Alessandra Sussulini   *Editor*

Metabolomics: From Fundamentals to Clinical Applications

Springer

Advances in Experimental Medicine and Biology

Proteomics, Metabolomics, Interactomics and Systems Biology

Series Editor
Daniel Martins-de-Souza
University of Campinas (UNICAMP)
Institute of Biology
Laboratory of Neuroproteomics
Campinas, Brazil

This series of volumes focuses on concepts, techniques and recent advances in the field of proteomics, interactomics, metabolomics and systems biology. Recent advances in various 'omics' technologies enable quantitative monitoring of myriad various biological molecules in a high-throughput manner, and allow determination of their variation between different biological states on a genomic scale. Now that the sequencing of various genomes, from prokaryotes to humans, has provided the list and linear sequence of proteins and RNA that build living organisms, defining the complete set of interactions that sustain life constitutes one of the key challenges of the postgenomic era. This series is intended to cover experimental approaches for defining protein-protein, protein-RNA, protein-DNA and protein-lipid interactions; as well as theoretical approaches dealing with data analysis, integration and modeling and ethical issues.

More information about this series at http://www.springer.com/series/15040

Alessandra Sussulini
Editor

Metabolomics: From Fundamentals to Clinical Applications

Springer

Editor
Alessandra Sussulini
University of Campinas (UNICAMP)
Institute of Chemistry
Laboratory of Bioanalytics and Integrated Omics (LaBIOmics)
Campinas, Brazil

Preface

Metabolomics is an emerging scientific area that is rapidly growing in the last few years. It is a key strategy in systems biology that generates (quantitative) information about the intermediates or final products of the metabolism, thus indicating the metabolic pathways that are affected during a biological process, such as a disease or a drug treatment. Complementary to other omics strategies (genomics, transcriptomics, and proteomics), where extensive information about the genotype is obtained, metabolomics has an important role in connecting genotype-phenotype information.

This volume of *Proteomics, Metabolomics, Interactomics and Systems Biology* series provides a comprehensive view of metabolomics, from the basic concepts, through sample preparation and analytical methodologies, to data interpretation and applications in Medicine. This edition, entitled *Metabolomics: From Fundamentals to Clinical Applications*, is the first book to cover metabolomics clinical applications also emphasizing analytical and statistical aspects. Moreover, future trends and perspectives in clinical metabolomics are also presented. The book is organized in 13 chapters, where expert researchers in the field of metabolomics contributed, discussing and reviewing the most important and recently developed features of clinical metabolomics.

For researches already experienced in metabolomics, this book will be useful as an updated definitive reference. For beginners in the field and graduate/undergraduate students, this edition will provide detailed information about concepts and experimental aspects in metabolomics, as well as examples and perspectives of applications of this strategy to clinical questions.

Alessandra Sussulini
Campinas, Brazil

Contents

Contributors

Ahmed Ali RIKEN Quantitative Biology Center, Suita, Japan

Alessandra Sussulini University of Campinas, Campinas, Brazil

Aline Klassen Federal University of São Paulo, Diadema, Brazil

Aline Soriano Lopes Federal University of São Paulo, Diadema, Brazil

Ana Valéria Colnaghi Simionato University of Campinas, Campinas, Brazil

Andréa Tedesco Faccio University of São Paulo, São Paulo, Brazil

Andrew J. Chetwynd University of Birmingham, Birmingham, UK

Annalaura Mastrangelo University of San Pablo, Madrid, Spain

April Oga RIKEN Quantitative Biology Center, Suita, Japan

Banny S. Barbosa University of Campinas, Campinas, Brazil

Berin A. Boughton University of Melbourne, Melbourne, VIC, Australia

Brett Hamilton University of Queensland, St Lucia, QLD, Australia

Christian Ludwig University of Birmingham, Birmingham, UK

Clement Heude University of Birmingham, Birmingham, UK

Coral Barbas University of San Pablo, Madrid, Spain

Elisa Brietzke Federal University of São Paulo, São Paulo, Brazil

Elisa Castañeda Santa Cruz University of Campinas, Campinas, Brazil

Emily Grace Armitage University of Glasgow, Glasgow, UK

Giovanny Rodriguez-Blanco University of Birmingham, Birmingham, UK

Gisele André Baptista Canuto University of São Paulo, São Paulo, Brazil

Henrique Caracho Ribeiro University of Campinas, Campinas, Brazil

Ibrahim Karaman Imperial College London, London, UK

Jay Nath University of Birmingham, Birmingham, UK

Jianguo Xia McGill University, Montreal, QC, Canada

João G.M. Pontes University of Campinas, Campinas, Brazil

John Bosco Carrigan University of Birmingham, Birmingham, UK

Karina Trevisan Rodrigues University of São Paulo, São Paulo, Brazil

Ljubica Tasic University of Campinas, Campinas, Brazil

Marina Franco Maggi Tavares University of São Paulo, São Paulo, Brazil

Michal Ciborowski Medical University of Bialystok, Bialystok, Poland

Mirian A.F. Hayashi Federal University of São Paulo, São Paulo, Brazil

Pedro Luis Rocha da Cruz University of São Paulo, São Paulo, Brazil

Rui Climaco Pinto Imperial College London, London, UK

Samy Emara Misr International University, Cairo, Egypt

Sara Amer RIKEN Quantitative Biology Center, Suita, Japan

Sumit Sethi Federal University of São Paulo, São Paulo, Brazil

Tsutomu Masujima RIKEN Quantitative Biology Center, Suita, Japan

Warwick B. Dunn University of Birmingham, Birmingham, UK

Yasmine Abouleila RIKEN Quantitative Biology Center, Suita, Japan

Zuzana Cieslarová University of Campinas, Campinas, Brazil

Part I: Fundamentals and Analytical Methodologies in Metabolomics

Chapter 1
Metabolomics: Definitions and Significance in Systems Biology

Aline Klassen, Andréa Tedesco Faccio, Gisele André Baptista Canuto, Pedro Luis Rocha da Cruz, Henrique Caracho Ribeiro, Marina Franco Maggi Tavares, and Alessandra Sussulini

Abstract Nowadays, there is a growing interest in deeply understanding biological mechanisms not only at the molecular level (biological components) but also the effects of an ongoing biological process in the organism as a whole (biological functionality), as established by the concept of systems biology. Within this context, metabolomics is one of the most powerful bioanalytical strategies that allow obtaining a picture of the metabolites of an organism in the course of a biological process, being considered as a phenotyping tool. Briefly, metabolomics approach consists in identifying and determining the set of metabolites (or specific metabolites) in biological samples (tissues, cells, fluids, or organisms) under normal conditions in comparison with altered states promoted by disease, drug treatment, dietary intervention, or environmental modulation. The aim of this chapter is to review the fundamentals and definitions used in the metabolomics field, as well as to emphasize its importance in systems biology and clinical studies.

Keywords Metabolomics • Systems Biology • Targeted Metabolomics • Untargeted Metabolomics • Lipidomics • Clinical Metabolomics

A. Klassen
Department of Exact and Earth Sciences, Federal University of São Paulo (UNIFESP), Diadema, SP, Brazil

A.T. Faccio • G.A.B. Canuto • P.L.R. da Cruz • M.F.M. Tavares
Institute of Chemistry, University of São Paulo (USP), São Paulo, SP, Brazil

H.C. Ribeiro • A. Sussulini (✉)
Laboratory of Bioanalytics and Integrated Omics (LaBIOmics), Department of Analytical Chemistry, Institute of Chemistry, University of Campinas (UNICAMP), P.O. Box 6154, Campinas, SP, 13083-970, Brazil
e-mail: sussulini@iqm.unicamp.br

3

Abbreviations

CE   Capillary electrophoresis
DA   Discriminant analysis
DI   Direct infusion
GC   Gas chromatography
HPLC   High-performance liquid chromatography
IPLC   Ion-pairing liquid chromatography
LC   Liquid chromatography
MALDI   Matrix-assisted laser desorption ionization
MS   Mass spectrometry
MSI   Mass spectrometry imaging
NMR   Nuclear magnetic resonance
PCA   Principal component analysis
PLS   Partial least squares
OPLS   Orthogonal projections to latent structures
QC   Quality control
SRM   Selected reaction monitoring
UPLC   Ultra-performance liquid chromatography

1.1 A Brief History of Metabolomics and Its Relevance in Systems Biology

With the advent of the *systems biology* paradigm, which proposes to explore how interactions between biological components (biomolecules) affect the functionality (biological processes) of an organism as a whole [1], several bioanalytical methods have been proposed and/or improved. Formerly, molecular biology and physiology approaches were employed to acquire biomolecular and functional information, respectively. However, both strategies only provided limited data considering a target biomolecule and the directly related pathways, being incapable of characterizing a biological system in a complete and integrated way. For that reason, the development of the *omics* strategies caused a real revolution in this scientific area, and nowadays they are widely used in systems biology. *Omics* strategies aim at identifying the entire set of biomolecules (genes, proteins, metabolites, etc.) contained in a biological tissue, cell, fluid, or organism, thus generating a huge amount of data that are evaluated by biostatistics and bioinformatics tools. Figure 1.1 shows how the *omics* approaches are correlated and their respective objects (biomolecules) of study. Genomics, transcriptomics, and proteomics are beyond the scope of this book; therefore, we will herein focus on metabolomics as a key systems biology strategy.
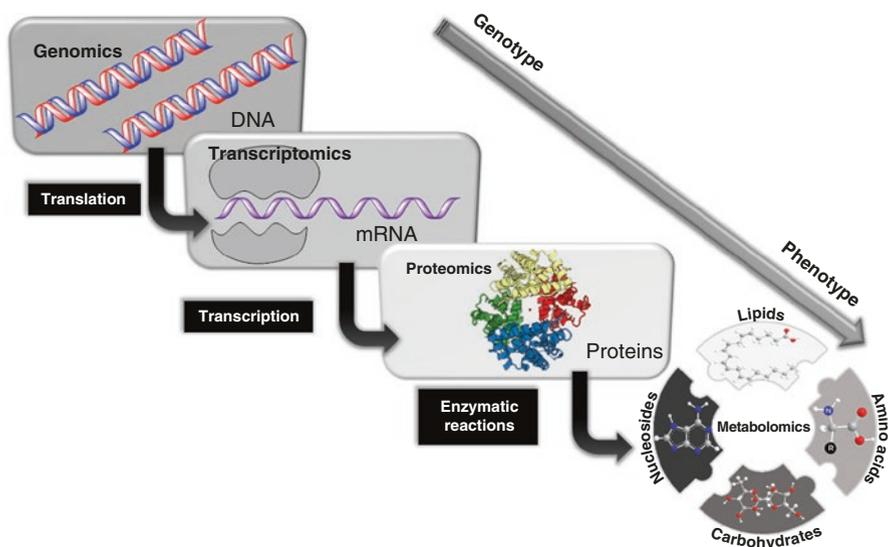
Fig. 1.1 A correlation between the main *omics* strategies used in systems biology studies

The term *metabolome* first appeared in the literature in 1998, when Oliver et al. [2] measured the change in the relative concentrations of metabolites as the result of deletion or overexpression of a gene. *Metabolome* is therefore used to address the entire set of metabolites an organism expresses. In 2001, *metabolomics* was defined by Fiehn as the comprehensive and quantitative analysis of all metabolites of the biological system under study [3]. Previously, in 1999, Nicholson et al. [4] used the term *metabonomics* to refer to the "quantitative measurement of the dynamic multi-parametric metabolic response of living systems to physiopathological stimuli or genetic modifications." Alterations in endogenous metabolite levels that may result from disease processes, drug toxicity, or gene function have been evaluated in cells, tissues, or biological fluids by metabonomics [5–8]. Latent biochemical information obtained from metabonomics may be used for diagnostic or prognostic purposes. Such information reflects actual biological events rather than the potential for disease, which gene expression data provide [9].

In the last decade, other terminologies have appeared in the literature to define and to classify metabolism studies. *Metabolite* (or *metabolic*) *profiling* was firstly described as the analysis of a small number of predefined metabolites for investigation of selected biochemical pathways and has its origin in early metabolism studies of Horning and Horning in 1971 [10]. *Metabolic fingerprinting* was defined by Fiehn [3] as "a rapid classification of samples according to their origin or their biological relevance." Finally, in 2005, Kell et al. [11] proposed the term *metabolic footprinting* to refer to the exometabolome, i.e., what a cell or system excrete under controlled conditions. Most recently, in 2015, the term *real-time metabolome profil-*

ing was proposed by Link et al. [12] referring to the direct injection of bacteria and cells in a high-resolution mass spectrometer and the monitoring of hundreds of metabolites in cycles of a few seconds over several hours.

However, before those terms were coined, studies involving metabolomics notions were firstly reported in the literature in the late 1940s by Williams and coworkers [13]. These studies were based in the data from over 200,000 paper chromatograms obtained from body fluid samples from different subjects, including alcoholics and schizophrenics, which produced evidence that there were characteristic metabolic patterns associated with each one of these groups, considering a hypothesis of "biochemical individuality." Gates et al. published, in 1978, a review compiling these historical events [14]. The development of novel analytical techniques and biostatistics improvements in the 1980's allowed an enormous progress of metabolic profiling studies. Then, at the end of the 1990s, many acronyms related to *omics* strategies appeared, and at that point, the terms *metabolome*, *metabolomics*, and *metabonomics* were proposed. A review from van der Greef and Smilde [15] discusses the symbiosis of metabolomics and chemometrics and presents an interesting timeline of the evolution of metabolomics.

Lipidomics is a subdivision of metabolomics defined as "the full characterization of lipid molecular species and their biological roles with respect to expression of proteins involved in lipid metabolism and function, including gene regulation" [16]. This term was proposed in 2003 by Han and Gross [17] to define the research area that focuses on identifying alterations in lipid metabolism and lipid-mediated signaling processes that regulate cellular homeostasis during health and disease. Currently, lipidomics research emphasis consists on identifying alterations in cells and body fluid lipid levels, revealing environmental disturbances, pathological processes, or response to drug treatments [18].

1.2   Definitions and the Metabolomics Workflow

Although several terms have been devised in the literature to classify metabolomics/metabonomics studies [2, 4, 19–21], there is still not an actual consensus regarding terminology. A much simpler general definition that relates to the fact whether the researcher knows a priori what kind of metabolites to search has been proposed here, and it will guide the decisions on the metabolomics workflow presented in Fig. 1.2. In this context, a *targeted metabolomics* approach is defined as a *quantitative analysis* (concentrations are determined) *or semiquantitative analysis* (relative intensities are registered) of a few metabolites and/or substrates of metabolic reactions that might be associated to common chemical classes or linked to selected metabolic pathways. *Metabolic profiling* as mentioned earlier thus belongs to this definition. An *untargeted metabolomics* approach is based primarily on the *qualitative or semiquantitative analysis* of the largest possible number of metabolites from a diversity of chemical and biological classes contained in a

biological specimen. Both *fingerprinting* and *footprinting metabolomics* belong to this definition.

Lipidomics could be considered as a targeted metabolomics strategy, since it involves the study of a subset of specific metabolites (lipids). However, due to the complexity of the lipids, lipidomics itself is categorized as targeted or untargeted lipidomics, when the objects of study are specific lipids or global exploratory analyses are performed, respectively. The term *focused lipidomics* was proposed in 2009 as a strategy "for detecting molecules in some categories while comprehensively utilizing specific fragments (product and precursor ion scanning) or neutral loss caused by a specific feature of the partial structures of the molecules (neutral loss scanning)" [22]. However, the execution of *focused lipidomics* is only possible when working with mass spectrometry techniques, and actually this is not so distant from the targeted lipidomics concept. More details about lipidomics can be found in Chap. 11.

The metabolomics workflow, shown in Fig. 1.2, comprises the sequential steps that underline both targeted and untargeted metabolomics analyses, which will be further described.

Biological problem and experimental design. The initial step of the metabolomics workflow relies on a clear and straightforward formulation of the biological problem to be addressed. This step is of crucial importance because it will govern the experimental design that follows. According to the biological problem, the type of metabolomics approach (targeted vs. untargeted metabolomics), sample type (biological fluids, tissues, cells, and/or intact organisms), sample size (number of specimens to be assessed), experimental conditions to which samples will be submitted, frequency of sample collection, metabolic quenching to interrupt enzymatic activity (addition of organic solvents and/or immediate freezing of samples by the use of dry ice or liquid nitrogen), storage conditions ($-80\ °C$ is usually preferred for long-term storage of biological fluids) [23], analytical platforms to be employed, and also sample preparation strategies must all be defined at this point, since they are somehow interrelated [24]. It is important to emphasize that metabolomics studies are always comparative in character; therefore, a group of control samples (samples that did not undertake the investigated condition) and test samples (carrying information on the investigated condition) are usually defined in the experimental design.

Sample preparation. Once the biological problem is defined and experimental conditions for sample collection and storage are established, a further step on sample preparation prior to analysis might be considered. Sample preparation is intimately related to the sample type (whether it is a cell, a tissue, or a biological fluid), the selected metabolomics approach (targeted vs. untargeted analysis), and the elected analytical platform.

For targeted metabolomics, the extraction procedure is usually optimized for the specific metabolites or metabolite chemical classes under consideration and may involve steps such as cleanup for removal of sample matrix interferents and/or pre-

Biological problem

Problem formulation

Metabolomics approach definition:

Targeted

Untargeted

Analysis of known compounds from specific chemical classes or specific pathways (*Profiling*)

Fingerprinting: intracellular metabolites
Footprinting: extracellular metabolites

Experimental design

Definition of groups: **TEST** samples versus **CONTROL** samples
Sample type & size and collection & storage conditions

Metabolism quenching / protein precipitation / centrifugation / dilution / use of internal standards / derivatization

Sample preparation

Clean-up & pre-concentration:
• Solid-phase extraction
• Liquid-liquid extraction

• Minimum sample preparation for biofluids
• Extraction procedures for tissue/cells

Data acquisition

Analytical platforms: LC-MS, GC-MS, CE-MS and/or NMR

MS: EIC / SRM / method validation / quantitation

GC-MS, LC-MS and *CE-MS:*
Decovolution / dataset creation / grouping / alignment / normalization / transformation
NMR:
Phasing / baseline correction / alignment / normalization

Data processing

Univariate statistics: t-test, Kruskal Wallis, Mann Whitney, ANOVA, etc.
Multivariate statistics: Data overview (PCA) / model building (PLS-DA, OPLS-DA, etc.) / model validation

Statistical analysis

Discriminant metabolites:
VIP, S-Plot, Jack-knife

MS / NMR: Putative metabolite search - (libraries)

MS: Spiking / fragmentation
NMR: 2D, Stocsy, ROC, etc.

Metabolite identification

Metabolic pathway relationships (libraries)

Metabolic pathways interpretation

Internal / external validation

Biological validation

Fig. 1.2 Analytical workflow for studies in metabolomics

concentration strategies, such as liquid-liquid and solid-phase extraction, to enhance the compound detectability [25].

For untargeted metabolomics of biofluids, sample preparation is usually minimal. Protein precipitation is sometimes considered as a mean to preserve column integrity in liquid chromatographic experiments or to prevent capillary clogging in capillary electrophoretic experiments. In general, a simple filtration and a few-fold dilution are often performed. Tissue and/or cell preparations require more elaborated extraction procedures, usually carried out by solid-phase extraction with pure solvents or mixtures, followed by centrifugation and dilution. Gas chromatographic analyses of biofluids and cell/tissue extracts demand further derivatization steps to convert polar metabolites into volatile adducts [26]. These steps are time-consuming and prone to errors, limiting the number of total samples to be processed in a single metabolomics experiment. Nuclear magnetic resonance experiments usually require sample dilution in proper deuterated solvents. More details about sample preparation in clinical metabolomics can be found in Chap. 2.

Data acquisition.   Differently from other *omics* sciences, metabolomics imposes a great analytical challenge due to the immense variety of chemical composition that biological samples exhibits, spanning from compounds with distinct chemical properties, structural features and functionality, as well as discrepant concentration levels. It is important to emphasize that currently no single analytical platform leads to a comprehensive identification and quantification of the entire metabolite set of a biological system [27, 28]. The chemical diversity of the metabolome, as well as its wide dynamic range [29], demands that different analytical techniques be combined to generate complementary results that will ultimately enhance metabolic coverage [5, 28].

The analytical techniques commonly employed for data acquisition in metabolomics studies are nuclear magnetic resonance (NMR) [23, 30] spectroscopy and mass spectrometry (MS) [31]. NMR spectroscopy can be considered as a universal metabolite detection technique, where samples can be analyzed directly with minimal manipulation and many classes of small metabolites can be measured simultaneously [5, 7, 32]. Major drawbacks in NMR for metabolomics are poor sensitivity and spectral complexity with superimposition of signals at certain spectral regions compromising clear identification. In a recent metabolomics study, NMR allowed characterization of 49 metabolites in human serum, with concentrations above $10\ \mu mol\ L^{-1}$ ("normal NMR-detected serum metabolome"), whereas MS techniques were able to characterize more than 90 metabolites with concentrations lower than $10\ \mu mol\ L^{-1}$ [33].

While MS is more sensitive and specific in comparison to NMR spectroscopy, it usually requires a previous separation step, using a hyphenated separation technique [27], such as gas chromatography (GC), high-performance liquid chromatography (HPLC) or ultra-performance liquid chromatography (UPLC), and capillary electrophoresis (CE). Separation techniques coupled to MS are important to reduce sample complexity and to minimize ionization suppression effects, thus enhancing the detection sensitivity and increasing the metabolome coverage [27].

GC-MS is a sound technique in the metabolomic arena [34–37]. However, the need for time-consuming sample derivatization schemes limits its applicability to small sets of samples. Nevertheless, structural specificity of the generated adducts makes it easy to build dedicated spectra libraries that aid metabolite identification. Furthermore, due to derivatization, quite distinct classes of important rather polar metabolites, such as amino acids, biogenic amines, and carboxylic acids, can be assessed in a single chromatographic run and ionization mode. Furthermore, GC-MS via headspace techniques cover the volatile portion of the metabolome.

Perhaps due to extensive sample manipulation GC-MS demands, LC-MS has been the technique of choice in many metabolomics studies [38–42], covering the moderately polar fraction of the metabolome. Several modes and different columns chemistries, including reversed-phase liquid chromatography (RPLC), hydrophilic interaction liquid chromatography (HILIC), and more rarely ion-pairing liquid chromatography (IPLC), allow LC-MS to cover a wide range of metabolite categories and polarities.

Although not as prominent as the chromatographic techniques, capillary electrophoresis coupled to mass spectrometry (CE-MS) has joined the metabolomic analytical arsenal due to its unique characteristics, particularly the ability to assess directly the most polar and/or ionic fraction of the metabolome [43–45].

It is also possible to perform a metabolomics analysis by direct infusion (DI) mass spectrometry, but a lot of information is lost, due to ionization suppression of many metabolites present at very low concentrations in complex biological matrices [46]. Matrix-assisted laser desorption ionization coupled to MS (MALDI-MS) and MALDI mass spectrometry imaging (MALDI-MSI) are increasingly being invoked for metabolomics studies specially those assessing tissues, cells, and their compartments [47]. The incapability of differentiating metabolite isomers is a shortcoming of the MS technology. More details about NMR spectroscopy in metabolomics can be found in Chap. 3. Chapters 4 and 5 describe MS coupled to chromatographic and electrophoretic techniques, respectively.

Analytical methodologies used for lipidomics were recently reviewed [18, 48]. Basically, the only difference from a metabolomic and a lipidomic experiment is sample preparation: for lipidomics is necessary to include a lipid extraction step, usually by liquid-liquid extraction or solid-phase extraction, prior to NMR or MS analysis [18].

Analytical platform stability issues often arise when untargeted metabolomic studies are performed, since all samples are analyzed just once in a series of randomized sequential runs. To circumvent these issues and to ensure data reliability for further processing, the use of a quality control (QC) pool sample, prepared by mixing small volumes of all control and test samples, is often employed. Instrumental stability is checked by running several times the QC sample upfront and during sample analyses by intercalating QC samples every four or five samples, depending on run time. Repeatability of QC spectra and/or mass chromatograms is inspected visually and statistically. The importance of QC samples in

metabolomics studies is thoroughly discussed in the review articles of Dunn et al. [49] and Theodoridis et al. [38].

Data processing. For untargeted metabolomics, the acquired raw data are submitted to a preprocessing step according to the type of analytical platform employed. For NMR, data treatment includes phasing, baseline correction, alignment, and normalization. Softwares and algorithms, such as PERCH (PERCH Solution Ltd.), Chenomx NMR Suite (Chenomx Inc.), MestReNova (MestreLab Research), MetaboLab [50], AutoFit [51], TopSpin (Bruker Corp.), and MATLAB (The MathWorks Inc.), are routinely used. For hyphenated MS techniques, data treatment includes spectral deconvolution, dataset creation, grouping, alignment, filling data gaps, normalization, and data transformation. Several free access and proprietary softwares are available to process MS data as discussed comparatively by Sugimoto et al. [52]: XCMS [53], Mass Profiler Professional (MPP, Agilent Technologies), MZmine [54], MetAlign [55], MassLynx (Waters Corp.), and AMDIS [56], among others [57].

For targeted metabolomics, analyte quantitation or semiquantitation is a relevant part of data processing and is more commonly carried out using MS spectroscopy [58, 59] rather than NMR, although recommendations are available [60]. Methods are usually developed and conditions optimized for the selected metabolite (s), and the proposed method undergoes extensive validation following regulated protocols for the parameters specificity/selectivity, precision, accuracy, linearity, limits of detection and quantification, and robustness before application is set out. In targeted analyses, the use of internal standards is recommended to improve precision and to handle matrix effects [61], specially isotope-labeled internal standards. Selected reaction monitoring (SRM) is a versatile tool for targeted metabolomics when triple-quadrupole mass spectrometers are used [61].

An interesting alternative approach for comprehensive targeted metabolomics has been put in practice lately using commercially available kits [62]. High-throughput quantitative MS analyses of hundreds of endogenous metabolites of a few chemical classes (acylcarnitines, amino acids, hexoses, phospho- and sphingo-lipids, biogenic amines, etc.) are performed upfront, and statistical evaluation of results selects the discriminant metabolites. Although a great deal of compounds is usually disregarded after statistical analyses, at the end the discriminant metabolites have already been quantified. This approach opposes the more traditional one where untargeted metabolomics indicates qualitatively the potentially discriminant metabolites, usually a small number of compounds, that are next quantified by targeted metabolomics, followed by statistical evaluation of importance. Drawbacks of the comprehensive targeted metabolomics relate to the fact that a lot of analytical effort is placed on the quantitation of hundreds of metabolites that might render no significant information, and the search of discriminant metabolites is carried out on a limited number of chemical classes imposed by the commercial kit composition, with no room for discovery of novel metabolites. Chapter 6 gives more details about data processing in metabolomics.

Statistical analysis. Metabolomic data are quite complex and require chemometric tools to reveal discriminant metabolites between control and test samples [63]. Multivariate analyses, comprising unsupervised methods, such as principal component analysis (PCA), and supervised methods, such as partial least square discriminant analysis (PLS-DA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA), are often employed for sample overview and classification. Univariate analysis based on Student's t-test, Mann-Whitney U test, etc. is also used to corroborate multivariate results. The mathematical models must be validated, which is carried out by cross validation procedures and permutation tests [63–65]. Chapter 7 brings more details about chemometrics in metabolomics.

Metabolite identification. Metabolite identification is required only for untargeted metabolomics studies, since for targeted metabolomics, the metabolite or metabolite class of interest is already defined. For such purpose, free databases and libraries, such as HMDB [66], KEGG [67], PubChem [68], Metlin [69], MassBank [70], LIPID MAPS [71], ChEBI [72], MMD [73], BioMagResBank [74], MetaboID [75], and Chenomx NMR Suite (Chenomx Inc.), are among the most commonly accessed. MassTRIX [76] is also a searching tool that uses some of the databases listed above. Once a putative metabolite has revealed identity, confirmation must be pursued. This can be accomplished by spiking techniques with authentic standards followed by comparison of fragmentation patterns between sample and standard (MS spectra) or 2-D NMR.

Metabolic pathways association. Biological interpretation is an important step of any metabolomics study, targeted or untargeted. Once putative metabolites are listed and their identification confirmed, corresponding metabolic pathways are next searched. Several databases are available for this purpose: KEGG [67], MetaCyc [77], SMPDB [78], MetaboLights [79], and Reactome [80], among others. For details on the information compiled in databases, the review of Karp and Caspi can be consulted [81]. When altered metabolites are associated to respective metabolic pathways, a rationale can be elaborated in attempt to answering the original biological question that guided the metabolomics study. In Chapter 8, more details about metabolite identification and pathways analysis can be found.

Biological validation. Although biological validation is not commonly pursued after a metabolomics study is completed, many authors consider that the results will only make a broader sense if proven by validation. Usually an external validation is recommended [63, 82], in which an entire new set of samples are collected and processed, as the work of Barbas et al. [83] exemplifies. Alternatively, the discriminant metabolites found preliminarily in the untargeted metabolomics study can be quantitatively analyzed in the same sample set (internal validation). Biological validation can also be reached by independent specific studies conducted with the discriminant metabolites found in the original metabolomics study. Ganti et al. [84] performed an untargeted metabolomics study using urine samples of kidney cancer patients and control subjects that revealed high levels of acylcarnitines associated with cancer status and kidney cancer grade. The study was then validated by in vitro experiments

establishing that acylcarnitines affect cell survival and are indicative of inflammation. Biological validation serves therefore to corroborate the results obtained preliminarily in the original metabolomics study and to consolidate the biological interpretation of results.

1.3   The Importance of Metabolomics in Clinical Studies

From the beginnings of metabolomics until nowadays, most of the applications are focused on plant metabolomics. Nevertheless, with the recent advent of precision medicine, clinical metabolomics is on the spotlight for being able to provide molecular phenotyping of biofluids, cells, or tissues. In this context, clinical metabolomics is increasingly being applied to diagnose diseases, understand disease mechanisms, identify novel drug targets, customize drug treatments, and monitor therapeutic outcomes [85, 86].

As metabolites indicate end points of the gene expression and cell activity, metabolomics can provide a holistic approach for understanding the phenotype of an organism, playing a fundamental role in systems biology [27]. The characterization of metabolic phenotypes supports precision medicine by pointing out the metabolic imbalances that underlie diseases, discovering new therapeutic targets, and indicating potential biomarkers that may be used to either diagnose disease or monitor action of therapeutics [87].

Clinical metabolomics is thus an area of intense investigation and has been revised periodically for different conditions and diseases [63, 88–97]. Chapters 9, 10, 11, 12, and 13 organized in this book compile many applications of clinical importance following the metabolomics framework.

1.4   References

 1. Westerhoff HV, Alberghina L. Systems biology: did we know it all along? In: Alberghina L, Westerhoff HV, editors. Systems biology – definitions and perspectives. Berlin: Springer; 2005. p. 3–9.
 2. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends Biotechnol. 1998;16:373–8.
 3. Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genomics. 2001;2:155–68.
 4. Nicholson J, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;11:1181–9.
 5. Dunn WB, Ellis DI. Metabolomics: current analytical platforms and methodologies. Trends Anal Chem. 2005;24:285–94.
 6. Lenz EM, Wilson ID. Analytical strategies in metabonomics. J Proteome Res. 2007;6:443–58.
 7. Nicholson JK, Lindon JC. Systems biology: metabonomics. Nature. 2008;455:1054–6.
 8. Kaddurah-Daouk R, Krishnan KRR. Metabolomics: a global biochemical approach to the study of central nervous system diseases. Neuropsychopharmacol. 2009;34:173–86.

9. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov. 2002;1:153–61.

10. Horning EC, Horning MG. Human metabolic profiles obtained by GC and GC/MS. J Chromatogr Sci. 1971;9:129–40.

11. Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG. Metabolic footprinting and systems biology: the medium is the message. Nat Rev Microbiol. 2005;3:557–65.

12. Link H, Fuhrer T, Gerosa L, Zamboni N, Sauer U. Real-time metabolome profiling of the metabolic switch between starvation and growth. Nat Methods. 2015;12:1091–7.

13. Williams RJ, Kirby H. Paper chromatography using capillary ascent. Science. 1948;107:481–3.

14. Gates SC, Sweeley CC. Quantitative metabolic profiling based on gas chromatography. Clin Chem. 1978;24:1663–73.

15. van der Greef J, Smilde AK. Symbiosis of chemometrics and metabolomics: past, present, and future. J Chemometrics. 2005;19:376–86.

16. Spener F, Lagarde M, Géloën A, Record M. What is lipidomics? Eur J Lipid Sci Technol. 2003;105:481–2.

17. Han X, Gross RW. Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. J Lipid Res. 2003;44:1071–9.

18. Sethi S, Hayashi MA, Sussulini A, Tasic L, Brietzke E. Analytical approaches for lipidomics and its potential applications in neuropsychiatric disorders. World J Biol Psychiatry. 2016;26:1–15.

19. Villas-Bôas S, Mas S, Akesson M, Smedsgaard J, Nielsen J. Mass spectrometry in metabolome analysis. Mass Spectrom Rev. 2005;24:613–46.

20. Lindon JC, Nicholson JK, Holmes E. The handbook of metabonomics and metabolomics. Amsterdam: Elsevier; 2007. ISBN 978-0-444-52841-4.

21. Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol. 2002;48:155–71.

22. Navas-Iglesias N, Carrasco-Pancorbo A, Cuadros-Rodríguez L. From lipid analysis towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part I: modern lipid analysis. Trends Anal Chem. 2009;28:393–403.

23. Beckonert O, Keun HC, Ebbels TMD, Bundy J, Holmes E, Lindon JC, Nicholson JK. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc. 2007;2:2692–703.

24. Yin P, Lehmann R, Xu G. Effects of pre-analytical processes on blood samples used in metabolomics studies. Anal Bioanal Chem. 2015;407:4879–92.

25. Álvarez-Sánchez B, Priego-Capote F, Luque de Castro MD. Metabolomics analysis II. Preparation of biological samples prior to detection. Trends Anal Chem. 2010;29:120–7.

26. Garcia A, Barbas C. Gas chromatography–mass spectrometry (GC-MS)-based metabolomics. In: Metz TO, editor. Metabolic profiling, methods in molecular biology 708. New York: Springer; 2011. p. 191–204.

27. Kuehnbaum NL, Britz-McKibbin P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. Chem Rev. 2013;113:2437–68.

28. Villas-Bôas SG, Bruheim P. The potential of metabolomics tools bioremediation studies. OMICS. 2007;11:305–13.

29. Bowen BP, Northen TR. Dealing with the unknown: metabolomics and metabolite atlases. J Am Soc Mass Spectrom. 2010;21:1471–6.

30. Ludwig C, Viant MR. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. Phytochem Anal. 2010;21:22–32.

31. Viant MR, Sommer U. Mass spectrometry based environmental metabolomics: a primer and review. Metabolomics. 2013;9:S144–58.

32. Wishart DS. Quantitative metabolomics using NMR. Trends Anal Chem. 2008;27:228–37.

33. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL,

Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS. The human serum metabolome. PLoS One. 2011;6:e16957.

34. Fiehn O. Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. Trends Anal Chem. 2008;27:261–9.

35. Jonsson P, Gullberg J, Nordstrom A, Kusano M, Kowalczyk M, Sjostrom M, Moritz T. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Anal Chem. 2004;76:1738–45.

36. Koek MM, Jellema RH, van der Greef J, Tas CA, Hankemeier T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics. 2011;7:307–28.

37. Hill CB, Roessner U. Metabolic profiling of plants by GC–MS. In: Weckwerth W, Kahl G, editors. The handbook of plant metabolomics. Weinheim: Wiley-VHC Verlag GmbH & Co. KGaA; 2013. doi:10.1002/9783527669882.ch1.

38. Theodoridis GA, Gika HG, Want EJ, Wilson ID. Liquid chromatography–mass spectrometry based global metabolite profiling: a review. Anal Chim Acta. 2012;711:7–16.

39. Forcisi S, Moritz F, Kanawati B, Tziotis D, Lehmann R, Schmitt-Kopplin P. Liquid chromatography–mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. J Chromatogr A. 2013;1292:51–65.

40. Zhou B, Xiao JF, Tuli L, Ressom HW. LC-MS-based metabolomics. Mol Biosyst. 2012;8:470–81.

41. Cubbon S, Antonio C, Wilson J, Thomas-Oates J. Metabolomic applications of HILIC–LC–MS. Mass Spectrom Rev. 2010;29:671–84.

42. Kloos D-P, Lingeman H, Niessen WMA, Deelder AM, Giera M, Mayboroda OA. Evaluation of different column chemistries for fast urinary metabolic profiling. J Chromatogr B. 2013;927:90–6.

43. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res. 2003;2:488–94.

44. Barbas C, Moraes EP, Viallaseñor A. Capillary electrophoresis as a metabolomics tool for non-targeted fingerprinting of biological samples. J Pharm Biomed. 2011;55:823–31.

45. Ramautar R, Somsen GW, Jong GJ. CE-MS for metabolomics: developments and applications in the period 2012–2014. Electrophoresis. 2015;36:212–24.

46. Han J, Danell RM, Patel JR, Gumerov DR, Scarlett CO, Speir JP, Parker CE, Rusyn I, Zeisel S, Borchers CH. Towards high-throughput metabolomics using ultrahigh-field Fourier transform ion cyclotron resonance mass spectrometry. Metabolomics. 2008;4:128–40.

47. Mirnezami R, Spagou K, Vorkas PA, Lewis MR, Kinross J, Want E, Shion H, Goldin RD, Darzi A, Takats Z, Holmes E, Cloarec O, Nicholson JK. Chemical mapping of the colorectal cancer microenvironment via MALDI imaging mass spectrometry (MALDI-MSI) reveals novel cancer-associated field effects. Mol Oncol. 2014;8:39–49.

48. Hyötyläinen T, Orešič M. Bioanalytical techniques in nontargeted clinical lipidomics. Bioanalysis. 2016;8:351–64.

49. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis. 2012;4:2249–64.

50. Ludwig C, Günther UL. MetaboLab –advanced NMR data processing and analysis for metabolomics. BMC Bioinformatics. 2011;12:366–71.

51. Mercier P, Lewis MJ, Chang D, Baker D, Wishart DS. Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. J Biomol NMR. 2011;49:307–23.

52. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M. Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. Curr Bioinform. 2012;7:96–108.

53. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006;78:779–87.

54. Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics. 2006;22:634–6.

55. Lommen A. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem. 2009;81:3079–86.

56. Automated mass spectral deconvolution and identification system. http://www.amdis.net. Accessed July 2016.

57. Misra BB, van der Hooft JJJ. Updates in metabolomics tools and resources: 2014-2015. Electrophoresis. 2016;37:86–110.

58. Koal T, Deigner H-P. Challenges in mass spectrometry based targeted metabolomics. Curr Mol Med. 2010;10:216–26.

59. Lu W, Bennett BD, Rabinowitz JD. Analytical strategies for LC–MS-based targeted metabolomics. J Chromatogr B. 2008;871:236–42.

60. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. Targeted profiling: quantitative analysis of ^{1}H NMR metabolomics data. Anal Chem. 2006;78:4430–42.

61. Guo B, Chen B, Liu A, Zhu W, Yao S. Liquid chromatography-mass spectrometric multiple reaction monitoring-based strategies for expanding targeted profiling towards quantitative metabolomics. Curr Drug Metab. 2012;13:1226–43.

62. Biocrates life sciences. http://www.biocrates.com. Accessed July 2016.

63. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics - A review in human disease diagnosis. Anal Chim Acta. 2010;659:23–33.

64. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, van Duijnhoven JPM, van Dorsten FA. Assessment of PLS-DA cross validation. Metabolomics. 2008;4:81–9.

65. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner H-P. Targeted metabolics for biomarker discovery. Angew Chem Int Ed. 2010;49:5426–45.

66. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. HMDB: the human metabolome database. Nucleic Acids Res. 2007;35:D521–6.

67. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 2006;34:D354–7.

68. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 2009;37:W623–33.

69. Tautenhahn R, Cho K, Uritboonthai W, Zhu ZJ, Patti GJ, Siuzdak G. An accelerated workflow for untargeted metabolomics using the METLIN database. Nat Biotechnol. 2012;30:826–8.

70. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Lida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010;45:703–14.

71. Fahy E, Sud M, Cotter D, Subramaniam S. LIPID MAPS online tools for lipid research. Nucleic Acids Res. 2007;35:W606–12.

72. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburne M. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2008;36:D344–50.

73. Brown M, Dunn WB, Dobson P, Patel Y, Winder CL, Francis-McIntyre S, Begley P, Carroll K, Broadhurst D, Tseng A, Swainston N, Spasic I, Goodacre R, Kell DB. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. Analyst. 2009;134:1322–32.

74. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL. Biomagresbank. Nucleic Acids Res. 2008;36:D402–8.

75. MacKinnon N, Somashekar BS, Tripathi P, Ge W, Rajendiran TM, Chinnaiyan AM, Ramamoorthy A. MetaboID: agraphical user interface package for assignment of ^{1}H NMR spectra of bodyfluids and tissues. J Magn Reson. 2013;226:93–9.
76. Suhre K, Schmitt-Kopplin P. MassTRIX: mass translator into pathways. Nucleic Acids Res. 2008;36:W481–4.
77. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2012;40:D742–53.
78. Frolkis A, Knox C, Lim E, Jewison T, Law V, Hau DD, Liu P, Gautam B, Ly S, Guo AC, Xia J, Liang Y, Shrivastava S, Wishart DS. SMPDB: The Small Molecule Pathway Database. Nucleic Acids Res. 2010;38:D-480–7.
79. Haug K, Salek RM, Conesa P, Hastings J, Matos P, Rijnbeek M, Mahendraker T, Williams M, Neumann S, Rocca-Serra P, Maguire E, González-Beltrán A, Sansone S-A, Griffin JL, Steinbeck C. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res. 2013;41:D781–6.
80. Haw R, Stein L. Using the reactome database. Curr Protoc Bioinformatics. 2012;38:8.7.1–23.
81. Karp PD, Caspi R. A survey of metabolic databases emphasizing the MetaCyc family. Arch Toxicol. 2011;85:1015–33.
82. Liesenfeld DB, Habermann N, Owen RW, Scalbert A, Ulrich CM. Review of mass spectrometry–based metabolomics in cancer research. Cancer Epidemiol Biomarkers Prev. 2013;22:2182–201.
83. Garcia-Perez I, Angulo S, Utzinger J, Holmes E, Legido-Quigley C, Barbas C. Chemometric and biological validation of a capillary electrophoresis metabolomic experiment of *Schistosoma mansoni* infection in mice. Electrophoresis. 2010;31:2338–48.
84. Ganti S, Taylor SL, Kim K, Hoppel CL, Guo L, Yang J, Evans C, Weiss RH. Urinary acylcarnitines are altered in human kidney cancer. Int J Cancer. 2012;130:2791–800.
85. Mastrangelo A, Armitage EG, García A, Barbas C. Metabolomics as a tool for drug discovery and personalised medicine. A review. Curr Top Med Chem. 2014;14:2627–36.
86. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov. 2016. doi:10.1038/nrd.2016.32.
87. Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harb Mol Case Stud. 2015;1:a000588.
88. Spratlin JL, Sekova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology; a review. Clin Cancer Res. 2009;15:431–40.
89. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Arch Toxicol. 2011;85:5–17.
90. Friedrich N. Metabolomics in diabetes research. J Endocrinol. 2012;215:29–42.
91. Becker S, Kortz L, Helmschrodt C, Thiery J, Ceglarek U. LC-MS-based metabolomics in the clinical laboratory. J Chromatogr B. 2012;883:68–75.
92. Zao Y-Y. Metabolomics in chronic kidney disease. Clin Chim Acta. 2013;422:59–69.
93. Ramautar R, Berger R, van der Greef J, Hankmeier T. Human metabolomics: strategies to understand biology. Chem Biol. 2013;17:841–6.
94. Bobrovnikova-Marjon E, Hurov JB. Targeting metabolic changes in cancer: novel therapeutic approaches. Annu Rev Med. 2014;65:157–70.
95. Kumar B, Prakash A, Ruhela RK, Medhi B. Potential of metabolomics in preclinical and clinical drug development. Pharmol Rep. 2014;66:956–63.
96. Dessì A, Marincola FC, Masili A, Gazzolo D, Fanos V. Clinical metabolomics and nutrition: the new frontier in neonatology and pediatrics. Biomed Res Int. 2014;2014:981219.
97. Trushina E, Mielke MM. Recent advances in the application of metabolomics to Alzheimer's disease. Biochem Biophys Acta. 2014;1842:1232–9.

Chapter 2
Collection and Preparation of Clinical Samples for Metabolomics

Andrew J. Chetwynd, Warwick B. Dunn, and Giovanny Rodriguez-Blanco

Abstract A wide range of biofluids (urine, serum, plasma, saliva, etc.) as well as cellular and tissue samples can be collected and investigated in clinical metabolomic studies. The choice of sample is dependent on the clinical question being investigated with biofluids typically studied to identify biomarkers, whereas tissues and primary/immortalised cells are typically studied to investigate mechanisms associated with pathophysiological processes. Methods applied to collect samples, quench metabolism and extract samples differ between sample types from simple collect, dilute and analyse methods for urine to complex washing, quenching and biphasic extraction methods for tissues. The range of sample collection and extraction methods are discussed with sample-specific considerations highlighted. Finally, methods for imaging of cells and tissues and for *in vivo* metabolomic analysis will also be introduced.

Keywords Serum • Plasma • Urine • Human cells • Human tissues • Metabolic quenching • Extraction • Imaging

A.J. Chetwynd • W.B. Dunn (✉) • G. Rodriguez-Blanco
School of Biosciences and Phenome Centre Birmingham, College of Life
and Environmental Sciences, University of Birmingham, Edgbaston,
Birmingham B15 2TT, UK
e-mail: w.dunn@bham.ac.uk

Abbreviations

ATCC	American Type Culture Centre
BMI	Body mass index
BSTFA	N,O-Bis(trimethylsilyl)trifluoroacetamide
CHO	Chinese hamster ovary
CSF	Cerebrospinal fluid
DESI-MS	Desorption electrospray ionisation-mass spectrometry
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionisation
GC-MS	Gas chromatography-mass spectrometry
HILIC	Hydrophilic interaction chromatography
HS-SPME	Headspace solid-phase microextraction
IPA	Isopropyl alcohol
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LLE	Liquid-liquid extraction
MALDI-MS	Matrix-assisted laser desorption/ionisation and mass spectrometry
MSTFA	N-Methyl-N-(trimethylsilyl) trifluoroacetamide
MTBE	Methyl tert-butyl ether
NMR	Nuclear magnetic resonance
PBS	Phosphate-buffered saline
RNA	Ribonucleic acid
SIMS	Secondary ion mass spectrometry
SOP	Standard operating procedure
SPE	Solid-phase extraction
SPME	Solid-phase microextraction
TCA	Tricarboxylic acid

2.1 Introduction

The choice, collection and preparation of biological samples in clinical metabolomic studies can have a significant impact on the metabolomic data acquired, the quality of the analytical data collected and the conclusions derived from the study. Therefore, careful consideration of which sample type to collect, how to collect the sample to provide a metabolic snapshot of the sample at the time of sampling and how to extract samples is required. In this chapter, we will discuss the different types of samples that can be collected and analysed, highlight appropriate methods for collection and sample extraction and discuss current advantages and limitations. These discussions will be focused on untargeted metabolomic studies to investigate hundreds or thousands of metabolites in a single sample. Sample imaging and in vivo real-time analysis will also be introduced.

Whole blood	Cardiac tissue
Plasma	Liver tissue
Serum	Kidney tissue
Urine	Veins/arteries
Amniotic fluid	Primary cells
Saliva	Immortalised cells
CSF	Bone marrow
Tissue dialysate	Exosomes
Faecal water	Others ....
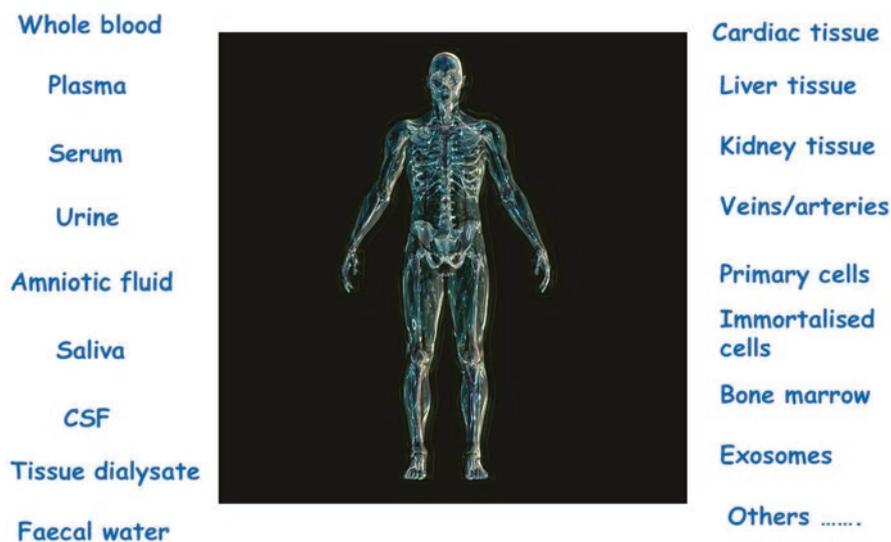
Fig. 2.1 A summary of the different biofluids, cells and tissues studied in clinical metabolomics

2.1.1 Sample Types Investigated in Clinical Studies and Associated Considerations

The range of samples that can be collected in a clinical study is wide ranging and can consist of biofluids, primary cells and tissues. Figure 2.1 provides an overview of different sample types accessible and applied in clinical metabolomic research, either currently or those that can be expected to be investigated in the future.

The sample type chosen can be defined by the clinical question being investigated. As examples:

(a) Biomarkers applied in the clinic are primarily assayed in biofluids, and therefore discovery studies to identify a single biomarker or a biomarker panel are typically also applied to biofluid samples.
(b) Studies to understand pathophysiological mechanisms are typically interested in a specific organ in the human body. To enable research to investigate the 'site of action' of a specific metabolic perturbation, then the organ should be sampled and studied.

However, in some studies, the preferred sample type cannot be collected, and a different 'surrogate' sample type has to be investigated. For example, the collection of 2000 kidney biopsies from healthy and diseased subjects to study metabolic mechanisms in kidney diseases is not feasible in relation to the expense of sample collection (e.g. surgery would be required for all subjects) and in performing an ethically acceptable study (e.g. the collection of kidney biopsies from healthy subjects is not ethically acceptable). However, the collection of urine or plasma samples could be performed as biofluid samples can be collected in an ethically acceptable

manner, and costs for collection would be significantly lower including for large-scale studies. Consideration as to whether the biofluid can identify metabolic changes related to the tissue of interest should also be undertaken. In this example, the analysis of urine, which is a by-product of the pathophysiological operation of the kidney, is a suitable 'surrogate' sample to determine metabolic or physiological changes in the kidney.

2.1.2   Sampling Considerations

The collection of samples requires a number of objectives to be met. The primary objective should be to ensure that the sample collected is qualitatively and quantitatively representative of the sample before it was collected. When this objective is met, then the data acquired from the sample is biologically representative of the question being asked, and any results can be viewed as valid. With this in mind, samples can be separated into two general classes: (i) metabolically active samples and (ii) metabolically inactive samples. Figure 2.2 shows the differences in relation to sample collection and extraction for these two classes of samples as defined for a mammalian cell culture. The extracellular metabolome can be typically viewed as a metabolically inactive sample, and the intracellular metabolome is viewed as the metabolically active sample.
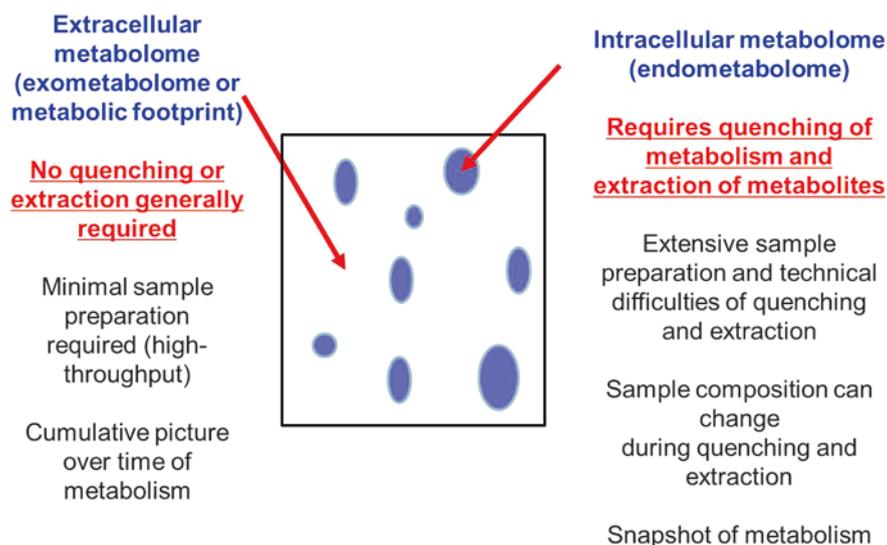

Fig. 2.2 A comparison of sampling, metabolic quenching and extraction required for intracellular and extracellular metabolome samples. The example here is shown for the culture of mammalian cells

All cells and tissues can be viewed as metabolically active, and therefore quenching of metabolism is required and will be discussed later in this chapter. Biofluids can be defined as metabolically inactive as they are extracellular, but care should be taken. Urine is metabolically inactive, but serum/plasma, CSF and saliva could be viewed as metabolically active, and care should be taken. As a minimum we recommend that biofluids are stored on ice while being processed and then stored following processing for long periods at −80 °C.

Importantly, tissues require metabolic quenching and extraction. Metabolic quenching of tissues typically involves rapidly washing in a saline or phosphate-buffered solution and rapid quenching by placing in liquid nitrogen [1]. Most tissues are collected in an operating theatre with the exception of the muscle and skin (as a punch biopsy), faeces and the placenta, which is expelled after pregnancy. Health and safety guidelines do not allow liquid nitrogen to be placed in an operating theatre, and so tissues have to be transported to a different location, preferably on ice. Therefore, rapid quenching of tissue in this environment is not always feasible, and this should be considered when the requirement to study metabolic pathways with a high metabolic flux is required (e.g. the glycolytic and TCA metabolic pathways in cancerous tissue).

2.2 Biofluids

2.2.1 Urine

Urine is one of the most widely studied biofluids in metabolomic research, primarily due to its ease of collection, as discussed later. The urinary metabolome has been used to investigate the metabolic consequences of disease for the entire body due to its being a major excretory route of water-soluble metabolites and xenobiotics [2]. The presence of both endogenous and exogenous metabolites means that urine metabolomics may be implemented for disease biomarker discovery [3], drug discovery and characterisation [4], determination of nutritional status [5] and effects of environmental toxicants [6]. To date over 3100 metabolites have been characterised in human urine, and this is considered the minimum number of metabolites present as new sample preparation and analytical techniques are likely to uncover other metabolites present at lower concentrations [7].

2.2.1.1 Sample Collection

When collecting urine samples the timing of collection can make an appreciable qualitative and quantitative difference in the urinary metabolome [8, 9]. Typically there are three types of urine sample that can be collected, first morning void, spot urine and a 24 h urine collection [10]. Generally first morning voids are a preferred sample type, following an overnight fast of several hours thus reducing the effect of the last

meal or medication [11, 12]. Spot urine samples are taken at a time point during the day and are particularly common when some form of intervention such as dietary or pharmaceutical has been administered [10]. However, it is well known that a number of metabolites are excreted in a diurnal rhythm or cosine rhythm [8, 9]. Ideally, a urine sample would comprise of a pooled sample of all voids within a 24 and h period thus reducing the impact of any circadian variation in the sample, defined as a 24 h urine and representing a complete 24 h circadian cycle. Spot urine samples at specific time points that are uniform between all subjects such as first morning void also help to counter this problem when a 24 h sample collection is not feasible.

Collection of urine samples is predominantly easy in the general population. Subjects are asked to provide a urine sample into a pot and return the sample. Urine samples can be collected in the home and transported to a clinic, which is appropriate for 24 h urine samples, or spot urines can be collected at the clinic. No highly trained staff is required for the collection of urine samples. For immobile subjects urine samples can be collected with a catheter, and for babies urine can be collected with absorbent pads in nappies.

2.2.1.2 Storage and Stability of Urine Samples

One concern with metabolomics is the time taken to collect all samples, extract samples and then collect a metabolic profile for each sample, as in most studies extended periods of sample storage is required, thus storage must have minimal/no effect on the metabolome of the sample being studied [13, 14]. Several studies utilising LC-MS, GC-MS and NMR have investigated the effect of different storage temperatures and number of freeze thaw cycles [13, 15–17]. It has been demonstrated that urine samples rapidly degrade when stored at room temperature even for short periods of time with glycolytic metabolites showing signs of degradation or metabolism [16]. For long-term storage, −20 °C has been demonstrated to have no effect on the urinary metabolome after storage for 6 months. However, −80 °C is recommended as no longer-term stability studies of length greater than 6 months have yet to be carried out applying untargeted metabolomics [2, 13, 16, 17]. Thus, it is optimal to freeze urine samples as soon as possible following collection to minimise the time spent at room temperature.

Analytical run times in large-scale clinical metabolomic studies can operate from hours to days in length during which time urine samples are kept in the autosamplers of LC-MS, GC-MS and NMR analysers at 4 °C. During this time, it is important that metabolic profiles at the end of the run are still representative of those in samples analysed at the start of the run. Studies suggest that 48 h at 4 °C does not significantly alter the urinary metabolome, as such it is recommended that only 48 h worth of samples should be stored at any one time in an autosampler [13, 17]. During the metabolomic workflow from sample collection to analysis, it is likely that samples will be frozen and thawed a number of times due to the time taken to collect and prepare samples. The impact of multiple freeze thaw cycles has been investigated using LC-MS, and it is reported that up to nine freeze thaw cycles

have no significant impact upon the urinary metabolome [13, 17]. However, it is prudent to limit the number of freeze thaw cycles to as few as reasonable possible [18].

A further complication to urine collection and storage is bacterial contamination and metabolism with a contribution to urinary metabolites by bacteria. While the urine itself is sterile in a healthy individual, bacterial contamination can be introduced via the urethra during urination. As such it is recommended to collect a midstream urine sample to minimise this risk [19]. Furthermore, a number of antibacterial additives such as sodium azide and sodium fluoride have been demonstrated to reduce metabolic variation as a result of bacterial contamination and metabolism [18, 20]. Storage of samples at −80 °C is also known to prevent any metabolism of urinary metabolites by contaminating bacteria with the added benefit of not adding any chemicals to the sample itself, which can be beneficial for GC-MS, NMR spectroscopy and LC-MS studies [18].

2.2.1.3 Sample Preparation

Sample preparation for urinary metabolomics is an often overlooked piece of the experimental workflow yet plays a vital role in the quality, reliability and coverage of the urinary metabolome [14, 21–23]. To date the most widespread preparation methods for LC-MS or NMR are either the injection of neat urine following centrifugation or injection of diluted urine following centrifugation [10, 19, 23]. Both neat and dilute urine preparations have benefits for global metabolomics. Neat is unmodified and therefore contains the unadulterated metabolome; however, when analysed by LC-MS, the sensitivity to low-abundance metabolites suffers as a result of high-abundance co-eluting peaks and subsequent ion suppression. The high salt content of urine encourages the formation of a range of adducts within the electrospray source, in addition to fouling the LC column and the ESI source [24–26]. The dilution of urine prior to analysis reduces this ion suppression and may potentially allow detection of some less abundant metabolites, which were previously hidden by a high-abundance co-eluting metabolite. However, it has been shown that many low-abundance peaks are diluted to below the limit of detection using a 'dilute-and-shoot' method [24, 27, 28].

The use of solid-phase extraction (SPE) sample preparation methods, which incorporate a sample cleanup and a sample concentration step, is becoming more popular [28–30]. These methods allow unwanted urinary salts, matrix effects and proteins to be removed and with sample pre-concentration low-abundance metabolites become more easily detected [23]. Importantly, the use of SPE has been shown to have a similar repeatability to 'dilute-and-shoot' sample preparation suggesting that a more comprehensive coverage of the metabolome can be achieved without compromising on data quality [28, 29]. The use of such an extensive sample cleanup may also have the added benefit of reducing source and chromatographic column fouling and extending column life.

In GC-MS analysis, it is typical to lyophilise the urine samples prior to a chemical derivatisation step in order to improve derivatisation efficiency and thus sensitivity [12, 31]. In these cases urine is typically derivatised using either BSTFA or MSTFA, although samples requiring storage are more stable following derivatisation with MSTFA [12]. Recently, a number of studies have begun to utilise liquid-liquid extraction and SPE as a sample cleanup step without the need for a derivatisation stage [32]. Furthermore, the use of headspace solid-phase microextraction (HS-SPME) has shown promise as a method of increasing sensitivity to volatile organic compounds in urine. Here urine is heated, and volatiles diffuse into the gas phase and become trapped on the SPME fibre before being thermally released into the GC-MS instrument [33, 34]. SPME has the benefit of not requiring any solvent and being a quick technique while incorporating a sample concentration step, which has benefits in terms of costs and environmental impact [33–35]. The high concentration of urea in urine can impact significantly on the quality of GC-MS data. The enzymatic removal of urea with urease is commonly applied to allow the detection of low-abundance metabolites, which co-elute with the broad urea peak [36]. However, some studies have reported a detrimental effect of urease treatment [37].

2.2.1.4   Pre-analysis Normalisation

The solute concentration of urine is known to vary by up to 15 times both between and within individuals providing separate samples [38]. This is due to a number of contributing factors such as fluid intake and health status [38, 39]. Consequently, without a normalisation method to correct for this variation in urine concentration between samples, metabolite variation may be mistakenly attributed to the case study, when in fact they are a result of differences in urine concentration between individuals. Recently a number of studies have shown that equalising urine concentration during the sample preparation process greatly reduces the effect on the biological results of varying urine concentration. This is achieved by measuring the urine concentration using methods such as osmolality or specific gravity and diluting samples down to the lowest concentration [40, 41].

2.2.2   Serum and Plasma

Blood serum and plasma are the second most frequently applied biofluid in metabolomic studies after urine. Blood provides a snapshot of metabolism that integrates many tissues in the human body through its interactions with these tissues and so provides a metabolic picture of global metabolism though with a lower level of specificity in comparison to urine. The choice of serum or plasma is an important but as of yet not fully answered question, and different research groups use serum or plasma. A number of studies have investigated the qualitative and quantitative

differences between serum and plasma, though only small differences have been identified, which do not providea a clear choice for either of the biofluids [42–44]. Indeed serum and plasma have been applied in a range of different applications related to cancer [45], endocrinology [46], inflammatory diseases [47] and diseases of the cardiovascular system [48].

2.2.2.1 Sample Collection

Unlike for the collection of urine and saliva, the collection of blood requires trained staff (phlebotomists) to collect and process blood in suitable volumes, and therefore collection is routinely applied in the clinic. The collection of dried blood spots is one exception that does not require trained staff. Sample collection can be performed by each subject, and samples can be transported at room temperature to the clinic by the subjects or via postal services (e.g. see reference [49]).

The main difference between serum and plasma is the presence or absence of clotting. For serum, whole blood is collected into tubes and is allowed to clot for a specified time and temperature before centrifugation to pellet the clot and cells and provides the liquid serum containing all metabolites and proteins not removed in the clotting process. Large differences in the time and temperature of clotting have been shown to influence the metabolite composition of serum, and standardised protocols should be applied [50, 51]. The authors of this chapter recommend allowing samples to clot at 4 °C to reduce any metabolic activity, which in the human body operates optimally at 37 °C. Plasma is the liquid volume of whole blood, and the collection of plasma does not involve a clotting process. Instead, whole blood is mixed with an anticoagulant to inhibit clotting followed by centrifugation to separate the liquid plasma from red and white blood cells. Typical anticoagulants include lithium heparin, EDTA and citrate. We recommend the use of lithium heparin, which is a high molecular mass biochemical unlike citrate and EDTA, which have similar molecular masses to metabolites, and indeed citrate is an endogenous metabolite.

For the collection of serum or plasma, whole blood is collected into different types of tubes following venepuncture [52]. The tubes allow for the collection of serum or plasma, with serum tubes containing no additive or a gel to aid clotting, and plasma tubes are coated with an anticoagulant to inhibit clotting. Tubes are inverted several times to allow mixing of anticoagulants with whole blood. The tubes are centrifuged, and the liquid fraction (serum or plasma) is transferred to separate tubes for storage [53].

2.2.2.2 Storage and Stability

It is recommended that serum or plasma is stored as 0.5 or 1.0 mL aliquots at −80 °C. Both biofluids contain proteins and enzymes, for example, released from cells and tissues in the human body before sampling, which can provide metabolic

activity. Processing of whole blood should be performed ideally at 4 °C with plasma and serum aliquots being thawed on ice before metabolite extraction. As discussed for urine samples, the time-extracted samples placed in an autosampler at 4 °C should also be considered, and the authors recommend a maximum time of 48 h in an autosampler.

2.2.2.3 Sample Preparation

Serum and plasma are a composite of water, metabolites and higher molecular mass biochemicals including proteins, RNA and DNA. Metabolites range from small ionic species like sodium and ammonium ions, through water-soluble metabolites to lipids, and the method of sample preparation is dependent on the metabolites to be investigated. All protocols have the objective to remove higher molecular mass biochemicals and extract metabolites into a suitable solvent system. The most frequently applied method to extract metabolites is liquid-liquid extraction (LLE) [54–56]. Here an organic solvent is added in excess to serum or plasma, which acts to precipitate the higher molecular mass biochemicals while allowing all or a subset of metabolites to remain in solution. A number of different solvents have been reported including methanol, acetonitrile, isopropyl alcohol (IPA) [57] and acetone [58], and the use of different temperatures and extraction times have been investigated [54, 59]. It is the choice of the researcher to define an appropriate solvent and temperature for their biological question and metabolites of interest; for example, the use of IPA for extraction of lipids is highly reported [57]. Following precipitation for a defined period of time, the samples are centrifuged to pellet the precipitate, and the supernatants are aliquoted to a different tube for further processing or analysis.

Whether the drying of samples is required should be considered in relation to the solvents applied. For GC-MS applications, sample drying is typically required prior to derivatisation with MSTFA. However, for LC-MS applications, sample drying can be avoided if an appropriate solvent is used for sample dilution. For example, for HILIC applications, which require injection of the sample in an organic-rich solvent, the samples can be injected directly or after dilution on to the chromatographic column without drying.

The presence of lipids, predominantly glycerophospholipids, can impact on the quality of data and the number of unique metabolites detected in untargeted studies [60, 61]. Selected extraction procedures or further sample manipulation steps can be applied to remove single or multiple lipid classes from extracted samples prior to analysis. The use of biphasic extraction methods to move lipids into an organic solvent and water-soluble metabolites into a water/methanol solvent, with each solvent immiscible in the other, provides the ability to analyse only the lipid fraction or the water-soluble fraction [58]. The use of complementary LC-MS assays can benefit from this approach, for example, the use of HILIC methods to analyse the water-soluble metabolites and the use of a C_{18} reversed-phase method for analysis of lipids. SPE can also be applied, typically the sorbent is a C_{18} phase which absorbs lipids while allowing water-soluble metabolites to be eluted and analysed [61, 62].

2.2.3   Other Biofluids

2.2.3.1   Saliva

Saliva has for the most part been an overlooked biological fluid for metabolomic analysis yet is thought to accurately reflect the plasma metabolome. With this in mind, saliva sampling could make a desirable surrogate for plasma as it does not require specialist collection and is easy to collect and is non-invasive [63, 64].

Consideration of the type of saliva sample either stimulated, whereby saliva production is stimulated with citric acid, or non-stimulated (resting) is required. A comparison of the two has shown that TCA cycle and amino acid profiles are disrupted and found at lower concentrations in stimulated saliva samples as a result of dilution [65]. As such the majority of studies into saliva metabolomics choose to use unstimulated samples, usually following a period of fasting and delayed oral hygiene to prevent contamination [66–68].

Saliva samples are typically collected and then frozen while awaiting analysis. Storage at −20 °C for up to 3 weeks has been shown to have no detrimental effect on the salivary metabolome [65]. Prior to any sample preparation, saliva samples are centrifuged to remove cellular and food debris [63].

The majority of salivary metabolomic studies have been performed using NMR, whereby samples are buffered in 0.2 mol L^{-1} phosphate buffer and diluted with D_2O [69, 70]. In LC-MS-based analysis, saliva has been hydrolysed using both NaOH and HCl in order to hydrolyse and release metabolites from proteins and cells present in the sample [63]. Hydrolysis via an ultrasonic probe has also been investigated as an enhanced hydrolysis method [63]. The use of any hydrolysis step has been demonstrated to significantly increase the number of metabolites detected; furthermore, a sample pre-concentration step is often required due to the low abundance of salivary metabolites [71].

2.2.3.2   CSF

Cerebrospinal fluid (CSF) is a fluid that fills the spinal column and brain ventricles and plays a role in fluid regulation and nutrient transport in the central nervous system [72]. Metabolomic analysis of CSF offers great promise for understanding neurological disorders such as Alzheimer's disease and amyotrophic lateral sclerosis [73].

CSF samples are collected via ahealthcare professional using a lumbar puncture. In a non-traumatic collection, the fluid should be clear and void of blood. Following collection, samples require centrifugation to remove any cell debris and then frozen at −20 °C for long-term storage. For short-term storage, the CSF metabolome is known to be stable for up to 2 days at 4 °C but known to be unstable at temperatures exceeding 5 °C [72, 74].

To date CSF samples have mainly been analysed using basic dilution sample preparation. However, lyophilisation and pre-concentration have been used in NMR to improve coverage of the metabolome at the loss of some volatile organic metabolites

[74]. For GC-MS analysis, samples may be derivatised using either BSTFA or MSTFA [75]. So far, very little work has been completed to develop new sample preparation methods for CSF metabolomics; given the sparse nature of these methods, it may be of interest to develop new methods utilising SPE or SPME in order to increase the ability to detect lower-abundance metabolites.

2.2.3.3   Sweat and Breast Milk

In addition to the biofluids already discussed, a number of others are used although much less frequently. Sweat composition is known to be modified in several disease states and as such is becoming a more popular matrix for metabolomic analysis [78]. Sweat collection is stimulated using a sweat inducer, which is applied to the skin, heats the area and collects the sweat [76, 77]. Few studies have analysed sweat but those that have used a basic sample dilution or neat sample following centrifugation for LC-MS or NMR analysis [76, 78]. Additionally, a sample cleanup method to remove the high salt content of sweat samples and allow a pre-concentration step using SPE has been developed [78].

Human breast milk has also begun to become a more popular biological fluid for metabolomic analysis particularly in the field of nutritional metabolomics for infants [79]. Care should be taken to record when the patient last fed to reduce the impact of the diet on the breast milk metabolome, and samples are cooled to as cold as possible to minimise any degradation [81]. Here for both LC-MS and NMR, a number of methods have been reported for extracting the polar and non-polar metabolites [80, 81].

2.3   Primary and Immortalised Cells

Metabolomic studies in human biological samples have mainly focused on the analysis of biofluids for clinical applications such as disease diagnosis or prognosis [82]. Metabolic profiling of a whole organism does not provide information about specific cell types under different conditions, which may be important for the development of drugs targeting specific cell phenotypes. In this regard, metabolomic studies of cell lines can be used to complement the information provided by whole organism metabolic phenotyping [83]. Metabolomic studies of mammalian cells are easier to perform and to interpret because there are no confounding factors to consider such as genotype, gender, age, BMI or alcohol intake, which are present in other clinical metabolomic applications [83, 84]. Another advantage of performing metabolomic experiments on cell lines is the relatively easy correlation with other 'omics' approaches such as transcriptomics or proteomics, thus enabling the construction of biological networks and pathway interactions at the system levels [85]. Cell culture metabolomics has been used in different areas including drug discovery and foodomics [83, 86, 87]. It has been explored recently for metabolic tracer and flux analysis [88, 89] and as a tool in basic biology to understand molecular mechanisms such as metabolic reprogramming in cancer cells and the Warburg

effect [90, 91]. Different cell lines from different organs, including the first human cancer cell lines, have been collected in cell culture biobanks and can be easily accessed from biological resource centres such as the American Type Culture Centre (ATCC, www.atcc.org). Cell culture systems are divided into primarily cell culture, cell lines and cell strains, which can grow in suspension or adherently [92].

2.3.1 Considerations

Metabolomic experiments in cell lines face different challenges in sample preparation in comparison with the analysis of body fluids. Some of these issues include variability of growth medium formulation, influence of number of passages, metabolic quenching and metabolite extraction, which are time-consuming and might lead to metabolite degradation and leakage from the cell before extraction [86, 93].

Some of the above-mentioned issues in cell metabolomics can be solved by appropriate experimental design and the development of standard operating procedures (SOPs) for metabolite extraction. Experimental design is important whichever sample is to be studied [94, 95]. For example, it is also recommended to randomise both sample extraction and sample analysis accordingly. Cell cultures grown in different flasks or wells, and subsequently treated in a similar way (i.e. drug treatment), are considered as biological replicates [83]. Different aliquots sampled from the same flask or well after a defined treatment can be considered as technical (and not biological) replicates. For cell culture metabolomics, six biological replicates (i.e. from 6-well plate cultures) are recommended.

Another concern in cell line metabolomics, especially when comparing different cell types, relies on variations in growth media formulation. It is recommended to use the same media (and batch of media) for all cell lines to reduce variability in metabolic profiles. However, the use of sub-optimal media can also affect the metabolic phonotype because cells might not achieve the same growth conditions [93]. Mammalian cell culture media, for example, are complex mixtures containing several buffers, amino acids and other variable components such as foetal bovine serum that can cause significant ion suppression effects in LC-MS. When analysing intracellular metabolites, the growth media must be efficiently washed from the cell pellet to avoid contamination with exogenous components [94]. On the other hand, when analysing extracellular metabolites, particular attention should be taken to method development and evaluation, to avoid leakage of intracellular metabolites into the extracellular medium during sampling [95].

2.3.2 Sampling and Extraction

Metabolic profiles should represent the physiological status of the cells at the time of sampling. Therefore, metabolic quenching constitutes the key step to minimise changes in metabolite levels, to improve the reproducibility and to avoid
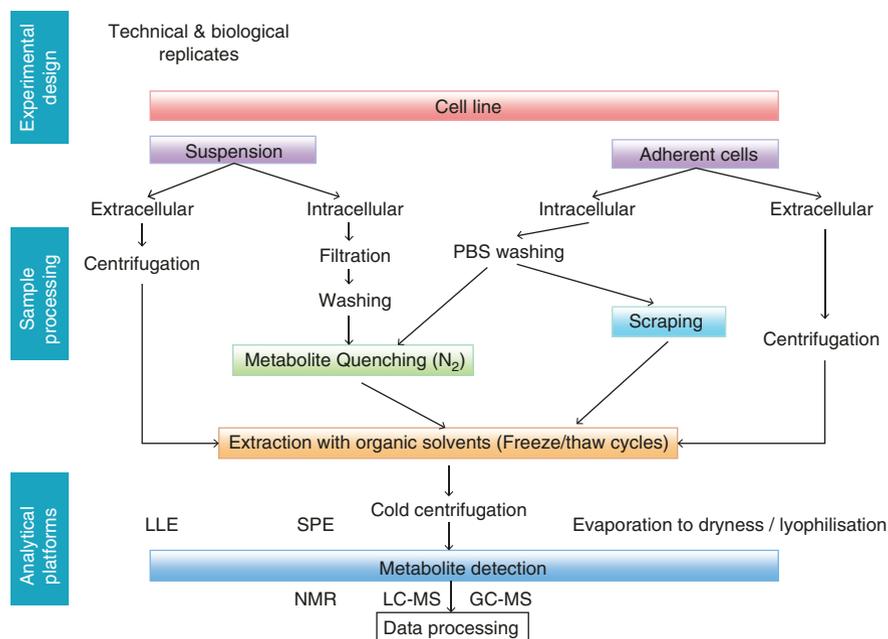
Fig. 2.3 Schematic methodological workflow proposed for metabolic quenching and metabolite extraction from mammalian cell cultures

misleading results [93, 96]. Therefore, metabolite extraction from cell lines must be performed as quickly as possible to avoid enzymatic reactions that can change the qualitative and quantitative composition of the sample. A schematic representation of the methodological steps in metabolite extraction from cell lines is presented in Fig. 2.3.

The analysis of the extracellular metabolome, also known as the exometabolome, metabolic footprint or spent culture media, can provide significant advantages as defined in Fig. 2.2. Sampling is relatively simple with separation of media from cells in suspension achieved either by cold centrifugation or by the use of low molecular mass cut-off filters [97]. It is recommended to perform these procedures rapidly in order to avoid metabolic activity and quantitative changes to the metabolic profile. Cold centrifugation usually takes 5–15 min to complete, and in some cases, a washing step is needed to remove salts interfering with the mass spectrometry analysis [93]. Filtration is quicker, but it is more expensive, and the membrane can get easily blocked [98]. A robust fast filtration sampling method has recently been reported called MxP® FastQuench, followed by lipid/polar extraction for cells in suspension. This method suggests an efficient metabolite recovery and the potential to be extendable to all mammalian cell types [99]. The collection of the extracellular metabolome in adherent cells is easier because the media can be collected by pipetting, followed by a cold centrifugation step to eliminate cell debris. Metabolite extraction is then performed by adding the extraction solvent. After centrifugation,

the supernatant is analysed directly, or evaporated for further reconstitution for the preferred analytical procedure [100].

Intracellular metabolite profiling of cells in suspension can be achieved following metabolic quenching by heating or by adding ice-cold solvents. Here the suspension (a composite of media and cells) is sampled followed by metabolic quenching and typically subsequent separation of cells from the media and solvents applied during quenching. The addition of ice-cold 100 % methanol at −40 °C for metabolic quenching should be avoided because it might lead to metabolite leakage after membrane solubilisation [86, 101]. Cold isotonic saline quenching solutions at 4 °C have arisen to prevent cell membrane damage. For cells in suspension, a mixture of acetonitrile/water (1:1) was demonstrated as optimal for metabolite extraction in a detailed study with different conditions for Chinese hamster ovary (CHO) cells [102].

Analysis of intracellular metabolites from adherent cells requires some additional and critical steps. The medium must first be discarded either by pouring it from all samples simultaneously or by aspiration with a pipette or a vacuum pump. Samples are then washed two to three times with ice-cold phosphate-buffered saline (PBS) to remove any medium residue. The excess of PBS is then washed out also by aspiration from all the samples; this is a critical step because the PBS can generate ion suppression in mass spectrometry analysis [103]. The extraction solvent (stored in the freezer at –20 °C) is then added to the cells in the range of $1–2 \times 10^6$ cells mL^{-1}, and this volume is consistent for all the samples. Plates/flasks can be incubated at −80 °C for 10 min, and then the cells are scraped to detach them from the growth surface for a final centrifugation step where the supernatant contains the metabolites of interest [103].

Different solvent compositions can be used for metabolite extraction from cell cultures. A mixture of methanol/water (4:1) has been reported, but different compositions containing methanol/acetonitrile/water can favour the extraction of highly polar metabolites such as nucleotide triphosphates [104]. Some authors have described the use of combined quenching and extraction procedures. A quick method for metabolite extraction involving a one-step washing with water, followed by direct addition of liquid nitrogen and a 1min solvent extraction step using a methanol/chloroform mixture (9:1), has been reported [105]. A recently proposed method for the analysis of cell metabolism using LC-MS and isotope tracers has been reported. They suggest a polar extraction solvent-containing methanol/acetonitrile/water (5:3:2), followed by a cell scraping procedure to fully wipe debris from the growth surface [89]. Others have reported cell scraping into an extraction solvent as an optimal procedure to simultaneously quench and harvest adherent cells [106].

For adherent cells, a trypsinisation procedure is commonly used to detach cells from wells or flasks. However, it has been reported that this procedure might affect the metabolic profiling because extra steps of washing and centrifugation can result in metabolite lost. In addition, this enzymatic procedure is cell type dependent, so if the cells are exposed to trypsin for more than 2–3 min, then cell lysis can occur with subsequent intracellular metabolite leaking [107]. Supernatants from simultaneous quenching and extraction can be directly injected into the mass spectrometer.

However, depending on both the number of cells in the experiment and the extraction method used for metabolite isolation, the concentration of some metabolites can fall below the limit of detection of the analytical instrument. A sample concentration procedure can be performed by either evaporating to dryness or lyophilising the samples and subsequent reconstitution with an appropriate solvent for the analytical platform of choice.

2.3.3   Normalisation

Normalisation of metabolomic data from cell line studies constitutes an important topic to consider when developing and evaluating a method. Cell counting can be a good practice, and it must be performed immediately prior to the incubation period. Separate experiments have to be performed exclusively to determine the number of cells. Alternatively, normalisation procedures including total protein content and total peak area have been used [108, 109]. DNA concentration has also been recently proposed as an efficient and robust method for normalising metabolomic data [110].

2.4   Tissues

A wide array of tissues have been studied applying metabolomic approaches including muscle [111], cardiac tissue [112], liver [113], cancerous lung tissue [114], placenta [115], arteries [116] and skin [117]. The selection of a tissue sample provides localised metabolic activity snapshots relevant to the tissue chosen in comparison to biofluids, which can reflect changes in multiple organs in the human body. The study of tissues is normally performed to investigate mechanistic differences related to pathophysiological processes.

2.4.1   Sample Collection, Storage and Stability

Human tissues are metabolically active and therefore require rapid metabolic quenching when they have been collected. Typically, tissues are collected, then are rapidly washed in a phosphate-buffered aqueous solution or in saline to remove as much blood as possible and are then rapidly frozen in liquid nitrogen to quench metabolism [118]. Washing of tissues to remove as much blood as possible is an important step as the blood metabolome will be different to the tissue metabolome and therefore contaminates the tissue metabolome. As for other sample types, tissue should be separated into appropriate size aliquots and stored frozen at −80 °C. For untargeted or targeted metabolomic studies, typically 20–100 mg of tissue is required to provide good coverage of the tissue metabolome [115, 119]. One exception is faeces, where up to 2 g of material is typically extracted [120].

Human tissue samples are normally collected via an invasive procedure with the exception of a small number of tissues, which can be collected with minimally invasive or non-invasive techniques. The placenta is a large pregnancy-related tissue, which is naturally expelled from the body following delivery of the foetus and which can be sampled and studied [115]. However, metabolic activity can still be in operation during this process, and different levels of oxygenation and nutrient delivery before and during sampling should be considered; it has been shown, for example, that the metabolic composition of tissue early and late in the first trimester is different and related to the level of blood delivery of oxygen and nutrients [115]. Skin and tissue biopsies can be collected with minimally invasive techniques; faeces can be collected without an invasive technique, while most other tissues are collected in a clinic or operating theatre.

One important aspect for tissues is that they are not homogenous when compared to biofluids. If small samples are collected from a large tissue, then careful consideration must be taken to ensure the same area of the tissue is collected from different subjects. For example, the adult human liver weighs approximately 1.5 kg and has four lobes, which are metabolically different. Therefore, the collection of 50 mg of tissue has to be carefully considered to ensure the same area of the liver is being sampled. Small metabolic differences have been observed in different parts of a tissue. For example, differences between the centre and edge of placenta have been reported [115].

The metabolic stability of tissue samples has not been studied in detail applying untargeted metabolomic approaches to our knowledge. We recommend similar storage times and range for all samples collected for a study and matching of storage times between different biological classes. As for other sample types, we recommend storage of the tissue at $-80\,°C$ and storage in an autosampler at $4\,°C$ following extraction for a maximum of 48 h.

2.4.2   Sample Preparation

The preparation of tissues can be separated into two processes. Tissue is normally homogenised applying physical techniques including manual mortar and pestle [115] or ball grinding with stainless steel or silica particles [121, 122]. Normally, this process is performed with the tissue and extraction solvent combined as the physical process of homogenisation also results in cell lysis and extraction of metabolites from the tissue into the solvent. The solvent or solvents applied vary depending on the assay to be applied. Extraction methods are either monophasic (one miscible solvent system) or biphasic (two immiscible solvent layers). Monophasic extraction methods provide an extract, which typically provides greater coverage of the metabolome in a single solution. Biphasic extractions have the advantage that water-soluble metabolites can be separated from lipids through the use of two immiscible solvents. Although each solvent system contains fewer metabolites, when analysed separately and the data combined, then a greater

coverage of the tissue metabolome is observed because of fewer interferences in the assays applied [123]. Monophasic extractions typically apply a water/methanol or water/acetonitrile solvent system through a combination of water/chloroform/methanol as a single-miscible solvent system has been used [124]. The use of IPA with other solvents to selectively extract lipids has also been applied [123]. For biphasic extractions, water and methanol are typically applied with a non-polar solvent and chloroform [115], dichloromethane [119] and MTBE [125] have all been applied. The most common technique applied is the Folch extraction, developed in the 1950s, which extracts into a single-miscible solution of water/chloroform/methanol followed by an addition of further water to create phase separation [126]. More recently, the Matyash method has grown in frequency of application, especially for the analysis of lipids [127]. One experimental limitation is observed in biphasic extractions. The two immiscible solvents are separated by a layer composed of the cellular debris. The aliquoting of the top layer is relatively easy, but the aliquoting of the lower layer requires the puncturing of the debris layer with needle or pipette which can cause some of the debris to enter the lower phase and contaminate it. Chloroform and dichloromethane are the lower layers when applied with methanol and water as their densities are greater than water. The water/methanol solution is the lower layer when applied with MTBE as MTBE has a density less than water. Therefore, the choice of solvent can depend on whether water-soluble metabolites or lipids are to be investigated. If lipids are to be investigated only then MTBE/methanol/water is appropriate, as the lipids will be present in the upper layer, whereas if water-soluble metabolites are to be investigated, then a chloroform/methanol/water extraction method is appropriate to ensure the water/methanol solution is the top layer.

2.4.3   Pre-analysis Normalisation

The normalisation of tissue mass extracted is an important aspect of tissue metabolomics. With only a 5–10 % difference in the masses of tissues to be extracted, then the same volume of extraction solution can be applied. However, with larger variations of tissue masses being extracted the volume of solvent should be normalised to the mass of tissue. For example, if you were extracting two tissues of mass 20 and 40 mg, then you would use 2× the volume of solvent for the 40 mg tissue compared to the 20 mg tissue. This ensures the ratio of tissue and solvent is identical for all samples as this ratio can influence the percentage recovery of metabolites [128].

2.5   Cell and Tissue Imaging

The sampling of cells and tissues typically involves the homogenisation of tissues and the lysis of cells during the extraction protocol. These processes remove relevant qualitative and quantitative information on the distribution of metabolites within a single cell or the collection of cells in a tissue. The subcellular location of

metabolites can provide further data in many clinical studies, which aid mechanistic interpretation. Imaging of intact cells and tissues can provide information on the spatial distribution of metabolites. A range of different imaging technologies can be applied dependent on the sample size and spatial resolution required.

Mass spectral imaging is frequently applied in metabolomics and includes MALDI-MS imaging [129], DESI-MS imaging [130] and SIMS imaging [131]. Thin tissue slices are prepared followed by analysis. Some controversy has been centred on whether paraffin-imbedded tissues, used commonly in pathology, could be applied for mass spectral imaging; recent work has shown the applicability of these sample types [132]. These imaging techniques raster the source across a tissue and collect a mass spectrum at each pixel of the sample. The mass spectral image is a composite of all the pixels and through computational analysis the distribution of different metabolites can be visualised. This approach will be discussed further in Chap. 12.

2.6   Studies Without the Need for Sampling

The importance of a suitable method for sample collection and preparation has been highlighted for the entire sample types discussed in this chapter. Cells and tissues require rapid metabolic quenching, which is not always feasible in the clinical environment. The time required for preparation and analysis of samples can be a number of hours, which is not ideal when data is being applied for clinical decision-making, especially during surgery. The ability to collect data *in vivo* and in real time during surgery and the use of these data for rapid clinical decision-making will move metabolomics into the operating theatre. The recent invention and translation of the intelligent knife (iKnife) is the most significant example of *in vivo* and real-time data collection being applied during surgery. Here the surgeon applies electrosurgical knives, which use an electrical current to rapidly heat tissue, cutting through it while minimising blood loss. This process vaporises the tissue and releases a smoke plume that can be sucked through a tube into a mass spectrometer placed in the operating theatre to provide real-time data for clinical decision-making [133]. For example, this technique can differentiate between tumour and healthy tissue allowing all of the tumours to be removed while not removing too much healthy tissues, both allowing a more positive clinical outcome [133]. Extensions into colonoscopy and other applications are expected in the next 5 years [134].

References

1. Castro C, Briggs W, Paschos GK, FitzGerald GA, Griffin JL. A metabolomic study of adipose tissue in mice with a disruption of the circadian system. Mol Biosyst. 2015;11(7):1897–906.
2. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, Holmes E, Nicholson JK. Global metabolic profiling procedures for urine using UPLC–MS. Nat Protoc. 2010; 5(6):1005–18.

3. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Arch Toxicol. 2011;85(1):5–17.
4. Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: a global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol. 2008;48:653–83.
5. Favé G, Beckmann ME, Draper JH, Mathers JC. Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? Genes Nutr. 2009;4(2):135–41.
6. Johnson CH, Patterson AD, Idle JR, Gonzalez FJ. Xenobiotic metabolomics: major impact on the metabolome. Annu Rev Pharmacol Toxicol. 2012;52:37–56.
7. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT. The human urine metabolome. PLoS One. 2013;8(9):e73076.
8. Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, Saude EJ, Lix B, Adamko DJ, Shah S, Greiner R. Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Anal Chem. 2007;79(18):6995–7004.
9. Giskeødegård GF, Davies SK, Revell VL, Keun H, Skene DJ. Diurnal rhythms in the human urine metabolome during sleep and total sleep deprivation. Sci Rep. 2015;5:14843.
10. Fernández-Peralbo MA, de Castro ML. Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis. Trends Anal Chem. 2012;41:75–85.
11. Adamko D, Rowe BH, Marrie T, Sykes BD. Variation of metabolites in normal human urine. Metabolomics. 2007;3(4):439–51.
12. Chan EC, Pasikanti KK, Nicholson JK. Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. Nat Protoc. 2011;6(10):1483–99.
13. Gika HG, Theodoridis GA, Wilson ID. Liquid chromatography and ultra-performance liquid chromatography–mass spectrometry fingerprinting of human urine: sample stability under different handling and storage conditions for metabonomics studies. J Chromatogr A. 2008;1189(1):314–22.
14. Álvarez-Sánchez B, Priego-Capote F, de Castro ML. Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. Trends Anal Chem. 2010;29(2):111–9.
15. Pasikanti KK, Ho PC, Chan EC. Development and validation of a gas chromatography/mass spectrometry metabonomic platform for the global profiling of urinary metabolites. Rapid Commun Mass Spectrom. 2008;22(19):2984–92.
16. Sykes BD. Urine stability for metabolomic studies: effects of preparation and storage. Metabolomics. 2007;3(1):19–27.
17. Gika HG, Theodoridis GA, Wingate JE, Wilson ID. Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. J Proteome Res. 2007;6(8):3291–303.
18. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, Pujos-Guillot E, Verheij E, Wishart D, Wopereis S. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics. 2009;5(4):435–58.
19. Emwas AH, Luchinat C, Turano P, Tenori L, Roy R, Salek RM, Ryan D, Merzaban JS, Kaddurah-Daouk R, Zeri AC, Gowda GN. Standardizing the experimental conditions for using urine in NMR-based metabolic studies with a particular focus on diagnostic studies: a review. Metabolomics. 2015;11(4):872–94.
20. Snyder ML, Lichstein HC. Sodium azide as an inhibiting substance for gram-negative bacteria. J Infect Dis. 1940;67(2):113–5.
21. Gika HG, Theodoridis GA, Plumb RS, Wilson ID. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. J Pharm Biomed Anal. 2014;87:12–25.
22. Theodoridis GA, Gika HG, Want EJ, Wilson ID. Liquid chromatography–mass spectrometry based global metabolite profiling: a review. Anal Chim Acta. 2012;711:7–16.
23. Chen Y, Xu J, Zhang R, Abliz Z. Methods used to increase the comprehensive coverage of urinary and plasma metabolomes by MS. Bioanalysis. 2016;8(9):981–97.

24. Waybright TJ, Van QN, Muschik GM, Conrads TP, Veenstra TD, Issaq HJ. LC-MS in meta-bonomics: optimization of experimental conditions for the analysis of metabolites in human urine. J Liq Chromatogr Relat Technol. 2006;29(17):2475–97.
25. Álvarez-Sánchez B, Priego-Capote F, de Castro ML. Metabolomics analysis II. Preparation of biological samples prior to detection. Trends Anal Chem. 2010;29(2):120–7.
26. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007;26(1):51–78.
27. Issaq HJ, Nativ O, Waybright T, Luke B, Veenstra TD, Issaq EJ, Kravstov A, Mullerad M. Detection of bladder cancer in human urine by metabolomic profiling using high performance liquid chromatography/mass spectrometry. J Urol. 2008;179(6):2422–6.
28. Chetwynd AJ, Abdul-Sada A, Hill EM. Solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionization mass spectrometry for improved global urine metabolomics. Anal Chem. 2015;87(2):1158–65.
29. Michopoulos F, Gika H, Palachanis D, Theodoridis G, Wilson ID. Solid phase extraction methodology for UPLC-MS based metabolic profiling of urine samples. Electrophoresis. 2015;36(18):2170–8.
30. Tulipani S, Mora-Cubillos X, Jáuregui O, Llorach R, García-Fuentes E, Tinahones FJ, Andres-Lacueva C. New and vintage solutions to enhance the plasma metabolome coverage by LC-ESI-MS untargeted metabolomics: the not-so-simple process of method performance evaluation. Anal Chem. 2015;87(5):2639–47.
31. Dunn WB, Broadhurst D, Ellis DI, Brown M, Halsall A, O'Hagan S, Spasic I, Tseng A, Kell DB. A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. Int J Epidemioly. 2008;37 Suppl 1:i23–30.
32. Woo HM, Kim KM, Choi MH, Jung BH, Lee J, Kong G, Nam SJ, Kim S, Bai SW, Chung BC. Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. Clin Chim Acta. 2009;400(1):63–9.
33. Silva C, Cavaco C, Perestrelo R, Pereira J, Câmara JS. Microextraction by packed sorbent (meps) and solid-phase microextraction (spme) as sample preparation procedures for the metabolomic profiling of urine. Metabolites. 2014;4(1):71–97.
34. Silva CL, Passos M, Câmara JS. Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers—a powerful strategy for breast cancer diagnosis. Talanta. 2012;89:360–8.
35. Bojko B, Reyes-Garcés N, Bessonneau V, Goryński K, Mousavi F, Silva EA, Pawliszyn J. Solid-phase microextraction in metabolomics. TrAC Trends in Anal Chem. 2014;61:168–80.
36. Michell AW, Mosedale D, Grainger DJ, Barker RA. Metabolomic analysis of urine and serum in Parkinson's disease. Metabolomics. 2008;4(3):191–201.
37. Kind T, Tolstikov V, Fiehn O, Weiss RH. A comprehensive urinary metabolomic approach for identifying kidney cancer. Anal Biochem. 2007;363(2):185–95.
38. Chen Y, Shen G, Zhang R, He J, Zhang Y, Xu J, Yang W, Chen X, Song Y, Abliz Z. Combination of injection volume calibration by creatinine and ms signals' normalization to overcome urine variability in LC-MS-based metabolomics studies. Anal Chem. 2013;85(16):7659–65.
39. Veselkov KA, Vingara LK, Masson P, Robinette SL, Want E, Li JV, Barton RH, Boursier-Neyret C, Walther B, Ebbels TM, Pelczer I. Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. Anal Chem. 2011;83(15):5864–72.
40. Chetwynd AJ, Abdul-Sada A, Holt SG, Hill EM. Use of a pre-analysis osmolality normalisation method to correct for variable urine concentrations and for improved metabolomic analyses. J Chromatogr A. 2016;1431:103–10.
41. Edmands WM, Ferrari P, Scalbert A. Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. Anal Chem. 2014;86(21):10925–31.
42. Wedge DC, Allwood JW, Dunn W, Vaughan AA, Simpson K, Brown M, Priest L, Blackhall FH, Whetton AD, Dive C, Goodacre R. Is serum or plasma more appropriate for intersubject

comparisons in metabolomic studies? An assessment in patients with small-cell lung cancer. Anal Chem. 2011;83(17):6689–97.

43. Dettmer K, Almstetter MF, Appel IJ, Nürnberger N, Schlamberger G, Gronwald W, Meyer HH, Oefner PJ. Comparison of serum versus plasma collection in gas chromatography–Mass spectrometry-based metabolomics. Electrophoresis. 2010;31(14):2365–73.

44. Yu Z, Kastenmüller G, He Y, Belcredi P, Möller G, Prehn C, Mendes J, Wahl S, Roemisch-Margl W, Ceglarek U, Polonikov A. Differences between human plasma and serum metabolite profiles. PLoS One. 2011;6(7):e21230.

45. Tenori L, Oakman C, Morris PG, Gralka E, Turner N, Cappadona S, Fornier M, Hudis C, Norton L, Luchinat C, Di Leo A. Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. Mol Oncol. 2015;9(1):128–39.

46. Drogan D, Dunn WB, Lin W, Buijsse B, Schulze MB, Langenberg C, Brown M, Floegel A, Dietrich S, Rolandsson O, Wedge DC. Untargeted metabolic profiling identifies altered serum metabolites of type 2 diabetes mellitus in a prospective, nested case control study. Clin Chem. 2015;61(3):487–97.

47. Zhang W, Sun G, Likhodii S, Liu M, Aref-Eshghi E, Harper PE, Martin G, Furey A, Green R, Randell E, Rahman P. Metabolomic analysis of human plasma reveals that arginine is depleted in knee osteoarthritis patients. Osteoarthritis Cartilage. 2016;24(5):827–34.

48. Cheng ML, Wang CH, Shiao MS, Liu MH, Huang YY, Huang CY, Mao CT, Lin JF, Ho HY, Yang NI. Metabolic disturbances identified in plasma are associated with outcomes in patients with heart failure: diagnostic and prognostic value of metabolomics. J Am College Cardiol. 2015;65(15):1509–20.

49. Vitamin D Blood Spot Assay, Pathology Department, City Hospital, Birmingham. [Cited 8 Aug 2016]. Available from: http://www.cityassays.org.uk/Vitamin%20D%20Blood%20Spot.html.

50. Hirayama A, Sugimoto M, Suzuki A, Hatakeyama Y, Enomoto A, Harada S, Soga T, Tomita M, Takebayashi T. Effects of processing and storage conditions on charged metabolomic profiles in blood. Electrophoresis. 2015;36(18):2148–55.

51. Yin P, Lehmann R, Xu G. Effects of pre-analytical processes on blood samples used in metabolomics studies. Anal Bioanal Chem. 2015;407(17):4879–92.

52. BD Vacutainer Venous Blood Collection, Tube Guide. [Cited 11 Aug 2016]. Available from: https://www.bd.com/vacutainer/pdfs/plus_plastic_tubes_wallchart_tubeguide_VS5229.pdf.

53. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011;6(7):1060–83.

54. Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kochhar S, Guy PA. Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. Anal Chem. 2009;81(9):3285–96.

55. Contrepois K, Jiang L, Snyder M. optimized analytical procedures for the untargeted metabolomic profiling of human urine and plasma by combining hydrophilic interaction (hilic) and reverse-phase liquid chromatography (RPLC)–Mass spectrometry. Mol Cell Proteomics. 2015;14(6):1684–95.

56. Reis A, Rudnitskaya A, Blackburn GJ, Fauzi NM, Pitt AR, Spickett CM. A comparison of five lipid extraction solvent systems for lipidomic studies of human LDL. J Lipid Res. 2013;54(7):1812–24.

57. Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, Marklund SL, Moritz T. Extraction and GC/MS analysis of the human blood plasma metabolome. Anal Chem. 2005;77(24):8086–94.

58. Patterson RE, Ducrocq AJ, McDougall DJ, Garrett TJ, Yost RA. Comparison of blood plasma sample preparation methods for combined LC–MS lipidomics and metabolomics. J Chromatogr B. 2015;1002:260–6.

59. Boernsen KO, Gatzek S, Imbert G. Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. Anal Chem. 2005;77(22):7255–64.

60. Want EJ, Smith CA, Qin C, Van Horne KC, Siuzdak G. Phospholipid capture combined with non-linear chromatographic correction for improved serum metabolite profiling. Metabolomics. 2006;2(3):145–54.
61. Michopoulos F, Lai L, Gika H, Theodoridis G, Wilson I. UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. J Proteome Res. 2009;8(4):2114–21.
62. David A, Abdul-Sada A, Lange A, Tyler CR, Hill EM. A new approach for plasma (xeno) metabolomics based on solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionisation mass spectrometry. J Chromatogr A. 2014;1365:72–85.
63. Álvarez-Sánchez B, Priego-Capote F, de Castro ML. Study of sample preparation for metabolomic profiling of human saliva by liquid chromatography–time of flight/mass spectrometry. J Chromatogr A. 2012;1248:178–81.
64. Zhang A, Sun H, Wang X. Saliva metabolomics opens door to biomarker discovery, disease diagnosis, and treatment. Appl Biochem Biotechnol. 2012;168(6):1718–27.
65. Takeda I, Stretch C, Barnaby P, Bhatnager K, Rankin K, Fu H, Weljie A, Jha N, Slupsky C. Understanding the human salivary metabolome. NMR Biomed. 2009;22(6):577–84.
66. Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. Metabolomics. 2010;6(1):78–95.
67. Dame ZT, Aziat F, Mandal R, Krishnamurthy R, Bouatra S, Borzouie S, Guo AC, Sajed T, Deng L, Lin H, Liu P. The human saliva metabolome. Metabolomics. 2015;11(6):1864–83.
68. Wang Q, Gao P, Wang X, Duan Y. The early diagnosis and monitoring of squamous cell carcinoma via saliva metabolomics. Sci Rep. 2014;4:6802.
69. Santone C, Dinallo V, Paci M, D'Ottavio S, Barbato G, Bernardini S. Saliva metabolomics by NMR for the evaluation of sport performance. J Pharm Biomed Anal. 2014;88:441–6.
70. Walsh MC, Brennan L, Malthouse JP, Roche HM, Gibney MJ. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. Am J Clin Nutr. 2006;84(3):531–9.
71. Bessonneau V, Bojko B, Pawliszyn J. Analysis of human saliva metabolome by direct immersion solid-phase microextraction LC and benchtop orbitrap MS. Bioanalysis. 2013;5(7):783–92.
72. Wishart DS, Lewis MJ, Morrissey JA, Flegel MD, Jeroncic K, Xiong Y, Cheng D, Eisner R, Gautam B, Tzur D, Sawhney S. The human cerebrospinal fluid metabolome. J Chromatogr B. 2008;871(2):164–73.
73. Zhang A, Sun H, Wang P, Han Y, Wang X. Recent and potential developments of biofluid analyses in metabolomics. J Proteomics. 2012;75(4):1079–88.
74. Maillet S, Vion-Dury J, Confort-Gouny S, Nicoli F, Lutz NW, Viout P, Cozzone PJ. Experimental protocol for clinical analysis of cerebrospinal fluid by high resolution proton magnetic resonance spectroscopy. Brain Res Protoc. 1998;3(2):123–34.
75. Mandal R, Guo AC, Chaudhary KK, Liu P, Yallou FS, Dong E, Aziat F, Wishart DS. Multi-platform characterization of the human cerebrospinal fluid metabolome: a comprehensive and quantitative update. Genome Med. 2012;4(4):1.
76. Mena-Bravo A, de Castro ML. Sweat: a sample with limited present applications and promising future in metabolomics. J Pharm Biomed Anal. 2014;90:139–47.
77. Calderón-Santiago M, Priego-Capote F, Jurado-Gámez B, de Castro ML. Optimization study for metabolomics analysis of human sweat by liquid chromatography–tandem mass spectrometry in high resolution mode. J Chromatogr A. 2014;1333:70–8.
78. Kutyshenko VP, Molchanov M, Beskaravayny P, Uversky VN, Timchenko MA. Analyzing and mapping sweat metabolomics by high-resolution NMR spectroscopy. Plos One. 2011;6(12):e28824.
79. Smilowitz JT, O'Sullivan A, Barile D, German JB, Lönnerdal B, Slupsky CM. The human milk metabolome reveals diverse oligosaccharide profiles. J Nutr. 2013;143(11):1709–18.
80. Praticò G, Capuani G, Tomassini A, Baldassarre ME, Delfini M, Miccheli A. Exploring human breast milk composition by NMR-based metabolomics. Nat Prod Res. 2014;28(2):95–101.

81. Villaseñor A, Garcia-Perez I, Garcia A, Posma JM, Fernández-López M, Nicholas AJ, Modi N, Holmes E, Barbas C. Breast milk metabolome characterization in a single-phase extraction, multiplatform analytical approach. Anal Chem. 2014;86(16):8245–52.

82. Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z, Lindon JC. Metabolic phenotyping in clinical and surgical environments. Nature. 2012;491(7424):384–92.

83. Čuperlović-Culf M, Barnett DA, Culf AS, Chute I. Cell culture metabolomics: applications and future directions. Drug Discov Today. 2010;15(15):610–21.

84. Dunn WB, Lin W, Broadhurst D, Begley P, Brown M, Zelena E, Vaughan AA, Halsall A, Harding N, Knowles JD, Francis-McIntyre S. Molecular phenotyping of a UK population: defining the human serum metabolome. Metabolomics. 2015;11(1):9–26.

85. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG. A community-driven global reconstruction of human metabolism. Nat Biotechnol. 2013;31(5):419–25.

86. Halama A. Metabolomics in cell culture—a strategy to study crucial metabolic pathways in cancer development and the response to treatment. Arch Biochem Biophys. 2014;564:100–9.

87. Zhang A, Sun H, Xu H, Qiu S, Wang X. Cell metabolomics. Omics J Integrative Biol. 2013;17(10):495–501.

88. Kim DH, Achcar F, Breitling R, Burgess KE, Barrett MP. LC–MS-based absolute metabolite quantification: application to metabolic flux measurement in trypanosomes. Metabolomics. 2015;11(6):1721–32.

89. Mackay GM, Zheng L, van den Broek NJF, Gottlieb E. Analysis of cell metabolism using LC-MS and isotope tracers. In: Metallo CM, editor. Methods in enzymology. Metabolic analysis using stable isotopes. 1st ed. Academic Press; Waltham, USA 2015.

90. Vander Heiden MG. Targeting cancer metabolism: a therapeutic window opens. Nat Rev Drug Discovery. 2011;10(9):671–84.

91. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324(5930):1029–33.

92. Schaeffer WI. Usage of vertebrate, invertebrate and plant cell, tissue and organ culture terminology. In Vitro. 1984;20(1):19–24.

93. León Z, García-Cañaveras JC, Donato MT, Lahoz A. Mammalian cell metabolomics: experimental design and sample preparation. Electrophoresis. 2013;34(19):2762–75.

94. Hounoum BM, Blasco H, Nadal-Desbarats L, Diémé B, Montigny F, Andres CR, Emond P, Mavel S. Analytical methodology for metabolomics study of adherent mammalian cells using NMR, GC-MS and LC-HRMS. Anal Bioanal Chem. 2015;407(29):8861–72.

95. Vuckovic D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. Anal Bioanal Chem. 2012;403(6):1523–48.

96. Paglia G, Hrafnsdóttir S, Magnúsdóttir M, Fleming RM, Thorlacius S, Palsson BØ, Thiele I. Monitoring metabolites consumption and secretion in cultured cells using ultra-performance liquid chromatography quadrupole–time of flight mass spectrometry (UPLC–Q–ToF-MS). Anal Bioanal Chem. 2012;402(3):1183–98.

97. Mercier P, Lewis MJ, Chang D, Baker D, Wishart DS. Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. J Biomol NMR. 2011;49(3-4):307–23.

98. Dietmair S, Timmins NE, Gray PP, Nielsen LK, Krömer JO. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. Anal Biochem. 2010;404(2):155–64.

99. Bordag N, Janakiraman V, Nachtigall J, Maldonado SG, Bethan B, Laine JP, Fux E. Fast filtration of bacterial or mammalian suspension cell cultures for optimal metabolomics results. PLoS One. 2016;11(7):e0159389.

100. Hounoum BM, Blasco H, Emond P, Mavel S. Liquid chromatography–high-resolution mass spectrometry-based cell metabolomics: experimental design, recommendations, and applications. Trends Anal Chem. 2016;75:118–28.

101. Sellick CA, Hansen R, Stephens GM, Goodacre R, Dickson AJ. Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. Nat Protoc. 2011;6(8):1241–9.
102. Dietmair S, Hodson MP, Quek LE, Timmins NE, Chrysanthopoulos P, Jacob SS, Gray P, Nielsen LK. Metabolite profiling of CHO cells with different growth characteristics. Biotechnol Bioeng. 2012;109(6):1404–14.
103. Han W, Li L. Matrix effect on chemical isotope labeling and its implication in metabolomic sample preparation for quantitative metabolomics. Metabolomics. 2015;11(6):1733–42.
104. Bi H, Krausz KW, Manna SK, Li F, Johnson CH, Gonzalez FJ. Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI-MS-based metabolomic analysis of adherent mammalian cancer cells. Anal Bioanal Chem. 2013;405(15):5279–89.
105. Lorenz MA, Burant CF, Kennedy RT. Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. Anal Chem. 2011;83(9):3406–14.
106. Dettmer K, Nürnberger N, Kaspar H, Gruber MA, Almstetter MF, Oefner PJ. Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. Anal Bioanal Chem. 2011;399(3):1127–39.
107. Teng Q, Huang W, Collette TW, Ekman DR, Tan C. A direct cell quenching method for cell-culture based metabolomics. Metabolomics. 2009;5(2):199–208.
108. Panopoulos AD, Yanes O, Ruiz S, Kida YS, Diep D, Tautenhahn R, Herrerías A, Batchelder EM, Plongthongkum N, Lutz M, Berggren WT. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. Cell Res. 2012;22(1):168–77.
109. Munger J, Bajad SU, Coller HA, Shenk T, Rabinowitz JD. Dynamics of the cellular metabolome during human cytomegalovirus infection. PLoS Pathog. 2006;2(12):e132.
110. Silva LP, Lorenzi PL, Purwaha P, Yong V, Hawke DH, Weinstein JN. Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines. Anal Chem. 2013;85(20):9536–42.
111. Fazelzadeh P, Hangelbroek RW, Tieland M, de Groot LC, Verdijk LB, van Loon LJ, Smilde AK, Alves RD, Vervoort J, Müller M, van Duynhoven JP. The muscle metabolome differs between healthy and frail older adults. J Proteome Res. 2016;15(2):499–509.
112. Mayr M, Yusuf S, Weir G, Chung YL, Mayr U, Yin X, Ladroue C, Madhu B, Roberts N, De Souza A, Fredericks S. Combined metabolomic and proteomic analysis of human atrial fibrillation. J Am College Cardiol. 2008;51(5):585–94.
113. Schönfels W, Patsenker E, Fahrner R, Itzel T, Hinrichsen H, Brosch M, Erhart W, Gruodyte A, Vollnberg B, Richter K, Landrock A. Metabolomic tissue signature in human non-alcoholic fatty liver disease identifies protective candidate metabolites. Liver Int. 2015;35(1):207–14.
114. Rocha CM, Barros AS, Goodfellow BJ, Carreira IM, Gomes A, Sousa V, Bernardo J, Carvalho L, Gil AM, Duarte IF. NMR metabolomics of human lung tumours reveals distinct metabolic signatures for adenocarcinoma and squamous cell carcinoma. Carcinogenesis. 2015;36(1):68–75.
115. Dunn WB, Brown M, Worton SA, Davies K, Jones RL, Kell DB, Heazell AE. The metabolome of human placental tissue: investigation of first trimester tissue and changes related to preeclampsia in late pregnancy. Metabolomics. 2012;8(4):579–97.
116. Anwar MA, Vorkas PA, Li JV, Shalhoub J, Want EJ, Davies AH, Holmes E. Optimization of metabolite extraction of human vein tissue for ultra performance liquid chromatography-mass spectrometry and nuclear magnetic resonance-based untargeted metabolic profiling. Analyst. 2015;140(22):7586–97.
117. Randhawa M, Sangar V, Tucker-Samaras S, Southall M. Metabolic signature of sun exposed skin suggests catabolic pathway overweighs anabolic pathway. PLoS One. 2014;9(3):e90367.
118. Allwood JW, Winder CL, Dunn WB, Goodacre R. Considerations in sample preparation, collection, and extraction approaches applied in microbial, plant, and mammalian metabolic profiling. In: Lutz NW, Sweedler JV, Wevers RA, editors. Methodologies for metabolomics: experimental strategies and techniques. 1st ed. Cambridge: Cambridge University Press; 2013.

119. Want EJ, Masson P, Michopoulos F, Wilson ID, Theodoridis G, Plumb RS, Shockcor J, Loftus N, Holmes E, Nicholson JK. Global metabolic profiling of animal and human tissues via UPLC-MS. Nat Protoc. 2013;8(1):17–32.
120. Brown DG, Rao S, Weir TL, O'Malia J, Bazan M, Brown RJ, Ryan EP. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. Cancer Metab. 2016;4(1):1.
121. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics. 2012;8(1):133–42.
122. Chouchani ET, Pell VR, Gaude E, Aksentijević D, Sundier SY, Robb EL, Logan A, Nadtochiy SM, Ord EN, Smith AC, Eyassu F. Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. Nature. 2014;515(7527):431–5.
123. Vorkas PA, Isaac G, Anwar MA, Davies AH, Want EJ, Nicholson JK, Holmes E. Untargeted UPLC-MS profiling pipeline to expand tissue metabolome coverage: application to cardiovascular disease. Anal Chem. 2015;87(8):4184–93.
124. Gehmlich K, Dodd MS, Allwood JW, Kelly M, Bellahcene M, Lad HV, Stockenhuber A, Hooper C, Ashrafian H, Redwood CS, Carrier L. Changes in the cardiac metabolome caused by perhexiline treatment in a mouse model of hypertrophic cardiomyopathy. Mol Bio Syst. 2015;11(2):564–73.
125. Chen S, Hoene M, Li J, Li Y, Zhao X, Häring HU, Schleicher ED, Weigert C, Xu G, Lehmann R. Simultaneous extraction of metabolome and lipidome with methyl tert-butyl ether from a single small tissue sample for ultra-high performance liquid chromatography/mass spectrometry. J Chromatogr A. 2013;1298:9–16.
126. Folch J, Lees M, Sloane-Stanley GH. A simple method for the isolation and purification of total lipids from animal tissues. J Biol Chem. 1957;226(1):497–509.
127. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res. 2008;49(5):1137–46.
128. Wu H, Southam AD, Hines A, Viant MR. High-throughput tissue extraction protocol for NMR-and MS-based metabolomics. Anal Biochem. 2008;372(2):204–12.
129. Rao S, Walters KB, Wilson L, Chen B, Bolisetty S, Graves D, Barnes S, Agarwal A, Kabarowski JH. Early lipid changes in acute kidney injury using SWATH lipidomics coupled with MALDI tissue imaging. Am J Physiol Renal Physiol. 2016;310(10):F1136–47.
130. Jarmusch AK, Pirro V, Baird Z, Hattab EM, Cohen-Gadol AA, Cooks RG. Lipid and metabolite profiles of human brain tumors by desorption electrospray ionization-MS. Proc Natl Acad Sci U S A. 2016;113(6):1486–91.
131. Park JW, Jeong H, Kang B, Kim SJ, Park SY, Kang S, Kim HK, Choi JS, Hwang D, Lee TG. Multi-dimensional TOF-SIMS analysis for effective profiling of disease-related ions from the tissue surface. Sci Rep. 2015;5:5.
132. Ly A, Buck A, Balluff B, Sun N, Gorzolka K, Feuchtinger A, Janssen KP, Kuppen PJ, van de Velde CJ, Weirich G, Erlmeier F. High-mass-resolution MALDI mass spectrometry imaging of metabolites from formalin-fixed paraffin-embedded tissue. Nat Protoc. 2016;11(8):1428–43.
133. Balog J, Sasi-Szabó L, Kinross J, Lewis MR, Muirhead LJ, Veselkov K, Mirnezami R, Dezső B, Damjanovich L, Darzi A, Nicholson JK. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. Sci Translational Med. 2013;5(194):194ra93.
134. Kinross JM, Muirhead L, Alexander J, Balog J, Guallar-Hoya C, Speller A, Golff O, Goldin R, Darzi A, Nicholson J, Takats Z. iKnife: rapid evaporative ionization mass spectrometry (REIMS) enables real-time chemical analysis of the mucosal lipidome for diagnostic and prognostic use in colorectal cancer. Cancer Res. 2016;76(14 Suppl):3977.

Chapter 3
Nuclear Magnetic Resonance Strategies for Metabolic Analysis

Clement Heude, Jay Nath, John Bosco Carrigan, and Christian Ludwig

Abstract NMR spectroscopy is a powerful tool for metabolomic studies, offering highly reproducible and quantitative analyses. This burgeoning field of NMR metabolomics has been greatly aided by the development of modern spectrometers and software, allowing high-throughput analysis with near real-time feedback. Whilst one-dimensional proton (1D-^{1}H) NMR analysis is best described and remains most widely used, a plethora of alternative NMR techniques are now available that offer additional chemical and structural information and resolve many of the limitations of conventional 1D-^{1}H NMR such as spectral overlay. In this book chapter, we review the principal concepts of practical NMR spectroscopy, from common sample preparation protocols to the benefits and theoretical concepts underpinning the commonly used pulse sequences. Finally, as a case study to highlight the utility of NMR as a method for metabolomic investigation, we have detailed how NMR has been used to gain valuable insight into the metabolism occurring in kidneys prior to transplantation and the potential implications of this.

Keywords NMR • Metabolism • Metabolomics • Tracer • Kidney • Transplantation • HMP • Hypothermic • Perfusion

C. Heude
NERC Metabolomics Facility, School of Biosciences, University of Birmingham, Birmingham, UK

Institute of Metabolism and Systems Research (IMSR), University of Birmingham, Birmingham, UK

J. Nath • J.B. Carrigan • C. Ludwig (✉)
Institute of Metabolism and Systems Research (IMSR), College of Medical and Dental Sciences, University of Birmingham, Birmingham, B152TT, UK

Abbreviations

AsLS	Asymmetric least square smoothing
AST	Aspartate transaminase
BML-NMR	Birmingham Metabolite Library Nuclear Magnetic Resonance
BMRB	BioMagResBank
CluPA	Cluster-based peak alignment
COW	Correlation optimized warping
CPMG	Carr-Purcell-Meiboom-Gill
DGF	Delayed graft function
DTW	Dynamic time warping
EDTA	Ethylene diamine tetra acetate
FFT	Fast Fourier transform
FID	Free induction decay
FW	Iterative fuzzy warping
GST	Glutathione-S-transferase
HR-MAS	High-resolution magic angle spinning
HMBC	Heteronuclear multiple-bond correlation
HMDB	Human Metabolome Database
HMP	Hypothermic machine perfusion
HSQC	Heteronuclear single quantum coherence
ICA	Independent component analysis
Icoshift	Interval correlated shifting
IGF	Immediate graft function
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOH	Potassium hydroxide
LC-MS	Liquid chromatography-mass spectrometry
LDH	Lactate dehydrogenase
LOWESS	Locally weighted scatterplot smoothing
MMCD	Madison Metabolomics Consortium Database
MTBE	Methyl tert-butyl ether
NMR	Nuclear magnetic resonance
NUS	Non-uniform sampling
OPLS-DA	Orthogonal partial least square – discriminant analysis
PARS	Peak alignment using reduced set mapping
PCA	Principal component analysis
PLS	Partial least square
PLS-DA	Partial least square – discriminant analysis
PQN	Probabilistic quotient normalization
SCS	Static cold storage
SPE	Solid phase extraction
TMAO	Trimethylamine-N-oxide
TMSP	Trimethylsilylpropanoate
TOCSY	Total correlation spectroscopy
TSA	Total spectral area
Vast	Variable stability

3.1 Introduction

Metabolomics, like other 'omics' studies, has been the subject of great scientific interest in recent years. NMR spectroscopy is a valid method for identification of constituent metabolites within a bio-sample, with proponents highlighting the highly sensitive, reproducible nature of this technique. Metabolic pathways are reflective of the genome and proteome with up to 10,000-fold increase in metabolite concentration resulting from single amino acid change in a protein or base change in a gene [1]. Furthermore, metabolic changes are apparent within minutes of a biological event and therefore provide an almost 'real-time' feedback [2]. Whilst NMR metabolomic experiments have been described since the early 1980s [3], the development of high-throughput NMR spectrometers with the necessary software for analysis has been relatively recent and has greatly improved its scientific utility [2]. Indeed, the development of modern NMR spectrometers combined with the setup of standardized sample preparation and acquisition protocols [4] has enabled the analysis of thousands of molecules with high sensitivity and reproducibility within a few minutes [5].

NMR metabolomic techniques have been used to gain clinical insight into a multitude of pathological conditions, ranging from cancer [6] to neurodegeneration [7], and a full range of the clinical applications of NMR is outside the remit of this chapter. However, in order to highlight some of the applications of NMR spectroscopy, we have used renal transplantation as a case study.

3.2 NMR of Biological Samples

Whilst variation in spectrometer configuration does exist between groups, efforts have been made to standardize NMR techniques to optimize reproducibility and enable valid comparison of results. Multinational projects such as the International Phenome Center Network (IPCN) have served to further highlight this, with the principal instrument manufacturer suggesting a standardized setup for this purpose. This configuration is a 600 MHz spectrometer, equipped with the latest generation of console, a 5 mm room temperature probe and a temperature-controlled automated sample-changing robot. For the purpose of this review, we will mainly focus on this standard spectrometer configuration, and whilst this is sufficient for the vast majority of NMR metabolic studies, there are situations (e.g. limited sample size) where this is inappropriate and alternative configurations should be used (e.g. cryogenic or fine bore probes).

Fortunately, sample preparation of biological samples prior to NMR analysis is usually relatively straightforward with example protocols for commonly used samples discussed below. Common to all sample preparation protocols is the addition of a phosphate buffer solution and of an internal chemical shift standard such as TMSP or DSS. Deuterated water (D_2O) is added to all aqueous NMR samples to a final concentration of 5 %. NMR systems use the D_2O signal as a lock frequency to compensate for long-term magnetic field drifts. Once samples are transferred into NMR

tubes, the caps should be sealed with a polyoxometalate (POM) ball to avoid solvent hydrogenation and evaporation. If there is any delay between sample preparation and data acquisition, it is a good practice to store samples at −80 °C during any interim period.

3.2.1 Biofluids: Plasma/Serum (Blood-Derived Samples)

In contrast to serum, plasma is usually collected into vials containing an anticoagulant to prevent clot formation and maintain samples in a fluid state. The choice of anticoagulant (typically either lithium heparin or *e*thylene *d*iamine *t*etra *a*cetate (EDTA)) [8] depends on the NMR analysis purpose. For lipoprotein analysis, EDTA is advantageous as lithium heparin and lipoprotein signals overlap making the interpretation complex. Similarly, for the analysis of small metabolites, lithium heparin is preferred as the Mg^{2+} and Ca^{2+} complexes within EDTA results in signal overlay.

A widely endorsed sample preparation protocol [4] for blood-derived (either plasma or serum) samples using 5 mm NMR tubes consists of sample centrifugation (typically 15 min) with subsequent collection of the supernatant. 350 μL of sample (plasma or serum) is mixed with 350 μL of a pH 7.4 75 mmol L^{-1} NaH_2PO_4 phosphate NMR buffer solution in an Eppendorf tube with a solvent composition of H_2O:D_2O of 90:10 as well as containing the chemical shift standard DSS at a concentration of 0.6 mmol L^{-1}.

3.2.2 Biofluids: Urine

As urine samples are prone to bacterial contamination, a small volume of aqueous NaN_3 solution (0.05 % m/v) is commonly added to each sample before storage (−80 °C) and sample preparation. 600 μL of urine is centrifuged at high speed, and 540 μL of supernatant is vigorously mixed with 60 μL of a 1.5 mol L^{-1} pH 7.4 KH_2PO_4 phosphate buffer solution containing sodium *t*ri*m*ethyl*s*ilyl*p*ropanoate (TMSP). As for the plasma samples, the tubes are sealed and placed in an automatic sample changer prior to NMR analysis.

3.2.3 Tissue/Cells: Extraction Procedures

Analyses of tissues or cells for NMR metabolomics commonly require extraction procedures to disrupt tissue and cellular integrity and ensure consistent metabolite distribution within samples. Although the utility of tissues or cells for NMR metabolomics analyses is self-evident, there is no standardized extraction procedure for this and is likely to reflect the different characteristics of tissues studied. However, there are some steps common to all extraction methods.

The first is the cessation of enzyme activity and therefore ongoing metabolic processes. This is commonly achieved by quenching the samples [9] in liquid N_2 or in chilled methanol ($-40\ °C$) with subsequent storage at $-80\ °C$ until extraction.

Homogenization is necessary to release intracellular metabolite stores. The two principal methods for this are fast homogenization using an electric tissue homogenizer [10] or manual homogenization using a liquid nitrogen-cooled mortar and pestle [11]. However, the second method is cumbersome, labour intensive and time-consuming and requires great caution to avoid partially thawing of the samples during grinding.

Because the large variety of metabolites contained within cells and tissues, and the wide range of polarity and physicochemical features of these, there is no ideal method to extract the entire metabolome. However, as NMR-based metabolic studies mostly focus on the polar compounds, the optimal method should be able to extract the highest amount of polar metabolites from the sample.

Furthermore, the method must be robust, reliable and, depending on the scientific purpose, sometimes be able to remove lipids very efficiently. Considering the high number of extraction methods, only the main procedures used in metabolomics will be described in this section.

One widely used extraction method is the perchloric acid method [12]. This procedure allows proteins to precipitate and extracts hydrophilic metabolites. The samples are centrifuged after the addition of the acid to remove proteins from the sample, and the supernatant is collected after being neutralized with potassium hydroxide (KOH). Whilst this method is very effective for extracting amines [13], acid treatment may damage the chemical structure of other metabolites of interest. Hence, monophasic methods, such as the acetonitrile/water [14], methanol/water or methanol/acetonitrile/water [15], and biphasic procedures such as methanol/chloroform/water [16] are widely used. An alternative biphasic extraction method has been developed using *m*ethyl *t*ert-*b*utyl *e*ther (MTBE) [17] instead of chloroform to reduce solvent toxicity. For all mono-/biphasic methods, the polar phase is recovered after centrifugation and dried using a vacuum concentrator. Once dried, samples are usually resuspended in phosphate buffer and centrifuged before the supernatant is placed in an NMR tube. Although biphasic methods are more labour intensive, the results are superior, compared to monophasic techniques, in terms of yield and reproducibility [18] and provide hydrophilic and hydrophobic metabolite separation.

In addition, although mainly used for *l*iquid *c*hromatography-*m*ass *s*pectrometry (LC-MS) analysis [19, 20], *s*olid *p*hase *e*xtraction (SPE) can also be successfully applied for NMR analysis as a separation technique or in order to concentrate the metabolic content and thus enhancing the sensitivity. Recently, SPE has been incorporated into LC-NMR systems as an interface between LC and NMR, the SPE cartridges trapping the peaks of interest.

3.2.4 Tissue: HR-MAS Samples

Although most NMR-based metabolic tissue studies are performed following extraction, an elegant alternative method for tissue analysis is *h*igh-*r*esolution *m*agic *a*ngle *s*pinning (HR-MAS) spectroscopy. In addition to the avoidance of time-consuming

extraction procedures, HR-MAS is non-destructive enabling sample recovery after analysis, which is often particularly useful in clinical studies where tissue volume is critical and needed for different investigative modalities (e.g. histopathology). Another advantage of HR-MAS is that it allows the absolute quantification of all the different metabolites by working directly on the raw material [21], avoiding inconsistencies inherent to all extraction methods. However, unlike 1D liquid-state NMR, additional pulse sequence elements such as CPMG (*C*arr-*P*urcell-*M*eiboom-*G*ill) are necessary to remove broad signals from macromolecules from the HR-MAS NMR spectrum. Unfortunately, the use of a CPMG element, which is a T_2-filter, leads to a loss of magnetization. Hence, metabolite concentrations derived from CPMG spectra have to be corrected to account for magnetization loss during the CPMG sequence. Various methods have been developed to minimize the problem of magnetization loss [21–23]; however, T_2 relaxation times of different metabolites are tissue dependent; therefore a universal solution addressing all different tissue types used is not available.

For HR-MAS tissue studies, the most common NMR sample preparation method [24, 25] involves acquiring a tissue sample using a 2 mm biopsy punch and placing this (typically between 10 and 15 mg) into a 33 µL-disposable Kel-F insert with 10 µL of deuterium oxide (D_2O) containing TMSP. Contrary to liquid-state NMR, for HR-MAS, homogenization of the magnetic field has to be done manually prior to data acquisition. The solvent has also an important role by increasing the degree of mobility of the sample which facilitates the averaging of dipolar interactions at the magic angle [26].

The use of a biopsy punch allows work with either frozen or fresh tissue. The insert is then sealed with a conical plug and a screw cap before being placed into a standard 4 mm zirconium oxide (ZrO_2) rotor and closed with a cap.

Another way to prepare HR-MAS samples is to introduce the sample directly into a 4 mm ZrO_2 rotor fitted with a 12 or 50 µL cylindrical insert [27] and adding a variable amount of solvent, depending on rotor and actual sample size. Generally, all the different steps of the sample preparation protocol are done on a cooled plate (3–4 °C) to minimize tissue degradation.

Disposable inserts are advantageous during sample preparation as their use avoids long and fastidious washing procedures necessary to avoid cross-contamination between samples. After the sample insertion into the NMR rotor, the rotor is then either immediately inserted into the HR-MAS probe pre-cooled to 3 °C or placed in a cooled HR-MAS sample changer.

3.3 NMR Pulse Sequences

There are several NMR spectroscopic methods used to identify and quantify the metabolome. For quantification, the simplest method is one-dimensional proton NMR spectroscopy (1D-^{1}H NMR). Another commonly used method is two-dimensional proton J-resolved NMR spectroscopy (2D-^{1}H Jres NMR). For the purpose of signal annotation (i.e. peak identification), NMR spectroscopic techniques

such as two-dimensional proton-proton homonuclear *to*tal *c*orrelation *s*pectroscop*y* (2D-^{1}H,^{1}H TOCSY) and two-dimensional proton-carbon *h*eteronuclear *s*ingle *qu*antum *c*oherence spectroscopy (2D-^{1}H,^{13}C HSQC) are commonly used. The 2D-^{1}H,^{13}C HSQC method is also commonly used in tracer-based metabolism analysis.

3.3.1   1D ^{1}H NMR Spectroscopy

Because of the reproducibility and suitability for high-throughput sample analysis, 1D-^{1}H NMR spectroscopy is widely used for NMR-based metabolomic studies [4]. Another advantage of this approach is the linearity of the NMR signal with respect to the metabolite concentration, i.e. 1D-^{1}H NMR is perfectly suited to quantify metabolite concentrations in complex mixtures with a large range of concentrations of the different molecules.

Most samples for NMR analysis are water based (e.g. urine / plasma / CSF), and even dried extracts are usually resuspended in aqueous solutions. As the metabolite concentrations within samples are usually quite small (50 μmol L^{-1} to 50 mmol L^{-1}), compared with proton concentration of water (approximately 110 mol L^{-1}), water suppression is needed to suppress the otherwise domineering signal of water-related protons within the NMR spectrum. Several methods for water suppression are commonly used and briefly discussed.

The simplest approach is to irradiate the sample with a weak (i.e. not to affect other NMR resonances apart from the water signal) radio frequency (RF) pulse for an extended period of time (usually 4–5 s) prior to an excitation pulse with a much higher RF strength to generate observable magnetization. Although this approach, also known as pre-saturation, is the simplest form of water suppression, it can be greatly enhanced by a NOESY pre-saturation pulse sequence [28]. In this, the excitation pulse is replaced by three consecutive pulses where a short (typically 10 ms) mixing time is introduced between the second and the third pulse. The three pulses effectively replace the single excitation pulse in the pre-saturation experiment and result in a more uniform excitation of the NMR sample, which leads to enhanced water suppression. Pulsed *f*ield *g*radients are commonly used to enhance water suppression using this NMR technique.

Other water-suppression techniques, such as SOGGY [29], are potentially superior in terms of water suppression but introduce artefacts into the NMR spectra, which lead to non-linearity of the NMR signal with respect to the metabolite concentration. Given the RF homogeneity modern NMR probes can achieve, the NOESY pre-saturation experiment is the best-suited NMR pulse sequence for ^{1}H-NMR-based metabolomic studies.

The area under each signal in a 1D-^{1}H NMR spectrum is proportional to the number of nuclei contributing to this signal, which can be used to accurately quantify metabolite concentrations. Using high-powered spectrometers, metabolite concentrations as low as 50 μmol L^{-1} can be reliably quantified within complex samples.

3.3.2 2D J-Resolved NMR Spectroscopy

Each proton in a molecule within a unique chemical environment results in a peak in a 1D-^{1}H NMR spectrum. Even with a small number of constituent molecules within a sample, NMR spectra often have hundreds of different peaks resulting in areas of overlapping peaks termed 'spectral congestion'. This congestion phenomenon is exacerbated by signal splittings through spin-spin or J-coupling. Depending on the number of protons attached to adjacent carbons, NMR signals are split into several components. Collectively, this compromises unambiguous peak identification and accurate metabolite quantification. To resolve any spectral discrepancies, the introduction of an independent second dimension can be invaluable (2D NMR). Figure 3.1 depicts a 2D-Jres spectrum of glutamate.

One of the simplest two-dimensional pulse sequences is 2D J-resolved NMR (2D-Jres NMR) spectroscopy [31]. The acquisition of a 2D-Jres spectrum is comparatively quick, compared to most 2D NMR methods, and can be achieved in as little as 5 min [32]. 2D-Jres spectroscopy reduces spectral overlap by separating
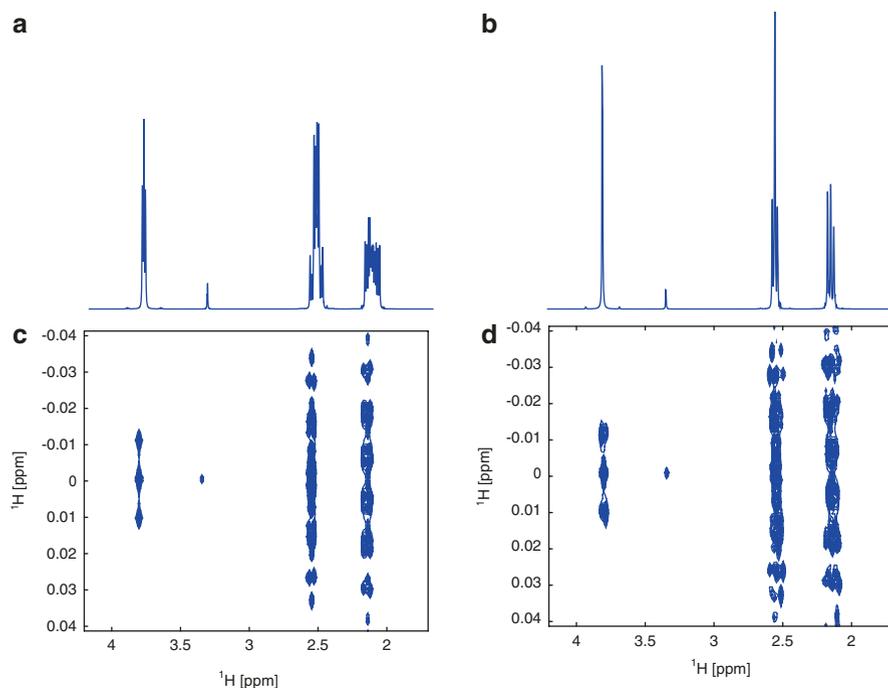
Fig. 3.1 One-dimensional projected Jres (**a**, **b**) and two-dimensional Jres spectra (**c**, **d**) of glutamate. The spectra in **a** and **c** are obtained by Fourier transformation of the apodized 2D-FID, whereas the spectra in **b** and **d** are the result of the same processing as the spectra on the *left-hand side* followed by a TILT and symmetrization procedure, which effectively removes homonuclear ^{1}H couplings from the 1D projection. All spectra were processed and plotted using the MetaboLab software package [30]

chemical shift and signal splitting. Whilst the horizontal dimension of the 2D spectrum contains contributions from both chemical shift and J-coupling, the vertical dimension of the spectrum contains only information from J-coupling. Through a procedure termed tilting, followed by symmetrisation, the J-coupling information exclusively appears in the vertical dimension of the spectrum, whilst the horizontal dimension only contains chemical shift information. The projection of the resulting 2D spectrum (pJres) onto the chemical shift axis appears as a proton-decoupled 1D NMR spectrum with all signal splittings being removed.

Another benefit of this spectroscopic technique is that it is possible to directly compare spectra acquired at different magnetic field strengths. Whilst resonance frequencies change linearly with the external magnetic field strength, J-coupling constants are fixed and independent of the external magnetic field strength. Through the separation of chemical shift and J-coupling into two independent dimensions, spectra acquired at different magnetic field strengths can be directly compared. This is not possible using a 1D ^{1}H NMR spectroscopic approach.

Although this experimental approach reduces overlap in the resulting NMR spectra, there can still be substantial spectral congestion present. Magnetisation transfer throughout the entire molecule (e.g. 2D-TOCSY) or the introduction of different nuclei (e.g. ^{13}C using a 2D-^{1}H,^{13}C HSQC approach) can be very helpful to further reduce spectral overlap. However, ^{12}C, by far the most abundant carbon isotope, is not NMR active. The natural abundance of ^{13}C is only 1.07 %, whereas it is practically 100 % for ^{1}H, and therefore limits detection to metabolites with reasonably high concentrations. Most metabolomic studies employ spectral techniques such as 1D-^{1}H or 2D-Jres using only the most sensitive nucleus, ^{1}H.

Both 1D-^{1}H and 2D-Jres NMR are untargeted approaches. No assumptions about molecular composition, etc. are made. The acquisition times are reasonably short so that spectra can be acquired in high throughput. Hardware development in recent years (e.g. automated sample filling, sample changing, tuning and matching) supports this approach. Using these NMR approaches, it is possible to acquire spectra for approximately 50–100 samples in a day using a single spectrometer. There are several spectral databases and numerous software tools available to facilitate the analysis of these spectra. At this stage, these NMR approaches are beneficial for metabolic profiling. Metabolites soluble in aqueous solutions are quantifiable using this methodology. Such metabolites include sugars, organic acids, amino acids, phenolic compounds.

3.3.3 2D ^{1}H-^{1}H-TOCSY NMR Spectroscopy

The 2D ^{1}H-^{1}H-TOCSY NMR experiment can help to reduce overlap further. In a 2D TOCSY spectrum, both frequency axes are chemical shift axes. The magnetization is transferred around the entire spin system (which is any part of a molecule where protons are not separated by a quaternary carbon). Therefore, in addition to signals with the same chemical shift in both dimensions (also known as diagonal peaks), additional signals appear where the chemical shifts in the horizontal and the vertical

dimension, respectively, are different (cross peaks, see Fig. 3.2). These cross peaks can help to annotate peaks in overcrowded regions of a 1D spectrum. As long as one of the resonances in the molecule is far enough away, the cross peak will be isolated in the 2D spectrum. Even if both diagonal signals, giving rise to the cross peak, are located in overlapped regions of the spectrum, it is still possible to obtain isolated cross peaks, as long as there is no other molecule that contains protons with very similar resonance frequencies in a single spin system.

The key element for the TOCSY experiment is the isotropic mixing sequence. This pulse sequence element creates an environment where the magnetization transfer via the J-coupled proton network can occur. There is a variety of different isotropic mixing sequences available with the DIPSI-2 and several variants of the MLEV being amongst the most commonly used. Whilst DIPSI-2 is a sequence developed for an optimal polarization transfer without shaped pulses, MLEV/16 is purely based on 180° pulses and therefore suffers considerably from off-resonance arte-facts. These artefacts can be avoided if the pulses used cover a very large bandwidth, which is why this mixing scheme is often used in conjunction with adiabatic pulses [33]. This is particularly attractive when using a HR-MAS probe, because the adia-batic pulses can be easily synchronized with the rotor speed, eliminating potential
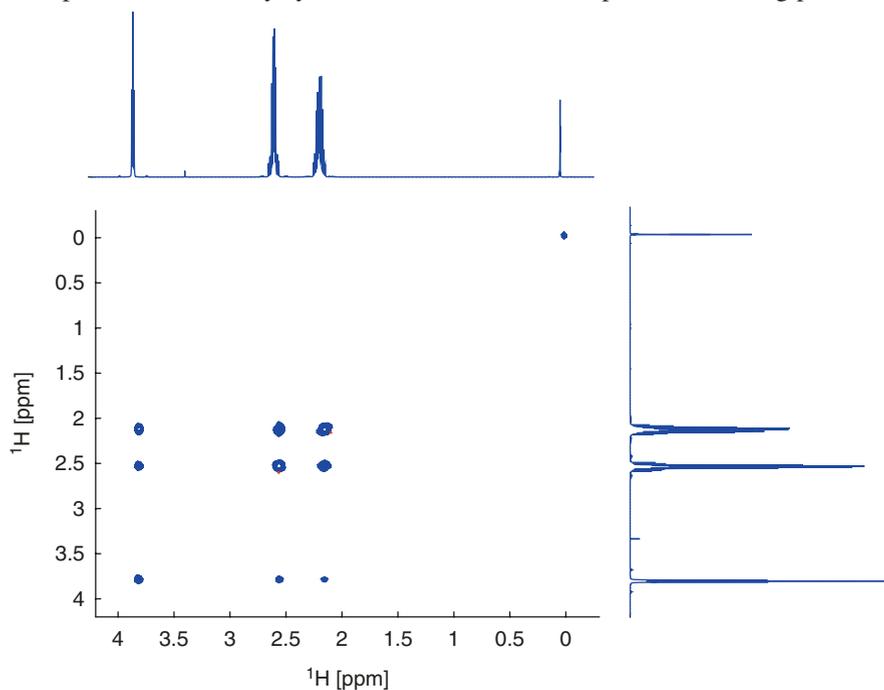
Fig. 3.2 Two-dimensional TOCSY spectrum of glutamate with 1D-NOESY-presat spectra of glu-tamate shown on *top* and the *right side*. Magnetization is distributed across the entire spin system inside the glutamate molecule. The NMR signal at the *upper right-hand corner* without any cross peaks to the glutamate signal belongs to a different molecule (DSS). All spectra were processed and plotted using the MetaboLab software package [30]

magnetic susceptibility-based artefacts as well as magnetization loss through RF inhomogeneity [34].

2D ^{1}H-^{1}H-TOCSY NMR spectra can also be used to determine relative pathway information in tracer studies. In any ^{1}H NMR spectrum, a proton signal is split into two peaks if the proton is bound to a ^{13}C nucleus. Whilst in a 1D-^{1}H spectrum this signal splitting exacerbates the spectral congestion, cross peaks in a 2D ^{1}H-^{1}H-TOCSY spectrum are well enough resolved to cope with the additional spectral complexity. Because each cross peak in a 2D-TOCSY spectrum belongs to two different protons, the splitting pattern of the cross peaks carries information on ^{13}C isotopomer distribution in the molecule [35].

3.3.4 2D ^{1}H-^{13}C-HSQC NMR Spectroscopy

The 2D ^{1}H-^{13}C-HSQC spectrum compared to the 2D ^{1}H-^{1}H-TOCSY spectrum shows a further decrease in spectral overlap. The HSQC pulse sequence transfers magnetization between protons and directly bounds ^{13}C nuclei, therefore correlating their respective resonance frequencies. Each pair of carbon and proton nuclei gives rise to a signal in a 2D-HSQC spectrum, with the ^{13}C chemical shift usually displayed in the vertical and the ^{1}H chemical shift in the horizontal dimension. The correlation of the two chemical shifts can help to annotate signals where there is ambiguity in 1D-^{1}H spectra.

Because of the natural abundance of ^{13}C (1.07 %), its use in standard metabolomic studies is quite limited due to its reduced sensitivity, which is vital for unambiguous metabolite identification. However, the 2D-^{1}H,^{13}C HSQC experiment is invaluable for tracer-based approaches [36, 37]. Rise in signal intensity and ^{13}C-^{13}C signal splittings, arising in the vertical (^{13}C) dimension of high-resolution HSQC spectra can be used to conduct a model-free isotopomer analysis in order to study metabolic events in great detail [37].

To retain multiplicity information, HSQC spectra need to be acquired at high resolution, such as 16,384 increments, which leads to prolonged acquisition times. Even when using fast acquisition techniques such as *n*on-*u*niform *s*ampling schemes (NUS), acquisition times can be as long as 5 h per spectrum.

3.4 Data Preprocessing

Most NMR-based metabolic studies, especially in a clinical setting, are performed with a large sample number. Differences between sample groups are not known in advance and can be quite subtle. Hence, the use of multivariate statistical analysis is essential in order to highlight differences or correlations between samples or groups of samples.

However, these multivariate statistical analyses are extremely sensitive to data preprocessing, which is an intermediate step between raw NMR spectra and statisti-

cal data analysis. Given the potential for errors during the multiple steps between raw NMR data and statistical outputs, robust and accurate processing is essential to ensure differences between the sample groups are biologically meaningful. Key stages of NMR data processing and preprocessing are detailed below.

3.4.1 Automated NMR Spectra Processing

An essential step after acquisition of the FID (*f*ree *i*nduction *d*ecay) is Fourier *t*ransformation (FT). This operation converts a signal from the time domain (seconds) to the frequency domain (Hertz). This procedure is fully automated on modern NMR spectrometers. Prior to FT, weighting functions, also called apodization functions, and zero filling are performed. The zero-filling process is based on the addition of zeroes to the existing data points of the FID to double their number ('zero filling once'). This is done to increase the resolution of the spectrum. The weighting functions are used to enhance either the sensitivity or the resolution of the resulting NMR spectrum, depending on the type acquired (e.g. 1D-^{1}H vs. 2D-Jres NMR). This procedure consists of multiplying the FID by a mathematical function (e.g. a decaying exponential, Gaussian, sine or squared sine bell or a combination of those window functions). The most commonly used apodization function for 1D-^{1}H NMR is the exponential window function, generally applied with a line broadening of 0.3 Hz. For 2D JRES experiments the most widely used window functions are sine or SEM [38].

Usually, only the real part of the frequency domain is displayed but it contains a mixture of absorption and dispersion line shapes. An absorption Lorentzian line shape is always positive and is centred at the frequency of the signal, whilst a dispersion Lorentzian line shape is broader and made of positive and negative parts.

Unfortunately, the spectrometer produces time-domain data with an arbitrary phase and so the real part of the spectrum will not be in pure absorption mode. Hence, it is necessary to adjust the phase of the spectrum until all the peaks appear to have the required absorption line shape. This procedure is called 'phasing the spectrum' and can be done automatically or manually using NMR software. This crucial step has to be done for 1D and multidimensional NMR spectra alike. It is important to point out that the process of manual phase correction is operator dependent with different spectroscopists producing different results based on the same raw data. Accordingly, automated phase correction is preferable, particularly for room temperature systems, where instruments produce data with very stable phases and high-quality baselines. In contrast, on cryogenic NMR systems, which generally produce spectra with less phase stability and less favourable baseline properties, phase correction has to been done manually.

3.4.2 Baseline Correction

Once all spectra have been phased, the next important preprocessing step is baseline correction. Indeed, most NMR spectra present baseline distortions due to different factors originating from instability or imperfections of the spectrometers or from the

nature of the sample (heterogeneity of the sample and chemical exchange). The presence of macromolecules (long chain lipids or proteins), which have very short relaxation times compared to small metabolites, gives rise to very broad signals in the NMR spectrum. Therefore, macromolecular signals appear as baseline distortions in the spectrum. As baseline distortions affect the area under each NMR peak, it impacts on metabolite quantification. As most multivariate statistical analysis methods treat baseline signals (noise) and real signals in the same way, these are directly affected when the baseline is not perfectly flat and can lead to overfitting of the data.

To correct these distortions, baseline correction can be performed either on the raw data before FT or on the NMR spectrum. Although the latter is the most commonly used way to adjust the baseline, it is possible to reconstruct the time domain by oversampling [39] or extrapolation [40] of the data. Digital filters have also been integrated into the acquisition parameters on modern NMR spectrometers to obtain a flat baseline with a very low residual first order phase correction by shortly delaying the acquisition and calculating the missing data points during this delay. For state-of-the-art NMR-based metabolic studies, use of the digital filters is highly recommended.

Baseline correction in the frequency domain is based on a subtraction of a modelled baseline of the experimental NMR spectrum. The most common methods are iterative polynomial fitting with automatic threshold [41], *a*symmetric *l*east *s*quare smoothing (AsLS) [42], *lo*cally *we*ighted *s*catterplot *s*moothing (LOWESS) [43] and automatic peak recognition followed by fitting a fifth degree polynomial and its subtraction from the original spectrum [44]. It is also possible to correct the baseline in an automated way using commercially available NMR software packages such as the Chenomx NMR suite [45].

For major baseline distortions, in particular when using cryogenically cooled probes, manual baseline correction is recommended. However, on standard metabolic systems with room temperature probes, automated processing is recommended to achieve the highest degree of reproducibility.

3.4.3 Alignment

Signal shifts between spectra from different samples lead to major problems in subsequent data analysis. These shifts can occur even if all the samples have been prepared following the same strict sample preparation protocol. Chemical shift referencing using the internal standard only allows correcting for global shifts between samples. Peaks can shift to different extents and in opposite directions for different signals in the NMR spectrum. Chemical shift is influenced by a multitude of factors, including concentration of salts or specific ions, instrumental factors or temperature and pH variations. Even small variations can bias multivariate statistical analysis and obscure the discovery of biomarkers or the pattern of metabolic profiles. Therefore, it is necessary to use alignment algorithms as an additional preprocessing step to correct for local signal shifts.

One of the most common NMR alignment methods is *c*orrelation *o*ptimized *w*arping (COW) [46], which aligns different intervals by optimizing the overall correlation

between the sample and a reference spectrum. This algorithm divides the spectrum into equal-sized segments and aligns them independently by compression or stretching. Although initially created to correct for misalignment of chromatographic datasets, this method works reliably for NMR spectra.

Another widely used method is *i*nterval *c*orrelated *s*hifting (icoshift) [47]. This warping method consists of segmenting the NMR spectra into regularly spaced or different length segments and aligning the different segments to a reference spectrum. The algorithm calculates and maximizes the cross-correlation of each interval by a *f*ast *F*ourier *t*ransform (FFT) and aligns all the segments simultaneously. By aligning the entire set of spectra concurrently, it shortens calculation time compared to the COW algorithm significantly.

Peak alignment represents a key challenge in NMR spectroscopy. Various methods to correct for signal misalignment have been developed such as hierarchical *clu*ster-based *p*eak *a*lignment (CluPA) [48], *d*ynamic *t*ime *w*arping (DTW) [46] and *p*eak *a*lignment using *r*educed *s*et mapping (PARS) [49]. For urine datasets, an iterative *f*uzzy *w*arping (FW) [50] algorithm, initially developed for chromatography, has been introduced and successfully implemented for NMR spectra.

Despite being widely used, these alignment methods may bias signal areas and can compromise the accuracy of metabolite quantification. For this reason, absolute quantification based on raw data is advised. Software packages such as Chenomx allow for signal shifts in individual NMR spectra and can therefore be used for reliable metabolite quantification without the need for signal alignment.

3.4.4 Binning

To apply multivariate statistical analysis, it is necessary to transform all the information contained in the spectra into a table containing as many rows as observations (samples) and as many columns as the number of variables (NMR data points). However, in metabolic studies, an NMR spectrum contains tens of thousands of data points, and a bucketing or binning procedure is commonly applied to reduce the number of variables and simplify data analysis. Binning consists of segmenting the NMR spectrum into a number of bins and measuring the integral (area under the curve) of every segment. Signal shifts smaller than the bin size will vanish after binning if they occur just within a single bin. Binning leads to a loss of resolution, and the size of the different bins has to be determined carefully. There are two different spectral binning procedures: all bins can be of equal size, typically between 0.005 and 0.05 ppm. Alternatively an 'intelligent' binning procedure [51–55] can be applied, which determines the size of each interval in such a way that every bin represents a peak or a group of peaks. The main drawback of equidistant binning is the risk of splitting a peak when a signal is assigned to different bins in different NMR spectra. Indeed, even if all the spectra are correctly aligned, due to different shimming qualities, the boundaries of the peak may vary slightly. Therefore, the use of variable bin size is recommended.

Regions of the spectrum containing only noise should be removed from the NMR spectrum to avoid overfitting of the data during statistical analysis. It is also necessary to remove regions corresponding to the solvent signals used during extraction or sample preparation. When a solvent suppression method (e.g. pre-saturation of one or multiple frequencies) is used, it is very important to exclude not only the region corresponding to the residual solvent peak but the entire region affected by the saturation.

3.4.5   Data Normalization

As we have seen in a previous part of this chapter, although NMR sample preparation is straightforward and most of the protocols are standardized and highly reproducible, small variations linked to this step may appear between the samples. For example, small variations may be caused during multiple stages of sample preparation (pipetting, weighing tissue). Furthermore, in the case of metabolic urine studies, the urine concentration, which depends in particular on the amount of water ingested, could also create variations between the samples. All these factors introduce variations in signal intensities which are not correlated to metabolism and which interfere with multivariate statistical analysis. To correct these effects, different normalization methods have been developed. The first one is normalization of the entire spectral area, which consists of dividing the intensity of every bucket by the *t*otal *s*pectral *a*rea (TSA) normalization. However, this method has limitations in terms of robustness and accuracy especially in the case of an important variation of one or few metabolites by downscaling the normalized spectra [56]. Thus, the standard method used in NMR-based metabolic study, which is able to address this issue, has been established and called *p*robabilistic *q*uotient *n*ormalization (PQN) [56]. This algorithm relies on the determination of the most probable dilution factor for all signals of the spectrum compared to a reference spectrum. Once the most probable quotient (k) has been calculated, a k-fold normalization is applied to all the variables.

The two above-mentioned procedures are the most common in NMR metabolomics, but other normalizing methods may also be used including normalization to creatinine concentration (for urine samples), vector length normalization and histogram matching normalization [57].

At this stage of the data processing, the final step is variance stabilization of the different variables. The concentration ranges of all the metabolites detected by NMR may be very important and vary a lot between the different samples. Although the most abundant metabolites usually display larger variations, it frequently happens that small signals are responsible for the discrimination between two groups of samples. If the data is not scaled, higher signals will have the largest influence on the statistical multivariate results and can obscure the contribution of the smaller signals. The most common scaling methods in NMR metabolomics are mean centring, autoscaling, Pareto scaling, vast scaling and glog transformation. It is important to point out that contrary to normalization methods, which are carried out independently on each sample, scaling operations are performed simultaneously on all spectra in the dataset.

Mean centring transforms all bins in such a way that they vary around zero instead of around the mean value and corrects variations between high-concentrated and low-concentrated metabolites. From a mathematical point of view, it consists of subtracting the mean intensity of each column for each observation and is typically applied prior to the application of another scaling method.

Autoscaling, also called standardization, involves the reduction of the data by dividing every variable by its standard deviation, whilst Pareto scaling entails a division by the square root of the standard deviation. Autoscaling converts all the values in such a way that all the metabolites have unit variance, so that they are equally important and have a comparable scale. Great care has to be taken when applying autoscaling. This method gives a large influence to variables containing only noise and could lead to overfitting of the data. Therefore, Pareto scaling is a good alternative by being an intermediate between standardization and no scaling, hence staying closer to the original data. However, this procedure has the disadvantage to decrease the importance of large variations, the latter being more reduced than small variations. Vast scaling [58], which corresponds to the acronym of *va*riable *st*ability (Vast), is an extension of autoscaling, but an additional scaling step is done using also the coefficient of variation as a scaling factor. This method has the opposite effect of Pareto scaling: it increases the influence of metabolites with small standard deviation and decreases the one of metabolites with large standard deviation. Finally, one of most suitable scaling method for NMR-based metabolic studies is the generalized logarithm (glog) transform [59]. It transforms the intensity for each variable to a new value dependant on the transform parameter and the value of original intensity, to stabilize the variability in the dataset. This method has been reported to be more suitable and more efficient than the other ones above-mentioned [59].

At this stage, the data are ready to be analysed by unsupervised multivariate statistical analysis such as *p*rincipal *c*omponent *a*nalysis (PCA) [60] or *i*ndependent *c*omponent *a*nalysis (ICA) [61] as an exploratory tool to determine some trends between the samples and in some favourable case to discriminate groups of samples. When the separation obtained with unsupervised methods is not good enough, it is possible to apply supervised methods (with pre-assignment of classes to the different samples) such as *p*artial *l*east *s*quare – *d*iscriminant *a*nalysis (PLS-DA) or *o*rthogonal PLS-DA (OPLS-DA) [62].

Figure 3.3 shows an example of a statistical data analysis applied to a series of 30 NMR spectra, which underwent data processing and preprocessing as outlined in this section.

3.5 NMR Databases for Metabolic Research

With the growth of NMR-based metabolomics over the past decade, there has been a concomitant need for appropriate spectral databases to aid in the identification of metabolites. During the early days of metabolomics or metabolite analysis by NMR, laboratories had to construct their own private databases and run a range of
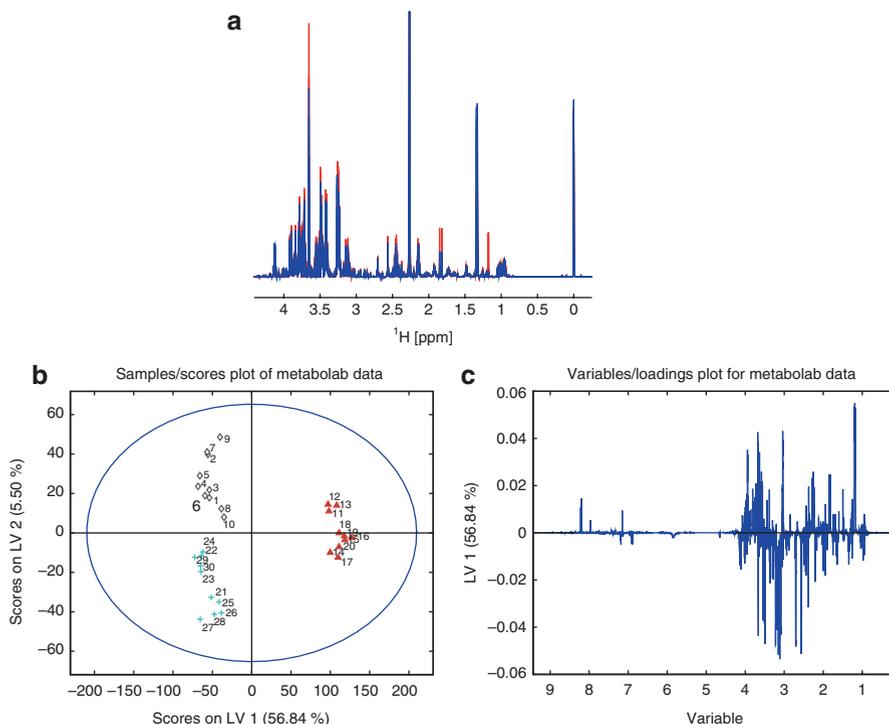
Fig. 3.3 Possible outcome of a statistical data analysis of NMR spectra. Panel **a** shows an overlay of 30 NMR spectra of blood samples belonging to three different sample groups (control and two different disease states). A score plot of the PLS-DA analysis of these samples is shown in panel **b**, where the control group on the *right-hand side* is well separated from the two diseased groups. Both disease groups separate from the control group on the second latent variable. Panel **c** shows a loading plot of latent variable number 1, which is a graphical representation of how much each data point in the NMR spectra contributes to the separation between the different sample groups. All spectra were processed and plotted using the MetaboLab software package [30]

reference spectra of the most common metabolites they expected to find. This targeted approach also involved consulting with published work to search for it. Apart from being time-consuming, this was generally unsatisfactory given the many methods by which samples were prepared and analysed. The need for databases containing not only reference spectra but also information pertaining to how this data was obtained (in addition to other information, such as cross reference) has led to an enormous increase in confidence within the discipline. Most modern spectral databases vary in the detail of their content, may be freely available and may vary in sophistication and indeed in price in some circumstances. Whilst they contain, as a minimum, chemical shift reference information, others may provide complex spectral-fitting algorithms, allowing for proper quantification of metabolites within a complex matrix. 1D NMR data remains the most abundant data accessible, but, in recent years, a surge in HSQC and TOCSY experimental spectral libraries has been

apparent as the use of these techniques within the community has increased. This is a brief list of the more recognizable freely available databases.

3.5.1   HMDB

The Human Metabolome Database (HMDB) was first introduced in 2007 [63] and currently remains the world's largest and most comprehensive, organism-specific metabolomic database. Standardized NMR spectra (such as from 1D-^{1}H, 1D-^{13}C, 2D-^{1}H^{13}C HSQC, 2D-^{1}H^{1}H TOCSY) has risen from 385 spectra in the original release to 1,054 in 2013 [64]. Many of these are linked to MS data, as well as offering other useful information such as the role of a given metabolite in metabolic pathways or any link it may have to specific diseases. This increase in metabolite number is primarily a result of the significant expansion of both 'detected' metabolites and 'expected' metabolites (those for which biochemical pathways are known or human intake/exposure is frequent but the compound has yet to be detected in the body). In the 3.0 release of HMDB, a clear distinction has been made between these categories.

The new chemical ontology also distinguishes the source (or probable source) of these compounds using an 'origin' data field. Using this data field (for both detected and expected metabolites), compounds are further classified as being microbial, endogenous and drug, toxin or food derived. The database can be accessed by browsing, chemical shift search or bulk downloading. However, a significant disadvantage of the database is the lack of batch-processing capabilities for spectral matching.

3.5.2   Madison Metabolomics Consortium Database (MMCD)

This database, which is maintained by the National Magnetic Resonance Facility at Madison, is a resource for metabolomics research based on nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) [65]. Specifically, one-dimensional (1D) and two-dimensional (2D) NMR data are available, including 1D-^{1}H, 1D-^{13}C, 2D-^{1}H^{1}H-TOCSY, 2D-^{1}H^{13}C HSQC, 2D-^{1}H^{13}C HMBC (*h*eteronuclear *m*ultiple-*b*ond *c*orrelation) and 2D-HSQC-TOCSY. Peak lists can be typed in manually, or files can be uploaded in a variety of the common formats used by NMR spectroscopists. This database gives an emphasis to *Arabidopsis thaliana* but is not species specific.

Similar to the HMDB, the MMCD provides chemical shifts, chemical formula, names and synonyms, structure and physical and chemical properties, in addition to NMR (and MS) data on pure compounds under defined conditions in some cases. NMR searches can use any one of MMCD's three chemical shift databases: experimental, empirically predicted from structure or quantum chemical calculated.

NMR-based searches give users considerable flexibility with regard to the type and quality of data entered. Chemical shifts can be combined with filters that search for complex multinuclear spin topologies. For example, users can specify chemical shift and atom connectivity (e.g. number of attached hydrogens).

As well as experimental NMR data collected by the MMC, its compounds contain links to NMR data collected by the HMDB. Although the HMDB and MMC collect data under different conditions (HMDB, H_2O, 50 mmol L^{-1} phosphate buffer, pH 7.0; MMC, 99.9 % D_2O, containing 50 mmol L^{-1} phosphate buffer, pH 7.4), the chemical shifts for compounds common to the two are as expected in good general agreement.

The database also contains information on the presence of the metabolite in different biological species and extensive links to images, references and other public databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and PubChem. The MMCD search engine supports versatile data mining and allows users to make individual or bulk queries on the basis of experimental NMR and/or MS data plus other criteria. The site supports complex queries from any combination of its five basic search engines: text, structure, NMR, mass and miscellanea.

3.5.3   BioMagResBank

The BioMagResBank (BMRB) is the central repository for experimental NMR spectral data, primarily for macromolecules [66]. However, the BMRB also contains a subsection specifically for metabolite data. The database as a whole also contains structures, structure viewing applets, nomenclature data, extensive 1D and 2D spectral peak lists (from 1D, TOCSY, DEPT, HSQC experiments), raw spectra and FIDs for several hundred molecules. The data is both searchable and downloadable. Based at the University of Wisconsin, BMRB mirror sites exist at Osaka University, Japan, and at CERM in Florence, Italy, with the Osaka facility also being a data deposition and processing site. BMRB also collaborates closely with NMR metabolomic/metabonomic groups and the National Magnetic Resonance Facility at Madison (NMRFAM) and with many other groups in the NMR community.

3.5.4   BML-NMR

The Birmingham Metabolite Library Nuclear Magnetic Resonance (BML-NMR) database is centred at the University of Birmingham (UK) and the initiative acts as a freely available resource containing over NMR spectra of 208 common metabolite standards [67]. This database also includes both 2D-^{1}H J-resolved spectra and 1D-^{1}H spectra. The spectra have been recorded at 500 MHz using various water-suppression methods and acquisition parameters, for solutions at pH values of 6.6, 7.0 and 7.4. Library data can be accessed freely and searched through a custom-written web

interface, whilst FIDs, NMR spectra and associated metadata can be downloaded according to a Microsoft installer compatible XML schema.

3.5.5 COLMAR

The COLMAR database as developed in the Brüschweiler laboratory is a web server that allows one to query a range of NMR data. The general approach is the unification of the NMR spectroscopic information of two of the largest public metabolomic databases, namely, the BMRB and the HMDB discussed before. COLMAR ^{13}C-^{1}H HSQC sorts HSQC spectra of metabolites into their individual isomeric states, which permits improved identification of metabolites, because it is isomer-population insensitive [68]. According to the authors, this, together with an improved query algorithm, allows COLMAR ^{13}C-^{1}H HSQC metabolomic database to increase the accuracy of metabolite identification by more than 37 % and decreases the false positive identification rates by more than 82 % over existing ^{13}C-^{1}H HSQC metabolomic databases.

Also included on the homepage is the C-TOCCATA customized database, which specializes the querying of ^{13}C-^{13}C TOCSY spectra of uniformly ^{13}C-labelled metabolomic samples, and the $^{1}H(^{13}C)$-TOCCATA customized database, which permits the querying of $^{1}H^{1}H$ TOCSY and $^{13}C^{1}H$ HSQC-TOCSY spectra of complex metabolite mixtures at natural ^{13}C abundance [69]. The innovative element of such databases is that they sort the spectral information of each metabolite into its individual spin systems and, where applicable, its slowly interconverting isomers. Since selected cross sections of the 2D TOCSY spectrum reflects the 1D spectrum of spin systems rather than the entire 1D spectrum, this increases the accuracy of metabolite identification over existing 1D ^{1}H and 1D ^{13}C NMR metabolomic databases, respectively.

3.6 Metabolic Analysis of Kidneys Prior to Transplantation Using NMR Techniques

3.6.1 Introduction to Transplantation

The evolution of renal transplantation is a triumph of modern medicine, with patients undergoing organ transplant in the modern era expecting a 90 % 1-year graft survival [70]. Transplantation is now unquestionably the optimal treatment for patients with end stage renal failure, with distinct survival and quality of life advantages over remaining on dialysis, even in patients with complex medical comorbidity [71–73]. Kidneys from live donors have the best outcomes, but those not fortunate enough to have a suitable donor must rely on a cadaveric organ.

For cadaveric organs, there is an inevitable time duration between the organ removal (retrieval) operation and implantation into the recipient, owing to both logistical (e.g. transportation) and immune compatibility matching processes.

Hypothermic machine perfusion (HMP) and static cold storage (SCS) are the two methods of kidney preservation that are widely used in clinical practice during this time period between organ retrieval and implantation [74]. Static cold storage (SCS) is the simplest and remains the most commonly used [75]. The kidney is essentially stored 'on ice' in a hypertonic solution within a polystyrene storage container.

Hypothermic machine perfusion (HMP) is the alternative to SCS and involves the recirculation of specialized perfusion fluid through the renal vasculature at sub-physiological pressures (e.g. 30 mmHg for the LifePort kidney transporter). HMP is associated with improved short- and longer-term function following transplantation of cadaveric kidneys [76–78]. Although the beneficial effects of HMP were initially thought to be largely mechanical, there is increasing evidence that facilitative metabolism may also have a beneficial role [79, 80]. NMR spectroscopy is one method for detailing renal metabolism during SCS and HMP conditions, and ^{1}H 1D NMR has been used to analyse urine, perfusion fluid recipient serum and tissue samples in kidney transplant models [81–88].

An in-depth understanding of the metabolism occurring during HMP has clear translational value. Despite the ever-growing demand, transplantation is limited by the paucity of available organs. Resultantly, increasing numbers of marginal organs are being transplanted, many of which would have been previously deemed to be non-transplantable [89]. Such marginal kidneys have inferior outcomes compared with standard criteria cadaveric organs, and therefore metabolic optimization could be a useful therapeutic target during the preservation window to ensure the best outcome for these high-risk kidneys. Secondly, there is a clinical need for a reliable test to determine post-transplant graft outcome and thus determine the 'usability' of such marginal organs. However to date, there remains no single reliable pre-transplant predictor of post-transplant graft function, and a metabolic biomarker could serve to highlight such non-viable organs.

Hypothermic machine perfusion seems particularly well suited for metabolomic analysis as perfusion fluid samples, or perfusate, are readily available, are non-invasive and offer insight into the cellular components within the organ. Furthermore, machine perfusion also offers the opportunity for repeated measurements, and the rapidity of ^{1}H spectrum acquisition makes the utilization in real time feasible [83].

3.6.2 Metabolism During Organ Preservation

The governing principles of SCS are simple and have undergone little change since the inception of renal transplantation some 60 years ago, namely, the deceleration of metabolism that occurs as the temperature of the kidney is reduced. At storage temperatures below 4 °C, there is a reduction of normal metabolic function to 5–8 %

[90]. This is thought to preserve organ ATP reserves and limit the accumulation of harmful byproducts such as lactic acid.

However, hypothermia does not cause a uniform deceleration of all metabolic pathways [79], and, furthermore, the effect on cellular metabolism of hypothermia within different organs is not uniform [91]. If indeed certain enzymes are particularly thermo-sensitive, then manipulation of such pathways may offer a mechanism for protective metabolic support.

In a recent study of urine in patients post kidney transplant, using both ^{1}H NMR and GC-MS [84], there were clear differences between the urinary metabolome at day 7 compared with 12 months postoperatively. These early differences included increased taurine, hypotaurine and D-glucose and decreased citric acid. The authors suggested (based on metabolic pathway analysis software) that these changes were due to the predomination of glycolytic pathways and reduction in TCA activity in the post-transplant period. In a previous study of urine ^{1}H NMR spectra, Foxall et al. demonstrated differing spectral patterns for kidneys with good function compared to those with delayed function, highlighting the link between the urinary metabolome post-transplant and functional outcome [88]. High levels of the medullary trimethylamine-N-oxide (TMAO), in particular, correlated with poorer function and are an indicator of renal medullary injury [81, 88, 92].

In a similar (^{1}H NMR) study of serum from kidney transplants, the metabolic profile was not found to vary as dramatically over the first week but was markedly different to that from healthy controls [93].

In a murine transplant model, Serkova et al. demonstrated increased levels of allantoin from animals with kidney transplants exposed to greater ischemic insults [81]. During periods of oxidative stress, there is increased xanthine oxidase activity and degradation of xanthines, resulting in accumulation of end pathway products such as allantoin. Therefore, the authors suggested allantoin could be used as a surrogate marker for oxidative stress and that levels correlated with the severity of cold ischemia. Interestingly, the metabolic phenotype identified in this study had better correlation with degree of histological injury compared with the non-specific but commonly used measurement of transplant function serum creatinine [81].

Cellular levels of ATP have been proposed as a valid determinant of organ viability during the preservation period which can be measured directly and indirectly using ^{31}P NMR [94–101]. This technique has been employed to study whole organs ex vivo, obviating the problems associated with the rapid degradation of ATP in vitro in both oxygenated [94, 101] and non-oxygenated conditions [99]. The prognostic value of ATP measurement to denote organ viability appears preserved even several years following the index transplantation [102].

Bon et al. reported the metabolite panel present in the perfusion fluid of porcine kidneys during HMP using ^{1}H 1D NMR [83]. The concentration of metabolites increased during perfusion and included central metabolites such as lactate and acetate, amino acids (e.g. alanine, valine, glutamine and glutamate), creatinine and TMAO. This metabolic perfusion fluid panel has been corroborated in subsequent studies [85, 86]. In the study by Bon et al., the metabolites in the perfusion fluid

identified by ^{1}H NMR was different when a modified perfusion fluid was used, highlighting that metabolism can be manipulated during HMP.

We sought to determine the amount of new metabolism that occurs during HMP and SCS conditions using ^{1}H NMR (as yet unpublished). Pairs of porcine kidneys were preserved using either SCS or HMP conditions for 24 h. Overlay spectra of both tissue and fluid are depicted in Fig. 3.4. The total amount of each metabolite was calculated in the perfusate in each condition as well as the total amount in each kidney for these closed systems. This was compared with control kidneys and the net metabolic gain ascertained. As expected, the total amount of metabolites in the circulating HMP fluid was greater compared with the static conditions of SCS. However, the amount of metabolites in the tissue also varied dramatically between the two conditions. The net gain of central metabolites such as lactate, glutamate, aspartate, fumarate and acetate was greater for HMP compared with SCS conditions although not exclusively in the tissue. This study served to highlight several important concepts. Firstly, that the net metabolic gain during preservation
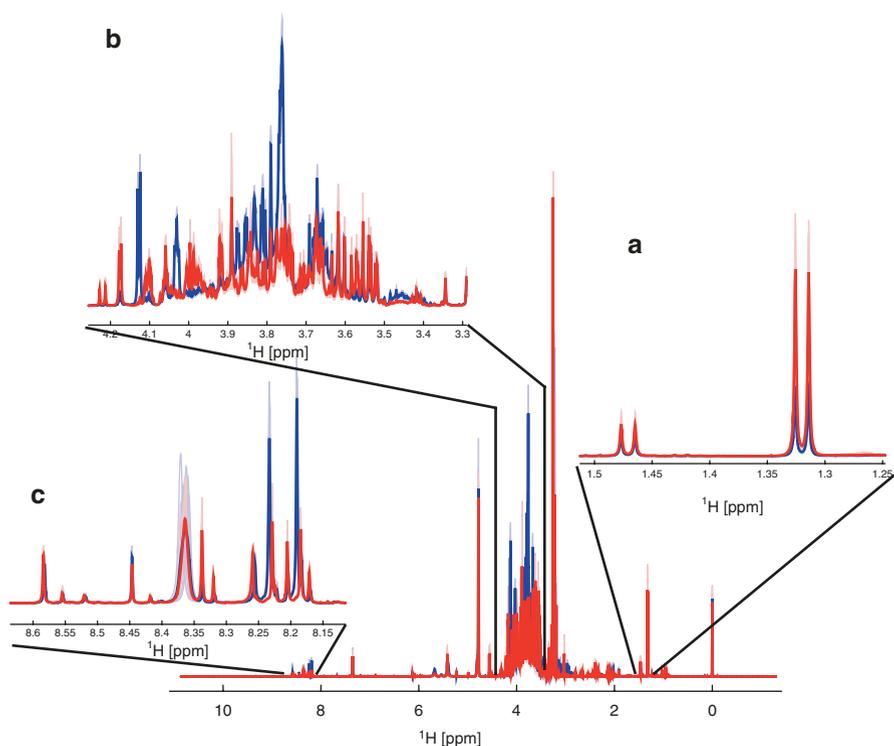
Fig. 3.4 ^{1}H-NMR spectral overlay plot demonstrating marked differences in spectra between kidneys stored using HMP (*red*) and SCS (*blue*) conditions for 24 h. Samples are from extracted paired porcine medulla samples with the whole spectrum, and focused regions of interest (**a–c**) are displayed. All spectra were processed and plotted using the MetaboLab software package [30]

is different for HMP and SCS conditions, secondly that ^{1}H NMR can be utilized to determine these differences and thirdly that even in these extreme non-physiological conditions such as HMP, efficient mechanisms are active to transport metabolites from the intracellular arena into the extracellular perfusion fluid.

Porcine kidneys are considered the best model for human organs in transplantation studies owing to their similar physiological and anatomical properties [103–106]. In order to determine whether porcine organs are a valid metabolic model during HMP, we compared the metabolic profiles of perfusate for human and porcine organs using ^{1}H 1D NMR [86]. Of the 30 metabolites that were identifiable using NMR and present in both species, 16 (53.3 %) were present in comparable concentrations in the pig and human kidney perfusates. For 29 metabolites (96.7 %), there was no difference in the rate of change of concentration between pig and human samples suggesting that porcine kidneys are indeed valid metabolic models for human HMP studies. This is important, as experimental studies are needed in order to establish the optimal conditions of perfusion with additional therapies such as supplemental oxygen still under investigation.

3.6.3   Biomarker Studies

There is a clinical need for a reliable test to determine post-transplant graft outcome and thus determine the 'usability' of marginal organs. Although scoring systems have been proposed for this purpose, often using histological parameters from kidney biopsy [107], they are not used widely. In practice, donor information, such as patient age, comorbidity or terminal serum creatinine, along with kidney biopsy data are used to judge the 'transplantability' of kidney, but they still have a limited capacity for accurate prediction of graft outcomes [108, 109]. Furthermore, clinically used measurements for the monitoring of post-transplant graft function also have limitations as rejection can occur in the presence of normal creatinine values, and structural changes on renal biopsy only present after significant injury has occurred [110].

The identification of a biomarker during the pre-transplant period would be attractive and useful but to date has proved elusive. Low intrarenal resistance has been shown to correlate with improved post-transplant function, but several studies have demonstrated that organs should not be rejected solely on flow dynamic information which strengthens the need for a perfusate biomarker [111–114].

In a systematic review of perfusate biomarker analysis by Bhangoo et al. [115], there was found to be a paucity of high-quality biomarker studies. Elevated levels of glutathione-S-transferase (GST), lactate dehydrogenase (LDH) and aspartate transaminase (AST) were significantly associated with delayed graft function in the majority of studies. However, there was insufficient evidence to recommend any single parameter as a sensitive pre-transplant biomarker. AST is expressed from injured renal parenchymal cells. GST is a marker of renal tubular injury and LDH is a non-specific marker of cellular injury [115].

In order to determine whether ^{1}H 1D NMR could be used to highlight a useful biomarker to predict transplant outcome, the perfusion fluid of 26 cadaveric kidneys was analysed during HMP at two time points (45 min and 4 h) [85]. We found that the metabolic profile of kidneys that functioned immediately (IGF) differed to those in which delayed graft function (DGF) was observed. Four metabolites appeared particularly predictive of graft function: glucose, inosine, leucine and gluconate. This study served to highlight that the metabolic profile of perfusion fluid in cadaveric human kidneys for transplantation is predictive of functional outcome. The idea of a metabolic panel identified using ^{1}H NMR serving as a biomarker of transplant viability has been corroborated by previous animal studies [83].

3.6.4 Metabolic Pathway Analysis and 2D Studies

1D NMR methods (both ^{1}H and ^{31}P) have highlighted an array of metabolites that can be used both to understand metabolic processes within the organ and potentially serve as biomarkers of post-transplant function. Whilst metabolic pathway software is available to try and highlight possible pathways involved, such tools have inherent limitations, which are probably exaggerated in the highly non-physiological conditions of ex vivo organ preservation.

Although multiple time point analysis can provide further evidence to demonstrate active pathways, ^{1}H 1D NMR essentially details a metabolic snapshot at the time of sampling. It is often difficult to draw meaningful mechanistic metabolic information from this. For example, the detection of lactate within the perfusate of a machine-perfused kidney could be secondary to the metabolism or release of pre-existing intracellular substrate stores as well as de novo metabolism of substrates derived from the perfusion fluid. Thus, the appearance of a particular metabolite within the perfusion fluid does not confirm de novo metabolism.

Metabolic tracer analysis is an alternative way to highlight active metabolic pathways and can elegantly and unequivocally demonstrate the presence of de novo metabolism within these complex systems. NMR spectroscopy is a powerful tool to analyse complex ^{13}C isotopomer/isotopologue distributions in metabolites derived from labelled tracer molecules. There are various spectroscopic methods available of which the simplest is 1D-^{13}C approach. This is not a new concept with initial studies using ^{13}C in NMR tracer experiments reported over 40 years ago [116] and validation of this model in ex vivo organ perfusion models (e.g. heart and lung) [117–120].

Whilst the data requirement (and therefore acquisition time) is much greater for two-dimensional (2D) ^{13}C NMR tracer studies, they do have significant advantages compared to 1D ^{13}C NMR experiments. In addition to the greater sensitivity, there is increased spectral dispersion using 2D NMR, and the utility of this has been demonstrated in human studies [121].

In an effort to determine metabolism in porcine kidneys during HMP, we incorporated [U-^{13}C] glucose into the cooled recirculating perfusion fluid [122]. Analysis

of perfusion fluid and kidney tissue extracts was performed using both 1D ^{1}H and 2D ^{1}H, ^{13}C heteronuclear single quantum coherence NMR (2D-^{1}H,^{13}C HSQC). This approach enables both metabolite quantification and proportionate distribution of ^{13}C isotopologues to be calculated. In this small study, we found that there was significant enrichment of ^{13}C in central metabolites such as [U-^{13}C] lactate which provides unequivocal evidence of de novo glycolytic pathway activity.

3.6.5   Conclusion

There appears to be a strong correlation between the metabolic phenotype of transplant kidneys during the period prior to transplantation and the functional outcome for that organ. The implications of this are twofold; firstly that a metabolic 'panel' could be used as a useful biomarker to determine which organs are non-viable and should not be transplanted. Secondly, the metabolism could potentially be optimized during perfusion to improve the function of highly damaged organs. NMR spectroscopy is a valid technique for metabolic characterization of organs during perfusion and is likely to be used increasingly for this purpose. Both 1D-^{1}H NMR and 2D studies offer metabolic insights into this complex system with the high-throughput, reproducible quantification of metabolites determined by 1D-^{1}H analysis complemented by mechanistic information from 2D-^{13}C studies. We hope that the utility of NMR in kidney transplantation can be translated to multiple clinical scenarios.

References

1. Bory C, Boulieu R, Chantin C, Mathieu M. Diagnosis of alcaptonuria: rapid analysis of homogentisic acid by HPLC. Clin Chim Acta. 1990;189(1):7–11.
2. Wishart DS. Metabolomics: the principles and potential applications to transplantation. Am J Transplant. 2005;5(12):2814–20.
3. Bales JR, Bell JD, Nicholson JK, Sadler PJ. 1H NMR studies of urine during fasting: excretion of ketone bodies and acetylcarnitine. Magn Reson Med. 1986;3(6):849–56.
4. Dona AC, Jiménez B, Schäfer H, Humpfer E, Spraul M, Lewis MR, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem. 2014;86(19):9887–94.
5. Griffiths WJ, Wang Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. Chem Soc Rev. 2009;38(7):1882–96.
6. Ludwig C, Ward DG, Martin A, Viant MR, Ismail T, Johnson PJ, et al. Fast targeted multidimensional NMR metabolomics of colorectal cancer. Magn Reson Chem MRC. 2009;47 Suppl 1:S68–73.
7. Smolinska A, Blanchet L, Buydens LMC, Wijmenga SS. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Anal Chim Acta. 2012;750:82–97.
8. Sadagopan NP, Li W, Cook JA, Galvan B, Weller DL, Fountain ST, Cohen LH. Investigation of EDTA anticoagulant in plasma to improve the throughput of liquid chromatography/tandem mass spectrometric assays. Rapid Commun Mass Spectrom. 2003;17(10):1065–70.

9. Faijes M, Mars AE, Smid EJ. Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum. Microb Cell Factories. 2007;6(1):1.

10. Warne MA, Lenz EM, Osborn D, Weeks JM, Nicholson JK. Comparative biochemistry and short-term starvation effects on the earthworms Eisenia veneta and Lumbricus terrestris studied by 1 H NMR spectroscopy and pattern recognition. Soil Biol Biochem. 2001;33(9):1171–80.

11. Rosenblum ES, Viant MR, Braid BM, Moore JD, Friedman CS, Tjeerdema RS. Characterizing the metabolic actions of natural stresses in the California red abalone, Haliotis rufescens using 1H NMR metabolomics. Metabolomics. 2005;1(2):199–209.

12. Viant MR, Rosenblum ES, Tjeerdema RS. NMR-based metabolomics: a powerful approach for characterizing the effects of environmental stressors on organism health. Environ Sci Technol. 2003;37(21):4982–9.

13. Bouchereau A, Guénot P, Larher F. Analysis of amines in plant materials. J Chromatogr B Biomed Sci App. 2000;747(1):49–67.

14. Anastassiades M, Lehotay SJ, Štajnbaher D, Schenck FJ. Fast and easy multiresidue method employing acetonitrile extraction/partitioning and "dispersive solid-phase extraction" for the determination of pesticide residues in produce. J AOAC Int. 2003;86(2):412–31.

15. Mackay GM, Zheng L, van den Broek NJ, Gottlieb E. Chapter five: analysis of cell metabolism using LC-MS and isotope tracers. Methods Enzymol. 2015;561:171–96.

16. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37(8):911–7.

17. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res. 2008;49(5):1137–46.

18. Lin CY, Wu H, Tjeerdema RS, Viant MR. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. Metabolomics. 2007;3(1):55–67.

19. David A, Abdul-Sada A, Lange A, Tyler CR, Hill EM. A new approach for plasma (xeno) metabolomics based on solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionisation mass spectrometry. J Chromatogr A. 2014;1365:72–85.

20. Chetwynd AJ, Abdul-Sada A, Hill EM. Solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionization mass spectrometry for improved global urine metabolomics. Anal Chem. 2015;87:1158–65.

21. Martínez-Bisbal MC, Monleon D, Assemat O, Piotto M, Piquer J, Llacer JL, Celda B. Determination of metabolite concentrations in human brain tumour biopsy samples using HR-MAS and ERETIC measurements. NMR Biomed. 2009;22(2):199–206.

22. Bharti SK, Sinha N, Joshi BS, Mandal SK, Roy R, Khetrapal CL. Improved quantification from 1H-NMR spectra using reduced repetition times. Metabolomics. 2008;4(4):367–76.

23. Rabeson H, Fauvelle F, Testylier G, Foquin A, Carpentier P, Dorandeu F, Graveron-Demilly D. Quantitation with QUEST of brain HRMAS-NMR signals: application to metabolic disorders in experimental epileptic seizures. Magn Reson Med. 2008;59(6):1266–73.

24. Elbayed K, Berl V, Debeuckelaere C, Moussallieh FM, Piotto M, Namer IJ, Lepoittevin JP. HR-MAS NMR spectroscopy of reconstructed human epidermis: potential for the in situ investigation of the chemical interactions between skin allergens and nucleophilic amino acids. Chem Res Toxicol. 2013;26(1):136–45.

25. Heude C, Lemasson E, Elbayed K, Piotto M. Rapid assessment of fish freshness and quality by 1H HR-MAS NMR spectroscopy. Food Anal Methods. 2015;8(4):907–15.

26. Lippens G, Bourdonneau M, Dhalluin C, Warras R, Richert T, Seetharaman C, Boutillon C, Piotto M. Study of compounds attached to solid supports using high resolution magic angle spinning NMR. Curr Org Chem. 1999;3:147–69.

27. Martínez-Bisbal MC, Martí-Bonmatí L, Piquer J, Revert A, Ferrer P, Llácer JL, et al. 1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas. NMR Biomed. 2004;17(4):191–205.

28. Guéron M, Plateau P, Decorps M. Solvent signal suppression in NMR. Prog Nucl Magn Reson Spectrosc. 1991;23(2):135–209.

29. Nguyen BD, Meng X, Donovan KJ, Shaka AJ. SOGGY: solvent-optimized double gradient spectroscopy for water suppression. A comparison with some existing techniques. J Magn Reson. 2007;184(2):263–74.

30. Ludwig C, Günther UL. MetaboLab – advanced NMR data processing and analysis for metabolomics. BMC Bioinformatics. 2011;12:366.

31. Aue WP, Karhan J, Ernst RR. Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy. J Chem Phys. 1976;64(10):4226–7.

32. Ludwig C, Viant MR. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. Phytochem Anal PCA. 2010;21(1):22–32.

33. Peti W, Griesinger C, Bermel W. Adiabatic TOCSY for C, C and H, H J-transfer. J Biomol NMR. 2000;18(3):199–205.

34. Piotto M, Bourdonneau M, Furrer J, Bianco A, Raya J, Elbayed K. Destruction of magnetization during TOCSY experiments performed under magic angle spinning: effect of radial B1 inhomogeneities. J Magn Reson. 2001;149(1):114–8.

35. Lane AN, Fan TW-M, Bousamra M, Higashi RM, Yan J, Miller DM. Stable isotope-resolved metabolomics (SIRM) in cancer research with clinical application to nonsmall cell lung cancer. Omics J Integr Biol. 2011;15(3):173–82.

36. Hollinshead KER, Williams DS, Tennant DA, Ludwig C. Probing cancer cell metabolism using NMR spectroscopy. Adv Exp Med Biol. 2016;899:89–111.

37. Lussey-Lepoutre C, Hollinshead KER, Ludwig C, Menara M, Morin A, Castro-Vega L-J, et al. Loss of succinate dehydrogenase activity results in dependency on pyruvate carboxylation for cellular anabolism. Nat Commun. 2015;6:8784.

38. Tiziani S, Lodi A, Ludwig C, Parsons HM, Viant MR. Effects of the application of different window functions and projection methods on processing of 1 H J-resolved nuclear magnetic resonance spectra for metabolomics. Anal Chim Acta. 2008;610(1):80–8.

39. Wider G. Elimination of baseline artifacts in NMR spectra by oversampling. J Magn Reson. 1990;89(2):406–9.

40. Marion D, Bax A. Baseline correction of 2D FT NMR spectra using a simple linear prediction extrapolation of the time-domain data. J Magn Reson. 1989;83(1):205–11.

41. Gan F, Ruan G, Mo J. Baseline correction by improved iterative polynomial fitting with automatic threshold. Chemom Intell Lab Syst. 2006;82(1):59–65.

42. Eilers PH, Boelens HF. Baseline correction with asymmetric least squares smoothing. Leiden Univ Med Cent Rep. 2005;1:1.

43. Xi Y, Rocke DM. Baseline correction for NMR spectroscopic metabolomics data analysis. BMC Bioinformatics. 2008;9(1):1.

44. Dietrich W, Rüdel CH, Neumann M. Fast and precise automatic baseline correction of one- and two-dimensional NMR spectra. J Magn Reson. 1991;91(1):1–11.

45. Chang D, Banack CD, Shah SL. Robust baseline correction algorithm for signal dense NMR spectra. J Magn Reson. 2007;187(2):288–92.

46. Tomasi G, Van Den Berg F, Andersson C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. J Chemom. 2004;18(5):231–41.

47. Savorani F, Tomasi G, Engelsen SB. Icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson. 2010;202(2):190–202.

48. Vu TN, Valkenborg D, Smets K, Verwaest KA, Dommisse R, Lemière F, Laukens K. An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. BMC Bioinformatics. 2011;12(1):405.

49. Torgrip RJO, Aberg M, Karlberg B, Jacobsson SP. Peak alignment using reduced set mapping. J Chemom. 2003;17(11):573–82.

50. Wu W, Daszykowski M, Walczak B, Sweatman BC, Connor SC, Haselden JN, Lutz M. Peak alignment of urine NMR spectra using fuzzy warping. J Chem Inf Model. 2006;46(2):863–75.

51. Davis RA, Charlton AJ, Godward J, Jones SA, Harrison M, Wilson JC. Adaptive binning: an improved binning method for metabolomics data using the undecimated wavelet transform. Chemom Intell Lab Syst. 2007;85(1):144–54.

52. Anderson PE, Reo NV, DelRaso NJ, Doom TE, Raymer ML. Gaussian binning: a new kernel-based method for processing NMR spectroscopic data for metabolomics. Metabolomics. 2008;4(3):261–72.

53. De Meyer T, Sinnaeve D, Van Gasse B, Tsiporkova E, Rietzschel ER, De Buyzere ML, et al. NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. Anal Chem. 2008;80(10):3783–90.

54. Anderson PE, Mahle DA, Doom TE, Reo NV, DelRaso NJ, Raymer ML. Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. Metabolomics. 2011;7(2):179–90.

55. Sousa SAA, Magalhães A, Ferreira MMC. Optimized bucketing for NMR spectra: three case studies. Chemom Intell Lab Syst. 2013;122:93–102.

56. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem. 2006;78(13):4281–90.

57. Dong J, Cheng KK, Xu J, Chen Z, Griffin JL. Group aggregating normalization method for the preprocessing of NMR-based metabolomic data. Chemom Intell Lab Syst. 2011;108(2):123–32.

58. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006;7(1):1.

59. Parsons HM, Ludwig C, Günther UL, Viant MR. Improved classification accuracy in 1-and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. BMC Bioinformatics. 2007;8(1):234.

60. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab Syst. 1987;2(1–3):37–52.

61. Lee TW. Independent component analysis. Springer US, Boston MA, USA; 1998. p. 27–66.

62. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemom. 2006;20(8–10):341–51.

63. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007;35(Database issue):D521–6.

64. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0 – the Human Metabolome Database in 2013. Nucleic Acids Res. 2013;41(Database issue):D801–7.

65. Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, et al. Metabolite identification via the Madison metabolomics consortium database. Nat Biotechnol. 2008;26(2):162–4.

66. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, et al. BioMagResBank. Nucleic Acids Res. 2008;36(Database issue):D402–8.

67. Ludwig C, Easton JM, Lodi A, Tiziani S, Manzoor SE, Southam AD, et al. Birmingham Metabolite Library: a publicly accessible database of 1-D 1H and 2-D 1H J-resolved NMR spectra of authentic metabolite standards (BML-NMR). Metabolomics. 2011;8(1):8–18.

68. Bingol K, Li D-W, Bruschweiler-Li L, Cabrera OA, Megraw T, Zhang F, et al. Unified and isomer-specific NMR metabolomics database for the accurate analysis of (13)C-(1)H HSQC spectra. ACS Chem Biol. 2015;10(2):452–9.

69. Bingol K, Zhang F, Bruschweiler-Li L, Brüschweiler R. TOCCATA: a customized carbon total correlation spectroscopy NMR metabolomics database. Anal Chem. 2012;84(21):9395–401.

70. NHS Blood and Transplant. Organ donation and transplantation activity report 2014/15 [Internet]. 2015. Available from: http://nhsbtmediaservices.blob.core.windows.net/organ-donation-assets/pdfs/activity_report_2014_15.pdf.

71. Wolfe RA, Ashby VB, Milford EL, Ojo AO, Ettenger RE, Agodoa LY, et al. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. N Engl J Med. 1999;341(23):1725–30.

72. Jofré R, López-Gómez JM, Moreno F, Sanz-Guajardo D, Valderrábano F. Changes in quality of life after renal transplantation. Am J Kidney Dis Off J Natl Kidney Found. 1998;32(1):93–100.

73. Fiebiger W, Mitterbauer C, Oberbauer R. Health-related quality of life outcomes after kidney transplantation. Health Qual Life Outcomes. 2004;2:2.

74. Lee CY, Mangino MJ. Preservation methods for kidney and liver. Organogenesis. 2009;5(3):105–12.

75. Opelz G, Döhler B. Multicenter analysis of kidney preservation. Transplantation. 2007;83(3):247–53.

76. Moers C, Smits JM, Maathuis M-HJ, Treckmann J, van Gelder F, Napieralski BP, et al. Machine perfusion or cold storage in deceased-donor kidney transplantation. N Engl J Med. 2009;360(1):7–19.

77. Moers C, Pirenne J, Paul A, Ploeg RJ, Machine Preservation Trial Study Group. Machine perfusion or cold storage in deceased-donor kidney transplantation. N Engl J Med. 2012;366(8):770–1.

78. O'Callaghan JM, Morgan RD, Knight SR, Morris PJ. Systematic review and meta-analysis of hypothermic machine perfusion versus static cold storage of kidney allografts on transplant outcomes. Br J Surg. 2013;100(8):991–1001.

79. Taylor MJ, Baicu SC. Current state of hypothermic machine perfusion preservation of organs: the clinical perspective. Cryobiology. 2010;60(3 Suppl):S20–35.

80. Fuller BJ, Lee CY. Hypothermic perfusion preservation: the future of organ preservation revisited? Cryobiology. 2007;54(2):129–45.

81. Serkova N, Fuller TF, Klawitter J, Freise CE, Niemann CU. H-NMR-based metabolic signatures of mild and severe ischemia/reperfusion injury in rat kidney transplants. Kidney Int. 2005;67(3):1142–51.

82. Hauet T, Baumert H, Gibelin H, Hameury F, Goujon JM, Carretier M, et al. Noninvasive monitoring of citrate, acetate, lactate, and renal medullary osmolyte excretion in urine as biomarkers of exposure to ischemic reperfusion injury. Cryobiology. 2000;41(4):280–91.

83. Bon D, Billault C, Claire B, Thuillier R, Hebrard W, Boildieu N, et al. Analysis of perfusates during hypothermic machine perfusion by NMR spectroscopy: a potential tool for predicting kidney graft outcome. Transplantation. 2014;97(8):810–6.

84. Kienana M, Lydie N-D, Jean-Michel H, Binta D, Matthias B, Patrick E, et al. Elucidating time-dependent changes in the urinary metabolome of renal transplant patients by a combined (1)H NMR and GC-MS approach. Mol Biosyst. 2015;11(9):2493–510.

85. Guy AJ, Nath J, Cobbold M, Ludwig C, Tennant DA, Inston NG, et al. Metabolomic analysis of perfusate during hypothermic machine perfusion of human cadaveric kidneys. Transplantation. 2015;99(4):754–9.

86. Nath J, Guy A, Smith TB, Cobbold M, Inston NG, Hodson J, et al. Metabolomic perfusate analysis during kidney machine perfusion: the pig provides an appropriate model for human studies. Bruns H, editor. PLoS ONE. 2014;9(12):e114818.

87. Faure JP. Polyethylene glycol reduces early and long-term cold ischemia-reperfusion and renal medulla injury. J Pharmacol Exp Ther. 2002;302(3):861–70.

88. Foxall PJ, Mellotte GJ, Bending MR, Lindon JC, Nicholson JK. NMR spectroscopy as a novel approach to the monitoring of renal transplant function. Kidney Int. 1993;43(1):234–45.

89. Port FK, Bragg-Gresham JL, Metzger RA, Dykstra DM, Gillespie BW, Young EW, et al. Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors. Transplantation. 2002;74(9):1281–6.

90. Southard JH, Belzer FO. Organ preservation. Annu Rev Med. 1995;46:235–47.

91. Boutilier RG. Mechanisms of cell survival in hypoxia and hypothermia. J Exp Biol. 2001;204(Pt 18):3171–81.

92. Hauet T, Gibelin H, Godart C, Eugene M, Carretier M. Kidney retrieval conditions influence damage to renal medulla: evaluation by proton nuclear magnetic resonance (NMR) spectroscopy. Clin Chem Lab Med. 2000;38(11):1085–92.

93. Li L, Sui W, Che W, Li W, Chen J, Li H, et al. 1H NMR-based metabolic profiling of human serum before and after renal transplantation. ASAIO J. 2013;59(3):268–93.

94. Buchs J-B, Lazeyras F, Ruttimann R, Nastasi A, Morel P. Oxygenated hypothermic pulsatile perfusion versus cold static storage for kidneys from non heart-beating donors tested by in-line ATP resynthesis to establish a strategy of preservation. Perfusion. 2011;26(2):159–65.

95. von Elverfeldt D, Niekisch M, Quaschning T, El Saman A, Kirste G, Krämer-Guth A, et al. Kinetics of PME/Pi in pig kidneys during cold ischemia. NMR Biomed. 2007;20(7):652–7.
96. Bretan PN, Baldwin N, Novick AC, Majors A, Easley K, Ng T, et al. Pretransplant assessment of renal viability by phosphorus-31 magnetic resonance spectroscopy. Clinical experience in 40 recipient patients. Transplantation. 1989;48(1):48–53.
97. Möller HE, Gaupp A, Vestring T, Dietl KH, Vermathen P, Buchholz B, et al. Kinetics of metabolism in human kidney transplants measured by dynamic 31P NMR spectroscopy. Z Für Naturforschung C J Biosci. 1995;50(5–6):439–50.
98. Kurková D, Herynek V, Gintelová J, Táborský P, Hájek M. Potential of 31P magnetic resonance spectroscopy in monitoring the viability of human renal grafts stored in Euro-Collins perfusion solution. Physiol Res Acad Sci Bohemoslov. 1995;44(5):327–32.
99. Hené RJ, van der Grond J, Boer WH, Mali WP, Koomans HA. Pre-transplantation assessment of renal viability with 31P magnetic resonance spectroscopy. Kidney Int. 1994;46(6):1694–9.
100. Pomer S, Hull WE, Röhl L. Assessment of renal viability for transplantation by high field 31P-NMR. Transplant Proc. 1988;20(5):899–901.
101. Lazeyras F, Buhler L, Vallee J-P, Hergt M, Nastasi A, Ruttimann R, et al. Detection of ATP by "in line" 31P magnetic resonance spectroscopy during oxygenated hypothermic pulsatile perfusion of pigs' kidneys. Magma N Y N. 2012;25(5):391–9.
102. Seto K, Ikehira H, Obata T, Sakamoto K, Yamada K, Kashiwabara H, et al. Long-term assessment of posttransplant renal prognosis with 31 P magnetic resonance spectroscopy. Transplantation. 2001;72(4):627–30.
103. Giraud S, Favreau F, Chatauret N, Thuillier R, Maiga S, Hauet T. Contribution of large pig for renal ischemia-reperfusion and transplantation studies: the preclinical model. J Biomed Biotechnol. 2011;2011:1–14.
104. Sachs DH. The pig as a potential xenograft donor. Vet Immunol Immunopathol. 1994;43(1–3):185–91.
105. Hannon JP, Bossone CA, Wade CE. Normal physiological values for conscious pigs used in biomedical research. Lab Anim Sci. 1990;40(3):293–8.
106. Swindle MM. The development of swine models in drug discovery and development. Future Med Chem. 2012;4(14):1771–2.
107. Cockfield SM, Moore RB, Todd G, Solez K, Gourishankar S. The prognostic utility of deceased donor implantation biopsy in determining function and graft survival after kidney transplantation. Transplantation. 2010;89(5):559–66.
108. Jochmans I, Pirenne J. Graft quality assessment in kidney transplantation: not an exact science yet! Curr Opin Organ Transplant. 2011;16(2):174–9.
109. Louvar DW, Li N, Snyder J, Peng Y, Kasiske BL, Israni AK. "Nature versus nurture" study of deceased-donor pairs in kidney transplantation. J Am Soc Nephrol. 2009;20(6):1351–8.
110. Bohra R, Klepacki J, Klawitter J, Klawitter J, Thurman JM, Christians U. Proteomics and metabolomics in renal transplantation-quo vadis? Transpl Int. 2013;26(3):225–41.
111. Jochmans I, Moers C, Smits JM, Leuvenink HGD, Treckmann J, Paul A, et al. The prognostic value of renal resistance during hypothermic machine perfusion of deceased donor kidneys. Am J Transplant. 2011;11(10):2214–20.
112. Sonnenday CJ, Cooper M, Kraus E, Gage F, Handley C, Montgomery RA. The hazards of basing acceptance of cadaveric renal allografts on pulsatile perfusion parameters alone. Transplantation. 2003;75(12):2029–33.
113. Mozes MF, Skolek RB, Korf BC. Use of perfusion parameters in predicting outcomes of machine-preserved kidneys. Transplant Proc. 2005;37(1):350–1.
114. Guarrera JV, Goldstein MJ, Samstein B, Henry S, Reverte C, Arrington B, et al. "When good kidneys pump badly": outcomes of deceased donor renal allografts with poor pulsatile perfusion characteristics. Transpl Int. 2010;23(4):444–6.
115. Bhangoo RS, Hall IE, Reese PP, Parikh CR. Deceased-donor kidney perfusate and urine biomarkers for kidney allograft outcomes: a systematic review. Nephrol Dial Transplant. 2012;27(8):3305–14.

116. Dorman D, Roberts J. Nuclear magnetic resonance spectroscopy: 13C spectra of some common nucleotides. Proc Natl Acad Sci U S A. 1970;65:19–26.
117. Peltz M, He T-T, Adams GA, Chao RY, Meyer DM, Jessen ME. Characterizing lung metabolism with carbon-13 magnetic resonance spectroscopy in a small-animal model: evidence of gluconeogenesis during hypothermic storage. Transplantation. 2005;80(3):417–20.
118. Gilbert NF, Meyer PE, Tauriainen MP, Chao RY, Patel JB, Malloy CR, et al. Effects of hypothermia on myocardial substrate selection. Ann Thorac Surg. 2002;74(4):1208–12.
119. Peltz M, He T-T, Adams GA, Koshy S, Burgess SC, Chao RY, et al. Perfusion preservation maintains myocardial ATP levels and reduces apoptosis in an ex vivo rat heart transplantation model. Surgery. 2005;138(4):795–805.
120. Rosenbaum DH, Peltz M, DiMaio JM, Meyer DM, Wait MA, Merritt ME, et al. Perfusion preservation versus static preservation for cardiac transplantation: effects on myocardial function and metabolism. J Heart Lung Transplant. 2008;27(1):93–9.
121. Fan TWM, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, et al. Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). Mol Cancer. 2009;8:41.
122. Nath J, Smith T, Hollis A, Ebbs S, Canbilen S, Tennant D, et al. 13C glucose labelling studies using 2D NMR is a useful tool for determining ex vivo whole organ metabolism during hypothermic machine perfusion of kidneys. Transpl Res. 2016;5(7).

Chapter 4
Metabolomic Strategies Involving Mass Spectrometry Combined with Liquid and Gas Chromatography

Aline Soriano Lopes, Elisa Castañeda Santa Cruz, Alessandra Sussulini, and Aline Klassen

Abstract Amongst all omics sciences, there is no doubt that metabolomics is undergoing the most important growth in the last decade. The advances in analytical techniques and data analysis tools are the main factors that make possible the development and establishment of metabolomics as a significant research field in systems biology. As metabolomic analysis demands high sensitivity for detecting metabolites present in low concentrations in biological samples, high-resolution power for identifying the metabolites and wide dynamic range to detect metabolites with variable concentrations in complex matrices, mass spectrometry is being the most extensively used analytical technique for fulfilling these requirements. Mass spectrometry alone can be used in a metabolomic analysis; however, some issues such as ion suppression may difficultate the quantification/identification of metabolites with lower concentrations or some metabolite classes that do not ionise as well as others. The best choice is coupling separation techniques, such as gas or liquid chromatography, to mass spectrometry, in order to improve the sensitivity and resolution power of the analysis, besides obtaining extra information (retention time) that facilitates the identification of the metabolites, especially when considering untargeted metabolomic strategies. In this chapter, the main aspects of mass spectrometry (MS), liquid chromatography (LC) and gas chromatography (GC) are discussed, and recent clinical applications of LC-MS and GC-MS are also presented.

Keywords Metabolomics • Mass spectrometry • Liquid chromatography • Gas chromatography • Mass analysers

A.S. Lopes • A. Klassen (✉)
Department of Exact and Earth Sciences, Federal University of São Paulo (UNIFESP), Diadema, SP, Brazil
e-mail: aline.klassen@unifesp.br; aline.klassen@gmail.com

E.C.S. Cruz • A. Sussulini
Laboratory of Bioanalytics and Integrated Omics (LaBIOmics), University of Campinas (UNICAMP), Campinas, SP, Brazil

Abbreviations

AMDIS	Automated Mass Spectra Deconvolution and Identification System
APCI	Atmospheric pressure chemical ionisation
APPI	Atmospheric pressure photoionisation
BSTFA	N,O-bis-(trimethylsilyl)-trifluoroacetamide
CE	Capillary electrophoresis
CI	Chemical ionisation
DB-5MS	Equivalent to a (5 %-phenyl)-methylpolysiloxane
DC	Direct current
DIMS	Direct infusion mass spectrometry
EI	Electron ionisation
ERHILIC	Electrostatic repulsion hydrophilic interaction chromatography
ESI	Electrospray ionisation
FTICR	Fourier transform ion cyclotron resonance
FTMS	Fourier transform mass spectrometry
FWHM	Full-width half-maximum
GC	Gas chromatography
GC-MS	Gas chromatography mass spectrometry
GCxGC	Comprehensive two-dimensional gas chromatography
HILIC	Hydrophilic interaction liquid chromatography
IT	Ion trap
LC	Liquid chromatography
LSER	Linear solvation energy relationship
MS	Mass spectrometry
MSTFA	N-methyl-N-(trimethylsilyl)-trifluoroacetamide
MTBSTFA	N-methyl-N-tertbutyldimethylsilyltrifluoroacetamide
NIST	National Institute of Standards and Technology
NPLC	Normal-phase liquid chromatography
OT	Orbitrap
Q	Quadrupole
QIT	Quadrupole ion trap
QqQ	Triple quadrupole
QTOF	Quadrupole time of flight
RF	Radio frequency
RPLC	Reverse-phase liquid chromatography
RTL	Runtime library
SIM	Single-ion monitoring
SRM	Selected reaction monitoring
THF	Tetrahydrofuran
TOF	Time of flight
UPLC	Ultra-performance liquid chromatography

4.1 Introduction

Metabolomic analysis in biological systems has become more and more important nowadays in different research areas, including the search for biomarkers that can support the understanding of the aetiology and biological/molecular bases of complex diseases, such as diabetes, coronary heart disease and cancer [1].

Mass spectrometry (MS) is the most suitable analytical technique in metabolomics for clinical studies, and the use of this technique has been growing in the last decades. In the clinical area, biological fluids, such as urine, blood and saliva, are commonly the object of study used for metabolomic analysis. Those biofluids are complex samples, with variable compound concentrations; consequently, the direct injection of the samples in the mass spectrometer compromises the detection of many metabolites due to ion suppression. Thus, separation techniques, such as gas chromatography (GC) and liquid chromatography (LC), coupled to the mass spectrometer, are required. In this way, all compounds, previously separated by one of the techniques mentioned before, will be introduced one after another into the mass spectrometer, which facilitates the metabolite identification by retention time and using authentic standards, in addition to the structural information (mass-to-charge ratio and relative abundance of the molecular ion/fragments) obtained by mass spectrometry.

In this chapter, we discuss the main aspects of mass spectrometry, including the most widely used mass analysers in targeted or untargeted metabolomics, and also GC- and LC-hyphenated techniques, considering the proper MS ionisation techniques in each case.

4.2 Mass Spectrometry

Mass spectrometry (MS) is often the technique of choice in clinical metabolomics for identifying/quantifying different classes of metabolites. Briefly, MS consists in producing gas-phase ions that are further detected and characterised by their mass and charge [2].

A sample inlet, an ion source, a mass analyser and a detector compose a mass spectrometer. The sample inlet has the function of introducing the sample into the mass spectrometer, the ion source generates gas-phase ions via an ionisation technique, the mass analyser separates the ions according to their mass-to-charge ratio (m/z), and the detector generates an electric current from the incident ions that is proportional to their abundances [3].

As considered by this chapter, the sample inlet is a separation technique (either gas or liquid chromatography), and proper ionisation techniques will be discussed for each chromatography-MS coupling in the following sections. Of course, the sample can be also directly injected into a mass spectrometer, being this technique called direct infusion mass spectrometry (DIMS). However, the major drawbacks are ion suppression

effects, which cause an enormous loss of metabolite information, and the requirement of a high-resolution mass analyser, which increases the cost of the analysis.

Mass analysers can be used either alone or combined. This combination can be made between the same type of mass analyser or between different mass analysers (hybrid instruments) and is called tandem mass spectrometry (MS/MS). In MS/MS, the ions that arrive at the first mass analyser (precursor ions) are isolated, subsequently fragmented, and finally those fragment ions are separated according to their *m/z* in a second mass analyser and detected. For some types of mass analysers, the number of mass analysis steps can be increased, i.e. the fragment ions can be refragmented and further detected. In this case, the experiment is termed multiple-stage mass spectrometry (MS^n, where n refers to the number of mass analysis steps). Tandem mass spectrometry and multiple-stage mass spectrometry improve the identification of a molecule, because not only the molecular ions are detected but also the fragments generated from precursor ions.

The main performance characteristics of a mass analyser are [2–4]:

(a) Mass accuracy (accuracy of the measured *m/z* provided by the mass analyser, directly related to the mass resolving power and stability of the mass analyser)
(b) Mass resolving power (ability of a mass spectrometer to provide a specified value of mass resolution, i.e. generate distinct signals for two ions with a small *m/z* difference)
(c) Mass range (limits of *m/z* over which a mass spectrometer can detect ions or is operated to record a mass spectrum)
(d) Transmission efficiency (ratio of the number of ions reaching the detector and the number of ions leaving the mass analyser, related to the sensitivity of the mass spectrometer, i.e. the minimal concentration of a compound leading to a peak intensity greater than a specified signal-to-noise ratio)
(e) Scan speed (rate at which the analyser measures over a certain mass range)
(f) Scan cycle time (the time required to obtain a mass spectrum, also called duty cycle)

Mass analysers that can be used for LC-MS or GC-MS are the same. The particularity that guides the mass analyser choice is the type of metabolomic analysis to be performed, either targeted or untargeted.

4.2.1   Mass Analysers Used for Targeted Metabolomics

As previously described in Chap. 1, the goal of targeted metabolomics is to perform a quantitative analysis of specific metabolites (or a defined set of metabolites). The main features for a mass analyser to be used in targeted metabolomics are transmission efficiency, scan cycle time and scan speed. In this case, single quadrupole (Q), triple quadrupole (QqQ), quadrupole ion trap (QIT) and Orbitrap (OT) are the most employed mass analysers.

Single quadrupole (Q) is the simplest mass analyser, which operates either in the single-ion monitoring (SIM) or scan mode. For targeted metabolomics, the SIM

mode is preferred because it provides a significantly better sensitivity. In the SIM mode, the quadrupole parameters (RF and DC voltages) are adjusted to filter and select only one specific *m/z* [5].

Triple quadrupole (QqQ) instruments have the advantage over single quadrupole ones for being able to perform selected reaction monitoring (SRM) experiments. In the first mass analysis stage, a specific precursor ion is selected by the first quadrupole. Then, fragmentation occurs in the collision cell (second quadrupole or other multipole, operating at RF only) by collisions with an inert gas. Finally, specific fragment ions are monitored in the third quadrupole to increase both the sensitivity and the selectivity compared to the single quadrupole operating in the SIM mode [3, 5].

Quadrupole ion trap (QIT) has similar principles as the single quadrupole mass analyser, which uses an electric field applied in the electrodes for ion separation by mass-to-charge ratio. Once the ions, with certain *m/z*, enter into the electrode area, the applied field promotes the orbit of these ions. As the radio frequency is increased, the ions with higher *m/z* become more stabilised, whereas the ions with lower *m/z* become less stabilised and are not detected due to the collisions with the walls of the mass analyser [6]. This analyser is not suitable to be used in combination with UPLC due to its low sensitivity, once the injection volume is reduced in this chromatography type.

The most recently developed mass analyser, with high acquisition speed, is the Orbitrap (OT). It operates at acquisition rates of 12 Hz. The analyser operation is based on harmonic ion oscillations in electrostatic field. The ions around a central electrode are trapped, and the *m/z* values are measured from the frequency of ion oscillations. As the ions are tangentially introduced into a logarithmic electric DC field between these two electrodes, they start to oscillate radially around the wire and are eventually ejected at the ends of the trap. This device provides high mass resolution (>100,000 FWHM), high mass accuracy (2–5 ppm) and acceptable dynamic range (10^3). However, the scan speed is inversely related to mass resolution, i.e. one scan per second can be acquired when selecting 100,000 mass resolution; as a result, the reproduction of the correct chromatographic peak shape is affected. Whereas, when faster scanning is selected (10 scans s^{-1}), mass resolution is decreased (10,000 FWHM) [7]. Recently, a modification of the Orbitrap Elite™ instrument has provided a resolution above 1,000,000 at a transient length of 3 s [8, 9].

Thus, for targeted metabolomics, high sensitivity can be achieved with OT mass analyser operating at high acquisition speed. In addition, the high-resolution power helps to associate the fragment ions to precursor ions in complex mixtures.

4.2.2 Mass Analysers Used for Untargeted Metabolomics

Untargeted metabolomics consists in an exploratory analysis that aims to identify the entire set (or at least the majority) of metabolites contained in a biological sample, as previously described in Chap. 1. The main features for a mass analyser to be used in untargeted metabolomics are mass resolution power, mass range and mass accuracy. In this case, time of flight (TOF), quadrupole time of flight (QTOF),

Fourier transform ion cyclotron resonance (FTICR) and Orbitrap (OT) are the most employed mass analysers.

Time of flight and quadrupole time of flight are the most used mass analysers for untargeted metabolomics due to the data acquisition over a wide mass range with high mass accuracy and resolving power. Their performance involves the time measurement that ions take to travel from the beginning to the end of a field-free flight tube. Ions are accelerated in an electric field reaching a terminal linear velocity, which depends on their *m/z* ratio. About 10,000 consecutive scanning events per second at sampling rate of 50 Hz can be achieved with a mass error of 5 ppm, whereas a resolution of 40,000 at *m/z* 956 and maximum acquisition speed of 30 Hz can be achieved by the integration of ion mobility separation to TOF. QTOF mass analyser is distinguished of TOF by the possibility to integrate MS/MS at the same resolution of the precursor ion [9].

FTICR is a high-resolution mass analyser that employs cyclotron frequency in a fixed magnetic field for the determination of the ions *m/z*. The disadvantage of FTICR instruments is their relatively slow acquisition rates. At a scan rate of 1 Hz with mass resolution of 100,000 at *m/z* 4000, the number of points over the chromatographic peak, especially if additional MS/MS scans are required, is low when FTMS is combined with modern fast chromatography systems. This limits the application of FTICR in liquid chromatography mass spectrometry (LC-MS)- and capillary electrophoresis mass spectrometry (CE-MS)-based metabolomics [10].

About OT mass analysers, as its performance was already explained in Sect. 4.2.1, it is important to mention that an OT mass analyser allows untargeted metabolomic experiments due to the high resolution at high speed acquisition and mass accuracy (<5 ppm), which is much higher in comparison to a TOF mass analyser. According to the literature, OT has become a mainstream instrument for metabolomics, for providing more complete results by LC-OT in comparison to LC-FTICR and LC-TOF MS [9].

4.3   Liquid Chromatography Coupled to Mass Spectrometry in Clinical Metabolomics

Amongst all separation techniques that can be coupled to mass spectrometry in metabolomics, liquid chromatography (LC) is the most employed, mainly due to its versatility, i.e. the possibility of separating different classes of compounds, from very polar up to very non-polar compounds. This versatility is possibly owed to the many chromatographic columns with a variety of stationary phases available [11].

The separation in the chromatographic system depends, basically, on properties such as hydrophobicity, molecular size and polarity of the compounds. The separation of compounds occurs into a chromatographic column composed by a stationary phase with polar or non-polar properties. In chromatography using polar stationary phase columns, the solvent used to elute the compounds from the stationary phase

(mobile phase) presents higher polarity than the stationary phase, which is called normal-phase liquid chromatography (NPLC). However, in chromatography using non-polar stationary phase columns, the mobile phase presents lower polarity than the stationary phase, which is called reversed-phase liquid chromatography (RPLC). Then, non-polar compounds, such as lipids, elute first in NPLC, whereas polar compounds, such as amino acids, elute first in RPLC [12].

Clinical samples contain very polar compounds (amino acids) and also compounds with high hydrophobicity (phospholipids). Thus, the stationary phase can be chosen based on the compound classes of interest, if the aim of the study is targeted metabolomics. However, if the interest is to reach the most information as possible (untargeted metabolomics), more than one type of column is necessary [13]. Table 4.1 summarises LC-MS applications in clinical metabolomics, considering different column types.

Table 4.1 Applications of LC-MS in clinical metabolomics, considering the most used column types

Column	LC-MS system	Metabolites	Biological matrix	Reference(s)
C18	UPLC-QTOF MS	Amino acids, sugars, peptides, lipids, nitrogenous bases, organic acids, nucleotides, phospholipids	Human liver cancer cell line, mammary cancer cell line, breast cancer cell line	[14–17]
	UPLC-TOF MS, HPLC-QTOF MS, UPLC-QqQ MS	Amino acids, α-hidroxy esters, sterol lipids, sugars, organic acids	Diabetic mouse kidney and liver tissue	[18]
		Lipids, sugars, organic acids, amino acids, sphingolipids	Diabetic mouse plasma, rat plasma, human myocardial ischemia plasma, rat haemolytic and aplastic anaemia plasma	[18–22]
		Amino acids, lipids, organic acids	Mouse hair	[18]
		Organic acids	Rat faeces	[19]
	HPLC-FTICR MS, HPLC-QTOF MS, UPLC-QqQ MS, UPLC-QTOF MS	Nucleosides, amino acids, organic acids, nitrogenous bases, sphingolipids	Human bladder cancer urine, human urine, human type 2 diabetic urine, rat haemolytic and aplastic anaemia urine	[20, 22–24]
	UPLC-IT-FTICR MS, HPLC-TOF MS, HPLC-QTOF MS	Steroids, lipids, amino acids, dipeptides, glycerolipids, nitrogenous bases, organic acids	Human hepatitis disease serum, rat myocardial infarction serum, human oesophageal cancer serum	[25–27]

<div align="right">(continued)</div>

Table 4.1 (continued)

Column	LC-MS system	Metabolites	Biological matrix	Reference(s)
C8	UPLC-QTOF MS	Lipids, amino acids, organic acids	Diabetic rat liver tissue	[28]
	HPLC-QTOF MS	Amino acids, organic acids, lipids	Human oesophageal cancer serum	[27]
	HPLC-TOF MS, HPLC-QTOF MS	Lipids, steroid lipids, glycophospholipids, sugars, amino acids, sphingolipids	Human plasma, rat diabetic plasma	[29, 30]
	HPLC-QTOF MS	Organic acids, sterol lipids	Human urine	[20]
HILIC	UPLC-QTOF MS, HPLC-QqQ MS	Sugars, amino acids, nucleosides, organic acids, nitrogenous bases, peptides	Human liver cancer cell line, mouse mammary tumour model, human pancreatic cancer cells, human colon cancer cells	[14, 31–33]
	HPLC-IT MS, UPLC-TOF MS, UPLC-QqQ MS	Organic acids, amino acids, nucleosides, amino sugars, sugars, nitrogenous bases	Rat urine, human urine	[34–37]
	UPLC-QqQ MS, HPLC-TOF MS, UPLC-QTOF MS, UPLC-OT MS, HPLC-QqQ MS	Amino acids, organic acids, phospholipids	Rat plasma, human plasma, human cardiovascular disease plasma	[21, 29, 38]
		Amino acids, lipids, glycerolipids, nitrogenous bases	Rat brain and liver tissue	[39, 40]

In order to obtain information of non-polar and weakly polar compounds, C18 and C8 are the most used columns, mainly due to their robustness, ease to handle, fast conditioning, versatility, the ability to cover a wide range of chemical classes and good performance for clinical untargeted metabolomics [41, 42]. However, for hydrophilic, ionic and polar compounds, which are poorly retained in C18 or C8 columns, or insufficiently charged to be retained by ion-exchange chromatography, hydrophilic interaction liquid chromatography (HILIC) is recommended. HILIC is similar to NPLC; the difference comes from the mobile phase, which is composed of polar and/or aprotic organic solvent miscible in water.

HILIC combines highly hydrophilic stationary phase, such as unmodified, chemically modified silica (amino, cyano, amide, diol, zwitterionic and/or polar polymer), with mobile phase, such as any polar organic solvents (acetonitrile, methanol,

isopropanol, etc.) and aprotic solvents (tetrahydrofuran, THF and dioxane) containing small amount of water (<5 %).

In comparison to RPLC, in HILIC the strong solvent is water; thus, the gradient starts with high percentage of organic solvent, and the separation begins after increasing the percentage of aqueous phase, ending with high percentage of this phase.

In metabolomics, polar compounds are preferentially determined by HILIC, mainly due to the mobile-phase compatibility to mass spectrometry, in comparison to NPLC, high-detection sensitivity and high amount of organic solvent in the mobile-phase composition, which allows desolvation in the mass spectrometer ionisation process. In addition, a low back pressure due to minor viscosity of organic solvent used in the organic phase and the possibility to inject the sample dissolved in organic solvents are advantageous [43]. However, a disadvantage in comparison to RPLC is low injection volume capability, which reduces the sensitivity and generates wider peaks, resulting in low-peak resolution. Applications of RPLC and HILIC in metabolomics field have been reported in the literature [44, 45].

About HILIC, the retention mechanisms are based by partitioning, hydrogen bonds and electrostatic interactions [46]. Partitioning phenomenon is considered as a liquid-liquid separation system because polar compounds are partitioned between two liquid phases, acetonitrile-rich phase and water-rich layer immobilised on the hydrophilic stationary phase, which is formed by strong water attraction by the polar groups present in the stationary phase. The lower the water concentration in mobile phase, the higher the water layer. For mobile phases with water at concentration lower than 20 %, an excess of water adsorbed at stationary phase is created; as a result, a multilayer is formed. At high concentrations of organic phase, only water layer is adsorbed (closer) to the stationary phase, because only molecules of water can interact with the residual silanols of the stationary phase. Therefore, the polar compounds are solubilised in the water layer; thus, the more hydrophilic compounds are, the more solubilised in the water layer, the more retained in the stationary phase. On the other hand, the elution of polar compounds happens at high concentrations of water due to the water layer reduction. A condition for the water layer formation is the presence of small amount of water (lower than 5 %) in the organic phase [42].

In addition to the partitioning mechanism, hydrogen bonds (H-bond) of the analyte with stationary phase are considered. This mechanism was related by LSER (linear solvation energy relationship) and proved that compounds that contain functional groups with hydrogen donor or hydrogen acceptor can interact with stationary phase by H-bond [47]. This separation mechanism is pronounced if low quantity of water is used in mobile phase.

Finally, electrostatic interactions of basic and acid compounds can be considered. The ionised residual silanols, present in the stationary phase, which were not eliminated by hydrogen bond or steric effects, can interact with basic compounds, such as occurs in ion-exchange mechanism, whereas acid compounds (e.g. acid phosphopeptides) can be retained due to hydrophilic interaction or electrostatic repulsion hydrophilic interaction chromatography (ERHLIC), generated when enough organic solvent is employed in a mobile phase [48].

The extension of each mechanism depends mainly on the kind of stationary and mobile phase. As explained above, the percentage of water in the organic phase affects directly the water layer formation; as a result, the partitioning mechanism is affected. However, the salt present in aqueous phase decreases the electrostatic attraction or repulsion. On the other hand, analyte retention occurs due to the water layer increasing, which is affected by presence of salt in mobile phase. In addition, the pH of the mobile phase influences the extension of retention because the pKa of the analytes are affected; thus, the charge of the analytes can be altered [42].

As the stationary phase is classified in neutral, charged and zwitterionic, the separation mechanism extension can be affected distinctly. Neutral stationary phase contains polar functional groups, which are uncharged at pH 3 up to 8; thus, in this case, hydrophilic interaction is the main separation mechanism [42]. However, if charged stationary phase is employed, which contains polar groups, and is dependent on the pH of the mobile phase, the extension of the separation mechanism is based on ion-exchange mechanism and hydrophilic partitioning [49].

Finally, the three separation mechanisms (partition, electrostatic and hydrogen bond) can occur in a zwitterionic stationary phase, mainly due to the presence of zwitterionic ligands (sulfobetaines) that have negative and positive charges. Many applications of HILIC have been presented in the literature and specific to the clinical field, as well as the use of RPLC and HILIC simultaneously [13, 50–53].

Currently, RPLC and HILIC columns with lower internal diameter (e.g, 1 mm) and lower length have been proposed to improve metabolite detection. Therefore, an introduction of instruments that are able to operate at very high pressure – ultra-performance liquid chromatography (UPLC) – coupled to mass spectrometry has been proposed to improve metabolite detection. This technology allows an increased resolution, better sensitivity and ion suppression reduction. As a result, more analytes are detected in a sample in comparison with conventional HPLC. In addition, lower solvent consumption is observed, due to the low flow rate (150–250 μL min^{-1}), which is possible because of the internal column diameter reduction (e.g. 1 mm). This kind of columns allows the same linear velocities obtained in conventional columns (i.d. 2.1 mm). The use of UPLC technology for biofluid analysis is increasing [54, 55].

In order to enhance the column lifetime, as to RPLC as to HILIC, sample preparation is necessary for clinical samples. The introduction of the samples without sample preparation can clog the chromatographic column and in the mass spectrometer can cause ion suppression and ion source deterioration because of the presence of salts in relatively high concentrations. The sample preparation method depends on the aim of the analysis. The main challenge in sample treatment for LC-MS in clinical applications is to isolate a compound and/or compound classes (targeted metabolomics) from the biofluids or the variety of concentrations of different classes (untargeted metabolomics). Different sample cleanup techniques for clinical applications, for targeted metabolomics, have been reviewed [56]. However, for untargeted analysis, as the interest is to obtain the most information as possible, only the protein precipitation procedure is recommended. More details about sample preparation for clinical metabolomics can be found in Chap. 2.

4.3.1 MS Ionisation Techniques Used in LC-MS

Mass spectrometry is considered a powerful technique to detect metabolites and has high sensitivity to quantify them, as mentioned in Sect. 4.2. Metabolites present in biofluids must be ionised before MS detection. Hence, an interface between LC and MS is necessary, and this is usually the ion source. Three ionisation techniques are commonly used in metabolomics: electrospray ionisation (ESI), atmospheric pressure chemical ionisation (APCI) and atmospheric pressure photoionisation (APPI). Amongst them, ESI is the most widely used ionisation technique for untargeted metabolomics, mainly because ESI is considered a soft ionisation technique, e.g. generates ions with little or no fragmentation, which can help in the identification of unknown metabolites, and also for being able to ionise compounds at an extensive polarity range. This approach has been extensively implemented in clinical metabolomics [57, 58]. In addition, ESI requires no sample derivatisation, ionises a large mass range of compounds and is suitable for non-volatile and polar compound analyses with high sensitivity.

Electrospray is a process that creates or transfers intact ions from solution to gas phase at atmospheric pressure. A spray is induced after a high-voltage application on a capillary. Thus, charged droplets are formed and emerge from a Taylor cone (jet of charged particles formed by an electric field after voltage application) to the mass analyser after solvent evaporation, which release ions with multiple charges (z) to the gas phase, even though metabolomics has interest in $z = 1$.

Other ionisation techniques are employed for LC-MS systems in metabolomics, such as APCI and APPI. Similar to ESI, APCI and APPI are considered soft ionisation techniques. Both can be used in positive and negative modes and are employed for non-polar and thermally stable compounds (e.g. lipids), respectively.

Each ionisation technique is able to detect compounds with different polarities and mass ranges; therefore, complementary information is obtained if more than one ionisation technique is employed. In addition, to provide complementary metabolomic information, analyses in the positive and negative ionisation modes are required. Comparing ESI analysis alone, an increase of 20 % of detected metabolites in human blood was observed when APCI analysis was employed [59]. Some MS instruments can provide analysis in positive and negative modes and detect compounds with different polarities at the same mass spectrometry analysis [60].

4.4 Gas Chromatography Coupled to Mass Spectrometry in Clinical Metabolomics

Gas chromatography (GC) is a well-established analytical technique used routinely in metabolomics platforms when coupled with some types of mass spectrometers. This instrumentation is based on separation of volatile (or made more volatile by chemical derivatisation) and thermally stable metabolites. The chemical classes that

can be considered naturally volatile are [61]: ketones, aldehydes, alcohols, esters, furan and pyrrole derivatives, heterocyclic compounds, sulphides, some lipids, isocyanates, isothiocyanates and hydrocarbons with 1–12 carbons. The classes that can be made volatile by derivatisation are sugars, sugar phosphates, amino acids, lipids, peptides, long-chain alcohols, amines, amides, alkaloids, sugar alcohols and organic acids [61].

When the sample is injected in the GC instrument (usually a volume of 0.5–2 µL), the metabolites are volatilised immediately, and an inert gas (helium or nitrogen) carries the sample from a heated injection system (200–250 °C) to a coated capillary column. Capillary column coating is composed by a solid or liquid phase (called as stationary phase), where an inert gas flows through, carrying the metabolites. The capillary column is maintained within an oven, which has a fine temperature control. As the temperature increases, the compounds that have low boiling points elute from the column sooner than those that have higher boiling points. Columns with varying chemical composition of stationary phases have been utilised in clinical metabolomic analysis; however, DB-5MS (chemically bonded with 5 % diphenyl cross-linked 95 % dimethylpolysiloxane) columns or columns with equivalent stationary phase (HP-5MS and RTX-5MS) are more commonly used [62, 63]. The separation of metabolites occurs by a chemical interaction between the stationary phase (polarity) and the temperature (volatility). A result of this complex mechanism of separation when coupled to a mass spectrometer is high-resolution power and sensitivity. However, the main limitation of the use of gas chromatography is the derivatisation step, necessary for some metabolite classes. For many applications of GC, the metabolites are not naturally volatile, and it is necessary to add this sample preparation step, which is time consuming, low throughput and can be error prone, introducing variability and artefacts [64].

The derivatisation step in metabolomic studies is usually made by a two-step process, which encompasses oximation and silylation/chloroformate reagent. The oximation protects ketone functional groups from keto-enol tautomerism and decarboxylation and inhibits the ring formation of reducing sugars [65]. Derivatisation based on silylation is the most popular; nevertheless, reagents based on chloroformate are also used for clinical metabolites [66]. Derivatisation based on silylation includes the following agents: N,O-bis-(trimethylsilyl)-trifluoroacetamide (BSTFA), N-methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) and N-methyl-N-tertbutyldimethylsilyltrifluoroacetamide (MTBSTFA) that can react with nearly all polar functional groups, including –COOH, −OH, −NH and –SH, increasing the compound volatility by replacing the active hydrogen with an alkylsilyl group [65, 67]. Before the silylation step, it is necessary to check for complete dryness of the sample to avoid hydrolysis that is typical of these reagents, whose efficiency indeed depends upon the preservation of the anhydrous environment [66]. Xiong et al. [68] used the oximation-silylation (with BSTFA reagent) method to derivative metabolites in urine samples, focusing to discriminate patients with phenylketonuria. The authors detected, simultaneously, amino acids, organic acids, carbohydrates, amides and fatty acids. Begley et al. [69] also used the oximation-silylation method, but with MSTFA reagent for untargeted metabolomics of human serum samples.

Derivatisation based on chloroformate has the advantage of being conducted in aqueous media. Zheng et al. [70] quantified, simultaneously, short-chain fatty acids and branched-chain amino acids using propyl chloroformate reagent in complex biological samples, including faeces, plasma and urine, from animal and human subjects. Other derivatisation reagents can also be used for specific metabolites on targeted studies [71, 72].

After the separation by GC, the metabolites enter into the mass spectrometer for identification and/or quantification. For this purpose, the molecule (metabolite) needs to be ionised, in order to be further separated according to the *m/z* ratio, which is detected and converted into electronic signal. The production of charged metabolites separated by GC occurs by two forms: electron ionisation (EI) or chemical ionisation (CI). EI is performed in a high-vacuum ion source (10^{-7} to 10^{-5} mbar, 200–250 °C) where the gas-phase molecules are bombarded by a fixed electron voltage, typically -70 eV [73]. This electron bombarding gives the sample molecules excess of energy, and many fragment ions are formed. Fragmentation pattern is characteristic to a particular molecule and therefore can be useful in determining the structure of the analyte, which is easily compared to available databases [62]. However, some compounds fragment completely and do not provide the molecular ion; thus, CI can be utilised as an alternative ionisation technique for these specific metabolites. In CI, a gas (methane or ammonia) reacts with the metabolite resulting in a charged molecule. CI is a relatively softer ionisation technique, producing spectra with reduced fragmentation when compared to EI [62]; in conjunction with exact mass, it can aid in the confirmation or identification of metabolites and hence potential biomarker candidates [74]. EI is the most commonly used ionisation technique in GC-MS-based metabolomic studies, generating reproducible mass spectra with minimal instrument-to-instrument variations. CI can produce molecular ions for some volatile compounds that do not give molecular ions in EI. The main use of CI is to confirm the molecular mass of some compounds [74, 75].

After the fragmentation in the ion source, the fragments are separated in mass analysers and detected. The generated mass spectrum is characteristic for each molecule and can be compared with mass libraries available in databases together with the retention time (or retention index) reported from chromatograms. Data treatment involves computational tools used to validate the metabolite identification and can usually be made using software packages provided by the instrument's manufacturer or using free access softwares. Nowadays, there are many software packages available [76, 77]. In general, after analysis, data treatment includes data preprocessing, data processing, statistical analysis and validation.

Data preprocessing includes the following important steps: deconvolution, library-based identification and alignment [65]. Deconvolution is a very important step for an untargeted metabolomic study, extracting only valuable signals from a complex mixture of signals in the chromatogram, and, in addition, performs treatment of noise, correction for baseline drift and extraction of co-eluting components [65]. Amongst the softwares, the AMDIS (Automated Mass Spectra Deconvolution and Identification System) presents free access and is commonly used for this task by many research groups. Subsequently, metabolite identification by GC-MS is usually

made by library databases. The two most used libraries for GC-MS metabolite identification are Fiehn RTL library and NIST (National Institute of Standards and Technology) mass spectra database. Fiehn library has the advantage of including retention index and retention time information that can be compared with experiments performed following the same analytical method [78]. This additional information increases the reliability in the analysis and decreases false positive results. The availability of spectral libraries to metabolite identification is the main advantage of the GC-MS over LC-MS methods for clinical metabolomic studies. Therefore, few preprocessing software packages and available databases are used for both low- and high-resolution data. The study performed by Peralbo-Molina et al. [79] represents the difficulty of identifying compounds using databases: NIST database (used by the authors) does not contain high-resolution MS information as provided by the TOF analyser; thus, more steps in the data processing were necessary to validate the identification of each compound. Finally, alignment is needed for correcting retention time differences between chromatographic runs and matching data from different samples. This preprocessing is based on retention time and mass spectra similarity: compounds from different samples are compared together by computing a spectral score from their respective spectra [65, 77]. Some software packages are available to perform alignment alone or as consequential process of peak finding and deconvolution. Bioinformatics tools for GC-MS data preprocessing were compared recently on the literature [80].

After data preprocessing, it is necessary to explore the data and to remove any mystifying information, meanly for untargeted metabolomics [81]. This step is known as data processing, which includes discard of contaminants (derivatisation reagents, compounds from column bleeding, etc.), normalisation, scaling and transformation [65]. Finally, the statistical analysis (univariate or multivariate data analysis) is applied to the conclusion of the research goal.

The mass analysers that can be coupled to gas chromatography are single quadrupole (Q), triple quadrupole (QqQ), time of flight (TOF) and ion trap (IT). Recent GC-MS applications using these analysers can be seen in Table 4.2. The coupling between Orbitrap (OT) and GC was recently developed [106]; nevertheless, until the present moment, it was only applied for plant extract [107], and there are expectations that it will be applied to clinical metabolomics.

Although the new developments of fast and high-resolution power mass analysers coupled to GC, the tendency of this chromatography separation technique in metabolomics includes an increase in the use of comprehensive two-dimensional gas chromatography (GCxGC) using columns with different selectivities, thus enhancing the resolution power. GCxGC combining two columns with orthogonal separation characteristics yields a multiplicative increase in peak capacity [108]. A thermal- or pressure-based modulator is located between the columns to periodically focus the effluent from the first column and transfer it to the second column in small concentrated segments [109]. GCxGC has been applied to targeted and untargeted metabolomics in the clinical area and is preferentially coupled to TOF-MS [110, 111].

Table 4.2 Applications of GC-MS in clinical metabolomics

Column	GC-MS system	Metabolites	Biological matrix	Reference
DB-5MS	GC-QTOF MS	Organic acids, esters, alcohols, lipids	Exhaled breath condensate	[79]
DB-5	GC-TOF MS	Amino acids, esters, steroids, sugar, organic acids, sugar alcohols	Serum	[82]
VF-1 ms and HP-1[a]	GC-Q MS and GC-IT MS	Steroids	Urine	[83]
AT-5MS	GC-IT MS	Steroids	Plasma	[84]
DB-5MS	GC-QTOF MS	Alcohols, organic acids, esters, sugar alcohols, sugars, amino acids, piperidines, non-metal oxoanionic compounds, lipids	Human sweat	[85]
CP-SIL 8 CB	GC-Q MS	Organic acids, amino acids, sugars, nitrogenous bases, amides	Serum	[86]
Rtx5Sil-MS	GC-TOF MS	Amines, sugars, organic acids, amino acids, steroids, lipids, amino acids, non-metal oxoanionic compounds	Cerebrospinal fluid	[87]
DB-50[a]	GC-TOF MS	Untargeted analysis (no identification performed)	Plasma	[88]
Rtx5Sil-MS	GC-TOF MS	Sugars, amines, sugar alcohols, lipids, organic acids	Blood (plasma/ serum)	[89]
Rxi-1 ms	GC-TOF MS	Amines, alcohols, nucleosides, organic acids, nitrogenous bases, sugars	Saliva	[90]
RTX-5Sil MS[a]	GC-TOF MS	Amino acids, organic acids, sugars, lipids, amino acids, nucleosides, amines	Breast tissue	[91]
ZB-5MS	GC-QqQ MS	Amino acids, organic acids, sugar alcohols, sugars, amines, alcohols	Urine	[92]
TR-5MS	GC-QqQ MS	Organic acids	Urine	[93]
DB-5 + DB-17	GCxGC-TOF MS	Organic acids, amino acids, lipids	Serum	[94]
BPX-5 + BPX-50	GCxGC-TOF MS	Organic acids	Urine	[95]
DB-5 + DB-1701	GCxGC-TOF MS	Sugars, lipids, non-metal oxoanionic compounds	Plasma	[96]
DB-5MS and HP-5	GC-Q MS and GC-TOF MS	Amino acids, organic acids, vitamins, sugars, steroids	Plasma	[97]

(continued)

Table 4.2 (continued)

Column	GC-MS system	Metabolites	Biological matrix	Reference
DB-5MS	GC-Q MS	Amino acids, organic acids, amines	Urine	[98]
HP-5MS	GC-Q MS	Amino acids, sugars, amines, nucleosides, inorganic acids, organic acids, alcohols, amides, heterocyclic compounds, steroids	Mucosal tissue	[99]
DB-50[a]	GC-TOF MS	Nucleosides, organic acids, sugar alcohols, alcohols, sugars	Serum	[100]
DB-5MS	GC-TOF MS	Lipids, organic acids derivatives, amino acids, amines, sugars, organic acids, sugars, sugar alcohols	Serum	[101]
VF-5 ms	GC-IT MS	Lipids, amino acids, sugars, sugar alcohols, organic acids derivatives, heterocyclic compounds	Plasma	[102]
DB-5[a]	GC-Q MS	Pyrimidine bases	Urine	[103]
DB-1MS[a]	GC-Q MS	Amino acids, organic acids	Urine	[104]
DB-5MS	GC-IT MS	Amino acids, organic acids, esters, amides	Urine	[105]

[a]Cited in a reference therein

4.5 Conclusions and Future Prospects

The use of chromatographic techniques coupled to mass spectrometry is a perfect combination for clinical metabolomics field. When combining efficient separation techniques with high sensitivity and/or high-resolution mass spectrometers, the quantification/identification of metabolites contained in complex biological samples can be successfully achieved for targeted and untargeted clinical metabolomics.

LC-MS and GC-MS are the most employed analytical platforms in clinical metabolomics, and the tendency is that they will remain with this status, since many improvements have been developed in the last decade, both in chromatography (new stationary phases for columns, two-dimensional separations, etc.) and in mass spectrometry (mass analysers with increased resolution power, nanospray ion sources, etc.).

References

1. Becker S, Kortz L, Helmschrodt C, Thiery J, Ceglarek U. LC-MS-based metabolomics in the clinical laboratory. J Chromatogr B: Analyt Technol Biomed Life Sci. 2012;883–884:68–75.
2. Murray KK, Boyd RK, Eberlin MN, Langley GJ, Li L, Naito Y. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). Pure Appl Chem. 2013;85:1515–609.

3. De Hoffmann E, Stroobant V. Mass spectrometry – principles and applications. 3rd ed. vol. 29, Mass spectrometry reviews. England: John Wiley & Sons Ltd; 2007. p. 489.

4. Makarov A, Scigelova M. Coupling liquid chromatography to orbitrap mass spectrometry. J Chromatogr A. 2010;1217:3938–45.

5. Rodriguez-Aller M, Gurny R, Veuthey J-L, Guillarme D. Coupling ultra high-pressure liquid chromatography with mass spectrometry: constraints and possible applications. J Chromatogr A. 2012;1292:2–18.

6. March RE. Quadrupole ion traps. Mass Spectrom Rev. 2009;28:961–89.

7. Hernández F, Sancho JV, Ibáñez M, Abad E, Portolés T, Mattioli L. Current use of high-resolution mass spectrometry in the environmental sciences. Anal Bioanal Chem. 2012;403:1251–64.

8. Denisov E, Damoc E, Lange O, Makarov A. Orbitrap mass spectrometry with resolving powers above 1,000,000. Int J Mass Spectrom. 2012;325–327:80–5.

9. Forcisi S, Moritz F, Kanawati B, Tziotis D, Lehmann R, Schmitt-Kopplin P. Liquid chromatography-mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. J Chromatogr A. 2013;1292:51–65.

10. Ghaste M, Mistrik R, Shulaev V. Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics. Int J Mol Sci. 2016;17:816.

11. Kuehnbaum NL, Britz-Mckibbin P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. Chem Rev. 2013;113:2437–68.

12. Snyder LR, Kirkland JJ, Glajch JL. Practical HPLC method development. 2nd ed. New York: John Wiley & Sons, Inc; 1997. p. 767.

13. Rainville PD, Theodoridis G, Plumb RS, Wilson ID. Advances in liquid chromatography coupled to mass spectrometry for metabolic phenotyping. Trends Anal Chem. 2014;61:181–91.

14. García-Cañaveras JC, Jiménez N, Gómez-Lechón MJ, Castell JV, Donato MT, Lahoz A. LC-MS untargeted metabolomic analysis of drug-induced hepatotoxicity in HepG2 cells. Electrophoresis. 2015;36:2294–302.

15. Sheikh KD, Khanna S, Byers SW, Fornace A, Cheema AK. Small molecule metabolite extraction strategy for improving LC/MS detection of cancer cell metabolome. J Biomol Tech. 2011;22:1–4.

16. Willmann L, Schlimpert M, Hirschfeld M, Erbes T, Neubauer H, Stickeler E, et al. Alterations of the exo- and endometabolite profiles in breast cancer cell lines: a mass spectrometry-based metabolomics approach. Anal Chim Acta. 2016;925:34–42.

17. Shao X, Gao D, Wang Y, Jin F, Wu Q, Liu H. Application of metabolomics to investigate the antitumor mechanism of flavopiridol in MCF-7 breast cancer cells. J Chromatogr B: Analyt Technol Biomed Life Sci. 2016;1025:40–7.

18. Tsutsui H, Maeda T, Min JZ, Inagaki S, Higashi T, Kagawa Y, et al. Biomarker discovery in biological specimens (plasma, hair, liver and kidney) of diabetic mice based upon metabolite profiling using ultra-performance liquid chromatography with electrospray ionization time-of-flight mass spectrometry. Clin Chim Acta. 2011;412:861–72.

19. Raju B, Ramesh M, Borkar RM, Srinivas R, Padiya R, Banerjee SK. In vivo metabolic investigation of moxifloxacin using liquid chromatography/electrospray ionization tandem mass spectrometry in combination with online hydrogen/deuterium exchange experiments. Rapid Commun Mass Spectrom. 2012;26:1817–31.

20. Waybright TJ, Van QN, Muschik GM, Conrads TP, Veenstra TD, Issaq HJ. LC-MS in metabonomics: optimization of experimental conditions for the analysis of metabolites in human urine. J Liq Chromatogr Relat Technol. 2006;29:2475–97.

21. Zhao L, Xiong Z, Lu X, Zheng S, Wang F, Ge L, et al. Metabonomic evaluation of chronic unpredictable mild stress-induced changes in rats by intervention of fluoxetine by HILIC-UHPLC/MS. PLoS One. 2015;10:e0129146.

22. Li S, Lin H, Qu C, Tang Y, Shen J, Li W, et al. Urine and plasma metabonomics coupled with UHPLC- QTOF/MS and multivariate data analysis on potential biomarkers in anemia and hematinic effects of herb pair Gui-hong. J Ethnopharmacol. 2015;170:175–83.

23. Peng J, Chen YT, Chen CL, Li L. Development of a universal metabolome-standard method for long-term LC-MS metabolome profiling and its application for bladder cancer urine-metabolite- biomarker discovery. Anal Chem. 2014;86:6540–7.

24. Huo T, Xiong Z, Lu X, Cai S. Metabonomic study of biochemical changes in urinary of type 2 diabetes mellitus patients after the treatment of sulfonylurea antidiabetic drugs based on ultra-performance liquid chromatography/mass spectrometry. Biomed Chromatogr. 2015;29:115–22.

25. Fitian A, Nelson D, Liu C, Xu Y, Ararat M, Cabrera R. Integrated metabolomic profiling of hepatocellular carcinoma in hepatitis C cirrhosis through GC/MS and UPLC/MS-MS. Liver Int. 2014;34:1428–44.

26. Yao H, Shi P, Zhang L, Fan X, Shao Q, Cheng Y. Untargeted metabolic profiling reveals potential biomarkers in myocardial infarction and its application. Mol Biosyst. 2010;6:1061–70.

27. Zhang J, Bowers J, Liu L, Wei S, Gowda GAN, Hammoud Z, et al. Esophageal cancer metabolite biomarkers detected by LC-MS and NMR methods. PLoS One. 2012;7(1):e30181.

28. Huang Q, Yin P, Wang J, Chen J, Kong H, Lu X, et al. Method for liver tissue metabolic profiling study and its application in type 2 diabetic rats based on ultra performance liquid chromatography-mass spectrometry. J Chromatogr B: Analyt Technol Biomed Life Sci. 2011;879:961–7.

29. Yang Y, Cruickshank C, Armstrong M, Mahaffey S, Reisdorph R, Reisdorph N. New sample preparation approach for mass spectrometry-based profiling of plasma results in improved coverage of metabolome. J Chromatogr A. 2013;1300:217–26.

30. Godzien J, Ciborowski M, Whiley L, Legido-Quigley C, Ruperez FJ, Barbas C. In-vial dual extraction liquid chromatography coupled to mass spectrometry applied to streptozotocin-treated diabetic rats. Tips and pitfalls of the method. J Chromatogr A. 2013;1304:52–60.

31. Lu X, Bennet B, Mu E, Rabinowitz J, Kang Y. Metabolomic changes accompanying transformation and acquisition of metastatic potential in a syngeneic mouse mammary tumor model. J Biol Chem. 2010;285:9317–21.

32. Bi H, Krausz KW, Manna SK, Li F, Johnson CH, Gonzalez FJ. Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI-MS-based metabolomic analysis of adherent mammalian cancer cells. Anal Bioanal Chem. 2013;405:5279–89.

33. Ibáñez C, Simó C, García-Cañas V, Gómez-Martínez Á, Ferragut JA, Cifuentes A. CE/LC-MS multiplatform for broad metabolomic analysis of dietary polyphenols effect on colon cancer cells proliferation. Electrophoresis. 2012;33:2328–36.

34. Lin Y, Si D, Zhang Z, Liu C. An integrated metabonomic method for profiling of metabolic changes in carbon tetrachloride induced rat urine. Toxicology. 2009;256:191–200.

35. Michopoulos F, Gika H, Palachanis D, Theodoridis G, Wilson ID. Solid phase extraction methodology for UPLC-MS based metabolic profiling of urine samples. Electrophoresis. 2015;36:2170–8.

36. Spagou K, Wilson ID, Masson P, Theodoridis G, Raikos N, Coen M, et al. HILIC UPLC MS for exploratory urinary metabolic profiling in toxicological studies. Anal Chem. 2011;83:382–90.

37. Sen A, Knappy C, Lewis MR, Plumb RS, Wilson ID, Nicholson JK, et al. Analysis of polar urinary metabolites for metabolic phenotyping using supercritical fluid chromatography and mass spectrometry. J Chromatogr A. 2016;1449:141–55.

38. Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engström G, et al. A diabetes-predictive amino acid score and future cardiovascular disease. Eur Heart J. 2013;34:1982–9.

39. Tang H, Wang X, Xu L, Ran X, Li X, Chen L, et al. Establishment of local searching methods for orbitrap-based high throughput metabolomics analysis. Talanta. 2016;156–157:163–71.

40. Ebshiana AA, Snowden SG, Thambisetty M, Parsons R, Hye A, Legido-Quigley C. Metabolomic method: UPLC-q-ToF Polar and non-polar metabolites in the healthy rat cerebellum using an in-vial dual extraction. PLoS One. 2015;10:e0122883.

41. Denoroy L, Zimmer L, Renaud B, Parrot S. Ultra high performance liquid chromatography as a tool for the discovery and the analysis of biomarkers of diseases: a review. J Chromatogr B: Analyt Technol Biomed Life Sci. 2013;927:37–53.
42. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. Nat Protoc. 2010;5:1005–18.
43. Greco G, Letzel T. Main interactions and influences of the chromatographic parameters in HILIC separations. J Chromatogr Sci. 2013;51:684–93.
44. Ivanisevic J, Zhu Z, Plate L, Tautenhahn R, Chen S, Johnson CH, et al. Toward omic scale metabolite profiling: a dual separation mass spectrometry approach for coverage of lipids and central carbon metabolism. Anal Chem. 2013;85:6876–84.
45. Ramakrishnan P, Nair S, Rangiah K. A method for comparative metabolomics in urine using high resolution mass spectrometry. J Chromatogr A. 2016;1443:83–92.
46. Buszewski B, Noga S. Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique. Anal Bioanal Chem. 2012;402:231–47.
47. Abraham MH, Ibrahim A, Zissimos AM. Determination of sets of solute descriptors from chromatographic measurements. J Chromatogr A. 2004;1037:29–47.
48. Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal Chem. 2008;80:62–76.
49. Dinh NP, Jonsson T, Irgum K. Probing the interaction mode in hydrophilic interaction chromatography. J Chromatogr A. 2011;1218:5880–91.
50. Gika HG, Theodoridis GA, Vrhovsek U, Mattivi F. Quantitative profiling of polar primary metabolites using hydrophilic interaction ultrahigh performance liquid chromatography-tandem mass spectrometry. J Chromatogr A. 2012;1259:121–7.
51. Spagou K, Tsoukali H, Raikos N, Gika H, Wilson ID, Theodoridis G. Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. J Sep Sci. 2010;33: 716–27.
52. Cubbon S, Antonio C, Wilson J, Thomas-Oates J. Metabolomic applications of HILIC-LC-MS. Mass Spectrom Rev. 2010;29:671–84.
53. Zhang T, Watson DG. Evaluation of the technical variations and the suitability of a hydrophilic interaction liquid chromatography-high resolution mass spectrometry (ZIC-pHILIC-Exactive orbitrap) for clinical urinary metabolomics study. J Chromatogr B: Analyt Technol Biomed Life Sci. 2016;1022:199–205.
54. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011;6:1060–83.
55. Want EJ, Masson P, Michopoulos F, Wilson ID, Theodoridis G, Plumb RS, et al. Global metabolic profiling of animal and human tissues via UPLC-MS. Nat Protoc. 2013;8:17–32.
56. Leung KS, Fong BM. LC-MS/MS in the routine clinical laboratory: has its time come? Anal Bioanal Chem. 2014;406:2289–301.
57. Iadarola P, Fumagalli M, Bardoni AM, Salvini R, Viglio S. Recent applications of CE- and HPLC-MS in the analysis of human fluids. Electrophoresis. 2016;37:212–30.
58. Mil'man BL, Zhurkovich IK. Mass spectrometric analysis of medical samples and aspects of clinical diagnostics. J Anal Chem. 2015;70:1179–91.
59. Nordstrom A, Want E, Northen T, Lehtio J, Siuzdak G. Multiple ionization mass spectrometry strategy reveals the complexity of metabolomics. Anal Chem. 2008;80:421–9.
60. Yuan M, Breitkopf S, Yang X, Asara J. A positive/negative ion–switching, targeted mass spectrometry–based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. Nat Protoc. 2012;7:872–81.
61. Trivedi DK, Iles RK. Do not just do it, do it right: urinary metabolomics -establishing clinically relevant baselines. Biomed Chromatogr. 2014;28:1491–501.
62. Pasikanti KK, Ho PC, Chan ECY. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. J Chromatogr B: Analyt Technol Biomed Life Sci. 2008;871:202–11.

63. Qiu Y, Su M, Liu Y, Chen M, Gu J, Zhang J, et al. Application of ethyl chloroformate derivatization for gas chromatography–mass spectrometry based metabonomic profiling. Anal Chim Acta. 2007;583:277–83.

64. Little JL. Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. J Chromatogr A. 1999;844:1–22.

65. Mastrangelo A, Ferrarini A, Rey-Stolle F, García A, Barbas C. From sample treatment to biomarker discovery: a tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. Anal Chim Acta. 2015;900:21–35.

66. Qiu Y, Reed D. Gas chromatography in metabolomics study. In: Guo X, editor. Advances in gas chromatograph. 2014. p. 83–101. http://www.intechopen.com/books/advances-in-gas-chromatography/gas-chromatography-in-metabolomics-study.

67. Abbiss H, Rawlinson C, Maker GL, Trengove R. Assessment of automated trimethylsilyl derivatization protocols for GC-MS-based untargeted metabolomic analysis of urine. Metabolomics. 2015;11:1908–21.

68. Xiong X, Sheng X, Liu D, Zeng T, Peng Y, Wang Y. A GC/MS-based metabolomic approach for reliable diagnosis of phenylketonuria. Anal Bioanal Chem. 2015;407:8825–33.

69. Begley P, Francis-McIntyre S, Dunn WB, Broadhurst DI, Halsall A, Tseng A, et al. Development and performance of a gas chromatography-time-of-flight mass spectrometry analysis for large-scale nontargeted metabolomic studies of human serum. Anal Chem. 2009;81:7038–46.

70. Zheng X, Qiu Y, Zhong W, Baxter S, Su M, Li Q, et al. A targeted metabolomic protocol for short-chain fatty acids and branched-chain amino acids. Metabolomics. 2013;9:818–27.

71. Chiu HH, Tsai SJ, Tseng YJ, Wu MS, Liao WC, Huang CS, et al. An efficient and robust fatty acid profiling method for plasma metabolic studies by gas chromatography–mass spectrometry. Clin Chim Acta. 2015;451:183–90.

72. Kloos D, Lingeman H, Mayboroda OA, Deelder AM, Niessen WMA, Giera M. Analysis of biologically-active, endogenous carboxylic acids based on chromatography-mass spectrometry. Trends Anal Chem. 2014;61:17–28.

73. Garcia A, Barbas C. Gas chromatography–mass spectrometry (GS-MS)-based metabolomics. In: Metabolic profiling – methods and protocols. New York: Humana Press & Springer Science Business Media; 2011. p. 191–204.

74. Major H, Williams R, Wilson A, Wilson ID. A metabonomic analysis of plasma from zucker rat strains using gas chromatography/mass spectrometry and pattern recognition hilary. Rapid Commun Mass Spectrom. 2006;20:3295–302.

75. Gao X, Pujos-Guillot E, Sébédio JL. Development of a quantitative metabolomic approach to study clinical human fecal water metabolome based on trimethylsilylation derivatization and GC/MS analysis. Anal Chem. 2010;82:6447–56.

76. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M. Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. Curr Bioinforma. 2012;7:96–108.

77. Katajamaa M, Orešič M. Data processing for mass spectrometry-based metabolomics. J Chromatogr A. 2007;1158:318–28.

78. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, 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;81:10038–48.

79. Peralbo-Molina A, Calderón-Santiago M, Priego-Capote F, Jurado-Gámez B, De Luque Castro MD. Development of a method for metabolomic analysis of human exhaled breath condensate by gas chromatography–mass spectrometry in high resolution mode. Anal Chim Acta. 2015;887:118–26.

80. Niu W, Knight E, Xia Q, McGarvey BD. Comparative evaluation of eight software programs for alignment of gas chromatography–mass spectrometry chromatograms in metabolomics experiments. J Chromatogr A. 2014;1374:199–206.

81. Godzien J, Ciborowski M, Angulo S, Barbas C. From numbers to a biological sense: how the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LCMS. Electrophoresis. 2013;34:2812–26.

82. Che N, Cheng J, Li H, Zhang Z, Zhang X, Ding Z, et al. Decreased serum 5-oxoproline in TB patients is associated with pathological damage of the lung. Clin Chim Acta. 2013;423:5–9.
83. Christakoudi S, Cowan DA, Taylor NF. Steroids excreted in urine by neonates with 21-hydroxylase deficiency. 3. Characterization, using GC-MS and GC-MS/MS, of androstanes and androstenes. Steroids. 2012;77:1487–501.
84. Tsakalof AK, Gkagtzis DC, Koukoulis GN, Hadjichristodoulou CS. Development of GC-MS/MS method with programmable temperature vaporization large volume injection for monitoring of 17β-estradiol and 2-methoxyestradiol in plasma. Anal Chim Acta. 2012;709:73–80.
85. Delgado-Povedano MM, Calderón-Santiago M, Priego-Capote F, De Luque Castro MD. Development of a method for enhancing metabolomics coverage of human sweat by gas chromatography–mass spectrometry in high resolution mode. Anal Chim Acta. 2016;905:115–25.
86. Kohashi M, Nishiumi S, Ooi M, Yoshie T, Matsubara A, Suzuki M, et al. A novel gas chromatography mass spectrometry-based serum diagnostic and assessment approach to ulcerative colitis. J Crohns Colitis. 2014;8:1010–21.
87. Motsinger-Reif AA, Zhu H, Kling MA, Matson W, Sharma S, Fiehn O, et al. Comparing metabolomic and pathologic biomarkers alone and in combination for discriminating Alzheimer's disease from normal cognitive aging. Acta Neuropathol Commun. 2013;1:28.
88. Kenny LC, Dunn WB, Ellis DI, Myers J, Baker PN, Kell DB. Novel biomarkers for preeclampsia detected using metabolomics and machine learning. Metabolomics. 2005;1:227–34.
89. Miyamoto S, Taylor S, Barupal D, Taguchi A, Wohlgemuth G, Wikoff W, et al. Systemic metabolic changes in blood samples of lung cancer patients identified by gas chromatography time-of-flight mass spectrometry. Metabolites. 2015;5:192–210.
90. Mueller DC, Piller M, Niessner R, Scherer M, Scherer G. Untargeted metabolomic profiling in saliva of smokers and nonsmokers by a validated GC-TOF-MS method. J Proteome Res. 2014;13:1602–13.
91. Budczies J, Denkert C, Müller BM, Brockmöller SF, Klauschen F, Györffy B, et al. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue - a GC-TOFMS based metabolomics study. BMC Genomics. 2012;13:334.
92. Struck-Lewicka W, Kordalewska M, Bujak R, Yumba Mpanga A, Markuszewski M, Jacyna J, et al. Urine metabolic fingerprinting using LC-MS and GC-MS reveals metabolite changes in prostate cancer: a pilot study. J Pharm Biomed Anal. 2015;111(2015):351–61.
93. Monteleone M, Naccarato A, Sindona G, Tagarelli A. A reliable and simple method for the assay of neuroendocrine tumor markers in human urine by solid-phase microextraction-gas chromatography-triple quadrupole mass spectrometry. Anal Chim Acta. 2013;759:66–73.
94. Asiago VM, Alvarado LZ, Shanaiah N, Gowda GAN, Owusu-Sarfo K, Ballas RA, et al. Early detection of recurrent breast cancer using metabolite profiling. Cancer Res. 2010;70:8309–18.
95. Kouremenos KA, Pitt J, Marriott PJ. Metabolic profiling of infant urine using comprehensive two-dimensional gas chromatography: application to the diagnosis of organic acidurias and biomarker discovery. J Chromatogr A. 2010;1217:104–11.
96. Li X, Xu Z, Lu X, Yang X, Yin P, Kong H, et al. Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabonomics: biomarker discovery for diabetes mellitus. Anal Chim Acta. 2009;633:257–62.
97. Ranjbar MRN, Luo Y, DiPoto C, Varghese RS, Ferrarini A, Zhang C, et al. GC-MS based plasma metabolomics for identification of candidate biomarkers for hepatocellular carcinoma in Egyptian cohort. PLoS One. 2015;10:e0127299.
98. Xie GX, Chen TL, Qiu YP, Shi P, Zheng XJ, Su MM, et al. Urine metabolite profiling offers potential early diagnosis of oral cancer. Metabolomics. 2012;8:220–31.
99. Wu H, Xue R, Lu C, Deng C, Liu T, Zeng H, et al. Metabolomic study for diagnostic model of oesophageal cancer using gas chromatography/mass spectrometry. J Chromatogr B: Analyt Technol Biomed Life Sci. 2009;877:3111–7.
100. Dunn WB, Broadhurst DI, Deepak SM, Buch MH, McDowell G, Spasic I, et al. Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. Metabolomics. 2007;3:413–26.

101. Jia W, Jia W, Lu J, Zhou J, Bao Y, Chen T, et al. Serum metabolic signatures of fulminant type 1 diabetes. J Proteome Res. 2012;11:4705–11.
102. Vallejo M, García A, Tuñón J, García-Martínez D, Angulo S, Martin-Ventura JL, et al. Plasma fingerprinting with GC-MS in acute coronary syndrome. Anal Bioanal Chem. 2009;394:1517–24.
103. Kuhara T, Ohse M, Inoue Y, Shinka T. Five cases of β-ureidopropionase deficiency detected by GC/MS analysis of urine metabolome. J Mass Spectrom. 2009;44:214–21.
104. Smuts I, van der Westhuizen FH, Louw R, Mienie LJ, Engelke UFH, Wevers RA, et al. Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach. Metabolomics. 2013;9:379–91.
105. Tian JS, Peng GJ, Wu YF, Zhou JJ, Xiang H, Gao XX, et al. A GC-MS urinary quantitative metabolomics analysis in depressed patients treated with TCM formula of Xiaoyaosan. J Chromatogr B: Anal Technol Biomed Life Sci [Internet]. 2015. Available from: http://dx.doi.org/10.1016/j.jchromb.2015.12.026.
106. Peterson AC, Hauschild JP, Quarmby ST, Krumwiede D, Lange O, Lemke RAS, et al. Development of a GC/quadrupole-orbitrap mass spectrometer, part I: design and characterization. Anal Chem. 2014;86:10036–43.
107. Peterson AC, Balloon AJ, Westphall MS, Coon JJ. Development of a GC/quadrupole-orbitrap mass spectrometer, part II: new approaches for discovery metabolomics. Anal Chem. 2014;86:10044–51.
108. Almstetter MF, Oefner PJ, Dettmer K. Comprehensive two-dimensional gas chromatography in metabolomics. Anal Bioanal Chem. 2012;402:1993–2013.
109. Bertsch W. Two-dimensional gas chromatography. Concepts, instrumentation, and applications – part 2: comprehensive two-dimensional gas chromatography. J High Resolut Chromatogr. 2000;23:167–81.
110. Tranchida PQ, Costa R, Donato P, Sciarrone D, Ragonese C, Dugo P, et al. Acquisition of deeper knowledge on the human plasma fatty acid profile exploiting comprehensive 2-D GC. J Sep Sci. 2008;31:3347–51.
111. Dettmer K, Almstetter MF, Appel IJ, Nürnberger N, Schlamberger G, Gronwald W, et al. Comparison of serum versus plasma collection in gas chromatography - mass spectrometry-based metabolomics. Electrophoresis. 2010;31:2365–73.

Chapter 5
Strategies Involving Mass Spectrometry Combined with Capillary Electrophoresis in Metabolomics

Karina Trevisan Rodrigues, Zuzana Cieslarová, Marina Franco Maggi Tavares, and Ana Valéria Colnaghi Simionato

Abstract This chapter focuses on the important contribution of CE-MS in metabolomics, describing the nature of CE-MS coupling and the technical improvements that have led to the interfaces used in modern instrumentation. Moreover, it will discourse how the variety of electrolyte compositions and additives, which has conferred CE the exceptional selectivity of its multiple separation modes, has been handled to allow interfacing with MS without compromising ionization efficiency and the spectrometer integrity. Finally, the methodologies of CE-MS in current use for metabolomics will be discussed in detail. To verify the scope of CE-MS in clinical metabolomics, a myriad of representative applications has been compiled.

Keywords Metabolomics • Targeted metabolomics • Untargeted metabolomics • Clinical metabolomics • CE-MS • Capillary electrophoresis • CE-MS interfaces

K.T. Rodrigues • M.F.M. Tavares (✉)
Institute of Chemistry, University of São Paulo (USP), 05513-970 São Paulo, SP, Brazil
e-mail: mfmtavar@iq.usp.br

Z. Cieslarová
Department of Analytical Chemistry, Institute of Chemistry, University of Campinas (UNICAMP), 13083-970 Campinas, SP, Brazil

A.V.C. Simionato (✉)
Department of Analytical Chemistry, Institute of Chemistry, University of Campinas (UNICAMP), 13083-970 Campinas, SP, Brazil

National Institute of Science and Technology in Bioanalytics – INCTBio,
Campinas, SP, Brazil
e-mail: avsimionato@iqm.unicamp.br

Abbreviations

ACE	Affinity capillary electrophoresis
ACS	Acute coronary syndrome
AD	Alzheimer's disease
ANN	Artificial neural network
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
APFO	Ammonium perfluorooctanoate
BGE	Background electrolyte
CE	Capillary electrophoresis
CEC	Capillary electrochromatography
p-CEC	Pressure-assisted CEC
CESI	Capillary electrophoresis integrated to electrospray ionization
CGE	Capillary gel electrophoresis
CIEF	Capillary isoelectric focusing
CITP	Capillary isotachophoresis
CKD	Chronic kidney disease
CMC	Critical micelle concentration
CoA	Co-enzyme A
CRPS	Complex regional pain syndrome
CSF	Cerebrospinal fluid
CTAB	Cetyltrimethylammonium bromide
CVA	Canonical variate analysis
CZE	Capillary zone electrophoresis
DA	Discriminant analysis
DS	Dextran sulfate
EOF	Electroosmotic flow
ESI	Electrospray ionization
FAB	Fast atom bombardment
FDR	False discovery rate
FTICR	Fourier-transform ion cyclotron resonance
GC	Gas chromatography
HCA	Hierarchical cluster analysis
HF	Hydrogen fluoride
HILIC	Hydrophilic interaction liquid chromatography
^{1}H NMR	Proton nuclear magnetic resonance
HPLC	High-performance liquid chromatography
I.D.	Inner diameter
IS	Internal standard
IT	Ion trap
kNN	k-Nearest neighbors
LC	Liquid chromatography

LDA	Linear discriminant analysis
LLE	Liquid-liquid extraction
LOO-CV	Leave-one-out cross validation
M	Molecule
MALDI	Matrix-assisted laser desorption ionization
MCR-ALS	Multivariate curve resolution alternating least squares
MEKC	Micellar electrokinetic chromatography
MLR	Multiple linear regression
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSc	Multiple sclerosis
MSI	Multi-segment injection
O.D.	Outer diameter
OPLS	Orthogonal projections to latent structures
PB	Polybrene
PCA	Principal component analysis
PF	Partial filling
PKD	Polycystic kidney disease
PLS	Partial least square
PVA	Polyvinyl alcohol
PVS	Poly(vinyl sulfonate)
Q	Quadrupole
QC	Quality control
QqQ	Triple quadrupole
qRT-PCR	Quantitative real-time polymerase chain reaction
RPLC	Reversed-phase liquid chromatography
SAM	Significance analysis of microarrays
SDS	Sodium dodecyl sulfate
SHL	Sheath liquid
SPE	Solid-phase extraction
SRM	Selected reaction monitoring
SVM	Support vector machine
TEA	Triethylamine
TEDETAMA-co-HPMA	Copolymers of N-(2-hydroxypropyl) methacrylamide (HPMA) and the dendronic methacrylic monomer 2-(3-(Bis(2-(diethylamino)ethyl)amino)propanamido) ethyl methacrylate (TEDETAMA, derived from N,N,N′,N′-tetraethyldiethylenetriamine, TEDETA)
TOF	Time of flight
UPLC	Ultra-performance liquid chromatography
UTI	Urinary tract infection
UV	Ultraviolet radiation
VIP	Variable importance in the projection
VUR	Vesicoureteral reflux

5.1   The Niche of CE-MS in Metabolomics

Metabolomics, the analysis of the entire set of metabolites (metabolome), or a partial set of selected metabolites and/or substrates, expressed by an organism in preestablished conditions, via comparative experiments, has been the subject of irrefutable attention by the scientific community, since its inception in the late 1990s by Nicholson et al. and Fiehn [1, 2]. Both formats, untargeted (hypothesis generating) and targeted (hypothesis driven) metabolomics, are possible and have helped characterizing systemic responses of organisms to disease, pharmaceutical intervention, and dietary modulation [3–6]. Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy [7, 8] and mass spectrometry (MS) hyphenated with high- or ultra-performance liquid chromatography (HPLC or UPLC) and gas chromatography (GC) are the analytical platforms with prevalent use in the characterization of the metabolome [9–14]. A plethora of applications [15] with natural product-related [16, 17], nutritional [18–20], pharmaceutical [21], and clinical [22–26] importance have been compiled periodically. The choice of analytical technology applied in such studies is typically dependent upon the assessed class of chemical compounds, the cost of analysis, ease of sample preparation, and the requirement for sensitivity, specificity, and robustness. No single method enables complete coverage of the holistic metabolic information, and increasingly, metabolomics studies are adopting more than one analytical platform to augment the number of identified metabolites.

The particularities of metabolomics within the context of systems biology, as well as a general workflow of metabolomics studies from experimental design to biological validation, have been discussed thoroughly in Chap. 1. It is important to detail here the extent by which different analytical platforms approach the metabolome contents and how capillary electrophoresis is inserted in this context.

NMR has been the precursor technique for metabolomics and made a relevant contribution in the variety of application areas cited and referenced so far. This is mostly due to suitable performance characteristics, such as robustness, ease of data acquisition, and fairly wide metabolic coverage [1, 7, 8]. However, sensitivity and spectrum complexity have been issues in NMR metabolomics studies.

Nowadays, the high selectivity and sensitivity offered by MS platforms allowed MS to have conquered a sizeable niche in metabolomics, especially when the mass analyzer is hyphenated up front to a separation instrument [14]. Temporal separation of metabolites prior to detection is a desirable feature when complex matrices such as biological fluids and tissues are assessed.

GC-MS has been comprehensively explored for metabolomics since the very beginning [2, 14], with early studies in the context of plant metabolomics [16]. Although GC-MS is suited to assess the volatile portion of the metabolome, sample derivatization schemes [27] aiming primarily at volatility enhancement allowed GC

to reach a rather polar fraction in water-rich biofluids; for instance, carboxylic acids, amino acids, and biogenic amines can all be analyzed simultaneously in a single chromatographic run and ionization mode. Moreover, the high specificity associated with the resulting adducts allows the use of low-resolution mass spectrometers and the building of dedicated spectra libraries for compound identification [28]. Nevertheless, necessary sample derivatization schemes are time-consuming tasks and have limited the application of GC-MS to clinical protocols where only a small set of samples is under consideration.

Liquid chromatography-mass spectrometry (LC-MS) has been the premier technique in metabolomics for many years [9–14], despite the fact that to achieve the same metabolic coverage NMR does, multiple column chemistries must be screened. There are plenty of systematic studies where the information acquired from the more traditional reversed-phase (RPLC-MS) to the recently revisited hydrophilic interaction (HILIC-MS) modes is combined to promote a more thorough metabolic coverage (from nonpolar and/or moderately polar metabolites up to the ionic/polar ones) [29, 30]. The completion of human serum and urine metabolomes is a good example of the complementary information NMR and hyphenated MS analytical platforms offer [31, 32].

Considering the orthogonal separation mechanism provided by capillary electrophoresis (CE), it has emerged as a promising complementary technique to both liquid and gas chromatography for metabolic profiling of biological fluids as an impressive series of periodic review articles attest [33–56]. Intrinsic characteristics, such as high efficiency and resolution power, rapid analyses, and, most importantly, the ability to assess, without derivatization, the most polar and/or ionic compounds in the metabolome, have placed CE in an advantageous position. This chapter will therefore give a comprehensive overview of the state of the art in CE-MS technology, describing the methodologies in use for metabolomics and compiling representative applications of CE-MS in clinical metabolomics.

5.2   CE-MS

5.2.1   Onset and Pioneer Work

Although MS has currently achieved remarkable capacity to screen the composition of complex samples, in order to obtain relevant information about any biological system in a comprehensive manner as metabolomics studies do, it is recommendable to couple MS with different separation techniques and benefit from the three-dimensional information the hyphenated system imparts (retention and/or migration time, peak intensity, and mass-to-charge ratio).

Within the context of coupling separation techniques to MS, CE-MS was the last to be established, and interfacing the two platforms followed a timeline. While GC and HPLC were firstly registered around the 1950s and 1960s [57, 58], the first reports on electrophoresis effectively performed on capillary tube dimensions were registered in 1981, by Jorgenson and Lukacs [59, 60]. Before that, some authors had published electrophoretic separation on "quasi-capillary" dimensions, namely, Hjérten (using 300 μm i.d. capillary for the separation of inorganic ions, nucleotides, and proteins), Virtanen (using 200–500 μm i.d. capillaries), and Everaerts and collaborators, who first reported a completely automated CE system using 100 μm i.d. capillaries [61–63].

Capillary zone electrophoresis (CZE) is the simplest and most commonly used CE mode due to the straightforwardness of background electrolyte (BGE) composition, principle of separation, and broad application to the analysis of diverse samples, containing from small ions to large biomolecules [64]. In CZE, analytes are separated according to differences in electrophoretic mobilities, which are dependent on the molecule/species charge-to-radius ratio and the medium viscosity. Neutral analytes are thus not separated by this mode, constituting one of the CZE main drawbacks. To overcome such limitations and to expand CE applicability, other CE modes have been developed, such as micellar electrokinetic chromatography (MEKC, where micelles are used as carriers to assess primarily the separation of neutral compounds) [65], capillary isoelectric focusing (CIEF, where separation of amphiprotic substances is conducted in a pH gradient) [66–68], capillary isotachophoresis (CITP, where discontinuous leading and terminating electrolytes are used to separate small molecules and ions) [69], capillary gel electrophoresis (CGE, which uses gels or entangled polymers to assess large molecules and polymers) [70], capillary electrochromatography (CEC, where packed capillary columns are used to explore additional solute-stationary phase interactions) [71], and affinity capillary electrophoresis (ACE, which explores biospecific interactions) [72], among other modes. In fact, the versatility of performing almost all different modes (CEC and certain formats of CGE are a few exceptions) in the same capillary format and in the same equipment, only requiring alteration of the BGE composition, constitutes one of the major advantages of CE as a separation technique. Additional characteristics of CE include high resolution and efficiency (a million plates can be achieved), low consumption of BGE (few μL per run), small sample volume (few nL per run), and relatively fast separations (less than 5 min in favorable cases).

The most frequently used detection scheme available in almost all commercial CE equipments is based on absorption of UV-visible radiation (CE-UV). The UV detector is usually built on-capillary. By removing a narrow portion of the polyimide that coats externally the capillary, a detection window of tenths of millimeters is created. Any compound containing a chromophore group that passes the detection window will absorb the UV-visible radiation focused on the capillary and give a signal. However, the optical length available for absorption is the capillary inner diameter (usually 50 or 75 μm). Comparing such dimensions with those presented by HPLC detection cells (in the order of cm), allied to the reduced sample volume

introduced into the CE capillary (in the order of nL), puts in evidence the reduced concentration sensitivity posed by CE-UV systems.

The analysis of biological samples generally requires the use of sensitive, selective, and universal detectors. CE-MS coupling has arisen as a valuable alternative to overcome sensitivity and also selectivity issues associated with the UV detector, since the MS detector provides online spectral information; in addition, it is universal, and rather sensitive, depending on the interface used in the coupling. However, two main issues must be addressed when coupling CE to MS, namely, the separation mode and the interface design.

The CE-MS coupling was registered for the first time in 1987 by Smith and collaborators [73], constituting the first description of what is now known as sheathless interface. In their pioneer work, a capillary electrophoresis system (operated under CZE mode) was coupled to a quadrupole mass spectrometer using electrospray ionization (ESI) for the analysis of quaternary ammonium salts. In their instrumental arrangement, the cathode (or low voltage end) of the CE capillary was inserted into a stainless steel capillary in order to establish both the CE and the ESI electrical circuits. The capillary inlet was immersed in a BGE reservoir, and an electroosmotically induced flow allowed the CE effluent to be directly introduced into the MS. Nitrogen gas was used as drying gas in countercurrent to the CE effluent to assist droplet desolvation. Higher separation efficiency was obtained by decreasing the sample plug and concentration. However, several restrictions were imposed by this first CE-MS arrangement: low flow requirements for spray stability, limited BGE composition, and issues related to the capillary preparation process, such as the need of several steps for metal deposition and erosion of the deposited metal, requiring replacement after a few days of operation. By addressing these problems, in the following year, Smith and col. published a new manuscript, where some improvements were reported [74]. The CE capillary cathode end was again inserted into a stainless steel tube, and few millimeters of the capillary were protruded outside. In addition, silver vapor was used to produce the metal deposit at the capillary end, providing a system with better mechanical strength and extended lifetime. As a result, improved efficiencies of ESI sampling and ion transmission were achieved. A broader range of compounds, such as amino acids, polypeptides, quaternary ammonium salts, and water-soluble vitamins, was analyzed, presenting better separation efficiency than reported previously. In the same year, Smith and col. described a completely new and improved ESI interface, where the contact at the CE terminus was replaced by a thin sheath of flowing liquid [75]. It was the birth of the sheath liquid interface. With this new arrangement, a qualitative improvement in ESI stability was reached, and more importantly, no special treatment was required to establish the electrical contact at the capillary end, allowing easy replacement of the capillary. This design also constituted the basis for implementation of other CE modes. It is important to recognize the merit and contribution of Smith's research group, who proposed in 1 year apart both CE-MS interface designs, which evolved into today's modern instrumentation [76].

Figure 5.1 depicts a schematic representation of the variety of CE mechanisms, ionization modes, and types of mass analyzers reported so far in CE-MS coupling.
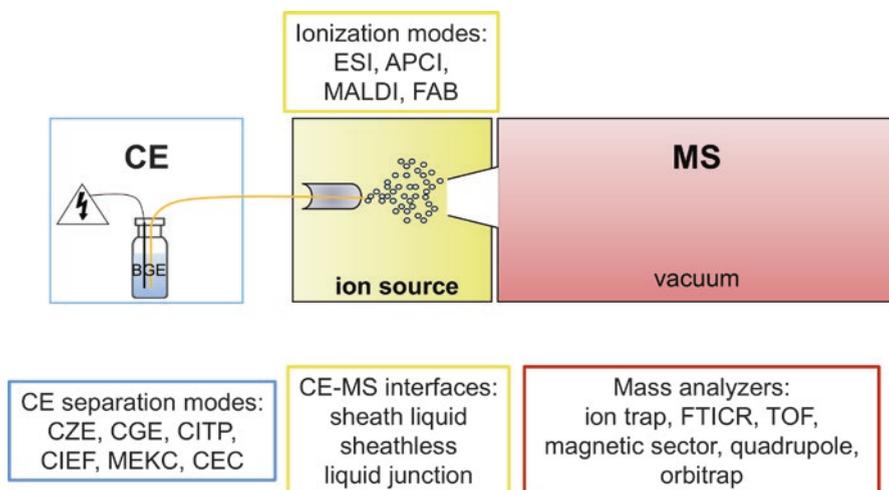
Fig. 5.1 Schematic diagram of CE-ESI-MS systems

CZE is undoubtedly the most commonly used separation mode in the CE-MS coupling, due to the easy manipulation of BGE composition (vide Table 5.1 for examples). In general, a simple buffer solution composed of volatile or semi-volatile acids or bases and corresponding salts is employed, e.g., formic or acetic acids, ammonia, and/or ammonium formate or acetate. However, the use of such simple BGE compositions restricts the pH range in which CE-MS separations can be performed. BGE additives may be required to improve the separation quality when resolution is compromised. The use of organic modifiers and cyclodextrins has been invoked, although the latter is known to hinder ionization efficiency and to contaminate the ion source and/or the mass analyzer ion optics, restricting its applicability.

MEKC has also been coupled to MS and provide the concurrent separation of neutral and ionic compounds, which interact with micelle compartments and/or surface to different extents. The solute McGown volume and the solute's ability to interact with electrolyte components via hydrogen bonding seem to be the determinant factors that explain retention. The surfactants commonly used in MEKC (SDS, CTAB, bile salts, etc.), which are added to the BGE in a concentration above their critical micelle concentrations (CMC), are generally nonvolatile species. Thus, depending on the concentration used, they might cause ionic suppression at the interface and contamination of the ion optics or capillary clogging if introduced into the MS system [147]. To circumvent these problems, the addition of volatile surfactants to the BGE has been recommended. Moreno-González et al. have employed ammonium perfluorooctanoate (APFO) as a semi-volatile surfactant for the separation of amino acids in human urine by MEKC-MS [148]. Separation of 20 amino acids (including leucine and isoleucine) with detectability in the ng mL^{-1} range was achieved with a BGE composed of 150 mmol L^{-1} APFO aqueous solution, adjusted to pH 9.0 with 14.2 mol L^{-1} ammonium hydroxide. At these pH

Table 5.1 Representative applications of CE-MS in clinical metabolomics

Ref.	Application	Matrix	Metabolomics	Metabolites	BGE	SHL	MS	Sample preparation	Data analysis	Other method
[77]	Acute coronary syndrome	Serum	Untargeted	Acetylcarnitines, amino acids	$0.8\ mol\ L^{-1}$ HFor in 10% MeOH	$4\ \mu L$ HFor in 50% MeOH	ESI(+)-TOF	LLE, ultracentrifugation	t-test, PCA, PLS-DA, OPLS-DA	LC-MS/MS
[78]	Aging	TTD mice urine	Untargeted	Amino acids and derivatives and acetylspermidine	$2\ mol\ L^{-1}$ HFor in 20% MeOH	50% MeOH with 0.1% HFor	ESI(+)-TOF	Mixed with MeOH, H_2O, and BGE; centrifuged	XCMS, PCA, PLS-DA	–
[79]	Alzheimer's disease *Progression*	CSF	Untargeted	Choline, arginine, histidine, dimethy-L-arginine, carnitine, creatine, valine, serine, and proline	$0.5\ mol\ L^{-1}$ HFor at pH 1.8	IPOH/H_2O (50%)	ESI(+)-TOF	Internal standards added to the CSF followed by ultrafiltration	Data analysis (Bruker Daltonics), Mzmine, PCA, LOO-CV, LDA, CVA	–
[80]	Alzheimer's disease *Progression*	Serum	Untargeted	Polar metabolites	$0.8\ mol\ L^{-1}$ HFor in 10% MeOH	$1\ mmol\ L^{-1}$ HFor in 50% MeOH	ESI(+)-TOF	LLE, ultracentrifugation	PLS-DA, VIP	–
[81]	Cancer *Bladder*	Urine	Untargeted	Amino acids and derivatives	$0.8\ mol\ L^{-1}$ HFor in 10% MeOH	$1\ mmol\ L^{-1}$ HFor in 50% MeOH	ESI(+)-TOF	Dilution in H_2O (1:5)	ANOVA, OPLS-DA	LC-QTOF-MS
[82]	Cancer *Breast, mouth, and pancreas*	Saliva	Untargeted	Amino acids and derivatives, carboxylic acids, carnitine, amines, bile acid	$1\ mol\ L^{-1}$ HFor	50% MeOH with $0.5\ \mu mol\ L^{-1}$ reserpine	ESI(+)-TOF	Dilution in H_2O (9:1) with IS	MassHunter (Agilent Technologies), XCMS, Douglas-Peucker algorithm, Steel-Dwass test, ANN, PCA, SVM, MLR	–
[83]	Cancer *Colon*	Cell	Untargeted	Amino acids and derivatives, amines, nucleoside, sugar	$1\ mol\ L^{-1}$ HFor	IPOH/H_2O (50%)	ESI(+)-TOF	LLE, ultrafiltration and lyophilization	XCMS, Welch t-test, Student's t-test	LC-MS (HILIC and RP)

(continued)

Table 5.1 (continued)

Ref.	Application	Matrix	Metabolomics	Metabolites	BGE	SHL	MS	Sample preparation	Data analysis	Other method
[84]	Cancer *Colon and stomach*	Tumor tissue	Targeted	Metabolites involved in glycolysis, pentose phosphate pathway, TCA, urea cycles, and amino acid and nucleotide metabolisms	1 mol L⁻¹ HFor; 50 mmol L⁻¹ AmAc, pH 8.5 (SMILE (+)-coated capillary) and 50 mmol L⁻¹ AmAc, pH 7.5	50% MeOH with 0.5 μmol L⁻¹ reserpine; 50% MeOH with 1 μmol L⁻¹ reserpine and 5 mmol L⁻¹ AmAc in 50% MeOH	ESI(+/−)-TOF	LLE, ultrafiltration and lyophilization	Wilcoxon test, multiExperiment Viewer	–
[85]	Cancer *Lung and prostate*	Tumor and surrounding tissues	Untargeted	Several	Cations: 1 mol L⁻¹ HFor Anions: 50 mmol L⁻¹ AmAc, pH 8.5	Commercial SHL	ESI(+/−)-TOF	LLE, ultracentrifugation	PCA	nanoLC-MS/MS
[86]	Cancer *Stomach*	Urine	Untargeted	Amino acids	1 mol L⁻¹ HFor	0.1 % HFor in 50% MeOH	ESI(+)-IT	Centrifugation, filtration	Mann-Whitney, ANOVA Kruskal-Wallis test, PCA	–
[87]	Cancer *Prostate*	Urine	Targeted	Sarcosine, L-proline, l-cysteine, L-leucine, L-glutamic acid, and L-kynurenine	0.4% HFor in 50% MeOH and 0.2% HFor in 50% MeOH	–	ESI(+)-IT	SPE (Strata-X strong cation cartridge) and lyophilization	–	–
[88]	Chronic kidney disease	Plasma	Untargeted		1 mol L⁻¹ HFor; 50 mmol L⁻¹ AmAc, pH 8.5 (SMILE (+)-coated capillary)	5 mmol L⁻¹ AmAc in 50% MeOH	ESI(+/−)-TOF	LLE, ultrafiltration and lyophilization	Spearman's rank correlation, Likehood function	–

Ref	Application	Sample	Targeting	Analytes	BGE	Sheath liquid	Ionization	Sample preparation	Data analysis	Comparison
[89]	Complex regional pain syndrome	Urine	Untargeted	Amino acids and derivatives, carboxylic acids, amines	1 mol L⁻¹ HFor pH 1.8 (PB-PVS coating)	0.1% HFor in 50% MeOH	ESI(+)-TOF	Dilution 1:1 with BGE	XCMS, PCA, PLS-DA	MEKC
[90]	Data processing *Clustering algorithm*	Urine	Untargeted	Amino acids, organic acids, nucleotides	1 mol L⁻¹ HFor adjusted with NH₄OH to pH 2.4 (TEDETAMA-co-HPMA copolymer-coated capillary)	50% IPOH	ESI(−)-TOF	Filtration	Hierarchical agglomerative cluster analysis (new algorithms)	–
[91]	Data processing *Missing values*	Plasma	Untargeted	Several	0.8 mol L⁻¹ HFor in 10% MeOH	1 mmol L⁻¹ HFor in 50% MeOH	ESI(+)-TOF	LLE, ultracentrifugation	Zero, median, ½ minimum and kNN imputation methods, t-test, and Mann-Whitney test	–
[92]	Diabetes mellitus *Type 1*	Urine	Untargeted	Protein and amino acid metabolism	0.8 mol L⁻¹ HFor in 10% MeOH	1 mmol L⁻¹ HFor in 50% MeOH	ESI(+)-TOF	Dilution with H₂O (10x) and centrifugation	PCA, OPLS-DA, t-test	LC-MS (plasma)
[93]	Diabetes mellitus *Type 2*	Serum	Untargeted	Amino acids, citrulline, acetylcarnitine	0.8 mol L⁻¹ HFor in 10% MeOH	4 µL HFor in 50% MeOH	ESI(+)-TOF	LLE, ultracentrifugation	t-test, Mann-Whitney test, Mann-Kendall trend analysis	–
[94]	Dilated cardiomyopathy	Heart tissue	Untargeted	Charged metabolites	Cations: 1 mol L⁻¹ HFor Anions: 50 mmol L⁻¹ AmAc, pH 8.5	5 mmol L⁻¹ AmAc in 50% MeOH containing 0.1 µmol L⁻¹ hexakis	ESI(+/−)-TOF	Homogenization with MeOH	PCA, t-test	LC-MS/MS LC-TOF-MS

(continued)

Table 5.1 (continued)

Ref.	Application	Matrix	Metabolomics	Metabolites	BGE	SHL	MS	Sample preparation	Data analysis	Other method
[95]	Exercise training	Plasma	Untargeted	L-carnitine, glutathionyl-L-cysteine, hypoxanthine, O-acetyl-L-carnitine	1 mol L⁻¹ HFor in 15% ACN	0.1% HFor in 60% MeOH	ESI(+)-TOF	Dilution with 200 mmol L⁻¹ AmAc, ultrafiltration	t-test, PCA, HCA, PLS-DA, two-way ANOVA	–
[96]	Fatty liver disease *Nonalcoholic related*	Serum	Untargeted	Sulfated steroids	50 mmol L⁻¹ AmAc pH 8.5 (COSMO (+)-coated capillary)	5 mmol L⁻¹ AmAc in 50% MeOH containing 0.1 µmol L⁻¹ hexakis	ESI(−)-TOF	LLE, ultracentrifugation	Steel-Dwass test	LC-TOF-MS
[97]	Gastric injury *Aspirin induced*	Serum, and stomach tissue	Untargeted	TCA cycle, β-oxidation, collagen metabolism	Cations: 1 mol L⁻¹ HFor Anions: 50 mmol L⁻¹ AmAc, pH 8.5 (COSMO (+)-coated capillary)	5 mmol L⁻¹ AmAc in 50% MeOH containing 0.1 µmol L⁻¹ hexakis	ESI(+/−)-TOF	LLE, ultrafiltration	PCA, PLS-DA, ANOVA, Dunnett's test	–
[98]	Hepatitis	Tissue	Untargeted	Compounds related to glutathione biosynthesis	1 mol L⁻¹ HFor; 50 mmol L⁻¹ AmAc, pH 8.5 (SMILE (+)-coated capillary)	5 mmol L⁻¹ AmAc in 50% MeOH containing 20 µmol L⁻¹ PEPIS and 1 µmol L⁻¹ reserpine	ESI(+/−)-TOF	Centrifugation, filtration and lyophilization	Douglas-Peucker algorithm, Reijenga function	–
[99]	Hepato-cellular carcinoma	Serum	Untargeted	Amino acids, organic acids, amines, sugar phosphates	Cations: 1 mol L⁻¹ HFor Anions: 50 mmol L⁻¹ AmAc pH 8.5	5 mmol L⁻¹ AmAc in 50% MeOH containing 0.1 µmol L⁻¹ hexakis	ESI(+/−)-TOF	LLE, ultracentrifugation	PLS-DA, HCA, Wilcoxon Mann-Whitney test, FDR	–

Ref	Disease	Sample	Approach	Analytes	BGE	Sheath/additive	ESI	Sample preparation	Statistical analysis	Other technique
[100]	Huntington's disease *Progression*	Plasma	Untargeted	Protein metabolism, prostaglandins, thromboxanes, lipoxins, and leukotrienes	50 mmol L^{-1} HAc and 50 mmol L^{-1} HFor adjusted to pH 3.5 with ammonia	0.05% HFor in 60% IPOH	ESI(+)-TOF	Protein precipitation, ultracentrifugation, SPE online	MCR-ALS, PLS-DA	–
[101]	Hypercholesterolaemia *Diet induced*	Plasma	Untargeted	Amino acids and derivatives, fatty acids esters	0.8 mol L^{-1} HFor in 10% MeOH	1 mmol L^{-1} HFor in 50% MeOH	ESI(+)-TOF	LLE, ultracentrifugation	PCA, PLS-DA	LC-MS, GC-MS
[102]	Inborn errors of metabolism	Dried blood spot	Targeted	Amino acid and acylcarnitine	1.4 mol L^{-1} HFor; pH 1.8	1:1 MeOH/H$_2$O with 0.1% HFor	ESI(+)-IT	LLE, ultrafiltration and dilution with AmAc solution	–	–
[103]	Liver disease	Serum	Untargeted	γ-glutamyl, dipeptides, transaminases, and methionine sulfoxide	50 mmol L^{-1} AmAc, pH 8.5 (COSMO (+)-coated capillary)	5 mmol L^{-1} AmAc in 50% MeOH and 0.1 μmol L^{-1} hexakis	ESI(−)-TOF		Kruskal-Wallis test and Dunn's post-test, Mann-Whitney test, MLR	LC-TOF-MS and LC-MS/MS
[104]	Liver disease	Serum, and liver tissue	Targeted	γ-glutamyl peptides	200 mmol L^{-1} HAc pH 3.3	0.5 mmol L^{-1} AmAc in 50% MeOH	ESI(+)-QqQ	LLE, ultracentrifugation	–	LC-MS/MS
[105]	Liver injury *Alcohol related*	Plasma	Untargeted	Amino acids, guanidinosuccinate	Cations: 1 mol L^{-1} HFor Anions: 50 mmol L^{-1} AmAc, pH 8.5 (COSMO (+)-coated capillary)	Cations: 0.5 μmol L^{-1} reserpine in 50% MeOH Anions: 5 mmol L^{-1} AmAc in 50% MeOH containing 0.1 μmol L^{-1} hexakis	ESI(+/−)-TOF	LLE, ultracentrifugation	Linear regression analysis, Benjamini and Hochberg's FDR	–

(continued)

Table 5.1 (continued)

Ref.	Application	Matrix	Metabolomics	Metabolites	BGE	SHL	MS	Sample preparation	Data analysis	Other method
[106]	Lung injury *Ventilator induced*	Plasma	Untargeted	Organic amines, amino acids, and their derivatives, carnitines	0.8 mol L⁻¹ HFor in 10% MeOH	4 µL HFor in 50% MeOH	ESI(+)-TOF	SPE (phospholipids and proteins removal) Ultracentrifugation	t-test, PLS-DA	
[107]	Metabolic disorders	Urine and blood spots	Targeted	Carnitines, carboxylic acid, creatinine, and galactose	20 mmol L⁻¹ AmAc, pH 8.5	2 mmol L⁻¹ AmAc in 50% MeOH	ESI(−)-QqQ	Blood spots: solid-liquid extraction; urine: filtration (0.45 µm)	–	–
[108]	Metabolite profiling *Estrogen speciation*	Urine	Targeted	Estrogens	50 mmol L⁻¹ ammonium bicarbonate, pH 9.5	5 mmol L⁻¹ ammonium bicarbonate in 80% MeOH	ESI(−)-TOF	Dilution (10x)	–	–
[109]	Metabolite profiling *Isobaric labeling*	Urine	Targeted	Amine-containing metabolites	0.2% HFor in 50% MeOH	0.2% HFor in 50% MeOH	ESI(+)-QTOF	Ultrafiltration, labeling with 4-plex DiLeu	–	nanoLC-ESI-MS/MS
[110]	Metabolite profiling *Kynurenic pathway*	CSF	Targeted	Tryptophan metabolites	5 mmol L⁻¹ AmAc in 5% ACN pH 9.7	50% MeOH	ESI(+)-TOF	Diluted (5x), add 50% ACN	–	–
[111]	Metabolite profiling *Thiol speciation*	Plasma	Targeted	Thiols	1 mol L⁻¹ HFor, pH 1.8	MeOH/H₂O (60:40%) with 0.1% HFor	ESI(+)-IT	Diluted 3-fold with 200 mmol L⁻¹ AmAc, pH 5 and 20 mol L⁻¹ Ala-Ala	RRF, MLR, k-fold cross validation	–
[112]	Method optimization *Additives for BGE and SHL*	Urine	Targeted	Amino acids and derivatives, carboxylic acids, nucleoside, sugar phosphate, purine	25 mmol L⁻¹ TEA (pH 11.7) (PB-DS-PB-coated capillary)	50% MeOH with 5 mmol L⁻¹ TEA	ESI(−)-TOF	Centrifuged and mixed with BGE (1:1)	DataAnalysis (Bruker Daltonics)	–
[113]	Method optimization *Capillary coating*	CSF, plasma and urine	Untargeted	Organic acids, amino acids	1 mol L⁻¹ HFor pH 1.8 (PB-PVS-coated capillary)	50% MeOH with 0.1% HFor	ESI(+)-TOF	CSF and plasma: no preparation; urine: mixed with BGE and centrifuged	–	CE-UV

Ref	Method optimization	Sample	Mode	Analytes	BGE	Sheath liquid	ESI	Sample preparation	Software	Comparison
[114]	Method optimization *Capillary coating*	Urine	Untargeted/targeted	Nucleosides, amino acids, carboxylic acids, nucleotides, amines	1 mol L^{-1} HFor; Anion: 25 mmol L^{-1} AmAc, pH 9 (PB-PVS and PB-DS-PB coated capillaries for both modes)	50% MeOH with 0.1% HForand 50% MeOH with 0.1% concentrated NH$_4$OH	ESI(+/−)-TOF	Mixed with BGE (1:1) for cation analysis or with H$_2$O (1:1) for anion	–	CE-UV
[115]	Method optimization *Capillary coating*	Urine	Untargeted/targeted	Tyramine, dopamine, creatinine, hippuric acid, glutathione, proline betaine, and amino acids	1 mol L^{-1} HFor pH 2 (PB-DS-PB coating)	50% MeOH with 0.1% HFor	ESI(+)-TOF	Mixed with BGE (1:1) and centrifuged	XCMS, PLS-DA, PCA	UPLC-MS
[116]	Method optimization *Interface*	Urine	Untargeted	Amino acids and derivatives, amines, nucleic acids, and small peptides	1.7 molL^{-1} (10%) HAc	Sheathless	ESI(+)-TOF	Mixed with IS and H$_2$O (1:1:8) and ultrafiltered	MassHunter (Agilent Technologies), XCMS, Douglas-Peucker algorithm, Steel-Dwass test	–
[117]	Method optimization *Interface*	Urine	Untargeted/targeted	Amino acids and derivatives, carnitines, nucleosides, creatinine, vitamins	10% HAc	Sheathless	ESI(+)-TOF	Mixed with BGE (1:1) and centrifuged	DataAnalysis (Bruker Daltonics)	–
[118]	Method optimization *Interface*	Mouse CSF, plasma and urine	Untargeted/targeted	Purine, amino acids and derivatives, choline, creatinine	10% HAc	Sheathless	ESI(+)-TOF	Diluted with H$_2$O (1:1)	–	–
[119]	Method optimization *Overall conditions*	Urine	Targeted	Dopamine, norepinephrine, epinephrine, 3-methoxytyramine, normetanephrine, metanephrine	1% HAc, pH 2.8 (PVA-coated capillary)	MeOH:H$_2$O (75:25) with 0.1% HAc	ESI(+)-TOF	SPE on Oasis MXC cation-exchange cartridge	–	CE-UV

Table 5.1 (continued)

Ref.	Application	Matrix	Metabolomics	Metabolites	BGE	SHL	MS	Sample preparation	Data analysis	Other method
[120]	Method optimization *Overall conditions*	Urine	Untargeted/targeted	Nucleosides, amino acids, carboxylic acids, sugars, nucleotides, monoamine	50 mmol L⁻¹ HAc and 50 mmol L⁻¹ HFor at pH 2.5	MeOH:H₂O (80:20) with 0.1 % HAc and IPOH/ H₂O (60:40 %) with 0.5 % ammonia	ESI(+/−)-IT	Centrifugation, filtration and lyophilization	–	CE-UV
[121]	Method optimization *Overall conditions*	Plasma	Targeted	Carnitines	200 mol L⁻¹ ammonium formate, pH 2.5	MeOH:H₂O (70:30) with 0.1 % HFor	ESI(+)-IT	Deproteinized with cold ACN (1:5) and centrifuged	–	–
[122]	Method optimization *Pt needle sprayer*	Mouse liver tissue	Targeted	Metabolites in glycolysis, pentose phosphate, and TCA pathways	50 mmol L⁻¹ AmAc, pH 8.5 (COSMO (+) coated capillary)	5 mmol L⁻¹ AmAc in 50% MeOH with 0.1 µmol L⁻¹ hexakis	ESI(−)-TOF	LLE, ultrafiltration and lyophilization	MZmine	–
[123]	Method optimization *Overall conditions*	Urine	Targeted	Amino acids	1 mol L⁻¹ HFor	5 mmol L⁻¹ AmAc in 50% MeOH	ESI(+)-QqQ	Dilution with H₂O (1:5)	–	–
[124]	Migraine *Cortical spreading depression*	Plasma	Untargeted	Lysine, pipecolic acid	10 % HAc (neutrally coated capillary)	Sheathless	ESI(+)-TOF	Protein precipitation with ethanol	PCA, PLS-DA, OPLS-DA	LC-MS/MS
[125]	Model organism *Bacteria*	*Bacillus subtilis*cells	Targeted	Intermediates of glycolysis and the TCA cycle	50 mmol L⁻¹ AmAc, pH 9 (SMILE (+)-coated capillary)	5 mmol L⁻¹ AmAc in 50% MeOH	ESI(−)-IT	LLE, ultrafiltration and lyophilization	–	–
[126]	Model organism *Bacteria*	*Bacillus subtilis*cells	Targeted	Citrate isomers, nucleotides, dinucleotides, and coenzyme A compounds	50 mmol L⁻¹ AmAc, pH 7.5 (DB-1-coated capillary)	5 mmol L⁻¹ AmAc in 50% MeOH	ESI(−)-MS pressure assisted	LLE, ultrafiltration and lyophilization	–	–

Ref	Type	Organism/cells	Targeted/Untargeted	Analytes	BGE	Sheath/Interface liquid	ESI mode	Sample preparation	Data analysis	Other
[127]	Model organism Bacteria	Bacillus subtilis cells	Untargeted/targeted	Nucleosides, amino acids, organic acids, sugars, nucleotides, monoamine; compounds involved in glycolysis, TCA cycle, and pentose phosphate pathways	1 mol L^{-1} HFor; 50 mmol L^{-1} AmAc, pH 8.5 (SMILE (+)-coated capillary) and 50 mmol L^{-1} AmAc, pH 7.5	5 mmol L^{-1} AmAc in 50% MeOH	ESI(+/−)-IT	LLE, ultrafiltration and lyophilization	–	–
[128]	Model organism Bacteria	Escherichia coli cells	Untargeted	Nucleosides, amino acids, organic acids, sugars, nucleotides, monoamine	80% 20 mmol L^{-1} AmAc pH 9.5: 20% IPOH	Sheathless	ESI(−)-QIT	LLE, ultrafiltration and lyophilization	–	Direct infusion ESI-MS
[129]	Model organism Bacteria	Escherichia coli cells	Untargeted/targeted	Nucleosides, amino acids, carboxylic acids, nucleotides, amines, sugars	50 mol L^{-1} AmAc, pH 8.7 in 5% MeOH	20 mmol L^{-1} NH_4OH in 50% IPOH	ESI(−)-TOF	LLE, ultrafiltration and lyophilization	QuantAnalysis (Bruker Daltonics) and MZmine	GC-MS
[130]	Model organism Cell - single cell	Thalamic tissue	Untargeted	GABA	1% HFor	0.1% HFor in 50% MeOH	ESI(+)-QTOF	LLE	–	–
[131]	Model organism Cell	Colon adenocarcinoma HT-29 cell line	Untargeted	Polyamines pathway	3 mol L^{-1} HFor	50% IPOH	ESI(+)-TOF	LLE, ultracentrifugation	PCA, ANOVA	–
[132]	Model organism Cell	U-87 MG glioblastoma cell line	Untargeted	Organic acids, sugar phosphates, and nucleotides	10% HAc pH 2.2	Sheathless	ESI(−)-TOF	Lysis, LLE	–	–
[133]	Model organism Fish	Zebrafish embryo	Untargeted	Cationic metabolites	10% HAc	0.1% HFor in 50% MeOH or 0.1% HAc in 75% IPOH (flow-through micro-vial interface)	ESI(+)-QTOF	Mechanical lysis, ultrafiltration	–	–

(continued)

Table 5.1 (continued)

Ref.	Application	Matrix	Metabolomics	Metabolites	BGE	SHL	MS	Sample preparation	Data analysis	Other method
[134]	Model organism *Parasite*	*Fasciola hepatica* tissue	Untargeted/targeted	Organic amines, amino acids	0.8 mol L⁻¹ HFor at pH 1.8 in 20% MeOH	0.5% HFor in 70% MeOH	ESI(+)-QTOF	LLE	XCMS, in-house script Matlab,	UPLC-MS (RP and HILIC)
[135]	Multiple sclerosis *Effect of methionine enkephalin*	Glioma cell line (C6, RG2, H4, U251,U87)	Untargeted	Amino acids, glycylglycine	Cations: 1 mol L⁻¹ HFor Anions: 50 mmol L⁻¹ AmAc, pH 8.5	50% MeOH containing 0.1 µmol L⁻¹ hexakis	ESI(+/−)-TOF	LLE	HCA, SAM	qRT-PCR, flow cytometry analysis
[136]	Myopia	Vitreous humor	Untargeted	Methylation process metabolites, acetylcarnitines, amino acids	0.8 mol L⁻¹ HFor in 10% MeOH	1 mmol L⁻¹ HFor in 50% MeOH	ESI(+)-TOF	Dilution with H₂O (5x)	OPLS-DA	LC-MS
[137]	Neurodegenerative dementia	Serum, saliva	Untargeted	TCA cyclic metabolites, amino acids and derivatives, creatinine	1.7 mol L⁻¹ HAc	Sheathless	ESI(+)-TOF	Ultracentrifugation	PCA, PLS-DA	–
[138]	Osteoarthritis	Urine	Targeted	Amino acids	2 mol L⁻¹ HFor with 20% MeOH	50% IPOH with 0.1% HFor	ESI(+)-TOF	Spiked urine	PCA	–
[139]	Polycystic kidney disease	Plasma	Untargeted	Amino acids and derivatives, carboxylic acids, nucleoside, amines, carnitines	1 mol L⁻¹ HFor, and 50 mmol L⁻¹ AmAc, pH 8.5 (COSMO (+)-coated capillary)	50% MeOH with 0.1 µmol L⁻¹ hexakis	ESI(+)-TOF	LLE, ultrafiltration and lyophilization	Unpaired *t*-test	–

[140]	Sample handling *Multi-segment injection*	Plasma	Untargeted	Amino acids and derivatives	1 mol L^{-1} HFor with 15% ACN	MeOH:H$_2$O (60:40) with 0.1% HFor	ESI(+)-TOF	Dilution with AmAc (4x) and ultracentrifugation	–	–
[141]	Sample handling *Processing and storage*	Serum, plasma	Untargeted	Charged metabolites	Cations: 1 mol L^{-1} HFor Anions: 50 mmol L^{-1} AmAc, pH 8.5 (COSMO (+) capillary)	Cations: 50% MeOH containing 0.1 μmol L^{-1} hexakis Anions: 5 mmol L^{-1} AmAc in 50% MeOH containing 0.1 μmol L^{-1} hexakis	ESI(+/−)-TOF	LLE, ultrafiltration	PCA, *t*-test	–
[142]	Schizophrenia	Plasma	Untargeted	Creatine, betaine, organic acids, homocysteine	Cations: 1 mol L^{-1} HFor Anions: 50 mmol L^{-1} AmAc, pH 8.5 (COSMO (+)-coated capillary)	Cations: 0.5 μmol L^{-1} reserpine in 50% MeOH Anions: 5 mmol L^{-1} AmAc in 50% MeOH containing 0.1 μmol L^{-1} hexakis	ESI(+/−)-TOF	LLE, ultracentrifugation	Mann-Whitney test, step-wise DA	–
[143]	Sepsis	Rat lung tissue	Untargeted	Amino acids, amines, carnitines	0.8 mol L^{-1} HFor in 10% MeOH	4 μL HFor in 50% MeOH	ESI(+)-TOF	LLE, ultracentrifugation	*t*-test, PLS-DA	LC-MS, GC-MS
[144]	Urinary tract infection	Urine	Untargeted/ targeted	Amino acids and derivatives	1 mol L^{-1} HFor pH 1.8 (PB-PVS-coated capillary)	0.1% HFor in 50% MeOH	ESI(+)-TOF	Mixed with BGE and centrifuged	ANOVA, PLS-DA, VIP	–

(continued)

Table 5.1 (continued)

Ref.	Application	Matrix	Metabolomics	Metabolites	BGE	SHL	MS	Sample preparation	Data analysis	Other method
[145]	Vesicoureteral reflux	Urine	Targeted	Amino acids	0.80 mol L^{-1} HFor pH 1.96 in 15% MeOH	0.50% HFor in 60% MeOH	ESI(+)-IT	Dilution	–	–
[146]	Xenobiotic exposure	Urine	Untargeted	Paracetamol sulfate, glucoronide, cystein, and paracetamol glucoronide	20 mmol L^{-1} ionic strength HFor/ammonium formate buffer, pH 3 (PolyE-323-coated capillary) and 20 mmol L^{-1} ionic strength NH$_4$OH/AmAc buffer, pH 9	80% IPOH and 20% BGE	ESI(+/−)-QqQ	–	In-house developed Matlab procedures and PCA	–

All percentages and proportions are expressed in v/v

ACN acetonitrile, *AmAc* ammonium acetate, *HAc* acetic acid, *HFor* formic acid, *IPOH* isopropanol, *MeOH* methanol

and concentration conditions, a strong electroosmotic flow (EOF) is observed, the surfactant is totally deprotonated (perfluorooctanoic acid has a pK_a of 2.8), and APFO micelles are formed (CMC is 25 mmol L^{-1}). With the exception of lysine and arginine, at the pH range from 7 to 9, amino acids are either negatively charged or neutral, which enhances their interaction with micelles. The authors observed that analyte resolution under the optimized conditions is a result of micelle partitioning and electrophoresis. Therefore, the most negatively charged amino acids were attracted to the anode, presenting low mobilities, while the positively charged amino acids interact electrostatically with the micelle surface, also showing long migration times. Finally, this method presented an improved selectivity when compared to a standard CZE-MS method and required a simple dilution of the urine sample with BGE prior to introduction into the CE system.

An interesting strategy to couple MEKC to MS is to partially fill the CE capillary with a BGE containing surfactant, whereas the remaining portion of the capillary is filled with a regular BGE, compatible with the MS system. In the partial filling technique (PF-MEKC-MS), the analytes are then primarily separated by interactions with micelles in the surfactant BGE length and reach the CE-MS interface before the surfactant does, when the current is interrupted. Sirén et al. have developed both PF-MEKC-UV and PF-MEKC-MS methods for the separation of endogenous low-hydrophilic steroids in plasma and urine samples [149]. For the PF-MEKC-MS method, the micellar solution was composed of 29.3 mmol L^{-1} SDS and 1.1 mmol L^{-1} sodium taurocholate in 20 mmol L^{-1} ammonium acetate at pH 9.68. Analytes that partially co-migrated after passing the surfactant BGE length were resolved in the MS operated at selected reaction monitoring (SRM) mode, resulting in the separation of eight analytes within 9 min.

CEC-MS has also been successfully used for the analysis of clinical samples. CEC development aimed at joining the high efficiency of CE (since the mobile phase flows through the capillary by EOF action, instead of pump pressure) with the high selectivity and peak capacity offered by the stationary phase in liquid chromatography. Therefore, a stationary phase must be introduced into the CE capillary, which may be one of the main experimental challenges of the CEC technique. The literature reports some options to this task, such as slurry packing (which may form bubbles in the separation bed and demands insertion of frits into the capillary ends to retain the stationary phase), open-channel CEC (where the inner capillary wall is functionalized), and in situ polymerization of a monolithic phase, which is the preferred strategy. Blas and McCord have performed the analysis of urine samples by CEC-MS to quantify traces of ten benzodiazepines [150]. A monolith based on porous acrylate was used. An online pre-concentration step (stacking consisting of the injection of a large amount of sample – 15 min at 12 bar – dissolved in aqueous medium) and the use of a TOF mass analyzer were required to obtain high sensitivity and specificity. To ensure that only hydrophobic interactions between analytes and the monolith occur, analyses were performed at pH 7.0, using 5 mmol L^{-1} ammonium acetate as BGE. Under these conditions, analytes at 1 ng mL^{-1} in urine samples could be quantified. One of the main problems encountered in CEC is bubble formation due to Joule heating, which may cause column dry out and current

interruption. To circumvent these problems, pressure-assisted CEC (*p*-CEC), where an extra pressure flow matches the EOF, coupled with ESI-QTOF-MS via a sheathless interface has been proposed for metabolomics profiling of urine samples [151]. The optimized method was successfully applied in the contrast of lung cancer patients and healthy subjects. Among 16 discriminant metabolites, three glutamine conjugates, including phenylacetylglutamine, acylglutamine C8:1, and acylglutamine C6:1, were identified.

To our knowledge, other CE modes in CE-MS couplings, such as CIEF-MS and CGE-MS, have not been applied in clinical metabolomics studies, and they will not be covered in this chapter.

5.2.2   CE-MS Interfaces

An important aspect to be considered when coupling CE to MS is the interface itself. Although many ionization schemes have been tested to date (Fig. 5.1), electrospray ionization (ESI) has been the ionization mode of choice, since it transfers ionizable analytes from the liquid phase to the gas phase, and it allows the analyses of high molecular-mass molecules by inducing formation of multiple charges (reduced *m/z* values). The development of CE-ESI-MS interfaces has mirrored the established LC-ESI-MS couplings. However, the reduced CE flow and the CE electric circuit (which must be closed at the CE capillary outlet or at the MS entrance) had to be regarded.

There are three main configurations for coupling CE to MS: coaxial sheath liquid interface, liquid junction interface, and sheathless interface. The formers are also called microspray interfaces while the latter is referred as nanospray interface [55, 152–155]. Some of the modern interface couplings used in CE-MS technology are schematically represented in Fig. 5.2.

The coaxial sheath liquid interface has gained great acceptance in CE-MS applications because it promotes a good spray stability, resulting in great robustness (Fig. 5.2a, b). Basically, the CE capillary outlet is introduced into a concentric tube where a sheath liquid (SHL) is pumped at nano- to microliter min^{-1} range. A third concentric tube may be introduced in order to conduct a nebulizer gas, assisting in the spray formation. One main drawback of sheath liquid interfaces is that the electrophoretic effluent (typical flow rate between 10 and 300 μL min^{-1}) is mixed to the SHL (typically 1–10 μL min^{-1}) at the capillary end, leading to sample dilution and consequent reduced detection sensitivity (one to two orders of magnitude decrease, depending on the BGE pH). Optimization of BGE composition as well as SHL flow and composition must therefore be carefully investigated in order to improve the ESI process and overall system detectability [55]. Nevertheless, the selection of both BGE and SHL is restricted to volatile or semi-volatile compounds, and routinely used salts in CE separations, such as borate and phosphate, must be avoided. Mixtures of organic solvents –
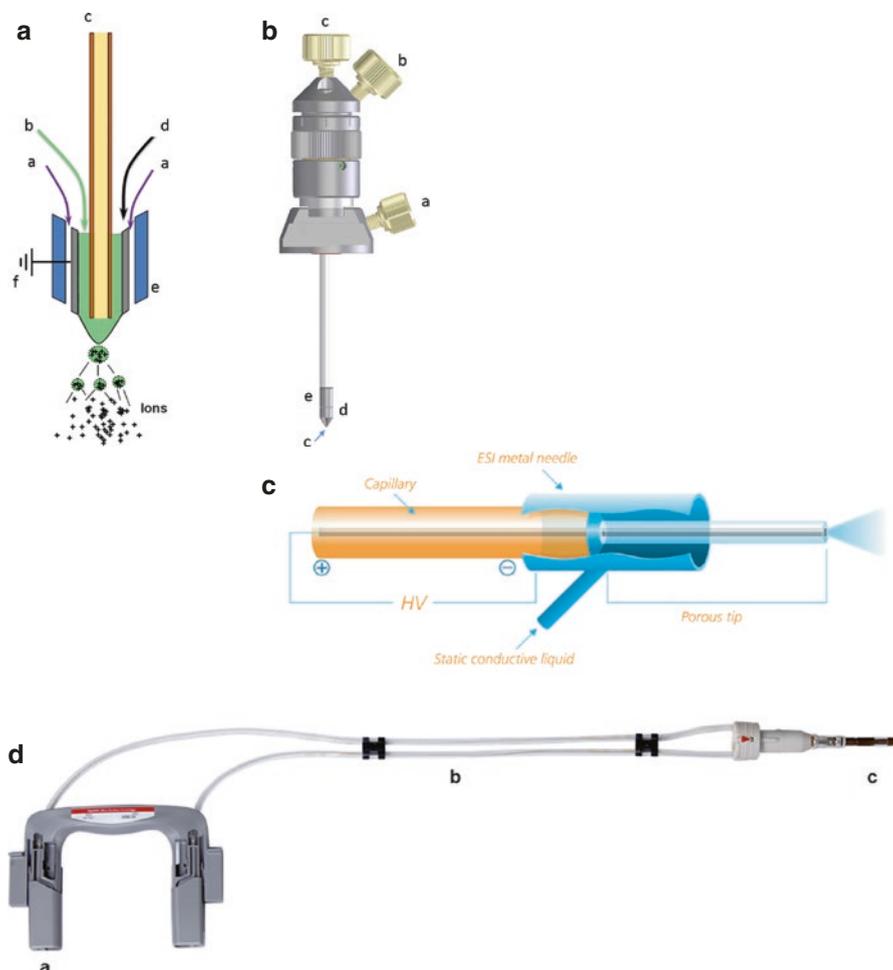
Fig. 5.2 Modern interfaces for CE-MS: coaxial sheath liquid (**a**, **b**) and sheathless (**c**, **d**) designs. Legends: (**a**) pictorial representation of Agilent coaxial sheath liquid CE-MS interface, (*a*) nebulizing gas, (*b*) sheath liquid, (*c*) CE capillary with BGE, (*d*) stainless steel spray needle with 0.4 mm i.d. and 0.5 mm o.d., (*e*) outer tube, and (*f*) ground connection; (**b**) engineering sketch of the coaxial sheath liquid CE-MS interface (graphics courtesy from Agilent Technologies) (reprinted with permission from Ref. [55]); (**c**) pictorial representation of Sciex sheathless CE-MS interface, (*a*) CE capillary inlet, (*b*) static conductive liquid capillary, and (*c*) sprayer porous tip; (**d**) engineering sketch of the CESI interface and cartridge (Photos are provided courtesy of AB Sciex Pte. Ltd. Operating as Sciex)

such as acetonitrile, methanol, and isopropanol – water, and weak acids or bases solutions (for positive and negative ESI, respectively) are generally considered for SHL [156]. Evaluation of the stability of CE-ESI-MS methods with sheath liquid interface may be derived from inter- and intraday precision measurements,

where the variation of peak areas of analytical standards spiked in body fluids is considered [157].

The sheath liquid interface position may be linear or orthogonal to the MS system. The main advantage offered by the latter geometry is that contamination or clogging of the MS inlet is less prone to occur and the choice of BGE composition is less critical, allowing the use of less volatile salts. Moreover, since the ESI voltage is applied at the MS entrance, charged species from the CE system are directed to the MS by electrostatic interaction, leading to higher detectability than in linear interfaces [158]. In fact, this is the most suitable and robust way to isolate the CE and the ESI electrical circuits.

The liquid junction interface is also based on a system supported by a sheath liquid [159]. However, mixing of BGE and SHL occurs far from the MS entrance, within a reservoir. Actually, the CE capillary ends inside this reservoir, and, in the opposite side, an electrospray needle is positioned within a distance ranging from 10 to 25 μm. With this geometry, the CE and the ESI electrical circuits operate individually, and the BGE selection may be performed independently of the MS restrictions. In addition, replacement of the ESI needle may be easily accomplished and does not affect the CE capillary. However, there are three main disadvantages that cause the scarce application of this interface: (i) the right alignment between the CE capillary and the ESI needle is laborious, (ii) the dead volume within the reservoir leads to band broadening (with consequent loss of separation efficiency), and finally, (iii) bubble formation on the CE capillary outlet often occurs, due to electrolysis reactions, resulting in current drop.

A recent variation of sheath liquid interfaces has been proposed by Chen and collaborators, named flow-through micro-vial interface [160]. The main characteristic of this interface is that the electrical circuits and flow rate requirements of the separation and ionization processes are decoupled. The authors have used a stainless steel hollow needle with optimized geometry to surround the CE capillary end. Therefore, the inner side of the needle works as the CE outlet vial, while the outer side is used as the ESI emitter. The CE capillary end is inserted into the needle until its outer diameter meets the dimension of the inner side of the needle. Both needle and CE capillary are connected to a tee union, where a second capillary is orthogonally attached in order to deliver the SHL into the needle. Typical flow rates of the SHL are as low as 0.1 mL min^{-1}, which reduces considerably the dilution of the CE effluent at the capillary end, when compared to regular sheath liquid interfaces, improving sensitivity. Another characteristic of the proposed interface is the possibility of using capillaries with any type of surface modification (such as neutral-coated capillaries for protein analysis, for instance) or even no pretreatment. The performance of the flow-through micro-vial interface comparatively to the conventional sheath liquid interface was evaluated by Lindenburg et al. in the profiling of cationic metabolite standards, exhibiting a fivefold improvement in terms of detection limits [133].

Sheathless interfaces transfer directly the CE effluent into the MS system, avoiding sample dilution and, consequently, present the best detectability among the CE-MS interfaces [161]. The main requirement of such interface is to close the CE

electrical circuit at the capillary end and simultaneously to afford electrical potential to the ESI. Considering that coaxial sheath flow interfaces have been developed after LC-MS interfaces, the possibility to develop an interface exclusively used in CE-MS coupling has arisen the interest of many research groups. For this reason, the literature reports several different ways to couple CE to MS by a sheathless interface focusing on creating a distinct ESI electrical contact. Application of a conductive coating to the emitter tip, joining a conductive emitter tip to the CE capillary, insertion of a wire into the CE capillary end, and positioning a metal sleeve around a porous etched CE capillary wall are among the many propositions [162, 163]. Although sheathless interfaces present better detectability due to the absence of a sheath liquid, allow closer positioning of the CE capillary to the MS (increasing the effective analyte mass transfer), and exhibit improved ionization and droplet desolvation, it still poses some limitations. The disadvantages of sheathless CE-MS interfaces comprise: (i) The absence of commercially available apparatus (except for the recently launched interface based on the work of Moini, discussed below). (ii) EOF variation. (iii) Low robustness. (iv) Limited lifetime of the emitter tip. (v) Limited BGE composition selection, which must comprise volatile compounds, since the CE effluent is directly inserted into the MS system. Therefore, routine analysis with sheathless interfaces may be jeopardized because of the constant need for emitter tip substitution. In addition, low system repeatability is generally observed.

In 2007, Moini has shown for the first time a robust sheathless CE-MS interface, commercialized some years afterward by Sciex [164]. Nowadays, this is still the only sheathless interface that is commercialized with a CE-MS equipment. In Moini's design, the CE capillary tip has been etched with a 49 % HF solution (after removing the polyimide external coating) to obtain a porous tip to be inserted into the ESI needle, filled with BGE (Fig. 5.2c, d). The porous junction is necessary to allow ion transport for closing the CE electrical circuit and concomitantly to supply the ESI voltage. This interface has overcome the limitations imposed by the previously reported sheathless interfaces in many aspects: (i) Its fabrication is reproducible and automated. (ii) A single-step etching process makes the capillary tip porous and with a smaller outer diameter – the inner wall is preserved during fabrication by flowing nitrogen gas. (iii) Any tip disruption may be easily fixed by removing a small section of the capillary. (iv) Eventual electrolysis reactions occur outside the CE capillary, avoiding bubble formation, which would harm CE separation due to current interruption. Ramautar et al. have explored this interface configuration for profiling human urine metabolites [117].

A comparison of the performance of CE-ESI-MS sheath liquid and sheathless interfaces in terms of detectability for the analysis of intact proteins has been reported by Haselberg et al. [165]. Capillaries with a porous tip were inserted into a stainless steel needle filled with static conductive liquid and installed in a conventional ESI source. The same porous tip capillaries were used in a sheath liquid interface with isopropanol as SHL, resulting in fairly similar responses in terms of protein signals. However, limits of detection obtained with the sheath liquid interface were substantially higher than those obtained with the sheathless interface (from 82 to 136 times

higher), due to increased baseline noise levels in the former. Detection limits were overall improved by a factor of 6.5–20 with sheathless CE-MS.

5.2.3   CE-MS Methodologies for Metabolomics

CE-MS metabolomics studies are often conducted under electrospray ionization (ESI) with triple coaxial sheath flow interfaces and time-of-flight (TOF) mass analyzers [33–56]. Due to simplicity, CZE is the preferred CE mode in metabolomics, generating robust methods. Unlike LC-MS, the mobile phase or more precisely the BGE composition changes according to the selected ionization mode. The addition of low percentages of organic solvents to volatile BGEs is often sought to improve resolution. Baseline separation of leucine/isoleucine/*allo*-isoleucine isomers in methanol-modified formic acid BGE is a landmark [138, 145]. In addition, a sheath liquid that may be of distinct composition for each ionization mode is used to promote and/or enhance ionization at the ion source. Small cationic and anionic charged species are the expected metabolite targets visualized by CE separations. In CE-MS with positive ionization mode ($[M]^+$, $[M+H]^+$, $[M-H_2O+H]^+$, $[M+Na]^+$, etc.; M stands for molecule), it is possible to inspect amino acids, biogenic amines, and nucleosides, whereas the negative ionization mode ($[M-H]^-$, $[M+HCOO]^-$, $[2M+Na-2H]^-$, etc.) reveals carboxylic acids, phosphorylated carboxylic acids, phosphorylated saccharides, nucleotides, nicotinamide and flavin adenine coenzymes, as well as citrate isomers, dinucleotides, and CoA compounds [125].

A schematic representation of the overall possibilities CE-MS offers for the analysis of cationic and anionic metabolites is depicted in Fig. 5.3. Typically, cationic metabolites are screened in uncoated fused-silica capillaries with low pH volatile electrolytes, such as formic acid or acetic acid, generating a small but normal electroosmotic flow (EOF, flow toward the cathode due to little ionization of the capillary wall silanol groups). The CE system is operated under positive high voltage, applied at the capillary inlet, and it is connected to the MS via ESI in positive ionization mode (Fig. 5.3a). Anionic metabolites may also be screened in this format; however, high pH volatile electrolytes, such as ammonia/ammonium salt buffers (ammonium formate, acetate, or carbonate being the most commonly used), are mandatory to generate an EOF high enough to conduct the compounds which passed the interface toward the MS entrance. Nevertheless, since a positive voltage is applied at the capillary inlet, the anions will migrate counter-electroosmotically. Moreover, the ESI voltage needs to be set appropriately (negative ionization mode). This approach is not preferential for anionic metabolite analysis because by setting the migration of anions against EOF, long analytical runs are imposed and migration time repeatability might be compromised, which is already a concern in CE separations in bare fused-silica capillaries.

A more elegant solution for the CE-MS analysis of anionic metabolites is reached with the use of coated capillaries and high pH electrolytes (Fig. 5.3b, c). The CE system is now operated under a negative high voltage, applied at the capillary inlet,
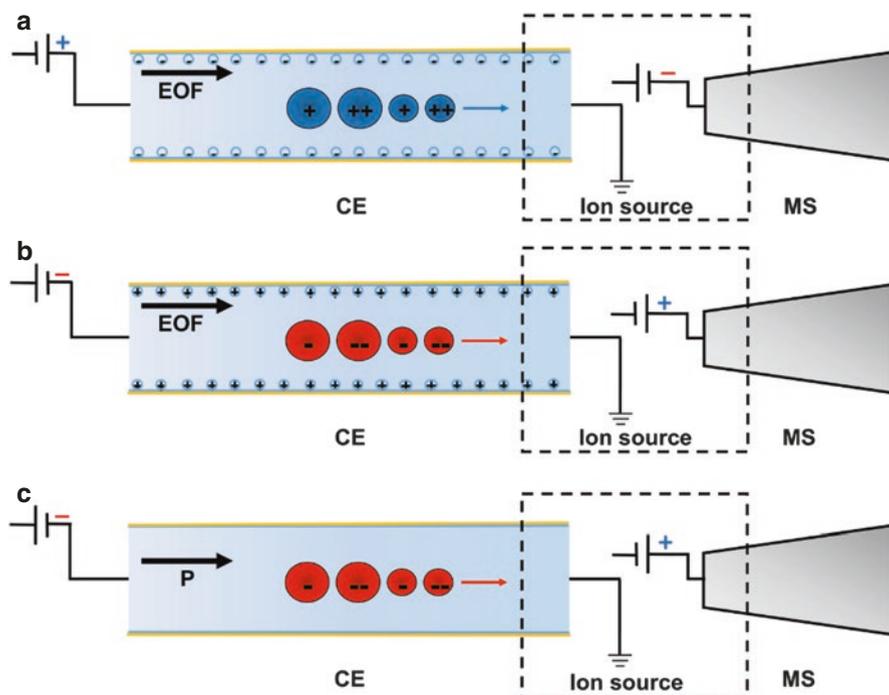
Fig. 5.3 CE-MS methodologies for untargeted metabolomics of cationic metabolites (**a**) and anionic metabolites using cationic polymer-coated capillaries (**b**) and neutral capillaries (**c**)

and it is connected to the MS via ESI in negative ionization mode. Either a cationic polymer coating (Fig. 5.3b) to reverse EOF (flow directed toward the anode) or a neutral coating (Fig. 5.3c) to eliminate EOF can be chosen. In the former, the negative species electrophoretic velocity and the electroosmotic flow velocity are in the same direction (toward the anode or positive pole), resulting in additionally faster separations. In the latter, since EOF is eliminated, a pressure-driven flow is usually implemented to prevent that sheath liquid components enter the separation capillary.

With these simple approaches, Soga and collaborators introduced untargeted metabolomics of biological cells using CE-MS platforms for the first time [125, 127]. By using all schemes of Fig. 5.3, a thorough evaluation of the metabolome of *Bacillus subtilis* cells upon the onset of sporulation was pursued. For the cationic metabolites screening, a BGE composed of 1 mol L^{-1} formic acid and a SHL comprised of 5 mmol L^{-1} ammonium acetate in 50 % methanol/water were used. Sets of 30 protonated [M+H]$^+$ ions were analyzed successively by SIM mode to cover the entire range of *m/z* from 70 to 1027 (Fig. 5.4). Anionic metabolites were screened in a BGE composed of 50 mmol L^{-1} ammonium acetate at pH 8.5 with a SHL comprised of 5 mmol L^{-1} ammonium acetate in 50 % methanol/water in a cationic polymer-coated capillary, SMILE(+). Nucleotides and coenzyme A compounds
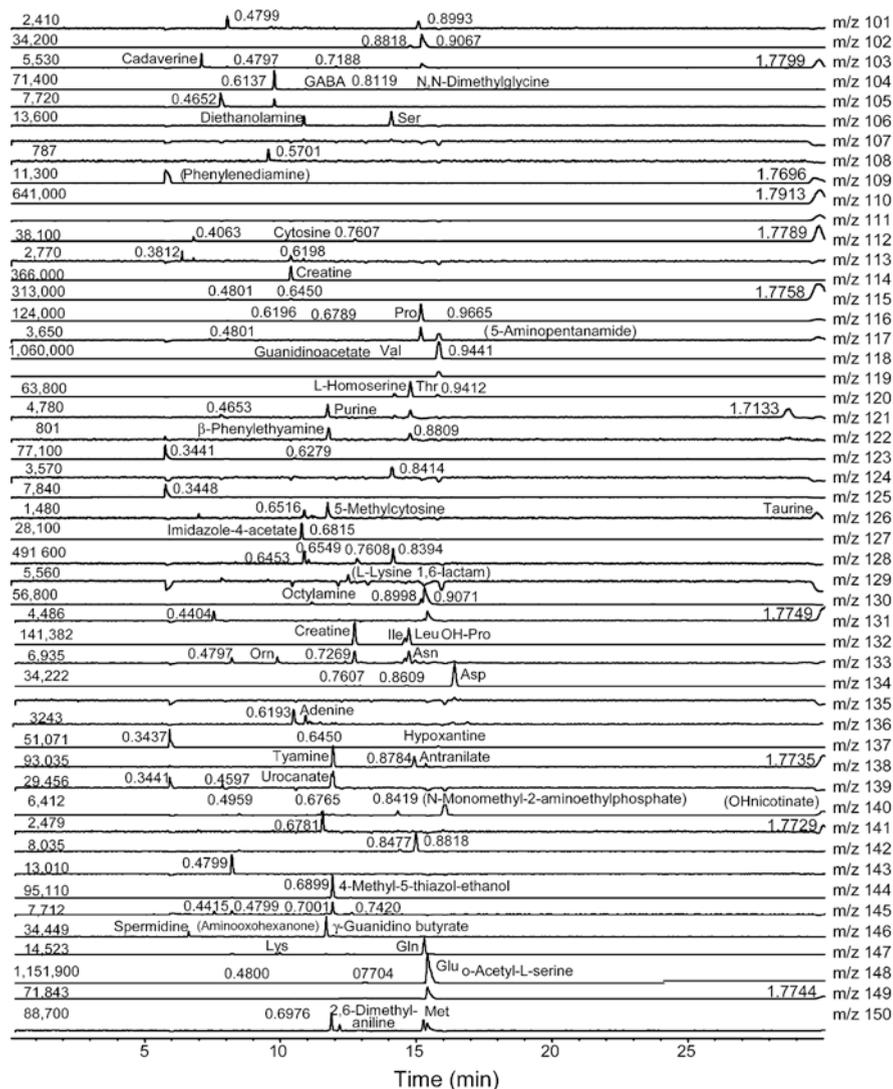
Fig. 5.4 Selected ion electropherograms for cationic metabolites of *Bacillus subtilis* in the range of *m/z* 101–150. The numbers in the upper left corner of each trace are the abundances associated with the tallest peak in the electropherogram, for each *m/z*, and the numbers on top of peaks are relative migration times normalized with methionine sulfone (IS) (Reprinted with permission from Ref. [127])

were screened in a GC-coated capillary (polydimethylsiloxane, DB-1) in a slightly lower pH BGE, 50 mmol L⁻¹ ammonium acetate at pH 7.5, with the same SHL. To prevent entrance of the SHL into the separation channel, a pressure of 50 mbar was applied to the capillary inlet promoting a flow of solution toward the anode. Exactly

1692 metabolites were catalogued, 150 were positively identified, and 83 were assigned based on the expected charge state and isotopic distribution. Later on, Soga and col. advocated the use of platinum ESI spray needle to replace stainless steel spray needles in the analysis of anionic metabolites [122]. It was observed that stainless steel was prone to oxidation and corrosion at the anodic electrode due to electrolysis; the resulting precipitation of iron oxides plugged the capillary outlet. Moreover, eventual complexation of anionic metabolites with iron and nickel ions generated by corrosion would reduce significantly detection sensitivity because the formed complexes are positively charged and move backward to the cathode (capillary inlet).

Within the context of CE-MS methodologies for metabolomics, a few parameters of interest will be examined. Because metabolomics studies are comparative in nature, they demand high-precision measurements. Migration time repeatability, essential in untargeted metabolomics studies, and/or peak area repeatability, crucial in targeted metabolomics quantitation studies, must be addressed properly. It is well accepted that migration time variability is a consequence of EOF variability, which in turn is related to the capillary inner surface state and integrity. Thus, adsorption of solutes to the capillary wall and/or any sort of adverse solute-wall interactions, electrolyte components-wall interactions, etc. will compromise the EOF magnitude and consequently affect migration time repeatability. Many authors have addressed this issue by using covalently coated capillaries, such as the work of Soga and col. referred above [122, 125, 127]. De Jong and collaborators have proposed to modify the capillary walls dynamically with charged polymers [166]. Bilayers constituted of polybrene (PB) and poly(vinyl sulfonate) (PVS) or triple layers constituted of PB, dextran sulfate (DS), and PB have been extensively investigated to the metabolic profiling of biofluids [113, 114, 144]. Overall, covalently bound polymers are still preferred in metabolomics studies due to stability and durability. Moreover, any leakage of polymer during CE operation cannot be tolerated, especially if it results in contamination of the mass analyzer.

Full coverage of metabolites by any hyphenated technique to MS demands the use of both positive and negative ionization modes. It is worth mentioning that a large fraction of metabolites in biological fluids is acidic in nature and can only be ionized efficiently using negative ionization. However, signal-to-noise ratios in negative ionization mode are often low by two to three orders of magnitude when compared to positive ionization, thereby limiting sensitivity in metabolomics applications [112]. Reduced MS signals for anions have been attributed to analyte ionization suppression by the presence of acetate ions in the BGE and/or SHL [114, 167]. To circumvent this loss of sensitivity, the transformation of anionic metabolites into cationic compounds by derivatization or complexation has been proposed, allowing positive ionization mode to be applied [167, 168]. With these methodologies, sensitivity indeed improved for anionic compounds, and more favorable detection limits were achieved. However, derivatization procedures increase sample pretreatment complexity, and losses of metabolites can occur due to incomplete derivatization. Furthermore, not every anionic compound can be derivatized efficiently. Therefore, a great deal of development is still necessary for

the CE-MS analysis of the metabolome in negative ionization mode, despite the efforts toward the testing of new BGE additives [112].

Still regarding ionization modes, an alternative strategy to simplify procedures during data acquisition in CE-MS metabolomics has been proposed by Gulersonmez et al. [132]. A single BGE at an intermediary pH is used for both positive and negative ionization modes, e.g., pH 3.0 acetic acid. This pH is low enough to protonate most of the metabolites exhibiting basic moieties (biogenic amines, amino acids, etc.) generating cationic compounds, and at the same time, it is high enough to promote partial dissociation of those metabolites with acidic moieties (carboxylic acids, nucleotides, etc.) generating anionic compounds. Therefore, the same BGE is used to screen both cationic and anionic portions of the metabolome, in consecutive runs (TOF mass analyzers), by a simple switch of the ESI voltage.

Analytical frequency is another parameter of concern in metabolomics studies, since runs are usually long to ensure that a large variety of metabolites of differing properties is inspected. Multiple sequential injections of samples (volumetric transfer of sample to the separation capillary by applying pressure at capillary inlet), intercalated by injections of BGE zones, before the high voltage has been set, are a classical CE strategy to improve analytical frequency, and it has been implemented advantageously in CE-MS metabolomics by Britz-McKibbin and collaborators [140]. The authors developed a multi-segment injection (MSI) as a multiplexed CE-MS platform in which a serial injection of seven or more discrete human plasma sample segments could be performed within a single capillary without compromising the separation quality and/or quantitative performance. The overall MSI scheme is depicted in Fig. 5.5 and increased sample throughput by one order of magnitude. By using a seven-segment sample injection for single-step acquisition, building of external and processed analytical curves for quantitation of polar metabolites and isomers in plasma, with acceptable accuracy and precision, use and/or selection of internal standards, running recovery tests samples, identification strategies via pattern recognition, etc., have all been successfully demonstrated.

5.3   Representative Applications of CE-MS in Clinical Metabolomics

Capillary electrophoresis as an analytical platform to assess metabolites in biological samples has been used for decades. The work of Jellum and collaborators in the profiling of organic acids in biofluids, using UV detectors to screen metabolic diseases, may be referred as the first CE-based clinical metabolomics [169–171]. By the same token, the work of Barbas and collaborators, who investigated metabolic disorders, known as inborn errors of metabolism, by screening short-chain carboxylic acids in human urine, is another innovatory example of the diagnostic power of CE technology [172, 173]. The pioneerism of global metabolic fingerprinting or untargeted metabolomics using CE-MS platforms has been attributed unequivocally to Soga's research group [125, 127], as described previously in this chapter and summarized by the
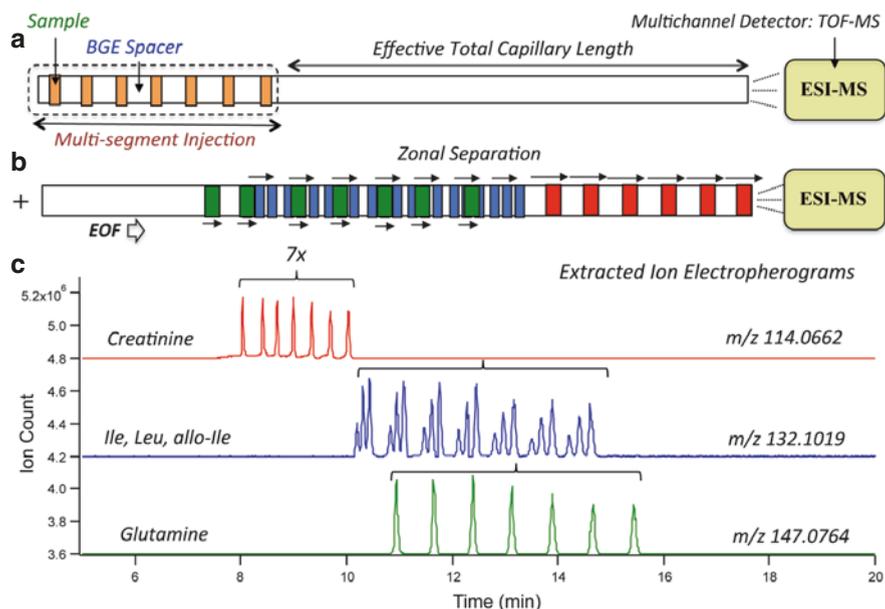
Fig. 5.5 Multiplexed separation based on serial injection of seven discrete sample segments within a single capillary by MSI-CE-MS (**a**), where (**b**) ions migrate as a series of zones in free solution prior to ionization. This format enables reliable quantification of polar metabolites and their isomers in different samples since ionization occurs within a short time interval (≈2–6 min) under steady-state conditions when using a high mass resolution TOF-MS (**c**) (Reprinted with permission from Ref. [140])

group reviews [36, 174]. At this point, it is worth mentioning the important contribution of Mishak's research group, who has established the reliability of CE-MS platforms for the initial diagnosis and prognosis of the progression of numerous diseases via biomarker discovery by mapping endogenous peptides in human urine [175, 176].

Table 5.1 compiles many examples of CE-MS in clinical metabolomics organized by studied condition or disease, biological matrix, metabolomics approach, and type of metabolites screened (targeted metabolomics) or revealed (untargeted metabolomics). A few details of the analytical methods, such as BGE and/or SHL composition, type of mass analyzer, sample preparation procedures, algorithm and/or software used in data treatment, and finally whether the study was exclusively conducted by CE-MS or data was acquired in a multiplatform setup, were also provided.

Table 5.1 was meant to present a comprehensive revision of the literature in the period from 2001 to 2016. Based on the relatively small number of applications Table 5.1 brings, associated with an even smaller number of groups researching in the field, it is fair to conclude that CE-MS in the clinical metabolomics scenario has still much room for growth. A possible explanation for the rather limited use of CE-MS technology in this field might be related to the fact that CE-MS is still considered a novelty compared to other much more established techniques, such as GC-MS, LC-MS, and NMR, and there is a certain resistance to consider its use in metabolomics. Issues such as migration time variability, sample loadability and throughput,

low concentration sensitivity, etc. are still of concern by many metabolomics leading groups, despite the relevant advancements made over decades to overcome CE-MS technical and methodological difficulties, as thoroughly discussed in this chapter. Another relevant aspect that might hinder the use of CE-MS in clinical metabolomics studies is the lack of standardized operating protocols. As Table 5.1 contents sustain, each research group develops and implements its own method, with small but tangible variations of capillary coatings, dimensions and conditioning, BGE and/or SHL composition, as well as MS type and parameters. There is no convergence toward a single optimized strategy to perform untargeted metabolomics studies, for instance, or a complete detailed protocol for metabolomics, as it is the case for GC-MS-, LC-MS-, and NMR-based metabolomics. Such protocol would boost sales of CE-MS instrumentation, allowing the creation of an universal database for metabolite identification, and stimulate applications in clinical metabolomics, among other areas.

Metabolomics studies follow a general workflow, comprising problem formulation, experimental design, sample preparation, data acquisition and processing, statistical analysis, metabolite identification, association to metabolic pathways, and biological validation. All these steps were critically discussed in Chap. 1. Notably, a large number of applications compiled in Table 5.1 were conducted under the premise of untargeted clinical metabolomics, i.e., to improve the knowledge on the onset and progression of a given disease at metabolic level and to search discriminant metabolites that could be used further on for diagnosis and/or prognosis purposes. Furthermore, most of the reported CE-MS methods in Table 5.1 were conducted with high-resolution mass spectrometers using sheath liquid interfaces (available commercially much longer than sheathless interfaces), using low pH BGE and aqueous methanolic SHL. Untargeted metabolomics studies comprise simpler sample treatments, involving protein precipitation, followed by filtration and dilution. Sample procedures for targeted metabolomics are of course dependent on the identity of the metabolites under investigation. Overall sample preparation strategies for metabolomics and their impact on results have been revised by many authors [42, 106, 177, 178]. Quality control samples (QC), pool of all control and test samples under consideration in a given study, have often been considered to attest platform stability during data acquisition. Method validation concepts for untargeted metabolomics have also been reviewed [157].

Another feature of Table 5.1 is that data preprocessing often relies on in-house developed algorithms or free access softwares with PCA, PLS-DA, and OPLS-DA being the preferred multivariate data analysis. Peak alignment in such algorithms is usually of great importance because it deals with the intrinsic migration time variability of CE-MS data and it has long been a topic of investigation [179–182]. A great challenge in metabolomics studies in general is metabolite identification. In CE-MS platforms, several authors have demonstrated strategies of peak identification using mobility in conjunction with accurate m/z values [183–185]. Finally, analytical multiplatform studies in Table 5.1 are rare.

An illustration of the contents of Table 5.1, discriminating conditions and diseases studied so far by CE-MS under the metabolomics perspective, is depicted in Fig. 5.6a. It is readily observed that several types of cancer received a great deal of attention by the scientific community. Fig. 5.6b shows the temporal evolution of publications in the field denoting the growing interest CE-MS has drawn in the last decades.
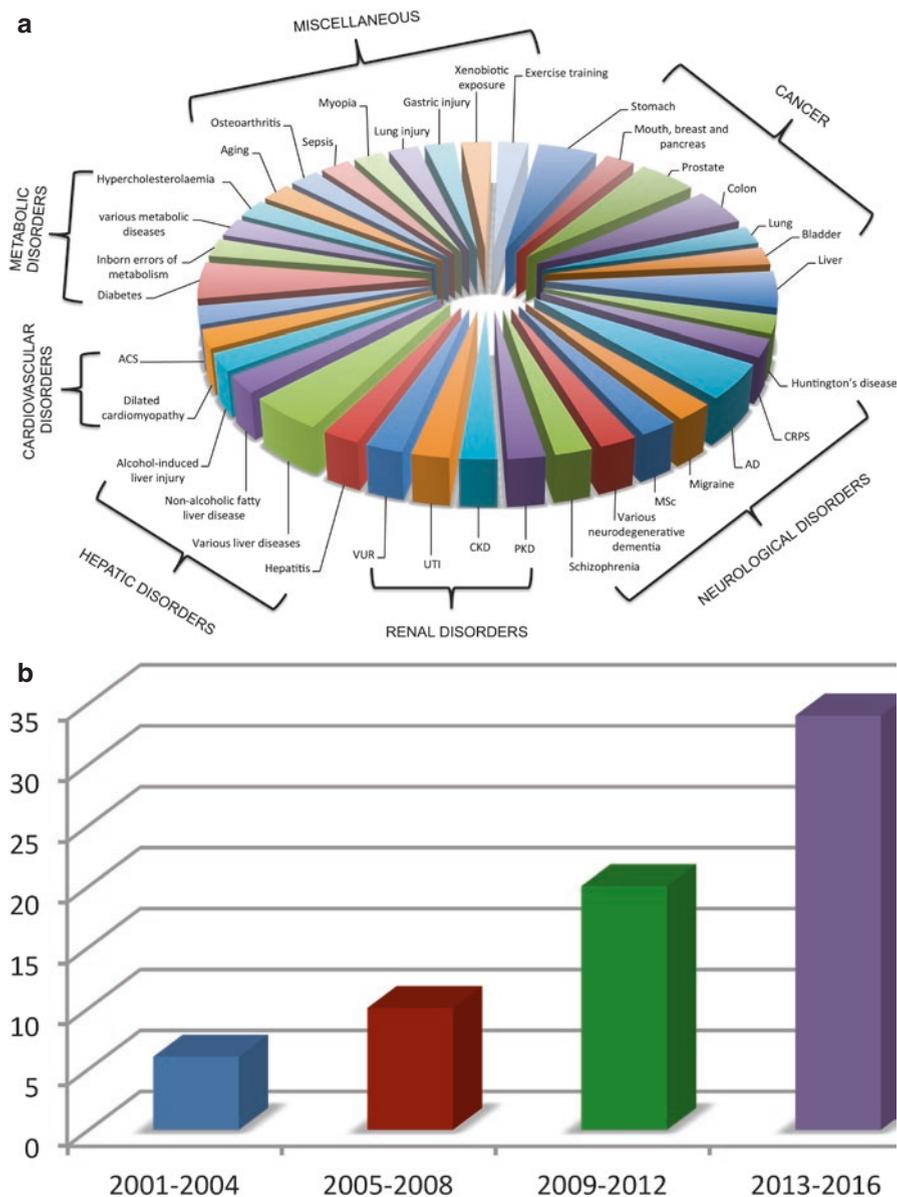
Fig. 5.6 CE-MS in clinical metabolomics. Literature publications compiled in Table 5.1 were organized by type of condition and/or disease (**a**) and temporal progression of articles (**b**). Legends: *ACS* acute coronary syndrome, *AD* Alzheimer's disease, *CKD* chronic kidney disease, *CRPS* complex regional pain syndrome, *MSc* multiple sclerosis, *PKD* polycystic kidney disease, *VUR* vesicoureteral reflux, *UTI* urinary tract infection

5.4   Conclusions and Perspectives

From its inception in the late 1980s, CE-MS has matured into a resourceful technique that encompasses the analysis of compounds from many different chemical classes, especially those with ionic and/or highly polar character that constitute an important subset of the human metabolome. Relevant features of CE, such as high efficiency and resolution power, fast analysis time, multiple separation modes, use of aqueous-based electrolytes, compatibility with biofluids, small sample volume, etc. were combined with the remarkable detection sensitivity, extra selectivity, and spectral information provided by MS technologies. Several technical difficulties related to early CE-MS interface designs and platform stability issues have been tackled and improved considerably, especially in the last decade that was testimony to the commercial launching of complete CE-MS systems. Intrinsic aspects related to the technique performance have also been addressed properly, allowing CE-MS to grow into a robust technology for metabolomics.

Despite the clear adequacy of CE-MS for clinical metabolomics and the technical improvements evidenced over the years, the field is still underrepresented when compared to the contribution of other well-established NMR and chromatography-based platforms, showing a rather limited number of research leading groups actively working in the area. Implementation of analytical multiplatform approaches, necessary to establish a more comprehensive coverage of the metabolome, analysis of larger clinical cohorts, expansion of the applicability to key diseases and conditions, and setting up interlaboratorial validation studies are a few strategies that should boost the use of CE-MS in clinical metabolomics and build user confidence in the technology.

References

1. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29:1181–9.
2. Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol. 2002;48:155–71.
3. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biochem. 2004;22:245–52.
4. Patti GJ, Yanes O, Siuzdak G. Metabolomics: the apogee of the omics trilogy. Nature Rev. 2012;13:263–9.
5. Ramautar R, Berger R, van der Greef J, Hankmeier T. Human metabolomics: strategies to understand biology. Curr Op Chem Biol. 2013;17:841–6.

6. Misra BB, van der Hooft JJJ. Updates in metabolomics tools and resources: 2014–2015. Electrophoresis. 2016;37:86–110.
7. Lenz EM, Wilson ID. Analytical strategies in metabonomics. J Proteome Res. 2007;6:443–58.
8. Dona AC, Jiménez B, Schäfer H, Humpfer E, Spraul M, Lewis MR, Pearce JTM, Holmes E, Lindon JC, Nicholson JK. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem. 2014;86:9887–94.
9. Lu E, Bennett BD, Rabinowitz JD. Analytical strategies for LC-MS-based targeted metabolomics. J Chromatogr B. 2008;871:236–42.
10. Theodoridis G, Gika HG, Wilson ID. LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. Trends Anal Chem. 2008;27:251–60.
11. Xiayan L, Legido-Quigley C. Advances in separation science applied to metabonomics. Electrophoresis. 2008;29:3724–36.
12. Issaq HJ, Abbott E, Veenstra TD. Utility of separation science in metabolomic studies. J Sep Sci. 2008;31:1936–47.
13. Roux A, Lison D, Junot C, Heilier J-F. Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: a review. Clin Biochem. 2011;44:119–35.
14. Kuehnbaum NL, Britz-McKibbin P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. Chem Rev. 2013;113:2437–68.
15. Putri SP, Nakayama Y, Matsuda F, Uchikata T, Kobayashi S, Matsubara A, Fukusaki E. Current metabolomics: practical applications. J Biosci Bioeng. 2013;115:579–89.
16. Sheth BP, Thaker VS. Plant systems biology: insights, advances and challenges. Planta. 2014;240:33–54.
17. Simó C, Ibáñez C, Valdés A, Cifuentes A, García-Cañas V. Metabolomics of genetically modified crops. Int J Mol Sci. 2014;15:18941–66.
18. Jones DP, Park Y, Ziegler TR. Nutritional metabolomics progress in addressing complexity in diet and health. Annu Rev Nutr. 2012;32:183–202.
19. Astarita G, Langridge J. An emerging role for metabolomics in nutrition. J Nutrigenet Nutrigenomics. 2013;6:179–98.
20. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, Rappaport SM, van der Hooft JJJ, Wishart DS. The food metabolome: a window over dietary exposure. Am J Clin Nutr. 2014;99:1286–308.
21. Monte AA, Brocker C, Nebert DW, Gonzalez FJ, Thompson DC, Vasiliou V. Improved drug therapy: triangulating phenomics with genomics and metabolomics. Hum Genomics. 2014;8:16.
22. Reaves ML, Rabinowitz JD. Metabolomics in systems microbiology. Curr Op Biotech. 2011;22:17–25.
23. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Arch Toxicol. 2011;85:5–17.
24. Dessì A, Marincola FC, Masili A, Gazzolo D, Fanos V. Clinical metabolomics and nutrition: the new frontier in neonatology and pediatrics. BioMed Res Int. 2014;id981219.
25. Xu Y-J, Wang C, Ho WE, Ong CN. Recent developments and applications of metabolomics in microbiological investigations. Trends Anal Chem. 2014;56:37–48.
26. Mastrangelo A, Armitage EG, García A, Barbas C. Metabolomics as a tool for drug Discovery and personalised medicine. A review. Curr Top Med Chem. 2014;14:1617–36.
27. Garcia A, Barbas C. Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. In: Metz TO, editor. Metabolic profiling. Methods in molecular biology 708. Springer, Totowa, NJ; 2011.
28. Kind T, Fiehn O. Advances in structure elucidation of small molecules using mass spectrometry. Bioanal Rev. 2010;2:23–60.
29. Lewis MR, Pearce JT, Spagou K, Green M, Dona AC, Yuen AH, David M, Berry DJ, Chappell K, Horneffer-van der Sluis V, Shaw R, Lovestone S, Elliott P, Shockcor JP, Lindon JC, Cloarec O, Takats Z, Holmes E, Nicholson JK. Development and application of UPLC-ToF MS for

precision large scale urinary metabolic phenotyping. Anal Chem. 2016. doi:10.1021/acs.analchem.6b01481.

30. Jackson F, Georgakopoulou N, Kaluarachchi MR, Kyriakides M, Andreas NJ, Przysiezna N, Hyde MJ, Modi N, Nicholson JK, Wijeyesekera A, Holmes E. Development of a pipeline for exploratory metabolic profiling of infant urine. J Proteome Res. 2016. doi:10.1021/acs.jproteome.6b00234.

31. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS. The human serum metabolome. PLoS One. 2011;6:e16957. doi:10.1371/journal.pone.0016957.

32. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT, Poelzer J, Huynh J, Yallou FS, Psychogios N, Dong E, Bogumil R, Roehring C, Wishart DS. The human urine metabolome. PLoS One. 2013;8:e73076. doi:10.1371/journal.pone.0073076.

33. Huck CW, Bakry R, Bonn GK. Progress in capillary electrophoresis of biomarkers and metabolites between 2002 and 2005. Electrophoresis. 2006;27:111–25.

34. Ramautar R, Demirci A, de Jong GJ. Capillary electrophoresis in metabolomics. Trends Anal Chem. 2006;25:455–66.

35. Suresh Babu CV, Song EJ, Babar SME, Wi MH, Yoo YS. Capillary electrophoresis at the omics level: towards systems biology. Electrophoresis. 2006;27:97–110.

36. Monton MRN, Soga T. Metabolome analysis by capillary electrophoresis-mass spectrometry. J Chromatogr A. 2007;1168:237–46.

37. Song EJ, Babar SM, Oh E, Hasan MN, Hong HM, Yoo YS. CE at the omics level: towards systems biology – an update. Electrophoresis. 2008;29:129–42.

38. García-Pérez I, Vallejo M, García A, Legido-Quigley C, Barbas C. Metabolic fingerprinting with capillary electrophoresis. J Chromatogr A. 2008;1204:130–9.

39. Issaq HJ, Blonder J. Electrophoresis and liquid chromatography/tandem mass spectrometry in disease biomarker discovery. J Chromatogr B. 2009;877:1222–8.

40. Raumatar R, Somsen GW, de Jong GJ. CE-MS in metabolomics. Electrophoresis. 2009;30:276–91.

41. Oh E, Hasan MN, Jamshed M, Park SH, Hong HM, Song EJ, Yoo YS. Growing trend of CE at the omics level: the frontier of systems biology. Electrophoresis. 2010;31:74–92.

42. Barbas C, Moraes EP, Villasenora A. Capillary electrophoresis as a metabolomics tool for non-targeted fingerprinting of biological samples. J Pharm Biom Anal. 2011;55:823–31.

43. Ramautar R, Mayboroda OA, Somsen GW, de Jong GJ. CE-MS for metabolomics: developments and applications in the period 2008–2010. Electrophoresis. 2011;32:52–65.

44. Ban E, Park SH, Kang M-J, Yoo YS. Growing trend of CE at the omics level: the frontier of systems biology – an update. Electrophoresis. 2012;33:2–13.

45. Wang X, Li K, Adams E, Schepdael AV. Capillary electrophoresis–mass spectrometry in metabolomics: the potential for driving drug discovery and development. Curr Drug Metab. 2013;14:807–13.

46. Kok MGM, Somsen GW, de Jong GJ. The role of capillary electrophoresis in metabolic profiling studies employing multiple analytical techniques. Trends Anal Chem. 2014;61:223–35.

47. Hirayama A, Wakayama M, Soga T. Metabolome analysis based on capillary electrophoresis-mass spectrometry. Trends Anal Chem. 2014;61:215–22.

48. Robledo VR, Smyth WF. Review of the CE-MS platform as a powerful alternative to conventional couplings in bio-omics and target-based applications. Electrophoresis. 2014;35:2292–308.

49. Zhong X, Zhang Z, Jiang S, Li L. Recent advances in coupling capillary electrophoresis-based separation techniques to ESI and MALDI-MS. Electrophoresis. 2014;35:1214–25.

50. Poinsot V, Ong-Meang V, Gavard P, Couderc F. Recent advances in amino acid analysis by capillary electromigration methods, 2011-2013. Electrophoresis. 2014;35:50–68.

51. Naz S, dos Santos DCM, García A, Barbas C. Analytical protocols based on LC-MS, GC-MS and CE-MS for nontargeted metabolomics of biological tissues. Bioanalysis. 2014;6:1657–77.

52. Buzatto AZ, de Sousa AC, Guedes SF, Cieslarová Z, Simionato AV. Metabolomic investigation of human diseases biomarkers by CE and LC coupled to MS. Electrophoresis. 2014;35:1285–307.

53. Ramautar R, de Jong GJ. Recent developments in liquid-phase separation techniques for metabolomics. Bioanalysis. 2014;6:1011–26.

54. Ramautar R, Somsen GW, de Jong GJ. CE-MS for metabolomics: developments and applications in the period 2012–2014. Electrophoresis. 2015;36:212–24.

55. Lindenburg PW, Haselberg R, Rozing G, Ramautar R. Developments in interfacing designs for CE–MS: towards enabling tools for proteomics and metabolomics. Chromatographia. 2015;78:367–77.

56. Iadarola P, Fumagalli M, Bardoni AM, Salvini R, Viglio S. Recent applications of CE- and HPLC-MS in the analysis of human fluids. Electrophoresis. 2016;37:212–30.

57. James AT, Martin AJP. Gas-liquid partition chromatography: the separation and micro estimation of volatile fatty acids from formic acid to dodecanoic acid. Biochem J. 1952;50:679–90.

58. Horvath CG, Preiss BA, Lipsky SR. Fast liquid chromatography. Investigation of operating parameters and the separation of nucleotides on pellicular ion exchangers. Anal Chem. 1967;39:1422–8.

59. Jorgenson JW, Lukacs KD. Zone electrophoresis in open-tubular glass capillaries. Anal Chem. 1981;53:1298–302.

60. Jorgenson JW, Lukacs KD. High-resolution separations based on electrophoresis and electro-osmosis. J Chromatogr. 1981;218:209–16.

61. Hjérten S. Free zone electrophoresis. Chromatogr Rev. 1967;9:122–219.

62. Virtanen R. Zone electrophoresis in a narrow-bore tube employing potentiometric detection. Theoretical and experimental study. Acta Polytech Scand Chem Incl Metall Ser. 1974;123:1–67.

63. Mikkers FEP, Everaerts FM, Verheggen PEM. High performance zone electrophoresis. J Chromatogr. 1979;169:11–20.

64. Landers JP. Handbook of capillary and microchip electrophoresis and associated microtechniques. New York: CRC Press; 2007.

65. Terabe S, Otsuka K, Ichikawa K, Tsuchiya A, Ando T. Electrokinetic separations with micellar solutions and open-tubular capillaries. Anal Chem. 1984;56:111–3.

66. Righetti PG. Determination of the isoelectric point of proteins by capillary isoelectric focusing. J Chromatogr A. 2004;1037:491–9.

67. Silvertand LHH, Torao JS, Bennekom WP, de Jong GJ. Recent developments in capillary isoelectric focusing. J Chromatogr A. 2008;1204:157–70.

68. Righetti PG, Sebastiano R, Citterio A. Capillary electrophoresis and isoelectric focusing in peptide and protein analysis. Proteomics. 2013;13:325–40.

69. Malá Z, Gebauer P, Boček P. Recent progress in analytical capillary isotachophoresis. Electrophoresis. 2015;36:2–14.

70. Cohen AS, Paulus A, Karger BL. High-performance capillary electrophoresis using open tubes and gels. Chromatographia. 1987;24:15–24.

71. Knox JH, Grant IH. Electrochromatography in packed tubes using 1.5 to 50 silica gels and ODS bonded silica gels. Chromatographia. 1991;32:317–28.

72. Heegaard NHH. Affinity in electrophoresis. Electrophoresis. 2009;30:S229–39.

73. Olivares JA, Nguyen NT, Yonker CR, Smith RD. On-line mass spectrometric detection for capillary zone electrophoresis. Anal Chem. 1987;59:1230–2.

74. Smith RD, Olivares JA, Nguyen NT, Udseth HR. Capillary zone electrophoresis-mass spectrometry using an electrospray ionization interface. Anal Chem. 1988;60:436–41.

75. Smith RD, Barinaga CJ, Udseth HR. Improved electrospray ionization interface for capillary zone electrophoresis-mass spectrometry. Anal Chem. 1988;60:1948–52.

76. Smith RD, Udseth HR. Capillary zone electrophoresis-MS. Nature. 1988;331:639–40.

77. Naz S, Calderón AA, García A, Gallafrio J, Mestre RT, González EG, de Cabo CM, Delgado MCM, Balanza JAL, Simionato AVC, Vaeza NN, Barbas C, Rupérez FJ. Unveiling differences between patients with acute coronary syndrome with and without ST elevation through fingerprinting with CE-MS and HILIC-MS targeted analysis. Electrophoresis. 2015;36:2303–13.

78. Nevedomskaya E, Ramautar R, Derks R, Westbroek I, Zondag G, van der Pluijm I, Deelder AM, Mayboroda OA. CE-MS for metabolic profiling of volume-limited urine samples: application to accelerated aging TTD mice. J Proteome Res. 2010;9:4869–74.

79. Ibáñez C, Simó C, Martín-Álvarez PJ, Kivipelto M, Winblad B, Cedazo-Mínguez A, Cifuentes A. Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry metabolomics. Anal Chem. 2012;84:8532–40.

80. González-Domínguez R, García A, García-Barrera T, Barbas C, Gómez-Ariza JL. Metabolomic profiling of serum in the progression of Alzheimer's disease by capillary electrophoresis-mass spectrometry. Electrophoresis. 2014;35:3321–30.

81. Alberice JV, Amaral AFS, Armitage EG, Lorente JA, Algaba F, Carrilho E, Márquez M, García A, Malats N, Barbas C. Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry metabolomics approach. J Chromatogr A. 2013;1318:163–70.

82. Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. Metabolomics. 2010;6:78–95.

83. Ibáñez C, Simó C, García-Cañas V, Gómez-Martínez A, Ferragut JA, Cifuentes A. CE/LC-MS multiplatform for broad metabolomic analysis of dietary polyphenols effect on colon cancer cells proliferation. Electrophoresis. 2012;33:2328–36.

84. Hirayama A, Kami K, Sugimoto M, Sugawara M, Toki N, Onozuka H, Kinoshita T, Saito N, Ochiai A, Tomita M, Esumi H, Soga T. Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. Cancer Res. 2009;69:4918–25.

85. Kami K, Fujimori T, Sato H, Sato M, Yamamoto H, Ohashi Y, Sugiyama N, Ishihama Y, Onozuka H, Ochiai A, Esumi H, Soga T, Tomita M. Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry. Metabolomics. 2013;9:444–53.

86. Chen J-L, Fan J, Lu X-J. CE-MS based on moving reaction boundary method for urinary metabolomics analysis of gastric cancer patients. Electrophoresis. 2014;35:1032–9.

87. Soliman LC, Hui Y, Hewavitharana AK, Chen DD. Monitoring potential prostate cancer biomarkers in urine by capillary electrophoresis-tandem mass spectrometry. J Chromatogr A. 2012;1267:162–9.

88. Toyohara T, Akiyama Y, Suzuki T, Takeuchi Y, Mishima E, Tanemoto M, Momose A, Toki N, Sato H, Nakayama M, Hozawa A, Tsuji I, Ito S, Soga T, Abe T. Metabolomic profiling of uremic solutes in CKD patients. Hypertens Res. 2010;33:944–52.

89. Ramautar R, van der Plas AA, Nevedomskaya E, Derks RJ, Somsen GW, de Jong GJ, van Hilten JJ, Deelder AM, Mayboroda OA. Explorative analysis of urine by capillary electrophoresis-mass spectrometry in chronic patients with complex regional pain syndrome. J Proteome Res. 2009;8:5559–67.

90. Erny GL, Acunha T, Simó C, Cifuentes A, Alves A. Algorithm for comprehensive analysis of datasets from hyphenated high resolution mass spectrometric techniques using single ion profiles and cluster analysis. J Chromatogr A. 2016;1429:134–41.

91. Armitage EG, Godzien J, Alonso-Herranz V, López-Gonzálvez Á, Barbas C. Missing value imputation strategies for metabolomics data. Electrophoresis. 2015;36:3050–60.

92. Balderaz C, Rupérez FJ, Ibáñez E, Señorans J, Guerrero-Fernández J, Casado IG, Gracia-Bouthelier R, García A, Barbas C. Plasma and urine metabolic fingerprinting of type 1 diabetic children. Electrophoresis. 2013;34:2882–90.

93. Ciborowski M, Adamska E, Rusak M, Godzien J, Wilk J, Citko A, Bauer W, Gorska M, Kretowski A. CE-MS-based serum fingerprinting to track evolution of type 2 diabetes mellitus. Electrophoresis. 2015;36:2286–93.

94. Maekawa K, Hirayama A, Iwata Y, Tajima Y, Nishimaki-Mogami T, Sugawara S, Ueno N, Abe H, Ishikawa M, Murayama M, Matsuzawa Y, Nakanishi H, Ikeda K, Arita M, Taguchi R, Minamino N, Wakabayashi S, Soga T, Saito Y. Global metabolomic analysis of heart tissue in a hamster model for dilated cardiomyopathy. J Mol Cell Cardiol. 2013;59:76–85.

95. Keuhnbaum NL, Gillen JB, Kormendi A, Lam KP, DiBattista A, Gibala MJ, Britz-McKibbin P. Multiplexed separations for biomarker discovery in metabolomics: elucidating adaptive responses to exercise training. Electrophoresis. 2015;36:2226–36.

96. Tokushige K, Hashimoto E, Kodama K, Tobari M, Matsishita N, Kogiso T, Taiai M, Torii N, Shiratori K, Nishizaki Y, Ohga T, Ohashi T, Sato T. Serum metabolomic profile and potential biomarkers for severity of fibrosis in nonalcohol fatty liver disease. J Gastroenterol. 2013;48:1392–400.

97. Takeuchi K, Ohishi M, Endo K, Suzumura K, Naraoka H, Ohata T, Seki J, Miyamae Y, Honma M, Soga T. Metabolomic analysis of the effects of omeprazole and famotidine on aspirin-induced gastric injury. Metabolomics. 2014;10:995–1004.

98. Soga T, Baran R, Suematsu M, Ueno Y, Ikeda S, Sakurakawa T, Kakazu Y, Ishikawa T, Robert M, Nishioka T, Tomita M. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. J Biol Chem. 2006;281:16768–76.

99. Zeng J, Yin P, Tan Y, Dong L, Hu C, Huang Q, Lu X, Wang H, Xu G. Metabolomics study of hepatocellular carcinoma: discovery and validation of serum potential biomarkers using capillary electrophoresis-mass spectrometry. J Proteome Res. 2014;13:3420–31.

100. Pont L, Benavente F, Jaumot J, Tauler R, Alberch J, Ginés S, Barbosa J, Sanz-Nebot V. Metabolic profiling for the identification of Huntington biomarkers by on-line solid-phase extraction capillary electrophoresis-mass spectrometry combined with advanced data analysis tools. Electrophoresis. 2016;37:795–808.

101. González-Peña D, Dudzik D, Colina-Coca C, de Ancos B, García A, Barbas C, Sánchez-Moreno C. Evaluation of onion as a functional ingredient in the prevention of metabolic impairments associated to diet-induced hypercholesterolaemia using a multiplatform approach based on LC-MS CE-MS and GC-MS. J Funct Foods. 2015;19:363–75.

102. Chalcraft KR, Britz-McKibbin P. Newborn screening of inborn errors of metabolism by capillary electrophoresis-electrospray ionization-mass spectrometry: a second-tier method with improved specificity and sensitivity. Anal Chem. 2009;81:307–14.

103. Soga T, Sugimoto M, Honma M, Mori M, Igarashi K, Kashikura K, Ikeda S, Hirayama A, Yamamoto T, Yoshida H, Otsuka M, Tsuji S, Yatomi Y, Sakuragawa T, Watanabe H, Nihei K, Saito T, Kawata S, Suzuki H, Tomita M, Suematsu M. Serum metabolomics reveals γ-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. J Hepatol. 2011;55:896–905.

104. Hirayama A, Igarashi K, Tomita M, Soga T. Development of quantitative method for determination of γ-glutamyl peptides by capillary electrophoresis tandem mass spectrometry: an efficient approach avoiding matrix effect. J Chromatogr A. 2014;1369:161–9.

105. Harada S, Takebayashi T, Kurihara A, Akiyama M, Suzuki A, Hatakeyama Y, Sugiyama D, Kuwabara K, Takeuchi A, Okamura T, Nishiwaki Y, Tanaka T, Hirayama A, Sugimoto M, Soga T, Tomita M. Metabolomic profiling reveals novel biomarkers of alcohol intake and alcohol-induced liver injury in community-dwelling men. Environ Health Prev Med. 2016;21:18–26.

106. Naz S, Garcia A, Rusak M, Barbas C. Method development and validation for rat serum fingerprinting with CE-MS: application to ventilator-induced-lung-injury study. Anal Bioanal Chem. 2013;405:4849–58.

107. Katja BP, Elgstoen J, Zhao Y, Anacleto JF, Jellum E. Potential of capillary electrophoresis, tandem mass spectrometry and coupled capillary electrophoresis-tandem mass spectrometry as diagnostic tools. J Chromatogr A. 2001;914:265–75.

108. Kuehnbaum NL, Britz-McKibbin P. Comprehensive profiling of free and conjugated estrogens by capillary electrophoresis time-of-flight/mass spectrometry. Anal Chem. 2011;83:8063–8.

109. Hao L, Zhong X, Greer T, Ye H, Li L. Relative quantification of amine-containing metabolites using isobaric N, N-dimethyl leucine (DiLeu) reagents via LC-ESI-MS/MS and CE-ESI-MS/MS. Analyst. 2015;140:467–75.

110. Arvidsson B, Johannesson N, Citterio A, Righetti PG, Bergquist J. High throughput analysis of tryptophan metabolites in a complex matrix using capillary electrophoresis coupled to time-of-flight mass spectrometry. J Chromatogr A. 2007;1159:154–8.

111. D'Agostino LA, Lam KP, Lee R, Britz-McKibbin P. Comprehensive plasma thiol redox status determination for metabolomics. J Proteome Res. 2011;10:592–603.

112. Kok MGM, de Jong GJ, Somsen GW. Sensitivity enhancement in capillary electrophoresis-mass spectrometry of anionic metabolites using a triethylamine containing background electrolyte and sheath liquid. Electrophoresis. 2011;32:3016–24.

113. Ramautar R, Mayboroda OA, Deelder AM, Somsen GW, de Jong GJ. Metabolic analysis of body fluids by capillary electrophoresis using noncovalently coated capillaries. J Chromatogr B. 2008;871:370–4.

114. Ramautar R, Toraño JS, Somsen GW, de Jong GJ. Evaluation of CE methods for global metabolic profiling of urine. Electrophoresis. 2010;31:2319–27.

115. Ramautar R, Nevedomskaya E, Mayboroda OA, Deelder AM, Wilson ID, Gika HG, Theodoridis GA, Somsen GW, de Jong GJ. Metabolic profiling of human urine by CE-MS using a positively charged capillary coating and comparison with UPLC-MS. Mol Biosyst. 2011;7:194–9.

116. Hirayama A, Tomita M, Soga T. Sheathless capillary electrophoresis-mass spectrometry with a high-sensitivity porous sprayer for cationic metabolome analysis. Analyst. 2012;137:5026–33.

117. Ramautar R, Busnel JM, Deelder AM, Mayboroda OA. Enhancing the coverage of the urinary metabolome by sheathless capillary electrophoresis-mass spectrometry. Anal Chem. 2012;84:885–92.

118. Ramautar R, Shyti R, Schoenmaker B, de Groote L, Derks RJ, Ferrari MD, van den Maagdenberg AM, Deelder AM, Mayboroda OA. Metabolic profiling of mouse cerebrospinal fluid by sheathless CE-MS. Anal Bioanal Chem. 2012;404:2895–900.

119. Peterson ZD, Collins DC, Bowerbank CR, Lee ML, Graves SW. Determination of catecholamines and metanephrines in urine by CE-ESI-TOF mass spectrometry. J Chromatogr B. 2002;776:221–9.

120. Benavente F, van der Heijden R, Tjaden UR, van der Greef J, Hankemeier T. Metabolite profiling of human urine by CE-ESI-MS using separation electrolytes at low pH. Electrophoresis. 2006;27:4570–84.

121. Desiderio C, De Rossi A, Inzitari R, Mancinelli A, Rossetti DV, Castagnola M, Messana I. Optimization of a rapid capillary electrophoresis ESI-IT tandem mass spectrometry method for the analysis of short-chain carnitines in human plasma. Anal Bioanal Chem. 2008;390:1637–44.

122. Soga T, Igarashi K, Ito C, Mizobuchi K, Zimmermann HP, Tomita M. Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. Anal Chem. 2009;81:6165–74.

123. Soga T, Kakazu Y, Robert M, Tomita M, Nishioka T. Qualitative and quantitative analysis of amino acids by capillary electrophoresis-electrospray ionization-tandem mass spectrometry. Electrophoresis. 2004;25:1964–72.

124. Shyti R, Kohler I, Schoenmaker B, Derks RJE, Ferrari MD, Tolner EA, Mayboroda OA, van den Maagdenberg AMJM. Plasma metabolic profiling after cortical spreading depression in a transgenic mouse model of hemiplegic migraine by capillary electrophoresis-mass spectrometry. Mol Biosyst. 2015;11:1462–71.

125. Soga T, Ueno Y, Naraoka H, Ohashi Y, Tomita M, Nishioka T. Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem. 2002;74:2233–9.

126. Soga T, Ueno Y, Naraoka H, Matsuda K, Tomita M, Nishioka T. Pressure-assisted capillary electrophoresis electrospray ionization mass spectrometry for analysis of multivalent anions. Anal Chem. 2003;74:6224–9.

127. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res. 2003;2:488–94.

128. Edwards JL, Chisolm CN, Shackmana JG, Kennedy RT. Negative mode sheathless capillary electrophoresis electrospray ionization-mass spectrometry for metabolite analysis of prokaryotes. J Chromatogr A. 2006;1106:80–8.

129. Timischl B, Dettmer K, Kaspar H, Thieme M, Oefner PJ. Development of a quantitative, validated capillary electrophoresis-time-of-flight-mass spectrometry method with integrated high-confidence analyte identification for metabolomics. Electrophoresis. 2008;29:2203–14.
130. Aerts JT, Louis KR, Crandall SR, Govindaiah G, Cox CL, Sweedler JV. Patch clamp electrophysiology and capillary electrophoresis-mass spectrometry metabolomics for single cell characterization. Anal Chem. 2014;86:3203–8.
131. Ibáñez C, Simó C, Valdés A, Campone L, Piccinelli AL, García-Cañas V, Cifuentes A. Metabolomics of adherent mammalian cells by capillary electrophoresis-mass spectrometry: HT-29 cell as case study. J Pharm Biomed Anal. 2015;110:83–92.
132. Gulersonmez MC, Lock S, Hankemeier T, Ramautar R. Sheathless capillary electrophoresis-mass spectrometry for anionic metabolic profiling. Electrophoresis. 2016;37:1007–14.
133. Lindenburg PW, Ramautar R, Jayo RG, Chen DDY, Hankemeier T. Capillary electrophoresis-mass spectrometry using a flow-through microvial interface for cationic metabolome analysis. Electrophoresis. 2014;35:1308–14.
134. Saric J, Want EJ, Duthaler U, Lewis M, Keiser J, Shockcor JP, Ross GA, Nicholson JK, Holmes E, Tavares MFM. Systematic evaluation of extraction methods for multiplatform-based metabotyping: application to the *Fasciola hepatica* metabolome. Anal Chem. 2012;84:6963–72.
135. Zhao C, Du H, Xu L, Wang J, Tang L, Cao Y, Li C, Wang Q, Liu Y, Shan F, Feng J, Xu F, Gao P. Metabolomic analysis revealed glycylglycine accumulation in astrocytes after methionine enkephalin administration exhibiting neuron protective effects. J Pharm Biomed Anal. 2015;115:48–54.
136. Barbas-Bernardos C, Armitage EG, García A, Mérida S, Navea A, Bosch-Morell F, Barbas C. Looking into aqueous humor through metabolomics spectacles - exploring its metabolic characteristics in relation to myopia. J Pharm Biomed Anal. 2016;127:18–25.
137. Tsuruoka M, Hara J, Hirayama A, Sugimoto M, Soga T, Shankle WR, Tomita M. Capillary electrophoresis-mass spectrometry-based metabolome analysis of serum and saliva from neurodegenerative dementia patients. Electrophoresis. 2013;34:2865–72.
138. Mayboroda OA, Neussus C, Pelzing M, Zurek G, Derks R, Meulenbelt I, Kloppenburg M, Slagboom EP, Deelder AM. Amino acid profiling in urine by capillary zone electrophoresis - mass spectrometry. J Chromatogr A. 2007;1159:149–53.
139. Toyohara T, Suzuki T, Akiyama Y, Yoshihara D, Takeuchi Y, Mishima E, Kikuchi K, Suzuki C, Tanemoto M, Ito S, Nagao S, Soga T, Abe T. Metabolomic profiling of the autosomal dominant polycystic kidney disease rat model. Clin Exp Nephrol. 2011;15:676–87.
140. Kuehnbaum NL, Kormendi A, Britz-McKibbin P. Multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform for metabolomics with high data fidelity. Anal Chem. 2013;85:10664–9.
141. Hirayama A, Sugimoto M, Suzuki A, Hatakeyama Y, Enomoto A, Harada S, Soga T, Tomita M, Takebayashi T. Effects of processing and storage conditions on charged metabolomic profiles in blood. Electrophoresis. 2015;36:2148–55.
142. Koike S, Bundo M, Iwamoto K, Suga M, Kuwabara H, Ohashi Y, Shinoda K, Takano Y, Iwashiro N, Satomura Y, Nagai T, Natsubori T, Tada M, Yamasue H, Kasai K. A snapshot of plasma metabolites in first-episode schizophrenia: a capillary electrophoresis time-of-flight mass spectrometry study. Transl Psychiatry. 2014;4:e379.
143. Naz S, García A, Barbas C. Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. Anal Chem. 2013;85:10941–8.
144. Ramautar R, Mayboroda OA, Derks RJE, van Nieuwkoop C, van Dissel JT, Somsen GW, Deelder AM, de Jong GJ. Capillary electrophoresis-time-of-flight-mass spectrometry using noncovalently bilayer-coated capillaries for the analysis of amino acids in human urine. Electrophoresis. 2008;29:2714–22.
145. Rodrigues KT, Mekahli D, Tavares MFM, Van Schepdael A. Development and validation of a CE-MS method for the targeted assessment of amino acids in urine. Electrophoresis. 2016;37:1039–47.

146. Ullsten S, Danielsson R, Bäckström D, Sjöberg P, Bergquist J. Urine profiling using capillary electrophoresis-mass spectrometry and multivariate data analysis. J Chromatogr A. 2006;1117:87–93.
147. Somsen GW, Mol R, de Jong GJ. On-line micellar electrokinetic chromatography-mass spectrometry: feasibility of direct introduction of non-volatile buffer and surfactant into the electrospray interface. J Chromatogr A. 2003;1000:953–61.
148. Moreno-González D, Toraño JS, Gámiz-Gracia L, García-Campana AM, de Jong GJ, Somsen GW. Micellar electrokinetic chromatography–electrospray ionization mass spectrometry employing a volatile surfactant for the analysis of amino acids in human urine. Electrophoresis. 2013;34:2615–22.
149. Sirén H, Seppanen-Laakso T, Oresic M. Capillary electrophoresis with UV detection and mass spectrometry in method development for profiling metabolites of steroid hormone metabolism. J Chromatogr B. 2008;871:375–82.
150. Blas M, McCord BR. Determination of trace levels of benzodiazepine in urine using capillary electrochromatography-time of flight mass spectrometry. Electrophoresis. 2008;29:2182–92.
151. Wu Q, Yu X, Wang Y, Gu X, Ma X, Lv W, Chen Z, Yan C. Pressurized CEC coupled with QTOF-MS for urinary metabolomics. Electrophoresis. 2014;35:2470–8.
152. Schmitt-Kopplin P, Frommberger M. Capillary electrophoresis - mass spectrometry: 15 years of developments and applications. Electrophoresis. 2003;24:3837–67.
153. Hernández-Borges J, Neusüß C, Cifuentes A, Pelzing M. On-line capillary electrophoresis-mass spectrometry for the analysis of biomolecules. Electrophoresis. 2004;25:2257–81.
154. Assunção NA, Bechara EJH, Simionato AVC, Tavares MFM, Carrilho E. Eletroforese capilar acoplada à espectrometria de massas (CE-MS): vinte anos de desenvolvimento. Quim Nova. 2008;31:2124–33.
155. Klampfl CW. CE with MS detection: a rapidly developing hyphenated technique. Electrophoresis. 2009;30:S83–91.
156. Causon TJ, Maringer L, Buchberger W, Klampfl CW. Addition of reagents to the sheath liquid: a novel concept in capillary electrophoresis-mass spectrometry. J Chromatogr A. 2014;1343:182–7.
157. Naz S, Vallejo M, García A, Barbas C. Method validation strategies involved in non-targeted metabolomics. J Chromatogr A. 2014;1353:99–105.
158. Reiter SM, Buchberger W, Klampfl CW. Investigations on the influence of interface-geometry in CE-ESI-TOF-MS. Chromatographia. 2010;71:715–9.
159. Wachs T, Sheppard RL, Henion J. Design and applications of a self-aligning liquid junction-electrospray interface for capillary electrophoresis-mass spectrometry. J Chromatogr B Biomed Appl. 1996;685:335–42.
160. Maxwell EJ, Zhong X, Zhang H, van Zeijl N, Chen DDY. Decoupling CE and ESI for a more robust interface with MS. Electrophoresis. 2010;31:1130–7.
161. Issaq HJ, Janini GM, Chan KC, Veenstra TD. Sheathless electrospray ionization interfaces for capillary electrophoresis–mass spectrometric detection: advantages and limitations. J Chromatogr A. 2004;1053:37–42.
162. Zamfir AD. Recent advances in sheathless interfacing of capillary electrophoresis and electrospray ionization mass spectrometry. J Chromatogr A. 2007;1159:2–13.
163. Maxwell EJ, Chen DDY. Twenty years of interface development for capillary electrophoresis-electrospray ionization-mass spectrometry. Anal Chim Acta. 2008;627:25–33.
164. Moini M. Simplifying CE-MS operation. 2. Interfacing low-flow separation techniques to mass spectrometry using a porous tip. Anal Chem. 2007;70:4241–6.
165. Haselberg R, Ratnayake CK, de Jong GJ, Somsen GW. Performance of a sheathless porous tip sprayer for capillary electrophoresis–electrospray ionization-mass spectrometry of intact proteins. J Chromatogr A. 2010;1217:7605–11.
166. Ramautar R, Somsen GW, de Jong GJ. Direct sample injection for capillary electrophoretic determination of organic acids in cerebrospinal fluid. Anal Bioanal Chem. 2007;387:293–301.

167. Yang WC, Regnier FE, Adamec J. Comparative metabolite profiling of carboxylic acids in rat urine by CE-ESI MS/MS through positively pre-charged and 2H-coded derivatization. Electrophoresis. 2008;29:4549–60.
168. Lin X, Gerardi AR, Breitbach ZS, Armstrong DW, Colyer CL. CE-ESI-MS analysis of singly charged inorganic and organic anions using a dicationic reagent as a complexing agent. Electrophoresis. 2009;30:3918–25.
169. Jellum E, Thorsrud AK, Time E. Capillary electrophoresis for diagnosis and studies of human disease, particularly metabolic disorders. J Chromatogr. 1991;559:455–65.
170. Jellum E, Dollekamp H, Brunsvig A, Gislefoss R. Diagnostic applications of chromatography and capillary electrophoresis. J Chromatogr B Biomed Sci Appl. 1997;689:155–64.
171. Jellum E, Dollekamp H, Blessum C. Capillary electrophoresis for clinical problem solving: analysis of urinary diagnostic metabolites and serum proteins. J Chromatogr B Biomed Sci Appl. 1996;683:55–65.
172. Barbas C, Adeva N, Aguilar R, Rosillo M, Rubio T, Castro M. Quantitative determination of short-chain organic acids in urine by capillary electrophoresis. Clin Chem. 1998;44:1340–2.
173. García A, Barbas C, Aguilar R, Castro M. Capillary electrophoresis for rapid profiling of organic acidurias. Clin Chem. 1998;44:1905–11.
174. Wakayama M, Hiramaya A, Soga T. Capillary electrophoresis-mass spectrometry. Methods Mol Biol. 2015;1277:113–22.
175. Pejchinovski M, Hrnjez D, Ramirez-Torres A, Bitsika V, Mermelekas G, Vlahou A, Zürbig P, Mischak H, Metzger J, Koeck T. Capillary zone electrophoresis on-line coupled to mass spectrometry: a perspective application for clinical proteomics. Proteomics Clin Appl. 2015;9:453–68.
176. Stalmach A, Husi H, Mosbahi K, Albalat A, Mullen W, Mischak H. Methods in capillary electrophoresis coupled to mass spectrometry for the identification of clinical proteomic/peptidomic biomarkers in biofluids. Methods Mol Biol. 2015;1243:187–205.
177. Simó C, Ibanez C, Gomez-Martinez A, Ferragut JA, Cifuentes A. Is metabolomics reachable? Different purification strategies of human colon cancer cells provide different CE-MS metabolite profiles. Electrophoresis. 2011;32:1765–77.
178. Garcia A, Naz S, Barbas C. Metabolite fingerprinting by capillary electrophoresis-mass spectrometry. Methods Mol Biol. 2014;1198:107–23.
179. Barbas C, Vallejo M, García A, Barlow D, Hanna-Brown M. Capillary electrophoresis as a metabolomic tool in antioxidant therapy studies. J Pharm Biomed Anal. 2008;47:388–98.
180. Nevedomskaya E, Derks R, Deelder AM, Mayboroda OA, Palmblad M. Alignment of capillary electrophoresis–mass spectrometry datasets using accurate mass information. Anal Bioanal Chem. 2009;395:2527–33.
181. Sugimoto M, Hirayama A, Ishikawa T, Robert M, Baran R, Uehara K, Kawai K, Soga T, Tomita M. Differential metabolomics software for capillary electrophoresis-mass spectrometry data analysis. Metabolomics. 2010;6:27–41.
182. Kok MG, Ruijken MM, Swann JR, Wilson ID, Somsen GW, de Jong GJ. Anionic metabolic profiling of urine from antibiotic-treated rats by capillary electrophoresis-mass spectrometry. Anal Bioanal Chem. 2013;405:2585–94.
183. Sugimoto M, Hirayama A, Robert M, Abe S, Soga T, Tomita M. Prediction of metabolite identity from accurate mass, migration time prediction and isotopic pattern information in CE-TOFMS data. Electrophoresis. 2010;31:2311–8.
184. Lee R, Britz-McKibbin P. Differential rates of glutathione oxidation for assessment of cellular redox status and antioxidant capacity by capillary electrophoresis-mass spectrometry: an elusive biomarker of oxidative stress. Anal Chem. 2009;81:7047–56.
185. Chalcraft KR, Lee R, Mills C, Britz-McKibbin P. Virtual quantification of metabolites by capillary electrophoresis-electrospray ionization-mass spectrometry: predicting ionization efficiency without chemical standards. Anal Chem. 2009;81:2506–15.

Part II: Statistical Analysis and Data Interpretation

Chapter 6
Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis

Ibrahim Karaman

Abstract From data acquisition to statistical analysis, metabolomics data need to undergo several processing steps, which are crucial for the data quality and interpretation of the results. In this chapter, methods for preprocessing, normalization, and pretreatment of metabolomics data generated from nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) are presented and discussed. Preprocessing is reported for both NMR and MS analysis. The challenges in preprocessing such complex data are highlighted. Subsequently, normalization methods such as total area normalization, probabilistic quotient normalization, and quantile normalization are explained. Finally, several scaling and data transformation methods are discussed for metabolomics data pretreatment, which is an important step prior to statistical analysis.

I. Karaman
Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, St. Mary's Campus, Norfolk Place, W2 1PG London, UK
e-mail: i.karaman@imperial.ac.uk

145

Abbreviations

ANOVA	Analysis of variance
CPMG	Carr-Purcell-Meiboom-Gill
GC	Gas chromatography
glog	Generalized log
LC	Liquid chromatography
LOESS	Locally estimated smoothing
m/z	Mass-to-charge ratio
MS	Mass spectrometry
NMR	Nuclear magnetic resonance spectroscopy
PCA	Principal component analysis
PLSR	Partial least squares regression
R^2	Linear regression coefficient
RSD	Relative standard deviation
RT	Retention time
QCs	Quality control samples
TSP	3-trimethylsilylpropionic acid

6.1   Introduction

Metabolomics analysis in clinical applications is often performed by either NMR or LC/GC-MS [1–4]. These platforms generate highly complex high-throughput data when biofluids are analyzed. NMR is a non-destructive and reproducible technique because the sample and the instrument do not physically interact [5]. In contrast, LC/GC-MS are destructive and less reproducible techniques [6]. However, LC/GC-MS have higher sensitivity compared to NMR [7]. Advanced technologies in the instrumentation offer fast and inexpensive solutions for metabolomics analysis and provide the opportunity of analyzing more than a thousand samples in an experimental run [8, 9]. The complexity of the data escalates, and consequently the data must go through various preprocessing and quality control steps prior to statistical analysis. There are many available methods and softwares for preprocessing, and new methods are developed or novel softwares are released constantly. The recent methodologies in omics data preprocessing can be followed via online web-based tools [10]. In Fig. 6.1, a generalized workflow for metabolomics data analysis is shown, considering the steps from data acquisition to statistical analysis. After the samples are analyzed by the instrument, the raw data of each sample need to be converted and processed in order to be summarized in a data table. The rows and the columns must be as comparable as possible after all the processing steps. This chapter focuses on the three blocks at the middle of this workflow and aims to give better understanding to the reader about the various steps of preprocessing and pretreatment. The following sections are fashioned according to Fig. 6.1 as Sects. 6.2, 6.3, 6.4, and 6.5.
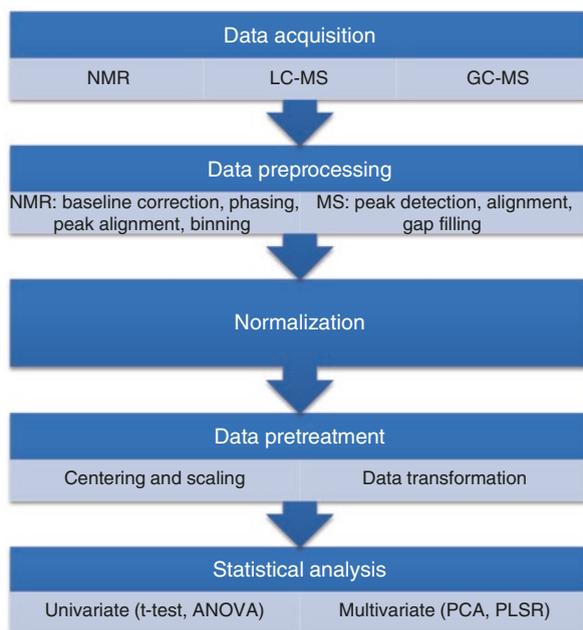
Fig. 6.1 General processing steps of metabolomics data analysis, from data acquisition to statistical analysis

6.2   Preprocessing of LC/GC-MS Data

In MS-based analyses, the measured variables are mass-to-charge ratios (*m/z*). When MS is combined with LC or GC, an additional dimension is added to the variable space, which is the chromatographic retention time. Therefore, raw LC/GC-MS data consist of a 3D structure of *m/z*, retention time (RT), and intensity count. In Fig. 6.2, an LC-MS profile of a blood serum sample is demonstrated in a specific retention time interval. Raw LC-MS data have many data points in one sample, as the data are often acquired in high-resolution instruments. Most of the data are generally either spectral noise or not biologically relevant (column material, contaminants, etc.). Therefore, it is necessary to convert each 2D sample profile into a 1D vector of peak areas/intensities.

The aim of preprocessing is to generate a 2D data table of features where the rows correspond to the study samples and the columns to *m/z*-RT pairs. There are several preprocessing steps in order to achieve this, and various softwares are available to perform the preprocessing, such as MarkerLynx (Waters), MassHunter (Agilent), MarkerView (AB Sciex), XCMS [11], MZmine 2 [12], and Progenesis QI (Waters). The LC/GC-MS data preprocessing steps are:

(a) *Peak picking/detection and deconvolution*: Peak picking is a crucial step of the preprocessing pipeline. It aims to detect each measured ion in a sample and to

assign to a feature (*m/z*-RT pair). In this step, the peak picking algorithm captures and deconvolutes peaks from the extracted ion chromatograms taking possible baseline and noise structures into account. If necessary, smoothing such as moving average or Savitzky-Golay filters can be applied during this step.

(b) *Alignment*: Improvements in the technology of mass spectrometers provide good reproducibility in the *m/z* dimension; however, reproducibility may be a problem in the RT dimension especially for LC-MS experiments. During chromatographic separation, RT shifts can occur due to changes in the mobile phase and the column stationary phase, variations in temperature and pressure, column aging, or effects related to sample matrix. Therefore, a metabolite can be eluted in slightly different retention times across the samples. This problem is crucial when hundreds or thousands of samples are analyzed in a long experimental run. Alignment algorithm aims to group detected peaks across the samples with respect to a *m/z* and a RT window. The grouped peaks are subsequently integrated as peak height or peak area and assigned to a feature in the data table.

(c) *Gap filling*: The data table after peak picking and alignment will contain missing values (gaps) in some of the samples. The reason of the presence of missing values is generally the existence of badly shaped peaks, which can be missed during the peak picking process, and peaks with low intensity, which cannot be detected during the peak picking process. Some of the preprocessing algorithms have gap-filling algorithms where peak structures are searched in the raw data on the defined *m/z* and RT window. This approach is useful when large peaks are missed during peak picking. There are also missing value estimation methods in literature [13], such as k-nearest neighbor imputation method. Care must be taken when using such methods because the imputations are based on the complete part of the whole data, which may not be the best representation for imputing the missing values.

After the initial preprocessing steps, the data table is complete without missing values. The next step is to assess the quality of the features in the data table. For metabolomics studies, it is recommended to analyze quality control samples (QCs) after every couple of (between 5 and 10) study samples in the entire sample run in order to monitor the experiment [14, 15]. The QCs are prepared by pooling the study samples; therefore, they represent the whole sample set. By looking at the QCs, it is possible to assess each feature in the data table for:

(a) *Presence in the QCs*: Some preprocessing softwares provide the number of samples, which are present in a predefined sample group (the QCs in this case). Features that are not present in a certain number of QCs can be filtered out from the data table. This filtering step assumes the sample set is well represented by the QCs.

(b) *Intensity drifts*: As data acquisition takes a significant amount of time, it is common to observe intensity drifts, which cause intra- and inter-batch variation throughout the analysis. These drifts are specific to each feature and cannot be handled by sample normalization. Therefore, each feature has to be examined separately. There are methods available to remove intensity drifts [14, 16, 17], and a common method is to fit a nonlinear locally estimated smoothing (LOESS) curve to the intensities of the QCs along the experimental run order. Thereafter, a correction factor for each study sample is estimated by interpolating the LOESS curve to the experimental run of the study samples. These correction factors are used to remove intensity drifts in each feature by dividing the intensity by the correction factor. In Fig. 6.3, the effect of the drift correction on the data is demonstrated. This drift correction step is important, and care must be taken when applying, because there should not be outliers among the QCs, and they may require normalization beforehand.

(c) *Repeatability*: Each feature in the QCs should have low relative standard deviation (RSD) across the QCs throughout the experimental run in order to have a good repeatability. RSD for each feature is calculated by dividing the sample standard deviation by the sample mean. The features with high percent RSD values should subsequently be removed from the data table. The suggested
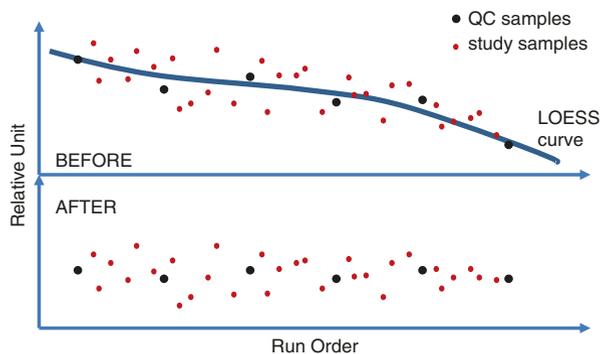
Fig. 6.3 Graphical representation of intensity drift correction using LOESS curve

threshold is 20 % for LC-MS and 30 % for GC-MS [14], but it may be more flexible depending on the size of the sample set.

(d) *Linearity*: A series of QC samples with varying dilutions can be prepared and analyzed within the experimental run [18]. The dilution factors can thus be regressed against the corresponding intensities of each feature in the data table. The features with low R^2 and with negative beta coefficients are thus removed from the data table. An R^2 threshold between 0.5 and 0.7, which is not very stringent, is suggested; nonetheless, it depends on the study sample size. Inspecting the distribution of the R^2 values may provide help in deciding the threshold.

At the end of the preprocessing steps, the data table is generated by features of *m/z*-RT pairs after filtering based on the QCs and correcting for instrumental drifts. The columns of the data table are to the best extent made comparable for further analysis.

6.3   Preprocessing of ^{1}H NMR Data

Preprocessing of metabolomics data acquired using ^{1}H NMR is crucial and challenging in clinical studies when blood (serum/plasma) and urine samples are analyzed. In NMR metabolomics, sample spectra can be acquired by different NMR experiments, such as standard 1D ^{1}H NMR experiment, 1D ^{1}H Carr-Purcell-Meiboom-Gill (CPMG) spin-echo NMR experiment, and 2D ^{1}H-^{1}H J-resolved NMR experiment. Each of these experiments contains water presaturation. CPMG experiment is specifically used for blood samples because it removes the broad baselines with respect to the macromolecules, such as the phospholipids and lipoproteins, in the blood.

In general, initial steps of preprocessing in ^{1}H NMR experiments involve apodization, Fourier transform, phasing, baseline correction, and chemical shift calibration. These steps are currently automated by the instrument vendor software and can be applied either manually or automatically according to the scientific problem.

Figure 6.4 demonstrates representative 1D ^{1}H NMR spectra for blood serum and urine after proper initial preprocessing steps. The spectral data acquisition range for these samples is δ -0.50–10.00 ppm by the instrumental setting because no bona fide metabolite signals are expected outside this region. By looking at the CPMG (Fig. 6.4a) and standard 1D (Fig. 6.4b) ^{1}H NMR spectra of blood serum, the broad baselines under the sharp peaks on the standard 1D ^{1}H NMR spectrum draw attention. The latter spectrum contains several broad resonances from macromolecules, and these broad resonances are highly overlapped with the low molecular mass metabolites with sharp peaks. Nevertheless, the information captured from both experiments is complementary. On the other hand, standard 1D ^{1}H NMR spectrum of urine (Fig. 6.4c) exhibits numerous sharp peaks throughout the spectral range. The broad baselines are observed only locally.
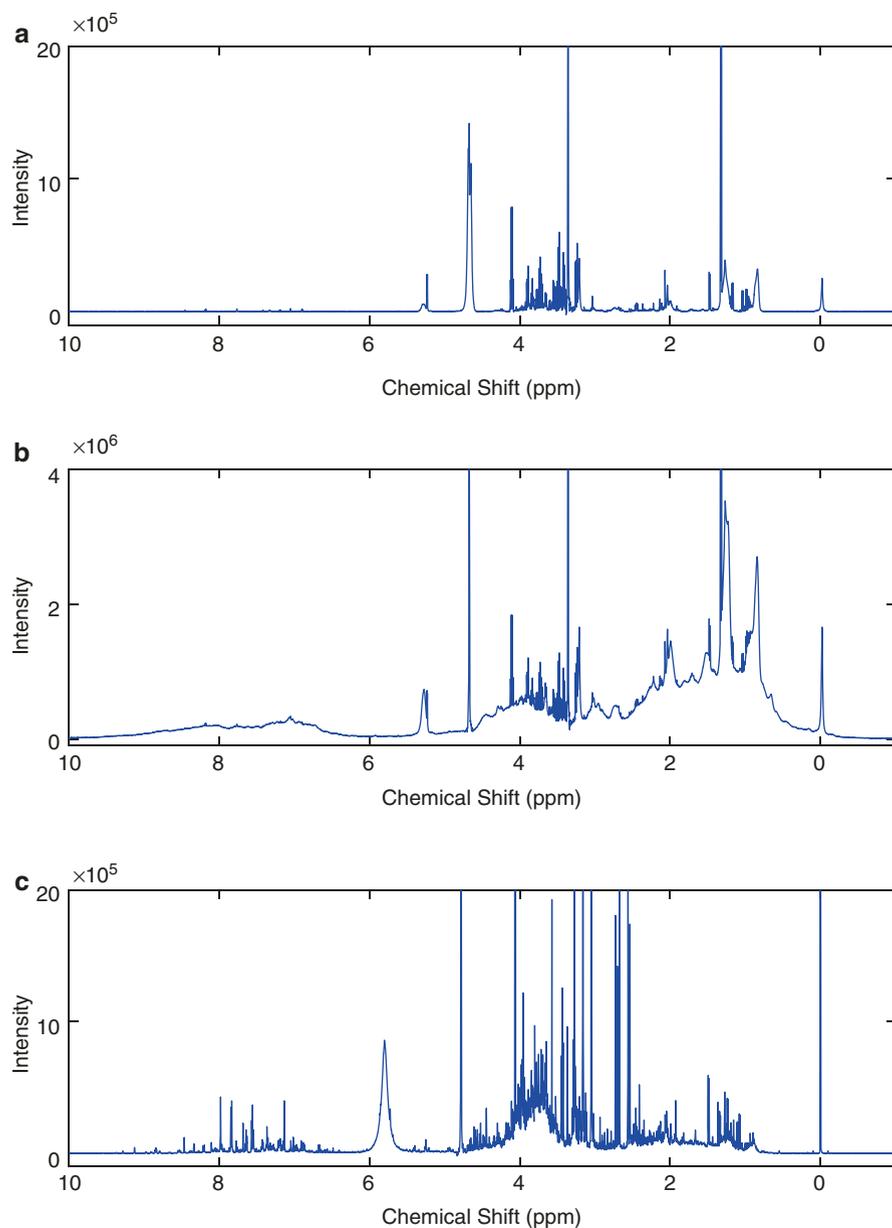
Fig. 6.4 Representative (**a**) ¹H NMR CPMG spectrum of blood serum, (**b**) ¹H NMR standard 1D spectrum of blood serum, and (**c**) ¹H NMR standard 1D spectrum of urine

Once all of the sample spectra are available, one can generate a data table where the rows correspond to the study samples and the columns to spectral data points. The columns of the data table should correspond to the same information after the preprocessing. During spectral data acquisition, peak shifts can be observed due to changes in pH, temperature, or fluctuations in the magnetic field. This will cause the columns of the data incomparable along the samples. Therefore, metabolite signals should be aligned and made comparable prior to statistical analysis. Below, some of the most commonly used approaches for NMR preprocessing are highlighted:

(a) *Using high-resolution ^{1}H NMR spectra*: The data can be analyzed as raw spectra or after applying a peak alignment algorithm. When peak shifts are systematic, they can be corrected by calibrating the spectra toward a reference peak, such as the singlet at δ 0.00 ppm due to the internal standard TSP added to every sample (see the sharp peak at δ 0.00 on the spectra in Fig. 6.4). For blood serum/plasma, the glucose doublet at δ 5.23 ppm is a suitable alternative to TSP since TSP may bind to proteins in serum/plasma samples, which will cause changes in peak shape and position. However, there may still be small but significant shifts in the peak positions between the samples. Applying peak alignment algorithms can correct shifts in the peak positions to some extent. Figure 6.5 depicts a hypothetical example of spectral data alignment. There are several algorithms available in the literature [19–21]. Most of these methods locate the position of the peaks in a sample spectrum and fit the corresponding chemical shift in a reference spectrum. Some of the common ones are icoshift and recursive segment-wise peak alignment algorithms. Both algorithms require a reference spectrum. A reference spectrum can be randomly selected, or a sample spectrum, which is the closest to the rest of the sample spectra, can be used. Alternative to using a sample spectrum as reference is creating a reference spectrum by calculating the mean or median spectrum from the entire sample set or the QCs. The major drawback of peak alignment methods is that they may not handle overlapping peaks correctly, especially when two adjacent peaks overlap
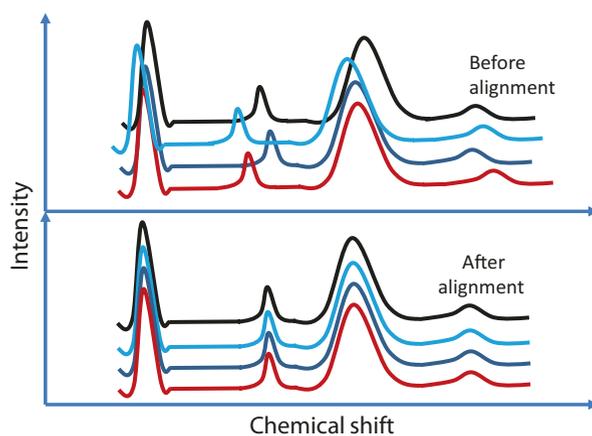
Fig. 6.5 Schematic representation of spectral data alignment

or even swap in position between samples due to different amount of peak shifts. Nonetheless, statistical analysis of spectra after carefully applying peak alignment algorithms can be a good compromise to analyzing raw spectra.

(b) *Using binned ¹H NMR spectra*: Binning or bucketing can be applied to raw spectra or to aligned spectra in order to correct for shifts in the peak positions on the raw spectra or small misalignments on the aligned spectra. By binning, spectral resolution is lowered by converting segments of the spectrum into a bin where the spectral data inside each segment are summed as area under the curve and represented by one single value. The data also become more compact and easy to handle computationally. There are several methods available in literature for binning [22, 23]. Although binning is attractive for lowering the resolution and handling the misalignments, care must be taken when applying. The binning method and the parameters should be selected with caution in order to obtain good data. Otherwise, peaks may fall into the wrong bins, peaks may be split, and obviously binning does not handle overlapping peaks.

(c) *Quantifying known metabolites*: When targeted metabolomics data analysis is the case, converting the spectral data table into a table of annotated and quantified metabolites is convenient even though new metabolites are not generally available for analysis. Automated quantification of metabolite levels from spectra is not an easy process due to peak overlapping, variations in peak positions, and spectral noise. In addition, building a calibration model is time consuming and requires standard sample spectra for calibration. There are methods in literature such as Bayesian automated metabolite analyzer [24] where peaks from 1D ¹H NMR spectra are deconvoluted and assigned to specific metabolites from a known metabolite list.

As one can see above, all of the approaches have benefits and weaknesses. They aim to make the columns of the data table comparable between the samples. It is up to the analyst to decide how to proceed with the preprocessing, keeping in mind the various consequences. Assuming the columns of the data table are comparable after the preprocessing steps, the analyst can move to the next step, which is the normalization of the samples. If high-resolution or binned ¹H NMR spectra are used for further analysis, it is important to remove interfering spectral regions related to water suppression residual (δ 4.40–5.00 ppm) and possible contaminants in the samples. Peak of urea at around δ 5.80 ppm should be also removed from urine spectra. The reason is that they are not changing proportionally with the changes in concentration, and they may adversely influence the normalization of the samples.

6.4   Normalization

The term normalization here is used for the division of each row of data table by a normalization factor. Normalization procedure removes unwanted variation between the samples and allows quantitative comparison of the samples. In metabolomics,

study samples are biofluids in most cases, and they exhibit differences in the concentration of metabolites due to varying dilution factors for different samples. For example, metabolite concentrations in different urine samples may differ with respect to the amount of water as the solvent. Therefore, the measured metabolite concentrations will reflect to dilution instead of the changes in metabolic responses. In order to remove such variations between the samples, a normalization factor should be computed for each row of the data table. There are several ways for performing normalization [25–29]:

(a) *Addition of internal/external standard(s)*: A standard with known concentration can be added to every sample, and the samples can be normalized using the peak area/intensity of this standard. However, this method is not convenient for untargeted metabolomics because the source of unwanted variation is not only related to sample introduction to the instrument but also the variations in the dilution factors.

(b) *Total area normalization*: The normalization factor for each sample is computed by summing all of the features in the corresponding row. The disadvantage of this normalization is that changes in metabolite concentrations across the samples will affect the normalization factor because the technique assumes the total metabolite concentration in a sample does not change across the samples. High-concentration metabolites contribute to the total area, i.e., the normalization factor, more than the small-concentration metabolites. In the presence of a significant change in the peak intensity/area of a high-concentration metabolite, the normalization factor will be affected.

(c) *Probabilistic Quotient Normalization*: This method assumes that metabolite peaks affected by dilution will have the same fold changes between two samples. Fold changes of a sample are computed for every feature against a target spectrum/profile, which can be the median spectrum/profile. The normalization factor for that sample is the median value of the fold changes. This method is not affected by large changes in a few metabolites because it uses the median of many fold change values instead of an estimated single sum as for total area normalization.

(d) *Quantile normalization*: This method forces all samples in a sample set to have identical peak intensity/area distribution. The difference of quantile normalization from the previous ones is that there is no estimated normalization factor for each sample. First, each row of the data table is sorted from lowest to highest. Thereafter, mean/median of each column is calculated from the sorted data table. These mean/median values form the target spectrum/profile. All rows of the data table are replaced with the target spectrum/profile. Finally, the data table is restored into its original order before sorting. The rows of the new data table are composed by the normalized samples. This method can be problematic with high-value features in the data table because they can dramatically differ from sample to sample.

6.5 Data Pretreatment

When clean and normalized metabolomics data are ready for statistical analysis, it is important to use the appropriate data pretreatment method before starting [30, 31]. The data are converted into different forms by data pretreatment. The effects of technical and measurement errors are aimed to be reduced, whereas the relevant biological variations are aimed to be enhanced.

The choice of data pretreatment method depends on the scientific question and the data analysis method to be used. If univariate analysis is used, generally there is no need for a pretreatment. However, when multivariate analysis methods are considered, data pretreatment plays an important role in obtaining and interpreting the results. In Sects. 6.5.1 and 6.5.2, ways for data pretreatment are explained with a few example methods. A publically available exemplar LC-MS data set [32] was used for demonstrating the outcome of each data pretreatment method described. The preprocessed data table consists of 28 rows/samples and 168 columns/features. Although metabolomics data from clinical applications contain thousands of samples and features, the exemplar data set used here is sufficient for the readers to understand how data pretreatment works.

6.5.1 Centering and Scaling

In untargeted metabolomics studies with a purpose of biomarker discovery, multivariate analysis techniques based on latent variable projections such as PCA or PLSR are used. Such methods extract information from the data by projecting onto the direction of the maximum variance. Analyzing the data from NMR and MS platforms directly by latent variable projection techniques will focus on the average spectrum/profile, and any type of biological variation in the data will be masked. Therefore, mean-centering the data table, where the mean of a feature is subtracted from each element of the feature vector, is a common practice before PCA and PLSR, and generally it is applied by default. By mean centering, it is aimed to remove the offset from the data and focus on the biological variation, as well as similarities/dissimilarities among the samples in the data.

Metabolites that are more abundant will exhibit high values in the data table and subsequently show large differences among samples compared to the low-abundant metabolites. NMR and MS platforms are effective in quantifying low-abundant metabolites, as well as the highly abundant metabolites. As PCA and PLSR are focusing on the maximum variance, centering the data alone may not be enough to find biomarkers because the highly abundant metabolites will dominantly contribute to the model. The biologically important but low-abundant metabolites thus can be masked, and the results of the statistical analysis may become biased. Consequently, scaling each feature in the data table, which potentially corresponds

to a metabolite, needs to be carefully considered. In the following, the scaling operations are explained for one feature, i.e., column, in the whole set of features in the data set.

(a) *Auto-scaling (unit variance scaling)*: The mean and the standard deviation of the feature are calculated. The feature is first mean-centered. Thereafter each element in the mean-centered feature is divided by the standard deviation. The aim of auto-scaling is to give equal weights to all of the features. Therefore, metabolites with both low and high abundance will equally contribute to the multivariate model. The drawback of auto-scaling is that noisy and uninformative features will also be as important as the interesting features. Moreover, the measurement errors on the metabolites with low abundance will inflate as they are more affected. One needs to make sure that the features in the data table have good quality, i.e., noisy features or features with low repeatability/linearity are filtered in case of analyzing MS data. When NMR data analysis is considered, auto-scaling may be better used after removing noisy and outlying/contaminant regions from the spectra. Auto-scaling can be also useful when multivariate analysis is combined with variable selection.

(b) *Pareto scaling*: This is similar to auto-scaling but in this case, each element in the mean-centered feature is divided by the square root of the standard deviation. Pareto scaling is a compromise between mean-centering and auto-scaling because Pareto-scaled metabolites with high abundance are less dominant compared to the corresponding mean-centered ones. Nonetheless, the Pareto-scaled data are kept closer to the mean-centered data, and the drawbacks of using only mean-centering count also for Pareto scaling. Therefore, multivariate analysis may be still prone to focus on the metabolites with high abundance.

(c) *Range scaling*: The mean and the range of the feature are calculated. The range is defined as the difference between the minimum and the maximum values in the feature. Each element in the mean-centered feature is divided by the range in range scaling. Using the range as the scaling factor is risky as it is sensitive to only a few outlying samples in a large sample set. It can still be an alternative to auto-scaling when range is estimated robustly.

(d) *Vast (variable stability) scaling*: Each element in the auto-scaled feature is divided by the coefficient of variation, which is the ratio of the standard deviation and the mean. In contrast to auto-scaling where each feature equally contributes to the statistical model, the focus falls onto the more stable features after vast scaling. The assumption here is that important metabolic features should have small coefficient of variation, i.e., relative standard deviation, so that they will be more stable.

In Fig. 6.6, the effect of centering and scaling to features in sample 17 from the exemplar LC-MS data set is depicted. In panel (a), most of the features seem to have low abundance. There are a few very highly abundant features. In panel (b), mean-centering moved the features to distribute around zero, but the same highly abundant features are still present and dominating the data. This is also visible in panel (d) after the features were Pareto-scaled even though low-abundant features were
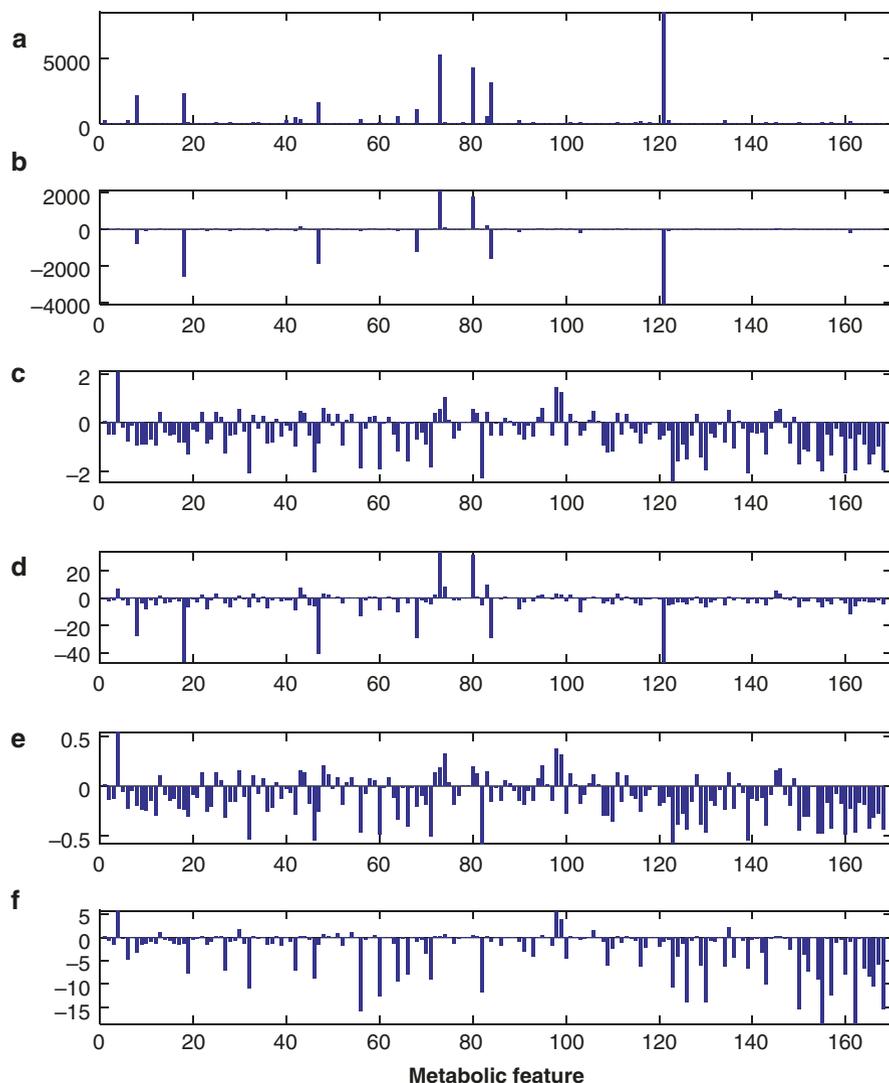
Fig. 6.6 Graphical representation of (**a**) untreated, (**b**) mean-centered, (**c**) auto-scaled, (**d**) Pareto-scaled, (**e**) range-scaled, and (**f**) vast-scaled features of sample 17 from the exemplar data set. Y-axes of each plot were left unlabeled because they are varying with respect to the pretreatment method

inflated to some extent. On the other hand, the features seem to be more comparable with each other after auto-scaling and range scaling in panel (c) and (e). In panel (f), the vast-scaled features seem to be more comparable compared to mean-centered and Pareto-scaled data; however, care must be taken because the features with high coefficient of variation were penalized.

6.5.2   Data Transformation

The data from NMR and MS platforms are generally subject to heteroscedastic noise from various sources where the amount of noise increases as a function of increased signal intensity. Statistical analysis tools assume the noise is homoscedastic where the noise is consistent across all features. Therefore, the data table may need to be transformed into a form in which the noise structure is no more heteroscedastic. Furthermore, the distributions of the features can be skewed and may need to be made close to normal prior to any type of statistical analysis. Transformations aim to correct for heteroscedasticity and skewness. They also have pseudo scaling effect on the features because the differences of the features with high and low abundances are substantially diminished. Notwithstanding, it may still be necessary to apply centering and scaling after transformation. In the following, a few common transformation operations are explained for each element of the entire data table.

(a) *Log transformation*: Logarithm of each element in the feature is calculated and replaced with the original data. In case of the presence of values between 0 and 1, 1 can be added to each element in the feature before the logarithm operation. Log transformation aims to convert multiplicative noise into additive noise.
(b) *Glog transformation*: This is similar to log transformation but logarithm operation is applied to $x + \sqrt{x^2 + \lambda}$ instead of x directly where x is the untransformed element in the data, and λ is the transform parameter. Glog transformation can be used as a scaling method after optimizing the transform parameter using a series of technical replicate samples [33]. Therefore, only biological variation will predominantly remain in the data table after glog transformation.
(c) *Power transformation*: Square root of each element in the feature is calculated and replaced with the original data. Although it does not convert the multiplicative noise into additive noise, it has similar effects as log transformation.

In Fig. 6.7, the effect of data transformation to features from the exemplar LC-MS data set is depicted. Homoscedastic data are supposed to have a flat distribution on such plots. In panel (a), some features with high average seem to have high standard deviation as well. This means the data set is heteroscedastic and needs to be made homoscedastic by data transformation. In panel (b) and (c), log and glog transformations seem to work well on this data set because the transformed data have a flat distribution. On the other hand, the power-transformed data do not have flat distribution, as can be seen on panel (d). The reason might be the presence of multiplicative error.

6.6   Concluding Remarks

In this chapter, the main preprocessing steps involved in metabolomics data analysis for NMR and LC/GC-MS platforms were summarized. Descriptions for commonly used methods for each step were briefly provided with
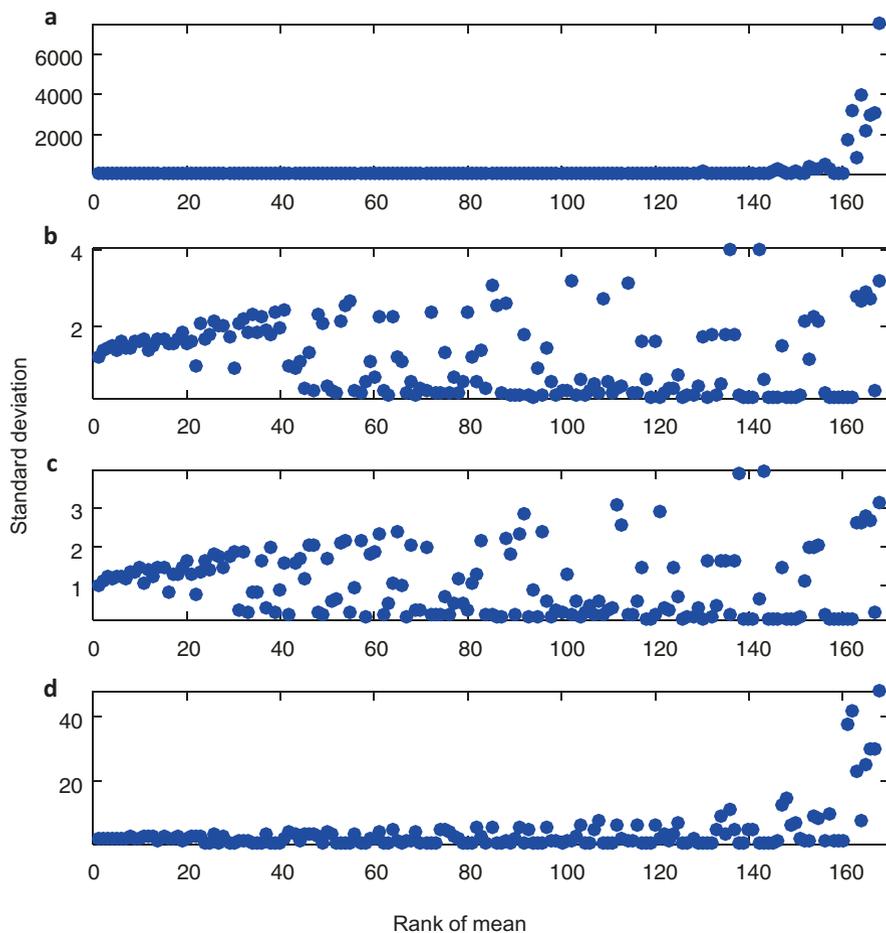
Fig. 6.7 Graphical representation of (**a**) untreated, (**b**) log-transformed, (**c**) glog-transformed, and (**d**) power-transformed features from the exemplar data set. Transform parameter was set to 10 when glog transformation was applied

discussions on their advantages and disadvantages. The purpose of preprocessing and normalization procedures is to extract clean and comparable data across the samples from the raw data. Pretreatment aims to focus on the biologically relevant information in the data. It is important to use methods that are convenient to the data set at hand in order to remove artifacts and variation without biological importance. The choice should not be biased according to the biological question; therefore, the methods must be chosen with respect to the assumptions and the limitations of the methods.

References

1. Emwas A-HM, Salek RM, Griffin JL, Merzaban J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. Metabolomics. 2013;9(5):1048–72. doi:10.1007/s11306-013-0524-y.

2. Robertson DG, Watkins PB, Reily MD. Metabolomics in toxicology: preclinical and clinical applications. Toxicol Sci. 2011;120(Suppl1):S146–S70. doi:10.1093/toxsci/kfq358.

3. Vermeersch KA, Styczynski MP. Applications of metabolomics in cancer research. J Carcinog. 2013;12:9. doi:10.4103/1477-3163.113622.

4. Yin P, Xu G. Current state-of-the-art of nontargeted metabolomics based on liquid chromatography–mass spectrometry with special emphasis in clinical applications. J Chromatogr A. 2014;1374:1–13. doi:http://dx.doi.org/10.1016/j.chroma.2014.11.050.

5. Lacy P, McKay RT, Finkel M, Karnovsky A, Woehler S, Lewis MJ, et al. Signal intensities derived from different NMR probes and parameters contribute to variations in quantification of metabolites. PLoS One. 2014;9(1):e85732. doi:10.1371/journal.pone.0085732.

6. Gika HG, Theodoridis GA, Wingate JE, Wilson ID. Within-day reproducibility of an HPLC−MS-based method for metabonomic analysis: application to human urine. J Proteome Res. 2007;6(8):3291–303. doi:10.1021/pr070183p.

7. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Anal Bioanal Chem. 2007;387(2):525–7. doi:10.1007/s00216-006-0687-8.

8. Lewis MR, Pearce JTM, Spagou K, Green M, Dona AC, Yuen AHY, et al. Development and application of ultra-performance liquid chromatography-TOF MS for precision large scale urinary metabolic phenotyping. Anal Chem. 2016. doi:10.1021/acs.analchem.6b01481.

9. Dona AC, Jiménez B, Schäfer H, Humpfer E, Spraul M, Lewis MR, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem. 2014;86(19):9887–94. doi:10.1021/ac5025039.

10. Henry VJ, Bandrowski AE, Pepin A-S, Gonzalez BJ, Desfeux A. OMICtools: an informative directory for multi-omic data analysis. Database. 2014. doi:10.1093/database/bau069.

11. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006;78(3):779–87.

12. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf. 2010;11:395.

13. Hrydziuszko O, Viant MR. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. Metabolomics. 2012;8(1):161–74. doi:10.1007/s11306-011-0366-4.

14. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011;6(7):1060–83. doi:http://www.nature.com/nprot/journal/v6/n7/abs/nprot.2011.335.html#supplementary-information.

15. Kamleh MA, Ebbels TMD, Spagou K, Masson P, Want EJ. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. Anal Chem. 2012;84(6):2670–7. doi:10.1021/ac202733q.

16. Fernández-Albert F, Llorach R, Garcia-Aloy M, Ziyatdinov A, Andres-Lacueva C, Perera A. Intensity drift removal in LC/MS metabolomics by common variance compensation. Bioinformatics. 2014. doi:10.1093/bioinformatics/btu423.

17. Kirwan JA, Broadhurst DI, Davidson RL, Viant MR. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. Anal Bioanal Chem. 2013;405(15):5147–57. doi:10.1007/s00216-013-6856-7.

18. Eliasson M, Rännar S, Madsen R, Donten MA, Marsden-Edwards E, Moritz T, et al. Strategy for optimizing LC-MS data processing in metabolomics: a design of experiments approach. Anal Chem. 2012;84(15):6869–76. doi:10.1021/ac301482k.

19. Veselkov KA, Lindon JC, Ebbels TMD, Crockford D, Volynkin VV, Holmes E, et al. Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery. Anal Chem. 2009;81(1):56–66. doi:10.1021/ac8011544.
20. Savorani F, Tomasi G, Engelsen SB. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson. 2010;202(2):190–202. doi:http://dx.doi.org/10.1016/j.jmr.2009.11.012.
21. Wong JWH, Durante C, Cartwright HM. Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets. Anal Chem. 2005;77(17):5655–61. doi:10.1021/ac050619p.
22. Blaise BJ, Shintu L, Elena B, Emsley L, Dumas M-E, Toulhoat P. Statistical recoupling prior to significance testing in nuclear magnetic resonance based metabonomics. Anal Chem. 2009;81(15):6242–51. doi:10.1021/ac9007754.
23. Sousa SAA, Magalhães A, Ferreira MMC. Optimized bucketing for NMR spectra: Three case studies. Chemom Intell Lab Syst. 2013;122:93–102. doi:http://dx.doi.org/10.1016/j.chemolab.2013.01.006.
24. Hao J, Liebeke M, Astle W, De Iorio M, Bundy JG, Ebbels TMD. Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. Nat Protoc. 2014;9(6):1416–27.
25. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem. 2006;78(13):4281–90. doi:10.1021/ac051632c.
26. Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19(2):185–93. doi:10.1093/bioinformatics/19.2.185.
27. Veselkov KA, Vingara LK, Masson P, Robinette SL, Want E, Li JV, et al. Optimized pre-processing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. Anal Chem. 2011;83(15):5864–72. doi:10.1021/ac201065j.
28. Sysi-Aho M, Katajamaa M, Yetukuri L, Orešič M. Normalization method for metabolomics data using optimal selection of multiple internal standards. BMC Bioinf. 2007;8(1):1–17. doi:10.1186/1471-2105-8-93.
29. Bijlsma S, Bobeldijk I, Verheij ER, Ramaker R, Kochhar S, Macdonald IA, et al. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. Anal Chem. 2006;78(2):567–74.
30. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006;7:142.
31. Bro R, Smilde AK. Centering and scaling in component analysis. J Chemom. 2003;17(1):16–33.
32. Acar E, Papalexakis EE, Gürdeniz G, Rasmussen MA, Lawaetz AJ, Nilsson M, et al. Structure-revealing data fusion. BMC Bioinf. 2014;15(1):1–17. doi:10.1186/1471-2105-15-239.
33. Parsons HM, Ludwig C, Günther UL, Viant MR. Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. BMC Bioinf. 2007;8(1):1–16. doi:10.1186/1471-2105-8-234.

Chapter 7
Chemometrics Methods and Strategies in Metabolomics

Rui Climaco Pinto

Abstract Chemometrics has been a fundamental discipline for the development of metabolomics, while symbiotically growing with it. From design of experiments, through data processing, to data analysis, chemometrics tools are used to design, process, visualize, explore and analyse metabolomics data.

In this chapter, the most commonly used chemometrics methods for data analysis and interpretation of metabolomics experiments will be presented, with focus on multivariate analysis. These are projection-based linear methods, like principal component analysis (PCA) and orthogonal projection to latent structures (OPLS), which facilitate interpretation of the causes behind the observed sample trends, correlation with outcomes or group discrimination analysis. Validation procedures for multivariate methods will be presented and discussed.

Univariate analysis is briefly discussed in the context of correlation-based linear regression methods to find associations to outcomes or in analysis of variance-based and logistic regression methods for class discrimination. These methods rely on frequentist statistics, with the determination of p-values and corresponding multiple correction procedures.

Several strategies of design-analysis of metabolomics experiments will be discussed, in order to guide the reader through different setups, adopted to better address some experimental issues and to better test the scientific hypotheses.

Keywords Metabolomics • Data analysis • Chemometrics • Multivariate • Exploratory • Regression • Classification • Discrimination • Discovery • Validation

R.C. Pinto

Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London,
St. Mary's Campus, Norfolk Place, W2 1PG, London, England, UK
e-mail: r.pinto@imperial.ac.uk

Abbreviations

ANOVA	Analysis of variance
ASCA	ANOVA-simultaneous component analysis
AUC	Area under the curve (in the context of ROC curves)
CV	Cross-validation
CV-ANOVA	Cross-validation – analysis of variance
FWER	Family-wise error rate
FDR	False discovery rate
GC-MS	Gas chromatography coupled to mass spectrometry
HCA	Hierarchical cluster analysis
ICA	Independent component analysis
iQC	Internal quality control (sample)
IS	Internal standard
LC-MS	Liquid chromatography coupled to mass spectrometry
LOO	Leave-one-out procedure in cross validation
MS	Mass spectrometry
MWAS	Metabolome-wide association studies
MWSL	Metabolome-wide significance level
OPLS	Orthogonal projections to latent structures
OPLS-DA	Orthogonal projections to latent structures – discriminant analysis
OPLS-EP	Orthogonal projections to latent structures – effect projection
PC	Principal component
PCA	Principal component analysis
PLS	Projections to latent structures
PRESS	Predicted residual error sum of squares
R2X	Fraction of variance in the data explained by each latent variable
R2Y	Fraction of variance of y/Y explained by each latent variable
ROC	Receiver operating characteristic (curve)
Q2	Model statistics to evaluate quality of model prediction
RMSECV	Root mean squared error of cross validation
RMSEP	Root mean squared error of prediction
SMART	Scaled-to-maximum, aligned and reduced trajectories
SUS	Shared and unique structures
VIP	Variable importance on projection

7.1   Introduction

Metabonomics [1] or metabolomics [2] concerns the study of the metabolome, a multivariate ensemble of small molecules that are intermediates and products of metabolism. Its main emphasis is on metabolite profiling, at the level of cells or organs, of endogenous and/or exogenous metabolites, and on the effects of perturbations of the metabolism caused by disease, environmental, or dietary influences.

Chemometrics can be defined as "the chemical discipline that uses mathematical, statistical, and other methods employing formal logic, to design or select optimal measurement procedures and experiments, and to provide maximum relevant chemical information by analysing chemical data" [3]. It differs from statistics in analytical chemistry mainly due to its computer intensive nature, and for being mostly multivariate analysis based [4]. Due to the nature of the signals in chemistry, namely in spectroscopy, chemometrics developed around the subject of multivariate analysis, because of its ease of interpretation. These are correlation-/projection-based methods, which require computationally intensive work. While bioinformatics and chemoinformatics are also used for data analysis, they are more related to data mining and use of databases. These disciplines have some overlap with chemometrics and methods like principal component analysis (PCA), for instance, are used by all of them.

Chemometrics is intensively used in the metabolomics context due to its experimental design component and to the fact that metabolic systems are multivariate in nature, with data mostly a product of ^{1}H NMR spectroscopy and gas/liquid chromatography coupled to mass spectrometry (GC/LC-MS). Metabolomics naturally relates to clinical research due to the fact that specific metabolite profiles express themselves in a living organism through a resulting health phenotype.

Clinical experiments exist in different areas and contexts, such as understanding biological processes and disease mechanisms, in vitro studies of materials of human origin, models of human disease processes, follow-up after surgery, epidemiological studies, diagnostic and therapeutic methods, effect and mechanism of vaccines and drugs, biomarker discovery and disease discrimination, among others [5]. Chemometrics may help unravel information from metabolomics in different aspects of each of these contexts. Although not specifically designed for clinical research, the methods presented in this chapter adapt to the field naturally, as they can be used to explore clinical metabolomics data.

This chapter is devoted to the uses of state-of-the-art chemometrics methods and their application to metabolomics data in clinical analysis.

7.2   Notation

Notation in the text is as follows: vectors are presented in bold lower-case (e.g. **y**), matrices in bold upper-case (e.g. **X**) and indexes in italic lower-case letters (e.g. i). The metabolite data matrix **X** consists of samples in i rows and metabolic features (or metabolites) in j columns. Each continuous or discrete outcome **y** (e.g. blood pressure) has the same length i as rows in **X**. To define classes for the two-class case, a dummy vector **y** (e.g. 0 = control; 1 = disease) is built. In case there are more than two classes, a dummy matrix **Y** with one vector per class is built. Confounder factors, when mentioned, are vectors **z** with the same length as the rows in **X**. Qualitative confounder vectors are transformed into dummy matrices the same way as described for multiple classes. In case there are two or more confounders, they are horizontally concatenated into a matrix **Z**. Transposed matrix is indicated by using the letter "T" in superscript, as in **X**^{T}.

7.3 Data Preprocessing

While using univariate analysis, there is no need for variable normalization (unless normal distribution is deemed necessary) because each metabolic feature is evaluated separately; however, in multivariate analysis, normalization is of utmost importance and depends on the analysis in question. As preprocessing, normalization, scaling and transformations of data are discussed in Chap. 6 of this book, they will not be herein discussed in detail. We assume the samples were already normalized with the objective of reducing magnitude effects (e.g. caused by different dilution levels), and the variables scaled in an appropriate way (e.g. ^{1}H NMR was Pareto scaled; LC-MS was centred and unit variance scaled) and potentially transformed adequately (log transformation or other). Both ^{1}H NMR and MS data are now considered a data matrix $\mathbf{X}$ of metabolic features ready for statistical analysis.

7.4 Chemometrics Contexts and Methods

The need for chemometrics tools arises around three decades ago, due to the development of more complex instruments with a consequent increase in the number of variables, and is propelled by the development of computational capacity. Large-scale dataset simultaneous visualization is more difficult in a univariate approach, and, for example, multiple regression modelling is constrained by variable colinearity. As referred previously, the chemometrics discipline is based on computing intensive methods, in general multivariate, which solves the colinearity problem in a covariance-/correlation-based framework.

There are many different multivariate methods for modelling data, as shown in previous literature reviews [6–8]. They can be unsupervised (no assumptions made on the samples) or supervised (samples are defined into classes, or each sample is associated to an outcome y_i value). Multivariate methods represent the samples as points in the space of the initial variables. The samples can then be projected into a lower dimensionality space – into components or latent variables – such as a line, a plane or a hyperplane, which can be seen as the "shadow" of the dataset viewed from its "best" viewpoint. The coordinates of the samples in the newly defined latent variables are defined as the scores, while the directions of variance to which they are projected are defined as the loadings. The loadings vector for each latent variable contains the weights of each of the initial variables in that latent variable. For a certain latent variable, the more a sample score is distant from its centre, the higher values it has in some of the initial variables (while potentially having lower values in others). Respectively, these initial variables have high weights in the loadings vector of that latent variable.

Projection-based linear methods are popular due to the simplicity of interpretation, thus used when understanding of a system is important. Nonlinear methods such as neural networks, support vector machines and random forests are less

common in metabolomics when interpretation is needed, and are used mostly for prediction of new samples in classification/regression contexts.

At the moment, due to the large amount of features involved in untargeted metabolomics, most of the statistical methods are applied previously to compound/metabolite identification. Only after finding a smaller number of important statistically significant metabolic features (putative metabolites), the analyst proceeds to the identification phase, as this may be very time-consuming. Bayesian networks have also been recently used in metabolomics but are not purely based on numerical metabolomics data. Because of their need for extra information, including metabolite identification and/or information from databases, these methods are considered to be more in the bioinformatics than in the chemometrics domain; thus, they will not be discussed here.

7.4.1 Multivariate Data Exploration (PCA)

The simplest correlation- and projection-based multivariate analysis linear method, and simultaneously the most widely used tool in chemometrics, is principal component analysis (PCA) [9–12]. It can be seen as the basis for other multivariate methods, thus being commonly used to introduce the concept of latent variables, and it is widely used as an exploration tool in metabolomics [13].

PCA is a non-supervised method. As it contains no assumptions on the data, it is used as a visualization and exploration tool at the start of any analysis, in order to detect trends, groups and outliers. It allows simpler global visualization by representing the variance in a small number of uncorrelated latent variables, which can then be understood to be information or random variation.

PCA decomposes the data matrix into principal components (latent variables or latent structures) that represent the underlying structure of the data. This allows one to represent the structured variance in the data by a smaller number of (latent) variables, while discarding the noise, thus making it appropriate for dimensional reduction. A matrix $\mathbf{X}$ (of e.g. metabolites) is decomposed by PCA using p components as follows: $\mathbf{X} = \mathbf{T}.\mathbf{P}^T + \mathbf{E}$, where $\mathbf{X}$ has dimensions $n \times m$, $\mathbf{T}$ is a $n \times p$ matrix of scores, $\mathbf{P}$ is a $m \times p$ matrix of loadings and $\mathbf{E}$ is a $n \times m$ matrix with residual variance, i.e. not included in the latent variable model. Depending on the objective of the analysis, the number of components in the model can be decided arbitrarily (e.g. a number "large enough"), according to a certain percentage of variance described with that number of components (e.g. 95 % of cumulative variance), or by using cross-validation strategies (which are later described).

An example of a PCA analysis is depicted in Fig. 7.1. Scores are coloured according to some meta-information after PCA calculations, in order to understand the reasons for the clusters. Samples in the same cluster are similar in the components represented, while variables in the same clusters are correlated with each other. To see, e.g. which variables are higher/lower in group B, draw a line passing in the centre of group B and through zero and then draw a line in the same direction
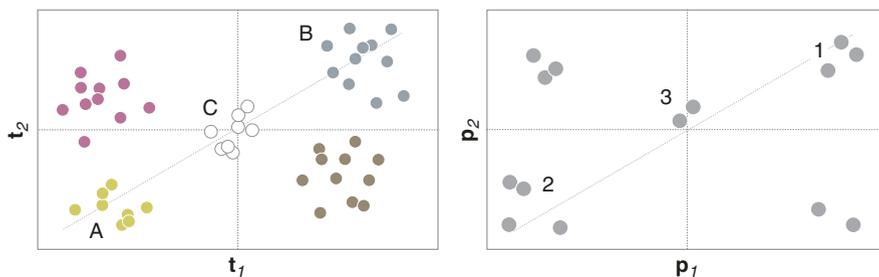
Fig. 7.1 PCA scatter plots of scores $\mathbf{t}_1$ vs $\mathbf{t}_2$ (*left*) and loadings $\mathbf{p}_1$ vs $\mathbf{p}_2$ (*right*) should be inspected simultaneously in order to understand the relations between trends and groups observed in the samples (score plot) and which variables – metabolites – are responsible for it (loadings plot)

in the loadings plot. Variables 1 are over that line in the same area as samples from group B; thus they are in higher values in B than in, e.g. group A (which is on the opposite side). Inversely, variables 2 have higher values in group A. Samples C, located close to the origin, have average behaviour between A and B. Variables 3 have no influence in this component, as their weights in the loadings of PC1 and PC2 are close to zero. Note that PC1 vs PC2 are being shown, but due to PCA's orthogonality of components, any PC can be plotted perpendicular to each other. In addition, sometimes plots of three components (xyz) are used, although they may become too complex to visualize due to the number of features involved.

Mathematically, the first principal component is the line that better approximates the data, in the least-squares sense. It represents thus the direction of the largest variance in the dataset, or in other words, the direction in which the variance of the coordinates of the samples is maximized. The dataset information explained by the first component can be subtracted from the initial data, and a second component can then be calculated from the residuals. Each principal component (PC) represents a fraction of the variance in the data – a pattern that can be in higher or lower magnitude in each sample – and is unrelated (orthogonal in a linear algebra sense, perpendicular in a geometrical sense) to the others (thus can be drawn perpendicularly to each and every other). The orthogonality property of PCA can be easily understood if one considers the calculation of each PC at a time. After PC_i is calculated from a data matrix $\mathbf{X}$, the information it represents is deleted from $\mathbf{X}$. Thus, for the calculation of the next component PC_{i+1}, that information is not available anymore.

Apart from helping at visualizing trends and groups in the data, an important application of PCA is to look for outliers in the samples. Outliers are samples that have scores very distant (thus different) from the others. They can be found by inspecting the scores or a model's cumulative measure of distance such as Hotelling T^2 [14], as well as by inspecting the residuals of the model (large residuals may indicate mild outliers). Due to their high leverage during model creation, special care must be taken in order to remove them or not, prior to defining a model and interpreting it. It may make sense to remove outliers, if one understands they are caused by gross errors during sample preparation or instrumental analysis. More

difficult decisions arise for less extreme samples, in which the large score distance to other samples cannot be justified by that, but is the result of correctly measured high or low values in some variables. Many different ways exist to look for multivariate outliers [15–18]. Robust algorithms, which can better at handling outliers, have been developed for PCA [19]. Note: an extensive literature list on PCA can be found on http://www.stats.org.uk/pca.

7.4.2 Multivariate Regression (OPLS)

Projection to latent structures (PLS) [20] is a supervised multivariate linear regression method similar in concept to PCA, which finds the relations between two matrices (data $\mathbf{X}$ and response $\mathbf{Y}$), by maximizing the covariance of their latent variables. It allows to understand which variables (e.g. metabolites) of $\mathbf{X}$ are more correlated to the response (e.g. calcium levels in blood) and to make predictions for new samples.

Orthogonal projection to latent structures (OPLS) [21] is a modification of the PLS method. OPLS has the same predictive power as PLS but provides better interpretation of the relevant variables than PLS. It does so by decomposing the data in so-called "predictive" information related to the response $\mathbf{Y}$ (as concentrations, classes), "orthogonal" structured information not related to the response (as instrumental, biological variations) and residual variation.

The decomposition of a matrix $\mathbf{X}$ by OPLS for the single-y case using p latent variables is as follows [22]: $\mathbf{X} = 1.\bar{x}^T + \mathbf{t}_p.\mathbf{p}_p^T + \mathbf{T}_o.\mathbf{P}_o^T + \mathbf{E}$, where the data matrix $\mathbf{X}$ has dimensions $n \times m$, 1 is a vector of dimension $n \times 1$ with ones in all positions, $\bar{x}$ is a vector $n \times 1$ with the column averages of $\mathbf{X}$, $\mathbf{t}_p$ is a vector of $n \times 1$ predictive scores, $\mathbf{p}_p$ is a vector of $n \times 1$ predictive loadings, $\mathbf{T}_o$ is a $n \times p - 1$ matrix of orthogonal scores, $\mathbf{P}_o$ is a $m \times p - 1$ matrix of orthogonal loadings and $\mathbf{E}$ is a $n \times m$ matrix with residual variance, not included in the latent variable model, as it contains only residual, nonstructured variation.

The model prediction of a y variable by OPLS is obtained by $\mathbf{y} = \bar{y} + \mathbf{t}_p.\mathbf{q}_p^T + \mathbf{r}$, in which $\mathbf{y}$ is a response vector of dimensions $n \times 1$, $\bar{y}$ is a vector of dimension $n \times 1$ with the average of $\mathbf{y}$ in all positions, $\mathbf{t}$ is the predictive scores vector from $\mathbf{X}$ and $\mathbf{q}$ is a vector $n \times 1$ of predictive loadings from $\mathbf{y}$, while $\mathbf{r}$ is a $n \times 1$ vector of y residuals.

Notice that for the single-y case, there can be only one predictive component, although many orthogonal ones may exist. Because of the predictive and orthogonal variance decomposition, one can look at the predictive score direction from negative to positive as an increase in the magnitude of $\mathbf{y}$, which is positively correlated with variables in the positive side of the predictive loadings (and inversely correlated with variables on the negative side). For the multiple-y case, there may be multiple predictive components, reflecting the overlap in information between the matrices $\mathbf{X}$ and $\mathbf{Y}$. Figure 7.2 illustrates single-y OPLS analysis.

OPLS is the multivariate linear method of choice to, e.g. find metabolic biomarkers correlated with a continuous variable, such as calcium score or blood pressure.
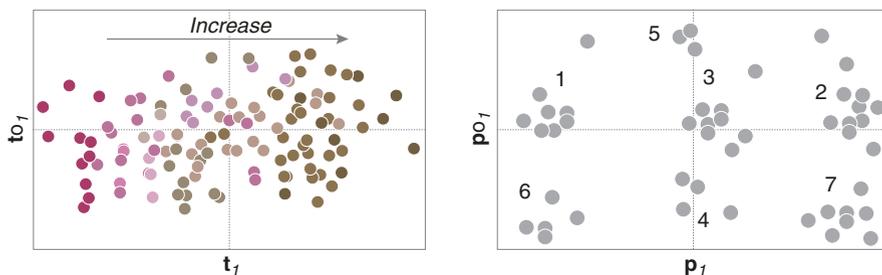
Fig. 7.2 Single-y OPLS scatter plots of scores (*left*) and loadings (*right*) for predictive component 1 vs orthogonal component 1, with scores coloured by **y** variable (e.g. blood pressure). OPLS models the **y** variable in the predictive component; thus samples with positive score t_1 (*right side* of the scores plot) are more concentrated on variables on the positive side of $\mathbf{p}_1$ (clusters 2 and 7) and less concentrated in variables with negative $\mathbf{p}_1$ (clusters 1 and 6). The orthogonal variation that is seen in the orthogonal scores to_1 (up–down) can also be inspected by colouring the scores according to different meta-information (e.g. gender, age) or the loadings (e.g. compound class). Variables related to a trend in the orthogonal scores are found along the orthogonal component loadings po_1

7.4.3 Multivariate Classification/Class Discrimination (OPLS-DA)

OPLS discriminant analysis (OPLS-DA) [23] has been largely used in the metabolomics context, and it is now the multivariate linear model of choice for classification/discrimination [24]. The term classification is used when the objective is to classify new objects into one of two or more possible classes (e.g. control, disease A, disease B). The term discrimination is used for the two-class case, in which the objective is to separate two classes and investigate the causes for class separation (e.g. biomarker discovery or which metabolites are in higher/lower concentration in a disease class in relation to a control class). Figure 7.3 shows an OPLS-DA example.

Notice that in OPLS the vector *y* is a continuous variable; in two-class discrimination, OPLS-DA **y** is categorical and, thus, defined as a dummy vector of 0/1 for the two-class case (for the multiple-y case, it is a dummy matrix with a 0/1 vector per class), describing class belonging. Although multi-class OPLS-DA can be calculated, most of the applications in metabolomics use a two-class model, as the interpretation is much more straightforward. Strategies for multiple class comparison using OPLS-DA are presented later in the chapter.

7.4.4 Note on Orthogonality

PLS was the method of choice for multivariate regression for many years, but OPLS has lately seen an increase in metabolomics data analysis, especially for discrimination and biomarker discovery. The reason is that although the methods explain the same variance in both **X** and **Y** matrices and have the same predictive capability,
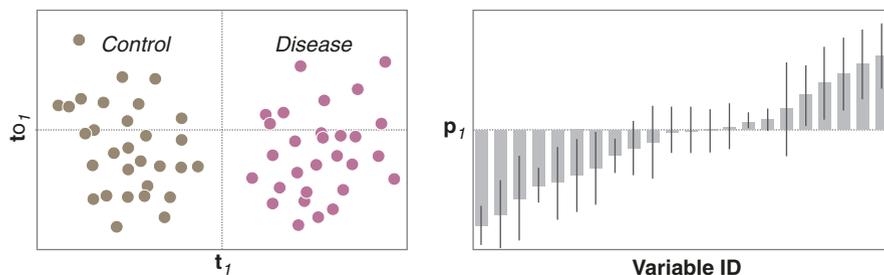
Fig. 7.3 OPLS-DA predictive vs orthogonal scores (*left*) and predictive loadings (*right*) for a two-class separation (e.g. control vs disease). The disease group, with positive predictive scores, has higher values than the control group in the variables with positive $\mathbf{p}_1$ (on the *right* of the loadings plot); it has lower values than the control group in the variables with negative $\mathbf{p}_1$ (*left side* of the loadings plot). The loadings weights were ordered according to magnitude, for easy visualization of its importance, and also the existence of confidence intervals which indicate their statistical significance

PLS computes latent variables that contain mixed sources of variation, while OPLS decomposes the structured variation into predictive and orthogonal. In the simplest case, OPLS with only one y variable – or OPLS-DA with two classes – the information related to **y** is contained in the first predictive component, while the orthogonal components contain information related to other sources of structured variance, while discarding residual variance or noise.

It is important to realize though that orthogonal components contain information that is not noise [25] and should be investigated in order to bring more understanding of an experiment. With that in mind, the datasets should be accompanied of the most complete amount of meta-information regarding unintended sources of variation such as sample preparation, experimental conditions and characterization of samples and variables as possible. In some cases, patterns and groups of samples (or variables) can be seen in the orthogonal scores (colouring them according to the meta-information may help), which can be related to that variation, e.g. sample batch, gender, age, sample dilution or other stratifications of the data. Then the orthogonal loadings should be investigated to see which variables have influence in the orthogonal score trends and groups. As all components in the model have their variation quantified, that may allow additional understanding of the relative variation in the phenomenon in study in comparison to others and, e.g. allow better tuning of experimental conditions in future experiments.

7.4.5   Cluster Analysis

Many cluster analysis methods exist, because as some authors consider, "clustering is in the eye of the beholder" [26]. Nonetheless, due to its simplicity and usefulness, hierarchical cluster analysis (HCA) has been widely used and will thus be presented. This is a non-supervised clustering method, used to put in

evidence natural clustering of samples and/or variables, in the dataset. In case both samples and variables are clustered, one can see which clusters of variables are defining the clustering of the samples. Although the method is generally used for multivariable analysis, its nature is not multivariate, as no latent variables are defined.

In one of its forms, the method starts by considering that each single object is a cluster. On the first iteration, it finds the minimal distance between two of these (single object) clusters and clusters them. In the second iteration, it finds again the minimal distance between the updated clusters and clusters them. It proceeds the same way until all objects are part of the same cluster. Thus, since the beginning (after appropriate normalization/transformation of the objects in study), two parameters must be defined: the distance metrics to use and the linkage type. Distance metrics is related to how one measures "closeness" of two objects, and commonly used metrics are the Euclidean and Mahalanobis distances, or the Pearson and Spearman correlations. Linkage type is related to which objects in the current groups are used to calculate those distances, and common types are "single" (minimum distance between one object in each group), "average" (distance between averages of the objects in the groups) and "Ward" (minimum model error increase for merging two clusters). A dendrogram of the clustering process can be plotted, in which the length of the bars represent the distance between the clusters, together with a heat map of the actual data values (see Fig. 7.4).

Considering the samples, and depending on the study context and objective, the method can be applied to the actual data (metabolic features values), to its PCA scores, PCA distances to model, or any other meaningful transformation of the data.

The major advantages of the method are that it is easy to understand and its application is straightforward. The major disadvantage is the difficulty in interpreting the data when there are too many samples or too many variables (most common in metabolomics).

7.4.6 Independent Component Analysis (ICA)

ICA is a blind source separation method used in signal processing, and it separates multivariate signals into additive subcomponents. Its interpretation is similar to PCA, but instead of orthogonal components, it calculates non-Gaussian, mutually independent ones. Contrary to PCA, ICA does not order the components according to variance, and the number of components influences the structure of the components themselves; thus an adequate determination of the right number of components is of utmost importance. ICA algorithms have been used for analysis of metabolomics data, to detect metabolic patterns [27], phenotypes [28], and in class discrimination [29].
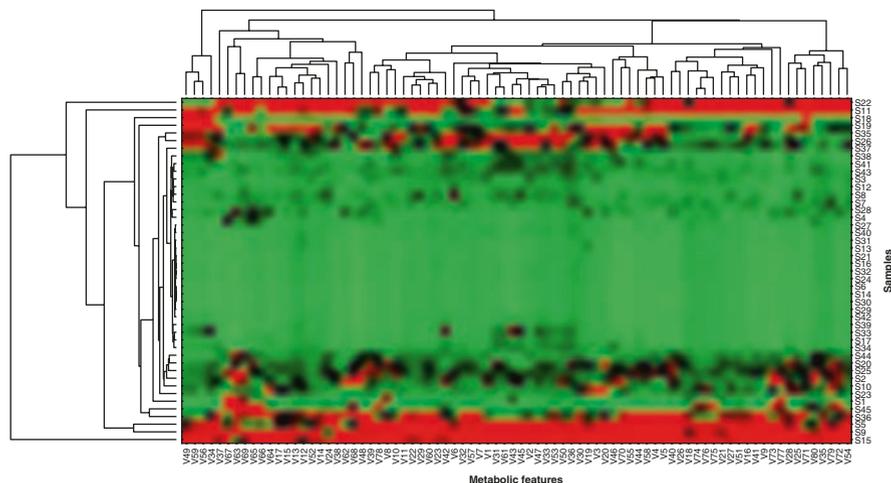
Fig. 7.4 Example of heat map and dendrograms clustering samples and metabolic features in a dataset after HCA using Euclidean distance and average linkage. By looking at the heat map, one can have visual clues about which metabolic features are aggregating the samples in clusters. In the case of the samples, the horizontal lines in their dendrogram (on the *left*) are proportional to cluster distance; for metabolic features (on the *top*), it is the vertical bars

7.5 Complementary Strategies for Data Analysis

Most applications of multivariate analysis in metabolomics use PCA for data exploration and then OPLS or OPLS-DA for regression or class discrimination/biomarker discovery, respectively. The methods are applied directly to the dataset itself, after appropriate preprocessing. Nonetheless, these same methods can be used to analyse or visualize the data in creative and helpful ways, depending on the experimental design and the objectives of the study. Below we present some of those examples.

7.5.1 OPLS-DA Strategies for Comparison of Discriminant Metabolites

As referred in Sect. 7.4.3, OPLS-DA can be used for multiple class discrimination and classification, but the direction of class separation may not align over the latent variables axes, thus turning interpretation less straightforward. In case the objective of the study is to understand the difference of multiple classes (treatments, conditions or disease states) to the same control class, it may be preferable to model each of the classes against control separately (e.g. control vs disease A and control vs

disease B) and then compare the results. Two suggestions on how to do that are presented below:

(a) *Comparison of models of two classes vs same control*:

 In this case, one can use the shared and unique structures (SUS) plot [30]. For the two-class case OPLS-DA models always represent the class discrimination along the predictive ("abscissa" axis) component, which allows straightforward loadings interpretation. For more than two classes/properties, this representation may not be possible to do using only one predictive component; thus the class separation may not be along a single axis. For this reason, the most convenient way of comparing two models is to create individual models for each of the possibilities (e.g. control vs disease A and control vs disease B) and then compare their loadings against each other.

 From the OPLS (DA) models, different loadings vectors can be obtained. The correlation between the predictive score vector and each of the **X** variables is defined as the p_{correl} loadings. Being composed of actual correlations, its values vary between -1 and 1, thus appropriately standardized for inter-model comparison. The SUS plot is simply a scatter plot of the p_{correl} of two individual models. It should be visualized simultaneously with a p-loadings plot with confidence intervals (or any other measure of variable significance), so one can also see which variables are significant. Three different situations may arise for each of the significant variables:

 (i) If aligned along a positive "/" diagonal, they show the same behaviour in both models (e.g. increased concentration of metabolite $\mathbf{X}_i$ in disease A vs control as well as in disease B vs control).

 (ii) If aligned along a negative "\" diagonal, they show opposite behaviour in each model (e.g. increased metabolite concentration of metabolite $\mathbf{X}_i$ in disease A vs control, but decreased in disease B vs control).

 (iii) Aligned along the horizontal/vertical axis shows an effect in one of the models, but not in the other.

(b) *Comparison of models of more than two classes vs same control*:

With more than two classes, the SUS plot approach gets complicated. A better visualization approach can be made using a network approach, plotting together the significant variables from each of the models. When considering, e.g. a small number of different diseases in relation to the same control class, the following definitions may be used (see Fig. 7.5):

 (i) The different diseases are represented as central nodes, in a different shape/size and colour than the metabolites.

 (ii) For each disease, its significant metabolites are represented as peripheral nodes, connected through directed edges to the disease.

 (iii) The direction (or colour) of the edges indicates if the metabolite is more concentrated in the disease (pointing to the disease) or in the control (pointing to the metabolite).

 (iv) The edge width can be used to denote the degree of variable significance (*p*-value, correlation, fold change).
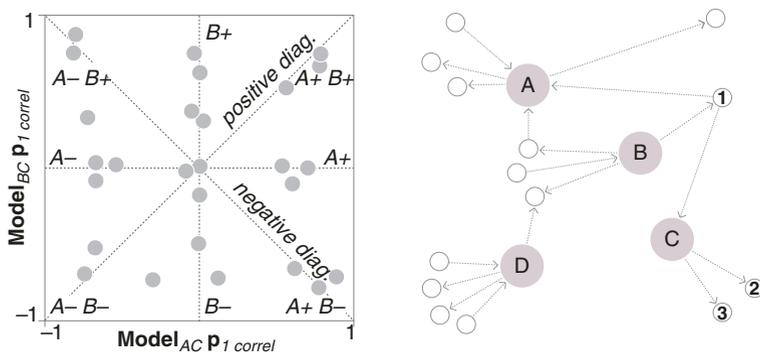
Fig. 7.5 Plots for comparing metabolites coming from models from, e.g. different diseases vs the same control group

(v) Colour codes can be used for each metabolite, reflecting the number of diseases they are common to or any other relevant information (e.g. chemical class).

As shown in Fig. 7.5 (left), SUS plot is a scatter plot of p_{correl} of two models of, e.g. diseases A and B vs control (C). Each p_{correl} is a correlation value, thus varying between -1 and 1 in each model. Variables (features or metabolites) over the positive diagonal are upregulated (A + B+) or downregulated in both diseases (A−B−) in relation to control. Variables over the negative diagonal have inverse behaviour in each of the diseases (A+B−) or (A−B+) in relation to control. Variables over the *x*-/*y*-axis are only up-/downregulated in one of the diseases (e.g. A+ or B+), but not significantly different from control in the other.

Figure 7.5 (right) shows a network representation of relevant metabolites (small circles) obtained for OPLS-DA for comparisons of, e.g. different disease classes (A–D, large circles) to the same control. The colour and size of the nodes differentiate the disease classes from the metabolites. The edge arrows pointing from a metabolite to one of the disease nodes (A–D) indicate that the metabolite level is higher in that disease than in the control group and vice versa. As examples, metabolite 1 is statistically significant in the models of diseases A, B and C. Looking at the arrow directions, it is upregulated in A and C and downregulated in B, in relation to control. Metabolites 2 and 3 are only downregulated in the model of disease C, in relation to control.

In case there are not many variables, but many classes, one can change the roles of classes with metabolites in the plots.

7.5.2   Comparisons of Trajectories and Profiles

The modelling of multivariate metabolic trajectories has been used mostly to follow time series processes using different strategies or in different subjects. It can be applied when the design of experiments uses groups of samples that have a

sequential dependence (e.g. follow multiple individuals during time). The main idea is to use multivariate modelling to follow the evolution of each of the groups of dependent samples and then compare their score trajectories. Although not all these examples are in the clinical context, they may be adapted to it, if the right experimental design is used.

Using a method called "scaled-to-maximum, aligned and reduced trajectories" (SMART) [31], two groups of animals were studied in relation to the effects of drugs against control using ^{1}H NMR. The average of the initial time point of each individual is subtracted from all the samples of all individuals, to achieve a similar start point. The data is then scaled to a common magnitude by using the largest magnitude value for each treatment group, prior to using PCA (scores) to visualize the average trajectories for each treatment. Similar trajectories correspond to similar behaviour of the groups and vice versa. The same strategy can be used in other experimental settings.

In another type of application, urine samples of patients following a kidney transplant were collected in time and studied using ^{1}H NMR, in order to identify profiles for toxicity/rejection or normal recovery [32]. Grafts take different time in different individuals until actually start working properly/incorrectly. In this type of analysis, each patient was used as their own reference; thus the specific changes occurring during time for each individual could be examined. As the samples of each individual were not separated into "before" and "after" graft classes, the first objective was to identify time samples related to those two moments in the procedure. For that the researchers first used PCA in each patient and selected grouped samples in the extreme sides of the scores into the two classes ("before" and "after" graft functioning). Then they performed OPLS-DA to discriminate between those classes for each patient and collected the predictive discriminant loadings into a matrix. Finally, these discriminant loadings – which represent the profile that discriminates "before" and "after" graft for each patient – were used to represent each of the patients. PCA was performed in this matrix of loadings profiles to find a common effect profile and the most important metabolites for good patient recovery or kidney rejection.

Finally, we discuss a high-throughput multiple comparison study of transgenic tree lines against a common wild type done over several years [33]. The study objective was to understand which mutant lines (around five biological replicates per line) were more affected by the genetic engineering and which ones were similar in metabolic features. A batch effect due to biological and experimental variation did not allow the global comparison of phenotypes using the original data. The authors used an integrated chemometrics pipeline for the analysis of the data, which had the additional advantage of reducing those batch effects. This pipeline started with PCA for quality control of the wild-type samples of each of the batches separately, to detect outliers. Then, they investigated the existence of outliers in the mutants, by projecting the mutant plant samples of each batch in their respective batch PCA and looking at how they differed from the ones in their group. After the data was cleaned of outliers, they used OPLS-DA for class discrimination between the control samples and each of the mutant lines. This had

the objective of finding the pattern (OPLS-DA predictive loadings) for differentiation from control for each of the mutants and also of reducing the batch differences for global comparison. The loadings representing the differentiation pattern were used in representation of each mutant line and visualized using PCA and HCA, where clusters of mutants could be visualized, together with correlated groups of metabolic features.

7.5.3   Modelling Designed Data (ASCA)

The well-established analysis of variance (ANOVA) is an ensemble of univariate methods used to analyse differences among group means. It partitions the variation of designed factor treatments and interactions, to evaluate if any of the levels in a factor or interaction is statistically different from the others. ANOVA-simultaneous component analysis (ASCA) [34, 35], in which ANOVA and PCA work together, is designed with similar intent, generalized for the multivariate case. In experiments where there is the possibility to design a balanced experiment, it can be used to evaluate which factors and interactions are statistically significant and to find which metabolites are relevant in each of those factors. Different types of data scaling may be used to amplify some aspects of the data, thus giving rise to different solutions [36].

An example of this type of design would be an experiment in which different drug formulations are given to individuals and their metabolic profile is evaluated at several time points, in order to understand if a formulation level at a specific time is statistically different from the others.

ASCA tests for the statistical significance of levels' difference in multivariate factors using a random permutation approach [37] for the factor(s) of interest. The rationale is that if no level is different from the others, the averaging process will make the factors approach zero, which should happen in the permuted models, but not in a statistically significant factor. For each factor testing, the sample group is randomly changed a large number of times and each time the factors are recalculated. A *p*-value for the significance of each factor can be calculated, based on the frequency of (number of times) factors that have a sum of squares (SSQ) larger than the original, non-permuted, factor.

7.5.4   Evaluate Effects on Matched Samples

In case the objective is to study the effect of a treatment, and there are matched samples of the type "before" and "after", the strategy OPLS-effect projections (OPLS-EP) [38, 39] can be used.

The method can be seen as a generalization of t-test for paired samples, thus similar to investigating if an average if different from zero (in opposition to t-test for

unpaired samples, in which the difference between averages of two classes is evaluated). It simply uses subtraction of the "before" sample from the respective "after" sample and considers that this difference will reflect the effect of the treatment. If MS instruments are used, it has also additional advantages on the reduction of batch and drift effects, attained through running the paired samples close to each other in the run order, while randomizing its relative position. The assumption is that if paired samples are ran close to each other, there is no significant drift between them.

In this strategy, the resulting "after-minus-before" subtracted data is modelled using OPLS, using (in general) metabolite data divided by its standard deviation, against a y vector with 1 in all its positions. While OPLS-DA on the same data would model class discrimination (comparable to unpaired t-test), OPLS-EP models effect difference (comparable to paired t-test). By plotting the predicted effect ($\mathbf{Y}_{hat}$, target value of 1) for each sample, one can understand which samples had larger ($\mathbf{Y}_{hat} > 1$) or smaller ($\mathbf{Y}_{hat} < 1$) effect, while looking at the predictive loadings indicates which variables were more important in that effect. The advantage of the method is that it looks for an effect, without being in reality a supervised method, as the samples have no need for class definition.

7.6   OPLS-Type Model Validation

PCA is mostly used as an exploratory method, and as the inclusion of more components has no influence in the previous ones, most times there is no need to decide the appropriate number of latent variables to use in the model. The same is not true for OPLS-type models. Furthermore, once an OPLS-type model is established, before it can be used for prediction, or to decide on the significance of the discriminant variables, rigorous validation must be performed [40]. It is worth to mention that many times the score plot will look like indicating a good class separation, but later validation does not confirm that.

The methods described in the following sections are used in the context of multivariate analysis, mostly of the OPLS-type, for the selection of an appropriate number of latent variables and for model validation.

7.6.1   Internal Cross Validation (CV)

During model building, CV is generally used in order to decide on the appropriate number of latent variables to include in the model. For each number of latent variables desired, X is divided into subsets of samples and then one model is built at a time, containing all samples except the ones in the corresponding subset. The subset samples are then predicted in the corresponding model, and the difference between the expected and the predicted value is saved. By doing the same for all subsets, one can obtain the predicted residual error sum of squares (PRESS) and additional

statistics that allow model quality comparison. The number of latent variables that gives the least prediction error is selected for the final model. A root mean squared error of cross validation (RMSECV) can be calculated and expressed in the same units as the **Y** variable. One should then evaluate three important statistics:

  (i) R2X: fraction of variance of **X** explained by each latent variable. Always increases with increasing number of components, even when overfitting by modelling noise in **X**. Answers questions (in OPLS-DA) of the type "how much of the variation in **X** is related to the difference between the classes?"

 (ii) R2Y: fraction of variance of **y/Y** explained by each latent variable. It always increases with increasing number of components, as it starts modelling noise in **X** in order to explain **y/Y**. This statistics answers questions (in OPLS-DA) of the type "how good is the separation between the two classes?"

(iii) Q2: The most important statistic to decide on the quality of the model, it varies between[-INF, 1]. It is the fraction of variance of **y/Y** predicted by each latent variable. Because it is based on prediction, the inclusion of noise should not increase Q2. However, although it is expected to stop increasing, or to start decreasing, after all structured information was modelled, that is not always observed. This statistic answers questions (in OPLS-DA) of the type "how well can we predict the two classes?"

CV yields different statistic results, depending on how CV groups are defined. A commonly used strategy of leaving one sample out (LOO) at each CV round is not advised [41], as the perturbation in the data may be too weak to have a significant effect. CV should also not be used with replicate samples, as the inclusion of one of the replicates allows better predictions of its sisters, and this will inflate the statistics, showing better results than it should. Designed data may also have its problems, as the removal of some influential samples from a model may destroy the structure of the data and not allow them to be well predicted (e.g. as for samples in the extremes of the design factors).

In general, CV rounds should be defined in a balanced way, with each round containing samples from all quadrants of the experimental design. When dealing with datasets containing multiple individuals and samples of, e.g. different disease phases of the same individual, a practical advice would be to leave all samples from one individual out in each of the CV rounds, due to risk of autocorrelation. This strategy removes the contribution of each individual to build a model in each round of CV, while evaluating how the samples of that individual match the other individuals for each of its disease phases.

7.6.2 Cross Validation Scores (CV-Scores)

During CV, the different samples' subsets are predicted, and scores can be obtained. Because they are predicted from samples that were not used in the same round to build an OPLS-DA model, classes in the CV-scores always look less separated than

when visualizing the scores (obtained from the model including all samples). In the case of an OPLS regression model, the CV-scores will look less correlated to **y**. Nevertheless, CV-scores should always be visualized as they give a more realistic figure of future model prediction quality. Evaluation of CV-scores is in most literature just visual, but numerical measures can also be adopted [42].

7.6.3 Cross Validation-ANOVA (CV-ANOVA)

Analysis of variance can be used to compare the size of the residuals of two models applied to the same data [43]. Here it is adapted as a diagnostic method to evaluate the reliability of an OPLS-type model. It compares the *y* predicted residuals of the model with the variation around the global average, using an F-test for comparison of variances. In case the model predicted residuals are significantly smaller than the variation around the average, the null hypothesis of equal residuals of the two models is rejected, with a certain confidence level (e.g. 0.05). It is a rapid method as it uses values calculated during cross validation, and easy to evaluate, as it provides a significance *p*-value. Nevertheless, according to the author's experience, due to biases and unidentified structured information in the data, if CV-ANOVA indicates a bad model, that is most certainly true, while if it indicates a good model, that may not necessarily be the case.

7.6.4 Permutation Test

Permutation test is a method to evaluate the statistical significance of the estimated predicted power, Q2 [22, 44, 45]. In this method, the R2Y (self-prediction) and Q2 (cross-validated prediction) of a defined model are compared to the ones from a large number of models in which the *y* vector has been randomized, and no good prediction capability is expected. The evaluation of the model validity is done by looking at the number of "random" models that present better statistics than the one being evaluated, or by looking at the intercept of the linear regression of each of those two statistics [22].

Notice that the R2Y and Q2 values are plotted in function of the correlation of each of the randomized *y* vectors and the actual true *y* vector. In case high Q2 values are found for some "random *y*" models, one should examine if that correlation is high, in which case means that the randomization process created a "randomized" y that is very similar to the actual *y* vector.

7.6.5 External Validation

While the above validation methods can provide an idea of the quality of the models, prediction of an external data can elevate our confidence in the model quality to a higher level. Depending on the objective of the experiment, more confidence can be

deposited in a model that can predict samples that were acquired or processed in different times, machines, by different operators, etc. and were not used for model construction. Sensitivity and specificity can be evaluated in case of OPLS discriminant analysis/classification, while continuous measures of prediction error can be calculated in case of OPLS regression (root mean squared error of prediction, RMSEP).

7.6.6   Comparison of Model Loadings

It is itself an external validation; one can evaluate the validity of models by confirming the statistical significance of the discriminant variables. If an experiment is repeated, the class discrimination should be influenced by the same variables. An SUS plot of the two models may be used for that, in which case the same discriminant variables should be aligned along a positive diagonal.

7.6.7   Receiver Operating Characteristic (ROC) Curves

A well-established technique in clinical essays [46], it is very useful to compare over different classification models (using the area under the curve, AUC) or to evaluate thresholds for better sensitivity or specificity (using graphical representation). In the case of two-class OPLS-DA, after a model is calculated and the predicted classes obtained, a ROC curve can be calculated by incrementally moving the discrimination threshold between 0 and 1 and plotting the results for each incremental value.

7.7   Significance of Variables in OPLS-Type Regression/ Discrimination

Once an OPLS-type model has been determined and adequately validated, it is in general of interest to find out which variables (features or metabolites) are more influential in the model and to decide on its statistical significance. In order to visualize the influence of a variable in the model (regression or class discrimination), one can just sort the relevant vectors (p-loadings or VIP) by magnitude. To determine the validity of each of the variables, several strategies are described in the following sections.

7.7.1   p-loadings with Confidence Intervals

The *p*-loadings of the predictive latent variable represent the influence of the variable in the OPLS regression/discrimination. Furthermore, during internal cross validation (CV), multiple OPLS-type models are produced, while leaving some samples

out in these different CV rounds. The p-loadings obtained in each of these CV rounds can be averaged and a standard error (error bars, confidence intervals) calculated with a predefined level of significance (e.g. 0.05). Some authors sort the p-loadings by magnitude just for model-influence visualization. Then, for statistical significance consider that, for a certain variable, if the error bars do not cross zero, the variable is statistically significant (in other words, the absolute value of variable minus standard error is larger than zero).

7.7.2 p_{correl} and Correlation Threshold

The *p*-loadings can be rescaled as the correlation coefficient between the variables in X and the scores (t) of the OPLS-type model (here defined as $\mathbf{p}_{correl}$). These $\mathbf{p}_{correl}$ are correlation values and, thus, have values between the limits $[-1–1]$. A correlation threshold for a desired level of significance (e.g. 0.05), dependent on the number of samples, can be obtained from a table of critical values for Pearson correlation. The statistically significant variables are the ones which absolute $\mathbf{p}_{correl}$ larger than the adequate critical correlation threshold.

7.7.3 *Variable Importance on the Projection (VIP)*

This is an established and compact parameter used to summarize the importance of each of the **X** variables in a PLS with >2 components. Important variables have VIP larger than 1, while a variable is more irrelevant the lower than 1 is its VIP. There are different VIP measures for OPLS [47, 48], and researchers adopt in general the one defined in their software package.

7.7.4 *Note on Significance of Variables*

Some authors choose to select statistically valid features or metabolites only if they obey multiple criteria, including some of the ones described above plus others coming from univariate testing. These can be a minimum correlation needed, *p*-values after some multiple testing correction or fold change.

7.8 Univariate Analysis

Univariate analysis has been used in conjunction with multivariate analysis to study variation and to test statistical significance of parameters and variables in metabolomics studies. Notice that while multivariate analysis can handle certain amounts of

batch and drift variation, univariate analysis should only be used if correction for these effects is satisfactory. Nevertheless, new attention has been given to the analysis of metabolomics data using univariate analysis [49], especially in the field of epidemiology [50]. Until recently, rare – if any – metabolomics studies were composed of thousands of samples due to its cost but also to issues related to process automatization, data handling, processing and reproducibility, among others. However, some large-scale metabolite profiling studies have now been done [51], mostly in epidemiological research, as metabolomics is expected to measure environmental and exogenous exposures more precisely than traditional questionnaires. In these studies, linear or generalized linear models are used in univariate analysis fashion, while correcting for confounders. These confounders are experimental factors that may be correlated with the outcome and in that case are not removed using OPLS-type multivariate methods.

A word should be said in relation to the use of parametric (e.g. t-tests, Pearson correlation) or nonparametric (e.g. Mann–Whitney U test, Spearman correlation) strategies. For normal populations, parametric tests are more powerful than nonparametric ones, but that is not the case for non-normal populations, unequal variances and unequal small sample sizes, where using a nonparametric test would be advantageous. While testing for normality distribution in four datasets, some authors found in average 65 % of metabolic features met normality and equality of variance assumptions. Still, as it was dependent on the dataset, they suggest to use both strategies, and if there is a large difference in the results, one should look for outliers in the dataset [49].

7.9   Multiple Testing Corrections

To decide on the statistical significance of a feature or a metabolite, e.g. if it is or not correlated with an outcome or if it has discriminant capacity between two classes, univariate methods rely on p-values. Because in metabolomics untargeted studies one is looking after thousands of variables, multiple testing corrections must be applied. The Bonferroni correction was commonly used in the past, when the number of variables was small, but as it corrects for the family-wise error rate (FWER) – the probability of at least one false positive – it ends up being too conservative. Benjamini–Hochberg and other corrections that control the false discovery rate (FDR) [52, 53] are less conservative and widely applied. Nonetheless, due to the high degree of correlation observed in metabolomics data (e.g. adjacent intensities in NMR peaks), existence of multiple features for the same metabolite (e.g. LC-MS dimers, adducts; NMR signal multiplicity), they are also considered not appropriate. Thus, permutation strategies such as the metabolome-wide significance level (MWSL) [54] have been developed to control for the FWER, which determine more robust p-value thresholds for discovery than the above methods.

7.10 Practical Aspects of Chemometrics in the Context of Preprocessing, Pretreatment and Experimental Design

Many decisions must be taken when designing a metabolomics experiment, regarding sample type and number, cost, time, human resources, instruments and data analysis methods.

The decisions taken will provide answers to different questions; thus a very well-defined idea about the methods that will be used for data analysis and interpretation is fundamental, in order to be able to pose objective questions and obtain correspondingly appropriate answers. This has a retrospective impact on the design of experiments itself. Although not the main focus of this chapter, it seems appropriate to include a brief description of different options that can be made, which condition the chemometrics data analysis and may be used for batch correction in MS data.

7.10.1 ^{1}H NMR Data: Types of Data Matrices

Different strategies can be used for the preprocessing of ^{1}H NMR data, depending on the type of sample (e.g. urine or serum), because of chemical shifts due to physicochemical sample differences. Blood samples are expected not to change much due to homeostatic regulation, while urine is known to change more in concentration as well as in properties, as it is more affected by microbiota, drugs, diet and disease [50, 55].

Direct analysis after alignment: if the alignment is good enough, the data can be immediately analysed. When using multivariate methods for the data analysis, data tend to be scaled by mean centring or Pareto normalization, with optional log transformation, so the spectral structure is kept. In this case, unit variance normalization is not used as it gives the same importance to variation in the signal and in the noise region, and the loadings do not show spectral structure similar to the data. If large "saw-tooth" (inverted peaks) effects are found in the loadings, it is a sign that the alignment was not performed perfectly.

Analysis after alignment and binning: if there are some issues with the alignment, but the whole spectrum is to be analysed, binning adjacent values can be used. If multivariate analysis is used, data should be normalized by centring or Pareto normalization, to keep the original spectral structure, with optional log transformation.

Analysis after alignment and peak picking: a similar strategy to binning is to integrate peaks, but in a more targeted way, thus rejecting noise regions. If multivariate analysis is to be used, unit variance scaling can be used, as noise regions are not supposed to exist. When using this strategy, the spectral structure is lost.

7.10.2   MS Data: Reducing Batch and Drift Effects

GC/LC-MS instruments are prone to batch and time drift effects, due to changes in instrument sensitivity and intensity, among other effects. Targeted methods correct batch and drift effects with the inclusion of labelled internal standards, with which a ratio between the target compound and the internal standard can be calculated. For untargeted GC-MS and LC-MS, several strategies have been used to correct for these effects: inclusion of periodic quality control samples (pooled from all or from a group of samples in the experimental set) that are expected to yield the same results along time, addition of internal standards to the samples (heavily labelled and/or not occurring in the samples) representing different compound classes, and experimental design using paired samples (when having samples with and without effect). These methods present both advantages and disadvantages, which are tentatively explained below.

Periodic Internal Quality Control (iQC) Samples [56] An adequate volume of pooled sample is built by pooling an amount of each of the biological samples. Subsamples of this main sample, assumedly with equal composition and concentration, are interspersed (e.g. every fifth sample) with the biological samples and ran in the instrument. For each individual variable, these iQC samples are then modelled using locally weighted regression (e.g. robust loess). Once a model is established, one can calculate the ratio between each biological sample and the LOESS curve (while the iQC samples should all equal 1) for each variable. This method is used both for batch and drift correction. While potentially the most adequate correction method, its major disadvantage is the increase in the total number of samples, with impact on time and cost of the experiment.

Internal Standards (IS) A certain number of quality control labelled compounds, not expected to have endogenous expression in the samples, representing different chemical/biological classes (e.g. amino acids, fatty acids) are added in the same concentration to each biological sample. These compounds are assumed to be in the same concentration in each sample. The advantage of the method is that there is no need for additional samples. The disadvantage is that the internal standards may not be the adequate ones to normalize the data. To correct for batch/drift using internal standards, the following strategies have been applied:

(i) *Correction by single or multiple internal standards* (*IS*): if a single IS was used, the ratio or log2 ratio between each variable and the IS can be calculated, and the variable is considered corrected. If multiple IS were used, the decision of which IS shall be used to normalize a certain variable can be done according to maximum IS correlation [57] or by minimal retention time difference [58].

(ii) *Correction using multiple IS and PCA*: the features representing the IS are normalized dividing each one by their corresponding standard deviation. PCA is calculated on this dataset, and the scores of the first component are obtained, representing the major batch/drift effect on the dataset. Then the procedure is the

same as in the previous case. For each variable, the ratio between each sample and the corresponding score value is calculated, normalizing the data. The major disadvantage of this method is that it can only correct one major batch/drift effect and may even wrongly correct features that were less affected by those effects.

(iii) *Correction using OPLS*: a procedure similar in concept to orthogonal signal correction has been applied to microchannel microarray data using OPLS [59], cleaning the data from orthogonal variation that is not common within sets of biological replicates. The method uses the data in matrix *X* and a dummy matrix identifying the replicates in a matrix *Y*. After it identifies the orthogonal information, it builds the corrected data matrix using matrix multiplication of predictive scores and loadings, plus the residual variance. The strategy seems applicable to metabolomics, and a variation of it, using information from the IS samples, has been used in the context of batch normalization and drift correction [60].

Experimental Design Using Paired Samples In case there are reference samples, like "before" (baseline) and "after" treatment for the same individual, or when using multiple time points, or matched case–control, and the objective is to study an effect (e.g. of a drug, or disease). In these cases the matched samples can be ran close to each other, and the drift between them is assumed as negligible. The baseline sample can then be subtracted from the effect(s) sample(s), with the result being the difference between the two (the effect itself). Local randomization of the matched samples is used, to minimize for any sequential bias. While not needing additional samples, the major disadvantage of this method is that it is only applicable in situations where a reference sample exists. Additionally, one will be studying not the current metabolite relative levels but the effect's metabolite relative levels in relation to baseline. The OPLS-EP method previously mentioned [38] is an example of this strategy in practice, for paired samples "before" and "after" effect. Alternatively to subtracting a "before" sample (baseline), the average per group of paired samples could be subtracted from each of the respective paired samples, in which case the baseline sample would be kept for analysis.

7.11 Internet Resources and Software

Following the developments in other omics fields, efforts have been put into creating internet platforms for automated and semi-automated metabolomics data analysis. Some very good resources are now available, which may require a minimum of knowledge of the methods on the side of the researcher to output meaningful data analysis results. Among the most well known and commonly used are MetaboAnalyst [61–65], metaP-server [66], Workflow4metabolomics [67] and Galaxy-M [68], and work is in progress in the large-scale computing for medical metabolomics website PhenoMeNal [69]. Finally, the website OMICtools [70] provides a comprehensive description of software that can be used for metabolomics data analysis, as well as a number of sites that can be used for different purposes in the omics fields.

7.12   Concluding Remarks

Chemometrics has been heavily used in all steps of metabolomics studies and has here been discussed in the context of data analysis in clinical metabolomics contexts. Its relevance in this field is due to the complexity and number of variables in metabolomics datasets and the simplicity of interpretation of its results. While other strategies in bioinformatics start appearing that gather information from databases, thus needing previous identification of metabolites, chemometrics methods are purely numerical, thus finding its own place in the data analysis pipeline. The possibility of automation has brought to light some websites that provide statistical calculations, chemometrics methods included, without major input from the analyst.

References

1. Nicholson JK, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181–9.
2. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. Metabolite profiling for plant functional genomics. Nat Biotechnol. 2000;18(11):1157–61.
3. Massart DL, Deming SN, Michotte Y, Kaufman L, Vandeginste BGM. Chemometrics: a textbook. New York: Elsevier Sciences Ltd.; 1988.
4. Brereton RG. A short history of chemometrics: a personal view. J Chemom. 2014;28(10):749–60.
5. Piantadosi S. Clinical trials: a methodologic perspective, second edition. 2nd ed. New Jersey: John Wiley & Sons; 2005. p. 720.
6. Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. J Proteome Res. 2007;6(2):469–79.
7. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics – a review in human disease diagnosis. Anal Chim Acta. 2010;659(1–2):23–33.
8. Bartel J, Krumsiek J, Theis FJ. Statistical methods for the analysis of high-throughput metabolomics data. Comput Struct Biotechnol J. 2013;4:e201301009.
9. Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag. 1901;2(series 6, 11):559–72.
10. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab. 1987;2(1–3):37–52.
11. Jackson JE. A user's guide to principal components. New York: John Wiley & sons; 1991.
12. Jolliffe IT. Principal component analysis. 2nd ed. New York: Springer-Verlag New York, Inc.; 2002.
13. Pinto J, Barros AS, Domingues MR, Goodfellow BJ, Galhano E, Pita C, et al. Following healthy pregnancy by NMR metabolomics of plasma and correlation to urine. J Proteome Res. 2015;14(2):1263–74.
14. Hotelling H. The generalization of student's ratio. Ann Math Stat. 1931;2(3):360–78.
15. Berkane M, Bentler PM. Estimation of contamination parameters and identification of outliers in multivariate data. Sociol Methods Res. 1988;17(1):55–64.

16. Filzmoser P, Ruiz-Gazen A, Thomas-Agnan C. Identification of local multivariate outliers. Stat Pap. 2014;55(1):29–47.
17. Magis D, De Boeck P. Identification of differential item functioning in multiple-group settings: a multivariate outlier detection approach. Multivar Behav Res. 2011;46(5):733–55.
18. Rocke DM, Woodruff DL. Identification of outliers in multivariate data. J Am Stat Assoc. 1996;91(435):1047–61.
19. Hubert M, Rousseeuw PJ, Vanden BK. ROBPCA: a new approach to robust principal component analysis. Technometrics. 2005;47(1):64–79.
20. Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemom Intell Lab. 2001;58(2):109–30.
21. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemom. 2002;16(3):119–28.
22. Eriksson L, Byrne T, Johansson E, Trygg J, Wikstrom C. Multi- and megavariate data analysis basic principles and applications, third revised edition. Malmo: MKS Umetrics AB; 2013.
23. Bylesjo M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemom. 2006;20(8–10):341–51.
24. Yap IK, Brown IJ, Chan Q, Wijeyesekera A, Garcia-Perez I, Bictash M, et al. Metabolome-wide association study identifies multiple biomarkers that discriminate north and south Chinese populations at differing risks of cardiovascular disease: INTERMAP study. J Proteome Res. 2010;9(12):6647–54.
25. Pinto RC, Trygg J, Gottfries J. Advantages of orthogonal inspection in chemometrics. J Chemom. 2012;26(6):231–5.
26. Estivill-Castro V. Why so many clustering algorithms: a position paper. ACM SIGKDD Explorations Newsletter. 2002;4(1):65–75.
27. Li X, Hansen J, Zhao XJ, Lu X, Weigert C, Haring HU, et al. Independent component analysis in non-hypothesis driven metabolomics: Improvement of pattern discovery and simplification of biological data interpretation demonstrated with plasma samples of exercising humans. J Chromatogr B. 2012;910:156–62.
28. Liu Y, Smirnov K, Lucio M, Gougeon RD, Alexandre H, Schmitt-Kopplin P. MetICA: independent component analysis for high-resolution mass-spectrometry based non-targeted metabolomics. BMC Bioinf. 2016;17:114.
29. Monakhova YB, Godelmann R, Kuballa T, Mushtakova SP, Rutledge DN. Independent components analysis to increase efficiency of discriminant analysis methods (FDA and LDA): Application to NMR fingerprinting of wine. Talanta. 2015;141:60–5.
30. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Anal Chem. 2008;80(1):115–22.
31. Keun HC, Ebbels TM, Bollard ME, Beckonert O, Antti H, Holmes E, et al. Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles. Chem Res Toxicol. 2004;17(5):579–87.
32. Stenlund H, Madsen R, Vivi A, Calderisi M, Lundstedt T, Tassini M, et al. Monitoring kidney-transplant patients using metabolomics and dynamic modeling. Chemom Intell Lab. 2009;98(1):45–50.
33. Pinto RC, Gerber L, Eliasson M, Sundberg B, Trygg J. Strategy for minimizing between-study variation of large-scale phenotypic experiments using multivariate analysis. Anal Chem. 2012;84(20):8675–81.
34. Smilde AK, Jansen JJ, Hoefsloot HC, Lamers RJ, van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. Bioinformatics. 2005;21(13):3043–8.
35. Jansen JJ, Hoefsloot HCJ, van der Greef J, Timmerman ME, Westerhuis JA, Smilde AK. ASCA: analysis of multivariate data obtained from an experimental design. J Chemom. 2005;19(9):469–81.

36. Timmerman ME, Hoefsloot HC, Smilde AK, Ceulemans E. Scaling in ANOVA-simultaneous component analysis. Metabolomics. 2015;11(5):1265–76.
37. Vis DJ, Westerhuis JA, Smilde AK, van der Greef J. Statistical validation of megavariate effects in ASCA. BMC Bioinf. 2007;8:322.
38. Jonsson P, Wuolikainen A, Thysell E, Chorell E, Stattin P, Wikstrom P, et al. Constrained randomization and multivariate effect projections improve information extraction and biomarker pattern discovery in metabolomics studies involving dependent samples. Metabolomics. 2015;11(6):1667–78.
39. Bjorkblom B, Wibom C, Jonsson P, Moren L, Andersson U, Johannesen TB, et al. Metabolomic screening of pre-diagnostic serum samples identifies association between alpha- and gamma-tocopherols and glioblastoma risk. Oncotarget. 2016; 7(24):37043–37053.
40. Szymanska E, Saccenti E, Smilde AK, Westerhuis JA. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics. 2012;8 Suppl 1:3–16.
41. Shao J. Linear-model selection by cross-validation. J Am Stat Assoc. 1993;88(422):486–94.
42. Worley B, Halouska S, Powers R. Utilities for quantifying separation in PCA/PLS-DA scores plots. Anal Biochem. 2013;433(2):102–4.
43. Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS (R) models. J Chemom. 2008;22(11–12):594–600.
44. Van der Voet H. Comparing the predictive accuracy of models using a simple randomization test. Chemom Intell Lab. 1994;25:313–23.
45. Eigenvector Research I. PLS toolbox: Permutation Test: Eigenvector Research, Manson, WA, USA, Inc.; 2014. http://wiki.eigenvector.com/index.php?title=Tools:_Permutation_Test
46. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561–77.
47. Galindo-Prieto B, Eriksson L, Trygg J. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). J Chemom. 2014;28(8):623–32.
48. Galindo-Prieto B, Eriksson L, Trygg J. Variable influence on projection (VIP) for OPLS models and its applicability in multivariate time series analysis. Chemom Intell Lab. 2015;146:297–304.
49. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. Metabolites. 2012;2(4):775–95.
50. Tzoulaki I, Ebbels TM, Valdes A, Elliott P, Ioannidis JP. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. Am J Epidemiol. 2014;180(2):129–39.
51. Dunn WB, Lin W, Broadhurst D, Begley P, Brown M, Zelena E, et al. Molecular phenotyping of a UK population: defining the human serum metabolome. Metabolomics. 2015;11:9–26.
52. Benjamini Y, Hochberg Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. J Roy Stat Soc B Methodol. 1995;57(1):289–300.
53. Benjamini Y, Cohen R. Weighted false discovery rate controlling procedures for clinical trials. Biostatistics. 2016.
54. Chadeau-Hyam M, Ebbels TMD, Brown IJ, Chan Q, Stemler J, Huang CC, et al. Metabolic profiling and the metabolome-wide association study: significance level for biomarker identification. J Proteome Res. 2010;9(9):4620–7.
55. Bictash M, Ebbels TM, Chan Q, Loo RL, Yap IKS, Brown IJ, et al. Opening up the "Black Box": metabolic phenotyping and metabolome-wide association studies in epidemiology. J Clin Epidemiol. 2010;63(9):970–9.
56. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011;6(7):1060–83.
57. Sysi-Aho M, Katajamaa M, Yetukuri L, Oresic M. Normalization method for metabolomics data using optimal selection of multiple internal standards. BMC Bioinf. 2007;8:93.
58. Bijlsma S, Bobeldijk I, Verheij ER, Ramaker R, Kochhar S, Macdonald IA, et al. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. Anal Chem. 2006;78(2):567–74.

59. Bylesjo M, Eriksson D, Sjodin A, Jansson S, Moritz T, Trygg J. Orthogonal projections to latent structures as a strategy for microarray data normalization. BMC Bioinf. 2007;8:207.

60. Mattsson A, Karrman A, Pinto R, Brunstrom B. Metabolic profiling of chicken embryos exposed to perfluorooctanoic acid (PFOA) and agonists to peroxisome proliferator-activated receptors. PLoS One. 2015;10(12):e0143780.

61. Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res. 2009;37(Web Server issue):W652–60.

62. Xia J, Wishart DS. Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. Curr Protoc Bioinf. 2011;Chapter 14:Unit 14 0.

63. Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protoc. 2011;6(6):743–60.

64. Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0 – a comprehensive server for metabolomic data analysis. Nucleic Acids Res. 2012;40(Web Server issue):W127–33.

65. Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0 – making metabolomics more meaningful. Nucleic Acids Res. 2015;43(W1):W251–7.

66. Kastenmuller G, Romisch-Margl W, Wagele B, Altmaier E, Suhre K. metaP-server: a web-based metabolomics data analysis tool. J Biomed Biotechnol. 2011; Volume 2011, Article ID 839862, 7 pages.

67. Giacomoni F, Le Corguille G, Monsoor M, Landi M, Pericard P, Petera M, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. Bioinformatics. 2015;31(9):1493–5.

68. Davidson RL, Weber RJ, Liu H, Sharma-Oates A, Viant MR. Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. Gigascience. 2016;5:10.

69. Kale N, Steinbeck C, Consortium P. PhenoMeNal – an e-infrastructure for analysis of metabolic phenotype data: Metabonews. 2016. Available from: http://www.metabonews.ca/Jan2016/MetaboNews_Jan2016.htm.

70. Henry VJ, Bandrowski AE, Pepin AS, Gonzalez BJ, Desfeux A. OMICtools: an informative directory for multi-omic data analysis. Database (2014) 2014: bau069 doi:10.1093/database/bau069.

Chapter 8
Computational Strategies for Biological Interpretation of Metabolomics Data

Jianguo Xia

Abstract Biological interpretation of metabolomics data relies on two basic steps: metabolite identification and functional analysis. These two steps need to be applied in a coordinated manner to enable effective data understanding. The focus of this chapter is to introduce the main computational concepts and workflows during this process. After a general overview of the field, three sections will be presented: the first section will introduce the main computational methods and bioinformatics tools for metabolite identification using spectra from common analytical platforms; the second section will focus on introducing major bioinformatics approaches for functional enrichment analysis of metabolomics data; and the last section will discuss the three main workflows in current metabolomics studies, including the chemometrics approach, the metabolic profiling approach and the more recent chemo-enrichment analysis approach. The chapter ends with summary and future perspectives on computational metabolomics.

Keywords Metabolomics • Chemometrics • Metabolic profiling • Metabolite set enrichment analysis • Chemo-enrichment analysis

Abbreviations

AMDIS   Automated mass spectral deconvolution and identification system
BATMAN   Bayesian automated metabolite analyzer for NMR

J. Xia
Institute of Parasitology, and Department of Animal Science, McGill University,
Sainte Anne de Bellevue, QC H9X 3V9, Canada

Department of Microbiology and Immunology, McGill University,
Montreal, QC H3A 2B4, Canada
e-mail: jeff.xia@mcgill.ca

GC-MS   Gas chromatography mass spectrometry
CSF   Cerebral spinal fluid
GO   Gene ontology
GSEA   Gene set enrichment analysis
LC-MS   Liquid chromatography mass spectrometry
MSEA   Metabolite set enrichment analysis
NIST   National Institute of Standards and Technology
NMR   Nuclear magnetic resonance
PCA   Principal component analysis
PLS-DA   Partial least squares discriminant analysis
OPLS-DA   Orthogonal partial least squares discriminant analysis
ORA   Overrepresentation analysis
PCR   Polymerase chain reaction

8.1   Introduction

Measuring metabolites and interpreting their biological relevance within the contexts of different experimental conditions are the primary objective in metabolomics researches. To achieve this objective, two basic steps need to be performed: metabolite identification and functional analysis, with the former providing the necessary inputs for the latter operation. These two steps need to be executed in a coordinated manner to promote efficient biological understanding. However, significant challenges remain in both steps.

The ultimate goal of metabolomics is to achieve comprehensive and high-throughput metabolome measurement. This goal is hampered by at least three major obstacles: (1) small compounds have diverse chemical properties, making it difficult to assay many metabolites simultaneously using a single analytical platform; (2) there is no effective amplification technique available to facilitate detection of low-abundance metabolites (such as using PCR for DNA molecules); and (3) many metabolites lack unique spectral signatures to allow unambiguous compound asignment. Nuclear magnetic resonance (NMR) spectroscopy and gas or liquid chromatography coupled with mass spectrometry (GC- or LC-MS) are commonly used in combination to improve the metabolome coverage. Metabolite identification is mainly performed by searching the spectral features against a reference spectral library. However, searching a comprehensive spectral database often leads to many potential hits with similar matching scores, and researchers often need to manually choose the most probable identities based on the context and domain knowledge. This step represents a key bottleneck in current metabolomics studies. Better algorithms and more context-specific databases are needed to enable high-throughput and high-accurate metabolite identifications.

Knowing compound identities is the first step toward biological interpretation of metabolomics data. The conventional procedure after this step involves manually looking up the metabolites of interest in different compound databases, reading rel-

evant literature, and finally synthesizing the information into a justifiable biological "story" based on the overall information obtained. This approach is subjective and time-consuming. Over the past decade, many computer-assisted data interpretation strategies have been developed. Among them, functional enrichment analysis using a predefined knowledge database has gained wide acceptance in omics data interpretation. The basic idea is to shift the unit of analysis from a single molecule to groups of functionally related molecules (i.e., those within the same pathway or biological process). This approach directly connects statistical significance with biological interpretation. More advanced algorithms have also been recently implemented that are able to integrate the dependencies and connectivities among different molecules to further reveal the biological insight and to improve system understanding.

Based on their strategies in dealing with metabolite identification and functional analysis, current metabolomics workflows can be summarized into three general categories: the chemometrics approach (also known as untargeted metabolomics), the metabolic profiling approach (also known as targeted or quantitative metabolomics), and the chemo-enrichment analysis approach (Fig. 8.1). The chemometrics approach focuses on identifying and interpreting a subset of spectral features that are found to have changed significantly during the experimental studies; the metabolic profiling approach aims to comprehensively characterize all metabolites in the spectra before subsequent statistical and functional analysis; and the more recent chemo-enrichment analysis approach directly maps spectral features into metabolic pathways/networks and then tests the enrichment of the collective chemical signals generated from these biological processes, which largely avoids the time-consuming step for accurate compound assignment.

This chapter is organized into three sections. The first section introduces the main computational approaches for metabolite identification from common analytical platforms (Fig. 8.1, Step 1); the second section describes the three main bioinfor-
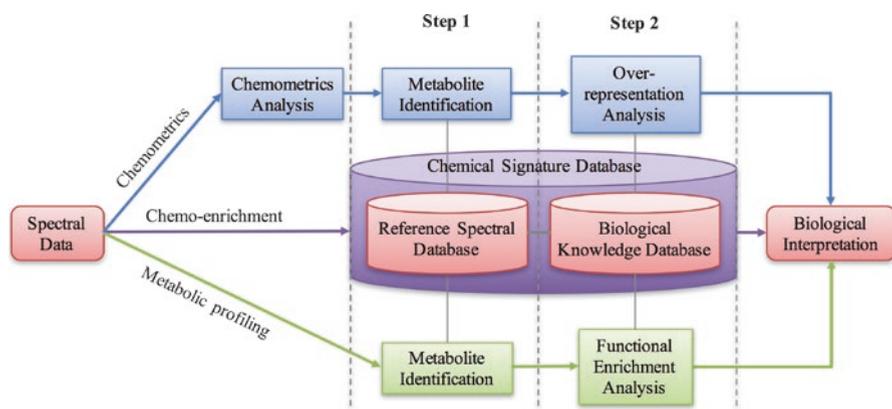
Fig. 8.1 The diagram summarizes the three computational strategies for metabolomics data interpretation: the chemometrics approach (*top*), the metabolic profiling (*bottom*), and the chemo-enrichment analysis (*middle*). The dotted lines delineate the two major steps in the process: metabolite identification and functional analysis. Note that these two steps are integrated into a single one in the chemo-enrichment approach

matics approaches for functional enrichment analysis (Fig. 8.1; Step 2); and the last section compares the three metabolomics workflows for biological interpretation. Each section is further organized under subtitles describing the computational concepts, the available bioinformatics tools, and their main features.

8.2 Metabolite Identification Methods

Although it is possible to determine the identity of a single metabolite *de novo* through labor-intensive NMR or MS-based methods, this approach is generally infeasible in metabolomics in which hundreds to thousands of compound species are measured simultaneously. In practice, compound identification is based on matching features from sample spectra against a reference spectral database, and a closely matched hit will be considered as the putative identity of the corresponding spectral peaks. However, many metabolites do not produce unique, detectable signatures in their NMR or MS spectra to permit unambiguously determination of their identities. The situation is further complicated by peak shifts and overlaps typical in the spectra of complex biological samples. Direct database search tends to yield high percentage of false positives, and further labor-intensive manual refinement is usually necessary. To improve the efficiency of metabolite identification, two general computational strategies have been employed: (1) limiting the search space to only those biologically and biochemically possible candidates by developing more context-specific spectral databases, and (2) improving the peak assignment algorithms by incorporating prior knowledge based on spectral dependencies, biochemical connectivities and biological relationships.

8.2.1 Compound Identification from NMR Spectra

Proton NMR spectroscopy has been widely used in metabolomics studies involving human biofluids. Multiple small-molecule metabolites can be measured simultaneously without prior separation, which greatly simplifies the sample preparation requirements. NMR spectra are highly reproducible, and samples analyzed from one spectrometer will generate near-identical results to those measured on other types of spectrometers. These features have made NMR spectroscopy a platform of choice for large-scale collaborative metabolomics projects.

The Chenomx NMR Suite (Chenomx, Canada) is a widely used metabolomics tool for processing and profiling one-dimensional (1D) proton NMR spectra. The main feature of Chenomx is the integration of a powerful interactive visualization interface with a reference spectral library for over 600 metabolites that are detectable by NMR in common biofluids. Metabolite identification and quantification are achieved through manual peak fitting against those reference spectra. Another widely used commercial tool is the AMIX software package (Bruker Biospin GmbH, Germany), which offers similar features. The company has recently implemented a

software (FoodScreener) that supports automated high-throughput targeted metabolomics profiling for wine, honey, and juice using defined spectra libraries.

Compared to commercial tools, public bioinformatics tools for NMR-based metabolomics tend to focus on spectral alignment, binning and batch processing [1, 2]. They usually lack user-friendly interface or comprehensive spectra libraries to support manual compound identification. As public NMR spectra libraries become increasingly available [3, 4], this situation has begun to change. For instance, the Bayesian automated metabolite analyzer for NMR (BATMAN) is an R package designed for deconvolution and quantification of metabolites from 1D proton NMR spectra of complex mixtures [5, 6]. The Bayesian model incorporates characteristic peak patterns of metabolites and also accounts for peak shifts commonly seen in NMR spectra of biological samples. BATMAN can compute relative concentrations of the compounds together with associated uncertainty estimates using a Markov chain Monte Carlo algorithm. The procedure is computationally intensive and usually requires hours of CPU time to process a single spectrum of common biofluids. Bayesil is a web-based tool that supports automated phasing, referencing, baseline correction, metabolite identification, and quantification for 1D proton NMR metabolomics spectra [7]. The algorithm is implemented based on probabilistic graphical models and a prior knowledge of probable biofluid compositions with built-in support for cerebral spinal fluid (CSF), serum, and plasma. Compared to BATMAN, Bayesil can process a spectrum in a few minutes with high precision and recall. For excessively overlapped NMR spectra of complex biofluid mixtures, two-dimensional (2D) NMR is often used to help resolve spectra ambiguities for metabolite identification purpose. The Bruker AMIX package (Bruker Biospin GmbH, Germany) can also support 2D NMR analysis. The Java desktop application MetaboMiner and the R package rNMR are two public bioinformatics tools for metabolite identification from 2D NMR spectra [8, 9].

8.2.2 Compound Identification from GC-MS Spectra

GC-MS offers a high degree of chromatographic resolution and reproducibility. The platform is suitable for measuring volatile, low-molecular mass (<500 Da), and thermally stable compounds such as sugars, fatty acids, and amino acids. For large and polar compounds, chemical derivatization is often employed to improve their volatility and thermal stability. The most commonly used ionization technique in GC-MS is electron ionization, which is very robust and reproducible. The characteristic mass spectral fragmentation patterns can be used to build a spectral library for metabolite identification.

Many software tools are available for metabolite identification and quantification from GC-MS-based metabolomics data. The automated mass spectral deconvolution and identification system (AMDIS) coupled with the National Institute of Standards and Technology (NIST) database is probably the most widely used software package for GC-MS data analysis [10]. The AnalyzerPro (SpectralWorks, UK) and ChromaTOF (LECO, USA) are the two widely used commercial tools for processing and profiling

the GC-MS spectra for metabolomics studies. Compared to NMR-based metabolomics data, more public bioinformatics tools are available for GC-MS spectral processing, deconvolution, alignment, as well as compound identification. Popular tools include BinBase [11], MetaQuant [12], MetabolomeExpress [13], MetaboliteDetector [14], TagFinder [15], *etc*. With the availability of public GC-MS spectral databases [16, 17] and our improved knowledge on the metabolite compositions of common biofluids such as CSF, serum, and urine [18–20], the GC-MS-based metabolomics is expected to be the most promising platform to deliver automated compound identification and quantification for a broad range of biofluids.

8.2.3 Compound Identification from LC-MS Spectra

Compared to GC-MS, LC-MS typically has lower chromatographic resolution and reproducibility. However, LC-MS techniques can access a much broader mass range (100–2000 Da) because volatilization or derivatization is not necessary. LC-MS is also a better choice for separating and identifying polar and nonvolatile compounds. Electrospray ionization and atmospheric pressure chemical ionization are the two most common ionization methods used in LC-MS. Both techniques will generate a molecular ion whose mass can be searched against a spectral database of known metabolites for possible identification. However, due to the finite mass accuracy of the MS equipment and the large number of potential formulas, using mass information alone is usually insufficient for metabolite identification [21].

To address this issue, many bioinformatics tools employ extra information to improve peak assignment and metabolite identification from LC-MS metabolomics data. One approach incorporates known chemical reactions among candidate compounds based on the metabolic pathways/networks to improve annotation, as certain combinations would make more biochemical sense when they are detected together. For instance, the MI-Pack and the ProbMetab are able to use the metabolic pathway information obtained from MetaCyc or KEGG to improve metabolite identification [22, 23]. The second approach takes into consideration of the dependency structures of multiple peaks (isotopologues, adducts, molecular fragments, and multiply charged ions) derived from each metabolite in a LC-MS spectrum to improve peak annotation. The MetAssign tool has implemented this approach [24]. The core algorithms used in these tools are based on graphical models, with most of them using a Bayesian approach to perform probabilistic annotation of metabolites.

8.3 Functional Analysis Approaches

Most metabolites can potentially participate in multiple functional roles within a biological system, and it is difficult to pinpoint the biological processes responsible for the profiles observed in a metabolomics experiment. A biological process is

typically made of a group of molecules. If a biological process is changed in a study, the molecules involved should have a higher potential to be identified as significant by the omics platform. Motivated by this concept, functional analysis has shifted the unit of analysis from a single molecule to a group of functionally related molecules. Instead of testing a single gene or metabolite, researchers now directly evaluate whether a group of molecules (representing a biological process) is consistently changed (enriched). This approach greatly simplifies the omics data interpretation and is more sensitive in detecting subtle but consistent changes occurred in a biological process.

The functional analysis requires two components: a knowledge database defining functionally related molecule groups and a statistical algorithm to perform enrichment tests. The popular gene set enrichment analysis (GSEA) tool is shipped with a comprehensive collection of gene sets in the form of Molecular Signature Database (MSigDB), which greatly facilitates the subsequent development of tools for enrichment analysis [25, 26]. In metabolomics, except the public metabolic pathway databases such as KEGG [27] or MetaCyc [28], a comprehensive collection of functionally related metabolite groups was unavailable until very recently. The first large collection of metabolite sets appeared in 2010 with the publication of the MSEA tool containing >6000 groups of metabolites based on pathways, diseases, genetic variants, and cellular compartments [29]. The other useful resource is the ConceptMetab database containing >16,000 biologically defined metabolite sets developed based on GO, KEGG, and Medical Subject Headings [30]. The ongoing developments of ontologies for systematic metabolite annotations are expected to greatly facilitate the development of enrichment analysis tools for metabolomics [31, 32]. Below I will introduce the three main categories of statistical approaches for functional analysis for metabolomics data: over-representation analysis (ORA), metabolite set enrichment analysis (MSEA), and metabolic pathway/network analysis.

8.3.1   Over-representation Analysis (ORA)

The ORA approach is a traditional strategy for enrichment analysis. It starts with a list of metabolites of interest and tests whether certain metabolite groups appear more often than would be expected by random chance. This type of analysis can be performed using Fisher's exact test, a chi-square test, a hypergeometric test, or its binomial approximation. To perform ORA, researchers need to first perform a statistical comparison such as t-tests or ANOVA and then select significant metabolites using a certain threshold or criterion (i.e., adjusted p-values <0.05). Fold change values are also considered sometimes during the selection process.

The ORA approach is very flexible to use and is simple to implement. It has been implemented in many metabolomics tools and databases including MSEA, MBRole, MetaPA, IMPaLA, MPEA, BiNChE, and ConceptMetab [29–31, 33–36]. A common critic of the approach is related to its somewhat arbitrary threshold to decide

whether a metabolite is significant or not. For instance, different cutoffs sometimes lead to different interpretations, and ORA cannot be applied if no significant metabolites are found in a given study. Another limitation is that all metabolites are treated equally after the selection, ignoring their quantitative differences. Despite these shortcomings, ORA remains widely used in omics data interpretation [37].

8.3.2   Metabolite Set Enrichment Analysis (MSEA)

The MSEA approach has been developed to address the shortcomings associated with ORA. It directly tests the enrichment of functional groups using the complete concentration data without preselection of significant metabolites. The MSEA is named after the popular GSEA developed for gene expression data interpretation [26]. The original GSEA approach first uses a univariate method to rank all the genes and then tests whether the ranks in the gene set differ from a uniform distribution, using a weighted Kolmogorov-Smirnov test. The p-value for each gene set is calculated via permutation tests. Since then, many different variations of the GSEA have been developed with different performance characteristics [38]. For instance, the GlobalTest method has shown a general improved performance in terms of sensitivity, versatility, and computational efficiency and works especially well if most of the molecules within a group are associated with the phenotype in a modest way [38]. The algorithm is based on a generalized linear model to test whether a group of molecules is significantly associated with a specific phenotype [39].

Several bioinformatics tools have been implemented to support MSEA for metabolomics data. The web-based MSEA program (now part of MetaboAnalyst) is the first tool with such capacity to support functional analysis for quantitative metabolomics data [29, 40]. Like the original GSEA tool, it contains built-in libraries of defined metabolite sets associated with metabolic pathways, diseases, genetic variations, cellular compartments, *etc*. The GlobalTest algorithm is used for quantitative enrichment analysis directly from a metabolite concentration table. Another metabolomics tool with MSEA capacity is the MeltDB, which uses a modified GSEA method against the metabolite sets defined by the KEGG metabolic pathways [41]. With improved functional annotations for metabolite sets such as the ConceptMetab and metabolite ontologies [30, 31], more metabolomics tools with MSEA support will be developed in the near future.

8.3.3   Metabolic Pathway and Network Analysis

In the MSEA approach, groups of molecules labeled with biologically meaningful names are used to organize a large body of our current knowledge, making it a popular approach to aid in omics data interpretation. However, this "flat" representation of knowledge followed by enrichment tests based on group memberships ignores

the connectivities and dependencies among molecules as well as the inherent overlaps/hierarchies among different groups. For instance, changes at a central location within a pathway tend to have a larger impact on its overall functions compared to changes at the very downstream. Integrating the functional analysis with pathway/network topology analysis will help improve the accuracy in ranking the resulting list of biological processes.

In gene expression data analysis, the TopGO is probably the first method that integrates knowledge about relationships between different GO terms into calculating the statistical significances to increase the explanatory power of GO enrichment analysis [42]. The signaling pathway impact analysis (SPIA) is another approach that combines the evidence obtained from classical enrichment analysis with a novel type of evidence that utilize the pathway topology to measure the impact on a given pathway [43, 44]. Both approaches have been shown to provide increased sensitivity and specificity when compared to other methods based solely on enrichment analysis. Many more tools have been implemented to take into consideration of pathway topology for enrichment analysis of gene expression data [45]. Applications of similar approaches to metabolomics are currently hampered by two obstacles: firstly, metabolomics typically can only measure a small fraction of any given metabolic pathway at the moment, which greatly limits our ability to evaluate the impact on the overall pathway; secondly, the development of a hierarchical ontology system for metabolite annotation has not been well established to allow easy plug-in by different bioinformatics tools, as is the case of gene ontology system. Therefore, current metabolomics tools focus primarily on enrichment analysis and visualization of metabolic pathways. The web-based tool MetPA (now part of MetaboAnalyst) is the first tool that supports both enrichment analysis and topology analysis within the context of KEGG metabolic pathways [36]. The MetScape is another tool implemented as a Cytoscape plug-in that is able to incorporate prior knowledge of pathways and molecular interactions for metabolomics pathway analysis and network visualization [46].

8.4 Metabolomics Workflows for Biological Interpretation

As indicated in Fig. 8.1, current metabolomics workflows can be largely divided into three general categories based on their strategies in metabolite identification and functional analysis: chemometrics approach, metabolic profiling approach, and chemo-enrichment analysis approach. The chemometrics approach focuses on identifying and interpreting a subset of spectral features that are found to be important within the study. It is relatively high throughput, as only the significant features need to be characterized. This approach is widely used in exploratory metabolomics studies and for discovery of novel biomarkers. A main drawback associated with this approach is the difficulties in biological interpretation, as a limited number of compounds are usually insufficient to pinpoint the underlying biological processes. In contrast to the chemometrics approach, the metabolic profiling approach aims to

characterize all detectable metabolites from the spectral data before subsequent functional analysis. It generally yields better sensitivity, selectivity, and interpretability but is of very limited use for novel biomarker discovery. The main drawback associated with this approach is that the metabolite identification is usually time-consuming and labor intensive. The chemo-enrichment analysis approach has been recently developed to address the limitations associated with both chemometrics and metabolic profiling. It aims to estimate biological activities directly from the spectral features by mapping all possible metabolite matches to metabolic pathways/networks and then comparing the resulting profiles to identify the enriched biological processes.

8.4.1   The Chemometrics Approach

Chemometrics methods are a class of multivariate statistical methods heavily used in analytical chemistry and later metabolomics. These methods are especially useful for analysis and modeling of high-dimensional complex spectral data in untargeted metabolomics, where features (peaks or spectral bins) are highly correlated. The two most commonly used chemometrics methods are principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). PCA aims to project a high-dimensional data into a low-dimensional space that captures the most variance of the data. The direction of projection is computed based on the data (X) only, without referring to the experimental conditions (Y). PCA is suitable for data overview and to understand the inherent patterns within the data. There is no guarantee that the directions of maximum variance will be the same as the directions of the variance associated with the experimental conditions. In contrast, PLS-DA aims to project a high-dimensional data X into a low-dimensional space that capture the most covariance between X and Y. It is often used to identify the spectral features that are different across experimental conditions. Orthogonal PLS-DA (OPLS-DA) is a variant of PLS-DA which uses orthogonal signal correction to maximize the explained covariance between X and Y on the first component, and the remaining components capture variance in X which is orthogonal to Y [47].

The chemometrics approach is composed of three general steps. A chemometrics method such as PLS-DA or OPLS-DA is first applied to analyze the spectral data to identify significant features associated with the experimental conditions. This step can be performed using several commercial or public tools. The SIMCA-P program (Umetrics, Sweden) is widely used by the metabolomics community. It offers excellent graphic capabilities and comprehensive analysis options for chemometrics methods including PCA, PLS/OPLS-DA, and SIMCA (soft independent modeling of class analogy). MetaboAnalyst is a web-based tool that supports comprehensive metabolomics data processing, normalization, and chemometrics analysis (PCA, PLS-DA, and more recently, Orthogonal PLS-DA [40, 48, 49]. For users who know how to program in R, many R packages are available for chemometrics analysis [50, 51]. After selection of significant spectral features, the second step is to perform

compound identification using the tools and resources as described in Sect. 8.2. In the third step, the list of identified metabolites will be subject to ORA to find out which pathways or biological processes are significantly enriched biological processes are significantly enriched for biological interpretation (Sect. 8.3).

8.4.2   The Metabolic Profiling Approach

Metabolic profiling is often used to validate and expand upon results obtained from untargeted analysis. It is also increasingly applied to study variations of metabolite concentrations in relatively well-characterized biofluids such as CSF, blood, urine, etc. Although the process of metabolite identification and quantification is currently a rate-limiting step, this approach offers several distinctive advantages. For instance, metabolic profiling significantly improves statistical power by reducing the number of features from 1000–10,000 of features peaks to hundreds of metabolites. The manual process also largely removes missing values and spectral noises, which greatly facilitates downstream statistical analysis and biomarker discovery.

The biggest advantage of metabolic profiling is the ease of data interpretation. The complete metabolite concentration table can be directly used for MSEA, metabolic pathway, or network analysis using the tools described in Sect. 8.3. The web-based tool MetaboAnalyst provides extensive functions for functional analysis and interpretation for data generated from metabolic profiling approach. Importantly, the metabolite concentration data is very compatible with other omics data and can be analyzed together to help pinpoint the biological pathways involved in the experimental conditions. There are several bioinformatics tools that provide support for integrated analysis of metabolomics data with transcriptomics data. For instance, the MetaCore (Thomson Reuters, USA) allows joint analysis and visual exploration within its comprehensive collections of pathway and network [52]. The public tools IMPaLA and MetScape can accept a list of metabolites and a list of genes for joint analysis and visualization on metabolic networks [34, 46]. INMEX is a web-based tool that supports statistical analysis and joint enrichment analysis for data sets from transcriptomics and metabolic profiling studies [53].

8.4.3   The Chemo-enrichment Analysis Approach

The chemo-enrichment analysis approach is a more recent strategy developed to facilitate high-throughput interpretation of metabolomics data generated from high-resolution LC-MS platforms. The key idea is to redefine the metabolite sets, metabolic pathways, or networks using the spectral features (i.e., m/z) of the corresponding metabolites and then test the enrichment of these "collective chemical signals" within the untargeted metabolomics data. Accurate compound identification is not necessary because errors (i.e., incorrect peak assignments) tend to will be randomly

distributed, while the true biological signals will be consistent, which can be detected by testing the enrichment of their collective chemical signals. The chemo-enrichment approach directly connects spectral features with biological interpretations without explicit compound identification. In practice, the metabolite identification is performed *post hoc* for those enriched biological processes of interest. The approach is useful in metabolomics studies for organisms with well-annotated metabolic pathways and networks.

There are a few tools that offer support for chemo-enrichment analysis. The *mummichog* is probably the first bioinformatics tool that implemented the concept [54]. It accepts two lists of spectral peaks (i.e., *m/z* values) – a significant peak list (i.e., those identified using *t*-tests) and a reference peak list (all features detected in the MS experiment). The significant peak lists are then searched against a database to find all potential matches to metabolic pathways and networks. The result is compared with those obtained based on peak lists randomly drawn from the reference peaks to compute statistical significance. The tool is available as a Python program. It has been recently implemented in the popular web-based tool XCMS Online to reach a broader audience [55]. MarVis-Pathway is a more recent stand-alone bioinformatics tool with chemo-enrichment analysis feature. It employs a hypergeometric-based approach to evaluate the enrichment of metabolic pathways directly from the untargeted metabolomics data [56].

8.5 Summary and Future Perspectives

This chapter introduces several key concepts and recent developments in computational strategies for metabolomics data interpretation. Compound identification constitutes a major bottleneck in current metabolomics studies. Accurate metabolite identification requires manual intervention and additional laboratory experiments. Advances in both analytical platforms and algorithms are making ways to enable high-throughput data interpretation. Integrating high-resolution analytics, context-specific reference spectral databases, together with advanced algorithms that incorporate chemical and biological information, we will be able to achieve accurate and high-throughput metabolite identification and biological interpretation.

Identification of metabolites (accurately or approximately) is a prerequisite for data interpretation. The list of compounds needs to be put into proper biological context by identifying their roles in metabolic pathways, their interconnectivity with other metabolites, links to genetic variations, or associations with pathophysiological conditions. The group-based functional enrichment analysis has been developed to address this issue. This is an active research area with a wide range of tools and implementations available. Given the current limitations of the knowledge databases and the statistical algorithms, the resulting enrichment *p*-values should be treated as a ranking system for data exploration and hypothesis generating rather than an absolute cutoff for decision-making purpose.

Compared to transcriptomics, metabolomics is closer to an organism's phenotype and is more sensitive to environmental perturbations. Small compounds represent the final products of complex interactions between the host genetics and environment. The metabolome includes both the endogenous metabolites produced directly by the host organism and the compounds derived from microbial, xenobiotic, dietary, and other exogenous sources. As a result, metabolomics is increasingly applied to study the impact of diet, gut microbiota, and environmental exposures. Developing novel bioinformatics tools and specialized knowledge databases to support these applications are the new frontiers in the current computational metabolomics.

References

1. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe – a multidimensional spectral processing system based on Unix pipes. J Biomol NMR. 1995;6(3):277–93. doi:10.1007/Bf00197809.
2. Zhao Q, Stoyanova R, Du SY, Sajda P, Brown TR. HiRes – a tool for comprehensive assessment and interpretation of metabolomic data. Bioinformatics. 2006;22(20):2562–4. doi:10.1093/bioinformatics/btl428.
3. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0–the human metabolome database in 2013. Nucleic Acids Res. 2013;41(Database issue):D801–7. doi:10.1093/nar/gks1065.
4. Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, et al. Metabolite identification via the Madison metabolomics consortium database. Nat Biotechnol. 2008;26(2):162–4. doi:10.1038/nbt0208-162.
5. Hao J, Liebeke M, Astle W, De Iorio M, Bundy JG, Ebbels TMD. Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. Nat Protoc. 2014;9(6):1416–27. doi:10.1038/nprot.2014.090.
6. Hao J, Astle W, De Iorio M, Ebbels TMD. BATMAN-an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. Bioinformatics. 2012;28(15):2088–90. doi:10.1093/bioinformatics/bts308.
7. Ravanbakhsh S, Liu P, Bjorndahl TC, Mandal R, Grant JR, Wilson M, et al. Accurate, fully-automated NMR spectral profiling for metabolomics (vol 10, e0124219, 2015). Plos One. 2015;10(7). doi:10.1371/journal.pone.0132873.
8. Xia JG, Bjorndahl TC, Tang P, Wishart DS. MetaboMiner – semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. BMC Bioinformatics. 2008;9:507. doi:10.1186/1471-2105-9-507.
9. Lewis IA, Schommer SC, Markley JL. rNMR: open source software for identifying and quantifying metabolites in NMR spectra. Magn Reson Chem. 2009;47:S123–6. doi:10.1002/mrc.2526.
10. Stein SE. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J Am Soc Mass Spectr. 1999;10(8):770–81. doi:10.1016/S1044-0305(99)00047-1.
11. Fiehn O, Wohlgemuth G, Scholz M. Setup and annotation of metabolomic experiments by integrating biological and mass spectrometric metadata. Lect Notes Comput Sci. 2005;3615:224–39. doi:10.1007/11530084_18.
12. Bunk B, Kucklick M, Jonas R, Munch R, Schobert M, Jahn D, et al. MetaQuant: a tool for the automatic quantification of GC/MS-based metabolome data. Bioinformatics. 2006;22(23):2962–5. doi:10.1093/bioinformatics/btl526.

13. Carroll AJ, Badger MR, Millar AH. The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. BMC Bioinformatics. 2010;11. doi:10.1186/1471-2105-11-376.
14. Hiller K, Hangebrauk J, Jager C, Spura J, Schreiber K, Schomburg D. MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. Anal Chem. 2009;81(9):3429–39. doi:10.1021/ac802689c.
15. Luedemann A, Strassburg K, Erban A, Kopka J. TagFinder for the quantitative analysis of gas chromatography – mass spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics. 2008;24(5):732–7. doi:10.1093/bioinformatics/btn023.
16. Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, et al. GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett. 2005;579(6):1332–7. doi:10.1016/j.febslet.2005.01.029.
17. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, 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;81(24):10038–48. doi:10.1021/ac9019522.
18. Wishart DS, Lewis MJ, Morrissey JA, Flegel MD, Jeroncic K, Xiong YP, et al. The human cerebrospinal fluid metabolome. J Chromatogr B. 2008;871(2):164–73. doi:10.1016/j.jchromb.2008.05.001.
19. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. Plos One. 2011;6(2). doi:10.1371/journal.pone.0016957.
20. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The human urine metabolome. Plos One. 2013;8(9). doi:10.1371/journal.pone.0073076.
21. Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics. 2006;7. doi:10.1186/1471-2105-7-234.
22. Weber RJM, Viant MR. MI-pack: increased confidence of metabolite identification in mass spectra by integrating accurate masses and metabolic pathways. Chemometr Intell Lab. 2010;104(1):75–82. doi:10.1016/j.chemolab.2010.04.010.
23. Silva RR, Jourdan F, Salvanha DM, Letisse F, Jamin EL, Guidetti-Gonzalez S, et al. ProbMetab: an R package for Bayesian probabilistic annotation of LC-MS-based metabolomics. Bioinformatics. 2014;30(9):1336–7. doi:10.1093/bioinformatics/btu019.
24. Daly R, Rogers S, Wandy J, Jankevics A, Burgess KEV, Breitling R. MetAssign: probabilistic annotation of metabolites from LC-MS data using a Bayesian clustering approach. Bioinformatics. 2014;30(19):2764–71. doi:10.1093/bioinformatics/btu370.
25. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739–40. doi:10.1093/bioinformatics/btr260.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50. doi:10.1073/pnas.0506580102.
27. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue):D199–205. doi:10.1093/nar/gkt1076.
28. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2016;44(D1):D471–80. doi:10.1093/nar/gkv1164.
29. Xia JG, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res. 2010;38:W71–7. doi:10.1093/nar/gkq329.
30. Cavalcante RG, Patil S, Weymouth TE, Bendinskas KG, Karnovsky A, Sartor MA. ConceptMetab: exploring relationships among metabolite sets to identify links among biomedical concepts. Bioinformatics. 2016;32(10):1536–43. doi:10.1093/bioinformatics/btw016.
31. Moreno P, Beisken S, Harsha B, Muthukrishnan V, Tudose I, Dekker A, et al. BiNChE: a web tool and library for chemical enrichment analysis based on the ChEBI ontology. BMC Bioinformatics. 2015;16. doi:10.1186/s12859-015-0486-3.

32. Hastings J, Chepelev L, Willighagen E, Adams N, Steinbeck C, Dumontier M. The chemical information ontology: provenance and disambiguation for chemical data on the biological semantic web. Plos One. 2011;6(10):e25513. doi:10.1371/journal.pone.0025513.

33. Kankainen M, Gopalacharyulu P, Holm L, Oresic M. MPEA-metabolite pathway enrichment analysis. Bioinformatics. 2011;27(13):1878–9. doi:10.1093/bioinformatics/btr278.

34. Kamburov A, Cavill R, Ebbels TMD, Herwig R, Keun HC. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. Bioinformatics. 2011;27(20):2917–8. doi:10.1093/bioinformatics/btr499.

35. Chagoyen M, Pazos F. MBRole: enrichment analysis of metabolomic data. Bioinformatics. 2011;27(5):730–1. doi:10.1093/bioinformatics/btr001.

36. Xia JG, Wishart DS. MetPA: a web-based metabolomics tool for pathway analysis and visualization. Bioinformatics. 2010;26(18):2342–4. doi:10.1093/bioinformatics/btq418.

37. da Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13. doi:10.1093/nar/gkn923.

38. Tarca AL, Bhatti G, Romero R. A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. Plos One. 2013;8(11). doi:10.1371/journal.pone.0079217.

39. Goeman JJ, Van De Geer SA, Van Houwelingen HC. Testing against a high dimensional alternative. J R Stat Soc Ser B (Stat Methodol). 2006;68(3):477–93. doi:10.1111/j.1467-9868.2006.00551.x.

40. Xia JG, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Res. 2015;43(W1):W251–7. doi:10.1093/nar/gkv380.

41. Persicke M, Ruckert C, Plassmeier J, Stutz LJ, Kessler N, Kalinowski J, et al. MSEA: metabolite set enrichment analysis in the MeltDB metabolomics software platform: metabolic profiling of Corynebacterium glutamicum as an example. Metabolomics. 2012;8(2):310–22. doi:10.1007/s11306-011-0311-6.

42. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22(13):1600–7. doi:10.1093/bioinformatics/btl140.

43. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, et al. A novel signaling pathway impact analysis. Bioinformatics. 2009;25(1):75–82. doi:10.1093/bioinformatics/btn577.

44. Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, et al. A systems biology approach for pathway level analysis. Genome Res. 2007;17(10):1537–45. doi:10.1101/gr.6202607.

45. Mitrea C, Taghavi Z, Bokanizad B, Hanoudi S, Tagett R, Donato M, et al. Methods and approaches in the topology-based analysis of biological pathways. Front Physiol. 2013;4. doi:10.3389/fphys.2013.00278.

46. Karnovsky A, Weymouth T, Hull T, Tarcea VG, Scardoni G, Laudanna C, et al. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. Bioinformatics. 2012;28(3):373–80. doi:10.1093/bioinformatics/btr661.

47. Bylesjo M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemometr. 2006;20(8–10):341–51. doi:10.1002/cem.1006.

48. Xia JG, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0-a comprehensive server for metabolomic data analysis. Nucleic Acids Res. 2012;40(W1):W127–33. doi:10.1093/nar/gks374.

49. Xia JG, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res. 2009;37:W652–60. doi:10.1093/nar/gkp356.

50. Thevenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the human adult urinary metabolome variations with Age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. J Proteome Res. 2015;14(8):3322–35. doi:10.1021/acs.jproteome.5b00354.

51. Edoardo G, Francesca C, Dimitrios S, Andrea S, Michela G, Jose MG-M, et al. muma, an R package for metabolomics univariate and multivariate statistical analysis. Curr Metabolomics. 2013;1(2):180–9. doi:10.2174/2213235X11301020005.

52. Jennen D, Ruiz-Aracama A, Magkoufopoulou C, Peijnenburg A, Lommen A, van Delft J, et al. Integrating transcriptomics and metabonomics to unravel modes-of-action of 2,3,7,8-tetrachlo

rodibenzo-p-dioxin (TCDD) in HepG2 cells. BMC Syst Biol. 2011;5:139. doi:10.1186/1752-0509-5-139.

53. Xia J, Fjell CD, Mayer ML, Pena OM, Wishart DS, Hancock RE. INMEX–a web-based tool for integrative meta-analysis of expression data. Nucleic Acids Res. 2013;41(Web Server issue):W63–70. doi:10.1093/nar/gkt338.

54. Li SZ, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, et al. Predicting network activity from high throughput metabolomics. Plos Comput Biol. 2013;9(7). doi:10.1371/journal.pcbi.1003123.

55. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS Online: a web-based platform to process untargeted metabolomic data. Anal Chem. 2012;84(11):5035–9. doi:10.1021/ac300698c.

56. Kaever A, Landesfeind M, Feussner K, Mosblech A, Heilmann I, Morgenstern B, et al. MarVis-Pathway: integrative and exploratory pathway analysis of non-targeted metabolomics data. Metabolomics. 2015;11(3):764–77. doi:10.1007/s11306-014-0734-y.

Part III: Applications and Future Trends of Metabolomics in Clinical Cases

Chapter 9
Applications of Metabolomics in Cancer Studies

Emily Grace Armitage and Michal Ciborowski

Abstract Since the start of metabolomics as a field of research, the number of studies related to cancer has grown to such an extent that cancer metabolomics now represents its own discipline. In this chapter, the applications of metabolomics in cancer studies are explored. Different approaches and analytical platforms can be employed for the analysis of samples depending on the goal of the study and the aspects of the cancer metabolome being investigated. Analyses have concerned a range of cancers including lung, colorectal, bladder, breast, gastric, oesophageal and thyroid, amongst others. Developments in these strategies and methodologies that have been applied are discussed, in addition to exemplifying the use of cancer metabolomics in the discovery of biomarkers and in the assessment of therapy (both pharmaceutical and nutraceutical). Finally, the application of cancer metabolomics in personalised medicine is presented.

Keywords Cancer • Metabolomics • Biomarker • Chemotherapy • Pharmaceutical • Nutraceutical • Effect of treatment • Personalised medicine

E.G. Armitage (✉)
Centre for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad CEU San Pablo, Campus Monteprincipe, Madrid, Spain

Wellcome Trust Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, College of Medical Veterinary and Life Sciences, Sir Graeme Davies Building, University of Glasgow, Glasgow, UK

Glasgow Polyomics, Wolfson Wohl Cancer Research Centre, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK
e-mail: emily.armitage@glasgow.ac.uk

M. Ciborowski
Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland

Abbreviations

9-AA	9-Aminoacridine
ALL	Acute lymphoblastic leukaemia
ATP	Adenosine triphosphate
CE	Capillary electrophoresis
CLL	Chronic lymphocytic leukaemia
ECOG	Eastern Cooperative Oncology Group
ESI	Electrospray ionisation
GC	Gas chromatography
GC×GC	Comprehensive two-dimensional gas chromatography
GPC	Glycerophosphocholine
HIF	Hypoxia inducible factor
HILIC	Hydrophilic interaction chromatography
LC	Liquid chromatography
MALDI	Matrix-assisted laser desorption ionisation
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NEDC	*N*-(1-naphthyl)ethylenediamine dihydrochloride
NMR	Nuclear magnetic resonance
NSCLC	Non-small cell lung cancer
p53	Cellular tumour antigen p53
PC	Phosphocholine
PTC	Papillary thyroid carcinoma
QqQ-MS	Triple quadrupole mass spectrometry
TCA	Tricarboxylic acid
tCho	Total choline
TOF	Time-of-flight

9.1   Introduction

Differences in central carbon metabolism between cancerous and normal cells were first demonstrated by Otto Warburg in the 1930s. By the 1950s, Warburg demonstrated that cancer cells preferentially used glycolysis over oxidative phosphorylation even in the presence of oxygen [1]. Furthermore, elevated glucose levels can suppress both glycolysis and oxidative phosphorylation via the 'Crabtree effect', a short-term, reversible response to glucose availability [2]. Although the phenotype of reliance on glycolysis rather than oxidative phosphorylation is not efficient for ATP production, glycolysis can provide intermediary precursors to feed into many biosynthetic pathways that ultimately generate nucleotides, amino acids and lipids, as well as ATP. The mechanisms by which this is achieved are multiple.
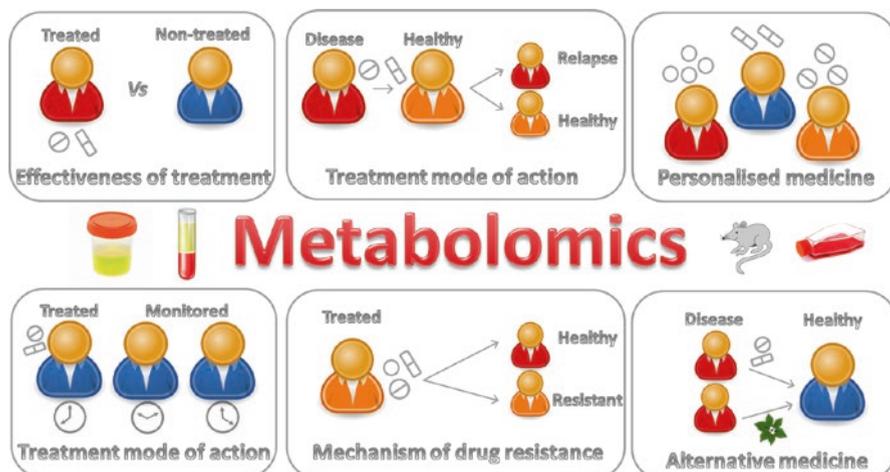
Fig. 9.1 Schematic of the applications and approaches of metabolomics used to study cancer. In each case, different samples from human biofluids to animal models or cell culture can be used

A number of metabolic enzymes are affected through the alteration of oncogenes (PTEN, RAS, ERK, etc.) and onco-transcription factors (p53, c-MYC, HIF, etc.), contributing to the drive in metabolic shifts observed in cancer. Cancer metabolism, the closest level related to the cancer phenotype, has therefore been a focus of research for decades (for reviews, see Armitage and Barbas, 2014 [3], and Boroughs and DeBerardinis, 2015 [4]). Interestingly, different alterations within the cell result in similar downstream metabolic effects, demonstrating the importance metabolism plays in the cancer cell. Typical responses include elevation of glycolytic flux where a high proportion of glucose is metabolised to lactate, even under oxygenated conditions; use of glucose and lactate carbon in the synthesis of nucleic acids, proteins and lipids; disruption or truncation of the TCA cycle; orchestration of an alternative supply of ATP (e.g. from glutamine or fatty acids); higher affinity for NADPH and glutathione production and increased tolerance to oxidative stress and reactive oxygen species-mediated apoptosis. An altered metabolism in cancer has been described as an increasingly acknowledged important aspect of the disease and a potentially fertile area for the identification of novel therapeutic targets [5].

Metabolomics, due to its ability to detect changes in numerous metabolites simultaneously and with no prior hypothesis on the region of metabolome required, is currently one of the fastest developing disciplines in cancer research. Some of the most notable approaches within cancer metabolomics are depicted in Fig. 9.1.

Over the last decade, over 2000 original research articles were published on cancer metabolomics (Web of Science). This involved the research of a range of cancers in different types of sample (cells, tissues, biofluids, etc.) to make discoveries on metabolic signatures or biomarkers. Such markers have potential use in improving sensitivity and selectivity in the detection, prognosis and diagnosis of cancer, in addition to revealing why treatments work (or don't work) and to propose new

potential drug targets in cancer and more. In the last 5 years, the number of articles and citations of original research in cancer metabolomics has grown dramatically, moving cancer metabolomics into a field of its own. In 2015 alone, almost 400 original research articles were published in the field of cancer metabolomics. Figure 9.2 shows the Web of Science citation report for 'cancer' & 'metabolomics' research articles over the last decade.

Once per year, the Metabolomics Society organises the largest international conference devoted to metabolomics and selects research to be showcased on significant advancements made in the field. After each event, highlights of the conference as decided by highly respected senior researchers in the field are published in the metabolomics journal. From 2015, a number of the featured highlights were specifically related to cancer metabolomics, now one of the largest subdisciplines in metabolomics [6]. One of the most significant of these was the reflection on research into the iKnife, a technology developed by Zoltan Takats at Imperial College London, UK. This highly innovative technique shows just how clinically relevant metabolomics is becoming in cancer research and is the perfect demonstration of how far metabolomics-based research has come in the transition from bench to bedside. The iKnife uses rapid evaporative ionisation mass spectrometry (MS) coupled to an electrosurgical knife allowing metabolomics-based tissue typing in real time and thus provides an alternative to conventional histological tests during cancer surgery [6]. Other highlights included the metabolomics and proteomics approach to identify metabolic enzymes differentially regulated in small cell lung cancer tissues, the elucidation of the metabolic effect of coculturing ovarian cancer cells with adipocytes, and how cancer cells deal with toxic-free ammonia generated during glutaminolysis (determined through isotope tracer experiments) [6]. In fact, these highlights cover some of the hottest topics in cancer metabolomics: the combination of different 'omics' approaches to reveal system properties of cancer and the linkage
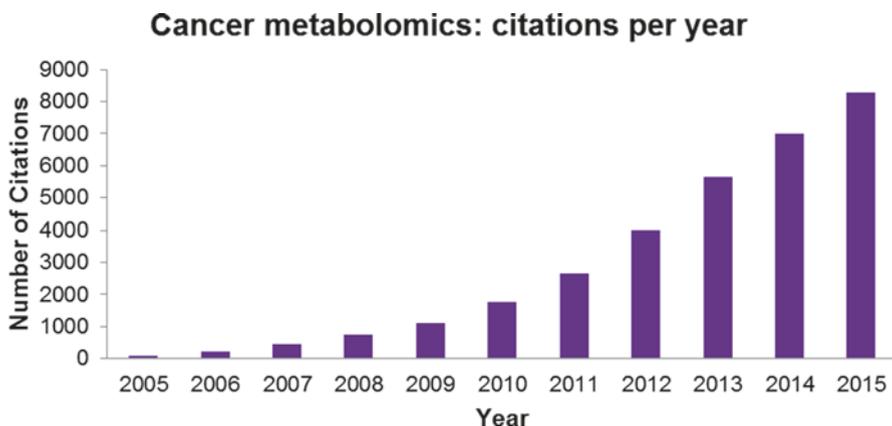
Fig. 9.2 The number of citations for research articles related to cancer metabolomics in each year over the last decade. Information obtained from the Web of Science, counts collated considering only original research articles.

of cancer with other metabolic lifestyle conditions such as obesity and isotope tracer experiments to better understand metabolic fluxes. All of this in addition to the previously mentioned advancement from bench to bedside making metabolomics a real clinical tool in cancer research sums up the topical applicability of cancer metabolomics.

In this chapter, techniques for the analysis of the cancer metabolome will be discussed in addition to exploring the role of metabolomics in the revelation of cancer biomarkers and in the assessment of cancer therapy. In line with the current approaches and elucidations from the field of cancer metabolomics, the final part of this chapter will cover one of the key subtopics of cancer metabolomics that is personalised medicine. In fact, personalised medicine is highly interesting for a number of diseases, but the advancement of knowledge gained through cancer metabolomics has been remarkable in the last few years.

9.2 Analysis of the Cancer Metabolome

Two analytical techniques currently dominate the global measurement of metabolites. These are nuclear magnetic resonance (NMR) and MS. NMR was the first analytical platform used for comprehensive measurement of metabolites present in a biological sample [7]. This technique is highly reproducible, selective and non-destructive and requires minimal sample preparation [8]. Furthermore, it is generally accepted as the 'gold standard' tool in metabolite structural elucidation [9]. Considering changes in energy metabolic pathways which are disturbed by developing cancer, NMR is a useful tool to measure metabolites involved in glycolysis (e.g. glucose, lactate, pyruvate) or the TCA cycle (e.g. cis-aconitate, citrate, succinate, pyruvate). To date, NMR-based metabolomics has already been successfully applied to search for biomarkers of several types of cancer including prostate [10], gastric [11], renal [12], cervical [13], oral [14], lung [15] and many others [16]. Moreover, several studies have shown utility of NMR metabolic fingerprinting for cancer risk prediction [17, 18], early diagnosis [19, 20] and staging [14, 15, 21]. Outside of biomarker discovery, this technique has also been found useful to study known and potential anticancer agents [22, 23], mechanisms of chemotherapy resistance [24] or effects of therapy [25]. The drawback of NMR is its relatively low sensitivity in comparison to MS [26]. MS-based metabolomics offers quantitative analyses with high selectivity and sensitivity and the potential to identify metabolites. In comparison to NMR, MS analysis usually requires a sample preparation step, which mostly consists of protein precipitation and solid- or liquid-phase extraction [27, 28]. For metabolomics, MS detection of metabolites is usually preceded by their separation to reduce the complexity of the acquired mass spectra, to provide isobar separation and to deliver additional information on the physico-chemical properties of the metabolites [29]. Separation techniques commonly used in metabolomics studies include liquid chromatography (LC), gas chromatography (GC) and capillary electrophoresis (CE) [30]. Sample treatment
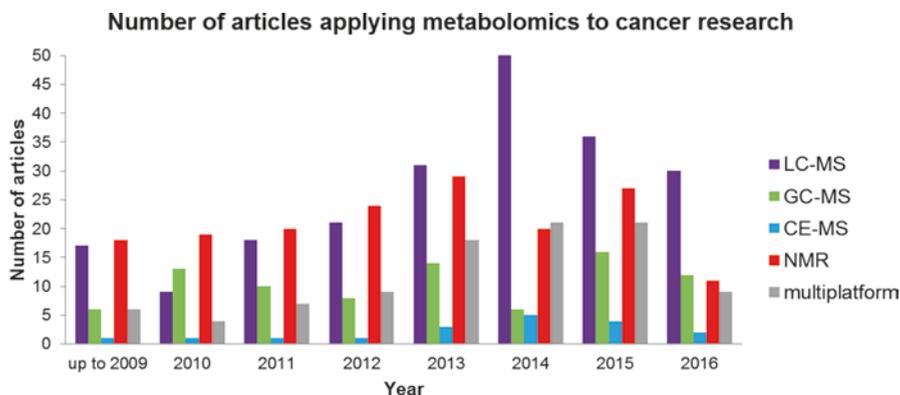
Fig. 9.3 Number of articles applying metabolomics in cancer research divided by different analytical platforms (or multiplatform studies) and year of publication. These data were obtained based on the search of terms "metabolomics + cancer" at PubMed page (http://www.ncbi.nlm.nih.gov/pubmed) performed on June 7, 2016. Out of 2083 records, reviews and studies just mentioning metabolomics were excluded.

procedures and consequently measured metabolites are dependent on the chosen separation technique. GC-MS is suitable for volatile and thermally stable analytes, while CE-MS is suitable for polar and charged molecules. LC-MS is the most versatile technique and with use of the appropriate columns – non-polar (reversed-phase chromatography) or polar (HILIC chromatography) – a huge array of metabolites can be detected [31]. All of these techniques have been used in combination with MS to study cancer metabolism and for discovery of potential biomarkers; however, the most often utilised are GC-MS and LC-MS (Fig. 9.3). LC-MS has been used to study lung [32], biliary tract [33], gastric [34], bladder [35] and other genitourinary cancers [36], as well as sarcoma [37], hepatocellular carcinoma [38], B-cell malignancies [39] and renal cell carcinoma [40]. LC-MS-based metabolomics has also been used to study treatment effects in acute lymphoblastic leukaemia [41] and prostate cancer [42] patients. While for LC-MS metabolomics sample preparation is rather simple and usually simultaneous protein precipitation and metabolite liquid extraction are performed [43], for GC-MS, a complex and time-consuming sample derivatisation procedure is necessary in order to make metabolites volatile [31]. GC-MS as a stand-alone technique has been already applied to study hepatocellular carcinoma [44] and gliomas [45], as well as lung [46], colorectal [47], bladder [48], breast [49], gastric [50], oesophageal [51] and thyroid [52] cancers. In comparison to GC-MS, CE-MS does not require any chemical derivatisation procedures. It has the potential for rapid analysis and efficient resolution of ionic metabolites including amino acids, organic acids, nucleotides and sugar phosphates, thus detecting numerous metabolites involved in central metabolic pathways. The sample preparation procedure for CE-MS metabolomics analysis is simple, rapid and common to many types of compounds; however, the migration time of each compound is often less

reproducible due to the difference of sample matrix and temperature in the environment [53]. Despite the above-mentioned advantages of CE-MS, it is not utilised as often in cancer metabolomics studies as other metabolomics platforms. Examples of its application have been shown for hepatocellular carcinoma [54], gastric cancer [55], lung and prostate tumour tissues [56], in addition to colon cancer cells [57]. Each analytical technique has its own drawbacks and advantages for metabolomics. NMR has lower sensitivity but has a huge potential in metabolite identification. MS is much more sensitive in comparison to NMR; however, depending on the chosen separation technique, different limitations appear. Due to the highly reproducible mass spectra of metabolites and availability of universal structural and mass spectral libraries for GC-MS, identification of metabolites is rather automated, though the derivatisation step can introduce artefacts and affect reproducibility [31]. In the case of LC-MS, identification is the most challenging step. The assignment of a measured *m/z* value to a real metabolite can be performed by the analysis of an authentic standard or comparison of obtained MS/MS spectra with one available in the literature or internet databases [58]. For many biological compounds, standards are still not available. Less-confident identification can be performed by interpretation of fragmentation spectra, but still a putative identification based on accurate mass is necessary. Otherwise, acquired MS/MS spectra can be useless [59]. Sometimes a charged molecule with different adducts may produce different fragmentation spectra, making identification more challenging [54]. Low reproducibility of migration times and the possibility to separate only charged molecules limit the utility of CE-MS in metabolomics. However, the most important limitation of any analytical platform used in metabolomics studies is the inability to measure all metabolites present in a biological sample. Actually, no single analytical technique is capable of measuring and identifying all metabolites; therefore, comprehensive metabolomics data needs to be assessed by bringing together data from different platforms [60]. Consequently, as it can be seen also in cancer research (Fig. 9.3), the number of studies in which several analytical platforms are applied together to achieve the scientific goal is increasing. Multiplatform approaches have also been used to study several types of cancer including lung [61], breast [62], colorectal [63] and prostate [64].

The method chosen to study the metabolome depends also on the approach that one wants to use. The following approaches are applied in metabolomics research: target analysis of metabolites; metabolic profiling, i.e. measurement of selected compounds which belong to one class or one metabolic pathway, and untargeted analysis, often termed metabolic fingerprinting, which aims to detect and semi-quantify all metabolites (if possible) present in a biological sample in order to define the unique metabolic pattern which characterises the biological system under particular conditions [65]. Metabolic fingerprinting has the greatest potential for discovery of novel findings. In cancer research, this approach has potential in biomarker discovery and interventional studies in order to evaluate the effect of the treatment or search for novel therapeutic targets [66]. Fingerprinting is not focused on particular metabolite(s); therefore it allows the discovery of novel metabolic pathways, which are disturbed by the disease or studied stimulus. Metabolic

fingerprinting can be performed by all above-mentioned analytical platforms. While NMR-based metabolic fingerprinting allows the detection of 20–50 metabolites [67], MS-based approaches (especially LC-MS) can detect 100 or 1000 of metabolites [68]. However, due to the technical limitations related to ion source cleaning, a sequence of fingerprinting analyses with MS detection can be performed on only a limited number of samples in one go [43]. Therefore, a discovery study with the use of an untargeted method should be supported by a target methodology (or metabolic profiling method) in order to validate the obtained results on a bigger cohort. Such a strategy was recently published for a cancer-related study. Piszcz et al. applied serum metabolic fingerprinting to find biomarkers for treatment indication in chronic lymphocytic leukaemia patients. In this research, a panel of biomarkers selected by LC-MS fingerprinting was succeeded by triple quadrupole MS (QqQ-MS) analysis. Validation by QqQ-MS allowed the proposal of a potentially highly specific and sensitive diagnostic approach composed of selected acylcarnitines and fatty acid amides [69]. A similar strategy was applied to study papillary thyroid carcinoma (PTC) tissue samples in order to find diagnostic markers and identify altered metabolites [70]. In this study, a GC-MS-based discovery phase was followed by an LC-QqQ-MS- and GC-TOF-MS-based validation. Targeted metabolomics proved that galactinol, melibiose and melatonin were differentially expressed between PTC and healthy tissues. Galactose metabolism was found to be an important factor influencing PTC development by affecting energy metabolism. Alpha-galactosidase was proposed as a potential target for PTC therapy [70]. An interesting study was performed for lung cancer using multiplatform metabolomics in order to select serum metabolites for improved staging of non-small cell lung cancer (NSCLC) patients. Out of 29 metabolites exhibiting a significant trend in levels away from normal individuals to early- and late-stage patients, bilirubin and λ-glutamylalanine (the most significant) were selected for LC-MS/MS validation. Bilirubin emerged as a metabolite that consistently showed a statistically significant trend with increasing NSCLC stage [71]. Amongst others, a validation step has also been included in studies on colorectal [63], ovarian [72] and oesophageal cancer [73].

Classical measurement of metabolite levels alone provides only a very static view on metabolism. For a fuller understanding of metabolism, the underlying metabolic fluxes are much more important and informative because they provide a much closer functional link to an observed phenotype [74]. Flux-based metabolomics (fluxomics) was first proposed around 20 years ago and is based on utilisation of stable, ^{13}C isotope-labelled substrates used for accurately tracking changes in the distribution of metabolites in biochemical pathways [75]. ^{13}C-labelled glucose is often used in fluxomics studies [76]; however, it is also possible to use other labelled compounds as substrates. Application of flux analysis in the context of cancer research has revealed metabolic alterations of several metabolic pathways in a wide variety of tumours. GC-MS-based fluxomics experiments are often performed in order to trace central carbon metabolism. This approach was used to study glutamine dynamics in pancreatic ductal adenocarcinoma [77], as well as glutamine-associated changes in glioma cells during impaired mitochondrial pyruvate transport [78].

The approach has also been used to reveal that the reductive metabolism of α-keto-glutarate contributes to de novo lipogenesis [76]. LC-MS has been applied to study the metabolic alterations associated with the M2 isoform of pyruvate kinase, show-ing significant differences in glycolytic intermediates. It was revealed that these glycolytic metabolites feed into serine synthesis, allowing them to proliferate in serine-depleted medium [79].

The other approach for the analysis of tumour samples is matrix-assisted laser desorption ionisation mass spectrometry (MALDI-MS) imaging. Although this ana-lytical platform is rather used for the analysis of larger molecules such as proteins [80], with some modifications in the sample preparation and operating approaches, it can be used to measure small molecules including lipids, carbohydrates, hor-mones, nucleotides/nucleosides and drugs and drug metabolites [81]. With this technology, it is possible to measure the distribution of diverse molecular species in a tissue section without destroying the tissue or the use of target-specific molecular labelling reagents [82]. The most popular application of MALDI-MS imaging in oncology is drug distribution analysis [83, 84]. By the use of MALDI-MS, the detection of an orally administered drug compound in mouse tumour tissue surface has been demonstrated [85]. High-resolution MS imaging significantly improved the localisation of drug metabolites; therefore, this technique has been found to be important to study pharmacokinetics and pharmacodynamics of drugs. Regarding anticancer drug studies on localisation of tamoxifen in human breast cancer tumours [84], the distribution of alectinib in murine brain [82] or localisation of sunitinib and its metabolite tumour-bearing mice [83] by use of MALDI-MS can be mentioned. Outside of pharmacological studies, MALDI-MS has also been applied to measure endogenous metabolites. With the application of 9-aminoacridine (9-AA) as the matrix, the detection of low-mass metabolites and lipids directly from cancer tissues has been demonstrated. Applications include lactate and pyruvate for studying the Warburg effect, as well as succinate and fumarate, metabolites whose accumulation is associated with specific syndromes. It has been possible to use this approach to identify regions within tumour tissue samples with distinct metabolic signatures that are consistent with known tumour biology [86]. Also, 9-AA-supported MALDI-MS imaging has also been used to compare metabolic profiles between control livers and those bearing metastatic foci of human colon cancer. Differences in nucleotides, lipids and several amino-sugars were observed between the tissues studied. The metastatic human colon cancer xenografts displayed remarkable accu-mulation of UDP-*N*-acetyl hexosamines and glutathione in vivo [87]. MS and mag-netic resonance imaging have also been used to analyse total choline (tCho) and phosphocholine (PC) in metastatic breast tumour model. MR imaging showed that high tCho levels, consisting of free choline, PC and glycerophosphocholine (GPC) displayed a heterogeneous spatial distribution in the tumour. MS imaging performed on tumour sections has detected the spatial distributions of individual PC, Cho and GPC. PC and Cho intensities were increased in viable compared with necrotic regions of MDA-MB-231 tumours but were relatively homogeneously distributed in MCF-7 tumours [88]. Wang et al. showed that *N*-(1-naphthyl)ethylenediamine dihy-drochloride (NEDC) could act as a matrix for MALDI imaging yielding many more

endogenous compounds than 9-AA. NEDC-assisted imaging was applied to study colorectal cancer liver metastasis, showing that NEDC is especially well suited for examining distributions of glycerophospholipids and low-molecular-weight metabolites below *m/z* 400 [89].

Metabolomics as a field encompasses not only the high-throughput measurement of a number of metabolic variables simultaneously – a large part of the metabolomics discipline involves computational approaches to study metabolism. One very interesting example of this with a particular reference to cancer research was published in 2015 and concerned the chemometric fusion of metabolic profiling data from plasma samples with auxiliary patient lifestyle information. Through this method, authors created a biocontour, defined as a complex pattern of relevant biological and phenotypic information that provided a forecast, described as being on par with how well most current biomarkers can diagnose current cancer [17]. It was shown that this method offers sensitivity and specificity well above 80 % as compared with mammography, which has a diagnosis capacity associated with around 75 % sensitivity and specificity. This retrospective study compared two sets of data in order to develop a suitable model for the forecast of breast cancer.

9.3 Cancer Metabolomics in the Discovery of Biomarkers

In 2014, Armitage and Barbas reviewed the current trend and future perspectives of the cancer biomarker discovery field, collating research on potential biomarkers in different cancers revealed by employing a range of different analytical platforms [3]. Since then, research has continued to grow globally to find new potential biomarkers. Figure 9.4 shows the cancers that are most commonly studied by metabolomics, highlighting for each cancer the most cited article to date in each field.

Aiming to improve prognosis/diagnosis of the disease is one of the main applications of metabolomics in cancer research. This usually involves the discovery of biomarkers and the assessment of their sensitivity and selectivity powers in disease prognosis/diagnosis. Metabolomics has become increasingly popular in postulating potential biomarkers, typically by the comparison of control subjects and cancer subjects or cancer subjects before and after initiation of disease or new stage of disease (to study progression). The latter is usually performed by comparing samples from patients with the disease for which there are retrospective samples available from before the disease presented or from an earlier stage of the disease. Similarly, biomarkers can also be revealed to show metabolic features of risk of recurrence or relapse in patients by comparing pretreatment samples to samples collected at later time points with knowledge of whether or not cancer recurred in those patients. Finally, cancer biomarkers can be revealed by observing the effects of mutations or knockouts that theoretically reduce or remove the chance of metastasis by comparison to wild-type cancer controls.

The comparison of control and cancer subjects can include the study of cells, adjacent tissues or patients. For example, in a comparison of cancer cells to controls,
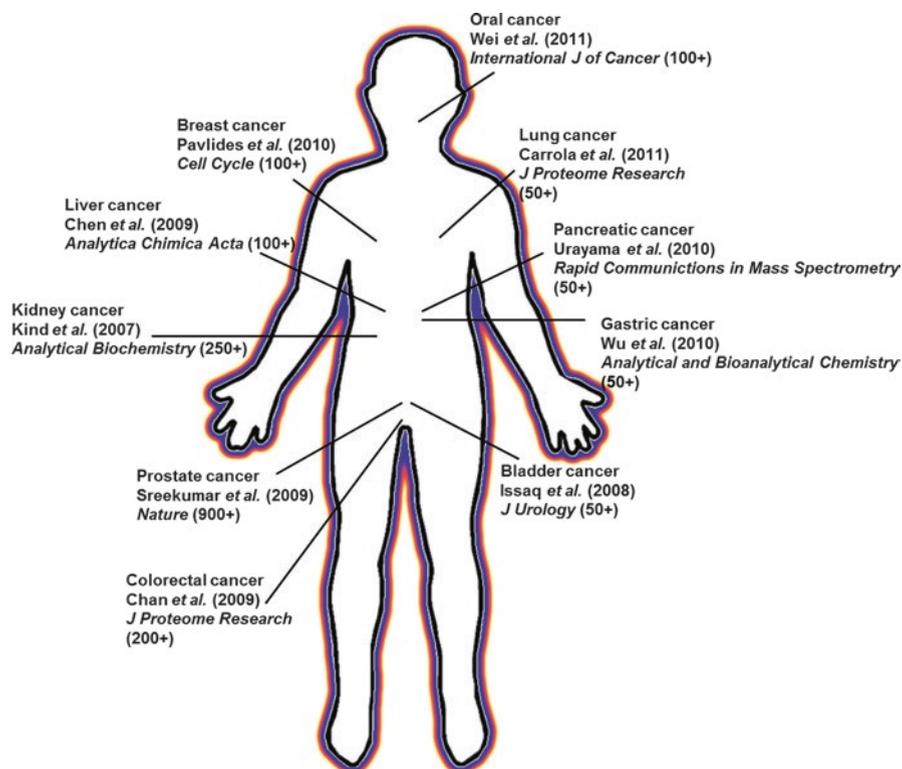
Fig. 9.4 The most common cancers studied by metabolomics and the highest cited paper published to date (June 2016). Approximate number of citations are displayed.

volatile metabolic signatures in human breast cancer cell lines versus normal human mammary cells have been unveiled [90]. Novel analyses of volatile compounds in the headspace of conditioned culture medium were directly fingerprinted by secondary electrospray ionisation (ESI)-MS. Samples could be classified based on a characteristic odour released by cancer cells whose constituents could be used as disease markers. Similarly, comparing adjacent tissues can yield candidate biomarkers of cancer. For example, 13 tumours and seven normal tissue markers were identified that separated cancer from normal tissues with >80 % sensitivity and specificity, mainly due to two metabolite classifiers: cytidine-5-monophosphate and pentadecanoic acid. The ratio between these was the most significant discriminator between cancer and normal tissues and allowed detection of cancer with sensitivity as well as specificity around 94 %.

In recent years, metabolomics has been performed more readily on the comparison of samples from clinical patients. For example, using ESI-MS/MS with no prior separation technique, serum profiling of cancer patients and control subjects has revealed potential biomarkers of colorectal cancer. Quantitative profiles of the concentrations of 26 amino acids elucidated 11 to be significantly different between

cancer and control subjects [91]. All except one had been previously reported as potential colorectal cancer biomarkers in previous literature covering a range of different analytical platforms. In a recently published article on the detection of early-stage ovarian cancer, LC-MS was employed to interrogate the serum metabolome of patients and age-matched controls in order to establish a linear support vector machine model of 16 diagnostic metabolites [92]. This panel of lipid-related identifiers was successful at a rate of 100 % accuracy for the patient cohort implicating the importance of lipid and fatty acid metabolism in ovarian cancer.

The study of disease initiation, progression, recurrence or relapse using metabolomics usually applies to the analysis of plasma or serum from patients. A key example of this was published in Cancer Research in 2010 that focussed on the development of a metabolic profiling test using GC×GC-MS and NMR to monitor recurrent breast cancer [93]. Metabolite profiles of 257 retrospective serial serum samples from 56 previously diagnosed and surgically treated breast cancer patients were analysed, whereby 116 were derived from 20 patients with recurrent breast cancer and 141 samples from 36 patients with no clinical evidence of the disease during ∼6 years of sample collection. Eleven metabolic markers – formate, histidine, proline, choline, tyrosine, 3-hydroxybutyrate, lactate, glutamic acid, N-acetyl-glycine, 3-hydroxy-2-methyl-butanoic acid and nonanedioic acid – were shortlisted using logistic regression that provided around 85 % specificity, as well as sensitivity leading to 55 % accuracy rate of prediction of recurrence more than 12 months before recurrence could be clinically diagnosed through alternative methods.

In another example, Lodi et al. published interesting research on the NMR-based metabolomics analysis of archived serial paired serum and urine samples from myeloma patients at different stages of disease to revel metabolic patterns correlated with progression that identified markers useful to individual patients [94]. In this way, patients were distinguished based on relapse or remission, and this difference in metabolic profile is thought to be applicable to provide prognoses in future cases.

In the application to distinguish between early and metastatic breast cancer patients and as a tool to potentially predict disease relapse, metabolomics has been applied to study serum metabolomes of women with metastatic and predominantly oestrogen receptor-negative early-stage breast cancer using high-resolution NMR spectroscopy [95]. In this study, a model of prognosis was created using sophisticated statistical approaches that showed that it was possible to distinguish between early and metastatic disease in the metabolome at a rate of 83.7 % with 90 % sensitivity, 67 % specificity and 73 % predictive accuracy. Results were subsequently reproduced in an independent sample with 82 % sensitivity, 72 % specificity and 75 % predictive accuracy, and a further confirmation study is said to be underway. From these studies, it was shown that the relapse of this disease is associated to lower levels of histidine and increase levels of glucose and lipids.

In 2013, Alberice et al. published a research on the recurrence of bladder cancer [96]. Pretreatment urine samples were collected from 48 patients diagnosed with urothelial bladder cancer and analysed by untargeted LC-MS- and CE-MS-based metabolomics. After the analysis, patients were followed up through hospital pathological

charts to identify whether and when the disease recurred or progressed. Patients were classified based on their proposed risk (depending on original tumour grade, stage) and their final outcome (recurrence of cancer or not), and data were analysed considering groups based on these factors. Authors reported a total of 27 metabolites that significantly distinguished patients based on these groups, some of which had been previously related to bladder cancer but not to the progression or recurrence of the disease. It was suggested that these candidate biomarkers could be useful in the diagnosis of bladder cancer and also in prognostics by characterising the stage of the disease. Moreover, based on pretreatment profiling, this information could be useful in predicting outcome of treatment or to decide which course of treatment could be most appropriate to reduce the risk of recurrence.

Zhu et al. have recently published a proof of concept study utilising LC-MS/MS applied to monitor disease progression in colorectal carcinoma patients [97]. This involved sequential metabolite ratio analysis of serially acquired samples from 20 patients in order to identify the metabolites that correlate to the status of the disease. This novel approach reduces patient-to-patient variability and allows the generation of a prognostic model through a defined panel of biomarkers. Although this was a relatively small study, it served as a sound proof of principal for further investigation and was strengthened by the reduction of sample variability that usually makes a huge demand on the number of patients required in a metabolomics study of human subjects.

Heterogeneity in the clinical course of patients is rife in many cancers. Chronic lymphocytic leukaemia (CLL) is one such cancer in which patients require different action: some patients present an aggressive form of the disease that requires immediate therapy, while others remain without treatment for years. In recent years, research has been directed into finding new prognostic markers to distinguish between stable and progressive forms of the disease and subsequently to forecast patient survival and select the appropriate course for treatment. A metabolomics study to discriminate serum samples from patients with indolent or aggressive forms of CLL, highly specific and sensitive diagnostic markers were postulated as discussed in Sect. 9.2 [69]. Ten markers of discrimination from an initial study were validated in a new set of patients and forming a panel of acylcarnitines and fatty acid amides; it was possible to find the best method of distinction between indolent and aggressive CLL.

Analysing the effects of a knockout or mutation of an onco-feature (gene, transcription factor, protein) to observe how this feature promotes survival of cancer cells and therefore how cancer is avoided by this knockout or mutation or how cancer cells overcome this removal in order to continue surviving can be very specific but highly topical. This is a particularly interesting type of approach if the onco-feature is suspected to be a reasonable target for future therapy. One example of this was published in 2015, whereby the effect of hypoxia inducible factor (HIF), a transcription factor promoting survival of tumour cells in low-oxygen environments, on cellular metabolism, was explored [98]. Wild-type cells were compared to HIF-deficient cells at different oxygen tensions with the aim of highlighting the potential role of this factor under low oxygen in addition to alternative mechanisms

of survival employed in the absence of HIF function. GC-MS was used to compare the metabolic profiles of extracts from 30 samples of cells with and without HIF exposed to normoxia (21 % oxygen), hypoxia (1 % oxygen) and anaoxia (0 % oxygen). The study revealed different metabolic hubs (highly connected metabolites in the pathways as revealed in the metabolic profiles) when HIF was functional or not. Cells were observed to be strongly reliant on 4-hydroxyproline to survive in hypoxic conditions, while fructose took over this role in HIF-deficient cells. This showed both the primary function of HIF but also highlighted why HIF may not be the best single target for therapy since cells are able to survive through alternative mechanisms centred on fructose in the absence of HIF.

9.4 Cancer Metabolomics in the Assessment of Therapy

Metabolomics is increasingly employed in the determination of the effectiveness of disease therapy as well as in the exploration of pharmaceutical mechanism of action (for a recent review, see Wishart 2016 [99]). In many cases, including different cancers, these studies have led to, or have the potential to lead to, the discovery of new therapeutic targets. Moreover, such studies have also been useful for looking into the ineffectiveness of treatments and to propose their improvement.

Metformin has been a particularly well-studied cancer therapy by metabolomics. Neoadjuvant studies of potential metformin biomarkers and its effect in breast cancer patients through metabolomics-based approaches have recently been reviewed [100]. In 2014, the experiment of a preoperative window study of metformin in endometrial cancer was published, evaluating the anti-proliferative, molecular and metabolic effects of the drug [101]. In this study, 20 obese women with endometrioid endometrial cancer were treated with 850 mg metformin daily for up to 4 weeks prior to surgical staging. With particular emphasis on the metabolomics assay, pre- and post-treatment serum samples and matched tumour samples collected post-treatment were analysed by LC-MS global profiling. Responders to the treatment were found to be sensitive to metformin's effects on induction of lipolysis that correlated with increased fatty acid oxidation and glycogen metabolism in matched tumours relative to the nonresponders. Metformin induced a pronounced shift in lipid and glycogen metabolism in the serum and tumours of responders, and thus it was suggested from this study that metformin could be a viable treatment for endometrial cancer, at least in obese individuals.

Metformin has been studied by cancer metabolomics not only to investigate its effectiveness or ineffectiveness but also to better understand its mechanism of action. In a recent study published by He et al. (2015), metformin was studied through a combined metabolomics and transcriptomics approach to study its effect over time (8, 24 and 38 h following exposure) in human-derived LoVo cells treated with the drug [102]. More than 40 differential metabolites in carbohydrate, lipid, amino acid, vitamin and nucleotide metabolic pathways between controls and treated cells at each time point were identified showing net global upregulation at

8 h followed by downregulation by 24 h. The transcriptome revealed 100–1000 of differentially expressed genes with exposure to metformin involving cancer signalling and cell energy metabolism mechanisms. It was concluded from this study that metformin supresses the proliferation of LoVo cells, likely through the modulation of cell energy metabolism at both transcriptomic and metabolomic levels.

Another frequently used drug in cancer is docetaxel, used in the treatment of breast cancer. However, challenges of resistance or incomplete response to this drug make its use under-optimal. Using a magnetic resonance spectroscopy-based metabolomics approach, potential biomarkers of docetaxel resistance in a mouse model for BRCA1-mutated breast cancer have been identified [103]. In this study, significant metabolic differences between sensitive and resistant tissue samples were revealed elucidating mechanisms of resistance to the compound. Choline metabolites were identified as markers of resistance in that their concentrations were generally stably higher in docetaxel resistant than in sensitive BRCA1-mutated mouse mammary tumours, but that the first days after docetaxel treatment, choline metabolites were increased only in the sensitive tumours. From this, it was concluded that both pre- and post-treatment tissue levels of choline compounds have the potential to predict response to docetaxel treatment.

In a study of ovarian cancer, LC-MS- and GC-MS-based metabolomics have recently been applied to study the mechanisms of platinum sensitivity or resistance in cancer cell lines [104]. Almost 180 metabolites were identified as being significantly different as a function of sensitivity/resistance including 70 increases and 109 decreases in platinum-resistant cells. The most altered pathway was determined to be cysteine and methionine metabolism, specifically the methionine degradation super-pathway and cysteine biosynthesis were the top two canonical pathways implicated by Ingenuity Pathway Analysis.

In another study of resistance, untargeted metabolomics has been employed to study the therapeutically relevant danorubicin in P-glycoprotein overexpressing T-cell acute lymphoblastic leukaemia (ALL) cells [105]. In this case, resistant ALL cells were found to exhibit a 'rewired' central metabolism with reduced dependence on glutamine combined with a higher demand for glucose and an altered rate of fatty acid β-oxidation accompanied by a decreased capacity for pantothenic acid uptake. Untargeted metabolomics analysis utilising LC-MS was performed for comparative analysis of sensitive and resistant cells. Findings were validated by selectively targeting components of the metabolic switch away from fatty acid β-oxidation with low pantothenic acid availability, using approved drugs and starvation approaches followed by cell viability analyses. These were also later tested in an acute myeloid leukaemia sensitive/resistant cell line pair.

Immunotherapy is another 'hot topic' in cancer therapy for which cancer metabolomics has been useful in discovering its capabilities as well as its flaws. The past 5 years have been described to see 'something of a revolution in immunologically targeted approaches to treatment' and 'the area in which there has been the most promising evidence of effective translation' [106]. In a recent study by Wettersten et al., metabolic alterations occurring in renal cell carcinoma have been explored, revealing the likelihood of ineffective treatment by means of immunotherapies such

as IFNγ [107]. More than 200 metabolic profiles were analysed from primary renal tumours and normal renal tissue, leading to key discoveries: (i) increased utilisation of aerobic glycolysis at the expense of oxidative metabolism via the TCA cycle in higher-grade tumours; (ii) glutamine is predominantly metabolised to generate glutathione to attenuate oxidative stress rather than being metabolised to preferentially fuel fatty acid synthesis; and (iii) tryptophan metabolism is upregulated by the disease, causing a subsequent increase in kynurenine, an immunosuppressive metabolite. The latter explains the likely ineffectiveness of immunotherapies but points towards a therapeutic target in indoleamine 2,3-dioxygenase as the enzyme that catalyses the metabolism of tryptophan to *N*-formylkynurenine. One such inhibitor, 1-D-MT, is currently in clinical trial (NCT00567931), while inhibitors of their other findings are also currently under investigation [108].

Another topic of recent interest has been in the effectiveness of commonly used drugs in the prevention of cancer. For example, in 2016, Liesenfeld et al. showed that aspirin reduces the concentration of onco-metabolite 2-hydroxyglutarate [109]. This wasn't an investigation of the effectiveness of treatment in cancer patients, rather, the effect of a commonly used over-the-counter drug was tested in healthy individuals as part of a randomised, double-blind crossover trial was experimented using a metabolomics approach to show aspirin's potential in cancer prevention.

Cancer metabolomics has been useful for advancing research not only in pharmaceutical therapies but also in nutraceuticals. In general, the utility of nutraceutical compounds to combat disease is becoming more and more favourable, and metabolomics has played a significant role in the elucidation of the value of such compounds, particularly in cancer therapy. Table 9.1 summarises some of these notable advancements.

9.5   Personalised Medicine and the Future for Cancer Metabolomics

Metabolomics has been described as the 'fast line for personalised medicine' [117]. Individuality in drug metabolism has been recognised as an important factor in prescription of cancer therapies [118]. Individuality in metabolism can lead to drastic differences between cancer patients with respect to the efficiency or inefficiency of currently available treatments. Key examples are tamoxifen and its efficiency as determined by patients exhibiting different polymorphisms of CYP2D6 that metabolises tamoxifen to biologically active metabolites [119] and irinotecan and its metabolism via glucuronosyl transferase encoded by UGT1A1 that can differ from patient to patient.

Personalised medicine stems from the theory and utility of stratified medicine. The latter involves subdividing patients with a particular disease into groups based on the likely mechanisms of the disease and consequently the likelihood of effectiveness of treatment. A key example of this in cancer is the use of stratification for patients with chronic myeloid or acute lymphoblastic leukaemia based on whether

Table 9.1 The study of nutraceuticals as potential therapies in cancer through metabolomics

Nutraceutical	Sample	Findings	Reference
Silymarin extracted from *Silybum marianum*	Human-derived liver and T-cell cultures	Nontoxic doses induce cytoprotection via suppression of metabolism, activation of stress pathways and downregulation of inflammatory signalling (observed in both models)	Lovelace et al. (2015) [110]
Oil extracted from *Saussurea lappa,* costunolide and dehydrocostus lactone isolated from it	Serum and urine from MCF-7 breast carcinoma xenograft mice	Metabolic transformation by elevated glycolysis and steroid hormone metabolism; reduction in fatty acid metabolism	Peng et al. (2015) [111]
Supercritical fluid extract from olive leaf ('El Hor' variety)	JIMT-1 human breast cancer cells	Diosmetin is the most abundant flavone – sixfold higher concentration than apigenin and 170-fold more than luteolin in treated cells. Diosmetin concluded to be main driver of anti-proliferative effects. Analysis of these flavones in isolation showed none as potent as extract itself; therefore, synergistic action within extract is required for maximal effect	Barrajón-Catalán et al. (2015) [112]
Halofuginone extracted from *Dichroa febrifuga*	In vitro and in vivo models of colorectal cancer (various cell lines)	Treatment caused Akt/mTORC1-mediated aerobic glycolysis downregulation, reduction of glycolysis, GLUT 1 activity, TCA cycle intermediates, phosphatidylcholine, ceramide, sphingomyelin, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, phosphatidic acid and expression of fatty acid synthase	Chen et al. (2015) [113]
Flexibilide isolated from *Sinularia flexibilis*	HCT-116 human colorectal carcinoma cells	19 distinct metabolites involved in sphingolipid, alanine, aspartate and glutamate, D-glutamine and glutamate, glycerophospholipid and pyrimidine metabolism involved in mechanism of action. Likely effects attributed to treatment causing cell membrane lesions, apoptosis, downregulation of TCA cycle, and decreased mitochondrial transmembrane potential.	Gao et al. (2016) [114]
Vitamin C	Human colorectal carcinoma cells K-RAS and BRAF mutations (various cell lines)	Mutant cells accumulate glycolytic intermediates upstream of glyceraldehyde 3-phosphate dehydrogenase and deplete those downstream implying GAPDH inhibition. Increase in PPP metabolites implicates shift in glycolytic flux towards oxidative PPP as a response mechanism.	Yun et al. (2015) [115]

(continued)

Table 9.1 (continued)

Nutraceutical	Sample	Findings	Reference
Nutmeg	Serum from mice harbouring adenomatous polyposis coli gene mutation, another important inducer of colon cancer	Reduced levels of uremic toxins cresol sulphate, cresol glucuronide, indoxyl sulphate and phenyl sulphate	Li et al. (2015) [116]

or not they carry the Philadelphia chromosome: carriers may be successfully treated with tyrosine kinase inhibitors such as imatinib while noncarriers may not [120].

Nowadays there are many examples of studies using metabolomics in the field of personalised medicine for cancer. For example, in 2014, Navarette et al. published research on the success of mitomycin C treatment for pancreatic cancer in a personalised treatment designed for a patient resistant to other treatments [121]. The motive for this study was that mitomycin C proved more successful than rapamycin or even the combination of the two. Mitomycin C was selected as the single treatment for this patient based on trials in a murine xenograft tumour model encompassing pancreatic adenocarcinoma cells extracted from the patient. However, it was not until this metabolomics study that the mechanism of its success was explored. Untargeted metabolomics utilising GC-MS and LC-MS revealed mitomycin C's effect on the TCA cycle, purine metabolism and fatty acid biosynthesis, in addition to many significant lipid and amino acid alterations that apparently lead to its success over other treatments.

It is clear that personalised medicine offers an interesting avenue for the future in cancer therapy. Whether or not its utility as a widespread approach in medicine will be feasible from the economical point of view remains to be demonstrated [106]. The requirement of funding at different stages of developing a personalised medicine is huge, but on the other hand, if all patients were to receive the most effective treatment for their case at the outset, there is potential for a net saving considering the lack of necessity to fund subsequent healthcare costs and socio-economic costs from loss of earnings for individuals, etc. In any way, it is clear that cancer metabolomics has played and will continue to play a vital role in the discovery of new approaches in personalised or stratified medicine.

Metabolomics has scope as a tool in the personalised medicine pipeline not only to reveal how or why a treatment has been successful in specific cases. It offers a highly sensitive and selective approach in prognosis that could lead to determination of a treatment strategy either by stratification or on a case-by-case basis.

Prognosis and development of a plan of action for treatment in cancer is generally performed by 'overall survival'. For example, overall survival in colorectal cancer is usually predicted based on the presence or absence of K-RAS mutations,

blood cell counts, serum levels of proteins such as lactate dehydrogenase and metabolites such as bilirubin [16]. However, it is already known that these markers in addition to the Eastern Cooperative Oncology Group (ECOG) performance status may not always be reliable in predicting overall survival. For example, from the study of serum samples from 153 metastatic colorectal patients and 139 healthy controls, Bertini et al. showed that cancer patients presented metabolic alterations interpreted as perturbations in energy metabolism, in addition to an even more pronounced inflammatory response within their metabolome that proved to be statistically more relevant than either ECOG or K-RAS mutations [122].

In an example of exploring personalised response to medicine, Tenori et al. published research whereby pretreatment and serial on-treatment serum samples of 579 women with metastatic breast cancer randomised to paclitaxel plus either a targeted anti-HER2 treatment (lapatinib) or placebo were analysed by NMR [123]. Metabolic profiles were compared with time to progression, overall survival and treatment toxicity to reveal that a subgroup of patients with HER2-positive disease treated with paclitaxel plus lapatinib, metabolic profiles from patients in the upper and lower thirds of the dataset showed significant differences for time to progression and overall survival, indicating the applicability of metabolomics in sub-selecting patients with HER2-positive disease with greater sensitivity to paclitaxel plus lapatinib [123].

Rather than testing for one or a few specific biomarkers in isolation, a patient's metabolic profile could offer a greater whelm of information that could guide a better treatment strategy, beyond the point of diagnosis. Since tumours rarely constitute more than 1 % of total body weight, it is highly unlikely anyway that changes observed in body fluids such as serum, plasma or urine are due to cancer itself [16]. However, the changes observed can reveal information about the host response to the disease and based on that infer how a patient may respond to treatment. Moreover, the gut microbiome is becoming of greater importance in understanding diseases like cancer and predicting responses of patients to treatment. Key future research applications for metabolomics and lipidomics will likely be to investigate the role of gut microbiota in cancer and to better understand how metabolic therapies can be tailored using a personalised medicine approach. Understanding gut microbiota in cancer is particularly important given that this can alter the metabolic response to drug therapies [124] and also the efficacy anticancer treatment [125].

9.6   Conclusions

Over the past decades, metabolomics has grown exponentially and so has the number of articles related to metabolomics. During this time, cancer metabolomics has become a field of its own. In this chapter, cancer metabolomics from its history to the latest developments have been discussed. Significant advancements have been made in biomarker discovery for prognosis, diagnosis and assessment of treatment in cancer. This is owed to the continually enhanced methodologies and applications in metabolomics and selection of the appropriate experimental design.

Experimental design is highly critical for the success and impact of any metabolomics study, but it is particularly relevant in disease exploration. The choice of samples, analytical platforms, structure of experiment and nature of study strongly influence the outcome, and therefore only results from the most robust experimental designs should be scrutinised and considered in the implication of new prognostics/diagnostics or therapeutic strategies. As discussed, a range of different approaches has been applied to date in cancer metabolomics, and new topics are trending. For example, as the link between diet and lifestyle with diseases like cancer becomes more apparent, not to mention the interaction between gut microbiota and host in the likelihood of disease or success of treatment, metabolomics analyses of food (foodomics) and the microbiome are likely to increase. This will open an avenue of new information and opportunities in cancer research.

References

1. Warburg O. Injuring of respiration the origin of cancer cells. Science. 1956;123(3191):309–14.
2. Diaz-Ruiz R, Uribe-Carvajal S, Devin A, Rigoulet M. Tumor cell energy metabolism and its common features with yeast metabolism. Biochim Biophys Acta. 2009;1796(2):252–65.
3. Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. J Pharm Biomed Anal. 2014;87:1–11.
4. Boroughs LK, DeBerardinis RJ. Metabolic pathways promoting cancer cell survival and growth. Nat Cell Biol. 2015;17(4):351–9.
5. Dhakshinamoorthy S, Dinh N-T, Skolnick J, Styczynski MP. Metabolomics identifies the intersection of phosphoethanolamine with menaquinone-triggered apoptosis in an in vitro model of leukemia. Mol Bio Syst. 2015;11(9):2406–16.
6. Fiehn O, Putri SP, Saito K, Salek RM, Creek DJ. Metabolomics continues to expand: highlights from the 2015 metabolomics conference. Metabolomics. 2015;11(5):1036–40.
7. Nicholson JK, Lindon JC, Holmes E. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181–9.
8. Brennan L. NMR-based metabolomics: from sample preparation to applications in nutrition research. Prog Nucl Magn Reson Spectrosc. 2014;83:42–9.
9. Johnson SR, Lange BM. Open-access metabolomics databases for natural product research: present capabilities and future potential. Front Bioeng Biotechnol. 2015;3:22.
10. Kumar D, Gupta A, Mandhani A, Sankhwar SN. NMR spectroscopy of filtered serum of prostate cancer: a new frontier in metabolomics. Prostate. 2016;76:1106–19.
11. Hu JD, Tang HQ, Zhang Q, Fan J, Hong J, Gu JZ, et al. Prediction of gastric cancer metastasis through urinary metabolomic investigation using GC/MS. World J Gastroenterol. 2011;17(6):727–34.
12. Gil AM. NMR metabolomics of renal cancer: an overview NMR metabolomics of renal cancer: an overview. Bioanalysis. 2015;7(18):2361–74.
13. Ye N, Liu C, Shi P. Metabolomics analysis of cervical cancer, cervical intraepithelial neoplasia and chronic cervicitis by 1H NMR spectroscopy. Eur J Gynaecol Oncol. 2015;36(2):174–80.

14. Gupta A, Gupta S, Mahdi AA. 1H NMR-derived serum metabolomics of leukoplakia and squamous cell carcinoma. Clin Chim Acta. 2015;441:47–55.

15. Deja S, Porebska I, Kowal A, Zabek A, Barg W, Pawelczyk K, et al. Metabolomics provide new insights on lung cancer staging and discrimination from chronic obstructive pulmonary disease. J Pharm Biomed Anal. 2014;100:369–80.

16. Palmnas MSA, Vogel HJ. The future of NMR metabolomics in cancer therapy: towards personalizing treatment and developing targeted drugs? Metabolites. 2013;3(2):373–96.

17. Bro R, Kamstrup-Nielsen MH, Engelsen SB, Savorani F, Rasmussen MA, Hansen L, et al. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. Metabolomics. 2015;11(5):1376–80. Springer US.

18. Jiménez B, Mirnezami R, Kinross J, Cloarec O, Keun HC, Holmes E, et al. 1H HR-MAS NMR spectroscopy of tumor-induced local metabolic "field-effects" enables colorectal cancer staging and prognostication. J Proteome Res. 2013;12(2):959–68.

19. Lin Y, Ma C, Liu C, Wang Z, Yang J. NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. Oncotarget. 2016. doi:10.18632/oncotarget.8762.

20. Tiziani S, Lopes V, Günther UL. Early stage diagnosis of oral cancer using 1H NMR-based metabolomics. Neoplasia. 2009;11(3):269–76.

21. Puchades-Carrasco L, Jantus-lewintre E, Pérez-Rambla C, García-García F, Lucas R, Calabuig S, et al. Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer. Oncotarget. 2016;7(11):12904–16.

22. Lefort N, Brown A, Lloyd V, Ouellette R, Touaibia M, Culf AS, et al. 1H NMR metabolomics analysis of the effect of dichloroacetate and allopurinol on breast cancers. J Pharm Biomed Anal. 2014;93:77–85.

23. Liu S, Wang W, Zhou X, Gu R, Ding Z. Dose responsive effects of cisplatin in L02 cells using NMR-based metabolomics. Environ Toxicol Pharmacol. 2014;37(1):150–7.

24. Wang H, Chen J, Feng Y, Zhou W, Zhang J, Yu Y, et al. 1H nuclear magnetic resonance-based extracellular metabolomic analysis of multidrug resistant Tca8113 oral squamous carcinoma cells. Oncol Lett. 2015;9(6):2551–9.

25. Puchades-Carrasco L, Lecumberri R, Martínez-López J, Lahuerta JJ, Mateos MV, Prósper F, et al. Multiple myeloma patients have a specific serum metabolomic profile that changes after achieving complete remission. Clin Cancer Res. 2013;19(17):4770–9.

26. Lei Z, Huhman DV, Sumner LW. Mass spectrometry strategies in metabolomics. J Biol Chem. 2011;286(29):25435–42.

27. Godzien J, Ciborowski M, Armitage EG, Jorge I, Camafeita E, Burillo E, et al. A single invial dual extraction strategy for the simultaneous lipidomics and proteomics analysis of HDL and LDL fractions. J Proteome Res. 2016;15(6):1762–75.

28. Calderón-Santiago M, Priego-Capote F, de Castro MDL. Enhancing detection coverage in untargeted metabolomics analysis by solid-phase extraction on-line coupled to LC-MS/MS. Electrophoresis. 2015;36(18):2179–87.

29. Dettmer K, Aronov PA, Hammock BD. Mass Spectrom Rev. 2007;26(1):51–78.

30. Moco S, Vervoort J, Moco S, Bino RJ, De Vos RCH, Bino R. Metabolomics technologies and metabolite identification. Trends Anal Chem. 2007;26(9):855–66.

31. Bujak R, Struck-Lewicka W, Markuszewski MJ, Kaliszan R. Metabolomics for laboratory diagnostics. J Pharm Biomed Anal. 2015;113:108–20.

32. Li Y, Song X, Zhao X, Zou L, Xu G. Serum metabolic profiling study of lung cancer using ultra high performance liquid chromatography/quadrupole time-of-flight mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 2014;966:147–53.

33. Xu X, Cheng S, Ding C, Lv Z, Chen D, Wu J, et al. Identification of bile biomarkers of biliary tract cancer through a liquid chromatography/mass spectrometry-based metabolomic method. Mol Med Rep. 2015;11(3):2191–8.

34. Liang Q, Wang C, Li B. Metabolomic analysis using liquid chromatography/mass spectrometry for gastric cancer. Appl Biochem Biotechnol. 2015;176(8):2170–84.

35. Peng J, Chen YT, Chen CL, Li L. Development of a universal metabolome-standard method for long-term LC-MS metabolome profiling and its application for bladder cancer urine-metabolite- biomarker discovery. Anal Chem. 2014;86(13):6540–7.

36. Lin L, Huang Z, Gao Y, Chen Y, Hang W, Xing J, et al. LC-MS-based serum metabolic profiling for genitourinary cancer classification and cancer type-specific biomarker discovery. Proteomics. 2012;12(14):2238–46.

37. Kelly AD, Breitkopf SB, Yuan M, Goldsmith J, Spentzos D, Asara JM. Metabolomic profiling from formalin-fixed, paraffin-embedded tumor tissue using targeted LC/MS/MS: application in sarcoma. PLoS One. 2011;6(10):e25357.

38. Bowers J, Hughes E, Skill N, Maluccio M, Raftery D. Detection of hepatocellular carcinoma in hepatitis C patients: biomarker discovery by LC-MS. J Chromatogr B Analyt Technol Biomed Life Sci. 2014;966:154–62.

39. Piszcz J, Lemancewicz D, Dudzik D, Ciborowski M. Differences and similarities between LC-MS derived serum fingerprints of patients with B-cell malignancies. Electrophoresis. 2013;34(19):2857–64.

40. Lin L, Huang Z, Gao Y, Yan X, Xing J, Hang W. LC-MS based serum metabonomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery. J Proteome Res. 2011;10(3):1396–405.

41. Bannur Z, Teh LK, Hennesy T, Rosli WRW, Mohamad N, Nasir A, et al. The differential metabolite profiles of acute lymphoblastic leukaemic patients treated with 6-mercaptopurine using untargeted metabolomics approach. Clin Biochem. 2014;47(6):427–31.

42. Huang G, Liu X, Jiao L, Xu C, Zhang Z, Wang L, et al. Metabolomic evaluation of the response to endocrine therapy in patients with prostate cancer. Eur J Pharmacol. 2014;729(1):132–7.

43. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis. 2012;4(18):2249–64.

44. Ranjbar MRN, Luo Y, DiPoto C, Varghese RS, Ferrarini A, Zhang C, et al. GC-MS based plasma metabolomics for identification of candidate biomarkers for hepatocellular carcinoma in Egyptian cohort. PLoS One. 2015;10(6):e0127299.

45. Nakamizo S, Sasayama T, Shinohara M, Irino Y, Nishiumi S, Nishihara M, et al. GC/MS-based metabolomic analysis of cerebrospinal fluid (CSF) from glioma patients. J Neurooncol. 2013;113(1):65–74.

46. Cameron SJS, Lewis KE, Beckmann M, Allison GG, Ghosal R, Lewis PD, et al. The metabolomic detection of lung cancer biomarkers in sputum. Lung Cancer. 2016;94:88–95.

47. Qiu Y, Cai G, Zhou B, Li D, Zhao A, Xie G, et al. A distinct metabolic signature of human colorectal cancer with prognostic potential. Clin Cancer Res. 2014;20(8):2136–46.

48. Pasikanti KK, Esuvaranathan K, Hong Y, Ho PC, Mahendran R, Raman Nee Mani L, et al. Urinary metabotyping of bladder cancer using two-dimensional gas chromatography time-of-flight mass spectrometry. J Proteome Res. 2013;12(9):3865–73.

49. Budczies J, Denkert C, Müller BM, Brockmöller SF, Klauschen F, Györffy B, et al. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue – a GC-TOFMS based metabolomics study. BMC Genomics. 2012;13:334.

50. Yu L, Aa J, Xu J, Sun M, Qian S, Cheng L, et al. Metabolomic phenotype of gastric cancer and precancerous stages based on gas chromatography time-of-flight mass spectrometry. J Gastroenterol Hepatol. 2011;26(8):1290–7.

51. Wu H, Xue R, Lu C, Deng C, Liu T, Zeng H, et al. Metabolomic study for diagnostic model of oesophageal cancer using gas chromatography/mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 2009;877(27):3111–7.

52. Wojakowska A, Chekan M, Marczak Ł, Polanski K, Lange D, Pietrowska M, et al. Detection of metabolites discriminating subtypes of thyroid cancer: molecular profiling of FFPE samples using the GC/MS approach. Mol Cell Endocrinol. 2015;417:149–57.

53. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, et al. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics. 2009;5(4):435–58.

54. Zeng J, Yin P, Tan Y, Dong L, Hu C, Huang Q, et al. Metabolomics study of hepatocellular carcinoma: discovery and validation of serum potential biomarkers by using capillary electrophoresis-mass spectrometry. J Proteome Res. 2014;13(7):3420–31.
55. Chen JL, Fan J, Lu XJ. CE-MS based on moving reaction boundary method for urinary metabolomic analysis of gastric cancer patients. Electrophoresis. 2014;35(7):1032–9.
56. Kami K, Fujimori T, Sato H, Sato M, Yamamoto H, Ohashi Y, et al. Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry. Metabolomics. 2013;9(2):444–53.
57. Simó C, Ibáñez C, Gómez-Martínez Á, Ferragut JA, Cifuentes A. Is metabolomics reachable? Different purification strategies of human colon cancer cells provide different CE-MS metabolite profiles. Electrophoresis. 2011;32(13):1765–77.
58. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;3(3):211–21.
59. Lynn KS, Cheng ML, Chen YR, Hsu C, Chen A, Lih TM, et al. Metabolite identification for mass spectrometry-based metabolomics using multiple types of correlated ion information. Anal Chem. 2015;87(4):2143–51.
60. Berg M, Vanaerschot M, Jankevics A, Cuypers B, Breitling R, Dujardin J-C. LC-MS metabolomics from study design to data-analysis – using a versatile pathogen as a test case. Comput Struct Biotechnol J. 2013;4(5):e201301002.
61. Hao D, Sarfaraz MO, Farshidfar F, Bebb DG, Lee CY, Card CM, et al. Temporal characterization of serum metabolite signatures in lung cancer patients undergoing treatment. Metabolomics. 2016;12(3):1–9.
62. Fan Y, Zhou X, Xia T, Chen Z, Li J, Liu Q, et al. Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer. Oncotarget. 2016;7(9):9925–38.
63. Gao P, Zhou C, Zhao L, Zhang G, Zhang Y. Tissue amino acid profile could be used to differentiate advanced adenoma from colorectal cancer. J Pharm Biomed Anal. 2016;118:349–55.
64. Giskeødegård GF, Hansen AF, Bertilsson H, Gonzalez SV, Kristiansen KA, Bruheim P, et al. Metabolic markers in blood can separate prostate cancer from benign prostatic hyperplasia. Br J Cancer. 2015;113:1712–9.
65. Ellis DI, Dunn WB, Griffin JL, Allwood JW, Goodacre R. Metabolic fingerprinting as a diagnostic tool. Pharmacogenomics. 2007;8(9):1243–66.
66. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology: a review. Clin Cancer Res. 2009;15(2):431–40.
67. Noto A, Cibecchini F, Fanos V, Mussap M. NGAL and metabolomics: the single biomarker to reveal the metabolome alterations in kidney injury. Biomed Res Int. 2013:612032, 6.
68. Zhang A, Sun H, Yan G, Wang P, Wang X. Mass spectrometry-based metabolomics: applications to biomarker and metabolic pathway research. Biomed Chromatogr. 2016;30(1):7–12.
69. Piszcz J, Armitage EG, Ferrarini A, Rupérez FJ, Kulczynska A, Bolkun L, et al. To treat or not to treat: metabolomics reveals biomarkers for treatment indication in chronic lymphocytic leukaemia patients. Oncotarget. 2016;7(16):22324–38.
70. Shang X, Zhong X, Tian X. Metabolomics of papillary thyroid carcinoma tissues: potential biomarkers for diagnosis and promising targets for therapy. Tumor Biol. 2016;37:11163–75.
71. Wen CP, Zhang F, Liang D, Wen C, Gu J, Skinner H, et al. The ability of bilirubin in identifying smokers with higher risk of lung cancer: a large cohort study in conjunction with global metabolomic profiling. Clin Cancer Res. 2015;21(1):193–200.
72. Chen J, Zhang X, Cao R, Lu X, Zhao S, Fekete A, et al. Serum 27-nor-5β-cholestane-3,7,12,24,25 pentol glucuronide discovered by metabolomics as potential diagnostic biomarker for epithelium ovarian cancer. J Proteome Res. 2011;10(5):2625–32.
73. Sanchez-Espiridion B, Liang D, Ajani JA, Liang S, Ye Y, Hildebrandt MAT, et al. Identification of serum markers of esophageal adenocarcinoma by global and targeted metabolic profiling. Clin Gastroenterol Hepatol. 2015;13(10):1730–7.

74. Brown DG, Rao S, Weir TL, O'Malia J, Bazan M, Brown RJ, et al. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. Cancer Metab. 2016;4(1):11.
75. Crown SB, Antoniewicz MR. Parallel labeling experiments and metabolic flux analysis: past, present and future methodologies. Metab Eng. 2013;16:21–32.
76. Metallo CM, Walther JL, Stephanopoulos G. Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. J Biotechnol. 2009;144(3):167–74.
77. Son J, Lyssiotis CA, Ying H, Wang X, Hua S, Ligorio M, et al. Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. Nature. 2013;496(7443):101–5.
78. Yang C, Ko B, Hensley CT, Jiang L, Wasti AT, Kim J, et al. Glutamine oxidation maintains the TCA cycle and cell survival during impaired mitochondrial pyruvate transport. Mol Cell. 2014;56(3):414–24.
79. Ye J, Mancuso A, Tong X, Ward PS, Fan J, Rabinowitz JD, et al. Pyruvate kinase M2 promotes de novo serine synthesis to sustain mTORC1 activity and cell proliferation. Proc Natl Acad Sci U S A. 2012;109(18):6904–9.
80. Caldwell RL, Caprioli RM. Tissue profiling by mass spectrometry: a review of methodology and applications. Mol Cell Proteomics. 2005;4(4):394–401.
81. Duncan MW, Nedelkov D, Walsh R, Hattan SJ. Applications of MALDI mass spectrometry in clinical chemistry. Clin Chem. 2016;62(1):134–43.
82. Aikawa H, Hayashi M, Ryu S, Yamashita M, Ohtsuka N, Nishidate M, et al. Visualizing spatial distribution of alectinib in murine brain using quantitative mass spectrometry imaging. Sci Rep. 2016;6:23749.
83. Torok S, Vegvari A, Rezeli M, Fehniger TE, Tovari J, Paku S, et al. Localization of sunitinib, its metabolites and its target receptors in tumour-bearing mice: a MALDI-MS imaging study. Br J Pharmacol. 2015;172(4):1148–63.
84. Végvári Á, Shavkunov AS, Fehniger TE, Grabau D, Niméus E, Marko-Varga G. Localization of tamoxifen in human breast cancer tumors by MALDI mass spectrometry imaging. Clin Transl Med. 2016;5(1):10.
85. Reyzer ML, Hsieh Y, Ng K, Korfmacher WA, Caprioli RM. Direct analysis of drug candidates in tissue by matrix-assisted laser desorption/ionization mass spectrometry. J Mass Spectrom. 2003;38(10):1081–92.
86. Dekker TJA, Jones EA, Corver WE, van Zeijl RJM, Deelder AM, Tollenaar RAEM, et al. Towards imaging metabolic pathways in tissues. Anal Bioanal Chem. 2015;407(8):2167–76.
87. Kubo A, Ohmura M, Wakui M, Harada T, Kajihara S, Ogawa K, et al. Semi-quantitative analyses of metabolic systems of human colon cancer metastatic xenografts in livers of super-immunodeficient NOG mice. Anal Bioanal Chem. 2011;400(7):1895–904.
88. Van Hove ERA, Blackwell TR, Klinkert I, Eijkel GB, Heeren RMA, Glunde K. Multimodal mass spectrometric imaging of small molecules reveals distinct spatio-molecular signatures in differentially metastatic breast tumor models. Cancer Res. 2010;70(22):9012–21.
89. Wang J, Qiu S, Chen S, Xiong C, Liu H, Wang J, et al. MALDI-TOF MS imaging of metabolites with a N-(1-naphthyl) ethylenediamine dihydrochloride matrix and its application to colorectal cancer liver metastasis. Anal Chem. 2015;87(1):422–30.
90. He J, Sinues PM-L, Hollmén M, Li X, Detmar M, Zenobi R. Fingerprinting breast cancer vs. normal mammary cells by mass spectrometric analysis of volatiles. Sci Rep. 2014;4:5196.
91. Leichtle AB, Nuoffer JM, Ceglarek U, Kase J, Conrad T, Witzigmann H, et al. Serum amino acid profiles and their alterations in colorectal cancer. Metabolomics. 2012;8(4):643–53.
92. Gaul DA, Mezencev R, Long TQ, Jones CM, Benigno BB, Gray A, et al. Highly-accurate metabolomic detection of early-stage ovarian cancer. Sci Rep. 2015;5:16351.
93. Asiago VM, Alvarado LZ, Shanaiah N, Gowda GAN, Owusu-Sarfo K, Ballas RA, et al. Early detection of recurrent breast cancer using metabolite profiling. Cancer Res. 2010;70(21):8309–18.
94. Lodi A, Tiziani S, Khanim FL, Günther UL, Viant MR, Morgan GJ, et al. Proton NMR-based metabolite analyses of archived serial paired serum and urine samples from myeloma patients at different stages of disease activity identifies acetylcarnitine as a novel marker of active disease. PLoS One. 2013;8(2):e56422.

95. Tenori L, Oakman C, Morris PG, Gralka E, Turner N, Cappadona S, et al. Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. Mol Oncol. 2015;9(1):128–39.
96. Alberice JV, Amaral AFS, Armitage EG, Lorente JA, Algaba F, Carrilho E, et al. Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry metabolomics approach. J Chromatogr A. 2013;1318:163–70.
97. Zhu J, Djukovic D, Deng L, Gu H, Himmati F, Abu Zaid M, et al. Targeted serum metabolite profiling and sequential metabolite ratio analysis for colorectal cancer progression monitoring. Anal Bioanal Chem. 2015;407:7857–63.
98. Armitage EG, Kotze HL, Allwood JW, Dunn WB, Goodacre R, Williams KJ. Metabolic profiling reveals potential metabolic markers associated with Hypoxia Inducible Factor-mediated signalling in hypoxic cancer cells. Sci Rep. 2015;5:15649.
99. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov. 2016;9(5):307–22.
100. Lord SR, Patel N, Liu D, Fenwick J, Gleeson F, Buffa F, et al. Neoadjuvant window studies of metformin and biomarker development for drugs targeting cancer metabolism. J Natl Cancer Inst Monogr. 2015;2015(51):81–6.
101. Schuler KM, Rambally BS, DiFurio MJ, Sampey BP, Gehrig PA, Makowski L, et al. Antiproliferative and metabolic effects of metformin in a preoperative window clinical trial for endometrial cancer. Cancer Med. 2015;4(2):161–73.
102. He J, Wang K, Zheng N, Qiu Y, Xie G, Su M, et al. Metformin suppressed the proliferation of LoVo cells and induced a time-dependent metabolic and transcriptional alteration. Sci Rep. 2015;5:17423.
103. van Asten JJA, Vettukattil R, Buckle T, Rottenberg S, van Leeuwen F, Bathen TF, et al. Increased levels of choline metabolites are an early marker of docetaxel treatment response in BRCA1-mutated mouse mammary tumors: an assessment by ex vivo proton magnetic resonance spectroscopy. J Transl Med. 2015;13:114.
104. Poisson LM, Munkarah A, Madi H, Datta I, Hensley-Alford S, Tebbe C, et al. A metabolomic approach to identifying platinum resistance in ovarian cancer. J Ovarian Res. 2015;8(1):13.
105. Stäubert C, Bhuiyan H, Lindahl A, Broom OJ, Zhu Y, Islam S, et al. Rewired metabolism in drug-resistant leukemia cells: a metabolic switch hallmarked by reduced dependence on exogenous glutamine. J Biol Chem. 2015;290(13):8348–59.
106. Willis JCD, Lord GM. Immune biomarkers: the promises and pitfalls of personalized medicine. Nat Rev Immunol. 2015;15(5):323–9.
107. Wettersten HI, Hakimi AA, Morin D, Bianchi C, Johnstone ME, Donohoe DR, et al. Grade-dependent metabolic reprogramming in kidney cancer revealed by combined proteomics and metabolomics analysis. Cancer Res. 2015;75(12):2541–52.
108. Minton DR, Nanus DM. Kidney cancer: novel targets in altered tumour metabolism in kidney cancer. Nat Rev Urol. 2015;12(8):428–9.
109. Liesenfeld DB, Botma A, Habermann N, Toth R, Weigel C, Popanda O, et al. Aspirin reduces plasma concentrations of the oncometabolite 2-hydroxyglutarate: results of a randomized, double-blind, crossover trial. Cancer Epidemiol Biomarkers Prev. 2016;25(1):180–7.
110. Lovelace ES, Wagoner J, MacDonald J, Bammler T, Bruckner J, Brownell J, et al. Silymarin suppresses cellular inflammation by inducing reparative stress signaling. J Nat Prod. 2015;78(8):1990–2000.
111. Peng Z xiao, Wang Y, Gu X, Xue Y, Wu Q, Zhou J yi, et al. Metabolic transformation of breast cancer in a MCF-7 xenograft mouse model and inhibitory effect of volatile oil from Saussurea lappa Decne treatment. Metabolomics. Springer: US. 2015;11(3):636–56.
112. Barrajón-Catalán E, Taamalli A, Quirantes-Piné R, Roldan-Segura C, Arráez-Román D, Segura-Carretero A, et al. Differential metabolomic analysis of the potential antiproliferative mechanism of olive leaf extract on the JIMT-1 breast cancer cell line. J Pharm Biomed Anal. 2015;105:156–62.

113. Chen GQ, Tang CF, Shi XK, Lin CY, Fatima S, Pan XH, et al. Halofuginone inhibits colorectal cancer growth through suppression of Akt/mTORC1 signaling and glucose metabolism. Oncotarget. 2015;6(27):24148–62.
114. Gao D, Wang Y, Xie W, Yang T, Jiang Y, Guo Y, et al. Metabolomics study on the antitumor effect of marine natural compound flexibilide in HCT-116 colon cancer cell line. J Chromatogr B Analyt Technol Biomed Life Sci. 2016;1014:17–23.
115. Yun J, Mullarky E, Lu C, Bosch KN, Kavalier A, Rivera K, et al. Vitamin C selectively kills KRAS and BRAF mutant colorectal cancer cells by targeting GAPDH. Science. 2015;350(6266):1391–6.
116. Li F, Yang XW, Krausz KW, Nichols RG, Xu W, Patterson AD, et al. Modulation of colon cancer by nutmeg. J Proteome Res. 2015;14(4):1937–46.
117. Trifonova O, Knight RA, Lisitsa A, Melino G, Antonov AV. Exploration of individuality in drug metabolism by high-throughput metabolomics: the fast line for personalized medicine. Drug Discov Today. 2016;21(1):103–10.
118. Schilsky RL. Personalized medicine in oncology: the future is now. Nat Rev. 2009;9:363–6.
119. Madlensky L, Natarajan L, Tchu S, Pu M, Mortimer J, Flatt SW, et al. Tamoxifen metabolite concentrations, CYP2D6 genotype, and breast cancer outcomes. Clin Pharmacol Ther. 2011;89(5):718–25.
120. Goldman JM, Melo JV. Chronic myeloid leukemia – advances in biology and new approaches to treatment. N Engl J Med. 2003;349:1451–64.
121. Navarrete A, Armitage EG, Musteanu M, García A, Mastrangelo A, Bujak R, et al. Metabolomic evaluation of Mitomycin C and rapamycin in a personalized treatment of pancreatic cancer. Pharmacol Res Perspect. 2014;2(6):e00067.
122. Bertini I, Cacciatore S, Jensen BV, Schou JV, Johansen JS, Kruhøffer M, et al. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. Cancer Res. 2012;72(1):356–64.
123. Tenori L, Oakman C, Claudino WM, Bernini P, Cappadona S, Nepi S, et al. Exploration of serum metabolomic profiles and outcomes in women with metastatic breast cancer: a pilot study. Mol Oncol. 2012;6(4):437–44.
124. Li H, He J, Jia W. The influence of gut microbiota on drug metabolism and toxicity. Expert Opin Drug Metab Toxicol. 2016;12(1):31–40.
125. Vétizou M, Pitt JM, Daillère R, Lepage P, Waldschmitt N, Flament C, et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science. 2015;350(6264):1079–84.

Chapter 10
Chronic Diseases and Lifestyle Biomarkers Identification by Metabolomics

Annalaura Mastrangelo and Coral Barbas

Abstract Chronic diseases, also known as noncommunicable diseases (NCDs), are complex disorders that last for long periods of time and progress slowly. They currently account for the major cause of death worldwide with an alarming increase in rate both in developed and developing countries. In this chapter, the principal metabolomic-based investigations on chronic diseases (cardiovascular diseases, diabetes, and respiratory chronic diseases) and their major risk factors (particularly overweight/obesity) are described by focusing both on metabolites and metabolic pathways. Additional information on the contribution of metabolomics strategies in the ambit of the biomarker discovery for NCDs is also provided by exploring the major prospective studies of the last years (i.e., Framingham Heart Study, EPIC, MONICA, KORA, FINRIK, ECLIPSE). The metabolic signature of diseases, which arises from the metabolomic-based investigation, is therefore depicted in the chapter by pointing out the potential of metabolomics to explain the pathophysiological mechanisms underlying a disease, as well as to propose new therapeutic targets for alternative treatments.

Keywords Diabetes • Cardiovascular diseases • Chronic respiratory diseases • Biomarker • Obesity • Risk factors • Metabolomics • Metabolic signature • Longitudinal studies • Metabotypes

A. Mastrangelo • C. Barbas (✉)
Center for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte 28668, Madrid, Spain
e-mail: cbarbas@ceu.es

Abbreviations

2AA	2-Aminoadipic acid
ArAA	Aromatic amino acids
AUC	Area under the curve
BA	Bile acid
BAIBA	Beta-aminoisobutyric
BCAA	Branched-chain amino acids
BCKDH	Branched-chain alpha-keto acid dehydrogenase
BMI	Body mass index
BWHHS	British Women's Heart and Health Study cohort
CE	Capillary electrophoresis
COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular disease
DM-AA	Diabetes-predictive amino acid
ECLIPSE	Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points
EPIC	European Prospective Investigation into Cancer and Nutrition
FA	Fatty acids
FAHFA	Fatty acid esters of hydroxy fatty acid
FAO	Fatty acids oxidation
FIA	Flow injection analysis
FINRISK	National FINRISK study
FSH	Framingham Heart Study
GC	Gas chromatography
GD	Gestational diabetes
HFA	Hydroxy fatty acids
IFG	Impaired fasting glycemia
IGT	Impaired glucose tolerance
IR	Insulin resistance
KORA	Cooperative Health Research in the Region Augsburg
LC	Liquid chromatography
LDLs	Low-density lipoproteins
LysoPC	Lysophosphocholine
LysoPEs	Lysophosphoethanolamines
MDC-CC	Malmö Diet and Cancer Study-Cardiovascular Cohort
MONICA	Multinational monitoring of trends and determinants in cardiovascular disease
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
mTOR	Mammalian target of rapamycin
NCDs	Noncommunicable diseases
NGT	Normal glucose tolerance
NMR	Nuclear magnetic resonance
PC	Phosphocholine

PCa   Alkyl-phosphatidylcholines
PL   Phospholipids
ROC   Receiver-operating characteristic
SABRE   Southall And Brent REvisited cohort
S-AMP   Adenylosuccinate
T1D   Type 1 diabetes
T2D   Type 2 diabetes
TMAO   Trimethylamine N-oxide

10.1   Chronic Diseases

Chronic diseases, also known as noncommunicable diseases (NCDs), are medical conditions that last for long periods of time and progress slowly. Often less visible than communicable diseases, they have noninfectious and non-transmissible cause. NCDs are currently the major cause of death worldwide (32 %, in 2012), more than all other causes combined (68 %, in 2012) [1, 2]. Contrary to common perception, the majority of all NCD deaths occur before the age of 70 and mainly in low- and middle-income countries where the access to affordable treatment and effective health-care services is limited [2]. The rapidly increasing burden of chronic diseases is a global threat for the population, not only for the high percentage of deaths but also for its economic, psychological, and social impact. Notably, the indirect costs of chronic diseases (e.g., inability to work, loss of productivity, cost of caregivers, among others) in the USA amounted to five times their direct costs (i.e., treatments, hospitalization) [3, 4].

The WHO has classified the major chronic diseases in four types as:

- Cardiovascular diseases (CVDs)
- Cancers
- Chronic respiratory diseases
- Diabetes

Altogether they accounted for the leading causes of NCD deaths in 2012 (see Fig. 10.1) [2]. In the following sections, the principal chronic diseases will be described, except for cancers, which are fully discussed in Chap. 9. Additionally, the major risk factors leading to chronic diseases will be explored, focusing on overweight/obesity as one of the biggest contributors.

10.1.1   Cardiovascular Diseases

Cardiovascular diseases (CVDs) comprise several disorders of the heart and blood vessels including coronary heart disease (which leads to heart attack), cerebrovascular disease (which leads to stroke), rheumatic heart diseases, and other conditions. CVDs, particularly heart attack and stroke, are the first cause of death globally, with

Leading causes of deaths by noncommunicable diseases (NCDs)
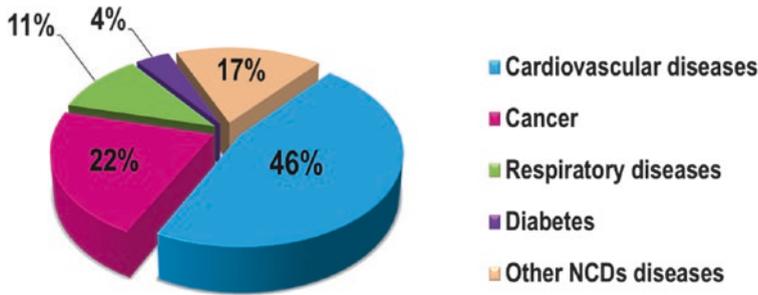
Source: Global status report on noncommunicable diseases 2014 - World Health Organization (WHO)

Fig. 10.1 Pie chart displaying the leading causes of noncommunicable diseases (NCDs) deaths in 2012 [2] [Source: Global status report on noncommunicable diseases 2014 – World Health Organization (WHO)]

the low- and middle-income countries showing a substantial increased mortality over the years. Currently, over 80 % of cardiovascular deaths occur in developing countries with a projection to increase [5].

Heart attack and stroke are usually acute events resulting from inadequate blood supply to a portion of myocardium (myocardial ischemia) or of the brain (cerebral ischemia); they are strongly associated with atherosclerosis, which consists of lipid accumulation in large arteries, that narrows the inner surface of the vessels by blocking or severely reducing the normal blood flow. The resulting lack of oxygen and glucose induces the death of the cells, thereby damaging the tissue.

Atherosclerosis has a complex etiology; it is initiated by inflammation in the endothelial layer of the artery that allows the low-density lipoproteins (LDLs) to accumulate in the inner layer of the artery, the intima. LDLs, and their oxidized form, then trigger the transmigration of immune cells, particularly monocytes, into the intima by creating plaques that become progressively larger with time. The plaque formation is a slow and silent process that develops over the years and eventually results in the plaque break or in the complete coronary/cerebral artery blockage (heart attack/stroke) causing premature death if untreated.

Although the recovery from the damage is possible, ischemic events often evolve into chronic disabilities that markedly affect the individual long life both emotionally and physically. Most importantly, patients who have suffered a heart attack and stroke have increased likelihood for second coronary and cerebral events [6].

Hence, prevention of atherosclerosis and CVDs is the most effective measure to prevent from premature morbidity, mortality, and disability. Indeed, although the drug therapy (i.e., combining aspirin, statins, beta-blockers, and diuretics) is effective in reducing the number of ischemic events, the identification of high-risk subjects and the preventions from complications remain the best option, both for people with established disease and for those at high risk of developing disease.

10.1.2 Chronic Respiratory Diseases

Chronic respiratory diseases are a group of diseases affecting the airways and the other structures of the lungs. They include asthma and respiratory allergies, chronic obstructive pulmonary disease (COPD), occupational lung diseases, sleep apnea syndrome, and pulmonary hypertension [7]. According to the WHO, hundreds of millions of people are affected by chronic respiratory diseases, with asthma and COPD as the most prevalent lung diseases and major causes of morbidity and mortality worldwide. Indeed, it was estimated that currently 235 million people have asthma, whereas 64 million people suffer from COPD. Besides, in 2002, COPD has been the fifth leading cause of death globally, and it is expected to become the third in 2030 [8].

Asthma and COPD are multifactorial and complex diseases. They are characterized by a remarkable heterogeneity both in the clinical course and in their pathophysiological phenotypes that makes them frequently under-recognized, underdiagnosed, undertreated, and insufficiently prevented.

Asthma is a chronic inflammatory disorder mostly common among children where it appears with the same incidence as cancer and diabetes [9]. The typical symptoms include episodes of wheezing, coughing, chest tightness, and shortness of breath, generally in response to environmental exposure to various stimuli (allergens, viral respiratory infections, irritant fumes or gases). Along with a genetic predisposition, they trigger an inflammatory and immune response in the lungs' airways that causes an abnormal narrowing of the airways leading to the typical asthma symptoms [7].

In contrast, *COPD* is a multicomponent and systemic syndrome that affects both lungs and organs outside the lungs. It includes conditions such as emphysema and chronic bronchitis and is characterized by shortness of breath, cough, and sputum production. The principal underlying cause is cigarette smoking both from primary and secondhand exposure that together with occupational dust and chemicals (in high-income country) and indoor and outdoor pollution (mainly in low- and middle-income countries) damages the lungs progressively and irreversibly. COPD progresses slowly and is mostly asymptomatic until the frequent exacerbations and further reductions in airflow make it clinically apparent, generally by the age of 40 [10].

Up to now, no precise diagnosis or definitive therapy is available both for asthma and COPD. The diagnosis is typically based on the pattern of symptoms and the response to therapy over time and is eventually confirmed by the spirometry test. Concerning the therapy, the medicaments commonly employed are bronchodilators (long- and short-acting beta-agonist) and corticosteroids that reduce the inflammation and relieve the symptoms and oxygen administration for patients with chronic respiratory failure. Furthermore, avoiding asthma triggers reduces the severity of the asthmatic attack [11].

Although the management of these diseases is possible, they remain a health threat that need to be monitored over the life span. Indeed, the failure to use appropriate medications or to adhere to treatment can lead to death. Hence, the development

of reliable tests for an early and accurate diagnosis, the reduction of the exposure to the major risk factors, and prevention strategies to control the progression, exacerbation, and complications of the disease are the essential measure to efficiently manage these serious long-term diseases.

10.1.3 Diabetes

Diabetes is a collection of metabolic diseases characterized by chronic high blood glucose levels (hyperglycemia) that, if not well controlled, causes serious damages to the whole body (i.e., the heart, blood vessels, eyes, kidneys, and nerves) and other long-term consequences that impair the quality of life significantly [12].

The WHO estimates that in 2014 diabetes has affected 422 million people in the world, mainly Southeast Asia and Western Pacific Regions, with prevalence among adult population. Over the past few decades, diabetes showed a steady rise (the incidence of diabetes has quadrupled since 1980), particularly in low- and middle-income countries, with an increased frequency in children and young people [13]. Moreover, in 2012, diabetes was the eighth leading cause of death globally with 3.7 million deaths, 1.5 million of which directly caused by diabetes and additional 2.2 million deaths from diseases (i.e., cardiovascular diseases, chronic kidney disease, and tuberculosis) related to higher-than-optimal blood glucose [13].

According to the different etiology underlying the insulin deficiency that causes hyperglycemia, diabetes has been classified in type 1 diabetes (T1D) and type 2 diabetes (T2D) [14]. Other conditions characterized by higher-than-optimal blood glucose have been described including impaired glucose tolerance (IGT) and impaired fasting glycemia (IFG); they are intermediate conditions of hyperglycemia that may result in diabetes (mainly T2D). Although these conditions are not established diseases, they increase the risk for complications (e.g., CVDs) and have to be adequately monitored for life [12]. Additionally, gestational diabetes (GD), which is characterized by hyperglycemia and hyperinsulinemia that occur during pregnancy and usually remits after pregnancy, has been described as a temporary form of diabetes that can be responsible for adverse outcomes during pregnancy, childbirth, and future susceptibility to T2D [14, 15].

Concerning T1D, it is characterized by a selective autoimmune destruction of the pancreatic β-cells that reduces and eventually eliminates insulin production. Commonly with a juvenile onset and with a lesser incidence, it occurs in genetically susceptible individuals that are exposed to environmental factors still not well-defined (hypothetically viral infections, gut microbiota, and specific diet) [16]. In contrast, T2D results from gradual depletion in pancreatic β-cells mass and functions in response to peripheral insulin resistance that makes the body unresponsive to insulin and stimulates its secretion, thereby leading to β-cell exhaustion from failing to compensate the increased insulin demand. Genetic predisposition, ethnicity, older age, and environmental risk factors (i.e., overweight/obesity, inadequate physical activity, smoking, and poor diet) are the major underlying causes of T2D [13].

While no prevention strategies have yet been successful for T1D, since its etiology is still unknown, T2D is potentially preventable through diet and physical activity (remarkably more effective than medication) [13]. Besides, even though they are chronic progressive diseases, several measures can be carried out to assure long and healthy lives for diabetic subjects, including the access to insulin and a strict control of glycemia for T1D and lifestyle interventions and early diagnosis for T2D [13]. Early diagnosis is particularly important for T2D since its symptomatology is less marked than in T1D, and the diagnostic assays generally employed (i.e., fasting blood glucose test, oral glucose tolerance tests, measuring glycated hemoglobin) do not provide prediabetic and diabetic threshold values [14, 17]. It is noteworthy that despite the availability of several diagnostic tests, up to 62 % of T2D cases are undiagnosed and untreated [18]. This underscores the need for enhanced diagnostic tools to allow the delay or even the prevention of the disease onset and its complications.

10.1.4   Risk Factors

Risk factors for chronic diseases can be gathered into three strongly interrelated groups: underlying factors (e.g., globalization, urbanization, socioeconomic determinants, aging), behavioral risk factors (e.g., physical inactivity, alcohol abuse, unhealthy diet, tobacco use), and metabolic/physiological risk factors (e.g., hypertension, hyperglycemia, hyperlipidemia, and overweight/obesity) [19].

Urbanization and globalization have greatly influenced the habits of the developing countries by promoting the rise of untraditional diets; the use of processed foods high in saturated fats, salt, and sugar; an increased tobacco and alcohol use; urban air pollution; and a more sedentary lifestyle, among others. The chronical exposure to these behavioral risk factors then represents the main underlying cause of NCDs and premature death. In 2012, alcohol abuse was responsible for 3.3 million deaths, with NCDs being responsible for more than half. In addition, tobacco causes six million preventable deaths every year, whereas 3.2 million annual deaths have been attributed to insufficient physical activity, and 1.7 million annual deaths from CVDs were attributed to excess of salt intake [2].

Behavioral risk factors are also responsible for metabolic/physiological alterations including hyperglycemia, hypertension, hyperlipidemia, and overweight/obesity that, in turn, contribute to the progression of the disease toward life-impairing complications and premature death.

However, despite the alarming incidence of chronic diseases worldwide, their slow evolution and the dependence on modifiable risk factors have influenced preventing measures which are expected to reduce the prevalence of NCDs by 25 % by 2015 [2].

Among the risk factors for chronic diseases, the condition of being overweight and obese is one of the biggest contributors. It can be considered as a model for the simultaneous investigation of several risk factors underlying the chronic diseases by

providing insights on the interaction patterns that may be responsible for the onset of such complex diseases. Indeed, the excess of body weight, which characterizes the medical condition of being overweight and obese, results from the interaction of genetic and environmental factors and includes at the same time the underlying and behavioral risk factors typical of NCDs. Besides, it is usually associated to the metabolic and physiological changes, such as hypertension, hyperlipidemia, and hyperglycemia, generally present in the NCDs, thereby posing a greater risk for their development. Importantly, the increase of NCDs over the years has mirrored the prevalence of obesity and overweight (e.g., in 2014, overweight and obesity accounted for about 65–80 % of the new cases of T2D in Europe).

10.1.4.1   Overweight and Obesity

Obesity is a complex condition that affects virtually all age and socioeconomic groups, thereby being a global health threat, the "globesity." In 2014, more than 1.9 billion adults worldwide were overweight, and over 600 million of which were obese, with a predominance of women. Furthermore, according to WHO estimates, obesity causes 3.4 million deaths every year, that along with deaths caused by diseases of which obesity is a leading factor, and its strong social and psychological impact has placed obesity at the forefront of public health concern [20].

In clinical and epidemiological practice, the body mass index (BMI) is the parameter internationally recommended to categorize adult underweight, normal weight, overweight, and obesity (see Fig. 10.2). However, since it is an ethnic-independent measurement, the possibility to employ alternative BMI cutoffs in Asia and the Pacific Regions, where the risk of developing chronic diseases is at a lower BMI level than populations of European origin, is under evaluation [21].

Moreover, the waist circumference has been employed as additional measurement of obesity for its relationship with the visceral fat which is independent of the

$$BMI = \frac{Weight\ (kg)}{Height\ (m)^2}$$

Underweight	Normal weight	Overweight	Obese
BMI ≤ 18.49	18.49 ≤ BMI ≤ 24.99	25.00 ≤ BMI ≤ 29.99	BMI ≥ 30.00
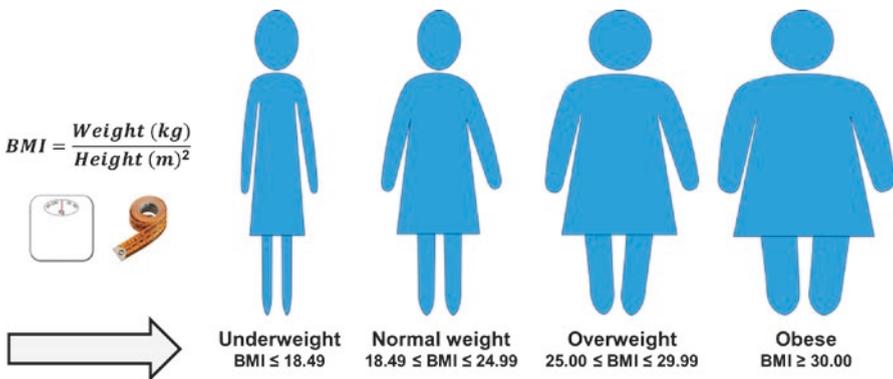
Fig. 10.2 International classification of adult underweight, overweight, and obesity according to BMI [21]

changes in BMI. It provides sex-specific cutoffs that combined with BMI have demonstrated to have a potential ability to predict the risk of chronic diseases. Indeed, higher waist circumference and BMI have been associated with increased risk of CVDs and T2D by allowing timely interventions [22].

Currently, behavioral strategies (diet and physical activity) and taxing policies (e.g., increased price for sugary beverages and unhealthy food) are the most efficient measures to treat obesity [2]. However, the rapid increase of obesity, particularly in children and developing countries, has boosted the need of strategies aimed to preventing and controlling obesity especially in these vulnerable populations. It is noteworthy that the incidence of childhood overweight is increasing worldwide with 42 million of children (<5 years old) overweight in 2013 [20]. Moreover, childhood obesity is strongly associated with higher cardiometabolic risks in adolescence and higher morbidity and mortality from NCDs, mainly T2D and CVDs, in adulthood [23–25]. Hence, reducing the prevalence of obesity in children would have a long-term effect on reducing the prevalence in adults as well as on the susceptibility to chronic diseases later in life. Importantly, considerable evidences have highlighted that chronic disease risk is present from fetal life and continues cumulatively during the life span [26–29]; life-course investigation and interventions are thus essential in order to face and control the incidence and the premature mortality from NCDs.

10.2   Metabolomics and Chronic Diseases

The development of chronic diseases is a complex process. NCDs are characterized by a progressive dysfunction of metabolic and physiological functions in response to chronical exposure to lifestyle factors. From an evolutionary perspective, the rapid cultural change has far outpaced the genetic adaptation by generating a mismatch between the human evolution and the daily life, thereby increasing the susceptibility to chronic diseases [30]. Moreover, because of the polygenic nature of human traits and their adaptive nature, the phenotypic expression of such diseases turned out to be heavily affected by the environment; NCDs are indeed considered as a physiological adaptation of the body homeostasis to harmful lifestyle behaviors.

Thus, metabolomics, which measures the entire set of metabolites of a wide range of biological specimen in a certain time and under particular conditions [31, 32], has emerged as a versatile and valuable tool to investigate the etiology and the pathophysiology of such complex diseases [33]. Indeed, since it is the most proximal to the phenotype among the omics, it offers the possibility to investigate metabolic pathways that play a role in the overall metabolic dysfunction underlying NCDs (either before their onset or during their progression).

Over the years, metabolomics has addressed the investigation of chronic diseases by providing an integrated perspective on how metabolites interact in response to specific exposures by characterizing metabolic signatures of the diseases [34–37].

Besides, metabolomics has demonstrated predictive, diagnostic, and prognostic capabilities that have enabled the study of factors influencing the onset and progression of chronic diseases [38–40]. Metabolomics studies can be indeed classified according to these qualities depending on whether the focus of the study was on the identification of the subjects more susceptible to develop a certain disease in the future, on the early detection of a currently occurring disease, and on the determination of features able to predict the disease outcome or the efficacy of a treatment, respectively [40].

Thus, while predictive studies have the power to tackle the growth of the chronic diseases by anticipating their onset, diagnostic and prognostic studies are able to improve the management of an already overt disease, by preventing its adverse outcome. To achieve this goal, predictive studies require large cohorts (i.e., several thousands of participants), where initially healthy subjects are monitored through a large period of time (i.e., over 10 years) in prospective study designs. During the follow-up period, then, a limited number of healthy subjects will develop the disease by allowing the identification of risk factors, which are strongly related to the disease onset [41] (see Sect. 10.5.3 for more details on the major prospective studies of chronic diseases investigated by metabolomics approaches). Concerning diagnostic and prognostic studies, they employ instead cross-sectional studies aimed to identify biomarkers that allow tracking of the disease state in order to achieve a more effective patient stratification and a more accurate characterization of the disease outcome and the monitoring of the treatment's effectiveness [42].

Novel biomarkers, thus, hold the promise to be relevant tools in the clinical setting (in combination, or not, with traditional biomarkers) by driving a more effective decision-making process that helps the physician in the daily clinical practice. Additional information on the contribution of metabolomics strategies in the ambit of biomarker discovery for chronical diseases will be discussed in detail in the following section.

10.3 Metabolomics and Biomarker Discovery

Biomarkers are classified as screening, diagnostic, and prognostic according to their capability on detecting a future disease, a suspected disease, and the progression or remission of overt disease, respectively [43]. Since many diseases result in characteristic changes in the metabolite profiles, several metabolites have been employed as reliable biomarkers for decades [44–47]. Over the last few years, high-throughput technologies such as metabolomics, which broaden the coverage of the metabolome, have been applied with more frequency in the field of the biomarker discovery [33, 41, 48].

An ideal biomarker should be safe and easy to measure, cost-effective during both the discovery and the follow-up processes, and consistent across genders and different ethnicities [43]. Regarding the use of metabolites as diagnostic markers, one of the major challenges in metabolomics is the validation of the compounds statistically significant in small sets of well-selected samples, in a big cohort. While there are numerous screening studies in metabolomics research producing potential

biomarkers, most of the identified biomarkers have failed to replace existing clinical tests. To become a clinically approved test, a potential biomarker should be confirmed and validated using hundreds of individuals and should be reproducible, specific, and sensitive. The reproducibility is assured by validating the biomarkers in other study samples, preferably from an independent cohort. In contrast, concerning sensitivity and specificity, they are essential features of a biomarker as they measure the biomarker's ability to correctly detect subjects with the target condition (true positive rate) and without the target condition (true negative rate), respectively. They are generally computed through the receiver operating characteristic (ROC) curve analysis which provides the C-statistics or area under the curve (AUC) as a measure of the predictive ability of the biomarker model with values that range from 0.5 (random classification) to 1.0 (perfect classification) [49, 50].

However, in multifactorial disorders such as NCDs, single biomarkers rarely own high values of specificity and sensitivity; therefore, a multiple biomarker approach has been increasingly employed over the years to select the simplest combination of biomarkers that produces an effective predictive outcome [43].

Biomarkers (alone and more frequently in combination) can be further employed to generate risk scores as an estimate of the individual's risk of developing a certain disease in the future. The risk scores are usually generated within prospective studies that allow exploring the contribution of a new biomarker in an already existing predictive model [51]. This assessment is carried out by evaluating the discrimination power of the new model (model discrimination), the agreement between the observed outcome and the expected risk (model calibration), and the possibility to refine the stratification of the population into more pragmatic risk categories (i.e., reclassification of the subjects from an intermediate risk level to either an upper or lower risk level, risk reclassification) [52, 53].

Indeed, it is important to point out that for a metabolite to be employable as a biomarker other than in the clinical research, it has to prove to strengthen the predictive model beyond that achieved by conventional biomarkers that are employed in the clinical practice [43].

The discovery of new biomarkers is therefore a challenging task that metabolomics has addressed only recently by providing promising findings mainly in the field of hypothesis-generating biomarkers. This typology of biomarkers, which is focused on explaining the pathophysiological mechanisms underlying a disease, aims to understand the metabolic alteration associated with a disease with the ultimate goal of driving the discovery of a more efficient and personalized treatment or the design of new drugs from an informed perspective.

10.4 Study Design and Analytical Considerations

Regarding the workflow in metabolomics, researchers in the field do not agree upon the terms, not only for metabolomics and metabonomics, which were originally considered as different definitions that nowadays are used indistinctively [31, 32,

54], but also for the approach employed (i.e., fingerprinting, global profiling, profiling, among others). Hence, in order to simplify, in the present chapter, they will be referred to as targeted or untargeted metabolomics.

In the targeted approach, specific metabolites of known identity are analyzed. In mass spectrometry (MS), this often involves the addition of multiple stable isotope-labeled standards to the biological sample prior to the extraction and derivatization steps to control for differences in analyte loss during sample processing and to compensate for ionization-suppression effects. Advantages of targeted methods are that (i) identification of compounds is straightforward and (ii) metabolites can be quantified. A disadvantage is their limited metabolite coverage that can include from a small set to several hundred metabolites.

In contrast, untargeted metabolomics involves the simultaneous measurement of as many metabolites as possible in a biological specimen. This approach is generally used in differential analysis of two or more biological or clinical states/treatments; the report consists of differences between the states and is based on signal abundances of raw spectral data. The chemical identity of the signals is not known a priori, and significant chemical/spectral analysis must be performed to define the molecular species. It is noteworthy to point out that while semiquantitative data can be employed in the discovery phase, quantitative data is paramount for implementation in the clinical practice.

In a standard metabolomics workflow, metabolites can come from any biofluid or tissue after convenient extraction and can be detected using various chemical detection platforms including MS and nuclear magnetic resonance (NMR) as the most important. Notably, due to the chemical diversity of the metabolites, no single analytical technique is able to cover the entire metabolome; therefore, whenever possible, a combination of platforms has been increasingly applied over the last few years. The multiplatform approach, indeed, broadens the metabolite coverage and at the same time allows a mutual cross validation of the metabolites that are detected in more than one analytical technique. Concerning NMR, it has the potential for high-throughput fingerprinting, minimal requirements for sample preparation, robustness of the response, and nondestructive nature of the technique. However, only medium to high abundance metabolites will be detected with this approach, and the identification of individual metabolites based on chemical shift signals, which cause sample clustering in multivariate analysis, is challenging in complex mixtures. MS-based metabolite detection is instead a powerful tool for investigations of metabolism due to its sensitivity for low-abundant molecules and flexibility for the detection of multiple chemical molecular classes. MS detection platforms are biased in their compatibility of a particular molecule with a mode of ionization or detection. The ability to globally profile highly complex mixtures of plant extracts is enhanced by coupling chromatography with MS detection. Thus, a "metabolomics platform" refers to the combination of a separation technique and MS. The most commonly utilized metabolomics platforms include liquid chromatography–mass spectrometry (LC-MS), gas chromatography–mass spectrometry (GC-MS), and capillary electrophoresis–mass spectrometry (CE-MS). Following data acquisition and processing, MS-metabolomics data is often expressed as a matrix of molecular features defined by (i) elution time, (ii) mass

(mass/charge ratio), and (iii) abundance of the mass signal. Annotating the detected molecular feature as a metabolite is the major bottleneck in MS-metabolomics workflows [55].

10.5   Metabolic Signatures of Chronic Diseases

In the past decade, metabolomics has made remarkable progress in providing new insights into the systemic alteration underlying NCDs: (1) disease-related metabotypes have been described that reflect changes in metabolites (i.e., amino acids, lipids, and organic acids) in body fluids, organs, and/or tissues as consequence of disease or disease-related conditions; (2) the role of new contributors (i.e., gut microbiome) in the development and progression of NCDs has been unveiled; (3) markers of the disorder's onset, progression, and prognosis have been identified in prospective metabolomic-based studies. The metabolic signature of chronic diseases that arises from these discoveries (Fig. 10.3) will be described in the following sections by focusing both on metabolites and metabolic pathways.

10.5.1   Metabotypes of NCDs

In 2009, Newgard et al. described for the first time a metabotype of obesity and insulin resistance (IR) characterized by the increase in branched-chain amino acids (BCAA, i.e., leucine, isoleucine, and valine) and related metabolites (i.e., propionylcarnitine (C3), isovalerylcarnitine (C5), glutamate) in mice ingesting a high-fat diet [36]. This finding was then corroborated by subsequent studies in obese and/or diabetic humans and rodents, by identifying BCAA and their by-products, mainly short-chain acylcarnitines, as sensitive metabolic marker of obesity, IR, and future T2D [56–58]. Interestingly, consistent with these studies, an improvement in insulin sensitivity associated with lower BCAA levels was described for subjects undergoing weight-loss interventions (i.e., dietary, behavioral, bariatric surgery) [59, 60].

Various hypotheses have been proposed for the increase of BCAA and related catabolites in obese and or/diabetic subjects including an increased protein intake, increased proteolysis, reduced protein anabolism, or impaired mitochondrial catabolism. While several studies have ascribed only a marginal role to the first four processes [61–63], the altered BCAA catabolism has been suggested as the principal mechanism underpinning such changes [58]. Recent findings have highlighted a decrease in BCAA-catabolizing enzymes (e.g., branched-chain alpha-keto acid dehydrogenase, BCKDH) in the fat and liver of obese genetically modified mice and rats; insulin was also linked to BCAA catabolism through its action on the hypothalamus [64]. Shin et al. indeed pointed out an inducing effect on the hepatic BCKDH, mediated by the insulin signaling in the brain that was
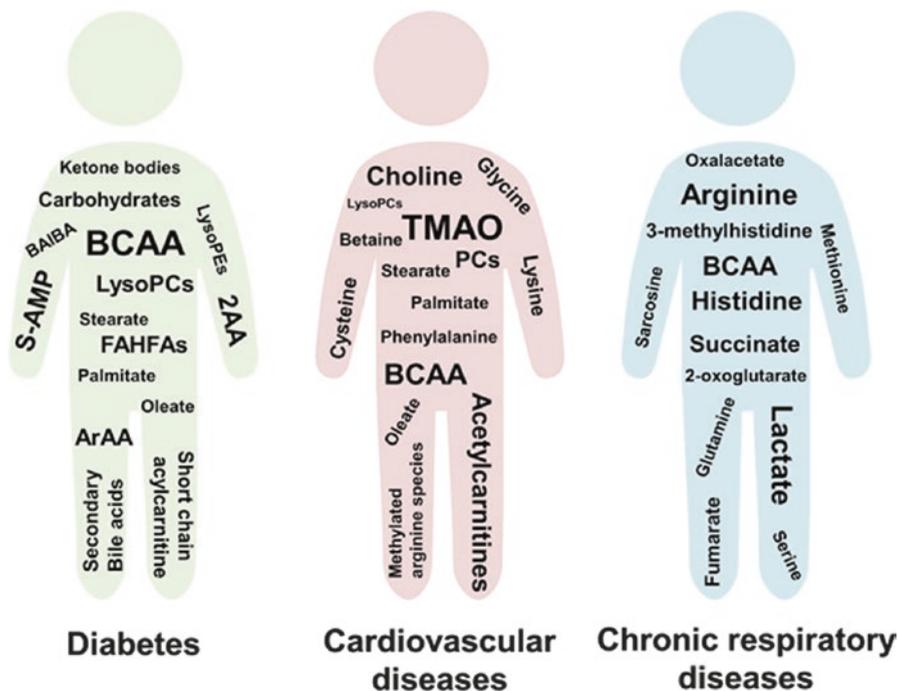
Fig. 10.3 Metabolic signatures of chronic diseases unveiled by metabolomics. Abbreviations: *BCAA* branched-chain amino acid, *S-AMP* adenylosuccinate, *BAIBA* beta-aminoisobutyric, *ArAA* aromatic amino acids, *PC* phosphocholine, *LysoPCs* lysophosphocholines, *LysoPEs* lysophospho-ethanolamines, *FAHFAs* fatty acid esters of hydroxy fatty acids, *2AA* 2-aminoadipic acid, *TMAO* trimethylamine N-oxide

found to be responsible for lowering the plasma levels of BCAA, thereby suggesting for the BCAA a role as marker of hypothalamic IR.

The detrimental effects mediated by BCAA have been attributed to their role on the overactivation of the mTOR (mammalian target of rapamycin) pathway which induces hepatic IR, thus worsening the systemic insulin signaling [65]. High levels of BCAA were also found to affect the fatty acids oxidation (FAO); the hepatic BCKDH is indeed involved in the catabolism of both BCAA and acylcarnitines, and in the case of high BCAA levels, it resulted to be overloaded, thereby producing incomplete FAO by-products (i.e., short-chain acylcarnitines) [66]. These metabolites then have been related to the mitochondrial stress and impaired insulin signaling that characterize T2D [67].

Further alterations in the amino acid metabolism were found to be associated to CVDs and/or related conditions. Wang et al. postulated that an increase in methylated arginine species (i.e., N-mono-methylarginine, asymmetrical dimethylarginine, and symmetrical dimethylarginine), which are related to the inhibition of the nitric oxide production, may serve as a marker of increased risk of coronary artery disease, myocardial infarction, and stroke [68], whereas Wang et al. revealed

changes in the levels of amino acids (glycine, lysine, and cysteine, particularly) in young hypertensive men by shedding light on metabolic variations taking place at an early stage of hypertension [69].

Changes in amino acid metabolism have been reported also for chronic respiratory diseases: Wedes et al. identified the urinary metabolite bromotyrosine, which is generated by the enzyme eosinophil peroxidase, as a noninvasive marker of future asthma exacerbation in children [70]; Jung et al. described alteration in metabolites (i.e., increase in methionine, glutamine, and histidine and decrease in acetate, choline, and arginine) in serum of asthma patients [71]; finally, several studies reported a decrease in plasma BCAA in COPD patients concomitant to cachexia [72–74].

A further example of metabotype of respiratory chronic diseases comprises of metabolites of the TCA cycle (i.e., succinate, fumarate, oxalacetate, cis-aconitate, and 2-oxoglutarate) that were found to be increased in urine and/or serum of asthmatic patients [71, 75]. High levels of lactate were also found in this patients by supporting the hypothesis of an upregulation of the TCA cycle due to a greater effort to breathe for the patients with a reduced oxygenation concomitant to the disease exacerbation [75].

An additional metabotype made up of lipids (i.e., mainly phospholipids and fatty acids) and illustrative of T2D, CVDs, and related conditions has emerged over the years by metabolomics and lipidomics approaches.

Phospholipids (PL) are an important class of lipids involved in NCDs. Generally described as the main components of the cellular membranes and lipoproteins (HDL and LDL mainly), they are involved in various metabolic pathways including signaling events and inflammation that are usually underlying NCDs and their related conditions. For instance, Ha and colleagues reported an altered lipid profile comprising of several PL metabolites, namely, six lysophosphocholines (LysoPCs C14:0, C16:1, C18:1, C18:3, C20:5, and C22:6) and three lysophosphoethanolamines (LysoPEs C18:1, C18:2, and C22:6), in case of diabetes; the lipid profile was also found to correlate to inflammation, oxidative stress, and future diabetes-related complications (i.e., arterial stiffness) [76]. Interestingly, also other LysoPCs (predominantly with long-chain acyl groups, $C \geq 16$) were found to be elevated in prehypertensive young men; these lipids then were described to be highly associated with oxidized LDLs, thereby featuring an increased oxidative stress and inflammation process as potential predictors of future hypertension, atherosclerosis, and CVDs [77].

High levels of fatty acids (i.e., palmitic acid, stearic acid, and oleic acid, among others) have been also associated to increased risk for T2D and CVDs. Yang et al. proposed a link between serum docosahexaenoic, palmitic, and palmitoleic acids and prevalence of hypertension [78]. Increased levels of free fatty acids and their oxidized by-product (beta-hydroxybutyrate, acetoacetate, and acetone) have been also associated to T2D and heart failure [79, 80]. Of note, a new class of fatty acids has been recently discovered by untargeted lipidomics, namely, the fatty acid esters of hydroxy fatty acids (FAHFAs) that consist of a combination of four fatty acids (FA) and four hydroxy fatty acids (HFA) [81]. FAHFAs were described to be present in food, synthesizable by mammalian, and at low levels in obese/insulin-resistant

humans and mice. Besides, conversely to other fatty acids, FAHFAs were described to exert a plethora of beneficial effects on diabetic-related conditions including the enhancement of glucose uptake from the bloodstream, improvement of insulin secretion and sensitivity, and reduction of inflammation. The FAHFAs' discovery represents therefore an important breakthrough in the field of NCDs and a great example of the potential of metabolomics for opening new avenues for the investigation of uncharacterized biochemical pathways in human physiology and diseases as well as proposing therapeutic targets for an alternative treatment of metabolic diseases.

10.5.2   New Contributors in Chronic Diseases

The gastrointestinal tract comprises around 1013 cells (1,183–3,180 bacterial phylotypes) in adult's intestinal microbiome [82], which means 3×10^6 genes (130-fold higher than the number in human body) for a metabolically active organ that has been proposed as one of the major contributors to human health and disease. Indeed, accumulating evidences highlighted the crucial role of the gut microbiota on the development of chronic diseases (mainly T2D and CVDs) and related conditions (obesity, IR, and atherosclerosis, among others) by its action in several metabolic pathways including lipid metabolism, inflammation, energy metabolism, and insulin signaling [83, 84]. In seminal work, Turnbaugh and colleagues demonstrated that the transplant of microbiota from obese mice to germfree recipients was able to transfer the obese phenotype to the recipients that indeed experienced an increased weight gain in comparison to the mice that received a "lean microbiota" [85]. This study represented an important new insights into the role of the microbiome in the development of a disease or diseases-related condition. Since then, several studies have been carried out to investigate the gut microbiota and its relationship with health and diseases. Concerning the metabolomic-based investigations, various metabolites mirroring the action of the microbiome have been uncovered by providing new insights into how the microbiota interacts with the host and which metabolic pathways are involved in the gut–host cross talk [86].

Wang et al. identified a novel metabolite, namely, the trimethylamine N-oxide (TMAO), with a pro-atherogenic action that was found to be generated by the action of the gut microbiota on the dietary phosphatidylcholine (PC) [68]. Dietary PC is indeed the main source of the TMAO's precursors (i.e., choline and betaine) that have been previously related to risk for CVDs (i.e., lower levels of choline and higher CVD risk). Together these metabolites were also described to increase the risk for future cardiovascular events, thereby unveiling an important link between dietary intake of lipids, gut microbiota, and future CVD events. Of note, high levels of TMAO were also found in the urine of T2D patients by highlighting the potential of these metabolites for alternative therapeutical approaches [87]. A further example of microbiota-derived metabolites that play a major role in the host metabolism is represented by secondary bile acids (BAs, deoxycholate, and lithocholate, among

others) which are generated in the gut by the action of the microbiota and reabsorbed from the distal ileum through the enterohepatic circulation. BAs have been described as signaling molecules through their interaction with the farsenoid X receptor and the G-protein-coupled receptor TGR5 in the liver and adipose tissue, thus involved in the lipid and glucose homeostasis of the host [88]. Besides, the altered bile acid pool has been described as an underlying condition of various disease and disease-related states. For instance, Zhao and colleagues reported high levels of glycochenodeoxycholic acid in plasma of impaired glucose-tolerant subjects [89]; Mastrangelo et al. identified in the increased of taurodeoxycholic acid and glycochenodeoxycholic acid in serum samples of obese children a marker of IR state [90]; and Shure et al. identified an altered bile acids pool (low levels of cholic and muricholic acids and increased deoxycholic acid) in diabetic patients of the KORA cohort [91] (see Sect. 10.5.3).

Together with the influence of the gut microbiota, other novel contributors to NCDs have been uncovered by metabolomics, namely, the adenylosuccinate (S-AMP) and the beta-aminoisobutyric acid (BAIBA) that have been associated with T2D and cardiometabolic risk factors, respectively [92, 93]. While Gooding et al. unveiled a novel action of S-AMP as a glucose-derived amplifying stimulus of insulin secretion, Roberts et al. showed a fascinating effect of BAIBA (by-product of the catabolism of thymine or valine) on the browning of the white fat and on the stimulation of the beta oxidation in hepatocytes via PPAR-alpha. Briefly, Gooding et al. demonstrated the effect of glucose on the production of S-AMP (intermediated of the purine/nucleotide pathway) via the pentose phosphate pathway; they have also highlighted the stimulating action mediated by S-AMP on insulin secretion from human pancreatic beta cells upon normal and diabetic conditions, thereby showing a striking ability on rescuing the T2D-impaired secretory function in beta cells and suggesting a novel target for therapies. Roberts and colleagues instead discovered a novel effect of the BAIBA on the expression of the genes coding for brown adipocytes in murine white adipocyte and in human pluripotent stem cells during the differentiation to mature adipocytes; they also found an increase of the BAIBA during physical activity and a further inverse correlation of the BAIBA to cardiometabolic risks by suggesting new metabolic pathways related to the beneficial effect of physical activity.

10.5.3   Metabolomics in the Epidemiological Setting

Prospective studies are important tools in the epidemiological setting to investigate the etiology of a disorder; indeed they offer the possibility to study a large cohort of subjects (i.e., thousands of participants) over a period of time (usually for years) by allowing the determination of the disease outcomes from initially healthy subjects and the eventual association with lifestyle risk factors to which they are exposed. An overview of the typical prospective study design is depicted in Fig. 10.4. The major prospective studies developed in the last years to address the investigation of NCDs
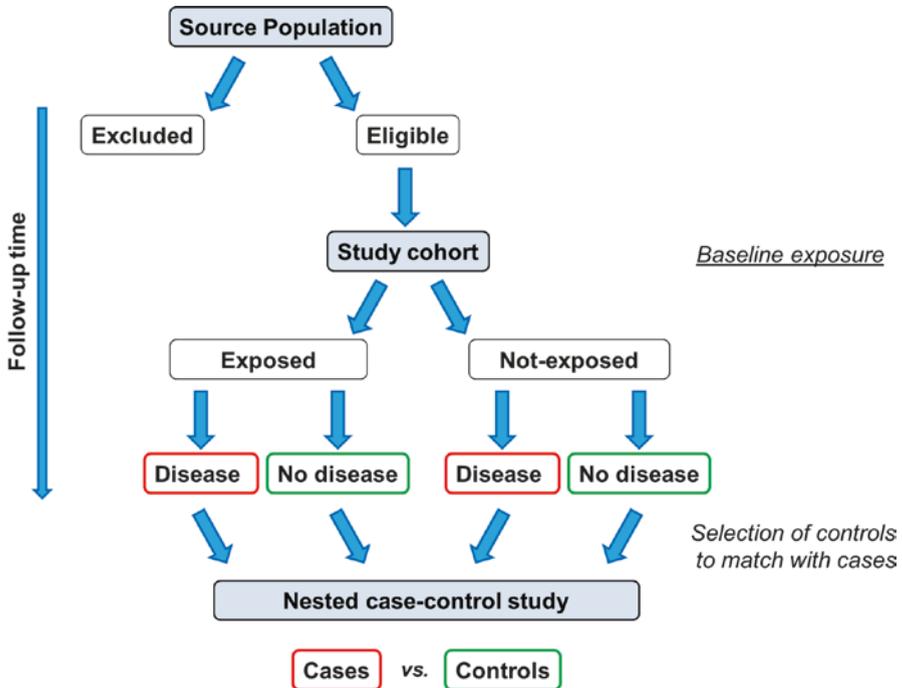
Fig. 10.4 Flowchart of a prospective study design

and lifestyle biomarkers by a metabolomics strategy are described in the following paragraphs, and their main characteristics and findings are summarized in Table 10.1 and Fig. 10.5, respectively.

The *Framingham Heart Study (FHS)* is the first longitudinal study aimed to identify the common factors that contribute to CVD. The original cohort (5,209 men and women between the ages of 30 and 62 from Framingham, Massachusetts) was recruited in 1948 and followed up every 2 years. Further cohorts were also included (the Offspring Cohort in 1971, the Third Generation Cohort in 2002, the Omni Cohort in 1994, and the Second Generation Omni Cohort in 2003), for a total of over 15,000 participants for a study that is still ongoing [105]. The study, led by the National Heart, Lung, and Blood Institute, in collaboration with Boston University, has generated a variety of graded risk scores to estimate the risk of several cardiovascular diseases 10–30 years in advance by using a sex-specific algorithm that includes smoking habits, blood pressure levels, age, and family history of CVD events, among others [106]. Over the years, new technologies have emerged and successfully employed in the investigation of the Framingham cohorts. Among the metabolomic-based studies, accurate predictors of future cardiovascular disease, diabetes, and metabolic syndrome (including obesity, dyslipidemia, and dysglycemia) were uncovered by studying the offspring and the third-generation cohorts. Concerning the risk assessment for diabetes, in 2011,

Table 10.1 Characteristics of the major prospective studies on NCDs and lifestyle biomarkers longitudinally investigated by a metabolomics approach

Reference	Cohort study	Study design	N, follow-up time	Sample type	Analytical strategy	Outcome	Validation cohort
Wang et al. [42]	FSH offspring	Nested case–control	378, 12 years	Plasma	Targeted LC-MS/MS	T2D	MCD-CC
Rhee et al. [94]	FSH offspring	Nested case–control	378, 12 years	Plasma	Targeted LC-MS/MS	T2D	–
Cheng et al. [95]	FSH offspring	Nested case–control	601, 12 years	Plasma	Targeted LC-MS/MS	T2D, CVDs	MCD-CC
Wang et al. [96]	FSH offspring	Nested case–control	376, 12 years	Plasma	Targeted LC-MS/MS	T2D	MCD-CC
Yin et al. [97]	FSH offspring and third generation	Cohort prospective	554, 5–7 years	Plasma	Untargeted GC-MS	Obesity, dyslipidemia	BioImage
Magnusson et al. [98]	MCD-CC	Nested case–control	506, 12 years	Plasma	Targeted LC-MS/MS	CVDs	–
Floegel et al. [99]	EPIC-Potsdam	Case-cohort	2282 (subcohort) and 800 (T2D cases), 7 years	Serum	Targeted FIA-MS/MS	T2D	KORA
Drogan et al. [100]	EPIC-Potsdam	Nested case–control	300, 6 years	Serum	Untargeted LC-MS	T2D	–
Jacobs et al. [101]	EPIC-Potsdam	Case-cohort	1,610 (subcohort) and 417 (T2D cases), 7 years	Serum	Targeted FIA-MS/MS	T2D and coffee consumption	–

(continued)

Table 10.1 (continued)

Reference	Cohort study	Study design	N, follow-up time	Sample type	Analytical strategy	Outcome	Validation cohort
Wittenbecher et al. [102]	EPIC-Potsdam	Case-cohort	2,681 (subcohort) and 688 (T2D cases), 7 years	Serum	Targeted FIA-MS/MS	T2D and red meat consumption	–
Wang-Sattler et al. [38]	KORA S4/F4	Case-cohort	641 (NGT) and 118 (IGT), 7 years; 876 (subcohort) and 91 (T2D), 7 years	Serum	Targeted FIA-MS/MS	Prediabetes (IGT), T2D	EPIC-Potsdam
Mook-Kanamori et al. [103]	KORA S4/F4	Case-cohort	755 (subcohort) and 121 (cases), 7 years	Serum	Targeted LC-MS/MS	Hypertriglyceridemia	–
Wahl et al. [104]	KORA S4/F4	Cohort prospective	1,658, 7 years	Serum	Untargeted LC-MS/MS[2] GC-MS NMR	Weight change	–
Wurtz et al. [37]	FINRISK	Case-cohort	7,256 (subcohort) and 800 (CVD), 15 years	Serum	Targeted NMR LC-MS	CVDs	SABRE BWHHS
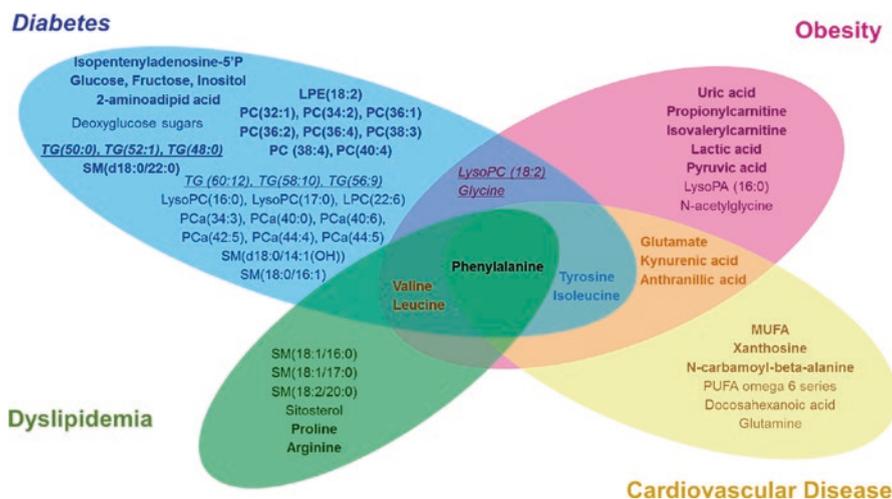
Fig. 10.5 Venn diagram illustrating the metabolites found to be associated with diabetes, cardiovascular disease, obesity, and/or dyslipidemia by prospective studies investigated by metabolomics. *Bold* metabolites were found to be increased in one of the four conditions understudy, whereas *underlined* metabolites were found to be highly associated with insulin resistance

Wang et al. described fasted levels of five amino acids (i.e., isoleucine, leucine, valine, tyrosine, and phenylalanine), at a baseline exposure, as highly associated with future onset of diabetes, particularly in predisposed subjects (i.e., obese and with high fasting glucose levels) [42]. They further uncovered that a combination of three amino acids (i.e., isoleucine, phenylalanine, and tyrosine), the so-called diabetes-predictive amino acid score (DM-AA score), predicted future diabetes up to 12 years in advance (four- to fivefold higher risk for individuals with the highest amino acids score). The findings were also replicated in an independent cohort study, the Malmö Diet and Cancer study (MCD, see below), by demonstrating their generalizability. Notably, a further link was uncovered in the same study population between the double-bond content and the carbon chain length of lipids (mainly triglycerides) and the risk of diabetes: lipids of lower/higher carbon number and double-bond content are associated to an increased/decreased risk of future diabetes (12 years in advance) [94]. In 2012, Cheng et al. identified an association between tryptophan metabolism by-products with future CVDs; besides, they confirmed the previous findings for the DM-AA score and uncovered a further metabolite (glutamine) as inversely related to future risk of diabetes [95]. In 2013, Wang et al. unveiled a further metabolite (2-aminoadipic acid, 2-AAA) as strongly associated with future diabetes (up to 12 years in advance), both in the discovery (FHS) and replication (MCD) cohorts [96]. Subsequent studies on cell-based and animal models have suggested that the 2-AAA might be involved in the stimulation of insulin secretion in pancreatic β-cells and the modulation of glucose homeostasis in vivo, respectively [96]. Finally, in 2016, Yin et al. investigated the relationship between metabolic profiles, at the baseline level,

with risk factors of the metabolic syndrome including obesity, dyslipidemia, and dysglycemia [97]. They discovered longitudinal associations between several metabolites, such as lipids [e.g., lysoPA(16:0), sphingomyelins, and sitosterol] and organic acids (e.g., quinic acid), with one or more features of the metabolic syndrome.

The *European Prospective Investigation into Cancer and Nutrition (EPIC)* is a prospective cohort with more than 521,000 study participants (men and women, between 35 and 70 years old) enrolled from 23 centers in ten Western European countries. Originally designed to explore the association between nutrition and cancer, it has included over the years the investigation of other chronic diseases such as CVDs and T2D. At the enrolment (1992–1999), detailed information on diet, lifestyle characteristics, anthropometric measurements, and medical history was collected; blood samples were also taken and stored in liquid nitrogen at the International Agency for Research on Cancer – World Health Organization [107]. Among the NCDs investigated, the association between cancer and diet has been the most studied, whereas upon the study of CVD and T2D and their risk factors, only two cohorts were used, namely, a selection of the MDC cohort, the MDC Cardiovascular Cohort (MCD-CC, 6,103 participants), and the EPIC-Potsdam cohort (27,584 participants), respectively [108, 109]. Concerning the MCD-CC, it was predominantly employed to replicate the findings of the FSH study, thereby describing a metabolic profile for diabetes and cardiovascular diseases' prediction. In 2013, the MCD-CC was also used by Magnusson et al. to investigate the predictive capability of the DM-AA score described by Wang et al. (see above) both for the onset and the consequences of a CVD event [98]. They found that the DM-AA score was able to predict CVD events (12 years in advance) by suggesting a possible link between diabetes and CVDs. Besides, a link between the amino acid score and an increased propensity toward atherosclerosis and inducible ischemia was unveiled. In the same year, Floegel et al. described for a subcohort of the EPIC-Potsdam study (2,500 selected randomly subjects and 800 T2D cases) a significant association between serum metabolites both with increased risk of T2D (e.g., hexose, phenylalanine, and diacyl-phosphatidylcholines) and decreased risk of T2D [i.e., glycine, SM(18:0/16:1), LPC(18:2), and alkyl-phosphatidylcholines] [99]. These metabolites were further included in the predictive model of the German Diabetes Risk score (i.e., ROC AUC from 0.847 to 0.912), thereby demonstrating their value as biomarkers. The results were then successfully replicated in the prospective KORA study (see below). The EPIC-Potsdam subcohort was further employed in 2015 to generate a nested case–control study for the investigation of the pathophysiology of T2D by using an untargeted approach [100]. Alteration in serum carbohydrates (e.g., hexoses), purines (e.g., isopentanyladenosine-5-monophosphate), and phospholipids [e.g., LPC(16:0)] was found to predict the onset of T2D up to 6 years in advance. Finally, the EPIC-Potsdam cohort was investigated to evaluate the effect of the specific food consumption (i.e., coffee and red meat) and the incidence of T2D [101, 102]. Sex-specific correlations were found by showing an inverse trend between coffee and T2D risk only in men and different metabolic profiles according to sex both for coffee and red meat consumption. Concerning the coffee consumption, only phenylalanine was

found to be slightly associated to T2D, whereas ferritin, glycine, and some lipids [i.e., PC(36:4), LPC(17:0), and SM(14:1)] were found to reflect both red meat consumption and increased risk for T2D.

The *multinational monitoring of trends and determinants in cardiovascular disease (MONICA)* is a WHO-funded project aimed to monitor the common risk factors (i.e., cigarette smoking, hypertension, obesity, total cholesterol) leading to CVD; a total of 38 populations and 21 countries from all over the word were included in the project, and more than ten million of men and women (25–64 years old) were surveyed (overall period covered: 1979–1996) [110]. Although the MONICA project ended in 1996, the survey on the Augsburg cohort continued and derived into the MONICA/Cooperative Health Research in the Region Augsburg (KORA) study (18,000 participants) that added the study of diabetes to the investigation of CVDs. The MONICA/KORA study comprises of four surveys (S1 to S4, from 1996 to 2001) that were performed with a 5-year interval and followed longitudinally between 4 and 20 years [111]. In 2010 the first metabolomic-based study was performed on a subset of the KORA F3 cohort (40 cases and 60 controls, males, over 54-year-old) [91]. This pilot study replicated the findings of known biomarkers of diabetes (i.e., BCAA, sugar metabolites, ketone bodies) and identified novel metabolites (i.e., 3-idroxyl sulfate, glycerophospholipids, free fatty acids, and bile acids) related to diabetes under subclinical condition. In 2012, then Wang-Sattler et al. investigated a subset of the KORA S4/F4 cohort (876 participants) by unraveling three metabolites (i.e., glycine, LPC(18:2), and acetylcarnitine) as markers of prediabetes [38]; the findings were replicated in the EPIC-Potsdam cohort by describing a role for glycine and LPC(18:2) as marker both of prediabetes and T2D. Besides, the KORA S4/F4 cohorts (S4 $n=4,261$, F4 $n=3,080$) were further investigated to explore the relationship of the metabolic profile with risk factors for T2D and CVDs including hypertriglyceridemia and obesity: Mook-Kanamori et al. highlighted increased levels of amino acids (i.e., leucine, valine, arginine, proline, and phenylalanine) as related to high levels of triglycerides both at the baseline and at 7 years follow-up [103], whereas Wahl et al. identified in dyslipidemia (altered lipoproteins and triglycerides) and modulated amino acids metabolism (mainly BCAA) the features of a potential mitochondrial dysfunction underlying long-term weight change [104]. Finally, a further prospective study that derives from the MONICA project and employs a metabolomics approach, namely, the National FINRISK Study (7,256 participants, overall period 1972–2012) [112], was explored by Wurtz et al. by identifying serum level of phenylalanine, monounsaturated fatty acids, omega-6 fatty acids, and docosahexaenoic acid as hallmarks for future CVDs. The results were further validated in two independent UK cohorts (i.e., Southall and Brent REvisited cohort and British Women's Heart and Health Study cohort) [37].

The Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) is a 3-year longitudinal study conducted at 46 centers in 12 countries. A total of 2,180 COPD patients (men and women aged 40–75, under medication) were surveyed every 3 months in order to identify predictors of the COPD progression and improve the discrimination of the COPD subtypes [113]. In 2012, Ubhi et al. employed a subset of the ECLIPSE cohort for two metabolomic-based investigations

that unraveled changes in the metabolism of several amino acids (e.g., serine, sarco-
sine, tryptophan, BCAAs, and 3-methylhistdine, among others) that enabled the strati-
fication of COPD patients (i.e., smoker vs. nonsmokers, patients with and without
emphysema or with and without cachexia) [73, 114]. Although these findings pro-
vided valuable information in the partially/merely explored field of respiratory chronic
diseases, none of the metabolomic-based studies investigated the ECLIPSE cohort
longitudinally, and markers of disease progression and patient's outcome are still
lacking to date.

10.6   Concluding Remarks

Through this chapter we have highlighted the versatility and the striking potential of
metabolomics to provide new advances in the field of chronic diseases: disease-
related metabotypes were described; crucial players involved in the NCDs are
unveiled; and finally, a long-term perspective on the disease's progression was
pointed out. Even though several limitations still need to be addressed (i.e., the
improvement of the metabolite identification, the exploiting of the synergies
between different omics, and the effective use of metabolomics in clinical practice,
among others), the metabolic signature of diseases that is revealed by the study of
NCDs is a clear demonstration of the importance of this discipline, not only for
NCDs but also in the wider context of the human health. The metabolic alterations
are indeed potentially detectable, understandable, and ultimately treatable by a
metabolomic-based strategy that thus holds the promise to drive a paradigm shift
toward the tailoring of the therapy on the altered metabolic pathways rather than on
the disease's symptomatology.

References

1. WHO. Noncommunicable diseases 2015. Available from: http://www.who.int/mediacentre/
   factsheets/fs355/en/. Accessed May 2016.
2. WHO. Global status report on noncommunicable diseases 2014. 2014. Available from: http://
   www.who.int/nmh/publications/ncd-status-report-2014/en/. Accessed May 2016.
3. Nugent R. Chronic diseases in developing countries: health and economic burdens. Ann N Y
   Acad Sci. 2008;1136:70–9.
4. National Center for Chronic Disease Prevention and Health Promotion. Chronic Disease
   Overview 2016. Available from: http://www.cdc.gov/chronicdisease/overview/. Accessed May
   2016.
5. WHO. Cardiovascular diseases 2016. Available from: http://www.euro.who.int/en/health-
   topics/noncommunicable-diseases/cardiovascular-diseases. Accessed May 2016.
6. WHO. Prevention of Recurrences of Myocardial Infarction and Stroke Study 2016. Available
   from: http://www.who.int/cardiovascular_diseases/priorities/secondary_prevention/country/
   en/index1.html. Accessed May 2016.
7. WHO. Chronic respiratory diseases 2016. Available from: http://www.who.int/respiratory/en/.
   Accessed May 2016.

8. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006;3(11):e442.
9. Centers for disease control and prevention. Child Health 2016. Available from: http://www.cdc.gov/nchs/fastats/child-health.htm. Accessed May 2016.
10. Pauwels RA, Rabe KF. Burden and clinical features of chronic obstructive pulmonary disease (COPD). Lancet. 2004;364(9434):613–20.
11. WHO. Global surveillance, prevention and control of chronic respiratory diseases. A comprehensive approach 2007. Available from: http://www.who.int/respiratory/publications/global_surveillance/en/. Accessed May 2016.
12. WHO. Diabetes 2016. Available from: http://www.who.int/mediacentre/factsheets/fs312/en/. Accessed May 2016.
13. WHO. Global report on diabetes 2016. Available from: http://www.who.int/diabetes/global-report/en/. Accessed May 2016.
14. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med. 1998;15(7):539–53.
15. WHO. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy 2013. Available from: http://www.who.int/diabetes/publications/Hyperglycaemia_In_Pregnancy/en/. Accessed May 2016.
16. van Belle TL, Coppieters KT, von Herrath MG. Type 1 diabetes: etiology, immunology, and therapeutic strategies. Physiol Rev. 2011;91(1):79–118.
17. WHO. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus 2011. Available from: http://www.who.int/diabetes/publications/diagnosis_diabetes2011/en/. Accessed May 2016.
18. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care. 2003;26(3):725–31.
19. WHO. Chronic diseases and their common risk factors 2016. Available from: http://www.who.int/chp/chronic_disease_report/information_sheets/en/. Accessed May 2016.
20. WHO. Obesity and overweight 2015. Available from: http://www.who.int/mediacentre/factsheets/fs311/en/. Accessed May 2016.
21. WHO. BMI classification 2016. Available from: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html. Accessed May 2016.
22. WHO. Waist circumference and waist–hip ratio. Report of a WHO expert consultation, Geneva, 8–11 December 2008. 2011. Available from: http://www.who.int/nutrition/publications/obesity/WHO_report_waistcircumference_and_waisthip_ratio/en/. Accessed May 2016.
23. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. N Engl J Med. 1997;337(13):869–73.
24. Skinner AC, Perrin EM, Moss LA, Skelton JA. Cardiometabolic risks and severity of obesity in children and young adults. N Engl J Med. 2015;373(14):1307–17.
25. Daniels SR. The consequences of childhood overweight and obesity. Future Child. 2006;16(1):47–67.
26. Barker DJ. Fetal origins of coronary heart disease. BMJ. 1995;311(6998):171–4.
27. WHO. Programming of chronic disease by impaired fetal nutrition. Evidence and implications for policy and intervention strategies Geneva 2002. Available from: http://www.who.int/nutrition/publications/obesity/WHO_NHD_02.3/en/. Accessed May 2016.
28. Godfrey KM, Barker DJ. Fetal nutrition and adult disease. Am J Clin Nutr. 2000;71(5 Suppl):1344s–52.
29. Wahlqvist ML, Krawetz SA, Rizzo NS, Dominguez-Bello MG, Szymanski LM, Barkin S, et al. Early-life influences on obesity: from preconception to adolescence. Ann N Y Acad Sci. 2015;1347:1–28.
30. Eaton SB, Konner M, Shostak M. Stone agers in the fast lane: chronic degenerative diseases in evolutionary perspective. Am J Med. 1988;84(4):739–49.
31. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181–9.

32. Fiehn O. Metabolomics–the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1–2):155–71.

33. German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. Metabolomics. 2005;1(1):3–9.

34. Hivert MF, Perng W, Watkins SM, Newgard CS, Kenny LC, Kristal BS, et al. Metabolomics in the developmental origins of obesity and its cardiometabolic consequences. J Dev Orig Health Dis. 2015;6(2):65–78.

35. Du F, Virtue A, Wang H, Yang XF. Metabolomic analyses for atherosclerosis, diabetes, and obesity. Biomark Res. 2013;1(1):17.

36. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab. United States. 2009;9:311–26.

37. Wurtz P, Makinen VP, Soininen P, Kangas AJ, Tukiainen T, Kettunen J, et al. Metabolic signatures of insulin resistance in 7,098 young adults. Diabetes. 2012;61(6):1372–80.

38. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol. 2012;8:615.

39. Nobakht MGBF, Aliannejad R, Rezaei-Tavirani M, Taheri S, Oskouie AA. The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. Biomarkers. 2015;20(1):5–16.

40. Klein MS, Shearer J. Metabolomics and type 2 diabetes: translating basic research into clinical application. J Diabetes Res. 2016;2016:3898502.

41. Roberts LD, Gerszten RE. Toward new biomarkers of cardiometabolic diseases. Cell Metab. 2013;18(1):43–50.

42. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. United States. 2011;17:448–53.

43. Parikh NI, Vasan RS. Assessing the clinical utility of biomarkers in medicine. Biomark Med. 2007;1(3):419–36.

44. Roe CR, Millington DS, Maltby DA. Identification of 3-methylglutarylcarnitine. A new diagnostic metabolite of 3-hydroxy-3-methylglutaryl-coenzyme A lyase deficiency. J Clin Invest. 1986;77(4):1391–4.

45. Jellum E, Kvittingen EA, Stokke O. Mass spectrometry in diagnosis of metabolic disorders. Biomed Environ Mass Spectrom. 1988;16(1–12):57–62.

46. Kim KR, Park HG, Paik MJ, Ryu HS, Oh KS, Myung SW, et al. Gas chromatographic profiling and pattern recognition analysis of urinary organic acids from uterine myoma patients and cervical cancer patients. J Chromatogr B Biomed Sci Appl. 1998;712(1–2):11–22.

47. Kimura M, Yamamoto T, Yamaguchi S. Automated metabolic profiling and interpretation of GC/MS data for organic acidemia screening: a personal computer-based system. Tohoku J Exp Med. 1999;188(4):317–34.

48. Zhang A, Sun H, Wang X. Power of metabolomics in biomarker discovery and mining mechanisms of obesity. Obes Rev. 2013;14(4):344–9.

49. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.

50. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med. 2004;23(13):2109–23.

51. Buijsse B, Simmons RK, Griffin SJ, Schulze MB. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. Epidemiol Rev. 2011;33:46–62.

52. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med. 2008;149(10):751–60.

53. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med. 2009;150(11):795–802.

54. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008;134(5):714–7.

55. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007;26(1):51–78.

56. Kim JY, Park JY, Kim OY, Ham BM, Kim HJ, Kwon DY, et al. Metabolic profiling of plasma in overweight/obese and lean men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC-Q-TOF MS). J Proteome Res. 2010;9(9):4368–75.

57. Mihalik SJ, Goodpaster BH, Kelley DE, Chace DH, Vockley J, Toledo FG, et al. Increased levels of plasma acylcarnitines in obesity and type 2 diabetes and identification of a marker of glucolipotoxicity. Obesity (Silver Spring). 2010;18(9):1695–700.

58. She P, Van Horn C, Reid T, Hutson SM, Cooney RN, Lynch CJ. Obesity-related elevations in plasma leucine are associated with alterations in enzymes involved in branched-chain amino acid metabolism. Am J Physiol Endocrinol Metab. 2007;293(6):E1552–63.

59. Laferrere B, Reilly D, Arias S, Swerdlow N, Gorroochurn P, Bawa B, et al. Differential metabolic impact of gastric bypass surgery versus dietary intervention in obese diabetic subjects despite identical weight loss. Sci Transl Med. United States. 2011;3:80re2.

60. Shah SH, Crosslin DR, Haynes CS, Nelson S, Turer CB, Stevens RD, et al. Branched-chain amino acid levels are associated with improvement in insulin resistance with weight loss. Diabetologia. 2012;55(2):321–30.

61. Halvatsiotis PG, Turk D, Alzaid A, Dinneen S, Rizza RA, Nair KS. Insulin effect on leucine kinetics in type 2 diabetes mellitus. Diabetes Nutr Metab. 2002;15(3):136–42.

62. Tessari P, Coracina A, Kiwanuka E, Vedovato M, Vettore M, Valerio A, et al. Effects of insulin on methionine and homocysteine kinetics in type 2 diabetes with nephropathy. Diabetes. 2005;54(10):2968–76.

63. Tai ES, Tan ML, Stevens RD, Low YL, Muehlbauer MJ, Goh DL, et al. Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men. Diabetologia. 2010;53(4):757–67.

64. Shin AC, Fasshauer M, Filatova N, Grundell LA, Zielinski E, Zhou JY, et al. Brain insulin lowers circulating BCAA levels by inducing hepatic BCAA catabolism. Cell Metab. 2014;20(5):898–909.

65. Lynch CJ, Adams SH. Branched-chain amino acids in metabolic signalling and insulin resistance. Nat Rev Endocrinol. England. 2014;10:723–36.

66. Newgard CB. Interplay between lipids and branched-chain amino acids in development of insulin resistance. Cell Metab. United States, 2012 Elsevier Inc. 2012;15:606–14.

67. Mihalik SJ, Michaliszyn SF, de las Heras J, Bacha F, Lee S, Chace DH, et al. Metabolomic profiling of fatty acid and amino acid metabolism in youth with obesity and type 2 diabetes: evidence for enhanced mitochondrial oxidation. Diabetes Care. 2012;35(3):605–11.

68. Wang Z, Tang WH, Cho L, Brennan DM, Hazen SL. Targeted metabolomic evaluation of arginine methylation and cardiovascular risks: potential mechanisms beyond nitric oxide synthase inhibition. Arterioscler Thromb Vasc Biol. 2009;29(9):1383–91.

69. Wang L, Hou E, Wang Y, Yang L, Zheng X, Xie G, et al. Reconstruction and analysis of correlation networks based on GC-MS metabolomics data for young hypertensive men. Anal Chim Acta. 2015;854:95–105.

70. Wedes SH, Wu W, Comhair SA, McDowell KM, DiDonato JA, Erzurum SC, et al. Urinary bromotyrosine measures asthma control and predicts asthma exacerbations in children. J Pediatr. 2011;159(2):248–55.e1.

71. Jung J, Kim SH, Lee HS, Choi GS, Jung YS, Ryu DH, et al. Serum metabolomics reveals pathways and biomarkers associated with asthma pathogenesis. Clin Exp Allergy. 2013;43(4):425–33.

72. Kutsuzawa T, Shioya S, Kurita D, Haida M. Plasma branched-chain amino acid levels and muscle energy metabolism in patients with chronic obstructive pulmonary disease. Clin Nutr. 2009;28(2):203–8.

73. Ubhi BK, Cheng KK, Dong J, Janowitz T, Jodrell D, Tal-Singer R, et al. Targeted metabolomics identifies perturbations in amino acid metabolism that sub-classify patients with COPD. Mol Biosyst. 2012;8(12):3125–33.

74. Wang L, Tang Y, Liu S, Mao S, Ling Y, Liu D, et al. Metabonomic profiling of serum and urine by (1)H NMR-based spectroscopy discriminates patients with chronic obstructive pulmonary disease and healthy individuals. PLoS One. 2013;8(6):e65675.

75. Saude EJ, Skappak CD, Regush S, Cook K, Ben-Zvi A, Becker A, et al. Metabolomic profiling of asthma: diagnostic utility of urine nuclear magnetic resonance spectroscopy. J Allergy Clin Immunol. 2011;127(3):757–64.e1–6.

76. Ha CY, Kim JY, Paik JK, Kim OY, Paik YH, Lee EJ, et al. The association of specific metabolites of lipid metabolism with markers of oxidative stress, inflammation and arterial stiffness in men with newly diagnosed type 2 diabetes. Clin Endocrinol (Oxf). 2012;76(5):674–82.

77. Kim M, Jung S, Kim SY, Lee SH, Lee JH. Prehypertension-associated elevation in circulating lysophosphatidylcholines, Lp-PLA2 activity, and oxidative stress. PLoS One. 2014;9(5):e96735.

78. Yang B, Ding F, Wang FL, Yan J, Ye XW, Yu W, et al. Association of serum fatty acid and estimated desaturase activity with hypertension in middle-aged and elderly Chinese population. Sci Rep. 2016;6:23446.

79. Xu F, Tavintharan S, Sum CF, Woon K, Lim SC, Ong CN. Metabolic signature shift in type 2 diabetes mellitus revealed by mass spectrometry-based metabolomics. J Clin Endocrinol Metab. 2013;98(6):E1060–5.

80. Du Z, Shen A, Huang Y, Su L, Lai W, Wang P, et al. 1H-NMR-based metabolic analysis of human serum reveals novel markers of myocardial energy expenditure in heart failure patients. PLoS One. 2014;9(2):e88102.

81. Yore MM, Syed I, Moraes-Vieira PM, Zhang T, Herman MA, Homan EA, et al. Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. Cell. 2014;159(2):318–32.

82. Claesson MJ, Cusack S, O'Sullivan O, Greene-Diniz R, de Weerd H, Flannery E, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4586–91.

83. Griffin JL, Wang X, Stanley E. Does our gut microbiome predict cardiovascular risk? A review of the evidence from metabolomics. Circ Cardiovasc Genet. 2015;8(1):187–91.

84. Boulange CL, Neves AL, Chilloux J, Nicholson JK, Dumas ME. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. Genome Med. 2016;8(1):42.

85. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027–31.

86. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. Science. 2012;336(6086):1262–7.

87. Salek RM, Maguire ML, Bentley E, Rubtsov DV, Hough T, Cheeseman M, et al. A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. Physiol Genomics. 2007;29(2):99–108.

88. Wahlstrom A, Sayin SI, Marschall HU, Backhed F. Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. Cell Metab. 2016;24(1):41–50.

89. Zhao X, Fritsche J, Wang J, Chen J, Rittig K, Schmitt-Kopplin P, et al. Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. Metabolomics. 2010;6(3):362–74.

90. Mastrangelo A, Martos-Moreno GA, Garcia A, Barrios V, Ruperez FJ, Chowen JA, et al. Insulin resistance in prepubertal obese children correlates with sex-dependent early onset metabolomic alterations. Int J Obes (Lond). 2016;40(10):1494–502.

91. 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(11):e13953.

92. Gooding JR, Jensen MV, Dai X, Wenner BR, Lu D, Arumugam R, et al. Adenylosuccinate is an insulin secretagogue derived from glucose-induced purine metabolism. Cell Rep. 2015;13(1):157–67.

93. Roberts LD, Bostrom P, O'Sullivan JF, Schinzel RT, Lewis GD, Dejam A, et al. beta-Aminoisobutyric acid induces browning of white fat and hepatic beta-oxidation and is inversely correlated with cardiometabolic risk factors. Cell Metab. 2014;19(1):96–108.

94. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest. 2011;121(4):1402–11.
95. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation. 2012;125(18):2222–31.
96. Wang TJ, Ngo D, Psychogios N, Dejam A, Larson MG, Vasan RS, et al. 2-Aminoadipic acid is a biomarker for diabetes risk. J Clin Invest. 2013;123(10):4309–17.
97. Yin X, Subramanian S, Willinger CM, Chen G, Juhasz P, Courchesne P, et al. Metabolite signatures of metabolic risk factors and their longitudinal changes. J Clin Endocrinol Metab. 2016;101(4):1779–89.
98. Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engstrom G, et al. A diabetes-predictive amino acid score and future cardiovascular disease. Eur Heart J. 2013;34(26):1982–9.
99. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. 2013;62(2):639–48.
100. Drogan D, Dunn WB, Lin W, Buijsse B, Schulze MB, Langenberg C, et al. Untargeted metabolic profiling identifies altered serum metabolites of type 2 diabetes mellitus in a prospective, nested case control study. Clin Chem. 2015;61(3):487–97.
101. Jacobs S, Kroger J, Floegel A, Boeing H, Drogan D, Pischon T, et al. Evaluation of various biomarkers as potential mediators of the association between coffee consumption and incident type 2 diabetes in the EPIC-Potsdam Study. Am J Clin Nutr. 2014;100(3):891–900.
102. Wittenbecher C, Muhlenbruch K, Kroger J, Jacobs S, Kuxhaus O, Floegel A, et al. Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. Am J Clin Nutr. 2015;101(6):1241–50.
103. Mook-Kanamori DO, Romisch-Margl W, Kastenmuller G, Prehn C, Petersen AK, Illig T, et al. Increased amino acids levels and the risk of developing of hypertriglyceridemia in a 7-year follow-up. J Endocrinol Invest. 2014;37(4):369–74.
104. Wahl S, Vogt S, Stuckler F, Krumsiek J, Bartel J, Kacprowski T, et al. Multi-omic signature of body weight change: results from a population-based cohort study. BMC Med. 2015;13:48.
105. Tsao CW, Vasan RS. Cohort Profile: The framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol. 2015;44(6):1800–13.
106. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014;383(9921):999–1008.
107. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002;5(6b):1113–24.
108. Persson M, Hedblad B, Nelson JJ, Berglund G. Elevated Lp-PLA2 levels add prognostic information to the metabolic syndrome on incidence of cardiovascular events among middle-aged nondiabetic subjects. Arterioscler Thromb Vasc Biol. 2007;27(6):1411–6.
109. Boeing H, Wahrendorf J, Becker N. EPIC-Germany–A source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. Ann Nutr Metab. 1999;43(4):195–204.
110. Evans A, Tolonen H, Hense HW, Ferrario M, Sans S, Kuulasmaa K. Trends in coronary risk factors in the WHO MONICA project. Int J Epidemiol. 2001;30 Suppl 1:S35–40.
111. Holle R, Happich M, Lowel H, Wichmann HE. KORA-a research platform for population based health research. Gesundheitswesen. 2005;67 Suppl 1:S19–25.
112. Borodulin K, Vartiainen E, Peltonen M, Jousilahti P, Juolevi A, Laatikainen T, et al. Forty-year trends in cardiovascular risk factors in Finland. Eur J Public Health. 2015;25(3):539–46.
113. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). Eur Respir J. 2008;31(4):869–73.
114. Ubhi BK, Riley JH, Shaw PA, Lomas DA, Tal-Singer R, MacNee W, et al. Metabolic profiling detects biomarkers of protein degradation in COPD patients. Eur Respir J. 2012;40(2):345–55.

Chapter 11
Lipidomics, Biomarkers, and Schizophrenia: A Current Perspective

Sumit Sethi, Mirian A.F. Hayashi, Banny S. Barbosa, João G.M. Pontes, Ljubica Tasic, and Elisa Brietzke

Abstract Lipidomics is a lipid-targeted metabolomics approach aiming at comprehensive analysis of lipids in biological systems. Recent technological progresses in mass spectrometry, nuclear magnetic resonance spectroscopy, and chromatography have significantly enhanced the developments and applications of metabolic profiling of lipids in more complex biological samples. As many diseases reveal a notable change in lipid profiles compared with that of healthy people, lipidomics have also been broadly introduced to scientific research on diseases. Exploration of lipid biochemistry by lipidomics approach will not only provide insights into specific roles of lipid molecular species in health and disease, but it will also support the identification of potential biomarkers for establishing preventive or therapeutic approaches for human health. This chapter aims to illustrate how lipidomics can contribute for understanding the biological mechanisms inherent to schizophrenia and why lipids are relevant biomarkers of schizophrenia. The application of lipidomics in clinical studies has the potential to provide new insights into lipid profiling and pathophysiological mechanisms underlying schizophrenia. The future perspectives of lipidomics in mental disorders are also discussed herein.

Keywords Biomarkers • Liquid Chromatography • Lipidomics • Mass Spectrometry • Nuclear Magnetic Resonance Spectroscopy • Metabolomics • Schizophrenia

S. Sethi • E. Brietzke (✉)
Department of Psychiatry, Federal University of São Paulo (UNIFESP),
Research Building II, Rua Pedro de Toledo 669-3, Floor Funds – Vila Clementino, 04039-032
São Paulo, Brazil
e-mail: elisabrietzke@hotmail.com

M.A.F. Hayashi
Department of Pharmacology, Federal University of São Paulo (UNIFESP), São Paulo, Brazil

B.S. Barbosa • J.G.M. Pontes • L. Tasic
Department of Organic Chemistry, Institute of Chemistry, University of Campinas
(UNICAMP), Campinas, Brazil

Abbreviations

AA	Arachidonic acid
APCI	Atmospheric pressure chemical ionization
BD	Bipolar disorder
BMI	Body mass index
CE	Cholesteryl ester
Cer	Ceramide
CNS	Central nervous system
COX	Cyclooxygenase
DG	Diacylglycerol
DHA	Docosahexaenoic acid
ELSD	Evaporative light-scattering detector
ESI	Electrospray ionization
FA	Fatty acyl
FFA	Free fatty acid
FID	Flame ionization detector
FTICR	Fourier transform ion cyclotron resonance
GC	Gas chromatography
GL	Glycerolipid
GP	Glycerophospholipid
GPA	Glycerophosphatidic acid
HDL	High-density lipoprotein
hexCer	Monohexosylceramide
HNE	4-Hydroxynonenal
HPLC	High-performance liquid chromatography
IM-MS	Ion mobility-mass spectrometry
LDL	Low-density lipoprotein
LOX	Lipoxygenase
LPC	Lysophosphatidylcholine
LPE	Lysophosphatidylethanolamine
LPO	Lipid peroxidation
MALDI	Matrix-assisted laser desorption/ionization
MS	Mass spectrometry
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NAPS	N-acyl-phosphatidylserine
NMR	Nuclear magnetic resonance
NPLC	Normal-phase liquid chromatography
PA	Phosphatidic acid
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphoglycerol
PI	Phosphatidylinositol

PK   Polyketide
Pl   Plasmalogen
PL   Phospholipid
PLA$_2$ Phospholipase A$_2$ PS: Phosphatidylserine
PR   Prenol lipid
PS   Phosphatidylserine
PUFA   Polyunsaturated fatty acid
Q   Quadrupole
RBC   Red blood cell
ROS   Reactive oxygen species
S1P   Sphingosine-1-phosphate
SCZ   Schizophrenia
SL   Saccharolipid
SM   Sphingomyelin
SP   Sphingolipid
SPE   Solid-phase extraction
ST   Sterol lipid
TG   Triacylglycerol
TLC   Thin-layer chromatography
TOF   Time of flight
UPLC   Ultra-performance liquid chromatography
VLDL   Very low-density lipoprotein

11.1   Introduction

With the progress of "omics," lipidomics, a branch of metabolomics, was first put forward by Han and Gross [1]. Lipidomics aims to characterize and quantify the range of intact lipid molecules in cells and biological fluids, allowing to correlate the lipid compositions to genomics, proteomics, diet, and diseases. The amount of genomic and proteomic data is greater than that in the lipidomics field, because of the complex nature of lipids and the limitations of tools available for such investigations. The key revolution that has incited advances in lipid analysis in the recent years was the development of new mass spectrometry techniques, particularly the "soft ionization" techniques, as the electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). Such developments that provided a high sensitivity and specificity, excellent mass, and chromatographic resolutions, which in addition to an increased accessibility to authentic synthetic lipid standards, coupled to the remarkable developments in data and bioinformatics analysis, have facilitated the analysis of a wide diversity of lipids, ranging from phospholipids (PLs) and triacylglycerols (TGs) to sterols and glycolipids [2]. Lipid metabolism may be of particular importance for the central nervous system (CNS), due to its characteristic high concentration of lipids. The complexity of such analysis is

highlighted by the recent characterization of over 500 different lipid species in a collective human serum sample conducted by the LIPID MAPS consortium (www.lipidmaps.org) [3].

The critical role of lipids in cell signaling and tissue physiology is demonstrated by the many neurological disorders, including bipolar disorder (BD) and schizophrenia (SCZ), and neurodegenerative diseases such as Alzheimer's, Parkinson's, and Niemann-Pick diseases, which present all deregulated lipid metabolism [4]. However, little is known about the molecular mechanisms that are altered in changes of states such as relapse and remission in mental illness patients. Lipidomics can be used in the search for biomarkers for specific diseases. Lipid-based biomarkers offer new prospects for precision medicine by providing sensitive diagnostic tools for disease forecast and monitoring.

In this book chapter, we describe the lipidomics approaches, summarize promising biomarkers reported in SCZ, and conclude with commentaries on the future contribution of the lipidomics approach within the larger biomarker discovery framework currently employed in the field of SCZ.

11.2   Lipid Classification

Lipids are water-insoluble compounds due to their hydrophobic features. The International Lipid Classification and the Nomenclature Committee, together with the Lipids Metabolites and Pathways Strategy (LIPID MAPS) Consortium, defined eight groups of lipids and divided them into classes and subclasses [2]. They classified the lipids by their chemically functional backbones and biochemical principles in:

1. Fatty acyls (FAs): fatty acids and conjugates, octadecanoids, eicosanoids, docosanoids, and fatty alcohols
2. Glycerolipids (GLs): monoradylglycerols, diradylglycerols, and triradyglycerols
3. Glycerophospholipids (GPs): glycerophosphocholines, glycerophosphoglycerols, glycerophosphoethanolamines, glycerophosphoglycerophosphates, glycerophosphoserines, and glycerophosphoinositols
4. Sphingolipids (SPs): sphingoid bases, ceramides, phosphosphingolipids, neutral glycosphingolipids, and acidic glycosphingolipids
5. Sterol lipids (ST): sterols
6. Prenol lipids (PR): isoprenoids
7. Saccharolipids (SL): acrylamide sugars
8. Polyketides (PK): linear polyketides

Fatty acids may be saturated, monounsaturated, or polyunsaturated. In animals, the residues of predominant fatty acids are the ones with a 16 or 18 carbon atoms chain – the palmitic and the stearic acids, which are saturated; oleic acid ($C^{18}\Delta^9$) and linoleic acid ($C^{18}\Delta^{9,12}$), which are unsaturated. The linolenic ($C^{18}\Delta^{9,12,15}$) and linoleic

acids form arachidonic, eicosapentaenoic, and docosahexaenoic acids and are essential fatty acids. TGs are the most important way to store energy in the organism, consisting of deposits in the adipose and muscle tissues.

11.3   The Molecular Biology of Lipids

Lipids are metabolites that play a significant role in different metabolic pathways. They are structural components of the cellular membranes, in which protein complexes, such as ion channels, receptors, and scaffolding complexes, are embedded, whether as a substratum, a product, or as a cofactor of biochemical reactions within a cell. Many of the biologically applicable lipids aggregate into macromolecular assemblies such as micelles or bilayers in aqueous environments of the human body. These aggregates comprise proteins, of course, in addition to lipid (e.g., apolipoproteins). Typically, the polar ends of lipid molecules face the aqueous milieu, whereas the more nonpolar fatty acyl moieties form a hydrophobic core (Fig. 11.1). Such aggregates therefore have unique biophysical properties and surface chemistries, which have important significances for the mechanisms of lipid function.
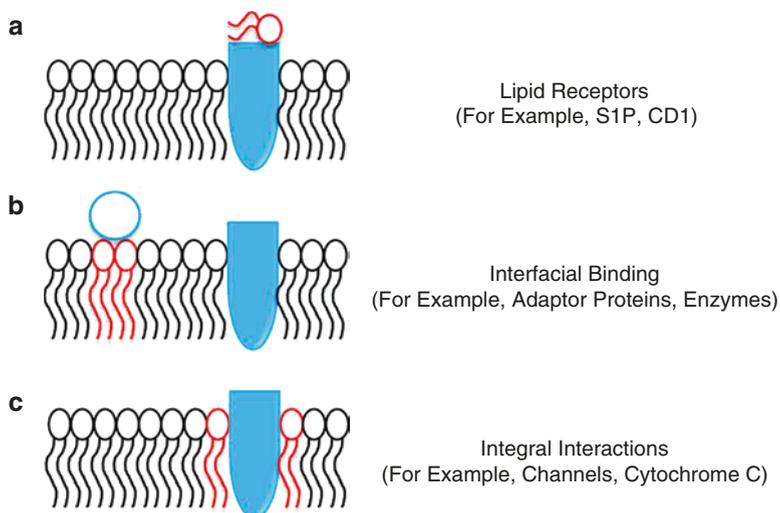
Fig. 11.1 Monolayers of lipid and protein arrangements: (**a**) binding of lipid ligands by CD1 protein occurs via the hydrophobic acyl chains of the lipid molecules; (**b**) protein modules that specifically interact with lipid head groups. Interfacial binding is a significant mode of interaction for many lipid enzymes, as well as effectors of lipids. It is directed by electrostatic interactions at the interfacial region and the characteristics of lipid head groups; (**c**) the interior portion of a lipid assembly (e.g., the bilayer interior) contributes with interactions that arise from the hydrophobic parts of lipids molecules that have a role in the regulation of membrane channels

The finding that membrane lipids can act as precursors for second messengers has added to a dramatically different view of lipid action however. Phospholipase-mediated hydrolysis of phosphatidylinositol 4,5-bisphosphate (PI(4,5)P_2) produces a variety of second messengers, such as diacylglycerol (DG), inositol 1,4,5-trisphosphate (IP(1,4,5)P_3), and arachidonic acid (AA), which by themselves are precursors of biologically active molecules. DG is rapidly phosphorylated to phosphatidic acid (PA) (glycerophosphatidic acid, GPA), which is an important intermediate in PL biosynthesis and a potent regulator of enzyme function and bilayer structure. IP(1,4,5)P_3 is metabolized by complicated enzymatic machinery, which leads to the generation of many different polyphosphorylated inositols. AA is the precursor for eicosanoids, which have an important and well-recognized role in inflammatory processes.

SPs are another example of highly bioactive membrane lipids and, like in the case of PI(4,5)P_2, various components of their structures exert different activities (Fig. 11.2). The ceramide backbone is found in many complex glycolipids, and ceramides are potent regulators of cellular growth and death. Sphingosine and its phosphorylated derivative, sphingosine-1-phosphate, control the migration of immune cells and act via binding to specific receptors.

Furthermore, and in addition to their role as precursors, it is now clear that many membrane lipids act as signaling components themselves. Phosphoinositides (PIs; phosphorylated derivatives of phosphatidylinositol, GPIns), of which PI(4,5)P_2 is a prominent representative, are an important class of such signaling lipids and contribute to a wide variety of cellular processes, including calcium homeostasis, membrane trafficking, and cytoskeletal dynamics. Indeed, it is becoming increasingly evident that understanding the biology of lipids often relates to understanding the responses that are mediated via lipids.

11.4   Lipids and the Central Nervous System (CNS)

A large number of diseases and neurological disorders in which lipid metabolism is altered confirm the crucial role of lipids in cell signaling and tissue physiology. Lipid metabolism may be of precise significance for the nervous system, as this organ has the second highest concentration of lipids, only after the adipose tissues. As mentioned, many neurological disorders, including BD and SCZ, and neurodegenerative diseases, such as Alzheimer's, Parkinson's, and Niemann-Pick diseases, involve deregulated lipid metabolism [4]. Altered lipid metabolism is also supposed to be an important event that contributes to CNS injury.

SCZ is noticeable by disturbances in thinking, emotional reactions, and social behavior, with delusions and hallucinations. Drugs that block dopamine receptors alleviate symptoms of SCZ, indicative of surplus dopaminergic function, while agents that block glutamate receptors induce some of the symptoms of SCZ in normal persons.

Fig. 11.2 Lipid classes and prominent representatives

Current concepts on the neurological deficits of SCZ have focused on aberrations in PL metabolism, mainly increased activity of phospholipase A_2 (PLA_2) enzymes, and reduced activity of the system which integrates polyunsaturated fatty acids (PUFAs) into PLs (a simultaneous increase in PL hydrolysis and

decrease in synthesis). Neither aberration alone produces SCZ but the presence of both does. These aberrations lead to changes in membrane structure and thus in the function of membrane-bound proteins, accessibility of cell signaling molecules, and in neurotransmitter systems. This assumption is reinforced by animal studies representing that application of PLA$_2$ into the brain produces alterations in the dopamine system [5]. Furthermore, since PL metabolism has a crucial role in neuronal and synaptic growth and remodeling, it is believable that defects in this system result in failure of normal neurodevelopment in SCZ. There is also proof that SCZ is related with changes in lipid transport proteins and membrane PL composition (increase in phosphatidylserine (PS) and decrease in PC and PE) [6]. Genome studies have found that numerous genes involved in myelination have decreased expression levels in SCZ. Recently, significantly increased myelin basic protein (MBP) expression was observed in first-episode psychosis patients compared to sex and age paired healthy controls [7].

11.5   Reactive Oxygen Species (ROS) and Lipid Peroxidation (LPO)

ROS, including superoxide anion radical and hydrogen peroxide, are produced by a number of cellular oxidative metabolic processes, including oxidative phosphorylation by the mitochondrial respiratory chain, which includes the xanthine oxidase, NAD(P)H oxidases, monoamine oxidases and metabolism of AA by cyclooxygenases/lipoxygenases (COX/LOX) [8]. Though current literature proposes that COX does not directly produce ROS during AA oxidative metabolism, COX does form free radicals (i.e., carbon-centered radicals on AA). There are numerous reports in the literature on ROS production by COX, and this is possibly due to a secondary ROS generation induced by several eicosanoids. Interruption of mitochondria during COX-2-associated apoptosis is a probable source of ROS production, as it has been established for a number of different cells [4, 9]. ROS then cause oxidative damage to nucleic acids, proteins, carbohydrates, and lipids. Beyond impairment to membranes, lipid peroxides give rise to reactive α,β-unsaturated aldehydes including malondialdehyde, 4-hydroxynonenal (HNE), and acrolein. These aldehydes covalently bind to proteins through reaction with thiol groups and modify their function. Although there are intracellular defenses against ROS, increased production of ROS or loss of antioxidant defenses leads to progressive cell damage and decline in physiological function. The "oxidative stress" results when generation of ROS exceeds the cell's capacity to detoxify them. The brain is believed to be particularly vulnerable to oxidative stress as it comprises high concentrations of PUFAs that are vulnerable to LPO, consumes relatively large amounts of oxygen for energy production, and has lower antioxidant defenses compared to other organs [8].

11.6 Systems-Level Approaches for Lipidomics

Traditional approaches for lipid analysis typically pre-fractionate lipids into classes using thin-layer chromatography (TLC), normal-phase liquid chromatography (NPLC), or solid-phase extraction (SPE), and then, distinct specific classes of lipids are fractionated into individual molecular species by high-performance liquid chromatography (HPLC) coupled with either ultraviolet (UV) or evaporative light-scattering (ELS) detector. With these traditional approaches, individual molecular species of many lipid classes can be evaluated. However, such "classical" techniques often lack sensitivity or need large sample volumes and multistep procedures for sample preparation; in addition, the resolution is inadequate, i.e., only a limited set of individual molecular species are evaluated. Alternatively, gas chromatography (GC) has been, and is still, frequently used for lipid analysis, although usually involving time-consuming procedures, which consist of hydrolysis and derivatization steps that are essential as most lipids are not GC amenable otherwise. Proper GC-based techniques meet the requirement of lipidomics with regards to the wider distribution of molecular component detection and physical properties and the wider dynamic range coverage of lipid concentrations; very often mass spectrometric (MS) detection is used.

11.6.1 MS as the First Choice and Successful Technique for Lipidomics

With the introduction of soft ionization techniques such as matrix-assisted laser desorption/ionization (MALDI), electrospray ionization (ESI), and atmospheric pressure chemical ionization (APCI) for MS, being the last two easily coupled to LC, the rapid and sensitive analysis for the majority or of a substantial fraction of lipids in one single experiment was made possible. Consequently, new soft ionization MS-based analytical strategies have been, and are still, emerging in lipidomics research. Strategies currently used in lipidomics include direct infusion ESI-MS and ESI-MS/MS, LC coupled with ESI-MS or MS/MS, and MALDI combined with Fourier transform ion cyclotron resonance MS (MALDI-FTICR-MS) or time-of-flight MS (MALDI-TOF-MS) [2].

Recently, novel multidimensional methodologies and ion mobility-MS (IM-MS) became available for lipidomics. A rapid separation of isomers, conformers, and enantiomers can be achieved by two-dimensional IM-MS, in addition to a resolving power analogous to that obtained with capillary GC. IM-MS has already shown enormous potential in lipid characterization and analysis of complex biological samples [10].

Identification of lipids can be achieved by using MS/MS, recorded in positive or in negative modes, as to get information on the head group (positive ESI), and on the carbon chain length and degree of unsaturation of the fatty acid chains of the

Table 11.1 Ions formed in ESI(+) and ESI(−) modes as well as the common fragment ions or neutral losses of different lipid groups in electrospray ionization MS/MS

Lipid class	Positive mode	Negative mode	m/z	Head group
PC	$[M+H]^+$, $[M+Na]^+$	$[M-H]^-$, $[M+HCOO]^-$, $[M+CH_3COO]^-$	184.0739	Choline
LPC	$[M+H]^+$, $[M+Na]^+$	$[M-H]^-$, $[M+HCOO]^-$, $[M+CH_3COO]^-$	184.0739	Choline
PE	$[M+H]^+$, $[M+Na]^+$	$[M-H]^-$	141.0191	Ethanolamine
LPE	$[M+H]^+$, $[M+Na]^+$	$[M-H]^-$	141.0191	Ethanolamine
PA	$[M+NH_4]^+$	$[M-H]^-$	152.9953	–
PG	$[M+NH_4]^+$	$[M-H]^-$	152.9953	Glycerol
PI	$[M+NH_4]^+$, $[M+H]^+$, $[M+Na]^+$	$[M-H]^-$	223.0008	Inositol
PS	$[M+H]^+$	$[M-H]^-$	185.0089	Serine
DG	$M+NH_4]^+$, $[M+Na]^+$	–	Fatty acid+NH_3	–
TG	$M+NH_4]^+$, $[M+Na]^+$	–	Fatty acid+NH_3	–
CE	$[M+NH_4]^+$, $[M+H]^+$, $[M+Na]^+$	–	Fatty acid+NH_3	–
SM	$[M+H]^+$	$[M+HCOO]^-$, $[M+CH_3COO]^-$	184.0739	Choline

A neutral loss (fatty acid+NH_3) means that the mass of the neutral loss is the sum of the mass of the fatty acid chain of the lipid and ammonia

lipid (negative ESI) (Table 11.1). Most GPs, such as phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), lysoPCs, ceramides, and cardiolipins, can be identified with both positive and negative ionization modes, whereas TGs and DGs are identified only in positive mode. TGs are mainly distinguished as ammonium and sodium adducts, whereas DGs form sodium adducts and display neutral loss of water caused by in-source fragmentation. The formation of these adducts varies depending on the chain length of the fatty acids. Shorter TG species ionize preferentially as $[M+Na]^+$ adducts, whereas longer chain length leads to predominant $[M+NH_4]^+$ species. In-house libraries are commonly created for specific lipids and use retention time, m/z value, and MS/MS data. All these tools that have recently been established facilitate automated lipid identification. However, it is difficult to get information on the double-bond positions, and typically, lipid structures are often stated as a single isomer. With the use of MS/MS, the sn-1 or sn-2 position can be determined; however, for the thorough data on the lipid identity, there is a need for using specific techniques, such as high-energy collision-induced dissociation, specific multistage fragmentation approaches, or ozone-induced dissociation. Another possibility is to gather specific lipid fractions and, after hydrolysis and

methylation, to examine the fatty acid composition with GC-electron ionization MS. Fatty acid methyl esters can be separated according to the carbon number and saturation, and positional isomers – that are n3, n6, n9, etc. fatty acid methyl esters – can be determined. In addition, with the use of appropriate stationary phases, *cis* and *trans* isomers can also be separated. Thorough structural information is important to comprehend the data in the biochemical context.

11.6.2 NMR as a Powerful Technique for Lipid Assessment

Nuclear magnetic resonance (NMR) spectroscopy has been used to study the physical properties of membrane components. This technique has been changed to study the properties of lipid mixtures in order to determine their structures and functions. Information from these studies is significant with regard to obtaining structure, composition of lipids in cells, turnover of lipids, and characterization of lipid synthesis/transport and degradation pathways [11].

Because NMR spectroscopy is an analytical technique widely employed for characterization and identification of many classes of substances, it also has emerged as suitable analytical platform for the study of biological systems. There are many attempts for optimization and development of pulse sequences and for methodologies of analysis that allow the identification and the classification of organic compounds in complex samples. A great variety of NMR experiments (e.g., HSQC, HMBC, TOCSY, etc.) are being used to solve biological issues where biofluid samples such as serum, plasma, urine, cerebrospinal fluids (CSF), and others are being investigated. Biofluids are constituted from a mixture of organic compounds as amino acids, carbohydrates, organic acids, lipids, etc. [12]. Beside of this, many of these molecules are present in different concentrations in samples and have different physicochemical properties (mass, mobility, functions, etc.), making very challenging lipid and lipoprotein analyses in metabolomics [13]. Nuclear magnetic resonance comes to facilitate the analysis of hundreds of metabolites in a single sample, with or without any previous treatment [14].

One-dimensional (1D) NMR spectroscopy using solvent suppression is often used for metabolomics analysis. Recently, the new approach is using lipidomics of biological samples from the isolated lipids, being one of the successful strategies for lipid biological system comprehension [15].

There are diverse types of clinical sample preparations for lipid analysis, from simple approaches such as single organic solvent extraction, liquid-liquid extraction, and solid-phase microextraction to advanced techniques such as supercritical fluid extraction, microwave-assisted extraction, and ultrasound-assisted extraction. The common approach is liquid-liquid extraction, which is a method involving the use of two immiscible organic solvents. In order to reach exhaustive extraction of lipid classes, comprehends from phospholipids and glycolipids to fatty acids, DAGs and TAGs, usually for lipids extraction a mixture of chloroform and methanol with water is used [16].

Modified methods aiming to reproduce all properties of the clinical samples, from those classical methods as by Folch et al. [17] (chloroform/methanol/water ratio 8:4:3 v/v/v) and by Bligh and Dyer [18] (chloroform/methanol/water ratio 1:2:0.8 v/v/v), are broadly used nowadays [16–20]. Tukiainen et al. [15] use NaCl 0.15 mol L^{-1} solution instead of water to extract lipids from plasma. For NMR analysis, samples do not have to pass through processes of derivatization or sequences of dilutions. Also, after lipid extraction, a solubilization using deutered solvent is performed, and this solution is placed into a NMR tube.

However, lipidomics analysis is still very challenging, once lipids show some properties that may bring difficulties in analysis performed by some analytical techniques. The typical difference in the polarity between the lipids and the lack of chromophores in their structures, for example, complicate the separation and identification of the classes [21]. In this context, NMR spectroscopy may help with the simplification in the sample preparation also, physicochemical properties of samples are not changed, and the data on diffusion coefficients though peaks at intensity measurements, using the pulsed field gradients (PFG) applied during the FT NMR experiment [13], can be done.

The use of NMR spectroscopy in lipidomics is not as widespread as in metabolomics because of its low sensitivity compared to MS. However, the application of NMR techniques that help in the discrimination of individual molecules, as well as in complex mixtures analysis, principally with the employment of two-dimensional NMR (DETOCSY, ^{31}P, ^{1}H COSY, among others), which have a detection limit around 4 nmol L^{-1}, is increasing in the lipidomics research based on NMR spectroscopy [14, 22].

The nuclei detected in NMR spectroscopy that commonly are used in the characterization of lipids are ^{1}H, ^{13}C, and ^{31}P, being the ^{13}C NMR spectroscopy very employed for triradylglycerols (GL03) and fatty acid analyses. ^{31}P NMR is more suitable for glycerophospholipid analysis, and ^{1}H NMR allows analysis of all types of lipids [21, 23]. In general, the advantages using NMR for lipid analysis when compared to other techniques are direct measurement; nondestructive techniques, which allow the recovery of the sample after analysis; and structural analysis of compounds. Disadvantages are low sensitivity, spectra dominated by very abundant lipids (cholesterol, PC) in ^{1}H NMR, and line broadening of lipids in aqueous solutions in ^{31}P NMR [14, 23].

Studies on ^{13}C and ^{31}P NMR spectroscopy data, and understanding of relations between the analytical parameters and physicochemical properties of lipids, help to achieve a reliable determination of the composition of phospholipids that constitute the matrix of cell membranes [24, 25].

In the ^{1}H NMR spectra acquisition of a blood serum sample, the resonances referred to lipid moieties show broad signals and high intensities of peaks similar to the signals of lipoproteins thus interfering the analyst in the assignments of the peaks [26, 27]. One method used to solve this problem is the application of diffusion NMR spectroscopy that is based on the diffusion process of molecules or ions, i.e., in the random translational motion also called Brownian motion of these molecules [28].

Liu et al. [13] developed a pulse sequence for spectra edition of biofluid samples based on the combination of relaxation time of spin, molecular diffusion, and water suppression (WATERGATE) [29], which was named diffusion and relaxation editing (DIRE) pulse sequence. In this work, different combinations between times of longitudinal relaxations (T_1) and transversal relaxations (T_2) and spin-echo values for attenuation of broad peaks from molecules of high molecular mass (or fast relaxing) as the lipoproteins and albumins were studied. These allowed the NMR analysis of samples constituted from molecules with different molecular masses without the need of a pretreatment as the dialysis.

Posteriorly, Liu et al. [30] studied the measurement of diffusion coefficients of individual molecules in blood plasma samples through ^{1}H-^{1}H diffusion-edited total-correlation NMR spectroscopy (DETOCSY) for better understanding of the transport of molecules in biological system. When they applied a low gradient strength, the cross peaks of small molecules were attenuated, while high gradient strength was responsible for attenuation of macromolecules or small molecules bound to them, and so enabled a more accurate measure of the diffusion coefficient through the pairs of cross peaks relating to the separation of molecules.

Lopes et al. [31] studied the lipids in plasma samples of overweight subjects that underwent Roux-en-Y gastric bypass surgery (RYGB), a clinical method for weight loss that helped the glycemic control induced by hormonal changes. The T_2-edited (Carr-Purcell-Meiboom-Gill pulse sequence) and diffusion-edited ^{1}H NMR spectra (Fig. 11.3) were acquired for monitoring the bariatric surgery and the gastric mixed-meal tolerance test (MMTT). By T_2-edited ^{1}H NMR spectrum, it was possible to monitor levels of small molecules (β-glucose, alanine, lactate, etc.) and some amino
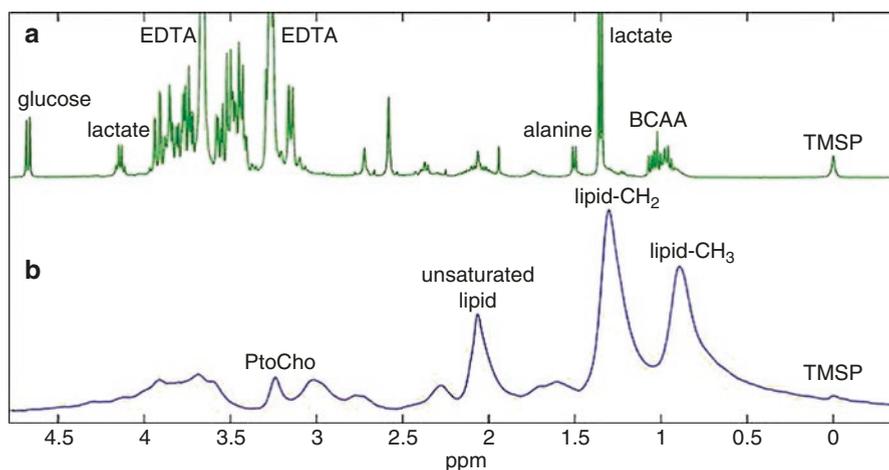
Fig. 11.3 (**a**) T_2-edited and (**b**) diffusion-edited ^{1}H NMR spectrum for overweight subjects before Roux-en-Y gastric bypass surgery. *BCAA* branched-chain amino acids, *EDTA* ethylenediaminetetraacetic acid, *PtoCho* phosphatidylcholine, *TMSP* 2,2,3,3-d_4-3-(trimethylsilyl)propionic acid (*Permission from Omics Journal*)

acids, while the diffusion-edited ^{1}H NMR enabled to measure the level changes in HDL, LDL, VLDL, phosphatidylcholine, etc.

Cai et al. [32] analyzed the lipoproteins in plasma and urine samples of first-episode neuroleptic-naive schizophrenia (FENNS) patients and after the treatment with risperidone drug using the diffusion-edited experiments and bipolar pulse pair longitudinal eddy current delay (BPP-LED) pulse sequence [33–35]. The LED is used to prevent spin relaxation during the diffusion, thus allowing to obtain information that can help in the determination of molecular diffusion coefficients [33, 36]. In their work, Cai et al. [32] observed a reduction in the levels of high-, low-, and very-low-density lipoproteins (HDL, LDL, and VLDL), phosphatidylcholine (PC), lipids, and unsaturated fatty acids (UFA) and an increase in lysophosphatidyl-cholines (LPC) in FENNS when compared to the control groups.

The ^{1}H NMR spectroscopy gives information about chemical shifts, intensities, and chemical environments regarding all compounds that have hydrogen atoms in their organic functions. The data from the NMR spectra such as chemical shift and intensities (variables) are preprocessed through different methods, as the normalization scaling and the transformation after do the correction of baseline and the standardization of reference peaks. The preprocessing passes the set of spectra in variables that are named of buckets or binning and used in chemometric analysis [37].

11.7   Lipidomics Data Analysis and Bioinformatics

Data processing, data mining, and identification are critical steps in the lipidomics analytical workflow. Numerous preprocessing tools that incorporate different approaches and possibilities are available for chromatography-MS data, and different approaches are typically used for shotgun data. The core functions achieved by most tools typically comprise of peak detection, filtering and artifact removal, alignment (for LC-MS data), normalization, etc. In addition, "gap filling" is a significant feature, i.e., finding "missing" peaks in the data to avoid zero values that make the statistical analysis difficult. Filtering and peak detection focus on detecting real chromatographic peaks in each data file, and peak alignment focuses on locating and listing detected peaks found in the sample files.

Also, the data processing tools have clear advantages but also limitations. In a recent study, four automated preprocessing tools for GC-MS and LC-MS data were assessed in terms of their ability to ascertain the greatest number of metabolites consistently in a set of samples, as well as the robustness of the methods [38]. The preprocessing tools selected were MetAlign, MZmine, XCMS, and SpectConnect. The results revealed that different tools accomplished better for the GC-MS data than for the LC-MS data and that the qualitative and quantitative performance also displayed perfect variances between the tools. For GC-MS data, MetAlign had the most component recognitions, followed by MZmine, SpectConnect, and XCMS, whereas for the accurate-mass LC-MS data, the order was MetAlign, XCMS, and

MZmine. The best presentation was obtained when two methods (e.g., MetAlign with MZmine or MetAlign and XCMS) were combined. However, in terms of quantitative presentation, SpectConnect and MetAlign achieved clearly worse than, for example, XCMS and MZmine for the accurate-mass LC-MS data. Both MZmine and XCMS provided acceptable quantitative results for GC-MS data as well as LC-MS data. All software tools stated a large number of false peaks, and thus manual examination of the chromatographic runs is required to eradicate those peaks. The study also indicated that although preprocessing tools have automated steps that are impractical to perform manually, a significant level of manual input is required in selecting the optimal parameters, processing peak tables, and validation. These outcomes clearly display that further expansion of the data processing algorithms is required.

Numerous software tools have been established for the identification of lipids in shotgun lipidomics, including LipidQA, LIMSA, FAAT, lipID, LipidSearch, LipidView, LipidInspector, LipidXplorer, LipidBlast, and ALEX, both for specific applications/instrumentation and for cross-platform software presenting user-specified instructions interrogating spectral data in an open-source format [38]. Typically, in the top-down approach, the data are investigated for specific molecular fragmentation. For example, in LipidXplorer individual spectra are systematized in a single flat file database that is further questioned by user-defined queries written in the molecular fragmentation query language (MFQL). In each MS spectrum, the lipid class-specific MFQL query checks if plausible precursor masses match the elemental composition likely for the corresponding molecular species. Optional search criteria can also be applied, for example, based on the odd or even number of carbon atoms in the fatty acid residues or the anticipated number of double bonds.

11.8 Physiological Factors Affecting Lipidomic Analysis

11.8.1 Lipid Composition in Blood-Based Samples

Blood is composed by (a) cellular components encompassing red and white blood cells and platelets and (b) a liquid carrier, named plasma. In blood, complex lipids are primarily found in the lipoprotein particles, while smaller, more polar lipids (e.g., free fatty acids (FFAs), bile acids, sterols) are found in free form or bound to protein carriers, such as albumin. Among these lipids, free cholesterol is the most abundant and is found primarily in lipoprotein particles. More complex lipids are also found in the red blood cells (RBCs) and platelets. Quehenberger et al. [3] accomplished a comprehensive characterization of lipids in human reference plasma, i.e., in pooled human plasma obtained from healthy individuals after overnight fasting and with gender balance and ethnic distribution that is symbolic of the US population. A summary of the foremost lipid subclasses and their concentrations is given in Table 11.2, with the most abundant lipids in each subclass noticeable. In

Table 11.2 Main lipid classes identified in human plasma

Lipid category	Sum (mmol L^{-1})	Most abundant lipids
Glycerolipids		
TGs	1,058	TG(16:1/18:1/18:1), TG(16:1/18:0/18:2), TG(16:0/18:1/18:2)
1,2-DGs	39	DG(36:3), DG(36:4), DG(36:2)
1,3-DGs	13	DG(36:3), DG(36:4), DG(34:1)
Total	1,110	
Glycerophospholipids		
PE	435	PE(38:5e)/PE(38:4p), PE(38:4), PE(38:6e)/PE(38:5p)
LPE	36.6	LPE(18:2), LPE(18:0), LPE(20:4)
PC	1,974	PC(34:2), PC(36:2), PC(38:4)
LPC	103	LPC(16:0), LPC(18:0), LPC(18:2)
OS	7	PS(36:0), PS(36:1), PS(40:6)
PG	6.12	PG(36:1), PG(38:6), PG(38:5)
PA	2.5	PA(34:0), PA(36:2), PA(36:0)
PI	31.5	PI(38:4), PI(36:2), PI(34:2)
NAPS	0.013	52:1, 54:2
Sphingolipids		
SMs	303.468	SM(d18:1/C16:0), SM(d18:1/c24:1), SM(d18:2/C24:1)
hexCers	2.3135	hexCer(d18:1/C22:0), hexCer(d18:1/C24:0), hexCer(d18:1/C16:0)
Cers	11.586	Cer(d18:1/C24:0), Cer(d18:0/C24:0), Cer(d18:1/C22:0)
Sphingoid bases	0.5678	S1P(d18:1-P), S1P(d18:0-P), SP(d18:1)
Sterol lipids		
Free sterols	3,800	Cholesterol, lathosterol, sitosterol
Esterified sterols	2,954	CE(18:2), CE(18:1), CE(20:4)

Referred from Quehenberger et al. [3] and modified

individual lipoprotein fractions, lipid profiles differ considerably. Specific lipids, such as ceramides, have been distinguished primarily in very low-density lipoprotein (VLDL) and low-density lipoprotein (LDL) fractions, whereas ethanolamine plasmalogens are found primarily in LDL and high-density lipoprotein (HDL) subfraction 2. LysoPCs and ether-linked PCs have been found in all lipoprotein fractions, with the greatest abundances in HDL subfraction 2, HDL subfraction 3, and LDL. Most large-scale studies are based on the extent of the lipids in total serum or plasma, because it is too time-consuming to first separate, for example, individual lipoprotein fractions, followed by lipid analysis. However, the lipid characterization of lipoprotein fractions can give a much more thorough view of the lipid metabolism. For example, the VLDL fraction can give improved perception into lipid metabolism in liver than the overall lipid composition [39].

The most abundant lipids in human blood are GLs, mainly TGs. The TGs are most abundant in chylomicrons and in VLDL and intermediate-density lipoprotein fractions. Most recent methods can detect and recognize about 50 or more TGs in

human plasma or serum; however, the definite amount of TGs is substantially higher than that. A large number of different GPs have been recognized in plasma and serum, including PA, PC, PE, PG, PI, and PS. By total class concentration, the overwhelming majority of GPs in human plasma are PCs and PEs (Table 11.2). These two classes also cover substantial amounts of ether-linked lipids. SPs in the blood correspond to a part of the circulating lipoprotein particles (VLDL, LDL, and HDL), they are carried by serum albumin, and they are also present in blood cells and platelets. Over 200 individual SPs were identified in human plasma, and the sphingomyelins (SMs) account for the largest fraction of SPs in plasma [40]. Cholesteryl esters (CEs) are one of the most abundant lipid classes (Table 11.2) in human serum and plasma, with CE (18:2) constituting a major fraction of the CE subclass. Some lipids are usually not identified in blood samples; for example, cardiolipins are present in blood-based samples at very low concentrations and are typically not detected with the current nontargeted profiling methods.

11.8.2   *Effect of Gender, Age, Diet, and Sampling Time on Lipid Levels*

Numerous phenotypic and physiological parameters, including medication and specific food supplements (e.g., fish oil, and vitamin D), have a significant influence on circulating lipid levels, and thus the medication should be documented and considered during the study planning and for the interpretation of the outcomes. Although the huge biological variation of lipid levels, the levels of lipids have been shown to have a relatively small difference in the same individual when analyzed under identical sampling conditions [41].

The levels of lipids such as SMs, PCs, and TGs in blood may vary between the genders. Most lipids do not present a strong association with gender, but for different age groups, gender does have an effect on specific lipids. Age also has a prominent effect on lipid concentrations, mainly in females. Numerous TGs along with a few other lipids have been shown to present significantly higher levels in elderly females than in young females in both serum and plasma samples [42]. The same study showed that female-specific age-associated changes of the levels of lysoPCs were detected in plasma but that were markedly less in serum.

Diet also affects the levels of lipids and individual fatty acids in blood, both in the long and short term. In fact, it has been proposed that about 50 % of metabolites measured are dependent on diet [43].

In addition to gender and age, the circadian system controls lipid and carbohydrate homeostasis, thus augmenting energy storage and its utilization during the day. Thus, the time of sample collection also has an effect on lipid profiles. Naturally, the diet and fasting conditions play a critical role, but even in fasting conditions, specific lipids have revealed difference due to the circadian time course. There are still an inadequate number of studies related to circadian time course of lipids in

humans; however, the studies reveal so far that specific lipids have a circadian cycle pattern [44].

Guidelines usually recommend that lipid profiles should be attained from individuals in the fasting state. However, it is not always promising to obtain fasting samples, for example, in the case of very young children. Also, fasting is uncomfortable predominantly in individuals with diabetes, due to the fear of either hyperglycemia or hypoglycemia. It is also essential to test the individuals' capacity and flexibility to respond with typical environmental factors, such as physical activity or eating. However, most studies to date are inadequate to the analysis of samples obtained in a fasted state.

The body mass index (BMI) of individuals also affects the lipid profile. In most studies, lipid profiles have been studied in relation to specific obesity-associated disease, and precise alterations were observed. However, a substantial portion of obese subjects does not present any of the well-known metabolic abnormalities. Thus, it is essential to recognize lipids that are associated with obesity rather than with a specific disease [45].

11.9 Lipidomics-Based Biomarkers

Since lipids retain a diversity of biological functions in the processes of life such as formation of cellular membranes, energy storage, and cell signaling, they can be anticipated to reflect much of the metabolic status in health and disease. In addition, several studies have demonstrated that lipid metabolic disorders or abnormalities can lead to various human diseases including diabetes, obesity, arteriosclerosis, coronary heart disease, and brain injuries [2, 46]. Therefore, monitoring the changes of lipid metabolites of certain molecular species in biological samples, as influenced by external stimuli or disturbance by disease processes, will be helpful for the discovery of lipid metabolites with potential of being indicative of metabolic disorders or diseases.

Biomarkers could be considered to extend all the way to include our fixed genomic characters. At the level of the subcellular and tissue, the research in this field has employed the transcriptomics, proteomics, metabolomics, lipidomics, immunological, and biological epigenetics [46, 47]. The recent attention in biomarker discovery is encouraged by new molecular biology techniques, which allows the rapid finding of relevant biomarkers, without specific comprehensive insight into the mechanisms of a disease. Lipidomics is the most appropriate approach for the study of pathways and networks of lipids in biological systems. There is no question that clinical requirement and complexity of lipid metabolism will pose a series of novel challenges for researchers devoted to lipidomics of various diseases. Numerous risk factors contributing to disease are suggested to influence lipid metabolism, and thus, they may be reflected in the lipid profile of an individual. One of the most extensively known lipid biomarkers is cholesterol, which, in the form of total blood cholesterol and/or HDL cholesterol, has been used for more than 50 years for the determination of risk for cardiovascular disease [48].

Apart from applications in human diseases, the approach of lipidomics-driven biomarker discovery has also been used in nutrition and health fields, aiming the health promotion and the disease prevention [49].

In spite of the numerous analytical strategies available to identify changes in lipid metabolism related to disorders or diseases, multivariate statistical analysis was almost invariably performed to assist the finding of novel lipid molecular species that could serve as potential biomarkers.

11.10 Lipidomics in Schizophrenia Research

One of the goals of schizophrenia (SCZ) research is to explain the disorder in clear and simple biological terms and, in doing so, relegate it to the index of mundane – and more amenable – human disorders. The existing pace of advances in large-scale biological data acquisition and data processing might make the task appear eminently achievable. However, even the most avid optimists would have to disclose that, despite this effort, we are still piecing together the edge of the jigsaw puzzle rather than seeing the full picture (in contraposition to the situation of Parkinson's disease – a cousin by neurotransmitter of SCZ). A cross-disciplinary understanding of environmental impacts, genetic risk, cellular pathology, anatomical pathology, network dysfunction, and outward symptomatology has converged to inform the development of novel therapies. Multiple explanations have been put forward for this qualitative difference in disorder complexity, with main focus on the idea that SCZ is not a unitary disorder but rather a common symptomatic endpoint of a great variety of brain dysfunctions and insults. In support to this, SCZ now belongs to a wide family of disorders that have proven interrelationships at the genetics level. These include BD, depression, intellectual disability, autism spectrum disorders, and, recently, multiple sclerosis [46]. Biomarker identification is one such arena of research that will have to meet these demands. The advancement in development of commercial SCZ biomarker tests depends on economic models, regulation, and perceived healthcare need, just as much as bioinformatics. Hence, there may be a more pragmatic short-term goal for biomarkers: to help the subdivision and classification of SCZ according to its principal molecular etiopathology.

Current technological developments for lipidomics analysis have provided remarkable outcomes in other areas of research, with potential to be useful for biomarker discovery in SCZ. The lipid biomarkers offer a new outlook for achieving better diagnosis either in preclinical experiments using mammal animal models or in human clinical evaluations. Lipidomics has the potential for the discovery of biomarkers for SCZ, not only in brain tissue but also in the blood (serum or plasma) and CSF (Table 11.3; referred from Sethi et al. [2]).

In one study, Kaddurah-Daouk et al. [19] used a specific lipidomics platform and found changes in different lipid classes (PC, PE, TG) in the plasma of SCZ patients after 2–3 weeks of treatment with atypical antipsychotic drugs. A recent study has also proved a significant downregulation of several n3 and n6 PUFAs compositions

Table 11.3 Application of lipidomics for discovering metabolites/biomarkers in animal model and human-based studies of schizophrenia

Model/subject	Sample	Analytical method(s)	Metabolites/biomarkers	Reference
Human studies	Postmortem brain tissues (left thalamus)	TLC	Phosphatidylcholine, sphingomyelin, galactocerebrosides 1 and 2, phosphatidylserine	[50]
	Plasma	HPLC-ELSD and GC-FID	Triacylglycerols, free fatty acids, phosphatidylcholine, phosphatidylethanolamine	[19]
	Postmortem brain tissues (white matter, gray matter)	UPLC-ESI-QTOF-MS	Free fatty acids, phosphatidylcholine, ceramide	[51]
	Red blood cells		Free fatty acids, ceramide	
	Postmortem brain tissue (hippocampus)	HPLC-ESI-Q-MS	Phosphatidylserine (n6), phosphatidylcholine (n6)	[52]
	Serum	UPLC-ESI-QTOF-MS	Triglycerides (lipid cluster, LC4 to LC9)	[53]
			Lysophosphatidylcholines (LPCs)	[54]
	Plasma	TLC and GC-FID	Phosphatidylcholine (n3, n6), phosphatidylethanolamine (n3, n6)	[55]
	Postmortem brain tissue (frontal cortex)	ESI-Orbitrap-MS	N-Acylphosphatidylserines, N-acylserines	[56]
		ESI-MS/MS	Glycosphingolipids, choline plasmalogens	[57]
	Plasma		Choline plasmalogen, ethanolamine plasmalogen, docosahexaenoic acid (DHA)	[58]
	Platelets			
	Postmortem brain tissue (frontal cortex)	ESI-Orbitrap-MS	Sulfatides, choline plasmalogen, ethanolamine plasmalogen, N-acyl-phosphatidylserines	[59]
Animal studies	Hippocampus		Glycosphingolipids, choline plasmalogens	[57]
	Serum	2D HPLC-ESI-QTOF-MS	Glycerophospholipid, sphingolipid, fatty acyls	[60]

in PE and in lipid classes in the blood plasma of first-episode SCZ patients [55]. These changes in lipid metabolism could indicate a metabolic vulnerability in patients with SCZ that may occur early in the development and onset of the disease.

Alterations in the peripheral tissue membrane PLs levels have also been detected [61] in RBCs of SCZ patients. Schwarz et al. [51] observed lipid changes in post-mortem brain samples from SCZ, and they found significant changes in the levels of

FFAs and PCs in the gray and white matter of SCZ patients compared to control samples. Ceramides were significantly increased in white matter of SCZ as compared to control levels. In addition, lipid profiling of RBCs of SCZ accused significantly decreased levels of FFAs and ceramides in drug-naïve first-onset patients. Reductions of PC levels have formerly been reported for different regions of the SCZ brain [50], which was linked to an increase in SM turnover, as PC is the choline donor to SM in neurons and oligodendrocytes.

Lipidomic analyses can be complicated due to the high occurrence of metabolic syndrome in SCZ and its induction/worsening by treatment with antipsychotics. However, in the case of plasmalogen (Pl) analyses, about 20 % decrease in circulating Pls was described in 20 first-episode and 20 recurrent SCZ patients [62], suggesting that this may represent an intrinsic biochemical deficit involved in this developmental disorder. A recent study also revealed significant decrease in choline plasmalogens, ethanolamine plasmalogens, and DHA in the plasma of patients with SCZ. In contrast, increased cellular levels of choline plasmalogens and decreased levels of ethanolamine plasmalogens and DHA were found in platelets of patients with SCZ [58]. Another recent study from the same research group demonstrated elevated level of sulfatides, choline plasmalogens, ethanolamine plasmalogens, and N-acyl-phosphatidylserines (NAPS) in the gray matter of postmortem frontal cortex SCZ subjects [59]. As major components of membranes, Pls are essential for membrane fluidity, lipid raft formation, membrane fusion for neurotransmitter release, ion transport, and regulation of cholesterol efflux. Pls are also essential in brain development, both for white and gray matters.

In conclusion, several studies showed significant changes in prefrontal cortex FA concentrations, particularly within cholesteryl ester (CE) and abnormalities in Pl levels of SCZ patients compared to controls [20, 62]. An increase in CE-FA concentration turnover in SCZ may reflect the excitotoxicity and neuronal loss reported in SCZ patients [63]. This is consistent with findings in postmortem brains from SCZ patients, suggesting a hypoglutamatergic and hyperdopaminergic neurotransmitter signaling, in association with neuronal loss and disease worsening over time [64]. Upcoming studies of Pls are required to explore their role in the pathogenesis of SCZ and to clarify whether restoration of normal Pl levels is linked with the therapeutic effects of antipsychotic drugs.

11.11   The Future of Clinical Lipidomics

Lipids offer the possibility of discovery novel key biomarkers for many areas and different diseases. They could also fuel diagnostic developments and, thus, support more personalized methodologies of treatment. Lipidomics can also be used for studying numerous experimental disease models, and this could offer an enormous boost for the translational medicine.

During the discovery stage, detailed lipidomic studies currently demand a repertoire of several different analytical platforms. Such a quest is currently time-consuming, and hence, it is appropriate only to a limited amount of samples. However,

technology and process improvements are developing and may enable to output thorough lipidome data sets based on large sample sets more rapidly in the near future. Another challenge is the bias of the final output arising from the chosen methodological setups and instrumentations. Although internal standards are typically applied to quantify the lipids of interest, owing to the lack of proper non-endogenous standards, only a subset of lipid species can be measured with acceptable precision. It is notable that the endpoint results depend on the internal standards applied, and, unfortunately, most aberrations among users are found in this aspect. Therefore, due to the lack of standardization, it is difficult to relate or combine lipidomic data from different laboratories. Although rigorous standardization and validation processes are required for conveying lipidomic assays to clinical practice, the benefits are assorted. Once set up, MALDI-MS assays are analytically robust and efficient clinical solutions. The assays can be scaled down from the discovery throughput mode involving long per sample scan times to just 1–2 min analysis time per sample, with no extensive washing or incubation steps in the overall assays. Lipidomic analyses are directed in a 96-well format and in a robot-assisted workflow, and therefore, this platform can attain substantial sample throughputs. More critical for wider assumption remain the issues centering on sample transportation, storage, number of freeze-thaw cycles, preparation, and sample handling during the analytical process.

Given the obtainability of better and more appropriate internal standards and MS analysis methods, the eminence of lipidomic outputs will be higher with more lipid species determined in absolute quantities. Bioinformatics solutions will permit the processing of all data and to put out lipidomic results instantly, in an accessible way. Concurrently, informatics setups can monitor all processes in the lipidomic workflow, identify automatically any failure in the process, and also monitor continuously the sample quality. Placing all these pieces together will guide lipidomic standardization, which will make lipidomic solutions attractive for preclinical and clinical studies and pertinent to the regulatory environment in a cost-effective manner.

11.12 Concluding Remarks and Perspectives

Abnormal lipid or metabolism dysfunction is considered to be the major influence aspect in many lifestyle-associated diseases and hereditary/genetic conditions. Lipidomics is an emerging approach for a widespread and systematic study of a variety of lipids. The field of lipidomics is under a continuous investigation to further explore the lipidome with the eventual goal to broaden our biological knowledge for different diseases. Currently, lipidomics has an enormous prospective in lipid research, in which different lipid profiling is associated to various diseases, and changes of lipid metabolism or pathway modulation can be identified in human complex diseases. This offers new perceptions into metabolic and inflammatory diseases. The combination of lipid profiles and multivariate statistics can help us in novel biomarker discovery, disease pathology explanation, drug-response monitoring in therapy and toxicity, translational medicine, and in-depth uncovering mechanisms of lipid-mediated disease.

MS and chromatography techniques have significantly encouraged the developments and applications of lipidomics in clinical chemistry. According to different research objectives, different MS and chromatography approaches can be selected, and selected approaches must be appropriate for application to the specific lipid species. Direct infusion ESI-MS has been used to distinguish whole lipid extracts, and it has revealed significant potential in the identification and quantification of PLs and fatty acid species in a quick and robust manner. However, direct infusion ESI-MS is vulnerable to ion suppression; this shortcoming can be overwhelmed to some extent by chromatographic techniques. LC-MS can also be applied to separation lipids from complex samples into individual lipid classes or separate the same lipid class. GC-MS is appropriate for the fatty acids and their derivatives, but it is limited by the necessity of analytes being volatile and due to its dynamic range. Combined analytical approaches can be acquired by overcoming the limitations of individual techniques for a wide range of the lipidome.

Although a number of lipidomic experiments has been carried out on exploring diseases through analyzing biomarkers and metabolic pathways in clinics, clinical lipidomics is still in its infancy compared to proteomics and metabolomics. The current researches on lipidomics focus predominantly on biomarker searching, which is apparently inadequate. There is no question that clinical requirement and complexity of lipid metabolism will pose researchers enthusiastic to lipidomics of various diseases with a series of novel challenges. The combined techniques will help advance our understanding of the physiological functions of lipid species and depict the etiology and pathophysiology of multiple lipid-related diseases, such as cancer, obesity, diabetes, and atherosclerosis. Integrated with other omics strategies, this platform will offer a new outlook for dissecting and improving disease diagnosis and prevention.

References

1. Han X, Gross RW. Electrospray ionization mass spectroscopic analysis of human erythrocyte plasma membrane phospholipids. Proc Natl Acad Sci U S A. 1994;91:10635–9.
2. Sethi S, Hayashi MA, Sussulini A, et al. Analytical approaches for lipidomics and its potential applications in neuropsychiatric disorders. World J Biol Psychiatry. 2016:1–15. doi:10.3109/15622975.2015.1117656.
3. Quehenberger O, Armando AM, Brown AH, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res. 2010;51:3299–305.
4. Adibhatla RM, Hatcher JF. Role of lipids in brain injury and diseases. Future Lipidol. 2007;2:403–22.
5. Horrobin D. The lipid hypothesis of schizophrenia. In: Skinner ER, editor. Brain lipids and disorders in biological psychiatry, vol. 35. Amsterdam: Elsevier Science; 2002. p. 39–52.
6. Berger GE, Smesny S, Amminger GP. Bioactive lipids in schizophrenia. Int Rev Psychiatry. 2006;18:85–98.

7. Ota VK, Noto C, Santoro ML, et al. Increased expression of NDEL1 and MBP genes in the peripheral blood of antispychotic-naïve patients with first-episode. Eur Neuropsychopharmacol. 2015;25:2416–25.

8. Maurya PK, Noto C, Rizzo LB, et al. The role of oxidative and nitrosative stress in accelerated aging and major depression disorder. Prog Neuropsychopharmacol Biol Psychiatry. 2016;65:134–44.

9. Kunz A, Anrather J, Zhou P, et al. Cyclooxygenase-2 does not contribute to postischemic production of reactive oxygen species. J Cereb Blood Flow Metab. 2007;27:545–51.

10. Paglia G, Kliman M, Claude E, et al. Applications of ion-mobility mass spectrometry for lipid analysis. Anal Bioanal Chem. 2015;407:4995–5007.

11. Vilella F, Ramirez LB, Simón C. Lipidomics as an emerging tool to predict endometrial receptivity. Fertil Steril. 2013;99:1100–6.

12. Smolinska A, Blanchet L, Buydens LMC, et al. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Anal Chim Acta. 2012;750:82–97.

13. Liu M, Nicholson JK, Lindon JC. High-resolution diffusion and relaxation edited one- and two-dimensional ^{1}H NMR spectroscopy of biological fluids. Anal Chem. 1996;68:3370–6.

14. Rolim AEH, Henrique-Araújo R, Ferraz EG, et al. Lipidomics in the study of lipid metabolism: current perspectives in the omic sciences. Gene. 2015;554:131–9.

15. Tukiainen T, Tynkkynen T, Mäkinen VP, et al. A multi-metabolite analysis of serum by 1H NMR spectroscopy: early systemic signs of Alzheimer's disease. Biochem Biophys Res Commun. 2008;375:356–61.

16. Teo CC, Chong WPK, Tan E, et al. Advances in sample preparation and analytical techniques for lipidomics study of clinical samples. Trends Anal Chem. 2015;66:1–18.

17. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipids from animal tissues. J Biol Chem. 1957;226:497–509.

18. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37:911–7.

19. Kaddurah-Daouk R, McEvoy J, Baillie RA, et al. Metabolomic mapping of atypical antipsychotic effects in schizophrenia. Mol Psychiatry. 2007;12:934–45.

20. Taha AY, Cheon Y, Ma K, et al. Altered fatty acid concentrations in prefrontal cortex of schizophrenic patients. J Psychiatr Res. 2013;47:636–43.

21. Carrasco-Pancorbo A, Navas-Iglesias N, Cuadros-Rodríguez L. From lipid analysis towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part I: modern lipid analysis. Trends Anal Chem. 2009;28:263–78.

22. Li M, Yang L, Bai Y, et al. Analytical methods in lipidomics and their applications. Anal Chem. 2014;81:161–75.

23. Wenk MR. The emerging field of lipidomics. Nat Rev Drug Discov. 2005;4:594–610.

24. Lutz NW, Cozzone PJ. Principles of multiparametric optimization for phospholipidomics by ^{31}P NMR spectroscopy. Biophys Rev. 2013;5:295–304.

25. Leftin A, Mologu TR, Job C. Area per lipid and cholesterol interactions in membranes from separated local-field ^{13}C NMR spectroscopy. Biophys J. 2014;107:2274–86.

26. Ala-Korpela M. ^{1}H NMR spectroscopy of human blood plasma. Prog Nucl Magn Reson. 1995;27:475–554.

27. Barrilero R, Llobet E, Mallol R, et al. Design and evaluation of standard lipid prediction models based on ^{1}H-NMR spectroscopy of human serum/plasma samples. Metabolomics. 2015;11:1394–404.

28. Nicolay K, Braun KPJ, de Graaf RA, et al. Diffusion NMR spectroscopy. NMR Biomed. 2001;14:94–111.

29. Piotto M, Saudek V, Sklenář V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR. 1992;2:661–5.

30. Liu M, Nicholson JK, Parkinson JA, et al. Measurement of biomolecular diffusion coefficients in blood plasma using two-dimensional ^{1}H-^{1}H diffusion-edited Total-Correlation NMR Spectroscopy. Anal Chem. 1997;69:1504–9.

31. Lopes TI, Geloneze B, Pareja JC, et al. "Omics" prospective monitoring of bariatric surgery: roux-en-Y gastric bypass outcomes using mixed-meal tolerance test and time-resolved (1)H NMR-based metabolomics. OMICS. 2016;20:415–23.

32. Cai HL, Li HD, Yan XZ, et al. Metabolomic analysis of biochemical changes in the plasma and urine of first-episode neuroleptic-naïve schizophrenia patients after treatment with Risperidone. J Proteome Res. 2012;11:4338–50.

33. Gibbs SJ, Johnson Jr CS. A PFG NMR experiment for accurate diffusion and flow studies in the presence of eddy currents. J Magn Reson. 1991;93:395–402.

34. Fordham EJ, Gibbs SJ, Hall LD. Partially restricted diffusion in a permeable sandstone: observations by stimulated echo PFG NMR. Magn Reson Imaging. 1994;12:279–84.

35. Wu D, Chen A, Johnson CS. An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. J Magn Reson A. 1995;115:260–4.

36. Beckwith-Hall BM, Thompson NA, Nicholson JK, et al. A metabonomic investigation of hepatotoxicity using diffusion-edited ^{1}H NMR spectroscopy of blood serum. Analyst. 2003;128:814–8.

37. Checa A, Bedia C, Jaumot J. Lipidomic data analysis: tutorial, practical guidelines and applications. Anal Chim Acta. 2015;885:1–16.

38. Hyötyläinen T, Orešič M. Optimizing the lipidomics workflow for clinical studies-practical considerations. Anal Bioanal Chem. 2015;407:4973–93.

39. Kotronen A, Velagapudi VR, Yetukuri L, et al. Saturated fatty acids containing triacylglycerols are better markers of insulin resistance than total serum triacylglycerol concentrations. Diabetologia. 2009;52:684–90.

40. Vieu C, Terce F, Chevy F, et al. Coupled assay of sphingomyelin and ceramide molecular species by gas liquid chromatography. J Lipid Res. 2002;43:510–22.

41. Breier M, Wahl S, Prehn C, et al. Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. PLoS One. 2014;9:e89728.

42. Ishikawa M, Maekawa K, Saito K, et al. Plasma and serum lipidomics of healthy white adults shows characteristic profiles by subjects' gender and age. PLoS One. 2014;9:e91806.

43. Zivkovic AM, Wiest MM, Nguyen U, et al. Assessing individual metabolic responsiveness to a lipid challenge using a targeted metabolomic approach. Metabolomics. 2009;5:209–18.

44. Gooley JJ, Chua EC. Diurnal regulation of lipid metabolism and applications of circadian lipidomics. J Genet Genomics. 2014;41:231–50.

45. Pietiläinen KH, Sysi-Aho M, Rissanen A, et al. Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects – a monozygotic twin study. PLoS One. 2007;2:e218.

46. Sethi S, Brietzke E. Omics-based biomarkers: application of metabolomics in neuropsychiatric disorders. Int J Neuropsychopharmacol. 2016;19(3):pyv096. doi:10.1093/ijnp/pyv096.

47. Sethi S, Chourasia D, Parhar IS. Approaches for targeted proteomics and its potential applications in neuroscience. J Biosci. 2015;40:607–27.

48. Meikle P, Barlow C, Weir J. Lipidomics and lipid biomarker discovery. Aus Biochemist. 2009;40:12–6.

49. Draisma HH, Reijmers TH, Bobeldijk-Pastorova I, et al. Similarities and differences in lipidomics profiles among healthy monozygotic twin pairs. OMICS. 2008;12:17–31.

50. Schmitt A, Wilczek K, Blennow K, et al. Altered thalamic membrane phospholipids in schizophrenia: a postmortem study. Biol Psychiatry. 2004;56:41–5.

51. Schwarz E, Prabakaran S, Whitfield P, et al. High throughput lipidomic profiling of schizophrenia and bipolar disorder brain tissue reveals alterations of free fatty acids, phosphatidylcholines, and Ceramides. J Proteome Res. 2008;7:4266–77.

52. Hamazaki K, Choi KH, Kim HY. Phospholipid profile in the postmortem hippocampus of patients with schizophrenia and bipolar disorder: no changes in docosahexaenoic acid species. J Psychiatr Res. 2010;44:688–93.

53. Orešič M, Tang J, Seppänen-Laakso T, et al. Metabolome in schizophrenia and other psychotic disorders: a general population-based study. Genome Med. 2011;3:19.

54. Orešič M, Seppänen-Laakso T, Sun D, et al. Phospholipids and insulin resistance in psychosis: a lipidomics study of twin pairs discordant for schizophrenia. Genome Med. 2012;4:1.

55. McEvoy J, Baillie RA, Zhu H, et al. Lipidomics reveals early metabolic changes in subjects with schizophrenia: effects of atypical antipsychotics. PLoS One. 2013;8:e68717.
56. Wood PL. Accumulation of N-acylphosphatidylserines and N-acylserines in the frontal cortex in schizophrenia. Neurotransmitter. 2014;1:e263.
57. Wood PL, Filiou MD, Otte DM, et al. Lipidomics reveals dysfunctional glycosynapses in schizophrenia and the G72/G30 transgenic mouse. Schizophr Res. 2014;159:365–9.
58. Wood PL, Unfried G, Whitehead W, et al. Dysfunctional plasmalogen dynamics in the plasma and platelets of patients with schizophrenia. Schizophr Res. 2015;161:506–10.
59. Wood PL, Holderman NR. Dysfunctional glycosynapses in schizophrenia: disease and regional specificity. Schizophr Res. 2015;166:235–7.
60. Weng R, Shen S, Burton C, et al. Lipidomic profiling of tryptophan hydroxylase 2 knockout mice reveals novel lipid biomarkers associated with serotonin deficiency. Anal Bioanal Chem. 2016;408:2963–73.
61. Ponizovsky AM, Modai I, Nechamkin Y, et al. Phospholipid patterns of erythrocytes in schizophrenia: relationships to symptomatology. Schizophr Res. 2001;52:121–6.
62. Kaddurah-Daouk R, McEvoy J, Baillie R, et al. Impaired plasmalogens in patients with schizophrenia. Psychiatry Res. 2012;198:347–52.
63. Rao JS, Kellom M, Reese EA, et al. Dysregulated glutamate and dopamine transporters in postmortem frontal cortex from bipolar and schizophrenic patients. J Affect Disord. 2012;136:63–71.
64. Moghaddam B, Javitt D. From revolution to evolution: the glutamate hypothesis of schizophrenia and its implication for treatment. Neuropsychopharmacology. 2012;37:4–15.

Chapter 12
Spatial Metabolite Profiling by Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Imaging

Berin A. Boughton and Brett Hamilton

Abstract Mass spectrometry imaging (MSI) is rapidly maturing as an advanced method for spatial metabolite profiling. Herein, we provide an introduction to MSI including types of instrumentation, detailed sample preparation, data collection, overview of data analysis steps, software, common standards, and new developments. Further, we provide an overview of MSI in the clinical space over the past 3 years where MSI has been deployed in diverse research areas including cancer, neurobiology, lipidomics, and metabolite profiling and mapping to name only a few. We provide several examples demonstrating the applicability of MSI to spatially profile metabolites in unique systems requiring special considerations outside of the norm.

Keywords Spatial metabolomics • Mass spectrometry imaging • MALDI • Matrix • High resolution

Abbreviations

3D-MSI   Three-dimensional mass spectrometry imaging
9-AA   9-Aminoacridine

B.A. Boughton (✉)
Metabolomics Australia, School of BioSciences, The University of Melbourne, Melbourne, VIC 3010, Australia
e-mail: baboug@unimelb.edu.au

B. Hamilton
Mater Research Institute, The University of Queensland, St Lucia, QLD, Australia

Pathology Department, Mater Health Services, South Brisbane, QLD, Australia

AP-MALDI	Atmospheric pressure matrix-assisted laser desorption ionization
CHCA (or HCCA)	α-Cyano-4-hydroxycinnamic acid
Cer	Ceramide
DAN	1,5-Diaminonapthalene
DESI	Desorption electrospray ionization
DHAP	2,5-Dihydroxyacetophenone
DHB	2,5-Dihydroxybenzoic acid
DMAN	1,8-Bis(dimethylamino)naphthalene
FA	Fatty acid
FFPE	Formalin-fixed paraffin embedded
fNPs	Functional iron nanoparticles
FT	Fourier transform
FTICR	Fourier transform ion cyclotron resonance
FT-IR	Fourier transform infrared spectroscopy
HCA	Hierarchical cluster analysis
Hex	Hexose
IR	Infrared
IR-MALDI	Infrared matrix-assisted laser desorption ionization
ITO	Indium tin oxide
kMSI	Kinetic mass spectrometry imaging
LDI	Laser desorption ionization
MALDI	Matrix-assisted laser desorption ionization
MIPC	Ceramide phosphoinositol
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MS^n	Multistage tandem mass spectrometry
MSI	Mass spectrometry imaging
MS/MS	Tandem mass spectrometry
m/z	Mass-to-charge ratio
NIMS	Nanostructure-initiator mass spectrometry
OCT	Optimal cutting temperature
PA	Phosphatidic acid
PCA	Principal component analysis
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PS	Phosphatidylserine
ROI	Region of interest
RP	Resolving power
SIMS	Secondary ion mass spectrometry
TIC	Total ion chromatogram
TOF	Time of flight

12.1 Introduction

Recent major technical advances in mass spectrometry (MS) have increased the scope, applicability, and adoption of the technology in a vast array of research areas [1]. The number and scope of approved diagnostic clinical applications utilizing mass spectrometry are increasing and broadening extremely rapidly. In particular, the application of MS to biochemical imaging via mass spectrometry imaging (MSI) has emerged as one of the leading spatial analysis technologies for high-throughput molecular imaging in biological systems. MSI has been employed to investigate a vast range of different spatial biological questions, and there have been many excellent comprehensive reviews published in recent years [1–14]. A recent survey of MSI users has identified matrix-assisted laser desorption ionization (MALDI) as the dominant ion source (95 %), and imaging of small molecules, including drugs, metabolites, and lipids, represents approximately 80 % of the application of MSI. In this chapter, we provide an introduction to MALDI-MSI used for biological-based research.

The "omics" technologies, genomics, transcriptomics, proteomics, and metabolomics (and others), have provided insights into biochemistry, physiology, and biology and are at the forefront of discovery in modern systems biology [15]. The exquisite specialization and compartmentalization of biological systems also require spatial approaches allowing examination of "where things are happening" to unveil the full complexity of the underlying biology.

Spatial analysis can be conducted using a number of different techniques, which can be broadly categorized into two approaches: (1) in vitro isolation and extraction of individual tissue/cell types and (2) in situ, including in vivo, analysis using an imaging approach. The suite of technologies available for in situ imaging is enormously powerful and varied, including Fourier transform infrared spectroscopy, magnetic resonance imaging, electron microscopy, histochemical and immunolabeling approaches coupled to optical and fluorescence microscopy, and X-ray fluorescence microscopy, with each approach taking advantage of different physical and chemical properties of the underlying tissue to provide unique insights. MSI has a number of advantages over other imaging modalities which are directly derived from the capabilities of modern MS instrumentation, which provide molecular specificity, high sensitivity for select analytes, and the ability to measure a broad range of analytes at high mass-resolving power with high mass accuracy across wide mass ranges. Even with these advantages, it is still a challenge to provide the depth of coverage that may be achieved from alternative approaches. MSI can provide very high lateral resolutions for imaging, giving the ability to distinguish the molecular nature of fine morphological features within tissues, even down to the single-cell level. Certain MSI approaches take advantage of minimal or no sample preparation steps with a number capable of ionization directly off sample surfaces. While our spatial resolution during MALDI imaging experiments has improved in recent years, other imaging modalities can achieve higher spatial resolution, and Caprioli and coworkers have sought to integrate, or fuse if you like, these different modalities to combine the

strengths of different imaging modes for a better outcome [16]. For example, they showed that correlating modest resolution MALDI imaging (100 μm) with optical scans of H&E-stained tissue allowed the prediction of regions of interest at higher spatial resolution – the prediction at 10 μm was verified by MSI acquired at 10 μm on a serial section. This advance has the potential to combine the specificity of MALDI-MSI with optical images and other modes of imaging. The other advantage of being able to use modest resolution MALDI-MSI is that the sample throughput can be higher, as very high spatial resolution MSI on large tissue sections is not fast enough to be considered high throughput on most MALDI instruments. Recent instrumental advances have also increased the speed of which data can be collected.

MSI was first applied to biomedical imaging [17, 18] in the mid-1990s corresponding with the introduction of soft ionization techniques, in particular MALDI. MSI has significantly advanced, providing both high lateral (spatial) and high mass resolution capabilities using a variety of different ion sources and approaches. MSI has found extensive use in molecular pathology and histology where the technique is used to map the spatial distribution of proteins and small molecules including drugs, lipids, and endogenous metabolites within tissues [1, 12]. MSI has been demonstrated to have a number of advantages, including a label-free analysis and the simultaneous multiplex measurement of 100 to possibly 1000 of analytes in a single imaging experiment, providing rich high density multidimensional data. Combination of MSI with advanced software and data analysis techniques now allows the virtual microdissection and interrogation of the molecular makeup of individual tissues. Lately, advances in spatial resolution have placed MSI at the forefront of single-cell metabolomics [19, 20], demonstrating an ability to measure the metabolism of an individual specialized cell within a subpopulation of cells. The development of novel data analysis techniques is opening the doors to conducting spatial metabolomics and comparative statistics across multiple samples allowing exploration of molecular changes during disease processes and identification of biomarkers [21, 22].

12.2   Mass Spectrometry Imaging

There are four essential steps in a basic MSI experiment: (1) sample selection and preparation, (2) desorption and ionization, (3) mass analysis, and (4) image registration and data analysis [14]. Careful control of each is essential to enable generation of high-quality images. In particular, sample selection, storage, and preparation have a disproportionate impact on the final results; there are many potential pitfalls that must be avoided as many sample preparation steps or techniques have the potential to contaminate the tissue section with exogenous material affecting reproducibility, ionization, and image quality. Fundamentally, the MSI process involves placing a suitable tissue section into an ion source, ionizing the sample and collecting a series of position-correlated mass spectra. This series of individual mass

spectra is collected in a two-dimensional (2D) array across surface of the sample using one of a range of different ion sources and mass analyzers. The most common approach is a microprobe approach where for each spatial coordinate, a single corresponding mass spectrum is collected. The resulting mass spectra represent the intensities of ionizable molecules present as their mass-to-charge ratios (*m/z*) which are then correlated with a high-resolution optical image of the tissue or histochemical stain with each spectrum assigned as an individual pixel for image generation. When the intensity value of each respective ion is plotted as an intensity map across the 2D array, the resultant reconstructed ion image represents the spatial distribution of the corresponding molecule(s). Three-dimensional (3D) approaches are also possible where serial 2D arrays from sequential tissue sections (or depth profiling) from the one tissue sample are measured and then a 3D volume is reconstructed computationally [23–25].

12.2.1 Ionization and Mass Analysis

MSI first relies on the ability to form ions that are then transferred under vacuum and measured by the mass analyzer. Currently, the dominant ion source and approach is MALDI, due to a range of commercially available instruments, which are capable of delivering high spatial and mass resolution, ease of use, and broad range of applicability to a variety of biological applications (Fig. 12.1). In practice, lateral resolutions for MALDI instruments are in the range 5–50 μm. The past 3–5 years have seen an explosion in different types of ion sources available, including atmospheric pressure MALDI (AP-MALDI) and other specialized sources for ambient ionization conditions [26]. Further, a number of popular alternative ion sources exist including SIMS, desorption electrospray ionization (DESI), nano-DESI, laser ablation electrospray ionization (LAESI), and atmospheric pressure MALDI. When undertaking MSI at very high spatial resolution, there is a significant trade-off with sensitivity, because the decreased sampling area will reduce the total number of ions available for detection. In short, MSI experiments will usually involve some kind of trade-off between spatial resolution and sensitivity; however, advanced mass analyzers and detectors are now allowing the measurement of very low numbers of ions that to some degree mitigates losses in sensitivity at high lateral resolution.

The mass analyzer is the core component of a mass spectrometer, enabling determination of mass-to-charge ratio (*m/z*) of an ion. The type of mass analyzer used and spectral resolution also have a direct impact on the ability to conduct MSI experiments (Table 12.1). The most common mass analyzers used on MSI instruments include time of flight (TOF) and Fourier transform (FT), encompassing both orbitrap and ion cyclotron resonance (FTICR) instruments. To distinguish differing metabolites in tissues, there is a clear need for accurate mass and high mass-resolving instruments and/or the use of tandem MS. Low mass resolution instruments can lead to misidentification or misinterpretation due to inability to resolve peaks of similar mass in MS scans. Ion traps have been used for imaging studies,
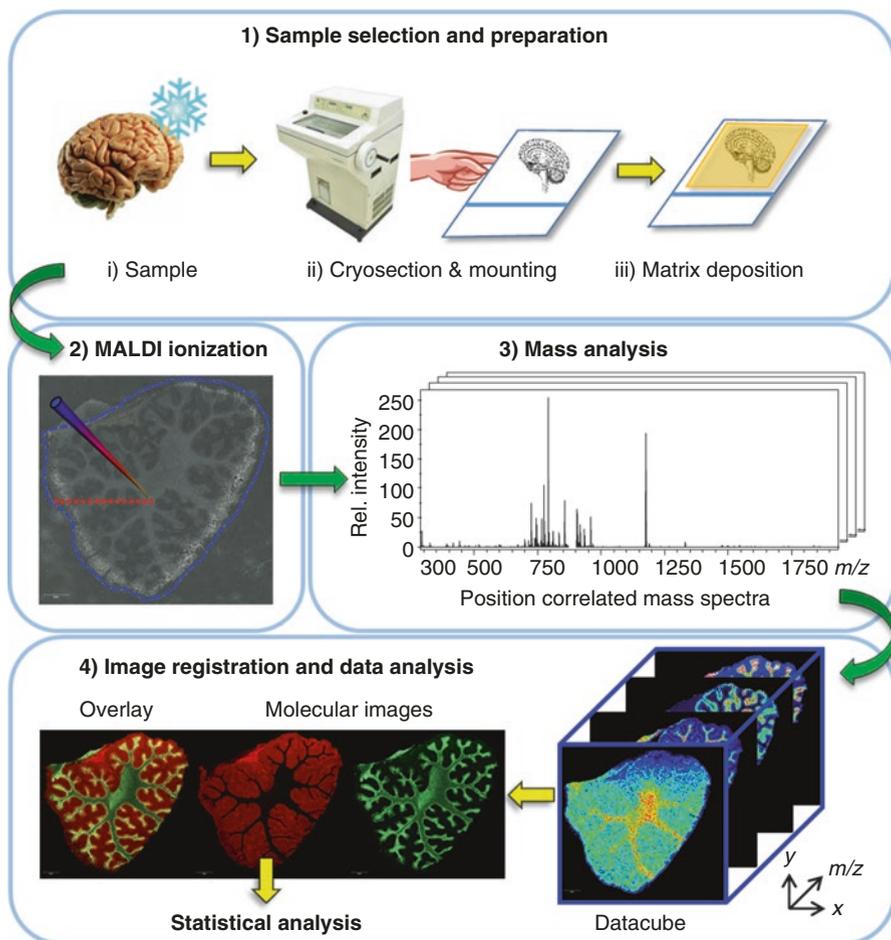
Fig. 12.1 Matrix-assisted laser desorption ionization (MALDI) approach. (*1*) Sample selection and preparation: sample tissues are first frozen and then cryo-sectioned with resulting thin sections mounted directly to a target. A thin layer of chemical matrix is typically applied across the surface of the tissue using a spray deposition or sublimation approach; (*2*) MALDI: molecules are desorbed from the surface by preferential absorption of UV or IR light energy by the matrix, localized phase transfer generates an evolving gas plume, ions may be pre-formed in the solid phase or generated in the gas phase by ion addition or abstraction from the respective analyte; (*3*) mass analysis: individual position-correlated mass spectra are collected in a uniform array; (*4*) image registration and data analysis: spectra are combined to generate a data cube. MSI data is then defined by *x,y* location, *m/z*, and ion signal intensity. Individual ion images are reconstructed by plotting the ion signal intensity for a single (or multiple) *m/z* as a false color image across the 2D grid correlated with an optical image. Further statistical analysis may be conducted to identify spatial segmentation or comparative analysis across sections

Table 12.1 List of common mass analyzers and instrument configurations detailing: mass resolution, approximate mass range, MS/MS capabilities, and acquisition speed

Mass analyzer/ configuration	Mass resolution	Mass range (Da)	MS/MS	MS^n	Acquisition speed
Ion trap	~1000	50–4000	Yes	Yes	Medium
TOF	2500–40,000	20–500,000	No	No	Fast
TOF/TOF	>20,000	20–500,000	Yes	No	Fast/very fast
IT-TOF	10,000	50–20,000	Yes	Yes	Fast
IT-orbitrap	>100,000	40–4000	Yes	Yes	Slow
FTICR	>200,000	10–10,000	Yes	Yes	Slow
Ion mobility QTOF	13,000/40,000	Up to 40,000	Yes	No	Fast

TOF time of flight, *TOF/TOF* tandem TOF, *IT* ion trap, *FTICR* Fourier transform ion cyclotron resonance, *QTOF* quadrupole time of flight, *Da* dalton

and while being lower resolution mass analyzers, they do offer the possibility to perform targeted MSI experiments at MS^2 or MS^n level, where product ions can monitored. This approach can be useful if there is sufficient signal for MS^n events, and even though the MSI would be displayed for a particular product ion, whole MS and MS^n spectra are acquired, meaning compound ID can be made by comparing full MS^n spectra acquired during MSI analysis to spectra from standard materials. Undoubtedly, higher mass resolution is very useful in MSI experiments, where sample purification and fractionation are not possible and sample cleanup is limited, but MS^n experiments using ion traps also have potential to identify compounds during MSI analysis.

The ability of a mass spectrometer to distinguish one mass peak from an ion close in mass is described by both mass resolution and resolving power (RP). MSI experiments are less sensitive than analyses that orthogonally separate analytes prior to measurement and detection; this is directly due to the extremely complex biological matrix of the tissues where vast concentration ranges of chemical entities are present with differing chemistries and molecular sizes (e.g., proteins, lipids, organic acids, amino acids, carbohydrates, inorganic ions, etc.). For MALDI experiments, the presence of high abundance low-molecular mass ions generated directly from the matrix employed can lead to significant interfering signal. Higher mass resolution allows easier identification of contributing ions and exclusion of interference from the presence of other chemical entities. Higher mass-resolving power is essential for high mass accuracy, whereby a higher RP allows identification of the center of peak and determination of the mass error, with low mass error allowing unambiguous assignment of a molecular formula aiding in identification. Modern high-resolution instruments are capable of <10 parts per million (ppm) mass error for TOF and <2 ppm mass error for FT instruments. Measurements conducted on low mass resolution instruments are typically operated in a targeted

tandem MS approach to provide molecular selectivity where specific fragment ions of single analytes are monitored, providing both molecular specificity and increased sensitivity. MSI measurements using higher resolution detectors provide the ability to unambiguously resolve a peak from the complex spectra that is generated allowing untargeted profiling-type techniques. The other exciting aspect of performing MSI with mass analyzers offering extreme resolution is that in addition to the mass accuracy measurement and the oft referred to isotope envelope, one can also measure the fine structure of each isotope peak. This fine structure is due to the very small differences in mass that exist in naturally occurring isotopes, for example, ^{34}S will be slightly lower in mass than ^{18}O, and if our analyzer is capable of very high resolution, these species can be resolved and observed separately. This fine structure is characteristic of the chemical composition and can be used along with mass accuracy and traditional isotope envelope measurement to confirm molecular formulae. This approach has shown great utility in the metabolite world. For protein samples, an on-tissue digestion would be required, given the mass range of FTICR analyzers. This approach is very exciting too, because it opens the possibility of "peptide mass fingerprint"-type experiments – whereby if we confirm the presence of several peptides that emanate from an individual protein, we essentially have protein identification during an MSI analysis. This approach is important because it removes the requirement for successful MS^2 experiments.

A hybrid approach that uses ion mobility coupled to mass spectrometry (IM-MS) that first separates ions by their mobility in a carrier gas followed by detection by MS has recently been developed [27, 28]. IM-MS offers the ability to orthogonally separate ions in the gas phase with similar *m/z* but different shapes via collisional cross section (CCS), providing a number of benefits including better signal to noise ratio (S/N) and the potential to separate isomers according to their shape and charge [29–32]. The application of IM-MS to MALDI-MSI experiments provides much promise for the analysis of lipids, peptides, and proteins; however, the benefits of IM-MS for small molecule analysis are slowly being unveiled as higher ion mobility resolving instruments are developed.

12.2.2   Sample Preparation

Prior to analysis, tissues must be collected and stored. The steps taken during both tissue collection and storage are critical for successful MSI analysis and often vary depending upon the analyte of interest. Most experiments will have a distinct timing mismatch between sample collection and analysis, requiring the storage of samples for a period of time. For most MSI analyses, tissue samples are typically flash-frozen to quench metabolism and to retain the spatial distribution of analytes and are sectioned or prepared at a later time point. Care must be taken to retain the tissue morphology during the freezing process and to preserve an accurate representation of the native tissue; soft tissues may deform and take the shape of the container

(tube or tray) within which they are frozen. Typically, to protect delicate tissues, structures, and small metabolites, a gentle freezing approach is recommended, including freezing in the atmosphere over liquid nitrogen or in cold carbon dioxide atmosphere over dry ice; alternatively, samples may be dipped into isopentane/liquid nitrogen or isopentane/dry ice slurries. Alternatively, a number of heat and microwave tissue stabilization methods have been developed for proteins and peptides [33–35].

Once samples are frozen, tissues and analytes are generally stable for months to years when stored at −80 °C. Embedding tissues within an external matrix is a common approach and is often required to ensure that suitable sections are generated from fragile tissue types that may have a tendency to fracture and crumble during sectioning. A number of different embedding media have been successfully demonstrated, including agarose [36], gelatin [20, 37–41], and aqueous carboxymethylcellulose solutions (1–5%) [42, 43]. In general, the easier the frozen matrix is to section or the closer the properties of the matrix are to the tissue being sectioned, the easier it will be to generate suitable sections of tissue for analysis.

Standard histological workflows utilize optimal cutting temperature (OCT) compound (a solution containing ~4% polyethylene glycols (PEG)) as an embedding medium, but this is strongly discouraged for MSI research due to absorption into the tissue and smearing of OCT across the tissue surface during cryo-sectioning, which has been shown to directly lead to ion suppression effects and loss of analyte signals [44].

While cryo-sectioning is the most commonly used method for sample preparation to access internal metabolites, there are other alternatives for tissue sectioning. Depending on the analysis method and instrument used, tissues must be prepared differently for imaging purposes, and a number of factors must be considered. External surfaces can be readily analyzed by mounting tissues directly to sample stages using double-sided tape, but for the measurement of internal distributions of metabolites, tissues must first be sectioned at an appropriate thickness to expose the underlying tissue. In particular, the type of analytes and their stability and turnover must be considered. Both the sample height and morphology may have a large effect upon the number of ions generated (due to laser focusing) and, for linear TOF instruments (LDI and MALDI), mass accuracy and resolution (due to changes in flight path length). Instruments where the detector is decoupled from the source, such as QIT, LIT, FTICR, and orbitrap instruments, are not reliant upon the sample thickness and are only limited by the physical configuration of the sample stage.

An established technique for generating thin sections from hard tissues has been recently adapted to MSI applications for delicate and difficult tissues [45]. The Kawamoto method uses an adhesive film to capture thin sections during cryo-sectioning. Once the tissue is adhered to the film, it can be transferred then fixed to a standard slide and prepared in the normal manner for MSI [46, 47].

For previously fixed tissue samples, there are a number of sample preparation protocols that have been developed for formalin-fixed paraffin-embedded (FFPE) mammalian tissue specifically for MSI analysis [48]. Previously, FFPE tissues have

been considered only suitable for examination of the distribution of metals, proteins, peptides, and other polymeric biomolecules in tissues due to the fixation extracting and degrading small molecules. More recently, the possibility of imaging small molecules from FFPE tissues has been demonstrated [49–51]. For proteins, peptides, and glycogens, further tissue preparation steps are required to retrieve antigens lengthening the sample preparation process.

Some tissue types can be very difficult to frozen section, such as secretory tissues, among other things. These tissue types benefit from a tissue fixation approach; however, FFPE fixation using formaldehyde renders the intact proteins inaccessible. This can be a problem especially for studies where transcriptome libraries do not exist – as the peptides that would be observed after antigen retrieval and enzymatic digest are meaningless in the absence of a transcriptome library. One approach for these types of sample is to utilize a fixation approach that does not involve the protein cross-linking caused by formaldehyde [52]. RCL2 and PAXgene Tissue are two products that can be used to fix tissue, which do not cross-link the proteins. Once fixed, the tissue is dehydrated (ethanol gradient) and cleared (xylene) prior to paraffin impregnation in much the same manner as routine tissue processing. This approach results in a paraffin-embedded tissue that can be sectioned very easily using a microtome. The only caveat is that the sectioned tissue cannot be floated on a water bath for mounting onto a slide, as the proteins are soluble. The carefully placed tissue section is heat mounted to a glass slide and then deparaffinized using xylene. At this stage, the tissue can have matrix applied in the same manner as any other tissue. For protein analysis, the samples are very good because the dehydration and clearing remove the lipids and other species that often reduce the sensitivity during an MSI analysis. However, the approach is clearly not ideal for analyte classes soluble in ethanol or xylene.

Once mounted to the sample carrier, the tissues are typically dehydrated under vacuum prior to either matrix deposition or direct analysis. Prior dehydration avoids any shrinkage of tissues leading to changes in sample morphology within the instrument. In MALDI-MSI using TOF detection, where a voltage is applied to the sample stage, samples are usually mounted either on glass slides coated with conductive indium tin oxide (ITO) or on reusable metal sample stages (steel or gold-coated steel). Samples are either directly freeze-thaw mounted to the surface or adhered using conductive double-sided tape [53]. Freeze-thaw mounting is generally performed by transferring the cut tissue section to the top of the sample holder (slide, plate) and then gently warming the holder from the underside using body heat. The tissue section quickly thaws and adheres to the surface of the holder. Once mounted, the sections are warmed and transferred to a vacuum desiccation chamber and dried under reduced pressure for at least 15 min before any further steps are conducted. Tissue sections may degrade rapidly and must either be stored under vacuum or, for longer periods, at −80 °C [54]. For MALDI-MSI, application of the matrix has been shown to stabilize analytes within the tissue to oxidation and degradation processes.

12.2.3 Tissue Washing

A commonly accepted principle of MSI analysis is to conduct the minimal amount of sample preparation steps, to avoid metabolite degradation, and to retain the distribution of analytes. However, a number of tissue washing steps can be conducted to either increase the sensitivity for certain analytes or to remove background salts to decrease salt adducts [55–57]. Mounted sections can be carefully dipped into washing solutions and then dried, before further processing such as enzymatic digestion or application of matrix. These steps have been successfully employed to increase the ionization of selected metabolites (including lipids, proteins, and peptides) in mammalian systems.

12.2.4 MALDI Matrix Application and In Situ Protein Digestion Strategies

MALDI relies upon an exogenous matrix, consisting typically of either small organic molecules or inorganic UV absorbent nanoparticles, which must be applied by one of a number of different techniques. Further, the achievable lateral resolution is dependent upon the size of the matrix crystals, which is in turn dependent upon the application technique employed. There are a number of approaches used to apply a MALDI matrix that can be separated into two different strategies, involving either dry deposition or wet deposition and extraction. The first, dry deposition strategy, deposits the matrix without any solvents to the top surface of a tissue section by one of two common techniques, employing handshaking of dry fine crystals of matrix onto the sample through a sieve or the use of a sublimation apparatus. A sublimation approach for deposition of matrix provides very uniform coatings with very small crystal sizes (typically in the range of 1–5 μm), allowing imaging with high spatial resolution. It is becoming one of the preferred approaches for small-molecule and lipid imaging [58].

Wet deposition strategies have also had significant attention, and there are many different techniques available for specific analyte classes. Wet deposition is one of the most common techniques for matrix deposition for MALDI-MSI analysis and is essential to conduct in situ protein digests. To conduct an in situ protein digestion, a protease, generally trypsin or α-chymotrypsin, is deposited in a buffered solution. Once uniform application of enzyme has been achieved, the sample is incubated in a humid atmosphere for a period of time, to allow localized digestion before drying and matrix application for MALDI-MSI. Matrix is first dissolved in a suitable solvent, then small droplets are applied to the surface of the tissue to be imaged, micro-extraction of endogenous molecules takes place at the solvent-tissue interface, and, as the solvent dries, analytes co-crystallize with the dissolved

matrix. The achievable lateral resolution of a wet deposition technique is predominantly dependent upon the droplet size maintained during matrix deposition. There are several different techniques reported in the literature, including homemade solutions and a range of commercially available instruments, ranging from manual airbrushing (where success is highly dependent upon the operator) to more controlled robotic spraying (HTX Imaging TM-Sprayer, HTX Technologies LLC, Carrboro, NC, USA; SunChrom SunCollect and SunCollect II plus+, SunChrom GmbH, Friedrichsdorf, Germany), automatic droplet deposition through piezo-electric vibration (ImagePrep, Bruker, Bremen, Germany), inkjet printing (ChIP 1000, Shimadzu Corp., Japan) with standard inkjet printers [59], robotic spotting (Labcyte Portrait 630 Spotter – no longer available), and automatic protein digestion robots (SunChrom SunDigest, SunCollect II plus+, SunChrom). Once deposition conditions have been optimized for specific solvents, matrix and concentration, number of passes or spray cycles, temperatures, and drying, it is possible to achieve very small crystal sizes of 5–20+ μm (in the longest dimension), allowing high-resolution imaging. A combination approach of initial dry deposition using sublimation followed by in situ "rehydration/recrystallization" by vapor exchange provides excellent results for protein and peptide imaging [7].

12.2.5 Matrices for MALDI Analysis

There are a large number of matrices that are either in common use or have been recently reported in the literature for MALDI, including the main stalwarts 2,5-dihydroxybenzoic acid (DHB) [60], 2,5-dihydroxyacetophenone (DHAP) [61], sinapinic acid (SA) [62, 63], and α-cyano-4-hydroxycinnamic acid (CHCA) [64–66], which are typically used for positive-mode MALDI analysis. Recently, lithium salts of DHB, SA, CHCA, and vanillin have been demonstrated as suitable matrices for imaging hydrocarbons as the lithiated adduct [67]. 9-Aminoacridine (9-AA) [41, 68], 1,8-bis-dimethylaminonaphthalene (DMAN) [38, 69], and 1,5-diaminonaphthalene (DAN) [20, 41, 60] were reported for negative-mode analyses. 2-Aminoethyl-N-2-aminonaphthalene has also been reported as a suitable matrix [70]. Recent use of the plant metabolites quercetin and morin [71], which are structural isomers, as matrices for both positive- and negative-mode analysis, has demonstrated vastly increased detection of phospholipids in mammalian tissues when using high-resolution FTICR-MS.

More recently, DAN has been adopted for plant-based imaging, which requires very low laser energy and very small crystal size [41]. DAN has been used for MSI imaging in both positive and negative modes at very high spatial resolution (however, caution is required when using DAN as it is suspected to be a carcinogen). Further, DAN is also chemically reactive with the ability to form gas phase radicals, to induce in-source decay, and to conduct gas phase reductions of disulfide bonds [72, 73]. The use of an ambient-pressure MALDI source allows the use of volatile matrices, including liquid ion matrices and also water in the form of ice for IR-MALDI within frozen tissues [74]. Nanoparticles and colloids have been

reported as suitable matrices for MALDI-MSI, including the use of silver and gold nanoparticles for the imaging of waxes and phospholipids [75–79]. Furthermore, functional iron nanoparticles (fNPs) have been demonstrated in mammalian tissues [80]. In the case of small-molecule matrices, these can be readily removed post-MSI acquisition, washed with a suitable solvent such as ethanol or aqueous solutions, and then subjected to histochemical staining [7].

12.3   Data Analysis

12.3.1   *Analytical Software and Data Analysis Techniques*

MSI experiments generate huge volumes and highly complex data; due to these properties, there is a requirement for advanced software and computational data analysis techniques to extract meaningful results from the data. Data analysis of MSI datasets was in the beginning largely limited to manual identification and mapping of individual ions but has in recent years advanced significantly and to incorporate advanced clustering and comparative visualization tools allowing spatial segmentation, identification, and comparison of multiple ions. Commercial data analysis packages include BioMap (Novartis, Basel, Switzerland), FlexImaging and ClinProTools (Bruker Daltonik, Bremen, Germany), HDI (high-definition MALDI MS imaging) coupled to MassLynx and MarkerLynx (Waters, Manchester UK), ImageQuest (Thermo Scientific, Waltham, MA, USA), MALDIVision (PREMIER Biosoft), SCiLS Lab (SCiLS Bremen, Germany), and TissueView (AB Sciex, based on BioMap). Recent adoption of the common mzML data format standard (www.imzml.org) [81] by instrument vendors and incorporation into a variety of tools or directly into the vendor software (such as FlexImaging) has allowed export of instrument-specific data into a common format, which has aided the development of vendor-independent tools for data analysis and application of advanced statistical techniques to identify underlying metabolite distributions and co-localizations. Open-source software packages include Datacube Explorer (FOM-AMOLF, Amsterdam, Netherlands) [82], Metabolite Imager (University of Texas) [83], MIRION (Justus Liebig University) [84], MSiReader (North Carolina State University) [85], OpenMSI (Lawrence Berkeley National Lab, CA, USA, http//openmsi.nersc.gov) [86], Cardinal [87], SpectViewer (www.maldi-msi.org), OmniSpect [88], MSIQuant [89], LabMSI [90], MSI.R [91], and MALDIquant [92]. Many of the current packages for MS image analysis have been developed incorporating only visualization and simple clustering techniques such as hierarchical cluster analysis (HCA) and principal component analysis (PCA).

Due to the inherent heterogeneity of MSI data, preprocessing and spectral "denoising" are recommended to obtain better results [93–95]. Preprocessing includes steps for baseline subtraction and smoothing, peak alignment and mass recalibration across the entire dataset, normalization of signal intensity, peak-picking, and data

reduction steps. A number of publications have provided detailed analysis pathways and suitable tools to examine MSI data [86, 93]. Once preprocessing steps are complete, there are three types of unsupervised approaches to identify hidden patterns and spatial distributions of metabolites: component analysis, spatial segmentation, and self-organizing maps. The first, component analysis, has been dominated by the use of principal component analysis (PCA), although other methods have been used to uncover the variation in MALDI-MSI data, including nonnegative matrix factorization, maximum autocorrelation factorization, and latent semantic analysis (see review by [93]). PCA represents the spatial patterns of molecules in terms of the set of score images, but PCA has a number of limitations including negative values (which are not present in the data) and difficulty in determining co-localized ion images for identified patterns of distribution. Spatial segmentation is a robust approach to examine MSI data where a segmentation map displays different regions in the tissues with distinct molecular composition [93]. A common approach is to use hierarchical cluster analysis (HCA), which is directly incorporated into FlexImaging. More recently, advanced spatial segmentation clustering techniques have been developed that cluster m/z values with distinct regions of the tissue [21, 94] and are incorporated directly into the commercial software SCiLS Lab. The third area is an emerging data analysis technique that makes use of unsupervised self-organizing maps (SOM) [96, 97] and growing self-organizing maps [98] that reduce the dimensionality of the data and allow identification of hidden patterns within the data.

Three-dimensional mass spectrometry imaging (3D-MSI) has been reported [99–101] and reviewed previously [5]. 3D-MSI is conducted using one of two approaches: (1) depth profiling on the same tissues by conducting sequential rastering events [5], which is common for SIMS [102, 103] but has also been reported for laser ablation electrospray ionization, which was used to depth profile plant leaf tissue [23], or (2) by combining multiple two-dimensional MSI measurements conducted on serial tissue sections from a single sample. Individual datasets are computationally reassembled to generate 3D volume reconstructions of individual ion distributions; for this purpose, researchers have used software such as Amira (www.fei.com), Image J (imagej.nih.gov/ij), MATLAB (www.mathworks.com), and more recently SCiLS Lab (www.scils.de) to generate 3D images.

12.3.2 Reporting Standards and Online Repositories

Recent guidelines for the reporting of MSI datasets have been published [104]. The article outlines the detailed metadata and contextualizing of information that is required to fully describe an MSI dataset, and it provides eight specific reportable areas: (1) tissue samples, including the type and how the tissue was sampled; (2) tissue preparation, including methods such as washing and matrix application steps; (3) optical image, detailing information about the corresponding optical images used for MSI analysis; (4) data acquisition, detailing the instrument and parameters used to acquire the data; (5) mass spectra preprocessing, detailing the parameters used

to baseline subtract, to smooth, and to align spectra, for intensity normalization methods, for peak picking, and for data reduction methods; (6) MSI visualization, including methods for peak picking and image generation parameters; (7) compound identification, including all procedures used to identify individual metabolites; and (8) data analysis, detailing procedures, methods, and software used. Current reporting standards for identification in metabolomics experiments, including definitions for tentative, putative, and confirmed identification, have been previously published [105] and at the time of publication are currently being reviewed and updated. For MSI experiments, the ability to confirm identifications is all the more difficult due to the inability to separate isobaric compounds. Future release of reporting standards for MSI experiments in 2017 will provide detailed guidelines for MSI identification strategies. A common public repository has also recently been announced, where MSI datasets can be deposited for storage and later retrieval [106]. More recently, SCiLS Lab has announced SCiLS in the Cloud (www.scils-lab.com), an online engine capable of sharing imaging and statistical analysis results in collaborative manner. A spatial metabolomics analysis server has been released by the Alexandrov group and is available at www.alpha.metasp.eu; the OpenMSI project also offers online data analysis and sharing (http://openmsi.nersc.gov).

12.4   Applications

MALDI-MSI has been extensively deployed in biomedical research with several 1000 studies published since the early 1990s. A PubMed (http://www.ncbi.nlm.nih.gov/pubmed) survey of recently published literature over the years 2013 to June 2016 returns 833 publications with MALDI mass spectrometry imaging. A selection of publications from the total are referenced below along with examples of imaging lipids, peptides, and special metabolites in novel, complex, and difficult tissue types. Further, filtering to within the clinical space shows a broad range of research areas, a large number of different applications, and a dominance of the use of MALDI instruments for analysis (Fig. 12.2). A third of publications focus upon cancer ($n=104$) with nearly another third on neurological disease ($n=84$), reflecting this breakdown a third of analytes studies are lipids (34 %) followed by proteins and peptides (29 %) and then drugs and small molecules (24 %). The breakdown reflects that lipids derived from biological membranes are readily abundant, require few sample preparation steps, and are easily observed by MALDI-MSI.

12.4.1   Lipids

Many disease models in cancer and neurobiology display significant changes in the lipid profile reflecting dramatic changes in lipid metabolism [6, 21, 107–112]. For these types of analysis, samples require relatively few preparation steps,
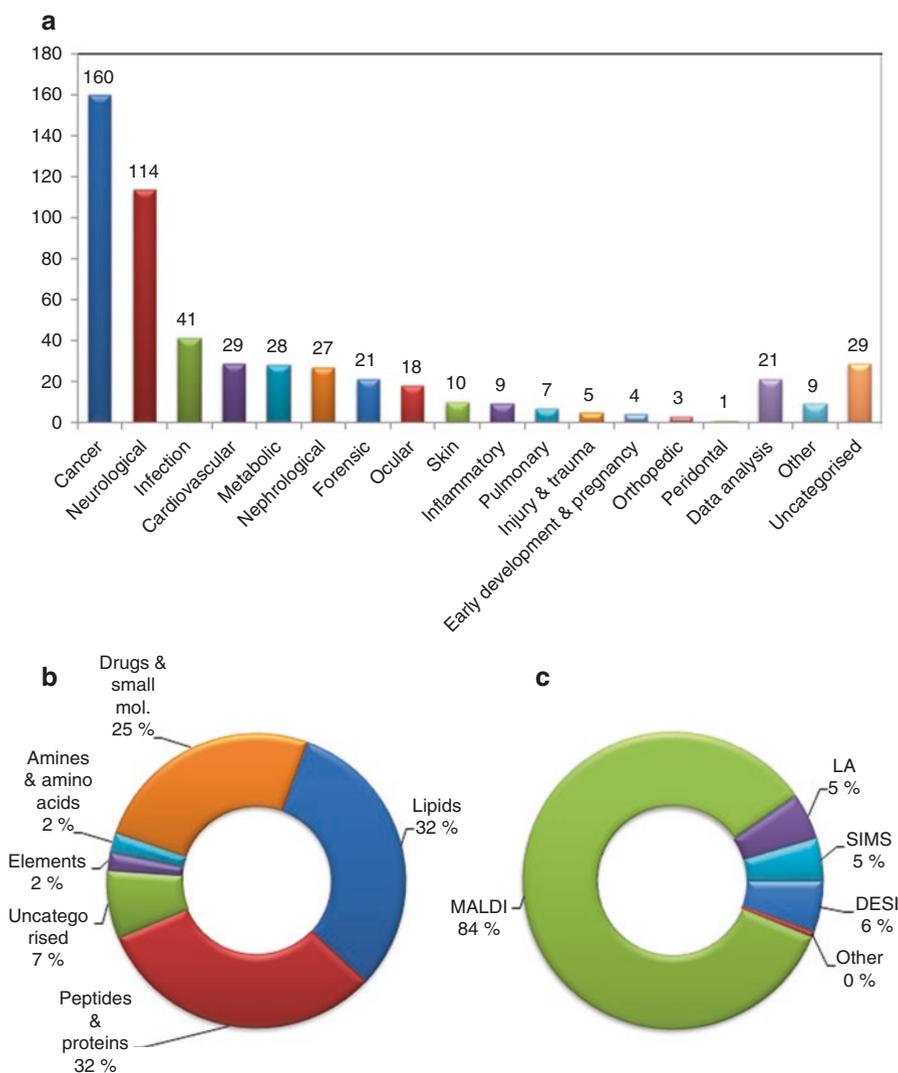
Fig. 12.2 (**a**) Total numbers of publications by disease type and research area over 2013–2016, (**b**) Type of analyte, (**c**) Type of ionization source

cryo-sectioning of fresh frozen tissues, mounting, dehydration, and application of matrix prior to imaging. An example of the complex distribution of lipid within class and between differing classes, including fatty acids, phospholipids, ceramides, and gangliosides in kangaroo cerebellum, is shown in Fig. 12.3. Analysis in negative ionization mode is capable of tentatively identifying up to 236 different lipids and metabolites. Lipids and metabolites were identified by using the Metaspace metabolite annotation engine (www.alpha.metasp.eu) using an accepted FDR of
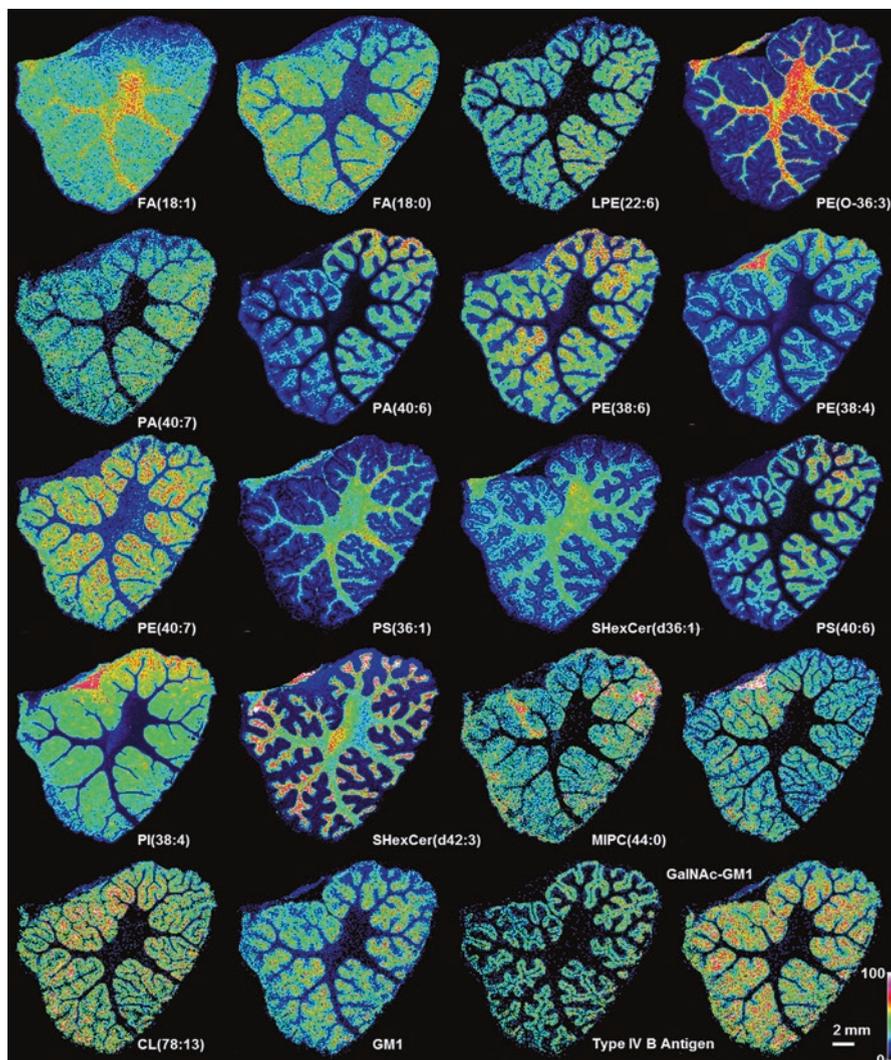
Fig. 12.3 Distribution of different lipid classes in kangaroo cerebellum. Lipids and metabolites were identified by using the Metaspace metabolite annotation engine (www.alpha.metasp.eu) using an accepted FDR of <0.2<0.1 and an accepted mass error of <5 ppm. Sagittal section of kangaroo brain cerebellum, 20 μm thick section, thaw mounted to glass slide with 1,8-bis (pyrro-lidinyl) naphthalene matrix (5 mg mL^{-1} in acetone) applied by spray deposition using a HTX TM-Sprayer (8 passes, 150 μL min^{-1} flow rate, 2 mm track spacing with 1 mm offset for repeat passes and 90° offset for alternate passes). Data generated on a 7 T Bruker SolariX XR MALDI-FTICR-MS in negative ionization mode, 150×150 μm spot array, 150,000 mass resolution at 400 m/z. Images were generated in Compass flexImaging 4.1 employing TIC normalization and scaled from 0 to 100% of maximum ion intensity for respective ions.

<0.1. There are distinct differences in the distribution of two of the most common and simple fatty acids FA (18:0) vs. FA (18:1), differing only by a single unsaturation. The unsaturated FA (18:1) is found in high amounts within the white matter and distributed throughout the gray matter vs. the fully saturated species FA (18:0) having a preferential distribution to the gray matter. Differences in the distribution of phospholipids, PA, PE, PI, PS, and cardiolipins, ceramides including sulfated species, MIPC, and various gangliosides are observed; in particular, the simpler sulfated hexose ceramides (SHexCer (d36:1), SHexCer (d42:3)) are found in the axon-rich white matter versus the more complex MIPC and ganglioside species (GM1, Type IV Antigen, GalNAc-GM1) found in the gray matter where the majority of the neuronal cell bodies are found.

12.4.2   Proteins and Peptides

Proteins, the biochemical engines of cells, and endogenous peptides including tachykinins, secretins, opioids, pancreatic peptides, and a range of other biochemically active peptides are the next most popular area of research in MSI [113–122]. Proteins can be imaged whole (but images tend to be dominated by the most abundant proteins), or for greater coverage, proteins are generally digested in situ to generate a series of peptides. A variety of different animal species produce venoms, which are cocktails of specialized peptide toxins, evolved for the capture of prey or defense against predators. Research into toxic venoms has developed into a significant area of study, in particular in the development of antivenoms and as a potential source of novel chemotherapeutics. More recently, spatial approaches have been applied to examine in vivo localization in the venom gland to better understand evolution of the toxic peptides and packaging for deployment. Within an organism venoms are generally generated in very delicate secretory tissues (glands) requiring highly specialized sample preparation methodology to preserve the structure and distribution of endogenous molecules. A recent study investigated the nature of the venoms present in the gland from a centipede, *Thereuopoda longicornis* (Fig. 12.4) [123]. Further, the study looked to determine whether compartmentalization in the gland existed. A venom gland was fixed using RCL2, dehydrated through an ethanol gradient, cleared with xylene, and impregnated with paraffin. The section was then deparaffinized using xylene prior to matrix application using an ImagePrep system (Bruker). Results demonstrated a heterogeneous distribution of differing venom peptides with the venom glands, providing insights into the evolution of venoms across centipede orders.

12.4.3   Endogenous Metabolites, Drugs, and Small Molecules

Imaging of endogenous metabolites and development of spatial metabolomics techniques and tools are a rapidly expanding area [124–126]. Special or highly specific metabolites from different species, labeled drugs, or compounds are
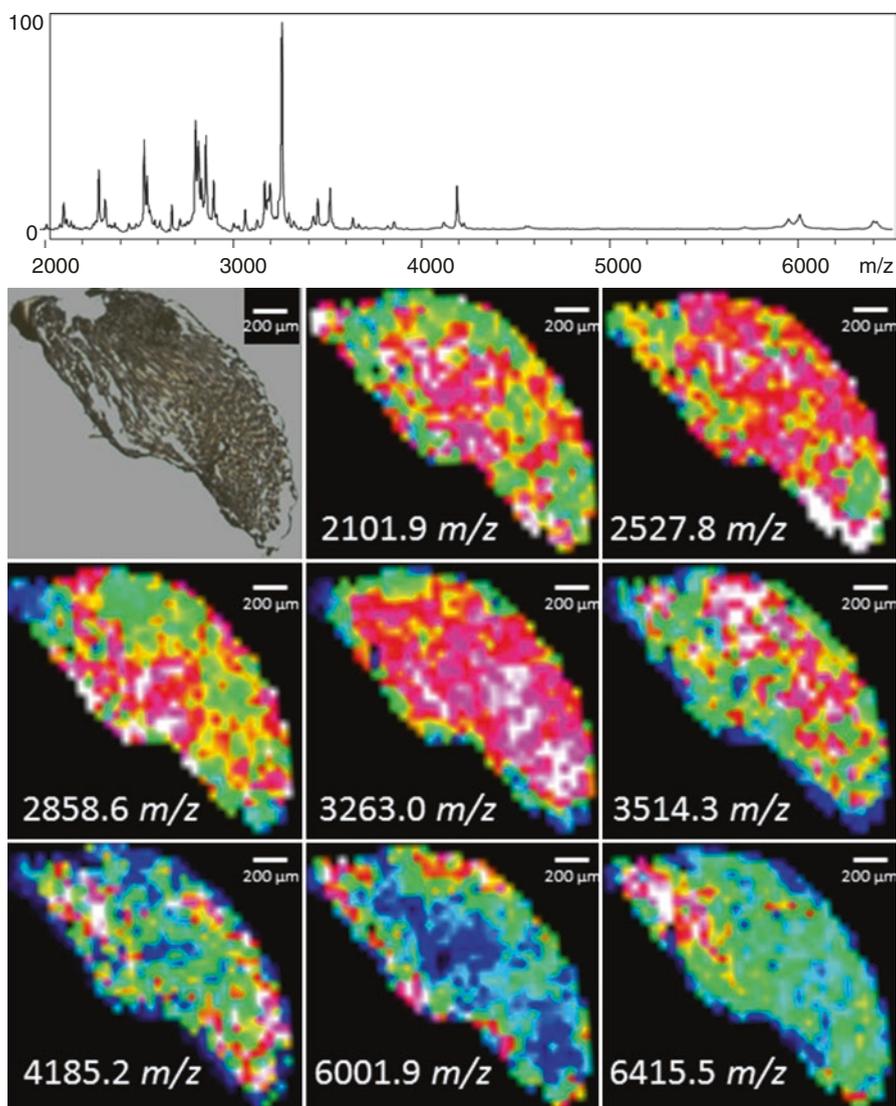
Fig. 12.4 MALDI-MSI experiment performed in linear positive mode at 50 μm resolution using CHCA as matrix. The sample is the venom gland from a centipede, *Thereuopoda longicornis*, which was fixed using RCL2, dehydrated through an ethanol gradient, cleared with xylene, and impregnated with paraffin. Section was deparaffinized using xylene prior to matrix application using an ImagePrep system (Bruker). The aim of the study was to investigate the nature of the venom present in the gland and to determine whether compartmentalization exists (Further details can be found in Undheim et al. [123], Reprinted by permission from PNAS 2015, Copyright © 2015)

attractive targets for small-molecule MSI [124, 127–139]. These types of compounds may be readily imaged by taking advantage of their specific chemical properties and use of high mass accuracy capabilities of MS. An example is brominated alkaloid analytes from marine sponge samples [140]. The brominated species were easy to observe due to the characteristic isotope pattern that bromine confers to organic molecules. Further, the sponge sample presented many difficulties to prepare for sectioning, as the brominated alkaloids were very soluble in organic solvents, making typical fixation approaches impossible. Sectioning was ultimately achieved by embedding in OCT, prior to frozen sectioning. Pieces of the frozen sponge were dropped into the OCT; thawed sponge sample allowed too much diffusion of OCT into the sample. The sections were washed in multiple rinses of water to remove as much OCT as possible due to the deleterious impact of OCT on collecting MS. Prepared sections had CHCA matrix applied using ImagePrep (Bruker). Figure 12.5 shows the average mass spectrum observed across the tissue along the distribution of two brominated analytes at *m/z* 619 and 574 – the third figure overlays these two compounds (619, blue; 574, green) to highlight the difference in their location across the tissue.

12.5 Future Directions

Kinetic mass spectrometric imaging (kMSI) has recently been developed as a new analytical approach to examine combined spatiotemporally resolved metabolism. A single MSI experiment provides only a static snapshot of the underlying molecular distribution of any metabolite. By incorporation of stable isotope labeling, metabolic flux within an organism can be examined and has been demonstrated for the turnover of and biosynthesis of lipids in a tumor model (Fig. 12.6) [141]. Multimodal imaging is an emerging theme, which involves combing two or more imaging modalities to provide deeper insights into biology. A simple form of multimodal imaging is already adopted in many MSI workflows which involves generating a histochemical stained section of tissue, either a serial section or in some cases the same piece of tissue on which an MSI measurement has been conducted and then co-registering a high-resolution optical images with the acquired MSI data. This approach provides more in-depth information (tissue/cell-type distribution) and can aid in sample interpretation. The combination of MALDI and SIMS has been used extensively in plant and animal MSI imaging [142–145], where the former has been used to generate lower resolution images across a wide area and SIMS used for very high-resolution imaging of a smaller subsection of the tissue. High-resolution magnetic resonance spectroscopic imaging (MRSI) has also been combined with MSI to examine choline metabolites and cations in tumor cells [146]. More recently, the hybrid predictive technique called image fusion has been reported and combines high spatial resolution but low chemical specificity information, such as images generated from optical microscopy at high magnification, coupled to lower spatial resolution but high chemical specificity
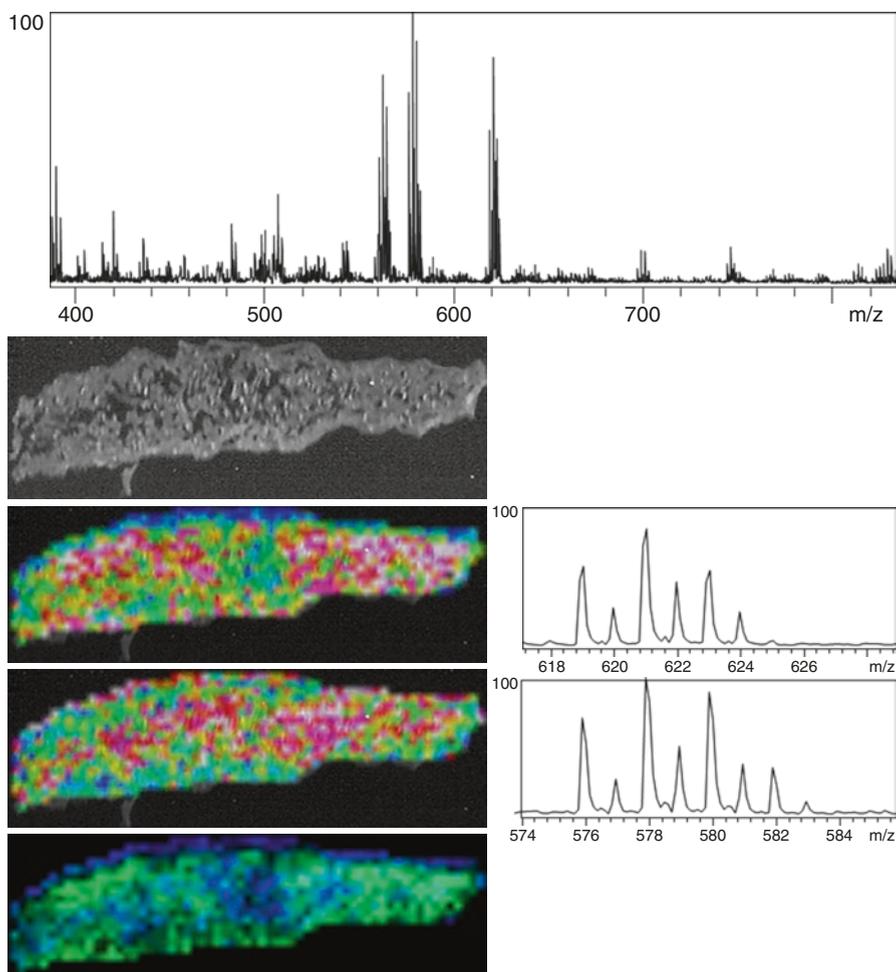
Fig. 12.5 MALDI-MSI of marine sponge, *Stylissa flabella*, at 200 μm spatial resolution over a mass range of 200–1100 *m/z* using CHCA as a matrix. The sponge sample presented many difficulties to prepare for sectioning; the analytes of interest, brominated alkaloids, were very soluble in organic solvents, making fixation approaches impossible. Sectioning was ultimately achieved by embedding in OCT, prior to frozen sectioning. Pieces of the frozen sponge were dropped into the OCT; thawed sponge sample allowed too much diffusion of OCT into the sample. The sections were washed in multiple rinses of water to remove as much OCT as possible. Prepared sections had CHCA matrix applied using ImagePrep (Bruker). Brominated analytes were easy to observe due to the characteristic isotope pattern that bromine confers to molecules. The figure above shows the average mass spectrum observed across the tissue along the distribution of two brominated analytes at *m/z* 619 and 574 – the third figure overlays these two compounds (619, blue; 574, green) to highlight the difference in their location across the tissue (Further information can be obtained in the following references – Yarnold et al. [140] (Reprinted with permission from Molecular Biosystems, Copyright © 2012)
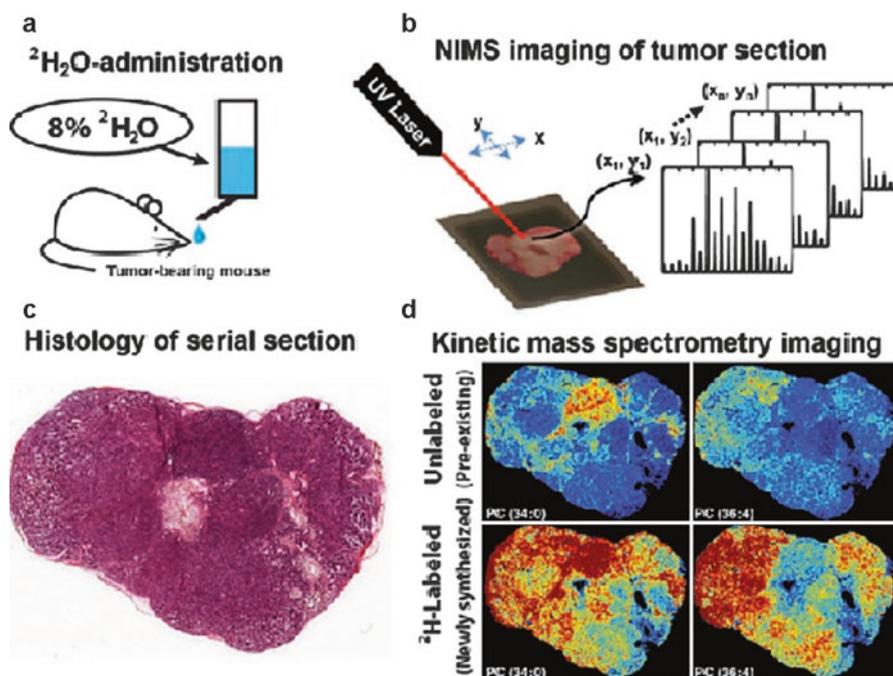
Fig. 12.6 Example of kinetic mass spectrometric imaging – experimental workflow for using kMSI to define spatial heterogeneity of lipid composition and biosynthesis. (**a**) A tumor-bearing mouse is administered ^2H_2O-enriched water to incorporate deuterium into tissue as a result of active metabolism. (**b**) The deuterium-enriched tumor is excised, sectioned, and imaged using NIMS. An individual mass spectrum is generated for each pixel every 50 μm, with spectra comprised of isotopologues from both ^2H-labeled and unlabeled lipid molecules. (**c**) Serial sections of the tumor are used for histopathology correlation with kMSI results. (**d**) Deconvolution of spectra is performed to separate ^2H-labeled and unlabeled lipids. Intensity images are generated to show the spatial distribution for both newly synthesized and preexisting lipids (Reprinted by permission from Macmillan Publishers Ltd: Scientific Reports, 3:1656 [141]. Copyright © 2013)

information, such as MSI data, to computationally predict the distribution of chemicals in the tissue sections [16]. New instrumentation is constantly being developed and recent developments include the Bruker RapiFlex TOF/TOF capable of high speed imaging (up to 80 pixels per second). Data generated from this instrument is being combined with ultrahigh mass resolution FTICR-MS imaging (relatively slow imaging) to take advantage of the benefits of each instrument to collect data quickly and provide molecular specificity [122]. Ion sources are being developed, including the MALDI-2-MS source, which incorporates a second post-ionization UV laser to generate gas phase photoionization of metabolites within the gas plume [147]. Data analysis remains a bottleneck; however, emerging MSI data analysis techniques that enable analysis of ultra-high-resolution MSI data and incorporate spatial segmentation will enhance discovery of spatially resolved metabolism. Further, development of unsupervised techniques that

utilize the spatial information within an MSI dataset and statistical techniques to discover co-occurring metabolites and significant differences in regions of tissue will unlock the power of MSI analysis speeding discovery processes.

12.6   Conclusion

MALDI-MSI has demonstrated application in a vast range of spatial biochemical and metabolomics research; the application of ultra-high-resolution and high mass accuracy MS provides the ability to distinguish molecular species very close in mass and accurately identifies molecular formula. High lateral resolution imaging is providing unique spatial insight into the distribution and function of many different analyte classes, and the rich, multidimensional, highly dense data is currently providing unique insights into the vast chemical complexity and specialization found within biological systems that is not possible using other methods. Challenges still exist including the development of technical methodology to examine specific classes of metabolites and advanced computational analysis to examine the data produced.

References

1. Spengler B. Mass spectrometry imaging of biomolecular information. Anal Chem. 2015;87(1):64–82.
2. Chaurand P. Imaging mass spectrometry of thin tissue sections: a decade of collective efforts. J Proteomics. 2012;75(16):4883–92.
3. Jungmann JH, Heeren RM. Emerging technologies in mass spectrometry imaging. J Proteomics. 2012;75(16):5077–92.
4. Miura D, Fujimura Y, Wariishi H. In situ metabolomic mass spectrometry imaging: recent advances and difficulties. J Proteomics. 2012;75(16):5052–60.
5. Seeley EH, Caprioli RM. 3D imaging by mass spectrometry: a new frontier. Anal Chem. 2012;84(5):2105–10.
6. Gode D, Volmer DA. Lipid imaging by mass spectrometry – a review. Analyst. 2013;138(5):1289–315.
7. Norris JL, Caprioli RM. Analysis of tissue specimens by matrix-assisted laser desorption/ ionization imaging mass spectrometry in biological and clinical research. Chem Rev. 2013;113(4):2309–42.
8. Rompp A, Spengler B. Mass spectrometry imaging with high resolution in mass and space. Histochem Cell Biol. 2013;139(6):759–83.

9. Wu C, Dill AL, Eberlin LS, Cooks RG, Ifa DR. Mass spectrometry imaging under ambient conditions. Mass Spectrom Rev. 2013;32(3):218–43.

10. Shariatgorji M, Svenningsson P, Andren PE. Mass spectrometry imaging, an emerging technology in neuropsychopharmacology. Neuropsychopharmacology. 2014;39(1):34–49.

11. Addie RD, Balluff B, Bovee JV, Morreau H, McDonnell LA. Current state and future challenges of mass spectrometry imaging for clinical research. Anal Chem. 2015;87(13):6426–33.

12. Aichler M, Walch A. MALDI imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. Lab Invest. 2015;95(4):422–31.

13. Chughtai K, Heeren RM. Mass spectrometric imaging for biomedical tissue analysis. Chem Rev. 2010;110(5):3237–77.

14. Boughton BA, Thinagaran D, Sarabia D, Bacic A, Roessner U. Mass spectrometry imaging for plant biology: a review. Phytochem Rev. 2016;15(3):445–88.

15. Sumner LW, Lei Z, Nikolau BJ, Saito K. Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. Nat Prod Rep. 2015;32(2):212–29.

16. Van de Plas R, Yang J, Spraggins J, Caprioli RM. Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. Nat Methods. 2015;12(4):366–72.

17. Spengler B, Kaufmann R. "Development of a new scanning UV-laser microprobe for Ion imaging and confocal microscopy", proceedings of the 42nd ASMS conference on mass spectrometry and allied topics: May 29-june 3, 1994. Chicago: ASMS; 1994.

18. Caprioli RM, Farmer TB, Gile J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal Chem. 1997;69(23):4751–60.

19. Thiery-Lavenant G, Zavalin AI, Caprioli RM. Targeted multiplex imaging mass spectrometry in transmission geometry for subcellular spatial resolution. J Am Soc Mass Spectrom. 2013;24(4):609–14.

20. Korte AR, Yandeau-Nelson MD, Nikolau BJ, Lee YJ. Subcellular-level resolution MALDI-MS imaging of maize leaf metabolites by MALDI-linear ion trap-Orbitrap mass spectrometer. Anal Bioanal Chem. 2015;407(8):2301–9.

21. Krasny L, Hoffmann F, Ernst G, Trede D, Alexandrov T, Havlicek V, et al. Spatial segmentation of MALDI FT-ICR MSI data: a powerful tool to explore the head and neck tumor in situ lipidome. J Am Soc Mass Spectrom. 2015;26(1):36–43.

22. Steurer S, Borkowski C, Odinga S, Buchholz M, Koop C, Huland H, et al. MALDI mass spectrometric imaging based identification of clinically relevant signals in prostate cancer using large-scale tissue microarrays. Int J Cancer. 2013;133(4):920–8.

23. Nemes P, Barton AA, Vertes A. Three-dimensional imaging of metabolites in tissues under ambient conditions by laser ablation electrospray ionization mass spectrometry. Anal Chem. 2009;81(16):6668–75.

24. Ye H, Greer T, Li L. From pixel to voxel: a deeper view of biological tissue by 3D mass spectral imaging. Bioanalysis. 2011;3(3):313–32.

25. Lanekoff I, Burnum-Johnson K, Thomas M, Cha J, Dey SK, Yang P, et al. Three-dimensional imaging of lipids and metabolites in tissues by nanospray desorption electrospray ionization mass spectrometry. Anal Bioanal Chem. 2015;407(8):2063–71.

26. Monge ME, Harris GA, Dwivedi P, Fernandez FM. Mass spectrometry: recent advances in direct open air surface sampling/ionization. Chem Rev. 2013;113(4):2269–308.

27. Stauber J, MacAleese L, Franck J, Claude E, Snel M, Kaletas BK, et al. On-tissue protein identification and imaging by MALDI-ion mobility mass spectrometry. J Am Soc Mass Spectrom. 2010;21(3):338–47.

28. Jackson SN, Barbacci D, Egan T, Lewis EK, Schultz JA, Woods AS. MALDI-ion mobility mass spectrometry of lipids in negative ion mode. Anal Methods. 2014;6(14):5001–7.

29. Uetrecht C, Rose RJ, van Duijn E, Lorenzen K, Heck AJR. Ion mobility mass spectrometry of proteins and protein assemblies. Chem Soc Rev. 2010;39(5):1633–55.

30. Liu FC, Kirk SR, Bleiholder C. On the structural denaturation of biological analytes in trapped ion mobility spectrometry ─ mass spectrometry. Analyst. 2016;141(12):3722–30.

31. Kyle JE, Zhang X, Weitz KK, Monroe ME, Ibrahim YM, Moore RJ, et al. Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry. Analyst. 2016;141(5):1649–59.

32. Groessl M, Graf S, Knochenmuss R. High resolution ion mobility-mass spectrometry for separation and identification of isomeric lipids. Analyst. 2015;140(20):6904–11.

33. Goodwin RJ, Dungworth JC, Cobb SR, Pitt AR. Time-dependent evolution of tissue markers by MALDI-MS imaging. Proteomics. 2008;8(18):3801–8.

34. Goodwin RJ, Lang AM, Allingham H, Boren M, Pitt AR. Stopping the clock on proteomic degradation by heat treatment at the point of tissue excision. Proteomics. 2010;10(9):1751–61.

35. Goodwin RJ, Iverson SL, Andren PE. The significance of ambient-temperature on pharmaceutical and endogenous compound abundance and distribution in tissues sections when analyzed by matrix-assisted laser desorption/ionization mass spectrometry imaging. Rapid Commun Mass Spectrom. 2012;26(5):494–8.

36. Marques JV, Dalisay DS, Yang H, Lee C, Davin LB, Lewis NG. A multi-omics strategy resolves the elusive nature of alkaloids in Podophyllum species. Mol Biosyst. 2014;10(11):2838–49.

37. Horn PJ, Silva JE, Anderson D, Fuchs J, Borisjuk L, Nazarenus TJ, et al. Imaging heterogeneity of membrane and storage lipids in transgenic Camelina sativa seeds with altered fatty acid profiles. Plant J. 2013;76(1):138–50.

38. Ye H, Gemperline E, Venkateshwaran M, Chen R, Delaux PM, Howes-Podoll M, et al. MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the Medicago truncatula-Sinorhizobium meliloti symbiosis. Plant J. 2013;75(1):130–45.

39. Gemperline E, Li L. MALDI-mass spectrometric imaging for the investigation of metabolites in Medicago truncatula root nodules. J Vis Exp. 2014(85). doi:10.3791/51434.

40. Horn PJ, Sturtevant D, Chapman KD. Modified oleic cottonseeds show altered content, composition and tissue-specific distribution of triacylglycerol molecular species. Biochimie. 2014;96:28–36.

41. Korte AR, Lee YJ. MALDI-MS analysis and imaging of small molecule metabolites with 1,5-diaminonaphthalene (DAN). J Mass Spectrom. 2014;49(8):737–41.

42. Bencivenni M, Faccini A, Zecchi R, Boscaro F, Moneti G, Dossena A, et al. Electrospray MS and MALDI imaging show that non-specific lipid-transfer proteins (LTPs) in tomato are present as several isoforms and are concentrated in seeds. J Mass Spectrom. 2014;49(12):1264–71.

43. Yoshimura Y, Zaima N, Moriyama T, Kawamura Y. Different localization patterns of anthocyanin species in the pericarp of black rice revealed by imaging mass spectrometry. PLoS One. 2012;7(2):e31285.

44. Schwartz SA, Reyzer ML, Caprioli RM. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. J Mass Spectrom. 2003;38(7):699–708.

45. Gorzolka K, Kölling J, Nattkemper TW, Niehaus K. Spatio-temporal metabolite profiling of the barley germination process by MALDI MS imaging. PLoS One. 2016;11(3):e0150208.

46. Kawamoto T, Kawamoto K. Preparation of thin frozen sections from nonfixed and undecalcified hard tissues using Kawamot's film method (2012). In: Hilton JM, editor. Skeletal development and repair: methods and protocols. Totowa: Humana Press; 2014. p. 149–64.

47. Kawamoto T. Use of a new adhesive film for the preparation of multi-purpose fresh-frozen sections from hard tissues, whole-animals, insects and plants. Arch Histol Cytol. 2003;66(2):123–43.

48. Powers TW, Neely BA, Shao Y, Tang H, Troyer DA, Mehta AS, et al. MALDI imaging mass spectrometry profiling of N-glycans in formalin-fixed paraffin embedded clinical tissue blocks and tissue microarrays. PLoS One. 2014;9(9), e106255.

49. Ly A, Buck A, Balluff B, Sun N, Gorzolka K, Feuchtinger A, et al. High-mass-resolution MALDI mass spectrometry imaging of metabolites from formalin-fixed paraffin-embedded tissue. Nat Protoc. 2016;11(8):1428–43.

50. Buck A, Balluff B, Voss A, Langer R, Zitzelsberger H, Aichler M, et al. How suitable is matrix-assisted laser desorption/ionization-time-of-flight for metabolite imaging from clinical formalin-fixed and paraffin-embedded tissue samples in comparison to matrix-assisted laser desorption/ionization-Fourier transform ion cyclotron resonance mass spectrometry? Anal Chem. 2016;88(10):5281–9.

51. Buck A, Ly A, Balluff B, Sun N, Gorzolka K, Feuchtinger A, et al. High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples. J Pathol. 2015;237(1):123–32.

52. Undheim EAB, Sunagar K, Hamilton BR, Jones A, Venter DJ, Fry BG, et al. Multifunctional warheads: diversification of the toxin arsenal of centipedes via novel multidomain transcripts. J Proteomics. 2014;102:1–10.

53. Burrell M, Earnshaw C, Clench M. Imaging matrix assisted laser desorption ionization mass spectrometry: a technique to map plant metabolites within tissues at high spatial resolution. J Exp Bot. 2007;58(4):757–63.

54. Patterson NH, Thomas A, Chaurand P. Monitoring time-dependent degradation of phospholipids in sectioned tissues by MALDI imaging mass spectrometry. J Mass Spectrom. 2014;49(7):622–7.

55. Seeley EH, Oppenheimer SR, Mi D, Chaurand P, Caprioli RM. Enhancement of protein sensitivity for MALDI imaging mass spectrometry after chemical treatment of tissue sections. J Am Soc Mass Spectrom. 2008;19(8):1069–77.

56. van Hove ER, Smith DF, Fornai L, Glunde K, Heeren RM. An alternative paper based tissue washing method for mass spectrometry imaging: localized washing and fragile tissue analysis. J Am Soc Mass Spectrom. 2011;22(10):1885–90.

57. Angel PM, Spraggins JM, Baldwin HS, Caprioli R. Enhanced sensitivity for high spatial resolution lipid analysis by negative ion mode matrix assisted laser desorption ionization imaging mass spectrometry. Anal Chem. 2012;84(3):1557–64.

58. Hankin JA, Barkley RM, Murphy RC. Sublimation as a method of matrix application for mass spectrometric imaging. J Am Soc Mass Spectrom. 2007;18(9):1646–52.

59. Baluya DL, Garrett TJ, Yost RA. Automated MALDI matrix deposition method with inkjet printing for imaging mass spectrometry. Anal Chem. 2007;79(17):6862–7.

60. Becker L, Carre V, Poutaraud A, Merdinoglu D, Chaimbault P. MALDI mass spectrometry imaging for the simultaneous location of resveratrol, pterostilbene and viniferins on grapevine leaves. Molecules. 2014;19(7):10587–600.

61. Meriaux C, Franck J, Wisztorski M, Salzet M, Fournier I. Liquid ionic matrixes for MALDI mass spectrometry imaging of lipids. J Proteomics. 2010;73(6):1204–18.

62. Anderson DM, Floyd KA, Barnes S, Clark JM, Clark JI, McHaourab H, et al. A method to prevent protein delocalization in imaging mass spectrometry of non-adherent tissues: application to small vertebrate lens imaging. Anal Bioanal Chem. 2015;407(8):2311–20.

63. Fraser PD, Enfissi EM, Goodfellow M, Eguchi T, Bramley PM. Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Plant J. 2007;49(3):552–64.

64. Debois D, Jourdan E, Smargiasso N, Thonart P, De Pauw E, Ongena M. Spatiotemporal monitoring of the antibiome secreted by bacillus biofilms on plant roots using MALDI mass spectrometry imaging. Anal Chem. 2014;86(9):4431–8.

65. Franceschi P, Dong Y, Strupat K, Vrhovsek U, Mattivi F. Combining intensity correlation analysis and MALDI imaging to study the distribution of flavonols and dihydrochalcones in golden delicious apples. J Exp Bot. 2012;63(3):1123–33.

66. Gemperline E, Li L. MALDI-Mass Spectrometric Imaging of Endogenous Metabolites in Biological Systems. eLS, Wiley; 2014. doi:10.1002/9780470015902.a0023207.

67. Horka P, Vrkoslav V, Hanus R, Peckova K, Cvacka J. New MALDI matrices based on lithium salts for the analysis of hydrocarbons and wax esters. J Mass Spectrom. 2014;49(7):628–38.

68. Shroff R, Schramm K, Jeschke V, Nemes P, Vertes A, Gershenzon J, et al. Quantification of plant surface metabolites by matrix-assisted laser desorption-ionization mass spectrometry imaging: glucosinolates on Arabidopsis thaliana leaves. Plant J. 2015;81(6):961–72.

69. Horn PJ, Korte AR, Neogi PB, Love E, Fuchs J, Strupat K, et al. Spatial mapping of lipids at cellular resolution in embryos of cotton. Plant Cell. 2012;24(2):622–36.

70. Cha S, Zhang H, Ilarslan HI, Wurtele ES, Brachova L, Nikolau BJ, et al. Direct profiling and imaging of plant metabolites in intact tissues by using colloidal graphite-assisted laser desorption ionization mass spectrometry. Plant J. 2008;55(2):348–60.

71. Wang X, Han J, Chou A, Yang J, Pan J, Borchers CH. Hydroxyflavones as a new family of matrices for MALDI tissue imaging. Anal Chem. 2013;85(15):7566–73.

72. Yang H, Liu N, Qiu X, Liu S. A New method for analysis of disulfide-containing proteins by matrix-assisted laser desorption ionization (MALDI) mass spectrometry. J Am Soc Mass Spectrom. 2009;20(12):2284–93.

73. Molin L, Seraglia R, Dani FR, Moneti G, Traldi P. The double nature of 1,5-diaminonaphthalene as matrix-assisted laser desorption/ionization matrix: some experimental evidence of the protonation and reduction mechanisms. Rapid Commun Mass Spectrom. 2011;25(20):3091–6.

74. Robichaud G, Barry J, Muddiman D. IR-MALDESI mass spectrometry imaging of biological tissue sections using ice as a matrix. J Am Soc Mass Spectrom. 2014;25(3):319–28.

75. Dufresne M, Thomas A, Breault-Turcot J, Masson J-F, Chaurand P. Silver-assisted laser desorption ionization for high spatial resolution imaging mass spectrometry of olefins from thin tissue sections. Anal Chem. 2013;85(6):3318–24.

76. Jackson SN, Baldwin K, Muller L, Womack VM, Schultz JA, Balaban C, et al. Imaging of lipids in rat heart by MALDI-MS with silver nanoparticles. Anal Bioanal Chem. 2014;406(5):1377–86.

77. Muller L, Kailas A, Jackson SN, Roux A, Barbacci DC, Schultz JA, et al. Lipid imaging within the normal rat kidney using silver nanoparticles by matrix-assisted laser desorption/ionization mass spectrometry. Kidney Int. 2015;88(1):186–92.

78. Wu HP, Yu CJ, Lin CY, Lin YH, Tseng WL. Gold nanoparticles as assisted matrices for the detection of biomolecules in a high-salt solution through laser desorption/ionization mass spectrometry. J Am Soc Mass Spectrom. 2009;20(5):875–82.

79. Jackson SN, Ugarov M, Egan T, Post JD, Langlais D, Albert Schultz J, et al. MALDI-ion mobility-TOFMS imaging of lipids in rat brain tissue. J Mass Spectrom. 2007;42(8):1093–8.

80. Taira S, Sugiura Y, Moritake S, Shimma S, Ichiyanagi Y, Setou M. Nanoparticle-assisted laser desorption/ionization based mass imaging with cellular resolution. Anal Chem. 2008;80(12):4761–6.

81. Schramm T, Hester A, Klinkert I, Both JP, Heeren RM, Brunelle A, et al. imzML-a common data format for the flexible exchange and processing of mass spectrometry imaging data. J Proteomics. 2012;75(16):5106–10.

82. Klinkert I, Chughtai K, Ellis SR, Heeren RMA. Methods for full resolution data exploration and visualization for large 2D and 3D mass spectrometry imaging datasets. Int J Mass Spectrom. 2014;362:40–7.

83. Horn PJ, Chapman KD. Metabolite Imager: customized spatial analysis of metabolite distributions in mass spectrometry imaging. Metabolomics. 2013;10(2):337–48.

84. Paschke C, Leisner A, Hester A, Maass K, Guenther S, Bouschen W, et al. Mirion--a software package for automatic processing of mass spectrometric images. J Am Soc Mass Spectrom. 2013;24(8):1296–306.

85. Robichaud G, Garrard KP, Barry JA, Muddiman DC. MSiReader: an open-source interface to view and analyze high resolving power MS imaging files on Matlab platform. J Am Soc Mass Spectrom. 2013;24(5):718–21.

86. Rübel O, Greiner A, Cholia S, Louie K, Bethel EW, Northen TR, et al. OpenMSI: a high-performance web-based platform for mass spectrometry imaging. Anal Chem. 2013;85(21):10354–61.

87. Bemis KD, Harry A, Eberlin LS, Ferreira C, van de Ven SM, Mallick P, et al. Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments. Bioinformatics. 2015;31(14):2418–20.

88. Parry RM, Galhena AS, Gamage CM, Bennett RV, Wang MD, Fernández FM. OmniSpect: an open MATLAB-based tool for visualization and analysis of matrix-assisted laser desorption/

ionization and desorption electrospray ionization mass spectrometry images. J Am Soc Mass Spectrom. 2013;24(4):646–9.

89. Källback P, Nilsson A, Shariatgorji M, Andrén PE. MsIQuant – quantitation software for mass spectrometry imaging enabling fast access, visualization, and analysis of large data sets. Anal Chem. 2016;88(8):4346–53.

90. Takahashi K, Kozuka T, Anegawa A, Nagatani A, Mimura T. Development and application of a high-resolution imaging mass spectrometer for the study of plant tissues. Plant Cell Physiol. 2015;56(7):1329–38.

91. Gamboa-Becerra R, Ramírez-Chávez E, Molina-Torres J, Winkler R. MSI.R scripts reveal volatile and semi-volatile features in low-temperature plasma mass spectrometry imaging (LTP-MSI) of chilli (Capsicum annuum). Anal Bioanal Chem. 2015;407(19):5673–84.

92. Gibb S, Strimmer K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinformatics. 2012;28(17):2270–1.

93. Alexandrov T. MALDI imaging mass spectrometry: statistical data analysis and current computational challenges. BMC Bioinformatics. 2012;13 Suppl 16:S11.

94. Alexandrov T, Chernyavsky I, Becker M, von Eggeling F, Nikolenko S. Analysis and interpretation of imaging mass spectrometry data by clustering mass-to-charge images according to their spatial similarity. Anal Chem. 2013;85(23):11189–95.

95. Norris JL, Cornett DS, Mobley JA, Andersson M, Seeley EH, Chaurand P, et al. Processing MALDI mass spectra to improve mass spectral direct tissue analysis. Int J Mass Spectrom. 2007;260(2–3):212–21.

96. Franceschi P, Wehrens R. Self-organizing maps: a versatile tool for the automatic analysis of untargeted imaging datasets. Proteomics. 2014;14(7–8):853–61.

97. Fonville JM, Carter CL, Pizarro L, Steven RT, Palmer AD, Griffiths RL, et al. Hyperspectral visualization of mass spectrometry imaging data. Anal Chem. 2013;85(3):1415–23.

98. Wijetunge CD, Saeed I, Halgamuge SK, Boughton B, Roessner U, editors. Unsupervised learning for exploring MALDI imaging mass spectrometry 'omics' data. Information and Automation for Sustainability (ICIAfS), 2014 7th International Conference on; 22–24 Dec. 2014.

99. Andersson M, Groseclose MR, Deutch AY, Caprioli RM. Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. Nat Methods. 2008;5(1):101–8.

100. Oetjen J, Veselkov K, Watrous J, McKenzie JS, Becker M, Hauberg-Lotte L, et al. Benchmark datasets for 3D MALDI- and DESI-imaging mass spectrometry. Gigascience. 2015;4:20.

101. Weaver EM, Hummon AB, Keithley RB. Chemometric analysis of MALDI mass spectrometric images of three-dimensional cell culture systems. Anal Methods. 2015;7(17):7208–19.

102. Fletcher JS, Lockyer NP, Vickerman JC. Developments in molecular SIMS depth profiling and 3D imaging of biological systems using polyatomic primary ions. Mass Spectrom Rev. 2011;30(1):142–74.

103. Fletcher JS, Vickerman JC, Winograd N. Label free biochemical 2D and 3D imaging using secondary ion mass spectrometry. Curr Opin Chem Biol. 2011;15(5):733–40.

104. McDonnell LA, Rompp A, Balluff B, Heeren RM, Albar JP, Andren PE, et al. Discussion point: reporting guidelines for mass spectrometry imaging. Anal Bioanal Chem. 2015;407(8):2035–45.

105. Creek DJ, Dunn WB, Fiehn O, Griffin JL, Hall RD, Lei Z, et al. Metabolite identification: are you sure? And how do your peers gauge your confidence? Metabolomics. 2014;10(3):350–3.

106. Rompp A, Wang R, Albar JP, Urbani A, Hermjakob H, Spengler B, et al. A public repository for mass spectrometry imaging data. Anal Bioanal Chem. 2015;407(8):2027–33.

107. Anderson DM, Ablonczy Z, Koutalos Y, Spraggins J, Crouch RK, Caprioli RM, et al. High resolution MALDI imaging mass spectrometry of retinal tissue lipids. J Am Soc Mass Spectrom. 2014;25(8):1394–403.

108. Fernandez R, Lage S, Abad-Garcia B, Barcelo-Coblijn G, Teres S, Lopez DH, et al. Analysis of the lipidome of xenografts using MALDI-IMS and UHPLC-ESI-QTOF. J Am Soc Mass Spectrom. 2014;25(7):1237–46.

109. Martin-Lorenzo M, Balluff B, Maroto AS, Carreira RJ, van Zeijl RJ, Gonzalez-Calero L, et al. Molecular anatomy of ascending aorta in atherosclerosis by MS imaging: specific lipid and protein patterns reflect pathology. J Proteomics. 2015;126:245–51.

110. Roux A, Muller L, Jackson SN, Post J, Baldwin K, Hoffer B, et al. Mass spectrometry imaging of rat brain lipid profile changes over time following traumatic brain injury. J Neurosci Methods. 2016. doi:10.1016/j.jneumeth.2016.02.004.
111. Ruh H, Salonikios T, Fuchser J, Schwartz M, Sticht C, Hochheim C, et al. MALDI imaging MS reveals candidate lipid markers of polycystic kidney disease. J Lipid Res. 2013;54(10):2785–94.
112. Uzbekova S, Elis S, Teixeira-Gomes AP, Desmarchais A, Maillard V, Labas V. MALDI mass spectrometry imaging of lipids and gene expression reveals differences in fatty acid metabolism between follicular compartments in porcine ovaries. Biology (Basel). 2015;4(1):216–36.
113. Chatterji B, Dickhut C, Mielke S, Kruger J, Just I, Glage S, et al. MALDI imaging mass spectrometry to investigate endogenous peptides in an animal model of Usher's disease. Proteomics. 2014;14(13–14):1674–87.
114. Gustafsson OJ, Eddes JS, Meding S, McColl SR, Oehler MK, Hoffmann P. Matrix-assisted laser desorption/ionization imaging protocol for in situ characterization of tryptic peptide identity and distribution in formalin-fixed tissue. Rapid Commun Mass Spectrom. 2013;27(6):655–70.
115. Hunt NJ, Phillips L, Waters KA, Machaalani R. Proteomic MALDI-TOF/TOF-IMS examination of peptide expression in the formalin fixed brainstem and changes in sudden infant death syndrome infants. J Proteomics. 2016;138:48–60.
116. Ljungdahl A, Hanrieder J, Bergquist J, Andersson M. Analysis of neuropeptides by MALDI imaging mass spectrometry. Methods Mol Biol. 2013;1023:121–36.
117. Meding S, Martin K, Gustafsson OJ, Eddes JS, Hack S, Oehler MK, et al. Tryptic peptide reference data sets for MALDI imaging mass spectrometry on formalin-fixed ovarian cancer tissues. J Proteome Res. 2013;12(1):308–15.
118. Park KM, Moon JH, Kim KP, Lee SH, Kim MS. Relative quantification in imaging of a peptide on a mouse brain tissue by matrix-assisted laser desorption ionization. Anal Chem. 2014;86(10):5131–5.
119. Sosnowski P, Zera T, Wilenska B, Szczepanska-Sadowska E, Misicka A. Imaging and identification of endogenous peptides from rat pituitary embedded in egg yolk. Rapid Commun Mass Spectrom. 2015;29(4):327–35.
120. Winderbaum LJ, Koch I, Gustafsson OJR, Meding S, Hoffmann P. Feature extraction for proteomics imaging mass spectrometry data. Ann Appl Statistics. 2015;9(4):1973–96.
121. Anderson DM, Van de Plas R, Rose KL, Hill S, Schey KL, Solga AC, et al. 3-D imaging mass spectrometry of protein distributions in mouse Neurofibromatosis 1 (NF1)-associated optic glioma. J Proteomics. 2016. doi:10.1016/j.jprot.2016.02.004.
122. Spraggins JM, Rizzo DG, Moore JL, Noto MJ, Skaar EP, Caprioli RM. Next-generation technologies for spatial proteomics: integrating ultra-high speed MALDI-TOF and high mass resolution MALDI FTICR imaging mass spectrometry for protein analysis. Proteomics. 2016;16(11–12):1678–89.
123. Undheim EA, Hamilton BR, Kurniawan ND, Bowlay G, Cribb BW, Merritt DJ, et al. Production and packaging of a biological arsenal: evolution of centipede venoms under morphological constraint. Proc Natl Acad Sci U S A. 2015;112(13):4026–31.
124. He J, Luo Z, Huang L, He J, Chen Y, Rong X, et al. Ambient mass spectrometry imaging metabolomics method provides novel insights into the action mechanism of drug candidates. Anal Chem. 2015;87(10):5372–9.
125. Wijetunge CD, Saeed I, Boughton BA, Spraggins JM, Caprioli RM, Bacic A, et al. EXIMS: an improved data analysis pipeline based on a new peak picking method for EXploring Imaging Mass Spectrometry data. Bioinformatics. 2015;31(19):3198–206.
126. Zhao YY, Miao H, Cheng XL, Wei F. Lipidomics: Novel insight into the biochemical mechanism of lipid metabolism and dysregulation-associated disease. Chem Biol Interact. 2015;240:220–38.
127. Barry JA, Robichaud G, Bokhart MT, Thompson C, Sykes C, Kashuba AD, et al. Mapping antiretroviral drugs in tissue by IR-MALDESI MSI coupled to the Q Exactive and comparison with LC-MS/MS SRM assay. J Am Soc Mass Spectrom. 2014;25(12):2038–47.

128. Bianga J, Bouslimani A, Bec N, Quenet F, Mounicou S, Szpunar J, et al. Complementarity of MALDI and LA ICP mass spectrometry for platinum anticancer imaging in human tumor. Metallomics. 2014;6(8):1382–6.
129. Huber K, Aichler M, Sun N, Buck A, Li Z, Fernandez IE, et al. A rapid ex vivo tissue model for optimising drug detection and ionisation in MALDI imaging studies. Histochem Cell Biol. 2014;142(4):361–71.
130. Huber K, Feuchtinger A, Borgmann DM, Li Z, Aichler M, Hauck SM, et al. Novel approach of MALDI drug imaging, immunohistochemistry, and digital image analysis for drug distribution studies in tissues. Anal Chem. 2014;86(21):10568–75.
131. Salphati L, Shahidi-Latham S, Quiason C, Barck K, Nishimura M, Alicke B, et al. Distribution of the phosphatidylinositol 3-kinase inhibitors Pictilisib (GDC-0941) and GNE-317 in U87 and GS2 intracranial glioblastoma models-assessment by matrix-assisted laser desorption ionization imaging. Drug Metab Dispos. 2014;42(7):1110–6.
132. Shariatgorji M, Nilsson A, Goodwin RJ, Kallback P, Schintu N, Zhang X, et al. Direct targeted quantitative molecular imaging of neurotransmitters in brain tissue sections. Neuron. 2014;84(4):697–707.
133. Connell JJ, Sugihara Y, Torok S, Dome B, Tovari J, Fehniger TE, et al. Localization of sunitinib in in vivo animal and in vitro experimental models by MALDI mass spectrometry imaging. Anal Bioanal Chem. 2015;407(8):2245–53.
134. Kamata T, Shima N, Sasaki K, Matsuta S, Takei S, Katagi M, et al. Time-course mass spectrometry imaging for depicting drug incorporation into hair. Anal Chem. 2015;87(11):5476–81.
135. Shobo A, Bratkowska D, Baijnath S, Naiker S, Bester LA, Singh SD, et al. Visualization of time-dependent distribution of rifampicin in rat brain using MALDI MSI and quantitative LCMS/MS. Assay Drug Dev Technol. 2015;13(5):277–84.
136. Fujiwara Y, Furuta M, Manabe S, Koga Y, Yasunaga M, Matsumura Y. Imaging mass spectrometry for the precise design of antibody-drug conjugates. Sci Rep. 2016;6:24954.
137. Goodwin RJ, Nilsson A, Mackay CL, Swales JG, Johansson MK, Billger M, et al. Exemplifying the screening power of mass spectrometry imaging over label-based technologies for simultaneous monitoring of drug and metabolite distributions in tissue sections. J Biomol Screen. 2016;21(2):187–93.
138. Hayasaka T. Application of imaging mass spectrometry for drug discovery. Yakugaku Zasshi: J Pharm Soc Japan. 2016;136(2):163–70.
139. Sun N, Fernandez IE, Wei M, Wu Y, Aichler M, Eickelberg O, et al. Pharmacokinetic and pharmacometabolomic study of pirfenidone in normal mouse tissues using high mass resolution MALDI-FTICR-mass spectrometry imaging. Histochem Cell Biol. 2016;145(2):201–11.
140. Yarnold JE, Hamilton BR, Welsh DT, Pool GF, Venter DJ, Carroll AR. High resolution spatial mapping of brominated pyrrole-2-aminoimidazole alkaloids distributions in the marine sponge Stylissa flabellata via MALDI-mass spectrometry imaging. Mol Biosyst. 2012;8(9):2249–59.
141. Louie KB, Bowen BP, McAlhany S, Huang Y, Price JC, Mao JH, et al. Mass spectrometry imaging for in situ kinetic histochemistry. Sci Rep. 2013;3:1656.
142. Seaman C, Flinders B, Eijkel G, Heeren RM, Bricklebank N, Clench MR. "Afterlife experiment": use of MALDI-MS and SIMS imaging for the study of the nitrogen cycle within plants. Anal Chem. 2014;86(20):10071–7.
143. Hanrieder J, Karlsson O, Brittebo EB, Malmberg P, Ewing AG. Probing the lipid chemistry of neurotoxin-induced hippocampal lesions using multimodal imaging mass spectrometry. Surf Interface Anal. 2014;46(S1):375–8.
144. Chughtai S, Chughtai K, Cillero-Pastor B, Kiss A, Agrawal P, MacAleese L, et al. A multimodal mass spectrometry imaging approach for the study of musculoskeletal tissues. Int J Mass Spectrom. 2012;325–327:150–60.

145. Ogrinc Potočnik N, Škrášková K, Flinders B, Pelicon P, Heeren RMA. Gold sputtered fiducial markers for combined secondary ion mass spectrometry and MALDI imaging of tissue samples. Anal Chem. 2014;86(14):6781–5.
146. Amstalden van Hove ER, Blackwell TR, Klinkert I, Eijkel GB, Heeren RMA, Glunde K. Multimodal mass spectrometric imaging of small molecules reveals distinct spatio-molecular signatures in differentially metastatic breast tumor models. Cancer Res. 2010;70(22):9012–21.
147. Soltwisch J, Kettling H, Vens-Cappell S, Wiegelmann M, Muthing J, Dreisewerd K. Mass spectrometry imaging with laser-induced postionization. Science. 2015;348(6231):211–5.

Chapter 13
Single-Cell Metabolomics

Samy Emara, Sara Amer, Ahmed Ali, Yasmine Abouleila, April Oga, and Tsutomu Masujima

Abstract The dynamics of a cell is always changing. Cells move, divide, communicate, adapt, and are always reacting to their surroundings non-synchronously. Currently, single-cell metabolomics has become the leading field in understanding the phenotypical variations between them, but sample volumes, low analyte concentrations, and validating gentle sample techniques have proven great barriers toward achieving accurate and complete metabolomics profiling. Certainly, advanced technologies such as nanodevices and microfluidic arrays are making great progress, and analytical techniques, such as matrix-assisted laser desorption ionization (MALDI), are gaining popularity with high-throughput methodology. Nevertheless, live single-cell mass spectrometry (LCSMS) values the sample quality and precision, turning once theoretical speculation into present-day applications in a variety of fields, including those of medicine, pharmaceutical, and agricultural industries. While there is still room for much improvement, it is clear that the metabolomics field is progressing toward analysis and discoveries at the single-cell level.

Keywords Single cell • Metabolomics • Mass spectrometry • Nanodevices • Matrix-assisted laser desorption ionization • Live single-cell mass spectrometry • Circulating tumor cell • Drug discovery

S. Emara
Faculty of Pharmacy, Misr International University, Cairo, Egypt

S. Amer • A. Ali • Y. Abouleila

Faculty of Pharmacy, Misr International University, Cairo, Egypt

RIKEN, Quantitative Biology Center, Single Cell Mass Spectrometry Lab, 6-2-4 Furue-dai, Suita, Osaka 565-0874, Japan

A. Oga • T. Masujima (✉)

RIKEN, Quantitative Biology Center, Single Cell Mass Spectrometry Lab, 6-2-4 Furue-dai, Suita, Osaka 565-0874, Japan

e-mail: masujima@riken.jp

Abbreviations

3D   Three dimensional
7-EC   7-Ethoxycoumarin
CTC   Circulating tumor cell
DNA   Deoxyribonucleic acid
ESI   Electrospray ionization
iMAP   Integrated microfluidic array plate
LC-MS   Liquid chromatography-mass spectrometry
LCSMS   Live single-cell mass spectrometry
MALDI   Matrix-assisted laser desorption ionization
MAMS   Microarrays for mass spectrometry
mRNA   Messenger ribonucleic acid
MS   Mass spectrometry
MS/MS   Tandem mass spectrometry
PDMA   Polydimethylsiloxane
TA   Tafluprost acid
TOF   Time of flight
UV   Ultraviolet

13.1   Introduction

Metabolomics, the study of the complete complement of all small molecules (<1500 Da) found in a specific cell, organ, or organism [1], is considered the most recent – and arguably the end point – of the omics cascade [2]. In fact, it has been mentioned by Patti et al. to be the "apogee of the omics trilogy" [3], because unlike genomics and proteomics, metabolomics depicts real-time biochemical activity and therefore is the key in phenotype association and offers a more reliable depiction of the dynamics of the sample in question. As science progresses and technology advances, it has now become clear that tissue-scale metabolomics, or even multicell metabolomics, yields averaged data that can oftentimes be misleading in making assumptions relating to a cell's condition. Cellular heterogeneity is dynamic and may result from a plethora of factors, including genetic, epigenetic, or phenotypic differences; morphological, biochemical, or functional changes; positional, exogenous, or endogenous mutations; and physical, chemical, or biological effects from the environment [4]. Even cells with identical genotypes can display phenotypical differences [5].

In addition to the above possible causes, stochasticity – induced phenotypic heterogeneity through gene and protein expression – was only recently added as another factor contributing to cell deviation [6, 7]. They explained that due to the low copy number of DNA and mRNA, "noise," or random irregularities in the rate of their respective reactions, occurs in gene or protein expression, which then causes variable molecule concentrations from cell to cell.

Alas, it is near impossible to find two cells with the same metabolome, even if they originate from the same precursor cell. Furthermore, metabolites and other small molecules cannot be amplified like DNA, which is why most metabolomic and proteomic studies are comprised of large cell number homogenization. The ability to detect, identify, and quantify metabolites within a single cell will open new doors to understanding the reason behind cell-to-cell heterogeneity, even within a seemingly homogeneous population.

This chapter discusses the importance of single-cell metabolomics and significant points to consider when sampling a single cell. It also highlights the foremost approaches in sampling and analysis and gives a perspective into the future for single-cell analysis.

13.1.1   Why Single-Cell Metabolomics?

In life sciences, the cell is regarded as the minimal functional unit, and its analysis has been undoubtedly crucial. Single-cell analysis represents qualitatively and exhaustively analyzing a wide range of molecular information carried on numerous biomolecules at the single-cell level. It can give insight into unknown processes such as cellular evolution, adaptation, and communication.

Single-cell studies are theoretically the only types of analyses that can give a depiction of real-time biochemical reactions that oftentimes only take seconds or minutes to occur. The realization of how rapid the kinetics actually is within a cell can only move science toward the direction of single-cell metabolomics.

Not only that, but cells behave non-synchronously and therefore must be studied independently when examining topics such as the cell cycle in order to obtain an accurate metabolic profile. Concepts such as stochasticity and heterogeneity would be lost in translation when analyzing cell populations due to averaging. Therefore, in cases of phenotypic studies, single-cell measurements are essential to produce unbiased metabolic models.

During the past few years, single-cell technologies have undergone rapid development and reached a critical point where they have become a valuable tool for systematic characterization of cellular heterogeneity, which in turn has important implications in a wide range of biomedical issues such as gene regulation, cell lineage differentiation, signaling response, and disease characterization. The medical field has evolved significantly, which has led toward a paradigm shift in medicine: moving the observation from patient and organ toward a more in-depth observation – single cells within the organ. This advancement has been enabled to some extent by engineering sciences, among which microsystem technologies were a major driving force.

Perhaps the best example for the need of single-cell metabolomics is cancer. It only takes one abnormal cell in the whole body that contains 30 trillion cells to cause cancer [8]. Single-cell analysis has the potential to help in early detection of medical conditions involving modifications in the cellular functions such as cancer genesis and progression. There is also a need to differentiate between cells with

different metabolomes within the same cancer in order to evaluate phenotype heterogeneity, and in effect, prevent drug resistance or discover new, more effective therapeutics.

13.2 Sample Considerations

13.2.1 *The Size of the Cell Matters*

The term "single cell" has no limitation on the actual size of the analyzed cell, and cells come in a wide spectrum in terms of size. To date, the largest single-cell organism documented is *Caulerpa taxifolia*, a member of the green algae family [9]. It ranges in size, but can grow up to several meters in length. Another much more common single cell is an egg. While the largest egg is that of an ostrich (measuring an average of 16 cm long and 13 cm in diameter and weighing 1.5 kg), chicken eggs (which are 24 times smaller than the ostrich egg on average) are of much more common use in science [10]. Mattsson et al. completed metabolic profiling of chicken embryos after exposure to perfluoroalkyl acids, a group of ecologically detrimental organic chemicals, by drilling holes in each "single cell" and injecting with the said substance [11]. Moving along the size spectrum, embryos of *Xenopus laevis*, or the South African clawed frog, have been studied to further understand embryonic development through single-cell metabolomics [12, 13]. The group of Sweedler has employed several techniques for extensive analysis of single neurons of the sea slug *Aplysia californica*, a cell that can grow up to 500 μm [14–19]. Concerning agricultural chemistry, large plant cells have often been the target of single-cell analysis due to their large size, including *Allium cepa*, commonly known as onion [20].

On the other side of the size spectrum, single-cell studies have been shifting toward smaller and smaller cells. The team of Masujima coined the term "live single-cell mass spectrometry" and have analyzed single plants cells – which can range from 10 to 100 μm – from the leaf, stem, and petal of *Pelargonium zonale* [21] and the leaf, stem, and root of *Raphanus sativus* [22]. The same group has done numerous studies on mammalian cells such as mouse embryonic fibroblasts Swiss 3 T3 [23], rat basophilic leukemia cell RBL-2H3 [24, 25], and hepatocellular carcinoma HepG2 [26], all ranging from 10 to 20 μm in size. Their most recent studies have been on single blood cells in relation to clinical studies, having successfully analyzed white blood cells and circulating tumor cells (CTC), ranging in size from 12 to 15 μm, and analyzed red blood cells, with an even smaller size of 6–8 μm [27]. Figure 13.1 shows actual size ratios of some of the most common cells used in single-cell analysis. The human oocyte is the largest single cell in the human body and is added as a point of comparison.

It seems that this may be the limit for true single-cell analysis. Ibáñez et al. recently published a study on single-cell yeast metabolomics [28], but their method involved using microarrays for mass spectrometry (MAMS) platform, in which each hydrophilic reservoir holds anywhere between one and 15 cells. In regard to bacteria, groups such as that of Tanaguchi have been successful in proteomics and
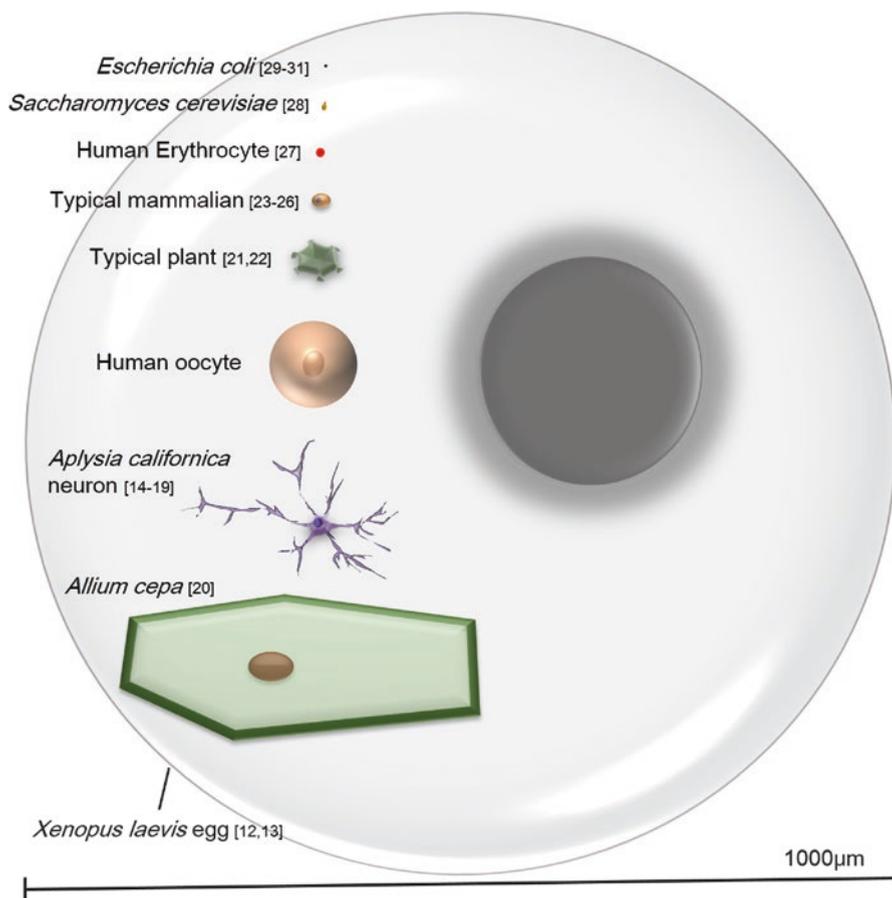
Fig. 13.1 Depiction of common cells in single-cell analysis magnified from actual size while maintaining the true ratio

transcriptomics of *Escherichia coli* [29–31], but no papers have been reported on single-cell bacteria metabolomics, revealing that the current instrumentation has reached a plateau in sensitivity. As previously stated, metabolites, in contrast to other biomolecules, cannot be amplified [32], and fluorescent labeling would negatively impact metabolic pathways within the cells, making sensitivity in smaller cells the major limiting factor in analysis.

13.2.2   The Condition of the Cell

Numerous techniques have been established for the isolation and analysis of a single cell, but one must ask themselves: what condition is the cell in? Even if the analysis is successful, will this be an accurate depiction of the metabolic profile? "A

major issue is a suitable sample preparation that does not upset the metabolism of the cells to be investigated. One way to cope with this problem is to keep the cell in a native environment as long as possible" [33].

Cell sampling is arguably the most critical step for single-cell analysis. Protocols used for metabolomic studies on a cell population are deemed useless on the single-cell level. For a single cell, many methods have been established, and many separation kits have been developed [34]; but these processes are often long and contain many steps, causing a disturbance in the natural microenvironment around the cell and increasing the possibility of metabolite distortion or exhaustion of molecules of low concentration. The time taken between sampling and analysis allows for, sometimes abnormal, enzymatic activity and biochemical processing to cope with the changing conditions. Methods such as shock freezing [28], or the immediate addition of organic solvent after sampling, have been used to prevent continual enzymatic activity [22].

To increase high throughput, newer approaches separate and grow single cells in individual wells or innovated trapping arrays. Because they are grown inside these devices, many argue that it is the perfect method for true single-cell metabolomics. On the other hand, others criticize these methods because the cell is not in its natural environment. Think of it this way: would a person in a room full of people act the same if he/she were in a room alone? Cell communication is a natural and essential factor, and completely isolating a cell puts it in an unnatural position, which could be projected into the metabolomic profile.

Moreover, it is worth noting that despite the recent innovations in single-cell isolation techniques, keeping the cell in its natural environment as much as possible until initiating analysis is the most important aspect of analysis. Most current methods require a degree of manipulation to the cell and isolate the cell from cell-cell interactions, thus making it difficult to study certain aspects of cell biology such as communication or signaling. The ideal sampling technique should be noninvasive, include an efficient quenching step, and have a short lysis/analysis time to prevent metabolite decomposition.

13.3   Methods and Approaches

13.3.1   Sampling Techniques

The process of single-cell analysis starts with sample preparation. The goal of sample preparation in single-cell analysis is to isolate the target cell in a high-throughput manner without affecting the normal metabolome of the cell. Moreover, to achieve such goal, many sampling techniques have been recently developed, each method having its own unique advantages, disadvantages, and possible applications, which will be discussed in this section.

There are two main approaches for isolation and preparation of the cell for analysis. The first approach is using nanodevices to manipulate the cell and isolate it for

further downstream analysis. The second approach utilizes microfluidic devices, which have the advantage of higher throughput, but also comes with its own set of limitations.

13.3.1.1   Nanodevices

Due to the recent advances in nanoscale fabrication, several techniques for single-cell metabolomics have been developed to isolate, introduce chemicals into the cell, or capture the cell itself by utilizing nanoscale devices.

   The use of micropipettes to gently isolate or sample the cells itself is a promising field in it of itself, and this can be done manually using a culture plate and a micromanipulator, which has the obvious challenge of achieving high throughput. In order to increase throughput, several automated systems have been introduced lately.

   One recent innovation is coupling a micropipette with an automated system that can target certain cell types or regions within the cell according to their visual characteristics by the aid of computer software and image analysis. After selecting the desired cell, a robot picks up the target cell using a glass micropipette with an internal diameter of 30 μm and positioned 5 μm away from the bottom of the Petri dish. The cell is picked up by a vacuum system connected to the micropipette. Finally, the cell is put on 3D printed miniature wells on a Petri dish for downstream analysis. This system succeeded in achieving relatively high throughput compared to other systems and has the inherent advantage of not using markers to isolate the cells (Fig. 13.2) [35]. Moreover, several methods were proposed to introduce chemicals into the cell itself or detect optical signals on a subcellular scale. One such method utilizes a nanowire attached to the end of an optical fiber, which guides visible light into subcellular components. This nanoscale endoscope can also be used to deliver payloads into the cell itself [36].

   Another interesting approach is using electroporation in which an induced electric field is applied to the cells to increase their cell membrane permeability, allowing chemicals or drugs to be introduced into the cell. The device proposed by Boukany et al. consists of two microchannels connected by a nanochannel where electroporation occurs [37]. The target cell is placed on one channel using optical tweezers, and the transfection agents are placed on the other microchannel. Transfection occurs by applying a voltage pulse between the channels resulting in an intense, localized electric field over a small area on the cell membrane, which allows a precise amount of the transfection agent to travel through the nanochannel, cell membrane, and into the cytoplasm by electrophoresis.

   Despite the recent innovations, there are still challenges to fully implement nanodevices as the mainstream cell isolation technique. Maintaining a delicate balance between high throughput, low loss percentage, precision, degree of invasiveness, and ease of use is not an easy task. Automated systems have the advantage of being high throughput, but they are also inherently more complex. Manual systems,
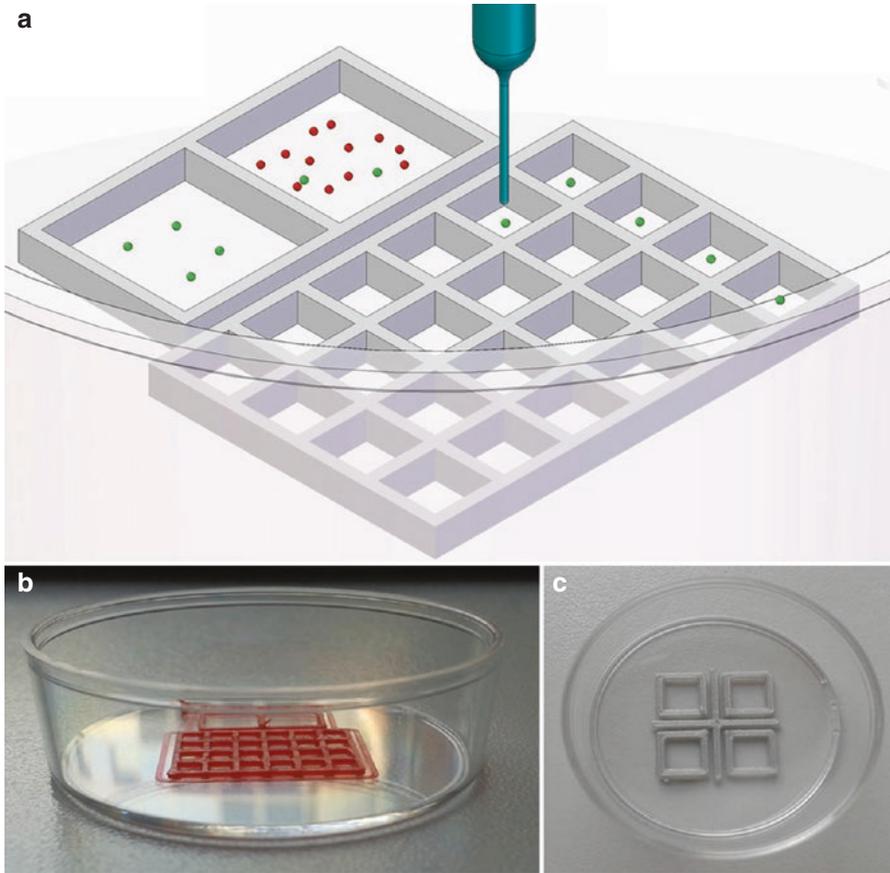
Fig. 13.2 Panel (**a**) represents the schematic of the method; the cells are selected by a software and then pipetted to a Petri dish containing miniature multiwall plates that are printed onto 35 mm plastic Petri dishes using a commercial 3D printer to reduce convection. Panel (**b**, **c**) show the multiwells (24, 2 × 2 mm² wells and 4, 5 × 5 mm² wells). The larger wells are used for stepwise, successive isolation in dense media. Reproduced from Ref 35 with permission from the Nature Publishing Group

on the other hand, are simple, but their efficiency depends on human factors and the throughput has a high degree of variability.

13.3.1.2 Microfluidic Arrays

Microfluidic arrays represent another recent approach for single-cell isolation. The goal of these methods is the same: to transport and isolate single cells for further downstream analysis. The separation or isolation occurs by passing the culture media containing the cells through microfluid channels or arrays that isolate the cells individually in a high-throughput manner, as shown in Fig. 13.3. There are
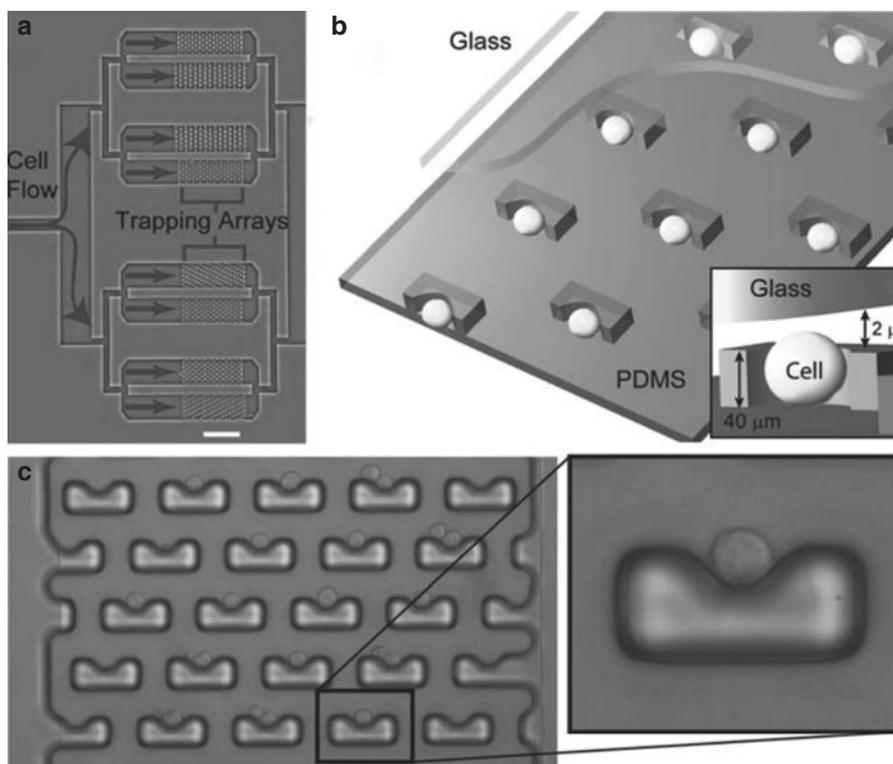
Fig. 13.3 Panel (**a**) shows an overview of the trapping system; cell flow is gently directed into a series of trapping arrays; the scale bar is 500 μm. A single trapping array is shown in panel (**b**). The traps are molded from PDMS and fixed on a glass surface; trap sizes are optimized to capture one or two cells at most. A bright-field micrograph of the array is shown with cells trapped inside in panel (**c**) with added magnification of a single cell trapped in the minimum potential of the well. Reproduced from Ref 38 with permission of The Royal Society of Chemistry

many recent innovations in single-cell isolation by using microfluidic arrays, among these are the integrated microfluidic array plate (iMAP) and dynamic single-cell culture array [38, 39]. Most, if not all, of these methods use polydimethylsiloxane (PDMS) as the construction material, along with utilizing soft lithography in fabrication. The iMAP, proposed by Dimov et al., utilizes gravity to guide the cells into their respective wells and has the advantage of greater capture rates than other methods (close to 100 %) [39]. Its design also allows for reagent addition and single-cell lysis depending on the analytical goals.

One of the challenges of using microfluidic arrays is isolating the cell without causing noticeable damage or changing normal cell behavior. The method established by Carlo et al. shows great promise in this regard [38]. The system works by channeling the cells gently into a branched array system that consists of U-shaped PDMS traps fixed on a glass surface. The geometry of the trap is optimized to isolate one or two cells; once a cell occupies the trap, the altered dynamic flow around it will minimize the chance of other cells joining it, so in a way, the trap is

self-regulating. The main advantage of such system is its ability to capture the cells with a high success rate (~90%) without causing significant disruptions to the normal cell behavior and environment.

13.3.2   Analytical Methods

Assuming the cell was successfully isolated with minimal pretreatment, the next logical step would be to uncover its metabolome. To this effect, a myriad of analytical methods have been developed to analyze biomolecules. However, there are two main hurdles for single-cell analysis, and both relate to the inherently small sample size. The first challenge is improving the ability of the instrument to distinguish between closely related molecules, i.e., the resolution. Since single-cell analysis deals with small sample size (pico- to nanoliters), conventional separation techniques are ill-equipped to deal with such low sample volume without diluting the sample excessively or causing significant sample loss. The second challenge is increasing the ability to detect lower and lower concentrations reliably, i.e., the sensitivity. As sensitivity increases, the viability of the instrument itself to perform analysis on the single-cell scale increases.

Keeping the previously mentioned challenges in mind, choosing the perfect analysis technique is no easy task, but among the available analytical methods, mass spectrometry has gained prominence lately as one of the best techniques used to analyze the chemical composition at the single-cell level due to its relatively high sensitivity and resolution. There are several approaches used in single-cell mass spectrometric analysis, but this chapter will focus on the two main methods that are matrix-assisted laser desorption ionization (MALDI) and live single-cell mass spectrometry (LSCMS).

13.3.2.1   Matrix-Assisted Laser Desorption Ionization

MALDI is considered a "soft" ionization method, meaning it does not cause extensive fragmentation to the sample ions. The desired sample is mixed with a solution of low-mass organic compounds called a matrix, which is essential for ionization because it acts as a proton supplier and as a support or scaffold by which ionization can occur. Then the resulting mixture is spotted onto a metal plate called the target, as shown in Fig. 13.4. After spotting, the mixture is left to dry out, and both the sample and the matrix co-crystallize to form a solid deposit on the target. The target plate is then loaded into the mass spectrometer – most commonly using a time-of-flight (TOF) mass analyzer – where it will be subjected to a vacuum while the solid crystals of the mixture are irradiated by a UV laser beam that causes ablation of said crystals into the gas phase, followed by ionization of the sample. TOF mass spectrometry analyzes and detects the ions and gives out signals with ion mass-to-charge ratio (*m/z*) forming a distinct mass spectrometric profile that can be matched to a database to identify the sample ions [40].
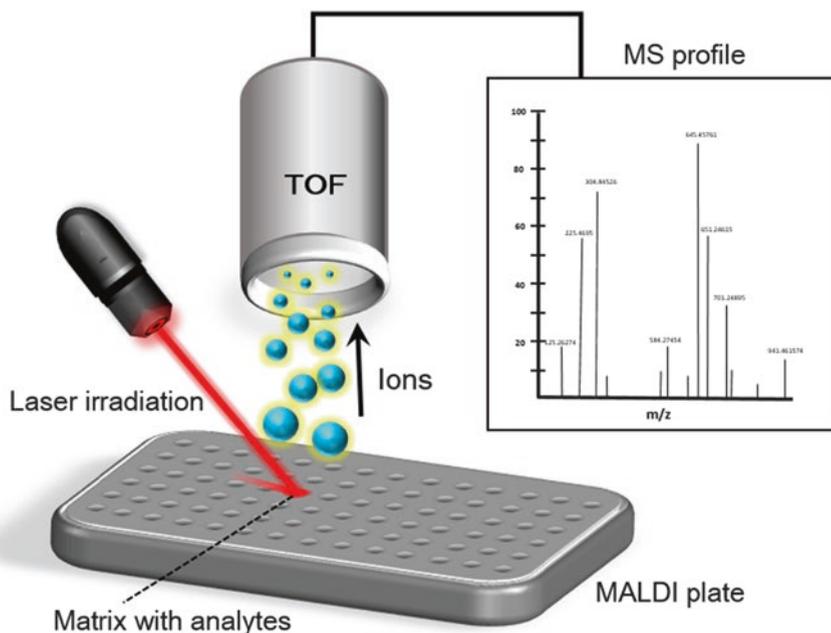
Fig. 13.4 A schematic representation of MS analysis using MALDI-TOF instrument; the sample-matrix mixture is left to dry out and then is irradiated with UV laser beam that causes its sublimation and subsequent sample ionization. The ions are then guided into the TOF by electrostatic attraction where they are analyzed depending on their mass-to-charge ratios (*m/z*); the MS spectra can then be matched to a database to identify unknown ions

Modern MALDI-TOF instruments have achieved enough sensitivity to be reliably used in single-cell analytical studies. However, there are still challenges associated with using MALDI on the single-cell level. For example, the extensive sample preparation and ionization under vacuum conditions is far away from the natural environment of the cell. Some studies aimed to alleviate this issue by performing the ionization process in normal atmospheric pressure and then transporting the ions by pneumatic assistance of a stream of nitrogen [41]. However, loss of sensitivity is inevitable.

Moreover, the matrix used in the ionization process has extensive molecular signals in the low molecular ranges (<500 Da), which incidentally is the region with most of the small molecule metabolites. This poses a significant challenge to single-cell metabolomics. Several methods have been proposed that forgo the use of a matrix all together and utilize nanophotonic effects for ionization [42]. Lipids, which dominate MALDI ionization, also mask metabolite detection as seen in many MALDI-TOF MS imaging results. Cell membranes contain a high percentage of lipids, also leading to difficulties in single-cell metabolomics. In conclusion, although MALDI-based approaches have a lot of potential in single-cell studies in regard to sensitivity, there are still challenges that need to be addressed to improve this method for use in single-cell analysis.

13.3.2.2 Live Single-Cell Mass Spectrometry

Understanding cell mechanisms and intracellular dynamics on the single-cell level is a tremendous challenge. The live single-cell mass spectrometry (LSCMS) method played an integral role in overcoming previous limitations, especially in the "omics biology." Over the past 10 years, the team of T. Masujima has been working on perfecting the live single-cell mass spectrometry protocol to aid in realization of direct and real-time molecular analysis with simultaneous visualization of a reacting single cell, which would elucidate clearly and specifically the molecular mechanisms of living systems [24].

As earlier discussed, there is a limited number of methods that can detect molecular signals, due to the low sensitivity at the cellular and subcellular levels. On the other hand, live single-cell mass spectrometry has succeeded in trapping these miniscule volumes and detecting 100–1000 of molecular peaks from a single living cell while observing the cell under video microscope. Figure 13.5 visually depicts the main steps of the LSCMS methodology.

Cells display task-oriented dynamic behavior that can be observed under a microscope [43]. While observing a cell, it exhibits various unanticipated and interesting behaviors, and discovering the mechanism behind those fascinating phenomena of life is very intriguing. Observations of cell behavior revealed that the behavior of each cell is not identical nor synchronized under the same conditions [44].

As discussed earlier, maintaining cells under the most natural conditions possible is paramount. In the LSCMS technique, in which cells are kept in their preferable medium until seconds before trapping and mass spectrometric analysis, this is the case. A metal-coated nanospray tip attached to a micromanipulator is used to suck a whole single cell or cellular contents from a specific micro-region using a tube-connected piston syringe. After capturing the cell contents at the top of the tip, it is very difficult to directly spray it into the mass spectrometer due to the high viscosity of the cells. To solve this, 2 μL of a standard ionization solvent is introduced from the rear end of the tip to aid sample quenching and ionization. Nanospray diameter sizes vary from 1 to 10 μm, which allows for greater flexibility in targeting the whole cell or its subcellular organelles or regions.

The nanospray tip's contents are fed into a nano-ESI attachment on a mass spectrometer. Nanospray ionization showed to be the most sensitive and exhaustive ionization method. The molecular contents of a cell can be extracted by nL min^{-1} level stream of an organic solvent through the nanospray tip's contents and sprayed out to the mass spectrometer [45]. Within minutes, the mass spectrometer detects 100 or 1000 of molecular peaks from the metabolites that were present in the cell under the specific conditions it faced at the time the contents were removed. These can then be identified by matching to databases to detect specific metabolites corresponding to the injected samples, which will be confirmed by their MS/MS fragmentation pattern.

Using this method, we can compare the molecular peaks of cells that are in different stages of growth, different locations, or responding to different circumstances using statistical analyses of the mass spectrometry data. If, for example, we find that certain metabolites are elevated in a specific strain, it implies that the enzyme or protein of this
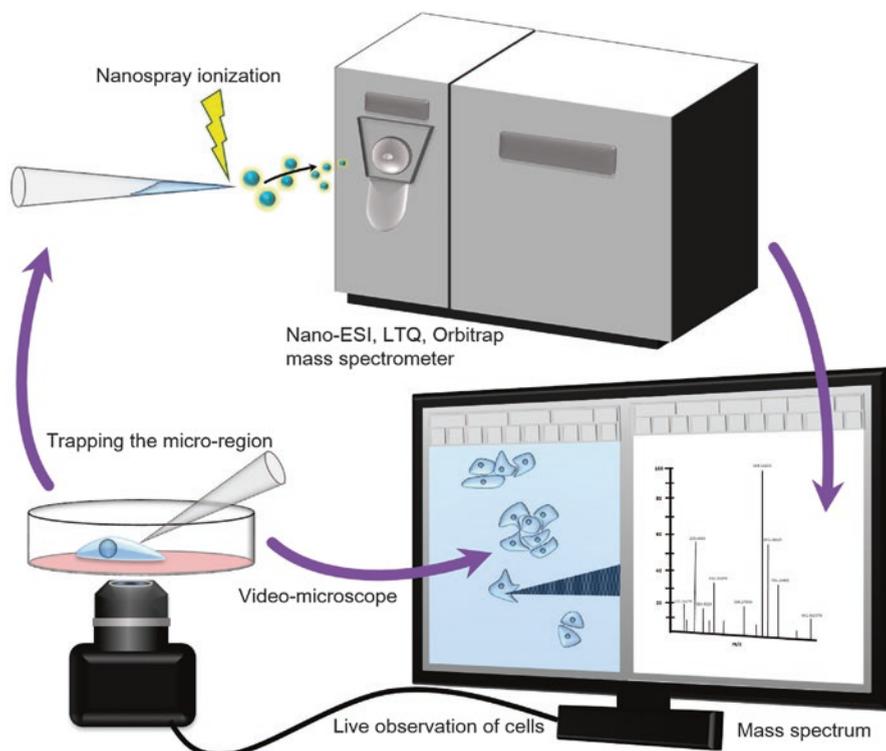
Fig. 13.5 Scheme of live single-cell mass spectrometry. The analysis is divided into two main parts. First, the cell behavior is monitored via video microscope, and the cell or organelle of interest was directly trapped inside a nanospray tip. Second, the tip's contents were then directly fed into a nano-ESI mass spectrometer after the addition of ionization solvent. Voltage is applied between the tip and mass spectrometer to obtain the mass spectra

specific metabolic pathway may be the key to the specificity of this strain and could also help us to identify new important pathways. The following sections highlight recent studies from three different fields to demonstrate possible applications of LSCMS.

Clinical

Live single-cell mass spectrometry has been mainly applied to adhesive cells, due to difficulty in sampling and isolation of suspended cells. However, live single-cell mass spectrometry succeeded in compiling the comprehensive metabolic profile of a single floating cell. At first, a single floating lymphocyte was directly trapped inside the nanospray tip from a single drop of blood from a healthy human after minimal dilution and sample treatment. Lymphocytes were chosen by visual comparison of their morphological and size differences to red blood cells. Then, efficient homogenization of the trapped cell was established by applying supersonic
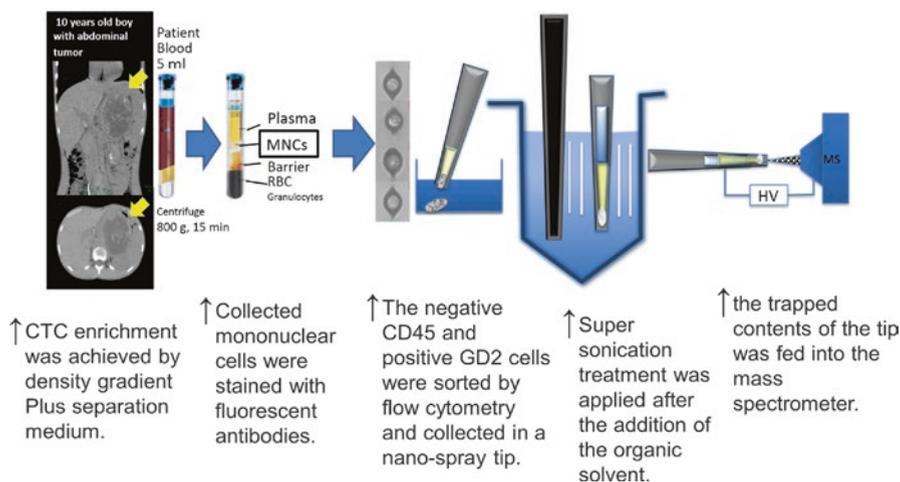
Fig. 13.6 Schematic diagram illustrating the processes of separation and analysis of circulating tumor cells using the "live single-cell mass spectrometry" method. Adapted from Ref 27 with permission from Analytical Sciences

waves, and the contents were fed into the mass spectrometer, which was conducive to acquiring a wider range of molecular peaks in the single-cell mass spectrum (Fig. 13.6). Molecular detection of higher intensities and larger number of peaks with a wider m/z range was obtained. Speculation would be that super-sonication causes outer lipid cell membrane distortion, which enhances extraction and ionization of the cell contents through mass spectrometry [27].

Discovering and perceiving the significance of cancer cells is one of the numerous potential promising applications of single-cell metabolomics. Detecting cancer cells that exhibit high metabolic rates within populations of normal cells that display normal metabolism, for example, CTCs that lead to metastasis, would be one such application. CTCs are cells shed from the primary tumor that circulate in the blood stream. Their primary function is still not clear, and their concentration in the circulating blood is usually very low (~2–10 cells per 10 mL of blood), which makes them a perfect candidate for single-cell studies. "Direct single-cell metabolomic" method was then applied to a single isolated CTC from a neuroblastoma patient's blood for a comprehensive detection of the metabolite and lipid profiles. CTCs were separated and sorted using the fluorescence flow cytometry technique. The metabolic profile of a single CTC was acquired along with detection of vital molecules such as amino acids, catecholamine metabolites, which are specific to neuroblastoma cancer and drugs from the patient's treatment regimen. This indicates that this method could be useful for monitoring drug delivery concentration levels to targeted cells. Site-specific and cell-specific metabolites were identified by matching corresponding peak numbers against the Human Metabolome Database and confirmed by establishing their MS/MS fragmentation patterns. "Direct single-cell metabolomic method" appears to have a role in future molecular diagnosis not only

for common cells but also for rare cells like CTCs that present in a very low concentration in the blood [27].

Pharmaceutical

There has been increased interest in the analysis of spatial distribution of drugs and their metabolites in various cultured cells, or in a target cell for drug discovery and development. Drug metabolism monitoring and analysis have been mostly carried out by LC-MS, which requires a large number of cells pretreated with sonication and homogenization. This leads to, in most cases, the loss of site-specific molecule identification and drug localization within a cell. However, live single-cell video MS has been developed and applied to the analysis of tamoxifen (anticancer drug) metabolism using a human hepatocellular carcinoma cell line. Cultured HepG2 cells were spiked with 5 µmol L^{-1} concentration of tamoxifen and then incubated. Using a nanospray tip, multiple organelle suctions took place from several different cells in the same incubator dish (cytoplasm, nucleus, and vacuole). Results showed detection of tamoxifen along with its five metabolites (N-desmethyl tamoxifen, 4-hydroxy tamoxifen, tamoxifen-N-oxide, 3,4-dihydroxy tamoxifen, 4-hydroxy tamoxifen N-oxide). N-Desmethyl tamoxifen, which is mainly metabolized by CYP3A4 enzyme, had the highest intensity and was preferably detected. This corroborates with the discovery that the major metabolite of tamoxifen is N-desmethyl tamoxifen and that CYP3A4 is the most expressed isozyme in P450 subfamilies of HepG2 cells. In vacuoles, tamoxifen was detected but none of its metabolites was detected. However, neither tamoxifen nor its metabolites were detected in the nucleus. Speculation would be that the cytoplasm contains a metabolizing organelle and the transport of metabolites into the nucleus and vacuoles is very limited. This study of monitoring drug metabolism on a single-cell level will pave the way for low-cost, rapid, precise, and site-specific drug monitoring and discovery [26].

Primary cultures of human hepatocytes are mainly used for predicting drug metabolism pathways in humans and detecting the differences between species' metabolic profiles. Therefore, hepatocytes were chosen as an appropriate cellular system for metabolic studies of tafluprost, an esterified prostaglandin F$_{2\alpha}$ and common drug for glaucoma. 7-Ethoxycoumarin (7-EC) was used to endorse the metabolic activity of hepatocytes in vitro.

LSCMS was applied to the analysis of tafluprost metabolites, including tafluprost acid (TA), dinor-tafluprost (dinor-TA), tetranor-TA and common phase I metabolites, hydroxylated 1,2,3,4-tetranor-TA (tetranor-TA-OH), and hydroxylated 1,2-dinor-TA (dinor-TA-OH) in a single hepatocyte. These data were compared with the averaged results obtained from multiple cells. A picoliter amount of cytoplasm and granules in the cell were captured in a nanospray tip, and the ionization solvent was added. The tip was then introduced to the nano-ESI interface of mass spectrometer to obtain a single-cell spectrum. Tafluprost metabolism results from the multiple hepatocyte analysis using LC-MS showed averaged metabolism to tafluprost acid (TA) and β-oxidized metabolites. On the other hand, LSCMS

indicated variation in tafluprost metabolism among individual cells showed significant variation in the quantity of TA and dinor-TA. In contrast, there was no significant variation of 7-ethoxycoumarin metabolism. This method succeeded in detecting the reported metabolic profile in the cytoplasm, and those metabolites matched a metabolic pathway and showed a variety of metabolic functions on the single-cell level. Therefore, LSCMS showed successful detection of drug metabolism heterogeneity in a single living hepatic cell. This approach has the potential for indicating the correlation between drug metabolism and the pharmacological as well as the toxicological effects taking place in cultured cells on single-cellular and subcellular levels [46].

Agricultural

Food is life, and maintaining a comprehensive metabolomic image of plant cells will unlock several pathways into improving crop yield, eliminating unwanted pests, and enhancing desired traits in crops. It is important to pursue studies of plant metabolomics and biochemistry because these will provide more insight on the natural molecular mechanisms and dynamic activities taking place inside plant cells. As a result of protein and enzyme activation, plant cell dynamical functioning produces metabolites corresponding to a specific enzyme or protein. Understanding those activities will impart an important outlook on the full image of how plants function, and recognizing plant genotypes will allow the regulation of such processes.

There are many factors to be taken into consideration in the single-cell analysis of plant tissues. Unlike cultured cells, plant tissues have an irregular surface, strong cellulose walls, and higher dilution of biomolecules inside the cell. As a result, LSCMS has been extended to obtain rapid, versatile, and noninvasive direct single-cell plant analysis, which is published in *Nature Protocols* [22]. This technique provides a molecular profile including metabolites, lipids, hormones, and nutrients of a single plant cell within minutes with minimal pretreatment (Fig. 13.7). If the plant tissues remained intact after analysis, morphological changes could be monitored along with metabolic pathways processes. This method was applied to leaves, stem, and petal from a healthy *Pelargonium zonale* plant. Collected data showed that there were specific metabolites, which present only in the leaf such as geranic acid, while methyl citronellate was detected in both the leaf and stem, but absent in the petal [21]. With the provided information, site-specific molecules and chemical composition of each site in the cell could be distinguished along with comparing between different plant samples. Furthermore, this method could be useful in several practical and industrial applications such as quality control of crop treatment and medicinal plants, food analysis, and controlling plant diseases. Finally, this will open a new outlook in the research done in agricultural sciences.
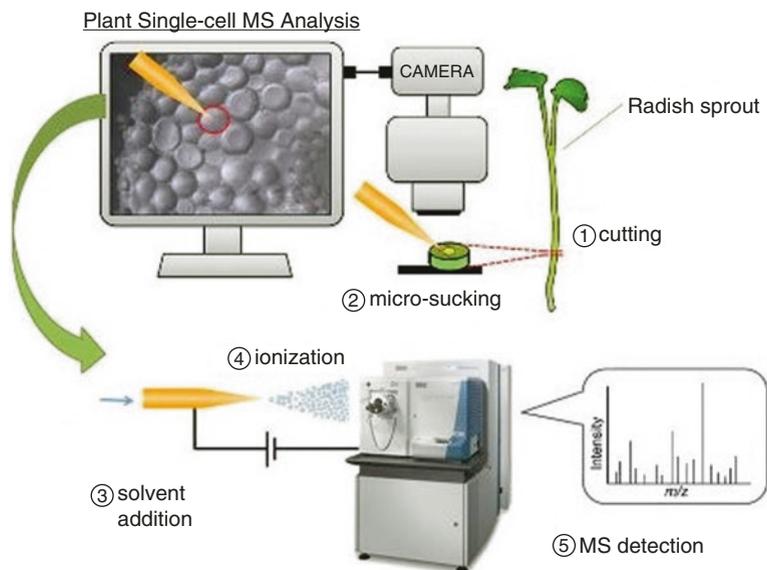
Fig. 13.7 Illustrative figure showing the steps of the "live single-cell mass spectrometry" technique. The fresh plant was cut with minimal sample pretreatment and observed under video microscope. The targeted cell was captured by a micropipette. Ionization solvent was added from the rear end, and the tip contents were introduced by electrospray ionization (ESI) to a mass spectrometer, thereby obtaining the metabolic profile of a single plant cell. Reproduced from Ref 22 with permission from the Nature Publishing Group

13.4   Future Prospects

Single-cell analysis is a rapidly growing field of biology with much room for improvement, but many challenges remain to be addressed; after all, the field is still in its infancy. As previously mentioned, there are several hurdles regarding sample considerations, isolation of single cells, and their subsequent analysis. In order to circumvent said challenges, LSCMS was developed, which combines nanoscale devices, ambient pressure ionization, and sensitive mass spectrometric measurements while maintaining high resolution by the use of Orbitrap technology. Despite all of that, there are still limitations and challenges ahead; since the sampling is performed manually and identification of metabolites by matching with databases is done offline, several concerns have been raised in regard to the throughput of the method itself. It might be argued that obtaining "high-quality" data by insuring that the cell was sampled in its natural environment, i.e., Petri dish, is better than increasing throughput by utilizing more aggressive isolation techniques. However, the fact still remains that improving throughput is a major challenge that needs to be addressed so that the method can be applied to large-scale studies of cell metabolomics.
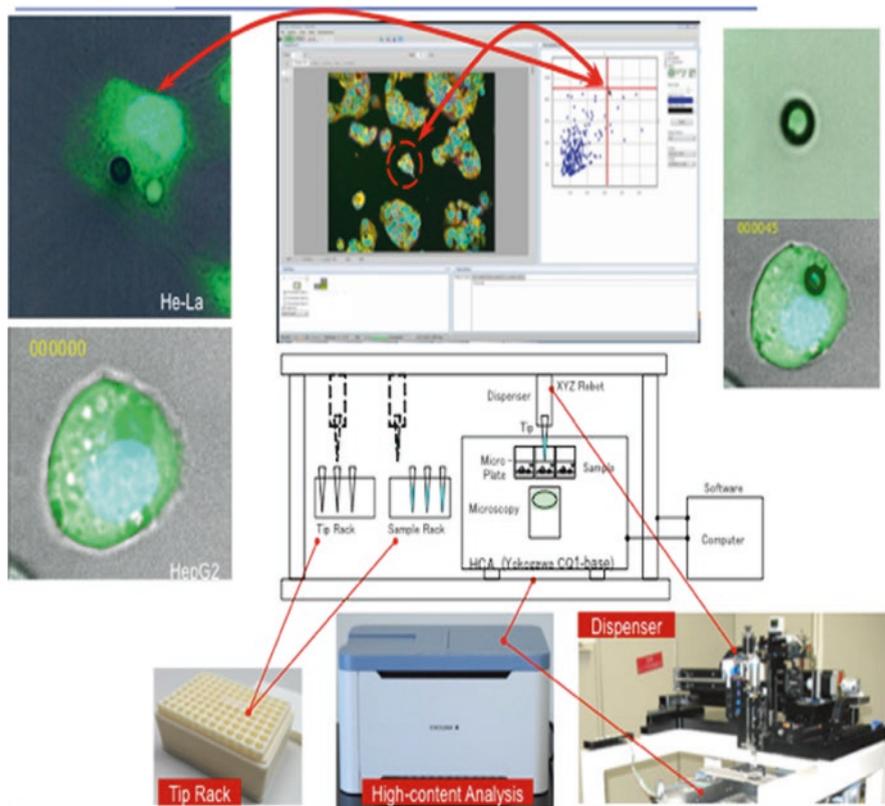
Fig. 13.8 High-content single-cell analysis system is shown; the cell is chosen by an image analysis algorithm according to its morphology or fluorescent markers. Then, the robot proceeds to sample a whole cell or part of a cell by using a specially made hollow nanospray tip. Finally, the trapped cell is stored in a special tip rack for downstream analysis or long-term storage

The sampling phase is the most time-consuming aspect of single-cell analysis, and LSCMS is no exception. In order to increase throughput, the process of selection of suitable cells and sampling was automatized in a collaborative study between the Masujima team and Yokogawa electric company. By coupling a motorized x-y stage to an automated software system and a dispenser robot, the newly developed system is capable of selecting the desired cells according to traits previously inputted and then proceed to automatically pick up the whole cell or an organelle, such as nucleus or cytoplasm. After that, the nanospray tips are stored in a specially made tip rack and can be frozen or analyzed immediately using ESI-MS as shown in Fig. 13.8.

Another challenge exists in the sample type itself; since single cells are considered as a complex biological matrix, sometimes the need for an efficient separation and enrichment method arises. Due to the low sample volume associated with single

cells, it is quite problematic to use a conventional separation method like high-performance liquid chromatography due to sample loss and dilution. It is also worth noting that mass spectrometry by itself cannot differentiate between optical isomers and ions with identical mass-to-charge (m/z) ratios. In the latter case, fragmentation and MS/MS studies can differentiate between identical m/z ions, but a strong signal is required for this to be done successfully – something that depends on the sample and the sensitivity of the instrument.

The ideal method for single-cell analysis would combine the minimal disruptiveness of nanoscale direct sampling of cells in their culture plates along with the high throughput of microfluidic devices coupled with an ionization source that operates in normal atmospheric conditions that also does not cause excessive heating or damage to the biomolecules themselves during the ionization process. The ionization source itself should incorporate a separation step according to ion mobility, for example, or other factors so that it is possible to differentiate between optical isomers. Finally, the mass analyzer used should have the highest sensitivity possible along with a sufficient resolution to differentiate between ions with closely similar mass-to-charge ratio.

Until now, no method proposed for single-cell analysis is perfect in all aspects. Balancing throughput, accuracy, and invasiveness of the isolation method while choosing the perfect analytical technique that combines high sensitivity, resolution, and selectivity is the ultimate goal to be achieved in order to contribute to a wide range of fields such as diagnostics, cancer treatment, agriculture, and many others. In principle, analysis on a single-cell level is an interdisciplinary science, and we hope that this overview will help in encouraging collaborative studies between different fields of life sciences so that we can gain a better understanding of the most important building block in our bodies.

References

1. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the human metabolome database. Nucleic Acids Res. 2007;35:D521–6. doi:10.1093/nar/gkl923.
2. Dettmer K, Hammock BD. Metabolomics – a new exciting field within the omics sciences. Environ Health Perspect. 2004;112:A396–7.
3. Patti GJ, Yanes O, Siuzdak G. Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13:263–9. doi:10.1038/nrm3314.
4. Wang D, Bodovitz S. Single cell analysis: the new frontier in 'omics'. Trends Biotechnol. 2010;28:281–90.
5. Snijder B, Pelkmans L. Origins of regulated cell-to-cell variability. Nat Rev Mol Cell Biol. 2011;12:119–25.
6. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297:1183–6. doi:10.1126/science.1070919.
7. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008;135:216–26.
8. Sender R, Fuchs S, Milo R. Preprint on bioRxiv. 2015. doi:10.1101/036103.
9. Ranjan A, Townsley BT, Ichihashi Y, Sinha NR, Chitwood DH. An intracellular transcriptomic atlas of the giant coenocyte *Caulerpa taxifolia*. PLoS Genet. 2015;11:e1004900. doi:10.1371/journal.pgen.1004900.

10. Moss S. Do birds have knees?: All your bird questions answered. London: Bloomsbury Publishing; 2016.
11. Mattsson A, Kärrman A, Pinto R, Brunström B. Metabolic profiling of chicken embryos exposed to perfluorooctanoic acid (PFOA) and agonists to peroxisome proliferator-activated receptors. PLoS One. 2015;10:e0143780. doi:10.1371/journal.pone.0143780.
12. Vastag L, Jorgensen P, Peshkin L, Wei R, Rabinowitz JD, Kirschner MW. Remodeling of the metabolome during early frog development. PLoS One. 2011;6:e16881. doi:10.1371/journal.pone.0016881.
13. Onjiko RM, Moody SA, Nemes P. Single-cell mass spectrometry reveals small molecules that affect cell fates in the 16-cell embryo. Proc Natl Acad Sci U S A. 2015;112:6545–50.
14. Fuller RR, Moroz LL, Gillette R, Sweedler JV. Single neuron analysis by capillary electrophoresis with fluorescence spectroscopy. Neuron. 1998;20:173–81. doi:10.1016/S0896-6273(00)80446-8.
15. Page JS, Rubakhin SS, Sweedler JV. Direct cellular assays using off-line capillary electrophoresis with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Analyst. 2000;125:555–61.
16. Rubakhin SS, Greenough WT, Sweedler JV. Spatial profiling with MALDI MS: distribution of neuropeptides within single neurons. Anal Chem. 2003;75:5374–80.
17. Rubakhin SS, Sweedler JV. Characterizing peptides in individual mammalian cells using mass spectrometry. Nat Protoc. 2007;2:1987–97.
18. Romanova EV, Rubakhin SS, Monroe EB, Sweedler JV. Single cell analysis: technologies and applications, vol. 56. Weinheim: Wiley-VCH; 2016.
19. Zimmerman TA, Rubakhin SS, Romanova EV, Tucker KR, Sweedler JV. MALDI mass spectrometric imaging using the stretched sample method to reveal neuropeptide distributions in aplysia nervous tissue. Anal Chem. 2009;81:9402–9.
20. Shrestha B, Vertes A. In situ metabolic profiling of single cells by laser ablation electrospray ionization mass spectrometry. Anal Chem. 2009;81:8265–71.
21. Tejedor ML, Mizuno H, Tsuyama N, Harada T, Masujima T. In situ molecular analysis of plant tissues by live single-cell mass spectrometry. Anal Chem. 2012;84:5221–8. doi:10.1021/ac202447t.
22. Fujii T, Matsuda S, Tejedor ML, Esaki T, Sakane I, Mizuno H, Tsuyama N, Masujima T. Direct metabolomics for plant cells by live single-cell mass spectrometry. Nat Protoc. 2015;10:1445–56. doi:10.1038/nprot.2015.084.
23. Tsuyama N, Mizuno H, Tokunaga E, Masujima T. Live single-cell molecular analysis by video-mass spectrometry. Anal Sci. 2008;24:559–61.
24. Mizuno H, Tsuyama N, Harada T, Masujima T. Live single-cell video-mass spectrometry for cellular and subcellular molecular detection and cell classification. J Mass Spectrom. 2008;43:1692–700. doi:10.1002/jms.1460.
25. Mizuno H, Tsuyama N, Date S, Harada T, Masujima T. Live single-cell metabolomics of tryptophan and histidine metabolites in a rat basophil leukemia cell. Anal Sci. 2008;24:1525–7.
26. Date S, Mizuno H, Tsuyama N, Harada T, Masujima T. Direct drug metabolism monitoring in a live single hepatic cell by video mass spectrometry. Anal Sci. 2012;28:201–3.
27. Hiyama E, Ali A, Amer S, Harada T, Shimamoto K, Furushima R, Abouleila Y, Emara S, Masujima T. Direct lipido-metabolomics of single floating cells for analysis of circulating tumor cells by live single-cell mass spectrometry. Anal Sci. 2015;31:1215–7. doi:10.2116/analsci.31.1215.
28. Ibáñez AJ, Fagerer SR, Schmidt AM, Urban PL, Jefimovs K, Geiger P, et al. Mass spectrometry-based metabolomics of single yeast cells. Proc Natl Acad Sci U S A. 2013;110:8790–4. doi:10.1073/pnas.1209302110.
29. Taniguchi Y. Genome-wide analysis of protein and mRNA copy numbers in single Escherichia coli cells with single-molecule sensitivity. Methods Mol Biol. 2015;1346:55–67.
30. Ohno M, Karagiannis P, Taniguchi Y. Protein expression analyses at the single cell level. Molecules. 2014;19:13932–47.

31. Taniguchi Y, Choi PJ, Li G, Chen H, Babu M, et al. Quantifying E-coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. 2010;329:533–8.
32. Essaka DC, Prendergast J, Keithley RB, Hindsgaul O, Palcic MM, Schnaar RL, Dovichi NJ. Single cell ganglioside catabolism in primary cerebellar neurons and glia. Neurochem Res. 2012;37:1308–14.
33. Zenobi R. Single-cell metabolomics: analytical and biological perspectives. Science. 2013;342:1243259. doi:10.1126/science.1243259.
34. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature. 2007;445:106–10. doi:10.1038/nature05372.
35. Ungai-Salánki R, Gerecsei T, Fürjes P, Orgovan N, Sándor N, Holczer E, et al. Automated single cell isolation from suspension with computer vision. Sci Rep. 2016;9:20375. doi:10.1038/srep20375.
36. Yan R, Park JH, Choi Y, Heo CJ, Yang SM, Lee LP, et al. Nanowire-based single-cell endoscopy. Nat Nanotechnol. 2011;7:191–6. doi:10.1038/nnano.2011.226.
37. Boukany PE, Morss A, Liao WC, Henslee B, Jung H, Zhang X, et al. Nanochannel electroporation delivers precise amounts of biomolecules into living cells. Nat Nanotechnol. 2011;6:747–54. doi:10.1038/nnano.2011.164.
38. Carlo DD, Wu LY, Lee LP. Dynamic single cell culture array. Lab Chip. 2006;6:1445–9. doi:10.1039/B605937F.
39. Dimov IK, Kijanka G, Park Y, Ducrée J, Kang T, Lee LP. Integrated microfluidic array plate (iMAP) for cellular and molecular analysis. Lab Chip. 2011;11:2701–10. doi:10.1039/c1lc20105k.
40. Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. 2013;26:547–603. doi:10.1128/CMR.00072-12.
41. Laiko VV, Baldwin MA, Burlingame AL. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem. 2000;72:652–7.
42. Walker BN, Stolee JA, Vertes A. Nanophotonic ionization for ultratrace and single-cell analysis by mass spectrometry. Anal Chem. 2012;84:7756–62. doi:10.1021/ac301238k.
43. Masujima T. Visualized single cell dynamics and analysis of molecular tricks. Anal Chim Acta. 1999;400:33–43. doi:10.1016/S0003-2670(99)00704-7.
44. Masujima T. Live single-cell mass spectrometry. Anal Sci. 2009;25:953–60.
45. Tejedor LM, Mizuno H, Tsuyama N, Harada T, Masujima T. Direct single-cell molecular analysis of plant tissues by video mass spectrometry. Anal Sci. 2009;25:1053–5.
46. Fukano Y, Tsuyama N, Mizuno H, Date S, Takano M, Masujima T. Drug metabolite heterogeneity in cultured single cells profiled by pico-trapping direct mass spectrometry. Nanomedicine. 2012;7:1365–74.

Index