gradient, ionization modalities (positive/negative), MS scan setup including accuracy of mass detection with efforts focussing on creating LC-MS bases mass spectral information including accurate mass, MS^n fragmentation pattern and more importantly, retention time under specific LC conditions [68]. More recently, work carried out by Wolfer *et. al.,* described an approach that enabled the generation of reliable quantitative structure retention relationship models tailored to specific chromatographic protocols. The methodology, applied to 442 experimentally characterized standards, employed a combination of random forest and support vector regression models with molecular interaction descriptors [68]. This retention time prediction framework could be replicated by different laboratories to suit their profiling platforms and enhance the value of standard library by providing a new tool for compound identification [68].

3.4 Post-analytics

The *push* for metabolomics standardization began over 10 years ago with the standard metabolic reporting structure initiative (SMRS) and the Architecture for Metabolomics consortium (Armet) which focused particularly on NMR based metabolomics [69]. In 2005, the metabolomics standards initiative (MSI) focused on a community-agreed minimum reporting standards providing initial efforts on the descriptions of the experimental metadata describing a metabolomics study [70]. Founded by the community, standards and infrastructure for metabolomics still require storage, exchange, comparison and re-utilization of metabolomics data. From this, 5 working groups (WG) were created focussing on: metabolomic pipeline; biological context metadata WG, chemical analysis WG, data processing WG, ontology WG and exchange format WG, with the task of collecting relevant metabolomics standards and a forum for discussion [71, 72]. In order to implement agreed and acceptable guidelines on reporting identified metabolites, an application platform such as a metabolomics repository in addition to a journal publication was required. In 2012, MetaboLights (http://www.ebi.ac.uk/metabolights) was the first general purpose database in metabolomics, developed and maintained by the European Bioinformatics Institute (EMBL-EBI) which combines small molecule 'reference' layer with information about individual metabolites, chemistries, spectrometry and biological roles with a study archive, where primary data and metadata from metabolomics studies are ontologically tagged and stored [73]. Such depositions receive a stable identifier for each study, which can be quoted in related publications and can be used to access the data on a long term. Making metabolomics data publicly accessible allows it to justify researchers' findings in a peer-reviewed publication, increases the possibility of wider collaborations within the metabolomics community and ultimately gives a study higher visibility

and increased citation [73]. More recently, the Framework Programme 7 EU Initiative 'coordination of standards in metabolomics' (COSMOS) is developing a robust data infrastructure and exchange standards for metabolomics data and metadata [72]. The data deposition and exchange workflow in the COSMOS consortium will be formally defined, agreed, and documented in relation with MetaboLights and all partnering databases in Europe and worldwide that would like to participate. The COSMOS consortium ultimately develops the standards and infrastructure for and with the metabolomics and fluxomics community. These efforts will directly enable the implementation of COSMOS important deliverable—that of a robust data infrastructure and mechanisms for standards metabolomics data representation and data/meta-data exchange that will enrich metabolomics science [72].

The general community agreement is that the challenge of metabolomics is the accurate identification of large numbers of metabolites in various untargeted profiling techniques. The metabolomics community has been discussing the challenges of metabolite identification and minimum reporting criteria for some time, and the Chemical Analysis Working Group of the MSI proposed some basic guidelines in 2007 [74]. While spectral standardization within a particular database such as METLIN is helpful, the diversity in acquisition is also beneficial for metabolite annotation (isomers), as it can highlight similar/dissimilar fragmentation processes across analytical conditions. Public databases are more often enough curated *in-house* in different laboratories (academia and commercial), applying a multitude of different analytical methods, reflecting the analytically diverse nature of the metabolomics community. Standardization of spectra acquisition using one particular ionization source, mass analyzer, and/or fragmentation technique would only be essential for a small percentage of groups.

3.5. Needs for future clinical applications

Targeted based metabolomics with respect to the detection of biomarker signatures for newborn screening, a *proof-of-concept* adopted more than 10 years ago will continue to be important in future (pre)clinical applications. Current studies provides new insights into numerous chronic and long-term developing diseases such as (cardiovascular disease, hypertension, cancer, metabolic disorders e.g. diabetes, inflammatory bowel disease, autoimmune disease, neurological diseases). Targeted metabolomics will continue to provide new knowledge about the commonalities, for example: insulin resistance, mitochondrial function, and inflammation and differences of these diseases and demonstrate strong metabolic causes to individual disease pathophysiology. The importance of the microbiome and the acquired dysbiosis of the gut microbiota [75-81] and immune system/acquired immune competence [82] are important drivers for the paradigm shift to the understanding of disease, wherein targeted metabolomics can be considered as the *gold standard* tool to measure and quantify related alterations in the metabolic phenotype. However, several factors need be considered for the successful translation of new biomarker signatures in routine analyses. (Pre-)clinical applications require a defined and standardized biological matrix and sample introduction. Furthermore, disease/metabolite associations has to be translated to understand causality and affected biological pathways providing understanding, which might assist in future therapy approaches. Nevertheless, there needs to be a *push* towards enhancing regulatory requirements with traceable and commutable reference materials [16], standards, methods, kits (medical device regulatory, FDA, CE/IVD), and/or reference laboratories (CLIA), external quality assurance programs to improve the robustness and validity of clinical data in future.

REFERENCES

1. Hunter, P., *Reading the metabolic fine print. The application of metabolomics to diagnostics, drug research and nutrition might be integral to improved health and personalized medicine.* EMBO Rep, 2009. 10: p. 20-23.

2. Guasch-Ferré, M., et al., *Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis.* Diabetes Care, 2016. 39(5): p. 833-846.

3. Lia, D., et al., *Plasma phospholipids and prevalence of mild cognitive impairment and/or dementia in the ARIC Neurocognitive Study (ARIC-NCS).* Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2016. 3(0): p. 73-82.

4. González-Domínguez, R., et al., *Metabolomic-Driven Elucidation of Serum Disturbances Associated with Alzheimer's Disease and Mild Cognitive Impairment.* Curr Alzheimer Res., 2016. 13(6): p. 641-653.

5. Fiandaca, M.S., et al., *Plasma 24-metabolite Panel Predicts Preclinical Transition to Clinical Stages of Alzheimer's Disease.* Front Neurol., 2015. 12(6): p. 237.

6. Zordoky, B.N., et al., *Metabolomic Fingerprint of Heart Failure with Preserved Ejection Fraction.* PLoS One, 2015. 10(0): p. e0124844.

7. Cheng, M.L., et al., *Metabolic disturbances identified in plasma are associated with outcomes in patients with heart failure: diagnostic and prognostic value of metabolomics.* J Am Coll Cardiol., 2015. 65(15): p. 1509-1520.

8. Jutley, G.S. and S.P. Young, *Metabolomics to identify biomarkers and as a predictive tool in inflammatory diseases.* Best Pract Res Clin Rheumatol., 2015. 29(6): p. 770-782.

9. Kelly, R.S., et al., *Metabolomic biomarkers of prostate cancer: prediction, diagnosis, progression, prognosis and recurrence.* Cancer Epidemiolog Biom & Prevent 2016: p. doi: 10.1158/1055-9965.epi-15-1223.

10. Miolo, G., et al., *Phamacometabolomics study identifies circulating spermidine and tryptophan as potential biomarkers associated with the complete pathological response to trastuzumab-paclitaxel neoadjuvant therapy in HER-2 positive breast cancer.* Oncotarget., 2016. doi: 10.18632/oncotarget.9489.

11. Ang, J.E., et al., *Plasma Metabolomic Changes following PI3K Inhibition as Pharmacodynamic Biomarkers: Preclinical Discovery to Phase I Trial Evaluation.* Mol Cancer Ther., 2016. 15(6): p. 1412-1424.

12. Bathe, O.F. and F. Farshidfar, *From Genotype to Functional Phenotype: Unraveling the Metabolomic Features of Colorectal Cancer.* Genes, 2014. 5(3): p. 536-560.

13. Phan, L.M., S.C. Yeung, and M.H. Lee, *Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies.* Cancer Biol Med., 2014. 11(1): p. 1-19.

14. Ramirez, T., et al., *Metabolomics in toxicology and preclinical research.* ALTEX, 2013. 30(2): p. 209-225.

15. Koal, T. and H.P. Deigner, *Challenges in Mass Spectrometry Based Targeted Metabolomics.* Curr Mol Med., 2010. 10(2): p. 216-226.

16. Greaves, R.F., et al., *Achievements ad Future Directions of the APFCB Harmonisation Projects on Serum Testosterone.* Clin Biochem Rev, 2016. 37(0): p. 15-32.

17. Atherton, H., et al., *A combined 1H-NMR spectroscopy- and mass spectrometry-based metabolomic study of the PPAR-alpha null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome.* Physiological Genomics, 2006. 27: p. 178-186.

18. Aretz, I. and D. Meierhofer, *Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology.* Int. J. Mol. Sci., 2016. 17(5): p. 1-14.

19. Sheedy, J.R., et al., *A sample preparation protocol for 1H nuclear magnetic resonance studies of water-soluble metabolites in blood and urine.* Anal Biochem, 2010. 398(2): p. 263-5.

20. Meyer, H., M. Liebeke, and M. Lalk, *A protocol for the investigation of the intracellular Staphylococcus aureus metabolome.* Anal Biochem, 2010. 401(2): p. 250-9.

21. De Souza, D.P., et al., *Progressive peak clustering in GC-MS Metabolomic experiments applied to Leishmania parasites.* Bioinformatics, 2006. 22(11): p. 1391-6.

22. Villas-Boas, S.G., et al., *Global metabolite analysis of yeast: evaluation of sample preparation methods.* Yeast, 2005. 22(14): p. 1155-69.

23. Ebbels, T.M.D. and R. Cavill, *Bioinformatic methods in NMR-based metabolic profiling.* Prog Nucl Magn Reson Spectros 2009. 55(4): p. 361-374.

24. Nicholson, J.K., J.C. Lindon, and E. Holmes, *Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data.* Xenobiotica, 1999. 29(11): p. 1181-1189.

25. Oliver, S.G., et al., *Systematic functional analysis of the yeast genome.* Trends Biotechnol 1998. 16(9): p. 373-378.

26. Shiratake, K. and M. Suzuki, *Omics studies of citrus, grape and rosaceae fruit trees.* Breed Sci., 2016. 66(1): p. 122-138.

27. Dias, D.A., et al., *Quantitative profiling of polar primary metabolites of two chickpeacultivars with contrasting responses to salinity.* J. Chromatogr. B, 2015. 1000(1): p. 1-13.

28. Roessner, U. and D.M. Beckles, *Metabolomics for salinity research.*, in *Methods Mol Biol.* 2012, Humana Press. p. 203-215.

29. Picard, B., et al., *Recent advances in omic technologies for meat quality management.* Meat Sci., 2015. 109(0): p. 18-26.

30. Johanningsmeier, S.D., G.K. Harris, and C.M. Klevorn, *Metabolomic Technologies for Improving the Quality of Food: Practice and Promise.* Annu. Rev. Food Sci. Technol., 2016. 7(0): p. 413-438.

31. Hughey, C.A., C.M. McMinn, and J. Phung, *Beeromics: from quality control to identification of differentially expressed compounds in beer.* Metabolomics, 2016. 12(11): p. 1-13.

32. Katz, L. and R.H. Baltz, *Natural product discovery: past, present, and future.* J Ind Microbiol Biotechnol, 2016. 43(2-3): p. 155-176.

33. David, B., J.-L. Wolfender, and D.A. Dias, *The pharmaceutical industry and natural products: historical status and new trends.* Phytochem Rev, 2015. 14(2): p. 299-315.

34. Guma, M., S. Tiziani, and G.S. Firestein, *Metabolomics in rheumatic diseases: desperately seeking biomarkers.* Nat Rev Rheumatol., 2016. 12(5): p. 269-281.

35. da Costa, J.P., T. Rocha-Santos, and A.C. Duarte, *Analytical tools to assess aging in humans: The rise of geriomics.* Trends Anal. Chem., 2016. 80(C): p. 204-212.

36. Piersigilli, F. and V. Bhandari, *Biomarkers in neonatology: the new "omics" of bronchopulmonary dysplasia.* J Matern Fetal Neonatal Med, 2016. 19(11): p. 1758-1764.

37. Castillo-Peinado, L.S. and M.D. Luquee de Castro, *Present and foreseeable future of metabolomics in forensic analysis.* Anal Chim Acta., 2016. 925(0): p. 1-15.

38. Danaher, J., et al., *The use of metabolomics to monitor simultaneous changes in metabolic variables following supramaximal low volume high intensity exercise.* Metabolomics, 2016. 12(0): p. 1-13.

39. Bhargava, P. and P.A. Calabresi, *Metabolomics in multiple sclerosis.* Mult Scler., 2016. 22(4): p. 451-460.

40. Roessner, U., et al., *Metabolic Profiling Allows Comprehensive Phenotyping of Genetically or Environmentally Modified Plant Systems.* The Plant Cell, 2001. 13(1): p. 11-29.

41. Kopka, J., et al., *GMD@CSB.DB: The Golm metabolome database.* Bioinformatics, 2005. 21: p. 1635-1638.

42. Kind, T., et al., *FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry.* Anal. Chem., 2009. 15: p. 10038-10048.

43. Psychogios, N., et al., *The human serum metabolome.* PLoS One, 2011. 6(2): p. e16957.

44. Schwarz, E., et al., *Biomarker discovery in human cerebrospinal fluid: the need for integrative metabolome and proteome databases.* Genome Med., 2012. 4(4): p. 39.

45. Bouatra, S., et al., *The human urine metabolome.* PLoS One, 2013. 8(9): p. e73076.

46. Boughton, B.A., et al., *Comprehensive profiling and quantitation of amine group containing metabolites.* Anal. Chem., 2011. 83: p. 7523-7530.

47. Xu, F., et al., *Enhancement of the capabilities of liquid chromatography-mass spectrometry with derivatization: general principles and applications.* Mass Spectrom. Rev., 2011. 30(6): p. 1143-1172.

48. Tang, Z. and F.P. Guengerich, *Dansylation of unactivated alcohols for improved mass spectral sensitivity and application to analysis of cytochrome P450 oxidation products in tissue extracts.* Anal. Chem., 2010. 82(18): p. 7706-7712.

49. Vinaixa, M., et al., *Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects.* Trends Anal. Chem., 2016. 78(0): p. 23-35.

50. Verpoorte, R., et al., *Metabolomics: back to basics.* Phytochem Rev 2008. 7(3): p. 525-537.

51. Roessner, U. and D.M. Beckles, *Metabolite measurements. In Plant metabolic networks (eds J. Schwender), 1st edn, pp. 39-69, Springer Verlag, New York.* 2009.

52. Villas-Bôas, S.G., et al., *(eds) Metabolome Analysis: An Introduction. John Wiley & Sons, Inc., New Jersey, USA.* 2007.

53. Hill, C.B. and U. Roessner, *Advances in high-throughput untargeted LC-MS analysis for plant metabolomics. In. Future Science eBook Series "Advanced LC-MS applications for metabolomics. Future Science Group, UK.* 2014.

54. Roessner, U. and D.A. Dias, *Plant tissue extraction for metabolomics. In: Methods in Molecular Biology, Metabolomics Tools for Natural Product Discoveries, (eds U. Roessner, U. and D.A. Dias), 1st edn, pp 21-28, Humana Press, New York, USA..* 2013.

55. Everett, J.R., *Pharmacometabonomics in Humans: A new Tool for Personalised Medicine. Pharmacogenomics.* Pharmacogenomics, 2015. 16(7): p. 737-754.

56. Wishart, D.S., *Emerging applications of metabolomics in drug discovery and precision medicine.* Nat Rev Drug Discov., 2016. 15(0): p. 473-484.

57. Aretz, I. and D. Meierhofer, *Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in System Biology.* Int. J. Mol. Sci., 2016. 17(5): p. 632.

58. Klupczyńska, A., P. Dereziński, and Z.J. Kokot, *Metabolomics in Medical Sciences - Trends, Challenges and Perspectives.* Acta Pol Pharm., 2015. 72(4): p. 629-641.

59. Wood, P.L., *Mass Spectrometry Strategies for Clinical Metabolomics and Lipidomics in Psychiatry, Neurology, and Neuro-Oncology.* Neuropsychopharmacology, 2014. 39(1): p. 24-33.

60. Trezzi, J.P., et al., *LacaScore: a novel plasma sample quality control tool based on ascorbic acid and lactic acid levels.* Metabolomics, 2016. 12(96): p. doi: 10.1007/s11306-016-1038-1.

61. Kamlage, B., et al., *Quality Markers Addressing Preanalytical Variations of Blood and Plasma Processing Identified by Broad and Targeted Metabolite Profiling.* Clin Chem., 2014. 60(2): p. 399-412.

62. Anton, G., et al., *Pre-analytical sample quality: metabolite ratios as an intrinsic marker for prolonged room temperature exposure of serum samples.* PLoS One, 2015. 10(3): p. e0121495.

63. Breier, M., et al., *Targeted Metabolomics Identifies Reliable and Stable Metabolites in Human Serum and Plasma Samples.* PLoS One, 2014. 9(2): p. doi: 10.1371/journal.pone.0089728.g001.

64. Davis, J.S., et al., *Ex-vivo changes in amino acid concentrations from blood stored at room temperature or on ice: implications for arginine and taurine measurements.* BMC Clin Pathol., 2009. 9(10): p. doi: 10.1186/1472-6890-9-10.

65. Bouhifd, M., et al., *Quality Assurance of Metabolomics.* ALTEX, 2015. 32(4): p. 319-326.

66. Cheema, A.K., et al., *The ABRF metabolomics Research Group 2013 Study: Investigations of Spiked Compound Differences in a Human Plasma Matrix.* J Biomol Tech., 2015. 26(3): p. 83-89.

67. Gika, H.G., I.D. Wilson, and G.A. Theodoridis, *LC–MS-based holistic metabolic profiling. Problems, limitations, LC–MS-based holistic metabolic profiling. Problems, limitations.* J Chromatogr B Analyt Technol Biomed Life Sci., 2014. 966(1-6): p. 966.

68. Wolfer, A.M., et al., *UPLC–MS retention time prediction: a machine learning approach to metabolite identification in untargeted profiling.* Metabolomics, 2016. 12: p. 1-13.

69. Jenkins, H., et al., *Nat Biotechnol.* . A proposed framework for the description of plant metabolomics experiments and their results, 2014. 22(12): p. 1601-1606.

70. Castle, A.L., et al., *Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results.* Brief Bioinform., 2006. 7(2): p. 159-165.

71. Goodacre, R., et al., *Proposed minimum reporting standards for data analysis in metabolomics.* Metabolomics, 2007. 3(3): p. 231-241.

72. Salek, R.M., et al., *COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access.* Metabolomics, 2015. 11(6): p. 1587-1597.

73. (Editorial), *Data producers deserve citation credit.* Nature Genetics, 2009. 41(10): p. 1045.

74. Sumner, L.W., et al., *Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).* Metabolomics, 2007. 3(3): p. 211-221.

75. Potgieter, M., et al., *The dormant blood microbiome in chronic, inflammatory diseases.* FEMS Microbiol Rev., 2015. 39(4): p. 567-291.

76. Cho, I. and M.J. Blaser, *The Human Microbiome: at the interface of health and disease.* Nat Rev Genet., 2012. 13(4): p. 260-270.

77. Smirnov, K.S., et al., *Challenges of metabolomics in human gut microbiota research.* Int. J. Med. Microbiol., 2016. S1438-4221(16): p. 30021-30022.

78. Zitvogel, L., et al., *Microbiome and Anticancer Immunosurveillance.* Cell, 2016. 165(2): p. 276-287.

79. Nicholson, J.K., et al., *Host-gut microbiota metabolic interactions.* Science, 2012. 336(6086): p. 1262-1267.

80. X., H., et al., *Gut Microbiota and Nonalcoholic Fatty Liver Disease: Insights on Mechanism and Application of Metabolomics.* Int. J. Mol. Sci., 2016. 17(3): p. E300.

81. Bhattacharjee, S. and W.J. Lukiw, *Alzheimer's disease and the microbiome.* Front Cell Neurosci., 2013. 7(0): p. 153.

82. Plaza, J.J.G., et al., *Role of metabolism during viral infections, and crosstalk with the innate immune system.* Intractable Rare Dis Res., 2016. 5(2): p. 90-96.