Handbook of Experimental Pharmacology 277

Veronica Ghini
Kathleen A. Stringer
Claudio Luchinat *Editors*

Metabolomics and Its Impact on Health and Diseases

Springer

Handbook of Experimental Pharmacology

Volume 277

Editor-in-Chief

Martin C. Michel, Dept of Pharmacology, Johannes Gutenberg Universität, Mainz, Germany

Editorial Board Members

James E. Barrett, Center for Substance Abuse Research, Lewis Katz School of Medicine at Temple University, Philadelphia, PA, USA

David Centurión, Dept. of Pharmabiology, Center for Research and Advanced Studies, Col. Granjas-Coapa, Mexico

Veit Flockerzi, Institute for Experimental and Clinical Pharmacology and Toxicology, Saarland University, Homburg, Germany

Pierangelo Geppetti, Headache Center, University of Florence, Florence, Italy

Franz B. Hofmann, Forschergruppe 923 Carvas, Technical University, München, Germany

Kathryn Elaine Meier, Dept. of Pharmaceutical Sciences, Washington State University Spokane, Spokane, USA

Clive P. Page, SIPP, Kings College London, London, UK

KeWei Wang, School of Pharmacy, Qingdao University, Qingdao, China

The *Handbook of Experimental Pharmacology* is one of the most authoritative and influential book series in pharmacology. It provides critical and comprehensive discussions of the most significant areas of pharmacological research, written by leading international authorities. Each volume in the series represents the most informative and contemporary account of its subject available, making it an unrivalled reference source.

HEP is indexed in PubMed and Scopus.

Veronica Ghini • Kathleen A. Stringer •
Claudio Luchinat
Editors

Metabolomics and Its Impact on Health and Diseases

Editors
Veronica Ghini
Department of Chemistry
University of Florence
Sesto Fiorentino, Florence, Italy

Kathleen A. Stringer
Department of Clinical Pharmacy
University of Michigan–Ann Arbor
Ann Arbor, MI, USA

Claudio Luchinat
Department of Chemistry
University of Florence
Sesto Fiorentino, Florence, Italy

ISSN 0171-2004 ISSN 1865-0325 (electronic)
Handbook of Experimental Pharmacology
ISBN 978-3-031-26858-8 ISBN 978-3-031-26859-5 (eBook)
https://doi.org/10.1007/978-3-031-26859-5

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume of the *Handbook of Experimental Pharmacology*, which celebrated its 100th anniversary in 2019, addresses the rapidly growing and evolving field of metabolomics. It has been compiled and designed to broaden and enrich your understanding as well as simplify a complicated picture of the diverse field of metabolomics. This is accomplished by chapters from experts in the field on basic principles as well as reviews and updates of analytical techniques. The variety and different perspectives of the nuclear magnetic resonance approaches are described in the chapters authored by David S. Wishart and co-authors, G. A. Nagana Gowda and Daniel Raftery, and Ryan T. McKay. Advances in mass spectrometry are covered by Charles R. Evans and co-authors and Stefan Kempa and co-authors. This book also reflects the state of the art in the application of metabolomics to cell biology (Ulrich L. Günther and co-authors) and chapters that share insights into the application of metabolomics. These include the assessment of treatment response (Paola Turano and co-authors) and the phenotyping of various diseases (Rachel S. Kelly and co-authors, Paige Lacy and co-authors, and Angela J. Rogers and co-author). Relationships of metabolomics and drugs are highlighted by Robert Verpoorte and co-authors, Oscar Millet and co-authors, and Daniel L. Hertz and co-author. Given the diverse topics addressed, we believe this book has interdisciplinary appeal and scholars with an interest in the role of metabolomics in achieving precision medicine will find it of particular or special interest.

We want to thank the authors for their contributions as this *Handbook* would not be possible without them. We also express our appreciation to the many investigators who work in the field of metabolomics and strive to advance the science. Its analytical advancement as well as its translation to the clinic is of vital importance to the field. It is our belief that metabolomics will continue to provide new and novel insights into complex illnesses and enable more accurate and precise therapies in the future. We would also like to express our sincere appreciation to Susanne Dathe, Springer Editor for Neurosciences/Pharmaceutical Sciences/Protocols, whose commitment and competence have helped to continue the tradition

of this remarkable series, and to the past and current editorial board members who have dedicated time and effort into establishing this series as one of the most recognized publications in pharmacology.

Florence, Italy Veronica Ghini
Ann Arbor, MI, USA Kathleen A. Stringer
Florence, Italy Claudio Luchinat

Contents

Practical Aspects of NMR-Based Metabolomics 1
David S. Wishart, Manoj Rout, Brian L. Lee, Mark Berjanskii,
Marcia LeVatte, and Matthias Lipfert

Compound Identification Strategies in Mass Spectrometry-Based Metabolomics and Pharmacometabolomics 43
Rylan Hissong, Kendra R. Evans, and Charles R. Evans

Metabolomics and NMR 73
Ryan T. McKay

Natural Products Drug Discovery: On Silica or In-Silico? 117
Hye Kyong Kim, Young Hae Choi, and Robert Verpoorte

Quantitative NMR Methods in Metabolomics 143
G. A. Nagana Gowda and Daniel Raftery

Advancements in Pulsed Stable Isotope-Resolved Metabolomics 165
Martin Forbes, Sabrina Geisberger, Matthias Pietzke, Guido Mastrobuoni,
and Stefan Kempa

Metabolomics in Cell Biology 181
Zuhal Eraslan, Marta Cascante, and Ulrich L. Günther

NMR-Based Metabolomics to Evaluate Individual Response to Treatments 209
Alessia Vignoli, Gaia Meoni, Veronica Ghini, Francesca Di Cesare,
Leonardo Tenori, Claudio Luchinat, and Paola Turano

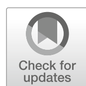
Pharmacometabolomics of Asthma as a Road Map to Precision Medicine 247
Rachel S. Kelly, Margaret F. Cote, Sofina Begum, and Jessica Lasky-Su

Prospective Metabolomic Studies in Precision Medicine: The AKRIBEA Project 275
Maider Bizkarguenaga, Rubén Gil-Redondo, Chiara Bruzzone,
Ganeko Bernardo-Seisdedos, Ana Laín, Beatriz González-Valle,
Nieves Embade, José M. Mato, and Oscar Millet

Chemotherapy-Induced Peripheral Neuropathy 299
Ciao-Sin Chen and Daniel L. Hertz

Metabolomics of Respiratory Diseases 339
Subhabrata Moitra, Arghya Bandyopadhyay, and Paige Lacy

The Metabolomics of Critical Illness 367
Ana E. Pacheco-Navarro and Angela J. Rogers



Practical Aspects of NMR-Based Metabolomics

David S. Wishart, Manoj Rout, Brian L. Lee, Mark Berjanskii, Marcia LeVatte, and Matthias Lipfert

Contents

1 Introduction ... 2
2 A Short History of NMR-Based Metabolomics ... 5
3 Practical Aspects of NMR-Based Metabolomics ... 6
3.1 Sample Preparation ... 6
3.2 Choosing the Right Pulse Sequence ... 10
3.3 Spectral Acquisition and Processing ... 14
4 Data Analysis ... 18
4.1 Data Analysis for Targeted Metabolomics ... 18
4.2 Data Analysis for Untargeted Metabolomics ... 22
4.3 Targeted or Untargeted? ... 24
5 Biological Interpretation ... 24
6 Conclusion and Future Prospects ... 27
References ... 31

D. S. Wishart (✉)
Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

Department of Computing Science, University of Alberta, Edmonton, AB, Canada

Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB, Canada

Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada
e-mail: dwishart@ualberta.ca

M. Rout · B. L. Lee · M. Berjanskii · M. LeVatte
Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada
e-mail: mkrout@ualberta.ca; bll@ualberta.ca; mb1@ualberta.ca; levatte@ualberta.ca

M. Lipfert
Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

Reference Standard Management & NMR QC, Lonza Group AG, Visp, Switzerland
e-mail: lipfert@ualberta.ca

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_613

Abstract

While NMR-based metabolomics is only about 20 years old, NMR has been a key part of metabolic and metabolism studies for >40 years. Historically, metabolic researchers used NMR because of its high level of reproducibility, superb instrument stability, facile sample preparation protocols, inherently quantitative character, non-destructive nature, and amenability to automation. In this chapter, we provide a short history of NMR-based metabolomics. We then provide a detailed description of some of the practical aspects of performing NMR-based metabolomics studies including sample preparation, pulse sequence selection, and spectral acquisition and processing. The two different approaches to metabolomics data analysis, targeted vs. untargeted, are briefly outlined. We also describe several software packages to help users process NMR spectra obtained via these two different approaches. We then give several examples of useful or interesting applications of NMR-based metabolomics, ranging from applications to drug toxicology, to identifying inborn errors of metabolism to analyzing the contents of biofluids from dairy cattle. Throughout this chapter, we will highlight the strengths and limitations of NMR-based metabolomics. Additionally, we will conclude with descriptions of recent advances in NMR hardware, methodology, and software and speculate about where NMR-based metabolomics is going in the next 5–10 years.

Keywords

Applications · Experimental methods · NMR spectroscopy · Targeted metabolomics · Untargeted metabolomics

1 Introduction

Metabolomics is a branch of analytical chemistry that comprehensively characterizes the molecules in various biofluids and tissues. Metabolites are the chemical constituents of the metabolome. The metabolome, therefore, can be defined as the complete collection of all chemicals or metabolites found within cells, biofluids, organs, or organisms (Oliver et al. 1998; Wishart 2005). These chemicals (most of which have a molecular weight <1,500 Da) include endogenously derived compounds (amino acids, nucleic acids, organic acids, carbohydrates, lipids, and minerals) and exogenously acquired compounds (vitamins, food additives, plant phytochemicals, drugs, cosmetic chemicals, dyes, detergents) or just about any other chemical that an organism can consume or to which it can be exposed.

Metabolites are essential building blocks for all life processes. They serve as the bricks and mortar for cells, being the small molecule constituents (proteins, RNA, DNA) for all basic cellular functions. Furthermore, metabolites are the fuel for cellular processes, the barriers to maintain cellular integrity, and messengers for signaling processes. Metabolites are the end-products of complex processes which

are encoded for and controlled by genes. Therefore, metabolites are exquisitely sensitive indicators of problems in the genome. Indeed, a single base change in a gene can lead to a 10,000-fold change in the concentrations of certain metabolites (Wishart et al. 2007). This remarkable sensitivity of metabolite levels to genetic variations led to one of the most common and widespread medical testing initiatives – newborn screening. For more than 100 years, metabolite testing has been used to identify and detect genetic diseases and inborn errors of metabolism or IEMs, such as phenylketonuria or alkaptonuria (Levy 2010). Metabolites are not only sensitive to genome-related processes, but also to what happens in the environment. In particular, metabolite concentrations are influenced by nutrition, exposure to workplace or household chemicals, physical activity, the time of day, or even the outside temperature (Bassini and Cameron 2014; Brown 2016).

Because metabolites are affected by what happens intra-cellularly (via the genome) and extra-cellularly (via the environment), metabolomics provides a detailed view of the gene–environment interactions. Metabolomics is therefore an ideal route for scientists to access and measure an organism's "chemical phenotype" (Fiehn 2002). This represents an important advantage of metabolomics over genomics analyses. While the genome can suggest what *might* happen, the metabolome actually indicates what *is* happening.

Continued advances in analytical chemistry and computational data analysis have made the study of metabolomics more accessible to a wider range of scientific disciplines. These advances have led to metabolomics being routinely used in disease screening, drug discovery, food and nutritional analysis, veterinary studies, crop assessment, biomaterial production, and environmental monitoring (Holmes et al. 2008; Viant 2008; Wishart 2008a, 2016; Kim et al. 2016). Indeed, metabolomics research has grown exponentially since 1999 which reported just two metabolomics papers to nearly 9,000 papers published in 2020.

Metabolomics experiments are relatively simple to perform. The general workflow to collect metabolomics data is shown in Fig. 1. The experiments begin with a biological sample which can be a biofluid or a tissue. For tissues, the metabolites must first be extracted or homogenized to produce a fluid. Once an appropriate metabolite extract or biofluid has been obtained, the liquid sample must be analyzed by one or more analytical chemistry platforms. The most popular platforms are liquid chromatography mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) spectroscopy. These analytical platforms are ultimately responsible for helping to identify and/or quantify the chemicals in the different biological mixtures. With the help of specialized software and carefully developed databases of compounds, the data generated from these platforms can be used to identify hundreds of compounds in the biological samples.

While LC-MS methods account for $>70\%$ of published metabolomics studies to date, NMR-based methods still garner considerable interest among metabolomics researchers. For instance, more than 1,200 NMR-based metabolomics papers were published in 2020, the most ever published in any given year. This suggests that NMR-based metabolomics is growing, and it still has plenty to offer to the metabolomics community. NMR has some unique advantages compared to other

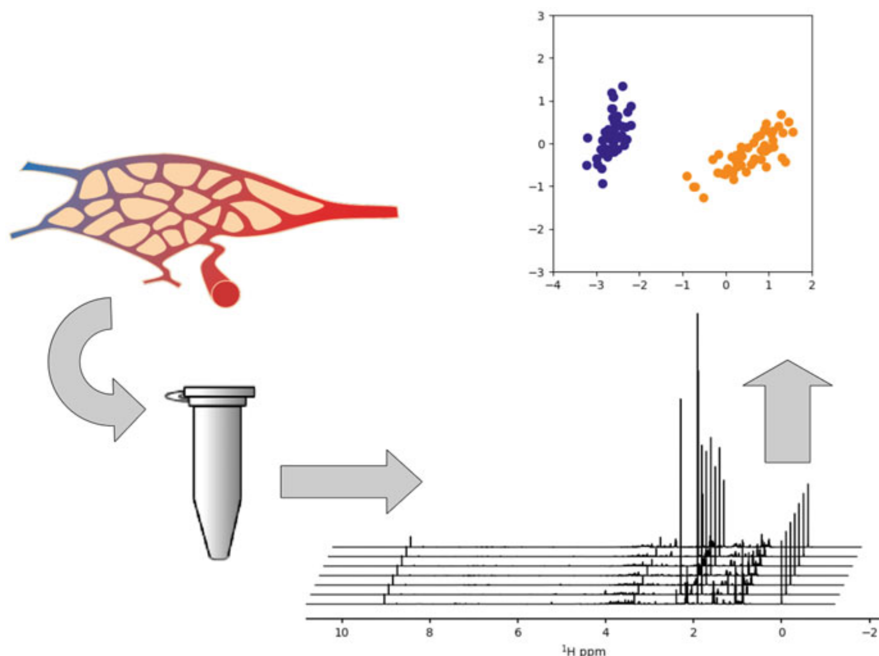


Fig. 1 A simplified workflow for metabolomics. Tissue samples may be obtained and homogenized, or biofluids such as urine or blood may be collected. Spectra of the liquid portion of the samples can be acquired using NMR or mass spectrometry, and the resulting data used for analysis

platforms such as LC-MS or gas chromatography mass spectrometry (GC-MS). The most obvious advantage is its non-destructive nature. Moreover, NMR is non-biased, supports accurate metabolite quantification, requires little or no compound separation, allows the identification of novel compounds, and does not require chemical derivatization. Furthermore, NMR is highly automatable and is exceptionally reproducible, making automated, high-throughput metabolomics studies with NMR spectroscopy much more feasible than with LC-MS or GC-MS. Moreover, NMR can detect and characterize compounds that are difficult for LC-MS, such as sugars, organic acids, alcohols, polyols, and other highly polar or low molecular weight compounds.

In this chapter, we will provide an overview of NMR-based metabolomics with an emphasis on the practical aspects of NMR-based applications. First, we will discuss a brief history of NMR-based metabolomics followed by a short discussion of the two different approaches to metabolomics, targeted vs. untargeted. Next, a discussion of the practical aspects of NMR sample preparation, as well as NMR spectral recording and acquisition will be provided. Then, we will explain how to interpret NMR spectra collected from biological samples using targeted or untargeted data analyses and provide references to several software packages to

help users process NMR spectra obtained via these two different approaches. We also illustrate several examples of where and how NMR-based metabolomics has been successfully implemented. Finally, we speculate on the future of NMR-based metabolomics and the potential areas of growth for this field.

2 A Short History of NMR-Based Metabolomics

For the past 40 years, NMR has played a central role in the understanding of metabolism and metabolomic processes. The first example demonstrating how NMR could be used in metabolic studies was published in 1974 when a ^{13}C isotope-tracer analysis was combined with NMR studies to decipher specific details of ethanol metabolism (Wilson and Burlingame 1974). Since then, steady improvements in NMR technology, NMR field strength, and other advancements have increased the popularity of NMR for many applications in biochemistry and metabolism. Indeed, NMR quickly became the tool of choice for many metabolism research areas, particularly those studying drug metabolism. During the 1970s and 1980s, NMR spectroscopy was used in studies to explore drug kinetics, drug metabolism, and the identification of drug metabolites (Midgley and Hawkins 1978; Williams et al. 1979). These studies were complemented by more traditional metabolic studies, focusing on cellular, microbial, plant, and animal metabolism using in vivo or in vitro ^{1}H, ^{13}C, and ^{31}P NMR techniques (Cohen et al. 1979; Weiner et al. 1989; Rothman et al. 2003). The widespread use of NMR for "classical" metabolic studies combined with its exceptional capacity to handle complex metabolomic mixtures made NMR the preferred analytical chemistry platform for launching the new field of metabolomics. Indeed, the very first metabolomics (or metabonomics) studies were conducted using NMR (Bock 1982; Yoshikawa et al. 1982; Bales et al. 1984a, b). By the late 1980s, NMR-based metabolomics studies of human plasma led to the identification of several putative biomarkers for cancer and coronary artery diseases (Fossel et al. 1986; Otvos et al. 1991). These studies were complemented by pioneering studies by Nicholson and colleagues who used NMR spectra from urine to characterize inborn errors of metabolism (IEM) and drug toxicity. These early studies proved that urine, a very complex biofluid, can be successfully analyzed by NMR (Bales et al. 1984a, b; Nicholson et al. 1984a, b).

The greatest challenge facing these early metabolomics researchers was the sheer complexity of the NMR spectra they were collecting from biofluids and tissue extracts. A metabolically rich biofluid such as urine can contain up to 5,000 detectable proton resonances in countless variations depending on the pH or concentrations of salts in the sample (Nicholson and Wilson 2003). This complexity of the spectral output led researchers to develop two different approaches for collecting, processing, analyzing, and interpreting metabolomics NMR data. One approach called "targeted metabolomics" uses spectral deconvolution software to identify and quantify fluid-specific or known metabolites in individual NMR spectra. The second approach called "untargeted metabolomics," or statistical spectroscopy, uses spectral alignment, spectral binning, and multivariate statistical analysis to

identify spectral features of interest. Once the key features are identified, the corresponding compounds and metabolites may or may not be identified. Both targeted and untargeted approaches have their advantages and disadvantages. While targeted metabolomics is more precise, highly quantitative, and much more reproducible, it is more time-consuming, more limited in scope and it does not allow one to identify novel chemicals. Untargeted metabolomics is faster, relatively open ended, and more useful for identification of novel compounds but it is not quantitative nor is it particularly reproducible. Given the importance of targeted and untargeted approaches in NMR-based metabolomics and given their requirements for different types of data analysis techniques, we will discuss them in more detail in Sect. 4.

3 Practical Aspects of NMR-Based Metabolomics

This section will provide an overview of the practical aspects and consensus recommendations for conducting liquid-state NMR metabolomics studies with a primary focus on ^{1}H NMR of biofluids or fluidized tissue extracts (McKay, "Metabolomics using NMR – avoiding the black box"; Raftery, "Quantitative NMR methods in metabolomics"). Readers interested in obtaining precise protocols or information about solid-state NMR or magic angle sample spinning (MAS) NMR should refer to other excellent reviews and book chapters (Weber et al. 2012; Wolak et al. 2012; Nagana Gowda and Raftery 2014; Nagana Gowda et al. 2015; Zhang et al. 2016; Mazzei and Piccolo 2017; Tilgner et al. 2019).

3.1 Sample Preparation

Every metabolomics experiment starts with a biological sample. Regardless of whether the sample is a tissue or a biofluid, it is important to remember that the sample is "alive." Unless the sample is frozen, dried, sterilized, or otherwise devoid of cells or enzymes, the sample is metabolically active. If not handled properly, this underlying metabolic activity can lead to deceptive results with large variations in metabolite composition and concentrations. Therefore, all metabolomics studies must include a metabolic quenching step. Metabolomic quenching uses either physical or chemical means to arrest all metabolic activities (Teng et al. 2009). The easiest method to quench a tissue sample is rapid freezing (using liquid nitrogen). For blood, the red and white blood cells should be first removed before freezing to prevent cell lysis upon thawing. For most other biofluids, moderate to rapid freezing is generally sufficient. In the frozen state, the biological sample may be stored for months or even years at −80°C (Sellick et al. 2009; Vuckovic 2012).

Prior to analysis, the sample must be thawed carefully and further extracted or purified before NMR analysis. Typically, different extraction methods or purification protocols are required for different samples. Tissue samples are often ground into a powder in a still-frozen state using a pestle and mortar and then the metabolites

are extracted with cold methanol or cold methanol/water or with chloroform (Wolak et al. 2012; Nagana Gowda and Raftery 2014). The use of an organic solvent, in addition to extracting the metabolites of interest, quenches unwanted metabolism by denaturing and precipitating almost all proteins/enzymes within the sample. The organic solvent extract is then centrifuged to separate the precipitated proteins from the sample. The sample then must be dried (by freeze-drying/lyophilization) and then re-dissolved in water or appropriate NMR solvents prior to NMR analysis.

For biofluids, the extraction protocols are much simpler. Cell-free biofluids such as serum, plasma, saliva, growth media, cell extracts, plant sap, or fecal water typically require ultrafiltration through a 3–5 kDa molecular weight cut-off filter that removes higher molecular weight proteins and enzymes (Psychogios et al. 2011; Nagana Gowda and Raftery 2014). Since no organic solvent is used, no lyophilization or evaporation is required, and volatile metabolites, such as formate, acetate, methanol and ethanol, can be easily detected by NMR. However, ultrafiltration can cause chemical contamination as the filters contain glycerol or other humectants as preservatives. These filters must be washed multiple times to remove any traces of these agents. Ultrafiltration may also artificially lower the concentrations of some metabolites, such as benzoic acid or tryptophan, which appear to bind to the filter or are bound to proteins that are removed by the filter (Psychogios et al. 2011; Nagana Gowda and Raftery 2014). Urine and cerebrospinal fluid samples (as well as other fluids, such as juice, wine, or beer, that are largely sterile and protein-free) are often simply filter-sterilized by passing the fluid through a 0.22 μm filter to remove any cells or organic debris. Often researchers also add a small amount of sodium azide (an NMR-invisible salt) to the sample to destroy any residual microbial activity.

An alternative method to ultrafiltration or filter-sterilization is organic solvent extraction. In general, methanol or acetonitrile (for hydrophilic compounds) and chloroform or methyl-tert-butyl ether (for hydrophobic compounds) can be added to a sample and used to simultaneously extract metabolites, sterilize the sample, and precipitate proteins. This method may be applied to serum, plasma, or urine (Beckonert et al. 2007; Nagana Gowda and Raftery 2014). Solvent extraction may also enrich for a particular class of chemical compounds (hydrophobic or hydrophilic compounds) depending on the choice of solvent and/or extraction protocol. As a general rule, methanol extraction works best for most NMR samples (Lin et al. 2007). However, as organic solvent extraction involves an evaporation step to remove the organic solvent (which is time consuming), volatile metabolites such as formic acid, ethanol, or acetic acid may be volatilized and no longer detectable.

Another method to enrich or concentrate particular classes of metabolites uses stable isotope chemical derivatization. Although more commonly used in MS-based metabolomics (Gowda et al. 2010), this technique is less widely utilized in NMR-based metabolomics. Stable isotope labeling can enhance the sensitivity and resolution by enabling heteronuclear NMR spectroscopy. It can also help to enrich certain classes of metabolites that contain a specific reactive chemical group (i.e., amines or carboxylate groups). Two isotopic tags have gained some popularity in the NMR community, ^{15}N-ethanolamine and ^{15}N-cholamine (see Fig. 2). Both react selectively with carboxyl groups (Ye et al. 2009; Tayyari et al. 2013). These isotope

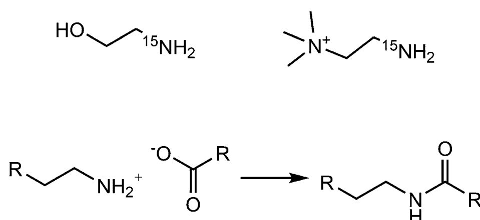


Fig. 2 The chemical structures of ^{15}N-labeled ethanolamine and cholamine (top). The amine group can react with carboxylic acid groups on metabolites, using DMT-MM (4-(4,6-dimethoxy[1,3,5] triazin-2-yl)-4-methylmorpholinium-chloride) as a catalyst (bottom). 2D ^{15}N-^{1}H HSQC spectra can then be acquired to detect the tagged metabolites

tags can be used to enhance the detection of organic acids and amino acids and can also be used in MS-based metabolomics studies.

After the extraction and/or enrichment step is complete, it is critical to use the right buffer system to adjust the salt concentrations and pH of the sample. For untargeted approaches, sample uniformity is absolutely required. Ideally all samples should have identical pH values and identical salt concentrations to ensure uniform chemical shifts among all metabolites. This uniformity greatly enhances the spectral alignment. For targeted metabolomics studies, pH and salt concentrations are not as critical. Nonetheless, buffering the sample with a 50–150 mM potassium phosphate buffer, maintaining the sample temperature at a constant value (say 25°C), adding a small amount of D_2O as a lock solvent, and ensuring the pH is near 7.0 are all recommended for targeted profiling with software tools such as Bayesil or Chenomx (to be discussed later) which have well-defined sample collection conditions (Mercier et al. 2011; Ravanbakhsh et al. 2015).

Sample preparation for NMR analysis always requires the addition of a chemical shift reference compound. The International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry and Molecular Biology (IUBMB) recommend the use of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as an internal chemical shift standard for aqueous samples (Wishart et al. 1995; Harris et al. 2002). Usually DSS is used in its deuterated form DSS-d_6. As the DSS peak at 0.00 ppm is usually well resolved and easily detected, it may also be used as an internal quantification reference. Trimethylsilylpropanoic acid (TSP) is also commonly used as a chemical shift standard but is not recommended for NMR-based metabolomics applications as the chemical shift varies substantially with pH. Both DSS and TSP can bind to macromolecules such as lipids or proteins (found in unfiltered serum or plasma), resulting in a broadened signal if these molecules are present in large quantities. If this occurs, the signal cannot be used for quantification. An alternative to the internal standard is an electronic reference signal, often called ERETIC (Electronic REference To access In vivo Concentrations) (Akoka et al. 1999). The ERETIC method is particularly appealing since the signal is electronically generated and can be placed in any position in the

spectrum. It can also be used to determine absolute concentrations (Watanabe et al. 2016).

After the sample preparation step is complete, the sample must be transferred to an NMR sample tube. Commonly, 5 mm borosilicate glass tubes (requiring volumes of 500 to 600 μL) are used for NMR experiments. More recently, many labs have converted to using 3 mm tubes as less volume (150 to 200 μL) is required while yielding almost the same signal-to-noise (S/N) ratio as 5 mm tubes. These narrower tubes also reduce the dielectric loss due to high salt concentrations, making them more amenable for use with cryo-probes. Shigemi tubes with susceptibility matching glass plugs can also be used to reduce the sample volume (~250 μL) while retaining the 5 mm tube diameter that is optimal for 5 mm NMR probes and 5 mm spinners. Microprobes are also available for volume-limited samples. These microprobes are optimized for use with 1.0 or 1.7 mm NMR tubes. However, 1.0 and 1.7 mm tubes are very delicate and difficult to clean (the 1.0 mm tubes are actually disposable since they are essentially impossible to clean). Nevertheless, with these tubes it is possible to work with volumes as low as 10 μL for a 1.0 mm tube and 35 μL when using a 1.7 mm tube.

Given the variety of ways biological samples can be collected, prepared, and processed prior to NMR-based metabolomics analysis, it has long been recognized that standardized protocols are needed to facilitate proper comparison of data between samples and between studies (Beckonert et al. 2007; Bernini et al. 2011; Emwas et al. 2015). As outlined in these papers, there are a relatively small number of preferred or optimal methods for sample collection, storage, and preparation. Some of these methods are quite specific to certain biosamples or biofluids, while others are nearly universal. Regardless of the methods used, it is vital that detailed, complete, and appropriate information about study design, sample types, sample collection methods, sample handling, sample processing, and sample storage be provided. Without this information, it can be very difficult for others to reproduce or interpret reported results. These factors are especially important if studies involve multiple locations and multiple laboratories. Toward this end, a number of initiatives have been launched to improve and standardize sample collection, analysis, and reporting. These include the Metabolomics Standards Initiative (Fiehn et al. 2007) along with recent updates (Spicer et al. 2017), COSMOS (Salek et al. 2015), and more recently SPIDIA and SPIDIA4P (Ghini et al. 2019). Such initiatives should help guide and encourage members of the metabolomics community to employ standard protocols and fully report standard procedures to acquire, store, prepare, process, and report metabolomics data. This kind of standardization helps ensure that NMR-based metabolomics experiments can be repeated, the collected data can be re-analyzed, and comparisons can be consistently made between samples, studies, and laboratories.

3.2 Choosing the Right Pulse Sequence

Once the sample is loaded into the NMR spectrometer, the NMR spectrum (or spectra) can be acquired. Most NMR-based metabolomics studies use 1D ^{1}H NMR because of its speed and simplicity with which spectra can be acquired, processed, and interpreted. Two types of pulse sequences or experiments are used for 1D NMR-based metabolomics: the *metnoesy* or 1D NOESY (Nuclear Overhauser Effect Spectroscopy) experiment and the Carr-Purcell-Meiboom-Gill (CPMG) experiment. The metnoesy experiment is a simple 1D NOESY pulse sequence that provides solvent suppression before the experiment and during the mixing time without the use of gradients (see Fig. 3). This simple pulse sequence can be used with almost any NMR probe (Mckay 2011). The metnoesy pulse sequence can be modified with more advanced water suppression techniques that use gradients (i.e., watergate, excitation sculpting) for more robust and effective solvent removal. With such modifications, more dilute samples can be recorded in less time (McKay 2009). Another advantage of the metnoesy sequence is that shaped pulses can be used to suppress not only water but additional other strong signals such as ethanol in wine and beer samples or organic solvents used in the extraction process. As most databases and deconvolution programs for NMR-based metabolomics were built using this simplistic metnoesy pulse sequence, this pulse sequence predominates most metabolomics studies with 1D ^{1}H-NMR.

The second most commonly used 1D NMR pulse sequence is the CPMG experiment (Fig. 4). This pulse sequence can spectroscopically remove signals of large molecules, such as proteins or lipoproteins, from the spectrum without the need for ultrafiltration or solvent extraction (Beckonert et al. 2007). The CPMG experiment takes advantage of the fact that small molecule metabolites and macromolecules (such as proteins) have different T_2 relaxation times. The T_2 relaxation time of macromolecules is very short (milliseconds) while the T_2 of metabolites is longer (seconds). With the CPMG sequence, all molecules with a short T_2 are suppressed, whereas those with a large T_2 are unaffected. Under ideal conditions, the CPMG pulse sequence would eliminate the need for solvent extraction and ultrafiltration, reducing time and resources required for sample preparation. However, the CPMG

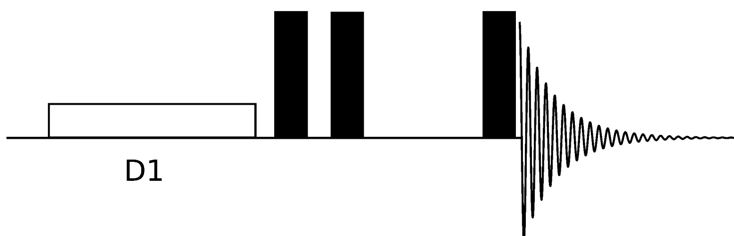


Fig. 3 The pulse sequence for the *metnoesy* or 1D NOESY experiment. This pulse sequence has a delay (D1) with a low power presaturation pulse, followed by two successive 90° pulses followed by a defined mixing time. The pulse sequence concludes with a final 90° pulse followed by the acquisition period

Fig. 4 The pulse sequence for the Carr-Purcell-Meiboom-Gill (CPMG) experiment. This experiment uses repeated pulses of 180° to select signals with long T_2 relaxation times and remove signals with short T_2 times

experiment is not perfect and all protein signals are not suppressed, increasing the time required to properly phase, process, and compare multiple CPMG spectra. As very few CPMG reference spectra are available in the NMR databases, it makes the CMPG pulse sequence very difficult to use for spectral deconvolution in targeted NMR-based metabolomics studies.

More recent developments in NMR-based metabolomics use "pure-shift" pulse sequences (Moutzouri et al. 2017; Lopez et al. 2019) to simplify overcrowded 1D ^{1}H-NMR spectra. Pure-shift NMR aims to convert all signals into singlets by refocusing homonuclear couplings and collapsing multiplet peaks into singlets, thereby reducing the overlap between compound peaks. However, this class of experiments are much less sensitive than standard ^{1}H experiments due to the use of spatial or frequency selective techniques. They also contain artifacts that result from the pseudo-2D method of data acquisition. The reduced sensitivity, the increased complexity, and the greater hardware requirements over more conventional 1D NMR experiments have likely prevented their widespread use in metabolomics. Despite these drawbacks, the improved resolution provided by these methods has seen their successful application in metabolite profiling (Lopez et al. 2019).

1D NMR pulse sequences are not the only experiments available to metabolomics researchers. With 2D NMR, multidimensional data can be recorded from the same type of nuclei (homonuclear) or different nuclei (heteronuclear). For metabolically complex samples such as urine with hundreds of different and variable metabolites, it can be advantageous to use 2D NMR experiments. Indeed homonuclear 2D experiments, such as 2D ^{1}H COSY (COrrelated SpectroscopY), 2D ^{1}H INADEQUATE (Incredible Natural Abundance DoublE QUAntum Transfer Experiment), or 2D J-resolved experiments have been widely used in NMR-based metabolomics studies (Martineau et al. 2011; Bingol and Brüschweiler 2014). 2D NMR spectra allow researchers to more easily identify unknown compounds, characterize novel compounds, and deconvolute overlapping peaks, which could be problematic for 1D spectral deconvolution programs. Several tools and databases are available to interpret 2D homonuclear metabolomics data (Bingol et al. 2014, 2016). For complex mixtures such as those found in metabolomics studies, the 2D J-resolved (JRES) experiment is particularly attractive due to its simplicity and relatively short

acquisition time relative to other 2D experiments (Ludwig and Viant 2010). The JRES experiment results in singlets in the ^{1}H dimension of the spectrum, effectively providing a robust, broadband decoupled ^{1}H spectrum, similar to the “pure-shift” experiments mentioned above. This provides a way to resolve the complex overlapping signals from a 1D ^{1}H spectrum. The multiplet patterns, however, are retained in the second dimension of the JRES spectrum, which can further aid in the identification of the metabolites. The JRES experiment can also be used for metabolite quantification. Further enhancement of metabolite identification can be achieved using heteronuclear 2D experiments, such as 2D ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Coherence Spectroscopy) and ^{1}H-^{13}C HSQC experiments. These experiments provide additional chemical shift information in the second dimension (^{15}N/^{13}C) and offer important structural detail and connectivity information (Lewis et al. 2007; Ye et al. 2009). Similar to the homonuclear 2D metabolomics data, spectral databases and programs are also available to facilitate the interpretation of 2D heteronuclear metabolomics data (Bingol et al. 2015, 2016).

While 2D NMR experiments offer tremendous advantages, there are at least three major disadvantages to using 2D or multidimensional NMR for metabolomics. The first disadvantage is the time required to collect, process, and interpret the data (hours compared to minutes for the 1D experiments). Second, 2D NMR experiments are less sensitive, with a lower limit of metabolite detection that is 5-10X higher than 1D NMR (often >50 μM compared to 10 μM). And lastly, obtaining robust absolute quantification of metabolites from 2D spectra is particularly challenging – although improvements are being made (Martineau and Giraudeau 2019; Martineau et al. 2020; Hansen et al. 2021).

Since the informational density in 2D NMR experiments is higher than 1D experiments, there is a considerable interest in developing techniques to reduce the time required to acquire 2D spectra. A number of different approaches are being investigated to reduce 2D spectral scanning time (Le Guennec et al. 2014). The first approach utilizes spectral folding or aliasing to reduce the spectral width. With this technique, peaks can be folded in empty spaces of the 2D spectrum. The reduction of the spectral width means less data points are required, shortening the overall acquisition time. However, additional NMR experiments have to be performed to determine the true chemical shift of folded peaks (Foroozandeh and Jeannerat 2010). If the sample pool is the same (such as urine), this only needs to be performed once with a single representative sample.

A second approach shortens the acquisition time by reducing the delay between scans or shortening the relaxation time. These experiments are known as band-selective optimized flip angle short transient (SOFAST) and band-selective excitation short transient (BEST) methods. However, as these experiments rely on spin diffusion as an effective relaxation mechanism (which is only true for macromolecules or small molecules in viscous solvents), SOFAST and BEST experiments can only be used in studies involving living cells, which have a viscous cytoplasm (Motta et al. 2010).

A third approach is available that dramatically shortens 2D NMR data acquisition time by combining multiple pulse sequences (such as COSY, NOESY, HSQC) into

one supersequence (Kupče and Claridge 2017; Hansen et al. 2021). An example of such a supersequence is called NOAH (NMR by Ordered Acquisition using ^{1}H detection). This pulse sequence has enabled 2D data collection in a single measurement and has been used to comprehensively characterize multiple metabolites within metabolically complex biofluids such as mouse urine.

A fourth method to shorten the experimental time of 2D NMR experiments uses non-uniform sampling (NUS). With this method, all data points in the indirect dimension are not recorded. Instead, the recorded data points are randomized and differentially weighted across the indirect dimension(s). The missing data points are reconstructed after the data is collected. With the NUS technique, the acquisition time of 2D and 3D experiments can be reduced by up to 75% while still retaining the same spectral resolution as a full-time multidimensional experiment. The quality of the spectrum mainly depends on the algorithm used to reconstruct it (Kazimierczuk et al. 2010).

The last technique is the ultrafast (UF) 2D NMR. This technique utilizes the sample height (or length) and generates slices which correlate with different time points in the second dimension (called spatial encoding). With the UF-NMR technique, a 2D spectrum can be acquired in a single scan, making it the fastest 2D NMR technique available. However, using this experiment, a compromise between spectral widths, resolution, and sensitivity is often necessary. Fortunately, the spectral width limitations of UF-NMR can be addressed by the folding/aliasing method (mentioned above) and the sensitivity can be improved by increasing the number of scans (Shrot and Frydman 2009; Tal and Frydman 2010; Pathan et al. 2011).

Compared to 1D NMR, 2D NMR techniques have a number of limitations with respect to metabolite quantification. Because 2D NMR techniques use more pulses, they are more sensitive to pulse imperfections. This may result in inconsistencies in peak intensity and peak volume, limiting the reliability of quantification. Furthermore, for heteronuclear NMR, the ^{13}C or ^{15}N nuclei for most metabolites have a very broad chemical shift range. An equal excitation of the ^{13}C or ^{15}N complete spectral region is difficult for higher field (>500 MHz) spectrometers. To achieve equal excitation, one could use complex pulses or pulse sequences, such as adiabatic pulses or shaped pulses. However, the use of these pulses leads to inconsistencies in spectral peak intensities and volumes, which makes quantification challenging. In addition, highly variable coupling and relaxation times also alter the peak volume in hard-to-predict ways, thereby limiting the use of 2D techniques for quantitative analysis. To overcome these limitations, large numbers of 2D spectral calibration curves must be collected for each type of 2D NMR experiment in order to use them for accurate metabolite quantification. However, gathering this kind of data is tedious, time-consuming, and difficult due to the long experimental acquisition times (Lewis et al. 2007).

In recent years, it has been shown that ^{13}C-HSQC data can be recorded in a way that is inherently quantitative. This can be done by extrapolation of the signal back to the initial excitation (time point zero), known as the $HSQC_0$ experiment (which requires three separate HSQC experiments) or with a quantitative sequence or the Q-HSQC experiment (which requires four times more scans to achieve the same

sensitivity as normal HSQC) or the quicker variant of the Q-HSQC experiment called the QQ-HSQC experiment (Peterson and Loening 2007; Hu et al. 2011; Martineau et al. 2013; Sette et al. 2013). These developments are encouraging and given the rapid progress in recent years, 2D homonuclear and heteronuclear NMR may become more appealing and more widely used in the near future.

3.3 Spectral Acquisition and Processing

As described above, the selection of the appropriate NMR experiment is obviously very important for obtaining high-quality NMR data. However, other aspects of spectral acquisition and data processing also have a significant impact on spectral quality. These aspects include the ability to obtain sharp, well-shimmed peaks; the level of digital resolution; the S/N ratio or signal quality; the presence of well-phased signals; and the ability to obtain flat baselines.

To obtain good quality NMR data, both the instrument and the sample must be appropriately locked, tuned, matched, and carefully shimmed. Most instruments are tuned for salt-free solvents (such as deuterochloroform or pure D_2O), but metabolomic samples usually contain relatively high salt concentrations (50–200 mM NaCl). If the spectrometer probe is improperly tuned or matched, these high salts could lead to poor performance, noticeable by a low (S/N) ratio, long excitation pulses, poor solvent suppression, etc. These effects become more significant with cryogenically cooled probes compared to room-temperature probes. Thus, with each NMR experiment, due care and effort must be applied to ensure that the sample is well locked, and optimal tuning and matching have been achieved. Often tuning and matching are performed manually. However, newer instruments support rapid automatic tuning and matching. This improved auto-tune/auto-match makes it possible to run multiple sample types with different solvents or salt concentrations in the same run.

Another prerequisite for obtaining good quality NMR spectra is good shimming. During the shimming process, small electromagnets (so-called shims) are adjusted to compensate for magnetic inhomogeneity in the superconducting magnet field. Inhomogeneities in the magnetic field can lead to distorted peaks in the spectrum (see Fig. 5). Most modern NMR spectrometers support automated shimming and can reliably adjust the magnet shims to achieve excellent line shapes. For both manual and automated shimming, the signal of the internal chemical shift standard (such as DSS and/or the solvent signal) is used to optimize the shims. After the shimming procedure, the line width of the reference standard (DSS) should be well below 1 Hz (0.5 Hz is typical). Adjustment of the field and the lock position during the shimming process can also improve peak shapes and the overall quality of the spectrum.

To obtain high-quality NMR spectra, good digital resolution and a large number of data points are required. Modern NMR spectrometers are equipped with high-speed, high memory computers that support rapid and high bandwidth analog to digital conversion (ADC). With digital oversampling techniques, 64,000 data points or even 128,000 data points (np) can be collected on modern high field NMR

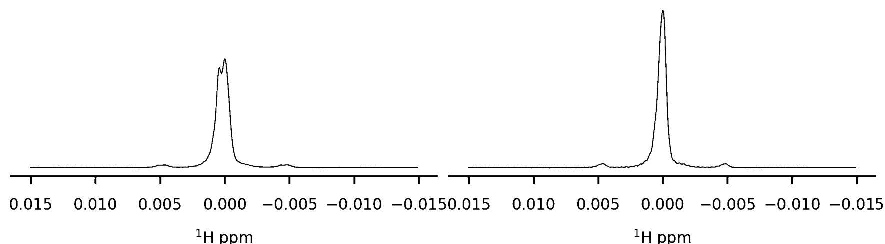


Fig. 5 A simple illustration of the effects of shimming on an NMR signal on the DSS reference peak. Poor shimming can cause peaks to appear unsymmetric or even appear like multiple peaks

spectrometers. This ensures that the digital resolution will typically be less than 0.25 Hz in the ^{1}H dimension on most high-resolution instruments (see Eq. 1; sweep widths (sw) range from 6,000 to 10,000 Hz):

$$\text{Resolution} = 2 \times \frac{\text{sw}}{np} \tag{1}$$

Good digital resolution ensures sharp resonances, but good S/N ensures good sensitivity. NMR is not known to be a particular sensitive technique. Thus, metabolomics researchers are always looking for ways to improve NMR sensitivity to extend the lowest detectable metabolite concentration. Increased sensitivity can be achieved by increasing the magnetic field strength (the higher the better), using cryogenically cooled probes (which have 2–4 times better S/N compared to room-temperature probes via the reduction of electronic noise), concentrating the sample, increasing sample volume, or optimizing the excitation flip angle. However, one of the simplest approaches to increasing the S/N and lowering the limit of detection is increasing the number of scans (ns) (as shown in Eq. 2):

$$\frac{S}{N} = \sqrt{ns} \tag{2}$$

Using Eq. 2, one can see that increasing the scans does not increase the S/N linearly. With four scans, the sensitivity (S/N) increases just twofold. With 100 scans, the sensitivity increases only 10-fold. The number of scans collected in a given NMR experiment must be tempered by the time it takes to collect each of those scans. That time is determined by the repetition period between scans, also known as relaxation delay. This delay is defined as the sum of the acquisition time and acquisition delay prior to the next scan. The relaxation delay should be five times longer than the longitudinal relaxation time (also known as T_1). For metabolites, T_1 is typically 2–3 s (Bloembergen et al. 1948). Therefore, a relaxation delay of about 15 s is usually enough time for a complete relaxation of all resonances (and consequently full recovery of signal intensity) between scans. However, employing such long delays would make data collection incredibly inefficient and severely limit the number of scans that could be collected. Therefore, shorter repetition times of

2–4 s are often used for the majority of NMR-based metabolomics studies (Beckonert et al. 2007). These shorter relaxation delays represent a reasonable compromise between trying to maximize signal recovery and maximizing the number of scans. However, this compromise means that spectral deconvolution algorithms used in targeted metabolomics cannot use "idealized" or theoretical reference NMR spectra but must, instead, use reference spectra that have been experimentally recorded using exactly the same type of acquisition parameters and short relaxation delays as used in the actual metabolomics experiment (Mercier et al. 2011; Worley and Powers 2014; Ravanbakhsh et al. 2015).

After an NMR spectrum has been collected, it must be properly phased. Phasing is an NMR spectral adjustment process that is designed to maximize the absorptive character of NMR peaks over all regions of an NMR spectrum. There are two types of phasing: zero-order phase correction (frequency-independent) and first-order (frequency-dependent) phase correction. While zero-order phase correction is independent of the position of the peaks, first-order phasing increases linearly with the offset from the carrier frequency. Both types of phase correction are usually needed to obtain symmetric, purely absorptive peaks. Accurate phase correction is an important step in spectral processing of metabolomics data as even small phase errors can lead to a cascade of problems in downstream spectral processing and post-spectral analysis affecting targeted or untargeted metabolomics techniques (Emwas et al. 2018). Several algorithms have been published for automated phasing of NMR spectra (Chen et al. 2002; de Brouwer 2009; Binczyk et al. 2015; Zorin et al. 2017; Steimers et al. 2020). Some of these routines have already been implemented in the operating systems of many modern NMR spectrometers. However, additional manual phasing is often required in NMR-based metabolomics studies since auto-phasing routines may have difficulty with more spectrally crowded metabolite spectra. Furthermore, auto-phasing programs can sometimes end up distorting the entire NMR spectrum while attempting to correct for the residual water signal. Despite these caveats, auto-phasing is still widely used in the NMR metabolomics community because it is fast, reasonably reliable and it avoids operator bias.

Baseline correction is another important step in NMR spectral data processing. Baseline correction yields a more pleasant looking NMR spectrum where signal-free regions appear as completely flat lines with zero intensity. While baseline correction is relatively easy for simpler NMR spectra with just a few peaks, it is much more difficult for NMR spectra containing thousands of peaks with large differences in peak intensities and peak widths. High-quality baseline correction is critical for proper spectral alignment (in untargeted metabolomics) and proper quantification or peak integration (in targeted metabolomics). Like phase correction, small errors in the baseline correction can lead to significant errors in the quantification of low abundance metabolites.

There are two general approaches to baseline correction: one involves correction in the time domain and the other involves correction in the frequency domain. Baseline correction through the time domain removes corrupted data in the free induction decay (FID) to decrease the effect of low frequencies. This can be done by discarding some of the initial data points and recreating them by a technique called

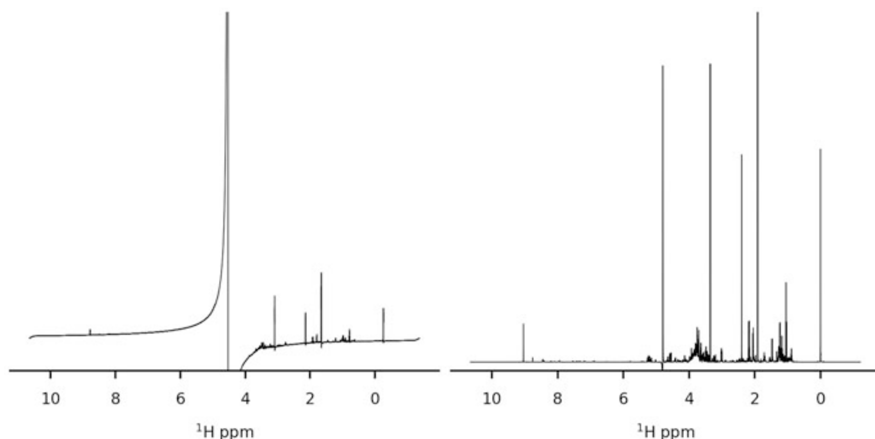


Fig. 6 An example of how good shimming, water suppression, phasing, baseline correction, and chemical shift referencing can make a significant difference to the quality and usability of a 1D NMR spectrum of a biofluid. The same sample and experimental parameters were used for both spectra, except shimming and water suppression were adjusted away from optimal for the left spectrum. The receiver gain was adjusted automatically

back-prediction (Heuer and Haeberlen 1989; Halamek et al. 1994). Baseline correction in the frequency domain involves selecting the valley or signal-absent regions of the spectra (either automatically or manually) and fitting these regions with a polynomial spline function. The corresponding baseline offset values based on this spline function are then subtracted from the spectrum to yield a corrected baseline (Golotvin and Williams 2000; Xi and Rocke 2008). Combining both frequency and time domain methods further improves the quality of the baseline. Baseline correction routines are available in the operating systems of most modern NMR instruments, although often the best baseline correction routines for complex spectra are found in spectral deconvolution tools designed specifically for NMR-based metabolomics (Weljie et al. 2006; Mercier et al. 2011; Worley and Powers 2014; Ravanbakhsh et al. 2015).

The final step of NMR data processing is spectra alignment. In this step, the ppm scale of all the experiments is calibrated by adjusting the position of the internal standard (DSS) to 0 ppm. Then, the intensities of all the peaks are normalized using the height of the internal standard peak. An example of a 1D NMR spectrum before and after proper chemical shift referencing, shimming, phasing, solvent removal, and baseline correction is shown in Fig. 6.

4 Data Analysis

After collecting a set of NMR spectra for a metabolomics study, the next step is data analysis. The type of data analysis one undertakes depends on the experimental design, the choice of the metabolomics experiment (targeted vs. untargeted), and the type of multivariate statistical techniques or software available to the researcher.

Many excellent reviews have been written on experimental design and multivariate statistics for metabolomics and readers who are interested in these topics should refer to these publications (Madsen et al. 2010; Smolinska et al. 2012; Saccenti et al. 2014; Ebbels et al. 2019; Percival et al. 2020). We would encourage new users or readers to explore a software tool called MetaboAnalyst to better understand and visualize multivariate statistics as used in metabolomics (Xia et al. 2009). MetaboAnalyst is an easy-to-use, web-based tool that is routinely used by approximately half of the global metabolomics community. It supports a wide range of downstream data analysis applications with extensive graphics support including principal component analysis (PCA), analysis of variance (ANOVA), partial least-squares discriminant analysis (PLS-DA), heat mapping and clustering, biomarker identification, pathway analysis, power analysis, and time series analysis (Verpoorte, "Natural products drug discovery: on silica or in-silico?"; Millet, "Prospective metabolomic studies in precision medicine. The AKRIBEA project"). A screenshot of the MetaboAnalyst package is shown in Fig. 7. MetaboAnalyst is configured to work with both targeted and untargeted metabolomics data, as well as with NMR and MS data. It is also quite unique in that it has a number of freely available NMR metabolomic data sets that users can test or explore.

4.1 Data Analysis for Targeted Metabolomics

For targeted metabolomics studies, the NMR spectra need to be analyzed using specialized peak fitting software. This software fits a reference set of NMR spectra obtained from pure compounds to the NMR spectra of the mixture of compounds found in the biofluid or extract. By matching the positions and intensities (or area) of peak clusters, the software identifies each reference spectrum and accurately determines their concentrations in the biofluid. This approach is often called spectral deconvolution. An illustration of how spectral deconvolution works is shown in Fig. 8. Once the compounds are identified and their concentrations have been precisely quantified, this information can then be compared against known or normal concentrations for that biofluid or further analyzed to detect significant differences between one sample (or group of samples) and the next.

The concepts underlying targeted NMR metabolomics emerged independently on three separate occasions – for three different applications. Targeted metabolomics was first described in 1991, as a novel approach to perform serum lipid and lipoprotein particle profiling (Otvos et al. 1991). The protocol uses a reference spectral library of different lipoprotein subclasses, to identify and quantify 15 different subcategories of VLDL (very low density lipoprotein), LDL (low density

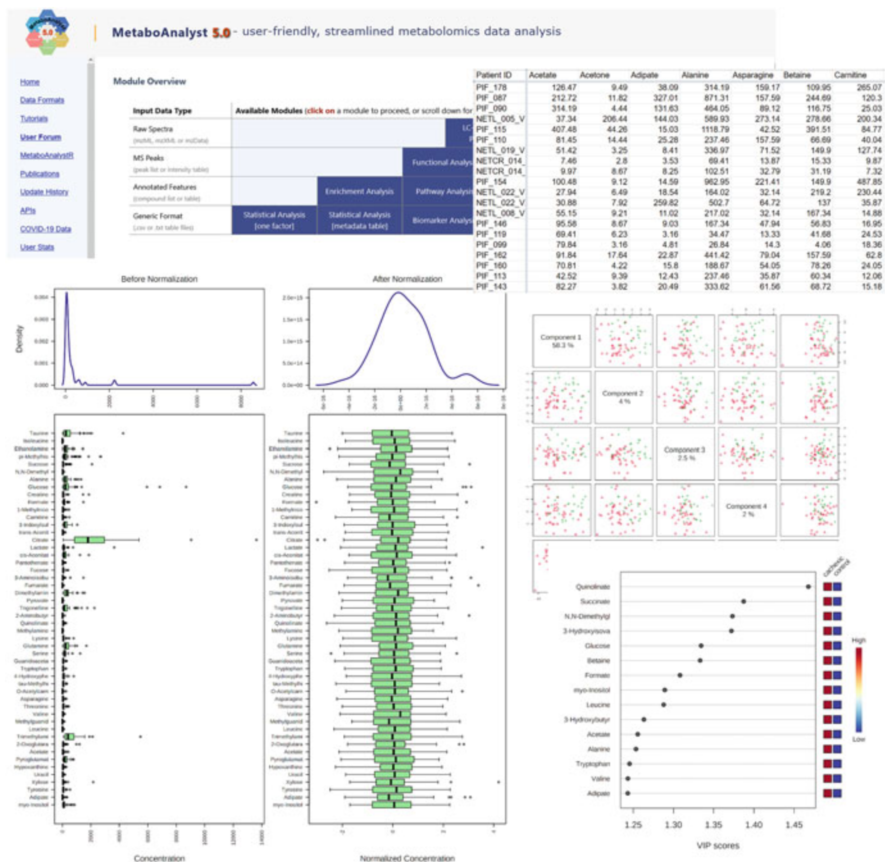


Fig. 7 A screenshot montage of the MetaboAnalyst website. More than a dozen statistical modules are available for analyzing metabolomics data

lipoprotein), and HDL (high density lipoprotein) particles (Jeyarajah et al. 2006). Several variations of the method have been described in the literature, with some methods using only 11 lipoprotein categories or others using more sophisticated wavelet deconvolution algorithms (Serrai et al. 1998; Ala-Korpela et al. 2007). Nevertheless, the method has proven to be particularly simple, fast, and robust compared to traditional methods of lipoprotein profiling which are less accurate, provide less information, and require time-consuming, manually intensive, multi-step separations (McNamara et al. 2006).

Targeted NMR metabolomics emerged for a second time in 1993, when it was used to identify brain metabolites from localized in vivo NMR spectroscopy (Provencher 1993). However it was not until 2001 that the method, known as LCModel, was made generally available (Provencher 2001). The central concept behind the LCModel approach for metabolite identification is to fit the broad peaks

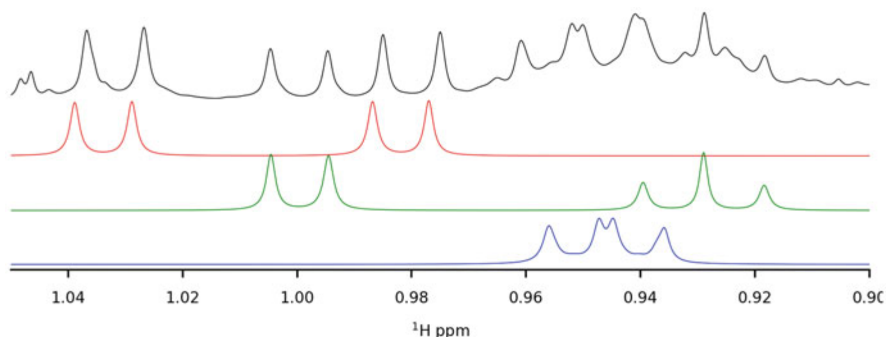


Fig. 8 A schematic illustration of the principles behind spectral deconvolution. The biofluid spectrum at the top contains a combination of the three individual reference spectra below it. The challenge in spectral deconvolution is to determine which combination of reference spectra, their scaling, and positioning, best produces the observed spectrum

obtained from in vivo NMR data to spectral libraries of pure metabolites (called in vitro basis sets) in which the spectra are artificially broadened and otherwise mathematically modified to look like those seen for in vivo NMR spectra. LCModel uses a constrained regularization method to handle differences in phase, baseline, and line shapes between the in vitro and in vivo spectra. It is able to identify between 10 and 15 metabolites and can accurately estimate the metabolite concentrations and their uncertainties (Simister et al. 2003; Marliani et al. 2007).

The third occasion in which targeted metabolomics emerged was in 2001, when it was specifically implemented for small molecule identification and quantification in biofluids and tissue extracts (i.e., traditional metabolomics) (Wishart et al. 2001). While similar in principle to the LCModel and lipoprotein characterization software mentioned above, the actual implementation is somewhat different. This is because spectral deconvolution of high-resolution NMR spectra consisting of dozens of small molecules requires the fitting of not just 15–20 broad peaks, but the fitting of hundreds to thousands of very sharp peaks. Small variations in position, line width, shape, or intensity due to pH or matrix effects can make the spectral fitting problem particularly challenging and "ill-conditioned." Likewise, because of spin-coupling effects, the NMR spectrum of a small molecule collected at 500 MHz often differs substantially from one collected at 800 MHz. To address these problems, thousands of reference NMR spectra must be collected for hundreds of reference compounds at different pH values and at different NMR spectrometer frequencies. Additionally, very sophisticated curve fitting programs that robustly handle sparse matrices and potential singularities must be employed.

Fortunately, a variety of software tools for small molecule NMR spectral deconvolution have been developed over the past 10 years that make this spectral fitting process relatively painless. These include commercial tools, such as the Chenomx NMR Suite (Mercier et al. 2011), Bruker's JuiceScreener (Monakhova et al. 2014), WineScreener (Spraul et al. 2015), FoodScreener (https://www.bruker.

com/en/products-and-solutions/mr/nmr-food-solutions/food-screener.html), and in vitro diagnostic research system (IVDr). Bruker's IVDr system expands the utility of automated metabolite quantification to several biofluids (urine, cerebrospinal fluid, plasma, and serum) by standardizing sample analyses through the incorporation of standard hardware, standard operating procedures, and automated sample handling and analysis (Bruker Corporation-Bruker 2013). The Bruker "X"-Screener and IVDr systems are normally bundled with specific NMR instruments or licensed on a per-sample basis, making them relatively expensive. A cheaper alternative to the commercial deconvolution and automated quantification tools are a number of freely available, non-commercial tools, such as Batman (Hao et al. 2014), Bayesil (Ravanbakhsh et al. 2015), an automated quantification algorithm (AQuA) (Röhnisch et al. 2018), an automatic method for identification and quantification of metabolites (ASICS) (Tardivel et al. 2017) and rDolphin (Cañueto et al. 2018). A screenshot of the Bayesil web server and its standard output is shown in Fig. 9. As can be seen in this figure, most deconvolution programs provide lists of compound identities and estimated concentrations along with an interactive display showing the fit between the observed NMR spectrum and the reference library NMR spectra. Some of these deconvolution tools are almost fully automated (such as Bayesil, Chenomx, and the Bruker products) while others require a fair bit of manual manipulation.

In addition to tools such as Bayesil, Chenomx, and the Bruker "X"-Screener and IVDr products, which identify and quantify organic compounds, there are now NMR-based tools for identifying and quantifying inorganic compounds. In many fields of metabolomics, metal ions are considered to be important metabolites and this subdiscipline of studying metal ions is called "metallomics." Normally metal ions are measured via inductively coupled plasma mass spectrometry (ICP-MS). However, it is also possible to identify and quantify metal ions and other inorganic ions from ^{1}H-NMR spectra. This can be done by taking advantage of the effect of dissolved inorganic ions (invisible by NMR) on organic compound chemical shifts (visible by NMR). This approach was described in detail by Takis et al. (2017). The technique required the measurement of 4,000 variable synthetic urine mixtures containing differing amounts of the most common organic compounds ($> 90\%$ occurrence) and differing amounts of the most abundant inorganic ions in urine and acquiring ^{1}H-NMR spectra of each of these mixtures. From these spectra, Takis et al. created an algorithm, called the Urine Shift Predictor, which is able to predict the concentration-dependent, inorganic ion-induced chemical shifts on different organic compounds. By measuring these chemical shift changes it is possible to estimate inorganic ion and metal ion concentrations from ^{1}H urine spectra.

Regardless of whether one measures organic or inorganic metabolite data (or both), once the list of metabolites and their concentrations has been obtained from a targeted metabolomics study, it is relatively easy to apply standard multivariate statistics such as PCA or PLS-DA (and other tools found in packages such as MetaboAnalyst) to identify significantly changed metabolites or to detect specific sets of metabolites as biomarkers.

Fig. 9 A series of screenshots taken from the Bayesil NMR server. Bayesil is a freely available spectral deconvolution package that can take high-resolution NMR spectra of biofluids, such as serum, and automatically identify and quantify the compounds in that biofluid. The top image (**a**) shows the Bayesil home page. The middle image (**b**) shows the interactive spectral viewer. The bottom image (**c**) shows a selection of the compounds identified from the spectrum

4.2 Data Analysis for Untargeted Metabolomics

Untargeted NMR metabolomics first emerged in the early 1990s when the complexity of high-resolution NMR spectra collected on biofluids seemed to be too daunting to allow routine compound identification (Gartland et al. 1990). Rather than attempting to identify compounds via spectral deconvolution as is done with targeted metabolomics, untargeted metabolomics exploits a field of science known as chemometrics to simplify, classify, and interpret groups of NMR spectra. More specifically, chemometrics is a branch of information science that uses mathematical and statistical methods to identify patterns and extract information from large data

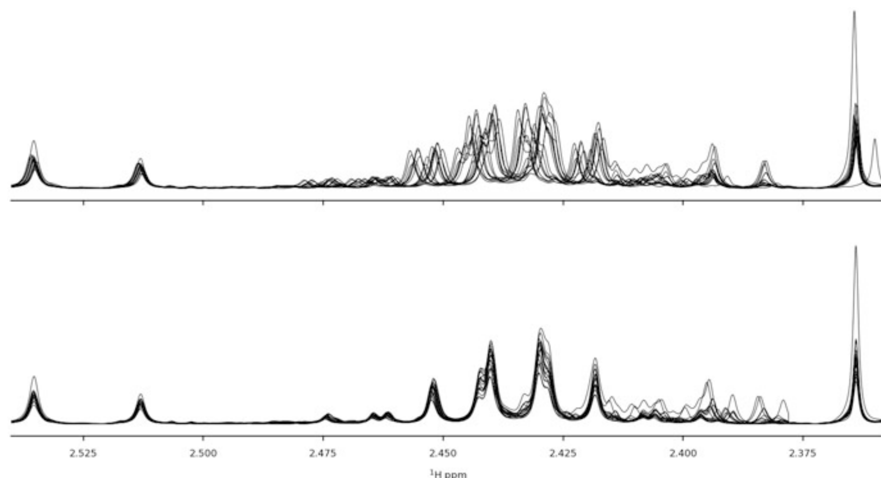


Fig. 10 An example of how spectral alignment of NMR spectra can be performed. Typically, multiple NMR spectra are collected, compared, and then aligned. The top figure shows multiple superimposed 1D 1H NMR spectra of human serum. The glutamine peak at 2.427 ppm shows a larger sensitivity to sample conditions between spectra compared to other nearby peaks. The bottom figure shows the same spectra aligned using a spectral alignment algorithm (Savorani et al. 2010). The glutamine peak is now better aligned between the spectra

sets collected on analytical instruments, such as UV, IR, and NMR spectrometers. When chemometrics is applied to NMR data, it is essential to have many NMR spectra already collected (generally dozens to hundreds of spectra, including both cases and controls). These spectra must then be aligned and binned using specially developed statistical approaches. A simple illustration of how spectral alignment is done is shown in Fig. 10. After the alignment step has been completed, the spectra must be scaled or normalized so that they can be easily compared. Once the scaling and normalization are complete, multivariate statistical techniques such as PCA and PLS-DA can be used to identify interesting spectral regions or clusters of peaks that differentiate one group of spectra from another (Beckonert et al. 2007; Lindon et al. 2007; Barton et al. 2008).

A method called statistical total correlation spectroscopy or STOCSY has also been developed which generates a pseudo-2D NMR spectrum representing the correlation among the peaks in a set of NMR spectra (Cloarec et al. 2005). STOCSY allows for the identification of peaks from the same compound or sets of compounds that co-vary among the spectra. In an untargeted analysis, sometimes only a set of statistically important spectral peaks or features is presented as the final result, without compound identification. In other cases, compound identification occurs only on the peaks which show the most significant changes in a particular study. The final peak identification step may use spectral deconvolution, compound spike-in methods, or peak look-up tables (Martínez-Arranz et al. 2015). A variety of software packages for NMR statistical spectroscopy have been developed over the past

10–15 years to make this kind of analysis more routine. These include MVAPack (Worley and Powers 2014), Automics (Wang et al. 2009), and KIMBLE (Verhoeven et al. 2018).

4.3 Targeted or Untargeted?

For relatively simple biofluids with fewer than 60–70 NMR-detectable compounds (such as serum, plasma, cerebrospinal fluid, fecal water, juice, or other fruit extracts) targeted NMR metabolomics techniques that use spectral deconvolution appear to work very well (Ravanbakhsh et al. 2015). Most of these methods focus on analyzing 1D ^{1}H NMR spectra, although methods have also been developed to analyze 2D ^{1}H NMR spectra and 2D heteronuclear spectra (Lewis et al. 2009; Bingol et al. 2014, 2015). Extensive spectral libraries now exist for essentially all the NMR-detectable compounds found in these biofluids and a number of the deconvolution software tools are becoming almost fully automated. Indeed, some software packages can be extremely fast and robust with >95% of the known compounds in a given biofluid being identified and accurately quantified within a few minutes (Mercier et al. 2011; Zheng et al. 2011; Hao et al. 2014; Ravanbakhsh et al. 2015).

On the other hand, for very complex biofluids, such as cell growth media, cell lysates, and urine, the corresponding NMR spectra are often too complex for any existing spectral deconvolution packages. These biofluids can contain between 70–150 NMR-detectable compounds and often less than 50% of the known compounds can be confidently identified or quantified using spectral deconvolution. Furthermore, the quality of the annotation is often highly dependent on the skill or experience of the operator (Sokolenko et al. 2013). Therefore, for complex biofluids, such as urine, statistical spectroscopy techniques or untargeted NMR approaches appear to offer the best option for spectral interpretation. These approaches allow useful results to be obtained with relatively little manual effort.

5 Biological Interpretation

The ultimate goal of acquiring and analyzing NMR metabolomics data is to use the results to reveal something about the biology of the system under study. This may involve the identification of biomarkers or biomarker panels, the tabulation of reference metabolite concentrations for specific samples or sample types, the characterization of activated or deactivated biological pathways, or the acquisition of insights into the underlying molecular metabolic or biochemical mechanisms associated with a particular biological condition. Rather than describing the software or methods used to perform biological interpretation, here we will focus on some of the more successful applications of NMR-based metabolomics toward biological interpretation.

One of the earliest applications of NMR in the field of metabolomics focused on drug toxicology (Midgley and Hawkins 1978; Cohen et al. 1979; Williams et al. 1979; Bock 1982; Nicholson et al. 1984a, b, 2002; Weiner et al. 1989; Rothman et al. 2003). This work led to the formation of the COnsoritum for MEtabonomic Toxicology (COMET) (Lindon et al. 2005) and became one of the main drivers for the development and growth of metabolomics (and especially NMR-based metabolomics) as a field. COMET helped establish baseline concentrations of metabolites in urine and blood in both human and lab animals and it helped to develop techniques and biomarker panels for high-throughput NMR-based toxicology screening. These pioneering studies demonstrated the potential of NMR and NMR-derived biomarkers to non-invasively identify and diagnose liver, brain, and kidney toxicity arising from specific drugs or drug metabolites.

NMR-based metabolomics has long been used to aid in the diagnosis, interpretation, and monitoring of inborn errors of metabolism (IEMs) (Iles et al. 1984; Yamaguchi et al. 1984; Griffiths and Edwards 1987; Lutz et al. 2013; Kostidis and Mikros 2015; Embade et al. 2019). IEMs are rare genetic disorders characterized by significant changes (several-fold increases or decreases) in the concentration of specific metabolites that arise from genetic disturbances in normal metabolism. According to the Human Metabolome Database (Wishart et al. 2007, 2018), there are more than 400 different metabolites associated with IEMs, of which at least 90 are detectable by NMR (Kostidis and Mikros 2015). While individually rare, as a group, IEMs have been estimated to occur at a rate of up to 1 in 800 births (Mak et al. 2013). Early detection and identification of IEM disorders is critical, as is the need to understand which metabolic pathways are affected. So too is the need to iteratively monitor and adjust treatments throughout a patient's life. Because NMR is so quantitative and reproducible, NMR-based metabolomics has been particularly useful for IEM detection, IEM pathway analysis, and disease monitoring.

More recently, NMR, in combination with mass spectrometry, has been used to determine the baseline concentrations of dozens to hundreds of metabolites in human biofluids including cerebrospinal fluid (Wishart et al. 2008), serum (Psychogios et al. 2011), urine (Bouatra et al. 2013), saliva (Dame et al. 2015), and feces (Karu et al. 2018). NMR has also been used to characterize the blood, urine, milk, and ruminal fluid of dairy cattle as part of a long-term project to characterize the metabolomes of cows (Saleem et al. 2013; Sundekilde et al. 2014; Foroutan et al. 2019, 2020) and other livestock (Chapinal et al. 2012; Hailemariam et al. 2014; Goldansaz et al. 2017). These referential values for each of the major metabolomes are being used by researchers around the world to help interpret metabolite measurements and to identify important biomarkers of diet, health, and disease.

One notable success story for NMR-based metabolomics has been in the area of medical diagnostics. Perhaps the first NMR-based metabolomics-related medical diagnostic test was the NMR LipoProfile test developed by LipoMed (now owned by LabCorp) in the 1990s. The LipoProfile test uses 1D ^{1}H NMR to measure the concentrations of lipoprotein particles in blood plasma samples (Otvos et al. 1991). Each lipoprotein class has a specific chemical shift and line shape, which can be used in a least-squares fit of the plasma samples to calculate the

concentrations of each lipoprotein class. Because of the success of the LipoProfile test, a number of other NMR-based lipid-profiling companies have emerged. These include Biosfer Teslab (https://biosferteslab.com/), Nightingale Health (https://nightingalehealth.com), and numares Health (https://www.numares.com/). Biosfer Teslab uses 2D diffusion ordered spectroscopy (DOSY) NMR to quantify lipoprotein concentrations. In a similar manner, numares Health combines NMR with machine learning to perform lipid-profiling as well as additional NMR-based renal, oncological, and neurological biomarker tests. Nightingale Health is the largest and most successful of these NMR-based metabolomics/diagnostic companies. It offers low-cost, high-throughput NMR-based metabolomics analysis of blood samples directly to customers or to healthcare service providers. Nightingale currently identifies and quantifies over 200 blood biomarkers including small molecule metabolites as well as lipids and lipoprotein particles (Soininen et al. 2015; Würtz et al. 2017). These biomarkers are further combined into "health indicators" that provide an individual with a quick summary of their overall health and their risks for particular diseases such as diabetes, hypertension, obesity, and atherosclerosis.

Another important application of NMR-based metabolomics is the tracking of small molecules as they travel through various metabolic pathways. The ability of NMR to detect isotopically labeled molecules, and to determine the position of these labels at atomic-resolution, makes NMR a particularly valuable tool in the study of metabolic pathways, their regulation, and the effects of disease on their integrity (Lane et al. 2011; Fan and Lane 2011a, b, 2016; Saborano et al. 2019). Isotopic labeling and the quantitative nature of NMR have also led to the development of NMR-based "fluxomics" which focused on determining the rates of metabolic reactions within a biological system. The unique needs of fluxomics have also spurred the development of rapid and quantitative 2D NMR experiments (Massou et al. 2007).

NMR-based metabolomics has been widely used in the study of both cancer and neurological diseases. Cancer is now widely regarded as a metabolic disease (Warburg 1956; Hanahan and Weinberg 2000; Seyfried and Shelton 2010; Wishart 2015). As a result, NMR-based metabolomics has become the technique of choice in many cancer studies (Bathe et al. 2011; Carrola et al. 2011; Namer et al. 2011; Teahan et al. 2011; Weljie et al. 2011; Cao et al. 2012; Farshidfar et al. 2012; Eisner et al. 2013; Wishart 2015; Kim et al. 2019). Other NMR-based studies have been used to classify tumors or to follow the efficacy of radiation or chemotherapy treatment (Blankenberg et al. 1997; Chan et al. 2009; Fong et al. 2011; Palmnas and Vogel 2013). NMR has also been used to identify potential biomarkers for early stages of Alzheimer's disease (Kork et al. 2012; Karamanos et al. 2015), amyotrophic lateral sclerosis (Blasco et al. 2010), or Parkinson's disease (Wu et al. 2016), or for early detection of schizophrenia (Kaddurah-Daouk 2006; Tasic et al. 2017). Non-invasive techniques such as NMR offer the promise of detecting and halting the progression of these diseases at an early stage.

NMR-based metabolomics is not limited to biomedical studies or biomedical applications. It has been widely applied to food safety and food origin studies

(Capitani et al. 2017; Sobolev et al. 2019), and to the composition and quality of foods, such as wine, beer, oil, juice, milk, honey, and fruit (Wishart 2008b; Melzer et al. 2013; Kim et al. 2016). NMR-based methods have also been applied to environmental metabolomics. In particular, NMR-based metabolomics has been used to explore how organisms respond to environmental changes, pollution, and climate change (Williams et al. 2009; Simpson and Bearden 2013; Sumner et al. 2015).

6 Conclusion and Future Prospects

The field of metabolomics has been around for 20 years. However, NMR has been a key part of metabolic studies for more than 40 years. The popularity of NMR as a metabolomics platform is largely due to its high instrument stability, reproducibility, simple non-destructive sample preparation, ease of quantification, and its amenability to automation. However, NMR is not without its limitations. Compared to MS, NMR has relatively poor sensitivity, a large instrument footprint, and high up-front and maintenance costs. NMR also lags behind MS in terms of available metabolomics libraries, easy-to-use data processing and analysis tools, and easy-to-use sample preparation kits.

However, progress is being made on almost all the fronts where NMR lags behind MS. To address the cost and space issues of NMR instrumentation, non-superconducting, bench-top NMR instruments with field strengths of up to 60 MHz are coming into routine use (Percival et al. 2019; Izquierdo-Garcia et al. 2020) and ones approaching or exceeding 200 MHz (Blümich and Singh 2018) are being developed. However, these lower field instruments have lower sensitivity and less spectral resolution than higher field instruments. These disadvantages could be compensated by adopting 2D UF-NMR techniques (Giraudeau and Frydman 2014) on bench-top instruments.

At the other extreme, the recent introduction of 1.2 GHz NMR spectrometers promises to greatly improve the sensitivity and resolution of many NMR-based metabolomics studies (Schwalbe 2017; Luchinat et al. 2021). These super-high field instruments will lower the limits of metabolite detection to the high nanomolar range (compared to 5 μM with lower field NMR instruments) and will likely double the number of metabolites detectable in biofluids such as serum and urine. Developments in NMR probe technology are also leading to some exciting improvements in sensitivity and compound structural elucidation. For instance, the recently developed ^{13}C-optimized 1.5-mm cryoprobe and a ^{1}H-^{13}C dual-optimized NMR probe permit 2D ^{1}H-^{13}C HSQC experiments to be collected at natural abundance (Ramaswamy et al. 2013, 2016; Clendinen et al. 2014, 2015).

Another route to improving the sensitivity of NMR is hyperpolarization where sensitivity enhancements of several thousand-fold (allowing detection of metabolites in the nM range) are technically possible. While hyperpolarization has been used for ^{13}C-labeled metabolites (Ardenkjaer-Larsen et al. 2003; Keshari et al. 2010; Lumata et al. 2015; Dey et al. 2020), the need for specialized equipment and sample

preparation have limited its uptake. Another hyperpolarization method called SABRE-SHEATH (Signal Amplification by Reversible Exchange in SHield Enables Alignment Transfer to Heteronuclei) hyperpolarizes ^{15}N spins at room temperature (Truong et al. 2015; Theis et al. 2016). SABRE-SHEATH could revolutionize metabolomics studies if it could be adapted to NMR-based metabolomics.

Advances in NMR software and NMR data libraries are also helping to move NMR-based metabolomics forward. The recent development of nmrML (Schober et al. 2018) as a universal standard for the exchange of NMR spectra could help to address the issue of limited or incompatible NMR data libraries and modest numbers of reference NMR spectra by providing a common, easily readable, and shareable file format. The release of a file format for mass spectrometry in 2011 (Martens et al. 2011) led to an explosion in the number of publicly available MS spectra. The release and adoption of nmrML by the NMR metabolomics community could potentially lead to the same result. Surprisingly, NMR has not had a modern data exchange standard for more than 30 years. Continued development of open-source or open-access software tools for spectral processing and automated or semi-automated NMR spectral convolution such as Bayesil (Ravanbakhsh et al. 2015), BATMAN (Hao et al. 2014), AquA (Röhnisch et al. 2018) and rDolphin (Cañueto et al. 2018) should also make NMR-based metabolomics more efficient, more user-friendly, and more appealing to non-NMR specialists.

Another area where software development is expected to have an impact on NMR-based metabolomics is in the application of machine learning techniques to analyzing and processing NMR spectra. With the availability of more NMR data and greater computing power, there has been significant growth in the applications of machine learning techniques to NMR. Deep-learning techniques, in particular, could greatly increase the sensitivity and accelerate the analysis of NMR-based metabolomics workflows. For example, deep neural networks have been used to denoise NMR spectra using a program called DN-Unet (Wu et al. 2021) and to efficiently reconstruct NUS NMR spectra (Hansen 2019). This has allowed the acquisition of very high S/N spectra in a fraction of the time. Bruker has developed a deep-learning algorithm called *sigreg* that enables signal recognition detection in ^{1}H-NMR spectra across multiple complex data sets (Paruzzo et al. 2020). Different deep-learning approaches have also been used for chemical shift prediction and molecular structure elucidation of small molecules and related metabolites (Cobas 2020). These deep-learning applications typically require massive training data sets to yield useful results. Given the current paucity of large NMR data sets, much work still needs to be done to maximize the benefits offered by machine learning to NMR-based metabolomics research.

One continuing challenge in NMR-based metabolomics lies in the identification of unknown compounds. As a rule, unknown identification for NMR-based metabolomics is not as difficult as it is for MS-based metabolomics. This is because most of the unknowns detected by NMR must have relatively high concentrations (>5 mM) and consequently are more likely to be “known unknowns” rather than “unknown unknowns.” The term “known unknowns” refers to compounds that have been previously characterized or which already exist in chemical databases, but for

which no reference NMR (or MS) spectra exist. On the other hand, "unknown unknowns" are compounds that are completely novel and have never been described before. To identify "known unknowns," a targeted NMR approach is often possible. For instance, one can use standard 2D NMR approaches to determine the unknown's molecular constituents or atomic connectivities which might lead to a possible match to an existing/known structure. Alternately, one can use a "suspect screening" approach where intelligent guesses (guided by chemical shift similarities) and the spiking in of pure standard of the suspected compound into the sample can help confirm its identity.

Obviously, the most appropriate solution to identifying "known unknowns" is adding more high-quality, experimentally collected, fully assigned reference NMR data for many more metabolites to NMR databases such as the HMDB, BMRB, or nmrshiftdb (McAlpine et al. 2019). Unfortunately, the pace at which reference experimental NMR spectra are being added to these databases has slowed considerably. In the absence of reference experimental NMR data, the use of predicted NMR spectra to identify "known unknowns" may be possible. Certainly, continued advancements in computing power along with improvements to the accuracy of quantum mechanical and machine learning based predictions of NMR chemical shifts and NMR coupling constants suggest that computational approaches could be an appealing new route (Borges et al. 2021).

To identify truly novel compounds (i.e., "unknown unknowns"), *de novo* structure determination would be required. De novo structure determination often requires partial or complete purification of the compound of interest, followed by a combination of mass spectrometry and 2D NMR analysis of the purified compound (Garcia-Perez et al. 2020). The process of characterizing novel compounds, however, can be particularly time consuming and also often limited by the low concentrations of many unknown compounds. Computational approaches for the analysis of spectral data and the proposal of candidate structures (through techniques such as computer-aided structure elucidation – also known as CASE) can accelerate the de novo structure determination process (Boiteau et al. 2018; Leggett et al. 2019).

The computational and experimental innovations being applied to the identification of unknowns by NMR is typical of the remarkable inventiveness and impressive creativity often seen in the NMR community. No doubt many other innovations in NMR instrumentation, in NMR spectral collection, or in NMR data analysis that will benefit NMR-based metabolomics are on the horizon. Certainly, for anyone who has been in the NMR field for more than a few years, one quickly learns to never underestimate the potential of NMR spectroscopists to come up with some remarkable, paradigm-changing innovations. It is probably fair to expect that the next 20 years of NMR-based metabolomics will be as interesting and fruitful as the past 20 years.

Acknowledgments The authors wish to thank Genome Alberta (a division of Genome Canada), The Canadian Institutes of Health Research (CIHR), Western Economic Diversification (WED), and Alberta Innovates Health Solutions (AIHS) for financial support.

Glossary

Baseline correction	A spectral processing technique that yields a more pleasant looking NMR spectrum where signal-free regions appear as completely flat lines with zero intensity
BEST	Band-Selective Excitation Transient
Chemometrics	A branch of information science that uses mathematical and statistical methods to identify patterns and extract information from large data sets collected on analytical instruments, such as UV, IR, and NMR spectrometers
CPMG	Carr-Purcell-Meiboom-Gill
CPMG experiment	A pulse that can filter out the signals arising from large molecules, such as proteins or lipoproteins, from the spectrum (without the need for ultrafiltration or solvent extraction)
COSY	COrrelated SpectroscopY
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid, a chemical shift reference compound
ERETIC	Electronic REference To access In vivo Concentrations, an electronic reference signal
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
IEM	Inborn error of metabolism. They are rare genetic disorders characterized by significant changes (several-fold increase or absence) in the concentration of specific metabolites that result from disturbances in normal metabolism
INADEQUATE	Incredible Natural Abundance DoublE QUAntum Transfer Experiment
IUPAC	The International Union of Pure and Applied Chemistry
IUBMB	The International Union of Biochemistry and Molecular Biology
Metabolomics	A branch of analytical chemistry that comprehensively characterizes the molecules in various biofluids and tissues
Metabolites	The chemical constituents of the metabolome
Metabolome	The complete collection of all chemicals or metabolites found within cells, biofluids, organs, or organisms
Metnoesy experiment	A simple 1D NOESY pulse sequence that provides solvent suppression before the experiment and during the mixing time without the use of gradients
NOESY	Nuclear Overhauser Effect Spectroscopy
Phasing	An NMR spectral adjustment process that is designed to maximize the absorptive character of NMR peaks over all regions of an NMR spectrum
Relaxation delay	The sum of the acquisition time and acquisition delay prior to the next scan
SABRE-SHEATH	Signal Amplification by Reversible Exchange in SHield Enables Alignment Transfer to Heteronuclei
SOFAST	Band-Selective Optimized Flip Angle Short Transient
STOCSY	Statistical total correlation spectroscopy
T_1	Longitudinal relaxation time
T_2	Transverse relaxation time
Targeted metabolomics	A metabolomics technique that uses spectral deconvolution software to identify and quantify fluid-specific or targeted metabolites in individual spectra

TSP	Trimethylsilylpropanoic acid, a chemical shift reference compound
Untargeted metabolomics	A metabolomics technique that uses spectral alignment, spectral binning, and multivariate statistical analysis to identify spectral features of interest

References

Akoka S, Barantin L, Trierweiler M (1999) Concentration measurement by proton NMR using the ERETIC method. Anal Chem 71(13):2554–2557. https://doi.org/10.1021/ac981422i

Ala-Korpela M et al (2007) The inherent accuracy of 1H NMR spectroscopy to quantify plasma lipoproteins is subclass dependent. Atherosclerosis 190(2):352–358. https://doi.org/10.1016/j.atherosclerosis.2006.04.020

Ardenkjaer-Larsen JH et al (2003) Increase in signal-to-noise ratio of >10,000 times in liquid-state NMR. Proc Natl Acad Sci 100(18):10158–10163. https://doi.org/10.1073/pnas.1733835100

Bales JR, Higham DP et al (1984a) Use of high-resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine. Clin Chem 30(3):426–432

Bales JR, Sadler PJ et al (1984b) Urinary excretion of acetaminophen and its metabolites as studied by proton NMR spectroscopy. Clin Chem 30(10):1631–1636

Barton RH et al (2008) High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. Int J Epidemiol 37(Suppl 1):i31–i40. https://doi.org/10.1093/ije/dym284

Bassini A, Cameron LC (2014) Sportomics: building a new concept in metabolic studies and exercise science. Biochem Biophys Res Commun 445(4):708–716. https://doi.org/10.1016/j.bbrc.2013.12.137

Bathe OF et al (2011) Feasibility of identifying pancreatic cancer based on serum metabolomics. Cancer Epidemiol Biomarkers Prev 20(1):140–147. https://doi.org/10.1158/1055-9965.EPI-10-0712

Beckonert O et al (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2(11):2692–2703. https://doi.org/10.1038/nprot.2007.376

Bernini P et al (2011) Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. J Biomol NMR 49(3):231–243. https://doi.org/10.1007/S10858-011-9489-1

Binczyk F, Tarnawski R, Polanska J (2015) Strategies for optimizing the phase correction algorithms in nuclear magnetic resonance spectroscopy. Biomed Eng Online 14(2):S5. https://doi.org/10.1186/1475-925X-14-S2-S5

Bingol K, Brüschweiler R (2014) Multidimensional approaches to NMR-based metabolomics. Anal Chem 86(1):47–57. https://doi.org/10.1021/ac403520j

Bingol K et al (2014) Customized metabolomics database for the analysis of NMR 1H-1H TOCSY and 13C-1H HSQC-TOCSY spectra of complex mixtures. Anal Chem 86(11):5494–5501. https://doi.org/10.1021/ac500979g

Bingol K et al (2015) Unified and isomer-specific NMR metabolomics database for the accurate analysis of 13 C – 1 H HSQC spectra. ACS Chem Biol 10(2):452–459. https://doi.org/10.1021/cb5006382

Bingol K et al (2016) Comprehensive metabolite identification strategy using multiple two-dimensional NMR spectra of a complex mixture implemented in the COLMARm web server. Anal Chem 88(24):12411–12418. https://doi.org/10.1021/acs.analchem.6b03724

Blankenberg FG et al (1997) Quantitative analysis of apoptotic cell death using proton nuclear magnetic resonance spectroscopy. Blood 89(10):3778–3786

Blasco H et al (2010) 1H-NMR-based metabolomic profiling of CSF in early amyotrophic lateral sclerosis. PLoS One 5(10):e13223. https://doi.org/10.1371/journal.pone.0013223
Bloembergen N, Purcell EM, Pound RV (1948) Relaxation effects in nuclear magnetic resonance absorption. Phys Rev 73(7):679–712. https://doi.org/10.1103/PhysRev.73.679
Blümich B, Singh K (2018) Desktop NMR and its applications from materials science to organic chemistry. Angew Chem Int Ed 57(24):6996–7010. https://doi.org/10.1002/anie.201707084
Bock JL (1982) Analysis of serum by high-field proton nuclear magnetic resonance. Clin Chem 28(9):1873–1877
Boiteau R et al (2018) Structure elucidation of unknown metabolites in metabolomics by combined NMR and MS/MS prediction. Metabolites 8(1):8. https://doi.org/10.3390/METABO8010008
Borges R et al (2021) Quantum chemistry calculations for cetabolomics. Chem Rev 121(10): 5633–5670. https://doi.org/10.1021/ACS.CHEMREV.0C00901
Bouatra S et al (2013) The human urine metabolome. PLoS One 8(9):e73076. https://doi.org/10.1371/journal.pone.0073076
Brown SA (2016) Circadian metabolism: from mechanisms to metabolomics and medicine. Trends Endocrinol Metab 27(6):415–426. https://doi.org/10.1016/j.tem.2016.03.015
Bruker Corporation-Bruker (2013) Bruker announces AVANCE™-IVDr as a standardized NMR platform for clinical screening and in vitro diagnostics (IVD) discovery and validation. https://ir.bruker.com/press-releases/press-release-details/2013/Bruker-Announces-AVANCE-IVDr-as-a-Standardized-NMR-Platform-for-Clinical-Screening-and-In-Vitro-Diagnostics-IVD-Discovery-and-Validation/default.aspx. Accessed 21 Sept 2021
Cañueto D et al (2018) rDolphin: a GUI R package for proficient automatic profiling of 1D 1 H-NMR spectra of study datasets. Metabolomics 14(3):24. https://doi.org/10.1007/s11306-018-1319-y
Cao M et al (2012) NMR-based metabolomic analysis of human bladder cancer. Anal Sci 28(5): 451–456. https://doi.org/10.2116/analsci.28.451
Capitani D, Sobolev AP, Mannina L (2017) Nuclear magnetic resonance – metabolomics. In: Georgiou CA, Danezis GP (eds) Food authentication: management, analysis and regulation. Wiley, Chichester, pp 177–197. https://doi.org/10.1002/9781118810224.ch6
Carrola J et al (2011) Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. J Proteome Res 10(1):221–230. https://doi.org/10.1021/pr100899x
Chan ECY et al (2009) Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). J Proteome Res 8(1):352–361. https://doi.org/10.1021/pr8006232
Chapinal N et al (2012) The association of serum metabolites in the transition period with milk production and early-lactation reproductive performance. J Dairy Sci 95(3):1301–1309. https://doi.org/10.3168/jds.2011-4724
Chen L et al (2002) An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization. J Magn Reson 158(1–2):164–168. https://doi.org/10.1016/S1090-7807(02)00069-1
Clendinen CS et al (2014) ^{13}C NMR metabolomics: applications at natural abundance. Anal Chem 86(18):9242–9250. https://doi.org/10.1021/ac502346h
Clendinen CS et al (2015) An overview of methods using (13)C for improved compound identification in metabolomics and natural products. Front Plant Sci 6:611. https://doi.org/10.3389/fpls.2015.00611
Cloarec O et al (2005) Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1 H NMR data sets. Anal Chem 77(5):1282–1289. https://doi.org/10.1021/ac048630x
Cobas C (2020) NMR signal processing, prediction, and structure verification with machine learning techniques. Magn Reson Chem 58(6):512–519. https://doi.org/10.1002/mrc.4989

Cohen SM, Ogawa S, Shulman RG (1979) 13C NMR studies of gluconeogenesis in rat liver cells: utilization of labeled glycerol by cells from euthyroid and hyperthyroid rats. Proc Natl Acad Sci U S A 76(4):1603–1609. https://doi.org/10.1073/pnas.76.4.1603

Dame ZT et al (2015) The human saliva metabolome. Metabolomics 11(6):1864–1883. https://doi.org/10.1007/s11306-015-0840-5

de Brouwer H (2009) Evaluation of algorithms for automated phase correction of NMR spectra. J Magn Reson 201(2):230–238. https://doi.org/10.1016/j.jmr.2009.09.017

Dey A et al (2020) Hyperpolarized NMR metabolomics at natural 13C abundance. Anal Chem 92(22):14867–14871. https://doi.org/10.1021/acs.analchem.0c03510

Ebbels TMD, De Iorio M, Stephens DA (2019) Statistical methods in metabolomics. In: Balding D, Moltke I, Marioni J (eds) Handbook of statistical genomics. Wiley, pp 949–975. https://doi.org/10.1002/9781119487845.ch34

Eisner R et al (2013) A machine-learned predictor of colonic polyps based on urinary metabolomics. Biomed Res Int 2013:1–11. https://doi.org/10.1155/2013/303982

Embade N et al (2019) NMR-based newborn urine screening for optimized detection of inherited errors of metabolism. Sci Rep 9(1):13067. https://doi.org/10.1038/s41598-019-49685-x

Emwas A-H et al (2015) Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. Metabolomics 11(4):872–894. https://doi.org/10.1007/S11306-014-0746-7

Emwas A-H et al (2018) Recommended strategies for spectral processing and post-processing of 1D 1H-NMR data of biofluids with a particular focus on urine. Metabolomics 14(3):31. https://doi.org/10.1007/s11306-018-1321-4

Fan TW-M, Lane AN (2011a) Erratum to: NMR-based stable isotope resolved metabolomics in systems biochemistry. J Biomol NMR 49:325. https://doi.org/10.1007/s10858-011-9503-7

Fan TWM, Lane AN (2011b) NMR-based stable isotope resolved metabolomics in systems biochemistry. J Biomol NMR 49(3–4):267–280. https://doi.org/10.1007/s10858-011-9484-6

Fan TWM, Lane AN (2016) Applications of NMR spectroscopy to systems biochemistry. Prog Nucl Magn Reson Spectrosc 92–93:18–53. https://doi.org/10.1016/j.pnmrs.2016.01.005

Farshidfar F et al (2012) Serum metabolomic profile as a means to distinguish stage of colorectal cancer. Genome Med 4(5):42. https://doi.org/10.1186/gm341

Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 48(1–2):155–171. https://doi.org/10.1023/A:1013713905833

Fiehn O et al (2007) The metabolomics standards initiative (MSI). Metabolomics 3(3):175–178. https://doi.org/10.1007/S11306-007-0070-6

Fong MY, McDunn J, Kakar SS (2011) Identification of metabolites in the normal ovary and their transformation in primary and metastatic ovarian cancer. PLoS One 6(5):e19963. https://doi.org/10.1371/journal.pone.0019963

Foroozandeh M, Jeannerat D (2010) Deciphered chemical shifts in aliased spectra recorded with two slightly different narrow windows or differential chemical shift evolution. ChemPhysChem 11(12):2503–2505. https://doi.org/10.1002/cphc.201000421

Foroutan A et al (2019) Chemical composition of commercial cow's milk. J Agric Food Chem 67(17):4897–4914. https://doi.org/10.1021/acs.jafc.9b00204

Foroutan A et al (2020) The bovine metabolome. Metabolites 10(6):233. https://doi.org/10.3390/metabo10060233

Fossel ET, Carr JM, McDonagh J (1986) Detection of malignant tumors. Water-suppressed proton nuclear magnetic resonance spectroscopy of plasma. N Engl J Med 315(22):1369–1376. https://doi.org/10.1056/NEJM198611273152201

Garcia-Perez I et al (2020) Identifying unknown metabolites using NMR-based metabolic profiling techniques. Nat Protoc 15(8):2538–2567. https://doi.org/10.1038/s41596-020-0343-3

Gartland KP et al (1990) Pattern recognition analysis of high resolution 1H NMR spectra of urine. A nonlinear mapping approach to the classification of toxicological data. NMR Biomed 3(4):166–172. https://doi.org/10.1002/nbm.1940030404

Ghini V et al (2019) NMR for sample quality assessment in metabolomics. N Biotechnol 52:25–34. https://doi.org/10.1016/J.NBT.2019.04.004
Giraudeau P, Frydman L (2014) Ultrafast 2D NMR: an emerging tool in analytical spectroscopy. Annu Rev Anal Chem 7(1):129–161. https://doi.org/10.1146/annurev-anchem-071213-020208
Goldansaz SA et al (2017) Livestock metabolomics and the livestock metabolome: a systematic review. PLoS One 12(5):e0177675. https://doi.org/10.1371/journal.pone.0177675
Golotvin S, Williams A (2000) Improved baseline recognition and modeling of FT NMR spectra. J Magn Reson 146(1):122–125. https://doi.org/10.1006/jmre.2000.2121
Gowda GAN et al (2010) Quantitative analysis of blood plasma metabolites using isotope enhanced NMR methods. Anal Chem 82(21):8983–8990. https://doi.org/10.1021/ac101938w
Griffiths RD, Edwards RH (1987) Magnetic resonance spectroscopy in the recognition of metabolic disease. J Inherit Metab Dis 10(Suppl 1):147–158
Hailemariam D et al (2014) Identification of predictive biomarkers of disease state in transition dairy cows. J Dairy Sci 97(5):2680–2693. https://doi.org/10.3168/jds.2013-6803
Halamek J, Vondra V, Kasal M (1994) The elimination of baseline distortions induced by audio filters. J Magn Reson Ser 110(2):194–197. https://doi.org/10.1006/jmra.1994.1204
Hanahan D, Weinberg R (2000) The hallmarks of cancer. Cell 100(1):57–70. https://doi.org/10.1016/S0092-8674(00)81683-9
Hansen DF (2019) Using deep neural networks to reconstruct non-uniformly sampled NMR spectra. J Biomol NMR 73(10–11):577–585. https://doi.org/10.1007/s10858-019-00265-1
Hansen AL et al (2021) 2D NMR-based metabolomics with HSQC/TOCSY NOAH supersequences. Anal Chem 93(15):6112–6119. https://doi.org/10.1021/acs.analchem.0c05205
Hao J et al (2014) Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. Nat Protoc 9(6):1416–1427. https://doi.org/10.1038/nprot.2014.090
Harris RK et al (2002) NMR nomenclature: nuclear spin properties and conventions for chemical shifts. IUPAC recommendations 2001. Solid State Nucl Magn Reson 22(4):458–483. https://doi.org/10.1006/snmr.2002.0063
Heuer A, Haeberlen U (1989) A new method for suppressing baseline distortions in FT NMR. J Magn Reson (1969) 85(1):79–94. https://doi.org/10.1016/0022-2364(89)90322-3
Holmes E, Wilson ID, Nicholson JK (2008) Metabolic phenotyping in health and disease. Cell 134(5):714–717. https://doi.org/10.1016/j.cell.2008.08.026
Hu K, Westler WM, Markley JL (2011) Simultaneous quantification and identification of individual chemicals in metabolite mixtures by two-dimensional extrapolated time-zero 1H-13C HSQC (HSQC 0). J Am Chem Soc 133(6):1662–1665. https://doi.org/10.1021/ja1095304
Iles RA et al (1984) Rapid screening of metabolic diseases by proton NMR. Lancet 2(8413): 1221–1222. https://doi.org/10.1016/S0140-6736(84)92783-1
Izquierdo-Garcia JL et al (2020) Discovery and validation of an NMR-based metabolomic profile in urine as TB biomarker. Sci Rep 10(1):22317. https://doi.org/10.1038/s41598-020-78999-4
Jeyarajah EJ, Cromwell WC, Otvos JD (2006) Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. Clin Lab Med 26(4):847–870. https://doi.org/10.1016/j.cll.2006.07.006
Kaddurah-Daouk R (2006) Metabolic profiling of patients with schizophrenia. PLoS Med 3(8): e363. https://doi.org/10.1371/journal.pmed.0030363
Karamanos TK et al (2015) Mechanisms of amyloid formation revealed by solution NMR. Prog Nucl Magn Reson Spectrosc 88–89:86–104. https://doi.org/10.1016/j.pnmrs.2015.05.002
Karu N et al (2018) A review on human fecal metabolomics: methods, applications and the human fecal metabolome database. Anal Chim Acta 1030:1–24. https://doi.org/10.1016/j.aca.2018.05.031
Kazimierczuk K et al (2010) Random sampling in multidimensional NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 57(4):420–434. https://doi.org/10.1016/j.pnmrs.2010.07.002
Keshari KR et al (2010) Hyperpolarized (13)C spectroscopy and an NMR-compatible bioreactor system for the investigation of real-time cellular metabolism. Magn Reson Med 63(2):322–329. https://doi.org/10.1002/mrm.22225

Kim S et al (2016) Food metabolomics: from farm to human. Curr Opin Biotechnol 37:16–23. https://doi.org/10.1016/j.copbio.2015.09.004
Kim E et al (2019) Urine-NMR metabolomics for screening of advanced colorectal adenoma and early stage colorectal cancer. Sci Rep 9(1):4786. https://doi.org/10.1038/s41598-019-41216-y
Kork F et al (2012) A biomarker for severity of Alzheimer's disease: 1H-NMR resonances in cerebrospinal fluid correlate with performance in mini-mental-state-exam. Biomarkers 17(1): 36–42. https://doi.org/10.3109/1354750X.2011.635806
Kostidis S, Mikros E (2015) NMR studies of inborn errors of metabolism. In: eMagRes. Wiley, Chichester, pp 57–68. https://doi.org/10.1002/9780470034590.emrstm1400
Kupče Ē, Claridge TDW (2017) NOAH: NMR supersequences for small molecule analysis and structure elucidation. Angew Chem Int Ed 56(39):11779–11783. https://doi.org/10.1002/anie.201705506
Lane AN et al (2011) Stable isotope-resolved metabolomics (SIRM) in cancer research with clinical application to nonsmall cell lung cancer. OMICS 15(3):173–182. https://doi.org/10.1089/omi.2010.0088
Le Guennec A, Giraudeau P, Caldarelli S (2014) Evaluation of fast 2D NMR for metabolomics. Anal Chem 86(12):5946–5954. https://doi.org/10.1021/ac500966e
Leggett A et al (2019) Identification of unknown metabolomics mixture compounds by combining NMR, MS, and cheminformatics. Methods Enzymol 615:407–422. https://doi.org/10.1016/BS.MIE.2018.09.003
Levy PA (2010) An overview of newborn screening. J Dev Behav Pediatr 31(7):622–631. https://doi.org/10.1097/DBP.0b013e3181eedf01
Lewis IA et al (2007) Method for determining molar concentrations of metabolites in complex solutions from two-dimensional 1H-13C NMR spectra. Anal Chem 79(24):9385–9390. https://doi.org/10.1021/ac071583z
Lewis IA, Schommer SC, Markley JL (2009) rNMR: open source software for identifying and quantifying metabolites in NMR spectra. Magn Reson Chem 47(Suppl 1):S123. https://doi.org/10.1002/mrc.2526
Lin CY et al (2007) Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. Metabolomics 3(1):55–67. https://doi.org/10.1007/s11306-006-0043-1
Lindon JC et al (2005) The consortium for metabonomic toxicology (COMET): aims, activities and achievements. Pharmacogenomics 6(7):691–699. https://doi.org/10.2217/14622416.6.7.691
Lindon JC, Holmes E, Nicholson JK (2007) Metabonomics in pharmaceutical R & D. FEBS J 274(5):1140–1151. https://doi.org/10.1111/j.1742-4658.2007.05673.x
Lopez JM, Cabrera R, Maruenda H (2019) Ultra-clean pure shift 1 H-NMR applied to metabolomics profiling. Sci Rep 9(1):6900. https://doi.org/10.1038/s41598-019-43374-5
Luchinat E et al (2021) Protein in-cell NMR spectroscopy at 1.2 GHz. J Biomol 75(2–3):97–107. https://doi.org/10.1007/s10858-021-00358-w
Ludwig C, Viant M (2010) Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. Phytochem Anal 21(1):22–32. https://doi.org/10.1002/PCA.1186
Lumata L et al (2015) Hyperpolarized 13C magnetic resonance and its use in metabolic assessment of cultured cells and perfused organs. Methods Enzymol 561:73–106. https://doi.org/10.1016/bs.mie.2015.04.006
Lutz N, Sweedler J, Wevers R (eds) (2013) Methodologies for metabolomics: experimental strategies and techniques. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511996634
Madsen R, Lundstedt T, Trygg J (2010) Chemometrics in metabolomics – a review in human disease diagnosis. Anal Chim Acta 659(1–2):23–33. https://doi.org/10.1016/j.aca.2009.11.042
Mak CM et al (2013) Inborn errors of metabolism and expanded newborn screening: review and update. Crit Rev Clin Lab Sci 50(6):142–162. https://doi.org/10.3109/10408363.2013.847896

Marliani AF et al (2007) Quantitative proton magnetic resonance spectroscopy of the human cervical spinal cord at 3 tesla. Magn Reson Med 57(1):160–163. https://doi.org/10.1002/mrm.21113

Martens L et al (2011) mzML – a community standard form mass spectrometry data. Mol Cell Proteomics 10(1):R110.000133. https://doi.org/10.1074/mcp.R110.000133

Martineau E, Giraudeau P (2019) Fast quantitative 2D NMR for untargeted and targeted metabolomics. Methods Mol Biol 2037:365–383. https://doi.org/10.1007/978-1-4939-9690-2_20

Martineau E et al (2011) Fast and precise quantitative analysis of metabolic mixtures by 2D 1H INADEQUATE NMR. J Pharm Biomed Anal 54(1):252–257. https://doi.org/10.1016/j.jpba.2010.07.046

Martineau E et al (2013) Fast quantitative 1H-13C two-dimensional NMR with very high precision. Anal Chem 85(9):4777–4783. https://doi.org/10.1021/ac4005309

Martineau E, Dumez JN, Giraudeau P (2020) Fast quantitative 2D NMR for metabolomics and lipidomics: a tutorial. Magn Reson Chem 58(5):390–403. https://doi.org/10.1002/mrc.4899

Martínez-Arranz I et al (2015) Enhancing metabolomics research through data mining. J Proteomics 127(Pt B):275–288. https://doi.org/10.1016/j.jprot.2015.01.019

Massou S et al (2007) NMR-based fluxomics: quantitative 2D NMR methods for isotopomers analysis. Phytochemistry 68(16–18):2330–2340. https://doi.org/10.1016/j.phytochem.2007.03.011

Mazzei P, Piccolo A (2017) HRMAS NMR spectroscopy applications in agriculture. Chem Biol Technol Agric 4(1):11. https://doi.org/10.1186/S40538-017-0093-9

McAlpine J et al (2019) The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. Nat Prod Rep 36(1):35–107. https://doi.org/10.1039/C7NP00064B

McKay RT (2009) Chapter 2 recent advances in solvent suppression for solution NMR: a practical reference. In: Annual reports on NMR spectroscopy. Elsevier, pp 33–76. https://doi.org/10.1016/S0066-4103(08)00402-X

Mckay RT (2011) How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: an examination of the pulse sequence components and evolution. Concepts Magn Reson A 38(5):197–220. https://doi.org/10.1002/cmr.a.20223

McNamara JR, Warnick GR, Cooper GR (2006) A brief history of lipid and lipoprotein measurements and their contribution to clinical chemistry. Clin Chim Acta 369(2):158–167. https://doi.org/10.1016/j.cca.2006.02.041

Melzer N, Wittenburg D, Repsilber D (2013) Integrating milk metabolite profile information for the prediction of traditional milk traits based on SNP information for Holstein cows. PLoS One 8(8):e70256. https://doi.org/10.1371/journal.pone.0070256

Mercier P et al (2011) Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. J Biomol NMR 49(3–4):307–323. https://doi.org/10.1007/s10858-011-9480-x

Midgley I, Hawkins DR (1978) The use of 13C-nmr spectroscopy for the detection and identification of metabolites of carbon- 13 labelled amitriptyline. J Pharm Pharmacol 30(9):547–553. https://doi.org/10.1111/j.2042-7158.1978.tb13321.x

Monakhova YB et al (2014) Validation studies for multicomponent quantitative NMR analysis: the example of apple fruit juice. Accred Qual Assur 19(1):17–29. https://doi.org/10.1007/s00769-013-1026-3

Motta A, Paris D, Melck D (2010) Monitoring real-time metabolism of living cells by fast two-dimensional NMR spectroscopy. Anal Chem 82(6):2405–2411. https://doi.org/10.1021/ac9026934

Moutzouri P et al (2017) Ultraclean pure shift NMR. Chem Commun (Camb) 53(73):10188–10191. https://doi.org/10.1039/c7cc04423b

Nagana Gowda GA, Raftery D (2014) Quantitating metabolites in protein precipitated serum using NMR spectroscopy. Anal Chem 86(11):5433–5440. https://doi.org/10.1021/ac5005103

Nagana Gowda GA, Gowda YN, Raftery D (2015) Expanding the limits of human blood metabolite quantitation using NMR spectroscopy. Anal Chem 87(1):706–715. https://doi.org/10.1021/ac503651e

Namer IJ et al (2011) Metabolomic characterization of ovarian epithelial carcinomas by HRMAS-NMR spectroscopy. J Oncol 2011:174019. https://doi.org/10.1155/2011/174019

Nicholson JK, Wilson ID (2003) Opinion: understanding “global” systems biology: metabonomics and the continuum of metabolism. Nat Rev Drug Discov 2(8):668–676. https://doi.org/10.1038/nrd1157

Nicholson JK et al (1984a) Monitoring metabolic disease by proton NMR of urine. Lancet 2(8405):751–752. https://doi.org/10.1016/S0140-6736(84)92656-4

Nicholson JK et al (1984b) Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. Biochem J 217(2):365–375. https://doi.org/10.1042/bj2170365

Nicholson JK et al (2002) Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov 1(2):153–161. https://doi.org/10.1038/nrd728

Oliver S et al (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16(9):373–378. https://doi.org/10.1016/S0167-7799(98)01214-1

Otvos JD, Jeyarajah EJ, Bennett DW (1991) Quantification of plasma lipoproteins by proton nuclear magnetic resonance spectroscopy. Clin Chem 37(3):377–386

Palmnas MSA, Vogel HJ (2013) The future of NMR metabolomics in cancer therapy: towards personalizing treatment and developing targeted drugs? Metabolites 3(2):373–396. https://doi.org/10.3390/metabo3020373

Paruzzo F, Bruderer S, Janjar Y, Bjoern Heitmann B, Bolliger C (2020) Automatic signal region detection in 1H NMR spectra using deep learning [White paper]. Switzerland

Pathan M et al (2011) “Multi-scan single shot” quantitative 2D NMR: a valuable alternative to fast conventional quantitative 2D NMR. Analyst 136(15):3157–3163. https://doi.org/10.1039/c1an15278e

Percival BC et al (2019) Low-field, benchtop NMR spectroscopy as a potential tool for point-of-care diagnostics of metabolic conditions: validation, protocols and computational models. High-Throughput 8(1):2. https://doi.org/10.3390/ht8010002

Percival B et al (2020) Chapter 1: univariate and multivariate statistical approaches to the analysis and interpretation of NMR-based metabolomics datasets of increasing complexity. In: Computational techniques for analytical chemistry and bioanalysis. Royal Society of Chemistry, pp 1–40. https://doi.org/10.1039/9781788015882-00001

Peterson DJ, Loening NM (2007) QQ-HSQC: a quick, quantitative heteronuclear correlation experiment for NMR spectroscopy. Magn Reson Chem 45(11):937–941. https://doi.org/10.1002/mrc.2073

Provencher SW (1993) Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magn Reson Med 30(6):672–679. https://doi.org/10.1002/mrm.1910300604

Provencher SW (2001) Automatic quantitation of localized in vivo 1H spectra with LCModel. NMR Biomed 14(4):260–264. https://doi.org/10.1002/nbm.698

Psychogios N et al (2011) The human serum metabolome. PLoS One 6(2):e16957. https://doi.org/10.1371/journal.pone.0016957

Ramaswamy V et al (2013) Development of a 13C-optimized 1.5-mm high temperature superconducting NMR probe. J Magn Reson 235:58–65. https://doi.org/10.1016/j.jmr.2013.07.012

Ramaswamy V et al (2016) Development of a 1H-13C dual-optimized NMR probe based on double-tuned high temperature superconducting resonators. IEEE Trans Appl Supercond 26(3):1–5. https://doi.org/10.1109/TASC.2016.2522302

Ravanbakhsh S et al (2015) Accurate, fully-automated NMR spectral profiling for metabolomics. PLoS One 10(5):1–15. https://doi.org/10.1371/journal.pone.0124219

Röhnisch HE et al (2018) AQuA: an automated quantification algorithm for high-throughput NMR-based metabolomics and its application in human plasma. Anal Chem 90(3): 2095–2102. https://doi.org/10.1021/acs.analchem.7b04324

Rothman DL et al (2003) In vivo NMR studies of the glutamate neurotransmitter flux and neuroenergetics: implications for brain function. Annu Rev Physiol 65(1):401–427. https://doi.org/10.1146/annurev.physiol.65.092101.142131

Saborano R et al (2019) A framework for tracer-based metabolism in mammalian cells by NMR. Sci Rep 9(1):2520. https://doi.org/10.1038/s41598-018-37525-3

Saccenti E et al (2014) Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics 10(3):361–374. https://doi.org/10.1007/s11306-013-0598-6

Saleem F et al (2013) The bovine ruminal fluid metabolome. Metabolomics 9(2):360–378. https://doi.org/10.1007/s11306-012-0458-9

Salek R et al (2015) COordination of standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. Metabolomics 11(6):1587–1597. https://doi.org/10.1007/S11306-015-0810-Y

Savorani F, Tomasi G, Engelsen SB (2010) Icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson 202(2):190–202. https://doi.org/10.1016/j.jmr.2009.11.012

Schober D et al (2018) NmrML: a community supported open data standard for the description, storage, and exchange of NMR data. Anal Chem 90(1):649–656. https://doi.org/10.1021/acs.analchem.7b02795

Schwalbe H (2017) Editorial: new 1.2 GHz NMR spectrometers- new horizons? Angew Chem Int Ed 56(35):10252–10253. https://doi.org/10.1002/anie.201705936

Sellick CA et al (2009) Effective quenching processes for physiologically valid metabolite profiling of suspension cultured mammalian cells. Anal Chem 81(1):174–183. https://doi.org/10.1021/ac8016899

Serrai H et al (1998) Quantification of plasma lipoprotein fractions by wavelet transform time-domain data processing of the proton nuclear magnetic resonance methylene spectral region. NMR Biomed 11(6):273–280. https://doi.org/10.1002/(SICI)1099-1492(199810)11:6<273::AID-NBM523>3.0.CO;2-J

Sette M, Lange H, Crestini C (2013) Quantitative HSQC analyses of lignin: a practical comparison. Comput Struct Biotechnol J 6(7):e201303016. https://doi.org/10.5936/csbj.201303016

Seyfried T, Shelton L (2010) Cancer as a metabolic disease. Nutr Metab 7(1):1–22. https://doi.org/10.1186/1743-7075-7-7

Shrot Y, Frydman L (2009) Spatial/spectral encoding of the spin interactions in ultrafast multidimensional NMR. J Chem Phys 131(22):224516. https://doi.org/10.1063/1.3266422

Simister RJ et al (2003) A proton magnetic resonance spectroscopy study of metabolites in the occipital lobes in epilepsy. Epilepsia 44(4):550–558. https://doi.org/10.1046/j.1528-1157.2003.19102.x

Simpson MJ, Bearden DW (2013) Environmental metabolomics: NMR techniques. In: Harris RK, Wasylishen RL (eds) eMagRes. Wiley, pp 549–560. https://doi.org/10.1002/9780470034590.emrstm1352

Smolinska A et al (2012) NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Anal Chim Acta 750:82–97. https://doi.org/10.1016/j.aca.2012.05.049

Sobolev AP et al (2019) Use of NMR applications to tackle future food fraud issues. Trends Food Sci Technol 91:347–353. https://doi.org/10.1016/j.tifs.2019.07.035

Soininen P et al (2015) Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet 8(1):192–206. https://doi.org/10.1161/CIRCGENETICS.114.000216

Sokolenko S et al (2013) Understanding the variability of compound quantification from targeted profiling metabolomics of 1D-1H-NMR spectra in synthetic mixtures and urine with additional insights on choice of pulse sequences and robotic sampling. Metabolomics 9(4):887–903. https://doi.org/10.1007/s11306-013-0503-3

Spicer R, Salek R, Steinbeck C (2017) A decade after the metabolomics standards initiative it's time for a revision. Sci Data 4(1):170138. https://doi.org/10.1038/sdata.2017.138
Spraul M et al (2015) Wine analysis to check quality and authenticity by fully-automated 1 H-NMR. BIO Web Conf 5(23):02022. https://doi.org/10.1051/bioconf/20150502022
Steimers E et al (2020) Application of a new method for simultaneous phase and baseline correction of NMR signals (SINC). Magn Reson Chem 58(3):260–270. https://doi.org/10.1002/mrc.4964
Sumner LW et al (2015) Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. Nat Prod Rep 32(2):212–229. https://doi.org/10.1039/C4NP00072B
Sundekilde UK et al (2014) Association between the bovine milk metabolome and rennet-induced coagulation properties of milk. J Dairy Sci 97(10):6076–6084. https://doi.org/10.3168/jds.2014-8304
Takis PG et al (2017) Deconvoluting interrelationships between concentrations and chemical shifts in urine provides a powerful analysis tool. Nat Commun 8(1):1662. https://doi.org/10.1038/s41467-017-01587-0
Tal A, Frydman L (2010) Single-scan multidimensional magnetic resonance. Prog Nucl Magn Reson Spectrosc 57(3):241–292. https://doi.org/10.1016/j.pnmrs.2010.04.001
Tardivel P et al (2017) ASICS: an automatic method for identification and quantification of metabolites in complex 1D 1H NMR spectra. Metabolomics 13(10):109. https://doi.org/10.1007/s11306-017-1244-5
Tasic L et al (2017) Metabolomics and lipidomics analyses by1H nuclear magnetic resonance of schizophrenia patient serum reveal potential peripheral biomarkers for diagnosis. Schizophr Res 185:182–189. https://doi.org/10.1016/j.schres.2016.12.024
Tayyari F et al (2013) 15N-Cholamine – a smart isotope tag for combining NMR- and MS-based metabolite profiling. Anal Chem 85(18):8715–8721. https://doi.org/10.1021/ac401712a
Teahan O et al (2011) Metabolic signatures of malignant progression in prostate epithelial cells. Int J Biochem Cell Biol 43(7):1002–1009. https://doi.org/10.1016/j.biocel.2010.07.003
Teng Q et al (2009) A direct cell quenching method for cell-culture based metabolomics. Metabolomics 5(2):199–208. https://doi.org/10.1007/s11306-008-0137-z
Theis T et al (2016) Direct and cost-efficient hyperpolarization of long-lived nuclear spin states on universal 15N2-diazirine molecular tags. Sci Adv 2(3):e1501438. https://doi.org/10.1126/sciadv.1501438
Tilgner M et al (2019) High-resolution magic angle spinning (HRMAS) NMR methods in metabolomics. Methods Mol Biol 2037:49–67. https://doi.org/10.1007/978-1-4939-9690-2_4
Truong ML et al (2015) 15N hyperpolarization by reversible exchange using SABRE-SHEATH. J Phys Chem 119(16):8786–8797. https://doi.org/10.1021/acs.jpcc.5b01799
Verhoeven A, Giera M, Mayboroda OA (2018) KIMBLE: a versatile visual NMR metabolomics workbench in KNIME. Anal Chim Acta 1044:66–76. https://doi.org/10.1016/j.aca.2018.07.070
Viant MR (2008) Recent developments in environmental metabolomics. Mol Biosyst 4(10):980–986. https://doi.org/10.1039/b805354e
Vuckovic D (2012) Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. Anal Bioanal Chem 403(6):1523–1548. https://doi.org/10.1007/s00216-012-6039-y
Wang T et al (2009) Automics: an integrated platform for NMR-based metabonomics spectral processing and data analysis. BMC Bioinformatics 10(1):83. https://doi.org/10.1186/1471-2105-10-83
Warburg O (1956) On the origin of cancer cells. Science 123(3191):309–314. https://doi.org/10.1126/SCIENCE.123.3191.309
Watanabe R et al (2016) Quantitative nuclear magnetic resonance spectroscopy based on PULCON methodology: application to quantification of invaluable marine toxin, okadaic acid. Toxins 8(10):294. https://doi.org/10.3390/toxins8100294

Weber F et al (2012) 31P and 13C solid-state NMR spectroscopy to study collagen synthesis and biomineralization in polymer-based bone implants. NMR Biomed 25(3):464–475. https://doi.org/10.1002/nbm.1649
Weiner MW et al (1989) Clinical magnetic resonance spectroscopy of brain, heart, liver, kidney, and cancer. A quantitative approach. NMR Biomed 2(5–6):290–297. https://doi.org/10.1002/nbm.1940020519
Weljie AM et al (2006) Targeted profiling: quantitative analysis of 1H NMR metabolomics data. Anal Chem 78(13):4430–4442. https://doi.org/10.1021/ac060209g
Weljie AM et al (2011) 1H NMR metabolomics identification of markers of hypoxia-induced metabolic shifts in a breast cancer model system. J Biomol NMR 49(3–4):185–193. https://doi.org/10.1007/s10858-011-9486-4
Williams TH et al (1979) Novel application of proton nuclear magnetic resonance spectroscopy in the identification of 2'-chloronordiazepam metabolites in the dog. J Med Chem 22(4):436–440. https://doi.org/10.1021/jm00190a016
Williams TD et al (2009) Hepatic transcriptomic and metabolomic responses in the stickleback (Gasterosteus aculeatus) exposed to environmentally relevant concentrations of dibenzanthracene. Environ Sci Technol 43(16):6341–6348. https://doi.org/10.1021/es9008689
Wilson DM, Burlingame AL (1974) Deuterium and carbon-13 tracer studies of ethanol metabolism in the rat by 2H, 1H-decoupled 13C nuclear magnetic resonance. Biochem Biophys Res Commun 56(3):828–835. https://doi.org/10.1016/0006-291X(74)90680-9
Wishart DS (2005) Metabolomics: the principles and potential applications to transplantation. Am J Transplant 5(12):2814–2820. https://doi.org/10.1111/j.1600-6143.2005.01119.x
Wishart DS (2008a) Applications of metabolomics in drug discovery and development. Drugs R D 9(5):307–322. https://doi.org/10.2165/00126839-200809050-00002
Wishart DS (2008b) Metabolomics: applications to food science and nutrition research. Trends Food Sci Technol 19(9):482–493. https://doi.org/10.1016/j.tifs.2008.03.003
Wishart DS (2015) Is cancer a genetic disease or a metabolic disease? EBioMedicine 2(6):478–479. https://doi.org/10.1016/j.ebiom.2015.05.022
Wishart D (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15(7):473–484. https://doi.org/10.1038/nrd.2016.32
Wishart DS et al (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 6(2):135–140. https://doi.org/10.1007/BF00211777
Wishart DS et al (2001) Magnetic resonance diagnostics: a new technology for high-throughput clinical diagnostics. Clin Chem 47(10):1918–1921
Wishart DS et al (2007) HMDB: the human metabolome database. Nucleic Acids Res 35(Suppl 1): D521–D526. https://doi.org/10.1093/nar/gkl923
Wishart DS et al (2008) The human cerebrospinal fluid metabolome. J Chromatogr B Analyt Technol Biomed Life Sci 871(2):164–173. https://doi.org/10.1016/j.jchromb.2008.05.001
Wishart DS et al (2018) HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 46(D1):D608–D617. https://doi.org/10.1093/nar/gkx1089
Wolak J et al (2012) Noninvasive fluxomics in mammals by nuclear magnetic resonance spectroscopy. In: Fan TW-M, Lane A, Higashi R (eds) The handbook of metabolomics. Methods in pharmacology and toxicology. Humana Press, Totowa, pp 321–392. https://doi.org/10.1007/978-1-61779-618-0_11
Worley B, Powers R (2014) MVAPACK: a complete data handling package for NMR metabolomics. ACS Chem Biol 9(5):1138–1144. https://doi.org/10.1021/cb4008937
Wu J et al (2016) NMR analysis of the CSF and plasma metabolome of rigorously matched amyotrophic lateral sclerosis, Parkinson's disease and control subjects. Metabolomics 12(6): 101. https://doi.org/10.1007/s11306-016-1041-6
Wu K et al (2021) Improvement in signal-to-noise ratio of liquid-state NMR spectroscopy via a deep neural network DN-Unet. Anal Chem 93(3):1377–1382. https://doi.org/10.1021/acs.analchem.0c03087

Würtz P et al (2017) Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -Omic technologies. Am J Epidemiol 186(9):1084–1096. https://doi.org/10.1093/aje/kwx016

Xi Y, Rocke DM (2008) Baseline correction for NMR spectroscopic metabolomics data analysis. BMC Bioinform 9(1):324. https://doi.org/10.1186/1471-2105-9-324

Xia J et al (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 37(Web Server issue):W652–W660. https://doi.org/10.1093/nar/gkp356

Yamaguchi S et al (1984) Rapid screening of metabolic disease by proton NMR urinalysis. Lancet 2(8397):284

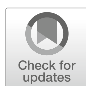
Ye T et al (2009) Chemoselective 15N tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. Anal Chem 81(12):4882–4888. https://doi.org/10.1021/ac900539y

Yoshikawa K, Matsushita K, Ohsaka A (1982) 1H-NMR spectroscopy in aqueous mediums. Examination of experimental conditions with human urine as a model sample. Physiol Chem Phys 14(4):385–389

Zhang B et al (2016) Nanoparticle-assisted removal of protein in human serum for metabolomics studies. Anal Chem 88(1):1003–1007. https://doi.org/10.1021/acs.analchem.5b03889

Zheng C et al (2011) Identification and quantification of metabolites in 1H NMR spectra by Bayesian model selection. Bioinformatics 27(12):1637–1644. https://doi.org/10.1093/bioinformatics/btr118

Zorin V, Bernstein MA, Cobas C (2017) A robust, general automatic phase correction algorithm for high-resolution NMR data. Magn Reson Chem 55(8):738–746. https://doi.org/10.1002/mrc.4586



Compound Identification Strategies in Mass Spectrometry-Based Metabolomics and Pharmacometabolomics

Rylan Hissong, Kendra R. Evans, and Charles R. Evans

Contents

1 Introduction 44
2 Step 1: Study Design: Define Compound Identification Goals 45
3 Step 2: Acquire Data – Experimental Strategies to Identify Compounds in Metabolomics Data 48
4 Step 3: Computational Strategies for Data Cleaning and Feature Annotation 54
5 Step 4: MS/MS Libraries and Compound Identification Using Library Search 57
6 Step 5: Assess and Report Identification Confidence 61
7 Conclusion: Pharmacology-Focused Compound Identification 63
References 64

Abstract

The metabolome is composed of a vast array of molecules, including endogenous metabolites and lipids, diet- and microbiome-derived substances, pharmaceuticals and supplements, and exposome chemicals. Correct identification of compounds from this diversity of classes is essential to derive biologically relevant insights from metabolomics data. In this chapter, we aim to provide a practical overview of compound identification strategies for mass spectrometry-based metabolomics, with a particular eye toward pharmacologically-relevant studies. First, we describe routine compound identification strategies applicable to targeted metabolomics. Next, we discuss both experimental (data

Rylan Hissong and Kendra R. Evans contributed equally to this work.

R. Hissong · C. R. Evans (✉)
University of Michigan, Ann Arbor, MI, USA
e-mail: chevans@med.umich.edu

K. R. Evans
University of Detroit Mercy, Detroit, MI, USA

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_617

acquisition-focused) and computational (software-focused) strategies used to identify unknown compounds in untargeted metabolomics data. We then discuss the importance of, and methods for, assessing and reporting the level of confidence of compound identifications. Throughout the chapter, we discuss how these steps can be implemented using today's technology, but also highlight research underway to further improve accuracy and certainty of compound identification. For readers interested in interpreting metabolomics data already collected, this chapter will supply important context regarding the origin of the metabolite names assigned to features in the data and help them assess the certainty of the identifications. For those planning new data acquisition, the chapter supplies guidance for designing experiments and selecting analysis methods to enable accurate compound identification, and it will point the reader toward best-practice data analysis and reporting strategies to allow sound biological and pharmacological interpretation.

Keywords

Compound identification · Identification confidence · LC-MS · Metabolomics · Molecular formula assignment · MS/MS search

1 Introduction

Metabolomics is a technique designed to provide a window into the small-molecule composition of a biological sample. In the context of pharmacological research, metabolomics can give insight into the uptake, metabolism, and clearance of a drug, delineate its impact on endogenous metabolism, or be used to discover biomarkers that predict which individuals will have a favorable response when it is administered. In all cases, metabolomics data is most meaningful when compounds in the sample can be accurately quantitated, confidently identified to the level of a unique chemical structure, and mapped to a biochemical or pharmacological pathway. Mass spectrometry-based metabolomics can be used to detect anywhere from a single compound of interest using a targeted method to tens of thousands of features using an untargeted workflow (Verpoorte, "Natural products drug discovery: on silica or in-silico?"; Günther, "Metabolomics in cell biology"). However, in a typical untargeted metabolomics study, a substantial portion of the features detected are artifacts and only a fraction of the presumed unique features can be readily identified (Mahieu and Patti 2017). For those features that can be identified, the degree of confidence with which chemical structures can be assigned varies and is determined by multiple factors. Therefore, selection of proper study design, data acquisition, and data analysis methods is essential to achieve the desired compound identification goals, and should be considered in the context of all the steps of a metabolomics study.

Most studies that use pharmacometabolomics rely on accurate compound identification, but many manuscripts provide limited detail about how compound identifications are made or the confidence with which they were assigned. It is therefore important for researchers in the field of pharmacology seeking to generate or interpret metabolomics data to be aware of strategies for compound identification and limitations of the methods. Numerous excellent reviews have already been published about compound identification strategies for metabolomics (Wishart 2009; Watson 2013; Blaženović et al. 2018; Chaleckis et al. 2019). In this chapter, we will summarize both basic strategies and innovative techniques that can be used to facilitate reliable compound identification when using metabolomics. We also include some discussion of compound identification strategies relevant to mass spectrometry-based lipidomics, which focuses on study of non-polar lipid and lipid-like molecules as opposed to polar metabolites. Whenever possible, we will note examples of studies that demonstrate varied compound identification strategies in the context of pharmacology and pharmacometabolomics. However, even though most of the literature focused on small molecule compound identification techniques does not specifically describe application to pharmacological research, the techniques apply to all classes of small molecules, whether endogenous, drug-derived, or of other origin.

This chapter is sub-divided into "steps" structured around a typical metabolomics (or pharmacometabolomics) study, focusing on how compound identification is relevant to or implemented at every stage in the workflow. In "Step 1," we begin by discussing how to select a workflow to achieve the investigator's compound identification-related goals, contrasting compound identification in targeted and untargeted metabolomics. "Step 2" focuses on experimental data collection strategies, describing the preparations needed to attempt to identify unknown features in untargeted metabolomics data. "Step 3" and "Step 4" turn our attention to data analysis, describing computational strategies developed to aid in identifying unknown features. The inter-connectedness and inter-dependency of computational and experimental methods will be highlighted. "Step 5" discusses strategies to assess confidence in the accuracy of compound identifications and best practices for reporting compound identifications in publications. The chapter concludes by summarizing considerations relating to compound identification of particular relevance to pharmacology. Throughout the chapter, we describe some of the major challenges that prevent universal and complete identification of all features that can be detected in metabolomics data, and we highlight current research and potential future developments that may bring this goal closer to fruition.

2 Step 1: Study Design: Define Compound Identification Goals

The first step of any metabolomics experiment is to define its goals; this also applies to compound identification. Is the project intended to perform routine identification and quantitation of a moderate number of well-known metabolic intermediates or a

drug and its known metabolites in human urine? Or, is it to detect and identify biomarkers of a disease in human plasma, including unknown compounds? Or, to determine if a particular metabolite profile predicts responsiveness to a treatment? All these scenarios, and many other uses of metabolomics, rely on robust compound identification. However, the way compound identification is best achieved is likely to differ depending on study design. One of the first choices that must be made is whether the metabolomics workflow to be used should be targeted, untargeted, or use a combination of both approaches. The distinctions between targeted and untargeted metabolomics are described in other chapters of this book and have been reviewed extensively (Patti et al. 2012; Cajka and Fiehn 2016; Fiehn 2016a; Schrimpe-Rutledge et al. 2016); Saigusa et al. describe advantages and disadvantages of targeted and untargeted methods in the context of pharmacometabolomics and drug discovery (Saigusa et al. 2021). Here, we briefly consider commonalities and distinctions of compound identification in targeted and untargeted metabolomics.

Compound Identification in Targeted Metabolomics In targeted metabolomics studies, most of the effort for compound identification is carried out up-front, before data from biological samples are collected. Methods are set up using authentic standards that allow detection of a pre-defined set of compounds, based on their mass, retention time, and in the case of MS/MS-based methods, specific fragment ions. Once targeted metabolomics methods are created, in principle no additional steps beyond routine instrumental and data analysis are needed to identify compounds. The tradeoff for the simplicity of compound identification in targeted metabolomics is the fact that the number of compounds that can be identified in a typical targeted analysis is limited. Most targeted assays are used to quantitate from one to a few hundred compounds, though recent targeted lipidomics methods have been reported to have the capability to detect and quantitate over 1,000 species using three separate LC–MS runs per sample (Contrepois et al. 2018). It is also important to note that identifications in targeted methods are not always free from interference or error. Structural isomers share the same molecular mass, are often indistinguishable by MS/MS fragmentation pattern, and may be difficult to resolve by chromatography. These factors may prevent unambiguous identification in targeted and untargeted methods alike.

Several guides and reviews have been written that describe how to develop targeted metabolomics methods (Griffiths et al. 2010; Parker et al. 2014; Zhou and Yin 2016; Roberts et al. 2012); but a few notes with relevance to compound identification merit mention. First, it is wise to investigate possible structural isomers of compounds of interest by using metabolite databases, described in “Step 3” in this chapter. In some cases, it may be prudent to purchase standards for potential interfering compounds and confirm they are chromatographically or spectrally resolved from targeted analytes. Secondly, if a method is to be adopted from the literature, it is not always safe to assume that provided parameters will produce results free from interference from isobaric (same nominal mass) species. Authentic standards should be used to validate methods, preferably by spiking a biological

sample with the expected compound at an appropriate concentration to observe an expected increase in peak area, with no evidence of alteration in peak shape or retention time that could indicate presence of an interferent. Finally, in the case of selected reaction monitoring or multiple reaction monitoring methods, it is beneficial to use both a primary product ion for quantification and secondary ion as a qualifier, and to monitor the abundance ratio between the two to ensure no deviation from that observed with an authentic standard.

Targeted metabolomics, or more broadly, targeted mass spectrometry-based small molecule assays, are widely used for pharmacological applications. Such applications include quantitation of drug molecules themselves or their direct metabolites within biological samples. Targeted approaches find use in pharmacokinetics studies (Kantae et al. 2017), forensic and clinical toxicology (Maurer 1998), assays for drugs of abuse (Zaitsu et al. 2016), along with many other applications. Prakash et al. reviewed strategies for identification of drug metabolites using mass spectrometry as the primary approach while defining the utility of complementary methods such as NMR and chemical derivatization (Prakash et al. 2007). Targeted metabolomics in pharmacological studies also frequently focuses on the study of endogenous metabolites, which can be used to assess the effect of a drug on specific endogenous metabolites or metabolic pathways of interest (Kantae et al. 2017; McCann et al. 2021).

Compound Identification in Untargeted Metabolomics and Lipidomics In contrast to targeted metabolomics, which focuses on analysis of specific pre-selected compounds, untargeted methods aim to quantitate as broad a swath of the metabolome or lipidome as possible using unbiased data acquisition methods. Thus, untargeted metabolomics is designed to detect all features in raw LC–MS or GC–MS data that may originate from a molecule in the sample. Initially, these features have no assigned chemical identity; they are represented by their measured physical or chemical properties such as their mass/charge ratio (m/z), retention time (RT), fragmentation pattern, and/or other measured data. Some features in untargeted metabolomics data may be identified by matching their m/z and RT, and when available their MS/MS spectrum or measured ion mobility drift time, to standards analyzed under identical conditions (preferably in the same lab, on the same instrument using the same chromatographic and mass spectrometric method). However, a substantial proportion of features detected in untargeted metabolomics data typically cannot be identified using this "targeted" strategy, even when many standards have been catalogued. Some portion of these unidentified features represent contaminants or degenerate signals; removal of as many of these artifacts as possible from the data is an important preliminary computational step in compound identification efforts and is described "Step 3" in this chapter.

After artifact removal, at least a portion of the unknown features may represent biologically or pharmacologically relevant compounds. Identification of these features is central to the discovery-driven nature of untargeted metabolomics, and is a primary goal of many pharmacometabolomics studies (Steuer et al. 2019). Strategies for unknown feature identification are the focus of the remainder of this

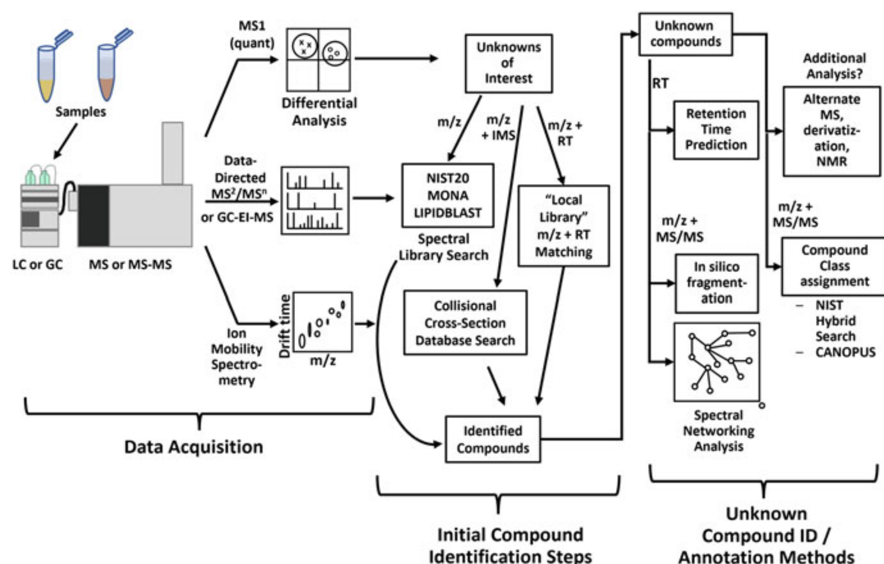


Fig. 1 Generalized workflow for compound identification in untargeted metabolomics illustrating both experimental and computational approaches

chapter. A generalized workflow for compound identification in untargeted metabolomics is illustrated in Fig. 1. It contains both experimental and computational components, which will be described in detail in "Steps 2–4" of this chapter.

3 Step 2: Acquire Data – Experimental Strategies to Identify Compounds in Metabolomics Data

Once a study design has been selected and goals for compound identification have been established, the next step is to perform experimental analysis. Here, we describe major steps of a metabolomics experiment relevant to compound identification, pointing out specific strategies that can be employed to improve the number and quality of identifications that can be obtained from the data.

Sample Preparation Sample preparation for metabolomics plays a substantial role in determining metabolite recovery, and by extension which metabolites can be detected and identified (Lu et al. 2008). Solvent extraction is a near-universal component of metabolomics sample preparation protocols and has been studied extensively in the context of recovery and quantitation of metabolites (Bruce et al. 2009; Lorenz et al. 2011), though some studies have also investigated how sample preparation methods modulate the number and chemical class of features that can be identified (Anderson et al. 2021; Lenz et al. 2007; Koek et al. 2008). One key choice that can impact compound identification is selecting between a single-phase or

biphasic extraction method. While single-phase extracts are simpler to prepare, multi-phase extracts separate polar and non-polar metabolites into separate solutions, allowing MS data acquisition parameters and database selection to be tailored to the polarity of each phase, potentially yielding more and more accurate identifications (Blaženović et al. 2018; Matyash et al. 2008). Other sample preparation methods such as solid phase extraction can be used to concentrate low-abundance compounds and potentially improve identification performance, though these methods' impact on compound identification has been less thoroughly studied (Wu et al. 2019). The simpler strategy of increasing column loading can also be used to enhance detection and identification of low-abundance compounds. Anderson et al. demonstrated that increasing sample concentration and injection volume to achieve column loading over 10-fold higher than typical loading for a reversed-phase LC–MS metabolomics experiment yielded more and higher-quality compound identifications, at the expense of a modest loss of chromatographic resolution (Anderson et al. 2021).

Chemical Derivatization The primary goal of chemical derivatization, a step used in some but not all metabolomics workflows, is usually to make compounds more amenable to detection and quantitation by GC–MS or LC–MS. However, it can also play a role in compound identification. GC–MS-based metabolomics relies on derivatization to convert non-volatile compounds to species that can be vaporized and chromatographed in the gas phase. The most widely used approach employs methoximation followed by addition of trimethylsilyl (TMS) or tert-butyldimethylsilyl (TBDMS) groups to derivatize ketones, alcohols, amines, and amides (Lisec et al. 2006; Lee et al. 2005). Fragmentation of TMS- or TBDMS-derivatized metabolites follow consistent patterns that can be interpreted to help identify unknown features. Lai et al. reviewed decades of publications describing GC–MS analysis of TMS-derivatized small molecules and compiled a series of rules and fragmentation trees useful for assigning substructures and facilitating unknown identification (Lai and Fiehn 2018). Silylated derivatives are also amenable to spectral library search, described in "Step 4" below; many GC–MS libraries contain spectra of an extensive catalog of silylated metabolite derivatives (Kind and Fiehn 2010; Halket et al. 2005). Silylated metabolites, like all compounds amenable to GC–MS analysis, can also be characterized by their retention index relative to a series of standards such as n-alkanes or fatty acid methyl esters (Kind et al. 2009; Strehmel et al. 2008). This allows chromatographic retention information to be included in GC–MS databases like FiehnLib and the NIST EI-MS library, both described in more detail below, adding orthogonal information for confirmation of compound identification (Kind and Fiehn 2010).

Chemical derivatization also has potential application to compound identification in LC–MS-based metabolomics (Zhao and Li 2020; Han et al. 2015). Multiple reagents with different selectivity are available. Phenylisothiocyanate (PITC) is a classic derivatization reagent commonly used for amino acid analysis by LC with ultraviolet absorbance or fluorescence detection, but has also been used for LC–MS analysis of species such as catecholamines (Zheng et al. 2018). Similarly, dansyl

chloride and benzoyl chloride react quantitatively with amines, improving ionization efficiency, chromatographic retention, and MS/MS fragmentation (Guo and Li 2009; Wong et al. 2016). The compound 3-nitrophenylhydrazine can be used to derivatize fatty acids, other carboxylic acids, and phosphoryl metabolites (Meng et al. 2021). Using a combination of derivatization techniques has been proposed as a strategy to facilitate characterization of the chemical composition of unknowns by revealing the presence of key functional groups and allowing focused analysis of these subsets of the metabolome (Zhao and Li 2020). Derivatization also has potential disadvantages: (1) derivatization of unknowns may not necessarily result in easier characterization of their structure, and (2) derivatization reactions may not always proceed to completion, resulting in a mixture of partially derivatized molecules and hence a more complex sample to analyze, and (3) some metabolites may not derivatize at all.

Stable Isotopes Compounds enriched with one or more stable isotopes of common elements (e.g., ^{2}H,^{13}C, ^{15}N, or ^{18}O) can be easily distinguished from their unlabeled counterparts by mass spectrometry, while their chemical properties remain essentially identical to the unlabeled compound. In addition to their applications as internal standards and as tracers to help quantitate metabolite flux, stable isotopes can also be used to aid compound identification in metabolomics. One approach using stable isotopes is termed isotope ratio outlier analysis (IROA), which uses paired samples isotopically labeled with either 5% or 95% ^{13}C and uses characteristic isotope patterns to help differentiate biological signals from artifacts and constrain potential molecular formulas assigned to unknowns (Clendinen et al. 2015). Another application of stable isotopes is to investigate the metabolic fate of individual metabolites or drugs (Kempa, “Advancements in pulsed stable isotope re-solved.

Metabolomics”; Günther, “Metabolomics in cell biology”). In one example, Chen et al. injected mice with 400 mg/kg of [acetyl-$^{2}H_3$]- or [2,3,5,6-$^{2}H_4$]-acetaminophen and used untargeted feature detection, followed by MS/MS-based structural elucidation, to identify three novel urinary acetaminophen metabolites potentially associated with its toxicity in overdose (Beyoğlu et al. 2018). Another potential use of stable isotopes is the use of hydrogen-deuterium exchange to help determine the chemical structure of unknown compounds (Majuta et al. 2019).

Chromatography Chromatographic methods for metabolomics are often optimized for quantitation and speed but can also be tailored to improve compound identification. One important choice is whether to use GC, LC, or both methods to characterize a sample (Gowda and Djukovic 2014). GC offers fast and reproducible chromatography and can be coupled to both electron impact (EI) and chemical ionization (CI) methods to obtain complementary fragmentation and molecular ion data, both of which contribute to compound identification (Misra and Olivier 2020), but samples typically require derivatization as described above. LC with electrospray ionization is more versatile for larger molecules or non-derivatizable species, and high-resolution accurate mass LC–MS/MS systems are more common than GC–MS systems with similar capabilities (Shackleton et al. 2018; Theodoridis et al. 2012). In both GC–MS and LC–MS methods, chromatographic conditions should be

evaluated in terms of their potential impact on compound identification, in addition to considerations regarding quantitation and throughput (Rainville et al. 2017). Combined optimization of gradient length, sample loading, and use of iterative precursor ion exclusion for LC–MS/MS analysis of the human plasma metabolome was determined to generate over a 10-fold improvement of probable unique compound identifications, including numerous trace-level drug compounds, compared to typical high-throughput methods (Anderson et al. 2021). Multi-hour gradients (Wang et al. 2015) and extreme column loading are not practical for quantitative analysis of large numbers of samples; however, features identified using such high-resolution methods can be aligned with features detected but not identified using faster run conditions (Habra et al. 2021).

Authentic Standard Libraries and Retention Time Databases As in targeted metabolomics, authentic standards can be used to help unambiguously identify features detected in untargeted metabolomics data by matching m/z, RT, and MS/MS spectrum to features detected in biological samples. Effective use of authentic standard libraries reduces the burden of unknown identification for the remainder of the features. An authentic standard library can be assembled by purchasing individual chemical standards from vendors or by purchasing pre-assembled kits or libraries. Of relevance to pharmacological research, several libraries containing hundreds to tens of thousands of known bioactive, drug-like, or drug candidate molecules can be obtained through commercial suppliers and/or government-supported repositories (Health in Northern Ireland 2022). Disadvantages of authentic standard libraries include the initial expense of purchasing, solvating, and/or analyzing the chemical standards, as well as the fact that a library may not be representative of a biological sample's constituents (Mahieu and Patti 2017). An alternative to in-house standard libraries are digital databases containing retention time or retention index information compiled on a specific analytical method. These provide some of the benefits of standard libraries without the need to purchase and analyze all the standards experimentally. One example is the FiehnLib GC–MS method, which uses retention indices to aid in compound identification (Kind and Fiehn 2010). Some online metabolite databases also contain RT information for LC methods, but inter-laboratory use of these data is less frequent as retention time alignment or indexing across labs is viewed as more challenging than for GC–MS. Spectral databases are described in more detail in "Step 3" of this chapter.

Mass Spectrometry: High Resolution, Accurate Mass Measurement While targeted metabolomics is dominated by single-quadrupole GC–MS and tandem-quadrupole LC–MS, untargeted analysis is more optimally performed using high-resolution accurate mass (HRAM) analyzers, including quadrupole-time of flight (QToF), Orbital Ion Trap (Orbitrap), and Fourier-Transform Ion Cyclotron Resonance (FT-ICR) instruments. Measurements performed on a well-calibrated QToF typically have a mass accuracy in the range of 3–5 ppm when internal calibration is used; Orbitrap instruments may achieve 0.5–1 ppm mass accuracy, whereas FT-ICR analyzers may achieve between 0.1 and 1 ppm (Balogh 2004; Fiehn 2016b). A

principal advantage of higher mass accuracy is that it reduces the number of possible molecular formulas that can be assigned to a feature; although, as demonstrated by Fiehn et al. even 0.1 ppm mass accuracy is not sufficient to unambiguously assign a molecular formula for species <500 Da, much less for larger compounds (Kind and Fiehn 2007). Spectral resolution does not directly determine the accuracy of molecular formula determination. Instead, higher resolution reduces the possibility of interference from co-eluting nearly isobaric species. The isotope distribution of an unknown is another feature of MS data that aids compound identification; if measured accurately, it is very effective in constraining molecular formula assignment (Böcker et al. 2009).

Mass Spectrometry: In-Source Fragmentation In GC–MS studies, electron impact ionization is performed at a standardized energy of 70 eV, which causes extensive in-source fragmentation in a manner that is reproducible from instrument to instrument (Taylor 2015). The fragmentation pattern reveals structural information about the compound and allows library searching to facilitate identification (as described in "Step 4" of this chapter). Co-eluting compounds may complicate interpretation by generating mass spectra that contain fragment ions from two or more compounds; one strategy to compensate for this is spectral deconvolution as implemented in NIST AMDIS software (Davies 1998) or GC–MS/MS analysis (Kvitvang et al. 2011). In-source fragmentation also occurs to a lesser extent in LC–ESI–MS but is usually considered undesirable and source settings are chosen to minimize it, although it is occasionally used to perform pseudo-MS/MS or pseudo-MS^3 for instruments lacking those capabilities. (Xue et al. 2020; Abdelhameed et al. 2014)

Mass Spectrometry: MS/MS and MS^n In LC–ESI–MS, tandem mass spectrometry (MS/MS or MS^n) is an important aspect of data collection in untargeted metabolomics studies, especially in the context of compound identification. MS/MS has also seen increasing use in GC–MS, though mainly for targeted analysis. Tandem MS involves selection and fragmentation of specific ions within the mass spectrometer; fragment ion spectra can be interpreted to discern information regarding the precursor ion's structure or used as a "fingerprint" for database searching. Multiple methods of fragmentation are available, depending on instrument capabilities. These include collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), electron transfer dissociation (ETD) and infrared multi-photon dissociation (IRMPD) (Ichou et al. 2014; Alley et al. 2009; Yoo et al. 2007). Recently, a strategy to obtain more extensive fragmentation of ESI-generated ions that uses an electron impact-type mechanism has been devised; one mode of operation has been termed electron impact excitation of ions from organics (EIEIO) (Baba et al. 2018; Ducati et al. 2021). There are two main methods that can be used to perform tandem MS data acquisition: data-dependent analysis (DDA), in which MS/MS is automatically triggered for the most abundant or otherwise selected features detected in the preceding MS1 scan, and data-independent analysis (DIA), in which MS/MS acquisition is performed according to pre-defined criteria not

influenced by the MS1 data. Details regarding these approaches and their advantages and disadvantages have been described in detail elsewhere (Guo and Huan 2020a, b).

It is good practice to acquire MS/MS data during any substantial untargeted metabolomics project. This allows both automated database searching and follow-up analysis to attempt identification or classification of features of interest, without returning to the instrument to acquire new data. In many cases, MS/MS data need not be acquired for each individual sample, but can be generated using representative pooled samples from the study. To acquire high-quality MS/MS spectra of as many features as possible, recent data acquisition software packages enable automatic generation of precursor ion exclusion lists that prevent serial LC–MS/MS runs of the same sample from acquiring MS/MS data on ions already fragmented in previous runs. This method, sometimes termed iterative DDA, is a powerful strategy to obtain deeper MS/MS coverage of detected features than is possible in a standard DDA workflow (Anderson et al. 2021; Koelmel et al. 2017).

Certain mass analyzers are capable of multiple stages of fragmentation, in which fragment ions are further fragmented. This is termed MS^n analysis; it enables construction of fragmentation trees for unknown compounds (Vaniya and Fiehn 2015). MS^n data are less well cataloged and interpretation of MS^n data is less widely supported by databases and software tools compared to MS/MS data. Nevertheless, under many circumstances MS^n data provide useful information to characterize and annotate unknown compounds (Vinaixa et al. 2016; Ridder et al. 2012).

Ion Mobility Spectrometry and Collisional Cross-Section Measurement Ion mobility spectrometry (IMS) separates gas phase ions based on differential mobility through low-pressure buffer gas (Kanu et al. 2008). Ions with a larger cross-sectional area experience more resistance and move more slowly than smaller, more compact ions. In the context of metabolomics, ion mobility is performed inside a mass spectrometer directly preceding the standard mass analyzer(s). It offers a semi-orthogonal separation to MS and resolves some isobaric species that are not separable by LC–MS, allowing generation of MS/MS spectra from a single precursor ion rather than from multiple co-eluting isobars (Rainville et al. 2017). With appropriate calibration, IMS data can also be used to compute collisional cross-section values (CCS), which are considered an intrinsic property of an ion and can be used to facilitate compound identification (Zhou et al. 2020). Various manufacturers have produced instruments with ion mobility spectrometry capabilities, and collisional cross-section values are beginning to be included with metabolite databases (Wishart et al. 2022). CCS measurements have been evaluated as a means of characterizing drug compound structure (Hines et al. 2017) and to predict their ability to cross the blood-brain barrier (Guntner et al. 2019), demonstrating the potential pharmacological relevance of this rapidly developing technology.

Nuclear Magnetic Resonance Spectroscopy for Metabolite Identification Although this chapter focuses on mass spectrometry-based methods, it is important to acknowledge the utility of nuclear magnetic resonance

spectroscopy (NMR)-based methods in identifying unknown compounds. NMR is an important and widely used technique for metabolomics and pharmacometabolomics in its own right; its strengths, limitations, and applications are thoroughly described in the literature (Emwas et al. 2019) and in other chapters of this book. In the context of compound identification, NMR is a touchstone method for the fields of synthetic chemistry, natural products chemistry, and drug development. For a pure sample of a small molecule compound analyzed in a modern high-field instrument using ^{1}H and ^{13}C NMR, it is typically possible to assign a definitive structure via computational modeling and/or manual interpretation (Willoughby et al. 2014). The primary challenge associated with applying the NMR to compound identification in metabolomics is its sensitivity. For higher-concentration compounds, NMR is effective at both quantification and identification, and detailed guides describing compound identification strategies in NMR metabolomics have been written (Dona et al. 2016). However, compounds present at lower concentrations (mid-low micromolar and below) are not amenable to identification without purification and concentration. Techniques to purify unknowns using chromatographic fractionation have been devised (van der Laan et al. 2021; Whiley et al. 2019) but are still not practical for lower-abundance features. While scale-up to semi-preparative or preparative chromatography or other refinements are possible, it remains challenging to obtain high-quality NMR spectra of a substantial portion of features detectable by mass spectrometry.

4 Step 3: Computational Strategies for Data Cleaning and Feature Annotation

Acquisition of high-quality data is not sufficient to identify unknown metabolites. Datasets are far too large to manually review, much less interpret, every spectrum collected in an experiment. Fortunately, numerous computational strategies can be employed to aid in identifying or annotating unknowns; when necessary, they can also help guide acquisition of additional data. Here, we present an overview of major computational strategies to aid the reader in finding and understanding available tools. Readers interested in a more detailed discussion of computational compound identification strategies are directed to one of several excellent reviews that have been published on the topic (Wishart 2009; Watson 2013; Blaženović et al. 2018) or to primary sources cited below.

Degeneracy Removal In the simplest scenario, each compound present in a sample would be represented by one feature in experimentally acquired untargeted metabolomics data. However, this is seldom the case. In GC-EI-MS, extensive in-source fragmentation is expected for each compound, but when peaks co-elute, it can be difficult to determine which fragment ions originate from which parent ion. In LC-ESI-MS, using positive ion mode as an example, singly-protonated $[M + H]^+$ ions are usually the most common form of ion, but other adducts such as $[M + Na]^+$, $[M + NH4]^+$, etc. or in-source fragment ions such as $[M\text{-}H2O + H]^+$ are often formed

and detected. More complex ion formation processes are frequently observed, including multiply charged ions, ions with multiple charge carriers, solvent adducts, adducted fragment ions, and dimers, multimers and heterodimers. (Mahieu and Patti 2017; Mahieu et al. 2016; Kachman et al. 2019; Nelson et al. 2022) Collectively, such features can be termed “degenerate” signals.

To avoid investing effort to identify features that are degenerate signals produced by other already-identified compounds, data clean-up steps including adduct and fragment annotation and/or removal are necessary. For GC-EI-MS, spectral deconvolution software such as AMDIS (Davies 1998) or MS-DIAL (Tsugawa et al. 2015) can help separate spectra of features that are only partially resolved by chromatography. For LC–MS, the process of adduct annotation can be at least partially automated by either instrument vendor or open-source data analysis software such as CAMERA, MZmine, or MS-DIAL (Tsugawa et al. 2015; Kuhl et al. 2012; Pluskal et al. 2010). The most rigorous approaches for degeneracy annotation also consider correlation of intensity for co-eluting features; those with high correlation are more likely to represent degenerate features (Kachman et al. 2019; Broeckling et al. 2014). Application of these clean-up steps reduces compound identification workload by decreasing the number of features that must be subjected to additional computational analysis and/or manual review.

Molecular Formula Assignment A fundamental step in identifying a compound, whenever possible, is to determine its molecular formula. The approach for this process differs depending on whether data are acquired on a low-resolution instrument, such as a GC–MS with a quadrupole mass analyzer, or a high-resolution accurate mass instrument, such as a QToF, orbital ion trap, or ion cyclotron resonance instrument. In the former case, molecular formula assignment is performed using a probability-based strategy (Scott 1992) such as that implemented in NIST MS Search software (Stein 1999). In the latter case, an effective strategy for molecular formula assignment from accurate mass data was developed by Kind et al. (Kind and Fiehn 2007) It uses “seven golden rules” developed to constrain potential chemical formulas based on characteristics shared by almost all common biological molecules containing some or all of the elements C, H, O, N, P, and S. In addition to the accurate m/z value measured for a compound, the natural isotope distribution of the compound is required to sufficiently constrain candidate formulas for all but the smallest metabolites. Fortunately, software tools to assign molecular formulas automatically using metabolomics data have been devised (Tolić et al. 2017; Ludwig et al. 2019; Dührkop et al. 2019) and are implemented in various instrument vendor software packages. These molecular formula assignment methods have also proven highly applicable to pharmaceuticals, returning a correct molecular formula as the top candidate with 88–99% probability for thousands of database spectra of drug molecules (Kind and Fiehn 2007). While these algorithms have become increasingly reliable, de-novo assignment of formula for larger compounds (MW $>$500) remains challenging and are best supplemented by support from compound database and spectral search.

Table 1 Selected small molecule compound databases. Table adapted from Blaženović et al. (2018)

Database	Contents	Additional compound ID-relevant features	Free web access/free download?
PubChem (Bolton et al. 2008)	All small molecules	Structure similarity search tool	Y/Y
ChemSpider (Pence and Williams 2010)	All small molecules		Y/N
ChEBI (Hastings et al. 2016)	Small molecules	Focus on compounds of biological interest	Y/ Y
KEGG (Kanehisa et al. 2006)	Metabolites	Curated pathway maps	Y/N
MetaCyc (Caspi et al. 2008)	Metabolites	Curated pathway maps	Y/ noncommercial
HMDB (Wishart et al. 2022)	Human metabolites/ exposome	Physiological concentration data, MS/MS spectra, text-mined literature context	Y/Y
Metlin (Guijas et al. 2018)	Metabolites		Y (MS1 only)/ N
RefMet (Fahy and Subramaniam 2020)	Metabolites	Name conversion tool	Y/Y
ChEMBL (Davies et al. 2015)	Bioactive drug-like molecules		Y/Y
DrugBank (Wishart et al. 2006)	Known drugs		Y/ noncommercial

Compound Databases Once degeneracy removal has been performed, and preferably after a molecular formula has been assigned, a feature's m/z or neutral mass can be searched against a compound database. Although m/z alone is by no means sufficient to identify an unknown metabolite, databases help provide a list of candidate compounds that can be evaluated using other means. Small molecule compound databases vary in their scope and application; major examples are listed in Table 1.

Databases range from those which focus on known endogenous metabolites to those which attempt to cover all known and plausible small molecule chemical structures. It is desirable to begin by querying the most specific applicable database first. Small molecule databases focused on pharmaceutical compounds, such as DrugBank (Wishart et al. 2018) and ChEMBL (Papadatos and Overington 2014; Mendez et al. 2019), are of particular use for putative annotation of drugs and their metabolites in pharmacometabolomics data. Likewise, using organism-specific databases or constraining taxonomy to the sample type being analyzed will generate

a narrower list of candidate compounds than an open-ended search against PubChem. Annotation of unusual or novel compounds may require use of broader databases.

Regardless of the database used, database hits against unknowns should not be treated or reported as confident identifications, even if the hit list contains only a single compound. Additional confirmation from other data sources is required, as described in "Steps 4 and 5" in this chapter.

Retention Time Prediction As already noted, some compound databases contain retention time information for specific compounds collected using defined chromatographic methods. However, in the frequent event that no experimentally determined RT information is available for these compounds, retention time prediction strategies and software tools can help fill the gap (Bonini et al. 2020; Stanstrup et al. 2015). These approaches require training a computational model using predicted chemical properties by inputting retention times and structures of known compounds. Once the model is complete, retention times can be predicted for candidate spectral matches and used to help confirm or refute compound assignments made by MS/MS or other data. No retention time prediction software yields precision comparable to experimental data; often, predicted retention times deviate 10% or more from the experimental value. Nevertheless, in many cases, this level of precision is enough to rule out a substantial portion of incorrect identifications.

Collisional Cross-Section Prediction Analogous to RT prediction, collisional cross-section values can be predicted based on compound structure and machine learning (Plante et al. 2019) or quantum chemistry-based (Colby et al. 2019) models. These values can then be compared against values experimentally determined for unknown features, which are beginning to appear in major databases like HMDB and Metlin. The desired result is either confirmation or refutation of the assigned compound identity. Computational collisional cross-section prediction is in the early phases of development but promises to impact compound identification as adoption of ion mobility spectrometry increases.

5 Step 4: MS/MS Libraries and Compound Identification Using Library Search

To move beyond candidate screening to true compound identification, the first-line strategy is to search the MS and/or MS/MS spectrum of the unknown against a spectral library. Mass spectral libraries contain experimentally collected or computationally predicted spectra of a database of small molecule compounds. For GC–MS, the most extensive libraries consist of EI-MS spectra acquired at a standard 70 eV ionization energy. For LC-ESI, databases often consist of MS/MS spectra that have been acquired for authentic standards on several different instrument types at a range of collision energy values. Certain libraries may focus on a particular class of

Table 2 Widely used spectral libraries. Table adapted from Blaženović et al. (2018)

Library	Data type	Type	Additional compound ID-relevant features	Free web access/free download?
NIST (Stein 1999)	EI-MS, CID-MS/MS	Experimental	Highly curated, includes search software, available from multiple vendors	N/N
Wiley (Solutions 2022)	EI-MS, CID-MS/MS	Experimental	Largest collection of EI-MS data, available from multiple vendors	N/N
METLIN (Guijas et al. 2018)	CID-MS/MS	Experimental	Developed for Q-ToF instruments, licensed annually	N/N
MoNA (Fiehn 2016b)	EI, MS/MS, MS^n	Experimental, user-contributed	Community database, automated curation	Y/Y
MassBank (Horai et al. 2010)	Metabolites	Experimental, user-contributed	Community database	Y/Y
mzCloud (LLC, H 2022)	MS/MS, MS^n	Experimental	Most complete MS^n database	Y/N
GNPS (Wang et al. 2016)	MS/MS	Experimental, user-contributed	Integrated with molecular networking tools	Y/Y
LipidBlast (Kind et al. 2013)	Bioactive drug-like molecules	Computational	Fully computational lipidomics database	Y/Y

molecule, be acquired on a specific instrument, or contain MS^n spectra (where $n > 2$). Several major spectral libraries exist (Table 2). The National Institutes of Standards and Testing (NIST) EI-MS and ESI-MS/MS databases contain carefully curated, periodically updated libraries of spectra of authentic standards; this and several other libraries are available from commercial distributors. Other libraries are public; some of these are derived from "crowdsourced" compilations of spectra and their assigned chemical structures, which typically receive community review to assess quality. Still other spectral libraries are generated by computationally predicting fragmentation of a list of chemical structures from a database, yielding an "in silico" spectral library.

To perform a library search, the mass spectrum of an unknown feature is computationally compared to entries in the spectral library. After an initial candidate screening step, each potential spectral match is ranked by a scoring function that generates a match score reflecting the similarity between the spectrum of the unknown compound and the library spectrum. Score ranges differ depending on the function that is used, but in general a higher score corresponds to a closer spectral match. The chemical structure of the compound associated with the highest-scoring hit is assigned as the most probable identification; additional lower-scoring

hits may be retained for user review. Many different scoring functions exist; among these, probability-based matching (PBM) was one of the first introduced scoring functions for EI-MS spectra (Stauffer et al. 1985), and the classic "dot-product" and "reverse dot-product" scoring algorithms have been widely used for MS/MS searching (Stein and Scott 1994). Many other scoring functions have also been evaluated, including a recent spectral entropy-based algorithm that demonstrated superior performance to dot-product scores (Li et al. 2021). Additional constraints, such as a narrow mass window for precursor ion match and a precursor ion isotope pattern match, can serve to further shorten a list of candidates or even select a single probable compound identity.

Many software tools exist to automate spectral library search. Most instrument manufacturers incorporate library search tools into their data analysis software; some also provide access to proprietary libraries searchable only using these tools. Among cross-platform tools, NIST MS Search is one of the best known (Stein 1999); it allows searching individual spectra against the NIST library and other user-loaded libraries with visualization of spectral matches using head-to-tail or difference plots. To enable much more rapid search of many spectra, a useful alternative is MSPepSearch (Zhang et al. 2018). Originally designed for peptide spectra, it has been adopted and extensively used for small molecule analysis; it uses a similar scoring algorithm but generates output in tabular form. Progenesis QI is a commercial data analysis tool that enables cross-vendor MS/MS search in addition to general data analysis. MS-DIAL is a free, open-source alternative that integrates feature finding, alignment, and spectral search in a unified workflow (Tsugawa et al. 2015). Other widely used untargeted metabolomics data analysis tools, including XCMS (Smith et al. 2006) and MZmine (Pluskal et al. 2010), also have some MS/MS search capabilities, though they are implemented in a less visual manner than in MS-DIAL. SIRIUS (Böcker et al. 2009; Dührkop et al. 2019) and GNPS (Wang et al. 2016), described later in this chapter, also use spectral search heavily in their workflow but due to other features are classified separately from typical library search tools.

It is important to note that while spectral searching provides a fast and often accurate means of assigning chemical structures to features in metabolomics data, it does not provide an objective means to assess the probability that these identifications are correct. Manual review of spectra, including matches beyond the top hit, can sometimes help clarify ambiguous assignments and determine appropriate score thresholds for identification. Well-defined strategies for assessing and reporting identification confidence are important to data analysis and are described in "Step 5" of our workflow.

Spectral Similarity Searching Searching an unknown feature against a library for a precise match is not likely to produce informative results when the compound in question is not in the library. As an alternative, it is possible to search for spectra that are not a direct match to the unknown but share some of its features. This approach is termed similarity searching and can be performed using several software tools. One of the most robust is "hybrid search," implemented in NIST MS Search and MSPepSearch (Cooper et al. 2019). This search method allows both direct peak

matching (as in identity searching) as well as matching of masses shifted by a neutral loss within a single spectrum. This accounts for cases in which the m/z of some fragments of an unknown are identical to those of a library compound, while others are shifted by a structural difference typically confined to a single region of the molecule. A second similarity search strategy is implemented in the software tool DeepMASS (Tiwary et al. 2019), which uses a machine deep-learning approach to perform structural similarity scoring. All similarity searches aim to generate matches that are likely to be structurally related to the unknown compound. The resulting matches can be used for compound class assignment as well as to help elucidate the structure of the unknown, with the aid of other techniques described below.

In Silico Libraries and In Silico Fragmentation Software To generate more extensive spectral libraries than is possible using authentic standards, several methods have been used to generate in silico spectra libraries. These predict how compounds will fragment, using techniques ranging from quantum-chemistry-based calculations (Wang et al. 2020) to rule-based methods (Tsugawa et al. 2016), and generate a theoretical fragmentation spectrum for each compound. Each strategy has advantages and limitations; these have been described and reviewed in detail elsewhere (Borges et al. 2021). In silico libraries are most widely used in lipidomics. The LipidBlast MS/MS spectral library enables identification of thousands of lipid species, and stands out in its accuracy and widespread use, because MS/MS fragmentation patterns are typically reproducible for all lipids within a lipid class (Kind et al. 2013). In silico spectral libraries of other small molecule classes are also available, but due to limited fragmentation, the high diversity of chemical structures, and the difficulty of predicting relative abundance of fragment ions, most non-lipid in silico spectral databases are of limited utility for automated spectral search.

An alternative strategy is to use the technique of in silico fragmentation to help interpret spectra of unknown compounds. Many software packages exist to help perform this task; among the most prominent are MetFrag (Ruttkies et al. 2016) and CSI:FingerID (Dührkop et al. 2015), which is now implemented in the SIRIUS data analysis package (Dührkop et al. 2019). To interpret an unknown spectrum, a user inputs information into the software regarding the unknown compound, including the observed precursor ion, adduct type or molecular formula if known, the obtained MS/MS spectrum of the unknown, and in some cases, additional data such as the isotope distribution of the precursor ion and the taxonomy of the organism from which the sample that generated the spectrum was obtained. The software then selects candidate compounds from a large compound database (PubChem or similar) that matches the precursor ion mass and any other metadata and then predicts fragments that would be formed from these precursors using rule-based, machine-learning, fragmentation tree generation, or other strategies. The predicted fragments are matched against the unknown spectrum, and the data are used to predict the structure, or at least key structural elements, of the unknown. In silico fragmentation and structure prediction does not always yield a definite compound identification for an unknown but nevertheless is one of the simplest strategies to help predict structure based on spectral data.

Spectral Networking Analysis Another technique that has gained prominence as a means of interpreting MS/MS spectra in untargeted metabolomics data is spectral networking analysis. The most widely used software tool in this area is the Global Natural Products Database (Wang et al. 2016). The MS/MS spectra of identified and unknown features in a dataset are searched against each other, in addition to against library spectra. The most similar spectra are grouped, and clustering methods are used to generate spectral networks that can be examined visually. Since the most strongly associated spectra are grouped together, neighboring features on the network may represent similar molecules with small structural differences. By calculating the precise mass shift difference between precursor ions, a molecular formula "difference" can often be assigned, which can help highlight functional groups that differentiate the compounds. If one of the two features is identified with high confidence, it is sometimes possible to predict the structure of the unknown feature from these data. Even when this is not possible, the networking analysis can reveal useful information about the structure of the unknowns.

6 Step 5: Assess and Report Identification Confidence

As evidenced by the range of techniques described in this chapter, compound identification in metabolomics ranges from routine to extremely challenging. Likewise, the evidence supporting an identification can range from unequivocal to uncertain. In the interest of scientific transparency, it is important that researchers report not only the assigned identity of a compound but also the method and data used to make identifications and an estimate of their confidence that the identification is correct at any level of information (structure, formula, compound class). As illustrated in Fig. 2, currently recommended methods for reporting compound identifications involve semi-subjective classification by the analyst, while methods under development have potential to improve accuracy, speed, and transparency of the compound identification and reporting process.

Reporting Identification Rigor Using Identification Levels The importance of data-reporting standards has long been clear to the metabolomics community. In 2007 an international consortium termed the Metabolomics Standards Initiative (MSI) published a set of recommended minimum data-reporting standards, which included four "identification levels." (Sumner et al. 2007) Level 1 represents the highest degree of rigor for compound identification, in which compound identity is established at the chemical structure level by matching at least two orthogonal forms of experimental data to reference data collected in the researcher's own laboratory using an authentic standard. These could include, for instance, accurate mass and retention time, or retention time and MS/MS spectrum. Level 2 is a putative annotation at the structure level based on data collected outside the researcher's laboratory, such as an MS/MS spectral match with a library, or a literature-based retention time or retention index. Level 3 signifies a putative compound class assignment based on spectral similarity searching and/or physiochemical property

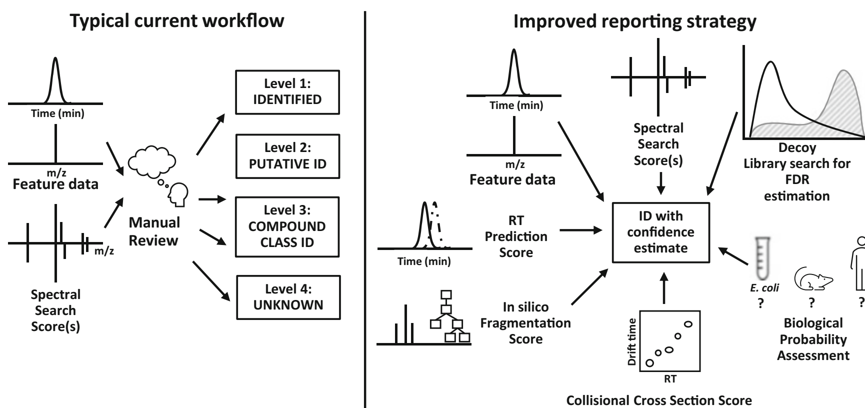


Fig. 2 Current typical current metabolite identification and reporting strategy compared to a hypothetical multi-input method for improving identification accuracy and assessing and reporting identification confidence

assessment. Level 4 represents unknown compounds. Further refinements to compound identification levels have been proposed and adopted by various groups or for particular applications (Schymanski et al. 2014). When applied, the MSI compound identification levels have served the community by allowing scientists to communicate about identifications using consistent terminology. One challenge is that MSI levels (or their equivalent) have not been universally adopted in the literature; many publications report metabolite identifications with minimal information about how they were made. Another key limitation is that acceptance or rejection of an identification at any identification level is, in large part, at the discretion of the analyst. For instance, no minimum spectral library search score is given for acceptance of a level 2 identification, nor is it easily possible to establish one since an appropriate score threshold may vary from one study or spectral library to the next. Manual review of compound identifications can help confirm correct and refute incorrect identifications, but this approach is too time consuming for experiments with tens of thousands of compounds detected.

False Discovery Estimation Using Decoy Metabolite Libraries An objective and automatable approach to assess and report compound identification confidence in the form of an estimated false discovery rate (FDR) would be useful to improve consistency and inter-lab comparability of metabolomics data (Scheubert et al. 2017). In proteomics, FDR estimation is performed by searching peptide MS/MS spectra against both normal and “decoy” in silico spectral libraries, the latter of which is generated by scrambling the amino acid sequence of all proteins in the library (Elias and Gygi 2010). By comparing the number of hits to the decoy library relative to the true library, an FDR can be estimated and reported with the data. Although no direct equivalent to amino acid sequence scrambling exists in metabolomics, several approaches for generation of decoy libraries have been

proposed and tested for small molecules, ranging from randomizing all peaks found in a typical library to approaches based on rearrangement of fragmentation trees (Scheubert et al. 2017; Wang et al. 2018; Alka et al. 2022). Decoy libraries have been used to help select spectral search score thresholds and other criteria to achieve a desired FDR. However, these attempts have demonstrated that decoy libraries are not very helpful to assess FDR for some classes of metabolites, particularly for compounds that produce very few fragment ions and those that have multiple structural isomers that produce similar fragmentation spectra. Thus, FDR estimation strategies for metabolomics need further refinement and will certainly require supplementation with strategies beyond decoy library searching before compound identification in metabolomics can achieve the level of automation and accuracy enjoyed by other omics sciences.

Integrating Multiple Strategies for Confident Compound Identification In this chapter, we have described a variety of experimental and computational strategies to help annotate and identify features in metabolomics data. Using present technologies, it is sometimes possible to arrive at a single correct structure-level identification for features of interest in the data, while in other cases, more limited annotation is all that is possible. Currently, integrating available information from the methods and tools we have described is the task of the analyst, who uses their judgment to set thresholds, finalize identifications and report methods used. Moving forward, a priority for the metabolomics community is to develop a strategy to integrate all available information in a consistent and automated manner to make compound identification assignments, with support from objective data that estimate identification confidence. A compound identification "meta-analysis" approach may prove useful (Fig. 2), but no comprehensive strategy has yet been devised. For the time being, compound identification in metabolomics is an exercise in defining aims appropriately, selecting methods and collecting sufficient relevant data, and applying available tools to help with interpretation. As the metabolomics research community repeats and refines this process, it continues to work toward more unifying methods for compound identification.

7 Conclusion: Pharmacology-Focused Compound Identification

As described throughout the chapter, considerations relevant to identification of small molecules in metabolomics data are, in general, fully applicable to pharmacology-focused metabolomics studies. A key reminder is to tailor strategies to study design. When only specific drug-derived or endogenous metabolites are of interest, a suitable targeted metabolomics workflow reduces the burden of assigning compound identities to the large number of features that would be detected in an untargeted study design. When untargeted analysis is desired, selection of a compound database or library to focus on the organism being studied or a drug class of interest can yield more meaningful results than when a broad, nonspecific database is

used. Researchers should also consider biological factors that may affect both compound identification and quantitation, such as sex-specific responses to drug metabolism (Soldin et al. 2011; Chary et al. 2022). Finally, the application of metabolomics to pharmacological studies is still a developing area of study. While at present most of the data relevant to compound identification in metabolomics is found in the analytical chemistry and bioinformatics literature, resources specific to the challenges of pharmacology can be expected to develop and strengthen over future years.

References

Abdelhameed A et al (2014) Pseudo-MS3 approach using electrospray mass spectrometry (ESI-MS/MS) to characterize certain (2E)-2-[3-(1H-Imidazol-1-yl)-1-phenylpropylidene] hydrazinecarboxamide derivatives. J Chem 10 p. https://doi.org/10.1155/2014/386301

Alka O et al (2022) DIAMetAlyzer allows automated false-discovery rate-controlled analysis for data-independent acquisition in metabolomics. Nat Commun 13:1347. https://doi.org/10.1038/s41467-022-29006-z

Alley WR, Mechref Y, Novotny MV (2009) Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. Rapid Commun Mass Spectrom 23:161–170. https://doi.org/10.1002/rcm.3850

Anderson BG, Raskind A, Habra H, Kennedy RT, Evans CR (2021) Modifying chromatography conditions for improved unknown feature identification in untargeted metabolomics. Anal Chem 93:15840–15849. https://doi.org/10.1021/acs.analchem.1c02149

Baba T, Campbell J, Le Blanc Y, Baker P, Ikeda K (2018) Quantitative structural multi-class lipidomics using differential mobility-electron impact excitation of ions from organics (EIEIO) mass spectrometry. J Lipid Res 59:jlr.D083261. https://doi.org/10.1194/jlr.D083261

Balogh MP (2004) Debating resolution and mass accuracy in mass spectrometry. Spectroscopy (Santa Monica) 19:34–39

Beyoğlu D, Zhou Y, Chen C, Idle JR (2018) Mass isotopomer-guided decluttering of metabolomic data to visualize endogenous biomarkers of drug toxicity. Biochem Pharmacol 156:491–500. https://doi.org/10.1016/j.bcp.2018.09.022

Blaženović I, Kind T, Ji J, Fiehn O (2018) Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites 8. https://doi.org/10.3390/metabo8020031

Böcker S, Letzel MC, Lipták Z, Pervukhin A (2009) SIRIUS: decomposing isotope patterns for metabolite identification. Bioinformatics 25:218–224

Bolton EE, Wang Y, Thiessen PA, Bryant SH (2008) Annual reports in computational chemistry, vol 4. Elsevier, pp 217–241

Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O (2020) Retip: retention time prediction for compound annotation in untargeted metabolomics. Anal Chem 92:7515–7522. https://doi.org/10.1021/acs.analchem.9b05765

Borges RM et al (2021) Quantum chemistry calculations for metabolomics. Chem Rev 121:5633–5670. https://doi.org/10.1021/acs.chemrev.0c00901

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE (2014) RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem 86:6812–6817. https://doi.org/10.1021/ac501530d

Bruce SJ et al (2009) Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. Anal Chem 81:3285–3296. https://doi.org/10.1021/ac8024569

Cajka T, Fiehn O (2016) Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and Lipidomics. Anal Chem 88:524–545. https://doi.org/10.1021/acs.analchem.5b04491

Caspi R et al (2008) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 36:D623–D631. https://doi.org/10.1093/nar/gkm900

Chaleckis R, Meister I, Zhang P, Wheelock CE (2019) Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. Curr Opin Biotechnol 55:44–50. https://doi.org/10.1016/j.copbio.2018.07.010

Chary S, Amrein K, Mahmoud SH, Lasky-Su JA, Christopher KB (2022) Sex-specific catabolic metabolism alterations in the critically ill following high dose vitamin D. Metabolites 12. https://doi.org/10.3390/metabo12030207

Clendinen CS et al (2015) An overview of methods using 13C for improved compound identification in metabolomics and natural products. Front Plant Sci 6. https://doi.org/10.3389/fpls.2015.00611

Colby SM et al (2019) ISiCLE: a quantum chemistry pipeline for establishing in silico collision cross section libraries. Anal Chem 91:4346–4356. https://doi.org/10.1021/acs.analchem.8b04567

Contrepois K et al (2018) Cross-platform comparison of untargeted and targeted Lipidomics approaches on aging mouse plasma. Sci Rep 8:17747. https://doi.org/10.1038/s41598-018-35807-4

Cooper BT et al (2019) Hybrid search: a method for identifying metabolites absent from tandem mass spectrometry libraries. Anal Chem 91:13924–13932. https://doi.org/10.1021/acs.analchem.9b03415

Davies AN (1998) The new Automated Mass Spectrometry Deconvolution and Identification System (AMDIS). Spectrosc Eur 10(3):22–26

Davies M et al (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res 43:W612–W620. https://doi.org/10.1093/nar/gkv352

Dona AC et al (2016) A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Comput Struct Biotechnol J 14:135–153. https://doi.org/10.1016/j.csbj.2016.02.005

Ducati AO et al (2021) Improved metabolite characterization by liquid chromatography – Tandem mass spectrometry through electron impact type fragments from adduct ions. Anal Chim Acta 1150:338207. https://doi.org/10.1016/j.aca.2021.338207

Dührkop K, Shen H, Meusel M, Rousu J, Böcker S (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci 112:12580–12585. https://doi.org/10.1073/pnas.1509788112

Dührkop K et al (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods 16:299–302. https://doi.org/10.1038/s41592-019-0344-8

Elias JE, Gygi SP (2010) Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol Biol 604:55–71. https://doi.org/10.1007/978-1-60761-444-9_5

Emwas A-H et al (2019) NMR spectroscopy for metabolomics research. Metabolites 9. https://doi.org/10.3390/metabo9070123

Fahy E, Subramaniam S (2020) RefMet: a reference nomenclature for metabolomics. Nat Methods 17:1173–1174. https://doi.org/10.1038/s41592-020-01009-y

Fiehn O (2016a) Metabolomics by gas chromatography-mass spectrometry: combined targeted and untargeted profiling. Curr Protoc Mol Biol 114:30.34.31–30.34.32. https://doi.org/10.1002/0471142727.mb3004s114

Fiehn O (2016b) Mass resolution and resolving power. https://fiehnlab.ucdavis.edu/projects/seven-golden-rules/mass-resolution

Gowda GAN, Djukovic D (2014) Overview of mass spectrometry-based metabolomics: opportunities and challenges. Methods Mol Biol 1198:3–12. https://doi.org/10.1007/978-1-4939-1258-2_1

Griffiths WJ et al (2010) Targeted metabolomics for biomarker discovery. Angew Chem Int Ed 49: 5426–5445. https://doi.org/10.1002/anie.200905579

Guijas C et al (2018) METLIN: a technology platform for identifying knowns and unknowns. Anal Chem 90:3156–3164. https://doi.org/10.1021/acs.analchem.7b04424

Guntner AS, Thalhamer B, Klampfl C, Buchberger W (2019) Collision cross sections obtained with ion mobility mass spectrometry as new descriptor to predict blood-brain barrier permeation by drugs. Sci Rep 9:19182. https://doi.org/10.1038/s41598-019-55856-7

Guo J, Huan T (2020a) Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics. Anal Chem 92:8072–8080. https://doi.org/10.1021/acs.analchem.9b05135

Guo J, Huan T (2020b) Evaluation of significant features discovered from different data acquisition modes in mass spectrometry-based untargeted metabolomics. Anal Chim Acta 1137:37–46. https://doi.org/10.1016/j.aca.2020.08.065

Guo K, Li L (2009) Differential (12)C/(13)C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. Anal Chem 81:3919–3932. https://doi.org/10.1021/ac900166a

Habra H et al (2021) metabCombiner: paired untargeted LC-HRMS metabolomics feature matching and concatenation of disparately acquired data sets. Anal Chem 93:5028–5036. https://doi.org/10.1021/acs.analchem.0c03693

Halket JM et al (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot 56:219–243. https://doi.org/10.1093/jxb/eri069

Han J, Lin K, Sequeira C, Borchers CH (2015) An isotope-labeled chemical derivatization method for the quantitation of short-chain fatty acids in human feces by liquid chromatography-tandem mass spectrometry. Anal Chim Acta 854:86–94. https://doi.org/10.1016/j.aca.2014.11.015

Hastings J et al (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. Nucleic Acids Res 44:D1214–D1219. https://doi.org/10.1093/nar/gkv1031

Health in Northern Ireland (2022) National Center for Advancing Translational Sciences Compound Management. https://ncats.nih.gov/preclinical/core/compound.

Hines KM, Ross DH, Davidson KL, Bush MF, Xu L (2017) Large-scale structural characterization of drug and drug-like compounds by high-throughput ion mobility-mass spectrometry. Anal Chem 89:9023–9030. https://doi.org/10.1021/acs.analchem.7b01709

Horai H et al (2010) MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 45:703–714. https://doi.org/10.1002/jms.1777

Ichou F et al (2014) Comparison of the activation time effects and the internal energy distributions for the CID, PQD and HCD excitation modes. J Mass Spectrom 49:498–508. https://doi.org/10.1002/jms.3365

Kachman M et al (2019) Deep annotation of untargeted LC-MS metabolomics data with Binner. Bioinformatics 36:1801–1806. https://doi.org/10.1093/bioinformatics/btz798

Kanehisa M et al (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34:D354–D357

Kantae V et al (2017) Integration of pharmacometabolomics with pharmacokinetics and pharmacodynamics: towards personalized drug therapy. Metabolomics 13. https://doi.org/10.1007/s11306-016-1143-1

Kanu AB, Dwivedi P, Tam M, Matz L, Hill HH (2008) Ion mobility-mass spectrometry. J Mass Spectrom 43:1–22. https://doi.org/10.1002/jms.1383

Kind T, Fiehn O (2007) Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8:105. https://doi.org/10.1186/1471-2105-8-105

Kind T, Fiehn O (2010) Advances in structure elucidation of small molecules using mass spectrometry. Bioanal Rev 2:23–60. https://doi.org/10.1007/s12566-010-0015-9

Kind T et al (2009) FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Anal Chem 81:10038–10048. https://doi.org/10.1021/ac9019522

Kind T et al (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. Nat Methods 10:755–758. https://doi.org/10.1038/nmeth.2551

Koek MM, Muilwijk B, van Stee LL, Hankemeier T (2008) Higher mass loadability in comprehensive two-dimensional gas chromatography-mass spectrometry for improved analytical performance in metabolomics analysis. J Chromatogr A 1186:420–429. https://doi.org/10.1016/j.chroma.2007.11.107

Koelmel JP et al (2017) Expanding lipidome coverage using LC-MS/MS data-dependent acquisition with automated exclusion list generation. J Am Soc Mass Spectrom 28:908–917

Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of LC/MS data sets. Anal Chem 84: 283–289. https://doi.org/10.1021/ac202450g

Kvitvang HFN, Andreassen T, Adam T, Villas-Bôas SG, Bruheim P (2011) Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids. Anal Chem 83: 2705–2711. https://doi.org/10.1021/ac103245b

Lai Z, Fiehn O (2018) Mass spectral fragmentation of trimethylsilylated small molecules. Mass Spectrom Rev 37:245–257. https://doi.org/10.1002/mas.21518

Lee HB, Peart TE, Svoboda ML (2005) Determination of endocrine-disrupting phenols, acidic pharmaceuticals, and personal-care products in sewage by solid-phase extraction and gas chromatography-mass spectrometry. J Chromatogr A 1094:122–129. https://doi.org/10.1016/j.chroma.2005.07.070

Lenz EM et al (2007) HPLC-NMR with severe column overloading: fast-track metabolite identification in urine and bile samples from rat and dog treated with [14C]-ZD6126. J Pharm Biomed Anal 43:1065–1077. https://doi.org/10.1016/j.jpba.2006.09.010

Li Y et al (2021) Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. Nat Methods 18:1524–1531. https://doi.org/10.1038/s41592-021-01331-z

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc 1:387–396. https://doi.org/10.1038/nprot.2006.59

LLC, H (2022) mzCloud advanced mass spectral database. https://www.mzcloud.org/

Lorenz MA, Burant CF, Kennedy RT (2011) Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. Anal Chem 83:3406–3414. https://doi.org/10.1021/ac103313x

Lu W, Bennett BD, Rabinowitz JD (2008) Analytical strategies for LC-MS-based targeted metabolomics. J Chromatogr B Analyt Technol Biomed Life Sci 871:236–242. https://doi.org/10.1016/j.jchromb.2008.04.031

Ludwig M et al (2019) ZODIAC: database-independent molecular formula annotation using Gibbs sampling reveals unknown small molecules

Mahieu NG, Patti GJ (2017) Systems-level annotation of a metabolomics data set reduces 25,000 features to fewer than 1000 unique metabolites. Anal Chem 89:10397–10406. https://doi.org/10.1021/acs.analchem.7b02380

Mahieu NG, Spalding JL, Gelman SJ, Patti GJ (2016) Defining and detecting complex peak relationships in mass spectral data: the Mz.Unity algorithm. Anal Chem 88:9037–9046. https://doi.org/10.1021/acs.analchem.6b01702

Majuta SN et al (2019) Rapid solution-phase hydrogen/deuterium exchange for metabolite compound identification. J Am Soc Mass Spectrom 30:1102–1114. https://doi.org/10.1007/s13361-019-02163-0

Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D (2008) Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res 49:1137–1146. https://doi.org/10.1194/jlr.D700041-JLR200

Maurer HH (1998) Liquid chromatography mass spectrometry in forensic and clinical toxicology. J Chromatogr B 713:3–25. https://doi.org/10.1016/s0378-4347(97)00514-8

McCann MR, George De la Rosa MV, Rosania GR, Stringer KA (2021) L-carnitine and acylcarnitines: mitochondrial biomarkers for precision medicine. Metabolites 11. https://doi.org/10.3390/metabo11010051
Mendez D et al (2019) ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 47: D930–d940. https://doi.org/10.1093/nar/gky1075
Meng X et al (2021) Simultaneous 3-nitrophenylhydrazine derivatization strategy of carbonyl, carboxyl and phosphoryl submetabolome for LC-MS/MS-based targeted metabolomics with improved sensitivity and coverage. Anal Chem 93:10075–10083. https://doi.org/10.1021/acs.analchem.1c00767
Misra BB, Olivier M (2020) High resolution GC-Orbitrap-MS metabolomics using both electron ionization and chemical ionization for analysis of human plasma. J Proteome Res 19:2717–2731. https://doi.org/10.1021/acs.jproteome.9b00774
Nelson AB, Chow LS, Hughey CC, Crawford PA, Puchalska P (2022) Artifactual fatty acid dimers mimic FAHFA signals in untargeted metabolomics pipelines. J Lipid Res 100201. https://doi.org/10.1016/j.jlr.2022.100201
Papadatos G, Overington JP (2014) The ChEMBL database: a taster for medicinal chemists. Future Med Chem 6:361–364. https://doi.org/10.4155/fmc.14.8
Parker CE et al (2014) In: Tang NLS, Poon T (eds) Chemical diagnostics: from bench to bedside vol. 336 topics in current chemistry-series. pp 117–137
Patti GJ, Yanes O, Siuzdak G (2012) Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 13:263–269. https://doi.org/10.1038/nrm3314
Pence H, Williams A (2010) ChemSpider: an online chemical information resource. J Chem Educ 87:1123–1124. https://doi.org/10.1021/ED100697W
Plante P-L et al (2019) Predicting ion mobility collision cross-sections using a deep neural network: DeepCCS. Anal Chem 91:5191–5199. https://doi.org/10.1021/acs.analchem.8b05821
Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11:395. https://doi.org/10.1186/1471-2105-11-395
Prakash C, Shaffer CL, Nedderman A (2007) Analytical strategies for identifying drug metabolites. Mass Spectrom Rev 26:340–369. https://doi.org/10.1002/mas.20128
Rainville PD et al (2017) Ion mobility spectrometry combined with ultra performance liquid chromatography/mass spectrometry for metabolic phenotyping of urine: effects of column length, gradient duration and ion mobility spectrometry on metabolite detection. Anal Chim Acta 982:1–8. https://doi.org/10.1016/j.aca.2017.06.020
Ridder L et al (2012) Substructure-based annotation of high-resolution multistage MSn spectral trees. Rapid Commun Mass Spectrom 26:2461–2471. https://doi.org/10.1002/rcm.6364
Roberts LD, Souza AL, Gerszten RE, Clish CB (2012) Targeted metabolomics. Curr Protoc Mol Biol. Chapter 30, Unit 30.32.31-24. https://doi.org/10.1002/0471142727.mb3002s98
Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S (2016) MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Chem 8:3. https://doi.org/10.1186/s13321-016-0115-9
Saigusa D, Matsukawa N, Hishinuma E, Koshiba S (2021) Identification of biomarkers to diagnose diseases and find adverse drug reactions by metabolomics. Drug Metab Pharmacokinet 37: 100373. https://doi.org/10.1016/j.dmpk.2020.11.008
Scheubert K et al (2017) Significance estimation for large scale metabolomics annotations by spectral matching. Nat Commun 8:1494. https://doi.org/10.1038/s41467-017-01318-5
Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA (2016) Untargeted metabolomics strategies – challenges and emerging directions. J Am Soc Mass Spectrom 27:1897–1905. https://doi.org/10.1007/s13361-016-1469-y
Schymanski EL et al (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol 48:2097–2098. https://doi.org/10.1021/es5002105

Scott DR (1992) Rapid and accurate method for estimating molecular weights of organic compounds from low resolution mass spectra. Chemom Intel Lab Syst 16:193–202. https://doi.org/10.1016/0169-7439(92)80037-5

Shackleton C, Pozo OJ, Marcos J (2018) GC/MS in recent years has defined the Normal and clinically disordered Steroidome: will it soon be surpassed by LC/tandem MS in this role? J Endocr Soc 2:974–996. https://doi.org/10.1210/js.2018-00135

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787. https://doi.org/10.1021/ac051437y

Soldin OP, Chung SH, Mattison DR (2011) Sex differences in drug disposition. J Biomed Biotechnol 2011:187103. https://doi.org/10.1155/2011/187103

Solutions WS (2022) Wiley registry 12th edition/NIST 2020 mass spectral library. https://sciencesolutions.wiley.com/solutions/technique/gc-ms/wiley-registry-12th-edition-nist-2020/

Stanstrup J, Neumann S, Vrhovšek U (2015) PredRet: prediction of retention time by direct mapping between multiple chromatographic systems. Anal Chem 87:9421–9428. https://doi.org/10.1021/acs.analchem.5b02287

Stauffer DB, McLafferty FW, Ellis RD, Peterson DW (1985) Probability-based-matching algorithm with forward searching capabilities for matching unknown mass spectra of mixtures. Anal Chem 57:1056–1060. https://doi.org/10.1021/ac00283a021

Stein SE (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J Am Soc Mass Spectrom 10:770–781

Stein SE, Scott DR (1994) Optimization and testing of mass spectral library search algorithms for compound identification. J Am Soc Mass Spectrom 5:859–866. https://doi.org/10.1016/1044-0305(94)87009-8

Steuer AE, Brockbals L, Kraemer T (2019) Metabolomic strategies in biomarker research-new approach for indirect identification of drug consumption and sample manipulation in clinical and forensic toxicology? Front Chem 7:319. https://doi.org/10.3389/fchem.2019.00319

Strehmel N, Hummel J, Erban A, Strassburg K, Kopka J (2008) Retention index thresholds for compound matching in GC-MS metabolite profiling. J Chromatogr B Analyt Technol Biomed Life Sci 871:182–190. https://doi.org/10.1016/j.jchromb.2008.04.042

Sumner LW et al (2007) Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). Metabolomics 3: 211–221. https://doi.org/10.1007/s11306-007-0082-2

Taylor T (2015) Understanding electron ionization processes for GC–MS. LCGC N Am 33:290

Theodoridis GA, Gika HG, Want EJ, Wilson ID (2012) Liquid chromatography-mass spectrometry based global metabolite profiling: a review. Anal Chim Acta 711:7–16. https://doi.org/10.1016/j.aca.2011.09.042

Tiwary S et al (2019) High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. Nat Methods 16:519–525. https://doi.org/10.1038/s41592-019-0427-6

Tolić N et al (2017) Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. Anal Chem 89:12659–12665. https://doi.org/10.1021/acs.analchem.7b03318

Tsugawa H et al (2015) MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods 12:523–526. https://doi.org/10.1038/nmeth.3393

Tsugawa H et al (2016) Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. Anal Chem 88:7946–7958. https://doi.org/10.1021/acs.analchem.6b00770

van der Laan T et al (2021) Fractionation platform for target identification using off-line directed two-dimensional chromatography, mass spectrometry and nuclear magnetic resonance. Anal Chim Acta 1142:28–37. https://doi.org/10.1016/j.aca.2020.10.054

Vaniya A, Fiehn O (2015) Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. Trends Anal Chem 69:52–61. https://doi.org/10.1016/j.trac.2015.04.002

Vinaixa M et al (2016) Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects. TrAC Trends Anal Chem 78:23–35. https://doi.org/10.1016/j.trac.2015.09.005

Wang H et al (2015) Systematic optimization of long gradient chromatography mass spectrometry for deep analysis of brain proteome. J Proteome Res 14:829–838. https://doi.org/10.1021/pr500882h

Wang M et al (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat Biotechnol 34:828–837. https://doi.org/10.1038/nbt.3597

Wang X et al (2018) Target-decoy-based false discovery rate estimation for large-scale metabolite identification. J Proteome Res 17:2328–2334. https://doi.org/10.1021/acs.jproteome.8b00019

Wang S, Kind T, Tantillo DJ, Fiehn O (2020) Predicting in silico electron ionization mass spectra using quantum chemistry. J Chem 12:63. https://doi.org/10.1186/s13321-020-00470-3

Watson DG (2013) A rough guide to metabolite identification using high resolution liquid chromatography mass spectrometry in metabolomic profiling in metazoans. Comput Struct Biotechnol J 4:e201301005. https://doi.org/10.5936/csbj.201301005

Whiley L et al (2019) Systematic isolation and structure elucidation of urinary metabolites optimized for the analytical-scale molecular profiling laboratory. Anal Chem 91:8873–8882. https://doi.org/10.1021/acs.analchem.9b00241

Willoughby PH, Jansma MJ, Hoye TR (2014) A guide to small-molecule structure assignment through computation of (1H and 13C) NMR chemical shifts. Nat Protoc 9:643–660. https://doi.org/10.1038/nprot.2014.042

Wishart DS (2009) Computational strategies for metabolite identification in metabolomics. Bioanalysis 1:1579–1596. https://doi.org/10.4155/bio.09.138

Wishart DS et al (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34:D668–D672. https://doi.org/10.1093/nar/gkj067

Wishart DS et al (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46:D1074–d1082. https://doi.org/10.1093/nar/gkx1037

Wishart DS et al (2022) HMDB 5.0: the human metabolome database for 2022. Nucleic Acids Res 50:D622–D631. https://doi.org/10.1093/nar/gkab1062

Wong JM et al (2016) Benzoyl chloride derivatization with liquid chromatography-mass spectrometry for targeted metabolomics of neurochemicals in biological samples. J Chromatogr A 1446:78–90. https://doi.org/10.1016/j.chroma.2016.04.006

Wu Q et al (2019) Enhancing coverage in LC-MS-based untargeted metabolomics by a new sample preparation procedure using mixed-mode solid-phase extraction and two derivatizations. Anal Bioanal Chem 411:6189–6202. https://doi.org/10.1007/s00216-019-02010-x

Xue J et al (2020) Enhanced in-source fragmentation annotation enables novel data independent acquisition and autonomous METLIN molecular identification. Anal Chem 92:6051–6059. https://doi.org/10.1021/acs.analchem.0c00409

Yoo HJ, Liu HC, Hakansson K (2007) Infrared multiphoton dissociation and electron-induced dissociation as alternative MS/MS strategies for metabolite identification. Anal Chem 79:7858–7866. https://doi.org/10.1021/ac071139w

Zaitsu K, Hayashi Y, Kusano M, Tsuchihashi H, Ishii A (2016) Application of metabolomics to toxicology of drugs of abuse: a mini review of metabolomics approach to acute and chronic toxicity studies. Drug Metab Pharmacokinet 31:21–26. https://doi.org/10.1016/j.dmpk.2015.10.002

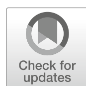
Zhang Z et al (2018) Reverse and random decoy methods for false discovery rate estimation in high mass accuracy peptide spectral library searches. J Proteome Res 17:846–857. https://doi.org/10.1021/acs.jproteome.7b00614

Zhao S, Li L (2020) Chemical derivatization in LC-MS-based metabolomics study. TrAC Trends Anal Chem 131:115988. https://doi.org/10.1016/j.trac.2020.115988

Zheng J, Mandal R, Wishart DS (2018) A sensitive, high-throughput LC-MS/MS method for measuring catecholamines in low volume serum. Anal Chim Acta 1037:159–167. https://doi.org/10.1016/j.aca.2018.01.021

Zhou JT, Yin YX (2016) Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. Analyst 141:6362–6373. https://doi.org/10.1039/c6an01753c

Zhou Z et al (2020) Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. Nat Commun 11:4334. https://doi.org/10.1038/s41467-020-18171-8



Metabolomics and NMR

Ryan T. McKay

Contents

1 Objectives 74
2 Brief History 75
 2.1 Potential 76
  2.1.1 A Warning 76
 2.2 Definitions 76
  2.2.1 Metabolome 76
  2.2.2 Metabo -Nomics or -Lomics 77
  2.2.3 Metabolites 78
3 Expectations 78
 3.1 Reality 78
  3.1.1 Key Considerations for NMR 78
4 NMR and MS 79
 4.1 Complementation: Can We Just Skip to the End? 79
 4.2 Metabolomics via Mass Spectrometry 80
 4.3 Metabolomics by NMR 81
5 Nuclear Magnetic Resonance Spectroscopy 81
 5.1 NMR Experiment for Metabolomics 81
  5.1.1 First Impressions 82
 5.2 Liquids NMR 82
  5.2.1 Progression to Metabolomics 83
 5.3 Samples 84
  5.3.1 Preparation 84
  5.3.2 NMR Tubes, Spinners, and Side-Bands 85
  5.3.3 Gathering and Handling 86
  5.3.4 In the Tube 89
 5.4 NMR Spectrometer 90
  5.4.1 Probe 90
  5.4.2 Console 92

R. T. McKay (✉)
Department Chemistry, College of Natural and Applied Sciences, University of Alberta, Edmonton, AB, Canada
e-mail: ryan.mckay@ualberta.ca

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_616

5.4.3 Host Computer ... 92
5.4.4 Experiment Pulse Sequence ... 93
5.4.5 Parameters ... 94
5.4.6 Phase Cycle ... 95
5.4.7 Temperature Calibration ... 95
5.4.8 Tune/Match ... 96
5.4.9 Excitation Pulse ... 97
5.4.10 Gain ... 98
5.4.11 Pulse Field Gradients ... 99
5.5 Manuscript Reporting Parameters ... 100
5.5.1 Pulse Width ... 100
5.5.2 Delays: Relaxation, Equilibrium ... 101
5.5.3 Post-Acquisition: Weighting Function ... 102
5.5.4 Corrections: Linear Prediction, Baseline, Phasing ... 102
5.5.5 Referencing ... 103
6 Future Ideas ... 103
6.1 Solids NMR Metabolomics ... 103
6.2 Working with Raw FID ... 104
6.3 PureShift ^{1}H-^{1}H J-Coupling Removal ... 104
7 Conclusion ... 105
Appendix A: Example Pulse Sequence: Metabolomics 1D-^{1}H "Metnoesy" ... 106
Appendix B: Example Experimental Section ... 106
References ... 108

Abstract

The purpose of this manuscript will be to convince the reader to dive deeper into NMR spectroscopy and prevent the technique from being just another "black-box" in the lab. We will try to concisely highlight interesting topics and supply additional references for further exploration at each stage. The advantages of delving into the technique will be shown. The secondary objective, i.e., avoiding common problems before starting, will hopefully then become clear. Lastly, we will emphasize the spectrometer information needed for manuscript reporting to allow reproduction of results and confirm findings.

Keywords

Automation · Biochemistry · Liquids · Metabolites · Metabolomics · NMR · Nuclear magnetic resonance · Quantitation · Small molecule · Solvent suppression · Spectrometry

1 Objectives

The purpose of this manuscript will be to convince the reader to dive deeper into NMR spectroscopy and prevent the technique from being just another "black-box" in the lab. We will try to concisely highlight interesting topics and supply additional references for further exploration at each stage. The advantages of delving into the technique will be shown. The secondary objective, i.e., avoiding common problems

before starting, will hopefully then become clear. Lastly, we will emphasize the spectrometer information needed for manuscript reporting to allow reproduction of results and confirm findings.

2 Brief History

Metabolomics involves a few key steps: hypothesis of perturbation observation by metabolomic measurement (e.g., identifying a disease), sample selection and handling, instrumentation setup/management, and finally processing and analysis. In any living system any number of molecules regularly and/or responsively change their prevalence and activity, and thus cellular function requires constant feedback/control of metabolites through kinetic and energetics. Therefore, at any stage, metabolites certainly change their presence and/or abundance, but can we detect them?

Seeking to elucidate, understand, and be able to apply our knowledge about these changes is the foundation of metabolomics, and there has been an astounding amount of development regarding metabolomics (Lindon et al. 1999; Wang and Li 2020; Tenori et al. 2020; Giraudeau 2020) spanning the last 20 years. Nuclear magnetic resonance (NMR) spectroscopy has been one of two primary tools from the start, with the second being mass spectrometry (MS), and while there are certainly other techniques, NMR and MS are arguably the most widely applied. Both NMR and MS are certainly powerful analytical techniques, each with advantages and disadvantages when compared directly or to other instrumental methods. This will be detailed later (e.g., see section NMR and MS Competition and Complementation).

Dramatic NMR spectrometer improvements including common access to higher magnetic fields, cryogenically cooled probes, new robotic sample handling with temperature control, automated software calibration/optimization/acquisition/processing and analysis have made the instrument "black-box" mode ever more seductive. There are many of these aspects that (with some interesting background and a little emphasis) can be efficiently optimized. Specifically understanding the fundamentals of sample management, instrument preparation, and analysis expectations can make project planning and setup easier with a concomitantly higher probability of success with a more reliable, comparable, and efficient study before even starting. Please note that by comparable, we mean the ability to validate and take data from other studies, instruments, and/or facilities for inclusion into your data analysis (or vice versa), not just the data collected in one location and/or one dedicated instrument (see (Lacy et al. 2014; Sokolenko et al. 2013) and references therein). It is often assumed that collecting data on one dedicated instrument removes multi-facility/multi-instrument complications. However while using a single dedicated instrument *should* achieve consistent precision, there is no guarantee of accuracy (all samples may be equally inaccurate) nor does this assumption consider unavoidable changes in instrumentation over time, e.g. repairs, replacements, updates, etc. (see Sect. 5.4 and other sections below).

2.1 Potential

Metabolomics has been utilized now for several decades (see (Emwas et al. 2020; Finco et al. 2016; Heather et al. 2013; Kenny et al. 2010; Li et al. 2012; Psychogios et al. 2011; Rasmussen et al. 2012) and references therein) but has not reached the promised potential for applicable personalized/precision medicine. While there are certainly specifically applied findings, there has not been the explosion of the predicted novel medical treatments. Very recently mRNA vaccine developments have introduced new concerns regarding the possible over-production of proteins/ metabolites (see review (Pardi et al. 2018)) and highlighted the application of metabolomics for monitoring.

2.1.1 A Warning

It is important to remind the reader that the promise of marketability drove much of the initial science. Metabolomics starting in the early 2000s, was envisioned as a novel tool with exceedingly high profitability that would be: rapid, reliable, reproducible, using easily acquired samples (e.g., urine) without extensive sample manipulation (Bingol et al. 2016; Tayyari et al. 2013), and contain directly interpretable, widely applicable, and useful results. Companies quickly jockeyed to be the first to sell and inhabiting their NMR spectrometers (and now perhaps benchtop units (Izquierdo-Garcia et al. 2020)) throughout clinical testing facilities across the globe. For any new marketable technique to survive it must either be unique and advantageous (i.e., provide a novel result) or be extremely competitively priced when compared to existing technology (i.e., faster, cheaper, etc.). The corporate emphasis was on finding unique biomarkers for dramatic high-profile diseases (e.g., cardiovascular disease, cancers, etc.) as quickly as possible and thus establish lucrative patents and contracts. This did not encourage systematic, calm evaluation, nor retesting, and that rush may have inadvertently hurt the entire field. Subsequent validation studies (Emwas et al. 2020; Lacy et al. 2014; Sokolenko et al. 2013; Markley et al. 2017) have begun to fill in the gaps, however questions regarding the cross-validation of NMR data persists, e.g., site to site and/or study to study along with assessment (Rocca-Serra et al. 2016; Cassiède et al. 2017).

2.2 Definitions

There has been ambiguity in the literature regarding key definitions and to avoid confusion we will quickly define our working interpretation below.

2.2.1 Metabolome

The metabolome is commonly defined as all small molecules in the mass range of 50 to 1,500 Da (not a hard limit as certainly lipids can quickly exceed), associated with a particular organism (Dunn et al. 2011; Psychogios et al. 2011; Wolfender et al. 2013; Zulyniak and Mutch 2011). This includes all the various complexes, sizes, and repetitive units of amino acids, lipids, carbohydrates, and other organic

molecules collectively termed “metabolites” and are involved in all stages of metabolism both from naturally internal (endogenous) starting points, and those originating from external (exogenous) sources. External include ingested foods or pharmaceuticals, and the gastrointestinal metagenome along with subsequent by-products, more colloquially referred to as the “gut flora and fauna” with these microbes apparently playing a diverse role as our understanding expands.

2.2.2 Metabo -Nomics or -Lomics

Here are two important terms that, depending on which manuscript you reference, can have subtle but important differences. Metabonomics was generally given a wider definition of studying all relevant interactions over an enter species (i.e., the subject’s metabonome). Metabolomics was mostly considered to be a more focused subset in a defined case. Also NMR and MS manuscripts tended to use one of the terms, but not the other. Metabolomics has become the most common and to avoid confusion we will use only metabolomics, defining it as studying the metabolites in a defined situation (e.g., human urine metabolites detectable by NMR). Readers interested in the distinctions are directed to excellent reviews such as (Bouatra et al. 2013; Psychogios et al. 2011; Macel et al. 2010; Gibney et al. 2005) and papers therein.

More “-Omics”?

For a review, one cannot simply ignore the “-omics” flood as it now also includes subtopics of metabolomics. Logically we start with the genome, then onto the regulation of transcription with subsequent modifications. Then we consider effective production/regulation/degradation (e.g., protein turn-over), and finally the basic building block level resulting in functional metabolism. These levels have an expanding and/or encompassing “-omics” (Ragguett and McIntyre 2020) associated with their study (i.e., genomics, transcriptomics, proteomics, metabolomics), and even more recent approaches such as pharmacogenomics or pharmacometabolomics (Van Der Wouden et al. 2020; Emwas et al. 2021; Lasky-Su et al. 2021; Vignoli et al. 2019; Sherlock and Mok 2019).

There are many new subgroupings of of “omics” including Lipidomics (see a recent full Nature review[1]), Foodomics (Balkir et al. 2021; Picone et al. 2022; Valdés et al. 2021), and Elemental Metabolomics (Andersson et al. 2021; Edison et al. 2020; Niziol et al. 2021; Zhang et al. 2018). In case readers have not had enough “-omics,” there is a relatively new though well established and pertinent sub-group focusing on the changes of metabolites over time called Fluxomics, see the comprehensive review by Giraudeau (2020) and also (Emwas et al. 2020) for additional references. Of key interest is the use of specific NMR labelling (Xu et al. 1999) in metabolites.

[1] https://www.nature.com/subjects/lipidomics.

2.2.3 Metabolites

To properly address the term metabolites, we will use the excellent definition provided by Tenori et al. in their concise review (Tenori et al. 2020).

> *Metabolites are the small molecules produced by, or taking part in, the chemical reactions due to biochemical activities (i.e., the metabolism) in living organisms, and their levels can change according to pathophysiological or environmental factors.*

3 Expectations

As metabolites encompass the end stage of cellular chemical regulation/function, it is postulated that this should be the easiest way to detect and ascertain the cause of fundamental upstream cellular changes. Essentially, we are hoping to observe an amplification of small changes from further up the "-omic" ladder therefore making disease detection and analysis faster and easier.

3.1 Reality

While DNA is a single chemical class comprising only four nucleotides, immense molecular stability, and a second strand as an immediately available backup copy for error correction, moving downstream to the metabolites suddenly expands to thousands if not tens of thousands of arguably critical molecules. These downstream molecules can be quickly and substantially impacted by even small changes at the DNA or transcription level. Metabolite concentrations can normally and dynamically range by several orders of magnitude (Bouatra et al. 2013). Immediately the reader can see that while there may indeed be amplification, i.e., a downstream effect any single "snap-shot" acquisition of metabolites (regardless of how detailed and accurate) may have limited practical applicability to any disease detection. This gets even more complicated regarding any practical translation to personalized-medicine ("bedside") treatment. Also metabolites have an immense range of molecular variability in terms of primary/secondary structure, function, modification, and lifespan.

3.1.1 Key Considerations for NMR

For NMR, concentration is perhaps *the* most crucial aspect due to inherent detection limitations. Any low concentration biomarkers exist in a literal sea of much higher concentration compounds, complexes, and aggregates. Essentially the detected signal's "dynamic range" (stealing an electrical engineer and audiophile term) is incredibly large spanning many orders of magnitude regardless of the technique selected. This is also directly relevant to the source/type of tissue involved that can result in vastly different concentrations of metabolites. Then there are complications regarding simple collection, e.g., considering human-based samples: the time of day, method of isolation/collection, fasting level of donor, female or male, age, activity,

genetic background, environmental influences, etc. We may also need to consider not a single biomarker, but instead a group or family of inter-related metabolites responding directly or indirectly to other metabolites. On top of all those considerations, when then have the metabolomic influence of the "gut metabolome." Specifically the natural beneficial and negative opportunistic organisms existing in our bodies consuming and producing metabolites as part of their normal functions (Visconti et al. 2019).

The key point is substantial planning, and experimental preparation will likely be required to minimize all these complications. We will either address these directly and/or provide references to assist the reader.

4 NMR and MS

4.1 Complementation: Can We Just Skip to the End?

Metabolomics by any method seeks a statistically relevant and consistent change (s) in a single or multiple observed variables. Ideally, we would like to see the novel presence or absence of a recognizable and preferably unique signal associated to a specific molecule which can be related to a disease, i.e., literally the lightbulb on/off. A disease "marker" that could be regularly, rapidly, and easily followed.[2]

There is far more likely a dynamic and complex series of cascading and/or interlocking molecular pathways responding to change(s) with dependency on the complexity of the organism being monitored. Feedback loops, changes of enzymatic expression levels, regulation of genetic expression, alterations of metabolism rates, changes in diet and/or activity due to the subject's phenotypic expression (e.g., how bad they may feel), self-medicating, the body actively trying to re-establish equilibrium; all should contribute to an interdependence if not system wide response.

So is there any single technique that can identify all possible metabolites in all samples regardless of origin? No, but there are several techniques offering different strengths and weaknesses, and there is always the added potential of linking together different techniques. The so-called "hyphenated" methods where one runs the sample consecutively through multiple instruments and gathers the acquired data to make a far more powerful determination.

The problems with "hyphenation," i.e., the myth of interconnected instruments with streamlined throughput, such as mass spectroscopy (MS), high-performance liquid chromatography (HPLC), fast protein liquid chromatography (FPLC), gas chromatography (GC), circular dichroism (CD) is that they may not always work well together, or the sequence is important to the result, or one needs multiple identical samples. An example is obtaining NMR data first, which then MS contaminates the sample with deuterium (for NMR "lock" see below), or capillary electrophoresis (CE) then MS (Qiu et al. 2020). There are always the concerns about

[2]This overly simplified one-molecule/one-marker situation is highly unlikely.

sample cross-contamination, how to adequately resolve signals in mixtures, incorrect assignment, or obscured/missing assignment. Then there is the practical problem of maintaining and constantly checking that the inter-related instruments are working consistently. Striving for any technique that provides fully automated sample preparation, instrument optimization, sample handling, data acquisition, sample storage, data processing and finally analysis is presently impossibly ambitious, but a semi-automated approach might be easier, faster, and more reliable. Like NMR structure-function protein backbone assignments – semi-automated, i.e., get the computer to do the routine work and then present the challenges to the experienced instrument operator who can use visual pattern recognition and experience to make the more difficult decisions seems to be the most efficient and reliable. As the title of this review implies, “black-boxing” the problem leads to (at best) consistent errors and more likely unusable data. At worst, false information makes it into the peer-reviewed literature taking a huge amount of work to recognize and correct later. Unfortunately, we cannot skip to the end and we will need to discuss and understand the instrumentation.

4.2 Metabolomics via Mass Spectrometry

The author approaches mass spectrometry (MS) with admittedly little practical experience surrendering any in-depth evaluation of MS to the experts, e.g., see highly recommended reviews (Wang and Li 2020; Alseekh et al. 2021)and references therein. While mass spectrometry has undeniably superior solute sensitivity (e.g., microlitre or sub-microlitre volumes with relatively low concentrations (Li et al. 2020) per unit instrument time, it does come with a literal financial cost and a cost in terms of total experimental length, monitoring/compensation for separation technique(s), necessity of quality/control sampling, and finally the destruction of the sample. It is important to note that MS instrumentation has a substantially smaller initial instrument cost, along with smaller maintenance costs (e.g., cryogens). A critical distinction is that MS usually requires some form of sample separation (Petrović et al. 2005; Korfmacher 2005; Alseekh et al. 2021): chromatographic (liquid or gas), electrophoretic, or based on ion mobility that can perturb the types and quantities of measured metabolites. The separation efficacy changes over time requiring continuous monitoring, evaluation, and correction during analysis, e.g., separatory columns degrade nonuniformly over time requiring a calibration of resulting metabolites. Sample separation and detection limitations (Wang and Li 2020) can also be linked to the particular detector such as[3]: Fourier transform ion cyclotron resonance (Marshall et al. 1998, 2007; Nikolaev et al. 2016), ion trap (Todd and March 1999), Orbitrap and linear ion (Perry et al. 2008), time-of-flight (Boesl 2017), and quadrupole (Linge and Jarvis 2009) to name some of the most common. Evaluation of the potential and realized application has been extensively

[3] Detector order is alphabetical only, not intended to imply frequency nor capabilities.

covered and readers are directed to representative articles (Trifonova et al. 2021; Alseekh et al. 2021; Lasky-Su et al. 2021; Wang and Li 2020). In the end MS has become an extremely successful method accurately detecting thousands of molecules and at much lower concentrations that can be practicably detected by NMR.

4.3 Metabolomics by NMR

Metabolomics by NMR has been vastly reviewed (Tenori et al. 2020; Giraudeau 2020; Giraudeau et al. 2014, 2015; Tavares et al. 2015; Halabalaki et al. 2014; Bingol and Brüschweiler 2014; Bouatra et al. 2013; Wolfender et al. 2013; Lubbe et al. 2013; Ellinger et al. 2013; Heather et al. 2013; Smolinska et al. 2012; Dunn et al. 2011; Macel et al. 2010; Lindon et al. 2007; Beckonert et al. 2007; Emwas et al. 2018, 2019; Stringer et al. 2016). There is also a highly recommended NMR book with an entire chapter dedicated to the practical aspects of metabolomics and NMR sample handling/data/processing (Teng 2012).

Metabolomics by NMR contains all the common instrumental concerns (e.g., consistency, detection, assignment of signals, etc.), but a unique aspect of NMR is that each atom, even in the same molecule, can have a specific resonance. This is both the strength, i.e., the ability to resolve an atom based on its magnetic environment, and weakness. Every atom resonating by itself provides little cumulative signal. Each relevant atomic signal must be detected, assigned, and then analyzed. All of this with the massive assumption that no unexpected external change(s), other than the central hypothesis of the study will perturb the measurement.

We will address that perturbation assumption and more below.

5 Nuclear Magnetic Resonance Spectroscopy

We will highlight key points in the subsections below to detail important information from our experience to hopefully improve the reader's future research studies. The points will use a special format indicated by **Note**.

5.1 NMR Experiment for Metabolomics

To start, a metabolomics NMR experiment (i.e., the "pulse sequence" itself) ideally must be easy to acquire, i.e., any required hardware is commonplace, minimal setup, robust, reliable, and reproducible. The instrument is assumed to be constantly maintained, tested, and repaired be experienced operators to ensure predictable performance. Next there are literally hundreds of NMR experiments (Berger and Braun 2004; Braun et al. 1998), and the number of individual and inter-related NMR parameters for each experiment is often overwhelming. The initial selection, testing, and maintenance can be a full-time occupation for facilities (Reynolds and Enriquez

2002). The person contemplating a metabolomics NMR project must either acquire equipment and become this, or already have the equipment and technical personnel available.

5.1.1 First Impressions

Users commonly first experience NMR through introductory organic chemistry courses (Derome 2013). This exposure can expand into some inorganic chemistry and certainly into biomolecular NMR (Ban et al. 2017; Göbl et al. 2014; Gardner and Kay 1998; Ziarek et al. 2018), but very rarely into the hardware itself. NMR is also commonly involved in natural products identification, analytical food ID/confirmation (e.g., wine, beer, scotch, juice, honey) (Spraul et al. 2015; Esslinger et al. 2015; Link et al. 2014; Kew et al. 2017; Link et al. 2014; Sandusky and Raftery 2005) and legal evaluation (i.e., spiking) (Lesar et al. 2011). Other experiences with NMR could perhaps involve academic spin-off pharmaceutical/commercial applications (Lindon et al. 2007; Duarte et al. 2014; Lepre 2011; Shuker et al. 1996). This type of experience lends itself well to metabolomics as most metabolites are small and will yield NMR spectra similar to what users have experienced. However, the sheer number of compounds in a common biological sample will certainly not be typical of an organic chemistry problem, nor does this address how the instrument operates.

NMR originally started as an interesting physics phenomena and expensive pH meter (Bull et al. 1964). The field has certainly grown since the 1950s and 1960s in terms of distribution/application of instruments, stable/achievable magnetic field strength, versatility, sensitivity, consistency of equipment performance, and reduced cost of operation. Instrument consoles have gotten smaller while magnets have become increasingly powerful and shielded thus reducing the laboratory footprints for installation and/or the number of instruments required.

Note

While the accepted "NMR standard" is 0.1% ethylbenzene with 0.01% TMS in deuterated chloroform, Benchtop NMR manufacturers have created their own standard using 1% ethylbenzene instead. Therefore, the reported benchtop ^{1}H NMR signal-to-noise ratios must be either divided by a factor of 10 to compare to previously published standards and/or the user must recognize a factor of 100 for required experiment time to achieve comparable results.

5.2 Liquids NMR

While it is fun to teach and delve into the world of NMR theory (e.g., providing useful analogies describing each atomic nuclei as a little bar magnet spinning at its own frequency depending on the magnetic field strength etc.), we simply do not have the space in this review; especially considering the wealth of previously published

information by extremely articulate lectures/authors/presenters on the subject. The interested reader is first guided to very well-established books[4] on the subject, for example, Hore 2015; Hore et al. 2015; Vögeli 2014; Silverstein et al. 2014; Derome 2013; Keeler 2010; Bakhmutov 2005; Harris 1986; Freeman 1997; Brown 2016; Cavanagh et al. 2006; Levitt 2001; Zerbe and Jurt 2013; Wüthrich 1986 aimed at audiences ranging from the interested to the specialist/expert. These resources will then escort the reader into the vast fount of peer-reviewed manuscripts and reviews. It is worth mentioning that NMR can directly cover/monitor processes time scales covering many orders of magnitude (Ziarek et al. 2018). From the medical/biochemistry aspect protein structure function is one of the most recognized applications of NMR, but protein bioNMR is usually most effective with a solitary type of molecule and at relatively high concentrations (e.g.,$>$1 mM) and there is also the need (Amoureux et al. 2008) for isotopic enrichment, i.e., ^{13}C, ^{15}N, and even ^{2}H (Gardner and Kay 1998; Hiroaki 2013) or selectively enriched at key positions. For reviews see (Kay and Frydman 2014; Kay 2016).

Note
It is important to note that much of the background NMR material does not address complex or "strong" ^{1}H-^{1}H coupling regarding metabolomics assignments (see sections below).

Once the reader goes beyond the simple spin-1/2-coupling "tree" diagrams into the higher order (also called second order, complex, or strong coupling depending on the terminology used), we discover spectrometer/magnetic field-dependent patterns (Foroozandeh et al. 2014; Bain et al. 1994). This is important for software selection, analysis, and anyone aiming to assign biomarkers (Mercier et al. 2011; Tredwell et al. 2011; Weljie et al. 2006). Using a reference database that does not take this into account and/or have information for the magnetic field used will be more difficult to use and require operators to be more experienced (Lacy et al. 2014; Mercier et al. 2011; Tredwell et al. 2011; Weljie et al. 2006).

5.2.1 Progression to Metabolomics

Metabolomics evolved in earnest in the early 2000s (Lindon et al. 2007; Lauridsen et al. 2007; Wang et al. 2010; McGrath et al. 2007; Weljie et al. 2006; Saude et al. 2006; Tilgner et al. 2019). There were earlier endeavors seeking insight into what was then referred to as "in-born errors," but the technology needed time to catch up to the intellectual concepts (Lehnert and Hunkler 1986). The field[5] is now well established with typical superconducting magnets now allowing observation of

[4] No implied order, and naming a few of the author's personal favourites. There are many others certainly worth the reader's attention and the provided list is not exhaustive.

[5] Please forgive the puns.

hydrogen from 400 to 800 MHz (i.e., 9.39 to 18.8 Tesla static magnetic field strength) and offered from a range of vendors.[6]

While concerns about data processing and analysis are equally important, there are many publications and reviews focusing on issues involving time domain to frequency signals such as signal enhancement (apodization), deconvolution, integration, and then reliably identifying and assigning signals or signal components (Schönberger et al. 2015) to particular metabolite resonances. Readers seeking even more information are directed to a representative set of papers and references therein (Bartel et al. 2013; Beckonert et al. 2007; Bingol et al. 2016; Dudzik et al. 2018; Dunn et al. 2011; Eghbalnia et al. 2017; Ellinger et al. 2013; Emwas et al. 2016, 2018, 2019; Wang and Li 2020; Kohl et al. 2012; Krishnamurthy 2013; Lacy et al. 2014; Parsons et al. 2009). We will now focus on the practical aspects of sample and NMR instrumentation, e.g., see Chap. 3 of reference (Cavanagh et al. 2006) and Chap. 2 of reference (Zerbe and Jurt 2013).

Minimizing Problems

The detection and elimination of "confounders," i.e., signals or influences on signals that are not dependent on the hypothesized change (e.g., disease vs. healthy) but instead are artifacts and/or errors in sample preparation and/or instrumentation, must be a major focus for everyone involved in metabolomics. We wish to remove (or at least minimize) as many variables as possible prior to the acquisition of NMR data (Athersuch et al. 2013; Gibney et al. 2005; Meissner et al. 2014; Staab et al. 2010; Zulyniak and Mutch 2011). This requires prior awareness, continued attention, precise planning throughout, and finally careful preparation, and this leads us directly to the sections below regarding sample and instrumentation details and recommendations.

5.3 Samples

5.3.1 Preparation

Sample selection, preparation, consideration for repetitive sampling, and handling consistency are all incredibly important. This is only the first step, but the easiest to get wrong, and sometimes without being determinable until very late in the study.

> **Note**
>
> The Brian Sykes' First rule of NMR applies here, i.e., "Garbage" in equals "Garbage" out. Essentially no matter how good the operator nor expensive the instrumentation, if your sample is poor, your results are poor.

[6]Unfortunately in 2014, Agilent (who had purchased Varian Inc. in 2010) exited the NMR market. This has added a level of uncertainty for those still needing repair/support/parts for their massive equipment investments, especially in the ever-expanding austerity environments facing publicly funded academic institutions.

Any complication such as contamination in the sample solution, contamination on the outside of the tube,[7] material not dissolved and floating, i.e., hair, glass wool, debris etc., precipitated solute, a scratched NMR tube (see Sects. 3.2.3–3.2.5 in (Derome 2013)), and even just improperly positioned in the NMR spinner, can all have dramatic impacts on the overall quality and therefore the consistency of the recorded data. Without everyone involved in the study agreeing and following precise protocols (i.e., user/handling confounders), the study is doomed to problems.

Note

The key point here is absolute consistency, from *everyone*. Despite careful planning all the group members that will be handling samples must understand the stringent necessity of consistent sample treatment.

5.3.2 NMR Tubes, Spinners, and Side-Bands

Traditionally NMR tubes are placed in carriers called "spinners." The spinners allow the insertion/removal of a sample via compressed air shuttling the sample up to the top and down into the magnet core for observation. The spinner can also suspend the tube while inside the magnet on a slight cushion of air to facilitate rotation of the sample (i.e., parallel axis to the NMR magnet bore tube)(Harris 1986). The rotation of ~15–20 Hz was used to average or "spin-out" inhomogeneities in the magnetic field which were difficult to compensate using early NMR systems. Poorly manufactured tubes would cost less but have lower tolerances in concentricity (i.e., centering of inner bore of the tube properly in the glass) and/or camber (i.e., straightness of the tube). These imperfections during spinning would result in NMR spectral artifacts called spinning side-bands due to vibrations and imperfections in the magnetic field. High-quality NMR tubes can be purchased albeit with much higher prices.[8] For further information on shims (i.e., the small electro-magnets) used to optimize the magnetic field, please see references Liu et al. 2014; Maudsley et al. 1984; Van Zijl et al. 1994; van Zijl 1987. The basic results were narrower/taller peaks above the baseline noise with better resolution (less overlap). One would then logically wonder why we do not routinely "spin" metabolomics samples (or all NMR samples)? First the cost of the high precision tubes is a major barrier to large-scale studies, and even the best tubes would still experience some spinning-induced "side-bands" (i.e., individual or sets of symmetric artifactual peaks at specific distances from the real signal). These artifacts cannot be completely removed, resulting in additional peaks in the spectra. This causes confusion during interpretation, especially in heavily overlapped regions.

[7] https://blogs.umass.edu/weiguoh/?cat=81121.

[8] X and Y based shims (termed the non-spin shims) could be optimized more easily and the overall spectral line shape was normally much improved.

Modern instruments are now usually housed in excellent laboratory environments with exceptionally homogeneous magnetic fields due to refined manufacturing and extended shim sets, therefore spinning is no longer necessary. This has been published by Zerbe and Jurt (2013) and we confirm from our personal experiences. Lastly, many NMR experiments using pulse field gradients for coherence selection suffer dramatically from sample spinning.

Note

In our experience, the performance gains from spinning samples do not make up for the variability and problems associated with spinning side-bands. We do not use spinning nor recommend for metabolomics.

Additionally

The NMR tube manufacturing common practice of putting a magnetic field strength recommendation (e.g., eco tube for 400 MHz or below, or precision NMR tube for 600–800 MHz) assumes spinning of samples and is not necessary for non-spun and high throughput studies. In our experience, basic economical NMR tubes are sufficient in nearly all sample applications.

5.3.3 Gathering and Handling

Long before we can begin acquiring NMR data, we have to trust that all samples were handled uniformly prior to their arrival at the NMR preparation stage (Barton et al. 2008; Bernini et al. 2011; Dumas et al. 2006; Lauridsen et al. 2007; Pinto et al. 2014; Rist et al. 2013; van der Sar et al. 2015). Unfortunately, this is more difficult that may be initially considered, and we will present examples from our first human urine study.

Volunteers

First, can enough volunteers be found? How does one know how many or even how to appropriately attract volunteers for a study? While seemingly a trivial point it is no easy task with modern privacy laws, e.g., how do you legally even approach people without prior permission? A definite catch-22 situation. Our experience was this severely slowed down our study almost to the point of cancellation regardless of successful funding.

Note

How to get study volunteers? This is under jurisdictional control so carefully check your local requirements. We had to have research assistants set up tables on campus with signs and wait for individuals to approach. Privacy rules prevented pretty much every other idea we came up with (e.g., emails, offering prizes/incentives, telephone, etc.).

Another example of complication, volunteers needed to be pre-selected based on a predefined and often limited range of age, sex, weight, health, diet, medical history, fasting prior to sampling, a sufficient number of volunteers had to be selected for statistically relevance, and lastly enough extra volunteers needed to be found so that a sufficient number will successfully complete the study (there are always complications/withdrawing - see references above). All of these are no small feat. In our case physical activity was involved in the hypothesis, therefore volunteers needed to be selected for fitness/risk with full disclosure, etc. The question of number of samples per volunteer needs to be statistically addressed starting with power and sample size calculations (Jones et al. 2003) and we recommend a collaboration with someone specializing in this area along with analysis/interpretation of principle component data.

Long-Term Studies and Storage

For longer term studies, sufficient[9] samples must be acquired from each volunteer over the course of the study to properly address the hypothesis(es), e.g., establish baselines, and see statistically significant results. Additionally, appropriately confirmed preservatives and/or spiking using internal standards may be necessary for determination of time-dependent degradation. In our case, we needed to acquire samples to establish a minimal baseline prior to a physical stress-test, acquire samples during the test, and then re-establish return to equilibrium. This amplified the number of samples dramatically. Larger numbers of samples will require more long-term storage space, with suitable and stable temperature (i.e., −80°C), and complications on uniform handling/preparation thereafter. So many issues and we have not even gotten started yet with the NMR data.

If these storage requirements can be met, archived samples will still need absolute consistency in transport (time/conditions), preparation handling conditions, thawing temperature/time/exposure, addition of NMR referencing/internal concentration reference, pH balancing, addition of deuterated lock solvent, and then finally waiting time to acquire at the instrument. Extensive planning must be established prior to all stages. A full-time person handling all these aspects is ideal, but we know not always possible. Part-time or time-shared individuals must be even better organized in-order to coordinate their combined work. Sample preparation planning needs to include

[9]This is a loaded word, i.e., depending on study type, number of participants, metabolic reaction time, etc. just to name a few likely variables.

very precise details, e.g., how long does it take (and what temperature do samples reach?) for samples to transfer from long-term storage to "in the instrument." Prior agreement between individuals is key for consistency.

Note
While the final data acquisition and analysis (Schönberger et al. 2015) is often strongly emphasized in publications, we recommend focusing more on the initial preparation, handling, and organization as this will facilitate analysis.

Examples from practical experience; does one take all samples out of −80°C storage at the same time? We did choose to remove all samples at once, but immediately stored them in a normal +5°C refrigerator to thaw slowly and consistently (i.e., samples on the outside of the group did not thaw first etc.). We also had to address the preparation time difference between the first and last handled sample. Specifically, how long does the last sample sit in the spectrometer robot sample handling system waiting while all the previous samples are acquired, e.g., 7–8 min per sample with 100 submitted samples therefore the last sample sits ~800 min longer at room temperature than the first sample. Assuming bacterial inhibitors (Bernini et al. 2011; Lauridsen et al. 2007) are used, any chemical reaction and/or oxidation within the sample will undoubtedly be temperature and time sensitivity.

So there are really three key areas to consider. The first, what is the long-term storage available? Second, conditions (time, temp, etc.) from thawing to analysis. The last, how many times can the sample be frozen/re-frozen either for re-analysis or due to unforeseen delays (Pinto et al. 2014; Saude et al. 2006; Saude and Sykes 2007; Rist et al. 2013). The first concern is usually dealt with by storing at a consistent −80°C, and this has become the standard.

Note
Do not forget −80°C ultra-low temperature freezers are expensive, can have additional electrical requirements, produce a lot of residual room heating, and do fail. This means extensive infrastructure pre-planning and monitoring.

The second issue has a relatively simple solution, i.e., prep samples just prior to acquiring data and use lower ambient temperatures during prep. Specifically we recommend keeping samples at 5°C until just before instrument acquisition (and depending on your latitude/location you may also need to include humidity control). This includes appropriately selected robotic sample controls, and a consideration of total time waiting for acquisition (Saude and Sykes 2007) such as limiting the size and using a batch method for sample preparation. Preparing 40 samples at a time instead of 400 reduces the time differences, however this substantially increases the labor involved as more frequent/smaller batches must be prepared reducing lab efficiency and increasing costs. It was our experience that 20–30 samples per

batch, with two batches a day worked well allowing a single spectrometer to be utilized efficiently. This was especially important as we were restricted to weekend access and therefore had to maximize usage.

Note
While lowered temperature control while waiting on the instrument itself is beneficial, it adds complications for the NMR instrumentation. Thermal equilibrium (usually at room temperature) must be reached for the entire NMR sample prior to data acquisition in the spectrometer receiver coil.

Thermal equilibrium can take seconds to minutes depending on the instrument variable temperature controls. Pre-warming of samples may be necessary prior to insertion into the spectrometer to optimize instrument time (see Sect. 5.4.7 below).

5.3.4 In the Tube

Assuming the samples have been uniformly: selected, acquired, stored, and handled; the next step is preparation for NMR data acquisition. Liquid NMR samples are usually in 5 mm diameter NMR tubes, between 500 and 600 μL in volume, and have a concentration of somewhere in the 10's mM to μM range. Concentrations are approximate providing enough material to conclude ^{1}H NMR experiments in conveniently short periods (minutes to seconds, respectively). For a detailed review of relative and absolute NMR sensitivity see (Sanders and Hunter 1988). As we can add scans of the same sample together building up the *s*ignal-to-*n*oise (Hoult and Richards 1976) ratio over time NMR therefore has no theoretical detection limit. However, the S/N is proportional to the square of the number of scans taken (i.e., to double the S/N requires 2^2 times the number of scans). We cannot realistically expect anything resembling high throughput for samples if each requires multiple hours for a 1D experiment. NMR detection becomes even more limiting when considering multi-nuclear and/or multi-dimensional spectroscopy. Interested readers are referred to Bingol and Brüschweiler (2014), Dumez (2018), Gardner and Kay (1998), Hyberts et al. (2007), Silverstein et al. (2014), Zerbe and Jurt (2013), Ziessow (1990) and references therein.

Regarding the standard 5 mm NMR tube, various investigators have tried smaller tube diameters to reduce the sample volume. For example, Bruker Inc. has introduced a 1.7 mm NMR tube diameter probe (see NMR Probes section below) designed to maximize NMR signal detection from low volume samples. One of the purposes is to reduce sample storage/volume/preparation costs and increase "mass sensitivity." Other studies have adopted shorter and thinner 3 mm NMR tubes to reduce glass costs while optimizing throughput. Mass sensitivity is a bit of a tricky definition and assumes that the solute's solubility is not a limiting factor, not always the case.

For some samples, solubility is the key limiting factor, i.e., one cannot concentrate the solute further without precipitation and/or molecular changes occurring.

Here one does not gain “mass sensitivity” through concentrating into smaller diameter tubes. In these cases, a wider diameter sample tube physically allows more atoms (albeit at a lower “concentration” via a greater number of overall molecules) to be placed inside the observing receiver coils, and results in an improved signal. Conversely if solubility is not a limiting factor and the same molar mass of material can be concentrated into a smaller volume and smaller NMR tube, e.g., a 1.7 mm diameter NMR tube versus the larger 5 mm tube, then the atoms are physically closer to the receiver NMR coils and produce a larger relative signal. Smaller tubes also become advantageous with high salt samples and cryogenically cooled (i.e., cold) NMR probes. The S/N improvements of cold-probes are quickly lost (Kelly et al. 2002; Nausner et al. 2010; Xiao et al. 2009) as the ionic concentrations rise, but this phenomenon can be compensated for with narrower or specially shaped tubes.

Note
One substantial advantage for 5 mm NMR tubes is the increased ease of removing/recovering samples for further study and/or storage also the increased ability to wash and re-use the NMR tubes.

Groups may opt to dispose of 3 mm or thinner NMR tubes after each use due to cleaning/recovery difficulties and therefore the costs must be taken into consideration for large-scale samples studies.

5.4 NMR Spectrometer

5.4.1 Probe

The NMR probe is critical. The probe inductively delivers relatively high-power electromagnetic energy to the sample and influences the sample’s atomic state(s), therefore performing the NMR experiment (called a “pulse sequence”). The probe is also responsible for detecting the sample’s subsequent response(s), i.e., very low power inductive signals from the precessing atoms in the sample, due to the presence of a strong static external magnetic field (i.e., the NMR magnet).

One can easily imagine that the probe experiences a light impact from each sample and spinner as the air pressure is decreased and sample raised/lowered. Also contamination from any materials passed from the operator’s hands to the NMR sample, material (e.g., dust) from the air as the sample lowers, NMR tube breakage (e.g., multiple samples inserted without removal of the previous[10]) and wear due to high-power electromagnetic induction during pulse sequences to name but a few challenges. There are also physical moving electronic components (see

[10]For high throughout facilities under manual (i.e., non-robotic) sample handling, this problem occurs more often than you might think.

Sect. 5.4.8 below) (Halliday et al. 2013; Derome 2013) with the temperature, ionic strength, and dielectric constant of the solvent all coming into play. These aspects are important and cannot simply be considered “constant.” Even if the instrument is solely dedicated to the metabolomics user and not a multi-user facility (more common), the NMR instrumentation experiences wear-and-tear over time.

An improperly setup probe will not yield the expected signal amplitudes and may even rapidly devolve into relaxation and off-resonance effects, and far more complicated spectra outcomes.[11] On the receiver side, the resulting signal intensity will also suffer at best decreasing certainty of measurement, or worse preventing observation of the signal. Early metabolomics studies may not have had access to robotic sample handling that included automated tune/match (nor even regulated temperatures for samples awaiting data acquisition). Early studies assumed that the tune and match could be optimized for the first sample and would not deviate perceptively from sample to sample. In our experience, this assumption is not correct especially with human urine samples where the salt concentrations and therefore the dielectric matching conditions can change substantially with every sample.

Note

The practical result of improper tune/match (incorrect sample impedance matching) is inconsistent and inefficient energy transfer/excitation to the sample. Remember that metabolomics is all about consistency. It is our recommendation that any study include infrastructure capable of optimizing the tune/match for every entering NMR data acquisition.

Cryogenically Cooled Probes

Among some of the many recent NMR developments (Kovacs et al. 2005; Kupče 2007; Matsuki et al. 2015; Rovnyak et al. 2004; Webb 2006), one that has been particularly applicable to metabolomics is cryogenically cooled NMR probes. The cooling of the electronics has dramatically decreased the noise, and thereby improved the signal-to-noise ratio. Interestingly the same cooling technology utilized for the cryogenically cooled probes has also been leveraged into “cryogen-free” magnets, i.e., self-helium re-liquifying included inside the magnet, but unfortunately the electrical and annual maintenance costs of the cooling technologies (depending on size, manufacturer, and type of system) can make them financially impractical. The systems are variations on a compressor and/or gas expansion, each having advantages/disadvantages including vibration introduction into the spectrometer, and these vibrations can often be detrimental to the spectrometer performance.

For metabolomics the use of helium cryogenically cooled NMR probes (Webb 2006) and the associated increase in signal-to-noise (Schönberger et al. 2014) is often worth the complexity (Shishmarev and Otting 2011) and additional costly

[11] For those wishing a more detailed and mathematical description, see Chap. 23 especially Sect. 5 in (Brown 2016).

upkeep as the signal-to-noise ratio increases threefold–fourfold with a concomitant reduction in experimental time of 9–16 times. This is a substantial time savings and has been a major advance (Kovacs et al. 2005).

Note
Cryogenically cooled probes are often a logical and cost-effective upgrade to existing systems. However, probes and the application to magnets along with annual maintenance costs and supporting infrastructure requirements are not trivial so beware.

5.4.2 Console

The NMR console is the heart of the NMR spectrometer containing frequency generation, timing control, power supply, input/output controls, amplifiers, band selective equipment, relays, pneumatic controllers, thermal regulation, pulse field gradient generation, shim controls, etc. all the components necessary to perform the desired NMR experiments. The console, just like the NMR probe, requires constant monitoring, calibration, and inevitably some repair and/or replacement of parts. These must be done with full understanding of all consequences regarding instrument performance for metabolomics studies.

Note
We strongly recommend a professional staff person be utilized to initially calibrate and monitor the equipment at regular intervals to ensure consistency in the console performance.

5.4.3 Host Computer

The last supporting piece of equipment is the computer used by the operator to control and interact with the spectrometer. We will consider the communication hardware between the computer, console, and probe/magnet/supporting-infrastructure to be included. Like the previously mentioned equipment, the computer and software running the operating system and console are not always static. Anything from software updates incorporated for security and/or feature inclusion to spectrometer controlling software/firmware can go through changes. Any of these changes must also be taken into consideration by the metabolomics user as the spectrometer performance and/or saved NMR data may not be consistent and therefore introduce problems during processing and analysis.

5.4.4 Experiment Pulse Sequence

The order, number, repetition, carrier position, duration, magnitude (i.e., power), dynamics, and phase of any induced electromagnetic fields (called pulses), and any incorporated delays before/between/after these pulses along with the final read period to collect the observable data is collectively called the "pulse sequence" or "pulse program" (see example in Appendix A). This is complex and incredibly important regarding the consistency of the experiment used for data acquisition. Interested readers should see reference (Hore et al. 2015) and Chap. 13 of reference (Zerbe and Jurt 2013) for an excellent starting description. Then additionally references (Freeman 1997; Cavanagh et al. 2006) for further details. It is important to note that any change of the aforementioned settings, termed parameters (see this excellent review (Reynolds and Enriquez 2002) and our section below), can dramatically alter the experimental performance of the instrument (McKay 2011; Potts et al. 2001). This was also addressed by Saude et al. (2006) in 2006 when they carefully examined which pulse sequence to use for metabolomics. Types of pulse sequences for metabolomics were also included in the review by Beckonert et al. (2007) and more directly by others as interest in multi-dimensional and multi-nuclear experiments were explored (Van et al. 2003; Potts et al. 2001).

Pulse sequences can go through versions, and this can be a problem when software is updated as small changes can go unnoticed. Especially while the study data is being collected. It can be necessary to freeze a system in place and prevent software updates to ensure data collection integrity. For example, the most common metabolomics 1D-^{1}H NMR pulse sequence is the first increment of a 2D-^{1}H,^{1}H-NOESY (Bain et al. 1994; Blake and Summers 1990; Kumar et al. 1980), i.e., a one-dimensional data collection with no indirectly detected dimension. During one of our studies, the pulse sequence phase cycle (McClung 1999; Kingsley 1995; Kay 1995) (see also Sect. 5.4.6 below) was slightly altered, which dramatically changed the water solvent suppression (McKay 2009) and resulted in analog-to-digital overload errors, but only after the collection of the 16th free induction decay. Nowhere else in the data collection experienced a problem, and if one collected fewer than 16 scans (1/2 the phase cycle), there was no error and the data looked normal. This took quite a bit of time to discover and test the source of the issue. Eventually we had to retrieve an older pulse sequence version and rename it to prevent modification. This older sequence is now commonly known as the Chenomx "metabolomics-1D" (Lacy et al. 2014).

Note
Cursory comparisons of pulse sequence basics between versions or vendors will often miss internal phase cycle details. Even experienced NMR users can be caught unaware of the complexities involved.

5.4.5 Parameters

There are dozens to hundreds of parameters involved various pulse sequences. For example, even the simplest 1D "Pulse-Read" experiment (see Appendix A) requires the carrier position, power level, and pulse duration to be properly set. In addition, the length of the acquisition period and dwell time (i.e., time between sample-induced voltage readings in the receiver coil) will need to be properly selected so that required information is properly represented. Even this is a huge oversimplification. The more complex the manipulation of spins during the sequence the more parameters are involved, and the incorporation of multiple nuclei and/or multiple dimensions of indirectly detected dimensions will amplify the number of parameters exponentially.

We will focus on the most concise list of NMR parameters that need to be addressed for *each sample* involved in metabolomics NMR experiments. These are also necessary for testing and reproduction and should be included for publication of manuscript data.

We will first briefly address the practical outcome of parameters.

Error Bars and Statistical Relevance

One of the first metabolomics questions asked about NMR data is regarding error bars. This leads to a second question; how can a single NMR experiment be considered statistically relevant? For instrument error, we direct interested readers to reference (Sokolenko et al. 2013), where we attempted to address this question. In summary, the instrumental error is many orders of magnitude smaller than the error introduced during the assignment/integration/analysis phase depending on the operator experience and interest.

The second question is far more difficult. A single NMR experiment consists of tens or hundreds of thousands (or millions, e.g., solids NMR) of individual voltage measurements over several milliseconds or seconds of acquisition called the free induction decay or FID. These individual voltage observations are then repeated for each subsequent scan of the sample and added together. Example, a 500 MHz spectrometer with a 6,000 Hz sweep width, 4 s acquisition time, and 1 s recovery delay with presaturation of the solvent peak will typically have something on the order of 24 k complex acquire points (i.e., 24 k "real" and simultaneously 24 k "imaginary" collected at 90° offset from the first set making it even more complex) for quadrature detection. We will ignore the digital oversampling architecture of modern NMR consoles. Couple this with recording the experiment over-and-over on the same sample (~128 times) and then adding all those individual measurements together makes a statistical analysis surprisingly difficult. This does not even begin to touch on the self-artifact cancelling nature of phase cycling (see below) involved in the repetitive acquisition and how it relates to metabolomics (McKay 2009, 2011). Suffice to say that even a single NMR acquisition is statistically relevant and experimental error is extremely low (Sokolenko et al. 2013).

5.4.6 Phase Cycle

Though we have mentioned phase cycling before, it is important enough to further detail. While overall two NMR pulse sequences may look identical, the applied orientation (i.e., angle for each of the induced pulse or pulses and the receiver) can be individually controlled. This orientation of each component is termed the "phase"(Odedra and Wimperis 2012; McClung 1999; Kingsley 1995; Kay 1995; Cavanagh et al. 2006), and the total list of changes to any angle that each pulse/receiver is used for each FID over the course of the entire experiments is called the "phase cycle." This can also change with each repetition of the same experiment. Phase cycles are easily displayed in a table for each pulse/receiver on one axis with the scan number on the other axis. The phase cycle can be quite complex, e.g., 32 step phase cycle meaning that the experiment must be repeated 32 times to complete all phase changes. Experiments can usually be performed with far fewer (e.g., 4 acquired transients or scans) repeats of the data acquisition than the entire cycle requires, however there will be compromises and may result in instability in the resulting data. The larger phase cycle usually has more complicated and effective artifact suppression, e.g., collecting four scans of the sample and trying to compare to 32 scans of another sample with immediately have obvious S/N differences. There may also be differences due to partial phase cycle completion (McKay 2011).

Note

The phase cycle is used to cancel any general imperfections in the spectrometer receiver path, i.e., cyclops basic phase cycle. The higher orders of the phase cycle will usually attempt to remove smaller and smaller artifacts so one can usually use a subset or portion of the entire cycle. Consistency is again key, and one must determine the best selection prior to study start.

5.4.7 Temperature Calibration

Users often assume that the spectrometer displayed temperature is both accurate and precise for the sample once set via the spectrometer software. However, it has been our experience that this is almost never the case. Some instruments are certainly close, but there is always a benefit and often a substantial need for calibration (for examples, see Bernard et al. 2017; Raiford et al. 1979 and references therein). In our experience over the decades of instruments tested, once calibrated we see deviation of the set point to the actual temperature ranging from 1.6 to a few tenths of a degree Celsius in regulated temperature. This comparison is between what is measured via a calibrated sample and the software display. As metabolomics requires absolute consistency to identify small changes in sample composition and then equate those changes back to usable phenotypes, any variation in temperature will immediately create a problem.

Note

We are concerned with comparing data from instrument to instrument, but even if a single unchanging/dedicated instrument is used for the entire study, one would only have precision (i.e., consistent error). This may not be reproducible later even with the same spectrometer so relying on precision may not be sufficient. We highly recommend directly calibrating each instrument prior to study start, and after any hardware changes using referenced techniques (Holz et al. 2000; Karschin et al. 2022; Raiford et al. 1979).

Regarding sample temperature, Saude et al. (2006, 2007; Saude and Sykes 2007) published detailed tests of human urine stability versus storage time and temperature and we highly recommend interested readers consider this type of preparation for their samples to establish utmost consistency.

Note

To speed up data acquisition, we used a two-stage sample cooler. The first stage was long-term storage at ~5°C. The second stage was (using a golf analogy) termed "on-deck" and was a slightly higher temperature than the equilibrium temperature. See below for details.

Robotic Sample Handling

As the robot acquired the first sample, the otherwise idle robot would move the second sample from cold storage (e.g., 5°C) to the warmer on-deck region. The on-deck sample was set to slightly warmer than the NMR spectrometer (e.g., 30.4°C if samples were to be run at 30°C) so that as the sample was moved from the waiting on-deck position to the spectrometer, the slight cooling caused by transporting through the room air would be offset. The sample would arrive in the spectrometer receiver coil at the perfect temperature ready for equilibrium saving several minutes per sample. This could save hours of spectrometer data time per batch. The slightly warmer on-deck temperature was empirically calibrated based on average travel times and current room temperature so that the arriving sample temperature inside the spectrometer would be as close to ideal as possible. The on-deck temperature would vary throughout the year/seasons as the room temperature also changed.

5.4.8 Tune/Match

Frequency "tuning" and impedance "matching" of the sample and the spectrometer transmit/receive electronic circuit is critical for optimization. This ensures that the inductive moment or energy transfer going to, and the subsequent signal coming from the sample is as efficient as possible.

The concepts involved in impedance involve the resistors, frequencies, inductors, capacitance, mutual or self-inductance, etc. and goes far beyond this chapter (e.g.,

see Chaps. 32 and 33 dealing with these topics (Halliday et al. 2013)) and the authors capabilities. However in terms of practical NMR tuning and matching are still vitally important, and it is usually sufficient to use the common analogy that impedance matching involves adjusting the spectrometer circuitry to "match" that of the sample and the sample's dielectric constant of the solvent. Each solvent/sample can have different impendence characteristics (depending on temperature, salts, solvent(s), type of glass, ions in solution, etc.), so the probe must have an electronic range of capabilities for the nuclei to be observed and appropriate for the sample itself. Probes are usually designed with either manually manipulated, or automated movement of physical rods connecting deep inside the probe head. These rods move specific electronic components changing the transmitter and receiver's electronic circuit configuration.[12] A maximum transfer of power occurs when the two circuits (i.e., probe and sample) are identical in resistance, or in this case, due to alternating current, their impedance (Wilson 2007).

Note
With a poor "matching" condition, there can be substantial energy produced by the NMR console that does not enter the sample, but instead reflects, returning up the pathway and potentially damaging equipment.

This situation is commonly recognized by observing abnormally long NMR pulse widths that are inefficient along with poor S/N.

Interested readers are directed to excellent article from 1978 by Prof. David Hoult (1978) covering many aspects including the electronics of the probe and console design/function.[13] Other articles focus more on the practical implications and applications of matching (Bendet-Taicher et al. 2014; Nausner et al. 2010; Torchia 2009). An electronic-based description of impedance matching can be found in the ARRL handbook, e.g., Wilson 2007 or physics textbooks (Halliday et al. 2013; Feynman et al. 1965).

5.4.9 Excitation Pulse

Readers looking for further background are reminded of the previously reference books (Hoult and Richards 1976; Hore 2015; Silverstein et al. 2014; Keeler 2010; Levitt 2001; Freeman 1997; Harris 1986) and articles describing NMR and NMR theory. We will assume bulk coherence and transverse magnetization are understood.

[12] Capacitors and inductors (trying not to get confused with inductance nor impedance) of different sizes, types, and changeable physical positions are used to control analog frequency band widths and the subsequent matching of the circuitry.

[13] I had the pleasure of attending one of Dr. Hoult's talks as he was an invited speaker for the Alberta Cancer Foundation at the University of Alberta in 2015. His talk was amazing and the first time I really started to understand the difference between true NMR magnetic inductance of signal transmission versus the standard NMR analogy of radio signals and antennas.

For the simplest 1D-^{1}H direct detection NMR pulse sequence experiment (aka pulse-read) formed from a recovery delay, excitation pulse, and observation/data acquisition period, the only pulse power/duration consideration is the excitation pulse. As the reader may be aware, especially from the referenced literature, most spectroscopists use the maximum power (or close to it) that the probe can withstand. Therefore only the duration of the pulse application can be altered safely by the user. Consistent determination and application of the optimal excitation (i.e., movement of the bulk magnetization vector fully into the XY plane for detection by the receiver NMR coil in the probe) are critical for multi-pulse experiments, but the story can be more complex for the simplest pulse-read experiment. Examples of optimization are provided in these references (Bodenhausen et al. 1984; Breton and Reynolds 2013; Burrow et al. 2014; Reynolds and Enriquez 2002; Reynolds and Burns 2012). While some articles (Schönberger et al. 2014) recommend using only a 90° excitation pulse, there are complications due to relaxation. The "Ernst Angle"(Cavanagh et al. 2006; Freeman 1997; Keeler 2010; Levitt 2001) can be applied for simple experiments or more complex ones (Ogg et al. 1994; Zhang et al. 2000). The Ernst angle was originally developed to achieve the greatest S/N per unit time available. In our facility, the primary purpose of using less than a 90° pulse is to improve the accuracy of integration over all regions of the molecule. Different regions/atoms/ functional-groups may have different relaxation rates, and therefore additional scans (to improve the S/N) will result in different signal intensities/integrations based not just on the number of nuclei, but also their relative atomic and molecular mobility. Using a smaller pulse angle (e.g., 45°) reduces the time needed to fully relax, thus restoring the integrity of integration but with an S/N cost. For multi-pulse, e.g., multi-dimensional/multi-nuclear experiments involving subsequent magnetization precession and evolution with time, the optimal full 90° pulse is critical. Otherwise, there is a cumulative and compounding loss for each "imperfect" pulse.

Note

For metabolomics using the first increment of the 2D-^{1}H,^{1}H-NOESY experiment (aka metnoesy (McKay 2011)) which is standard for the Chenomx metabolomics software and database, it is essential that the calibrated 90° pulse is used consistently for all samples. Unless the solvent is highly consistent, unlikely with human urine/metabolomics samples, the 90° pulse needs to be determined for each sample prior to data acquisition.

5.4.10 Gain

Gain is essentially the volume control on the signal receiver. While we do not have time for the fundamental radio/receiver electronic aspects, suffice to say that too low a gain negatively impacts the experiment with poor S/N, while setting the gain too high risks overloading the receiver coil and damaging the FID data. Damaged data is not something that can be corrected later, and the experiment is forfeited. Overloads

occur when the voltage reading from the transverse recorded magnetization exceeds the electronic capability, and this is most often caused by a disproportionately strong solvent signal (e.g., water) or sample component (e.g., urea). Careful and consistent setting must be used to insure proper acquisition over a long-term metabolomics study.

Note
While it is not relevant to report the instrumental value of gain in the manuscript, it is important to note for the manuscript reviewer and readers how an appropriate gain was determined and if consistently used for all samples ensuring that receiver and/or analog-to-digital converter (ADC) errors were avoided.

5.4.11 Pulse Field Gradients

Pulse field gradients (PFGs) are an extraordinarily useful tool in NMR spectroscopy (Sakhaii et al. 2013; van Zijl and Hurd 2011; Zangger et al. 2001; Kay 1995; Keeler et al. 1994). The easiest analogy for PFGs is that we are slicing the length of the sample horizontally (think about slicing a sausage into thin circular cross-sections). Each cross-section experiences a unique magnetic field depending on the vertical position (i.e., isocromats). This allows spectroscopists to manipulate spins based on their relaxation, diffusion, or chemical shifts independently depending on where they started in the vertical sample.

While incredibly useful and often considered for solvent suppression in metabolomics, a common error occurs when excitation pulses and/or pulsed field gradient "eddy currents" have not subsided prior to data acquisition. If the "ringing" from the last pulse or gradient remains, the recording of the early data points will be distorted essentially by extra signal that does not originate from the sample, but instead the hardware. This is seen as baseline distortion and overall noise in the spectrum depending on how much hardware distortion leaks into the real data.[14]

We have found that PFGs disrupt the deuterium resonance peak used for the spectrometer deuterium "lock" (i.e., the automatic magnetic field compensation/stabilization). The instrument attempts to try and follow the deuterium signal during the initial application for PFGs, and then afterwards to re-gain the lock resonance causing field distortions. For many peaks, these distortions are small enough to ignore, but for the strongest/sharpest of the NMR peaks, the distortion appears as a dispersive NMR component. Unfortunately for DSS (Sheedy et al. 2010; Harris et al. 2008; Markley et al. 1998) commonly used as an internal reference standard and a peak shape normalization factor for the Chenomx software (Weljie et al. 2006), it

[14] A possible solution is to apply backwards linear prediction to the first few points, thus replacing the damaged data with realistic predictions based on the later non-distorted information. Not ideal but sometimes useful.

sees a large distortion on peak shape. Chenomx software uses the shape of the DSS methyl group to back-calculate corrections for any small shim imperfections. A distortion of the DSS methyl reference peak, however, can dramatically affect resonance integrations and the subsequent metabolite identification stage. There is no way to correct for PFG dispersive changes in the reference signal once collected (i.e., you cannot phase the distortion out without negatively impacting other resonances).

Note
For metabolomics, it is our experience that PFGs cause far more problems than they solve, and we highly recommend not utilizing pulse sequences with PFGs.

The easiest and therefore most common method is to remove all PFGs from the metabolomics NMR pulse sequence. This has been the case for nearly two decades now. While the use of alternating directional gradients can be used to reduce the artifacts (Nguyen et al. 2007; Sokolenko et al. 2013), validation of NMR metabolomics databases has not been done with any of the proposed pulse sequence changes.

5.5 Manuscript Reporting Parameters

To our knowledge, there is no standardized (e.g., IUPAC) recommendation for appropriately reporting NMR experimental parameters. Far too often when reviewing manuscripts, we find limited or nonexistent NMR experiment parameters (see Appendix B). An analogy would be stating you went on a trip in a "truck" and only specified the destination thus lacking all crucial information, e.g., what route was taken, how long were you gone, what type of truck, did you stop for fuel, if so how often, what speed, etc. to reproduce and confirm the journey. Stating in an experimental section, "Data was collected on a 500 MHz NMR spectrometer from manufacturer X" provides almost no usable information.

The pulse sequence/program used is essential and type/manufacturer of probe/ spectrometer is essential for the experimental section. Temperature, sweep/dwell time, etc. are all needed. The reporting should be considered as if a new group member would reproduce the exact experiment after reviewing the experimental write up. We detail certain key aspects below.

5.5.1 Pulse Width

In terms of reporting, it is easy to report the 90° pulse duration (usually in microseconds) however depending on the application (e.g., shaped pulses (Freeman 1998; Kupce and Freeman 1995; Morris and Freeman 1978; Prost et al. 2002))

Note
Without the 90° excitation pulse duration (often termed pulse width) or that a 90° pulse was used and stating the induced field strength(s) for the various pulses, we cannot reproduce the experiment. No reviewer should allow a paper into publication that does not have at least that minimum information.

5.5.2 Delays: Relaxation, Equilibrium

Relaxation, i.e., the return of signal to equilibrium between scans effects integrations. This makes molecule identification difficult and quantitation inaccurate. For metabolomics it will result in misassignment of molecules and/or missing the biomarker entirely.

Metabolomics relaxation has been studied, e.g., see Saude et al. 2006, Bakhmutov 2005 and references therein. Different pulse sequences will likely have different total relaxation times between scans. It is also important to understand that under these circumstances a single acquisition (i.e., 1 scan) will not necessarily give the same result when compared to the entire experiment (e.g., 32 scans) in terms of integration. The first point cannot be solely trusted to represent the final data that will be collected as the first scan usually has many seconds or minutes to establish spin equilibrium in the magnetic field.

Note
The key point here is consistency of relaxation times within the study. The number of scans, and other experimental parameters must be maintained for all the samples in the study.

The pulse sequence cannot be swapped out nor altered halfway through the study (e.g., choice of water suppression), as any change will affect the relaxation, water suppression, degree of excitation, etc. which will certainly be detected later in the processing/analysis phase.

Of equal importance is that one must conform to the pulse sequence delays required by the database(s) being used to determine sample content (see Sect. 5.4.6). For example, Chenomx Inc. (Mercier et al. 2011; Tredwell et al. 2011; Weljie et al. 2006) relies on the user running a 10 ms recovery time, with 990 ms solvent saturation at a controlled level (i.e., ~80 Hz gammaB_1 effective induced field strength see previous Sect. 5.4.9) precisely on resonance, followed by the excitation pulses, a 100 ms "mix" where saturation is turned back on, and finally the last 90° pulse and 4 s of acquisition. Total experiment time for a single scan is 5 s. Deviation from this will move the user away from accurate/useful integrations when comparing to the Chenomx database (Lacy et al. 2014; Mercier et al. 2011; Tredwell et al. 2011; Weljie et al. 2006).

5.5.3 Post-Acquisition: Weighting Function

The specific weighting function used should not be an issue, with the caveat that it is uniformly applied to all acquired data sets. We will recommend the selection of a weighting function matched to the observed decay of the acquired signal, and in our experience, a 0.5 Hz line broadening function works reasonably well, especially with the typical 1D-NOESY (metnoesy) NMR pulse sequence between 400 and 800 MHz (^{1}H). However, the selection of an appropriate apodization function will be left to the experts and undoubtedly some users will have strong reason for their selection.

Note
We have found that a great deal of problems can be avoided by using line broadening, e.g., 0.25–0.5 Hz.

Small shimming errors can be easily averaged or approximated out by the slight change. This must be included in the reported parameters for comparison and evaluation.

5.5.4 Corrections: Linear Prediction, Baseline, Phasing

We have found linear prediction to be of little value in 1D spectra, unless under specific hardware issues/circumstances usually outside metabolomics (e.g., long ring down times on cold-probes, etc.). We do not recommend, however, if used it must be explicitly stated in the manuscript.

Baseline correction can again be useful however as always in metabolomics, consistency is the key. We have found a general drift correction on a well-phased spectra to be relatively harmless. We typically avoid other corrections or try to correct the source problem either in the pulse sequence or spectrometer setup. For example, the pre-acquisition delays between the last excitation pulse and the opening of the receiver gate electronics.[15] Again any method must be included in the manuscript.

Phase correction (i.e., phasing) is commonly required, especially in 1D-NMR. There is a temptation to use automated phase corrections, but auto-phasing can be unreliable. It can also over emphasize the largest peak in the spectra which is most likely the solvent, and certainly not the most interesting and important resonance. Manual correction is more accurate, but introduces user error (precision), especially for inexperienced users.

[15] e.g. alfa and rof2 on older Varian/Agilent spectrometers with ddrtc coming into play on newer.

Note
Utmost effort should be taken to optimize NMR pulse sequence delays prior to the start of NMR acquisition. This care along with knowledgeable lock settings/setup should minimize the necessity of phasing and the introduction of possible user error.

5.5.5 Referencing

Referencing the zero point on the spectra is incredibly important for metabolomics. The databases typically use the chemical shift position of a peak to begin all auto-assignments. Even peak pattern recognition usually starts with the chemical shift of a unique resonance and tries to build out from there the possible coupling profiles. If the sample is referenced incorrectly the software will have infinitely more difficulty assigning peaks, let alone correctly.

As mentioned, the Chenomx software has several internal reference standards available checking the reference peak line shape for application of corrections to the rest of the spectra.

Note
The known amount of reference intensity also provides quantitative information for the integrations.

6 Future Ideas

6.1 Solids NMR Metabolomics

While metabolomics has been predominantly only liquids likely due to a focus on easily acquired samples (e.g., blood, urine, sputum, etc.), there is a growing amount of research into solids NMR utilizing high-resolution magic angle spinning (HRMAS) for metabolomics (Cheng 2007) and specifically Chap. 4 (Tilgner et al. 2019). We direct interested readers to the many excellent reviews of solid-state NMR (Laws et al. 2002; Ashbrook and Sneddon 2014; Reif et al. 2021) and references therein, solid-state cold-probes (Matsuki et al. 2015), and the relatively new field of dynamic nuclear polarization (DNP) enhancement (Albert et al. 2017; van Bentum et al. 2016; Matsuki et al. 2015). This is an exciting new area, and we look forward to developments.

6.2 Working with Raw FID

Dr. Krishnamurthy[16] has suggested a different route with *C*omplete *R*eduction to *A*mplitude *F*requency *T*able (CRAFT) NMR (Krishnamurthy 2013). As the author details, complexes and/or mixtures are very common and pose a difficulty for traditional NMR analysis. An example would be tailings "ponds" water used to deposit the remaining materials after oil extraction. We have seen spectra contain tens of thousands of compounds and those are only the ones in the lower molecular weight categories distinguishing themselves as sharp resonances.

Having a method that could identify and quantitate even complex mixtures would be amazingly beneficial and that is what Dr. Krishnamurthy has proposed in a 2013 publication. While risking oversimplifying, CRAFT does not use Fourier transformation to convert the time domain data into the commonly recognized NMR spectra. Instead it identifies the frequencies and amplitude of each of the raw components and reports them in a simple table ready for statistical analysis. Specifically they examine a fermentation broth for quantitative analysis and a spiked human blood plasma. The benefits of removing FT and operator error are certainly enticing. The method has yet to become prevalent with the majority of metabolomics analysis, likely due to users not being familiar with the technique nor confirmed when compared to traditional approaches.

6.3 PureShift ^{1}H-^{1}H J-Coupling Removal

One of the biggest potential new developments involves the PureShift style/family of NMR pulse sequences (Kiraly et al. 2018, 2021; Dumez 2018; Foroozandeh et al. 2018; Moutzouri et al. 2017; Castañar 2017; Mishra and Suryaprakash 2017; Kew et al. 2017; Zangger 2015; Aguilar et al. 2015; Foroozandeh et al. 2014, 2015; Mauhart et al. 2015; Reinsperger and Luy 2014; Kaltschnee et al. 2014; Paudel et al. 2013; Meyer and Zangger 2013; Aguilar et al. 2012; Aguilar et al. 2010; Zangger et al. 2001). There are many variations, iterations, and subsequent improvements, but all focus on removing the dipolar "coupling" or crosstalk between neighboring hydrogen atoms (as previously mentioned). PureShift sequences attempt to remove (or at least minimize) the couplings and complexity of the spectra. The variations attempt to compensate for the inherent weaknesses of the pulse sequence, i.e., signal-to-noise, and distortions in tightly coupled systems.

While these couplings are often crucial for organic chemistry and basic molecular identification in traditional NMR usage, for metabolomics they create greater spectral complexity with a multitude of overlapping congested spectral information. Chenomx software checks each peak position, relative integration, and coupling pattern when it attempts identification, and these have been validated for each

[16] "Krish" to friends and colleagues is an amazing resource for the NMR community and we gratefully acknowledge years of benefiting from this person's contributions.

molecule in question, in isolation, for each common spectrometer frequency. PureShift sequences could dramatically reduce the complexity and therefore the overlap of the NMR spectra. However, to use the Chenomx software with PureShift NMR, the molecule database would have to be entirely re-acquired. This would take a great deal of invested money and time.

The pulse sequences also require PFGs making the implementation more difficult for inexperienced users. Potentially more variability as well as upkeep and maintenance would be increased (see previous Sect. 5.4.11).

Lastly there is a large S/N reduction reducing the confidence in the data gathered and potentially missing weak signals. It is possible to compensate, but studies would be forced to substantially increase the number of scans (time) acquired for each sample. This would also re-introduce problems with sample storage/handling mentioned earlier as each sample would take more time on the instrument and robotic sample handling system.

7 Conclusion

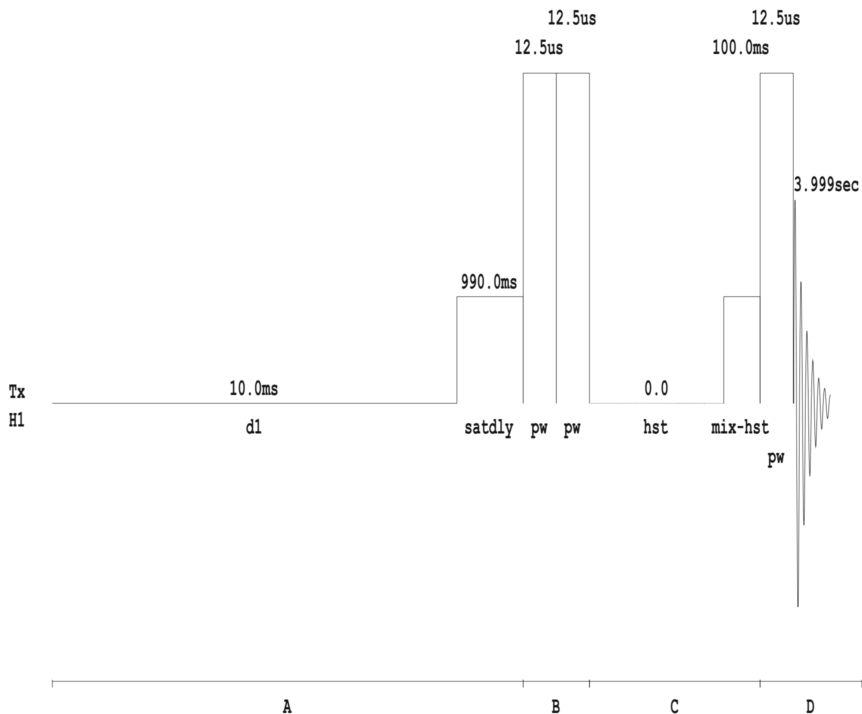
The most important point we can make is that there are many aspects requiring attention prior to initiating a metabolomics study via NMR. The information and experiences detailed above will hopefully help the reader avoid some of the common pitfalls that we encountered. We also hope that we have convinced new users that every single metabolomics study can benefit from having experienced knowledgeable spectroscopists involved, especially in the earliest planning stages. The chances of acquiring reliable/reproducible data increase exponentially when including the NMR operators. Unfortunately, these problems are often unrealized until the statistical analysis stage, when the errors due to confounders can exceed the actual data. The entire study may thus be ruined.

Consistency in every aspect is paramount. From sample acquisition, storage, preparation, handling, data acquisition, storage and/or retesting, to processing and analysis. There are many steps that can go astray, making the data more difficult to assess. However, organized from the start, carefully monitored throughout, and diligently systematically checking while acquiring the data will give the research group their best chance of having a useful data set at the end. Spectrometer consistency is paramount and mentioned above extensively, therefore users may want to explore the possibility of establishing quality control checks are regular intervals in their study.

Finally, we hope that the excellent referenced publications and books referenced will be as useful to your study as they have been to our education, though we are sure there are many others we have not yet discovered. Best of luck.

Acknowledgments The author would like to thank Professor Vladimir Michaelis (Department of Chemistry, Univ. of Alberta) for critical reading and suggestions on the writing of this document.

Appendix A: Example Pulse Sequence: Metabolomics 1D-^{1}H "Metnoesy"



Appendix B: Example Experimental Section

Experimental Section

NMR Samples

NMR samples were dissolved in 90% D_2O (~600 μL in volume) and sample data acquired in 5 mm XXXmodel NMR tubes purchased from XXXmanufacturer. All NMR solvents were purchased from XXX. NMR tubes were washed using three rinses of 95% ethanol with a single final D_2O rinse, and then inverted to air dry overnight.

NMR Spectroscopy

NMR experiments were collected on either a 14.1T (600 MHz) Varian/Agilent VNMRS with an Agilent 7,620 automatic sample handling system, or a 9.39T

(400 MHz) Varian Inova NMR spectrometer, both at 27°C (calibrated using methanol[17]). On the 600 MHz spectrometer, an HCN BioNMR probe (i.e., direct detect ^{1}H inner coil) was used, while on the 400 MHz instrument an AutoDB broadband (i.e., broadband inner coil) 5 mm probe was available. All spectra were run "locked" on the ^{2}H resonance signal and chemical shifts were referenced using the residual proton ^{1}HOD signal position[18] (i.e., 4.7 ppm) prior to saturation. One dimensional ^{1}H data was acquired using either presaturation[19] followed by a single excitation pulse/acquire spectrometer sequence, or the first dimension of a -2D-^{1}H,^{1}H-NOESY (metnoesy). The metnoesy (see Appendix A) uses a recovery delay of 10 ms, then 990 ms of presaturation followed by two 90° pulses, a mixing time of 100 ms with saturation, a final 90° pulse and lastly a 4 s acquisition period. The saturation pulse and carrier position were manually optimized and placed on the water resonance. Saturation was applied with a gammaB_1 induced field strength of 100 Hz (600) or 30 Hz (400) depending on water suppression efficacy and to avoid receiver overloads. For the simple 1D-^{1}H experiments, the duration of the saturation pulse was 2 s on both instruments. Parameter settings for all experiments were: sweep width of 7,183 Hz, acquisition time 2 s, with 28,736 real plus imaginary acquired (600), or a sweep width of 4,801 Hz. For the 1D-1H, an acquisition time of 3 s with 28,812 real and imaginary points for the 400 MHz NMR was utilized, and an ~30° (i.e., 3.4 μs) excitation pulse angle (applied at ~24 kHz gammaB_1) was used following the concept of the improved integration and reduced relaxation times (commonly known as the "Ernst Angle"),[20] however on some of the initial highly concentrated samples an extremely short pulse length was needed to avoid receiver overflows (e.g., 1us excitation pulse).

For processing of all NMR data, the acquired points were zero-filled to twice the number of acquired points, and a line-broadening apodization function of 0.5 Hz was

[17] D. S. Raiford, C. L. Fisk, E. D. Becker, *Anal. Chem. 51*, 2050 (2002).

[18] Wishart, D., Bigam, C., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J., and Sykes, B. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR, *J Biomol NMR 6*, 135–140.

Trainor, K., Palumbo, J. A., MacKenzie, D. W. S., and Meiering, E. M. (2020) Temperature dependence of NMR chemical shifts: Tracking and statistical analysis, *Protein Science 29*, 306–314.

[19] Hoult, D. I. (1976). Solvent Peak Saturation with SIngle Phase and Quadrature Fourier Transformation. *Journal of Magnetic Resonance*, *21*, 337–347.

Campbell, I. D., Dobson, C. M., Jeminet, G., & Williams, R. J. P. (1974). Pulsed NMR methods for the observation and assignment of exchangeable hydrogens: Application to bacitracin. *FEBS Letters*, *49*(1), 115–119.

[20] Lauridsen, M., Maher, A. D., Keun, H., Lindon, J. C., Nicholson, J. K., Nyberg, N. T. et al. (2008). Application of the FLIPSY pulse sequence for increased sensitivity in 1H NMR-based metabolic profiling studies. *Anal Chem*, *80*(9), 3365-3371.

Waugh, J. S. (1970). Sensitivity in Fourier transform NMR spectroscopy of slowly relaxing systems. *Journal of Molecular Spectroscopy*, *35*(2), 298–305.

Ernst, R. R., & Anderson, W. A. (1966). Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments*, *37*(1), 93–102.

then applied. Manual phasing and referencing to the solvent peak were used to confirm referencing based on the lock solvent (when available) and previously determined carrier position. Spectra were analyzed using VNMRJ 4.2 patch110 software.

References

Aguilar JA et al (2010) Pure shift 1H NMR: a resolution of the resolution problem. Angew Chem Int Ed Engl 49(23):3901–3903

Aguilar JA et al (2012) Decoupling two-dimensional NMR spectroscopy in both dimensions: pure shift NOESY and COSY. Angew Chem Int Ed Engl 51(26):6460–6463

Aguilar JA et al (2015) Minimising research bottlenecks by decluttering NMR spectra. Chemistry 21(17):6623–6630

Albert BJ et al (2017) Instrumentation for cryogenic magic angle spinning dynamic nuclear polarization using 90L of liquid nitrogen per day. J Magn Reson 283:71–78

Alseekh S et al (2021) Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat Methods 18(7):747–756

Amoureux JP et al (2008) Rapid analysis of isotopically unmodified amino acids by high-resolution (14)N-edited (1)H-(13)C correlation NMR spectroscopy. Chem Commun (Camb) 48:6525–6527

Andersson ER et al (2021) Identifying metabolic alterations associated with coral growth anomalies using 1H NMR metabolomics. Coral Reefs 40(4):1195–1209

Ashbrook SE, Sneddon S (2014) New methods and applications in solid-state NMR spectroscopy of quadrupolar nuclei. J Am Chem Soc 136(44):15440–15456

Athersuch TJ et al (2013) Evaluation of 1H NMR metabolic profiling using biofluid mixture design. Anal Chem 85(14):6674–6681

Bain AD, Burton IW, Reynolds WF (1994) Artifacts in two-dimensional NMR. Prog Nucl Magn Reson Spectrosc 26:59–89

Bakhmutov VI (2005) Practical nuclear magnetic resonance relaxation for chemists. Wiley, p 216

Balkir P, Kemahlioglu K, Yucel U (2021) Foodomics: a new approach in food quality and safety. Trends Food Sci Technol 108:49–57

Ban D et al (2017) Recent advances in measuring the kinetics of biomolecules by NMR relaxation dispersion spectroscopy. Arch Biochem Biophys 628:81–91

Bartel J, Krumsiek J, Theis FJ (2013) Statistical methods for the analysis of high-throughput metabolomics data. Comput Struct Biotechnol J 4:e201301009

Barton RH et al (2008) High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. Int J Epidemiol 37 Suppl 1:i31–i40

Beckonert O et al (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2(11):2692–2703

Bendet-Taicher E, Müller N, Jerschow A (2014) Dependence of NMR noise line shapes on tuning, matching, and transmission line properties. Concepts Magn Reson B Magn Reson Eng 44(1):1–11

Berger S, Braun S (2004) 200 and more NMR experiments: a practical course. Wiley-VCH, p 838

Bernard GM et al (2017) Methylammonium lead chloride: a sensitive sample for an accurate NMR thermometer. J Magn Reson 283:14–21

Bernini P et al (2011) Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. J Biomol NMR 49(3–4):231–243

Bingol K, Brüschweiler R (2014) Multidimensional approaches to NMR-based metabolomics. Anal Chem 86(1):47–57

Bingol K et al (2016) Emerging new strategies for successful metabolite identification in metabolomics. Bioanalysis 8(6):557–573

Blake PR, Summers MF (1990) NOESY-1-1-Echo spectroscopy with eliminated radiation damping. J Magn Reson 86:622–625

Bodenhausen G, Kogler H, Ernst RR (1984) Selection of coherence-transfer pathways in NMR pulse experiments. J Magn Reson 58(3):370–388

Boesl U (2017) Time-of-flight mass spectrometry: introduction to the basics. Mass Spectrom Rev 36(1):86–109

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

Braun S et al (1998) 150 and more basic NMR experiments: a practical course. Wiley-VCH, p 596

Breton RC, Reynolds WF (2013) Using NMR to identify and characterize natural products. Nat Prod Rep 30(4):501–524

Brown KC (2016) Essential mathematics for NMR and MRI Spectroscopists. Royal Society of Chemistry

Bull HB et al (1964) The pH of urea solutions. Arch Biochem Biophys 104(2):297–304

Burrow TE et al (2014) CRAPT: an improved version of APT with compensation for variations in JCH. Magn Reson Chem 52(5):195–201

Cassiède M et al (2017) Assessment of ^{1}H NMR-based metabolomics analysis for normalization of urinary metals against creatinine. Clin Chim Acta 464:37–43

Castañar L (2017) Pure shift ^{1} H NMR: what is next. Magn Reson Chem 55(1):47–53

Cavanagh J et al (2006) Protein NMR spectroscopy: principles and practice, 2nd edn. Academic Pres, Toronto, pp 1–885

Cheng LL (2007) Tissue and cell samples by HRMAS NMR. eMagRes

Der Wouden V, Cathelijne H et al (2020) Generating evidence for precision medicine: considerations made by the ubiquitous pharmacogenomics consortium when designing and operationalizing the PREPARE study. Pharmacogenet Genomics 30(6):131

Derome AE (2013) Modern NMR techniques for chemistry research. Elsevier, p 299

Duarte IF, Diaz SO, Gil AM (2014) NMR metabolomics of human blood and urine in disease research. J Pharm Biomed Anal 93:17–26

Dudzik D et al (2018) Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. J Pharm Biomed Anal 147:149–173

Dumas ME et al (2006) Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP study. Anal Chem 78(7):2199–2208

Dumez JN (2018) Spatial encoding and spatial selection methods in high-resolution NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 109:101–134

Dunn WB et al (2011) Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chem Soc Rev 40(1):387–426

Edison AS et al (2020) NMR: unique strengths that enhance modern metabolomics research. Anal Chem 93(1):478–499

Eghbalnia HR et al (2017) Increasing rigor in NMR-based metabolomics through validated and open source tools. Curr Opin Biotechnol 43:56–61

Ellinger JJ et al (2013) Databases and Software for NMR-based metabolomics. Curr Metabolomics 1

Emwas AH et al (2016) Recommendations and standardization of biomarker quantification using NMR-based metabolomics with particular focus on urinary analysis. J Proteome Res 15(2): 360–373

Emwas AH et al (2018) Recommended strategies for spectral processing and post-processing of 1D ^{1}H-NMR data of biofluids with a particular focus on urine. Metabolomics 14(3):31

Emwas AH et al (2019) NMR spectroscopy for metabolomics research. Metabolites 9(7)

Emwas AH et al (2020) NMR as a “Gold Standard” method in drug design and discovery. Molecules 25(20):E4597

Emwas A-H et al (2021) Pharmacometabolomics: a new horizon in personalized medicine. In: Metabolomics – methodology and applications in medical sciences and life sciences. IntechOpen
Esslinger S, Fauhl-Hassek C, Wittkowski R (2015) Authentication of wine by1H-NMR spectroscopy: opportunities and challenges. In: Ebeler SB et al (eds) ACS symposium series: advances in wine research, ACS symposium series. American Chemical Society, Washington, pp 85–108
Feynman RP et al (1965) The Feynman lectures on physics; vol I. Am J Physiol 33(9):750–752
Finco G et al (2016) Can urine metabolomics be helpful in differentiating neuropathic and nociceptive pain? A proof-of-concept study. PLoS One 11(3):e0150476
Foroozandeh M et al (2014) Ultrahigh-resolution total correlation NMR spectroscopy. J Am Chem Soc 136(34):11867–11869
Foroozandeh M et al (2015) Measuring couplings in crowded NMR spectra: pure shift NMR with multiplet analysis. Chem Commun 51(84):15410–15413
Foroozandeh M, Morris GA, Nilsson M (2018) PSYCHE pure shift NMR spectroscopy. Chemistry 24(53):13988–14000
Freeman R (1997) Spin choreography: basic steps in high resolution NMR. Oxford University Press, Oxford, pp 1–391
Freeman R (1998) Shaped radiofrequency pulses in high resolution NMR. Prog Nucl Magn Reson Spectrosc 32(1):59–106
Gardner KH, Kay LE (1998) The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct 27:357–406
Gibney MJ et al (2005) Metabolomics in human nutrition: opportunities and challenges. Am J Clin Nutr 82(3):497–503
Giraudeau P (2020) NMR-based metabolomics and fluxomics: developments and future prospects. Analyst 145(7):2457–2472
Giraudeau P et al (2014) Reference and normalization methods: essential tools for the intercomparison of NMR spectra. J Pharm Biomed Anal 93:3–16
Giraudeau P, Silvestre V, Akoka S (2015) Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. Metabolomics 11(5):1041–1055
Göbl C et al (2014) NMR approaches for structural analysis of multidomain proteins and complexes in solution. Prog Nucl Magn Reson Spectrosc 80:26–63
Halabalaki M et al (2014) Recent advances and new strategies in the NMR-based identification of natural products. Curr Opin Biotechnol 25:1–7
Halliday D, Resnick R, Walker J (2013) Fundamentals of physics. Wiley
Harris RK (1986) Nuclear magnetic resonance spectroscopy
Harris RK et al (2008) Further conventions for NMR shielding and chemical shifts (IUPAC recommendations 2008). Magn Reson Chem 46(6):582–598
Heather LC et al (2013) A practical guide to metabolomic profiling as a discovery tool for human heart disease. J Mol Cell Cardiol 55:2–11
Hiroaki H (2013) Recent applications of isotopic labeling for protein NMR in drug discovery. Expert Opin Drug Discovery 8(5):523–536
Holz M, Heil SR, Sacco A (2000) Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. Phys Chem Chem Phys 2(20):4740–4742
Hore PJ (2015) Nuclear magnetic resonance. Oxford University Press, p 112
Hore P, Jones J, Wimperis S (2015) NMR – the toolkit - how pulse sequences work. Oxford University Press, p 120
Hoult DI (1978) The NMR receiver: a description and analysis of design. Prog Nucl Magn Reson Spectrosc 12(1):41–77
Hoult DI, Richards RE (1976) The signal-to-noise ratio of the nuclear magnetic resonance experiment. Journal of Magnetic Resonance (1969) 24(1):71–85

Hyberts SG et al (2007) Ultrahigh-resolution (1)H-(13)C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. J Am Chem Soc 129(16):5108–5116
Izquierdo-Garcia JL et al (2020) Discovery and validation of an NMR-based metabolomic profile in urine as TB biomarker. Sci Rep 10(1):1–13
Jones S, Carley S, Harrison M (2003) An introduction to power and sample size estimation. Emerg Med J 20(5):453
Kaltschnee L et al (2014) "Perfecting" pure shift HSQC: full homodecoupling for accurate and precise determination of heteronuclear couplings. Chem Commun (Camb) 50(99):15702–15705
Karschin N et al (2022) Extension and improvement of the methanol-d4 NMR thermometer calibration. Magn Reson Chem 60(2):203–209
Kay LE (1995) Field gradient techniques in NMR spectroscopy. Curr Opin Struct Biol 5:674–681
Kay LE (2016) New views of functionally dynamic proteins by solution NMR spectroscopy. J Mol Biol 428(2 Pt A):323–331
Kay LE, Frydman L (2014) A special "JMR perspectives" issue: foresights in biomolecular solution-state NMR spectroscopy – from spin gymnastics to structure and dynamics. J Magn Reson 241:1–2
Keeler J (2010) Understanding NMR spectroscopy. Wiley
Keeler J et al (1994) Pulsed-field gradients: theory and practice. Methods Enzymol 239:145–207
Kelly AE et al (2002) Low-conductivity buffers for high-sensitivity NMR measurements. J Am Chem Soc 124(40):12013–12019
Kenny LC et al (2010) Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. Hypertension 56(4):741–749
Kew W et al (2017) Advanced solvent signal suppression for the acquisition of 1D and 2D NMR spectra of Scotch Whisky. Magn Reson Chem 55(9):785–796
Kingsley PB (1995) Product operators, coherence pathways, and phase cycling. Part III: phase cycling. Concepts Magn Reson 7(3):167–192
Kiraly P, Nilsson M, Morris GA (2018) Semi-real-time acquisition for fast pure shift NMR at maximum resolution. J Magn Reson
Kiraly P et al (2021) Single-scan selective excitation of individual NMR signals in overlapping multiplets. Angew Chem 133(2):676–679
Kohl SM et al (2012) State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics 8(Suppl 1):146–160
Korfmacher WA (2005) Foundation review: principles and applications of LC-MS in new drug discovery. Drug Discov Today 10(20):1357–1367
Kovacs H, Moskau D, Spraul M (2005) Cryogenically cooled probes – a leap in NMR technology. Prog Nucl Magn Reson Spectrosc 46(2–3):131–155
Krishnamurthy K (2013) CRAFT (complete reduction to amplitude frequency table) – robust and time-efficient Bayesian approach for quantitative mixture analysis by NMR. Magn Reson Chem 51(12):821–829
Kumar A, Ernst RR, Wüthrich K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun 95(1):1–6
Kupče Ē (2007) NMR with multiple receivers. eMagRes:721–732
Kupce E, Freeman R (1995) Close encounters between soft pulses. J Magn Reson 112A:216–264
Lacy P et al (2014) Signal intensities derived from different NMR probes and parameters contribute to variations in quantification of metabolites. PLoS One 9(1):e85732
Lasky-Su J et al (2021) A strategy for advancing for population-based scientific discovery using the metabolome: the establishment of the metabolomics society metabolomic epidemiology task group. Metabolomics 17(5):1–6
Lauridsen M et al (2007) Human urine as test material in 1H NMR-based metabonomics: recommendations for sample preparation and storage. Anal Chem 79(3):1181–1186

Laws DD, Bitter HM, Jerschow A (2002) Solid-state NMR spectroscopic methods in chemistry. Angew Chem Int Ed Engl 41(17):3096–3129
Lehnert W, Hunkler D (1986) Possibilities of selective screening for inborn errors of metabolism using high-resolution 1H-FT-NMR spectrometry. Eur J Pediatr 145(4):260–266
Lepre CA (2011) Practical aspects of NMR-based fragment screening. Methods Enzymol 493:219–239
Lesar CT et al (2011) Report on the analysis of common beverages spiked with gamma-hydroxybutyric acid (GHB) and gamma-butyrolactone (GBL) using NMR and the PURGE solvent-suppression technique. Forensic Sci Int 212(1–3):e40–e45
Levitt MH (2001) Spin dynamics: basics of nuclear magnetic resonance
Li K et al (2012) Establishment of a blunt impact-induced brain injury model in rabbits. Chin J Traumatol 15(2):100–104
Li Y et al (2020) Sub-nanoliter metabolomics via mass spectrometry to characterize volume-limited samples. Nat Commun 11(1):1–16
Lindon JC, Nicholson JK, Everett JR (1999) NMR spectroscopy of biofluids. In: Annual reports on NMR spectroscopy: annual reports on NMR spectroscopy volume 38. Elsevier, pp 1–88
Lindon JC, Holmes E, Nicholson JK (2007) Metabonomics in pharmaceutical R&D. FEBS J 274(5):1140–1151
Linge KL, Jarvis KE (2009) Quadrupole ICP-MS: introduction to instrumentation, measurement techniques and analytical capabilities. Geostand Geoanal Res 33(4):445–467
Link M et al (2014) Food analysis to check quality, safety and authenticity by full-automated 1H-NMR. ll1.workcast.net
Liu G et al (2014) Fast 3D gradient shimming by only 2×2 pixels in XY plane for NMR-solution samples. J Magn Reson 248:13–18
Lubbe A et al (2013) NMR-based metabolomics analysis. In: Lämmerhofer M, Weckwerth W (eds) Metabolomics in practice, Metabolomics in practice: successful strategies to generate and analyze metabolic data. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp 209–238
Macel M, Van Dam NM, Keurentjes JJ (2010) Metabolomics: the chemistry between ecology and genetics. Mol Ecol Resour 10(4):583–593
Markley JL et al (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. J Biomol NMR 12(1):1–23
Markley JL et al (2017) The future of NMR-based metabolomics. Curr Opin Biotechnol 43:34–40
Marshall AG, Hendrickson CL, Jackson GS (1998) Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom Rev 17(1):1–35
Marshall AG et al (2007) Fourier transform ion cyclotron resonance: state of the art. Eur J Mass Spectrom 13(1):57–59
Matsuki Y et al (2015) Closed-cycle cold helium magic-angle spinning for sensitivity-enhanced multi-dimensional solid-state NMR. J Magn Reson 259:76–81
Maudsley AA, Simon HE, Hilal SK (1984) Magnetic field measurement by NMR imaging. J Phys E Sci Instrum 17(3):216–220
Mauhart J et al (2015) Faster and cleaner real-time pure shift NMR experiments. J Magn Reson 259: 207–215
McClung RED (1999) Coherence transfer pathways and phase cycles: the decoding of a pulse sequence. Concepts Magn Reson 11(1):1–28
McGrath BM et al (2007) Unlike lithium, anticonvulsants and antidepressants do not alter rat brain myo-inositol. Neuroreport 18(15):1595–1598
McKay RT (2009) Recent advances in solvent suppression for solution NMR: a practical reference. In: Webb G (ed) Annual reports on NMR spectroscopy. Elsevier, pp 33–76
McKay RT (2011) How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: an examination of the pulse sequence components and evolution. Concepts Magn Reson A 38A(5):197–220

Meissner A et al (2014) 1H-NMR metabolic profiling of cerebrospinal fluid in patients with complex regional pain syndrome-related dystonia. Pain 155(1):190–196
Mercier P et al (2011) Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. J Biomol NMR 49(3–4):307–323
Meyer NH, Zangger K (2013) Simplifying proton NMR spectra by instant homonuclear broadband decoupling. Angew Chem Int Ed Engl 52(28):7143–7146
Mishra SK, Suryaprakash N (2017) Pure shift edited ultra high resolution NMR spectrum with complete eradication of axial peaks and unwanted couplings. J Magn Reson 279:74–80
Morris GA, Freeman R (1978) Selective excitation in Fourier transform nuclear magnetic resonance. J Magn Reson 29:433–462
Moutzouri P et al (2017) Ultraclean pure shift NMR. Chem Commun (Camb) 53(73):10188–10191
Nausner M et al (2010) Signal enhancement in protein NMR using the spin-noise tuning optimum. J Biomol NMR 48(3):157–167
Nguyen BD et al (2007) SOGGY: solvent-optimized double gradient spectroscopy for water suppression. A comparison with some existing techniques. J Magn Reson 184(2):263–274
Nikolaev EN, Kostyukevich YI, Vladimirov GN (2016) Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: theory and simulations. Mass Spectrom Rev 35(2):219–258
Niziol J et al (2021) Metabolomic and elemental profiling of human tissue in kidney cancer. Metabolomics 17(3):1–15
Odedra S, Wimperis S (2012) Use of composite refocusing pulses to form spin echoes. J Magn Reson 214(1):68–75
Ogg RJ, Kingsley PB, Taylor JS (1994) WET, a T1- and B1-insensitive water-suppression method for in vivo localized 1H NMR spectroscopy. J Magn Reson B 104(1):1–10
Pardi N et al (2018) mRNA vaccines – a new era in vaccinology. Nat Rev Drug Discov 17(4):261
Parsons HM, Ludwig C, Viant MR (2009) Line-shape analysis of J-resolved NMR spectra: application to metabolomics and quantification of intensity errors from signal processing and high signal congestion. Magn Reson Chem 47 Suppl 1:S86–S95
Paudel L et al (2013) Simultaneously enhancing spectral resolution and sensitivity in heteronuclear correlation NMR spectroscopy. Angew Chem Int Ed Engl 52(44):11616–11619
Perry RH, Graham Cooks R, Noll RJ (2008) Orbitrap mass spectrometry: instrumentation, ion motion and applications. Mass Spectrom Rev 27(6):661–699
Petrović M et al (2005) Liquid chromatography–tandem mass spectrometry for the analysis of pharmaceutical residues in environmental samples: a review. J Chromatogr A 1067(1–2):1–14
Picone G, Mengucci C, Capozzi F (2022) The NMR added value to the green foodomics perspective: advances by machine learning to the holistic view on food and nutrition. Magn Reson Chem
Pinto J et al (2014) Human plasma stability during handling and storage: impact on NMR metabolomics. Analyst 139(5):1168–1177
Potts BC et al (2001) NMR of biofluids and pattern recognition: assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse. J Pharm Biomed Anal 26(3):463–476
Prost E et al (2002) A simple scheme for the design of solvent-suppression pulses. J Magn Reson 159(1):76–81
Psychogios N et al (2011) The human serum metabolome. PLoS One 6(2):e16957
Qiu D et al (2020) Analysis of inositol phosphate metabolism by capillary electrophoresis electrospray ionization mass spectrometry (CE-ESI-MS). bioRxiv
Ragguett R-M, McIntyre RS (2020) Metabolomics in psychiatry. In: Personalized psychiatry. Elsevier, pp 459–464
Raiford DS, Fisk CL, Becker ED (1979) Calibration of methanol and ethylene glycol nuclear magnetic resonance thermometers. Anal Chem 51(12):2050–2051
Rasmussen LG et al (2012) Assessment of the effect of high or low protein diet on the human urine metabolome as measured by NMR. Nutrients 4(2):112–131
Reif B et al (2021) Solid-state NMR spectroscopy. Nat Rev Methods Primers 1(2):1–23

Reinsperger T, Luy B (2014) Homonuclear BIRD-decoupled spectra for measuring one-bond couplings with highest resolution: CLIP/CLAP-RESET and constant-time-CLIP/CLAP-RESET. J Magn Reson 239:110–120

Reynolds WF, Burns DC (2012) Getting the most out of HSQC and HMBC spectra. In: Annual reports on NMR spectroscopy: annual reports on NMR spectroscopy volume 76. Elsevier, pp 1–21

Reynolds WF, Enriquez RG (2002) Choosing the best pulse sequences, acquisition parameters, postacquisition processing strategies, and probes for natural product structure elucidation by NMR spectroscopy. J Nat Prod 65(2):221–244

Rist MJ et al (2013) Influence of freezing and storage procedure on human urine samples in NMR-based metabolomics. Metabolites 3(2):243–258

Rocca-Serra P et al (2016) Data standards can boost metabolomics research, and if there is a will, there is a way. Metabolomics:12–14

Rovnyak D et al (2004) Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. J Biomol NMR 30(1):1–10

Sakhaii P et al (2013) Broadband homodecoupled NMR spectroscopy with enhanced sensitivity. J Magn Reson 233:92–95

Sanders JKM, Hunter BK (1988) Modern NMR spectroscopy: a guide for chemists

Sandusky P, Raftery D (2005) Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: application to the metabonomics of amino acids in honey. Anal Chem 77(8):2455–2463

Saude EJ, Sykes BD (2007) Urine stability for metabolomic studies: effects of preparation and storage. Metabolomics 3(1):19–27

Saude EJ, Slupsky CM, Sykes BD (2006) Optimization of NMR analysis of biological fluids for quantitative accuracy. Metabolomics 2(3):113–123

Saude EJ et al (2007) Variation of metabolites in normal human urine. Metabolomics 3(4):439–451

Schönberger T et al (2014) Guide to NMR method development and validation-part I: identification and quantification. Eurolab Tech Rep. 01/2014

Schönberger T et al (2015) Guide to NMR method development and validation–part II: multivariate data analysis

Sheedy JR et al (2010) A sample preparation protocol for 1H nuclear magnetic resonance studies of water-soluble metabolites in blood and urine. Anal Biochem 398(2):263–265

Sherlock L, Mok KH (2019) Metabolomics and its applications to personalized medicine. Europe-Korea Conf Sci Technol:25–42

Shishmarev D, Otting G (2011) Radiation damping on cryoprobes. J Magn Reson 213(1):76–81

Shuker SB et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534

Silverstein RM et al (2014) Spectrometric identification of organic compounds, 8th edn. Wiley, Hoboken, p 464

Smolinska A et al (2012) NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Anal Chim Acta 750:82–97

Sokolenko S et al (2013) Understanding the variability of compound quantification from targeted profiling metabolomics of 1D-1H-NMR spectra in synthetic mixtures and urine with additional insights on choice of pulse sequences and robotic sampling. Metabolomics 9(4):887–903

Spraul M et al (2015) Wine analysis to check quality and authenticity by fully-automated1H-NMR. BIO Web Conf 5:02022

Staab JM, O'Connell TM, Gomez SM (2010) Enhancing metabolomic data analysis with progressive consensus alignment of NMR Spectra (PCANS). BMC Bioinform 11 123

Stringer KA et al (2016) Metabolomics and its application to acute lung diseases. Front Immunol:7–44

Tavares LC et al (2015) Metabolic evaluations of cancer metabolism by NMR-based stable isotope tracer methodologies. Eur J Clin Invest 45 Suppl 1:37–43

Tayyari F et al (2013) 15N-cholamine – a smart isotope tag for combining NMR- and MS-based metabolite profiling. Anal Chem 85(18):8715–8721

Teng Q (2012) Structural biology: practical NMR applications. Springer Science & Business Media

Tenori L, Turano P, Luchinat C (2020) Metabolic profiling by NMR. eMagRes 9:199–204

Tilgner M et al (2019) High-resolution magic angle spinning (HRMAS) NMR methods in metabolomics. In: NMR-based metabolomics. Springer, pp 49–67

Todd JFJ, March RE (1999) A retrospective review of the development and application of the quadrupole ion trap prior to the appearance of commercial instruments. Int J Mass Spectrom 190:9–35

Torchia DA (2009) Slight mistuning of a cryogenic probe significantly perturbs the water 1H precession frequency. J Biomol NMR 45(3):241–244

Tredwell GD et al (2011) Between-person comparison of metabolite fitting for NMR-based quantitative metabolomics. Anal Chem 83(22):8683–8687

Trifonova OP et al (2021) Mass spectrometry-based metabolomics diagnostics–myth or reality. Expert Rev Proteomics 18(1):7–12

Valdés A et al (2021) Foodomics: analytical opportunities and challenges. Anal Chem 94(1): 366–381

van Bentum J et al (2016) Perspectives on DNP-enhanced NMR spectroscopy in solutions. J Magn Reson 264:59–67

van der Sar SA et al (2015) Ethanol contamination of cerebrospinal fluid during standardized sampling and its effect on (1)H-NMR metabolomics. Anal Bioanal Chem

van Zijl PCM (1987) The use of deuterium as a nucleus for locking, shimming, and measuring NMR at high magnetic fields. J Magn Reson (1969) 75(2):335–344

van Zijl PC, Hurd RE (2011) Gradient enhanced spectroscopy. J Magn Reson 213(2):474–476

Van Zijl PCM et al (1994) Optimized shimming for high-resolution NMR using three-dimensional image-based field mapping. J Magn Reson 111:203–207

Van QN, Chmurny GN, Veenstra TD (2003) The depletion of protein signals in metabonomics analysis with the WET-CPMG pulse sequence. Biochem Biophys Res Commun 301(4): 952–959

Vignoli A et al (2019) High-throughput metabolomics by 1D NMR. Angew Chem Int Ed 58(4): 968–994

Visconti A et al (2019) Interplay between the human gut microbiome and host metabolism. Nat Commun 10(1):1–10

Vögeli B (2014) The nuclear Overhauser effect from a quantitative perspective. Prog Nucl Magn Reson Spectrosc 78:1–46

Wang X, Li L (2020) Mass spectrometry for metabolome analysis. Mass Spectrom Lett 11(2): 17–24

Wang H et al (2010) Metabolomics and detection of colorectal cancer in humans: a systematic review. Future Oncol 6(9):1395–1406

Webb AG (2006) Advances in probe design for protein NMR. In: Annual reports on NMR spectroscopy. Elsevier, pp 1–50

Weljie AM et al (2006) Targeted profiling: quantitative analysis of 1H NMR metabolomics data. Anal Chem 78(13):4430–4442

Wilson MJ (2007) The ARRL operating manual for radio amateurs. American Radio Relay League

Wolfender JL et al (2013) Plant metabolomics: from holistic data to relevant biomarkers. Curr Med Chem 20(8):1056–1090

Wüthrich K (1986) NMR of proteins and nucleic acids, vol 17. Wiley, New York, pp 1–292

Xiao C et al (2009) An optimized buffer system for NMR-based urinary metabonomics with effective pH control, chemical shift consistency and dilution minimization. Analyst 134(5): 916–925

Xu R et al (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. Proc Natl Acad Sci 96(2):388–393

Zangger K (2015) Pure shift NMR. Prog Nucl Magn Reson Spectrosc 86–87:1–20

Zangger K, Oberer M, Sterk H (2001) Pure-phase selective excitation in fast-relaxing systems. J Magn Reson 152(1):48–56

Zerbe O, Jurt S (2013) Applied NMR spectroscopy for chemists and life scientists. Wiley, p 548

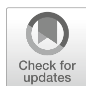
Zhang S, Yang X, Gorenstein DG (2000) Enhanced suppression of residual water in a "270" WET sequence. J Magn Reson 143(2):382–386

Zhang P, Georgiou CA, Brusic V (2018) Elemental metabolomics. Brief Bioinform 19(3):524–536

Ziarek JJ, Baptista D, Wagner G (2018) Recent developments in solution nuclear magnetic resonance (NMR)-based molecular biology. J Mol Med (Berl) 96(1):1–8

Ziessow D (1990) Understanding multiple-pulse experiments – an introduction to the product-operator description I. Starting with the vector model. Concepts in Magnetic Resonance 2:81–100

Zulyniak MA, Mutch DM (2011) Harnessing metabolomics for nutrition research. Curr Pharm Biotechnol 12(7):1005–1015



Natural Products Drug Discovery: On Silica or In-Silico?

Hye Kyong Kim, Young Hae Choi, and Robert Verpoorte

Contents

1 Introduction 118
2 Chemical Profiling Techniques in Metabolomics 122
3 Statistical Methods to Correlate Between Chemical Profiles and Bioactivity 125
3.1 Correlation with a Simplified Dataset: Discriminant Analysis 127
3.2 Correlation Analysis with Non-Discriminant Variables 128
4 Application of Metabolomics-Based Bioactivity to Various Disease Models 130
4.1 Antimicrobial Activity 130
4.2 Anticancer Activity 131
4.3 Anti-inflammatory Activity 135
5 Conclusions and Perspective 136
References 138

Abstract

Natural products have been the most important source for drug development throughout the human history. Over time, the formulation of drugs has evolved from crude drugs to refined chemicals. In modern drug discovery, conventional natural products lead-finding usually uses a top-down approach, namely bio-guided fractionation. In this approach, the crude extracts are separated by

H. K. Kim · R. Verpoorte (✉)
Natural Products Laboratory, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands
e-mail: h.k.kim@biology.leidenuniv.nl; verpoort@chem.leidenuniv.nl

Y. H. Choi
Natural Products Laboratory, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands

College of Pharmacy, Kyung Hee University, Seoul, Republic of Korea
e-mail: y.choi@chem.leidenuniv.nl

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_611

chromatography and resulting fractions are tested for activity. Subsequently, active fractions are further refined until a single active compound is obtained. However, this is a painstakingly slow and expensive process. Among the alternatives that have been developed to improve this situation, metabolomics has proved to yield interesting results having been applied successfully to drug discovery in the last two decades. The metabolomics-based approach in lead-finding comprises two steps: (1) in-depth chemical profiling of target samples, e.g. plant extracts, and bioactivity assessment, (2) correlation of the chemical and biological data by chemometrics. In the first step of this approach, the target samples are chemically profiled in an untargeted manner to detect as many compounds as possible. So far, NMR spectroscopy, LC-MS, GC-MS, and MS/MS spectrometry are the most common profiling tools. The profile data are correlated with the biological activity with the help of various chemometric methods such as multivariate data analysis. This *in-silico* analysis has a high potential to replace or complement conventional on-silica bioassay-guided fractionation as it will greatly reduce the number of bioassays, and thus time and costs. Moreover, it may reveal synergistic mechanisms, when present, something for which the classical top-down approach is clearly not suited. This chapter aims to give an overview of successful approaches based on the application of chemical profiling with chemometrics in natural products drug discovery.

Keywords

Antibiotics · anticancer · anti-inflammatory · Chemometrics · Correlation analysis · Discriminant analysis · *In-silico* · Metabolomics

1 Introduction

Since ancient times mankind has searched for medicinal plants in nature, in early times very likely by simple trial and error which eventually evolved into complex systems that included the documentation of the accrued knowledge which allowed its organized transmission from generation to generation. This is known as traditional medicine, some outstanding examples of which are the traditional Chinese Medicine (TCM) and Ayurveda in India. The traditional medicine of the Mediterranean region became the basis of Western medicine (Leonti and Verpoorte 2017). Thanks to the technological advances in processing and the accumulation of clinical information, the application of natural products developed from the direct use of a medicinal plant to the use of extracts, and then to pure compounds. Processing steps including grinding, drying, fermentation, extraction, heating, and the addition of other plants or materials (e.g., milk, butter, honey, sugar) have been, and still are, worldwide common practices in traditional medicine. This information is important for the transformation of the transmitted traditional knowledge into novel leads for medicines.

In Medieval times alchemists started to experiment with chemistry, looking for the essence of medicines, but it was not until the beginning of the nineteenth century that the first pure active compounds such as morphine, strychnine, quinine, caffeine, nicotine, atropine, and cocaine were isolated (Samuelsson and Bohlin 2009). However, the elucidation of the structures of these alkaloids took almost 150 years. At the end of the nineteenth century, the first synthetic medicines were produced, mostly based on models from nature, acetylsalicylate being the best known example. Thanks to the rapid advancement of technology in the twentieth century, many biologically active natural products have been isolated, some of which became important therapeutic aids per se, while many others served as a scaffold for the design of (semi)synthetic drugs.

No matter how we exploit nature for medicinal drugs, testing biological activity in various stages is an inevitable step for drug development, both for mixtures and isolated active components. In the twentieth century, pharmacologists developed various screening tools aimed at the detection of certain biological effects. Among these, the Hippocratic screening (Malone and Robichaud 1962) is still applied to screen plant extracts and pure compounds for a wide range of ailments in a systemic way directly on mice or rats. At some stage, the alleged pharmacological activity must be proven on animals and human beings. In the past 50 years, simple and fast in-vitro cell-, enzyme-, or receptor-based assays that can follow the activity during isolation or synthesis of novel active compounds have been developed. In the later stages of the evaluation of leads for drug development, detailed pharmacological studies of the mode of action are needed. This also includes *in-silico* docking of compounds in models for various receptors or enzymes.

The importance of bioactivity screening lies in the rapid activity mapping of large numbers of extracts or fractions that are highly complex mixtures of compounds. This can be done using bioassay-guided fractionation as an experimental design. The immanent paradigm of screening tests is a single target-single compound approach, e.g., measuring the degree of binding of a drug to a receptor or an enzyme. This approach was particularly successful in novel anticancer drug screening (Cragg and Newman 2005; Newman and Cragg 2007, 2020). This classical bioprospecting approach was one of the themes of the Nobel Prize in Medicine and Physiology in 2015, related to the random screening of soil bacteria and the resulting discovery of avermectin. This antibiotic in turn was the basis for a novel medication for river blindness. The other part of this Prize went to Chinese research that led to the identification of artemisinin, a novel antimalarial drug extracted from a millennial antimalarial traditional Chinese medicinal plant.

These conventional biological screening tests usually follow a top-down approach, starting from a mixture and narrowing down to a single compound which is eventually responsible for the alleged activity. However, a multitude of limitations persist. Among others, the long time it takes to find an active compound due to the number of steps involved in the identification of the features associated to activities. Moreover, in many cases the isolated active compound turns out to be an already-known compound. Furthermore, the long procedure required by conventional screening methods often leads to the loss of activity caused by chemical

degradation. But the greatest limitation of this method is its intrinsic ignorance of synergistic effects and prodrugs. The single target-single compound paradigm appears to contradict the base of traditional medicines that often relies on complex mixtures of plants in which every plant has a different function. That means that the multitarget-multicomponent approach is the key difference between traditional and modern drug development. It is obvious that such a systemic approach to drug discovery demands different tools. In the case of therapeutic activity it means a multitarget-testing system, i.e. the use of living organisms. For the identification of active compounds, a holistic approach such as metabolomics can provide the type of information required to identify correlations between activities and metabolites, thanks to the application of chemometric tools (see below).

Metabolomics aims at profiling all the metabolites in an organism. To these ends, it provides an overview of the metabolic profiles of the subject of interest in an untargeted manner, that is, ultimately, a systems biology approach (Wang et al. 2005). This systems biology approach using omics tools gained importance in the quest for biomarkers for diseases and other situations that require the identification of distinctive markers, e.g. quality standards. The experimental design for this approach includes processing a large number of representative samples (e.g., plant extracts or fractions) to obtain both metabolomics and activity data of each one. Numerous chemometric methods such as multivariate data analysis have been developed to identify potential correlations between all observations and identify indicators (e.g., NMR signals, LC-MS or GC-MS peaks) that correlate with a given activity. In this approach, any variable that has been measured can be used to test for possible correlations. Metabolomics, eventually in combination with proteomics, transcriptomics, and/or genomics, can also be used to identify potential targets in the tested organisms used in bioassays, e.g. cell lines or test organisms such as mice (Parng et al. 2002), zebra fish (*Danio rerio*) (Mushtaq et al. 2013; Akhtar et al. 2016), *Caenorhabditis elegans* (Salzer and Witting 2021), and brine shrimp (*Artemia salina*) (Ntungwe et al. 2020), it is possible to gain better insight into the possible mode(s) of action of an extract or a pure compound. This would be a molecular follow-up of the classic Hippocratic screening. The method must be validated by measuring the effect of various classes of drugs on the metabolome and a database of the changes in the metabolome can then be used to compare with the effects of extracts or novel compounds.

Finally, there is also the possibility of coupling a bioassay with the separation of extracts. For example, inhibition of acetylcholine esterase activity can be coupled with HPLC (Ingkaninan et al. 2000) or TLC (Rhee et al. 2003). Particularly in the case of TLC, its coupling with various bioactivity screening methods has shown to be a powerful tool for nano-scale identification of biologically active compounds (Klingelhofer et al. 2021; Morlock 2021; Morlock et al. 2021; Schreiner and Morlock 2021). In this case, though, the use of a single compound, single target approach, makes it somewhat less attractive for the study of complex traditional medicines, where synergy most likely plays a predominant role (Verpoorte et al. 2018). As seen in the study of *Artemisia annua*, the antimalarial activity observed in tea extract cannot be explained only by the already-known active compound

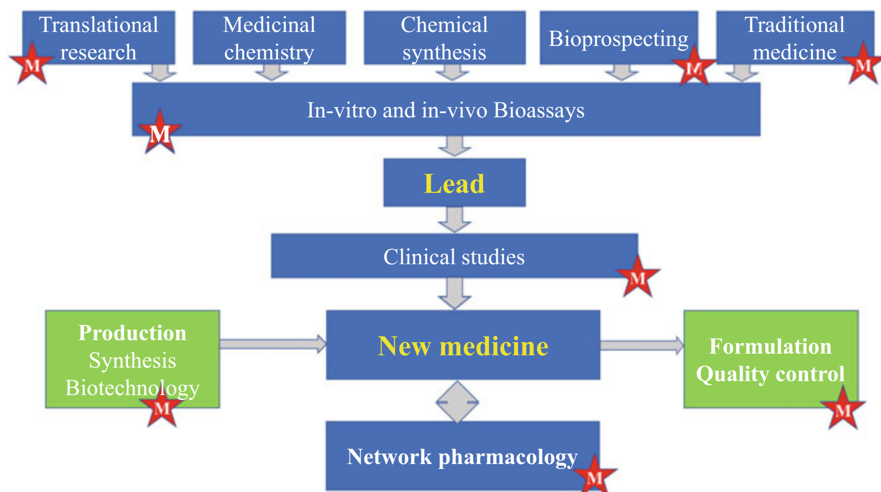


Fig. 1 Different steps in drug discovery and examples of metabolomics applied in this process. Adapted from Cuperlovic-Culf and Culf (2016)

artemisinin, as its concentration is too low in the tea extract to be responsible for the activity. The possible explanation is the presence of other active compounds such as flavonoids or saponins which could have a synergistic effect with artemisinin (De Donno et al. 2012). The limitation of this fully "on-silica" approach is that it is based on the single target-single compound approach, so any activity involving synergy and other interactions between molecules will not be detected, whereas the "*in-silico*" approach can reveal synergistic effects on other types of interaction-based mechanisms. The "in-silico" tools require the building of databases for fast identification of known compounds in extracts, and metabolomics databases of the changes caused by a standard set of major drugs in the in-vivo test systems. Network pharmacology would be useful to support such databases. Network pharmacology brings together all information on the effect of known drugs on the network of metabolism, signaling and diseases, on the level of genome, transcriptome, proteome, and metabolome. It can be used to obtain an insight into the mode of action of medicines and evaluate the differences and similarities of new drugs with existing ones (Boezio et al. 2017; Hopkins 2007, 2008; Jiang et al. 2021; Li and Zhang 2013; Ye et al. 2016). In Fig. 1 the various aspects of drug development are summarized, and the applications of metabolomics in these fields are highlighted. There are, basically, two types of applications. One applies metabolomics to various test organisms and patients aiming to identify markers for diseases and for the effects of (novel) medicines on their evolution. In the other case, metabolomics is used to identify active compounds in complex mixtures, e.g. traditional medicines. The systems biology approach opens new windows into the interactions between mixtures of active molecules and the test organisms providing even new insights

in the diseases themselves. This chapter describes the important features of metabolomics-based bioactivity screening using diverse activity models including antimicrobial, anticancer, and anti-inflammatory activities. The analysis of the chosen examples allows a discussion of the potential and the unresolved limitations of the method. Apart from the systemic approach chosen for the experimental design, we will also refer to new ideas for the fast dereplication of active compounds.

2 Chemical Profiling Techniques in Metabolomics

Chromatography became a major game changer in life sciences in the 1960s (TLC and GC) and 1970s (HPLC). An example of metabolomics avant-la-lettre is a paper by Baerheim Svendsen and Karlsen (1967) on the GC analysis of the essential oils of three different plants. However, in the 1980s this was considered to be like "collecting stamps." Fifteen years later, this sort of analysis was called metabolomics and included under the umbrella of "omics" technologies, in which the understanding of the function of genes was studied by the integration of genomics (DNA), transcriptomics (RNA), proteomics (proteins), and metabolomics data. The DNA and RNA molecules share similar physicochemical properties, enabling highly reproducible analyses based on strict robust extraction protocols. This is not the case of proteins that have a wide range of properties. In proteomics analysis, proteins are divided into two major classes, according to their solubility in water in hydrophilic and hydrophobic proteins and have thus different extraction protocols. In metabolomics, the situation is even more complex due to the great difference in physicochemical properties of metabolites in general. These can stem from their molecular size (e.g., polysaccharides, lignin) or even hydro/lipophilicity in the case of the small molecules. Among the "small" molecules there are primary metabolites, ubiquitous compounds that are found in all living cells but there are also specialized metabolites that are generally species-specific and related to the survival of an organism in its ecosystem. Altogether some 350,000 natural products are known from various sources (e.g., plants, microbes, insects) (Banerjee et al. 2015). Estimations of the number of specialized metabolites are mere speculations, but knowing that the number of species of living organisms in the world is somewhere between 10 and 100 million (Pimm et al. 1995) and assuming that every species produces one unique compound, there should be between 10 and100 million metabolites in the metabolome of the Earth. The total chemical space of natural products is thus very much larger than that covered by our present knowledge. There is still so much to be discovered!

This brings us to the problem of how to define a metabolome. We have just referred to the metabolome of the planet earth, i.e., the sum of all the metabolomes of all the organisms (dead or alive) and their environment. Each plant species has its own metabolome, which is the sum of the metabolomes of all its organs, i.e., roots, stem, and leaves, etc. These metabolomes correspond to that of the many different cells in the plant. Even within the cell the different organelles have different metabolomes. The metabolome of a species might be best defined as the measurable

set of metabolites that an organism is able to produce, i.e. some compounds might not always be present in certain plant parts, but their presence depends on the existence of certain circumstances. The metabolome of a species depends on the species' genome. Obviously the single cell analysis at all omics levels is the holy grail of the omics.

Clearly, any reference to metabolomics requires a clear documentation of the source of the "metabolome." The ultimate goal of metabolomics is to describe the changes in the metabolome under well-defined conditions. Eventually the fluxes in the organism through the various metabolic pathways should be analyzed (fluxomics), as changes in the homeostasis of an organism are connected with changes in fluxes.

To attain the goal of metabolomics, all applications have the same requirements, i.e. both qualitative and quantitative information about all metabolites in the biological sample must be generated. However, as per the discussion above it is clear that the requirements vary according to the application and so will the method needed to obtain the information. Firstly, when searching for biomarkers of diseases or bioactivities, it is essential to be able to count on a large database with metabolomic data of the test organisms to be used or material that has the desired activity, e.g. extracts from a specific plant species with an interesting activity. This database should ideally provide information on the effect of all biological variables on the studied material, for example, the diurnal variation of the metabolome. Existing databases with all kinds of information from previous experiments that provide abundant background information can help in the identification of compounds from their spectra and/or chromatographic data. However, bioprospecting requires the screening of large numbers of plants that are characterized by a highly variable and species-specific metabolome, particularly in the case of specialized metabolism, a situation that is clearly different from the case of mammalian cells, for example, in which most compounds are well-known primary metabolites.

It is now generally accepted that no single analytical tool can reveal the real metabolome (Wishart 2008; Emwas et al. 2019). At present, analytical chemists are moving in two possible directions to circumvent this limitation, i.e., improving the quality of individual techniques and/or integrating data from multiple methods.

In the case of NMR, the main limiting issues are its low sensitivity (μmol for detection limit) and complexity of signals. Sensitivity has been increased using cryo (cold)-probes and reducing the diameter of NMR tubes (capillary-, micro-, or nano-tubes). Nevertheless, the sensitivity is still far below that of MS-based methods (up to 10 times for conventional NMR probe). Generally speaking, NMR is considered to be the most powerful analytical tool for structure elucidation of pure compounds but not necessarily in the case of mixture analysis. The most attractive advantage of NMR as a metabolomics tool is its ease of quantitation. The height or area of ^{1}H NMR signals is directly proportional to the molar concentration of analytes, i.e. with a single internal standard all signals can be easily integrated and quantified. It is in this aspect that NMR has a clear edge over all other analytical techniques.

It should also be noted that NMR is currently the only metabolomics method that does not include a separation step.

In MS-MS, the separation is in the first MS dimension, based on molecular mass, after which the individual signals are subject to a second MS step in which the molecules are subject to fragmentation. In hyphenated chromatography-MS techniques, the retention time, molecular mass, and the fragmentation pattern can be used to search databases for already-known compounds. For example, it is possible to identify unknown peaks with high-resolution mass measurements in combination with the "seven golden rules" or other "dereplication" approaches. Though (stereo) isomers will be difficult to identify with 100% certainty, unfortunately too often the identity of a compound is accepted ignoring the inherent level of uncertainty of the used method. For a full identification of new or rare compounds, their UV, MS, ^{1}H NMR, ^{13}C NMR, and various multidimensional NMR spectral information must be analyzed. The 1D-NMR data alone are insufficient. Most cases require multidimensional NMR methods such as COSY, TOCSY, HSQC, and HMBC spectroscopy to determine distances and interactions between protons, e.g. to confirm stereochemistry. Known compounds can be identified with a reasonable certainty if isolated, but the full structural elucidation of new specialized metabolites from natural products mixtures is still enormously challenging. The technical limitations of the instrumental NMR analysis were partially solved by a statistical signal correlation of ^{1}H NMR resonances, the so-called statistical total correlation spectroscopy (STOCSY) that has been used to identify novel metabolites in urine, enabling the selection of signals characteristic of one molecule; there are, as well, a multitude of deconvolution methods for NMR data (Cloarec et al. 2005). As mentioned before, most of the chemical diversity of metabolites includes specialized metabolites of which there are no reliable comprehensive NMR databases. However, there are a number of available NMR spectra databases of primary metabolites (Human Metabolome Database (HMDB, http://www.hmdb.ca), Biological Magnetic Resonance Bank (BMRB, http://www.bmrb.wisc.edu/metabolomics/), NMRshiftDB (http://nmrshiftdb.ice.mpg.de/).

Mass spectrometry (MS) is a highly sensitive method (pmol level) and is also selective due to the high resolution and level of accuracy of the determined molecular mass and the different fragmentation patterns of the metabolites. The number of detected signals in MS-based platforms is 10–100-fold that of NMR. In terms of identification, however, it has inherent limitations. The mass spectrometer only detects ions formed in the ion source, but there can be a large variation in the sensitivity for the formation of ions. That means that absolute quantitation is only possible by running calibration curves for every single compound within certain ranges of concentrations. Identification is possible by comparison of the exact molecular mass of the compounds with databases and comparison of fragmentation patterns with possible candidates and closely related compounds. The MS data is insufficient for the structural elucidation of new compounds. For low molecular mass values there are scores of isomers, thus for the elucidation of the full structure of a molecule, including its stereochemistry, further spectral data are needed (see above).

Recent improvements of MS in metabolomics have risen from signal deconvolution of MS data rather than sensitivity or accuracy which is already remarkably high. This means that the focus has moved to the identification of compounds by increasing resolution and developing algorithms that can deduce molecular formulae from adduct ions based on a classification of molecules. Increasing resolution is mostly associated with improved mass analyzers, such as a quadrupole, TOF, Orbitrap, or iontrap. In the early times, the key issue was to improve accuracy of ion mass by suppressing fragmented ions. But the trend is changing. The goal is now to keep fragment signals together with molecular ones to use them for fingerprinting. This additionally offers the possibility of comparing the spectra to reveal the shared fragments of compounds which may point to a basic structure that is similar to all. Recently, many manufacturers offer improved ion trap analyzers to produce robust MS/MS signals that are the key for the identification with matching techniques. Unlike NMR signals, MS data do not provide any indication of similarity between signals. Jeffryes et al. (2015) have suggested the use of Metabolic *In-silico* Network Expansions (MINEs) as a new tool for identifying metabolites from an LC-MS dataset. MINEs used generalized biochemical transformations to propose structures, leading thus to the suggestion of putative metabolite structures.

Another dereplication approach is the use of molecular networking (MN) (Yang et al. 2013). This statistical correlation of MS/MS data is based on a generated MS/MS database of metabolites which can be searched for fragments of known metabolites in the biosynthetic network. The construction of a molecular network is based on the analysis of MS/MS spectra of compounds, presuming that molecules with similar structures should display similar fragmentation patterns. Allard et al. (2016) demonstrated the use of MN combined with other *in-silico* MS/MS fragmentation database as a dereplication strategy of the metabolites from natural sources.

The comparison of different profiling tools is summarized in Table 1.

3 Statistical Methods to Correlate Between Chemical Profiles and Bioactivity

The key step of OMICS-based chemical profiling is data mining using statistical analysis. The metabolomics datasets generated by NMR or MS are vast, requiring chemometrics to extract any useful information from the data. The chemometrics methods usually include a statistics-focused approach and/or a computer-dominant approach such as machine learning (Wishart 2008). Paul and de Boves Harrington (2021) summarized the basic concepts and applications of those methods very clearly.

In this section, we will focus on the statistical approach which is generally used to detect the active compounds in natural products extracts. This approach uses two different methods to link signals of compounds with activity: discriminant analysis or correlation analysis.

Table 1 Comparison of different technologies used in metabolomics

Technology	Advantages	Disadvantages
NMR spectroscopy	• Non-destructive • Rapid (ca. 5 min per sample) • Simple sample preparation – Requires no derivatization – Requires no extraction in case of biologic fluids • Detects broad range of compounds • Quantitative • Strong structural elucidation and identification power • Robust and reproducible • Compatible with liquids and solids	• Low sensitive • Requires large sample size • Cannot detect inorganic ions and non-protonated compounds • Expensive equipment
GC-mass spectroscopy	• Robust • High resolution and sensitivity • Requires modest sample size • Detects most organic and some inorganic molecules • Excellent database for identification	• Destructive (Sample not recoverable) • Requires sample derivatization • Requires individual calibration curve for absolute quantitation • Takes longer time (ca 30 min per sample) • Limitation on the novel compound identification
LC-mass spectroscopy	• High sensitivity • Requires minimal sample size • Relatively easy sample preparation • Deconvolution overlapping features	• Destructive • Requires individual calibration curve for absolute quantitation • Takes longer time (ca 30 min per sample) • Limitation on the novel compound identification • Limited software available and databases for (secondary) metabolite identification
MS and MS/MS	• High resolution & reproducibility • Relative quantitation • Large databases for metabolite identification • Can be used in metabolite imaging single cells (MALDI)	• Destructive • Requires individual calibration curve for absolute quantitation • Limited software available and databases for (secondary) metabolite identification

NMR nuclear magnetic resonance, *GC* gas chromatography, *LC* liquid chromatography, *MALDI* matrix-assisted laser desorption/ionization

3.1 Correlation with a Simplified Dataset: Discriminant Analysis

Discriminant analysis is nowadays the most popular approach for the identification of active compounds in complex mixtures. Partial least squares discriminant analysis (PLS-DA) is frequently applied to detect differences by making qualitative classifications, such as active and non-active samples; it is based on the definition of preferably two classes, and any type of classes can be formulated. PLS-DA, a supervised method, models the variance within the dataset by statistically discriminating groups of observations. Orthogonal PLS (OPLS) was developed to improve the correlation power of PLS by orthogonalizing non-related variables through orthogonal signal correction (OSC filtering). This facilitates the interpretation of the model because the variables (metabolome) are related to the targeted property (e.g., high or low bioactivity).

A good example of this application, published by Cardoso-Taketa et al. (2008), is the correlation of the sedative effect of *Galphimia glauca* Cav. and galphimine, a triterpenoid already known to occur in this plant. Working on six different collections of *G. glauca* they found two collections that were highly active, while the rest did not show much activity. Applying PLS-DA to the NMR metabolomic data and the sedative effects found in the animal model, they were able to correlate the biological activity with galphimine.

A similar approach was applied to identify biologically active metabolites obtained with different extraction methods and solvents from *Ocotea odorifera* using MS-based metabolomics. Alcântara et al. (2021) investigated the anti-inflammatory activity of a decoction of *Ocotea odorifera* and different fractions from its ethanolic extracts using dual inhibition of edema and neurophil recruitment. The chemical profiling data obtained by UPLC-HRMS or GC-MS were correlated with the anti-inflammatory activity by PLS-DA, resulting in the identification of S-(+)-reticuline as an active principle. Its activity was confirmed by testing the pure compound after isolation.

Other examples of the successful application of this approach are the studies of the antimicrobial activity of essential oils (Maree et al. 2014) and the identification of an antitussive active compound from *Tussilago farfara* L. (Lia et al. 2013).

The work of Maree et al. (2014) showcases the strength of the chemometrics and in particular of the OPLS-DA method. Based on the data obtained from the analysis of 158 different essential oils by GC-MS and their antimicrobial activities on several different strains of microorganisms, two classes were defined: active (MIC <2 mg/ml) and non-active (MIC >2 mg/ml). This allowed the detection of eugenol as a putative marker for activity, and an apparent synergistic effect with geraniol. α-Pinene, limonene, and sabinene, other components of the essential oils correlated with no activity. The study also revealed both antagonistic and synergistic antimicrobial effects between monoterpenes and eugenol on some microbes.

The correlation approach is clearly more efficient than conventional bioactivity-guided fractionation, to identify active compounds in the initial stage of drug development. The limitation of this application, however, lies in a tendency to overfit if the number of variables considerably exceeds the number of samples

(Gromski et al. 2015). In other words, in many cases there is a risk of a model showing a significant separation mainly by chance. This problem can be partially solved by increasing the number of samples which in turn increases the demand on the robustness of the analytical methods to, i.e. their ability to provide highly reproducible results.

The possibility of overfitting in PLS- and OPLSD-DA can be tested by proper validation methods. Commonly used validation tests are the permutation test or the CV-ANOVA test (Szymańska et al. 2012). Once the model is validated, corresponding discriminating components can be further investigated to consider the responsible metabolites for the activity, preferably at single compound level. The validation methods for the correlation tests have been well reviewed by Westerhuis et al. (2008) providing details of their advantages and limitations.

Other limitations are associated with the quality of biological data. Ideally, the number of data processing steps should be reduced as it has been shown that there is an inverse relationship between the number of steps and closeness to an intact correlation. However, it is not always possible to have two clearly activity-distinguished groups since most biological data have a much lower degree of resolution than chemical data. Thus, original data are often grouped into simple classes, for example, active and non-active groups that severely misrepresent reality. For example, Maree et al. (2014) considered the samples with MIC values below 2 mg/ml as active and non-active above this value. It is challenging to set up the proper criteria, in fact, in most cases it is based on trial and error. Although the simplified variables can reduce the ambiguity in the correlation between two datasets, there is clearly some degree of overfitting. Instead of dividing groups by different criteria, it could be better to work with the quantitative data obtained from the biological activity tests and correlate the activity directly without the input of supervised group information.

3.2 Correlation Analysis with Non-Discriminant Variables

In a correlation analysis, the relationship between two different datasets, e.g. metabolome data (mostly X-variables) and biological data (Y-variables, quantitative), is determined without simplifying original variables. Among the available methods, partial least squares to latent structures (PLS) modeling is the most popular. In this method, biological data are the quantitative and continuous variables, if necessary with some post-processing, e.g. logarithmic transformation of the biological data. Many studies have been done using this approach to identify active metabolites.

One example is the study published by Yuliana et al. (2011) who identified two flavonoids as responsible for the adenosine A1 receptor binding activity of *Orthosiphon aristatus* (Blume) Miq. (synonym *Orthosiphon stamineus*) based on the correlation of its NMR-based metabolome and the tested activity. In their study, a novel gradient extraction method was applied. In this method, instead of a single solvent, a stepwise polarity gradient from, e.g., ethyl acetate via methanol to water

was pumped through the ground dry plant material mixed with kieselguhr (7:1), yielding between 20 and 30 fractions. Each fraction was profiled by ^{1}H NMR and its potential anti-obesity activity was assessed based on an adenosine A1 receptor binding assay. Examination of the loading plots of the PLS and OPLS models led to the identification of two active methoxyflavonoids, 4′,5,6,7-tetramethoxyflavone (tetramethyl scutellarein) and 3′,4′,5,6,7- pentamethoxyflavone (sinensetin).

Another example is the work of Ali et al. (2013), who applied PLS and OPLS analyses to correlate the metabolome of sponges obtained using NMR spectroscopy with their effect in the adenosine A1 receptor binding activity assay. The metabolome of over 200 sponge samples was obtained using ^{1}H NMR, and their adenosine A1 receptor binding activity was measured. OPLS analysis was performed with two datasets. From the loading plots several signals were selected as activity-associated. Most signals were unknown or difficult to identify only by ^{1}H-NMR, requiring the isolation of compounds. Having isolated several compounds which matched well with activity-associated signals, these active metabolites were then identified as the sesterterpenes, halisulfate -1, -3, -4, -5, and suvanine.

De Melo et al. (2020) studied the spasmolytic activity of *Cissampelos sympodialis* Eichler leaf extract in a trachea preparation. While warifteine, a bisbenzylisoquilonine alkaloid, is allegedly the main bioactive substance in this species, its low solubility in polar solvents suggested the presence of another bioactive compound in the aqueous extracts. To investigate this alternative, PLS modeling was used to study the correlation of the metabolome obtained by NMR of diverse polar extracts with spasmolytic activity. The PLS model showed that the signals from flavonoids were positively correlated with the activity. Therefore, they concluded that identified flavonoids such as kaempferol and quercetin might be the important contributors to the activity.

Many studies have applied successfully the combination of metabolomics and bioassays with chemometrics in the identification of biologically active compounds in complex mixtures. Researchers can identify active metabolites from the crude extracts without isolation of single compound with this approach. However, from a practical point of view, the metabolomics approach itself is not enough to identify active molecules. Ultimately, the bioactivity of a compound must be confirmed with tests on the pure compound. In any case, and particularly in case of novel compounds, proper identification requires the determination of the full set of physicochemical properties and spectral data. Nonetheless, the advantages of the metabolomics approach are clear. To begin with, it allows the fast dereplication in case of already-known active compounds. Secondly, there is no need for large-scale isolation of the active compounds, nor bioassays and, if necessary, fractionation can be guided by the NMR or (LC- or GC-) MS spectra of the fractions. Particularly in the case of in-vivo bioassays (including clinical trials) only one first round of bioassays is needed, after which chemistry-guided fractionation is sufficient to isolate the active compound. In many cases this may have the additional advantage of reducing the number of animal experiments.

4 Application of Metabolomics-Based Bioactivity to Various Disease Models

4.1 Antimicrobial Activity

Most infectious diseases are easy to diagnose and humans have been tirelessly searching for medicines in nature to treat infections since ancient times. With the discovery of penicillin in 1928 by Fleming, microorganisms have been targeted for the discovery of other novel antibiotics. After a period of reduced interest, in recent years there is a revival in antibiotic bioprospecting, mainly because of the rapid spread of antibiotic-resistant pathogenic microorganisms and the failure to develop active derivatives of the existing ones. The conventional lead-finding of antimicrobial drugs is bioactivity-guided fractionation much like for other disease models. Aided by chemometrics, diverse *in-silico* methods have been recently introduced into this field, most of which are applied in the post lead-finding steps such as hit characterization and hit optimization (Xu et al. 2009). The application of metabolomics in the process of lead-finding involves:

1. metabolic profiling of crude extracts of selected organisms with antimicrobial activity, identification of active compounds
2. metabolomics-guided fractionation to isolate an active compound(s)
3. metabolic profiling of pathogenic microorganisms after treatment with potentially antimicrobial extracts or leads to identify their potential mode of action.

Tang et al. (2015) identified 10 antimicrobial compounds from burdock (*Arctium lappa* L.) leaves. The chemical composition of the leaf extracts was analyzed by UPLC-MS. For the activity, the inhibition of biofilm formation was measured. Biofilms are communities of microorganisms that are attached to a surface and are considered to be a survival mechanism of bacteria (Donlan and Costerton 2002). Anti-biofilm compounds can be used as antimicrobials. The obtained chemical and biological data were correlated by PLS-DA and resulted in the identification of 10 activity-correlated metabolites: chlorogenic acid, caffeic acid, *p*-coumaric acid, quercetin, ursolic acid, rutin, luteolin, crocin, benzoic acid, and tenacissoside. After evaluating the activity of pure samples of these compounds, chlorogenic acid and quercetin were determined to be the main anti-biofilm active compounds isolated from burdock leaf.

Dos Santos et al. (2018) used a similar approach to examine the antimicrobial activities of the volatile oils of several plant species. Eight volatile oils representing different levels of antimicrobial activity (from inactive to very active) were selected and profiled using GC-MS for their metabolomic analysis. The subsequent OPLS-DA revealed a high correlation of several metabolites 7,8-epoxy-1-octene, cis-α-bergamotene, methyl linolelaidate, alloaromadendrene, and veridiflorol with a significant antimicrobial activity. Interestingly, evidence of a specific chemical interaction between bornyl acetate and 4-terpineol was also observed, since they

were negatively correlated with each other in terms of activity, suggesting antagonism of those compounds.

Streptomyces species are the most studied Actinomycetes for the production of antibiotics. Wu et al. (2015a) applied NMR-based metabolomics to Actinomycetes and rapidly discovered novel antibiotics. The metabolomic comparison of wild-type and streptomycin-resistant actinomycetes strains revealed a stronger antimicrobial activity of the latter. Using PLS analysis, 7-prenylisatin was identified as the bioactive compound. The antimicrobial activity was confirmed by studies with the pure isolated compound (Fig. 2).

Because of bacterial resistance to current antibiotics, it is extremely important to find active compounds with a different mode of action to that of the current antibiotics. Thus, the mode of action should be identified in an early stage of lead-finding. A metabolomics approach in which the effect on the pathogenic microorganisms of known antibiotics and novel candidates is compared could provide an insight into the mode of action (Hoerr et al. 2016). Clearly, metabolomics is an important tool that should speed up the process of finding novel antibiotics.

While not directly connected to the correlation between chemical ingredients and antibiotic activity, the possibility of the induction of new antibiotics in a co-culture system (*Streptomyces* and *Aspergillus*) was probed using an analogous approach. Though well-known as a source of antibiotics, the Streptomyces metabolome is very sensitive to a number of external conditions. By co-culturing two different species, the interactions between the organisms may induce the production of novel compounds in either species. Moreover, catabolism of the compounds formed may add even further chemodiversity. Wu et al. (2015b) showed that the co-culture of *Streptomyces* with *Aspergillus* produced significant modifications in their individual biosynthetic processes yielding many new metabolites which could not be detected in the individual organisms.

4.2 Anticancer Activity

Natural products and their derivatives have a tremendous potential for the development of anticancer drugs. More than 50% of all anticancer drugs approved for therapeutic use in the past 70 years are either natural product or semi-synthetic analogs (Butler et al. 2014). Metabolomics has been used not only for drug development, but also in many fields related to cancer research including cancer prognosis, diagnosis, and treatment efficacy. The *in-silico* approach described above has been applied to the search of new lead compounds in many natural resources. An interesting example of its application in cancer research is reported by Graziani et al. (2018). Extracts of 14 legumes were screened against an array of human colorectal cancer cell lines in the search for bioactive natural products. Two plant extracts were selected for their strong antiproliferative activity using PCA and Hierarchical Cluster Analysis (HCA) (Fig. 3). The 2D NMR analysis of these plant extracts allowed the identification of two putative active compounds: a cycloartane glycoside and a protodioscin derivative. To confirm their bioactivity, these two compounds were

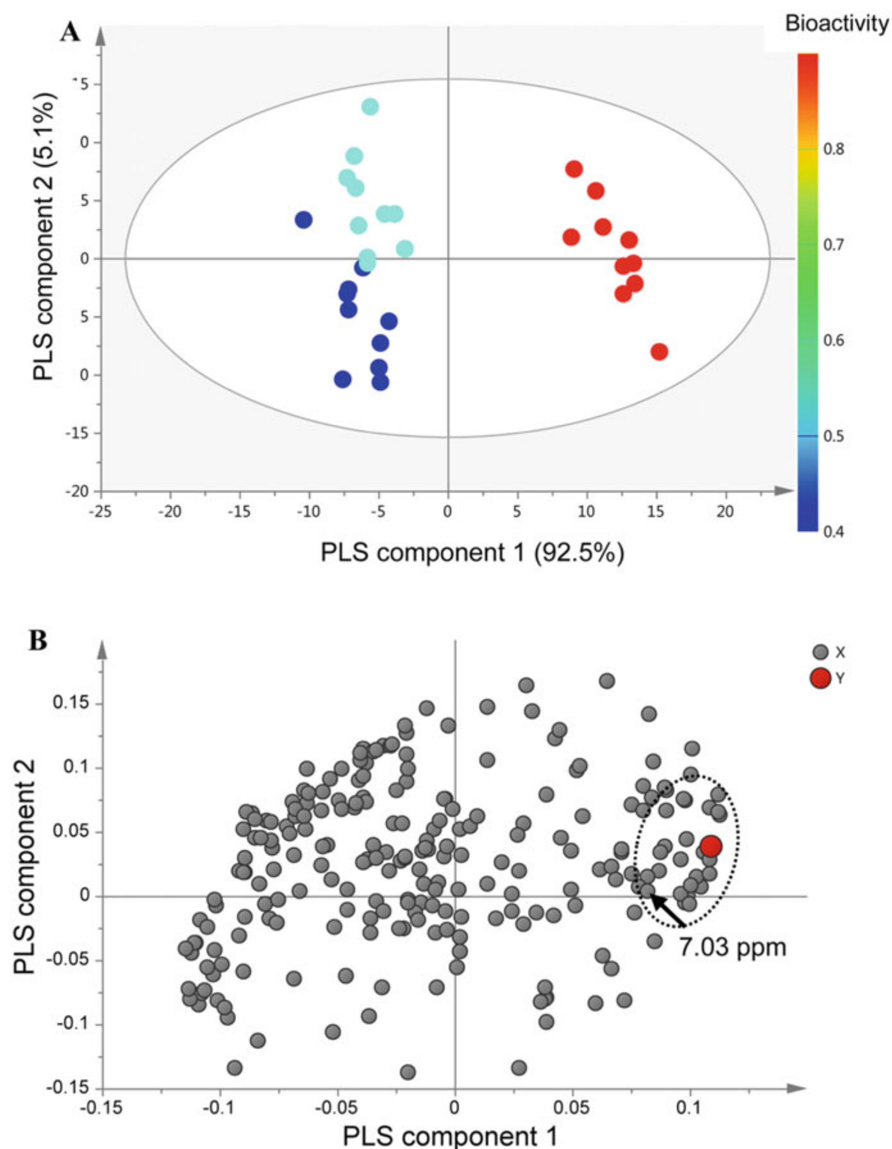


Fig. 2 Example of multivariate data analysis. PLS score plot (**a**) shows a good separation between wild-type *Streptomyces* sp. (blue) and its two mutant derivatives (MBT28-30: light blue and MBT28-91: red). The corresponding loading plot (**b**) presents the NMR signals (dashed circles) which contribute mostly to the separation of the samples and their bioactivity. The arrow refers to the characteristic proton signal at δ 7.03 (t, $J = 7.2$ Hz). X, primary variable of the chemical shift; Y, bioactivity. (Adopted from Wu et al. 2015a)

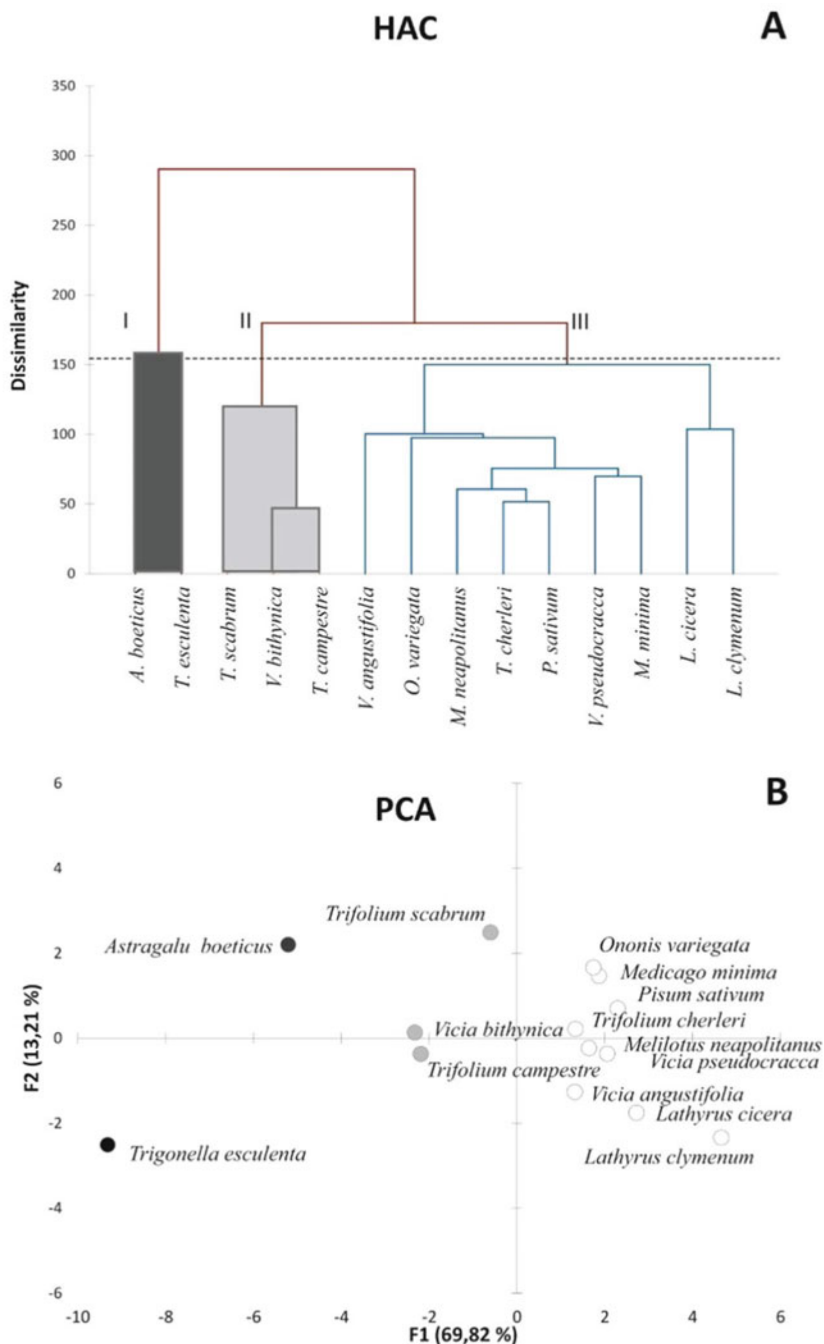


Fig. 3 Data of cell growth percentage were analyzed by HCA dendrogram (**a**) and PCA (**b**). Control and colon cancer cell lines which were treated with the plant extracts were distinguished. These analyses enabled classification of the species into 3 subsets; Groups I – the active species

isolated and tested positively for antiproliferative activity against colon cancer cells. These results are clear evidence of the efficiency of metabolomics as a tool to speed up the lead-finding process, since the active compounds are already identified in the crude extracts avoiding the time-consuming on-silica isolation procedure.

Using a similar approach, Bao et al. (2018) applied UPLC-MS-based metabolomics to identify anticancer compounds in Forsythiae fructus (from *Forsythia suspensa* Vahl). Crude extracts obtained with diverse solvents were profiled and the results were correlated with their anticancer activity against murine melanoma B16-F10 cell lines. OPLS-DA was applied to maximize the discrimination of extracts with different degrees of anticancer activity resulting in the selection of betulinic acid as a potential anticancer compound in Forsythiae fructus.

Gao et al. (2010) performed antiproliferative activity tests against human lung cell line SK-MES-1 on *Scutellaria baicalensis* Georgi root extracts. The data provided by the HPLC-UV and ^{1}H NMR analysis of the extracts were combined with PCA and PLS analysis for more complete metabolomics data, allowing the identification of baicalin, baicalein, and wogonin as the compounds responsible for the cell growth inhibition activity of the extracts.

Tawfike et al. (2019) adopted a different approach to study anticancer activity of the endophytic fungus *Aspergillus flocculus*, isolated from the stem of the medicinal plant *Markhamia lutea* (Benth.) K.Schum. (*Markhamia platycalyx* is a synonym). Its in-vitro culture was found to yield high levels of anticancer compounds active against the chronic myelogenous leukemia cell line K562. Several fractions of the fungal culture were analyzed by LC-MS. A combination of a molecular interaction network and OPLS-DA of the chemical data resulted in the identification of five active metabolites, namely, cis-4-hydroxymellein, 5-hydroxymellein, diorcinol, botryoisocoumarin A, and mullein. This study clearly shows the advantage of combining different forms of *in-silico* analysis prior to any purification attempts.

Another successful study using this approach was recently published by Ory et al. (2019) involving an anti-breast cancer compound (tested on MCF-7 cells) from a marine-derived *Penicillium chrysogenum* extract. Using different correlation and discriminant analysis including PLS-DA and PLS modeling, ergosterol was found to have anticancer activity with an antiproliferative activity on MCF-7 cells with an IC50 at 0.10 mM concentration. An interesting aspect of this report is the description of a workflow that allows the combination of both the chemical and biological data, a key step in this sort of application. Sometimes, a metabolomics approach alone is not sufficient to identify active metabolites in the crude extracts due not only to the complexity of the mixture but also to the low concentration levels of potentially interesting metabolites. In that case, conventional bio-guided fractionation can be useful as a supplementary tool. The work of Graziani et al. (2021) is a good example of this. They evaluated the anticancer activity of two plant species, *Ononis diffusa*

Fig. 3 (continued) (black), group II – active only at the highest tested doses (gray), and group III – no significant effect (white). Adapted from Graziani et al. (2018)

Ten. and *O. variegata* L. using colorectal cancer cell lines. When the NMR-based metabolomics data were correlated with the anticancer activity, most of the potentially interesting activity related signals were concentrated in the aliphatic region of the NMR spectra, a region which is difficult to annotate. The partial purification of the extracts using column chromatography and the application of different 2D NMR techniques led to the identification of a mixture of oxylipins as putative bioactive compounds. However, the report provides no information on activity confirmation of pure isolated compounds.

Gao et al. (2010) reported the application of metabolomics to study the mode of action of (−)-5-hydroxyequol, an isoflavone metabolite obtained by microbial biotransformation known to affect lung cancer cells through an unknown mechanism of action. Applying ^{1}H NMR-based metabolomics it was possible to observe rapid changes in the metabolism in human lung cancer cells, especially in the glycometabolism.

4.3 Anti-inflammatory Activity

Inflammation is involved in complex diseases such as autoimmune diseases, metabolic syndrome, neurodegenerative diseases, cancers, and cardiovascular diseases (Chen et al. 2017). The development of new inflammatory modulators to treat such diseases is therefore of great interest in the pharmaceutical industry. However, their development has been hampered by several issues regarding their potency, efficacy, and adverse effects. Since the introduction of acetylsalicylic acid in the nineteenth century, based on the use of salicylic acid-containing medicinal plants, nature has been regarded as an important resource to screen in the search for anti-inflammatory drugs resulting in multiple reports of natural products with alleged anti-inflammatory effects both in vitro and in vivo. Unsurprisingly, metabolomics has also been applied frequently in recent times to detect novel lead compounds from nature. A good example was recently reported for the methanolic extracts of *Cyrtanthus contractus* N.E.Br. (Amaryllidaceae) bulbs by Rafova et al. (2019). These bulb extracts showed a significant anti-inflammatory activity decreasing the level of E-selectin, a key player in the initiation of inflammation in a dose-dependent manner. Fractions from the extracts were profiled using LC-MS and evaluated for their biological activity. Subsequently, the correlation between biological activity and metabolite levels was calculated, resulting in the identification of narciclasine as a putative active compound. The bioactivity of pure narciclasine confirmed the findings.

Alcântara et al. (2021) reported a similar approach to study the leaves of *Ocotea odorifera* (Vell.) Rohwer. This plant has been used traditionally for the treatment of rheumatism. Chemical profiles of leaf extracts and their fractions were obtained using LC-HRMS and their in vivo anti-inflammatory activity was tested. PLS-DA of the chemical data suggested that the activity is correlated with S-(+)-reticuline, a known alkaloid from this plant. Yet another unknown compound, which was not identified in this study was also found to be active. The researchers proved their aim,

which was to show the competitiveness of metabolomics as a rapid tool for dereplication.

Domingos et al. (2019) demonstrated the successful identification of a new anti-inflammatory compound using this approach. The extracts from *Poincianella pluviosa* var. *peltophoroides* (Benth.) L.P.Queiroz (unresolved name) showed significant in vivo anti-inflammatory activity. The UHPLC-HRMS profiling of the extracts allowed the isolation of four compounds using a metabolomics-guided chromatographic process. Of these, two were previously unknown compounds and were identified as 4‴-methoxycaesalpinioflavone and 7-methoxycaesalpinioflavone while the other two were the known compounds, rhuschalcone VI and caesalpinioflavone. The activity of all four was confirmed using pure compounds.

The combination of metabolomics and bio-guided fractionation has thus proved to be very efficient for the identification of active metabolites. A further example of its successful application is the LC-MS-based metabolomics of *Actinidia arguta* (Siebold & Zucc.) Planch. ex Miq. leaves which combined with bio-guided fractionation resulted in the identification of the active compounds caffeoylthreonic acid and danshensu (Kim et al. 2019). The identification of known compounds is thus possible by using spectral data obtained from NMR and/or MS spectroscopy. Thanks to the development of analytical techniques and technological improvements in existing instrumentation, it is now much easier to analyze mixtures and identify the biologically active components within them.

5 Conclusions and Perspective

We have selected representative examples among the hundreds of published reports that describe the role of metabolomics in natural products lead-finding. Nature has been and will most likely continue to be a relevant, if not the most relevant, source for novel drugs as reflected in, among others, the review by Newman and Cragg (2007). Biodiversity is essential for chemodiversity. This was naturally appreciated by our ancestors, who understood its power and were able to take advantage of it. Medicinal plants such as *Atropa belladonna* L. and *Papaver somniferum* L. are just two among the many examples of plants that have provided humanity with a number of important therapeutic solutions that are part of the core of western medicine.

Analyzing natural products research for novel drugs, it is clear that there have been two ways to go about it. One consists in the at-random screening of plants and other organisms for a certain activity. The other is to use the accrued traditional knowledge on medicinal plants as a starting point. Whichever the approach chosen, researchers will have to deal with complex mixtures of compounds among which a few active ones will need to be identified. Active compounds might be present in very low concentrations, might be very labile, or present in an inactive form that may require some kind of transformation to acquire activity. Furthermore, experience has shown that in many cases, the alleged activity of a traditional medicine is due to a particular combination of compounds that act through synergistic and/or

antagonistic interactions. Bioprospecting throughout the past decades used the classical approach of at-random screening for a certain activity followed by bioassay-guided fractionation. The present industrial approach is at-random high-throughput screening using fast molecular level bioassays. However, this approach is an elaborate, time- and energy-consuming process with results that do not often enough match with the invested resources. This experimental design is based on the single target-single compound paradigm and having been undoubtedly successful in some cases was used for many years and continues to be used. An example of this approach was the NCI program for antitumor drugs which yielded some very active molecules and leads. However, the present day perspective of big pharmaceutical companies regarding novel drugs development is one of caution, considering the level of risk of the investment required, leaving the lead-finding process in hands of small start-up companies, universities, or public research institutions. The biodiversity treaty (Nagoya protocol on Access and Benefit-sharing, https://www.cbd.int/abs/about/) did not really encourage the pharmaceutical industry to invest in bioprospecting, whereas on national level investment in such research is limited. The studies on local traditional medicines have often higher priority. It is clear that to increase the chances of success, the whole process has to be speeded up, and particularly in connection with traditional medicine sources, more in-vivo assays should be performed to validate their use, using a systems biology-based approach. Whichever the chosen approach, metabolomics can be used to shorten the time to identify the active compounds. Metabolomics of different extracts from a plant in combination with the bioassay data can in both designs be used to identify the signals from the metabolomics data that correlate with activity. Based on the information obtained from the set of analytical data (e.g., NMR signals, Mass spectra, retention time), a preliminary identification is possible by comparing the data with already available data from a metabolomics database. This information allows an informed decision on whether it is worthwhile to isolate the presumably active compound(s) for further studies. If the compound is deemed of interest, the isolation procedure must be scaled up to obtain sufficient amounts of the pure compound(s) to submit it to full spectral data analysis and the relevant bioassay(s). For this metabolomics-guided fractionation should be preferred over the bioassay-guided fractionation process since it is faster and may reduce the use of animal experiments.

In conclusion, the “on-silica” approach remains an excellent tool to fractionate extracts on a small scale for metabolomics analysis and on a larger scale for the isolation of sufficient amounts of pure active compounds for full spectral analysis for identification or for structure elucidation of novel compounds.

The “in-silico” approach, based on the systems biology paradigm of holistic observations, accelerates the dereplication of known active compounds, among others, by reducing the required number of bioassays and animal experiments. Moreover the “in-silico” approach will allow deeper insight into the mode of action by measuring the metabolic changes observed in living cells and organisms when treated with drugs. A further not inconsiderate advantage of the application of “in-silico” tools is the reduction of the time needed to identify an active compound, but

most importantly in reduction of the amount of toxic, contaminating solvents and consumables required by early stage dereplication of known active compounds. In other words, the “in-silico” approach is not only faster but also a greener procedure for finding novel leads.

References

Akhtar MHT, Mushtaq MY, Verpoorte R et al (2016) Zebrafish as a model for systems medicine R&D: rethinking the metabolic effects of carrier solvents and culture buffers determined by ^{1}H NMR metabolomics. OMICS 20:42–52. https://doi.org/10.1089/omi.2015.0119

Alcântara BGV, Oliveira FP, Katchborian-Neto A et al (2021) Confirmation of ethnopharmacological anti-inflammatory properties of *Ocotea odorifera* and determination of its main active compounds. J Ethnopharmacol 264:113378. https://doi.org/10.1016/j.jep.2020.113378

Ali K, Iqbal M, Yuliana ND et al (2013) Identification of bioactive metabolites against adenosine A1 receptor using NMR-based metabolomics. Metabolomics 9:778–785. https://doi.org/10.1007/s11306-013-0498-9

Allard PM, Peŕesse T, Bisson J et al (2016) Integration of molecular networking and *in-silico* MS/MS fragmentation for natural products dereplication. Anal Chem 88:3317–3323. https://doi.org/10.1021/acs.analchem.5b04804

Baerheim Svendsen A, Karlsen J (1967) Gaschromatographie von Monoterpenkohlenwasserstoffen aus Aetherischen Oelen an gepackten Trennsaulen mit niedrigem Gehalt an fluessiger stationaere Phase. Planta Med 15:1–5. https://doi.org/10.1055/s-0028-1099949

Banerjee P, Erehman J, Gohlke BO et al (2015) Super natural II – a database of natural products. Nucleic Acids Res 43:D935–D939. https://doi.org/10.1093/nar/gku886

Bao JL, Ding RB, Jia XJ et al (2018) Fast identification of anticancer constituents in Forsythiae Fructus based on metabolomics approaches. J Pharm Biomed Anal 154:312–320. https://doi.org/10.1016/j.jpba.2018.03.020

Boezio B, Audouze K, Ducrot P et al (2017) Network-based approaches in pharmacology. Mol Informatics 36:1700048. https://doi.org/10.1002/minf.201700048

Butler MS, Robertson AAB, Cooper MA (2014) Natural product and natural product derived drugs in clinical trials. Nat Prod Rep 31:1612–1661. https://doi.org/10.1039/c4np00064a

Cardoso-Taketa AT, Pereda-Miranda R, Choi YH et al (2008) Metabolic profiling of the Mexican anxiolytic and sedative plant *Galphimia glauca* using nuclear magnetic resonance spectroscopy and multivariate data analysis. Planta Med 74:1295–1301. https://doi.org/10.1055/s-2008-1074583

Chen L, Deng H, Cui H et al (2017) Inflammatory responses and inflammation-associated diseases in organs. Oncotarget 9:7204–7218. https://doi.org/10.18632/oncotarget.23208

Cloarec O, Dumas ME, Craig A et al (2005) Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. Anal Chem 77:1282–1289. https://doi.org/10.1021/ac048630x

Cragg GM, Newman DJ (2005) Plants as a source of anti-cancer agents. J Ethnopharmacol 100:72–79. https://doi.org/10.1016/j.jep.2005.05.011

Cuperlovic-Culf M, Culf AS (2016) Applied metabolomics in drug discovery. Expert Opin Drug Discovery 11(8):759–770. https://doi.org/10.1080/17460441.2016.1195365

De Donno A, Grassi T, Idolo A et al (2012) First-time comparison of the in vitro antimalarial activity of *Artemisia annua* herbal tea and artemisinin. Trans R Soc Trop Med Hyg 106:696–700. https://doi.org/10.1016/j.trstmh.2012.07.008

de Melo ICAR, de Souza ILL, Vasconcelos LHC et al (2020) Metabolomic fingerprinting of *Cissampelos sympodialis* Eichler leaf extract and correlation with its spasmolytic activity. J Ethnopharmacol 253:112678. https://doi.org/10.1016/j.jep.2020.112678

Domingos OD, Alcântara BGV, Santos MFC et al (2019) Anti-inflammatory derivatives with dual mechanism of action from the metabolomic screening of *Poincianella pluviosa*. Molecules 24:4375. https://doi.org/10.3390/molecules24234375

Donlan RM, Costerton JW (2002) Biofilms: survival mechanisms of clinically relevant microorganisms. Clin Microb Rev 15:167–193. https://doi.org/10.1128/CMR.15.2.167-193.2002

dos Santos FA, Sousa IP, Furtado NAJC et al (2018) Combined OPLS-DA and decision tree as a strategy to identify antimicrobial biomarkers of volatile oils analyzed by gas chromatography–mass spectrometry. Rev Bras Farm 28:647–653. https://doi.org/10.1016/j.bjp.2018.08.006

Emwas AH, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, Raftery D, Alahmari F, Jaremko L, Jaremko M, Wishart DS (2019) NMR spectroscopy for metabolomics research. Metabolites 9(7):123. https://doi.org/10.3390/metabo9070123

Gao JY, Zhao HY, Hylands PJ et al (2010) Secondary metabolite mapping identifies *Scutellaria* inhibitors of human lung cancer cells. J Pharm Biomed Anal 53:723–728. https://doi.org/10.1016/j.jpba.2010.04.019

Graziani V, Scognamiglio M, Belli V et al (2018) Metabolomic approach for a rapid identification of natural products with cytotoxic activity against human colorectal cancer cells. Sci Rep 8:5309. https://doi.org/10.1038/s41598-018-23704-9

Graziani V, Potenza N, D'Abrosca B et al (2021) NMR profiling of *Ononis diffusa* identifies cytotoxic compounds against cetuximab-resistant colon cancer cell lines. Molecules 26:3266. https://doi.org/10.3390/molecules26113266

Gromski PS, Muhamadali H, Di E et al (2015) A tutorial review: metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. Anal Chim Acta 879:10–23. https://doi.org/10.1016/j.aca.2015.02.012

Hoerr V, Duggan GE, Zbytnuik L et al (2016) Characterization and prediction of the mechanism of action of antibiotics through NMR metabolomics. BMC Microbiol 16:82. https://doi.org/10.1186/s12866-016-0696-5

Hopkins AL (2007) Network pharmacology: network biology illuminates our understanding of drug action. Nat Biotechnol 25:1110–1111. https://doi.org/10.1038/nbt1007-1110

Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4:682–690. https://doi.org/10.1038/nchembio.118

Ingkaninan K, de Best CM, van der Heijden R et al (2000) HPLC with on-line coupled UV, mass spectrometric and biochemical detection for identification of acetylcholinesterase inhibitors from natural products. J Chromatogr A 872:61–73. https://doi.org/10.1016/s0021-9673(99)01292-3

Jeffryes JG, Colastani RL, Elbadawi-Sidhu M et al (2015) MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. J Chem Inform 7:44. https://doi.org/10.1186/s13321-015-0087-1

Jiang H, Hu C, Chen M (2021) The advantages of connectivity map applied in traditional Chinese medicine. Front Pharmacol 12:474267. https://doi.org/10.3389/fphar.2021.474267

Kim GD, Lee JY, Auh JH (2019) Metabolomic screening of anti-inflammatory compounds from the leaves of *Actinidia arguta* (Siebold & Zucc.) Planch. ex Miq. (Hardy Kiwi). Foods 8:47. https://doi.org/10.3390/foods8020047

Klingelhofer I, Ngoc LP, van der Burg B et al (2021) A bioimaging system combining human cultured reporter cells and planar chromatography to identify novel bioactive molecules. Anal Chim Acta 1183:338956. https://doi.org/10.1016/j.aca.2021.338956

Leonti M, Verpoorte R (2017) Traditional mediterranean and European herbal medicines. J Ethnopharmacol 199:161–167. https://doi.org/10.1016/j.jep.2017.01.052

Li S, Zhang B (2013) Traditional Chinese medicine network pharmacology: theory, methodology and application. Chin J Nat Med 11:0110–0120. https://doi.org/10.1016/S1875-5364(13)60037-0

Lia ZY, Zhi HJ, Zhang FS et al (2013) Metabolomic profiling of the antitussive and expectorant plant *Tussilago farfara* L. by nuclear magnetic resonance spectroscopy and multivariate data analysis. J Pharm Biomed Anal 75:158–164. https://doi.org/10.1016/j.jpba.2012.11.023

Malone MH, Robichaud RC (1962) A Hippocratic screen for pure or crude drug materials. Lloydia 25:320–332

Maree J, Kamatou G, Gibbons S et al (2014) The application of GC–MS combined with chemometrics for the identification of antimicrobial compounds from selected commercial essential oils. Chemom Intel Lab Syst 130:172–181. https://doi.org/10.1016/j.chemolab.2013.11.004

Morlock GE (2021) High-performance thin-layer chromatography combined with effect directed assays and high-resolution mass spectrometry as an emerging hyphenated technology: a tutorial review. Anal Chim Acta 1180:338644. https://doi.org/10.1016/j.aca.2021.338644

Morlock GE, Drotleff L, Brinkmann S (2021) Miniaturized all-in-one nanoGIT(+active) system for on-surface metabolization, separation and effect imaging. Anal Chim Acta 1154:33830. https://doi.org/10.1016/j.aca.2021.338307

Mushtaq MY, Verpoorte R, Kim HK (2013) Zebrafish as a model for systems biology. Biotechnol Gen Engin Rev 29:187–205. https://doi.org/10.1080/02648725.2013.801238

Newman DJ, Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. J Nat Prod 70:461–477. https://doi.org/10.1021/np068054v

Newman DJ, Cragg GM (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod 83:770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

Ntungwe NE, Dominguez-Martin EM, Roberto A et al (2020) *Artemia* species: an important tool to screen general toxicity samples. Curr Pharm Design 26:2892–2908. https://doi.org/10.2174/1381612826666200406083035

Ory L, Nazih EH, Daoud S et al (2019) Targeting bioactive compounds in natural extracts – development of a comprehensive workflow combining chemical and biological data. Anal Chim Acta 1070:29–42. https://doi.org/10.1016/j.aca.2019.04.038

Parng C, Seng WL, Semino C et al (2002) Zebrafish: a preclinical model for drug screening. Assay Drug Dev Technol 1:41–48. https://doi.org/10.1089/154065802761001293

Paul A, de Boves Harrington P (2021) Chemometric applications in metabolomic studies using chromatography-mass spectrometry. Trends Anal Chem 135:116165. https://doi.org/10.1016/j.trac.2020.116165

Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. Science 269:347–350. https://doi.org/10.1126/science.269.5222.347

Rafova L, Ncube B, Van Staden J et al (2019) Identification of narciclasine as an in vitro anti-inflammatory component of *Cyrtanthus contractus* by correlation-based metabolomics. J Nat Prod 82:1372–1376. https://doi.org/10.1021/acs.jnatprod.8b00973

Rhee IK, van Rijn RM, Verpoorte R (2003) Qualitative determination of false-positive effects in the acetylcholinesterase assays using thin layer chromatography. Phytochem Anal 14:127–131. https://doi.org/10.1002/pca.675

Salzer L, Witting M (2021) Quo Vadis *Caenorhabditis elegans* metabolomics-a review of current methods and applications to explore metabolism in the nematode. Metabolites 11:284. https://doi.org/10.3390/metabo11050284

Samuelsson G, Bohlin L (eds) (2009) Drugs of natural origin: a treatise of pharmacognosy. Swedish Academy of Pharmaceutical Sciences, Stockholm

Schreiner T, Morlock GE (2021) Non-target bioanalytical eight-dimensional hyphenation including bioassay, heart-cut trapping, online desalting, orthogonal separations and mass spectrometry. J Chromatogr 1647:462154. https://doi.org/10.1016/j.chroma.2021.462154

Szymańska E, Saccenti E, Smilde AK et al (2012) Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics 8:3–16. https://doi.org/10.1007/s11306-011-0330-3

Tang Y, Lou Z, Yang L et al (2015) Screening of antimicrobial compounds against *Salmonellaty phimurium* from burdock (*Arctium lappa*) leaf based on metabolomics. Eur Food Res Technol 240:1203–1209. https://doi.org/10.1007/s00217-015-2423-0

Tawfike TAF, Romli M, Clements C et al (2019) Isolation of anticancer and anti-trypanosome secondary metabolites from the endophytic fungus *Aspergillus flocculus* via bioactivity guided isolation and MS based metabolomics. J Chromatogr B 1106–1107:71–83. https://doi.org/10.1016/j.jchromb.2018.12.032

Verpoorte R, Kim HK, Choi YH (2018) Synergy: easier to say than to prove. Synergy 7:34–35. https://doi.org/10.1016/j.synres.2018.10.004

Wang M, Lamers RJAN, Korthout HAAJ et al (2005) Metabolomics in the context of systems biology: bridging traditional Chinese medicine and molecular pharmacology. Phytother Res 19: 173–182. https://doi.org/10.1002/ptr.1624

Westerhuis JA, Hoefsloot HCJ, Smit S et al (2008) Assessment of PLSDA cross validation. Metabolomics 4:81–89. https://doi.org/10.1007/s11306-007-0099-6

Wishart DS (2008) Applications of metabolomics in drug discovery and development. Drugs R D 9:307–322. https://doi.org/10.2165/00126839-200809050-00002

Wu C, Du C, Gubbens J et al (2015a) Metabolomics-driven discovery of a prenylated isatin antibiotic produced by *Streptomyces* species MBT28. J Nat Prod 78:2355–2363. https://doi.org/10.1021/acs.jnatprod.5b00276

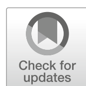
Wu C, Zacchetti B, Ram AFJ et al (2015b) Expanding the chemical space for natural products by *Aspergillus niger* and *Streptomyces coelicolor* co-cultivation and biotransformation. Nat Sci Rep 4:10868. https://doi.org/10.1038/srep/10868

Xu EY, Schaefer WH, Xu Q (2009) Metabolomics in pharmaceutical research and development: metabolites, mechanisms and pathways. Curr Opin Drug Disc 12:40–52. https://doi.org/10.1016/j.copbio.2015.04.004

Yang JY, Sanchez LM, Rath CM et al (2013) Molecular networking as a dereplication strategy. J Nat Prod 76:1686–1699. https://doi.org/10.1021/np400413s

Ye H, Wei J, Tang KL et al (2016) Drug repositioning Ttrough network pharmacology. Curr Top Med Chem 16:3646–3656. https://doi.org/10.2174/1568026616666160530181328

Yuliana ND, Khatib A, Choi YH et al (2011) Comprehensive extraction integrated with NMR metabolomics: a new way of bioactivity screening methods for plants, adenosine A1 receptor binding compounds in *Orthosiphon stamineus* Benth. Anal Chem 83:6902–6906. https://doi.org/10.1021/ac201458n



Quantitative NMR Methods in Metabolomics

G. A. Nagana Gowda and Daniel Raftery

Contents

1 Introduction 144
2 Quantitation Approaches Using NMR 146
2.1 Internal Reference Standards for Absolute Quantitation 146
2.2 Alternative Reference Standards for Absolute Quantitation 149
2.3 Quantitation of Metabolites Using Intact Samples 149
2.3.1 Intact Serum and Plasma Analysis 150
2.3.2 Intact Urine Analysis 150
2.3.3 Intact Tissue Analysis 151
2.4 Metabolite Quantitation Using Processed Samples 152
2.4.1 Analysis of Aqueous Metabolites 152
2.4.2 Analysis of Coenzymes and Antioxidants 154
2.4.3 Analysis of Lipids 156
2.5 Quantitation Methods Using Stable Isotope Labeling 157
2.5.1 Isotope Labeling Focused on Metabolic Fluxes and Pathways 157
2.5.2 Isotope Labeling Focused on Metabolite Analysis 157
3 Conclusion 158
References 159

G. A. Nagana Gowda (✉)
Northwest Metabolomics Research Center, University of Washington, Seattle, WA, USA

Mitochondria and Metabolism Center, Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA
e-mail: ngowda@uw.edu

D. Raftery (✉)
Northwest Metabolomics Research Center, University of Washington, Seattle, WA, USA

Mitochondria and Metabolism Center, Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA

Fred Hutchinson Cancer Research Center, Seattle, WA, USA
e-mail: draftery@uw.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_612

Abstract

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the two major analytical platforms in the field of metabolomics, the other being mass spectrometry (MS). NMR is less sensitive than MS and hence it detects a relatively small number of metabolites. However, NMR exhibits numerous unique characteristics including its high reproducibility and non-destructive nature, its ability to identify unknown metabolites definitively, and its capabilities to obtain absolute concentrations of all detected metabolites, sometimes even without an internal standard. These characteristics outweigh the relatively low sensitivity and resolution of NMR in metabolomics applications. Since biological mixtures are highly complex, increased demand for new methods to improve detection, better identify unknown metabolites, and provide more accurate quantitation continues unabated. Technological and methodological advances to date have helped to improve the resolution and sensitivity and detection of a larger number of metabolite signals. Efforts focused on measuring unknown metabolite signals have resulted in the identification and quantitation of an expanded pool of metabolites including labile metabolites such as cellular redox coenzymes, energy coenzymes, and antioxidants. This chapter describes quantitative NMR methods in metabolomics with an emphasis on recent methodological developments, while highlighting the benefits and challenges of NMR-based metabolomics.

Keywords

Fast NMR methods · Isotope tagging · Metabolomics · Nuclear magnetic resonance (NMR) · Quantitation

1 Introduction

The field of metabolomics represents the parallel analysis of large numbers of metabolites in biological systems. Metabolites provide information on the instantaneous biological state of an organism or system along with the functions of upstream cellular molecular species such as genes, transcripts, and proteins in health and pathological conditions. Using a variety of advanced methodologies, comprehensive analysis of metabolite data enables understanding biological phenotypes, deciphering mechanisms, and identifying disease biomarkers or drug targets (Raftery 2014; Nagana Gowda and Raftery 2019). Metabolomics applications span a wide range of disciplines including human health and diseases, pharmacology, drug development, toxicology, environment, plants, food, and nutrition. However, a majority of the studies to date are focused on improving the mechanistic understanding, prevention, early diagnosis, and management of human diseases (Kodama et al. 2020; Goldman et al. 2019; Johnson et al. 2016; Wishart 2016).

Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are the two most widely used methods in the metabolomics field. MS typically provides two to three orders of magnitude higher sensitivity than NMR and thereby enables analysis of several hundreds to thousands of metabolites from a single measurement. Generally, in MS analysis, metabolites from biological mixtures are subjected to separation using methods such as liquid chromatography, gas chromatography, or electrophoresis prior to detection. Separation using liquid chromatography, however, is the most popular and nearly 80% of the metabolomics methods use liquid chromatography resolved MS method (Edison et al. 2021). Absolute quantitation of metabolites in MS involves using internal or external standards, ideally, for each metabolite. However, finding isotopically labeled internal standards for each metabolite is challenging and hence, one standard that represents a class of metabolites is often used (Djukovic et al. 2020). This approach, however, can result in a loss of accuracy. In contrast, and as will be described below, NMR provides several approaches for accurate quantitation.

Although NMR spectroscopy is less sensitive than MS, it exhibits numerous unique and favorable characteristics that are beneficial to the field of metabolomics (Edison et al. 2021; Wishart 2019; Nagana Gowda and Raftery 2014a, 2015, 2017a, 2019). Notably: (1) NMR is highly reproducible and has excellent linearity (Mo and Raftery 2008); (2) NMR provides absolute quantitation of all metabolites in the spectrum using a single internal standard or even without the need for an internal standard; (3) it provides the gold standard approach in establishing the identity of unknown metabolites; (4) it enables the analysis of intact biofluid and tissue samples with little to no need for sample preprocessing; (5) it is non-destructive, which means the sample remains intact after the analysis and can be reused for analysis using NMR or using other methods such as MS; (6) it enables tracing of metabolic pathways and measuring metabolic fluxes utilizing stable isotope-labeled precursors; (7) using NMR, the same metabolites can be detected through one or more types of atomic nuclei such as ^{1}H, ^{13}C, ^{31}P, or ^{15}N, which provides flexibility to measure metabolite levels; (8) NMR's ability to detect essentially all molecular species with a given nucleus makes it extremely useful for following methods development; and (9) NMR offers new avenues to measure unstable metabolites that are fundamental to cellular functions. Such characteristics far outweigh the poor sensitivity and resolution of NMR and have been exploited extensively in the metabolomics field.

Human blood serum/plasma, urine, and tissue continue to be the most widely used biological specimens in the metabolomics field. However, other biological specimens including saliva (Lohavanichbutr et al. 2018), cerebrospinal fluid (Albrecht et al. 2020), gut aspirate (Bala et al. 2006), bile (Nagana Gowda 2011), amniotic fluid (Orczyk-Pawilowicz et al. 2016), synovial fluid (Anderson et al. 2020), fecal samples (Zierer et al. 2018), exhaled breath condensate (Maniscalco et al. 2020), tear (Yazdani et al. 2019), and sperm-seminal fluid (Engel et al. 2019) have also been analyzed. In addition, specimens from animal models, cell lines, yeast (Airoldi et al. 2015), bacteria (Lussu et al. 2017), tumor cells (Lane et al. 2017), tumor spheroids (Kalfe et al. 2015), exosomes (Zebrowska et al. 2019), and isolated mitochondria (Xu et al. 2018) have been used.

The key steps involved in almost all metabolomics investigations include metabolite detection, unknown peak identification, and quantification. Relative or absolute concentrations of metabolites thus obtained are then subjected to statistical and/or metabolic pathway analysis focused on a wide variety of applications in the areas of basic and medical sciences. Typically, metabolite data are analyzed using univariate and multivariate statistical analysis focused on the discovery and validation of putative metabolite biomarkers. Alternatively, metabolite levels or isotope labeled metabolites are used for identifying the perturbed metabolic pathways, which provide mechanistic understanding of cellular functions including information on drug targets for therapeutic development.

2 Quantitation Approaches Using NMR

The quantitative ability of NMR makes it an important platform complementary to MS in metabolomics. NMR can be used for quantitative analysis of metabolites in intact samples, extracted samples, live organisms, cells, or subcellular organelles such as mitochondria. In NMR, generally, metabolite peaks are identified prior to their relative or absolute quantitation. The identities of metabolites are established using databases of standard compounds, the comprehensive analysis of 1D and 2D NMR spectra, and/or spiking with authentic compounds.

Quantitation generally involves either (a) relative quantitation, in which metabolite levels are measured relative to one another; or (b) absolute quantitation, in which molar concentrations of metabolites are determined using an internal or external standard. Currently, relative quantitation is the most widely used approach owing to its ease of use combined with challenges associated with absolute quantitation, especially for some sample types, such as cells, tissue, fecal samples, etc. However, absolute quantitation promises a number of benefits. Importantly, it provides a basic platform of metabolite levels for a specific type of biological specimen. This is important for assessment of data quality, such as to compare samples across different geographical regions, different batches or analysis times, or perhaps most importantly to compare to known values, such as clinical ranges for blood or urine metabolites. Considering the increased interest for absolute quantitation, there have been numerous efforts in recent years focused on establishing reference standards for absolute quantitation using NMR as described in the following section.

2.1 Internal Reference Standards for Absolute Quantitation

Many compounds (>25) have been evaluated as potential internal standards for applications in numerous areas including organic chemistry, natural product chemistry, agriculture, drug discovery, and pharmaceuticals (Maniara et al. 1998; Holzgrabe 2010; Pauli et al. 2012; Rundlöf et al. 2010; Salem and Mossa 2012). These compounds exhibit favorable physical characteristics, such as unique chemical shift, purity, stability, solubility, and suitability for accurate gravimetry.

However, most of these are not suitable for metabolomics due to aqueous solubility concerns or chemical shift overlap. Chemical shift reference compounds such as TSP (trimethylsilylpropionic acid) and DSS (trimethylsilylpropanesulfonic acid) have been used as internal standards for absolute quantitation of metabolites. It was realized some years ago, however, that these compounds are unsuitable for quantitation owing to their peak suppression arising from the interaction with proteins. One alternative, formic acid, was evaluated as an alternative to TSP for quantitation of metabolites in intact serum many years ago (Kriat et al. 1992). However, formic acid is unsuitable as a reference since it is an endogenous metabolite; the endogenous concentration in serum varies significantly from person to person (~40 to 350 μM) and hence, it interferes with externally added formic acid (Kubáň and Boček 2013; d'Alessandro et al. 1994; Kapur et al. 2007). In another study, DSA (4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate), which is a derivative of DSS, was evaluated as a potential internal standard using intact rat plasma (Alum et al. 2008). However, it is also unsuitable as a reliable internal standard since several factors including the increased line broadening by a factor of >2 at pH 7.4 relative to pH 3.0 indicate that DSA interacts with sample matrix. One remedy for analysis of metabolites in samples such as blood serum/plasma that contain copious macromolecules is to remove the macromolecules effectively by ultrafiltration; in such a case, TSP or DSS can still be used as standards for absolute quantitation (Psychogios et al. 2011; Barding et al. 2012; Simón-Manso et al. 2013). The challenge with ultrafiltration, however, is that it attenuates many metabolite peaks (Nagana Gowda and Raftery 2014b), requires larger sample volumes, and is cumbersome for large-scale studies. In addition, ultrafiltration cannot be used for analysis of samples such as tissue and whole blood. Ultrafiltration is also incompatible with MS analysis, the other major analytical platform used in metabolomics, since MS analysis invariably employs protein precipitation to remove macromolecules, prior to analysis (Nagana Gowda et al. 2018).

Protein precipitation that removes macromolecules from samples provides an alternative approach to quantitate metabolites and is well suited for large-scale studies. However, even in such samples, peaks from the traditional internal standards, TSP and DSS, are attenuated by up to 35% and hence they are unsuitable as internal standards. More recently, two compounds, maleic acid and fumaric acid, were evaluated for their utility as potential internal standards for quantitation of metabolites since both provide a single peak in NMR spectrum and their peaks do not overlap with peaks from bio-specimen spectra (Nagana Gowda et al. 2021) (Fig. 1). It was shown that fumaric acid is a robust standard for protein precipitated serum, plasma, and whole blood; and maleic acid is suitable for plasma and serum, but it overlaps with coenzyme peaks in whole blood samples. These findings provide new opportunities for improved and accurate quantitation of metabolites in human plasma, serum, and whole blood using NMR spectroscopy. The potential utility of maleic acid and fumaric acid as internal standards may be extended to other biological specimens, as long as they do not overlap with bio-specimen peaks.

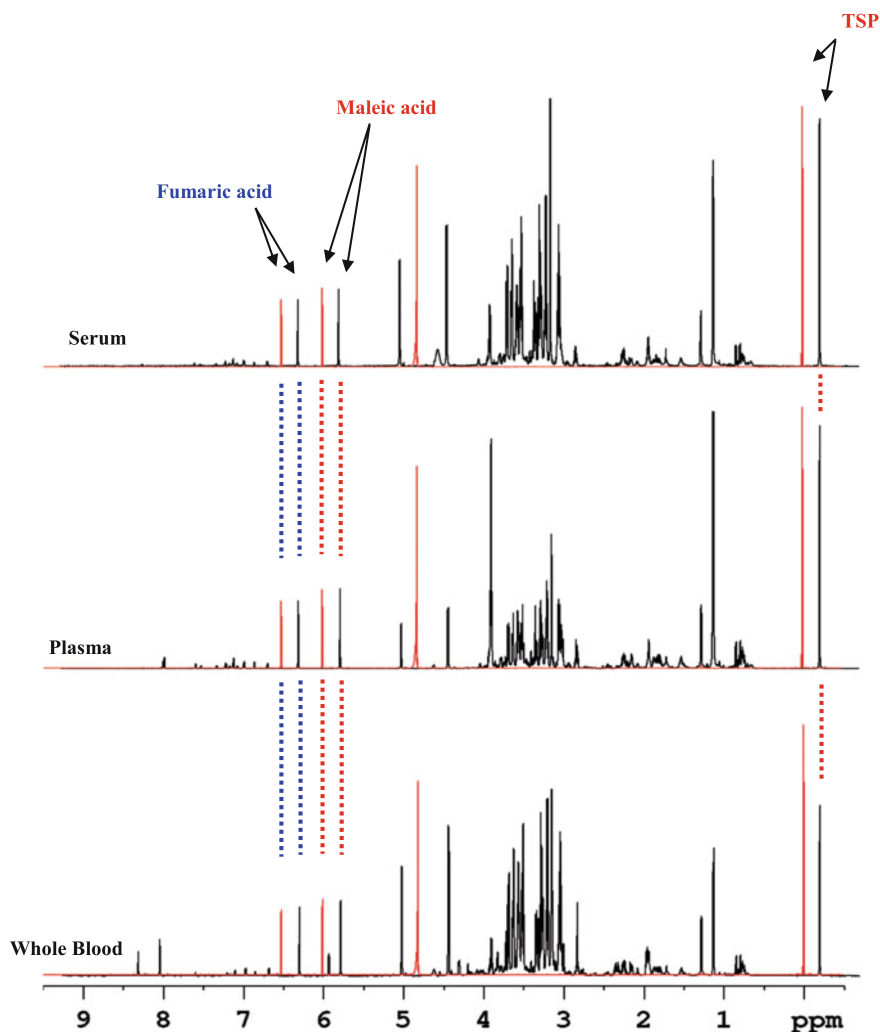


Fig. 1 Typical 800 MHz ^{1}H NMR spectra of a protein precipitated sample of human (**a**) whole blood, (**b**) plasma, and (**c**) serum, solubilized in D_2O buffer containing a mixture of three internal standards (TSP, 238 μM; maleic acid, 350 μM; fumaric acid, 293 μM). Each spectrum is overlaid with a spectrum from the blank buffer consisting of the same three standards (spectrum shown in red) to enable the visualization of peak heights for the three internal standards; the spectra of the bio-specimens are slightly right shifted relative to the blank spectrum for clarity. Peaks from the blank are marked with asterisks. Heights for the fumaric acid peaks from the bio-specimen and blank are approximately matched; however, a significant attenuation of the TSP peak in all three bio-specimens spectra is noticeable. TSP: Trimethylsilylpropionic acid-d_4. (reproduced with permission from Nagana Gowda et al. 2021)

2.2 Alternative Reference Standards for Absolute Quantitation

An altogether different approach is to determine metabolite concentrations without the need for an internal standard. One such method is ERETIC (electronic reference to access in vivo concentrations) (Akoka et al. 1999). In the ERETIC method, a synthetic signal is generated in NMR spectra with the desired peak intensity, line width, and chemical shift. This peak is then calibrated and used as a reference for quantitation. A drawback of this method is that the quantitation error can be large when NMR peaks are attenuated due to altered pulse widths arising, for example, from lossy biological solutions. More recently, a method known as PULCON (pulse length based concentration determination) alleviates the limitation of ERETIC and provides a more robust approach to quantitation without the need for an internal standard (Wider and Dreier 2006). PULCON, also known as ERETIC 2, works based on the principle of reciprocity (Hoult and Richards 1976; Hoult 2000; Van der Klink 2001) and allows the correlation of signal strength from a reference spectrum with the spectrum of interest. This method shows immense promise for metabolomics applications (Jiménez et al. 2018; Goldoni et al. 2016). The method, however, requires that reference and test spectra are obtained using the same probe, probe tuning and matching for the test samples should be identical to the reference sample, and the same RF power needs to be delivered to the coil for each NMR spectrum. Poor probe matching for salty test samples, for example, will lead to inaccurate quantitation of metabolites. Despite this limitation, however, by using standardized operating procedures quantitative data can be obtained that can potentially enable sharing of inter-laboratory results.

Another approach for quantitation is to use the solvent signal as a concentration reference (Mo and Raftery 2008). Most solvents can be observed by NMR and solvent concentrations can be readily determined independently. In particular, a widely used solvent such as water can serve as a primary concentration standard for metabolite quantitation. The potential problems of radiation damping associated with a strong NMR signal can be alleviated by small pulse angle excitation. The fact that the solvent signal can be detected by the NMR receiver with the same efficiency as analytes enables their accurate quantitation. It is shown by this approach that analyte concentration can be accurately determined from 4 μM to more than 100 M.

2.3 Quantitation of Metabolites Using Intact Samples

The ability to analyze intact samples with no need for sample preparation or separation using NMR is an important characteristic that continues to drive NMR-based metabolomics. Initially, widely used bio-specimens including blood serum and plasma were used only in their intact forms and this approach continues to be widely used. In the following sections, analyses of intact bio-specimens that are most widely used such as serum/plasma, urine, and tissue are described. The methods presented here are also applicable for other specimen types.

2.3.1 Intact Serum and Plasma Analysis

Analysis of intact serum or plasma enables quantitation of aqueous metabolites as well as lipids and various classes of lipoproteins in serum and plasma (Würtz et al. 2017). Two widely used, one-dimensional (1D) NMR pulse techniques are 1D NOESY (nuclear Overhauser enhancement spectroscopy) and CPMG (Carr-Purcell-Meiboom-Gill) with water signal suppression (often using presaturation) (Nicholson et al. 1995). The 1D NOESY detects both small molecules such as metabolites and macromolecules such as lipids and lipoproteins. On the other hand, the CPMG experiment detects only small molecules; the macromolecule signals from proteins and lipoproteins are suppressed based on a T_2 (transverse relaxation) filter (Beckonert et al. 2007); metabolites exhibit longer T_2 relaxation times compared to macromolecules and hence they are selectively retained in the CPMG spectra. Numerous large-scale epidemiological studies have demonstrated quantitation of 50–70 metabolite peaks and over 200 metabolic measures (which include ratios of metabolite peaks) on a routine basis (Soininen et al. 2015; Würtz et al. 2017). As described above, recent advances in NMR enable absolute quantitation using an external reference, with no need for an internal standard (Wider and Dreier 2006). This is remarkable considering that internal standards largely cannot be used for absolute quantitation since they interact with copious proteins present in the samples. However, a challenge for reliable analysis of metabolites in intact samples is that metabolite binding to proteins causes signal attenuation (Nicholson and Gartland 1989; Chatham and Forder 1999; Bell et al. 1988; Nagana Gowda and Raftery 2014b). Moreover, exchange of metabolites between free and protein bound forms results in broader NMR peaks. Further, residual macromolecule signals cause distorted spectral baseline in CPMG spectra, which together adversely affect metabolite quantitation.

2.3.2 Intact Urine Analysis

Urine provides a rich source of information as it contains a significantly higher number of detectable metabolites, compared to serum/plasma, and with a vast concentration range (~10^6). In addition, urine has a relatively low concentration of proteins and hence macromolecular interference is minimal for metabolite analysis. A step-by-step procedure for NMR analysis of urine is provided as a guide for routine applications (Beckonert et al. 2007; Emwas et al. 2016). The pH of normal human urine varies widely, from approximately 5 to 8 (Hernandez et al. 2001; Rylander et al. 2006; Welch et al. 2008) and the salt concentration also varies significantly from sample to sample. Such pH and salt concentration variations alter chemical shifts of many peaks in the urine NMR spectra. Such peak shifts are significant for metabolites with functional groups with pKa's near the physiological pH. This causes a challenge for peak identification, comparison of different spectra, and quantitation of metabolites. Therefore, urine samples are generally mixed with buffer solution typically in a 1:1 (v/v) ratio (at pH = 7.4). Using Chenomx software and authentic compound spiking, >200 metabolites in urine have been identified (Bouatra et al. 2013). However, considering the high complexity of the urine NMR spectrum and the sensitivity of chemical shifts to factors such as pH and salt

concentration, the number of metabolites that can be analyzed on a routine basis is restricted to ~60 to 70. Factors such as diet, medications, physical activity, smoking, gender, age, gut microbe diversity greatly affect the metabolome and they should be carefully accounted for disease biomarker identification (Emwas et al. 2015, 2016). Large-scale (>1,000 samples) high-throughput studies now enable quantitative analysis of urinary metabolites using automated or semi-automated regression-based spectral analysis (Tynkkynen et al. 2019). In a large and impressive study, it was shown that prediction of metabolite concentrations, including many invisible inorganic ions, could be made based on the interrelationships between chemical shifts and concentrations, for automated urine analysis (Takis et al. 2017). Such advances promise new applications to areas including clinical, epidemiological, and pharmaceutical research.

2.3.3 Intact Tissue Analysis

NMR spectra of intact tissue are obtained using high-resolution magic angle spinning (HR-MAS) techniques (Tilgner et al. 2019). HR-MAS provides highly resolved spectra, which are comparable to those of bio-fluids. Tissue specimens typically collected from a surgical procedure or biopsy are often snap-frozen and stored for later analyses. The use of fresh samples for direct analysis, however, is advantageous for sensitive and structurally delicate biopsy samples. Resected or biopsied tissue is washed by quickly rinsing, typically with D_2O, to remove any blood contamination prior to freezing or direct analysis. The use of fresh samples avoids any deleterious effects caused by the freeze/thaw process and protects tissue integrity. Care should, however, be exercised to ensure fresh samples specifically from biopsy are kept under cold and humid conditions until the analyses are performed to retain the integrity of the metabolite profiles and reduce the possibility of metabolic changes. The ability to recover tissue after NMR analysis provides an opportunity to use the same specimens for other studies such as proteomic and genomic analysis or even histology. Advances in probe technologies with a ^2H field-frequency lock channel and a magnetic field gradient coil offer spectral stability and resolution sufficient for routine metabolomics studies of tissue samples as small as a few ng (Wong et al. 2012). Such capabilities, combined with minimal sample preparation and fast data acquisition, promise to extend the application of metabolic profiling of biopsied tissue to clinical applications. As examples, studies have shown that HR-MAS NMR of core needle biopsy tissue can predict breast tumor aggressiveness prior to surgery (Choi et al. 2012). Tissue metabolite profiles offer numerous benefits owing to the close association of tissue with disease pathologies. For example, alteration in tissue metabolite profiles has been shown to differentiate breast cancer tumors from normal tissue (Paul et al. 2018; Sitter et al. 2010). Importantly, HR-MAS NMR potentially enables diagnosis, prognosis, and staging of cancers (Dinges et al. 2019; Chen et al. 2017).

2.4 Metabolite Quantitation Using Processed Samples

Sample processing involves separation of metabolites from the macromolecular matrix. Such an approach enables detection of a significantly expanded pool of metabolites. A number of sample processing methods exist, to date, focused on analysis of aqueous metabolites or lipids or both. Further, new methods are being continuously developed that focus on improving the extraction efficiency, preserving metabolite integrity and simplifying the extraction process.

2.4.1 Analysis of Aqueous Metabolites

Methods such as ultrafiltration, solid phase extraction, and protein precipitation using organic solvents such as methanol, acetonitrile, acetone, perchloric acid, or trichloroacetic acid have been explored for many years to extract metabolites (Wevers et al. 1994; Daykin et al. 2002; Tiziani et al. 2008; Fan 2012). Among them, ultrafiltration using low molecular weight (~3 kDa) cut-off filter removes proteins most effectively. Using this method, nearly 50 aqueous metabolites could be identified and quantified (Psychogios et al. 2011). However, ultrafiltration attenuates many metabolite peaks (Nagana Gowda and Raftery 2014b), requires larger sample volumes, and is particularly cumbersome for large-scale studies. Nearly half of the detected metabolites in ultrafiltered serum exhibited lower concentrations ranging from nearly 10 to 75% (Nagana Gowda and Raftery 2014b). Further, ultrafiltration is incompatible with analysis of samples such as whole blood and tissue as well as with analysis using MS, which generally employs protein precipitation using organic solvents (Nagana Gowda et al. 2018).

Detailed studies have focused on increasing the number of detected metabolites, identifying unknown metabolites and optimizing their quantitation in blood serum and plasma (Nagana Gowda and Raftery 2014b; Nagana Gowda et al. 2015). These studies have shown that protein precipitation using methanol in a 2:1 ratio (v/v) with the sample offers an optimal approach for analysis of aqueous metabolites in blood serum/plasma. The use of acetonitrile for protein precipitation, on the other hand, revealed a surprisingly poor performance; one-third of the detected metabolites were attenuated by up to 70% compared to methanol precipitation at the same solvent to serum ratio of 2:1 (v/v) (Fig. 2). A further attenuation of nearly two-third of the metabolites was observed for an acetonitrile to serum ratio of 4:1 (v/v). As the analysis of metabolites using MS invariably employs protein precipitation prior to analysis, methods developed for NMR analysis also help analysis using MS. The performance of sample processing for MS analysis is typically evaluated using the total number of ions detected, which is problematic (Ivanisevic et al. 2013) and is an inaccurate approach as far as quantitation is concerned.

Protein precipitation, however, does not remove macromolecules completely and the residual macromolecules (~2%) are water-soluble, which cause broad baselines in NMR spectra when obtained using the one-pulse or 1D NOESY pulse sequence (Nagana Gowda et al. 2015, 2021). The use of the CPMG sequence helps to suppress signals from these residual proteins (~2%) and provides a flat baseline. The CPMG sequence, however, causes a small attenuation for many signals due to differential T_2

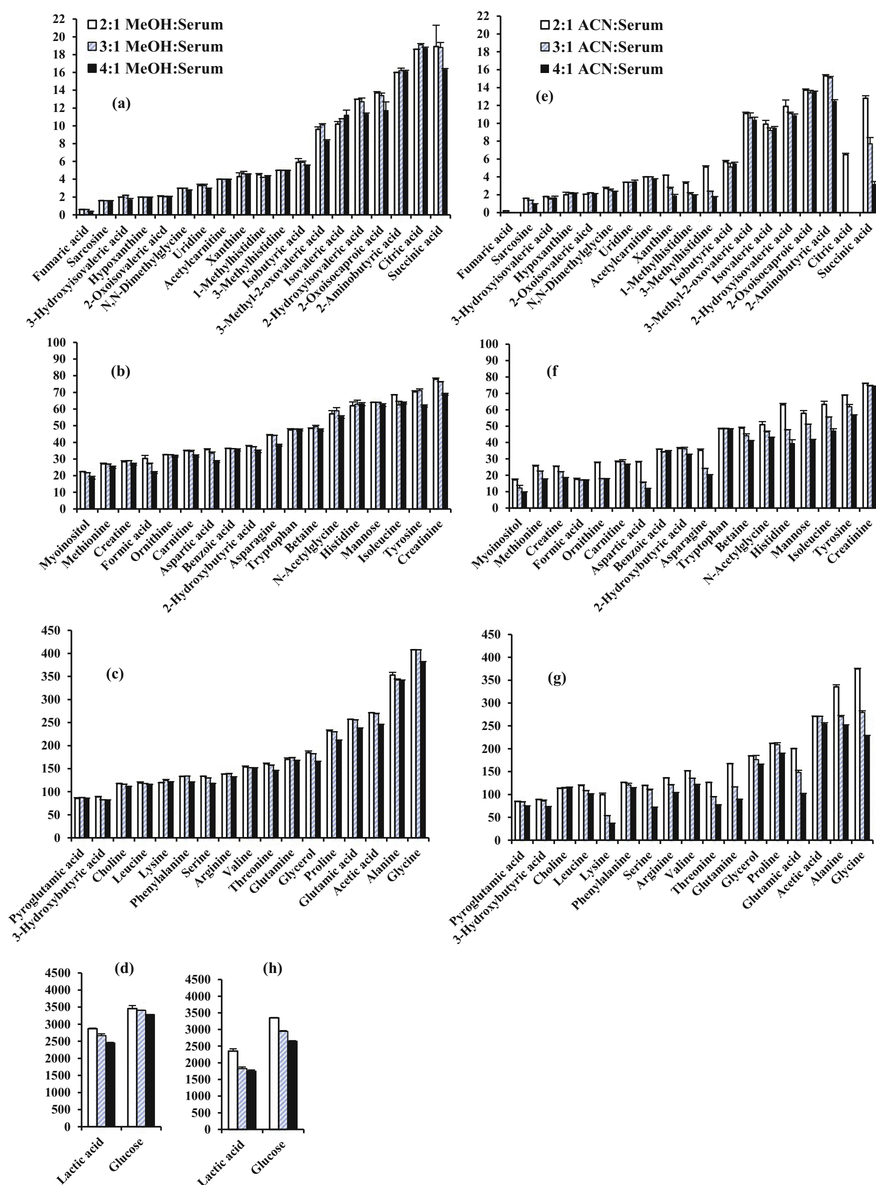


Fig. 2 Comparison of absolute concentrations (in μM) of metabolites detected in pooled human blood serum and quantitated using 800 MHz NMR spectroscopy after protein precipitation using methanol (MeOH) (**a**, **b**, **c** and **d**) or acetonitrile (ACN) (**e**, **f**, **g** and **h**) at a solvent to serum ratios of 2:1, 3:1, and 4:1. Methanol performs most optimally over a wide range and a methanol to serum ratio of 2:1 provides the best performance (reproduced with permission from Nagana Gowda et al. 2015)

relaxation rates; for example, an evaluation of 20 metabolite peaks revealed an average of ~6% attenuation for plasma and serum when a 32 ms CPMG 180° echo pulse train was used. The peak attenuation increased with increasing duration of the echo pulse train and it exceeded 10% for a 256 ms echo pulse train. Hence, for accurate quantitation, signal attenuation due to T_2 relaxation in the CPMG spectra should be carefully accounted. Potential alternatives to the CPMG sequence, including the use of stimulated echo (STE) pulse sequence (Lucas et al. 2005) have proved unsuitable for metabolomics applications.

2.4.2 Analysis of Coenzymes and Antioxidants

Coenzymes, including coenzyme A (CoA), acetyl coenzyme A (acetyl-CoA), coenzymes of redox reactions and energy, and antioxidants mediate biochemical reactions fundamental to the functioning of all living cells. The most common redox coenzymes include NAD^+ (oxidized nicotinamide adenine dinucleotide), NADH (reduced nicotinamide adenine dinucleotide), $NADP^+$ (oxidized nicotinamide adenine dinucleotide phosphate), and NADPH (reduced nicotinamide adenine dinucleotide phosphate). The coenzymes of energy include ATP (adenosine triphosphate), ADP (adenosine diphosphate), and AMP (adenosine monophosphate). Major antioxidants include GSSG (oxidized glutathione) and GSH (reduced glutathione). Conventional enzymatic assays are suboptimal, as separate protocols are needed for analysis of each coenzyme or their ratios. The interference from sample matrix and the finite linear range of these assays further add to the challenges. Although MS is extensively used, ion suppression, interference due to the unit mass difference in targeted analysis, and in-source fragmentation pose challenges for reliable coenzyme analysis (Evans et al. 2010; Trammell and Brennera 2013). Hence, the ability to analyze these coenzymes in one-step using NMR represents an important advancement in the metabolomics field. A major challenge unconnected with any analytical method, however, is the notoriously unstable nature of these compounds. Enzyme activity and oxidation affect their levels, deleteriously. Somewhat recently, sample harvesting, processing, and analysis conditions were optimized for heart tissue from mouse models and showed that a simple NMR experiment can simultaneously measure NAD^+, NADH, $NADP^+$, NADPH, ATP, ADP, and AMP in one step apart from other metabolites (Nagana Gowda et al. 2016, 2018). Later, the scope of NMR was extended to the analysis of CoA, acetyl-CoA, and antioxidants (GSH, GSSG) along with a large pool of other metabolites and coenzymes, in one step (Nagana Gowda et al. 2019) (Fig. 3). Further, as an important alternative to serum/plasma metabolomics, it was shown that using whole blood, the coenzymes and antioxidants can be measured simultaneously in addition to the nearly 70 metabolites that can be quantitated in serum/plasma with essentially no additional effort (Nagana Gowda and Raftery 2017b). The analysis protocols and the annotated characteristic fingerprints for these newly identified coenzymes and other metabolites are provided for easy identification and absolute quantification using a single internal reference. The ability to measure the unstable but ubiquitous coenzymes fundamental to cellular functions, simultaneously and reliably, offers a new avenue to investigate the mechanistic details of cellular function in health and diseases.

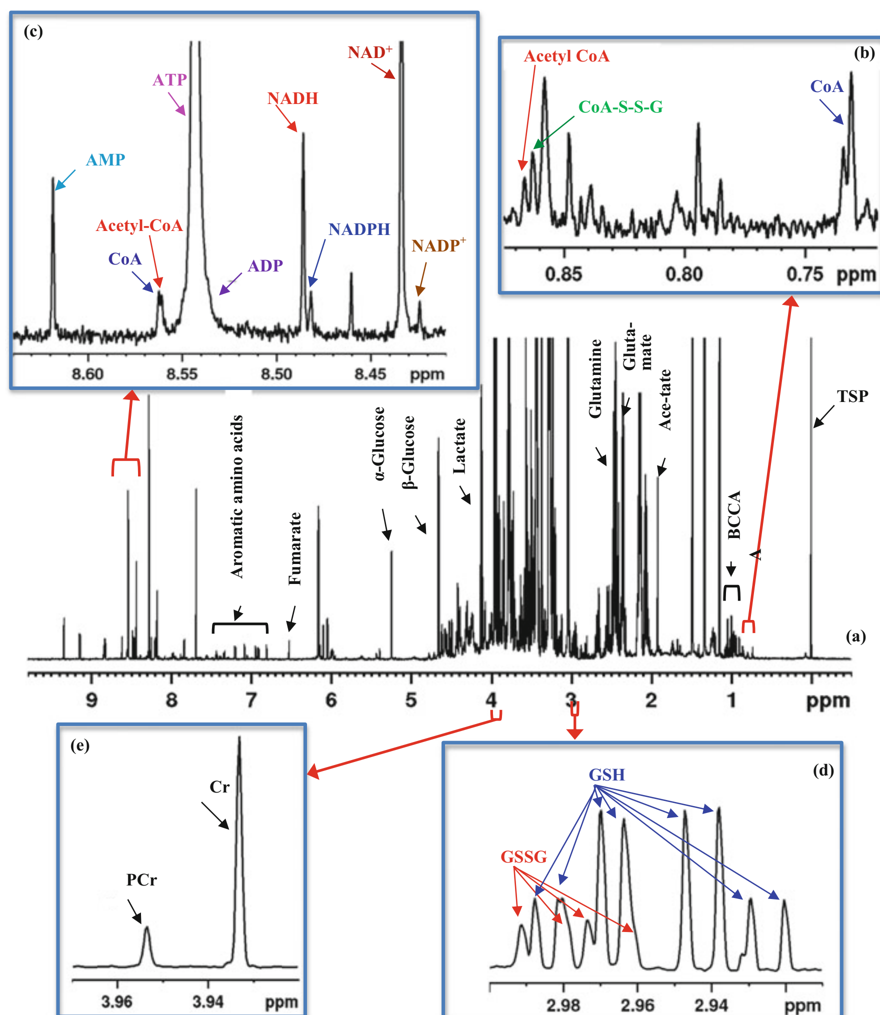


Fig. 3 (**a**) Typical 800 MHz ^{1}H NMR spectrum of a mouse heart tissue extract with labeling of some of the metabolites: BCCA: branched chain amino acids; TSP: reference peak; (**b**–**e**) expanded spectral regions highlighting characteristic peaks for (**b**) coenzyme A (CoA), acetyl coenzyme A (acetyl-CoA), and coenzyme A glutathione disulfide (CoA-S-S-G); (**c**) CoA, acetyl-CoA, oxidized nicotinamide adenine dinucleotide (NAD^{+}), oxidized nicotinamide adenine dinucleotide phosphate (NADP^{+}), reduced nicotinamide adenine dinucleotide (NADH), reduced nicotinamide adenine dinucleotide phosphate (NADPH), adenosine triphosphate (ATP), adenosine diphosphate (ADP), and adenosine monophosphate (AMP); (**d**) reduced glutathione (GSH) and oxidized glutathione (GSSG); and (**e**) creatine (Cr) and phosphocreatine (PCr) (reproduced with permission from Nagana Gowda et al. 2019)

2.4.3 Analysis of Lipids

NMR spectroscopy is widely used for analysis of lipids and lipoprotein particles in serum and plasma (Mallol et al. 2013). Identification and quantitation of lipoprotein particles by NMR exploits the characteristic chemical shifts of the methyl resonances of fatty acid chains of lipids from different particle sizes, with peaks from smaller particles appearing at lower frequencies. Methodologies used to characterize lipoprotein particles based on methyl resonances utilize either deconvolution (Jeyarajah et al. 2006, Kaess et al. 2008) or statistical (Soininen et al. 2009) methods. These methods have enabled determination of particle size and number for lipoprotein classes such as VLDL (very low-density lipoprotein), LDL (low-density lipoprotein), and HDL (high-density lipoprotein) and up to 14 (or more) lipoprotein subclasses. The ability to quantitate a variety of lipoprotein particles using NMR has opened avenues for clinical assessment and management of cardiovascular disease risk. In view of the fact that such lipoprotein classification and sub-classification using NMR is superior to the conventional methods, the method has been commercialized to manage the risk of heart diseases. Somewhat recently, a diffusion-based method was proposed to characterizing lipoprotein particles (Mallol et al. 2015). Here, two-dimensional diffusion-ordered ^{1}H NMR spectroscopy (DOSY) was used to measure diffusion coefficients, which provide information on the particle sizes of lipoproteins (Johnson 1999). The lipoprotein particle numbers are then calculated by dividing the peak volume by the size of lipoprotein particles. The ability to directly calculate lipoprotein sizes using the DOSY method was purported (Mallol et al. 2015) to provide a more accurate results for the particle numbers than the commercialized methods, which are based on 1D NMR.

After extraction, typically using a mixture of organic solvents, the analysis of tissue or blood samples provides quantitative information on individual lipids or lipid classes. The Folch extraction, consisting of chloroform/methanol/water in a volumetric ratio of 8:4:3 (v/v/v) is one of the earliest and most popular methods (Folch et al. 1957). Since then, numerous different lipid extraction protocols with modification to Folch et al. (1957) or Bligh and Dyer method (Bligh and Dyer 1959) have been proposed for biological specimens such as blood, tissue, and cells. A more recent method, involving butanol-methanol (BUME), eliminates the need for chloroform, which is an hazardous solvent (Löfgren et al. 2012; Cruz et al. 2016). More recently, the BUME method was modified to suit the analysis of lipids using NMR spectroscopy (Barrilero et al. 2018). This method replaces heptane with diisopropyl ether as the organic solvent, since peaks from the residual heptane overlap with lipid signals. Notably, this method has enabled identification and quantitation of 15 different lipid classes including fatty acids, triglycerides, phospholipids, and cholesterols in serum. A semiautomatic software, LipSpin, converts raw NMR data based on mathematical and reference spectral models and provides quantitative information on lipids (Barrilero et al. 2018). Detailed protocols for extraction and quantitative analysis of lipids in biological specimens such as serum, tissue, and cells are provided, which serve as a practical guide for beginners in the field (Gil et al. 2019).

2.5 Quantitation Methods Using Stable Isotope Labeling

Stable isotope incorporation in vivo or ex vivo offers opportunities to quantitate metabolites using NMR with improved resolution and sensitivity. In vivo analysis of metabolites in live systems enables monitoring of dynamic changes, measuring fluxes and monitoring metabolism in real time. The use of heteronuclear 2D (two-dimensional) NMR pulse techniques involving stable isotopes offers a combination of selectivity, sensitivity, and resolution and alleviates major challenges in NMR experiments involving nuclei with low natural abundance. To date, stable isotopes including ^{13}C, ^{15}N, ^{2}H, and/or ^{31}P have been employed for analysis of metabolites in biological mixtures and investigation of metabolic pathways.

2.5.1 Isotope Labeling Focused on Metabolic Fluxes and Pathways

Isotope labeling in vivo enables measurement of fluxes and tracing of metabolic pathways. Using this approach, the same metabolite that flows through multiple pathways can be distinguished. A growing number of pathways, including glycolysis, pentose phosphate pathway, glutaminolysis, fatty acid oxidation, and TCA cycle can be investigated using the combination of NMR and selective or uniformly isotope labeled substrates such as ^{13}C-glucose and $^{13}C/^{15}N$-glutamine (Lin et al. 2019). Quantitative analysis of in vitro or in vivo isotope labeled metabolites can be measured either ex vivo, after extraction of metabolites, or in live systems in vitro or in vivo. While analysis after extraction provides a snapshot of metabolite levels at a particular time point, in situ analysis using live systems enables the measurement of the dynamic changes in metabolite levels and monitoring of metabolism in real time. Analysis after extraction of metabolites has been widely used in the metabolomics field. However, the growing technological and methodological advances in NMR are witnessing an increasing number of in vitro or in vivo investigations using live systems such as *C. elegans*, cells, and isolated mitochondria (Nguyen et al. 2020; Wen et al. 2015; Xu et al. 2018). Isotope labeled studies using cells and subcellular organelles enable understanding of metabolic pathways under controlled conditions. And the use of organisms, animal models, or humans can translate the findings from studies of cells and subcellular organelles to investigate the pathogenesis of human diseases (Fan et al. 2009; Locasale et al. 2011; Lane et al. 2011).

2.5.2 Isotope Labeling Focused on Metabolite Analysis

Isotope labeling in vivo in plants and organisms such as bacteria and yeast offers significant enhancement to spectral resolution and the detection sensitivity (Zhang et al. 2012; Chikayama et al. 2008; Bingol et al. 2012, 2013). In particular, it alleviates the challenges invariably met with the analysis involving low natural abundance heteronuclei and enables analysis of a large number of metabolites using conventional high-resolution 2D NMR experiments such as HSQC and HMBC. The uniform labeling using nuclei such as ^{13}C also enables characterization of metabolites based on homonuclear 2D ^{13}C NMR experiments. Carbon-bond topology networks obtainable from such homonuclear 2D ^{13}C experiments provide

additional avenues for metabolite identification (Chikayama et al. 2008; Bingol et al. 2012).

An altogether different approach is to label different classes of metabolites based on the specific functional group (Shanaiah et al. 2007; Desilva et al. 2009; Ye et al. 2009). Chemical derivatization of metabolites using a substrate that contains isotope such as ^{13}C, ^{15}N, or ^{31}P offers both sensitivity and resolution enhancement, owing to the high isotopic abundance and wide chemical shift dispersion imparted by the incorporated isotope. The ^{1}H decoupled 1D or 2D NMR spectrum involving the isotope labeled heteronuclei provides a single peak for each metabolite, which further adds to the sensitivity and resolution. Metabolite classes including amines, carboxylic acids, and hydroxyls have thus been tagged with isotopes and analyzed using 1D or 2D NMR (Shanaiah et al. 2007; Desilva et al. 2009; Ye et al. 2009; Vicente-Muñoz et al. 2021). Owing to its high natural abundance, ^{31}P signals from metabolites, however, can show up as strong background peaks in the ^{31}P enriched experiments, unlike the other nuclei. Incorporation of a "smart isotope tag" such as ^{15}N-cholamine enables analysis of the carboxylic acid class of metabolites using both NMR and MS methods (Tayyari et al. 2013). The smart isotope tag possesses an NMR sensitive isotope (^{15}N) that offers good chemical shift dispersion and a permanent positive charge that improves MS sensitivity and enables quantitation of metabolites more accurately by both NMR and MS. Such analysis allows direct comparison of NMR and MS data, which is an important characteristic for biomarker discovery and biological interpretation in the metabolomics field.

3 Conclusion

The ability to identify unknown metabolites, absolute quantitation and analysis of intact bio-specimens including live cells and subcellular organelles, is expanding the application of NMR to new and exciting areas in metabolomics. Technological advances have provided significant improvements to sensitivity and resolution, which have led to the identification and quantitation of an expanded pool of metabolites. NMR spectroscopy offers opportunities to gain mechanistic insights into biochemical pathways in health and diseases, to discover biomarkers and potential therapy targets, and to translate laboratory findings to clinical applications. Continuing, multifaceted efforts to boost sensitivity, resolution, and the speed of data acquisition and to improve quantitative accuracy promise to alleviate the increasingly realized complexity of biological mixtures and large-scale metabolomics studies. Moreover, ongoing technical and methodological advances contribute to further expanding the routinely quantifiable metabolites in biological specimens and hence NMR-based metabolomics is anticipated to greatly improve and impact our understanding of systems biology and to help make progress in the treatment and management of human diseases.

Acknowledgments The authors gratefully acknowledge the financial support from the NIH grants R01GM138465, R01GM131491, P30CA015704, and P30DK035816.

References

Airoldi C, Tripodi F, Guzzi C, Nicastro R, Coccetti P (2015) NMR analysis of budding yeast metabolomics: a rapid method for sample preparation. Mol Biosyst 11(2):379–383

Akoka S, Barantin L, Trierweiler M (1999) Concentration measurement by proton NMR using the ERETIC method. Anal Chem 71(13):2554–2557

Albrecht B, Voronina E, Schipke C, Peters O, Parr MK, Díaz-Hernández MD, Schlörer NE (2020) Pursuing experimental reproducibility: an efficient protocol for the preparation of cerebrospinal fluid samples for NMR-based metabolomics and analysis of sample degradation. Metabolites 10(6):251

Alum MF, Shaw PA, Sweatman BC, Ubhi BK, Haselden JN, Connor SC (2008) 4,4-Dimethyl-4-silapentane-1-ammonium trifluoroacetate (DSA), a promising universal internal standard for NMR-based metabolic profiling studies of biofluids, including blood plasma and serum. Metabolomics 4:122–127

Anderson JR, Phelan MM, Rubio-Martinez LM, Fitzgerald MM, Jones SW, Clegg PD, Peffers MJ (2020) Optimization of synovial fluid collection and processing for NMR metabolomics and LC-MS/MS proteomics. J Proteome Res 19(7):2585–2597

Bala L, Ghoshal UC, Ghoshal U, Tripathi P, Misra A, Nagana Gowda GA, Khetrapal CL (2006) Malabsorption syndrome with and without small intestinal bacterial overgrowth: a study on upper-gut aspirate using 1H NMR spectroscopy. Magn Reson Med 56(4):738–744

Barding GA Jr, Salditos R, Larive CK (2012) Quantitative NMR for bioanalysis and metabolomics. Anal Bioanal Chem 404(4):1165–1179

Barrilero R, Gil M, Amigó N, Dias CB, Wood LG, Garg ML et al (2018) LipSpin: a new bioinformatics tool for quantitative 1 H NMR lipid profiling. Anal Chem 90:2031–2040

Beckonert O, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC, Nicholson JK (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2(11):2692–2703

Bell JD, Brown JC, Kubal G, Sadler PJ (1988) NMR-invisible lactate in blood plasma. FEBS Lett 235:81–86

Bingol K, Zhang F, Bruschweiler-Li L, Brüschweiler R (2012) Carbon backbone topology of the metabolome of a cell. J Am Chem Soc 134:9006–9011

Bingol K, Zhang F, Bruschweiler-Li L, Brüschweiler R (2013) Quantitative analysis of metabolic mixtures by two-dimensional 13C constant-time TOCSY NMR spectroscopy. Anal Chem 85: 6414–6420

Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. Can J Biochem Physiol 37(8):911–917

Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC et al (2013) The human urine metabolome. Plos One 8(9):e73076

Chatham JC, Forder JR (1999) Lactic acid and protein interactions: implications for the NMR visibility of lactate in biological systems. Biochim Biophys Acta 1426(1):177–184

Chen W, Lu S, Wang G, Chen F, Bai C (2017) Staging research of human lung cancer tissues by high-resolution magic angle spinning proton nuclear magnetic resonance spectroscopy (HRMAS 1 H NMR) and multivariate data analysis. Asia Pac J Clin Oncol 13(5):e232–e238

Chikayama E, Suto M, Nishihara T, Shinozaki K, Kikuchi J (2008) Systematic NMR analysis of stable isotope labeled metabolite mixtures in plant and animal systems: coarse grained views of metabolic pathways. PLoS One 3:e3805

Choi JS, Baek HM, Kim S, Kim MJ, Youk JH, Moon HJ, Kim EK et al (2012) HR-MAS MR spectroscopy of breast cancer tissue obtained with core needle biopsy: correlation with prognostic factors. PLoS One 7:e51712

Cruz M, Wang M, Frisch-Daiello J, Han X (2016) Improved butanol-methanol (BUME) method by replacing acetic acid for lipid extraction of biological samples. Lipids 51(7):887–896

d'Alessandro A, Osterloh JD, Chuwers P, Quinlan PJ, Kelly TJ, Becker CE (1994) Formate in serum and urine after controlled methanol exposure at the threshold limit value. Environ Health Perspect 102(2):178–181
Daykin CA, Foxall PJ, Connor SC et al (2002) The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by (1)H nuclear magnetic resonance spectroscopy. Anal Biochem 304(2):220–230
DeSilva MA, Shanaiah N, Nagana Gowda GA, Raftery MA, Hainline BE, Raftery D (2009) Application of 31P NMR spectroscopy and chemical derivatization for metabolite profiling of lipophilic compounds in human serum. Magn Reson Chem 47Suppl 1:S74–S80
Dinges SS, Vandergrift LA, Wu S, Berker Y, Habbel P, Taupitz M, Wu C-L et al (2019) Metabolomic prostate cancer fields in HRMAS MRS-profiled histologically benign tissue vary with cancer status and distance from cancer. NMR Biomed 32(10):e4038
Djukovic D, Raftery D, Nagana Gowda GA (2020) Chapter 16 – Mass spectrometry and NMR spectroscopy based quantitative metabolomics. In: Issaq HJ, Veenstra TD (eds) Proteomic and metabolomic approaches to biomarker discovery, 2nd edn. Academic Press, pp 289–311
Edison S, Colonna M, Gouveia GJ, Holderman NR, Judge MT, Shen X, Zhang S (2021) NMR: unique strengths that enhance modern metabolomics research. Anal Chem 93:478
Emwas AH, Luchinat C, Turano P et al (2015) Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. Metabolomics 11(4):872–894
Emwas AH, Roy R, McKay RT, Ryan D, Brennan L, Tenori L, Luchinat C, Gao X, Zeri AC, Nagana Gowda GA, Raftery D, Steinbeck C, Salek RM, Wishart DS (2016) Recommendations and standardization of biomarker quantification using NMR-based metabolomics with particular focus on urinary analysis. J Proteome Res 15(2):360–373
Engel KM, Baumann S, Rolle-Kampczyk U, Schiller J, von Bergen M, Grunewald S (2019) Metabolomic profiling reveals correlations between spermiogram parameters and the metabolites present in human spermatozoa and seminal plasma. PLoS One 14(2):e0211679
Evans C, Bogan KL, Song P, Burant CF, Kennedy RT, Brenner C (2010) NAD+ metabolite levels as a function of vitamins and calorie restriction: evidence for different mechanisms of longevity. BMC Chem Biol 10:2
Fan TW (2012) In: Fan TW, Higashi RM, Lane AN (eds) The handbook of metabolomics, methods in pharmacology and toxicology, vol 2012. Springer, New York, pp 7–27
Fan TW, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, Miller DH (2009) Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). Mol Cancer 8:41
Folch J, Lees M, Stanley G (1957) A simple method for the isolation and purification of total lipides from animal tissues. J Biol Chem 226:497–509
Gil M, Samino S, Barrilero R, Correig X (2019) Lipid profiling using 1 H NMR spectroscopy. Methods Mol Biol 2037:35–47
Goldman A, Khiste S, Freinkman E, Dhawan A, Majumder B, Mondal J, Pinkerton AB, Eton E, Medhi R, Chandrasekar V, Rahman MM, Ichimura T, Gopinath KS, Majumder P, Kohandel M, Sengupta S (2019) Targeting tumor phenotypic plasticity and metabolic remodeling in adaptive crossdrug tolerance. Sci Signal 12(595):eaas8779. https://doi.org/10.1126/scisignal.aas8779
Goldoni L, Beringhelli T, Rocchia W, Realini N, Piomelli D (2016) A simple and accurate protocol for absolute polar metabolite quantification in cell cultures using quantitative nuclear magnetic resonance. Anal Biochem 501:26–34
Hernandez ME, Lopez AC, Calatayud AG et al (2001) Vesical uric acid lithiasis in a child with renal hypouricemia. An Esp Pediatr 55(3):273–276
Holzgrabe U (2010) Quantitative NMR spectroscopy in pharmaceutical applications. Prog Nucl Magn Reson Spectrosc 57(2):229–240
Hoult DI (2000) The principle of reciprocity in signal strength calculations – a mathematical guide. Concepts Magn Reson 12:173–187

Hoult DI, Richards RE (1976) The signal-to-noise ratio of the nuclear magnetic resonance experiment. J Magn Reson 24:71–85
Ivanisevic J, Zhu ZJ, Plate L, Tautenhahn R, Chen S, O'Brien PJ, Johnson CH, Marletta MA, Patti GJ, Siuzdak G (2013) Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism. Anal Chem 85(14): 6876–6884
Jeyarajah EJ, Cromwell WC, Otvos JD (2006) Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. Clin Lab Med 26:847–870
Jiménez B, Holmes E, Heude C, Tolson RF, Harvey N, Lodge SL, Chetwynd AJ et al (2018) Quantitative lipoprotein subclass and low molecular weight metabolite analysis in human serum and plasma by 1H NMR spectroscopy in a multilaboratory trial. Anal Chem 90(20): 11962–11971
Johnson CS (1999) Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. Prog Nucl Magn Reson Spectrosc 34:203–256
Johnson CH, Ivanisevic J, Siuzdak G (2016) Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 17:451–459
Kaess B, Fischer M, Baessler A, Stark K, Huber F, Kremer W, Kalbitzer HR et al (2008) The lipoprotein subfraction profile: heritability and identification of quantitative trait loci. J Lipid Res 49:715–723
Kalfe A, Telfah A, Lambert J, Hergenröder R (2015) Looking into living cell systems: planar waveguide microfluidic NMR detector for in vitro metabolomics of tumor spheroids. Anal Chem 87(14):7402–7410
Kapur BM, Vandenbroucke AC, Adamchik Y, Lehotay DC, Carlen PL (2007) Formic acid, a novel metabolite of chronic ethanol abuse, causes neurotoxicity, which is prevented by folic acid. Alcohol Clin Exp Res 31(12):2114–2120
Kodama M, Oshikawa K, Shimizu H, Yoshioka S, Takahashi M, Izumi Y, Bamba T, Tateishi C, Tomonaga T, Matsumoto M, Nakayama KI (2020) A shift in glutamine nitrogen metabolism contributes to the malignant progression of cancer. Nat Commun 11:1320
Kriat M, Confort-Gouny S, Vion-Dury J, Sciaky M, Viout P, Cozzone PJ (1992) Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards. NMR Biomed 5(4):179–184
Kubáň P, Boček P (2013) Direct analysis of formate in human plasma, serum and whole blood by in-line coupling of microdialysis to capillary electrophoresis for rapid diagnosis of methanol poisoning. Anal Chim Acta 768:82–89
Lane AN, Fan TW, Bousamra M II, Higashi RM, Yan J, Milleret DM (2011) Stable isotope-resolved metabolomics (SIRM) in cancer research with clinical application to NonSmall cell lung cancer OMICS. J Integr Biol 15:173–182
Lane AN, Tan J, Wang Y et al (2017) Probing the metabolic phenotype of breast cancer cells by multiple tracer stable isotope resolved metabolomics. Metab Eng 43(Pt B):125–136
Lin P, Lane AN, Fan TW (2019) Stable isotope-resolved metabolomics by NMR. Methods Mol Biol 2037:151–168
Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ, Heffron G et al (2011) Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nat Genet 43:869–874
Löfgren L, Ståhlman M, Forsberg GB, Saarinen S, Nilsson R, Hansson GI (2012) The BUME method: a novel automated chloroform-free 96-well total lipid extraction method for blood plasma. J Lipid Res 53:1690–1700
Lohavanichbutr P, Zhang Y, Wang P, Gu H, Nagana Gowda GA, Djukovic D, Buas MF, Raftery D, Chen C (2018) Salivary metabolite profiling distinguishes patients with oral cavity squamous cell carcinoma from normal controls. PLoS One 13(9):e0204249
Lucas LH, Larive CK, Wilkinson PS, Huhn S (2005) Progress toward automated metabolic profiling of human serum: comparison of CPMG and gradient-filtered NMR analytical methods. J Pharm Biomed Anal 39(1–2):156–163

Lussu M, Camboni T, Piras C, Serra C, Del Carratore F, Griffin J, Atzori L, Manzin A (2017) 1H NMR spectroscopy-based metabolomics analysis for the diagnosis of symptomatic E. coli-associated urinary tract infection (UTI). BMC Microbiol 17(1):201
Mallol R, Rodriguez MA, Brezmes J, Masana L, Correig X (2013) Human serum/plasma lipoprotein analysis by NMR: application to the study of diabetic dyslipidemia. Prog Nucl Magn Reson Spectrosc 70:1–24
Mallol R, Amigó N, Rodríguez MA, Heras M, Vinaixa M, Plana N, Rock E et al (2015) Liposcale: a novel advanced lipoprotein test based on 2D diffusion-ordered 1H NMR spectroscopy. J Lipid Res 56(3):737–746
Maniara G, Rajamoorthi K, Rajan S, Stockton GW (1998) Method performance and validation for quantitative analysis by ^{1}H and ^{31}P NMR spectroscopy. Applications to analytical standards and agricultural chemicals. Anal Chem 70(23):4921–4928
Maniscalco M, Cutignano A, Paris D, Melck DJ, Molino A, Fuschillo S, Motta A (2020) Metabolomics of exhaled breath condensate by nuclear magnetic resonance spectroscopy and mass spectrometry: a methodological approach. Curr Med Chem 27(14):2381–2399
Mo H, Raftery D (2008) Solvent signal as an NMR concentration reference. Anal Chem 80(24): 9835–9839
Nagana Gowda GA (2011) NMR spectroscopy for discovery and quantitation of biomarkers of disease in human bile. Bioanalysis 3(16):1877–1890
Nagana Gowda GA (2018) Profiling redox and energy coenzymes in whole blood, tissue and cells using NMR spectroscopy. Metabolites 8(2):32
Nagana Gowda GA, Raftery D (2014a) Advances in NMR based metabolomics. In: Simó C, Cifuentes A, García-Cañas V (eds) Fundamentals of advanced Omics technologies: from genes to metabolites, comprehensive analytical chemistry, vol 63. Elsevier, New York, pp 187–211
Nagana Gowda GA, Raftery D (2014b) Quantitating metabolites in protein precipitated serum using NMR spectroscopy. Anal Chem 86(11):5433–5440
Nagana Gowda GA, Raftery D (2015) Can NMR solve some significant challenges in metabolomics? J Magn Reson 260:144–160
Nagana Gowda GA, Raftery D (2017a) Recent advances in NMR-based metabolomics. Anal Chem 89(1):490–510
Nagana Gowda GA, Raftery D (2017b) Whole blood metabolomics by 1H NMR spectroscopy provides a new opportunity to evaluate coenzymes and antioxidants. Anal Chem 89(8): 4620–4627
Nagana Gowda GA, Raftery D (eds) (2019) NMR based metabolomics: methods and protocols, Methods in molecular biology, vol 2037. Humana Press/Springer Science, New York
Nagana Gowda GA, Gowda YN, Raftery D (2015) Expanding the limits of human blood metabolite quantitation using NMR spectroscopy. Anal Chem 87(1):706–715
Nagana Gowda GA, Abell L, Lee CF, Tian R, Raftery D (2016) Simultaneous analysis of major coenzymes of cellular redox reactions and energy using ex vivo (1)H NMR spectroscopy. Anal Chem 88(9):4817–4824
Nagana Gowda GA, Djukovic D, Bettcher LF, Gu H, Raftery D (2018) NMR-guided mass spectrometry for absolute quantitation of human blood metabolites. Anal Chem 90(3): 2001–2009
Nagana Gowda GA, Abell L, Tian R (2019) Extending the scope of 1H NMR spectroscopy for the analysis of cellular coenzyme a and acetyl coenzyme A. Anal Chem 91(3):2464–2471
Nagana Gowda GA, Hong NN, Raftery D (2021) Evaluation of Fumaric acid and maleic acid as internal standards for NMR analysis of protein precipitated plasma, serum, and whole blood. Anal Chem 93(6):3233–3240
Nguyen TTM, An YJ, Cha JW, Ko YJ, Lee H, Chung CH, Jeon SM et al (2020) Real-time in-organism NMR metabolomics reveals different roles of AMP-activated protein kinase catalytic subunits. Anal Chem 92(11):7382–7387

Nicholson JK, Gartland KP (1989) 1H NMR studies on protein binding of histidine, tyrosine and phenylalanine in blood plasma. NMR Biomed 2(2):77–82

Nicholson JK, Foxall PJD, Spraul M, Farrant RD, Lindon JC (1995) 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. Anal Chem 67:793–811

Orczyk-Pawilowicz M, Jawien E, Deja S, Hirnle L, Zabek A, Mlynarz P (2016) Metabolomics of human amniotic fluid and maternal plasma during normal pregnancy. PLoS One 11(4): e0152740

Paul A, Kumar S, Raj A, Sonkar AA, Jain S, Singhai A, Roy R (2018) Alteration in lipid composition differentiates breast cancer tissues: a 1 H HRMAS NMR metabolomic study. Metabolomics 14(9):119

Pauli GF, Gödecke T, Jaki BU, Lankin DC (2012) Quantitative 1H NMR. Development and potential of an analytical method: an update. J Nat Prod 75(4):834–851

Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I et al (2011) The human serum metabolome. PLoS One 6(2):e16957

Raftery D (ed) (2014) Mass spectrometry in metabolomics: methods and protocols, Methods in molecular biology, vol 1198. Humana Press/Springer Science, New York

Rundlöf T, Mathiasson M, Bekiroglu S, Hakkarainen B, Bowden T, Arvidsson T (2010) Survey and qualification of internal standards for quantification by 1H NMR spectroscopy. J Pharm Biomed Anal 52(5):645–651

Rylander R, Remer T, Berkemeyer S, Vormann J (2006) Acid–base status affects renal magnesium losses in healthy, elderly persons. J Nutr 136(9):2374–2377

Salem AA, Mossa HA (2012) Method validation and determinations of levofloxacin, metronidazole and sulfamethoxazole in an aqueous pharmaceutical, urine and blood plasma samples using quantitative nuclear magnetic resonance spectrometry. Talanta 88:104–114

Shanaiah N, Desilva MA, Nagana Gowda GA et al (2007) Class selection of amino acid metabolites in body fluids using chemical derivatization and their enhanced 13C NMR. Proc Natl Acad Sci U S A 104(28):11540–11544

Simón-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, Mallard WG et al (2013) Metabolite profiling of a NIST standard reference material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. Anal Chem 85(24):11725–11731

Sitter B, Bathen TF, Singstad TE, Fjøsne HE, Lundgren S, Halgunset J, Gribbestad IS (2010) Quantification of metabolites in breast cancer patients with different clinical prognosis using HR MAS MR spectroscopy. NMR Biomed 23(4):424–431

Soininen P, Kangas AJ, Wurtz P, Tukiainen T, Tynkkynen T, Laatikainen R, Jarvelin MR et al (2009) High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst 134:1781–1785

Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M (2015) Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet 8(1):192–206

Takis PG, Schäfer H, Spraul M, Luchinat C (2017) Deconvoluting interrelationships between concentrations and chemical shifts in urine provides a powerful analysis tool. Nat Commun 8(1):1662

Tayyari F, Nagana Gowda GA, Gu H, Raftery D (2013) 15N-cholamine – a smart isotope tag for combining NMR- and MS-based metabolite profiling. Anal Chem 85(18):8715–8721

Tilgner M, Vater TS, Habbel P, Cheng LL (2019) High-resolution magic angle spinning (HRMAS) NMR methods in metabolomics. Methods Mol Biol 2037:49–67

Tiziani S, Emwas AH, Lodi A, Ludwig C, Bunce CM, Viant MR, Güntheret UL (2008) Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. Anal Biochem 377(1):16–23

Trammell SA, Brennera C (2013) Targeted, LCMS-based metabolomics for quantitative measurement of NAD(+) metabolites. Comput Struct Biotechnol J 4:e201301012

Tynkkynen T, Wang Q, Ekholm J, Anufrieva O, Ohukainen P, Vepsäläinen J, Männikkö M et al (2019) Proof of concept for quantitative urine NMR metabolomics pipeline for large-scale epidemiology and genetics. Int J Epidemiol 48(3):978–993
Van der Klink JJ (2001) The NMR reciprocity theorem for arbitrary probe geometry. J Magn Reson 148:147–154
Vicente-Muñoz S, Lin P, Fan TW, Lane AN (2021) NMR analysis of carboxylate isotopomers of 13C-metabolites by chemoselective derivatization with 15N-cholamine. Anal Chem 93 (17):6629–6637
Welch AA, Mulligan A, Bingham SA, Khaw K-T (2008) Urine pH is an indicator of dietary acid-base load, fruit and vegetables and meat intakes: results from the European prospective investigation into cancer and nutrition (EPIC)- Norfolk population study. Br J Nutr 99(6):1335–1343
Wen H, An YJ, Xu WJ, Kang KW, Park S (2015) Real-time molnitoring of cancer cell metabolism and effects of an anticancer agent using 2D in-cell NMR spectroscopy. Angew Chem Int Ed Engl 54(18):5374–5377
Wevers RA, Engelke U, Heerschap A (1994) High-resolution 1H-NMR spectroscopy of blood plasma for metabolic studies. Clin Chem 40(7 Pt 1):1245–1250
Wider G, Dreier L (2006) Measuring protein concentrations by NMR spectroscopy. J Am Chem Soc 128(8):2571–2576
Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15:473–484
Wishart DS (2019) NMR metabolomics: a look ahead. J Magn Reson 306:155–161
Wong A, Jiménez B, Li X, Holmes E, Nicholson JK, Lindon JC, Sakellariou D (2012) Evaluation of high resolution magic-angle coil spinning NMR spectroscopy for metabolic profiling of nanoliter tissue biopsies. Anal Chem 84(8):3843–3848
Würtz P, Kangas AJ, Soininen P, Lawlor DA, Smith GD, Ala-Korpela M (2017) Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -Omic technologies. Am J Epidemiol 186(9):1084–1096
Xu WJ, Wen H, Kim HS, Ko YJ, Dong SM, Park IS, Yook JI, Park S (2018) Observation of acetyl phosphate formation in mammalian mitochondria using real-time in-organelle NMR metabolomics. Proc Natl Acad Sci U S A 115(16):4152–4157
Yazdani M, Elgstøen KBP, Rootwelt H, Shahdadfar A, Utheim ØA, Utheim TP (2019) Tear metabolomics in dry eye disease: a review. Int J Mol Sci 20(15):3755. https://doi.org/10.3390/ijms20153755
Ye T, Mo H, Shanaiah N, Nagana Gowda GA, Zhang S, Raftery D (2009) Chemoselective 15N tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. Anal Chem 81(12):4882–4888
Zebrowska A, Skowronek A, Wojakowska A, Widlak P, Pietrowska M (2019) Metabolome of exosomes: focus on vesicles released by cancer cells and present in human body fluids. Int J Mol Sci 20(14):3461
Zhang F, Bruschweiler-Li L, Brüschweiler R (2012) High-resolution homonuclear 2D NMR of carbon-13 enriched metabolites and their mixtures. J Magn Reson 225:10–13
Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, Mohney RP, Small KS, Bell JT, Steves CJ, Valdes AM, Spector TD, Menni C (2018) The fecal metabolome as a functional readout of the gut microbiome. Nat Genet 50:790–795

Advancements in Pulsed Stable Isotope-Resolved Metabolomics

Martin Forbes, Sabrina Geisberger, Matthias Pietzke, Guido Mastrobuoni, and Stefan Kempa

Contents

1 Introduction: A Brief History of Isotopic Labelling 166
2 Time- and Isotope-Resolved Metabolomics 168
2.1 Stable Isotope-Resolved Metabolomics 168
2.2 Time Resolved Isotope Labelling Studies 169
2.3 Isotope Tracing at the Crossroad of Metabolic Pathways 172
2.4 The Application of Ultra-High Resolution Mass Spectrometry Allows the Tracing of Different Isotopic Species 174
3 Applications and Future Perspectives 176
3.1 Applications of pSIRM 176
3.2 Perspective Towards Single Cells 177
References 178

Abstract

The understanding of biochemical processes of metabolism is gained through the measurement of the concentration of intermediates and the rate of metabolite conversion. However, the measurement of metabolite concentrations does not give a full representation of this dynamic system. To understand the kinetics of metabolism, the system must be described and quantified in terms of metabolite flow as a function of time. In order to measure the metabolite flow, or more precisely the metabolic flux through a biological system, substrates of the cell are labelled with stable isotopes. The usage of these substrates by the cell leads to the incorporation of the isotopes into downstream intermediates.

M. Forbes · S. Geisberger · M. Pietzke · G. Mastrobuoni · S. Kempa (✉)
Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Berlin, Germany
e-mail: Martin.Forbes@mdc-berlin.de; Sabrina.Geisberger@mdc-berlin.de; pietzke@molgen.mpg.de; Guido.Mastrobuoni@mdc-berlin.de; Stefan.Kempa@mdc-berlin.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_621

The most important metabolic pathways are encompassed in the central carbon metabolism (CCM). According to the Kyoto Encyclopedia of Genes and Genomes (KEGG), the central carbon metabolism "is the most basic aspect of life". It includes all metabolites and enzymatic reactions within: glycolysis and gluconeogenesis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, oxidative phosphorylation (OXPHOS), amino acids and nucleotide metabolic pathways. Some molecules are at the crossroad of metabolic pathways, interconnecting diverse metabolic and therefore functional outcomes. Labelling these nodal metabolites and analysing their isotopic composition allows the precise determination of the metabolic flow within the biochemical networks that they are in.

Application of stable isotope labelled substrates allows the measurement of metabolic flux through a biochemical pathway. The rapid turnover of metabolites in pathways requires pulse-feeding cells with a labelled substrate. This method allows for the determination of different cell states. For example, the action of a drug from immediate impact until the compensatory response of the metabolic system (cell, organs, organisms). Pulsed labelling is an elegant way to analyse the action of small molecules and drugs and enables the analysis of regulatory metabolic processes in short time scales.

Keywords

Cancer metabolism · Isotope-resolved metabolomics · Mass spectrometry methods · Metabolic flux analysis

1 Introduction: A Brief History of Isotopic Labelling

The use of isotope labelled molecules in metabolic research began within the first decades of the last century. In the early days of metabolism research, radioactive isotopes were used to investigate the metabolic flow in bacteria, plants and animals. In order to investigate the structure of biochemical pathways and specifically metabolic cycles, e.g. tricarboxylic acid cycle, reverse tricarboxylic acid cycle or Calvin-Benson cycle, isotopes were applied in a time resolved manner. Time resolution allowed for the elucidation of the consecutive order of chemical reactions within the investigated pathways. The use of radioactive isotopes was of paramount importance to decipher the flow of carbon within cells and organisms.

In 1910, English chemist Frederick Soddy observed that "elements of different atomic weights may possess identical (chemical) properties" and therefore belong to the same position in the periodic table. This included not only radioactive elements but also stable isotopes of atoms, i.e. atoms with the same number of protons but different numbers of neutrons in their nucleus. We can refer to these elements as "hot" or "cold" isotopes of an atom in dependency of their radio- or non-radioactivity. The presence of radioactive isotopes was discovered by black spots occurring on photosensitive emulsions, as the decaying radioactive element

produced traceable radiation. This phenomenon gave rise to years of research into the isotopes of the periodic table and their many uses in a broad range of analytical fields, including: chemistry, geology, biology, physics and medicine (Wilikinson 2018).

Georg Charles de Hevesy was a Hungarian radio-chemist and Nobel Prize in Chemistry laureate, recognised in 1943 for his achievements in the development of radioactive tracers in the study of metabolism in animals. Hevesy is considered the first to use radioisotopes to measure metabolic flux in biological systems. One such (sadly unpublished) experiment of Hevesy and Rutherford describes the practices of a shrewd landlady in 1911, Manchester. Hevesy was convinced that his landlady was recycling food, much to the denial of such practice. In order to thwart the thrifty landlady, Hevesy "spiked" a portion of leftover meat with a tiny amount of a radioactive material. A few days later he brought an electroscope to the table to demonstrate to the indignant landlady that the food served that day was radioactive. Hevesy is thought to have enjoyed fresh hot meals as a result of his radioactive tracing experiment (Myers 1996). Hevesy continued his endeavours into the usage of tracers in biological systems. In 1934, Hevesy and Hofer first used tracers in medicine by using an enriched stable isotope to determine the rate of elimination of water from the human body. Hevesy and Hofer drank dilute deuterated water and assayed the isotopic dilution of the deuterium in their urine. From their results they concluded the average time a water molecule spent in their bodies was 13 $\pm$ 1.5 days (Hevesy and Hofer 1934).

The accounts of Hevesy are far more than comical anecdotes. They provide an understandable example for the use of radioisotopes in analytical chemistry. Through the development of mass spectrometry, stable isotopes replaced radioisotopes in tracing experiments. In mass spectrometry, a metabolite can be identified by the mass spectrum of its fragments, i.e. the intensity of the fragment's peaks at a defined mass-to-charge (m/z) ratio. The additional neutron in the atomic nucleus of isotopes makes the atom 1 Da heavier and increases the m/z ratio. The incorporation of isotopes into a metabolite changes the atomic composition of this intermediate and induces a shift in the respective fragment's mass spectrum. Specifically, stable isotopes of carbon (^{13}C), nitrogen (^{15}N), oxygen (^{18}O) and hydrogen (^{2}H) can be introduced into organic compounds. By tracing the incorporation of the isotope, the metabolic fate of these compounds within biological systems can be characterised (Wilikinson 2018). The incorporation of these isotopes can be resolved in molecules such as sugars, amino acids or nucleotides. This method of isotope labelling was used to describe the effect of oncogenes on cancer cell metabolism. For example, Le et al. applied ^{13}C-Glucose and ^{13}C,^{15}N-Glutamine to trace glucose and glutamine metabolism when the MYC oncogene was induced in P493 cancer cells. The authors showed that glutamine plays an essential role in the cells' proliferation and survival, highlighting targets of glutamine metabolism for cancer therapy (Le et al. 2012). The usage of isotope labelling provides insights into the dynamics and kinetics of metabolism, as a function of time and cell state.

2 Time- and Isotope-Resolved Metabolomics

2.1 Stable Isotope-Resolved Metabolomics

In order to measure the metabolite flow, or more precisely the metabolic flux through a biological system, substrates of the system are labelled with stable isotopes. These substrates may be, glucose or glutamine which provide carbon and nitrogen to the central metabolic pathways (Fig. 1). The usage of these substrates by the cell leads to the incorporation of the isotopes into downstream intermediates. After isotopically labelling a biological system and extracting the cellular metabolites, mass spectrometry is then employed to analyse the number of heavy atoms incorporated (isotopologues) and their positions (isotopomers) in detected metabolites (Bruntz et al. 2017). Isotopologues can be identified by increasing mass shifts, as every heavy atom incorporated rises the m/z ratio by one. Isotopomers, on the other hand, can be distinguished by the heavy atom incorporation visible in different fragments

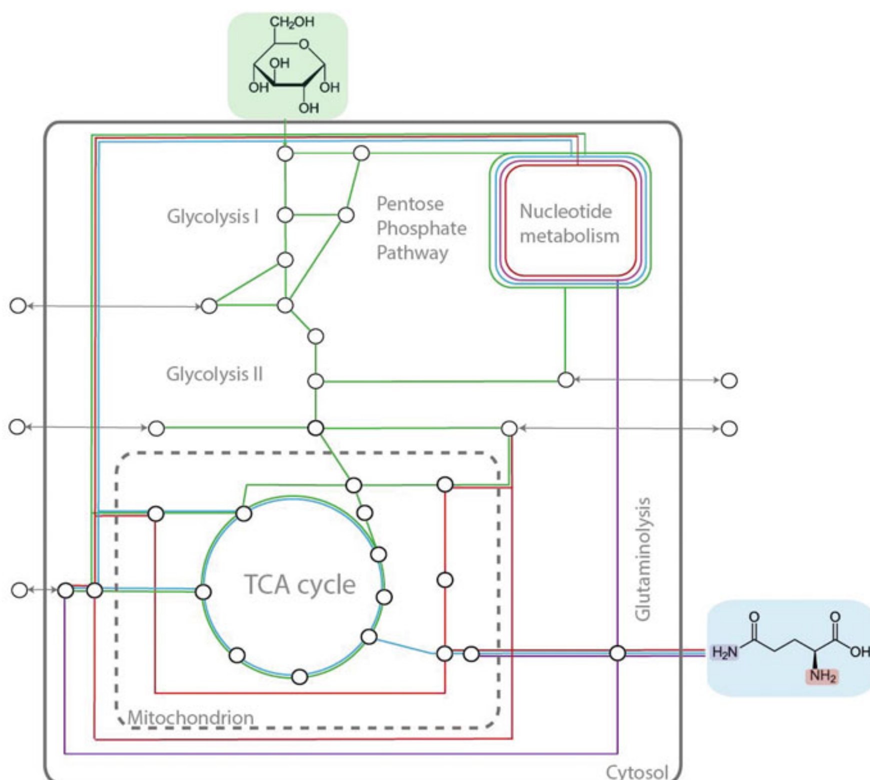


Fig. 1 Scheme of central metabolic pathways indicating carbon and nitrogen flow. The scheme displays main central metabolic pathways: glycolysis, pentose phosphate pathway, nucleotide metabolism, tricarboxylic acid cycle and glutaminolysis. The different colours depict the contribution of the distinct carbon and nitrogen atoms stemming from glucose or glutamine, respectively

of the same metabolite. In summary, the usage of isotope labelled substrates allows for the follow-up of interconnected anabolic and catabolic pathways. The tracing of isotope labels in metabolites, and the atomic position in which the label resides, is referred to as stable isotope-resolved metabolomics (SIRM).

Cellular metabolism is a rapid, dynamic process, in which both anabolism and catabolism of metabolites produce energy, build macromolecules for biomass and generate intermediates involved in intra- and inter-cellular signalling. The understanding of biochemical processes of metabolism is inferred through the concentration of intermediates and the rate of metabolite conversion. However, the measurement of metabolite concentrations does not give a full representation of this dynamic system. In order to understand the kinetics of metabolism, it must be described and quantified in terms of metabolite flow as a function of time (Sauer 2006; Pietzke et al. 2014; Buescher et al. 2015; Jang et al. 2018).

2.2 Time Resolved Isotope Labelling Studies

Metabolism is a highly dynamic system in which biochemical reactions occur rapidly. This allows the cell to constantly meet its needs and adapt to changing stimuli and micro-environments. The time for which cells are exposed to a labelled substrate determines the amount of isotopic label incorporated into its metabolites. Supplying a cell with an isotope labelled substrate for extended periods of time will lead to complete usage and saturation of downstream pathways. This condition is referred to as isotopic steady state and solely reflects the usage of metabolic pathways in relation to each other, i.e. in which pathways the given substrate is involved in. In other words, if one applies the labelled substrate for such a period of time that the label incorporation has reached saturation, then the model of flux would be stationary (Fig. 2, left (II)).

To understand the flux of labelled metabolites, we aim to assess metabolism in a non-stationary state. In order to measure the metabolic activity, the time a labelled substrate is offered to a cell is limited. A labelling time is chosen where the label incorporation into intermediates of interest is in linear relationship to time (Fig. 2, left (I)). At this chosen labelling time, one can analyse the speed of a specific pathway. It is also possible to assess how different conditions the cell is exposed to affect the velocity of the pathway in use.

By avoiding saturating the system, information about pathway preference and directionality is provided, as discussed in Sect. 2.3. The reduction in labelling time can be implemented by a pulsed labelling-quenching strategy, referred to as pulsed stable isotope-resolved metabolomics (pSIRM).

We may take glycolysis as a metabolic pathway that can be analysed by a tracing experiment. $^{13}C_6$-Glc is used as a substrate for the glycolytic pathway. This labelled tracer contains ^{13}C at every carbon position on the glucose molecule. As glycolysis proceeds ^{13}C is incorporated into downstream metabolites (Fig. 2, right). The rate of label incorporation can be described in both functions of time and quantity of labelled metabolite. The rate of label incorporation, or indeed the metabolic rate,

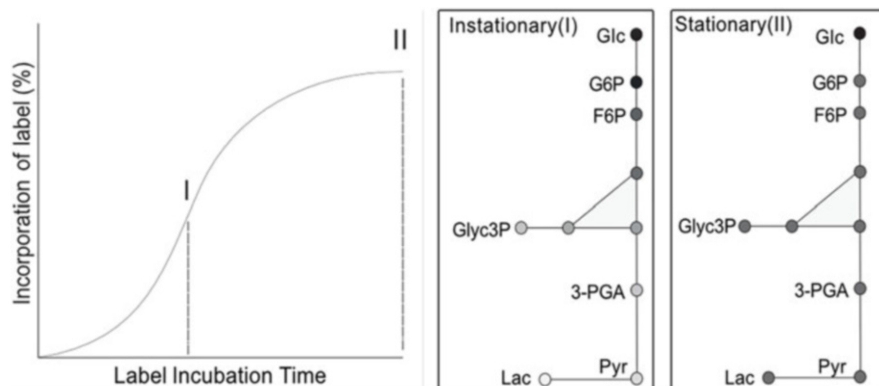


Fig. 2 Modelling the kinetics of metabolite labelling. As cells are incubated with a labelled substrate for a period of time, the quantity of label incorporated into metabolites increases. The rate of reaction, or the percentage of label in the metabolites relative to the unlabelled fraction, is dependent on the position of the metabolite in the pathway and the efficiency of associated enzymes. In order to understand the rate and kinetics of metabolic processes, we aim to observe labelled metabolites in the instationary phase (Fig. 2, left (I)). On the contrary, if all metabolites are saturated with label after a given period of time (Fig. 2, left (II)), then no kinetic information can be derived

of glycolysis is rapid. For instance, glucose-6-phosphate (G6P) – the primary step in glycolysis – is labelled to saturation in approximately 2–5 min (Pietzke et al. 2014). Longer labelling times are required for the label to reach saturation in metabolites downstream of G6P. Through a pulse-quench-harvest labelling strategy we can determine the rate of label incorporation into metabolites of the glycolytic pathway, as a function of the time cells are exposed to the labelled substrate. With this information we can determine how "glycolytic" a certain cell type is.

Furthermore, we can study the effects on inhibitors on pathways by monitoring the relative changes in label incorporation upon inhibitor treatment. 3-Bromopyruvate (BrPyr), for example, is a strong glycolytic inhibitor. More specifically, BrPyr inhibits glycolysis at the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) reaction (Pietzke et al. 2014). In order to localise BrPyr action on GAPDH, cells were pre-incubated with BrPyr for 12 min followed for a tracing with $^{13}C_6$-Glc for 3 min to label glycolytic metabolites. They showed that the carbon flow through metabolites downstream of GAPDH was almost completely inhibited in BrPyr treated cells (Fig. 3). The concentration of 3PGA drops below the detection limit and the dramatic decrease in the label incorporation into pyruvate, lactate and citrate proved bromopyruvate's inhibitory activity in the lower glycolysis. In contrast fructose-1,6-bisphosphate (above the blockage) accumulates roughly fourfold, while at the same time the label incorporation decreases by a factor 4, so the total flow into this compound remains constant. This illustrates the connection of metabolite pool size and label incorporation and highlights the importance of considering both readouts simultaneously during a pSIRM experiment.

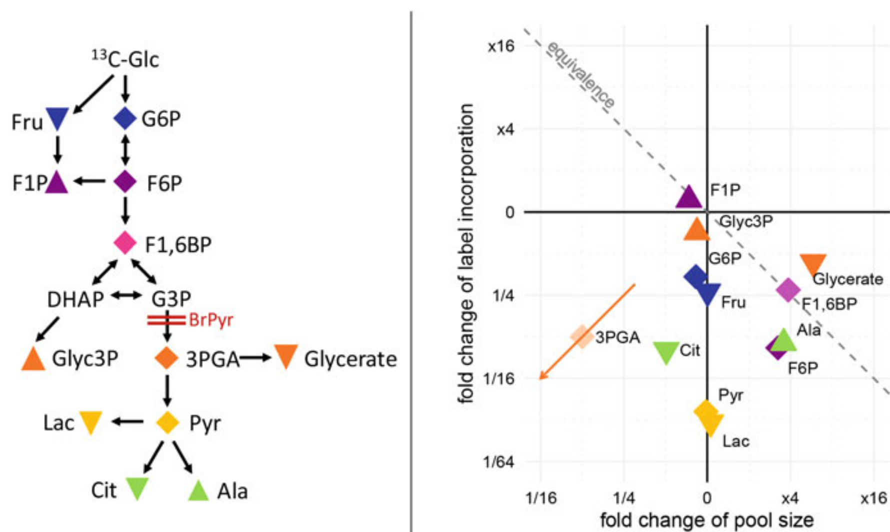


Fig. 3 pSIRM data in 2D. Each point represents one single metabolite of the CCM. On the x-axis fold-changes of metabolite concentrations and on the y-axis changes of isotope incorporation are shown. For this experiment cancer cells were incubated for a total time of 15 min with BrPyr and metabolites were measured using the pSIRM technology (see Pietzke et al. (2014)). The graph shows that in some metabolites a quantitative change in metabolite concentration can be observed. However, the effect on glycolytic inhibition can only be observed in the label incorporation of metabolites downstream of the GAPDH reaction

We have described the use of isotope labelled substrates for investigating the rate of reactions in metabolic pathways. One may also investigate the direction of metabolic pathways or deduce the relative dominance of a pathway by using substrates labelled at specific carbon positions. For example, the pentose phosphate pathway (PPP) – which links glycolysis to nucleotide synthesis – generates ribose-5-phosphate (R5P) from two reactions. The oxidative PPP utilises glucose-6-phosphate to produce 6-phosphogluconate which is de-carboxylated to ribulose-6-phosphate (Ru5P). In turn Ru5P is isomerised to R5P. The non-oxidative PPP cycles carbons from fructose-6-phosphate, glyceraldehyde-3-phosphate and erythrose-4-phosphate to produce xyulose-5-phosphate and R5P. This branch does not contain a de-carboxylation event and carbon number is maintained in the process. Therefore, we may employ a labelling strategy to decipher the ratio of the oxidative and non-oxidative branches by assessing the state of the de-carboxlyation event in the oxidative branch. By using 1,2-[^{13}C]-Glc as the labelling substrate we can measure the ratio of R5P present in the sample which contains a single (m/z + 1) or a double (m/z + 2) label. R5P which is derived from the oxidative PPP will only contain a single label as the labelled carbon is lost during the de-carboxylation event (Fig. 4).

Oxythiamine is an inhibitor of the non-oxidative PPP enzyme, transketolase (TKT). Boros et al. (1997) showed that oxythiamine inhibits Mia pancreatic adenocarcinoma cell growth by 39%. By using 1,2-[^{13}C]-Glc labelling the authors

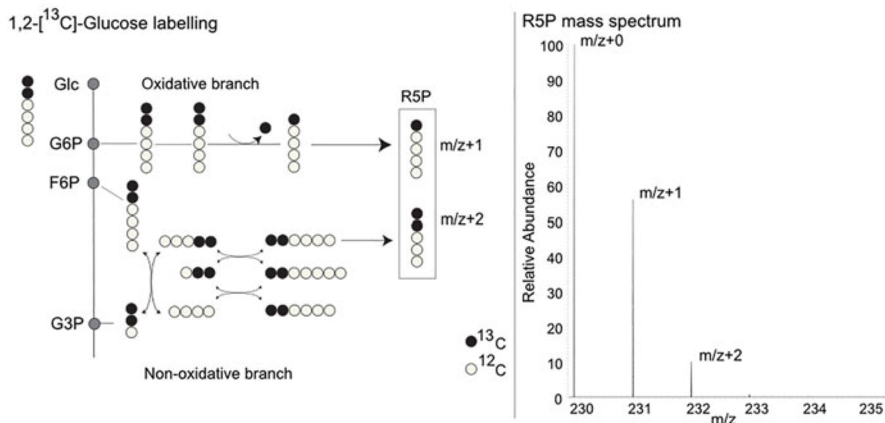


Fig. 4 Utilising 1,2-[^{13}C]-Glc in the analysis of the pentose phosphate pathway. The pentose phosphate pathway is comprised of the oxidative and non-oxidative branches. The oxidative branch observes a de-carboxylation event which removes a labelled carbon from glucose-6-phosphate. In the non-oxidative branch, labelled carbons are maintained as de-carboxylation does not occur. The ratio of single to double labelled R5P shows the relative dominance of the branches. One may observe the single or double labelled R5P in a mass spectrum generated by MS, by the relative abundance of R5P (m/z + 1) and R5P (m/z + 2). This is depicted in the simulated MS spectrum on the right

discovered that this tumour cell line generated R5P predominantly through transketolase and the non-oxidative PPP (85%). It was deduced that oxythiamine was an effective anti-tumour inhibitor due to the dependency of R5P synthesis, and downstream nucleotide synthesis, on the non-oxidative PPP (Boros et al. 1997). In summary, differential labelling of substrates permits not only the analysis of the kinetics of metabolism but also gives information on the relative dependencies of parallel reactions in different cell lines for targeted therapy development.

2.3 Isotope Tracing at the Crossroad of Metabolic Pathways

The most important metabolic pathways are encompassed in the central carbon metabolism (CCM). According to the Kyoto Encyclopedia of Genes and Genomes (KEGG), the central carbon metabolism “is the most basic aspect of life”. It includes all enzymatic reactions within: glycolysis and gluconeogenesis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle and oxidative phosphorylation (OXPHOS), amino acids and nucleotide metabolism pathways. Additionally, the CCM includes six known carbon fixation pathways (reductive pentose phosphate cycle (Calvin cycle), reductive citrate cycle, 3-hydroxypropionate bi-cycle, two variants of 4-hydroxybutyrate pathway and reductive acetyl-coenzyme A (CoA) pathway) as well as some pathways of methane metabolism, all not relevant in animal cells (Qiu 2013). Some molecules are at the crossroad of metabolic pathways,

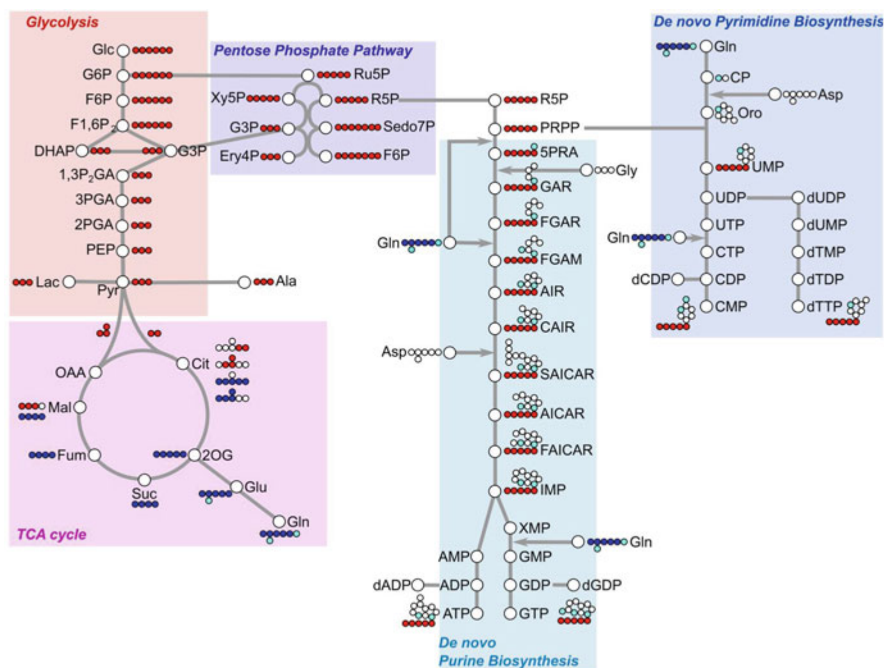


Fig. 5 ^{13}C and ^{15}N label integration into pathways of the central carbon metabolism. Pictogram of glycolysis, TCA cycle, pentose phosphate pathway, as well as de novo purine and pyrimidine biosynthesis. Red circles represent glucose-derived ^{13}C atoms. Blue represent the ^{13}C (dark blue) or ^{15}N (light blue) atoms derived from glutamine. Unlabelled C and N atoms are displayed in white

interconnecting diverse metabolic and therefore functional outcomes. Labelling these nodal metabolites and analysing their isotopic composition allows the precise determination of the metabolic flow within the biochemical networks that they are in (Fig. 5).

One example of such intersections is citrate. Glucose-derived pyruvate can be metabolised by pyruvate dehydrogenase to citrate or by pyruvate decarboxylase to oxaloacetate, which itself feeds into citrate during TCA cycling. On the other hand, glutamine can enter the TCA cycle via glutamic acid and 2-oxo-glutaric acid and feed into citrate via both, oxidative (classic) and reductive (reverse) TCA cycling. Via ATP-citrate lyase, citrate is further interconnected with fatty acid oxidation and fatty acid biosynthesis. When labelling with ^{13}C-glucose, ^{13}C-glutamine or ^{13}C-fatty acids one will be able to distinguish between different mass shifts in citrate, depending on the source of the ^{13}C-label as well as the directionality of TCA cycling. ^{13}C atoms derived from pyruvate dehydrogenase activity or fatty acid oxidation-derived acetyl-CoA will lead to an m/z + 2 mass shift in citrate, as two heavy carbon atoms are incorporated in the latter. On the contrary, an m/z + 3 mass shift is observed when pyruvate integrates into oxaloacetate and subsequently into citrate. Similarly, when providing cells with ^{13}C-glutamine, mass shifts of m/z + 4 and

m/z + 5 in citrate are the result of anaplerotic fuelling from glutamine into the TCA cycle and oxidative or reductive cycling, respectively.

Other examples of pathway interconnection are pyruvate and glutamine, as both can be metabolised into very different intermediates and play a role in distinct cellular functions. Pyruvate-derived ^{13}C-atoms can integrate, as described above, into citrate or oxaloacetate, but also into alanine or lactate. Thereby, metabolic flow through pyruvate merges amino acid metabolism, anaerobic glycolysis and TCA cycle. Glutamine, on the other hand, is an important nitrogen donor and is involved in amino acids metabolism and de novo biosynthesis of purine and pyrimidine nucleotides (Bayram et al. 2020). Additionally, via glutamic acid, the carbon backbone of glutamine feeds into TCA cycle, as well as polyamine and glutathione synthesis. The example of glutamine highlights the necessity of dual carbon and nitrogen labelling in order to greatly increase our understanding of pathway connectivity and metabolic fate. Combining ^{13}C- and ^{15}N-labelling allows the follow-up of glutamine usage, while distinguishing between the amino group, amido group and carbon backbone utilisation.

2.4 The Application of Ultra-High Resolution Mass Spectrometry Allows the Tracing of Different Isotopic Species

Using stable isotopes of different elements allows to determine the utilisation of two metabolic precursors (e.g. glucose and glutamine) in a single experiment simultaneously, or to follow the fate of different atoms from the same precursors (e.g. C and N from glutamine) into the downstream metabolic network.

For a long time such experiments have not been possible, since mass analysers with nominal mass resolution are not able to distinguish the mass increase due to the presence of different elemental isotopes. For example, the difference in the mass increase given by one ^{13}C or one ^{15}N is only 0.00632 Da (Fig. 6). According to the IUPAC definition (McNaught and Wilkinson 2008), the resolution needed to separate such mass difference for two molecules with nominal mass of 400 Da is around 60,000, far beyond the resolution of conventional quadrupole or ion trap analysers and barely reached by Time-of-Flight instruments. Only with the advent of (ultra) high resolution Fourier transform mass spectrometry mass analysers it became possible to analyse the isotopic fine structure of labelled compounds (Werner et al. 2008; Marshall and Hendrickson 2008).

One implementation of (ultra) high resolution MS for isotope tracing is the direct infusion of the metabolites in an Fourier-transform ion cyclotron resonance mass spectrometer (Le et al. 2012; Yang et al. 2017). This approach offers excellent results in terms of achievable resolution and accuracy in the determination of the ratios between different isotopes, because the ion signal in the ion cyclotron resonance cell can be averaged for long times. However, in direct infusion highly abundant analytes can suppress the ionisation of low abundant ones (Han et al. 2008). In addition, in absence of another mean of separation, isomers cannot be distinguished because they have the same accurate m/z ratio, like G6P and F6P. MS/MS experiments can help

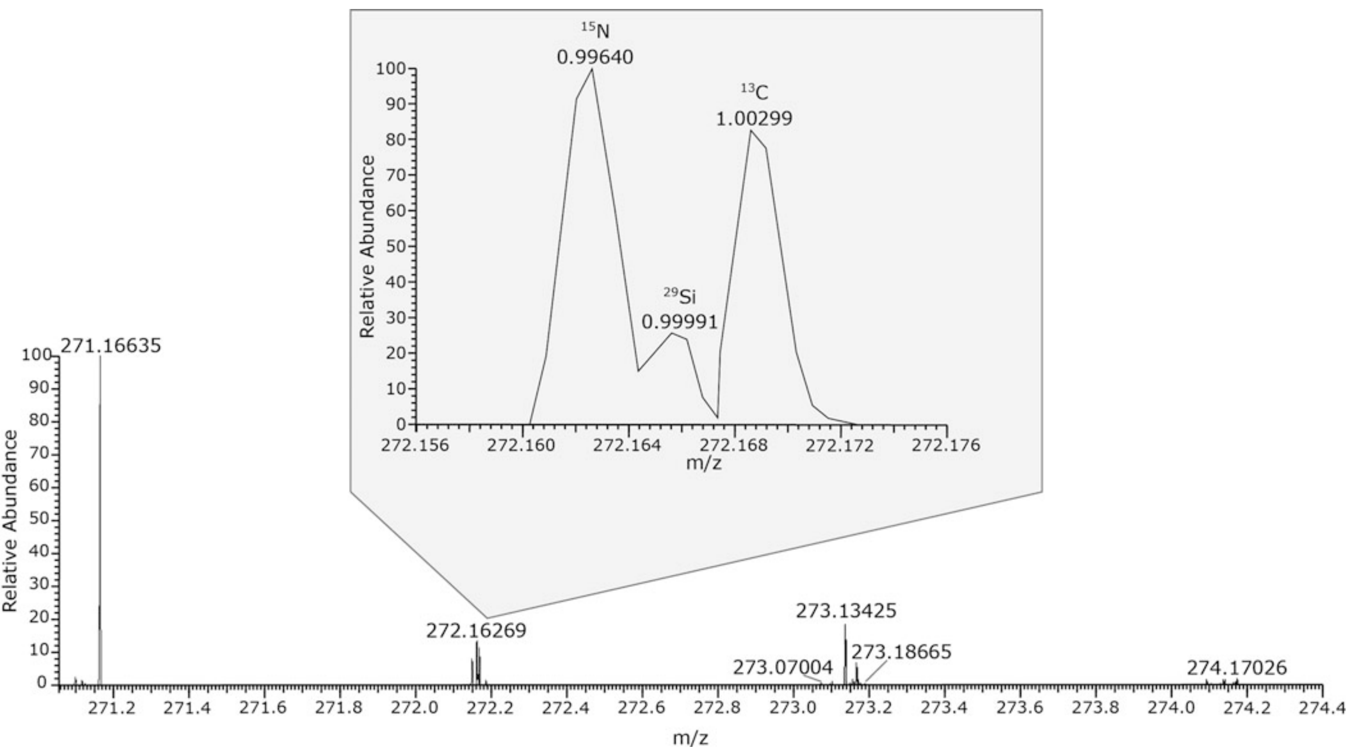


Fig. 6 MS spectrum of ^{15}N labelled Glutamine. The ion at m/z 271.16635 corresponds to the unlabelled fragment of TBDMS-derivatised glutamine with molecular formula C12H27ON2Si2. It contains 4 carbon and all nitrogen atoms of glutamine. Insert shows the magnification of the m/z + 1 peaks, revealing the separation of peaks due to the incorporation of ^{15}N or ^{13}C. The peak of ^{29}Si isotope, present in the derivatisation agent, is also resolved. The masses reported in the insert are the Δm relative to the monoisotopic peak

confirming or excluding the presence/absence of one or more isomers. Nevertheless, isomers often require different collision energies, making an accurate quantification difficult, if not impossible. Orbitrap mass analysers also can achieve high resolution with fast scanning rate (Makarov et al. 2006). With a scan rate up to 40 Hz (Kelstrup et al. 2018), this analyser is fast enough to be seamlessly hyphenated with GC or LC separation.

3 Applications and Future Perspectives

In line with the general "OMICS" concept, metabolomics aims to measure all metabolic components of a biological system at once in a quantitative manner. Unfortunately that's not possible so far. The chemical space of biomolecules regarding polarity and size as well as their quantities – ranging from a few molecules up to millimolar concentrations – is hindering its simultaneous detection. One step further is the measurement of the dynamics of metabolism. The combination of stable isotope labelling and mass spectrometric detection made a big contribution in this regard. More and more methods allow the simultaneous measurement of metabolite concentrations and isotope incorporation (Pietzke et al. 2014). Such integrated methods are a prerequisite when metabolic dynamics in vivo or in tissues will be analysed in a clinical context. Often a consecutive sampling and parallel quantification in additional samples is not possible. The application of pSIRM together with the present knowledge of the biochemical network allows to define certain metabolic nodes that contain superior information of the usage of metabolic pathways, e.g. citrate. Future method development may focus on such nodes, allowing for more directed pSIRM applications with tailored isotopically labelled substrates.

3.1 Applications of pSIRM

In the last years, pSIRM has been applied to successfully analyse metabolic changes within short and defined windows, e.g. during differentiation of cells. Delp et al. found that immature and mature neurons rely on different fuels. Precursor cells were found to be mainly glycolytic and strongly dependent on glutamine. During differentiation, however, they lost their glutamine dependency while gaining flexibility in energy production (Delp et al. 2018). By analysing the glutamine-derived carbon utilisation in high salt-treated differentiating macrophages a specific down-regulation of the succinate to fumarate conversion could be revealed. This observation led to the identification of sodium as strong regulator of complex II activity (Geisberger et al. 2021).

pSIRM can also be applied in vivo as well as ex vivo in tissue slices or organs. By administering $^{13}C_6$-glucose into the peritoneum of a hepatocellular carcinoma mouse model, Berndt et al. described enhanced glycolytic rates in tumours compared to normal liver. These data contributed to the creation of individualised metabolic profiles of tumours and modelling predictions of the efficacy of drug therapies

(Berndt et al. 2020). Similarly, application of stable isotopes in situ to cancerous and non-cancerous lung tissue revealed higher activity of glycolysis and the Krebs cycle in the tumour tissue (Fan et al. 2009). In brain tissue slices from naked mole rats, e.g. labelling with $^{13}C_6$-fructose uncovered the ability of these animals to metabolise fructose as fuel for neuronal tissues under hypoxic conditions (Park et al. 2017).

Beside the application of stable isotope labelled organic compounds, inorganic compounds such as CO_2 can be used for labelling experiments. For example, Rohwer et al. used $^{13}CO_2$ to demonstrate the reductive carboxylation in gastric cancer cells (Rohwer et al. 2016). These are only few examples of the wide range of applications pSIRM; a very versatile technique to describe metabolism in function of quantity and time.

3.2 Perspective Towards Single Cells

The advent of single cell "OMICS" has allowed for the analysis of the architecture of heterogenous cell samples at single cell resolution. Currently, single cell proteomics, genomics and transcriptomics dominate the single cell technology landscape, while single cell metabolomics is still in its infancy. The discrepancy of single cell metabolomics is mainly due to the hardware of mass spectrometry being incompatible with single cell resolution. A mammalian cell contains roughly 1 pL of analytical volume. Due to this extremely small sample volume GC-MS methods often use around two million cells per analysis. To further add complexity, metabolism is highly dynamic – temporally and spatially – in tissue. Therefore, acquisition of single cells and their processing presents many challenges in the context of metabolic profiling.

Firstly, we will address the process of sampling single cells. The sorting of cells from a sample is essential to single cell analysis. Fluorescence-activated cell sorting (FACS) employs fluorescent labels to sort cells of specified origin. FACS may be coupled to mass spectrometry methods to assess the profiles of sorted cells. However, the FACS process may interfere with metabolic profiles and therefore FACS coupled with MS is more suited to proteomic analysis (Bandura et al. 2009). To minimise the sampling time methods such as matrix-assisted laser desorption/ionisation (MALDI) employ a laser to ablate cells and their metabolites in situ. Following laser ablation the ionised metabolites are analysed via MS. Advances towards single cell resolution have been reported. By using a combination of computational imaging techniques and nuclei staining, the points of laser ablation during MALDI-MS analysis can be inferred as metabolite acquisition from a single cell (Rappez et al. 2021). However, this method is not quantitative and is biased to high abundant metabolites and proteins.

Current technologies are aiming to sample single cells through microfluidic trapping coupled with acquisition of intracellular metabolites through micro capillaries. However, efforts to move to a microfluidic environment present their own complications, mainly due to maintaining homeostatic environment in a miniaturised platform (Ali et al. 2019).

In conclusion, as mass spectrometers become more sensitive coupled with inventive single cell trapping techniques, metabolomics will be a valid and useful addition to the single cell "OMICS" universe.

References

Ali A, Abouleila Y, Shimizu Y, Hiyama E, Emara S, Mashaghi A, Hankemeier T (2019) Single-cell metabolomics by mass spectrometry: advances, challenges, and future applications. TrAC Trends Anal Chem 120:115436., ISSN 0165-9936. https://doi.org/10.1016/J.TRAC.2019.02.033

Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, Pavlov S, Vorobiev S, Dick JE, Tanner SD (2009) Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem 81(16):6813–6822., ISSN 00032700. https://pubs.acs.org/doi/full/10.1021/ac901049w. https://doi.org/10.1021/AC901049W

Bayram S, Fürst S, Forbes M, Kempa S (2020) Analysing central metabolism in ultra-high resolution: at the crossroads of carbon and nitrogen. Mol Metab 33:38–47., ISSN 2212-8778. https://doi.org/10.1016/J.MOLMET.2019.12.002

Berndt N, Egners A, Mastrobuoni G, Vvedenskaya O, Fragoulis A, Dugourd A, Bulik S, Pietzke M, Bielow C, van Gassel R, Damink SW, Erdem M, Saez-Rodriguez J, Holzhütter HG, Kempa S, Cramer T (2020. ISSN 15321827. https://www.nature.com/articles/s41416-019-0659-3) Kinetic modelling of quantitative proteome data predicts metabolic reprogramming of liver cancer. Br J Cancer 122(2):233–244. https://doi.org/10.1038/s41416-019-0659-3

Boros LG, Puigjaner J, Cascante M, Lee WNP, Brandes JL, Bassilian S, Yusuf FI, Williams RD, Muscarella P, Melvin WS, Schirmer WJ (1997) Oxythiamine and dehydroepiandrosterone inhibit the nonoxidative synthesis of ribose and tumor cell proliferation. Cancer Res. ISSN 00085472

Bruntz RC, Lane AN, Higashi RM, Fan TW-M (2017) Exploring cancer metabolism using stable isotope-resolved metabolomics (SIRM). J Biol Chem 292(28):11601–11609., ISSN 0021-9258. http://www.jbc.org/article/S0021925820370289/fulltext. http://www.jbc.org/article/S0021925820370289/abstract. https://doi.org/10.1074/JBC.R117.776054

Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, DeBerardinis RJ, Feron O, Frezza C, Ghesquiere B, Gottlieb E, Hiller K, Jones RG, Kamphorst JJ, Kibbey RG, Kimmelman AC, Locasale JW, Lunt SY, Maddocks OD, Malloy C, Metallo CM, Meuillet EJ, Munger J, Nöh K, Rabinowitz JD, Ralser M, Sauer U, Stephanopoulos G, St-Pierre J, Tennant DA, Wittmann C, Heiden MGV, Vazquez A, Vousden K, Young JD, Zamboni N, Fendt SM (2015) A roadmap for interpreting 13C metabolite labeling patterns from cells. Curr Opin Biotechnol 34:189–201. ISSN 0958-1669. https://doi.org/10.1016/J.COPBIO.2015.02.003

Delp J, Gutbier S, Cerff M, Zasada C, Niedenführ S, Zhao L, Smirnova L, Hartung T, Borlinghaus H, Schreiber F, Bergemann J, Gätgens J, Beyss M, Azzouzi S, Waldmann T, Kempa S, Nöh K, Leist M (2018) Stage-specific metabolic features of differentiating neurons: implications for toxicant sensitivity. Toxicol Appl Pharmacol 354:64–80. ISSN 1096-0333. https://pubmed.ncbi.nlm.nih.gov/29278688/. https://doi.org/10.1016/J.TAAP.2017.12.013

Fan TW, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, Miller DM (2009) Altered regulation of metabolic pathways in human lung cancer discerned by 13C stable isotope-resolved metabolomics (SIRM). Mol Cancer 8(1):1–19. ISSN 14764598. https://molecular-cancer.biomedcentral.com/articles/10.1186/1476-4598-8-41. https://doi.org/10.1186/1476-4598-8-41/TABLES/6

Geisberger S, Bartolomaeus H, Neubert P, Willebrand R, Zasada C, Bartolomaeus T, McParland V, Swinnen D, Geuzens A, Maifeld A, Krampert L, Vogl M, Mähler A, Wilck N, Markó L, Tilic E,

Forslund SK, Binger KJ, Stegbauer J, Dechend R, Kleinewietfeld M, Jantsch J, Kempa S, Müller DN (2021. ISSN 1524-4539. https://pubmed.ncbi.nlm.nih.gov/33906377/) Salt transiently inhibits mitochondrial energetics in mononuclear phagocytes. Circulation 144(2): 144–158. https://doi.org/10.1161/CIRCULATIONAHA.120.052788

Han J, Danell R, Patel J, Gumerov D, Scarlett C, Speir J, Parker C, Rusyn I, Zeisel S, Borchers C (2008) Towards high-throughput metabolomics using ultrahigh-field Fourier transform ion cyclotron resonance mass spectrometry. Metabolomics 4(2):128–140. ISSN 1573-3882. https://pubmed.ncbi.nlm.nih.gov/19081807/. https://doi.org/10.1007/S11306-008-0104-8

Hevesy G, Hofer E (1934) Elimination of water from the human body. Nature 134(3397):879–879., ISSN 1476-4687. https://www.nature.com/articles/134879a0. https://doi.org/10.1038/134879a0

Jang C, Chen L, Rabinowitz JD (2018) Leading edge primer metabolomics and isotope tracing. Cell 173:822–837. https://doi.org/10.1016/j.cell.2018.03.055

Kelstrup C, Bekker-Jensen D, Arrey T, Hogrebe A, Harder A, Olsen J (2018) Performance evaluation of the Q exactive HF-X for shotgun proteomics. J Proteome Res 17(1):727–738. ISSN 1535-3907. https://pubmed.ncbi.nlm.nih.gov/29183128/. https://doi.org/10.1021/ACS.JPROTEOME.7B00602

Le A, Lane AN, Hamaker M, Bose S, Gouw A, Barbi J, Tsukamoto T, Rojas CJ, Slusher BS, Zhang H, Zimmerman LJ, Liebler DC, Slebos RJ, Lorkiewicz PK, Higashi RM, Fan TW, Dang CV (2012) Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in b cells. Cell Metab 15(1):110–121. ISSN 15504131. https://pubmed.ncbi.nlm.nih.gov/22225880/. https://doi.org/10.1016/j.cmet.2011.12.009

Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal Chem 78(7):2113–2120. https://doi.org/10.1021/AC0518811. https://pubs.acs.org

Marshall A, Hendrickson C (2008) High-resolution mass spectrometers. Annu Rev Anal Chem 1(1):579–599. https://doi.org/10.1146/annurev.anchem.1.031207.112945. https://www.annualreviews.org/

McNaught D, Wilkinson A (2008) Mass resolving power in mass spectrometry. IUPAC Compend Chem Terminol. https://doi.org/10.1351/GOLDBOOK.M03730

Myers W (1996) Georg charles de hevesy: the father of nuclear medicine. J Nucl Med Technol 24: 291–294

Park TJ, Reznick J, Peterson BL, Blass G, Omerbašić D, Bennett NC, Kuich PHJ, Zasada C, Browe BM, Hamann W, Applegate DT, Radke MH, Kosten T, Lutermann H, Gavaghan V, Eigenbrod O, Bégay V, Amoroso VG, Govind V, Minshall RD, Smith ESJ, Larson J, Gotthardt M, Kempa S, Lewin GR (2017) Fructose-driven glycolysis supports anoxia resistance in the naked mole-rat. Science 356(6335):307–311. ISSN 1095-9203. https://pubmed.ncbi.nlm.nih.gov/28428423/. https://doi.org/10.1126/SCIENCE.AAB3896

Pietzke M, Zasada C, Mudrich S, Kempa S (2014) Decoding the dynamics of cellular metabolism and the action of 3-bromopyruvate and 2-deoxyglucose using pulsed stable isotope-resolved metabolomics. Cancer Metab 2(1):9. https://doi.org/10.1186/2049-3002-2-9

Qiu Y-Q (2013) KEGG pathway database. In: Encyclopedia of systems biology. Springer, New York, pp 1068–1069. https://doi.org/10.1007/978-1-4419-9863-7_472

Rappez L, Stadler M, Triana S, Gathungu RM, Ovchinnikova K, Phapale P, Heikenwalder M, Alexandrov T (2021) SpaceM reveals metabolic states of single cells. Nat Methods 18(7): 799–805. ISSN 1548-7105. https://www.nature.com/articles/s41592-021-01198-0. https://doi.org/10.1038/s41592-021-01198-0

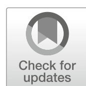
Rohwer N, Bindel F, Grimm C, Lin SJ, Wappler J, Klinger B, Blüthgen N, Du Bois I, Schmeck B, Lehrach H, De Graauw M, Goncalves E, Saez-Rodriguez J, Tan P, Grabsch HI, Prigione A, Kempa S, Cramer T (2016. ISSN 19492553. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4872743/) Annexin A1 sustains tumor metabolism and cellular proliferation upon stable loss of HIF1A. Oncotarget 7(6):6693. https://doi.org/10.18632/ONCOTARGET.6793

Sauer U (2006) Metabolic networks in motion: 13C-based flux analysis. Mol Syst Biol 2(1):62., ISSN 1744-4292. https://onlinelibrary.wiley.com. https://doi.org/10.1038/MSB4100109

Werner E, Heilier J, Ducruix C, Ezan E, Junot C, Tabet J (2008. ISSN 1570-0232. https://pubmed.ncbi.nlm.nih.gov/18672410/) Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. J Chromatogr B Analyt Technol Biomed Life Sci 871(2):143–163. https://doi.org/10.1016/J.JCHROMB.2008.07.004

Wilikinson D (2018) Historical and contemporary stable isotope tracer approaches to studying mammalian protein metabolism. Mass Spectrom Rev 37(1):57–80. ISSN 1098-2787. https://pubmed.ncbi.nlm.nih.gov/27182900/

Yang Y, Fan T, Lane A, Higashi RM (2017) Chloroformate derivatization for tracing the fate of amino acids in cells and tissues by multiple stable isotope resolved metabolomics (mSIRM). Anal Chim Acta 976:63–73. ISSN 1873-4324. https://pubmed.ncbi.nlm.nih.gov/28576319/. https://doi.org/10.1016/J.ACA.2017.04.014



Metabolomics in Cell Biology

Zuhal Eraslan, Marta Cascante, and Ulrich L. Günther

Contents

1 Introduction 182
1.1 Omics Context 182
1.2 History of Metabolomics Using Cellular Samples 183
2 Metabolomics Approaches: Targeted vs. Untargeted 184
2.1 Sample Types 186
2.2 Preparation of Cell Extracts for Metabolomics 186
2.3 Metabolomics Technologies 188
2.3.1 NMR-Based Methods 189
2.3.2 MS-Based Methods 191
3 Case Studies 193
3.1 Examples of Cancer Metabolomics 193
3.2 Real-Time Measurements of Cells 195
3.3 Metabolic Studies in Mycobacteria 197
3.4 Metabolic Fluxes by NMR and GC-MS 198
3.5 Examples for Fluxes in Cancer Cells by GC-MS 199
3.6 Fluxes and Metabolic Modelling 201
3.7 Metabolic Fluxes in Liver Disease 201

Z. Eraslan
Department of Dermatology, Weill Cornell Medicine, New York, NY, USA

M. Cascante
Department of Biochemistry and Molecular Biomedicine, University of Barcelona, Barcelona, Spain

Institute of Biomedicine of University of Barcelona (IBUB), University of Barcelona, Barcelona, Spain

CIBER of Hepatic and Digestive Diseases (CIBEREHD), Institute of Health Carlos III (ISCIII), Madrid, Spain

U. L. Günther (✉)
Institute of Chemistry and Metabolomics, University of Lübeck, Lübeck, Germany
e-mail: Ulrich.guenther@uni-luebeck.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_619

4 Conclusions 202
References 203

Abstract

Metabolomics has long been used in a biomedical context. The most typical samples are body fluids in which small molecules can be detected and quantified using technologies such as Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS). Many studies, in particular in the wider field of cancer research, are based on cellular models. Different cancer cells can have vastly different ways of regulating metabolism and responses to drug treatments depend on specific metabolic mechanisms which are often cell type specific. This has led to a series of publications using metabolomics to study metabolic mechanisms. Cell-based metabolomics has specific requirements and allows for interesting approaches where metabolism is followed in real-time. Here applications of metabolomics in cell biology have been reviewed, providing insight into specific technologies used and showing exemplary case studies with an emphasis towards applications which help to understand drug mechanisms.

Keywords

Cell biology · Metabolomics

1 Introduction

1.1 Omics Context

Metabolomics is the omics-science that analyses the small molecule compartment of biological samples. Originally, Nicholson coined the term metabonomics as the science that studies responses of living systems to metabolic changes (Nicholson et al. 1999). What distinguishes metabolomics from conventional analytical approaches is the way data are analysed, looking at many components of a sample at the same time in an untargeted manner, often in a high-through-put manner with subsequent statistical data analysis. Metabolomics is now often found in the context of other omics technologies, as part of multi-omics workflows, usually involving computational systems biology to integrate the different types of data. On the other hand, targeted analyses have been used to study mechanisms based on biological hypotheses in a targeted manner, often using isotopically labelled metabolic precursors as tracers.

Overall, metabolomics has become a broadly used technology in many fields of biomedical and nutritional sciences (Wishart 2016). Biomedical analyses are often based on body fluids with the goal to derive biomarkers. This often involves large sample numbers and high-throughput approaches and is increasingly linked to genetic and epigenetic features. Wishart, who has established the most

comprehensive repositories and online tools for metabolomics, has recently discussed the relevance of metabolomics to study disease mechanisms, to customise treatments and monitor treatment response and to identify new drugs (Wishart 2016). The expression phenotyping is often used, considering that metabolomics links genomics, epigenomics and transcriptomics with biological phenotypes.

1.2 History of Metabolomics Using Cellular Samples

Besides the use of metabolomics in a medical context, often with the intent to find biomarkers, there are an increasing number of applications in biological studies with the goal to identify specific mechanisms in cells. When looking at such applications of metabolomics in cellular systems the boundaries between targeted and non-targeted analysis are often more fluid. As a matter of fact, some of the earliest applications of what would now be called metabolomics were conducted in cellular systems. As early as 1978, Ugurbil and Shulman used NMR to study metabolic turnover in *Escherichia coli* cells (*E. coli*), first by ^{31}P-NMR and soon afterwards by ^{13}C-NMR. The first report used ^{31}P-NMR to measure glycolysis rates in *E. coli*, and also looked at the effect of ATPase inhibition (Ugurbil et al. 1978a). Even at a spectrometer with 360 MHz proton frequency, a 2 min time resolution was achieved. In a second publication, *E. coli* cells were incubated with [1-^{13}C]-labelled glucose in an NMR tube and followed the turnover of glucose via fructose bisphosphate into amino acids (alanine and valine) along with lactate, succinate and acetate using ^{13}C-NMR spectra (Ugurbil et al. 1978b). Upon oxygenation, the authors also observed glutamate, labelled at the C-4, C-3 and C-2 positions. Time-courses were reported over up to 3 h with a time resolution of 1 min, which is remarkable considering that the NMR instrumentation in 1978 was at 90 MHz proton frequency. This early work demonstrates the advantage of NMR to identify site-specific label incorporation. In a subsequent publication from Shulman's group, rat hepatocytes were used to quantitatively assess the metabolic flux of glycerol in gluconeogenesis, showing label incorporation in several glycolytic intermediates and products (Cohen et al. 1979) and to understand whether or to what extent hyperthyroid cells (after pre-treatment of rats with triiodothyronine hormone (T3)) increases the rate of glycerol consumption and glucose formation. In this work not only signal intensities were used but also scalar couplings between ^{13}C nuclei to quantitatively assess label incorporation. In hyperthyroid cells labelling was also reported in lactate, alanine, aspartate, glutamate and ketone bodies. By analysing time-courses of label incorporation in different positions of glucose from differently labelled glycerol precursors pentose phosphate pathway activity could be determined, along with activities of transaldolase and transketolase. These early findings which provide a historical background for cell-based metabolic studies are confirmed by more recent work taking similar approaches to measure fluxes by NMR (Jin et al. 2013).

Soon afterwards, Jardetzky and co-workers described a system for continuous-flow monitoring of metabolism in mammalian cells (Chinese hamster ovarian cells)

(Gonzalez-Mendez et al. 1982). These seminal experiments laid the ground for using NMR to look at metabolism at a time when nobody thought about metabolomics. They exemplify the virtues of NMR for examining metabolism in cellular systems, which include real-time analysis of intact cells, using isotopic labels as tracers and the ability to observe site-specific label incorporation in molecules. Although any analysis that uses labelled precursors is by definition not untargeted, NMR was always used to simultaneously detect many metabolites in one NMR spectrum.

Another milestone in this development has come from Szyperski et al. (Szyperski 1995; Szyperski et al. 1996) in 1995, using ^{1}H-^{13}C_HSQC spectra to analyse cellular label incorporation. Today HSQC spectra are often preferred as they offer higher sensitivity than^{13}C-observed spectra. Szyperski was able to make sense of highly complex J_{CC} couplings patterns which arise from mixtures of isotopomers. Today's most advanced NMR systems work at 1.2 GHz, commonly used metabolomics systems work at 600 MHz. Cryoprobes and other improvements of NMR technology have boosted the sensitivity of NMR experiments by at least an order of magnitude. There are now cryoprobes available that have been optimised for mass limited samples which provide immense sensitivity advantages for cell-based studies (Saborano et al. 2019). Microprobes hold great promise to work with very few cells (Finch et al. 2016).

Here we review approaches of metabolomics in cells and present the methods used, with a specific focus on advanced NMR methods. Exemplary case studies are presented which demonstrate how metabolomics can be used in cell-based investigations in the wider context of drug discovery.

2 Metabolomics Approaches: Targeted vs. Untargeted

Metabolomics can be carried out at different levels and using very different approaches (Fig. 1). Methods are typically classified as targeted or non-targeted. In early metabolomics non-targeted fingerprinting was commonly used where "features" or raw spectra were employed without specific assignments, often using multivariate statistical approaches to identify the most relevant features for given classes (Nicholson et al. 1999). The general advantage of fingerprinting is that data can be left unassigned. For NMR most initial work has used raw spectra for subsequent statistical, often multivariate analyses (Nicholson et al. 1999), although this approach is becoming increasingly uncommon as metabolites are now easily assigned in many types of samples, including cell extracts where it is extremely well-known which signals are unique identifiers for individual metabolites. In the case of GC-MS, the assignments of the signals are usually also well-known thus favouring targeted approaches. This has included tracer-based approaches for which GC-MS is equally well suited (Bruntz et al. 2017). Employing typical LC-MS methods, thousands of features can be detected which can either be used in univariate or multivariate statistical analyses in an untargeted manner. Alternatively in targeted approaches features are assigned and only fully assigned signals are used for further

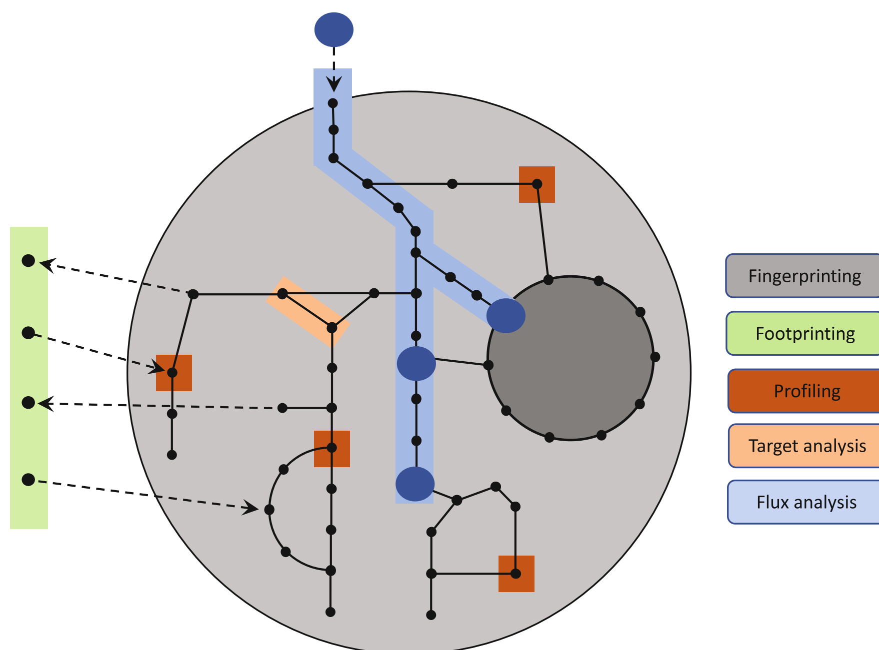


Fig. 1 Approaches towards metabolomics

analysis. In the special case of direct-injection MS, only a small percentage of signals can be uniquely assigned which makes the method better suitable for fingerprinting.

In today's applications of metabolomics in cell-based biological systems, targeted approaches have become much more common, although we also find high-throughput studies without assigning all signals. Whenever hypothesis-driven projects aim to decipher metabolic mechanisms in cells, targeted approaches are needed. This has often also included tracer-based metabolomics using ^{13}C- or ^{15}N-labelled metabolic precursors to study the fate of metabolites along often complex metabolic pathways. In the specific case of flux analysis, time-courses are used to derive kinetic data, or at least to determine the direction of metabolic fluxes (Kempa, "Advancements in pulsed stable isotope re-solved Metabolomics"). Commonly used isotopically labelled metabolic precursors have been glucose and glutamine, but there have also been reports using other metabolites (reviewed in (Saborano et al. 2019; Cascante and Marin 2008; Jang et al. 2018)). To a certain degree, it is also possible to obtain fluxes by looking at rates of consumption and production of metabolites in media. The expression "metabolic flux" is also often also used for modelling metabolic pathways using tracer-based data (Selivanov et al. 2006, 2020).

2.1 Sample Types

When working with cellular systems or organelles, there is a wide array of possible samples. Primary cultures are directly taken from organisms and can be subject to investigation for at least a short time. Cell lines arise from primary cells which have been immortalized. Cell cultures are available from public resources, such as ATCC (http://www.atcc.org), Coriell (http://ccr.coriell.org), DSMZ (http://www.dsmz.de), ECACC (https://www.culturecollections.org.uk), JCRB (https://cellbank.nibiohn.go.jp) or RIKEN Bioresource centre (https://web.brc.riken.jp/en), for which robust quality control protocols have been established (Yu et al. 2015). Most studies use immortalised cell lines which can be readily grown to obtain sufficient amounts of metabolites. *Cell extracts* represent the most common type of sample, but it is also possible and often reasonable to measure metabolite concentrations in the medium in which cells were grown, which can be used to calculate turnover for individual nutrients. Media-sampling can also be carried out in bioreactors where metabolic turnover can be directly detected. Moreover, NMR is also suitable for non-invasive analysis of metabolism in living cells and organelles which has led to several interesting studies focussed on specific metabolic mechanisms.

2.2 Preparation of Cell Extracts for Metabolomics

Several recent publications have reported protocols for the preparation of metabolomic samples from cells (Bhinderwala and Powers 2019; Halama 2014; Dietmair et al. 2012; Rais et al. 1999). A consensus protocol has been summarised in Fig. 2. Whether primary cells, or certain strains of cells or cell lines are used is less relevant for metabolomics sample preparation than the way the cells are grown. The protocol is different for cells grown in suspension or adherently on an inert surface (typically plastic or glass). Among mammalian cells, it is mainly haematological cancer cells which are grown in suspension. In the case of adherent cells it is useful to wash cells with ice cold buffer (usually PBS) as the first step of the preparation for metabolomics extraction. For suspension cells we wash with prewarmed PBS and then transfer cells into Eppendorf tubes. Suspension cells can be harvested by centrifugation or by filtration, in our lab we prefer centrifugation at 4°C as filters need to be washed to avoid contamination with small molecules from the filter. Cells grown adherently need to be removed from the surface on which they were grown. This can be achieved by scraping them off the surface or by trypsinization, where the latter is generally more common, as it is easier to be carried out reproducibly. Metabolism needs to be quenched by cooling cells rapidly, which is often carried out with precooled methanol and/or acetonitrile. Other authors prefer sonication and/or submerging cells in liquid nitrogen. Subsequently, chloroform can be added to obtain two phases. Often only the aqueous phase is used but it is also possible to analyse the lipids from the chloroform layer. Precipitated proteins are found at the interface between the two phases and the polar and non-polar extracts can readily be obtained with a syringe. We found it to be important to use glass vials from the first

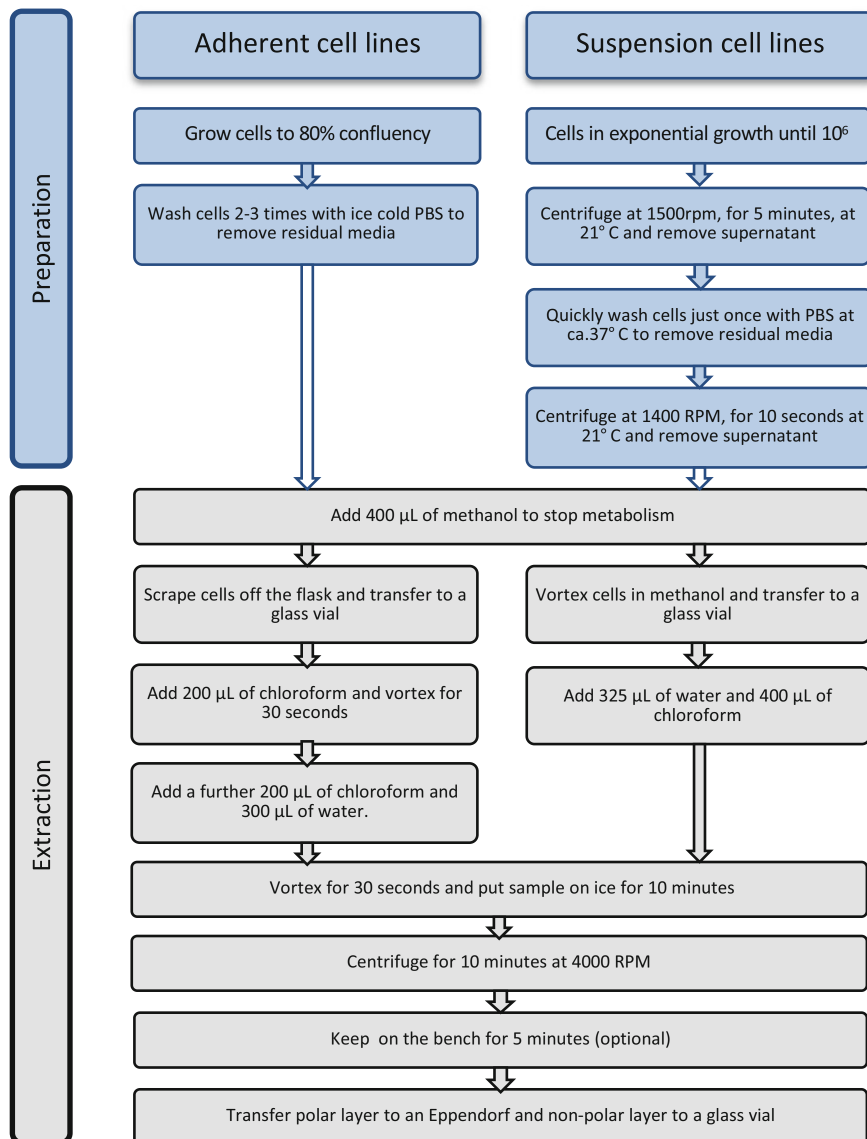


Fig. 2 Protocol for the preparation of metabolomics samples from cells

addition of methanol as tests with plastic vials showed compounds leaching from the plastic with methanol and chloroform. More recently plastic tubes have become available that are supposed to be chloroform proof.

Cell extracts recapitulate metabolic levels inside cells at one particular point of time. Such cell extracts can be examined by NMR or by MS. If NMR is used, a one-dimensional spectrum allows to reliably identify and quantify 30–40

metabolites using a 600 MHz spectrometer with a cryoprobe. The number of cells required for such measurements depends on the cell size. For reference, haematological cancer cells such as HL60 or K562, approximately 1–10 million cells are required for one-dimensional spectra, depending on the specific type of NMR experiments. For tracer-based flux experiments the number of cells required is larger and should be at least 10 million for ^{1}H-^{13}C-HSQC spectra or isotope-filtered spectra. For directly observed ^{13}C-spectra even higher concentrations of metabolites are required, although recently developed ^{13}C-optimised micro-cryoprobes offer huge sensitivity advantages with significant potential for metabolomics in cell-based systems (Ramaswamy et al. 2013; Clendinen et al. 2014; Thomas et al. 2021). Most MS studies of metabolic mechanisms used GC-MS which is somewhat more sensitive than NMR, >1 million cells are typically needed. LC-MS is much more sensitive and can work with very small numbers of cells, very much depending on the specific version of mass spectrometer that is used.

2.3 Metabolomics Technologies

As already introduced, the main technologies behind metabolomics are NMR and MS, each with specific advantages (Verpoorte, "Natural products drug discovery: on silica or in-silico?"; Evans, "Compound Identification Strategies in Mass Spectrometry-Based Metabolomics and Pharmacometabolomics"; Raftery, "Quantitative NMR Methods in Metabolomics; Wishart, "Practical Aspects of NMR-Based Metabolomics"; McKay, "Metabolomics using NMR- avoiding the "Black-Box"".). NMR methods are well established in drug discovery, mainly for studying the interaction of proteins with small molecules (Meyer and Peters 2003; Becker et al. 2018; Ludwig and Guenther 2009), and there is a range of methods to study cells. Table 1 lists the overall advantages of NMR and MS approaches along with specific sensitivities. For biologically driven users, it is particularly relevant to understand the level of application in a cellular context, translated into accessible metabolite concentrations, cell numbers and types of samples that can be studied. NMR requires millions of cells (assuming haematological cancer cells such as HL60 and K562 as a reference) whereas MS based methods are substantially more sensitive. NMR has a significant advantage for quantification and to detect small but relevant changes. Other advantages of NMR are the possibility to look a living cells non-invasively enabling the measurement of time-courses, and to identify site-specific label incorporation in tracer-based experiments. On the other hand, MS offers several orders higher sensitivity, enabling even single-cell metabolomics and providing access to a much wider range of metabolites.

MS can now be applied on a single-cell level, even combined with imaging (Rappez et al. 2021). Table 1 provides an overview over the different analytical methods that can be used providing information, their applicability, advantages and limitations. GC-MS has commonly preferred for tracer-based approaches and there are well-established protocols for this (Cascante and Marin 2008).

Table 1 Metabolomics methods used for cellular systems

Technology	^{1}H-NMR	GC-MS	LC-MS
Metabolites	Water-soluble metabolites; lipid fraction	Water-soluble metabolites which must be derivatized	Water-soluble metabolites; lipid fraction for lipidomics
Sample types	Whole cells, cell extracts, media, organoids	Cell extracts	Cell extracts
Number of cells[a]	Min 10 million	5–10 million	<1 million Single cell
Run-time	1D spectrum: 10 min	10 min	10–15 min
Detection limit	5–10 μM	<1 μM	pM–nM
Number of metabolites in cell-based samples	30–40	150	Thousands of features, ca. 250 assigned metabolites
	Excellent reproducibility; concentrations can be quantified	Good reproducibility; semi-quantitative	No simple quantification, except for BIOCRATES approach
	Simple sample preparation; HR-MAS can be used without preparation	Sample extraction and derivatization required	Sample extraction required

[a] Cell numbers for typical haematological cancer reference cells (HL60 or K562)

2.3.1 NMR-Based Methods

NMR in the context of metabolomics has been the subject of previous reviews which cover NMR methods and protocols used (Wishart 2016; Bhinderwala and Powers 2019; Halama 2014; Duarte et al. 2009; Powers 2009; ÄŒuperloviÄ 2010; Palmnas and Vogel 2013; Zhang et al. 2013; Markley et al. 2017; Vignoli et al. 2019). Usually, one-dimensional (1D) NMR spectra are used for metabolomics. For this, a pulse sequence called 1D-NOESY is most commonly employed, mainly because it allows for quantitative measurements with excellent suppression of the water signal. The 30–40 metabolites which can be quantified in 1D NMR spectra from cell extracts include amino acids, glutathione, taurine, several energy metabolites, such as UDP, ATP, UDP-glucose, lactate, pyruvate, succinate, fumarate and α-ketoglutarate, a range of sugars, myo-inositol, cholesterol, fatty acids (typically the CH, CH_2 and CH_3 groups), choline and phosphatidylcholine, phosphatidylethanolamine and triglycerides. While this list is short compared to what MS can observe, concentrations can be reproduced within 1–2% between repeats of cell cultures as illustrated by Tiziani et al. (2009). Raftery and co-workers have used a combination of 1D and 2D NMR methods to distinguish and quantify common phosphorylated coenzymes such as AMP, ADP and ATP, NADH, NADPH, NAD^+ and $NADP^+$ (Gowda et al. 2016).

1D-spectra can also be used for absolute quantification of metabolites and this in a highly reproducible manner. This can either be achieved by using an internal standard such as Trimethylsilylpropanoic acid (TMSP) or by using an electronic or software-generated reference signals using technologies which are now incorporated in most spectrometers such as ERETIC and PULCON (Akoka et al. 1999; Watanabe et al. 2016). These methods are now commonly used in NMR protocols for the analysis of blood samples (Dona et al. 2014).

NMR also offers a large number of alternative pulse sequences which each yield specific spectral features (reviewed in (Vignoli et al. 2019)). *J*-resolved (*J*-RES) spectra are frequently used, in particular in samples with significant overlap between signals. This is more common in plant-based samples than in samples arising from mammalian cell lines. For the assignment of *J*-RES spectra, the Birmingham Metabolite Library (Ludwig et al. 2011) provides a unique resource. *J*-RES spectra are usually processed in a manner where scalar couplings are removed thus providing considerably simplified spectra. *J*-RES is also a quantitative method suitable to calculate metabolite concentrations. Recently, a sophisticated ^{31}P-selective *J*-RES experiment has been used to identify several phosphometabolites from glycolysis (Cox et al. 2021). While this method has limited sensitivity, it can distinguish G6P/F6P or 2PG/3PG which is not possible using MS methods.

Many other NMR pulse sequences have been used to study metabolomics samples. The most important ones are the TOCSY experiment which links chemically connected protons and HSQC spectra which correlate ^{1}H and ^{13}C resonances. The two can also be combined in 2D-HSQC-TOCSY spectra which represent invaluable tools for the assignment of metabolites in metabolomics. Importantly, HSQC and HSQC-TOCSY spectra have been compiled in the COLMAR database which also offers online tools for metabolite identification (Robinette et al. 2008; Bingol et al. 2014, 2015a; Wang et al. 2020). HSQC-TOCSY spectra have also been used in conjunction with MS for the assignment of unknown metabolites in an approach that involves the simulation of NMR spectra using preliminary assignment from MS (Bingol et al. 2015b). Although generally important for metabolomics, this approach is less relevant for mammalian cells where the same 30–40 metabolites are observed in almost all cell extracts. Although TOCSY spectra cannot be used to derive absolute concentrations of metabolites, they have been used extensively for the analysis of tracer-based metabolism, mainly by Fan and Lane (2008). HSQC spectra offer considerable deconvolution of NMR spectra although at a much lower sensitivity than 1D spectra and with a loss of quantification. The reason for this lies in the nature of HSQCs which depend on scalar ^{1}H-^{13}C couplings ($^{1}J_{CH}$) which vary considerably between metabolites. As a consequence, the signal intensity of a HSQC spectrum depends not only on the concentration of a particular molecule but also on the size of the coupling constant. Wan et al. suggested to determine these coupling constants for a large number of metabolites and calculate the effect of $^{1}J_{CH}$ on the overall HSQC signal intensities (Wan et al. 2017), but this would require a significant effort of collecting such a spectral database.

Nevertheless, HSQC spectra have been used in metabolomics, predominantly in tracer-based metabolism (Saborano et al. 2019). For labelled samples, the effect of

the coupling constant on the HSQC intensity can be ignored when signals are compared for the same metabolite for samples grown under different conditions. In such cases, ratios between signals provide the relevant information on relative label incorporations. Alternatively specific filtering methods have been employed to quantify label incorporation. Such filters can be designed to become largely independent of the J_{CH} coupling constant (Reed et al. 2019).

A key advantage of NMR is that analyses are non-invasive, i.e. can be carried out using living cells. This has been exploited in various real-time approaches where metabolites have been measured over extended periods of time. Two main types of real-time experiments have been reported. High-Resolution Magic Angle Spinning (HR-MAS) stands for a technology which eliminates line broadening due to anisotropic interactions in samples (Raftery, "Quantitative NMR methods in metabolomics"). HR-MAS technologies for metabolomics have recently been reviewed in (Edison et al. 2021). HR-MAS requires almost no sample preparation and is non-destructive as long as cells can tolerate the high spinning forces. HR-MAS enables the measurement of metabolites in intact tissue samples, small animals (*C. elegans* (Blaise et al. 2007; Mobarhan et al. 2017) or drosophila (Sarou-Kanian et al. 2015)) or cells. The need for MAS arises from the limited mobility of molecules in semi-solid samples leading to broad lines in NMR spectra, either caused by anisotropic dipole–dipole interactions or by local variations in magnetic susceptibility. Both effects can be eliminated by rotating samples at the "magic angle" of 54.7° to the magnetic field. This requires placing the sample in a small rotor spinning at circa 6,000 Hz. For a 1.4 mm rotor at this spinning rate the sample experiences 200,000 g acceleration which induces serious stress to the sample (Edison et al. 2021). Recent developments may help to overcome this limitation, with rotor synchronised pulse sequences it has been possible to obtain good spectra of earth worms with as low as 100 Hz (Mobarhan et al. 2017). Micro-sized coils for 1 mm rotors were shown to further improve mass sensitivity (Lucas-Torres and Wong 2019). HR-MAS has been used, for example, to study bacterial cells (Righi et al. 2013), mycobacteria (Hanoulle et al. 2005a, 2006a; Lee et al. 2005) and also for examination of mammalian cells (Nyblom et al. 2008; Garcia-Alvarez et al. 2011; Gogiashvili et al. 2019; Judge et al. 2019; García-Álvarez et al. 2009; Vermathen et al. 2021) and recently a cellular fungus (Edison et al. 2021). Although such g-forces are seen as being too large for haematological cancer cells, Edison et al. demonstrated applicability to chronic lymphoid leukaemia cells (Edison et al. 2021). Recent works by Edison et al. were focussed towards real-time monitoring using HR-MAS (Edison et al. 2021). As an alternative approach, cells have been used in flow bioreactors (Hall et al. 2016) or have been embedded in matrices such as agarose for real-time measurements (Koczula et al. 2016; Alshamleh et al. 2020). Examples will be shown below.

2.3.2 MS-Based Methods

GC-MS has been used in the context of metabolomics for more than 50 years and protocols have been established which can identify and quantify more than 100 metabolites form cell cultures (reviewed in (Fiehn 2016)). First applications of

GC-MS were reported in the 1960s covering sugars, amino acids, sterols, hormones, hydroxyl acids, fatty acids, aromatics and many other small molecules (<650 Da), as reviewed in (Fiehn 2016), before the term metabolomics had been coined. GC-MS requires derivatization which has been established for a large range of small molecules and is also well suited for metabolic flux analysis. The method benefits greatly from large data bases (e.g. the NIST Mass Spectral Library (Babushok et al. 2007) and others reviewed in (Fiehn 2016)) and has played an important role in the study of cellular metabolism and metabolic fluxes (Fischer et al. 2004). Unlike NMR, GC-MS is not inherently quantitative but quantification is feasible using external or internal standards (Fiehn 2016).

Most modern applications rely on LC-MS which does not require any derivatization and is potentially truly global, covering thousands of features. LC-MS is a widely used technology which is now available in many bioanalytical laboratories. Unfortunately, there is much heterogeneity between LC-MS methods, different upfront chromatographic approaches, different ionisation sources and different MS hardware. The advantage of LC-MS for studying metabolomics in the context of drug discovery is well illustrated by a recent workflow which combined metabolomics, proteomics and transcriptomics datasets of 54 cancer cells to derive a map of metabolite–transcriptional regulator interactions which was also used to look at drug sensitivities (Ortmayr et al. 2019). Such studies typically involve a systems biology layer that is used to combine large amounts of data. It should be mentioned that direct flow-injection analysis time-of-flight mass spectrometry (FIA-TOFMS) shows a lot of potential, in particular because measurements are very fast, allowing for >5,000 samples to be run per day (Fuhrer et al. 2011) as illustrated by a recent study by Ortmayr et al. (2019) who screened 54 cancer cell lines linked with transcriptomic and proteomic data. This study reports 689 putative assignments based on HMDB (Ortmayr et al. 2019). Typically, 250–350 metabolites can be uniquely assigned using LC-MS, although not easily quantified.

For cell-based applications, the Biocrates kits approach has become popular as it offers excellent quantification by using isotopically labelled reference compounds (see (Illig et al. 2010) as an early example where this technology was used). BIOCRATES kits now cover a wide range of hydrophilic and lipophilic metabolites (Thompson et al. 2019) and offer a simple protocol for sample preparation.

Highly specialised MS metabolomics techniques have now reached a sensitivity where single cells become amenable to investigation, even combined with imaging (Rappez et al. 2021). This is particularly interesting to assess cancer cell heterogeneity and plasticity, manifested as metabolic heterogeneity. Several studies have attempted to characterise metabolic heterogeneity in cancers (Rappez et al. 2021; DeVilbiss et al. 2021; Lau et al. 2020), although this remains a challenging area of MS research.

3 Case Studies

3.1 Examples of Cancer Metabolomics

Tiziani and Lodi started to extract cells in 2009 to profile drug responses in Acute Myeloid Leukaemia (AML) cell lines through NMR spectroscopy (Tiziani et al. 2009). For this work, they established protocols for sample preparation and characterization which yield highly reproducible NMR spectra. PCA shows narrow clusters for HL-60, K562 and KG1a cell lines, perfectly separated from each other. Using these AML cell lines a combination of two repurposed drugs, bezafibrate (BEZ) and medroxyprogesterone acetate (MPA) was tested. This work has established the mechanism of action of these drugs to be in part related to the generation of reactive oxygen species (ROS) with a profound effect on cell viability. The mechanism behind this is that high ROS causes chemically driven decarboxylation of α-keto-acids, converting α-ketoglutarate into succinate, pyruvate to acetate and oxaloacetate into malonate, a mechanism that has recently also been observed in acute myeloid leukaemia cells interacting with stromal cells (Vilaplana-Lopera et al. 2021).

Tiziani et al. later developed this approach into a high-content drug screening tool (Tiziani et al. 2011). For this, he developed a new protocol which combined growing cells in 96-well plates with in situ lysing and quenching of metabolism via SDS addition and sonication. The cell extracts were subsequently analysed by *J*-RES-NMR spectra. Using this protocol, several inhibitors could be detected by monitoring changes of the lactate/pyruvate ratio. Validation in cell lines and in primary cancer cells was demonstrated. The key advantage of this approach is that it is not based on a single read-out but rather a high-content metabolic measurement reflecting increasingly well-understood metabolic changes in cells. Similarly, Lodi showed in a cellular model how a combinatorial treatment using phytochemical combinations affects prostate cancer cells (Lodi et al. 2017). In this work, flux measurements were used to monitor changes in glutamine metabolism in response to treatment.

Eraslan et al. have conducted metabolomic studies to reveal therapeutic metabolic targets for germinal centre-derived Burkitt lymphoma (BL) and Diffuse large B-cell lymphoma (DLBCL) by applying various 1D and 2D NMR techniques (Eraslan et al. 2021). A principal component analysis (PCA) that was performed on the 1D NMR spectra of the media showed a clear separation between the BL cells and DLBCL (Fig. 3a). The loading plot of the corresponding PCA analysis depicted that the separation of BL from DLBCL mostly derived from the difference in extracellular asparagine level of BL and DLBL (Fig. 3b, c). ^{13}C-tracer-based NMR metabolic analysis that was carried out with cell extracts derived from BL and DLBCL cells cultured in a medium with or no asparagine containing ^{13}C-stable-isotope labelled tracers depicted that asparagine regulates the synthesis of serine from glucose (Fig. 3d) and serine uptake (Fig. 3e) in both BL and DLBCL cells. By combining a metabolomic platform with a transcriptomic platform, they uncovered that BL cells express the genes involved in serine biosynthesis at a higher level than DLBCL cells do (Fig. 3f) and defined a new treatment model which solely works for BL. They

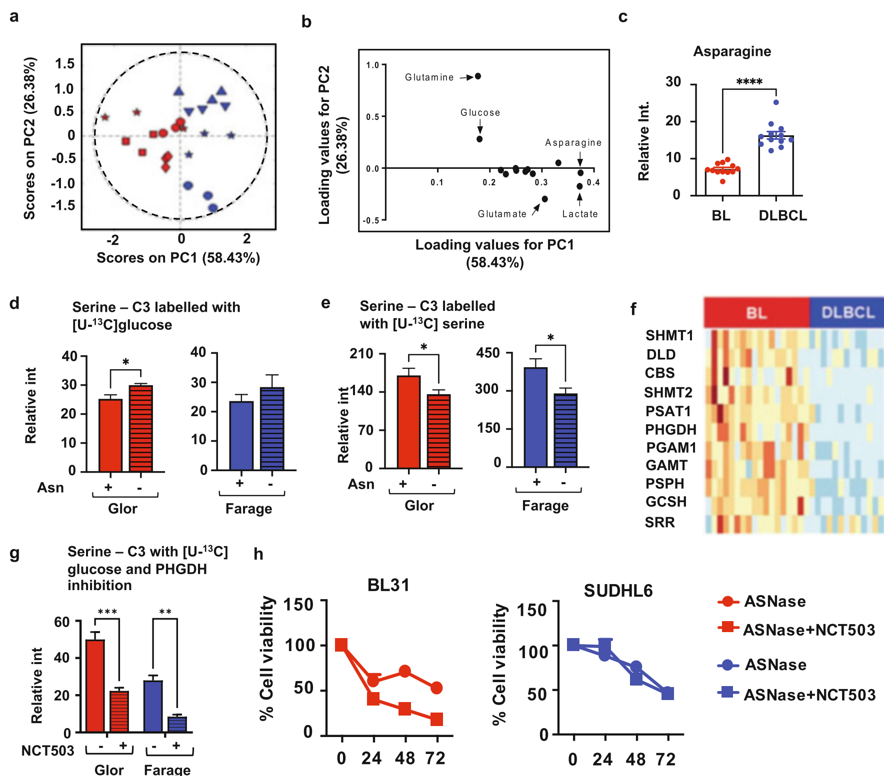


Fig. 3 The role of extracellular asparagine in regulation of serine metabolism in BL and DLBCL. (**a**) Principal Component Analysis (PCA) for 1D ^{1}H-NMR for media samples of Burkitt lymphoma cell lines (red) and diffuse large B-cell lymphoma cell lines (blue) (different symbols for different types of cells). (**b**) Representation of corresponding loadings plot showing metabolites which mostly contributed to separation of BL cells from DLBCL cells. (**c**) Relative 1D ^{1}H-NMR peak intensities for asparagine in the growth media of DLBCL and BL cells. (**d**) 2D ^{13}C-NMR analysis of the cell extraction of Glor and Farage cells cultured in a medium with or without asparagine cultured with [U-^{13}C]-glucose. (**e**) 2D ^{13}C-NMR analysis of the cell extraction of Glor and Farage cells cultured in a medium with or without asparagine cultured with [U-^{13}C]-serine. (**f**) Heatmap of statistically significant altered genes (FDR <0.01) associated with serine metabolism from differential expression analysis. (**g**) Analysis of 2D ^{13}C-NMR spectra of serine extracted from Glor and Farage cells cultured in a medium cultured [U-^{13}C]-glucose treated with the PHGDH inhibitor NCT-503 at 10 μM. (**h**) The viability of BL31 (BL cell line) and SUDHL6 (DLBCL cell line) cells after treatment with ASNase at 0.1 U/ml and ASNase at 0.1 U/ml plus NCT-503 for 24, 48 and 72 h

firstly performed a metabolomic study to assess the metabolic effect of an inhibitor (NCT503) that targets a rate-limiting enzyme PHGDH, in the serine biosynthesis pathway utilising a ^{13}C-tracer-based NMR metabolic approach (Fig. 3g). Then, they combined asparaginase (ASNase), which is used to treat acute lymphoblastic leukaemia (ALL) patients, at a very low dose with the PHGDH inhibitor. The combination of ASNase with the PHGDH inhibitor had a synergistic effect on cell viability

in BL cells while no synergistic effect on the viability of DLBCL cells (Fig. 3h). Thus, the integrated multi-omic approach with metabolomics and transcriptomic data has suggested an attractive new treatment model for BL.

To the best of our knowledge, the first application of HR-MAS NMR spectroscopy to study cells is a publication by Nyblom et al. studying fatty acid formation from [1-^{13}C]glucose in insulinoma cell lines (Nyblom et al. 2008)the effect of an antimitotic glycoside in glioma cells (García-Álvarez et al. 2009). Righi has demonstrated the applicability of HR-MAS for *Pseudomonas aeruginosa*, a pathogenic gram-negative bacterium (Righi et al. 2013). 25 metabolites could be assigned in 1D and 2D spectra. In an elegant experiment, Hanoulle et al. showed the activation of the anti-tuberculosis prodrug ethionamide in mycobacteria (Hanoulle et al. 2005b, 2006b). Judge has demonstrated the use of HR-MAS for human multiple myeloma cells showing that spectra can be obtained in 4 min thus allowing to measure time-courses for continuous in vivo monitoring of metabolism. In multiple myeloma cells, employing ^{13}C-labelled α-keto-isovalerate, conversion into valine could be monitored over a time-course of 60 min (Fig. 4) (Judge et al. 2018). The production of branched chain amino acids is an essential process for multiple myeloma cells which have a high demand for amino acids needed for protein production. For multicellular *Neurospora crassa* fungal cells, time-courses for central carbon metabolism, amino acids, TCA cycle intermediates, energy storage molecules and lipid and cell wall precursors were monitored for up to 10 h (Judge et al. 2018).

Vermathen et al. have used non-small cell lung cancer (NSCLC) cells to study Cisplatin (cisPt)-resistance metabolic adaptations employing HR-MAS (Vermathen et al. 2021). High-quality spectra allowed the detection of 53 metabolites. PCA showed close clustering of replicates and a clear separation with increasing resistance against cisPt. De-induced cells showed similar behaviour as cisPt-resistant cells, indicating a long-term memory after cisPt treatment. Metabolites predominantly changed in cisPt-resistant cells and their de-induced counterparts include glutathione and taurine (Vermathen et al. 2021). This study clearly demonstrates how HR-MAS NMR can be used to explain metabolic adaptations during drug resistance.

3.2 Real-Time Measurements of Cells

Koczula et al. described experiments where chronic lymphoid leukaemia (CLL) cells were embedded in a low-density agarose matrix to monitor metabolism over several hours yielding time-courses as shown in Fig. 5 (Koczula et al. 2016). The agarose matrix first helped to maintain cells afloat in a homogenous matrix and enable measurements with small numbers of cells (0.5–1.0 million) which would otherwise only have covered 1–2 mm at the bottom of the NMR tube. The concentration of cells in this experiment determines the overall turnover rate. Limited diffusion of the matrix caused oxygen depletion after circa 2 h which causes an abrupt increase of lactate, alanine and glutamate production along with changes in the rate of glutamine

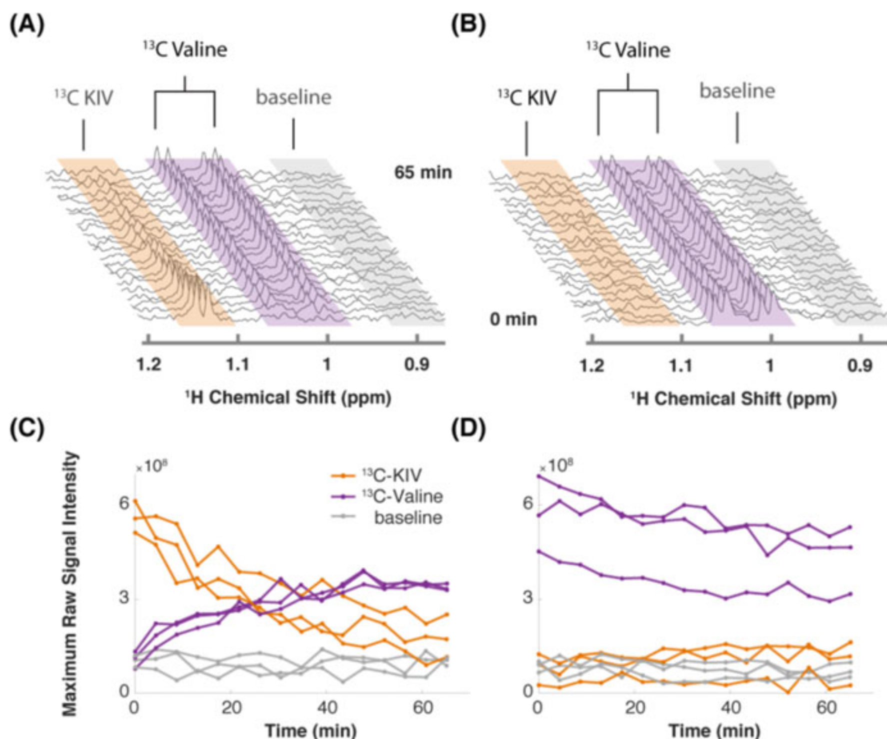


Fig. 4 Targeted isotopic CIVM-NMR measurements of metabolic flux in human leukaemia cells. (**a**) ^{13}C-labelled keto-isovalerate (KIV) was converted to valine. (**b**) ^{13}C-labelled valine was not converted to KIV, confirming unidirectional flux in ML cells. (**c**, **d**) Relative concentrations over time of ^{13}C-labelled KIV (orange) and 13 C-labelled valine (purple) compared to baseline noise (grey), from spectral intensities within each region of the representative experiments in (**a**, **b**), respectively, for 3 independent replicates. Reproduced from (Judge et al. 2018)

consumption (Koczula et al. 2016). By placing cells back in normoxic conditions before repeating the real-time NMR experiment, reversible metabolic adaptation of quiescent CLL cells to hypoxic conditions was demonstrated. By using chetomin as a HIF-1α inhibitor, it could be shown that this process of metabolic adaptation to hypoxia is HIF-1α dependent.

One advantage of this approach is that small numbers of cells are sufficient, the number of cells merely determines the overall rate of metabolic turnover. A major disadvantage of this approach is however that only the extracellular metabolome is observed because signals arising from intra-cellular metabolites are too weak and T_2-broadened to be observed. Alshamleh et al. designed a similar experiment to follow time-courses of metabolism using primary acute myeloid leukaemia (AML) cells and used methylcellulose as a more cell-friendly matrix, not affecting ATP levels as reported for agarose (Alshamleh et al. 2020).

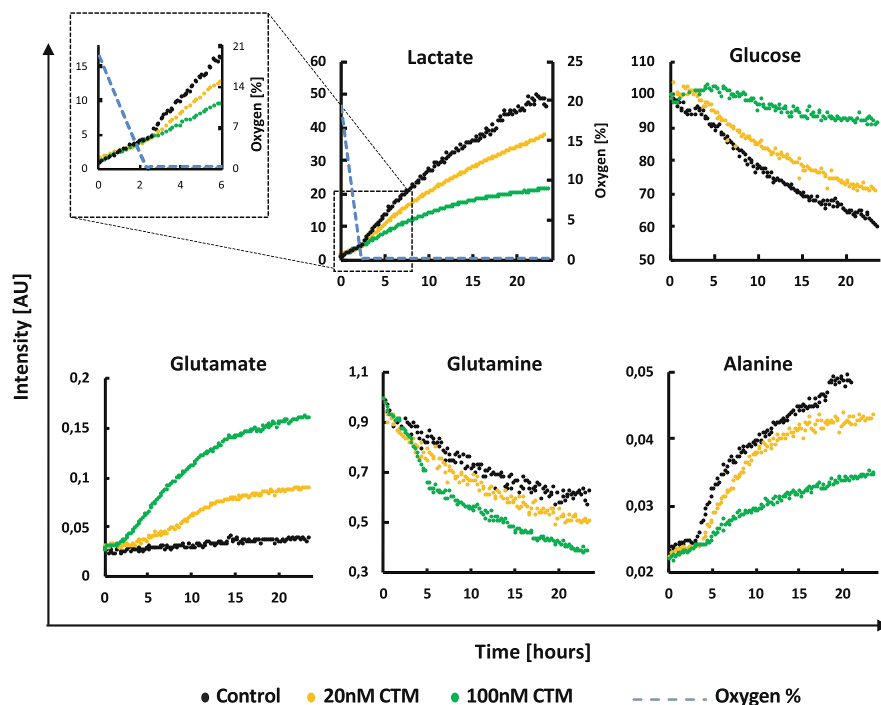


Fig. 5 Metabolic adaptation of CLL cells to hypoxia involves HIF-1α. Representative NMR time-course data for a CLL pre-treated for 3 h with either 0, 20 or 100 nM chetomin, before transferring into NMR for a further 24 h. Dashed lines on the lactate graph show oxygen levels inside the NMR tube. The top left panel shows an expanded view of lactate kinetics during the first 6 h with a visible shift after oxygen depletion which is inhibited by CTM. Data shown are representative of a minimum of N = 3 CLL samples. Reproduced from (Koczula et al. 2016)

3.3 Metabolic Studies in Mycobacteria

The applicability of HR-MAS for bacterial cells has been demonstrated by Li using *M. smegmatis* whole cells (Li 2006). Two early studies were focussed on mycobacteria where drug metabolites were monitored using HR-MAS. Ethionamide, the drug used in these studies needs to be activated by the monooxygenase EthA. HR-MAS spectra show nicely how this activation occurs when EthA is present (Hanoulle et al. 2005b, 2006b).

3D-cell cultures add a level of sophistication by allowing inter-cellular interactions which are known to be relevant in particular in cancer cells (Knitsch et al. 2021). Recently, Knitsch et al. (2021) designed a particularly interesting NMR experiment. Their experimental design allowed to acquire spatially resolved spectra from a 700 μm spheroid over a time-course of 112 h (Fig. 6). Profiles show spatially resolved depletion of, for example, glucose with concomitant build-up of lactate. Such 3D-culture systems are highly promising as they overcome main limitations of

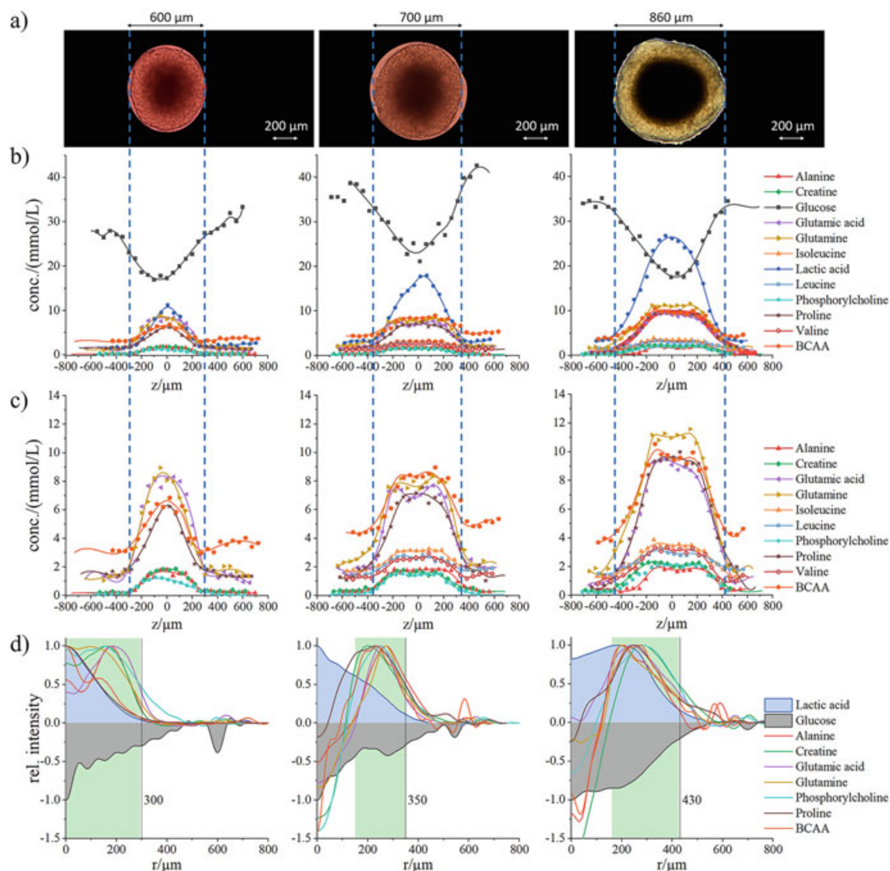


Fig. 6 (**a**) Light microscopy images of three spheroids aligned to the spatially resolved intensity profiles of NMR detected metabolites shown in (**b**) and enlarged (y-axis) in (**c**). Reproduced with permission from (Knitsch et al. 2021)

cell-based models. NMR is very well suited to study such systems in real time as illustrated and spacially resolved across the speroid.

3.4 Metabolic Fluxes by NMR and GC-MS

Both NMR and GC-MS have been used to study metabolic fluxes. Cascante has established GC-MS based fluxomics using mainly GC-MS, first by establishing the theoretically possible distribution of isotopes arising from selectively labelled glucose, comparing [1-^{13}C]glucose or [1,2-^{13}C]glucose (Cascante and Marin 2008). One of the key discoveries arising from this work has been the possibility to assess pentose phosphate pathway activity from the label distribution in pyruvate and

lactate (Boros et al. 1998). Chong et al. have shown how the complementary nature of GC-MS and NMR can be utilised to experimentally elucidate metabolic fluxes in mammalian cells (Chong et al. 2017). While GC-MS yields mass increments for each ^{13}C-atom that is incorporated into a metabolite, without positional information, NMR offers exactly the opposite, site-specific information of label incorporation. For this, high-resolution HSQC NMR spectra were used, which resolve adjacent ^{13}C-^{13}C couplings. These couplings can become rather complex in mixtures of metabolites with different label incorporations. An algorithm termed Combined Analysis of NMR and MS Spectra (CANMS) has been described to simulate such line shapes and use complementary GC-MS data to determine the exact label incorporation for different atoms in glycolytic or TCA cycle metabolites. Examples are shown for lactate and glutamate, for [1,2-^{13}C]glucose as a metabolic precursor. This approach is somewhat cumbersome but yields highly accurate site-specific label incorporation.

3.5 Examples for Fluxes in Cancer Cells by GC-MS

MYC addition has long been understood to be major driver for many cancers, associated with a strong glycolytic metabolism. Tarrado-Castellarnau et al. have used a tracer-based GC-MS and transcriptomics analysis to study metabolic consequences arising from the inhibition of CDK4/6 (Tarrado-Castellarnau et al. 2017). Cyclin-dependent kinases (CDK) are rational cancer therapeutic targets owing to their essential role in the regulation of cell cycle progression at the G1 restriction point. Upon inhibition of CDK4/6, an accumulation of MYC was observed. As a consequence, glutamine metabolism is upregulated, the mTOR pathway gets activated while HIF1α-mediated responses to hypoxia get disturbed. These MYC-driven adaptations to CDK4/6 inhibition render cancer cells highly sensitive to inhibitors of MYC, glutaminase or mTOR and to hypoxia. This targeted analysis was used to explore one avenue of a synergistic strategy by also inhibiting GLS1 demonstrating strong antiproliferative effects. This work exemplifies how metabolomics can be efficiently used in drug discovery to identify vulnerabilities to develop new synergistic treatment regimes.

Moonira et al. used GC-MS metabolomics to study the mechanism of the anti-diabetic drug metformin in hepatocytes (Moonira et al. 2020). They next used [1,2-$^{13}C_2$]glucose to measure partitioning of flux between glycolysis (via PFK1) and the pentose pathway by looking at mass isotopologues of lactate. Metabolism of [1,2-$^{13}C_2$]glucose by the pentose pathway generates m1 and m0 lactate, whereas glycolysis generates m2 and m0 lactate. Metformin increased the fraction of lactate derived from glucose and increased the m2/m1 ratio. An AMPK activator decreased the m2/m1 ratio. Cumulatively, metformin stimulates glycolysis via PFK1 in absolute terms and relative to flux via the pentose pathway, whereas the AMPK activator had the converse effect on the m2/m1 ratio. Metformin lowered cell G6P but not ATP and decreased G6pc mRNA at high glucose. In contrast, direct allosteric activators of AMPK were shown to have opposite effects compared to metformin

on glycolysis, gluconeogenesis and cell G6P. Overall this study showed how metformin acts in hepatocytes by lowering G6P when challenged with glucose by stimulation of glycolysis by an AMP-activated protein kinase-independent mechanism through changes in allosteric effectors of phosphofructokinase-1 and fructose bisphosphatase-1.

Selivanov has recently developed computational workflows based on their previously published software ISODYN (Selivanov et al. 2006) for the estimation of quantitative dynamic flux maps using stable-isotope resolved metabolomics (SIRM) (Selivanov et al. 2020) involving the extraction of mass spectra for metabolites of interest, correction of the spectra for natural isotope abundance and the evaluation of fluxes by simulation of mass spectra using a kinetic model based on ordinary differential equations (ODEs) for isotopomers of metabolites. The biochemical network is considered to derive a dynamic flux map. Foguet et al. has also extended a ^{13}C MFA-approach to seamlessly incorporate gene expression with metabolic flux data (Foguet et al. 2019).

Metabolic flux maps used for ^{13}C-MFA typically include less than 10% of the reactions contained in human cells Genome-Scale Metabolic Models (GSMMs) used to predict cell-level metabolic flux maps using human metabolic network reconstructions and constraint-based methods (Nilsson and Nielsen 2017). In recent years, it has been shown that GSMMs provide an advantageous platform for the integration of transcriptomics and metabolomics data to study the mechanisms underlying cell metabolic reprogramming and to identify new drug targets for cancer therapy (Karakitsou et al. 2021). Recently, Cascante's team developed genome-scale metabolic flux maps for colon cancer cell lines integrating multi-omics data on RECON2 human metabolic network reconstruction (Tarragó-Celada et al. 2021). In brief, transcriptomic data, growth rates, respiration parameters, rates of metabolite uptake and secretion, ^{13}C resolved metabolomics and targeted metabolomics have been used to parametrize the GSMMs. From the inferred metabolic flux maps, cystine transport xCT system and methylenetetrahydrofolate dehydrogenase 1 (MTHFD1) have been identified and validated as new potential targets for metastatic colorectal cancer.

Two studies from recent years exemplify how large-scale metabolomics screening can be used to map the larger landscape of cancer cell line metabolism. Li et al. have profiled 225 metabolites in 928 cell lines from more than 20 cancer types from the Cancer Cell Line Encyclopedia (CCLE) using LC-MS. Interrogation of metabolite/gene associations reiterated the well-known link between IDH1 mutations and the onco-metabolite 2-hydrogxyglutarate, which was also found in renal cell carcinoma cell lines that did not have a mutated IDH1. Interestingly, by combining a CRISPR–Cas9 knockout dataset with the Dependency Map (DepMap) exhibited that the reduced asparagine levels in cells are associated with hypermethylation of the promoter region of asparagine synthetase (ASNS). Further experiments showed that asparagine starvation depletes cells with lowered ASNS expression. This is in agreement with the targeted specific results of Eraslan et al. (2021) who showed that some Burkitt's lymphoma cell lines depend on asparagine for serine synthesis (needed for one-carbon metabolism). Both studies allow the conclusion that

asparaginase should be pursued as a cancer therapeutic beyond its current use in treating acute lymphoblastic leukaemia (ALL).

3.6 Fluxes and Metabolic Modelling

A second excellent example for high-throughput screening of cell lines used time-of-flight mass spectrometry (FIA-TOFMS) to obtain metabolic profiles of 54 cancer cell lines which were correlated with transcriptomic and proteomics data highlighting the role of transcription regulators (Ortmayr et al. 2019). FIA-TOFMS is by far the fastest metabolomics method which requires only 1 min per sample (Fuhrer et al. 2011). The disadvantage is that it is somewhat limited to fingerprinting, as it can yield multiple putative assignments for each feature. This particular study has used HMDB (Wishart et al. 2009) and RECON2 (Thiele et al. 2013) for the annotation of metabolites. The authors constructed a global network model incorporating the transcriptome, the proteome and the metabolome to predict regulatory associations with central metabolic pathways. One key finding was a global coordination between glucose and one-carbon metabolism, predicting a selective sensitivity to antifolate drugs in cell lines with low glucose uptake as a potential diagnostic marker for cancer cells that are more likely to respond to folate synthesis inhibitors.

Such high-throughput metabolomics approaches generally require a high level of computational support considering that the identification of metabolites is difficult.

3.7 Metabolic Fluxes in Liver Disease

Most cell-based metabolomics is currently focussed on cancer metabolism. Key mechanisms of cancer metabolomics along with a detailed list of drugs or drug candidates that target metabolism were recently published by Schmidt et al. (2021) One reason why cancer has been a major focus of metabolomics lies in the mere nature of cancer in which numerous tumour-associated metabolic alterations accumulate (Hanahan and Weinberg 2011). Most other disease models cannot be studied on a cellular basis but require at least tissue of organoids as models. For example, to study the action of ketohexokinase (KHK) inhibitors to improve steatosis, fibrosis and inflammation in the context of non-alcoholic fatty liver disease (NAFLD), human co-cultures are required (Shepherd 2021). While we struggled to use those for metabolomics studies, liver perfusion with [U-^{13}C]fructose worked well and allowed to obtain high-quality HSQC spectra (Fig. 7) which helped to decipher the mode of action of a KHK inhibitor showing reduction in fructose derived glycerol. These spectra illustrate the level of spectral quality that can be obtained using a micro-cryoprobe and HSQC spectra measured using non-uniform sampling with a 25% schedule. A time sequence of HSQC spectra of control and KHK inhibited samples confirmed uptake of labelled fructose for the control and

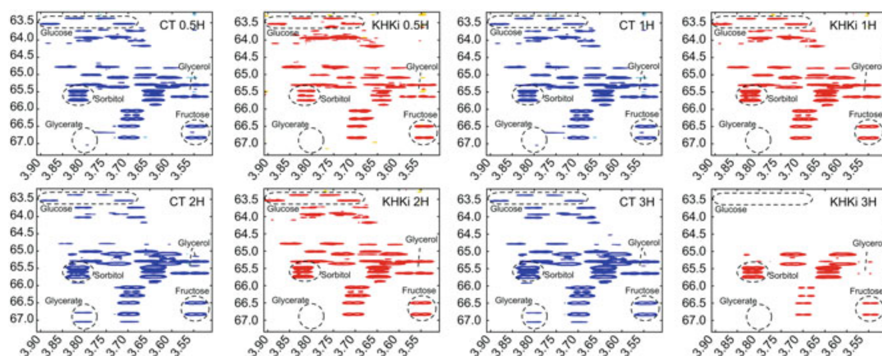


Fig. 7 NMR analysis of perfused human livers exposed to stable isotope-labelled fructose confirms ketohexokinase (KHK) inhibition reduces lipogenesis and glycolysis. ^{1}H-^{13}C-HSQC spectra from matched controls (CT: blue) and KHK inhibited (KHKi red) donor liver samples exposed to labelled fructose for indicated periods (0.5–3 h). Reproduced with permission from (Shepherd 2021)

inhibitor-treated livers, at a similar rate and accumulated over time. Gradual conversion of fructose to glycerol and glycerate was observed in control livers, as was conversion to labelled sorbitol. Although labelled sorbitol appeared in the KHK inhibited samples, there was a distinct lack of labelled glycerol and glycerate (Fig. 7) (Shepherd 2021).

4 Conclusions

Metabolomics has become an essential element in multi-omics studies, mainly because it builds a bridge between genetics and the phenotypical end point of biological systems. Cell lines remain a gold standard for many preclinical investigations despite obvious limitations. In many ways, cell lines are essential to investigate specific genetic features, for example in cancer. Metabolomics incorporates a wide range of methods, predominantly using NMR and mass spectrometry. We have shown several examples of high-throughput investigations that are probably only feasible using fast LC-MS methods. However, the lions' share of studies at this point focus on specific targets of mechanisms where targeted approaches are needed. To understand metabolic alterations in cancer cells, tracer-based analyses employing GC-MS and NMR metabolic data have played an important role. Moreover, NMR offers a unique angle to study metabolism in living cells using different types of real-time experiments. Such methods will likely be further developed to work under conditions better suitable to mimic physiological conditions. Overall, the methodologies reviewed here open new avenues for drug discovery by providing multi-parametric read-outs for cell-based experiments. The results of many targeted metabolic studies lend themselves to be integrated into now well-established Genome-Scale Metabolic Models to further predict drug targets.

Acknowledgments U. Günther and M. Cascante acknowledge the funding from the European Commission (HaemMetabolome [EC-675790]). M. Cascante acknowledges grant PID2020-115051RB-I00 funded by MCIN/AEI/10.13039/501100011033, AGAUR-Generalitat de Catalunya (2017SGR1033) and ICREA foundation (ICREA-Academia-2021).

References

Akoka S, Barantin L, Trierweiler M (1999) Concentration measurement by proton NMR using the ERETIC method. Anal Chem 71:2554–2557

Alshamleh I et al (2020) Real-time NMR spectroscopy for studying metabolism. Angew Chem Int Ed 59:2304–2308

ÄŒuperloviÄ M (2010) Cell culture metabolomics: applications and future directions. Drug Discov Today 15:12

Babushok VI et al (2007) Development of a database of gas chromatographic retention properties of organic compounds. J Chromatogr A 1157:414–421

Becker W, Bhattiprolu KC, Gubensäk N, Zangger K (2018) Investigating protein-ligand interactions by solution nuclear magnetic resonance spectroscopy. ChemPhysChem 19:895–906

Bhinderwala F, Powers R (2019) NMR metabolomics protocols for drug discovery. In: Gowda GAN, Raftery D (eds) NMR-based metabolomics, vol 2037. Springer, New York, pp 265–311

Bingol K, Bruschweiler-Li L, Li D-W, Brüschweiler R (2014) Customized metabolomics database for the analysis of NMR ^{1}H–^{1}H-TOCSY and ^{13}C–^{1}H-HSQC-TOCSY spectra of complex mixtures. Anal Chem 86:5494–5501

Bingol K et al (2015a) Unified and isomer-specific NMR metabolomics database for the accurate analysis of ^{13} C– ^{1} H HSQC spectra. ACS Chem Biol 10:452–459

Bingol K et al (2015b) Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid identification of new metabolites in complex mixtures. Anal Chem 87: 3864–3870

Blaise BJ et al (2007) Metabotyping of *Caenorhabditis elegans* reveals latent phenotypes. Proc Natl Acad Sci U S A 104:19808–19812

Boros LG et al (1998) Nonoxidative pentose phosphate pathways and their direct role in ribose synthesis in tumors: is cancer a disease of cellular glucose metabolism? Med Hypotheses 50:55–59

Bruntz RC, Lane AN, Higashi RM, Fan TW-M (2017) Exploring cancer metabolism using stable isotope-resolved metabolomics (SIRM). J Biol Chem 292:11601–11609

Cascante M, Marin S (2008) Metabolomics and fluxomics approaches. Essays Biochem 45:67–82

Chong M et al (2017) Combined analysis of NMR and MS spectra (CANMS). Angew Chem Int Ed 56:4140–4144

Clendinen CS et al (2014) ^{13}C NMR metabolomics: applications at natural abundance. Anal Chem 86:9242–9250

Cohen SM, Ogawa S, Shulman RG (1979) 13C NMR studies of gluconeogenesis in rat liver cells: utilization of labeled glycerol by cells from euthyroid and hyperthyroid rats. Proc Natl Acad Sci 76:1603–1607

Cox N, Millard P, Charlier C, Lippens G (2021) Improved NMR detection of phospho-metabolites in a complex mixture. Anal Chem 93:4818–4824

DeVilbiss AW et al (2021) Metabolomic profiling of rare cell populations isolated by flow cytometry from tissues. Elife 10:e61980

Dietmair S et al (2012) Metabolite profiling of CHO cells with different growth characteristics. Biotechnol Bioeng 109:1404–1414

Dona AC et al (2014) Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem 86:9887–9894

Duarte IF, Lamego I, Rocha C, Gil AM (2009) NMR metabonomics for mammalian cell metabolism studies. Bioanalysis 1:1597–1614
Edison AS et al (2021) NMR: unique strengths that enhance modern metabolomics research. Anal Chem 93:478–499
Eraslan Z, Papatzikas G, Cazier J-B, Khanim FL, Günther UL (2021) Targeting asparagine and serine metabolism in germinal centre-derived B cells non-Hodgkin lymphomas (B-NHL). Cell 10:2589
Fan TW-M, Lane AN (2008) Structure-based profiling of metabolites and isotopomers by NMR. Prog Nucl Magn Reson Spectrosc 52:69–117
Fiehn O (2016) Metabolomics by gas chromatography–mass spectrometry: combined targeted and untargeted profiling. Curr Protoc Mol Biol 114:30.4.1
Finch G, Yilmaz A, Utz M (2016) An optimised detector for in-situ high-resolution NMR in microfluidic devices. J Magn Reson 262:73–80
Fischer E, Zamboni N, Sauer U (2004) High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived 13C constraints. Anal Biochem 325:308–316
Foguet C, Jayaraman A, Marin S, Selivanov VA, Moreno P, Messeguer R, de Atauri P, Cascante M (2019) p13CMFA: Parsimonious 13C metabolic flux analysis. PLoS Comput Biol 15(9): e1007310
Fuhrer T, Heer D, Begemann B, Zamboni N (2011) High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection–time-of-flight mass spectrometry. Anal Chem 83: 7074–7080
García-Álvarez I, Garrido L, Doncel-Pérez E, Nieto-Sampedro M, Fernández-Mayoralas A (2009) Detection of metabolite changes in C6 glioma cells cultured with antimitotic Oleyl glycoside by ^{1} H MAS NMR. J Med Chem 52:1263–1267
Garcia-Alvarez I, Fernandez-Mayoralas A, Garrido L (2011) Effect of drugs in cells and tissues by NMR spectroscopy. Curr Top Med Chem 11:27–42
Gogiashvili M et al (2019) HR-MAS NMR based quantitative metabolomics in breast cancer. Metabolites 9:19
Gonzalez-Mendez R, Wemmer D, Hahn G, Wade-Jardetzky N, Jardetzky O (1982) Continuous-flow NMR culture system for mammalian cells. Biochim Biophys Acta 720:274–280
Gowda GAN, Abell L, Lee CF, Tian R, Raftery D (2016) Simultaneous analysis of major coenzymes of cellular redox reactions and energy using ex vivo 1H NMR spectroscopy. Anal Chem 8
Halama A (2014) Metabolomics in cell culture – a strategy to study crucial metabolic pathways in cancer development and the response to treatment. Arch Biochem Biophys 564:100–109
Hall AMR et al (2016) Practical aspects of real-time reaction monitoring using multi-nuclear high resolution flow NMR spectroscopy. Cat Sci Technol 6:8406–8417
Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144:646–674
Hanoulle et al (2005a) Monitoring of the ethionamide pro-drug activation .pdf
Hanoulle X et al (2005b) Monitoring of the ethionamide pro-drug activation in mycobacteria by 1H high resolution magic angle spinning NMR. Biochem Biophys Res Commun 331:452–458
Hanoulle et al (2006a) Selective intracellular accumulation of the major .pdf
Hanoulle X et al (2006b) Selective intracellular accumulation of the major metabolite issued from the activation of the prodrug ethionamide in mycobacteria. J Antimicrob Chemother 58:768–772
Illig T et al (2010) A genome-wide perspective of genetic variation in human metabolism. Nat Genet 42:137–141
Jang C, Chen L, Rabinowitz JD (2018) Metabolomics and isotope tracing. Cell 173:822–837
Jin ES, Sherry AD, Malloy CR (2013) Evidence for transaldolase activity in the isolated heart supplied with [U-13C3]glycerol. J Biol Chem 288:2914–2922
Judge MT et al (2018) Continuous in vivo metabolism by NMR. bioRxiv. https://doi.org/10.1101/501577
Judge MT et al (2019) Continuous in vivo metabolism by NMR. Front Mol Biosci 6:26

Karakitsou E et al (2021) Genome-scale integration of transcriptome and metabolome unveils squalene synthase and dihydrofolate reductase as targets against AML cells resistant to chemotherapy. Comput Struct Biotechnol J 19:4059–4066
Knitsch R, AlWahsh M, Raschke H, Lambert J, Hergenröder R (2021) In vitro spatio-temporal NMR metabolomics of living 3D cell models. Anal Chem 93:13485–13494
Koczula KM et al (2016) Metabolic plasticity in CLL: adaptation to the hypoxic niche. Leukemia 30:65–73
Lau AN et al (2020) Dissecting cell-type-specific metabolism in pancreatic ductal adenocarcinoma. Elife 9:e56782
Lee REB, Li W, Chatterjee D, Lee RE (2005) Rapid structural characterization of the arabinogalactan and lipoarabinomannan in live mycobacterial cells using 2D and 3D HR-MAS NMR: structural changes in the arabinan due to ethambutol treatment and gene mutation are observed. Glycobiology 15:139–151
Li W (2006) Multidimensional HRMAS NMR: a platform for in vivo studies using intact bacterial cells. Analyst 131:777
Lodi A et al (2017) Combinatorial treatment with natural compounds in prostate cancer inhibits prostate tumor growth and leads to key modulations of cancer cell metabolism. NPJ Precision Oncol 1:18
Lucas-Torres C, Wong A (2019) Current developments in μMAS NMR analysis for metabolomics. Metabolites 9:29
Ludwig C, Guenther UL (2009) Ligand based NMR methods for drug discovery. Front Biosci 14: 24
Ludwig C et al (2011) 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 8:8–18
Markley JL et al (2017) The future of NMR-based metabolomics. Curr Opin Biotechnol 43:34–40
Meyer B, Peters T (2003) NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew Chem Int Ed 42:864–890
Mobarhan YL, Struppe J, Fortier-McGill B, Simpson AJ (2017) Effective combined water and sideband suppression for low-speed tissue and in vivo MAS NMR. Anal Bioanal Chem 409: 5043–5055
Moonira T et al (2020) Metformin lowers glucose 6-phosphate in hepatocytes by activation of glycolysis downstream of glucose phosphorylation. J Biol Chem 295:3330–3346
Nicholson JK, Lindon JC, Holmes E (1999) ‘Metabonomics’: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29:1181–1189
Nilsson A, Nielsen J (2017) Genome scale metabolic modeling of cancer. Metab Eng 43:103–112
Nyblom HK, Nord LI, Andersson R, Kenne L, Bergsten P (2008) Glucose-induced de novo synthesis of fatty acyls causes proportional increases in INS-1E cellular lipids. NMR Biomed 21:357–365
Ortmayr K, Dubuis S, Zampieri M (2019) Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism. Nat Commun 10:1841
Palmnas M, Vogel H (2013) The future of NMR metabolomics in cancer therapy: towards personalizing treatment and developing targeted drugs? Metabolites 3:373–396
Powers R (2009) NMR metabolomics and drug discovery: NMR metabolomics and drug discovery. Magn Reson Chem 47:S2–S11
Rais B et al (1999) Oxythiamine and dehydroepiandrosterone induce a G(1) phase cycle arrest in Ehrlich’s tumor cells through inhibition of the pentose cycle. FEBS Lett 456:113–118
Ramaswamy V et al (2013) Development of a ^{13}C-optimized 1.5-mm high temperature superconducting NMR probe. J Magn Reson 235:58–65
Rappez L et al (2021) SpaceM reveals metabolic states of single cells. Nat Methods 18:799–805

Reed MAC, Roberts J, Gierth P, Kupče Ē, Günther UL (2019) Quantitative isotopomer rates in real-time metabolism of cells determined by NMR methods. Chembiochem 20(17):2207–2211. https://doi.org/10.1002/cbic.201900084

Righi V, Constantinou C, Kesarwani M, Rahme LG, Tzika AA (2013) Live-cell high resolution magic angle spinning magnetic resonance spectroscopy for in vivo analysis of Pseudomonas aeruginosa metabolomics. Biomed Rep 1:707–712

Robinette SL, Zhang F, Brüschweiler-Li L, Brüschweiler R (2008) Web server based complex mixture analysis by NMR. Anal Chem 80:3606–3611

Saborano R et al (2019) A framework for tracer-based metabolism in mammalian cells by NMR. Sci Rep 9(1):2520

Sarou-Kanian V et al (2015) Metabolite localization in living drosophila using high resolution magic angle spinning NMR. Sci Rep 5:9872

Schmidt DR et al (2021) Metabolomics in cancer research and emerging applications in clinical oncology. CA Cancer J Clin 71:333–358

Selivanov VA, Marin S, Lee PWN, Cascante M (2006) Software for dynamic analysis of tracer-based metabolomic data: estimation of metabolic fluxes and their statistical analysis. Bioinformatics 22:2806–2812

Selivanov VA et al (2020) Software supporting a workflow of quantitative dynamic flux maps estimation in central metabolism from SIRM experimental data. In: Nagrath D (ed) Metabolic flux analysis in eukaryotic cells, vol 2088. Springer, pp 271–298

Shepherd EL (2021) Ketohexokinase inhibition improves NASH by reducing fructose-induced steatosis and fibrogenesis. JHEP Rep 3:12

Szyperski T (1995) Biosynthetically directed fractional ^{13}C-labeling of Proteinogenic amino acids. Eur J Biochem 232:433–448

Szyperski T, Bailey JE, Wüthrich K (1996) Detecting and dissecting metabolic fluxes using biosynthetic fractional ^{13}C labeling and two-dimensional NMR spectroscopy. Trends Biotechnol 14:453–459

Tarrado-Castellarnau M, de Atauri P, Tarragó-Celada J, Perarnau J, Yuneva M, Thomson TM, Cascante M (2017) De novo MYC addiction as an adaptive response of cancer cells to CDK4/6 inhibition. Mol Syst Biol 13(10):940

Tarragó-Celada J et al (2021) Cysteine and folate metabolism are targetable vulnerabilities of metastatic colorectal cancer. Cancers 13:425

Thiele I et al (2013) A community-driven global reconstruction of human metabolism. Nat Biotechnol 31:419–425

Thomas JN et al (2021) Progress towards a higher sensitivity ^{13}C-optimized 1.5 mm HTS NMR probe. IEEE Trans Appl Supercond 31:1–4

Thompson JW et al (2019) International ring trial of a high resolution targeted metabolomics and lipidomics platform for serum and plasma analysis. Anal Chem 91:14407–14416

Tiziani S et al (2009) Metabolomic profiling of drug responses in acute myeloid Leukaemia cell lines. PLoS One 4:e4251

Tiziani S, Kang Y, Choi JS, Roberts W, Paternostro G (2011) Metabolomic high-content nuclear magnetic resonance-based drug screening of a kinase inhibitor library. Nat Commun 2:545

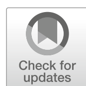
Ugurbil K, Rottenberg H, Glynn P, Shulman RG (1978a) 31P nuclear magnetic resonance studies of bioenergetics and glycolysis in anaerobic Escherichia coli cells. Proc Natl Acad Sci 75: 2244–2248

Ugurbil K, Brown TR, Den Hollander JA, Glynn P, Shulman RG (1978b) High-resolution ^{13}C nuclear magnetic resonance studies of glucose metabolism in Escherichia coli. Proc Natl Acad Sci 75:3742–3746

Vermathen M, von Tengg-Kobligk H, Hungerbühler MN, Vermathen P, Ruprecht N (2021) 1H HR-MAS NMR based metabolic profiling of lung cancer cells with induced and De-induced cisplatin resistance to reveal metabolic resistance adaptations. Molecules 26:6766

Vignoli A et al (2019) High-throughput metabolomics by 1D NMR. Angew Chem Int Ed 58: 968–994

Vilaplana-Lopera N et al (2021) Crosstalk between AML and stromal cells triggers acetate secretion through the metabolic rewiring of stromal cells. 2021.01.21.427406 Preprint. https://doi.org/10.1101/2021.01.21.427406
Wan Q, Wang Y, Tang H (2017) Quantitative ^{13}C traces of glucose fate in hepatitis B virus-infected hepatocytes. Anal Chem 89:3293–3299
Wang C et al (2020) COLMAR lipids web server and ultrahigh-resolution methods for two-dimensional nuclear magnetic resonance- and mass spectrometry-based lipidomics. J Proteome Res 19:1674–1683
Watanabe R et al (2016) Quantitative nuclear magnetic resonance spectroscopy based on PULCON methodology: application to quantification of invaluable marine toxin, okadaic acid. Toxins (Basel) 8:294
Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15:473–484
Wishart DS et al (2009) HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res 37:D603–D610
Yu M et al (2015) A resource for cell line authentication, annotation and quality control. Nature 520:307–311
Zhang A, Sun H, Xu H, Qiu S, Wang X (2013) Cell metabolomics. OMICS 17:495–501



NMR-Based Metabolomics to Evaluate Individual Response to Treatments

Alessia Vignoli, Gaia Meoni, Veronica Ghini, Francesca Di Cesare, Leonardo Tenori, Claudio Luchinat, and Paola Turano

Contents

1 Introduction 210
1.1 The Individual Metabolic Phenotype 211
1.2 Modulation of the Individual Metabolic Phenotype 212
2 Disease Fingerprinting and Individual Response to Pharmacological Therapy 214
2.1 Celiac Disease 214
2.2 Cancer 215
2.2.1 Breast Cancer 216
2.2.2 Lung Cancer 219
2.2.3 Colon Cancer 223
2.3 Viral Infections 226
2.3.1 Hepatitis 226
2.3.2 COVID-19 229
2.4 Chronic Obstructive Pulmonary Disease 232
2.5 Drug Toxicity 233
References 234

Alessia Vignoli and Gaia Meoni contributed equally to this work.

A. Vignoli · G. Meoni · V. Ghini · F. Di Cesare
Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Italy

Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

L. Tenori · C. Luchinat · P. Turano (✉)
Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Italy

Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

Consorzio Interuniversitario Risonanze Magnetiche MetalloProteine (CIRMMP), Sesto Fiorentino, Italy
e-mail: turano@cerm.unifi.it

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_618

Abstract

The aim of this chapter is to highlight the various aspects of metabolomics in relation to health and diseases, starting from the definition of metabolic space and of how individuals tend to maintain their own position in this space. Physio-pathological stimuli may cause individuals to lose their position and then regain it, or move irreversibly to other positions. By way of examples, mostly selected from our own work using ^{1}H NMR on biological fluids, we describe the effects on the individual metabolomic fingerprint of mild external interventions, such as diet or probiotic administration. Then we move to pathologies (such as celiac disease, various types of cancer, viral infections, and other diseases), each characterized by a well-defined metabolomic fingerprint. We describe the effects of drugs on the disease fingerprint and on its reversal to a healthy metabolomic status. Drug toxicity can be also monitored by metabolomics. We also show how the individual metabolomic fingerprint at the onset of a disease may discriminate responders from non-responders to a given drug, or how it may be prognostic of e.g., cancer recurrence after many years. In parallel with fingerprinting, profiling (i.e., the identification and quantification of many metabolites and, in the case of selected biofluids, of the lipoprotein components that contribute to the ^{1}H NMR spectral features) can provide hints on the metabolic pathways that are altered by a disease and assess their restoration after treatment.

Keywords

^{1}H-NMR · Cancer · Celiac disease · COPD · Drug toxicity · Fingerprinting · Profiling · Response to treatment · Viral infections

1 Introduction

The use of metabolomics in personalized medicine originates from two basic facts:

1. The existence of an individual metabolic phenotype characteristic for each individual, i.e. of an invariant part of the metabolome that allows each subject to be discriminated from all others.
2. The existence of a signature of the disease, which may range from very weak to very strong depending on the pathology or its severity. The disease transiently alters the individual metabolic phenotype, but this alteration disappears when the individual reverts to the “healthy” status following medical or pharmaceutical interventions.

Despite ^{1}H NMR features a lower sensitivity (detection limit in the order of μM) with respect to MS analysis, which downsizes the total number of measurable

metabolites, it is exquisitely amenable to the *untargeted fingerprinting* of the sample metabolome due to its very high reproducibility and capability for high throughput analysis (Griffiths 2008; Takis et al. 2019; Vignoli et al. 2019).

With the exception of inorganic ions, almost all small molecules contain hydrogen atoms that can be measured simultaneously, providing rapid and distinct global spectral patterns (or fingerprints) of the samples under investigation (Klupczyńska et al. 2015; Takis et al. 2019; Vignoli et al. 2019). Thus, the fingerprint of a sample is a recapitulation of its current metabolome, independently of the identification of the metabolites. Conversely, *metabolomic profiling* is a global evaluation of the metabolite content of all the samples in a comparative fashion. The final aim is to identify and accurately quantify as many compounds as possible. This approach enables the detection of changes in the concentration of the measurable metabolites related to specific physiological conditions.

Here, we first provide a series of examples demonstrating how untargeted ^{1}H NMR metabolomic fingerprinting can be used to identify the individual metabolic phenotype and assess its stability over a long timescale in the absence of important physio-pathological alterations. The existence of this individual fingerprint has been observed in several biosamples (such as urine, saliva, blood, breath condensate) that are commonly used in metabolomics. Then we provide a series of examples spanning from celiac disease to cancer and viral infections. These pathologies are very different in their etiologies and clinical manifestations and involve different organs. Likewise, in our lab we have observed metabolomic alterations that are clearly characteristic of each of them. Additionally, we have used NMR-based metabolomics to monitor the response to pharmaceutical treatments, and to identify, before treatment, features of the individual metabolome that are prognostics to discriminate responders from non-responders.

To help readers assess the potentiality of the NMR-based approach, we describe a series of case studies drawn on the many years of experience of our group.

1.1 The Individual Metabolic Phenotype

In 2008, our research group (Assfalg et al. 2008) demonstrated, for the first time, that the NMR detectable part of the metabolome of urine contains an invariant part, which can be considered as the chemical signature of each individual. Multivariate statistical analysis of multiple urine samples of different individuals enables the definition of their "metabolic space", where the metabolomic fingerprint of each subject can be visualized and discriminated from that of the other subjects with an accuracy close to 100%. Over a timescale of 10 years (Ghini et al. 2015a), in the absence of important stressful perturbations, each individual still occupies its specific region of the metabolic space; daily intra-individual variability leads to small fluctuations inside the individual metabolic space; conversely, shifts to other distinct regions are associated to significant changes of the individual metabolomic phenotype as a consequence of the occurrence of important physio-pathological conditions (Fig. 1).

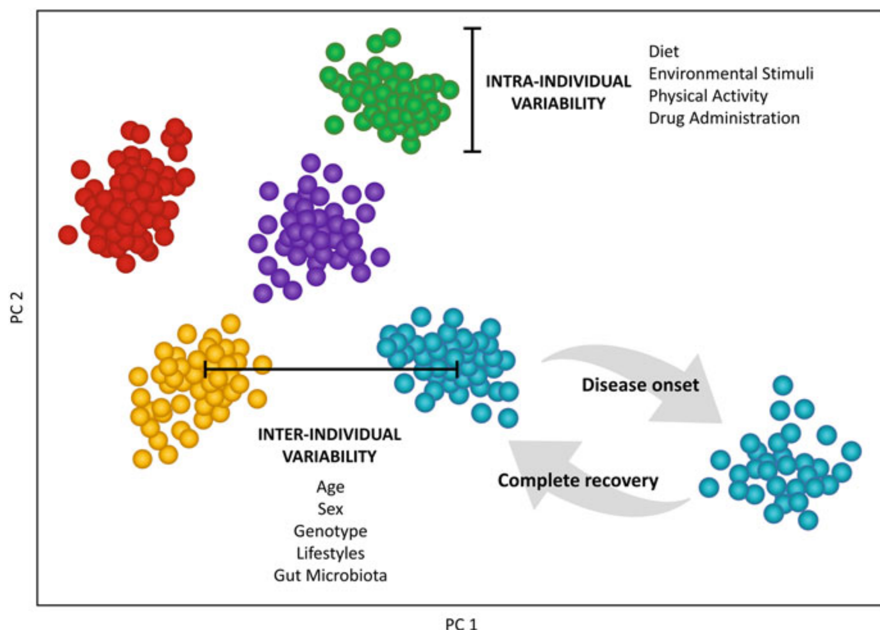


Fig. 1 A schematic representation of the individual metabolic phenotype. Each individual (represented by a different color code in the picture) has a stable phenotype, which distinguishes him/her from the other subjects (inter-individual variability). This phenotype is flexible enough to compensate for day-by-day changes and external stimuli (intra-individual variability). This capacity is lost at the onset of a disease but can be restored after a complete recovery

A strong individual phenotype also exists in saliva (Wallner-Liebmann et al. 2016) and blood (Holmes et al. 2008). For the latter two biofluids, their intrinsic nature causes daily intra-individual fluctuations in metabolome composition that are smaller than what occurs in urine; this is particularly true for blood, the composition of which is tightly regulated by homeostasis. Additionally, at variance with urine and saliva that have individual metabolomic fingerprints dominated by low-molecular weight molecules, in both plasma and serum the presence of lipids and lipoproteins significantly contributes to the individual fingerprint.

For serum/plasma, the stability of the individual phenotype has been reported over a period of 7 years (but is likely to be much longer), while no information about the long-term stability of the metabolic phenotype in saliva is yet available.

1.2 Modulation of the Individual Metabolic Phenotype

Beside the genetic components, all endogenous and exogenous metabolites derived from extrinsic factors such as diet, drugs, gut microflora, stressors, and pollutants contribute to the definition of the human metabolic phenotype. Thus, several studies have been performed to understand the effects of mild dietary interventions on the

metabolomic profile (Andersen et al. 2014; O'Sullivan et al. 2011). Importantly, despite the levels of a large set of metabolites have been reported to be modulated by the different dietary habits and to become more homogeneous in individuals undergoing diet standardization for a short period of time, diet regimes do not significantly change the individual signatures, which remain equally evident in all types of biofluids (Lenz et al. 2003; Marianne and Walsh 2006; Wallner-Liebmann et al. 2015; Winnike et al. 2009). The same effects are reported for probiotics assumption in 22 healthy volunteers for 8 weeks (Ghini et al. 2020b). Because of the probiotic treatment, significant modulations in the levels of a few urine and serum metabolites are observed, but the observed changes are not so strong as to hamper the individual recognition in the metabolic space. Another example is represented by the evaluation of the metabolomic effects induced by the administration of bioactive molecules in the serum profiles of volunteers at risk of metabolic syndrome. This study was performed in the context of the EC funded project PATHWAY-27 (FP-KBBE # 311876). In this framework, the beneficial effects of DHA (docosahexaenoic acid) administration on serum metabolomic profiles could be assessed. However, this was only possible by suppressing the intra-individual variations using paired approaches that compare the serum samples collected at the beginning and at the end of the trial for each subject (Ghini et al. 2017).

The above examples demonstrate that little or no alteration of the individual phenotype is brought about by mild treatments, whereas it can be profoundly affected by the presence of pathologies and major surgical interventions, such as bariatric surgery. Currently, bariatric surgery is the only available treatment to provide sustained weight loss (Buchwald et al. 2009), to improve glucose regulation, and to even promote complete remission of type 2 diabetes in severely obese patients (Meijer et al. 2011). All three different bariatric procedures, i.e. sleeve gastrectomy, proximal Roux-en-Y gastric bypass, and distal bypass, are associated with a strong alteration of the serum metabolomic fingerprints of the patients (Gralka et al. 2015). Within the common strong alterations, distal bypass patients could be discriminated from the other two groups of patients, suggesting a stronger impact of this procedure on the metabolic fingerprint. Different short-term and long-term alterations are also observed. The serum metabolomic profiles of severe obese patients are characterized by high levels of aromatic and branched-chain amino acids (AAA and BCAA, respectively), and of metabolites related to energy metabolism (pyruvate and citrate). Elevated levels of some metabolites related to the gut microbiota such as formate, methanol, and isopropanol are also associated with obesity. Interestingly, after bariatric surgery, independently of the type of procedure used, a significant reduction of AAA, BCAA, pyruvate, methanol, and isopropanol concentrations, along with an increase in arginine and glutamine levels are observed, indicating that surgically induced weight loss can, at least in part, normalize the alterations in amino acid metabolism associated with obesity (Gralka et al. 2015).

2 Disease Fingerprinting and Individual Response to Pharmacological Therapy

As anticipated, several pathologies have a strong metabolomic fingerprint at the systemic level. Below we provide some examples from our lab of the metabolomic characterization of different types of diseases. In these examples, NMR-based metabolomics successfully provided a medium-to-strong fingerprint of the disease and a definition of the molecular profile of the pathology in terms of small molecules and lipoprotein parameters. Additionally, NMR was used to monitor the change of the metabolomic fingerprint and profile along the disease evolution and in particular following pharmacological treatments. Indeed, pharmacometabolomics is defined as the determination of the individual metabolic phenotype to characterize signatures, both before and after drug administration, that might inform treatment outcomes, map the effects of drugs on metabolism and identify molecular pathways contributing to drug-response and drug-toxicity phenotypes (Kaddurah-Daouk et al. 2014).

2.1 Celiac Disease

The celiac disease (CD) originates from an aberrant adaptive immune response against gluten. The presence of a characteristic metabolomic fingerprint of CD has been demonstrated by multivariate analysis of the NMR spectra of serum and urine samples from affected subjects and healthy controls (Bertini et al. 2009, 2011a; Vignoli et al. 2019). For example, when comparing adult patients with sex- and age-matched controls, a discrimination accuracy in the range of 75–83% was obtained in urine, which rises to 81–94% in serum, depending on the study cohorts. The differences originate from variations in the levels of metabolites related to three main mechanisms: malabsorption, altered energy metabolism, and altered gut microflora (Bertini et al. 2009).

In serum samples, the main reported differences between CD patients and controls consist in lower levels of several amino acids (asparagine, isoleucine, methionine, proline, and valine) and of pyruvate, lactate, and lipids, and in higher levels of glucose and 3-hydroxybutyric acid. The decreased levels of pyruvate and lactate along with higher levels of glucose are indicative of an impaired glycolysis process. Enhanced beta-oxidation and malabsorption can instead explain the lower levels of amino acids and lipids observed. A possible increase in the use of ketonic bodies as a source of energy in celiac patients is consistent with the high levels of 3-hydroxybutyric acid. Energy conversion from lipids and catabolism of ketonic bodies is far less efficient than that from glucids; consistently, celiac patients often report symptoms of fatigue. Further, urine samples of CD patients are characterized by high levels of some metabolites related to the gut microbiota, i.e. indoxyl sulfate, meta-[hydroxyphenyl] propionic acid, phenylacetylglycine, and p-cresolsulfate. All these findings are consistent with the hypothesis that in CD patients the gut

microflora of the small bowel is altered or presents peculiar microbial species with their own metabolome.

Notably, the so-called potential celiac patients, i.e., those individuals who tested positive for the antibodies but have no evidence of intestinal damage, have an NMR fingerprint that is similar to that of the overt patients, suggesting that the dysmetabolism precedes the intestinal damage (Bernini et al. 2011a, b).

Celiac disease represents one example where dietary intervention induces relevant changes in the metabolomic fingerprint (Bernini et al. 2011a, b). The only available treatment for this disorder is to follow a gluten-free diet. One of the most interesting findings was that the metabolic fingerprint of CD patients reverts to normality after 12 months of a strict gluten-free diet. After the gluten-free diet, normal levels of glucose and 3-hydroxybutyric acids are restored, and the sense of fatigue tends to be reduced.

2.2 Cancer

Metabolomics can be used to derive information on cancer at various levels. Several studies concern the characterization of the metabolome of cancer cell lines as a complementary approach with respect to classical biochemical analyses and to other omics, with the aim to derive information on altered metabolic pathways (Cuperlovic-Culf et al. 2012; Li et al. 2017) and to conduct preclinical tests of new anticancer agents (D'Alessandro et al. 2019; Ghini et al. 2021; Li et al. 2019; Resendiz-Acevedo et al. 2021). NMR is particularly suitable for this purpose, as it allows for a fast untargeted characterization of the endo and exometabolome (i.e., the intracellular and extracellular composition in terms of metabolites), also at multiple post-drug treatment times. Additionally, isotopically labeled substrates (e.g., ^{13}C labeled glucose) can be used to define metabolic fluxes by following the metabolism of labeled substrates into their pathway products at specific time points (Antoniewicz 2018; Saborano et al. 2019).

Tumors can also be characterized at the level of the tissue metabolome; this approach provides a specific view of cancer cells and their cross-talk with the tumor microenvironment (Márquez and Matés 2021). Cancer tissues represent the localized site of the disease. ^{1}H NMR, in its HR-MAS version, is applicable to acquire spectra of small quantities of cryo-preserved tissues (of the order of 10 mg), provided they have not undergone any further sample manipulation (Vignoli et al. 2019). Indeed, the use of tissues for ex vivo metabolomics raises a number of criticalities, ranging from ethical to technical issues. Surgical specimens collected at different times of intraoperative (warm) or post-operative (cold) ischemia undergo significant molecular degradation, so that the measured "metabolome" no longer reflects the original physiological state of the tissues before intervention (Cacciatore et al. 2013). The need of suitable frozen material collected and handled under strict preanalytical conditions conflicts with ethics that assigns precedence, during intraoperative procedures, to the best surgical performances and afterward to the diagnostic needs of the pathologist.

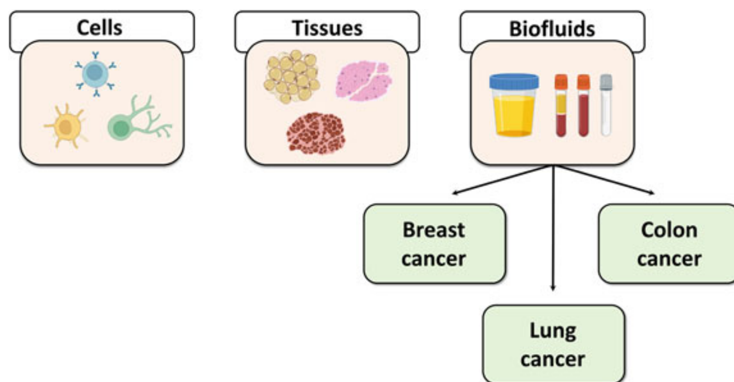


Fig. 2 Types of samples that can be analyzed via NMR-based metabolomics: cells, tissues, and biofluids. In particular, in Sect. 2.2, we will focus on the analysis of biofluids (i.e. serum, plasma, and urine) to investigate pathophysiological conditions associated with breast cancer, lung cancer, and colon cancer

Instead, the metabolomic analysis of biofluids, such as blood serum or plasma, or urine, presents several advantages. Sample collection is only minimally invasive, and multiple collections at different time points can be easily obtained to establish a metabolomic signature both before and after a given drug therapy. The preanalytical SOPs are simple and easily satisfied (Ghini et al. 2019; ISO 23118:2021; Vignoli et al. 2022). The resulting analyses might inform on the presence of specific features of prognostic value and on treatment outcomes, respectively.

Below we provide examples that have seen a significant contribution from our group and that regard the characterization of three of the most common types of cancer, namely breast cancer, colorectal cancer, and lung cancer; all of them are based on patients' serum/plasma (Fig. 2).

2.2.1 Breast Cancer

Female breast cancer (BC) is the most commonly diagnosed cancer, with an estimated 2.3 million new cases per year (11.7%), and is the leading cause of cancer death in women (Sung et al. 2021). For many years BC has been considered a unique clinical entity and treated with only one general approach. However, it has become extremely clear that there is a high degree of diversity between and within the BC subclasses (Polyak 2011). Based on their molecular characteristics, BC is classified into three major subtypes: luminal (positive for estrogen and progesterone receptors), human epidermal growth factor receptor 2+ (HER2+) enriched, and basal (the majority of the latter tumors are also called triple-negative BC). Currently, these different tumor subtypes are treated with specific therapies improving patient survival: hormone receptor-positive disease is generally treated with endocrine therapy (Early Breast Cancer Trialists' Collaborative Group (EBCTCG) 2005; Lerner et al. 1976; Wiggans et al. 1979); HER2+ disease is now cured with targeted agents, which have significantly improved outcomes in both the (neo)adjuvant

(Gianni et al. 2012; Martin et al. 2017; Slamon et al. 2011; von Minckwitz et al. 2017, 2019) and the metastatic setting (Swain et al. 2013; Verma et al. 2012).

Metabolomics can distinguish patients with BC with respect to healthy controls with high predictive accuracies: PLS models with $Q^2 > 0.6$ for studies on serum/ plasma (Cala et al. 2018; Lécuyer et al. 2018; Silva et al. 2019; Singh et al. 2017; Sitter et al. 2006; Slupsky et al. 2010; Suman et al. 2018; Tayyari et al. 2018; Wojtowicz et al. 2020; Zhou et al. 2017), and models with $Q^2 > 0.5$ for urinary studies (Silva et al. 2019; Slupsky et al. 2010; Zhou et al. 2017) were obtained.

Metabolomics proved useful in discriminating the plasma profiles of patients with different BC molecular subtypes in comparison to controls (Díaz-Beltrán et al. 2021): as compared to HER2-patients, the HER2+ group showed elevated aerobic glycolysis, gluconeogenesis, and increased fatty acid biosynthesis with reduced Krebs cycle. ER+ patients, as compared to ER- ones, showed elevated alanine, aspartate, and glutamate metabolism, decreased glycerol-lipid catabolism, and enhanced purine metabolism (Fan et al. 2016).

Given these premises, in the framework of BC precision oncology, it is crucial to identify, in each BC subtype, patients at higher risk of cancer recurrence and drug-response profiles able to guide patients' management (Bendinelli et al. 2021; Vignoli et al. 2021b). The NMR-based metabolomics has shown to be a valuable prognostic instrument and examples of its application with the abovementioned objectives are provided below.

Response to Chemotherapy

The possible association between different metabolic profiles and response to chemotherapy has been extensively investigated via NMR metabolomics of blood derivatives.

Pharmacometabolomics on blood plasma/serum with the aim of predicting which patients will benefit most from a specific treatment has provided significant results. First in 2012, our group demonstrated that NMR-based metabolomics may play a role in identifying the best responding patients with HER2+ metastatic breast cancer (MBC) treated with paclitaxel plus lapatinib (Tenori et al. 2012). With a similar NMR approach, Jiang and colleagues predicted the response to gemcitabine/ carboplatin chemotherapy (Jiang et al. 2018).

More recently, the general research attention has focused on neo-adjuvant chemotherapy (NAC), the preferred treatment strategy for patients with large primary tumors and/or locally advanced disease (Thompson and Moulder-Thompson 2012). NAC offers the unquestionable benefit of downstaging disease and reducing the tumor size prior to surgery, thus making patients with inoperable tumors candidates for surgical resection (Debik et al. 2019; Thompson and Moulder-Thompson 2012). Nonetheless, only less than 30% of patients overall exhibit pathological complete response (disappearance of all invasive cancer in the breast) to NAC (Wei et al. 2013), with lower rates of response observed in endocrine receptor-positive, HER2-disease.

In 2020, our research group, for the first time, investigated the capability of predicting pathological complete response, using the baseline host immune

cytokines and ^{1}H NMR metabolomic fingerprints in a highly homogeneous population of HER2+ BC patients enrolled for NAC treatment (Vignoli et al. 2020a). For this study, 43 HER2+ BC patients, stratified according to their ER status in 22 ER+ and 21 ER-, were enrolled at baseline prior to any interventions. The pathological complete response was obtained in 11 out of 22 ER+ patients and in 13 out of 21 ER- patients. The plasma metabolomic fingerprint can distinguish ER+ and ER- patients with 74.4% discrimination accuracy, suggesting a differential host–cancer interaction in these two subtypes of BC. Moreover, in the ER+ BC patients the baseline metabolomic fingerprint can be predictive of pathological complete response (72.7% accuracy). The good responders, as compared to poor responders, are characterized by lower concentrations of branched-chain amino acids isoleucine, and valine, as well as ethanol, several phospholipids and cholesterol associated to almost all classes of lipoproteins assigned by NMR. In the ER+ subgroup the combination of a cytokine (TNF-α) and a metabolite (valine) was found to significantly enhance discrimination between complete and partial response to NAC, yielding an area under the receiver operating characteristic (ROC) curve of 0.92, and an accuracy of 90.9%. Conversely, no predictivity was observed in ER- BC patients. In their pilot study, Wei et al. 2013 had already shown that metabolic profiling, performed by combining NMR and LC–MS method, can distinguish groups of BC patients with no, partial, or complete response; consistently, our study confirms this evidence and enriches this scenario by coupling NMR-based metabolomics with the analysis of a panel of 10 different cytokines.

Conversely, from blood serum analyses, NMR studies performed on breast biopsies did not reveal significant metabolic differences between complete pathological response and pathological non-responders before treatment (Choi et al. 2013; Euceda et al. 2017), implying that the host–cancer interaction at a systemic level plays a crucial role in the response to treatment. As a confirmation, a study performed via MS on both blood serum and tissue biopsies (collected before, during, and after NAC) showed that only marginally correlations are present between the two biospecimens, and that only the serum profile is predictive of NAC response (Debik et al. 2019).

Other relevant serum/plasma metabolomic studies have been targeted at characterizing the impact of NAC on the patient metabolism. The NAC induces relevant alterations in patient metabolism during and after treatment (Corona et al. 2021; Debik et al. 2019; Jobard et al. 2017). In particular, Jobard et al. (2017) have shown that the administration of trastuzumab and everolimus in combination induced systemic effects by altering lipid, glucose, and ketone bodies metabolisms. These alterations are observable on the metabolic profiles of patients even several weeks after the end of the drug intervention.

Disease Recurrence

Despite all efforts aimed at better stratifying BC patients, there is still a significant proportion of early breast cancer patients who are overtreated. Clinicians are looking for prognostic tools able to distinguish early BC patients at high risk of disease recurrence, who need to be treated with aggressive adjuvant therapies, with respect

to low-risk patients, who may be cured by locoregional therapy alone (McCartney et al. 2017, 2018). NMR-based metabolomics of blood derivatives has shown to play a role in this scenario.

The first evidence supporting the application of metabolomics as a potential prognostic tool for recurrence prediction was published by Asiago et al. in 2010 (Asiago et al. 2010). They utilized metabolic profiling approach, obtained by combining metabolites detected by both NMR and MS, to identify breast cancer relapse before it occurs. Over the past decade, our research group has pursued this research line using a fingerprinting approach. We have established a reproducible method, based on serum NMR metabolomic analysis, able to distinguish early and metastatic breast cancer patients with high discrimination accuracy. Furthermore, we demonstrated that this model can be used to predict cancer relapse: early BC patients "misclassified" as metastatic on the basis of their metabolomic fingerprint showed indeed high risk of recurrence, whereas early BC patients correctly classified as early BC can be considered at low risk. The recurrence prediction with this approach has been validated and reproduced in monocentric and multicentric cohorts of patients (Fig. 3) providing successful results (Tenori et al. 2015; Hart et al. 2017; McCartney et al. 2019). In our 2015 and 2019 studies, serum samples of early BC patients were collected before surgery, thus when the tumor was still in place, whereas in the multicentric study of 2017, samples were collected after surgery but before starting (when indicated) adjuvant chemotherapy or radiotherapy. Moreover, the early BC patients enrolled for the studies of 2017 and 2019 had ER+ breast cancer, whereas those of the study of 2015 were diagnosed with ER- breast cancer. The reproducibility of our approach, despite these differences in the study design, reinforces the evidence that NMR-based metabolomics is really a promising instrument for the stratification of patients with early breast cancer.

From our multicentric study (Hart et al. 2017) it emerged that, as compared with metastatic BC patients, patients with early ER+ BC are characterized by lower serum levels of citrate, choline, acetate, formate, lactate, glutamate, 3-hydroxybutyrate, phenylalanine, glycine, leucine, alanine, proline, tyrosine, isoleucine, creatine, creatinine, and methionine, and higher serum levels of glucose and glutamine. Interestingly, in the subgroup of early BC patients with either relapse or no-relapse (with a follow-up of at least 6 years), the patients who relapsed showed significantly higher serum levels of choline, phenylalanine, leucine, histidine, glutamate, glycine, tyrosine, valine, lactate, and isoleucine, thus resembling a "micro-metastatic" profile already at diagnosis. The decrease of tyrosine and lactate in early BC patients confirmed what already seen in the ER- negative cohort examined in 2015, whereas glucose and histidine showed opposite trends in ER- BC patients.

2.2.2 Lung Cancer

Lung cancer (LC) is the second most common cancer in both men and women and is by far the leading cause of cancer death worldwide, making up almost 25% of all cancer deaths. Most of the patients (>84%) with LC are affected by non-small cell lung cancer (NSCLC), with the majority of patients presenting with advanced, unresectable disease at the time of diagnosis (Siegel et al. 2021). Currently used

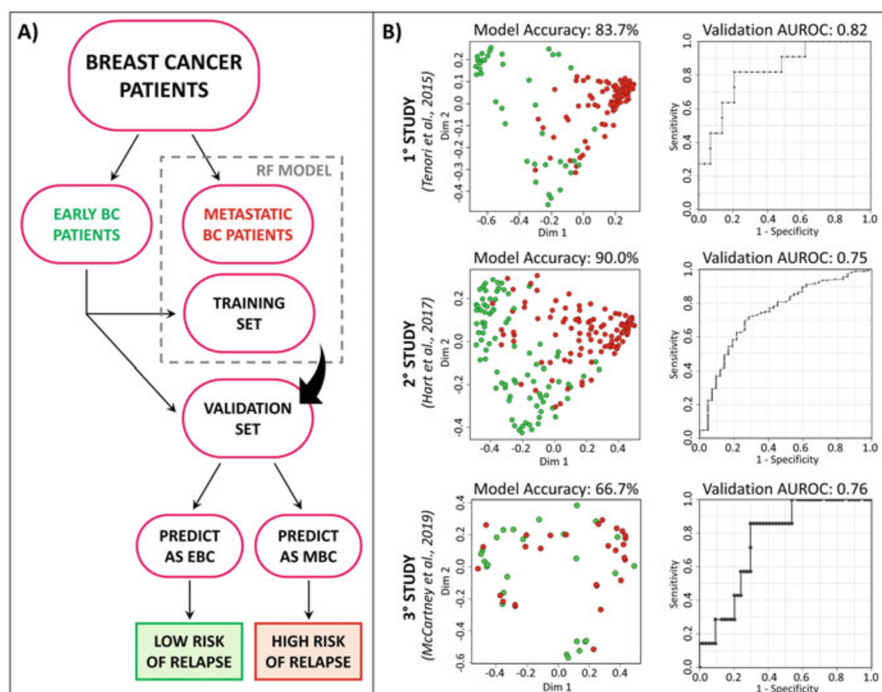


Fig. 3 (**a**) The study design of the three metabolomic studies performed in our laboratory (Tenori et al. 2015; Hart et al. 2017; McCartney et al. 2019) on serum samples of BC patients. Early BC patients were randomly divided into a training and a validation set. Using the early BC training set and the metastatic BC group, a random forest model was calculated on the serum NMR data. Then the early BC validation set was used to test the hypothesis that early BC patients at high risk of cancer recurrence could be classified as metastatic by the RF model. (**b**) The results of the three studies are summarized by reporting for each of them the RF proximity plot (early BC green dots, metastatic BC red dots) along with the RF model accuracy, and the area under the ROC curve of the validation set used for the recurrence prediction

treatments for advanced NSCLC include chemotherapy, targeted drug therapy, immunotherapy, or chemo-immunotherapy (Hirsch et al. 2017). Treatment options are based mainly on the tumor histology but other factors, such as certain cancer traits like PDL-1 TPS (tumor proportion score) and the presence of specific genomic mutations, are also important.

Response to Immunotherapy

Among all the therapeutic strategies, immunotherapy has become an attractive treatment modality for chemo-refractory solid tumors (Postow et al. 2015). NSCLC cells have the ability to evade the immune system by expressing on their surfaces certain "immune checkpoint" molecules that normally protect against autoimmunity and inflammation, such as cytotoxic T-lymphocyte antigen-4 (CTLA-4), programmed cell death protein 1 and its ligand (PD-1 and PDL-1,

respectively). Immune Checkpoint Inhibitor agents (ICIs), such as the monoclonal antibodies nivolumab and pembrolizumab, reactivate T lymphocyte-mediated immune response against the cancer cells by blocking the immune checkpoint molecules (Brahmer 2014; Hamada et al. 2018). Several ICIs have shown outstanding early success in many tumor types and have established an important role in the first line of treatment of advanced lung cancer as a monotherapy or in combination with chemotherapy, as well as in the second line after standard treatment (Borghaei et al. 2015; Gandhi et al. 2018; Gettinger et al. 2015; Herbst et al. 2016; Rittmeyer et al. 2017).

Unfortunately, not all patients respond to ICIs; the response rates are modest (approximately 30% or less in LC), the associated costs are high, and true predictive markers of treatment efficacy do not exist. Thus, the identification of biomarkers able to identify the patients that are most likely to respond to, and benefit from, ICIs treatment is of pivotal importance (Brahmer 2014). In this framework, metabolomic fingerprinting of biofluids may represent a timely tool to define metabolomic signatures that might inform on treatment outcomes.

In 2020, our research group conducted a pilot study based on ^{1}H NMR metabolomic investigation of sera samples from NSCLC patients treated with immune checkpoint inhibitors (Ghini et al. 2020a). The experimental scheme of the study is reported in Fig. 4. A total of 53 patients with advanced NSCLC were enrolled; 34 patients were treated with nivolumab (monoclonal antibody directed against PD-1) and 19 patients were treated with pembrolizumab (monoclonal antibody against PDL-1). All the analyzed samples were collected before the beginning of the treatment (T0) with the aim to a priori identify responder and non-responder subjects. Significantly, we could show that the metabolomic fingerprint of T0 serum acts as a predictive biomarker of immune checkpoint inhibitors response, being able to predict individual therapy outcome with >80% accuracy (Ghini et al. 2020a). In the serum samples of non-responder subjects, we detected significantly higher levels of pyruvate and alanine along with, even if not statistically significant, higher lactate and glycine levels and lower citrate levels. All these changes are evocative of increased glycolysis and decreased TCA pathway in non-responders. It is important to underline that the significance of the univariate analysis performed on single metabolite levels strongly relies on the number of subjects included in the study. Thus, further investigations enrolling much higher numbers of subjects are necessary to confirm these findings.

To the authors' knowledge, this represents the first study using NMR-based metabolomic fingerprinting of serum/plasma samples to predict the individual response to anti PD-1 therapy in NSCLC. Instead, a few other examples based on MS metabolomic profiling can be found in the literature. In 2020, Hatae et al. analyzed by GC-MS the plasma samples of 55 NSCLC patients treated with nivolumab and found that a combination of four plasma metabolites and several T cell markers could be used as a good biomarker for responder identification (area under the ROC curve = 0.96). The four selected metabolites include molecules related to (1) gut microbiota (hippuric acid), (2) fatty acid oxidation (butyryl-carnitine), and (3) redox-related metabolites (cystine and glutathione

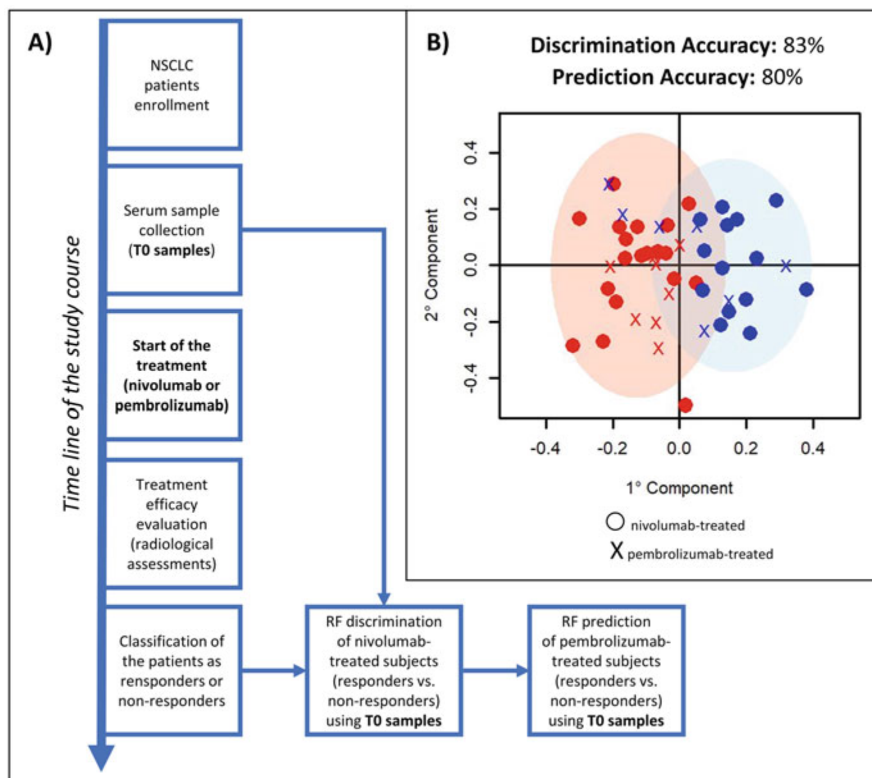


Fig. 4 (**a**) Experimental scheme to evaluate individual response to immunotherapy. (**b**) O-PLS discrimination between NSCLC responders and non-responders to immune check points inhibitors, adapted from Ghini et al. (2020a). Score plot, PC1 vs. PC2. Each symbol in the O-PLS score plot represents the fingerprint of the NMR spectrum of each patient recruited for the study. Red dots: nivolumab responder subjects; blue dots: nivolumab non-responder subjects. Red crosses: pembrolizumab subjects predicted as responders; blue crosses: pembrolizumab subjects predicted as non-responders

disulfide) (Hatae et al. 2020). One year later, in (2021), Nie X. et al. used LC-MS profiling of early on-treatment serum to explore predictors of clinical outcomes of anti-PD-1 treatment in 74 Chinese NSCLC patients. Serum samples were collected 2–3 weeks after the first infusion of PD-1 inhibitor. A small metabolite panel consisting of hypoxanthine and histidine was identified and validated as a predictor of treatment response, and high levels of both hypoxanthine and histidine were associated with improved progression-free survival and overall survival (Nie et al. 2021). The completely different metabolites observed in the reported investigations and proposed as treatment-efficacy biomarkers derive from different analytical platforms, which allow for the observation of different panels of metabolites.

The strength of our study relies on the uniqueness of the ^{1}H NMR fingerprinting, which takes advantage of its intrinsically untargeted nature and high reproducibility. This approach allowed us to identify a metabolomic signature associated with ICIs response that is independent of metabolite assignment and acts as a stronger "collective" biomarker with respect to a single molecule or to a panel of a few molecules.

2.2.3 Colon Cancer

Colorectal cancer (CRC) is the third most prevalent malignancy after breast and lung cancer, and the second most lethal disease in the world, with an anticipated 1.9 million new cases and 0.9 million deaths in 2020 (Xi and Xu 2021). The stage of CRC at diagnosis is the most important predictor of survival: a relative 5-year survival rate of around 90% has been demonstrated for patients diagnosed with localized-stage disease, declining to around 71% and 14% for those diagnosed with regional and distant metastasis respectively (Salmerón et al. 2022). Despite that, over the last 20 years, breakthroughs in CRC treatment have resulted in a steady increase in median overall survival (Dekker et al. 2019), CRC is still one of the most lethal diseases. CRC is known for its significant variety in clinical presentation and underlying tumor biology, as well as its relationship with numerous types of etiological variables (Cunningham et al. 2010). Metabolomics of several biofluids is increasingly used for successful patient classification in CRC, determining a strong signature of the disease. This allows for the discrimination of CRC patients from healthy subjects (Nannini et al. 2020; Turano 2014) and permits the prediction of the overall survival (OS) within a set of metastatic patients using serum samples (Bertini et al. 2012).

NMR-based metabolomics of minimally invasive biospecimens such as feces can, not only, correctly classify CRC patients from healthy subjects, as it was already shown by several studies (Le Gall et al. 2018; Lin et al. 2016, 2019; Monleón et al. 2009), but also discriminate CRC from patients with adenomatous polyps (AP) (Nannini et al. 2021). Some of the identified metabolites suggest that metabolic changes in CRC and adenoma are associated with different pathways, mainly involving amino acid metabolism. It is well known that 95% of CRCs begin as colonic AP or adenomas, and the possibility to correctly differentiate these two forms is of primary importance for the early detection of the tumor.

Disease Recurrence

Using metabolomics to determine the distinct profiles of CRCs might allow for more personalized or informed cancer therapy adjustments, contributing to precision medicine (Wishart 2015). Indeed, CRC shows different characteristics of clinical onset and individual response, even at the same pathological stage. Despite 80% of CRCs are diagnosed at early stage and immediately treated with surgery, 35% of these treated patients develop cancer relapse within 2–3 years after surgery (Guraya 2019). The assessment of recurrence risk in colon cancer primarily relies on the pathological stage defined by the TNM system, based on the depth of tumor invasion (T), lymph nodes involvement (N), and distant metastases (M). Risk stratification is of fundamental importance for the choice of adjuvant treatment; however, only a

small portion of patients benefits from it, with the majority being already cured by primary surgery, and others experiencing disease relapse despite having received adjuvant therapy. Improved identification of individuals who would benefit most from adjuvant chemotherapy is a critical aim, particularly in older patients who are more susceptible to treatment-related toxicity.

In 2021, our group demonstrated that NMR-based metabolomics on serum samples can improve risk stratification in elderly patients with early CRC (Di Donato et al. 2021). In this study, 169 serum samples, taken from three distinct clinical trials, were collected before treatment from elderly CRC patients. Of these, 94 were collected from patients with early CRC (65 relapse free and 29 relapsed) and 75 from patients with metastatic CRC (Fig. 5). The model was built using a supervised algorithm to discriminate the serum fingerprint of the relapse-free patients in the early CRC cohort and the patients in the metastatic CRC cohort, yielded 70% accuracy, 71% sensitivity, and 69% specificity. Then, with the hypothesis that the metabolomic fingerprint of relapsed early CRC patients would be more similar to that of patients with metastatic CRC, the model was used to predict the remaining early CRC who had disease recurrence. Among the early CRC patients, 69% were correctly predicted as metastatic (and therefore considered at high relapse risk). This suggests that even in the absence of clinically visible metastatic spread, the metabolomic fingerprint of individuals with early CRC, who may have cancer recurrence, has a potential “metastatic signature.”

Furthermore, when the metabolomic classification of all patients with early CRC was analyzed using Kaplan–Meier curves, a strong prognostic effect was observed, with patients classified at high risk by metabolomics (thus with a metabolomic profile more similar to that of metastatic CRC patients) having a significantly higher probability of disease relapse than those with a low risk metabolomic score (High vs. Low risk: Hazard Ratio = 3.68, p-value = 0.001). Histidine and glutamine were shown to be significantly decreased in the serum of metastatic CRC patients. Previous evidences already suggested an association of glutamine with cancer progression and poor cancer-specific survival (Bertini et al. 2012; Sirniö et al. 2019). Also, the downregulation of histidine was already observed in other studies (Tan et al. 2013; Zhu et al. 2014) and its alteration was correlated with an increased activity of histidine decarboxylase.

Our research group, in another recent paper (Vignoli et al. 2021a), investigated the CRC relapse in the serum samples of a group of pre-chemotherapy CRC patients undergoing surgery. We demonstrated that several differences between post- and pre-operative serum samples are associated with cancer recurrence, in particular an increment of HDL-Chol subfractions coupled with a decrement of VLDL-Chol subfractions. These results may corroborate the hypothesis that the development of CRC disrupts the physiological equilibrium of lipids and lipoproteins, leading to lipid metabolic disorders (Zhang et al. 2014).

Urine samples have been also used for an NMR-based metabolomic approach to predict relapse in a group of 62 CRC patients, yielding an area under the ROC curve of 0.650 for cancer recurrence (Dykstra et al. 2017). Interestingly, here the authors

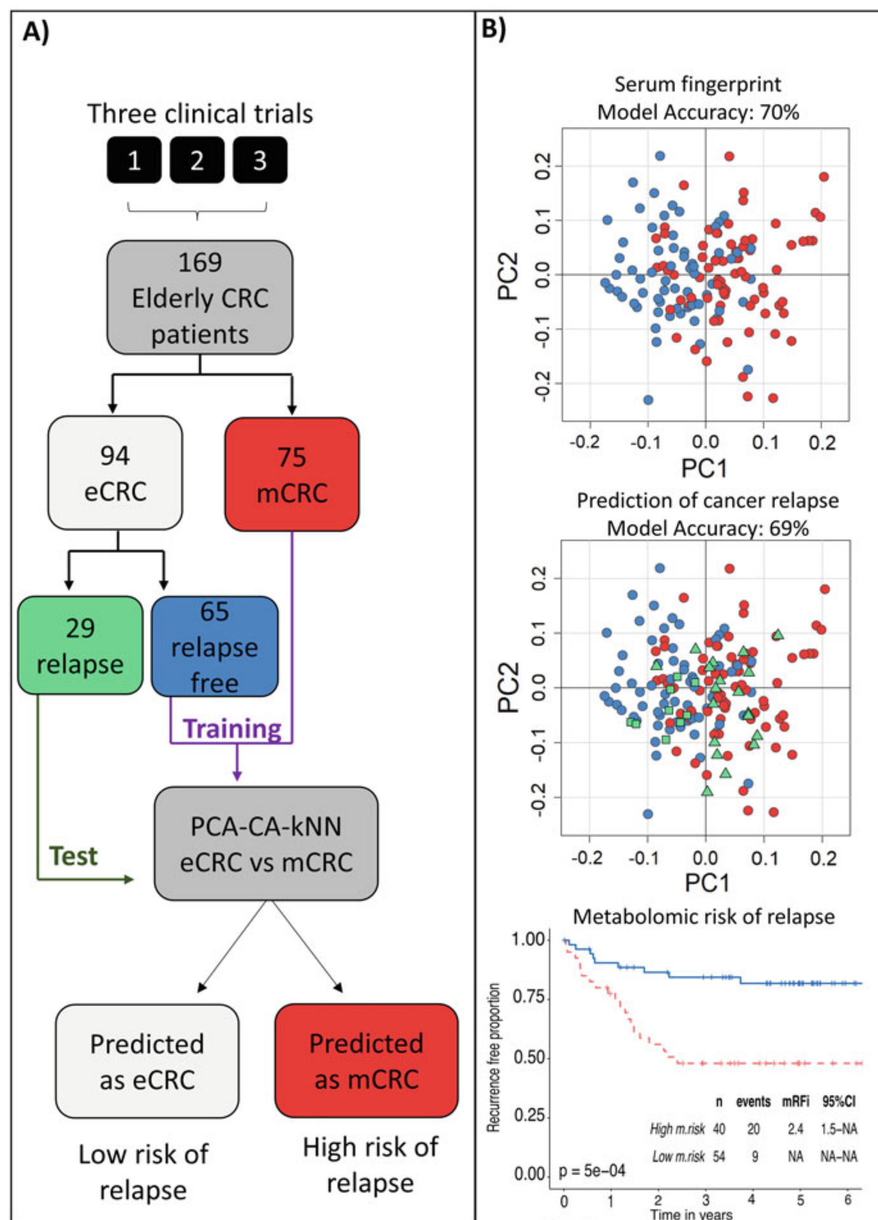


Fig. 5 Experimental scheme (**a**) and results (**b**) extracted from Di Donato et al. (2021). Dots in the score plots of the multivariate supervised PCA-CA kNN model represent the fingerprint of the NMR spectra of each patient recruited for the study. Red dots code for metastatic CRC (mCRC) patients, blue dots for early CRC (eCRC) relapse-free patients, green triangles for eCRC relapsed patients predicted as "metastatic," while green squares represent the CRC relapsed predicted as early CRC free from relapse. The Kaplan–Meier plot on the bottom right was used by the authors to estimate the outcome distribution

used the NMR data to predict treatment delay which could depend on reaction to chemotherapy, reaching an area under the ROC curve values of 0.750.

Pre-Surgical Effects of Anesthesia

The biofluids used by the various studies have been shown to contain important information detectable with metabolomics for the characterization of CRC patients. However, it is important to use the right preanalytical procedures on the samples to obtain reliable results. We have highlighted (Ghini et al. 2015b) that the moment of blood sample collection can strongly influence the plasma metabolomic profile. Blood samples were taken from 70 CRC patients (40 non-metastatic and 30 with liver metastasis) preoperatively, both prior and after anesthesia administration. Anesthesia depresses the metabolisms in uneven ways, thus reducing the information content of the metabolic profile and hence reducing the discrimination capability of the method. Consequently, post-anesthesia samples are not very suitable for standard metabolomics studies.

2.3 Viral Infections

Viruses utilize and/or rewire the host metabolism. Therefore, metabolomics is an excellent tool to study the effect of viral infection, either in vitro using infected cell cultures, or ex vivo in biosamples from infected animals or humans. In terms of practical applicability, obvious limitations arise from biosecurity, which requires the evaluation of the viral load and infectious risk of the different biological matrices as a function of their nature and of the nature of the infective agent, as well as of the biocontainment level of the laboratory.

On the other hand, identifying the metabolic pathways utilized by a virus has the potential to help revealing drug targets, to monitor the response to antiviral agents but also to evaluate the effect of vaccine administration, although very little is available on this aspect.

Here, we report examples from our laboratory on the metabolomic characterization of viral hepatitis and of SARS-CoV-2 infection.

2.3.1 Hepatitis

Viral hepatitis is a global health issue that affects millions of individuals and is associated with a high fatality rate. Except for the hepatitis A virus (HAV), all hepatotropic viruses, including hepatitis B, C, D, and E viruses (HBV, HCV, HDV, and HEV), can induce chronic infections. HBV and HCV are the most common viral causes of liver disease. According to the updated estimate of the WHO, there are about 300 million people suffering from chronic HBV and six million people suffering from chronic HCV (World Health Organization 2021).

Viral hepatitis chronic infection can cause progressive liver damage leading to fibrosis and cirrhosis. Cirrhosis is the end-stage of every chronic liver disease and is the major risk factor for hepatocellular carcinoma (HCC). HBV and HCV are the

leading cause of HCC worldwide, accounting for a significant mortality of more than 1.3 million death per year (Ringehan et al. 2017).

Prevention campaigns are one of the main weapons to limit the incidence of these viruses, especially for HCV for which a vaccine is not yet available. Direct-acting antivirals (DAAs) were approved in 2014, revolutionizing HCV therapy and allowing almost all patients to be cured. DAAs are very effective and well-tolerated and they constitute the gold standard for the treatment of HCV chronic infection in patients at all stages of liver disease.

Metabolomics analysis is being utilized to better understand host–pathogen interactions and screen host biospecimen for biomarkers that are characteristic of the viral infection. In 2019, an interesting picture emerged of the metabolomic fingerprint of HCV infection compared to both healthy subjects (HS) and HBV-infected patients (Meoni et al. 2019), suggesting that the two viruses exert a different impact on human metabolism. Indeed, by the comparison of the ^{1}H-NMR serum profiles of HCV- and HBV-infected patients, we identified characteristic metabolomic fingerprints of the two viral infections, obtaining an overall discrimination accuracy of 86% (OPLS-DA algorithm). As expected, the serum fingerprint of HCV- and HBV-infected patients resulted to be extremely different also from the serum fingerprint of HS, with a classification accuracy of 98.7% in the model built to discriminate HCV vs. HS, and a classification accuracy of 80% in the model built of HBV vs. HS (Fig. 6). Similarly, Godoy et al. (2010), using the ^{1}H-NMR fingerprinting approach, were able to accurately discriminate (95% predictive accuracy) the urine samples of HCV-infected patients from those of HCV-negative subjects, corroborating the potential of ^{1}H-NMR fingerprinting for the fast, non-invasive diagnosis of HCV infection using a urine sample.

The common changes we detected in the serum metabolomic profile of HBV- and HCV-infected patients when compared to HS (e.g., increased levels of 3-hydroxybutyrate, acetate, lactate, and pyruvate) support the hypothesis that these viruses preferentially stimulate glycolysis over oxidative phosphorylation, analogously to the Warburg effect in cancer (Okuda et al. 2002). Instead, in the comparison of HCV- and HBV-infected patients we noted a different behavior of several metabolites, suggesting that the perturbation could be attributable to a direct action of the two types of viruses rather than to the host response. Interestingly, the higher levels of 2-oxoglutarate and 3-hydroxybutyrate in HCV patients compared to HBV, also identified as biomarkers of cardiovascular disease and ketoacidosis, could explain why some extrahepatic manifestations, such as cardiovascular diseases and diabetes, are common in patients with chronic HCV and not in patients with HBV (Bernini et al. 2011a, b; Chen et al. 2014; Du et al. 2014).

Metabolomics proved useful also for disease staging and for characterizing the response to treatment. Anti-HCV treatment has advanced significantly in recent years, with direct-acting antivirals (DAAs) replacing pegylated interferon and ribavirin and providing effective treatment and less adverse effects. In our work, we characterized also the metabolomic fingerprint and the profile of HCV patients before and after effective DAA treatment. In this case, we identified a major contribution of the low-molecular weight molecules in characterizing the changes

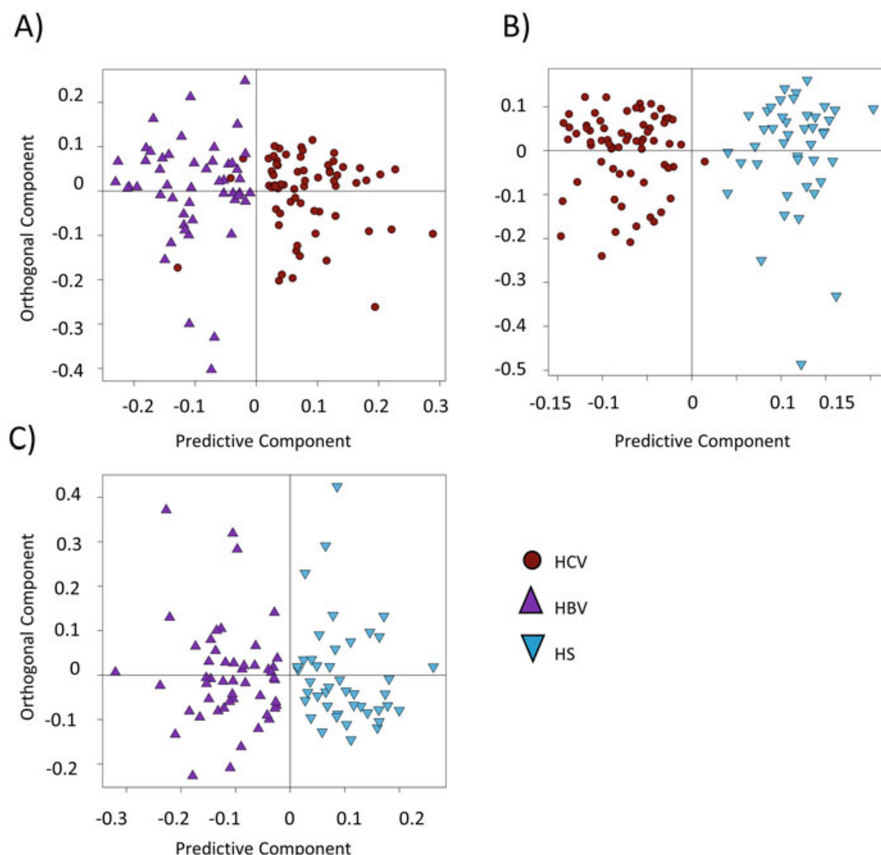


Fig. 6 Fingerprinting of the Hepatitis C and B viruses in serum samples resulting from OPLS-DA models. (**a**) HCV- (red dots) vs HBV- (purple triangles) infected patients; (**b**) HCV-infected patients (red dots) vs HS (sky-blue triangles); (**c**) HBV-infected patients (purple triangles) vs HS (sky-blue triangles). Adapted from Meoni et al. (2019)

introduced in the individual metabolomic profile by the therapy, suggesting also that the perturbation in lipid metabolism induced by the infection persists after viral eradication.

According to other studies (Cano et al. 2017; Nguyen et al. 2021; Sarfaraz et al. 2016) tyrosine and formate levels increase on passing from no/mild fibrosis to severe fibrosis. Furthermore, differences in metabolite levels between patients with higher and lower fibrosis scores were reduced after DAAs therapy, confirming that altered metabolites are restored, most likely due to liver damage regression after viral eradication.

Regarding the effect of treatment on sera of HBV-infected patients, Nguyen et al. 2021 recently published a study confirming that NMR-based metabolomics is capable of revealing in serum samples a gradual metabolic transition from

pretreatment to early treatment and then to a longer treatment period, as well as of accurately distinguishing the serum of patients who needed medical treatment (patients who would commence treatment within 6 months from sampling) from those who did not.

2.3.2 COVID-19

The COVID-19 pandemic has heavily reshaped research activities worldwide. Along with an incredibly fast development of vaccines that could bring the pandemic under control, efforts were also directed toward the development of antiviral drugs and monoclonal antibodies. Additionally, the pandemic has led to a rush to repurpose existing drugs. The pandemic has also stimulated researchers operating in fields complementary to pharmacology and vaccinology to contribute to the understanding of the physiopathology of the disease and to the characterization of risk factors and response to treatments. Coordinated efforts have been searched via the creation of networks that facilitate discussion among participants and communication of the key results prior to publication, together with the establishment of strategic transnational collaborations. Worth mentioning in this context are: (1) the Covid19-NMR project (https://covid19-nmr.de/), dealing with the determination of the structures of RNA and proteins of SARS-CoV-2 to investigate their drugability by small molecules; (2) the NMR international COVID-19 Research Network (CV19 Research Network), a metabolomics-based initiative consisting of several institutions that collaborate, using standardized NMR procedures, to detect the infection, predict outcomes during hospitalization, and direct efforts toward Long COVID.

Suitable biofluids from COVID-19 patients for metabolomics are serum, plasma, and urine, thanks to their low viral load, albeit collection of research samples during the worse phases of the COVID-19 pandemic was a further burden for the clinicians. In this frame, we present here some activities from our lab directed toward the use of NMR-based metabolomics and lipoproteomics to characterize the COVID-19 metabolomic fingerprints, to monitor the effect of repurposed drugs and vaccination follow-up (Ghini et al. 2022a; Meoni et al. 2021).

We and others (Ghini et al. 2022b; Ballout et al. 2021; Baranovicova et al. 2021; Bizkarguenaga et al. 2021; Bruzzone et al. 2020; Julkunen et al. 2021; Kimhofer et al. 2020; Lodge et al. 2021; Masuda et al. 2021) have shown that SARS-CoV-2 infection induces profound changes in the metabolic phenotype of the patients (Millet, "Prospective metabolomic studies in precision medicine. The AKRIBEA project."; Rogers, "The Metabolomics of Critical Illness"). Accordingly, ^{1}H NMR spectra of plasma samples of COVID-19 patients could be strongly discriminated from the spectra of both healthy subjects and COVID-19-recovered subjects, with a discrimination accuracy higher than 90% in both cases. The differences originate from significant alterations in the concentrations of several metabolites and of a panel of lipoprotein components. The metabolites and lipoprotein parameters that are significantly dysregulated in COVID-19 acute subjects are listed in Fig. 7.

Characteristic trends in metabolite and lipoprotein levels are also observed as a function of the disease severity (Ghini et al. 2022b). The analysis of the specific changes and correlations with clinical data enabled the identification of potential

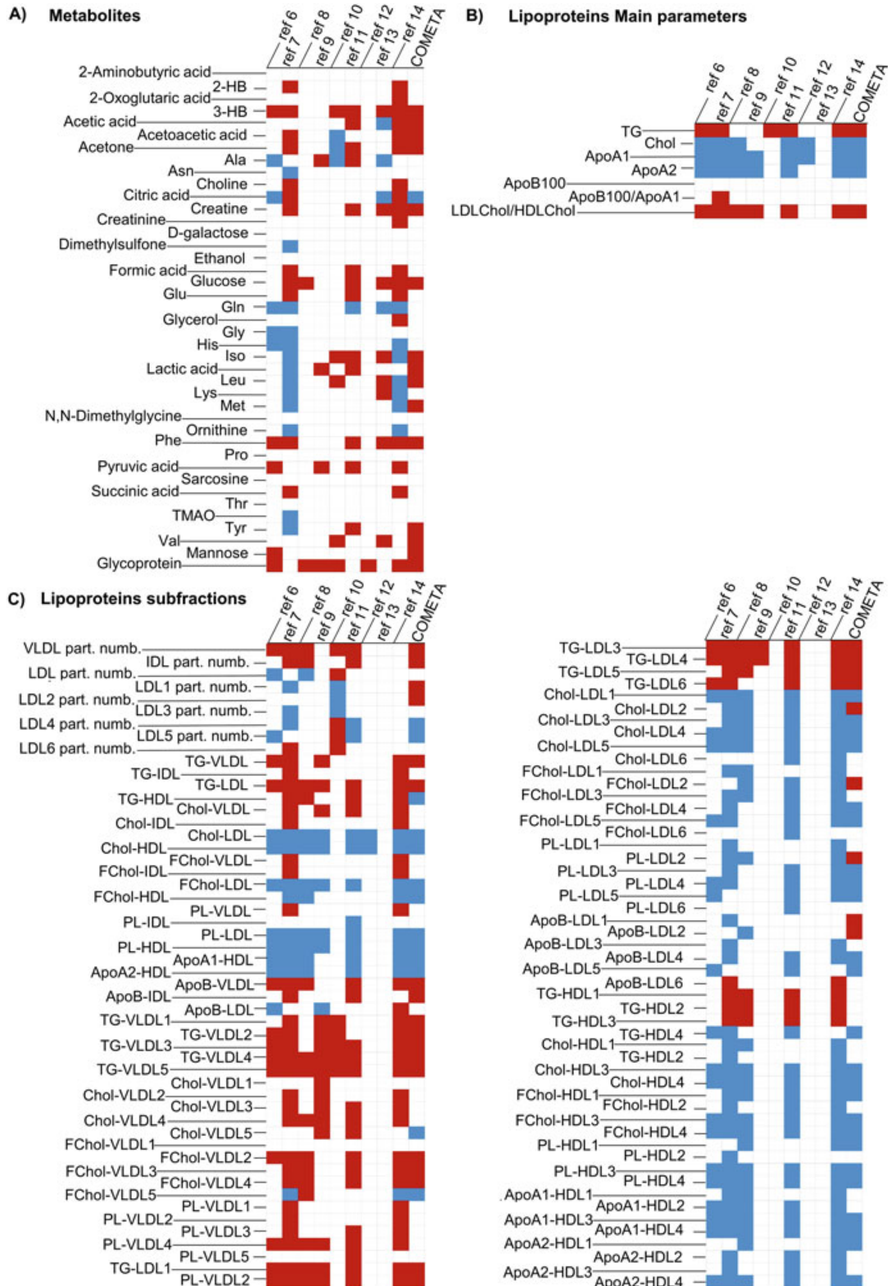


Fig. 7 Metabolites and lipoprotein parameters that are found significantly up- or down-regulated in COVID-19 patients with respect to healthy controls according to different Metabolomics papers (Meoni et al. 2021; Bruzzone et al. 2020; Kimhofer et al. 2020; Lodge et al. 2021; Ballout et al. 2021; Masuda et al. 2021; Julkunen et al. 2021; Baranovicova et al. 2021; Bruzzone et al. 2020) and COMETA project (Ghini et al. 2022b). Up-/down-regulated features are indicated by red/blue cells. (**a**) Metabolites; (**b**) Lipoprotein Main Parameters; (**c**) Lipoprotein Subfractions. Adapted from (Ghini et al. 2022b)

biochemical determinants of the disease fingerprint, which found confirmation in other studies performed at different centers worldwide and in some cases dealing with much larger cohorts. The parameters that are found altered in COVID-19 patients with respect to recovered individuals overlap with the acute infection biomarkers identified in the comparison with healthy subjects, indicating the substantial metabolic healing of COVID-19-recovered subjects. During the healing process, the metabolome and lipoproteome revert back to the “healthy” state with different rates; during either spontaneous healing or pharmacological treatments the metabolites are reverted faster than lipoproteins. Notably, several other metabolomics papers have been published, which identify common molecular features as characteristic of the COVID-19 profile. High convergence on common biomarkers from different metabolomics studies is not so common and denotes the presence of a strong profile, independent of confounding factors like place of origin, sex, age, and comorbidities.

Regarding the pharmacological treatments, we had the chance to analyze the effect of tocilizumab on a very small cohort (8 patients) treated at the Florence University hospital during the first wave of COVID-19 in spring 2020. Tocilizumab is a monoclonal antibody that attaches to the receptor of the cytokine interleukin-6, whose levels are elevated in response to systemic inflammation and plays an important role in severe COVID-19 disease and associated respiratory failure. On December 6th 2021, the European Medicines Agency (EMA) recommended extending the indication of RoActemra (tocilizumab) to include the treatment of adult COVID-19 patients who are receiving systemic treatment with corticosteroids and require supplemental oxygen or mechanical ventilation.

When measuring the post-treatment levels of metabolites and lipoproteins that are significantly altered by the infection, we found that eight metabolites (namely, acetone, citrate, glutamine, glycine, lactate, mannose, phenylalanine, and pyruvate) partially or completely revert toward the levels of CTR subjects. Some lipoprotein main- and sub-fractions were also significantly affected by the tocilizumab treatment, but, they did not revert back to “healthy” values, in line with the above-reported observation that the recovery of the lipoproteome is slower than that of the metabolome along the healing process (Meoni et al. 2021).

Finally, ^{1}H NMR spectra of sera have also been used to define the changes induced by vaccination with Pfizer-BioNTech vaccine in a cohort of 20 healthcare workers, 10 COVID-19 naïve, and 10 with a previous history of COVID-19 infection (infected in the period March–April 2020 with the Wuhan strain and recovered from the disease 208–280 days before vaccination). All of them received two doses of vaccine, 21 days apart, and their serum samples were collected at 6 different time points to monitor time-dependent changes induced by the vaccination. Importantly, the vaccination does not induce a major modification of the metabolic phenotype of the subjects; the intra-individual differences remain smaller than the inter-individual ones during all the course of the study, with an individual discrimination accuracy $>85\%$ (considering the six samples collected for each subject). Nevertheless, in response to vaccination we could observe some common changes that are consistently occurring in all subjects within each group. While vaccination does not induce any significant variation in the metabolome, it causes changes at the level of

lipoproteins that are smaller for COVID-19-recovered subjects with respect to naïve subjects, suggesting that a previous infection reduces the vaccine modulation of the lipoproteome composition. The differences between the two groups involve the nature and number of affected lipoprotein parameters. Additionally, the effect of the second dose is essentially negligible for the COVID-19-recovered subjects (Ghini et al. 2022a).

2.4 Chronic Obstructive Pulmonary Disease

As highlighted in chapter (Lacy, "Metabolomics of Respiratory Diseases"), chronic obstructive pulmonary disease (COPD) is a pathological condition characterized by the chronic, poorly reversible, and progressive development of airflow limitation often associated with parenchymal destruction and emphysema (Barnes 2000).

Although abnormal respiratory inflammation is crucial for COPD development, the complex COPD pathophysiology is not yet fully understood. Currently, no validated biomarker is accepted for disease prognosis or COPD therapy monitoring (Ghosh et al. 2016).

Several studies have identified serum, exhaled breath condensate (EBC), and urine ^{1}H NMR-based metabolomic fingerprints of COPD patients, also showing the ability of EBC metabolomics to assess airway inflammation (Airoldi et al. 2016; de Laurentiis et al. 2008; Motta et al. 2012; Wang et al. 2013; Ząbek et al. 2015). In this frame, a study from our group (Bertini et al. 2014) showed that NMR metabolomics of EBC could discriminate COPD patients from controls with an overall accuracy of 86%. As compared to controls, EBC from COPD patients featured significantly lower levels of acetone, valine, and lysine, and significantly higher levels of lactate, acetate, propionate, serine, proline, and tyrosine. Lower levels of valine and lysine (two essential amino acids) appear consistent with muscle wasting and weight loss that are known to occur in advanced COPD (Agusti et al. 2002). The hypothesis of a possible presence of subclinical malnutrition in COPD is also discussed in chapter (Lacy, "Metabolomics of Respiratory Diseases"), where the reader can find a more comprehensive overview of the metabolic alterations observed in sera of COPD patients.

The application of pharmacometabolomics in the context of COPD is aimed at identifying the appropriate treatment for each COPD patient, predicting his/her response to therapy. Our research group applied an original and holistic approach, dubbed "breathomics," to monitor the effects of treatment with and withdrawal from inhaled beclometasone/formoterol in patients with COPD (Montuschi et al. 2018). In our application, breathomics combined two electronic noses (carbon polymer sensor e-nose, quartz crystal sensor e-nose), EBC NMR-based metabolomics, sputum cell counts, sputum supernatant and EBC prostaglandin E2 (PGE2) and 15-F2t-isoprostane, fraction of exhaled nitric oxide, and spirometry data. Breathomics improves the identification of pharmacological treatment-induced effects as compared with standard spirometry. Furthermore, this approach provides insights into the anti-inflammatory effects of inhaled corticosteroids in COPD patients as reflected by reduced levels of sputum PGE2 and EBC acetate during

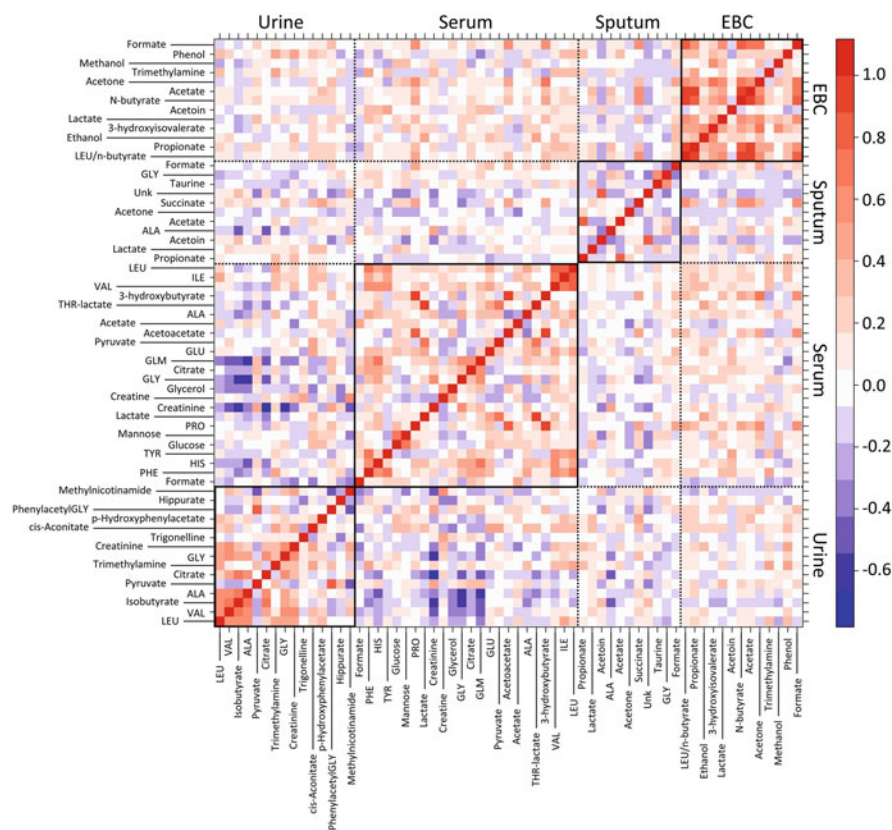


Fig. 8 Heatmap showing correlations among metabolites quantified in the urine, serum, exhaled breath condensate (EBC), and sputum supernatant samples of 14 patients with COPD. Correlation values (R) are reported as different degrees of color intensity (blue, negative correlation; red, positive correlations). The figure is adapted and reprinted with permission from Vignoli et al. (2020b). Copyright 2020 American Chemical Society

treatment with formoterol alone. This research line was further explored with the analysis of urine, serum and sputum supernatant, demonstrating that different biological matrices provide complementary information on the effects of beclometasone/formoterol administration in COPD patients, and thus their integration could be useful for elucidating the metabolic mechanism of action of inhaled corticosteroids (Vignoli et al. 2020b). An overview of the metabolite correlation patterns among the different biofluids is presented in Fig. 8.

2.5 Drug Toxicity

Pharmacometabolomics includes drug safety evaluation that is an important step in the drug pipeline and a main concern for regulatory agencies. Safety evaluation is

required across the whole process of drug development, from preclinical studies to clinical trials, as well as in post-approval safety surveillance. Biochemical and histological analyses are the major approaches used for drug safety evaluation. These approaches are effective in most cases to determine the safety profile of drug candidates (Wang et al. 2017). However, these methods can neither provide detailed information nor explore the mechanisms of drug toxicity (Nicholson et al. 2002). Metabolomics is ideally positioned to address the challenges of drug toxicity (Hertz, "Chemotherapy-induced Peripheral Neuropathy"). It represents a powerful tool for collecting mechanistic information, indicating not only the extent of a toxic insult but also its underlying mechanisms (Ramirez et al. 2013). Several examples exist in the literature of NMR-based metabolomic applications in toxicology and drug safety evaluation. Just to cite few of them, drugs and toxicants like flutamide (Choucha Snouber et al. 2013), hydrazine (Garrod et al. 2005; Lindon et al. 2003), and gentamicin (Lenz et al. 2005) have been assessed.

A typical example of a clinically useful class of molecules that presents high rates of therapy discontinuation due to acute side effects is represented by V-phosphodiesterase (PDE5) inhibitors (PDE5i) (Corona et al. 2016). These molecules are the first-line therapy for erectile dysfunction (ED), a widespread health problem in the general population of middle-aged men (Ayta et al. 1999; Yafi et al. 2016). Administration of PDE5i proved beneficial in 60–70% of patients with varying etiologies of sexual dysfunction (Yafi et al. 2016). However, adverse effects are a common drawback. In a study from our research group, we retrospectively evaluated serum and urine NMR-based metabolomic profiles to identify prognostic biomarkers of unfavorable efficacy/safety profile of PDE5i before drug administration. To the best of the authors' knowledge, this is the first and only NMR-based metabolomic study focused on PDE5i toxicity. Patients who are likely to experience adverse effects can be identified with an accuracy of 77% using pretreatment serum samples. Adverse drug reactions showed to be associated with high levels of LDL-lipoprotein subfractions at baseline (Rocca et al. 2020). The results of this pilot study underline how metabolomics may help in identifying the metabolic bases underlying efficacy/safety profile of the PDE5i therapy.

References

Agusti AG, Sauleda J, Miralles C, Gomez C, Togores B, Sala E, Batle S, Busquets X (2002) Skeletal muscle apoptosis and weight loss in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 166:485–489. https://doi.org/10.1164/rccm.2108013

Airoldi C, Ciaramelli C, Fumagalli M, Bussei R, Mazzoni V, Viglio S, Iadarola P, Stolk J (2016) 1H NMR To explore the metabolome of exhaled breath condensate in α1-antitrypsin deficient patients: a pilot study. J Proteome Res 15:4569–4578. https://doi.org/10.1021/acs.jproteome.6b00648

Andersen M-BS, Rinnan Å, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, Astrup A, Dragsted LO (2014) Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. J Proteome Res 13:1405–1418. https://doi.org/10.1021/pr400964s

Antoniewicz MR (2018) A guide to 13C metabolic flux analysis for the cancer biologist. Exp Mol Med 50:1–13. https://doi.org/10.1038/s12276-018-0060-y

Asiago VM, Alvarado LZ, Shanaiah N, Gowda GAN, Owusu-Sarfo K, Ballas RA, Raftery D (2010) Early detection of recurrent breast cancer using metabolite profiling. Cancer Res 70: 8309–8318. https://doi.org/10.1158/0008-5472.CAN-10-1319

Assfalg M, Bertini I, Colangiuli D, Luchinat C, Schäfer H, Schütz B, Spraul M (2008) Evidence of different metabolic phenotypes in humans. Proc Natl Acad Sci U S A 105:1420–1424. https://doi.org/10.1073/pnas.0705685105

Ayta IA, McKinlay JB, Krane RJ (1999) The likely worldwide increase in erectile dysfunction between 1995 and 2025 and some possible policy consequences. BJU Int 84:50–56. https://doi.org/10.1046/j.1464-410x.1999.00142.x

Ballout RA, Kong H, Sampson M, Otvos JD, Cox AL, Agbor-Enoh S, Remaley AT (2021) The NIH Lipo-COVID study: a pilot NMR investigation of lipoprotein subfractions and other metabolites in patients with severe COVID-19. Biomedicine 9:1090. https://doi.org/10.3390/biomedicines9091090

Baranovicova E, Bobcakova A, Vysehradsky R, Dankova Z, Halasova E, Nosal V, Lehotsky J (2021) The ability to normalise energy metabolism in advanced COVID-19 disease seems to be one of the key factors determining the disease progression – a metabolomic NMR study on blood plasma. Appl Sci 11:4231. https://doi.org/10.3390/app11094231

Barnes PJ (2000) Chronic obstructive pulmonary disease. N Engl J Med 343:269–280. https://doi.org/10.1056/NEJM200007273430407

Bendinelli B, Vignoli A, Palli D, Assedi M, Ambrogetti D, Luchinat C, Caini S, Saieva C, Turano P, Masala G (2021) Prediagnostic circulating metabolites in female breast cancer cases with low and high mammographic breast density. Sci Rep 11:13025. https://doi.org/10.1038/s41598-021-92508-1

Bernini P, Bertini I, Calabrò A, la Marca G, Lami G, Luchinat C, Renzi D, Tenori L (2011a) Are patients with potential celiac disease really potential? The answer of metabonomics. J Proteome Res 10:714–721. https://doi.org/10.1021/pr100896s

Bernini P, Bertini I, Luchinat C, Tenori L, Tognaccini A (2011b) The cardiovascular risk of healthy individuals studied by NMR metabonomics of plasma samples. J Proteome Res 10:4983–4992. https://doi.org/10.1021/pr200452j

Bertini I, Cacciatore S, Jensen BV, Schou JV, Johansen JS, Kruhøffer M, Luchinat C, Nielsen DL, Turano P (2012) Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. Cancer Res 72:356–364. https://doi.org/10.1158/0008-5472-11-1543

Bertini I, Calabrò A, De Carli V, Luchinat C, Nepi S, Porfirio B, Renzi D, Saccenti E, Tenori L (2009) The metabonomic signature of celiac disease. J Proteome Res 8:170–177. https://doi.org/10.1021/pr800548z

Bertini I, Luchinat C, Miniati M, Monti S, Tenori L (2014) Phenotyping COPD by H-1 NMR metabolomics of exhaled breath condensate. Metabolomics 10:302–311. https://doi.org/10.1007/s11306-013-0572-3

Bizkarguenaga M, Bruzzone C, Gil-Redondo R, SanJuan I, Martin-Ruiz I, Barriales D, Palacios A, Pasco ST, González-Valle B, Laín A, Herrera L, Azkarate A, Vesga MA, Eguizabal C, Anguita J, Embade N, Mato JM, Millet O (2021) Uneven metabolic and lipidomic profiles in recovered COVID-19 patients as investigated by plasma NMR metabolomics. NMR Biomed 35:e4637. https://doi.org/10.1002/nbm.4637

Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, Barlesi F, Kohlhäufl M, Arrieta O, Burgio MA, Fayette J, Lena H, Poddubskaya E, Gerber DE, Gettinger SN, Rudin CM, Rizvi N, Crinò L, Blumenschein GR, Antonia SJ, Dorange C, Harbison CT, Graf Finckenstein F, Brahmer JR (2015) Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med 373:1627–1639. https://doi.org/10.1056/NEJMoa1507643

Brahmer JR (2014) Immune checkpoint blockade: the hope for immunotherapy as a treatment of lung cancer? Semin Oncol 41:126–132. https://doi.org/10.1053/j.seminoncol.2013.12.014

Bruzzone C, Bizkarguenaga M, Gil-Redondo R, Diercks T, Arana E, García de Vicuña A, Seco M, Bosch A, Palazón A, San Juan I, Laín A, Gil-Martínez J, Bernardo-Seisdedos G, Fernández-Ramos D, Lopitz-Otsoa F, Embade N, Lu S, Mato JM, Millet O (2020) SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. iScience 23:101645. https://doi.org/10.1016/j.isci.2020.101645

Buchwald H, Estok R, Fahrbach K, Banel D, Jensen MD, Pories WJ, Bantle JP, Sledge I (2009) Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. Am J Med 122:248–256.e5. https://doi.org/10.1016/j.amjmed.2008.09.041

Cacciatore S, Hu X, Viertler C, Kap M, Bernhardt GA, Mischinger H-J, Riegman P, Zatloukal K, Luchinat C, Turano P (2013) Effects of intra- and post-operative ischemia on the metabolic profile of clinical liver tissue specimens monitored by NMR. J Proteome Res 12:5723–5729. https://doi.org/10.1021/pr400702d

Cala MP, Aldana J, Medina J, Sánchez J, Guio J, Wist J, Meesters RJW (2018) Multiplatform plasma metabolic and lipid fingerprinting of breast cancer: a pilot control-case study in Colombian Hispanic women. PLoS One 13:e0190958. https://doi.org/10.1371/journal.pone.0190958

Cano A, Mariño Z, Millet O, Martínez-Arranz I, Navasa M, Falcón-Pérez JM, Pérez-Cormenzana M, Caballería J, Embade N, Forns X, Bosch J, Castro A, Mato JM (2017) A metabolomics signature linked to liver fibrosis in the serum of transplanted hepatitis C patients. Sci Rep 7:10497. https://doi.org/10.1038/s41598-017-10807-y

Chen P-A, Xu Z-H, Huang Y-L, Luo Y, Zhu D-J, Wang P, Du Z-Y, Yang Y, Wu D-H, Lai W-Y, Ren H, Xu D-L (2014) Increased serum 2-oxoglutarate associated with high myocardial energy expenditure and poor prognosis in chronic heart failure patients. Biochim Biophys Acta 1842: 2120–2125. https://doi.org/10.1016/j.bbadis.2014.07.018

Choi JS, Baek H-M, Kim S, Kim MJ, Youk JH, Moon HJ, Kim E-K, Nam YK (2013) Magnetic resonance metabolic profiling of breast cancer tissue obtained with core needle biopsy for predicting pathologic response to neoadjuvant chemotherapy. PLoS One 8:e83866. https://doi.org/10.1371/journal.pone.0083866

Choucha Snouber L, Bunescu A, Naudot M, Legallais C, Brochot C, Dumas ME, Elena-Herrmann-B, Leclerc E (2013) Metabolomics-on-a-chip of hepatotoxicity induced by anticancer drug flutamide and its active metabolite hydroxyflutamide using HepG2/C3a microfluidic biochips. Toxicol Sci 132:8–20. https://doi.org/10.1093/toxsci/kfs230

Corona G, Di Gregorio E, Vignoli A, Muraro E, Steffan A, Miolo G (2021) 1H-NMR plasma lipoproteins profile analysis reveals lipid metabolism alterations in HER2-positive breast cancer patients. Cancers 13:5845. https://doi.org/10.3390/cancers13225845

Corona G, Rastrelli G, Burri A, Serra E, Gianfrilli D, Mannucci E, Jannini EA, Maggi M (2016) First-generation phosphodiesterase type 5 inhibitors dropout: a comprehensive review and meta-analysis. Andrology 4:1002–1009. https://doi.org/10.1111/andr.12255

Cunningham D, Atkin W, Lenz H-J, Lynch HT, Minsky B, Nordlinger B, Starling N (2010) Colorectal cancer. Lancet 375:1030–1047. https://doi.org/10.1016/S0140-6736(10)60353-4

Cuperlovic-Culf M, Ferguson D, Culf A, Morin P, Touaibia M (2012) 1H NMR metabolomics analysis of glioblastoma subtypes correlation between metabolomics and gene expression characteristics. J Biol Chem 287:20164–20175. https://doi.org/10.1074/jbc.M111.337196

D'Alessandro G, Quaglio D, Monaco L, Lauro C, Ghirga F, Ingallina C, De Martino M, Fucile S, Porzia A, Di Castro MA, Bellato F, Mastrotto F, Mori M, Infante P, Turano P, Salmaso S, Caliceti P, Di Marcotullio L, Botta B, Ghini V, Limatola C (2019) 1H-NMR metabolomics reveals the Glabrescione B exacerbation of glycolytic metabolism beside the cell growth inhibitory effect in glioma. Cell Commun Signal 17:108. https://doi.org/10.1186/s12964-019-0421-8

Debik J, Euceda LR, Lundgren S, Gythfeldt H v d L, Garred Ø, Borgen E, Engebraaten O, Bathen TF, Giskeødegård GF (2019) Assessing treatment response and prognosis by serum and tissue metabolomics in breast cancer patients. J Proteome Res 18:3649–3660. https://doi.org/10.1021/acs.jproteome.9b00316

Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB (2019) Colorectal cancer. Lancet 394: 1467–1480. https://doi.org/10.1016/S0140-6736(19)32319-0

Di Donato S, Vignoli A, Biagioni C, Malorni L, Mori E, Tenori L, Calamai V, Parnofiello A, Di Pierro G, Migliaccio I, Cantafio S, Baraghini M, Mottino G, Becheri D, Del Monte F, Miceli E, McCartney A, Di Leo A, Luchinat C, Biganzoli L (2021) A serum metabolomics classifier derived from elderly patients with metastatic colorectal cancer predicts relapse in the adjuvant setting. Cancers 13:2762. https://doi.org/10.3390/cancers13112762

Díaz-Beltrán L, González-Olmedo C, Luque-Caro N, Díaz C, Martín-Blázquez A, Fernández-Navarro M, Ortega-Granados AL, Gálvez-Montosa F, Vicente F, Pérez del Palacio J, Sánchez-Rovira P (2021) Human plasma metabolomics for biomarker discovery: targeting the molecular subtypes in breast cancer. Cancers 13:147. https://doi.org/10.3390/cancers13010147

Du Z, Shen A, Huang Y, Su L, Lai W, Wang P, Xie Z, Xie Z, Zeng Q, Ren H, Xu D (2014) 1H-NMR-based metabolic analysis of human serum reveals novel markers of myocardial energy expenditure in heart failure patients. PLoS One 9:e88102. https://doi.org/10.1371/journal.pone.0088102

Dykstra MA, Switzer N, Eisner R, Tso V, Foshaug R, Ismond K, Fedorak R, Wang H (2017) Urine metabolomics as a predictor of patient tolerance and response to adjuvant chemotherapy in colorectal cancer. Mol Clin Oncol 7:767–770. https://doi.org/10.3892/mco.2017.1407

Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365:1687–1717. https://doi.org/10.1016/S0140-6736(05)66544-0

Euceda LR, Haukaas TH, Giskeødegård GF, Vettukattil MR, Engel J, Silwal-Pandit L, Lundgren S, Borgen E, Garred Ø, Postma G, Buydens LMC, Børresen-Dale A-L, Engebråten O, Bathen TF (2017) Evaluation of metabolomic changes during neoadjuvant chemotherapy combined with bevacizumab in breast cancer using MR spectroscopy. Metabolomics 13:37. https://doi.org/10.1007/s11306-017-1168-0

Fan Y, Zhou X, Xia T-S, Chen Z, Li J, Liu Q, Alolga RN, Chen Y, Lai M-D, Li P, Zhu W, Qi L-W (2016) Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer. Oncotarget 7:9925–9938. https://doi.org/10.18632/oncotarget.7155

Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, Domine M, Clingan P, Hochmair MJ, Powell SF, Cheng SY-S, Bischoff HG, Peled N, Grossi F, Jennens RR, Reck M, Hui R, Garon EB, Boyer M, Rubio-Viqueira B, Novello S, Kurata T, Gray JE, Vida J, Wei Z, Yang J, Raftopoulos H, Pietanza MC, Garassino MC, KEYNOTE-189 Investigators (2018) Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 378:2078–2092. https://doi.org/10.1056/NEJMoa1801005

Garrod S, Bollard ME, Nicholls AW, Connor SC, Connelly J, Nicholson JK, Holmes E (2005) Integrated metabonomic analysis of the multiorgan effects of hydrazine toxicity in the rat. Chem Res Toxicol 18:115–122. https://doi.org/10.1021/tx0498915

Gettinger SN, Horn L, Gandhi L, Spigel DR, Antonia SJ, Rizvi NA, Powderly JD, Heist RS, Carvajal RD, Jackman DM, Sequist LV, Smith DC, Leming P, Carbone DP, Pinder-Schenck MC, Topalian SL, Hodi FS, Sosman JA, Sznol M, McDermott DF, Pardoll DM, Sankar V, Ahlers CM, Salvati M, Wigginton JM, Hellmann MD, Kollia GD, Gupta AK, Brahmer JR (2015) Overall survival and long-term safety of nivolumab (anti-programmed death 1 antibody, BMS-936558, ONO-4538) in patients with previously treated advanced non-small-cell lung cancer. J Clin Oncol 33:2004–2012. https://doi.org/10.1200/JCO.2014.58.3708

Ghini V, Di Nunzio M, Tenori L, Valli V, Danesi F, Capozzi F, Luchinat C, Bordoni A (2017) Evidence of a DHA signature in the lipidome and metabolome of human hepatocytes. Int J Mol Sci 18:359. https://doi.org/10.3390/ijms18020359

Ghini V, Laera L, Fantechi B, del Monte F, Benelli M, McCartney A, Tenori L, Luchinat C, Pozzessere D (2020a) Metabolomics to assess response to immune checkpoint inhibitors in patients with non-small-cell lung cancer. Cancers 12:3574. https://doi.org/10.3390/cancers12123574

Ghini V, Maggi L, Mazzoni A, Spinicci M, Zammarchi L, Bartoloni A, Annunziato F, Turano P (2022a) Serum NMR profiling reveals differential alterations in the lipoproteome induced by Pfizer-BioNTech vaccine in COVID-19 recovered subjects and Naïve subjects. Front Mol Biosci 9:839809. https://doi.org/10.3389/fmolb.2022.839809

Ghini V, Meoni G, Pelagatti L, Celli T, Veneziani F, Petrucci F, Vannucchi V, Bertini L, Luchinat C, Landini G, Turano P (2022b) Profiling metabolites and lipoproteins in COMETA, an Italian cohort of COVID-19 patients. PLoS Pathog 18:e1010443. https://doi.org/10.1371/journal.ppat.1010443

Ghini V, Quaglio D, Luchinat C, Turano P (2019) NMR for sample quality assessment in metabolomics. N Biotechnol 52:25–34. https://doi.org/10.1016/j.nbt.2019.04.004

Ghini V, Saccenti E, Tenori L, Assfalg M, Luchinat C (2015a) Allostasis and resilience of the human individual metabolic phenotype. J Proteome Res 14:2951–2962. https://doi.org/10.1021/acs.jproteome.5b00275

Ghini V, Senzacqua T, Massai L, Gamberi T, Messori L, Turano P (2021) NMR reveals the metabolic changes induced by auranofin in A2780 cancer cells: evidence for glutathione dysregulation. Dalton Trans 50:6349–6355. https://doi.org/10.1039/D1DT00750E

Ghini V, Tenori L, Pane M, Amoruso A, Marroncini G, Squarzanti DF, Azzimonti B, Rolla R, Savoia P, Tarocchi M, Galli A, Luchinat C (2020b) Effects of probiotics administration on human metabolic phenotype. Metabolites 10:396. https://doi.org/10.3390/metabo10100396

Ghini V, Unger FT, Tenori L, Turano P, Juhl H, David KA (2015b) Metabolomics profiling of pre-and post-anesthesia plasma samples of colorectal patients obtained via Ficoll separation. Metabolomics 11:1769–1778. https://doi.org/10.1007/s11306-015-0832-5

Ghosh N, Dutta M, Singh B, Banerjee R, Bhattacharyya P, Chaudhury K (2016) Transcriptomics, proteomics and metabolomics driven biomarker discovery in COPD: an update. Expert Rev Mol Diagn 16:897–913. https://doi.org/10.1080/14737159.2016.1198258

Gianni L, Pienkowski T, Im Y-H, Roman L, Tseng L-M, Liu M-C, Lluch A, Staroslawska E, de la Haba-Rodriguez J, Im S-A, Pedrini JL, Poirier B, Morandi P, Semiglazov V, Srimuninnimit V, Bianchi G, Szado T, Ratnayake J, Ross G, Valagussa P (2012) Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. Lancet Oncol 13:25–32. https://doi.org/10.1016/S1470-2045(11)70336-9

Godoy MGG, Lopes EPA, Silva RO, Hallwass F, Koury LCA, Moura IM, Gonçalves SMC, Simas AM (2010) Hepatitis C virus infection diagnosis using metabonomics. J Viral Hepat 17:854–858. https://doi.org/10.1111/j.1365-2893.2009.01252.x

Gralka E, Luchinat C, Tenori L, Ernst B, Thurnheer M, Schultes B (2015) Metabolomic fingerprint of severe obesity is dynamically affected by bariatric surgery in a procedure-dependent manner. Am J Clin Nutr 102:1313–1322. https://doi.org/10.3945/ajcn.115.110536

Griffiths WJ (2008) Metabolomics, metabonomics and metabolite profiling. RSC Publishing

Guraya SY (2019) Pattern, stage, and time of recurrent colorectal cancer after curative surgery. Clin Colorectal Cancer 18:e223–e228. https://doi.org/10.1016/j.clcc.2019.01.003

Hamada T, Kosumi K, Nakai Y, Koike K (2018) Surrogate study endpoints in the era of cancer immunotherapy. Ann Transl Med 6:S27. https://doi.org/10.21037/atm.2018.09.31

Hart CD, Vignoli A, Tenori L, Uy GL, To TV, Adebamowo C, Hossain SM, Biganzoli L, Risi E, Love RR, Luchinat C, Di Leo A (2017) Serum metabolomic profiles identify ER-positive early breast cancer patients at increased risk of disease recurrence in a multicenter population. Clin Cancer Res 23:1422–1431. https://doi.org/10.1158/1078-0432.CCR-16-1153

Hatae R, Chamoto K, Kim YH, Sonomura K, Taneishi K, Kawaguchi S, Yoshida H, Ozasa H, Sakamori Y, Akrami M, Fagarasan S, Masuda I, Okuno Y, Matsuda F, Hirai T, Honjo T (2020) Combination of host immune metabolic biomarkers for the PD-1 blockade cancer immunotherapy. JCI Insight 5:133501. https://doi.org/10.1172/jci.insight.133501

Herbst RS, Baas P, Kim D-W, Felip E, Pérez-Gracia JL, Han J-Y, Molina J, Kim J-H, Arvis CD, Ahn M-J, Majem M, Fidler MJ, de Castro G, Garrido M, Lubiniecki GM, Shentu Y, Im E, Dolled-Filhart M, Garon EB (2016) Pembrolizumab versus docetaxel for previously treated,

PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 387:1540–1550. https://doi.org/10.1016/S0140-6736(15)01281-7

Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Wu Y-L, Paz-Ares L (2017) Lung cancer: current therapies and new targeted treatments. Lancet 389:299–311. https://doi.org/10.1016/S0140-6736(16)30958-8

Holmes E, Loo RL, Stamler J, Bictash M, Yap IKS, Chan Q, Ebbels T, De Iorio M, Brown IJ, Veselkov KA, Daviglus ML, Kesteloot H, Ueshima H, Zhao L, Nicholson JK, Elliott P (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. Nature 453:396–400. https://doi.org/10.1038/nature06882

ISO 23118:2021. Molecular in vitro diagnostic examinations – specifications for pre-examination processes in metabolomics in urine, venous blood serum and plasma. https://www.iso.org/standard/74605.html

Jiang L, Lee SC, Ng TC (2018) Pharmacometabonomics analysis reveals serum formate and acetate potentially associated with varying response to gemcitabine-carboplatin chemotherapy in metastatic breast cancer patients. J Proteome Res 17:1248–1257. https://doi.org/10.1021/acs.jproteome.7b00859

Jobard E, Trédan O, Bachelot T, Vigneron AM, Aït-Oukhatar CM, Arnedos M, Rios M, Bonneterre J, Diéras V, Jimenez M, Merlin J-L, Campone M, Elena-Herrmann B (2017) Longitudinal serum metabolomics evaluation of trastuzumab and everolimus combination as pre-operative treatment for HER-2 positive breast cancer patients. Oncotarget 8:83570–83584. https://doi.org/10.18632/oncotarget.18784

Julkunen H, Cichońska A, Slagboom PE, Würtz P (2021) Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. Elife 10:e63033. https://doi.org/10.7554/eLife.63033

Kaddurah-Daouk R, Weinshilboum RM, Pharmacometabolomics Research Network (2014) Pharmacometabolomics: implications for clinical pharmacology and systems pharmacology. Clin Pharmacol Ther 95:154–167. https://doi.org/10.1038/clpt.2013.217

Kimhofer T, Lodge S, Whiley L, Gray N, Loo RL, Lawler NG, Nitschke P, Bong S-H, Morrison DL, Begum S, Richards T, Yeap BB, Smith C, Smith KCG, Holmes E, Nicholson JK (2020) Integrative modelling of quantitative plasma lipoprotein, metabolic and amino acid data reveals a multi-organ pathological signature of SARS-CoV-2 infection. J Proteome Res 19:4442–4454. https://doi.org/10.1021/acs.jproteome.0c00519

Klupczyńska A, Dereziński P, Kokot ZJ (2015) Metabolomics in medical sciences – trends, challenges and perspectives. Acta Pol Pharm 72:629–641

de Laurentiis G, Paris D, Melck D, Maniscalco M, Marsico S, Corso G, Motta A, Sofia M (2008) Metabonomic analysis of exhaled breath condensate in adults by nuclear magnetic resonance spectroscopy. Eur Respir J 32:1175–1183. https://doi.org/10.1183/09031936.00072408

Le Gall G, Guttula K, Kellingray L, Tett AJ, ten Hoopen R, Kemsley KE, Savva GM, Ibrahim A, Narbad A (2018) Metabolite quantification of faecal extracts from colorectal cancer patients and healthy controls. Oncotarget 9:33278–33289. https://doi.org/10.18632/oncotarget.26022

Lécuyer L, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson M-P, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srour B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Savarin P, Touvier M (2018) NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. Int J Epidemiol 47:484–494. https://doi.org/10.1093/ije/dyx271

Lenz EM, Bright J, Knight R, Westwood FR, Davies D, Major H, Wilson ID (2005) Metabonomics with 1H-NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. Biomarkers 10:173–187. https://doi.org/10.1080/13547500500094034

Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AFP (2003) A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal 33:1103–1115. https://doi.org/10.1016/S0731-7085(03)00410-2

Lerner HJ, Band PR, Israel L, Leung BS (1976) Phase II study of tamoxifen: report of 74 patients with stage IV breast cancer. Cancer Treat Rep 60:1431–1435

Li C, Li Z, Zhang T, Wei P, Li N, Zhang W, Ding X, Li J (2019) 1H NMR-based metabolomics reveals the antitumor mechanisms of triptolide in BALB/c mice bearing CT26 Tumors. Front Pharmacol 10:1175. https://doi.org/10.3389/fphar.2019.01175

Li Y, Wang C, Li D, Deng P, Shao X, Hu J, Liu C, Jie H, Lin Y, Li Z, Qian X, Zhang H, Zhao Y (2017) 1H-NMR-based metabolic profiling of a colorectal cancer CT-26 lung metastasis model in mice. Oncol Rep 38:3044–3054. https://doi.org/10.3892/or.2017.5954

Lin Y, Ma C, Bezabeh T, Wang Z, Liang J, Huang Y, Zhao J, Liu X, Ye W, Tang W, Ouyang T, Wu R (2019) 1 H NMR-based metabolomics reveal overlapping discriminatory metabolites and metabolic pathway disturbances between colorectal tumor tissues and fecal samples. Int J Cancer 145:1679–1689. https://doi.org/10.1002/ijc.32190

Lin Y, Ma C, Liu C, Wang Z, Yang J, Liu X, Shen Z, Wu R (2016) NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. Oncotarget 7: 29454–29464. https://doi.org/10.18632/oncotarget.8762

Lindon JC, Nicholson JK, Holmes E, Antti H, Bollard ME, Keun H, Beckonert O, Ebbels TM, Reily MD, Robertson D, Stevens GJ, Luke P, Breau AP, Cantor GH, Bible RH, Niederhauser U, Senn H, Schlotterbeck G, Sidelmann UG, Laursen SM, Tymiak A, Car BD, Lehman-McKeeman L, Colet J-M, Loukaci A, Thomas C (2003) Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. Toxicol Appl Pharmacol 187:137–146. https://doi.org/10.1016/S0041-008X(02)00079-0

Lodge S, Nitschke P, Kimhofer T, Coudert JD, Begum S, Bong S-H, Richards T, Edgar D, Raby E, Spraul M, Schaefer H, Lindon JC, Loo RL, Holmes E, Nicholson JK (2021) NMR spectroscopic windows on the systemic effects of SARS-CoV-2 infection on plasma lipoproteins and metabolites in relation to circulating cytokines. J Proteome Res 20:1382–1396. https://doi.org/10.1021/acs.jproteome.0c00876

Marianne C, Walsh LB (2006) Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. Am J Clin Nutr 84:531–539. https://doi.org/10.1093/ajcn/84.3.531

Márquez J, Matés JM (2021) Tumor metabolome: therapeutic opportunities targeting cancer metabolic reprogramming. Cancers (Basel) 13:314. https://doi.org/10.3390/cancers13020314

Martin M, Holmes FA, Ejlertsen B, Delaloge S, Moy B, Iwata H, von Minckwitz G, Chia SKL, Mansi J, Barrios CH, Gnant M, Tomašević Z, Denduluri N, Šeparović R, Gokmen E, Bashford A, Ruiz Borrego M, Kim S-B, Jakobsen EH, Ciceniene A, Inoue K, Overkamp F, Heijns JB, Armstrong AC, Link JS, Joy AA, Bryce R, Wong A, Moran S, Yao B, Xu F, Auerbach A, Buyse M, Chan A, ExteNET Study Group (2017) Neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): 5-year analysis of a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 18:1688–1700. https://doi.org/10.1016/S1470-2045(17)30717-9

Masuda R, Lodge S, Nitschke P, Spraul M, Schaefer H, Bong S-H, Kimhofer T, Hall D, Loo RL, Bizkarguenaga M, Bruzzone C, Gil-Redondo R, Embade N, Mato JM, Holmes E, Wist J, Millet O, Nicholson JK (2021) Integrative modeling of plasma metabolic and lipoprotein biomarkers of SARS-CoV-2 infection in Spanish and Australian COVID-19 patient cohorts. J Proteome Res 20:4139–4152. https://doi.org/10.1021/acs.jproteome.1c00458

McCartney A, Vignoli A, Biganzoli L, Love R, Tenori L, Luchinat C, Di Leo A (2018) Metabolomics in breast cancer: a decade in review. Cancer Treat Rev 67:88–96. https://doi.org/10.1016/j.ctrv.2018.04.012

McCartney A, Vignoli A, Hart C, Tenori L, Luchinat C, Biganzoli L, Di Leo A (2017) De-escalating and escalating treatment beyond endocrine therapy in patients with luminal breast cancer. Breast 34:S13–S18. https://doi.org/10.1016/j.breast.2017.06.021

McCartney A, Vignoli A, Tenori L, Fornier M, Rossi L, Risi E, Luchinat C, Biganzoli L, Di Leo A (2019) Metabolomic analysis of serum may refine 21-gene expression assay risk recurrence stratification. NPJ Br Cancer 5:26. https://doi.org/10.1038/s41523-019-0123-9

Meijer RI, van Wagensveld BA, Siegert CE, Eringa EC, Smulders YM (2011) Bariatric surgery as a novel treatment for type 2 diabetes mellitus: a systematic review. Arch Surg 146:744–750. https://doi.org/10.1001/archsurg.2011.134

Meoni G, Ghini V, Maggi L, Vignoli A, Mazzoni A, Salvati L, Capone M, Vanni A, Tenori L, Fontanari P, Lavorini F, Peris A, Bartoloni A, Liotta F, Cosmi L, Luchinat C, Annunziato F, Turano P (2021) Metabolomic/lipidomic profiling of COVID-19 and individual response to tocilizumab. PLoS Pathog 17:e1009243. https://doi.org/10.1371/journal.ppat.1009243

Meoni G, Lorini S, Monti M, Madia F, Corti G, Luchinat C, Zignego AL, Tenori L, Gragnani L (2019) The metabolic fingerprints of HCV and HBV infections studied by nuclear magnetic resonance spectroscopy. Sci Rep 9:4128. https://doi.org/10.1038/s41598-019-40028-4

Monleón D, Morales JM, Barrasa A, López JA, Vázquez C, Celda B (2009) Metabolite profiling of fecal water extracts from human colorectal cancer. NMR Biomed 22:342–348. https://doi.org/10.1002/nbm.1345

Montuschi P, Santini G, Mores N, Vignoli A, Macagno F, Shohreh R, Tenori L, Zini G, Fuso L, Mondino C, Di Natale C, D'Amico A, Barnes PJ, Higenbottam T (2018) BREATHOMICS for ASSESSING the effects of treatment and withdrawal with inhaled beclomethasone/formoterol in patients with COPD. Front Pharmacol 9:258. https://doi.org/10.3389/fphar.2018.00258

Motta A, Paris D, Melck D, de Laurentiis G, Maniscalco M, Sofia M, Montuschi P (2012) Nuclear magnetic resonance-based metabolomics of exhaled breath condensate: methodological aspects. Eur Respir J 39:498–500. https://doi.org/10.1183/09031936.00036411

Nannini G, Meoni G, Amedei A, Tenori L (2020) Metabolomics profile in gastrointestinal cancers: update and future perspectives. World J Gastroenterol 26:2514–2532. https://doi.org/10.3748/wjg.v26.i20.2514

Nannini G, Meoni G, Tenori L, Ringressi MN, Taddei A, Niccolai E, Baldi S, Russo E, Luchinat C, Amedei A (2021) Fecal metabolomic profiles: a comparative study of patients with colorectal cancer vs adenomatous polyps. World J Gastroenterol 27:6430–6441. https://doi.org/10.3748/wjg.v27.i38.6430

Nguyen HTT, Wimmer R, Le VQ, Krarup HB (2021) Metabolic fingerprint of progression of chronic hepatitis B: changes in the metabolome and novel diagnostic possibilities. Metabolomics 17:16. https://doi.org/10.1007/s11306-020-01767-y

Nicholson JK, Connelly J, Lindon JC, Holmes E (2002) Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov 1:153–161. https://doi.org/10.1038/nrd728

Nie X, Xia L, Gao F, Liu L, Yang Y, Chen Y, Duan H, Yao Y, Chen Z, Lu S, Wang Y, Yang C (2021) Serum metabolite biomarkers predictive of response to PD-1 blockade therapy in non-small cell lung cancer. Front Mol Biosci 8:472. https://doi.org/10.3389/fmolb.2021.678753

Okuda M, Li K, Beard MR, Showalter LA, Scholle F, Lemon SM, Weinman SA (2002) Mitochondrial injury, oxidative stress, and antioxidant gene expression are induced by hepatitis C virus core protein. Gastroenterology 122:366–375. https://doi.org/10.1053/gast.2002.30983

O'Sullivan A, Gibney MJ, Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. Am J Clin Nutr 93:314–321. https://doi.org/10.3945/ajcn.110.000950

Polyak K (2011) Heterogeneity in breast cancer. J Clin Invest 121:3786–3788. https://doi.org/10.1172/JCI60534

Postow MA, Callahan MK, Wolchok JD (2015) Immune checkpoint blockade in cancer therapy. J Clin Oncol 33:1974–1982. https://doi.org/10.1200/JCO.2014.59.4358

Ramirez T, Daneshian M, Kamp H, Bois FY, Clench MR, Coen M, Donley B, Fischer SM, Ekman DR, Fabian E, Guillou C, Heuer J, Hogberg HT, Jungnickel H, Keun HC, Krennrich G, Krupp E, Luch A, Noor F, Peter E, Riefke B, Seymour M, Skinner N, Smirnova L, Verheij E, Wagner S, Hartung T, van Ravenzwaay B, Leist M (2013) Metabolomics in toxicology and preclinical research. ALTEX 30:209–225. https://doi.org/10.14573/altex.2013.2.209

Resendiz-Acevedo K, García-Aguilera ME, Esturau-Escofet N, Ruiz-Azuara L (2021) 1H -NMR metabolomics study of the effect of cisplatin and CasiopeinaIIgly on MDA-MB-231 breast tumor cells. Front Mol Biosci 8:1135. https://doi.org/10.3389/fmolb.2021.742859

Ringehan M, McKeating JA, Protzer U (2017) Viral hepatitis and liver cancer. Philos Trans R Soc Lond B Biol Sci 372:20160274. https://doi.org/10.1098/rstb.2016.0274

Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, Gadgeel SM, Hida T, Kowalski DM, Dols MC, Cortinovis DL, Leach J, Polikoff J, Barrios C, Kabbinavar F, Frontera OA, De Marinis F, Turna H, Lee J-S, Ballinger M, Kowanetz M, He P, Chen DS, Sandler A, Gandara DR, OAK Study Group (2017) Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet 389:255–265. https://doi.org/10.1016/S0140-6736(16)32517-X

Rocca MS, Vignoli A, Tenori L, Ghezzi M, De Rocco Ponce M, Vatsellas G, Thanos D, Padrini R, Foresta C, De Toni L (2020) Evaluation of serum/urine genomic and metabolomic profiles to improve the adherence to sildenafil therapy in patients with erectile dysfunction. Front Pharmacol 11:602369. https://doi.org/10.3389/fphar.2020.602369

Saborano R, Eraslan Z, Roberts J, Khanim FL, Lalor PF, Reed MAC, Günther UL (2019) A framework for tracer-based metabolism in mammalian cells by NMR. Sci Rep 9:2520. https://doi.org/10.1038/s41598-018-37525-3

Salmerón AM, Tristán AI, Abreu AC, Fernández I (2022) Serum colorectal cancer biomarkers unraveled by NMR metabolomics: past, present, and future. Anal Chem 94:417–430. https://doi.org/10.1021/acs.analchem.1c04360

Sarfaraz MO, Myers RP, Coffin CS, Gao Z-H, Shaheen AAM, Crotty PM, Zhang P, Vogel HJ, Weljie AM (2016) A quantitative metabolomics profiling approach for the noninvasive assessment of liver histology in patients with chronic hepatitis C. Clin Transl Med 5:33. https://doi.org/10.1186/s40169-016-0109-2

Siegel RL, Miller KD, Fuchs HE, Jemal A (2021) Cancer statistics, 2021. CA Cancer J Clin 71:7–33. https://doi.org/10.3322/caac.21654

Silva CL, Olival A, Perestrelo R, Silva P, Tomás H, Câmara JS (2019) Untargeted urinary 1H NMR-based metabolomic pattern as a potential platform in breast cancer detection. Metabolites 9:269. https://doi.org/10.3390/metabo9110269

Singh A, Sharma RK, Chagtoo M, Agarwal G, George N, Sinha N, Godbole MM (2017) 1H NMR metabolomics reveals association of high expression of inositol 1, 4, 5 trisphosphate receptor and metabolites in breast cancer patients. PLoS One 12:e0169330. https://doi.org/10.1371/journal.pone.0169330

Sirniö P, Väyrynen JP, Klintrup K, Mäkelä J, Karhu T, Herzig K-H, Minkkinen I, Mäkinen MJ, Karttunen TJ, Tuomisto A (2019) Alterations in serum amino-acid profile in the progression of colorectal cancer: associations with systemic inflammation, tumour stage and patient survival. Br J Cancer 120:238–246. https://doi.org/10.1038/s41416-018-0357-6

Sitter B, Lundgren S, Bathen TF, Halgunset J, Fjosne HE, Gribbestad IS (2006) Comparison of HR MAS MR spectroscopic profiles of breast cancer tissue with clinical parameters. NMR Biomed 19:30–40. https://doi.org/10.1002/nbm.992

Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, Mackey J, Glaspy J, Chan A, Pawlicki M, Pinter T, Valero V, Liu M-C, Sauter G, von Minckwitz G, Visco F, Bee V, Buyse M, Bendahmane B, Tabah-Fisch I, Lindsay M-A, Riva A, Crown J, Breast Cancer International Research Group (2011) Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med 365:1273–1283. https://doi.org/10.1056/NEJMoa0910383

Slupsky CM, Steed H, Wells T, Dabbs K, Schepansky A, Capstick V, Faught W, Sawyer MB (2010) Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. Clin Cancer Res 16:5835–5841. https://doi.org/10.1158/1078-0432.CCR-10-1434

Suman S, Sharma RK, Kumar V, Sinha N, Shukla Y (2018) Metabolic fingerprinting in breast cancer stages through 1H NMR spectroscopy-based metabolomic analysis of plasma. J Pharm Biomed Anal 160:38–45. https://doi.org/10.1016/j.jpba.2018.07.024

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/caac.21660

Swain SM, Kim S-B, Cortés J, Ro J, Semiglazov V, Campone M, Ciruelos E, Ferrero J-M, Schneeweiss A, Knott A, Clark E, Ross G, Benyunes MC, Baselga J (2013) Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. Lancet Oncol 14:461–471. https://doi.org/10.1016/S1470-2045(13)70130-X

Takis PG, Ghini V, Tenori L, Turano P, Luchinat C (2019) Uniqueness of the NMR approach to metabolomics. TRAC-Trends Anal Chem 120:115300. https://doi.org/10.1016/j.trac.2018.10.036

Tan B, Qiu Y, Zou X, Chen T, Xie G, Cheng Y, Dong T, Zhao L, Feng B, Hu X, Xu LX, Zhao A, Zhang M, Cai G, Cai S, Zhou Z, Zheng M, Zhang Y, Jia W (2013) Metabonomics identifies serum metabolite markers of colorectal cancer. J Proteome Res 12:3000–3009. https://doi.org/10.1021/pr400337b

Tayyari F, Gowda GAN, Olopade OF, Berg R, Yang HH, Lee MP, Ngwa WF, Mittal SK, Raftery D, Mohammed SI (2018) Metabolic profiles of triple-negative and luminal A breast cancer subtypes in African-American identify key metabolic differences. Oncotarget 9:11677–11690. https://doi.org/10.18632/oncotarget.24433

Tenori L, Oakman C, Claudino WM, Bernini P, Cappadona S, Nepi S, Biganzoli L, Arbushites MC, Luchinat C, Bertini I, Di Leo A (2012) Exploration of serum metabolomic profiles and outcomes in women with metastatic breast cancer: a pilot study. Mol Oncol 6:437–444. https://doi.org/10.1016/j.molonc.2012.05.003

Tenori L, Oakman C, Morris PG, Gralka E, Turner N, Cappadona S, Fornier M, Hudis C, Norton L, Luchinat C, Di Leo A (2015) 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 9:128–139. https://doi.org/10.1016/j.molonc.2014.07.012

Thompson AM, Moulder-Thompson SL (2012) Neoadjuvant treatment of breast cancer. Ann Oncol 23:x231–x236. https://doi.org/10.1093/annonc/mds324

Turano P (2014) Colorectal cancer: the potential of metabolic fingerprinting. Expert Rev Gastroenterol Hepatol 8:847–849. https://doi.org/10.1586/17474124.2014.945912

Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, Pegram M, Oh D-Y, Diéras V, Guardino E, Fang L, Lu MW, Olsen S, Blackwell K, EMILIA Study Group (2012) Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 367:1783–1791. https://doi.org/10.1056/NEJMoa1209124

Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, Turano P, Luchinat C (2019) High-throughput metabolomics by 1D NMR. Angew Chem Int Ed 58:968–994. https://doi.org/10.1002/anie.201804736

Vignoli A, Mori E, Di Donato S, Malorni L, Biagioni C, Benelli M, Calamai V, Cantafio S, Parnofiello A, Baraghini M, Garzi A, Monte FD, Romagnoli D, Migliaccio I, Luchinat C, Tenori L, Biganzoli L (2021a) Exploring serum NMR-based metabolomic fingerprint of colorectal cancer patients: effects of surgery and possible associations with cancer relapse. Appl Sci 11:11120. https://doi.org/10.3390/app112311120

Vignoli A, Muraro E, Miolo G, Tenori L, Turano P, Di Gregorio E, Steffan A, Luchinat C, Corona G (2020a) Effect of Estrogen receptor status on circulatory immune and metabolomics profiles of HER2-positive breast cancer patients enrolled for neoadjuvant targeted chemotherapy. Cancers 12:314. https://doi.org/10.3390/cancers12020314

Vignoli A, Orlandini B, Tenori L, Biagini MR, Milani S, Renzi D, Luchinat C, Calabrò AS (2019) Metabolic signature of primary biliary cholangitis and its comparison with celiac disease. J Proteome Res 18:1228–1236. https://doi.org/10.1021/acs.jproteome.8b00849

Vignoli A, Risi E, McCartney A, Migliaccio I, Moretti E, Malorni L, Luchinat C, Biganzoli L, Tenori L (2021b) Precision oncology via NMR-based metabolomics: a review on breast cancer. Int J Mol Sci 22:4687. https://doi.org/10.3390/ijms22094687

Vignoli A, Santini G, Tenori L, Macis G, Mores N, Macagno F, Pagano F, Higenbottam T, Luchinat C, Montuschi P (2020b) NMR-based metabolomics for the assessment of inhaled pharmacotherapy in chronic obstructive pulmonary disease patients. J Proteome Res 19:64–74. https://doi.org/10.1021/acs.jproteome.9b00345

Vignoli A, Tenori L, Morsiani C, Turano P, Capri M, Luchinat C (2022) Serum or plasma (and which plasma), that is the question. J Proteome Res 21:1061–1072. https://doi.org/10.1021/acs.jproteome.1c00935

von Minckwitz G, Huang C-S, Mano MS, Loibl S, Mamounas EP, Untch M, Wolmark N, Rastogi P, Schneeweiss A, Redondo A, Fischer HH, Jacot W, Conlin AK, Arce-Salinas C, Wapnir IL, Jackisch C, DiGiovanna MP, Fasching PA, Crown JP, Wülfing P, Shao Z, Rota Caremoli E, Wu H, Lam LH, Tesarowski D, Smitt M, Douthwaite H, Singel SM, Geyer CE, Katherine Investigators (2019) Trastuzumab emtansine for residual invasive HER2-positive breast cancer. N Engl J Med 380:617–628. https://doi.org/10.1056/NEJMoa1814017

von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, Suter T, Arahmani A, Rouchet N, Clark E, Knott A, Lang I, Levy C, Yardley DA, Bines J, Gelber RD, Piccart M, Baselga J, Aphinity Steering Committee and Investigators (2017) Adjuvant pertuzumab and trastuzumab in early HER2-positive breast cancer. N Engl J Med 377:122–131. https://doi.org/10.1056/NEJMoa1703643

Wallner-Liebmann S, Gralka E, Tenori L, Konrad M, Hofmann P, Dieber-Rotheneder M, Turano P, Luchinat C, Zatloukal K (2015) The impact of free or standardized lifestyle and urine sampling protocol on metabolome recognition accuracy. Genes Nutr 10:441. https://doi.org/10.1007/s12263-014-0441-3

Wallner-Liebmann S, Tenori L, Mazzoleni A, Dieber-Rotheneder M, Konrad M, Hofmann P, Luchinat C, Turano P, Zatloukal K (2016) Individual human metabolic phenotype analyzed by (1)H NMR of saliva samples. J Proteome Res 15:1787–1793. https://doi.org/10.1021/acs.jproteome.5b01060

Wang L, Tang Y, Liu S, Mao S, Ling Y, Liu D, He X, Wang X (2013) Metabonomic profiling of serum and urine by 1H NMR-based spectroscopy discriminates patients with chronic obstructive pulmonary disease and healthy individuals. PLoS One 8:e65675. https://doi.org/10.1371/journal.pone.0065675

Wang P, Shehu AI, Ma X (2017) The opportunities of metabolomics in drug safety evaluation. Curr Pharmacol Rep 3:10–15. https://doi.org/10.1007/s40495-016-0079-5

Wei S, Liu L, Zhang J, Bowers J, Gowda GAN, Seeger H, Fehm T, Neubauer HJ, Vogel U, Clare SE, Raftery D (2013) Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer. Mol Oncol 7:297–307. https://doi.org/10.1016/j.molonc.2012.10.003

Wiggans RG, Woolley PV, Smythe T, Hoth D, Macdonald JS, Green L, Schein PS (1979) Phase-II trial of tamoxifen in advanced breast cancer. Cancer Chemother Pharmacol 3:45–48. https://doi.org/10.1007/BF00254419

Winnike JH, Busby MG, Watkins PB, O'Connell TM (2009) Effects of a prolonged standardized diet on normalizing the human metabolome. Am J Clin Nutr 90:1496–1501. https://doi.org/10.3945/ajcn.2009.28234

Wishart DS (2015) Is cancer a genetic disease or a metabolic disease? EBioMedicine 2:478–479. https://doi.org/10.1016/j.ebiom.2015.05.022

Wojtowicz W, Wróbel A, Pyziak K, Tarkowski R, Balcerzak A, Bębenek M, Młynarz P (2020) Evaluation of MDA-MB-468 cell culture media analysis in predicting triple-negative breast cancer patient sera metabolic profiles. Metabolites 10:173. https://doi.org/10.3390/metabo10050173

World Health Organization (2021) Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021: accountability for the global health sector strategies 2016–2021: actions for impact: web annex 1: key data at a glance. World Health Organization, Geneva

Xi Y, Xu P (2021) Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol 14:101174. https://doi.org/10.1016/j.tranon.2021.101174

Yafi FA, Jenkins L, Albersen M, Corona G, Isidori AM, Goldfarb S, Maggi M, Nelson CJ, Parish S, Salonia A, Tan R, Mulhall JP, Hellstrom WJG (2016) Erectile dysfunction. Nat Rev Dis Primers 2:16003. https://doi.org/10.1038/nrdp.2016.3

Ząbek A, Stanimirova I, Deja S, Barg W, Kowal A, Korzeniewska A, Orczyk-Pawiłowicz M, Baranowski D, Gdaniec Z, Jankowska R, Młynarz P (2015) Fusion of the 1H NMR data of serum, urine and exhaled breath condensate in order to discriminate chronic obstructive pulmonary disease and obstructive sleep apnea syndrome. Metabolomics 11:1563–1574. https://doi.org/10.1007/s11306-015-0808-5

Zhang X, Zhao X-W, Liu D-B, Han C-Z, Du L-L, Jing J-X, Wang Y (2014) Lipid levels in serum and cancerous tissues of colorectal cancer patients. World J Gastroenterol 20:8646–8652. https://doi.org/10.3748/wjg.v20.i26.8646

Zhou J, Wang Y, Zhang X (2017) Metabonomics studies on serum and urine of patients with breast cancer using 1 H-NMR spectroscopy. Oncotarget 5. https://doi.org/10.18632/oncotarget.16210

Zhu J, Djukovic D, Deng L, Gu H, Himmati F, Chiorean EG, Raftery D (2014) Colorectal cancer detection using targeted serum metabolic profiling. J Proteome Res 13:4120–4130. https://doi.org/10.1021/pr500494u

Pharmacometabolomics of Asthma as a Road Map to Precision Medicine

Rachel S. Kelly, Margaret F. Cote, Sofina Begum, and Jessica Lasky-Su

Contents

1 Introduction 248
2 Pharmacometabolomics 248
3 Applications of Pharmacometabolomics to Asthma 249
4 Metabolomic Responses to Asthma Therapeutics 250
4.1 Steroids (Inhaled and Oral) 252
4.2 Bronchodilators (Beta-Agonists) 254
4.3 Leukotriene Inhibitors 256
4.4 Combined Therapeutics 257
5 Therapy or Severity 257
6 Clinical Subgroups and Endotypes 258
6.1 Clinically Derived Subgroups (The "Bottom-Up" Approach) 259
6.2 Omic-Driven Endotypes (The "Top-Down" Approach) 261
6.3 Aspirin-Exacerbated Asthma 263
7 Prevention 264
8 Challenges of Pharmacometabolomics in Asthma 265
9 Conclusions 267
References 268

Abstract

Pharmacometabolomics applies the principles of metabolomics to therapeutics in order to elucidate the biological mechanisms underlying the variation in responses to drugs between groups and individuals. Asthma is associated with broad systemic effects and heterogeneity in treatment response and as such is

R. S. Kelly (✉) · M. F. Cote · S. Begum · J. Lasky-Su
Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
e-mail: hprke@channing.harvard.edu; mcote@college.harvard.edu; resof@channing.harvard.edu; rejas@channing.harvard.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_615

ideally suited to pharmacometabolomics. In this chapter, we discuss the state of the emerging field of asthma pharmacometabolomics, with a particular focus on studies of steroids, bronchodilators, and leukotriene inhibitors. We also consider those studies concerned with subtyping cases to better understand the pharmacology of those groups and those looking to leverage pharmacometabolomics for asthma prevention. We finish with a discussion of the challenges and opportunities of asthma pharmacometabolomics and reflect upon where this field must go next in order to realize its precision medicine potential.

Keywords

Asthma · Endotyping · Pharmacometabolomics · Steroids · Treatment response

1 Introduction

Approximately 334 million people are affected by asthma worldwide, attributing to nearly 400,000 deaths annually (GBD 2015 Chronic Respiratory Disease Collaborators 2017). It is predicted that in the next decade, over 100 million new cases will arise (Enilari and Sinha 2019). In addition to the public health burden associated with asthma, there is also a severe economic burden (Yaghoubi et al. 2019). It is estimated that by 2040 the direct costs of uncontrolled asthma will rise to 300.6 billion dollars in the USA alone (Yaghoubi et al. 2019). Consequently, the need for increased asthma control via improved therapeutics is paramount.

The definition of asthma has not changed in over 50 years (Hargreave and Nair 2009). However, studies have demonstrated its complex, heterogeneous nature (Darveaux and Busse 2015; Moore et al. 2010; Sinha et al. 2017; Tyler and Bunyavanich 2019). Asthma differs between individuals in its pathology, manifestation, molecular biology, and clinical response (Eder et al. 2006; Tyler and Bunyavanich 2019). Hence, there is a need for novel targeted therapeutics to replace the current "one-size-fits-all" approach (Tyler and Bunyavanich 2019). Precision medicine for asthma could improve management, reduce instances of serious outcomes, and prevent adverse drug responses through disease stratification (Moore et al. 2010).

2 Pharmacometabolomics

The metabolome reflects transcriptional and translational processes as well as environmental interactions (Fiehn 2002; Tyler and Bunyavanich 2019). Metabolomics utilizes spectroscopic techniques and methods for the comprehensive profiling of the metabolome in a biological specimen (Clish 2015). This powerful tool is central in precision medicine, facilitating deconvolution of the complex metabolic landscapes which underpin disease. Pharmacometabolomics utilizes the principles of metabolomics and extends these principles to therapeutics in order to

elucidate the underlying biological mechanisms behind complex diseases and provide insights into how individuals respond to the drugs designed for these conditions (Kaddurah-Daouk and Weinshilboum 2014). Through the implementation of pharmacometabolomic techniques and methods, we can improve understanding of why different individuals may respond differently to the same treatment (Kaddurah-Daouk and Weinshilboum 2014).

Asthma is associated with broad systemic effects including inflammation, oxidative stress, and tissue remodeling (Sahiner et al. 2011), which arise as a consequence of the complex interplay between genetic and environmental factors. Thus, it is ideal for the application of a systems approach, such as metabolomics and subsequently pharmacometabolomics (Zhu et al. 2019). Pharmacometabolomics offers the potential to improve our understanding of the actions of drugs on individuals across the spectrum of asthma and facilitate the development of novel targeted therapies with improved efficacy for all. Pharmacometabolomics offers a compelling and novel route to precision medicine for individuals with asthma.

3 Applications of Pharmacometabolomics to Asthma

Over the last two decades, there has been growing interest in the role of the metabolome in asthma. A large number of studies have demonstrated that the metabolome, as measured in a variety of biosamples including blood, urine, and exhaled breath consendate (EBC) can be leveraged to predict, diagnose, and assess the severity of asthma (Wang et al. 2021; Kelly et al. 2017; Papamichael et al. 2021). Importantly, several of the metabolomic perturbations reported as being associated with asthma and asthma phenotypes including pathways relating to hypoxia response, oxidative stress, immunity, inflammation, lipid metabolism, and the tricarboxylic acid cycle have been independently validated (Chung et al. 2014). More recently, a body of asthma metabolomics researchers have shifted their focus to study how an individual's metabolome can influence their response to therapeutic strategies. There is a growing interest in the utility of metabolomics for the development of more efficacious asthma treatments (McGeachie et al. 2015; Svenningsen and Nair 2017; Kelly et al. 2017, 2019b; Kuruvilla et al. 2019). Although only a small number of studies of asthma have specifically labelled themselves as "pharmacometabolomic" studies, many more have focused on the integration of asthma phenotypes, metabolomics, and therapeutics.

Pharmacometabolomic asthma studies tend to take one of the two broad approaches (Fig. 1): (1) the assessment of the metabolomic responses of individuals with asthma to various therapeutics; or (2) the subgrouping of individuals with asthma. These subgrouping or disease stratification studies can be further subdivided into those that take a "bottom-up" and those that take a "top-down" approach. The "bottom-up" approach clusters asthmatics based on their clinical characteristics as they relate to therapy and/or management and then investigates the metabolomic profiles of these subgroups (Fig. 1 [2a]). The "top-down" studies stratify disease based on their metabolome, clustering individuals with asthma into subgroups based

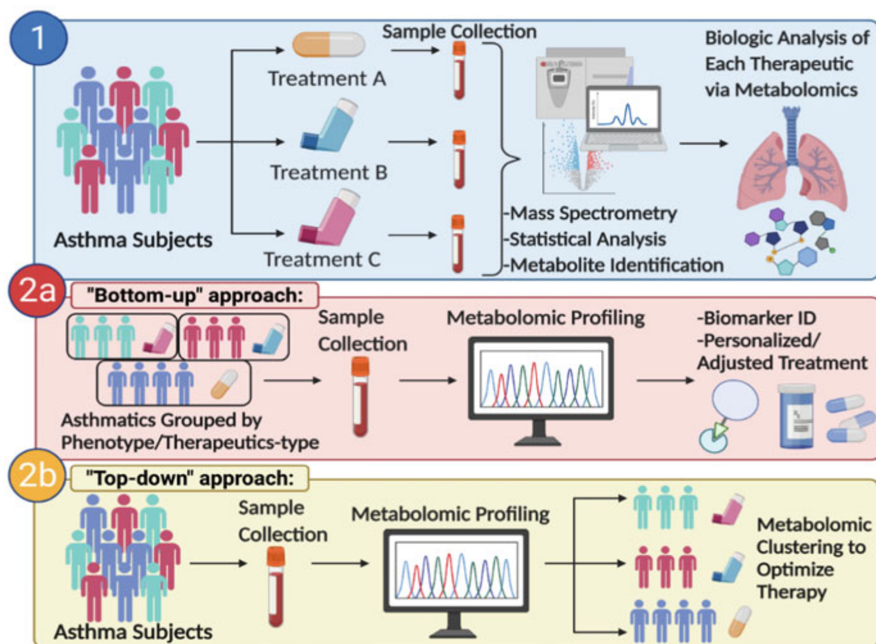


Fig. 1 Asthma pharmacometabolomics study design approaches

on their metabolic profile and investigating the therapeutic needs of these subgroups (Fig. 1 [2b]). In both instances, the end goal is to understand the mechanisms and biology underlying disease in the different subgroups in order to identify novel treatments targeted toward those underlying mechanisms.

In this chapter, we explore current perspectives and uses of pharmacometabolomics in the study of asthma, describe the findings of the studies to date, and reflect upon their potential roles in optimizing the treatment of individuals with asthma as a route toward precision medicine.

4 Metabolomic Responses to Asthma Therapeutics

Multiple therapeutics are currently utilized in the management of asthma and its symptoms. These include inhaled corticosteroids (ICSs), oral corticosteroids, long-acting β2-agonists (LABAs), short-acting β2-agonists (SABAs), leukotriene inhibitors (blockers of the CysLT receptor, specifically), long-acting muscarinic antagonists (LAMAs), short-acting muscarinic antagonists (SAMAs), and anti-IL and anti-IgE antibody therapies, theophylline and cromolyn (Wendell et al. 2020) (Table 1). These treatments, alone or in combination, comprise both long-term and quick-relief medications that aid in reducing symptoms among asthmatic patients by targeting different biological pathways and processes. However, their efficacy varies

Table 1 Common asthma therapeutics

Type	Function	Example medications
Steroids (inhaled)	Prohibits gene transcription of inflammatory genes	Beclomethasone Dipropionate Budesonide Ciclesonide Flunisolide Fluticasone Furoate Mometasone furoate Fluticasone Propionate
Steroids (Oral)	Prohibits gene transcription of inflammatory genes	Prednisolone Prednisone Methylprednisolone
Leukotriene inhibitors	Prevents the formation of leukotrienes and blocks leukotriene receptors	Zileuton Montelukast Zafirlukast Pranlukast
Anti-IL treatment: IL5	Binds to the IL-5 receptor to reduce eosinophil survival	Anti-IL5: Mepolizumab Reslizumab Benralizumab
Anti-IL treatment: IL-4	Binds to the IL-4 receptors to reduce airway hyperresponsiveness	Anti-IL4: Dupilumab
Bronchodilators: SABAS	Activates the beta2-adrenoreceptors present on the surfaces of airway smooth muscle cells for muscle relaxation	SABAS: Epinephrine Isoproterenol Salbutamol Pirbuterol Terbutaline
Bronchodilators: LABAS	Activates the beta2-adrenoreceptors present on the surfaces of airway smooth muscle cells for muscle relaxation	LABAS: Salmeterol Formoterol BI-167107 Vilanterol Indacaterol Olodaterol

across populations (Larsson et al. 2020), and many asthmatics do not respond to the "one-size-fits-all" management approach. Patients who do not respond to standard care are responsible for most of the asthma-related economic burden (Chung et al. 2014). Researchers are now beginning to leverage pharmacometabolomics to try and better understand the mechanisms of action of these therapeutics and the subgroups that may benefit from these treatments.

4.1 Steroids (Inhaled and Oral)

Steroids reduce the airway inflammation associated with asthma and are the most commonly prescribed drugs for this condition (Wendell et al. 2020). However, their efficacy varies between individuals; approximately 25–35% of asthma patients either do not respond or respond poorly to inhaled corticosteroids (ICS) (Kachroo et al. 2021). As such, they are often prescribed in combination with other therapeutics in order to optimize treatment. Identifying non- or poor-responders could help to improve treatment efficacy and to reduce the potential side effects associated with overtreatment (Kachroo et al. 2021), which can include weight gain, high blood pressure, muscle weakness, osteoporosis, and an increased risk of infection, in addition to the increasingly understood consequence of adrenal suppression (Grennan and Wang 2019).

There are two different categories of steroids that work to reduce asthma symptoms: long-term control and quick-relief medications (Wendell et al. 2020). Long-term control medications are typically ICS such as beclomethasone dipropionate, budesonide, ciclesonide, flunisolide, fluticasone furoate, mometasone furoate, and fluticasone propionate. Quick-relief medications include both oral and intravenous corticosteroids such as prednisolone, prednisone, and methylprednisolone (Wendell et al. 2020). The main action of corticosteroids is to prohibit transcription of inflammatory genes (Barnes 2006). They also induce transcription of anti-inflammatory and immune regulation genes including beta2-adrenergic receptors, secretory leukocyte inhibitory protein, and mitogen-activated protein kinase phosphatase-1 through the glucocorticoid receptor-mediated transactivation, further contributing to their anti-inflammatory properties (Newton and Giembycz 2016; Matera et al. 2020; Wendell et al. 2020).

Among the treatment options for asthma, steroids are by far the most commonly studied in a pharmacometabolomic framework. Within this category, studies of ICS dominate the research. Yet despite the relatively large body of literature, there is some disagreement regarding the influences of steroids on the metabolome.

Ferraro et al. performed breathomics (metabolomic profiling of the EBC) in 26 children with asthma before and after a 10-week course of inhaled beclomethasone dipropionate, a common corticosteroid, and in 16 children without asthma (Ferraro et al. 2020). Although the children with asthma demonstrated improvements in lung function and asthma control following treatment, there was no change in the EBC metabolome, and no differences in the urinary endogenous steroid profile. They also demonstrated that the EBC metabolome could distinguish between children with and without asthma and that steroid treatment did not affect this ability, with notable differences in prostaglandin, fatty acid, and glycerophospholipid metabolites between children with and without asthma. Interestingly, they identified more metabolomic differences between the asthmatic and non-asthmatic children before the steroid treatment. This could be interpreted to suggest that rather than causing further dysregulation to the metabolome the treatment actually returned it to a state more akin to a non-asthmatic control. (Ferraro et al. 2020). Another breathomics study investigated ethane, a product of lipid

peroxidation that occurs in response to oxidative stress, and it was further demonstrated that ethane levels in the non-treated individuals were correlated with exhaled nitric oxide, a marker of oxidative stress and inflammation (Paredi et al. 2000). In contrast to Ferraro et al.'s findings, which found no differences in the metabolome associated with treatment, this study of ethane in exhaled air (Paredi et al. 2000) determined that levels of ethane were higher in 12 patients not receiving steroids as compared to 14 steroid-treated patients and 12 non-smoking controls. This again points to the effectiveness of steroids in stabilizing treated subjects.

Oxidative stress was also a mechanism of interest in one (Loureiro et al. 2014) of two studies that considered the relationship between metabolites and exacerbations among individuals, with the hypothesis that exacerbation-associated metabolites may act as markers of treatment response(Kachroo et al. 2021; Loureiro et al. 2014). Loureiro et al. conducted a small study of 10 individuals with asthma on inhaled corticosteroids and long-acting b-agonists who had urinary metabolomics performed during exacerbations(Loureiro et al. 2014). They determined levels of threonine (and/or lactate), alanine, carnitine, acetylcarnitine, and trimethylamine-N-oxide, aldehydes and alkanes were increased during the exacerbated state compared to the stable condition, while acetate, citrate, malonate, hippurate, dimethylglycine, and phenylacetylglutamine were decreased (Loureiro et al. 2014). As such, these metabolites, several of which are involved in pathways of oxidative stress, can be considered as markers of steroid non-responsiveness. In a larger study of the plasma metabolome including 170 adults with asthma, Kachroo et al. identified a set of steroid-response associated metabolites. They reported eight plasma metabolites that were associated with episodes of exacerbation while on ICS after correction for multiple testing (Kachroo et al. 2021). Two of the top metabolites, hexadecanedioate and tetradecanedioate are involved in omega fatty acid oxidation, which is known to play a key role in immune function, airway remodeling, but which has not previously been linked to ICS response. They further identified associations with amino acids and urea cycle metabolites and reported inverse associations between cortisol and cortisone and exacerbations. The authors postulate that, as these metabolites are known markers of treatment adherence, this suggests individuals with asthma who consistently used their ICS medication also suffered less frequent exacerbations. Finally, the authors reported significant sex-metabolite and age-metabolite interactions for exacerbations. These findings raise a crucial point; the relationship between treatment and metabolite may not be consistent between the sexes or over in individuals of differing age. This underlies the importance of considering such variables when exploring the influence of any treatment on the metabolome.

One of the most intriguing aspects of the asthma pharmacometabolomic literature is the evidence regarding the long-term effects of ICS on the metabolome. In fact, the enduring effects of steroid treatment on the metabolome and therefore on overall health are a concern that goes beyond asthma to include any condition treated with steroids (Pandya et al. 2014). Exogenous corticosteroids have been shown to precipitate negative feedback via their action on the hypothalamus and anterior pituitary gland to reduce the release of cortisol from the adrenal cortex leading to shrinkage of the adrenals. The adrenals can therefore no longer produce the

necessary endogenous corticosteroids which can result in adrenal insufficiency and adrenal suppression (Pandya et al. 2014), conditions associated with multiple morbidities and increased mortality (Gurnell et al. 2021).

Within the asthma pharmacometabolomics literature, a recent analysis of over 14,000 individuals from four independent cohorts identified and validated 35 plasma metabolites significantly associated with asthma status. Of these, 34 metabolites were annotated to canonical curated pathways for corticosteroids, pregnenolone, and androgenic steroids and all 34 were detected at lower levels among asthmatics (Kachroo et al. 2022). Strikingly, these reductions were more pronounced for asthmatics on ICS as compared to asthmatics not taking ICS. Further exploration of these cohorts determined that, in agreement with the literature, long-term ICS use among asthmatics was associated with significant long-term reduction in cortisone and cortisol. This study is among the first to report that these reductions are evident at even the lowest dosages of ICS. Importantly, when tracking cortisol levels over the course of a 24-h period it was found that the decrease in cortisol levels was greatest in the early morning, when subjects were most susceptible to an asthma attack. Furthermore, 31% of asthmatics using ICS had cortisol levels low enough to meet the clinical threshold for adrenal suppression (Kachroo et al. 2022). These results confirm the long-term clinical implications of steroid treatment in asthmatics using a pharmacometabolomic framework, while highlighting the potential of pharmacometabolomics to generate and test novel hypotheses.

Evidence that treatment with steroids acts to alter the metabolome in the long term is further supported by studies in both children (Kannisto et al. 2001) and adults (Kannisto et al. 2004) with asthma who were found to have significantly reduced levels of serum dehydroepiandrosterone sulfate following treatment with inhaled glucocorticosteroids. As dehydroepiandrosterone sulfate is the most abundant androgen secreted by the adrenals, this lends further credence to the hypothesis that long-term steroid use may lead to adrenal suppression.

4.2 Bronchodilators (Beta-Agonists)

Bronchodilators, also known as beta-agonists, work to activate the β2-adrenoceptors (β2-ARs) that are present on the surfaces of airway smooth muscle cells, epithelium, and vascular cells, as well as submucosal glands (Ruffin et al. 1982; Carstairs et al. 1985). β2-ARs are G-protein coupled receptors that are activated by catecholamines, which transmit downstream signaling (typically resulting in calcium channel interaction) to mediate the relaxation of smooth muscle. β2-ARs also exist on the surface of inflammatory and immune cells including mast cells, macrophages, neutrophils, lymphocytes, eosinophils, epithelial and endothelial cells, and type I and type II alveolar cells (Cazzola et al. 2012). The purpose of bronchodilators is to utilize the β2-adrenoceptors' pathway to eventually open K+ channels that lead to the relaxation of airway smooth muscle (Jones et al. 1999), although the effects on inflammation and the inflammatory pathway have not yet been fully understood (Matera et al. 2020). For this reason, bronchodilators and ICS are commonly prescribed together to

ensure that both inflammation and smooth muscle relaxation are achieved (Wendell et al. 2020). The most common types of bronchodilators include short-acting-beta-agonists (SABAs) which are used when an individual needs an immediate muscle relaxation effect during an asthma exacerbation, while long-acting-beta-agonists are taken daily to maintain muscle relaxation over long periods of time (Wendell et al. 2020). Commonly prescribed SABAs include albuterol, epinephrine, isoproterenol, metaproterenol, salbutamol, pirbuterol, and terbutaline. The most commonly utilized LABAs are salmeterol, formoterol, BI-167107, vilanterol, indacaterol, and olodaterol (Wendell et al. 2020).

To date, only a small number of studies have considered beta-agonists and asthma in a metabolomics framework, with most investigating treatment related metabolomic changes. Urinary levels of bromotyrosine, a marker reactive brominating oxidants formed by eosinophil-catalyzed oxidation, have been shown to be associated with the use of beta-agonists (Wedes et al. 2011). Similarly increased serum levels of lactate have been significantly associated with albuterol use in adult asthmatics into two independent studies (Lewis et al. 2014; Rodrigo and Rodrigo 2005), which is thought to be a direct result of therapy (Rodrigo 2014).

McGeachie et al. also focused on SABAs, investigating differences in the levels of serum metabolites between eight individuals with asthma who reported SABA use in the last 7 days compared to 12 individuals with asthma who did not use SABAs in the same time period (McGeachie et al. 2015). While no metabolites survived stringent correction for multiple testing, they did observe a nominally significant decrease in a monohydroxy derivative of arachidonic acid in those individuals not using SABAs. Together with linoleic acid metabolism and sphingolipid metabolism, arachidonic acid metabolism ranked as the top dysregulated pathway with respect to SABA use as a proxy for asthma control (McGeachie et al. 2015). They then integrated the metabolomic data with genome-wide genotype, gene expression data and methylation data using a Conditional Gaussian Bayesian Network (CGBN) to identify the strongest predictors of asthma control across these omic types. Integrative over representation analyses of the top multi-omic predictors supported the importance of these pathways in SABA use/asthma control (McGeachie et al. 2015).

Taking a slightly different approach two studies considered the influence of age-related changes on the beta-agonist – metabolite relationship. Kelly et al. focused on the metric of bronchodilator response (BDR). BDR measures spirometry before and after beta-agonists/bronchodilator treatment to establish reversibility of airflow. As BDR is known to decline with age, this study aimed to determine whether the BDR-age relationship was mediated by underlying metabolite levels. They were able to identify and validate in two independent populations blood metabolites that interacted with age in BDR. Increased levels of 2-hydroxyglutarate in the blood of an individual with asthma may exacerbate the association between age and BDR decline, while increased levels of cholesterol esters, GABA, and ribothymidine may attenuate the age-associated BDR decline (Kelly et al. 2019b). These metabolites have been shown to play various roles in lung function, inflammation, and immune function, but have not previously been directly

related to treatment (Kelly et al. 2019b). A second study (Sordillo et al. 2020) also investigated whether circulating metabolites mediate age-related changes in bronchodilator response (BDR) for individuals with asthma. Sordillo et al. determined that a portion of the effect of age on BDR acted indirectly through phosphatidylcholine plasmalogens. It is hypothesized that plasmalogens which protect against oxidative stress may mediate lung function responses through their ability to alter the structural properties of lung surfactants, and they are known to be enriched in the smooth muscle that bronchodilator acts upon (Bozelli et al. 2021). As such, the authors concluded that plasmalogens may serve as potential pharmacologic targets for enhancement of lung function in individuals with asthma (Sordillo et al. 2020).

4.3 Leukotriene Inhibitors

Cysteinyl leukotrienes (CysLTs) are a class of bioactive fatty acids associated with airway inflammation (Wendell et al. 2020). These fatty acids are released by eosinophils during an asthma exacerbation as proinflammatory mediators, which contribute to the bronchoconstriction seen in asthma patients (Wendell et al. 2020). The receptors for these leukotrienes exist on the cell surface of monocytes, eosinophils, lung macrophages, and resident mast cells, and the interaction of the CysLTs with their receptors has been found to induce anaphylaxis and cause the prolonged contraction of bronchial smooth muscle, that is characteristic of asthma (Augstein et al. 1973; Yokomizo et al. 2018). Leukotrienes further contribute to the pathogenesis of asthma and airway obstruction by recruiting eosinophils and neutrophils in the mucosa of the airway (Bisgaard 2001; Foster and Chan 1991; Smith et al. 1993; Henderson et al. 1996).

There are two critical stages along the pathway at which the formation of leukotrienes and the subsequent inflammatory response in the airways can be prevented. The first is at the 5-LOX enzymatic step where arachidonic acid is converted into LTA4 by 5-lipoxygenase, where Zileuton blocks this pathway (Walter et al. 1987; Peters-Golden and Henderson 2007). The second is at the G-protein coupled receptor (GPCR) $CysLTR_1$ location, where LTC4 and LTD4 are the substrates for downstream signaling events (Wendell et al. 2020). Drugs including Montelukast, Zafirlukast, and Pranlukast act as leukotriene receptor antagonists (LTRAs) by targeting the $CysLTR_1$ to obstruct the signaling pathway (Peters-Golden and Henderson 2007; Wendell et al. 2020). Treatment with these drugs therefore results in a decrease in the level of leukotrienes present in the body, leading to fewer and less severe asthma exacerbations. Montelukast specifically was found to significantly reduce eosinophil infiltration in asthma patients, while Zileuton reduces the production of leukotrienes (Yokomizo et al. 2018; Bruno et al. 2018).

The actions of these therapeutics are supported by cell-based studies (Werner et al. 2019). Profiling of the lipid metabolome of healthy human macrophages demonstrated that Zileuton reduces the intensity of the immune response by blocking the pathway for the creation of the enzyme 5-LOX. It also showed a

decrease in leukotriene levels as a reaction to FLAP and 5-LOX inhibitors (Werner et al. 2019). However, to date only a single human study has investigated the effect of leukotriene inhibitors on the metabolome of asthmatics (Quan-Jun et al. 2017). This study focused on the serum and urine NMR metabolomic profile of children with asthma treated with Montelukast and concluded that it imparted no metabolomic effect (Quan-Jun et al. 2017).

4.4 Combined Therapeutics

In individuals with severe asthma, beta-agonists and ICS are commonly prescribed in combination. To date very few studies have explored the combined metabolomic effects of these therapeutics. In addition to the Loureiro et al. article (Loureiro et al. 2014), which was primarily concerned with steroid non-responsiveness in individuals with asthma on inhaled corticosteroids and long-acting b-agonists, Quan-Jun et al. investigated the impact on the metabolome of childhood asthmatics receiving both the glucocorticosteroid budesonide and the beta-2-agonist salbutamol (Quan-Jun et al. 2017). There was a clear difference in the serum metabolomic profile of children on the combined therapeutic when compared to children receiving neither, and this difference was driven by increased levels of 4-hydroxybutyrate, lactate, cis-aconitate, 5-HIAA, taurine, trans-4-hydroxy-l-proline, tiglylglycine, 3-hydroxybutyrate, 3-methylhistidine, and glucose and decreased levels of alanine, glycerol, arginine, glycylproline, 2-hydroxy-3-methylvalerate, creatine, citrulline, glutamate, asparagine, 2-hydroxyvalerate, and citrate. There was no difference in the metabolome when considering either treatment alone, suggesting it is the combination of therapies that drives the metabolomic shift (Quan-Jun et al. 2017). Furthermore, the authors determined there were no metabolomic effects of a further beta-2-agonist, procaterol. Similar results were observed when considering the urine metabolome of these children. Taken together these results suggested possible "metabolic reprogramming" as a result of inhaled budesonide and salbutamol in asthmatic children, that is particularly pronounced in the arginine and proline metabolism pathway.

5 Therapy or Severity

Several studies report metabolomic changes that associate with the use of commonly administered asthma therapeutics. However, when considering these results, the issue of cause and effect must be considered. Treatment is inherently linked to the disease itself, whereby more severe cases or cases with a particular phenotype are more likely to receive therapy (confounding by indication). Considering this, many pharmacometabolomic studies of the effect of a particular therapeutic on the metabolome may be confounded by the clinical indication for that therapy. This is reflected in the fact that a number of metabolites and metabolomic pathways found to be associated with treatment in the studies previously mentioned have also been

linked to both the presence of asthma and its severity. For example, linoleic acid metabolism and arachidonic acid metabolism have been associated with asthma through their roles as lipid mediators of inflammation (Wendell et al. 2014).

A number of the asthma pharmacometabolomics studies have addressed the question of cause versus effect directly. In Paredi et al.'s study of exhaled ethane in asthma, they determined that ethane was higher in the non-treated cases, but that even within this group it was highest in those with the greatest degree of lung function. This suggests oxidative stress is correlated with severity itself and that the observed association between ethane as a marker of oxidative stress and with steroid use is due to the fact that treatment is reducing oxidative stress in this group (Paredi et al. 2000). The fact that Ferraro et al. found few differences between the steroid-treated EBC metabolome and healthy controls, as compared to the untreated metabolome versus controls could support this hypothesis, particularly given no differences were seen within the same individuals before and after treatment (Ferraro et al. 2020). In further agreement McGeachie et al. remarked that in the confines of their study "whether these metabolomic changes reflect the drug or the phenotype cannot be determined" (McGeachie et al. 2015). It is also of interest to note that lactate was implicated in several different therapeutics, which may suggest that dysregulation of this metabolite and its pathways relates more to the need for treatment than to the treatment itself.

However, it should be noted that the long-term effects of ICS use on endogenous cortisol and cortisone levels and on dehydroepiandrosterone sulfate does suggest that treatments for asthma can cause metabolomic shifts independent of the disease. Furthermore, changes in lactate levels have been directly attributed to treatment (Kelly et al. 2019b). This suggests that whether or not therapy directly effects the metabolome is metabolite specific and therapy specific, it may occur in some instances but not others. Given findings from several studies that have considered interactions, these relationships are also likely to be impacted by external factors such as age and sex.

6 Clinical Subgroups and Endotypes

Given the potential challenges that accompany consideration of the influence of a treatment on the metabolome of a potentially heterogeneous group, increasingly, pharmacometabolomic studies of asthma have been leveraging the heterogeneous nature of this condition to explore clinical subtypes or derive endotypes using metabolomics. In this way, studies aim to explain why the metabolomic responses to therapeutics may differ between individuals and ultimately to identify the optimal therapeutics in each case. Although to date, there have been no instances where pharmacometabolomic findings in asthma have been translated into the clinic, important insights into the mechanisms and biology of treatment responses have already been reported.

6.1 Clinically Derived Subgroups (The "Bottom-Up" Approach)

The "bottom-up" approach argues that therapeutic approaches should be targeted toward groups of individuals with the same clinical characteristics and phenotypes (Wenzel 2012). Various asthma subtypes have been proposed based on key clinical features including symptoms, exacerbations, lung function, treatment response, and disease severity. Others have attempted to derive subtypes using laboratory parameters, such as peripheral blood or sputum immune cell count (most typically eosinophil and neutrophil counts), or percentage and fractional exhaled nitric oxide (FENO) (Wenzel 2012; Tyler and Bunyavanich 2019), characteristics which inform on underlying inflammation.

It is well known that differences in treatment related features between individuals, including therapeutic regimen, exacerbation frequency while on controller medication, and response to medication can correspond to differences in various omic profiles (Hastie et al. 2010; Loza et al. 2016; Bigler et al. 2017; Svenningsen and Nair 2017). Accordingly, several research efforts have been investigating the metabolomic profiles of the clinically relevant subgroups derived in their populations. Among these, only one reported no difference in the EBC metabolome as measured using NMR spectroscopy between their subgroups. Carro et al. could not distinguish children classified as non-severe asthma regularly treated with controller medications from those non-severe cases who were well controlled and steroid naïve based on their EBC NMR profile (Carraro et al. 2013).

In a study of 22 healthy subjects and 54 asthmatics between 18 and 70 years old, Reinke et al. classified patients into several subgroups based on their clinical characteristics and therapeutic regimens; healthy controls ($n = 22$), mild asthmatics with intermittent symptoms and only treated with β2-agonists alone ($n = 12$); moderate asthmatics with frequent symptoms and treated with ICS ($n = 20$); and severe asthmatics with persistent symptoms treated with ICS therapy ($n = 22$) or with ICS and oral corticosteroid (OCS) therapy ($n = 5$) (Reinke et al. 2017). Serum metabolomic profiling using targeted mass spectrometry revealed metabolomic differences between the asthma cases versus controls, but also differences between the asthma subgroups. Six metabolites were significantly different in between the healthy controls and the individuals in the three asthma subgroups with ICS or OCS use: DHEA-S, cortisone, ProHyp, pipecolate, N-palmitoyltaurine, and cortisol. Differences in these six metabolites were observed with even moderate ICS dosage, but with the exception of pipecolate, were greatest between the healthy controls and those who were on both ICS and OCS. This is in agreement with the findings from the steroids literature which found that cortisol and cortisone levels are lowest in asthmatics treated with high doses of exogenous steroids (Kachroo et al. 2022). To address the issue of potential confounding in their results, Reinke et al. further interrogated the data. While they did identify some metabolite shifts that were intrinsic to the disease process itself, for example Oleoylethanolamide increased with asthma severity independently of steroid treatment, they concluded that others were specific to treatment. Intriguingly, they reported that overall the association

between metabolic profile and ICS treatment is greater than that between metabolic profile and disease (Reinke et al. 2017).

In their study of the discriminatory ability of the urine metabolome for asthma phenotypes, Mattarucchi et al. addressed the question of whether the observed metabolomic changes were due to the treatment or due to the disease. They built two models, one which sought to distinguish between children with asthma who took controller medication and those who did not, and one which sought to distinguish cases from healthy controls. They determined that their controller medication model was highly effective, with only a single misclassification in a population of 42 (Mattarucchi et al. 2012). In keeping with the findings of Reinke et al., they found the discriminatory model was primarily driven by intermediates in the metabolism of C21-steroid hormones urocortisone and urocortisol. Importantly, this model was distinct from that that distinguished all 42 asthma cases from 12 age-matched controls demonstrating that the metabolomic differences between asthma cases and controls were not driven by consumption of controller medication (Mattarucchi et al. 2012).

Comhair et al. classified severe and non-severe asthmatics based on criteria outlined in the proceedings of the American Thoracic Society Workshop on Refractory Asthma, which were largely based on treatment regimen (Comhair et al. 2015). In their study all 10 severe adult asthmatics received a high dose of inhaled or oral corticosteroids either singly or in combination with long-acting beta-agonists, while the 10 non-severe asthmatics did not receive any corticosteroids or long-acting beta-agonists and received inhaled beta-agonist (rescue medication) infrequently but less than 2 times per week. A comparison of the plasma metabolomic profile of the severe and non-severe asthma cases identified 18 significant metabolites, all but one of which were lower in the severe asthmatics. Most of these metabolites were again related to steroid metabolism including 1-steraroyylglycerol, dehydroisoandrosterone sulfate, epi-androsterone sulfate, and androsterone sulfate. These findings are consistent with the suppression of adrenal steroids as a consequence of therapeutic use of corticosteroids as mentioned previously. They also observed differences in several amino acids, including beta-alanine, between the two groups of asthmatics, but further observed that the severe group shared many of the same metabolomic signatures as non-severe asthmatics, who were not on corticosteroids. The authors interpreted these findings as further evidence that corticosteroid therapy is not the driver of differences between asthma cases and healthy controls (Comhair et al. 2015). This hypothesis was further supported by Dallinga et al. also who showed that the discrimination between asthma cases and controls based on an analysis of volatile organic compound in breath samples was not driven by medication usage (Dallinga et al. 2010). They predicated this conclusion on the observation that the metabolomic products of known asthma medications were not reflected in the principal components that distinguished between asthma cases and controls.

Park et al. took a somewhat different approach, focusing on the metabolomic differences between those who respond to corticosteroids as compared to those who do not (Park et al. 2017). They performed high resolution mass spectrometry based

urinary metabolomic profiling of children with severe asthma; 15 children who did not responded to corticosteroids, defined by persistently poor asthma control with symptoms more than twice weekly and a less than 15% improvement in FEV_1 2 weeks after systemic triamcinolone administration, and 15 children who did respond. They identified 30 metabolomic features that differed significantly between the two groups, representing differences in the metabolism of tyrosine, glutathione, and the degradation of aromatic compounds. Of these five, 3,6-dihydronicotinic acid 3-methoxy-4-hydroxyphenyl(ethylene)glycol, 3,4-dihydroxy-phenylalanine, γ-glutamylcysteine, Cys-Gly, and Flavin mononucleotide were determined to contribute most to the distinction of corticosteroid responders and non-responders (Park et al. 2017). These results point to the reduced synthesis and increased degradation of the antioxidant glutathione, which in its thiol-reduced form (GSH) is the most abundant antioxidant in the airway epithelial lining fluid and plays an important role in the pathogenesis of asthma (Papamichael et al. 2021). The significant metabolites also included constituents of cigarette smoke, which were higher in the urine of corticosteroid-resistant children. This is consistent with evidence that smoking can impair the efficacy of corticosteroid treatment in asthma (Chaudhuri et al. 2003).

Fitzpatrick et al. compared the plasma metabolome of mild to moderate childhood asthmatics treated with ICS or ICS/ long-acting beta-agonists (LABA) combination therapy to that of severe asthmatics treated with high-dose ICS and LABA (Fitzpatrick et al. 2014). Interestingly, among the severe asthmatics they determined no children displayed complete corticosteroid responsiveness (defined as symptoms, lung function, bronchodilator reversibility, and exhaled nitric oxide values normalized 2 weeks after treatment with systemic triamcinolone acetonide). Among the severe asthmatics 89% demonstrated partial responsiveness, while the remainder had no discernable response, taken together these children were considered a corticosteroid refractory population. Their results indicated that the corticosteroid refractory children demonstrated a vast array of metabolic derangements relative to mild/moderate asthmatics. These differences again pointed to increased lipid peroxidation and dysregulation in thiol redox balance and oxidative stress related pathways, with dysregulation of the glycine, serine, and threonine metabolism pathway and the N-acylethanolamine and N-acyltransferase pathways observed. The results may support the hypothesis that oxidative stress is a contributory factor to corticosteroid refractory severe asthma in children. However, like many of the previous studies the authors determined that confounding by indication could not be ruled out, as individuals taking asthma medication for severe diseases are also more likely to have co-morbid conditions including obesity sinus disease and obstructive sleep apnea (Fitzpatrick et al. 2014).

6.2 Omic-Driven Endotypes (The "Top-Down" Approach)

The alternative to the bottom-up clinical subgrouping approach instead classifies patients based on the underlying mechanisms of their disease. This approach is predicated on the hypothesis that treating a disease based on its underlying biological

mechanisms may be more effective than treating it on its clinical manifestations (Svenningsen and Nair 2017; Tyler and Bunyavanich 2019). The inherently mechanistic nature of the metabolome makes it particularly compelling for the derivation of such endotypes (Tyler and Bunyavanich 2019).

Within U-BIOPRED (Unbiased Biomarkers for the PREDiction of respiratory disease outcomes), a pan-European cohort of severe asthma in adults and children, exhaled metabolomic fingerprints measured via electronic noses (eNoses) were available on 78 adults with severe asthma. Severe asthma was classified as having a prescription for high-dose ICS plus one other asthma control medication, a daily OCS prescription, two or more severe exacerbations, or a combination of these (Brinkman et al. 2019). Unsupervised clustering of these metabolite-derived profiles identified three clusters of asthmatics ($n = 26/33/19$) that differed in the clinically important metrics of circulating eosinophil and neutrophil percentages. The clusters also differed in the ratios of patients using oral corticosteroids; with the neutrophilic inflammation predominant cluster 2 containing the higher percentage of patients using OCS maintenance therapy. In contrast cluster 3 had the highest percentages of peripheral blood eosinophilia and the lowest percentage of subjects with chronic OCS use (Brinkman et al. 2019). These findings suggest that hypothesis-free clustering of individuals with asthma based on the metabolome can generate clinically meaningful subgroups that differ in their medication usage.

This hypothesis was further expanded upon by Kelly et al. in a larger study including over 2,000 children with asthma from two independent and well-characterized cohorts (Kelly et al. 2021). This allowed validation of their findings, a critical step when using metabolomic datasets and unsupervised approaches. In this study five metabolomic driven endotypes or “metabo-endotypes” were derived from unsupervised clustering of the plasma metabolome in a population of children with mild to moderate asthma. These metabo-endotypes differed in metrics of lung function and more specifically lung obstruction. There were also differences in medication uses across the five metabo-endotypes that were independent of lung function. Exploration of the metabolites contributing most to the formation of the metabo-endotypes identified multiple metabolites involved in the formation and homeostatic regulation of pulmonary surfactant, thereby pinpointing an actionable mechanism that precision therapeutics could be directed toward. Importantly, these metabo-endotypes were validated in an independent population of children with mild-moderate asthma (Kelly et al. 2021).

In contrast; Sinha et al. performed clustering on 61 asthmatics based on their EBC profile. Although they did identify phenotypic differences across these clusters, including significant differences in exacerbations and eosinophil count, they did not determine any differences in the number of patients using corticosteroids between the three clusters (Sinha et al. 2017).

Overall, the “top-down” metabolomics-driven approach seems promising, however it should be noted that given the cross-sectional nature of the sample collection for profiling and measurement of relevant clinical phenotypes, it is still challenging to determine causation; whether the profiles that are driving the clusters are driven by their need for a given therapy, or by the therapy itself. Kelly et al. reported that the

lung function related differences between their endotypes were robust to additional adjustment for medication use, suggesting medication is not driving the clusters. However, further work incorporating longitudinal data is necessary to explore this further.

6.3 Aspirin-Exacerbated Asthma

An estimated 5–15% of patients with asthma are considered to have aspirin-exacerbated respiratory disease (AERD), which is defined as the formation of nasal polyposis, the development of asthma and a respiratory reaction on ingestion of aspirin or another nonsteroidal anti-inflammatory drug (NSAID) (Haque et al. 2021). To date, only a small number of studies have utilized metabolomics to explore this subset of asthma cases characterized by their response to a common pharmacologic agent, with a focus on measured of leukotrienes. Ban et al. utilized untargeted metabolomic profiling of urine and serum from 45 AERD asthmatics and 44 aspirin-tolerant asthmatics (ATA) on urine and serum (Ban et al. 2017). Six metabolites were shown to differ between the AERD and the ATA subjects, including several members of the arachidonic acid pathway. This reflects the importance of the imbalance of the eicosanoid cascade in the pathogenesis of AERD. They further explored these results by targeted measurement of the arachidonic acid pathway and validation of their findings in an independent population. Their results suggested that serum baseline levels of Leukotriene E4 (LTE4) and LTE4/Prostaglandin F2alpha ratios can be useful diagnostic biomarkers for AERD (Ban et al. 2017).

The utility of LTE4, which provides a measure of the cysteinyl leukotrienes production associated with asthma attacks, has been further demonstrated in other studies. In a study of urinary levels in 240 cases of AERD and 226 cases of ATA significantly higher levels of urinary LTE4 were observed in the AERD population. Asthmatics with high levels of Urinary 3-bromotyrosine (uBrTyr) (> 0.101 ng/mg Cr), urinary LTE4 levels (> 800 pg/mg Cr), and blood eosinophils (> 300 cells/μl) were 7 times more likely to have AERD, however they did not observe a statistically significant difference between AERD and ATA for uBrTyr alone (Comhair et al. 2018).

Similarly, comparing the concentration of $uLTE_4$ in 247 AERD patients and 239 ATA patients, levels were significantly higher in the AERD cases, and the authors determined that at a concentration of 8,000 pg/mg creatinine,TE_4 could discriminate between the two groups with an area under the curve of 0.70 (95% CI: 0.66–0.74), a sensitivity of 49%, and a specificity of 81%. However, its discriminatory power was outperformed by a set of typical clinical parameters, calling into question its potential clinical translatability (Bochenek et al. 2018). Another study reported that at a uLTE threshold of 241 pg/mg Cr, specificity was even better at 92% (Divekar et al. 2016), but both these studies lacked validation of the predictive ability of their models in an independent population.

7 Prevention

A growing body of literature is focused on the prevention of asthma. Given the early life origins of this condition in most asthmatics much of this work is concentrated on maternal or prenatal interventions (Gern et al. 1999). These have largely been in the form of nutritional supplementations, such as vitamins D, C, and E (Litonjua et al. 2020), fish oil (Bisgaard et al. 2016), folic acid (Veeranki et al. 2015), pre and probiotics (Cuello-Garcia et al. 2016), and antioxidants (Gref et al. 2017). However, the results have been somewhat contradictory in terms of the efficacy of these interventions in preventing asthma (Gur et al. 2017). Here again, pharmacometabolomics, which in this instance can be expanded to include the use of metabolomics for understanding differing responses to the same nutritional intervention, can be leveraged to try and understand the mechanisms of success, or otherwise.

To date, there has been little work on preventative pharmacometabolomics of asthma. However, the existing literature does show that preventative measures administered to mothers can influence their offspring's metabolome (Blighe et al. 2017; Rago et al. 2019). For example, increased exposure to vitamin D in utero via supplementation has been found to be associated with lower concentrations of fatty acids, in particular linoleate, linolenate (18:3n–3/3n–6), myristate, oleate, palmitate, palmitoleate (16:1n–7), and stearate (18:0) in infants (Blighe et al. 2017). Similarly, prenatal supplementation with fish oil has been shown to result in lower levels of the n-6 LCPUFA pathway-related metabolites and saturated and monounsaturated long-chain fatty acids-containing compounds, lower levels of metabolites of the tryptophan pathway, and higher levels of metabolites in the tyrosine and glutamic acid pathway in early life (Rago et al. 2019). It has also been demonstrated that vitamin D and fish oil interventions can reduce the risk of asthma among offspring (Bisgaard et al. 2016; Wolsk et al. 2017). Given it is known that many of these altered metabolites play a direct role in the biological pathways and processes that underlie the development and progression of asthma, such as oxidative stress and inflammation, this leads to the hypothesis that an altered offspring metabolomic milieu resulting from a prenatal intervention may be mediating the association between that intervention and asthma risk. Indeed, a prenatal fish oil supplementation related metabolic profile at age 6 months was significantly associated with a reduced risk of asthma by age 5 and this profile could explain 24% of the observed asthma-protective effect of this supplementation (Rago et al. 2019). Similarly, the metabolomic changes in children accompanying vitamin D supplementation of their mothers have been linked to the reduced risk of asthma (Blighe et al. 2017).

However, these relationships can be complicated by the genetic profile of both the mother and child. Sphingolipids, which among their multitude of roles act as signaling molecules involved in immune response, inflammation, have been suggested to mediate the relationship between maternal vitamin D supplementation and child asthma risk (Kelly et al. 2019a). Vitamin D metabolites have been shown to be capable of activating the sphingolipid pathway and decreased synthesis of sphingolipids has been associated with increased airway hyperresponsiveness and

inflammation (Kelly et al. 2019a). Work within the VDAART (Vitamin D Antenatal Asthma Reduction Trial) prenatal supplementation trial seemed to confirm this relationship with one important caveat. There was no increase in child sphingolipid levels and therefore no protective effect for asthma among children with a key variant in a functional SNP in the region of *ORMLD3*, which regulates a rate-limiting step in the de novo synthesis of sphingolipids (Kelly et al. 2019a). Correspondingly, vitamin D supplementation conferred no protective effect for the children of mothers with this same variant (Knihtilä et al. 2021).

Further studies are required to explore the role that pharmacometabolomics may be able to play in the prevention of asthma.

8 Challenges of Pharmacometabolomics in Asthma

The utility of metabolomics to obtain insights into pathophysiology is clear (Fiehn and Kim 2014; Wishart 2019), and this utility has been demonstrated repeatedly for asthma (Kelly et al. 2017). The evidence that pharmacometabolomics can help to identify dysregulated metabolomic markers of disease, leading to more personalized and efficacious treatment options is growing. However, several important challenges remain. Many of these challenges pertain broadly to the field of metabolomics. While a battery of spectroscopic technologies is used in metabolomics, these are continually undergoing advancements for more accurate and improved resolution. Though NMR spectroscopy provides a reduced resolution in comparison with mass spectroscopy, the technique is highly reproducible and robust, not requiring separation or derivatization of molecules, factors that can sometimes hinder true resolution in mass spectroscopy with poor sample preparation or quality (Dona et al. 2014). Regardless of the technology used, the fact that the complexity and magnitude of the metabolomic map has yet to be completely characterized means that pertinent findings or critical information may still be missed (Franklin and Vondriska 2011; Mulvihill and Nomura 2014). While metabolite databases are continually being updated, curation and accuracy of identified metabolites remains an issue. A majority of studies use untargeted profiling which by definition does not include chemical standards and so only measures relative abundances of even those metabolites that have been characterized (Bowler et al. 2017), making it difficult to compare and combine studies.

Further issues relate to the fact that, although efforts are ongoing, there remains an overall lack of standardization in approaches, measurements, and analysis for metabolomic data (Bowler et al. 2017). One particular point of contention is the issue of imputing missing data, different studies take different approaches rendering comparison of the results challenging (Wei et al. 2018). Even more problematic is the fact that, as with several the studies included here, some do not even report upon the method used. However, it should be noted that such issues are not unique to metabolomics studies and are common to most epidemiological and particularly omic studies. The issue of multiple testing correction is similarly of concern to most omic studies, but is arguably a particular challenge for metabolomics given the

collinear nature of metabolites within coregulated biological pathways (Wei et al. 2018; Peluso et al. 2021). This is again evident among the pharmacometabolomics of asthma studies; while some applied the most stringent approaches for multiple testing, namely Bonferroni correction, others did not account for multiple testing at all, leading to a disconnect in the results presented and a complexity in comparing them. The highly correlated nature of metabolites within coregulated biological pathways has the additional effect of complicating biological interpretation and understanding (Bowler et al. 2017). Similarly, the dynamic nature of the metabolome and its sensitivity to environmental influences, which are in many ways one of the greatest strengths of metabolomics, can be considered a weakness when trying to deconvolute causation versus correlation and deal with confounding.

In addition to these broad metabolomic and epidemiological issues, there are also challenges that are applicable to asthma pharmacometabolomics, more specifically. Chiefly, due to the relatively recent development of the field, the body of literature is somewhat sparse, and several studies are based on very small sample sizes. The optimal biosample on which to conduct metabolomic profiling for pharmacometabolomic studies in individuals with asthma remains to be determined and various biosamples have been employed to date including sputum, plasma, serum, and exhaled breath condensate. The lung as the organ most proximal to the disease is of particular interest and can be considered to be uniquely informative in a metabolomic framework, but it is also uniquely challenging, with particular issues relating to the invasiveness of sample collection and the likelihood of sample contamination (Bowler et al. 2017). Similarly EBC, which has been widely used in the asthma pharmacometabolomics literature to date is limited by the lack of standardization in collection and the issue of appropriate dilution (Bowler et al. 2017). Consequently, much of the literature is focused on blood and urine; findings have been promising, but their relevance to lung disease has been at times called into question (Bowler et al. 2017). This issue is magnified when the tissue type of greatest relevance to the pharmacological agents and their biological processes may not be the same as that most relevant for asthma. Thus, the use of different biosamples across the asthma pharmacometabolomics literature further complicates its overall assessment.

Of the current pharmacometabolomics literature most studies are cross-sectional in nature, rendering the determination of a causal pathway between medications and metabolomic effects challenging. While several studies reported that the observed relationships between medication and therapy were independent of asthma, asthma phenotypes and severity, others cautioned that they could not rule out confounding by indication. Regardless, the majority of reported associations remain to be validated in independent populations or utilizing a randomized control design. Additional well-designed, large-scale longitudinal studies are required to address this further. Crucially, these studies need to recognize and consider the heterogeneous nature of asthma.

9 Conclusions

In conclusion, the field of pharmacometabolomics is an emerging area of research, with evidence that offers genuine clinical translational potential. Yet, much work remains to be done. The literature to date is somewhat disparate in terms of study design, biosample, characteristics of included asthma cases, therapeutics of interest, statistical approach, and metabolomic profiling technique.

Nevertheless, some common themes have emerged (Fig. 2). The largest number of studies focus on ICS use, and in several studies spanning different biosamples and metabolomic profiling techniques, it has been reported that adrenal suppression is a potential side effect of long-term ICS use. This is particularly important given recent changes to the GINA guidelines, which propose a shift from SABAs to ICS, given the risks of SABA-only treatment and SABA overuse (Reddel et al. 2021). Interestingly, the potential harms of SABA do not seem to be a focus of the current asthma pharmacometabolomics literature. However, as these new guidelines were only published at the end of 2021, we may see an increase in SABA related pharmacometabolomics literature moving forward.

There was also a consistent theme of oxidative stress, although there was some disagreement as to whether this was a consequence of the treatment or was instead related to phenotype. Further work is needed to clarify. There was similarly a lack of consensus regarding lactate. The discriminatory ability of leukotrienes for aspirin-exacerbated asthma was more definitive, although it should be noted that these were not metabolome-wide studies, rather they focused only on one or a small number of compounds.

The literature does suggest a growing interest in endotyping in asthma, which leverages the heterogeneity of this condition to derive subgroups based on mechanism. The small number of studies to date provide some compelling results and support the notion that endotyping may represent a particularly rich resource for the identification of therapeutic markers targeted to the mechanisms of a specific subgroup of disease (Hunt et al. 2002; Sinha et al. 2012, 2017; Brinkman et al. 2019). Endotypes derived via the metabolome are uniquely positioned for clinical

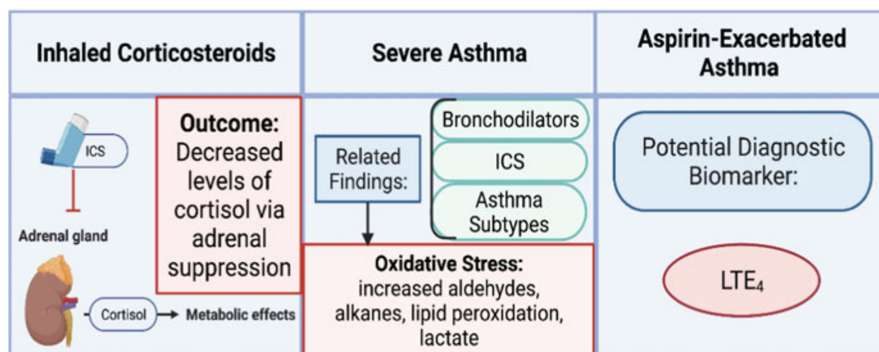


Fig. 2 Summary of findings

translational potential through both the assignment of individuals to subgroups that can receive treatment targeted to the mechanisms of their disease and through the identification of the therapeutic targets within those subgroups underlying that treatment (Kelly et al. 2021). However, it should be cautioned that the endotypes reported here are a long way from being used in clinical practice, and to date there are no examples of endotypes derived via any omic technology being used to determine treatment. Further investigation, including diverse validation populations, assessment of stability over time, and targeted quantification of metabolites that may have biomarker or therapeutic target potential, is vital (Kelly et al. 2021).

Pharmacometabolomics can be considered the natural successor of pharmacogenetics, which has several demonstrated successes in asthma, although validation and translation into clinical practice still remains somewhat of a bottleneck (Kersten and Koppelman 2017). It could therefore be hypothesized that combining these two technologies and other omics, such as proteomics, epigenetics, in a pharmaco-multi-omics approach could offer improved understanding of mechanisms in a systems biology framework. However, to date no such studies exist. This represents an area of untapped potential that is ripe for exploration.

A further development that could help propel the field of asthma pharmacometabolomics forward is advancements in the handling and analysis of electronic medical record data which can provide detailed information on medical and treatment history on huge numbers of well-phenotyped individuals. Similarly, exploiting the potential of pre-existing biobanks could support studies of individuals with asthma in far greater numbers than seen in the literature to date. This could also help to provide more diverse populations allowing further exploration of the influences of factors such as age sex and smoking status, that have been flagged in the literature as influencing the relationship between therapy and the metabolome.

Pharmacometabolomics is a nascent field, requiring further development and exploration. Although much can be taken from the parallel ongoing advancements in the field of metabolomics, there are many other issues that pertain specifically to the investigation of medications and therapies, which need to be addressed by the pharmacometabolomics community. Only then can we consider precision medicine initiatives underlined by pharmacometabolomics, as a realistic approach to combat the substantial public health burden of asthma.

References

Augstein J, Farmer JB, Lee TB et al (1973) Selective inhibitor of slow reacting substance of anaphylaxis. Nat New Biol 245:215–217. https://doi.org/10.1038/newbio245215a0

Ban G-Y, Cho K, Kim S-H et al (2017) Metabolomic analysis identifies potential diagnostic biomarkers for aspirin-exacerbated respiratory disease. Clin Exp Allergy 47:37–47. https://doi.org/10.1111/cea.12797

Barnes PJ (2006) Corticosteroid effects on cell signalling. Eur Respir J 27:413–426. https://doi.org/10.1183/09031936.06.00125404

Bigler J, Boedigheimer M, Schofield JPR et al (2017) A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts. Am J Respir Crit Care Med 195:1311–1320. https://doi.org/10.1164/rccm.201604-0866OC
Bisgaard H (2001) Pathophysiology of the cysteinyl leukotrienes and effects of leukotriene receptor antagonists in asthma. Allergy 56(Suppl 66):7–11. https://doi.org/10.1034/j.1398-9995.56.s66.2.x
Bisgaard H, Stokholm J, Chawes BL et al (2016) Fish oil-derived fatty acids in pregnancy and wheeze and asthma in offspring. N Engl J Med 375:2530–2539. https://doi.org/10.1056/NEJMoa1503734
Blighe K, Chawes BL, Kelly RS et al (2017) Vitamin D prenatal programming of childhood metabolomics profiles at age 3 y. Am J Clin Nutr 106:1092–1099. https://doi.org/10.3945/ajcn.117.158220
Bochenek G, Stachura T, Szafraniec K et al (2018) Diagnostic accuracy of urinary LTE4 measurement to predict aspirin-exacerbated respiratory disease in patients with asthma. J Allergy Clin Immunol Pract 6:528–535. https://doi.org/10.1016/j.jaip.2017.07.001
Bowler RP, Wendt CH, Fessler MB et al (2017) New strategies and challenges in lung proteomics and metabolomics. An official American Thoracic Society workshop report. Ann Am Thorac Soc 14:1721–1743. https://doi.org/10.1513/AnnalsATS.201710-770WS
Bozelli JC, Azher S, Epand RM (2021) Plasmalogens and chronic inflammatory diseases. Front Physiol 12. https://doi.org/10.3389/fphys.2021.730829
Brinkman P, Wagener AH, Hekking P-P et al (2019) Identification and prospective stability of electronic nose (eNose)-derived inflammatory phenotypes in patients with severe asthma. J Allergy Clin Immunol 143:1811–1820.e7. https://doi.org/10.1016/j.jaci.2018.10.058
Bruno F, Spaziano G, Liparulo A et al (2018) Recent advances in the search for novel 5-lipoxygenase inhibitors for the treatment of asthma. Eur J Med Chem 153:65–72. https://doi.org/10.1016/j.ejmech.2017.10.020
Carraro S, Giordano G, Reniero F et al (2013) Asthma severity in childhood and metabolomic profiling of breath condensate. Allergy 68:110–117. https://doi.org/10.1111/all.12063
Carstairs JR, Nimmo AJ, Barnes PJ (1985) Autoradiographic visualization of beta-adrenoceptor subtypes in human lung. Am Rev Respir Dis 132:541–547. https://doi.org/10.1164/arrd.1985.132.3.541
Cazzola M, Page CP, Calzetta L, Matera MG (2012) Pharmacology and therapeutics of bronchodilators. Pharmacol Rev 64:450–504. https://doi.org/10.1124/pr.111.004580
Chaudhuri R, Livingston E, McMahon AD et al (2003) Cigarette smoking impairs the therapeutic response to oral corticosteroids in chronic asthma. Am J Respir Crit Care Med 168:1308–1311. https://doi.org/10.1164/rccm.200304-503OC
Chung KF, Wenzel SE, Brozek JL et al (2014) International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. Eur Respir J 43:343–373. https://doi.org/10.1183/09031936.00202013
Clish CB (2015) Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harb Mol Case Stud 1:a000588. https://doi.org/10.1101/mcs.a000588
Comhair SAA, McDunn J, Bennett C et al (2015) Metabolomic endotype of asthma. J Immunol 195:643–650. https://doi.org/10.4049/jimmunol.1500736
Comhair SAA, Bochenek G, Baicker-McKee S et al (2018) The utility of biomarkers in diagnosis of aspirin exacerbated respiratory disease. Respir Res 19:210. https://doi.org/10.1186/s12931-018-0909-6
Cuello-Garcia CA, Fiocchi A, Pawankar R et al (2016) World allergy organization-McMaster University guidelines for allergic disease prevention (GLAD-P): prebiotics. World Allergy Organ J 9:10. https://doi.org/10.1186/s40413-016-0102-7
Dallinga JW, Robroeks CMHHT, van Berkel JJBN et al (2010) Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. Clin Exp Allergy 40:68–76. https://doi.org/10.1111/j.1365-2222.2009.03343.x

Darveaux J, Busse WW (2015) Biologics in asthma – the next step toward personalized treatment. J Allergy Clin Immunol Pract 3:152–160.; quiz 161. https://doi.org/10.1016/j.jaip.2014.09.014
Divekar R, Hagan J, Rank M et al (2016) Diagnostic utility of urinary LTE4 in asthma, allergic rhinitis, chronic rhinosinusitis, nasal polyps, and aspirin sensitivity. J Allergy Clin Immunol Pract 4:665–670. https://doi.org/10.1016/j.jaip.2016.03.004
Dona AC, Jiménez B, Schäfer H et al (2014) Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem 86:9887–9894. https://doi.org/10.1021/ac5025039
Eder W, Ege MJ, von Mutius E (2006) The asthma epidemic. N Engl J Med 355:2226–2235. https://doi.org/10.1056/NEJMra054308
Enilari O, Sinha S (2019) The global impact of asthma in adult populations. Ann Glob Health 85. https://doi.org/10.5334/aogh.2412
Ferraro VA, Carraro S, Pirillo P et al (2020) Breathomics in asthmatic children treated with inhaled corticosteroids. Metabolites 10. https://doi.org/10.3390/metabo10100390
Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 48: 155–171
Fiehn O, Kim J (2014) Metabolomics insights into pathophysiological mechanisms of interstitial cystitis. Int Neurourol J 18:106–114. https://doi.org/10.5213/inj.2014.18.3.106
Fitzpatrick AM, Park Y, Brown LAS, Jones DP (2014) Children with severe asthma have unique oxidative stress-associated metabolomic profiles. J Allergy Clin Immunol 133(258–61):e1–e8. https://doi.org/10.1016/j.jaci.2013.10.012
Foster A, Chan CC (1991) Peptide leukotriene involvement in pulmonary eosinophil migration upon antigen challenge in the actively sensitized guinea pig. Int Arch Allergy Appl Immunol 96: 279–284. https://doi.org/10.1159/000235508
Franklin S, Vondriska TM (2011) Genomes, proteomes, and the central dogma. Circ Cardiovasc Genet 4:576. https://doi.org/10.1161/CIRCGENETICS.110.957795
GBD 2015 Chronic Respiratory Disease Collaborators (2017) Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 5:691–706. https://doi.org/10.1016/S2213-2600(17)30293-X
Gern JE, Lemanske RF, Busse WW (1999) Early life origins of asthma. J Clin Invest 104:837–843. https://doi.org/10.1172/JCI8272
Gref A, Rautiainen S, Gruzieva O et al (2017) Dietary total antioxidant capacity in early school age and subsequent allergic disease. Clin Exp Allergy 47:751–759. https://doi.org/10.1111/cea.12911
Grennan D, Wang S (2019) Steroid side effects. JAMA 322:282–282. https://doi.org/10.1001/JAMA.2019.8506
Gur M, Hakim F, Bentur L (2017) Better understanding of childhood asthma, towards primary prevention – are we there yet? Consideration of pertinent literature. F1000Res 6(2152): 10.12688/f1000research.11601.1
Gurnell M, Heaney LG, Price D, Menzies-Gow A (2021) Long-term corticosteroid use, adrenal insufficiency and the need for steroid-sparing treatment in adult severe asthma. J Intern Med 290:240–256. https://doi.org/10.1111/joim.13273
Haque R, White AA, Jackson DJ, Hopkins C (2021) Clinical evaluation and diagnosis of aspirin-exacerbated respiratory disease. J Allergy Clin Immunol 148:283–291. https://doi.org/10.1016/j.jaci.2021.06.018
Hargreave FE, Nair P (2009) The definition and diagnosis of asthma. Clin Exp Allergy 39:1652–1658. https://doi.org/10.1111/j.1365-2222.2009.03321.x
Hastie AT, Moore WC, Meyers DA et al (2010) Analyses of asthma severity phenotypes and inflammatory proteins in subjects stratified by sputum granulocytes. J Allergy Clin Immunol 125:1028–1036.e13. https://doi.org/10.1016/j.jaci.2010.02.008

Henderson WR, Lewis DB, Albert RK et al (1996) The importance of leukotrienes in airway inflammation in a mouse model of asthma. J Exp Med 184:1483–1494. https://doi.org/10.1084/jem.184.4.1483

Hunt JF, Erwin E, Palmer L et al (2002) Expression and activity of pH-regulatory glutaminase in the human airway epithelium. Am J Respir Crit Care Med 165:101–107. https://doi.org/10.1164/ajrccm.165.1.2104131

Jones KA, Perkins WJ, Lorenz RR et al (1999) F-actin stabilization increases tension cost during contraction of permeabilized airway smooth muscle in dogs. J Physiol 519(Pt 2):527–538. https://doi.org/10.1111/j.1469-7793.1999.0527m.x

Kachroo P, Sordillo JE, Lutz SM et al (2021) Pharmaco-metabolomics of inhaled corticosteroid response in individuals with asthma. J Pers Med 11:1148. https://doi.org/10.3390/jpm11111148

Kachroo P, Stewart I, Kelly R et al (2022) The systematic use of metabolomic epidemiology, biobanks, and electronic medical records for precision medicine initiatives in asthma: findings suggest new guidelines to optimize treatment. https://doi.org/10.21203/rs.3.rs-268507/v1

Kaddurah-Daouk R, Weinshilboum RM (2014) Pharmacometabolomics: implications for clinical pharmacology and systems pharmacology. Clin Pharmacol Therap 95:154–167. https://doi.org/10.1038/clpt.2013.217

Kannisto S, Korppi M, Remes K, Voutilainen R (2001) Serum dehydroepiandrosterone sulfate concentration as an indicator of adrenocortical suppression in asthmatic children treated with inhaled steroids. J Clin Endocrinol Metab 86:4908–4912. https://doi.org/10.1210/jcem.86.10.7975

Kannisto S, Laatikainen A, Taivainen A et al (2004) Serum dehydroepiandrosterone sulfate concentration as an indicator of adrenocortical suppression during inhaled steroid therapy in adult asthmatic patients. Eur J Endocrinol 150:687–690. https://doi.org/10.1530/eje.0.1500687

Kelly RS, Dahlin A, McGeachie MJ et al (2017) Asthma metabolomics and the potential for integrative omics in research and the clinic. Chest 151:262–277. https://doi.org/10.1016/j.chest.2016.10.008

Kelly RS, Chawes BL, Guo F et al (2019a) The role of the 17q21 genotype in the prevention of early childhood asthma and recurrent wheeze by vitamin D. Eur Respir J 54. https://doi.org/10.1183/13993003.00761-2019

Kelly RS, Sordillo JE, Lutz SM et al (2019b) Pharmacometabolomics of bronchodilator response in asthma and the role of age-metabolite interactions. Metabolites 9. https://doi.org/10.3390/metabo9090179

Kelly RS, Mendez KM, Huang M et al (2021) Metabo-endotypes of asthma reveal differences in lung function: discovery and validation in two TOPMed cohorts. Am J Respir Crit Care Med. https://doi.org/10.1164/rccm.202105-1268OC

Kersten ETG, Koppelman GH (2017) Pharmacogenetics of asthma: toward precision medicine. Curr Opin Pulm Med 23:12–20. https://doi.org/10.1097/MCP.0000000000000335

Knihtilä HM, Kelly RS, Brustad N et al (2021) Maternal 17q21 genotype influences prenatal vitamin D effects on offspring asthma/recurrent wheeze. Eur Respir J 58. https://doi.org/10.1183/13993003.02012-2020

Kuruvilla ME, Lee FE-H, Lee GB (2019) Understanding asthma phenotypes, endotypes, and mechanisms of disease. Clin Rev Allergy Immunol 56:219–233. https://doi.org/10.1007/s12016-018-8712-1

Larsson K, Kankaanranta H, Janson C et al (2020) Bringing asthma care into the twenty-first century. NPJ Prim Care Respir Med 30:25. https://doi.org/10.1038/s41533-020-0182-2

Lewis LM, Ferguson I, House SL et al (2014) Albuterol administration is commonly associated with increases in serum lactate in patients with asthma treated for acute exacerbation of asthma. Chest 145:53–59. https://doi.org/10.1378/chest.13-0930

Litonjua AA, Carey VJ, Laranjo N et al (2020) Six-year follow-up of a trial of antenatal vitamin D for asthma reduction. N Engl J Med 382:525–533. https://doi.org/10.1056/NEJMoa1906137

Loureiro CC, Duarte IF, Gomes J et al (2014) Urinary metabolomic changes as a predictive biomarker of asthma exacerbation. J Allergy Clin Immunol 133:261-3.e1–261-3.e5. https://doi.org/10.1016/j.jaci.2013.11.004

Loza MJ, Djukanovic R, Chung KF et al (2016) Validated and longitudinally stable asthma phenotypes based on cluster analysis of the ADEPT study. Respir Res 17:165. https://doi.org/10.1186/s12931-016-0482-9

Matera MG, Page CP, Calzetta L et al (2020) Pharmacology and therapeutics of bronchodilators revisited. Pharmacol Rev 72:218–252. https://doi.org/10.1124/pr.119.018150

Mattarucchi E, Baraldi E, Guillou C (2012) Metabolomics applied to urine samples in childhood asthma; differentiation between asthma phenotypes and identification of relevant metabolites. Biomed Chromatogr 26:89–94. https://doi.org/10.1002/bmc.1631

McGeachie MJ, Dahlin A, Qiu W et al (2015) The metabolomics of asthma control: a promising link between genetics and disease. Immun Inflamm Dis 3:224–238. https://doi.org/10.1002/iid3.61

Moore WC, Meyers DA, Wenzel SE et al (2010) Identification of asthma phenotypes using cluster analysis in the severe asthma research program. Am J Respir Crit Care Med 181:315–323. https://doi.org/10.1164/rccm.200906-0896OC

Mulvihill MM, Nomura DK (2014) Metabolomic strategies to map functions of metabolic pathways. Am J Physiol Endocrinol Metab 307:E237–E244. https://doi.org/10.1152/ajpendo.00228.2014

Newton R, Giembycz MA (2016) Understanding how long-acting β2 -adrenoceptor agonists enhance the clinical efficacy of inhaled corticosteroids in asthma – an update. Br J Pharmacol 173:3405–3430. https://doi.org/10.1111/bph.13628

Pandya D, Puttanna A, Balagopal V (2014) Systemic effects of inhaled corticosteroids: an overview. Open Respir Med J 8:59–65. https://doi.org/10.2174/1874306401408010059

Papamichael MM, Katsardis C, Sarandi E et al (2021) Application of metabolomics in pediatric asthma: prediction, diagnosis and personalized treatment. Metabolites 11. https://doi.org/10.3390/metabo11040251

Paredi P, Kharitonov SA, Barnes PJ (2000) Elevation of exhaled ethane concentration in asthma. Am J Respir Crit Care Med 162:1450–1454. https://doi.org/10.1164/ajrccm.162.4.2003064

Park YH, Fitzpatrick AM, Medriano CA, Jones DP (2017) High-resolution metabolomics to identify urine biomarkers in corticosteroid-resistant asthmatic children. J Allergy Clin Immunol 139:1518–1524.e4. https://doi.org/10.1016/j.jaci.2016.08.018

Peluso A, Glen R, Ebbels TMD (2021) Multiple-testing correction in metabolome-wide association studies. BMC Bioinform 22:67. https://doi.org/10.1186/s12859-021-03975-2

Peters-Golden M, Henderson WR (2007) Leukotrienes. N Engl J Med 357:1841–1854. https://doi.org/10.1056/NEJMra071371

Quan-Jun Y, Jian-Ping Z, Jian-Hua Z et al (2017) Distinct metabolic profile of inhaled budesonide and salbutamol in asthmatic children during acute exacerbation. Basic Clin Pharmacol Toxicol 120:303–311. https://doi.org/10.1111/bcpt.12686

Rago D, Rasmussen MA, Lee-Sarwar KA et al (2019) Fish-oil supplementation in pregnancy, child metabolomics and asthma risk. EBioMedicine 46:399–410. https://doi.org/10.1016/j.ebiom.2019.07.057

Reddel HK, Bacharier LB, Bateman ED et al (2021) Global initiative for asthma (GINA) strategy 2021 – executive summary and rationale for key changes. J Allergy Clin Immunol Pract. https://doi.org/10.1016/j.jaip.2021.10.001

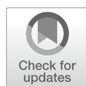
Reinke SN, Gallart-Ayala H, Gómez C et al (2017) Metabolomics analysis identifies different metabotypes of asthma severity. Eur Respir J 49. https://doi.org/10.1183/13993003.01740-2016

Rodrigo GJ (2014) Serum lactate increase during acute asthma treatment: a new piece of the puzzle. Chest 145:6–7. https://doi.org/10.1378/chest.13-2042

Rodrigo GJ, Rodrigo C (2005) Elevated plasma lactate level associated with high dose inhaled albuterol therapy in acute severe asthma. Emerg Med J 22:404–408. https://doi.org/10.1136/emj.2003.012039

Ruffin RE, McIntyre E, Crockett AJ et al (1982) Combination bronchodilator therapy in asthma. J Allergy Clin Immunol 69:60–65. https://doi.org/10.1016/0091-6749(82)90089-6

Sahiner UM, Birben E, Erzurum S et al (2011) Oxidative stress in asthma. World Allergy Organ J 4: 151–158. https://doi.org/10.1097/WOX.0b013e318232389e
Sinha A, Krishnan V, Sethi T et al (2012) Metabolomic signatures in nuclear magnetic resonance spectra of exhaled breath condensate identify asthma. Eur Respir J 39:500–502. https://doi.org/10.1183/09031936.00047711
Sinha A, Desiraju K, Aggarwal K et al (2017) Exhaled breath condensate metabolome clusters for endotype discovery in asthma. J Transl Med 15:262. https://doi.org/10.1186/s12967-017-1365-7
Smith LJ, Shamsuddin M, Houston M (1993) Effect of leukotriene D4 and platelet-activating factor on human alveolar macrophage eicosanoid and PAF synthesis. Am Rev Respir Dis 148:682–688. https://doi.org/10.1164/ajrccm/148.3.682
Sordillo JE, Lutz SM, Kelly RS et al (2020) Plasmalogens mediate the effect of age on bronchodilator response in individuals with asthma. Front Med (Lausanne) 7:38. https://doi.org/10.3389/fmed.2020.00038
Svenningsen S, Nair P (2017) Asthma Endotypes and an overview of targeted therapy for asthma. Front Med (Lausanne) 4:158. https://doi.org/10.3389/fmed.2017.00158
Tyler SR, Bunyavanich S (2019) Leveraging -omics for asthma endotyping. J Allergy Clin Immunol 144:13–23. https://doi.org/10.1016/j.jaci.2019.05.015
Veeranki SP, Gebretsadik T, Mitchel EF et al (2015) Maternal folic acid supplementation during pregnancy and early childhood asthma. Epidemiology 26:934–941. https://doi.org/10.1097/EDE.0000000000000380
Walter MA, Linsley PS, Cox DW (1987) Apa I polymorphism of a human immunoglobin VH3 subclass locus. Nucleic Acids Res 15:4697. https://doi.org/10.1093/nar/15.11.4697
Wang et al (2021) Research progress of metabolomics in asthma. Metabolites. PMID: 34564383
Wedes SH, Wu W, Comhair SAA et al (2011) Urinary bromotyrosine measures asthma control and predicts asthma exacerbations in children. J Pediatr 159:248–55.e1. https://doi.org/10.1016/j.jpeds.2011.01.029
Wei R, Wang J, Su M et al (2018) Missing value imputation approach for mass spectrometry-based metabolomics data. Sci Rep 8:663. https://doi.org/10.1038/s41598-017-19120-0
Wendell SG, Baffi C, Holguin F (2014) Fatty acids, inflammation, and asthma. J Allergy Clin Immunol 133:1255–1264. https://doi.org/10.1016/j.jaci.2013.12.1087
Wendell SG, Fan H, Zhang C (2020) G protein-coupled receptors in asthma therapy: pharmacology and drug action. Pharmacol Rev 72:1–49. https://doi.org/10.1124/pr.118.016899
Wenzel SE (2012) Asthma phenotypes: the evolution from clinical to molecular approaches. Nat Med 18:716–725. https://doi.org/10.1038/nm.2678
Werner M, Jordan PM, Romp E et al (2019) Targeting biosynthetic networks of the proinflammatory and proresolving lipid metabolome. FASEB J 33:6140–6153. https://doi.org/10.1096/fj.201802509R
Wishart DS (2019) Metabolomics for investigating physiological and pathophysiological processes. Physiol Rev 99:1819–1875. https://doi.org/10.1152/physrev.00035.2018
Wolsk HM, Chawes BL, Litonjua AA et al (2017) Prenatal vitamin D supplementation reduces risk of asthma/recurrent wheeze in early childhood: a combined analysis of two randomized controlled trials. PLoS One 12:e0186657. https://doi.org/10.1371/journal.pone.0186657
Yaghoubi M, Adibi A, Safari A et al (2019) The projected economic and health burden of uncontrolled asthma in the United States. Am J Respir Crit Care Med 200:1102–1112. https://doi.org/10.1164/rccm.201901-0016OC
Yokomizo T, Nakamura M, Shimizu T (2018) Leukotriene receptors as potential therapeutic targets. J Clin Invest 128:2691–2701. https://doi.org/10.1172/JCI97946
Zhu Z, Camargo CA, Hasegawa K (2019) Metabolomics in the prevention and management of asthma. Expert Rev Respir Med 13:1135–1138. https://doi.org/10.1080/17476348.2019.1674650



Prospective Metabolomic Studies in Precision Medicine: The AKRIBEA Project

Maider Bizkarguenaga, Rubén Gil-Redondo, Chiara Bruzzone, Ganeko Bernardo-Seisdedos, Ana Laín, Beatriz González-Valle, Nieves Embade, José M. Mato, and Oscar Millet

Contents

1 Metabolomics as a Tool for Precision Medicine 276
1.1 Precision Medicine and Molecular Medicine 276
1.2 Biomarkers and Metabolomics 277
1.3 Description of the Accessible NMR Metabolic Landscape 278
1.4 On the Cohort Size for Applications in Precision Medicine 282
2 Large-Cohort Sample Collection and Biobanking for Metabolomic Studies 283
3 Standard Operating Procedures for Metabolomics in Precision Medicine 285
4 Artificial Intelligence and Data Mining: Data Is the Way 286
4.1 Data Management 286
4.2 Data Analysis 288
4.3 Biological Contextualization 289
4.4 From the Laboratory to the Clinic 289
5 Applications of Precision Medicine: Non-oriented and Oriented Metabolomics Analysis 290
5.1 Non-oriented Metabolomics Study: Metabolic Syndrome 290
5.2 Oriented Metabolomics Study: COVID-19 293
6 Concluding Remarks 295
References 295

Abstract

For a long time, conventional medicine has analysed biomolecules to diagnose diseases. Yet, this approach has proven valid only for a limited number of metabolites and often through a bijective relationship with the disease

M. Bizkarguenaga · R. Gil-Redondo · C. Bruzzone · G. Bernardo-Seisdedos · A. Laín · B. González-Valle · N. Embade · J. M. Mato · O. Millet (✉)
Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Basque Research and Technology Alliance, Bizkaia Science and Technology Park, Derio, Bizkaia, Spain
e-mail: mbizcarguenaga@cicbiogune.es; rgil@cicbiogune.es; cbruzzone@cicbiogune.es; gbernardo.atlas@cicbiogune.es; alain@cicbiogune.es; bvalle@cicbiogune.es; nembade@cicbiogune.es; director@cicbiogune.es; omillet@cicbiogune.es

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*, Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_610

(i.e. glucose relationship with diabetes), ultimately offering incomplete diagnostic value. Nowadays, precision medicine emerges as an option to improve the prevention and/or treatment of numerous pathologies, focusing on the molecular mechanisms, acting in a patient-specific dimension, and leveraging multiple contributing factors such as genetic, environmental, or lifestyle. Metabolomics grasps the required subcellular complexity while being sensitive to all these factors, which results in a most suitable technique for precision medicine. The aim of this chapter is to describe how NMR-based metabolomics can be integrated in the design of a precision medicine strategy, using the Precision Medicine Initiative of the Basque Country (the AKRIBEA project) as a case study. To that end, we will illustrate the procedures to be followed when conducting an NMR-based metabolomics study with a large cohort of individuals, emphasizing the critical points. The chapter will conclude with the discussion of some relevant biomedical applications.

Keywords

Biomarker · NMR metabolomics · Precision medicine

1 Metabolomics as a Tool for Precision Medicine

1.1 Precision Medicine and Molecular Medicine

A silent revolution is steadily transforming traditional medicine into the novel field of *personalized medicine*, which tailors medical treatment to the individual characteristics of each patient (Kohler et al. 2017). Currently, medical practice is mostly based in evidence, organized at the organ level and administering a rather uniform treatment for a given disease. In contrast, precision medicine concentrates on the mechanistic aspects of the disease, is organized at the cellular and molecular levels, and aims to provide the most suitable personalized treatment for each patient. As a consequence of this innovative strategy, personalized medicine focuses on early detection and prospective evaluation, while canonical medical praxis is based on statistically based retrospective analyses to provide curative and palliative care.

To achieve this goal, the concept of *precision medicine*, first used in 2009 in the book *The Innovator's Prescription* by Clayton Christensen (Christensen et al. 2009), exploits the analysis of clinical and molecular data of patient subpopulations, which tries to classify a subject into a healthy individual or a patient by contrasting biological information with datasets from large population cohorts, enriched in biological diversity. This strategy allows to endotype a particular disease, ultimately leading to personalized care (Śliwczyński and Orlewska 2016). Furthermore, the generation of prediction models for prevention and diagnosis paves the way for application of precision medicine in a rationally designed drug treatment for genetically or metabolically stratified groups of patients (Authors 2021). In fact, this strategy not only tailors the appropriate drug treatment for each patient (Tranvåg

et al. 2021) but also provides an explanation for the non-responding population in clinical phase trials of drugs, where patient sub-classification also actively contributes to the better understanding of the mode-of-action of drugs (John et al. 2020; Puchades-Carrasco and Pineda-Lucena 2017).

As already mentioned, traditional medicine is rooted on medical evidence (measurable symptomatology), from which only a limited fraction of it arises from biochemical analyses (i.e. *molecular medicine*): glucose, cholesterol, high-density lipoprotein/low-density lipoprotein, transaminases, or creatinine, among others, to monitor diabetes, cardiovascular, liver, and renal disorders, respectively. Perhaps, the most powerful diagnostic use of metabolites currently occurs in the heel test to screen for neonatal congenital methabolopathies (Beger et al. 2016). In contrast, precision medicine heavily relies on genomics, proteomics, and metabolomics for an appropriate biochemical characterization of the individual. To that end, it integrates molecular information originated by all the abovementioned "omic" techniques with the medical records, environmental surroundings, diet, and lifestyle of the population under consideration. Altogether, this approach can provide predictive, diagnostic, and prognostic markers for a plethora of diseases, informing on their underlying molecular mechanisms and enabling their sub-classification and the stratification of patients based on the metabolic pathways involved (Śliwczyński and Orlewska 2016; Clish 2015).

1.2 Biomarkers and Metabolomics

Precision medicine is intimately coupled to biomarker validation. Indeed, biomarkers have a great potential as a diagnostic tool for the clinicians and to help deciding optimal treatments, dietary restrictions, or other major therapeutic lifestyle changes (Kohler et al. 2017; Beger et al. 2016). Biomarkers can be classified into two categories: *dynamic biomarkers*, that can monitor disease progression and treatment responses, and *static biomarkers*, used as a prognostic tool for the abovementioned conditions (Kohler et al. 2017; Hartl et al. 2021). In this context, metabolomics is based on the measurement of the low-molecular-weight molecules and metabolites of different cellular processes of a biological sample and has a crucial role in defining the molecular data that can have an impact in precision medicine (Kohler et al. 2017). Metabolomics quantitatively characterizes the intermediate- and end-products of the biochemical processes, which not only are associated to genetic diversity, but also to environmental events, and disease conditions or treatment responses (Kohler et al. 2017; Puchades-Carrasco and Pineda-Lucena 2017; Clish 2015). Furthermore, it provides a time-specific definition of the metabolic state since metabolome is in constant adaptation to gene expression, nutrient and drug intake, other external influences, variations in the gut microbiota and, importantly, to the natural history of the diseases.

A proper implementation of metabolomics into precision medicine and clinical practice optimally implies that the biomarkers monitorization must be conducted in a non-invasive and accessible matrix (i.e. biofluids). In theory, a biofluid should be

proximal to the organic focus of the disease but, in reality, most diseases also have significant systemic manifestations (Hartl et al. 2021). In this context, serum, plasma, and urine are the most suitable biofluids, characterized by optimal sensitivity to the time-dependent metabolome, disponible nature, and easy storage. Plasma and serum metabolomes undergo limited daily variation, also providing information of the extracellular metabolome, whereas urine is the most on-demand available biofluid (Kohler et al. 2017).

Nowadays, metabolomic research technology enables the identification and quantification of hundreds of distinct metabolites from biofluids, using technologies that can be easily extrapolated to large-scale cohorts of donors and thousands of samples. For the sake of reproducibility, Standardized Operating Procedures (SOPs) must be implemented for sample collection, storage, handling, analytical analyses, and data processing (Beger et al. 2016). Under these premises, research laboratories with standardized infrastructures will generate dataset that become *additive*, enabling large-scale and multicentred studies in metabolomics. Techniques that nowadays can fulfil these requirements include Nuclear Magnetic Resonance spectroscopy (NMR), Liquid Chromatography coupled to Mass Spectrometry (LC-MS), and Gas Chromatography with Mass Spectrometry (GC-MS). In here, we will focus on NMR spectroscopy, which is able to simultaneously quantify hundreds of metabolites in a biological sample with an acquisition time of less than 20 min.

1.3 Description of the Accessible NMR Metabolic Landscape

The human metabolome comprehends up to 100,000 metabolites that participate in more than 18,000 enzymatic reactions, which have to be in constant healthy equilibrium (Wishart et al. 2018). Metabolic phenotyping of biofluids and polar/apolar tissue extracts are powerful ways to monitor metabolite concentrations (Bernardo-Seisdedos et al. 2021; Dona et al. 2014). In 1974, ^{13}C-NMR spectroscopy was first used in metabolic studies to investigate metabolic processes (Wilson et al. 1974). Since then, NMR metabolomics experienced a huge development thanks to its ability to characterize the chemical composition of complex biological mixtures (Wishart 2019). Actually, in a few minutes, a one-dimensional proton NMR spectrum can detect tenths of molecules, which are in the micromolar concentration range, typically including amino acids, carbohydrates, alcohols, organic acids, the intermediates or end-products of the main metabolic pathways, and also the overall lipid composition of a biosystem (reviewed in (Vignoli et al. 2019)).

The metabolic landscape that is accessible to NMR depends on the investigated matrix and the experimental choice (i.e. the active nucleus to be monitored) (Fig. 1). Urine is an especially well suited biofluid for metabolomics since it perfectly reflects the body's metabolism. In addition, as a biological waste material, it typically contains catabolic products from a wide range of foods, drinks, drugs, environmental contaminants, endogenous waste metabolites, and bacterial by-products. Urine NMR analysis can routinely quantify over 100 metabolites including glucose and other carbohydrates, lipids, aromatic amino acids, and p-cresol sulphate, maltitol,

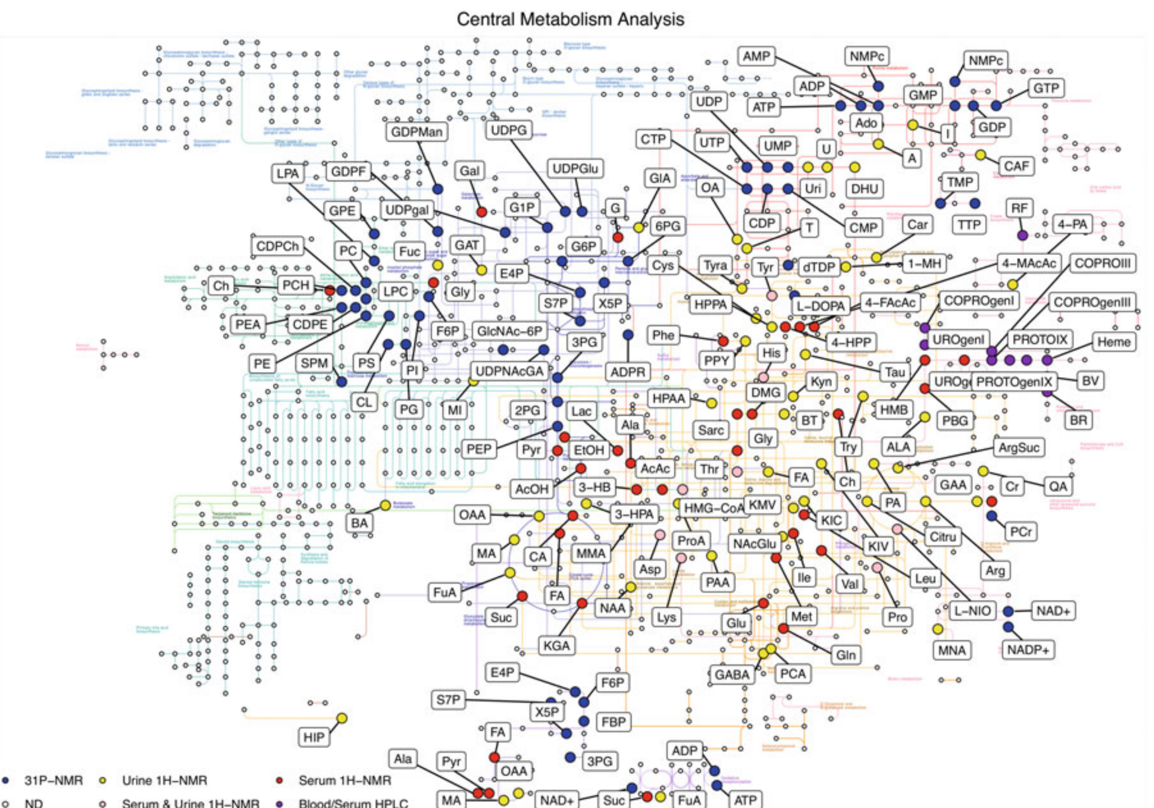


Fig. 1 Central metabolism pathways. NMR analysis allows the characterization of certain metabolites that are spread all over the pathways. The colour code represents the biofluid or methodology sensitive to the detected metabolite (see legend). The scheme is based on KEGG hsa01100 map. Abbreviations: (R)-3-Hydroxybutyric acid, 3-HB; (S)-3-Hydroxy-3-methylglutaryl-CoA, HMG-CoA; (S)-3-Methyl-2-oxopentanoic acid, KMV; 1,2-Diacylglycerol, LPC; 1-Methylnicotinamide, MNA; 2,5-Dihydroxyphenylacetic acid, 4-HPP; 2-Hydroxyphenylacetic acid, HPAA; 2-Hydroxytricarballylic acid, CA; 2-Oxobutanoic acid; FA; 2-Phospho-D-glycerate, 2PG; 3,4-Dihydroxy-L-phenylalanine, L-DOPA; 3',5'-Cyclic AMP, NMPc; 3',5'-Cyclic GMP, NMPc; 3-Methyl-2-oxobutanoic acid, KIV; 3-Phospho-D-glycerate, 3PG; 4-Fumarylacetoacetic acid, 4-FAcAc; 4-Maleylacetoacetate, 4-MAcAc; 4-Methyl-2-

among others (Fig. 1). The abnormal quantity of some of these metabolites can be associated with clinical pathologies and prevalent syndromes (Bruzzone et al. 2021). Additionally, the urine metabolome can also be scrutinized for the absence of metabolites, to confirm that the individual does not suffer a congenital metabolopathy, as an alternative to the newborn screening heel test (Embade et al. 2019).

Serum is the blood moiety devoid of cell and clotting factors, rich in proteins, and has been used in numerous diagnostic tests. Serum is metabolically very homeostatic and over 50 components can be easily identified and quantified by NMR spectroscopy (Fig. 1). Yet, the added value of NMR-based analysis of serum is the lipoprotein profiling (Raftery, "Quantitative NMR methods in metablomics"). Lipoproteins are a group of micelles consisting of lipids and proteins whose main function is to

Fig. 1 (continued) oxopentanoate, KIC; 4-Pyridoxate, 4-PA; 5,6-Dihydrouracil, DHU; 5-Aminolevulinate, ALA; 6-Phospho-D-gluconate, 6PG; Acetic acid, AcOH; Acetoacetic acid, AcAc; Adenine, A; Adenosine, Ado; Adenosine diphosphate, ADP; ADPribose, ADPR; Alanine, Ala; alpha-D-Galactose, Gal; alpha-Ketoglutaric acid, KGA; Adenosine monophosphate, AMP; Apolipoproteins, Apo; Arginine, Arg; Asparagine, Asp; Adenosine triphosphate, ATP; beta-D-Glucose, G; beta-D-Glucose 6-phosphate, G6P; Betaine, BT; Bilirubin, BR; Biliverdin, BV; Butyric acid, BA; C00570, CDPE; C04475, GPE; C05977, LPA; Caffeine, CAF; Carnosine, Car; Cytidine diphosphate, CDP; Choline, Ch; Choline phosphate, PCH; Cytidine monophosphate, CMP; Coproporphyrin III, COPROIII; Coproporphyrinogen I, COPROgenI; Coproporphyrinogen III, COPROgenIII; Creatine, Cr; Cytidine triphosphate, CTP; Cytidine 5'-diphosphocholine, CDPCh; D-Erythrose 4-phosphate, E4P; D-Fructose 1,6-bisphosphate, FBP; D-Fructose 6-phosphate, F6P; D-Gluconic acid, GlA; D-Glucose 1-phosphate, G1P; Diphosphatidylglycerol, CL; D-Sedoheptulose 7-phosphate, S7P; D-Sedoheptulose 7-phosphate, S7P; Deoxythymidine diphosphate dTDP; D-Xylulose 5-phosphate, X5P; Ethanol, EtOH; Ethanolamine phosphate, PEA; Ethylenesuccinic acid, Suc; Formic acid, FA; Fumaric acid, FuA; Galactitol, GAT; gamma-Aminobutyric acid, GABA; Guanosine diphosphate, GDP; GDP-L-fucose, GDPF; GDPmannose, GDPMan; Glutamic acid, Glu; Glutamine, Gln; Glycerol, Gly; Glycine, Gly; Guanosine monophosphate, GMP; Guanosine triphosphate, GTP, Guanidinoacetate, GAA; Heme, Heme; Hippurate, HIP; Histidine, His; Hydracrylic acid, 3-HPA; Hydroxymethylbilane, HMB; Inosine, I; Isoleucine, Ile; L-2-Hydroxybutanedioic acid, MA; L-Argininosuccinate, ArgSuc; L-Citrulline, Citru; L-Cystine, Cys; Leucine, Leu; L-Fucose, Fuc; L-Kynurenine, Kyn; L-Lactic acid, Lac; L-Ornithine, L-NIO; Lysine, Lys; Methionine, Met; Methylmalonic acid, MMA; myo-Inositol, MI; N(pai)-Methyl-L-histidine, 1-MH; N,N-Dimethylglycine, DMG; N-Acetyl-D-glucosamine 6-phosphate, GlcNAc-6P; N-Acetyl-L-aspartate, NAA; N-Acetyl-L-glutamic acid, NAcGlu; Nicotinamide adenine dinucleotide, NAD^+; Nicotinamide adenine dinucleotide phosphate, $NADP^+$; Oxosuccinic acid, OAA; Pantothenic acid, PA; Phenylacetic acid, PAA; Phenylalanine, Phe; Phenylpyruvate, PPY; Phosphatidylcholine, PC; Phosphatidylethanolamine, PE; Phosphatidylglycerol, PG; Phosphatidylinositol, PI; Phosphatidylserine, PS; Phosphocreatine, PCr; Phosphoenolpyruvate, PEP; p-Hydroxyphenylpyruvic acid, HPPA; Pidolic acid, PCA; Porphobilinogen, PBG; Proline, Pro; Propionic acid, ProA; Protoporphyrin, PROTOIX; Protoporphyrinogen IX, PROTOgenIX; Pyroracemic acid, Pyr; Quinolinic acid, QA; Riboflavin, RF; Sarcosine, Sarc; Sphingomyelin, SPM; Taurine, Tau; Thiamine monophosphate, TMP; Thiamine triphosphate, TTP; Threonine, Thr; Thymine, T; Tryptophan, Try; Tyramine, Tyra; Tyrosine, Tyr; Uridine diphosphate, UDP; UDP-D-galactose, UDPgal; UDPglucose, UDPG; UDPglucuronate, UDPGlu; UDP-N-acetyl-D-glucosamine, UDPNAcGA; Uridine monophosphate, UMP; Uracil, U; Uracil-6-carboxylic acid, OA, Uridine, Uri; Uroporphyrinogen I, UROgenI; Uroporphyrinogen III, UROgenIII; Uridine triphosphate, UTP; Valine, Val

transport insoluble plasma lipids such as cholesterol or triglycerides from absorption and/or synthesis organs such as the intestine or liver to the places where they need to be processed or used. These are chylomicrons, very low-density lipoproteins (VLDL), intermediate density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL) (Martin et al. 2015).

Proteins found in these compounds or apolipoproteins are necessary for the assembly, structure, and function of lipoproteins. They activate key enzymes for the metabolism of lipoproteins and are also capable of acting as ligands for cell surface receptors. ApoA-I, which is synthesized in the liver and intestine, is found in almost all HDL particles while ApoA-II is the second most abundant HDL apolipoprotein. ApoB is the main structural protein of chylomicrons, VLDL, IDL, and LDL and only one ApoB molecule is present in each lipoprotein particle. ApoB is the only apolipoprotein that does not transfer between different lipoprotein particles. ApoE is present in chylomicrons, VLDL, and IDL, and plays a critical role in the metabolism of triglyceride-rich particles. Three apolipoproteins of the C series (ApoC-I, ApoC-II, and ApoC-III) are also involved in the metabolism of triglyceride-rich lipoproteins.

NMR has been extensively used to identify the different classes of lipoproteins (Otvos et al. 1992) and to elucidate and quantify subclasses of the known lipoproteins (Jiménez et al. 2018). This strategy has proven especially useful to unravel the molecular mechanisms associated to cardiovascular risk, metabolic syndrome (Bruzzone et al. 2021), or infections such as SARS-CoV-2 virus (Bruzzone et al. 2020a).

One caveat for the metabolic characterization is its intrinsic overwhelming complexity. In this regard, our group hypothesized that the analysis of human metabolome could be simplified by just analysing the phosphorylated chemical compounds (Bernardo-Seisdedos et al. 2021). Due to the strategic deployment of phosphorylated metabolites (Fig. 1), the analysis of the phosphorylated moiety of the metabolism reduces complexity at an acceptable functional information loss. Specifically, phosphorylated metabolites occupy a prominent position in all metabolic and catabolic pathways that could be used as readouts or reporters of the balance between glycolysis, the tricarboxylic acid cycle (TCA), gluconeogenesis, the pentose phosphate pathway (PPP), the oxidative phosphorylation, and phospholipid metabolism. To this aim, we developed a strict SOP for the extraction of mouse liver and hepatocellular carcinoma tissues although it could be applied to different tissues, fluids, or biological systems. The detection and quantification of multiple hydrophilic and hydrophilic phosphorylated metabolites relies on ^{31}P-NMR–based methodology. This procedure worked both with murine and human liver samples and provides the absolute concentration of up to 54 phosphorylated metabolites (Fig. 1). Apart from the already mentioned pathways, obtained from the hydrophilic phase, the hydrophobic (lipophilic) phase provides information on the phospholipid metabolism such as phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylinositol (PI), phosphatidylserine (PS), sphingomyelin (SM), and cardiolipin (CL) (Bernardo-Seisdedos et al. 2021).

1.4 On the Cohort Size for Applications in Precision Medicine

Metabolomic studies investigating human populations can be carried out as a single interventional study or by longitudinal studies (Nabi et al. 2021; Mignot et al. 2021; Carrat et al. 2021; Ruiz et al. 2016), where different samples are recollected over time from the same individual. Ideally, the cohort must be large enough to ensure statistical power and to embrace the natural metabolic variability. Actually, a proper planning, optimal quality controls as well as unified data collection are essential to minimize biases. That said, working with large cohorts is challenging due to a number of reasons that include complex logistics to ensure the proper sample recollection and handling (i.e. abiding the SOPs), volunteer' adherence or change in their lifestyle during the longitudinal studies, and the associated economic burden, which may span over several years of sample collection. All these complications may explain why most of the studies in the literature (for metabolomics and other omics as well) scrutinized cohorts with a limited number of participants. Of course, exceptional studies with a large number of samples can also be found, including a very complete study (genomics, epigenomics, transcriptomics, genomics, and metabolomics) of a Dutch cohort of more than 1,000 participants (Tigchelaar et al. 2015). Another example targeted the Spanish Mediterranean population (1,011 individuals) to investigate genome-wide association studies (GWAS) (Ortega-Azorín et al. 2019). Also, the LIFE-Adult-Study recently completed the baseline examination of 10,000 randomly selected participants in Germany with the aim of investigating prevalence, markers, genetic predisposition, and lifestyle associated to the most common diseases (Loeffler et al. 2015). For this purpose, they used a mixture of the usual clinical trials for the diagnosis of numerous diseases combined with genomics and metabolomics. Finally, one of the large-scale projects developed in recent years is the Tohoku Medical Megabank (TMM) with the aim of conducting prospective cohort studies in several of the regions that were affected by the 2011 Japan earthquake (Koshiba et al. 2018). During the first stage of these studies (from 2013 to 2017), more than 150,000 participants have been successfully recruited.

At CIC bioGUNE we are developing a precision medicine initiative, vehiculated through a longitudinal study (AKRIBEA) that targets a cohort of 10,000 individuals from the Basque Country. The study is non-oriented and the only inclusion criteria is to belong to the working population of the Mondragón Cooperative (with about 40,000 employees), since the volunteer recruitment and the sample collection are carried out during their annual medical check-up. Available data includes general characteristics, biochemical and lifestyle data, the NMR-based identification and quantification of metabolites in urine and serum, the lipoprotein analysis of serum also obtained by NMR spectroscopy and, to a lesser extent, the genotyping and proteomic analysis of the population. This sample dataset can be interpreted as a reference cohort to investigate a plethora of diseases and syndromes, including metabolic syndrome, prostate cancer, Chron´s disease, Non-Alcoholic Steatohepatitis (NASH), and COVID-19, among others.

Due to the abovementioned recruitment characteristics, the cohort is geographically and culturally constrained, but it has been cross validated by comparing it to

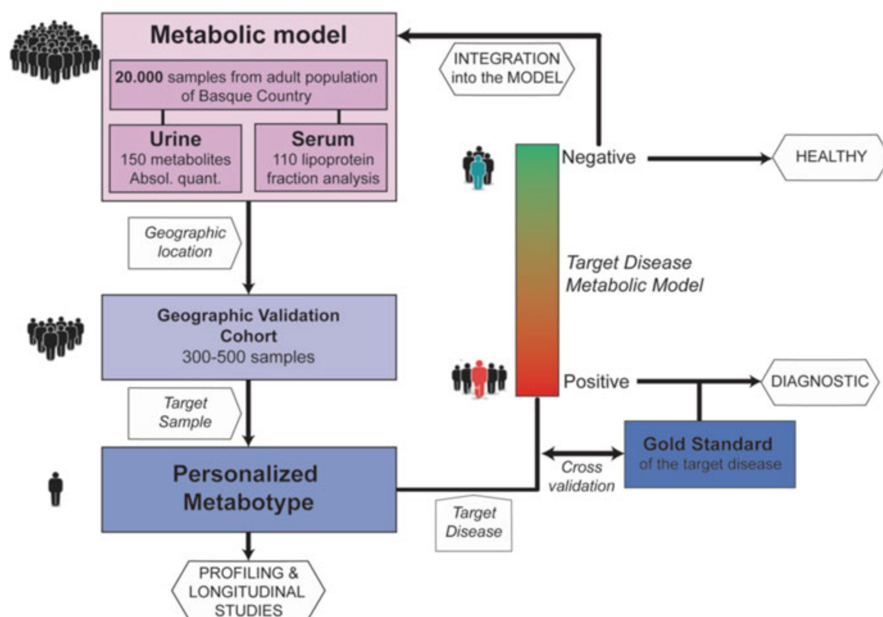


Fig. 2 Flowchart of the AKRIBEA initiative in the context of precision medicine

smaller dataset of samples from other unrelated cohorts measured under the same SOPs. There are two main applications of this large dataset (Fig. 2): (1) integral analysis of the dataset to characterize the cohort (i.e. environmental factors and habits, disease biomarkers for prevalent disease such as metabolic syndrome, density of rare diseases, big data and artificial intelligence analyses oriented to medicine and drug discovery), and (2) to compare the reference dataset to samples from smaller cohorts of patients from the pathology under study. Such cohorts must have been previously validated using an independent gold-standard methodology. In this chapter, we will use the AKRIBEA project as a case study to do an in-depth description of the required steps needed in precision medicine studies with large cohorts. Moreover, we will present illustrative examples for the two abovementioned exploration strategies: metabolic syndrome and COVID-19.

2 Large-Cohort Sample Collection and Biobanking for Metabolomic Studies

When considering large cohorts of subjects, it is essential to ensure collaboration between the different involved entities, the biobank coordination, and the implementation of a comprehensive, yet rigorous, methodology. In this context, the AKRIBEA project is a long-standing collaborative effort between CIC bioGUNE, the Basque Biobank, and Osarten Kooperatiba Elkartea, the prevention service of the

Mondragón Cooperative, whose mission is to promote the safety and health of the workers of the associated companies.

In projects involving human samples, ethical and government permits are required before starting the recollection of biofluids. It is important to plan this step quite in advance due to the considerable associated bureaucracy. For the AKRIBEA project, the approval was granted by the Euskadi Drug Research Ethics Committee (CEIm-E), an independent and multidisciplinary agency whose main purpose is to ensure the protection of the rights, safety, and well-being of people who participate in a biomedical research project, guaranteeing the correct application of ethical, methodological, and legal principles. At this stage, it is important to properly specify the intended use of the data in the informed consent and whether the data will be used in a coded or anonymized manner. It is convenient to circulate quite in advance a summary of the project among the potential donors. This document should describe the goals of the project and be redacted in an understandable way (i.e. an easy-to-read leaflet, distributed with the internal correspondence of the Corporation). Finally, a database must be created to correctly link the samples with the metadata (De Souza and Greenspan 2013). This is of particular importance in longitudinal studies.

Along AKRIBEA project, samples are collected using homogeneous disposable material and under the same conditions for all the involved patients: fasting conditions for serum and first urine in the morning. Extracted blood is conserved in the form of serum or plasma. Serum is richer in metabolites concentration, but at the expense of higher variability associated to the clotting procedure. Thus, pre-analytical SOPs (Vignoli et al. 2019) become critical as they properly define the clotting time and temperature, among other variables. In the case of plasma extraction, the use of anticoagulant is needed but this can lead also to the presence of contaminants in the samples (González-Domínguez et al. 2020) and, for example, EDTA becomes largely observable in the NMR spectra. In turn, urine samples are centrifuged to eliminate cell residuals and maintained at low temperature (4°C) for the shortest time possible until freezing to avoid bacterial contamination and to guarantee the stability of the metabolites present in this biofluid.

Automatization is key to minimize variability. In this line, the German National Cohort health study (https://www.dkfz.de/en/epidemiologie-krebserkrankungen/units/NAKO_Studienzentrum_eng/NAKO_engl.html), that will follow the long-term medical histories of 200,000 participants over 25–30 years, has been equipped with pipetting robots and all the aliquots are carried out automatically. In the AKRIBEA project, the efficiency of the personnel associated to the biobank has been instrumental to fulfil this goal.

In principle, biobanks can collect and store all kinds of human samples for analysis on a large scale. Samples must be collected and stored in a way that makes them suitable for metabolomic analysis, typically frozen at −80°C. Previous studies have demonstrated that negligible changes can be observed after long-term stored samples (Yin et al. 2015; Loo et al. 2020). Yet, a crucial point is to ensure dividing specimens into aliquots adjusting the final volumes and number of aliquots from every sample to the requirements of the project. This precaution will minimize

the number of freezing/thawing cycles experienced by the sample, which has drastic adverse effects in the metabolic integrity of the biofluids (Lodge et al. 2021a).

3 Standard Operating Procedures for Metabolomics in Precision Medicine

Standardized experiments use extensively validated NMR pulse programs and acquisition parameters to enable data exchange worldwide in order to undertake large-scale studies. Many SOPs have been specially designed to guarantee time and interlaboratory reproducibility (Wishart 2019; Vignoli et al. 2019; Emwas et al. 2019; Ghini et al. 2019; Monsonis Centelles et al. 2017), critically compared in an excellent recent review (Vignoli et al. 2019).

One valid strategy to ensure reproducibility and data transferability is to adopt analytical in vitro diagnostic research (IVDr) SOPs which, in NMR metabolomics studies imply the daily calibration of the spectrometer for temperature (MeOD sample), stability, water suppression and magnetic field homogeneity (2mM sucrose sample), and quantification performance (QuantRef routine) (Bruzzone et al. 2020a), prior to the sample measurement.

The SOPs also normally adapt to the given biological matrix. Urine samples are defrosted at room temperature for 30 min followed by centrifugation (at 4°C, for 5 min at 6,000 rpm). The supernatant is mixed with a specific buffer (1:10 buffer-to-sample ratio) to ensure pH stability and transferred into a 5 mm NMR tube for the measurement (Bruzzone et al. 2020b). For this biofluid, two experiments are acquired at 300 K: a one-dimensional ^{1}H spectrum with water presaturation for metabolite quantification (*noesypr1d*) and a two-dimensional (2D) J-resolved ^{1}H spectrum (*jresgpprqf*). In turn, serum samples are defrosted, mixed with phosphate buffer (1:1 ratio) and transferred into a 5 mm NMR tube for the measurement, where three different experiments are acquired at 310K: a Carr–Purcell–Meiboom–Gill (CPMG) spin-echo experiment (*cpmgpr*) and the *noesypr1d* and *jresgpprqf* experiments (Dona et al. 2014).

Chosen routine experiments are always time-optimized to enable high-throughput acquisition. However, it is very convenient to sparsely interleave 2D ^{1}H,^{1}H- TOCSY (Total Correlation SpectroscopY) for a subset of samples, for metabolite identification purposes.

Standard processing of spectra is followed by a quality control analysis that includes the monitorization of several parameters (i.e. linewidth, water suppression, proper referencing) and a comparison to a reference sample (obtained from a pool), which is measured by duplicate daily. In general, it may be necessary to perform a normalization to correct potential bias from several sources: batch effects, different machines, experimental conditions, etc. Typically, this is done either by division by total spectrum intensity or by a quantitative normalization with respect to a reference compound/signal (i.e. creatinine or DSS).

4 Artificial Intelligence and Data Mining: Data Is the Way

4.1 Data Management

In order to be applicable to metabolomic studies, the NMR spectra must be unambiguously linked to the relevant metainformation. This is typically done using unique identifiers (ID), which can be a single attribute or a combination of them. Yet, it is necessary to design a relational system with enough flexibility to deal with the multiplicity problem.

Since the focus is on precision medicine, the data focuses on people (donors or patients). An internal ID is normally associated to each donor, which also contains the invariant information (i.e. gender, ethnic, etc.). On the other hand, samples of different type (i.e.: urine, serum) and/or collected at different moments are associated to their own IDs, which is associated to the specific person ID via a timestamp (collection time, Fig. 3). This time stamp also includes anthropometric information (weight, height, etc.), questionaries, medical records, and all the time-dependent metadata. Finally, data must be curated for outliers (i.e. incompatible types, mixed units, typos, etc.).

In a similar way, NMR data also requires hierarchical organization and integration, as shown in Fig. 4. NMR data (the fid) can be processed in different ways (Processing types) that shall be stored independently, when necessary. In the AKRIBEA project, processed data goes through several quantification algorithms (i.e. one for small metabolites and another for blood lipoproteins) to yield a list of quantified metabolites with their values and associated units.

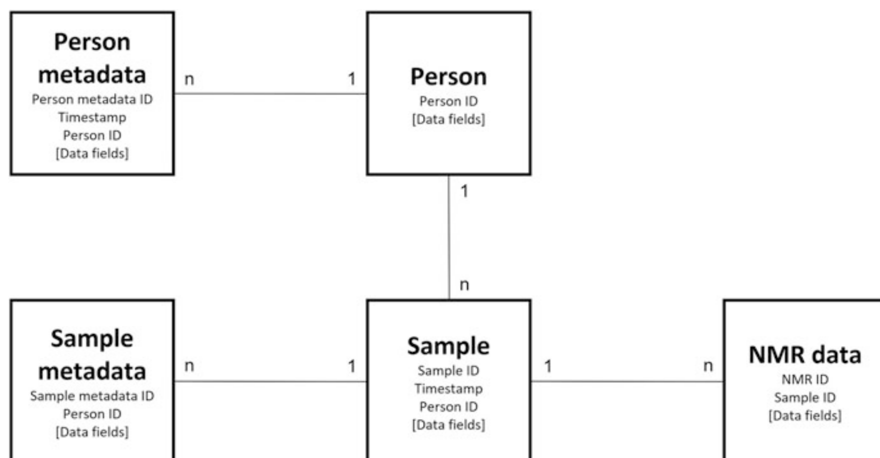


Fig. 3 Overview of main data entities and their relations

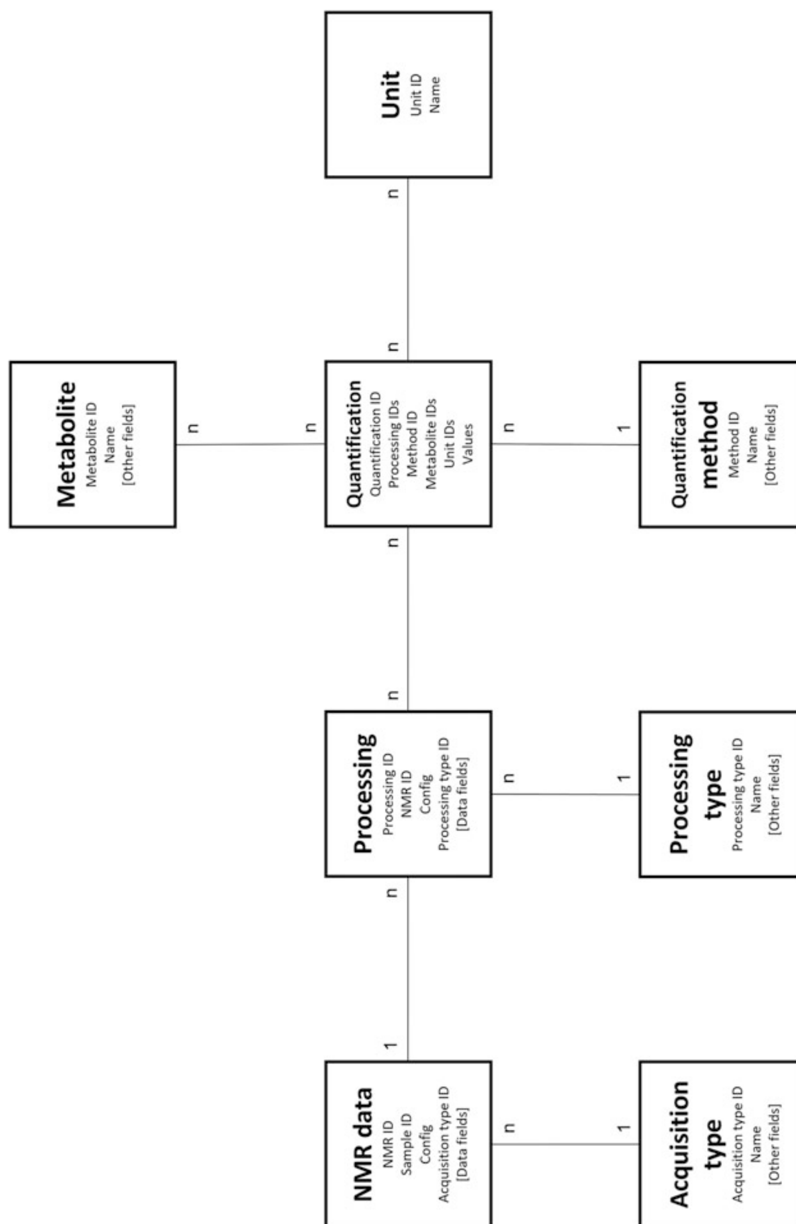


Fig. 4 Flowchart for NMR data acquisition and processing

4.2 Data Analysis

Data analysis may include unsupervised analysis (i.e. segmentation of the database into two or more non-previously defined subgroups which can be mutually compared) or supervised analysis (the segmentation groups have been previously defined according to metadata). The type of analysis may, respectively, help to respond to the questions whether there are metabolic differences in given classes or whether it is possible to identify subgroups of different metabolic profiles. Such analyses can be performed on each variable independently (univariate) or using two or more variables together (multivariate).

Univariate analysis involves the unidimensional analysis of a single variable, which usually results in straightforward biological interpretation. The analysis includes effect size (i.e. directional variation of the metabolite concentration) and the statistical significance of the variation, normally expressed as p-value. For the sake of comparison, effect size is normally expressed in fold-change as binary logarithm units. In univariate analysis, it is important to take precautions to control Type I error (i.e. the false positive rate) in multiple testing and to properly account for the confounding factors in the analysis, when necessary.

Multivariate analysis involves the concurrent consideration of several variables. In here, the distinction between unsupervised and supervised analyses becomes even stronger. The unsupervised multivariate analysis is often carried out by principal component analysis (PCA), a method for dimensionality reduction that comprises the maximum data variability in their first components. PCA also allows the visual representation of multidimensional data, helps to identify patterns and that can be combined with clustering algorithms (hierarchical or not) to discover groups of individuals with similar characteristics. In turn, multivariate supervised analyses identify the group or a specific value for each individual/sample (dependent variable) and fit the input (independent) variables to a model that explains/predicts the dependent variable. In metabolomics, it is common to use linear regression models or PLS (Partial Least Squares) because of their explanatory capacity through their coefficients. Instead, for classification problems, PLS Discriminant Analysis (PLS-DA) or orthogonal PLS-DA (OPLS-DA) is widely used (Verpoorte, "Natural products drug discovery: on silica or in-silico?"; Wishart, "Practical Aspects of NMR-Based Metabolomics"). Other machine learning methods such as support vector machines (SVM), random forest (RF), or neural networks are also commonly used.

In supervised techniques it is important to pay attention to method performance evaluation and tuning. Building a model is an iterative process where parameters shall be adjusted, and results must be evaluated. Sensitivity, specificity, accuracy, Area Under the Receiver Operating Characteristic curve (AUROC), coefficient of determination (R2), and Root Mean Square Error (RMSE) are among the performance metrics that can be used alone or in combination.

4.3 Biological Contextualization

The final goal for prospective any metabolomic study with human cohorts is to try to adequately contextualize the observed metabolic changes within the underlying biochemical pathways. Biological interpretation of metabolic changes is challenging and, consequently, largely depends on knowledge and experience. Yet, nowadays, there are databases and tools that help to support this task. They are mostly based on network analysis, where the main representative is metabolic pathway analysis. These computer programs use the identified metabolites (from metabolomics or indirectly from other omics) to pinpoint and analyse the involved pathways at both, the topological and the enrichment level. These analyses pave the way to formulate new hypothesis based on the experimental data interpretation, which can be further contrasted by searching for additional information in available online databases or by designing validation experiments.

4.4 From the Laboratory to the Clinic

Of course, the final goal of any putative biomarker is to end up validated and being used by clinicians for the diagnostic and/or prognostic of diseases. In general, for application in the clinic, the results should be validated at multiple levels, including analytical validation and using samples from many different sources. The validation method may involve conducting large corroboratory studies by directly testing each candidate biomarker with an independently collected new set of samples from the target population that will comprise a minimum of three groups (control, targeted disease group, and targeted disease group with comorbidities). Since this study includes significantly larger set of samples, the applied analytical techniques must have high-throughput capacity as, for instance, NMR can provide. Finally, it is worth noting that validated biomarkers are not directly used as direct diagnostic methods, but they can act as diagnostic tools to round-up the diagnostic.

An important technical aspect is that for a validated biomarker the methodology used for its determination has to fulfil the In Vitro Diagnostic (IVD) and/or the In Vitro Medical-devices Regulation (IVDR), a new harmonized regulatory framework to ensure the safety and performance of in vitro diagnostic medical devices on the European market. This implementation is technically complex and often requires to obtain several international and national certifications. Such regulatory standards will undoubtedly modify the laboratory operativity, accounting for many aspects such as the identification of the persons responsible for regulatory compliance, the implementation of unique device identification, and post-market surveillance, if that is finally the case. The implementation of this procedures often requires a risk analysis (typically a SWOT analysis, Strengths, Weaknesses, Threats, and Opportunities) for all processes that allows anticipating and minimizing the probability of failures and/or errors in the procedures. Moreover, the laboratory has to comply with some regulations such as ISO 31000 Risk Management standard and ISO 14971: Risk Management System for Medical Devices.

Unfortunately, the elaborated procedures and the difficult access to large cohort studies have prevented an avenue of validated biomarkers. In fact, in the past years, 150,000 studies have been conducted on biomarker, but only 100 biomarkers have been validated and implemented for clinical use.

5 Applications of Precision Medicine: Non-oriented and Oriented Metabolomics Analysis

5.1 Non-oriented Metabolomics Study: Metabolic Syndrome

The project "Metabotype of the active labour population of the Basque Country through the metabolomic analysis of serum and urine by NMR" recruited serum and urine samples from 10,000 subjects belonging to the working population of the Basque Country. This sample repository was scrutinized to generate a metabolomic (and to a less extent also a genomic) profiling for this community, that could set up the basis for studying different pathologies impacting in the metabolism. In here, we describe the application of the metabolomic analysis to investigate the Metabolic Syndrome (MetS).

MetS is a group of metabolic abnormalities whose existence is considered a predominant risk factor for the development of cardiovascular diseases (Day 2007). Currently, it is of great importance due to its high prevalence rate, since 35% of the adult population and 50% of the elderly population have MetS. Although the pathogenesis of MetS is not fully understood, it is generally diagnosed after the existence of 3 or more risk factors in an individual, altered glucose metabolism, obesity, dyslipidaemia, and/or hypertension (Neuhauser 2005). Yet, this definition is not adequate and is not supported at the molecular level by validated biomarkers. Many studies have metabolically investigated MetS or its participating risk factors (Reddy et al. 2018; Blouin et al. 2005), including studies based on NMR metabolomics (Monnerie et al. 2020). For instance, Würtz et al. (2012) analysed more than 7,000 serum samples from apparently healthy young adults using high-throughput NMR to investigate diabetes prevalence. The metabolic signatures of insulin resistance were modulated by obesity (Würtz et al. 2012) and found different for men and women.

In a recent study (Bruzzone et al. 2021), we have investigated MetS using NMR-based metabolomics of urine samples. The goal of the project was to investigate the natural history of the syndrome, evaluating the metabolic fingerprint at each step of the progression between apparently healthy individuals (at least devoid of any risk factor related to MetS) up to subjects that suffer all the MetS-associated risk factors. Indeed, the study was designed not only to investigate the contribution of each of the previously risk factors that can be involved in the development of MetS independently, but also to evaluate all their possible combinations between the presence or absence of them. Four different risk factors and their possible intermediate conditions were considered: obesity, dyslipidaemia, hypertension, and diabetes and all the possible intermediate conditions (Table 1). The reference AKRIBEA

Table 1 MetS conditions, definitions, and number of individuals allocated to each condition in the metabolic syndrome study (Bruzzone et al. 2021)

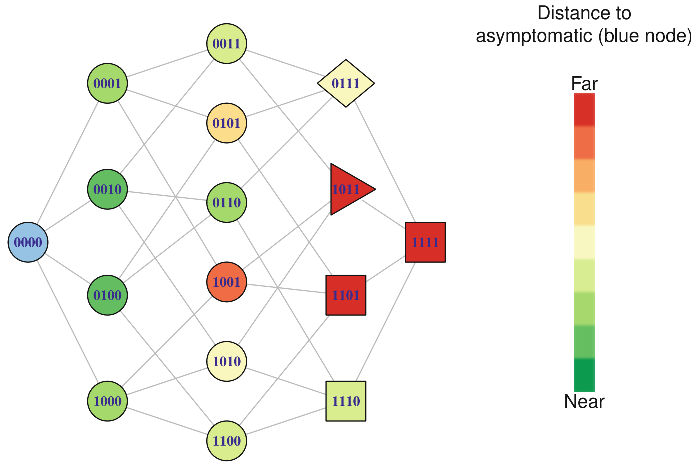
MetS condition		Total $N = 10{,}792$	Female $N = 4{,}351$	Male $N = 6{,}441$
0000	Apparently healthy	6,925 (64.17%)	2,935 (67.46%)	3,990 (61.95%)
0001	Hypertension	692 (6.41%)	276 (6.34%)	416 (6.46%)
0010	Dyslipidaemia	733 (6.79%)	120 (2.76%)	613 (9.52%)
0011	Obesity	170 (1.58%)	53 (1.22%)	117 (1.82%)
0100	Diabetes	504 (4.67%)	232 (5.33%)	272 (4.22%)
0101	Dyslipidaemia + hypertension	310 (2.87%)	169 (3.88%)	141 (2.19%)
0110	Obesity + hypertension	170 (1.58%)	37 (0.85%)	133 (2.06%)
0111	Obesity + dyslipidaemia	148 (1.37%)	62 (1.42%)	86 (1.34%)
1000	Diabetes + hypertension	282 (2.61%)	89 (2.05%)	193 (3.00%)
1001	Diabetes + dyslipidaemia	188 (1.74%)	83 (1.91%)	105 (1.63%)
1010	Diabetes + obesity	84 (0.78%)	18 (0.41%)	66 (1.02%)
1011	Obesity + dyslipidaemia + hypertension	90 (0.83%)	32 (0.74%)	58 (0.90%)
1100	Diabetes + dyslipidaemia + hypertension	92 (0.85%)	44 (1.01%)	48 (0.75%)
1101	Diabetes + obesity + hypertension	202 (1.87%)	111 (2.55%)	91 (1.41%)
1110	Diabetes + obesity + dyslipidaemia	62 (0.57%)	17 (0.39%)	45 (0.70%)
1111	Diabetes + obesity + dyslipidaemia + hypertension	140 (1.30%)	73 (1.68%)	67 (1.04%)

Legend: Diabetes as fasting plasma glucose >100 mg/dL, previously diagnosed type 2, impaired fasting glucose, impaired glucose tolerance, or insulin resistance taking medication for hyperglycaemia; obesity as BMI > 30 kg/m^2; dyslipidaemia as triglycerides >150 mg/dL HDL Cholesterol <34.75 mg/dL in men or <38.61 in women, previously diagnosed hypercholesterolaemia, hyperlipidaemia or hypertriglyceridaemia taking medication for dyslipidaemia; hypertension as blood pressure ≥140/90 mmHg, previously diagnosed hypertension taking medication for hypertension

cohort was complemented with three additional cohorts of individuals specially enriched in risk factors associated with MetS to ensure high statistical power for each intermediate condition. Thus, the less populated condition still remained with up to 62 associated samples (Table 1).

According to PCA analysis, NMR spectroscopy of urine can efficiently discriminate all the intermediate conditions, with a stronger power for diabetes and hypertension as compared to obesity and dyslipidaemia. Notably, the responsible metabolites for this statistical separation gradually adapt their concentrations when

Metabolic syndrome profile graph



Ordered profiles:
0000, 0010, 0100, 0110, 0001, 1000, 1100, 0011, 1110, 1010, 0111, 0101, 1001, 1101, 1011, 1111
Color by direct distance to asymptomatic (0000)

Fig. 5 Spearman correlation distances to the healthy condition for all 4 conditions. Colours represent the distance to the apparently healthy (0000) condition and the lines connect adjacent conditions. MetS definition according to WHO, European Group for the Study of Insulin (EGIR), and American Association of Clinical Endocrinology (AACE) is represented by squares and triangles; definition from National Cholesterol Education Program-Third Adult Treatment Panel (NCEP:ATPIII) is represented by squares, triangles, and rhombus; definition by International Diabetes Federation (IDF) is represented by squares and rhombus

moving from an apparently healthy condition (0000 according to our notation) towards MetS patients (i.e. 1111, 1011, etc.) (Bruzzone et al. 2021). This result strongly suggests that MetS is a continuous metabolic dysregulation, and a putative definition deducted from the metabolic analysis would clearly identify the conditions: 1001, 1011, 1101, and 1111 as MetS (Fig. 5). This definition is similar, but not identical, to other existing definitions such as the WHO or the harmonized definition, which are based on compatible symptomatology only.

The discriminant urine metabolic fingerprint contains metabolites associated to all the risk factors under consideration, but it is largely dominated by the high glucose levels, always associated to conditions proximal to MetS. Other significant metabolites were p-cresol sulphate, and 4-hydroxyphenylpyruvic acid, also involved in insulin resistance (Koppe et al. 2013).

Finally, the metabolic fingerprint was used to generate a model (AUROC of 0.87), able to discriminate the pool of samples (more than 13,000) according to MetS probability.

5.2 Oriented Metabolomics Study: COVID-19

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) (WHO, 2022). The spreading of the viral outbreaks around the world has been considered as a health problem associated to the current globalization, with frequent social contacts between people from different countries and continents, causing the virus to spread worldwide. As of August, 2021, more than 200 million cases have been reported in worldwide (WHO, 2022).

There is a plethora of symptoms associated with COVID-19 and patients can go through different stages of the disease. The initial infectious stage often involves an incubation period associated with mild and nonspecific symptoms, such as malaise, fever, and a dry cough (Ayres 2020). Symptomatic patients may undergo an acute phase, characterized by mild symptomatology: a pulmonary phase that may be accompanied by shortness of breath or, eventually, a severe hyperinflammatory phase that can lead to acute respiratory distress syndrome and/or heart failure (Ayres 2020). In many cases the prognosis is favourable, but approximately 20% of patients with COVID-19 require admission to the intensive care unit due to a life-threatening acute respiratory syndrome (ARDS) and an extrapulmonary systemic hyperinflammation syndrome, which could be accompanied by multiple organ failure (Niazkar et al. 2020). A growing number of studies in this area suggest that SARS-CoV-2 infection is not just an infection that affects the lungs but a systemic syndrome with numerous metabolic manifestations (Bruzzone et al. 2020a; Ayres 2020; Kimhofer et al. 2020; San Juan et al. 2020; Thomas et al. 2020).

In principle, a metabolomic approach, based on the study of serum from patients infected with SARS-CoV-2, should effectively address the problem of the great disparity in the evolution of COVID-19 between individuals, from the most general and complementary angle possible, similar to previous studies with other viruses. For instance, in dengue virus, a specific metabolomic fingerprint included altered metabolites associated with lipid metabolism, the regulation of inflammatory processes by the signal of fatty acids and phospholipids, as well as endothelial cell homeostasis and the vascular barrier (Voge et al. 2016). In this line, several studies have used NMR metabolomics to investigate COVID-19 patients (Bruzzone et al. 2020a; Holmes et al. 2021; Lodge et al. 2021b; Meoni et al. 2021) (Turano, "NMR-based metabolomics to evaluate individual response to treatment"; Rogers, "The metabolomics of critical illness"). In our laboratory, we hypothesized that the analysis of metabolic dysregulations measurable by NMR spectroscopy can, when combined with other specific inflammation markers, predict the different stages of SARS-CoV-2 infection associated to several biomarkers of the disease. To investigate that hypothesis, in collaboration with the Cruces University Hospital and Basurto University Hospital (Basque Country), we analysed 260 sera from patients in a mostly severe stage of the development of the disease and compared them to 280 pre-COVID sera (healthy population samples from the AKRIBEA cohort, that were collected in the two-year period prior to the COVID-19 outbreak), used as healthy reference cohort.

This study revealed that COVID-19 patients dysregulate ApoA-I and ApoA-II lipoproteins, which points to an increased risk of atherosclerosis. This is in addition to a singular profile of low-molecular-weight metabolites, which include high levels of ketone bodies (acetoacetic acid, 3-hydroxybutyric acid, and acetone) and 2-hydroxybutyric acid, related to liver damage, dyslipidaemia, and oxidative stress (Bruzzone et al. 2020a).

A follow-up study, in collaboration with the group of Jeremy Nicholson (ANPC, Murdoch University, Australia), reinforces the idea that COVID-19 disease dysregulates amino acids, biogenic amines, and tryptophan pathway, a bona fide biomolecular marker of the disease (Kimhofer et al. 2020). Also, in the acute phase of the disease the systemic inflammation markers appear to be elevated. This is in line with other COVID-19 studies (Kimhofer et al. 2020; Thomas et al. 2020), which show decreased counts of T-helper cells (regulators of T lymphocytes) and elevated inflammatory cytokines and biomarkers such as IL-2, IL-6, IL-7, macrophage inflammatory protein 1-a, tumour necrosis factor a, C-reactive protein, ferritin, and D-dimer.

The natural history of COVID-19 disease reveals that some people infected by SARS-CoV-2 do not return to a normal/healthy state, establishing a new deteriorated health-baseline (Ayres 2020). Long-term COVID-19 or post-acute COVID-19 syndrome patients may retain some symptoms and/or have developed unprecedented adverse effects (Holmes et al. 2021; Nalbandian et al. 2021; Yong 2021). Known manifestations include a wide variety of physical symptoms, such as severe fatigue, malaise, headaches, and other more worrisome phenomena such as increased risk of damage to the heart, lungs, and brain (López-León et al. 2021; Davis et al. 2021). It is estimated that around 1 out of 10 COVID-19 patients will undergo symptoms up to 12 weeks after having suffered the disease, according to a new report from the European Observatory of Health Systems and Policies promoted by the World Health Organization (WHO).

An important open question is whether post-acute sequela of COVID-19 are accompanied by incomplete metabolic phenoreversion. Metabolic alterations on a cohort of 27 recovered, non-hospitalized patients have been recently analysed using a combination of LC-MS spectrometry and NMR spectroscopy to show that a subset of these patients still display a biochemical pathophysiology (Holmes et al. 2021). Consistently, our lab studied a cohort of 97 post-COVID-19 recovered patients, 6 months after the infection on average, and compared them to a matched reference set ($n = 87$), created ad hoc from the AKRIBEA cohort. Our results show a bimodal distribution, with half the patient's population still experiencing a significant metabolic and lipidomic dysregulation after 6 months of the infection. Finally, a recent study analysed blood samples of PCR−/controls, PCR+/not-hospitalized, PCR+/hospitalized, and PCR+/intubated patients by mass spectrometry to show that kynurenine/tryptophan ratio, C26:0-lysoPC, and pyruvic acid can discriminate non-COVID controls from PCR+/not-hospitalized patients (López-Hernández et al. 2021). Moreover, C10:2-lysoPC, butyric acid, and pyruvic acid could distinguish between PCR+/ not-hospitalized and PCR+/hospitalized and PCR+/intubated. Altogether, these studies highlight the potential power of metabolomics for the

characterization of the pathophysiology of complex diseases such as COVID-19 and pave the way for their integration in the clinical practice, after exploring larger validation cohorts of patients.

6 Concluding Remarks

The optimization of the analytical techniques combined with the availability of large-scale logistics under strict SOPs enables the implementation of precision medicine into the clinical practice. In this not-so-futuristic scenario, NMR metabolomics may play a key role thanks to its analytical power, very large reproducibility, and the realistic possibility of deploying magnets in hospitals and surgery rooms. To that end, large spectral databases combined with AI techniques will become instrumental to validate the diagnostic value that metabolomics can potentially offer. Moreover, precision medicine could be applied in the clinic with this metabolism analysis, which is constantly changing, allowing the clinician to prescribe the right treatment and the specific dose of it at the right time of the disease state.

References

Authors (2021) Ultra-precision medicine. Nat Biotechnol 39:645. https://doi.org/10.1038/s41587-021-00967-8

Ayres JS (2020) A metabolic handbook for the COVID-19 pandemic. Nat Metab 2:572–585

Beger RD et al (2016) Metabolomics enables precision medicine: “A White Paper, Community Perspective”. Metabolomics 12

Bernardo-Seisdedos G et al (2021) Metabolic landscape of the mouse liver by quantitative 31P nuclear magnetic resonance analysis of the phosphorome. Hepatology 74:148–163

Blouin K et al (2005) Contribution of age and declining androgen levels to features of the metabolic syndrome in men. Metabolism 54:1034–1040

Bruzzone C et al (2020a) SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. iScience 23

Bruzzone C et al (2020b) 1H NMR-based urine metabolomics reveals signs of enhanced carbon and nitrogen recycling in prostate cancer. J Proteome Res 19:2419–2428

Bruzzone C et al (2021) A molecular signature for the metabolic syndrome by urine metabolomics. Cardiovasc Diabetol 20:155

Carrat F et al (2021) Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based “CONSTANCES” cohort. Eur J Epidemiol 36:219–222

Christensen CM, Grossman JH, Hwang J (2009) The innovator’s prescription: a disruptive solution for health care. McGraw-Hill

Clish CB (2015) Metabolomics: an emerging but powerful tool for precision medicine. Mol Case Stud 1:a000588

Davis HE et al (2021) Characterizing Long COVID in an international cohort: 7 months of symptoms and their impact. SSRN Electron J. https://doi.org/10.2139/ssrn.3820561

Day C (2007) Metabolic syndrome, or what you will: definitions and epidemiology. Diab Vasc Dis Res 4:32–38

De Souza YG, Greenspan JS (2013) Biobanking past, present and future: responsibilities and benefits. AIDS 27:303–312

Dona AC et al (2014) Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem 86:9887–9894

Embade N et al (2019) NMR-based newborn urine screening for optimized detection of inherited errors of metabolism. Sci Rep 9:1–9

Emwas AH et al (2019) NMR spectroscopy for metabolomics research. Metabolites 9

Ghini V, Quaglio D, Luchinat C, Turano P (2019) NMR for sample quality assessment in metabolomics. N Biotechnol 52:25–34

González-Domínguez R, González-Domínguez Á, Sayago A, Fernández-Recamales Á (2020) Recommendations and best practices for standardizing the pre-analytical processing of blood and urine samples in metabolomics. Metabolites 10:1–18

Hartl D et al (2021) Translational precision medicine: an industry perspective. J Transl Med 19:1–14

Holmes E, Wist J, Masuda R, Lodge S, Nitschke P, Kimhofer T, Loo RL, Begum S, Boughton B, Yang R, Morillon AC, Chin ST, Hall D, Ryan M, Bong SH, Gay M, Hartmut Lawler NG, Gray N, Whiley L, Nicholson J (2021) Incomplete systemic recovery and metabolic phenoreversion in post-acute-phase nonhospitalized COVID-19 patients: implications for assessment of post-acute COVID-19 syndrome. J Proteome Res

Jiménez B et al (2018) Quantitative lipoprotein subclass and low molecular weight metabolite analysis in human serum and plasma by (1)H NMR spectroscopy in a multilaboratory trial. Anal Chem 90:11962–11971

John A, Qin B, Kalari KR, Wang L, Yu J (2020) Patient-specific multi-omics models and the application in personalized combination therapy. Future Oncol 16:1737–1750

Kimhofer T et al (2020) Integrative modeling of quantitative plasma lipoprotein, metabolic, and amino acid data reveals a multiorgan pathological signature of SARS-CoV-2. J Proteome Res

Kohler I, Hankemeier T, van der Graaf PH, Knibbe CAJ, van Hasselt JGC (2017) Integrating clinical metabolomics-based biomarker discovery and clinical pharmacology to enable precision medicine. Eur J Pharm Sci 109:S15–S21

Koppe L et al (2013) p-Cresyl sulfate promotes insulin resistance associated with CKD. J Am Soc Nephrol 24:88–99

Koshiba S et al (2018) Omics research project on prospective cohort studies from the Tohoku Medical Megabank Project. Genes Cells 23:406–417

Lodge S et al (2021a) Low volume in vitro diagnostic proton NMR spectroscopy of human blood plasma for lipoprotein and metabolite analysis: application to SARS-CoV-2 biomarkers. J Proteome Res 20:1415–1423

Lodge S et al (2021b) NMR spectroscopic windows on the systemic effects of SARS-CoV-2 infection on plasma lipoproteins and metabolites in relation to circulating cytokines. J Proteome Res 20:1382–1396

Loeffler M et al (2015) The LIFE-adult-study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. BMC Public Health 15:1–14

Loo RL et al (2020) Quantitative in-vitro diagnostic NMR spectroscopy for lipoprotein and metabolite measurements in plasma and serum: recommendations for analytical artifact minimization with special reference to COVID-19/SARS-CoV-2 samples. J Proteome Res 19:4428–4441

López-Hernández Y et al (2021) Targeted metabolomics identifies high performing diagnostic and prognostic biomarkers for COVID-19. Sci Rep 11:14732

López-León S et al (2021) More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. SSRN Electron J 1–22. https://doi.org/10.2139/ssrn.3769978

Martin SS et al (2015) HDL cholesterol subclasses, myocardial infarction, and mortality in secondary prevention: the Lipoprotein Investigators Collaborative. Eur Heart J 36:22–30

Meoni G et al (2021) Metabolomic/lipidomic profiling of COVID-19 and individual response to tocilizumab. PLoS Pathog 17

Mignot S et al (2021) Correlates of premature pap test screening, under 25 years old: analysis of data from the CONSTANCES cohort study. BMC Public Health 21:1–9

Monnerie S et al (2020) Metabolomic and lipidomic signatures of metabolic syndrome and its physiological components in adults: a systematic review. Sci Rep 10:1–13
Monsonis Centelles S et al (2017) Toward reliable lipoprotein particle predictions from NMR spectra of human blood: an interlaboratory ring test. Anal Chem 89:8004–8012
Nabi O et al (2021) Comorbidities are associated with fibrosis in NAFLD subjects: a nationwide study (NASH-CO study). Dig Dis Sci. https://doi.org/10.1007/s10620-021-07032-z
Nalbandian A et al (2021) Post-acute COVID-19 syndrome. Nat Med 27:601–615
Neuhauser HK (2005) The metabolic syndrome. 366
Niazkar HR, Zibaee B, Nasimi A, Bahri N (2020) The neurological manifestations of COVID-19. Prat Neurol FMC 11:145–146
Ortega-Azorín C et al (2019) Candidate gene and genome-wide association studies for circulating leptin levels reveal population and sex-specific associations in high cardiovascular risk Mediterranean subjects. Nutrients 11:2751
Otvos JD, Jeyarajah EJ, Bennett DW, Krauss RM (1992) Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. Clin Chem 38:1632–1638
Puchades-Carrasco L, Pineda-Lucena A (2017) Metabolomics applications in precision medicine: an oncological perspective. Curr Top Med Chem 17:2740–2751
Reddy P, Leong J, Jialal I (2018) Amino acid levels in nascent metabolic syndrome: a contributor to the pro-inflammatory burden. J Diabetes Complications 32:465–469
Ruiz F et al (2016) High quality standards for a large-scale prospective population-based observational cohort: constances. BMC Public Health 16:1–10
San Juan I et al (2020) Abnormal concentration of porphyrins in serum from COVID-19 patients. Br J Haematol 190:e265–e267
Śliwczyński A, Orlewska E (2016) Precision medicine for managing chronic diseases. Pol Arch Med Wewn 126:681–687
Thomas T et al (2020) COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. JCI Insight 5
Tigchelaar EF et al (2015) Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: Study design and baseline characteristics. BMJ Open 5:1–10
Tranvåg EJ, Strand R, Ottersen T, Norheim OF (2021) Precision medicine and the principle of equal treatment: a conjoint analysis. BMC Med Ethics 22:1–9
Vignoli A et al (2019) High-throughput metabolomics by 1D NMR. Angew Chem Int Ed 58:968–994
Voge NV et al (2016) Metabolomics-based discovery of small molecule biomarkers in serum associated with dengue virus infections and disease outcomes. PLoS Negl Trop Dis 10:1–27
WHO (2022). https://www.who.int/health-topics/coronavirus
WHO (2022). https://covid19.who.int/
Wilson D, Burlingame A, Cronholm T, Sjövall J (1974) Deuterium and carbon-13 tracer studies of ethanol metabolism in the rat by 2H, 1H-decoupled 13C nuclear magnetic resonance. Biochem Biophys Res Commun 56:828–835
Wishart DS (2019) NMR metabolomics: a look ahead. J Magn Reson 306:155–161
Wishart DS et al (2018) HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 46:D608–D617
Würtz P et al (2012) Metabolic signatures of insulin resistance in 7,098 young adults. Diabetes 61: 1372–1380
Yin P, Lehmann R, Xu G (2015) Effects of pre-analytical processes on blood samples used in metabolomics studies. Anal Bioanal Chem 407:4879–4892
Yong SJ (2021) Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. Infect Dis 0:1–18

Chemotherapy-Induced Peripheral Neuropathy

Ciao-Sin Chen and Daniel L. Hertz

Contents

1 Introduction ... 300
2 Review of CIPN Metabolomics ... 302
2.1 Metabolomics Analyses Design ... 302
2.2 CIPN Study Design ... 303
2.3 Clinical CIPN Metabolomics Studies ... 303
2.4 Animal and Cellular CIPN Metabolomics Studies ... 306
2.5 Summary ... 308
3 Investigation of Peripheral Neuropathy Metabolomic Biomarker Pathways ... 308
3.1 Histidine ... 313
3.1.1 1-Methylhistidine ... 315
3.1.2 Carnosine ... 315
3.1.3 Ergothioneine ... 316
3.1.4 Histamine ... 316
3.1.5 Summary ... 316
3.2 Phenylalanine ... 317
3.2.1 Phenylalanine and Tyrosine ... 317
3.2.2 Phenethylamine ... 319
3.2.3 Summary ... 319
3.3 N-Acetylornithine ... 319
3.3.1 Ornithine and Arginine ... 320
3.3.2 Ornithine and Polyamines: Putrescine, Spermidine, and Spermine ... 321
3.3.3 Summary ... 321
3.4 Threonine ... 321
3.5 Glycogen ... 322
3.6 Adenosine Phosphates ... 323
4 Conclusion ... 324
References ... 324

C.-S. Chen · D. L. Hertz (✉)
Department of Clinical Pharmacy, University of Michigan College of Pharmacy, Ann Arbor, MI, USA
e-mail: cscchen@umich.edu; dlhertz@med.umich.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_609

Abstract

Chemotherapy-induced peripheral neuropathy (CIPN) is a debilitating side effect of many common anti-cancer agents that can lead to dose reduction or treatment discontinuation, which decrease chemotherapy efficacy. Long-term CIPN can interfere with activities of daily living and diminish the quality of life**.** The mechanism of CIPN is not yet fully understood, and biomarkers are needed to identify patients at high risk and potential treatment targets. Metabolomics can capture the complex behavioral and pathophysiological processes involved in CIPN. This chapter is to review the CIPN metabolomics studies to find metabolic pathways potentially involved in CIPN. These potential CIPN metabolites are then investigated to determine whether there is evidence from studies of other neuropathy etiologies such as diabetic neuropathy and Leber hereditary optic neuropathy to support the importance of these pathways in peripheral neuropathy. Six potential biomarkers and their putative mechanisms in peripheral neuropathy were reviewed. Among these biomarkers, histidine and phenylalanine have clear roles in neurotransmission or neuroinflammation in peripheral neuropathy. Further research is needed to discover and validate CIPN metabolomics biomarkers in large clinical studies.

Keywords

Chemotherapy · Histidine · Metabolomics · Peripheral neuropathy · Phenylalanine

1 Introduction

Chemotherapy-induced peripheral neuropathy (CIPN) is a debilitating side effect of many common anti-cancer agents, including taxane, vinca alkaloids, platinum, bortezomib, and thalidomide-related agents (Miltenburg and Boogerd 2014; Seretny et al. 2014). The prevalence and clinical characteristics of CIPN vary by regimen (Miltenburg and Boogerd 2014; Seretny et al. 2014) with prevalence estimates as high as 70% (Seretny et al. 2014). CIPN is a chronic distal symmetric polyneuropathy that predominantly presents as sensory symptoms, including numbness (stocking-glove), tingling, and shooting or burning pain (Miltenburg and Boogerd 2014; Quasthoff and Hartung 2002). Motor symptoms such as loss of balance or dexterity caused by muscle weakness can be found in severe cases, with prevalence up to 40% depending on the regimen (Argyriou et al. 2008; Haim et al. 1994; Lee and Swain 2006; Miltenburg and Boogerd 2014; Plasmati et al. 2007). Autonomic symptoms are usually rare, but constipation is common (up to 90%) in patients receiving certain neurotoxic agents including vincristine, bortezomib, and thalidomide-related agents (Miltenburg and Boogerd 2014).

CIPN symptoms can limit instrumental activities of daily living such as preparing meals, shopping for groceries or clothes, or using the phone. CIPN symptoms persist in about 40% of CIPN patients for at least 2 years and are irreversible in some patients (Bao et al. 2016; Hershman et al. 2018; Mustafa Ali et al. 2017). Irreversible CIPN diminishes long-term quality of life (Bandos et al. 2018; Mols et al. 2014), increases risk for falls, and causes disability in severe cases (Kolb et al. 2016; Winters-Stone et al. 2017). These life-long consequences are particularly concerning in patients with favorable long-term prognosis, such as those with early stage breast cancer (McGale et al. 2014).

The mechanism of CIPN is not yet fully understood and is likely to be multifactorial. The neurotoxicity caused by chemotherapy may involve DNA damage, microtubule disruption, mitochondrial damage, pain detection receptor remodeling, oxidative stress, and neuroinflammation in the neurons of the peripheral nervous system (Chan et al. 2019; Kerckhove et al. 2017; Zajączkowska et al. 2019). The lack of mechanistic understanding is one reason for the lack of effective strategies for CIPN prevention or treatment; the only guideline-recommended strategy for treatment of painful CIPN is duloxetine, which has limited efficacy (Loprinzi et al. 2020; Smith et al. 2013). Otherwise, guidelines recommend delaying, decreasing, or discontinuing neurotoxic chemotherapy treatment in patients experiencing CIPN, which reduces treatment efficacy (Hertz et al. 2021).

Biomarkers that could be used prior to or early in treatment to predict which patients are most at risk for severe, and perhaps irreversible, CIPN would be extremely clinically useful. The most established CIPN risk factors are cumulative treatment, prior neuropathy or neurotoxic treatment, age, race, diabetes, and obesity (Gu et al. 2021; Hershman et al. 2016; Kandula et al. 2016; Schneider et al. 2017; Seretny et al. 2014). There may be some confounding between these factors, as age, race, and obesity are all risk factors for diabetes, which is also a cause of peripheral neuropathy (Bao et al. 2016; Cox-Martin et al. 2017; Gu et al. 2021; Hershman et al. 2016; Petrovchich et al. 2019; Smith and Singleton 2013; Timmins et al. 2021; Wiggin et al. 2009). Modifiable nutritional factors such as diet and nutrient supplementation have also been reported to influence CIPN risk (Mongiovi et al. 2018; Stankovic et al. 2020; Zirpoli et al. 2017). More severe CIPN has been reported in patients with lower grain consumption, not taking vitamins or taking antioxidants (Mongiovi et al. 2018; Stankovic et al. 2020; Zirpoli et al. 2017), deficiency of vitamin D or saturated fatty acids (Grim et al. 2017; Jennaro et al. 2020), or a sedentary lifestyle (Bulls et al. 2020; Duregon et al. 2018; Greenlee et al. 2017; Kanzawa-Lee et al. 2020; Mols et al. 2015). Again, these risk factors may be due to shared mechanistic pathways with neuropathy caused by poor nutrition, obesity, and/or diabetes. Finally, substantial research has been undertaken to identify inherited genetic markers, including those that affect neurotoxic drug elimination or pharmacology, diabetes, and neuronal cell function or survival (Chan et al. 2019; Cliff et al. 2017; Ng et al. 2014; Sucheston-Campbell et al. 2018; Terrazzino et al. 2015), however, no genetic markers have been sufficiently validated for use in clinical practice (Chan et al. 2019).

Biomarkers that capture the complex behavioral and pathophysiological processes involved in CIPN are needed to identify patients with high CIPN risk or to gain further mechanistic understanding of CIPN to identify potential treatment targets. Metabolomics is a biomarker discovery strategy that analyzes the metabolome, which is the comprehensive endogenous metabolite profile of biospecimens. Metabolites are at the end of the DNA-RNA-protein central dogma axis. By providing a direct snapshot of the current metabolic status, metabolomics offers an opportunity to explain the inter-individual variability that cannot be detected by DNA sequence, such as environmental and nutritional factors (Clayton et al. 2009; O'Gorman and Brennan 2017; Shin et al. 2014). These factors are sometimes too complicated to be extracted from medical records and are often not fully examined or stratified (Everett 2019; Kantae et al. 2017; Rattray et al. 2018; Wishart 2016). Metabolomics can capture the end effect of these factors in quantifiable metabolic signatures, which have been found to be useful in predicting toxicity from cancer drugs (Backshall et al. 2011). In addition, compared to other omics, such as transcriptomics and proteomics, metabolomics is closer to the clinical phenotype and easier to quantify (Karczewski and Snyder 2018). In this chapter, we will review CIPN metabolomics studies to find metabolic pathways potentially involved in CIPN and then investigate whether there is evidence from studies of other neuropathy etiologies to support these pathways. CIPN biomarker studies that only investigated lipidomics are outside of the scope and will not be included (Kramer et al. 2015; Maekawa et al. 2019).

2 Review of CIPN Metabolomics

2.1 Metabolomics Analyses Design

The two most common techniques in metabolomics analyses are nuclear magnetic resonance (NMR) and mass spectrometry (MS). Each has its own advantages and disadvantages in profile coverage, sensitivity, stability, sample volume, sample preparation, quantification, high-throughput automation, time, and cost (Everett 2017; Wishart 2016). MS-based methods are 10–100 times more sensitive than NMR. Whereas a typical LC-MS study can detect more than 1,000 metabolites with concentration of 10–100 nM in human plasma, a typical NMR study can only detect 50–200 metabolites with concentration at the microM scale. On the other hand, only a few hundred of the metabolites detected by MS-based methods can be identified (Emwas et al. 2019). NMR sensitivity can be improved with digital spectrometers, cryoprobes, and low volume probes (Everett 2017). There is a trend toward increasing use of NMR due to its greater reproducibility and the ease of full automation (Everett 2017; Wishart 2016). Although NMR requires larger sample volume, sample preparation is minimal (Kirwan et al. 2018; Wishart 2016), and NMR can provide absolute quantification of metabolite concentrations in only a few minutes without reference standards (Wishart 2016). The start-up cost of NMR instrumentation is expensive, but the cost per sample is low (Everett 2017; Wishart

2016). There are two approaches to metabolite identification, targeted and untargeted metabolomics. Targeted approaches focus on a set of defined metabolites, whereas global untargeted approaches capture all metabolites that are present in a sample but may not provide the exact identification of some of the metabolites due to the limitations of databases or libraries. Untargeted metabolomics is usually qualitative, or can provide relative quantitation, and is used for hypothesis-agnostic discovery, while targeted metabolomics can provide absolute quantitation and is often used to confirm results from untargeted studies or confirm hypotheses (Roberts et al. 2012). Both NMR and MS-based methods can be used in untargeted and targeted analyses, but MS-based methods are generally superior for untargeted metabolomics discovery due to high sensitivity and for targeted metabolomics validation due to its selectivity, while NMR is used primarily in untargeted analyses (Emwas et al. 2019).

2.2 CIPN Study Design

Metabolomics studies of CIPN enroll patients with cancer who are going to receive neurotoxic chemotherapy. CIPN data are collected from clinician or patient assessment, and patients can be defined as whether or not they experienced CIPN, the severity of their CIPN, or the cumulative dose at the time that CIPN occurred (Hertz 2019). Biosamples are collected pretreatment (baseline) and on-treatment to acquire metabolite information. The pretreatment metabolite levels can be used to identify metabolic signatures that predict peripheral neuropathy, whereas the on-treatment changes from baseline can be used to find metabolic signatures that are associated with the neurotoxic effect of the chemotherapy, which can still be predictive if they are sufficiently early in treatment that CIPN has not yet become clinically observable. Metabolic changes at the time of clinically evident CIPN are not useful for prediction but may be useful to understand mechanisms underlying CIPN.

2.3 Clinical CIPN Metabolomics Studies

Two CIPN metabolomics studies have been conducted in patients receiving chemotherapy (Table 1). One untargeted whole blood NMR study by Sun et al. (2018) enrolled 48 patients with stage I-III or oligometastatic breast cancer planning to receive weekly paclitaxel and obtained metabolomics profiles prior to treatment (baseline), at the end of the first paclitaxel infusion, and 24 h after the first infusion. Paclitaxel-induced peripheral neuropathy was assessed by a patient-reported questionnaire at baseline and before each infusion, and CIPN severity was defined by the maximum change from baseline of the sensory symptom subscale. Without strict correction for multiple comparisons, lower pretreatment levels of three amino acids, histidine, phenylalanine, and threonine, were moderately correlated with more severe sensory neuropathy symptoms. Statistically uncorrected secondary analyses identified many metabolites that changed from pretreatment to post-infusion or 24-h later, however, none of these were associated with CIPN

Table 1 Human metabolomics studies of CIPN

Subjects	Neurotoxic chemotherapy	CIPN data	Metabolomics techniques	Sample type	Statistical methods	Findings	Reference
48 patients with breast cancer	Paclitaxel 80 mg/m^2 1-h infusion weekly for 12 weeks	Max change from baseline of CIPN8	NMR, untargeted 34 metabolites	Whole blood, non-fasting	Pearson correlation coefficient with false discovery rate $< 25\%$	Lower pretreatment levels of histidine ($r = -0.38$, $p = 0.01$), phenylalanine ($r = -0.34$, $p = 0.02$), and threonine ($r = -0.36$, $p = 0.01$) were moderately correlated with the severity of sensory peripheral neuropathy during treatment	(Sun et al. 2018)
32 pediatric patients with acute lymphoblastic leukemia	Vincristine 1.5 mg/m^2 (max 2 mg) during the induction (day 8 and 29) and consolidation phases (day 180)	Max of TNS-PV >8	LC-MS/MS, untargeted, total number of metabolites not reported	Plasma, non-fasting	Linear support vector machine with recursive feature elimination and five-fold cross validation	N-acetylornithine, glycogen, adenosine monophosphate, and adenosine diphosphate on day 29 were identified to be predictive (AUC = 0.75). Arginine, 1, 7-dimethylguanosine, gamma-glutamyl-isoleucine, glutathione disulfide, phenyl-alanyl-proline, pipecolate, 5-hydroxy-tryptophol, sphingomyelin SM (d18: 0/16: 1(9Z)), and glycerophosphocholine PC (16: 1(9Z)/0:0) on	(Verma et al. 2020)

						day 180 were identified indicative of CIPN development ($AUC = 0.92$). No metabolite concentrations were reported	

CIPN8, the subscale of the first eight sensory items of European Organization for Research and Treatment of Cancer Quality of Life Questionnaire for Chemotherapy-induced Peripheral Neuropathy 20-item scale
TNS-PV, total neuropathy score-pediatric vincristine

severity. A subsequent analysis identified metabolites that were moderately correlated with paclitaxel pharmacokinetics, but these statistically uncorrected discovery-phase associations did not mechanistically explain the CIPN metabolomics findings in the initial study (Chen et al. 2021).

The other clinical CIPN metabolomics study was an untargeted plasma LC-MS/MS study by Verma et al. (2020) which enrolled 32 pediatric patients receiving vincristine for precursor B-cell acute lymphoblastic leukemia. Patients with preexisting peripheral neuropathy or taking erythropoietin, itraconazole, or vitamin supplementation were excluded due to the possible interference with neurological assessment. Vincristine-induced peripheral neuropathy was assessed via physician-evaluation regularly throughout treatment, and patients were classified as peripheral neuropathy if they experienced a high symptom score (total neuropathy score-pediatric vincristine >8) at any time during treatment. The investigators analyzed the metabolome at days 8, 29, and 180 and used a linear support vector machine to build metabolomic models that predict (days 8 or 29) or indicate (day 180) the development of CIPN. The final models included 2, 14, and 21 metabolomics features, respectively, and all three models had excellent performance with recursive feature elimination and five-fold cross validation. However, metabolite concentrations were not reported, and some metabolites were not identified, for which only mass and retention time were reported in the supplementary information. The identified metabolites in the day 29 and 180 models can be found in Table 1. Pathway analyses did not identify any relevant metabolic pathways after multiple comparison correction.

2.4 Animal and Cellular CIPN Metabolomics Studies

Although clinical CIPN metabolomics studies are very scarce, animal and cellular studies may assist with identifying metabolic pathways of interest (Table 2). Two untargeted NMR studies by Ferrier et al. (2013, 2015) analyzed the metabolomics of central nervous system tissue in rat models of oxaliplatin-induced neurotoxicity. A single injection of oxaliplatin increased glutamine and decreased creatinine and adenosine phosphate in spinal dorsal horn tissues in rats that developed acute pain hypersensitivity compared to vehicle-treated rats, and these metabolic changes and pain hypersensitivity were abrogated by a polyamine-deficient diet (Ferrier et al. 2013). In a similar study of multiple oxaliplatin injections, higher choline and glycerophosphocholine and lower aspartate, glutamine, creatinine, and GABA were found in various brain tissues in rats that developed painful neuropathy compared to vehicle-treated rats (Ferrier et al. 2015). An untargeted LC-MS/MS study by Wu et al. (2018) analyzed the plasma metabolomics of a rat model of paclitaxel-induced peripheral neuropathy and found 19 lipids, including fatty acids, ketones, and glycerophosphocholine, which differentiated paclitaxel-treated from vehicle-treated rats. An untargeted study by Qin et al. (2012) used a different metabolomics technique, capillary electrophoresis (CE-MS), which is increasingly used in metabolomics due to its low sample volume requirement and high efficiency

Table 2 Animal and cellular metabolomics studies of CIPN

Subjects	Neurotoxic chemotherapy	CIPN data	Metabolomics techniques	Sample type	Statistical methods	Findings	Reference
Adult male rat model of oxaliplatin-induced acute pain hypersensitivity	Single intraperitoneal oxaliplatin injection	Behavioral tests of mechanical allodynia, pain hypersensitivity, and cold hypersensitivity	NMR, untargeted 18 metabolites	Spinal dorsal horn tissue, synthetic diet available ad libitum	Mann–Whitney U tests	Increased glutamine and decreased creatinine and adenosine phosphate were found in oxaliplatin-treated rats, and polyamine-deficient diet prevented such metabolic changes and oxaliplatin-induced acute pain hypersensitivity	(Ferrier et al. 2013)
Male rat model of oxaliplatin-induced painful neuropathy	Intravenous oxaliplatin injection twice a week for four and half weeks	Behavioral tests of mechanical allodynia, cold hypersensitivity	NMR, untargeted 17 metabolites	Various brain tissues, food available ad libitum	Mann–Whitney U tests	Higher choline and glycerophosphocholine and lower aspartate, creatinine, glutamine, and GABA were found in oxaliplatin-treated rats	(Ferrier et al. 2015)
Female rat model of paclitaxel-induced peripheral neuropathy	Intraperitoneal paclitaxel injection on three alternate days	Behavioral tests of mechanical allodynia	LC-MS/MS, untargeted, total number of metabolites not reported	Plasma, food available ad libitum	PLSDA and t tests	Fatty acids, ketones, and glycerophosphocholine differentiated between paclitaxel-treated rats and controls	(Wu et al. 2018)
Human neural progenitor cells	Low-dose thalidomide for 3 days	Immunofluorescence imaging of neuronal differentiation	CE-MS, untargeted 74 metabolites	Cell culture	Mann–Whitney U tests	Lower glutathione was found in thalidomide-treated cells during neural differentiation	(Qin et al. 2012)

Mechanical allodynia is a painful sensation caused by innocuous stimuli such as light touch of clothes

for separating polar metabolites (Harada et al. 2018; Kristoff et al. 2020; Zhang and Ramautar 2021). This study analyzed human neural progenitor cell cultures treated with low-dose thalidomide and found, compared to vehicle-treated controls, lower glutathione in thalidomide-treated cells during neural differentiation. Of the metabolites found in these animal and cellular studies, only adenosine phosphate was consistent with the metabolites identified in the CIPN clinical studies. This may be due to the use of vehicle-treated controls, as opposed to treated rats that did not develop CIPN, suggesting these results may reflect the direct effect of chemotherapy and be unrelated to CIPN.

2.5 Summary

Two pioneering metabolomics studies have been conducted in patients with cancer receiving neurotoxic chemotherapy to identify metabolic predictors of CIPN. These studies suggest six metabolites with putative pathways that may be pretreatment or early-in-treatment predictors of CIPN: histidine, phenylalanine, threonine (Sun et al. 2018), N-acetylornithine, glycogen, and adenosine mono/diphosphate (Verma et al. 2020). Except for adenosine phosphate, studies in animal models and neuronal progenitor cells found different metabolites that are affected by chemotherapy treatment, but their relationship to CIPN is unknown (Ferrier et al. 2013, 2015; Qin et al. 2012; Wu et al. 2018).

3 Investigation of Peripheral Neuropathy Metabolomic Biomarker Pathways

The rest of this chapter further investigates the putative associations between these six potential metabolomic biomarkers of CIPN, which will be listed in descending order of the strength of evidence (Table 3). These pathways are examined for any evidence of their involvement in human metabolomics studies of other types of peripheral neuropathy, including diabetic neuropathy, hereditary neuropathy, and peripheral neuropathy caused by autoimmune disorders. It can be difficult to collect metabolic or peripheral neuropathy information before the onset of these other etiologies of chronic peripheral neuropathy. Therefore, metabolomics studies usually compare metabolic profiles between patients with peripheral neuropathy and controls, who are either healthy volunteers or patients who share the underlying disease state but are not experiencing peripheral neuropathy. Some studies also analyzed the association between metabolite levels and the severity of peripheral neuropathy. Due to the large number of metabolites and the complex and unclear mechanisms underlying peripheral neuropathy, many studies also adjust for clinical covariates to try to identify the most representative metabolites.

In addition to reviewing metabolomics studies of other etiologies of peripheral neuropathy, for each of these six potential metabolomic biomarkers of CIPN, the putative mechanistic relationship with neuropathy is reviewed and any observational

Table 3 Evidence of association for each of the putative CIPN metabolomics biomarkers

Metabolites	Description	CIPN metabolomics	Other PN metabolomics	Other observational studies	Other interventional studies	Putative mechanisms
Histidine	Essential from food Precursor of histamine, which is a neurotransmitter Metabolites such as carnosine and ergothioneine have also been found to be associated with PN	Lower pretreatment plasma histidine associated with more severe sensory paclitaxel-induced peripheral neuropathy (Sun et al. 2018)	Higher plasma histidine associated with diabetic neuropathy (Lin et al. 2019) Lower plasma histidine correlated with worse cardiovascular autonomic neuropathy (Mathew et al. 2019) Lower skin fibroblast histidine levels associated with Leber hereditary optic neuropathy (Chao de la Barca et al. 2016) Higher plasma 1-methylhistidine levels associated with Leber hereditary optic neuropathy (Bocca et al. 2021)	Nominally higher blood ergothioneine in patients with less sensory CIPN (Winkels et al. 2020)	Carnosine supplementation prevents oxaliplatin CIPN (Yehia et al. 2019) Ergothioneine prevents platinum (Jong et al. 2011; Nishida et al. 2018; Song et al. 2010) CIPN in animals Histamine inhibitors prevent taxane (Gao et al. 2016; Nagano et al. 2012), oxaliplatin (Azevedo et al. 2013), or bortezomib (Tsukaguchi et al. 2013) CIPN pain in patients or animals	Histidine, carnosine, and ergothioneine have anti-oxidant and anti-inflammatory properties Ergothioneine inhibits oxaliplatin uptake in neurons Histamine inhibitors reduce pain detection in sensory neurons 1-Methylhistidine via unknown mechanism

(continued)

Table 3 (continued)

Metabolites	Description	CIPN metabolomics	Other PN metabolomics	Other observational studies	Other interventional studies	Putative mechanisms
Phenylalanine	Essential from food Precursor of tyrosine and catecholamine neurotransmitters such as levodopa Precursor of phenethylamine, which is a neurotransmitter	Lower pretreatment plasma phenylalanine associated with more severe sensory paclitaxel-induced peripheral neuropathy (Sun et al. 2018)	Lower skin fibroblast phenylalanine and tyrosine levels associated with Leber hereditary optic neuropathy (Chao de la Barca et al. 2016; Morvan and Demidem 2018)	Dietary phenylalanine restriction associated with sensory neuropathy in patients (Ludolph et al. 1992) and cats (Dickinson et al. 2004)	Levodopa relieves diabetic neuropathic pain (Ertas et al. 1998) A phenethylamine derivative prevents and relieves acute oxaliplatin-induced neurotoxicity (Durand et al. 2012) and diabetic neuropathic pain (Kadiroglu et al. 2008; Razazian et al. 2014; Rowbotham et al. 2004)	Deficiencies of phenylalanine and tyrosine cause sensory neuropathy
N-acetylornithine	Non-essential Metabolized to ornithine and then polyamines Ornithine predominantly from arginine	N-acetylornithine levels on day 29 predictive of vincristine-induced peripheral neuropathy (Verma et al. 2020)	Lower plasma ornithine correlated with worse cardiovascular autonomic neuropathy in diabetic patients (Mathew et al. 2019) Higher plasma N-acetylputrescine	Lower serum arginine associated with diabetic neuropathy (Ganz et al. 2017)	Arginine supplementation prevents diabetic neuropathy pain in rats (Rondón et al. 2018), but not significant in patients (Jude et al. 2010) A polyamine-deficient diet	Arginine is anti-oxidant Polyamine-deficient diet reduces neuronal excitotoxicity

			levels associated with peripheral neuropathy in obese patients (Guo et al. 2021) Lower skin fibroblast arginine and polyamines levels associated with Leber hereditary optic neuropathy (Chao de la Barca et al. 2016)		prevents pain hypersensitivity (Rivat et al. 2008) including oxaliplatin-induced acute pain hypersensitivity in rats (Ferrier et al. 2013)	
Threonine	Essential from food Small amount converts to glycine	Lower pretreatment plasma threonine associated with more severe sensory paclitaxel-induced peripheral neuropathy (Sun et al. 2018)	Lower skin fibroblast threonine and glycine levels associated with Leber hereditary optic neuropathy (Chao de la Barca et al. 2016; Morvan and Demidem 2018) Higher cerebrospinal fluid threonine levels found in patients with endemic optic neuropathy (Gonzalez-Quevedo et al. 2001)	Higher glycine levels in spinal cord tissues in mice with Charcot-Marie-Tooth type 2D neuropathy (Bais et al. 2016)	Glycine antagonist did not improve neuropathic pain (Wallace et al. 2002)	No mechanism determined

(continued)

Table 3 (continued)

Metabolites	Description	CIPN metabolomics	Other PN metabolomics	Other observational studies	Other interventional studies	Putative mechanisms
Glycogen	Glucose storage Glucose is its precursor and metabolite	Glycogen levels on day 29 predictive of vincristine-induced peripheral neuropathy (Verma et al. 2020)	N/A	Glycogen accumulation in neurons or Schwan cells of patients with late-onset Pompe disease (Finsterer et al. 2017) and diabetic neuropathy (Bischoff 1980; Powell et al. 1985; Yagihashi and Matsunaga 1979)	Anti-glycemic agents prevent diabetic neuropathy (El Mouhayyar et al. 2020) and possibly treat CIPN (Erdoğan et al. 2020; Fujita et al. 2015; Shigematsu et al. 2020)	Glycogen accumulation causes neuropathy Hyperglycemia causes neuronal injury and ischemia
Adenosine phosphates	DNA and RNA purine precursors Energy storage	Adenosine mono/diphosphate levels on day 29 predictive of vincristine-induced peripheral neuropathy (Verma et al. 2020) Decreased adenosine phosphate indicative of oxaliplatin-induced acute pain hypersensitivity (Ferrier et al. 2013)	N/A	Impaired adenosine phosphate signaling pathway in mice with diabetic neuropathy (Roy Chowdhury et al. 2012) Patients taking an adenosine phosphate signaling pathway activator had lower risk of diabetic neuropathy (Pop-Busui et al. 2009)	An adenosine phosphate signaling pathway activator prevents oxaliplatin CIPN in patients (El-Fatatry et al. 2018) and mice (Mao-Ying et al. 2014)	Adenosine phosphate signaling pathways promoting mitochondrial biogenesis and inhibiting autophagy and neuroinflammation

and interventional studies that support the association are described. Four of these metabolites are amino acids, which are the building blocks for protein synthesis. In humans, there are 20 proteinogenic amino acids, of which nine essential amino acids are not produced endogenously and must be supplied from the diet. Many amino acids participate in neurotransmitter biosynthesis, providing an apparent mechanism by which they may be involved in peripheral neuropathy (Dalangin et al. 2020; Kölker 2018). The other two metabolites are glycogen and adenosine phosphate. Glycogen plays a critical role in carbohydrate metabolism (Nordlie et al. 1999), and adenosine phosphate is the precursor of RNA and provides energy for cellular metabolic process (Hardie et al. 2012).

3.1 Histidine

Histidine is an essential amino acid supplied from dietary sources, such as meat. Histidine has an imidazole function group that can scavenge reactive oxygen species to achieve anti-oxidant and anti-inflammatory effects (Peterson et al. 1998). This section will also discuss four of its potentially peripheral neuropathy-related metabolites, 1-methylhistidine, carnosine, ergothioneine, and histamine (Fig. 1).

Similar to the paclitaxel-induced peripheral neuropathy metabolomics study (Sun et al. 2018), two metabolomics studies identified that histidine was associated with diabetic neuropathy, but the directions of associations were inconclusive, and the associations were not significant after adjusting for clinical covariates (Lin et al.

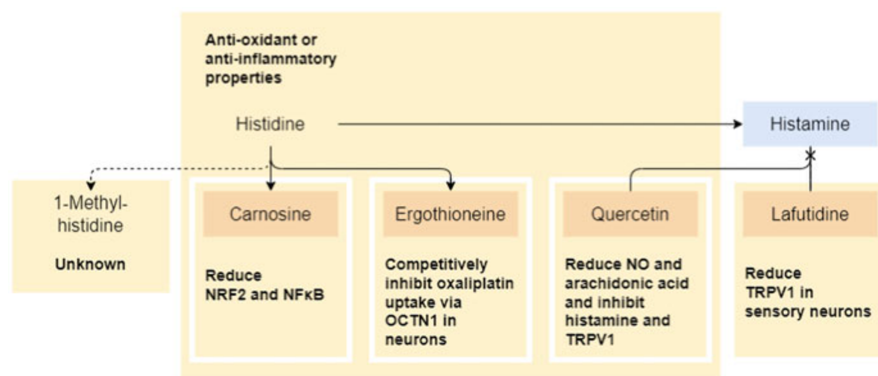


Fig. 1 Histidine and its peripheral neuropathy-related metabolites have several putative mechanisms (Bocca et al. 2021; Chao de la Barca et al. 2016; Jong et al. 2011; Lin et al. 2019; Mathew et al. 2019; Nishida et al. 2018; Song et al. 2010; Sun et al. 2018; Winkels et al. 2020; Yehia et al. 2019). Quercetin and lafutidine are not histidine metabolites, but they are histamine inhibitors (Azevedo et al. 2013; Gao et al. 2016; Nagano et al. 2012; Tsukaguchi et al. 2013). Blue boxes are neurotransmitters, such as histamine. Orange boxes are metabolites or compounds that have been found to prevent or treat CIPN. Solid arrows are metabolic reactions that can happen in humans, and dashed arrows are the ones that happen exclusively in animals. X-shaped arrows indicate inhibition

2019; Mathew et al. 2019). A longitudinal MS-based study by Mathew et al. (2019) measured plasma amino acids and tricarboxylic acid metabolites in patients with type I diabetes but without any signs of microvascular complications and age- and sex-matched healthy controls at baseline and for 3 years of observation. Correlation coefficients between baseline metabolite levels and end-of-study electrocardiogram parameters were calculated to identify metabolites that predict cardiovascular autonomic neuropathy. Lower histidine levels were moderately correlated with worse neuropathy but did not remain significant after adjusting for baseline clinical covariates, including blood glucose, HbA1c, years of diabetes, body mass index, cholesterol, estimated glomerular filtration rate, and urine microalbumin-to-creatinine ratio. An untargeted NMR study by Lin et al. (2019) enrolled patients with type II diabetes with and without microangiopathy and healthy controls. Orthogonal partial least squares discriminant analysis (OPLSDA) was used to identify predictive metabolites and build predictive models of diabetes and diabetic microangiopathy complications including sensory diabetic neuropathy. Higher plasma histidine was identified in multiple models that predicted the occurrence of diabetic neuropathy, but the association was not significant after adjusting for clinical covariates, including age, sex, body mass index, and the use of medication for hypertension and hyperlipidemia. The lack of association after covariate adjustment and seemingly conflicting findings in these studies does not eliminate the possibility of histidine being involved in the development of peripheral neuropathy, but indicates that the causal relationship between histidine, diabetes, and peripheral neuropathy may be complicated. For example, the two studies of diabetic neuropathy required fasting before sample collection (Lin et al. 2019; Mathew et al. 2019), but the paclitaxel-induced peripheral neuropathy study did not (Sun et al. 2018). Since diet can affect CIPN (Mongiovi et al. 2018), diabetic neuropathy (Bunner et al. 2015; Oza et al. 2021), and diabetes (Neuenschwander et al. 2019; Schwingshackl et al. 2017), and most of the histidine in the body is supplied by dietary sources, it is difficult to determine whether lower or higher plasma histidine is causally related to peripheral neuropathy or this relationship is due to statistical confounding. More studies are needed to elucidate the effect of histidine on, and its potential for use as a predictive biomarker of, peripheral neuropathy.

Several peripheral neuropathy metabolomics studies were focused on Leber hereditary optic neuropathy (LHON), which is an inherited disease that results in vision loss that can include other elements of peripheral neuropathy such as tremors and movement disorders. LHON is caused by mutations of mitochondrial DNA including *MT-ND1*, *MT-ND4*, *MT-ND4L*, or *MT-ND6*, which are critical for proper mitochondrial function. Studies have analyzed skin fibroblast metabolomics in patients with LHON to understand the metabolic changes caused by mitochondrial dysfunction without the confounding of the short-term effects of diet, medication, and circadian cycles, which complicate research of systemic metabolomics (Wilkins et al. 2019). A targeted LC-MS/MS metabolomics study by Chao de la Barca et al. (2016) used OPLSDA to compare skin fibroblast metabolomics between patients with LHON and healthy controls. Levels of almost all proteinogenic amino acids, including histidine, were significantly lower in patients with LHON (Chao de la

Barca et al. 2016), which is consistent with the known function of mitochondria synthesizing amino acids to release stored energy.

Histidine has not been tested for prevention or treatment of any types of peripheral neuropathy. However, histidine supplementation in obese women with metabolic syndrome has been demonstrated in a randomized controlled trial to suppress inflammation and oxidative stress (Feng et al. 2013), which may also protect against CIPN. Interventional studies of histidine supplementation for peripheral neuropathy prevention and treatment may be warranted.

3.1.1 1-Methylhistidine

1-Methylhistidine is a histidine metabolite that is found exclusively in red meat or poultry, making it a biomarker for meat consumption (Altorf-van der Kuil et al. 2013; Dragsted 2010; Fraser et al. 2016; Khodorova et al. 2019; Kochlik et al. 2018; Lloyd et al. 2011; Mitry et al. 2019; Sjölin et al. 1987; Yin et al. 2017), and has a half-life about half day (Sjölin et al. 1987). Another untargeted LC-MS metabolomics study of LHON by Bocca et al. (2021) found higher plasma 1-methylhistidine in patients than healthy controls in an OPLSDA multivariable model. This study required fasting before sample collection (Bocca et al. 2021), so this result is likely attributed to habitual meat consumption. However, prior studies have not found that higher meat consumption increases CIPN (Kenkhuis et al. 2021; Mongiovi et al. 2018). Alternatively, two small interventional studies demonstrated that patients with diabetic neuropathy had less pain after receiving a low-fat and plant-based diet (Bunner et al. 2015; Smith et al. 2006). Again, the causal relationships between diet, metabolites, and neuropathy are likely complex and not fully understood.

3.1.2 Carnosine

Carnosine is a dipeptide of histidine and alanine that is abundant in human and animal muscle and brain tissue and is a biomarker for red meat consumption (Altorf-van der Kuil et al. 2013; Cuparencu et al. 2019; Mitry et al. 2019). Observational metabolomics studies of meat intake often find increased carnosine (Cheung et al. 2017; Cuparencu et al. 2019; Vázquez-Fresno et al. 2015). Similar to histidine, carnosine has anti-oxidant and anti-inflammatory properties. In a randomized controlled trial, carnosine supplementation was protective against oxaliplatin-induced peripheral neuropathy, possibly by reducing nuclear factor erythroid-2 related factor-2 (NRF2) and nuclear factor kappa light chain enhancer of activated B cells (NFκB) (Yehia et al. 2019). Nrf2 is involved in the redox pathway and has been proposed as a potential treatment target for CIPN (Miao et al. 2019; Yang et al. 2018) and diabetic neuropathy (Ganesh Yerra et al. 2013; Gupta et al. 2021; Kumar and Mittal 2017; Xu et al. 2013). Similarly, NFκB contributes to CIPN (Janes et al. 2014; Li et al. 2015; Wang et al. 2017) and diabetic neuropathy (Cameron and Cotter 2008; Dewanjee et al. 2018; Ganesh Yerra et al. 2013) through its involvement in inflammatory pathways.

3.1.3 Ergothioneine

The other histidine downstream metabolite of interest, ergothioneine, is synthesized exclusively by fungi and bacteria, which makes it a biomarker for mushroom consumption (Pallister et al. 2016). An observational study of patients with colorectal cancer receiving chemotherapy found that patients with higher blood ergothioneine had nominally, but not significantly, less sensory CIPN symptoms (Winkels et al. 2020). Similar to histidine and carnosine, ergothioneine is a potent anti-oxidant that may protect injured tissue (Cheah and Halliwell 2012; Halliwell et al. 2016). In addition to its anti-oxidant properties, ergothioneine has also been found in animal studies to reduce oxaliplatin accumulation in dorsal root ganglion neurons via inhibition of organic cation/carnitine transporter (OCTN1)-mediated oxaliplatin transport (Fujita et al. 2019; Gründemann 2012; Jong et al. 2011; Nishida et al. 2018). Ergothioneine has been found to protect against cisplatin (Song et al. 2010) and oxaliplatin-induced peripheral neuropathy in animal studies (Jong et al. 2011; Nishida et al. 2018), but we are not aware of any interventional studies in humans.

3.1.4 Histamine

Histidine is a precursor of histamine, which is a neurotransmitter that participates in regulation of pain and neurogenic inflammation (Rosa and Fantozzi 2013; Yu et al. 2016). Histamine may have a role in diabetes and diabetic neuropathy (Pini et al. 2016; Wei et al. 2016), and there is a potential that histidine levels affect the nervous system through histamine pathways, but we are not aware of any observational peripheral neuropathy metabolomics studies that have directly assessed histamine. A flavonoid from plants, quercetin, has been found to inhibit mast cells from releasing antigen-induced histamine (Kimata et al. 2000; Pearce et al. 1984) and subsequently inhibit transient receptor potential cation channel subfamily V member 1 (TRPV1) (Gao et al. 2016), which has functions of detection and regulation of heat and pain (Benítez-Angeles et al. 2020; Cui et al. 2016; Frias and Merighi 2016; Gouin et al. 2017; Romanovsky et al. 2009). Quercetin has been found to have anti-oxidant properties via the nitric oxide pathways (Di Carlo et al. 1999; Gao et al. 2016) and anti-inflammatory property via the arachidonic acid pathways (Di Carlo et al. 1999; Ferrándiz and Alcaraz 1991) and has been found in animal studies to prevent pain from oxaliplatin-(Azevedo et al. 2013) and paclitaxel-induced peripheral neuropathy (Gao et al. 2016). A histamine H2 receptor antagonist, lafutidine, has been found in small clinical trials to potentially be effective for treating taxane-induced peripheral neuropathy (Nagano et al. 2012) and prevent bortezomib-induced peripheral neuropathy (Tsukaguchi et al. 2013). The pain relieving effect of lafutidine has been hypothesized to be similar to capsaicin, which activates and then reduces the expression of TRPV1 (TRPV1 defunctionalization) in sensory neurons (Onodera et al. 1995, 1999; Umeda et al. 1999; Yamamoto et al. 2001).

3.1.5 Summary

Lower histidine levels have been found to be associated with paclitaxel-induced peripheral neuropathy (Sun et al. 2018) and LHON (Chao de la Barca et al. 2016),

but the directions of associations in diabetic neuropathy were inconclusive (Lin et al. 2019; Mathew et al. 2019). The evidence suggests that histidine plays some role in peripheral neuropathy. However, the causal mechanistic relationship is unclear, which likely explains the lack of significance after adjusting for clinical covariates (Lin et al. 2019; Mathew et al. 2019).

Metabolomics studies of LHON found higher levels of 1-methylhistidine (Bocca et al. 2021), which has not been identified in observational dietary analyses (Kenkhuis et al. 2021; Mongiovi et al. 2018), but suggested that reducing meat consumption may be a possible dietary intervention of peripheral neuropathy (Bunner et al. 2015; Smith et al. 2006). If meat consumption is demonstrated to increase CIPN risk, 1-methylhistidine and carnosine are indicators of meat consumption that could be used as CIPN biomarkers (Altorf-van der Kuil et al. 2013; Cuparencu et al. 2019; Mitry et al. 2019).

Histidine, carnosine, and ergothioneine all have anti-oxidant and anti-inflammatory properties that can protect injured tissues and possibly protect against peripheral neuropathy. Carnosine supplementation may prevent oxaliplatin-induced peripheral neuropathy through anti-oxidant and anti-inflammatory pathways (Yehia et al. 2019), whereas ergothioneine may have an additional mechanism of reducing oxaliplatin accumulation in neurons (Fujita et al. 2019; Gründemann 2012; Jong et al. 2011; Nishida et al. 2018; Winkels et al. 2020).

Histidine may affect the nervous system through histamine, a neurotransmitter that regulates pain and neurogenic inflammation (Rosa and Fantozzi 2013; Yu et al. 2016). Histamine inhibitors such as quercetin (Azevedo et al. 2013; Gao et al. 2016) and lafutidine (Nagano et al. 2012; Tsukaguchi et al. 2013) may prevent or relieve CIPN pain. Further experimental work is needed to confirm these mechanisms and randomized interventional trials are needed to confirm the effectiveness for prevention or treatment of CIPN.

3.2 Phenylalanine

Phenylalanine is an essential aromatic amino acid found in high-protein food sources, such as meat and cheese. Phenylalanine is abundant in human brain tissue and plasma. Phenylalanine regulates insulin secretion (Floyd et al. 1966; Guasch-Ferré et al. 2016; Urpi-Sarda et al. 2019; Wang et al. 2011; Wishart 2019) and is associated with higher risk of diabetes (Guasch-Ferré et al. 2016; Park et al. 2018; Sun et al. 2020). This section will discuss phenylalanine and two of its potentially peripheral neuropathy-related metabolites, tyrosine and phenethylamine (Fig. 2).

3.2.1 Phenylalanine and Tyrosine

Phenylalanine is the precursor of tyrosine, which is also an essential aromatic amino acid. Tyrosine is rapidly metabolized to 3,4-dihydroxyphenylalanine (levodopa), which is the precursor of many catecholamine neurotransmitters, including dopamine, norepinephrine, and epinephrine. Tyrosine is associated with higher risk of diabetes (Guasch-Ferré et al. 2016; Park et al. 2018; Sun et al. 2020) and many

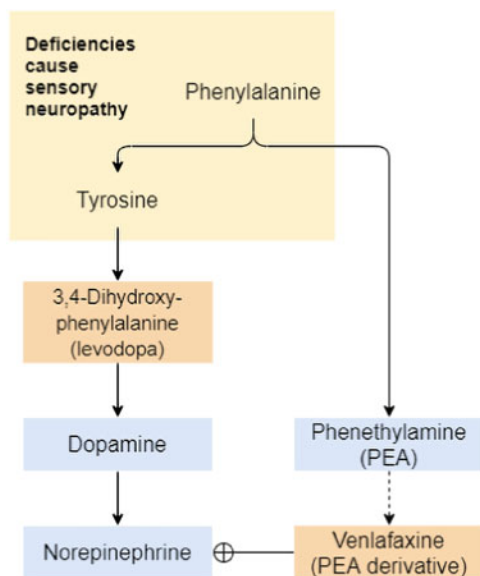


Fig. 2 Phenylalanine and its peripheral neuropathy-related metabolites (Chao de la Barca et al. 2016; Dickinson et al. 2004; Ertas et al. 1998; Ludolph et al. 1992; Morvan and Demidem 2018; Sun et al. 2018). Blue boxes are neurotransmitters. Orange boxes are metabolites or compounds that have been found to relieve neuropathic pain. Solid arrows are metabolic reactions that can happen in humans, and circle-shaped arrows indicate induction. Venlafaxine is a phenethylamine derivative that inhibits the reuptake of norepinephrine, which leads to increased norepinephrine levels (Durand et al. 2012; Kadiroglu et al. 2008; Razazian et al. 2014; Rowbotham et al. 2004)

disorders of the central nervous system, including headache (D'Andrea et al. 2019), depression (Bot et al. 2020), and Alzheimer's disease (Albrahim 2020), but that is beyond the scope of this chapter.

Even though phenylalanine and its metabolites seem to play a role in neurotransmission, only two peripheral neuropathy metabolomics studies identified differences in phenylalanine levels. These two LHON skin fibroblast metabolomics studies found most of the proteinogenic amino acids levels, including phenylalanine and tyrosine, were significantly lower in patients with LHON (Chao de la Barca et al. 2016; Morvan and Demidem 2018). In a previously mentioned diabetic neuropathy study, higher plasma tyrosine, but not phenylalanine, was identified in some multivariable models, but the performance of these models was not ideal, and tyrosine was not significant after adjusting for clinical covariates (Lin et al. 2019).

There are no interventional trials of phenylalanine supplementation to prevent or treat peripheral neuropathy, to the best of our knowledge. However, an observational study suggests that dietary phenylalanine restriction in patients with hereditary hyperphenylalaninemia causes sensory neuropathy (Ludolph et al. 1992). Another study in cats also showed that restricting dietary phenylalanine and tyrosine can cause sensory neuropathy (Dickinson et al. 2004). This strongly implicates

phenylalanine deficiency as a causal mechanism for sensory neuropathy. A meta-analysis found that tyrosine supplementation does not improve the neuropsychological performance in patients with phenylketonuria who receive a phenylalanine-restricted diet (Remmington and Smith 2021). Levodopa, which is within the previously mentioned catecholamine pathway, has been shown to relieve diabetic neuropathic pain in a small clinical trial (Ertas et al. 1998), which provides further support for this pathway and suggests the potential for repurposing levodopa for neuropathic pain.

3.2.2 Phenethylamine

The other phenylalanine metabolite, phenethylamine, is a brain neurotransmitter. Human metabolomics studies of the use of methamphetamine (Boxler et al. 2017; Kim et al. 2020), a phenethylamine derivative, have analyzed phenethylamine, but we are not aware of any observational peripheral neuropathy metabolomics studies that have assessed phenethylamine.

Another phenethylamine derivative, venlafaxine, is a serotonin-norepinephrine reuptake inhibitor antidepressant that has been found to prevent and relieve acute oxaliplatin-induced neurotoxicity (Durand et al. 2012), but not oxaliplatin-induced peripheral neuropathy (Zimmerman et al. 2016). Venlafaxine has also been found to reduce pain from diabetic neuropathy (Kadiroglu et al. 2008; Razazian et al. 2014; Rowbotham et al. 2004), but is less effective than duloxetine and pregabalin for treating CIPN (Farshchian et al. 2018) or diabetic neuropathy (Raskin et al. 2006; Razazian et al. 2014).

3.2.3 Summary

Similar to the findings in CIPN (Sun et al. 2018), metabolomics studies in LHON (Chao de la Barca et al. 2016; Morvan and Demidem 2018) indicate that phenylalanine or its metabolites may be involved in peripheral neuropathy. Two studies of phenylalanine or tyrosine dietary restriction confirm this causes sensory peripheral neuropathy (Dickinson et al. 2004; Ludolph et al. 1992), likely through the catecholamine neurotransmitter pathway. Catecholamines have demonstrated effectiveness in neuropathy-relevant conditions including the effectiveness of levodopa and venlafaxine for relieving neuropathic pain caused by chemotherapy (Durand et al. 2012) or diabetes (Ertas et al. 1998; Kadiroglu et al. 2008; Razazian et al. 2014; Rowbotham et al. 2004).

3.3 N-Acetylornithine

N-acetylornithine is an intermediate metabolic precursor of ornithine metabolism. There is minimal information of the source or physiological role of N-acetylornithine in humans. Other than the vincristine-induced peripheral neuropathy study (Verma et al. 2020), only one study found higher N-acetylornithine levels to be associated with the treatment effect of a tremor and involuntary movement disorder (Napoli et al. 2019). No other observational or interventional peripheral neuropathy studies

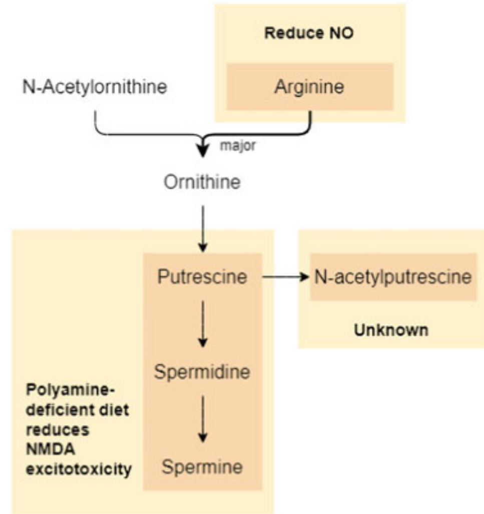


Fig. 3 N-acetylornithine and its peripheral neuropathy-related metabolites (Chao de la Barca et al. 2016; Ferrier et al. 2013; Ganz et al. 2017; Jude et al. 2010; Mathew et al. 2019; Rivat et al. 2008); Rondón et al. (2018); (Verma et al. 2020). Orange boxes are metabolites that have been found with effects in peripheral neuropathy. Solid arrows are metabolic reactions that can happen in humans

have assessed N-acetylornithine. The major metabolite of N-acetylornithine is ornithine. In this section, we will discuss ornithine, together with ornithine's major precursor, arginine, and a group of ornithine downstream metabolites, polyamines (Fig. 3).

3.3.1 Ornithine and Arginine

While ornithine is abundant in many plant food sources, ornithine is mostly produced from arginine in the urea cycle to dispose excess nitrogen. Only one metabolomics study has identified an association between ornithine and neuropathy. The longitudinal MS-based study by Mathew et al. (2019) found lower plasma ornithine levels were moderately correlated with worse cardiovascular autonomic neuropathy in patients with diabetes, and this association remained significant after adjusting for clinical covariates at baseline. Another study found lower serum arginine, but not ornithine, levels were associated with higher risk of peripheral neuropathy in patients with diabetes (Ganz et al. 2017). Two previously mentioned metabolomics studies found lower skin fibroblast levels of arginine, but not ornithine, in patients with LHON (Chao de la Barca et al. 2016; Morvan and Demidem 2018).

Ornithine has not been examined in interventional studies of peripheral neuropathy. Arginine supplementation has been found to prevent painful diabetic neuropathy in a rat diabetes model by reducing nitric oxide production (Rondón et al. 2018). However, in a randomized controlled trial arginine supplementation was not effective in reducing neuropathy in patients with diabetic neuropathy, though this could be due to the study being underpowered or confounding caused by glucose-lowering medications (Jude et al. 2010).

3.3.2 Ornithine and Polyamines: Putrescine, Spermidine, and Spermine

Polyamines, including putrescine, spermidine, and spermine, are synthesized from ornithine and are mainly from dietary intake and gut flora metabolism. Spermidine and spermine positively modulate N-methyl-D-aspartate (NMDA) receptors by shielding an inhibition site (Subramaniam et al. 1994; Traynelis et al. 1995), and a prolonged activation of NMDA receptors can lead to neuronal damage (Paschen 1992). An untargeted LC-MS/MS metabolomics study by Guo et al. compared glycemic status-matched obese patients with and without peripheral neuropathy via PLSDA and group least absolute shrinkage and selection operator (LASSO) regression after adjusting for age and sex. They found obese patients with peripheral neuropathy had higher plasma N-acetylputrescine, but the mechanisms of polyamines in PN have not been well studied (Guo et al. 2021). However, a previously mentioned skin fibroblast metabolomics study found lower levels of putrescine and spermidine in patients with LHON, but the mechanism for this seemingly inverse relationship was unclear (Chao de la Barca et al. 2016).

A polyamine-deficient diet has been found to protect against pain hypersensitivity (Rivat et al. 2008) including oxaliplatin-induced acute pain hypersensitivity in rats (Ferrier et al. 2013). An interventional randomized controlled trial of the efficacy of the polyamine-deficient diet in oxaliplatin-induced peripheral neuropathy is ongoing (Balayssac et al. 2015). We are not aware of any interventional studies testing the effect of restricting an individual polyamine.

3.3.3 Summary

The direct association between N-acetylornithine and peripheral neuropathy was only found in the vincristine-induced peripheral neuropathy metabolomics study (Verma et al. 2020), but lower levels of ornithine, arginine, and polyamines have been found in patients with other etiologies of peripheral neuropathy. The minimal data from interventional studies of arginine supplementation trials indicate limited efficacy. Despite the seemingly conflicting findings in observational studies of polyamines, polyamine-deficient diet may protect against peripheral neuropathic pain through NMDA pathways (Ferrier et al. 2013; Rivat et al. 2008). More studies with a larger sample size are needed to explore the role of ornithine, arginine, or polyamines in peripheral neuropathy.

3.4 Threonine

The association between lower threonine and peripheral neuropathy was found in the paclitaxel-induced peripheral neuropathy whole blood metabolomics study (Sun et al. 2018) Threonine is an essential amino acid found in high-protein food sources, such as meat and cheese. Threonine is abundant in human plasma and does not seem to be directly associated with neuropathy, but a small amount of threonine is converted to glycine (Darling et al. 2000; Edgar 2002; Zhao et al. 1986). This section will discuss threonine and glycine together.

Three metabolomics studies have reported an association between threonine and optic neuropathy (Chao de la Barca et al. 2016; Gonzalez-Quevedo et al. 2001; Morvan and Demidem 2018). The first two are previously mentioned skin fibroblast metabolomics studies, which found patients with LHON had lower levels of most of the proteinogenic amino acids, including threonine and glycine (Chao de la Barca et al. 2016; Morvan and Demidem 2018). The other study by Gonzalez-Quevedo et al. (2001) was an LC metabolomics study of endemic optic neuropathy, which is suspected to be due to nutrient insufficiency. Threonine levels were significantly higher in cerebrospinal fluid, but not serum, in patients with optic neuropathy. This study did not require fasting before sample collection, but the samples were collected in the early morning to minimize the influence from circadian rhythms and food. No other plasma metabolomics studies have found an association between threonine and peripheral neuropathy to the best of our knowledge.

In a mouse study of Charcot-Marie-Tooth type 2D neuropathy, which is an inherited neuropathy caused by mutations of glycyl-tRNA synthetase 1 (*GARS1*), mice with neuropathy had higher glycine levels in spinal cord tissues. However, glycine supplementation did not worsen or affect the neuropathy assessment (Bais et al. 2016), which implied that high glycine was not the cause of neuropathy. Threonine has not been tested in interventional studies of peripheral neuropathy, and a glycine antagonist did not improve pain or neuropathy symptoms in a clinical trial (Wallace et al. 2002). Overall, the evidence supporting an association of threonine or its metabolites is weak and no interventional studies demonstrate effectiveness of neuropathy prevention or treatment.

3.5 Glycogen

Glycogen is a polysaccharide that is the primary storage form of glucose. Other than the vincristine-induced peripheral neuropathy study (Verma et al. 2020), no other peripheral neuropathy metabolomics studies have assessed glycogen. An inherited disorder, Pompe disease, is caused by mutations of alpha glucosidase (GAA), which leads to glycogen accumulation in lysosomes, especially in heart and skeletal muscles. Neuropathy has been reported in patients with late-onset Pompe disease due to glycogen accumulation in neurons or Schwan cells (Finsterer et al. 2017). Aggregates of glycogen particles have been found in neurons and Schwann cells in patients with diabetic neuropathy as well (Bischoff 1980; Powell et al. 1985; Yagihashi and Matsunaga 1979). These studies strongly suggest that glycogen accumulation is associated with peripheral neuropathy, but the role of glycogen in CIPN is unclear.

Chronic elevation of glucose, the precursor and metabolite of glycogen, is the main cause of diabetes, and high glucose levels in peripheral nerves is the pathogenesis of diabetic neuropathy (Freeman et al. 2016). Hyperglycemia causes injury to neuronal cells and excessive glucose metabolism promotes thickening of capillary basement membrane, which leads to neuronal ischemia (Filla and Edwards 2016; Forbes and Cooper 2013). Numerous anti-glycemic agents have been approved for

reducing blood glucose in patients with diabetes and is expected to reduce the risk of diabetic microvascular complication including peripheral neuropathy (El Mouhayyar et al. 2020), and some of these drugs also showed efficacy in treating CIPN through glucagon-like peptide-1 signaling pathways (Erdoğan et al. 2020; Fujita et al. 2015; Shigematsu et al. 2020). However, blood glucose varies by dietary intake, and glycosylated hemoglobin is considered to be a more accurate long-term biomarker of glycemic control than short-term blood glucose (Koenig et al. 1976).

3.6 Adenosine Phosphates

Adenosine monophosphate and adenosine diphosphate were found in a plasma metabolomics study of vincristine-induced peripheral neuropathy, but the concentrations were not reported (Verma et al. 2020). Adenosine monophosphate and adenosine diphosphate are purine-based nucleotides with one and two phosphate groups, respectively. They are the precursors of RNA and can convert to each other or adenosine triphosphate, commonly referred to as ATP, which provides energy for a variety of signaling process in cells. A possible mechanistic pathway connecting these metabolites to neuropathy is that both adenosine monophosphate and adenosine diphosphate activate adenosine monophosphate activated protein kinase (AMPK), which has been found to play an important role in diabetic neuropathy (Madhavi et al. 2019; Shrikanth and Nandini 2020). Impaired AMPK signaling is a component of mitochondrial dysfunction and diabetic neuropathy in a diabetic mouse model (Roy Chowdhury et al. 2012). AMPK activation can prevent and reverse the neuronal injury by promoting mitochondrial biogenesis and inhibiting autophagy and neuroinflammation (Roy Chowdhury et al. 2012; Yerra et al. 2017), which suggests that activation of AMPK can be a prevention or treatment target for diabetic neuropathy and possibly other peripheral neuropathy (Madhavi et al. 2019; Price et al. 2016; Shaw et al. 2020; Shrikanth and Nandini 2020). This may explain the finding that patients receiving the AMPK activator metformin, the first-line treatment for diabetes, had lower risk of developing neuropathy than patients receiving insulin, and the association was independent of the duration of diabetes and glycemic control (Pop-Busui et al. 2009). Clinical and mouse studies also suggests metformin may protect against oxaliplatin-induced peripheral neuropathy (El-Fatatry et al. 2018; Mao-Ying et al. 2014). Reduced AMPK activity can result in upregulation of transient receptor potential cation channel subfamily A member 1 (TRPA1). TRPA1 is related to, but in a different receptor subfamily as, the TRVP1 receptor that is highly associated with neuropathic pain and was previously mentioned in the histamine subsection (Wang et al. 2018). In summary, adenosine phosphates were identified in both clinical (Verma et al. 2020) and animal (Ferrier et al. 2013) CIPN metabolomics studies, and may work via AMPK signaling pathway activation. AMPK activators, such as metformin, have shown potentials to prevent and treat diabetic neuropathy and CIPN (El-Fatatry et al. 2018; Mao-Ying et al. 2014; Pop-Busui et al. 2009), further supporting the possibility that this pathway has a role in CIPN.

4 Conclusion

Although scarce, there have been a few CIPN metabolomics biomarker studies in clinical, animal, and cellular models. Our review of the clinical studies identified six metabolites that may predict CIPN risk. We were able to find supportive evidence for these associations by reviewing metabolomics, observational, and interventional studies of CIPN and other etiologies of peripheral neuropathy. Histidine and phenylalanine have clear roles in neurotransmission or neuroinflammation in peripheral neuropathy, including CIPN. Furthermore, interventional studies of their metabolites, including carnosine and levodopa, have shown promise in preventing or treating CIPIN and other types of peripheral neuropathy, though definitive interventional trials are needed. Further research is needed to discover and validate CIPN metabolomics biomarkers in large clinical studies and to elucidate the mechanisms and causal relationships through animal or cell culture experiments. After retrospective validation, prospective biomarker-based interventional studies are essential to translate the biomarker findings into clinical practice. For example, metabolomics biomarkers could be used to identify patients at high CIPN risk before or early in treatment, in whom CIPN monitoring could be enhanced or investigational preventive interventions could be tested. Metabolomic biomarkers can also be indicative of effective interventions, such as nutrient supplementation, to prevent or reduce CIPN. The ultimate goal is to improve treatment outcomes and quality of life of patients with cancer.

References

Albrahim T (2020) The potential role of nutritional components in improving brain function among patients with Alzheimers disease: a meta-analysis of RCT studies. Neurosciences 25:4–17. https://doi.org/10.17712/nsj.2020.1.20190037

Altorf-van der Kuil W, Brink EJ, Boetje M, Siebelink E, Bijlsma S, Engberink MF, van't Veer P, Tomé D, Bakker SJ, van Baak MA, Geleijnse JM (2013) Identification of biomarkers for intake of protein from meat, dairy products and grains: a controlled dietary intervention study. Br J Nutr 110:810–822. https://doi.org/10.1017/s0007114512005788

Argyriou AA, Iconomou G, Kalofonos HP (2008) Bortezomib-induced peripheral neuropathy in multiple myeloma: a comprehensive review of the literature. Blood 112:1593–1599. https://doi.org/10.1182/blood-2008-04-149385

Azevedo MI, Pereira AF, Nogueira RB, Rolim FE, Brito GA, Wong DV, Lima-Júnior RC, de Albuquerque RR, Vale ML (2013) The antioxidant effects of the flavonoids rutin and quercetin inhibit oxaliplatin-induced chronic painful peripheral neuropathy. Mol Pain 9:53. https://doi.org/10.1186/1744-8069-9-53

Backshall A, Sharma R, Clarke SJ, Keun HC (2011) Pharmacometabonomic profiling as a predictor of toxicity in patients with inoperable colorectal cancer treated with capecitabine. Clin Cancer Res 17:3019–3028. https://doi.org/10.1158/1078-0432.Ccr-10-2474

Bais P, Beebe K, Morelli KH, Currie ME, Norberg SN, Evsikov AV, Miers KE, Seburn KL, Guergueltcheva V, Kremensky I, Jordanova A, Bult CJ, Burgess RW (2016) Metabolite profile of a mouse model of Charcot-Marie-Tooth type 2D neuropathy: implications for disease mechanisms and interventions. Biol Open 5:908–920. https://doi.org/10.1242/bio.019273

Balayssac D, Ferrier J, Pereira B, Gillet B, Pétorin C, Vein J, Libert F, Eschalier A, Pezet D (2015) Prevention of oxaliplatin-induced peripheral neuropathy by a polyamine-reduced diet-NEUROXAPOL: protocol of a prospective, randomised, controlled, single-blind and monocentric trial. BMJ Open 5:e007479. https://doi.org/10.1136/bmjopen-2014-007479

Bandos H, Melnikow J, Rivera DR, Swain SM, Sturtz K, Fehrenbacher L, Wade JL 3rd, Brufsky AM, Julian TB, Margolese RG, McCarron EC, Ganz PA (2018) Long-term peripheral neuropathy in breast cancer patients treated with adjuvant chemotherapy: NRG oncology/NSABP B-30. J Natl Cancer Inst 110. https://doi.org/10.1093/jnci/djx162

Bao T, Basal C, Seluzicki C, Li SQ, Seidman AD, Mao JJ (2016) Long-term chemotherapy-induced peripheral neuropathy among breast cancer survivors: prevalence, risk factors, and fall risk. Breast Cancer Res Treat 159:327–333. https://doi.org/10.1007/s10549-016-3939-0

Benítez-Angeles M, Morales-Lázaro SL, Juárez-González E, Rosenbaum T (2020) TRPV1: structure, endogenous agonists, and mechanisms. Int J Mol Sci 21. https://doi.org/10.3390/ijms21103421

Bischoff A (1980) Morphology of diabetic neuropathy. Horm Metab Res Suppl 9:18–28

Bocca C, Le Paih V, Barca JMC, Nzoughet JK, Amati-Bonneau P, Blanchet O, Védié B, Géromin D, Simard G, Procaccio V, Bonneau D, Leaners G, Orssaud C, Reynier P (2021) A plasma metabolomic signature of Leber hereditary optic neuropathy showing taurine and nicotinamide deficiencies. Hum Mol Genet. https://doi.org/10.1093/hmg/ddab013

Bot M, Milaneschi Y, Al-Shehri T, Amin N, Garmaeva S, Onderwater GLJ, Pool R, Thesing CS, Vijfhuizen LS, Vogelzangs N, Arts ICW, Demirkan A, van Duijn C, van Greevenbroek M, van der Kallen CJH, Köhler S, Ligthart L, van den Maagdenberg A, Mook-Kanamori DO, de Mutsert R, Tiemeier H, Schram MT, Stehouwer CDA, Terwindt GM, Willems van Dijk K, Fu J, Zhernakova A, Beekman M, Slagboom PE, Boomsma DI, Penninx B (2020) Metabolomics profile in depression: a pooled analysis of 230 metabolic markers in 5283 cases with depression and 10,145 controls. Biol Psychiatry 87:409–418. https://doi.org/10.1016/j.biopsych.2019.08.016

Boxler MI, Liechti ME, Schmid Y, Kraemer T, Steuer AE (2017) First time view on human metabolome changes after a single intake of 3,4-methylenedioxymethamphetamine in healthy placebo-controlled subjects. J Proteome Res 16:3310–3320. https://doi.org/10.1021/acs.jproteome.7b00294

Bulls HW, Hoogland AI, Small BJ, Kennedy B, James BW, Arboleda BL, Shahzad MMK, Gonzalez BD, Jim HSL (2020) Lagged relationships among chemotherapy-induced peripheral neuropathy, sleep quality, and physical activity during and after chemotherapy. Ann Behav Med. https://doi.org/10.1093/abm/kaaa101

Bunner AE, Wells CL, Gonzales J, Agarwal U, Bayat E, Barnard ND (2015) A dietary intervention for chronic diabetic neuropathy pain: a randomized controlled pilot study. Nutr Diabetes 5:e158. https://doi.org/10.1038/nutd.2015.8

Cameron NE, Cotter MA (2008) Pro-inflammatory mechanisms in diabetic neuropathy: focus on the nuclear factor kappa B pathway. Curr Drug Targets 9:60–67. https://doi.org/10.2174/138945008783431718

Chan A, Hertz DL, Morales M, Adams EJ, Gordon S, Tan CJ, Staff NP, Kamath J, Oh J, Shinde S, Pon D, Dixit N, D'Olimpio J, Dumitrescu C, Gobbo M, Kober K, Mayo S, Pang L, Subbiah I, Beutler AS, Peters KB, Loprinzi C, Lustberg MB (2019) Biological predictors of chemotherapy-induced peripheral neuropathy (CIPN): MASCC neurological complications working group overview. Support Care Cancer 27:3729–3737. https://doi.org/10.1007/s00520-019-04987-8

Chao de la Barca JM, Simard G, Amati-Bonneau P, Safiedeen Z, Prunier-Mirebeau D, Chupin S, Gadras C, Tessier L, Gueguen N, Chevrollier A, Desquiret-Dumas V, Ferré M, Bris C, Kouassi Nzoughet J, Bocca C, Leruez S, Verny C, Miléa D, Bonneau D, Lenaers G, Martinez MC, Procaccio V, Reynier P (2016) The metabolomic signature of Leber's hereditary optic neuropathy reveals endoplasmic reticulum stress. Brain 139:2864–2876. https://doi.org/10.1093/brain/aww222

Cheah IK, Halliwell B (2012) Ergothioneine; antioxidant potential, physiological function and role in disease. Biochim Biophys Acta 1822:784–793. https://doi.org/10.1016/j.bbadis.2011.09.017
Chen L, Chen CS, Sun Y, Henry NL, Stringer KA, Hertz DL (2021) Feasibility of pharmacometabolomics to identify potential predictors of paclitaxel pharmacokinetic variability. Cancer Chemother Pharmacol. https://doi.org/10.1007/s00280-021-04300-7
Cheung W, Keski-Rahkonen P, Assi N, Ferrari P, Freisling H, Rinaldi S, Slimani N, Zamora-Ros R, Rundle M, Frost G, Gibbons H, Carr E, Brennan L, Cross AJ, Pala V, Panico S, Sacerdote C, Palli D, Tumino R, Kühn T, Kaaks R, Boeing H, Floegel A, Mancini F, Boutron-Ruault MC, Baglietto L, Trichopoulou A, Naska A, Orfanos P, Scalbert A (2017) A metabolomic study of biomarkers of meat and fish intake. Am J Clin Nutr 105:600–608. https://doi.org/10.3945/ajcn.116.146639
Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK (2009) Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc Natl Acad Sci U S A 106:14728–14733. https://doi.org/10.1073/pnas.0904489106
Cliff J, Jorgensen AL, Lord R, Azam F, Cossar L, Carr DF, Pirmohamed M (2017) The molecular genetics of chemotherapy-induced peripheral neuropathy: a systematic review and meta-analysis. Crit Rev Oncol Hematol 120:127–140. https://doi.org/10.1016/j.critrevonc.2017.09.009
Cox-Martin E, Trahan LH, Cox MG, Dougherty PM, Lai EA, Novy DM (2017) Disease burden and pain in obese cancer patients with chemotherapy-induced peripheral neuropathy. Support Care Cancer 25:1873–1879. https://doi.org/10.1007/s00520-017-3571-5
Cui M, Gosu V, Basith S, Hong S, Choi S (2016) Polymodal transient receptor potential vanilloid type 1 Nocisensor: structure, modulators, and therapeutic applications. Adv Protein Chem Struct Biol 104:81–125. https://doi.org/10.1016/bs.apcsb.2015.11.005
Cuparencu C, Rinnan Å, Dragsted LO (2019) Combined markers to assess meat intake-human metabolomic studies of discovery and validation. Mol Nutr Food Res 63:e1900106. https://doi.org/10.1002/mnfr.201900106
D'Andrea G, Gucciardi A, Perini F, Leon A (2019) Pathogenesis of cluster headache: from episodic to chronic form, the role of neurotransmitters and neuromodulators. Headache 59:1665–1670. https://doi.org/10.1111/head.13673
Dalangin R, Kim A, Campbell RE (2020) The role of amino acids in neurotransmission and fluorescent tools for their detection. Int J Mol Sci 21. https://doi.org/10.3390/ijms21176197
Darling PB, Grunow J, Rafii M, Brookes S, Ball RO, Pencharz PB (2000) Threonine dehydrogenase is a minor degradative pathway of threonine catabolism in adult humans. Am J Physiol Endocrinol Metab 278:E877–E884. https://doi.org/10.1152/ajpendo.2000.278.5.E877
Dewanjee S, Das S, Das AK, Bhattacharjee N, Dihingia A, Dua TK, Kalita J, Manna P (2018) Molecular mechanism of diabetic neuropathy and its pharmacotherapeutic targets. Eur J Pharmacol 833:472–523. https://doi.org/10.1016/j.ejphar.2018.06.034
Di Carlo G, Mascolo N, Izzo AA, Capasso F (1999) Flavonoids: old and new aspects of a class of natural therapeutic drugs. Life Sci 65:337–353. https://doi.org/10.1016/s0024-3205(99)00120-4
Dickinson PJ, Anderson PJ, Williams DC, Powell HC, Shelton GD, Morris JG, LeCouteur RA (2004) Assessment of the neurologic effects of dietary deficiencies of phenylalanine and tyrosine in cats. Am J Vet Res 65:671–680. https://doi.org/10.2460/ajvr.2004.65.671
Dragsted LO (2010) Biomarkers of meat intake and the application of nutrigenomics. Meat Sci 84:301–307. https://doi.org/10.1016/j.meatsci.2009.08.028
Durand JP, Deplanque G, Montheil V, Gornet JM, Scotte F, Mir O, Cessot A, Coriat R, Raymond E, Mitry E, Herait P, Yataghene Y, Goldwasser F (2012) Efficacy of venlafaxine for the prevention and relief of oxaliplatin-induced acute neurotoxicity: results of EFFOX, a randomized, double-blind, placebo-controlled phase III trial. Ann Oncol 23:200–205. https://doi.org/10.1093/annonc/mdr045

Duregon F, Vendramin B, Bullo V, Gobbo S, Cugusi L, Di Blasio A, Neunhaeuserer D, Zaccaria M, Bergamin M, Ermolao A (2018) Effects of exercise on cancer patients suffering chemotherapy-induced peripheral neuropathy undergoing treatment: a systematic review. Crit Rev Oncol Hematol 121:90–100. https://doi.org/10.1016/j.critrevonc.2017.11.002

Edgar AJ (2002) The human L-threonine 3-dehydrogenase gene is an expressed pseudogene. BMC Genet 3:18. https://doi.org/10.1186/1471-2156-3-18

El-Fatatry BM, Ibrahim OM, Hussien FZ, Mostafa TM (2018) Role of metformin in oxaliplatin-induced peripheral neuropathy in patients with stage III colorectal cancer: randomized, controlled study. Int J Colorectal Dis 33:1675–1683. https://doi.org/10.1007/s00384-018-3104-9

El Mouhayyar C, Riachy R, Khalil AB, Eid A, Azar S (2020) SGLT2 inhibitors, GLP-1 agonists, and DPP-4 inhibitors in diabetes and microvascular complications: a review. Int J Endocrinol 2020:1762164. https://doi.org/10.1155/2020/1762164

Emwas A-H, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, Raftery D, Alahmari F, Jaremko L, Jaremko M, Wishart DS (2019) NMR spectroscopy for metabolomics research. Metabolites 9:123. https://doi.org/10.3390/metabo9070123

Erdoğan MA, Taşkıran E, Yiğittürk G, Erbaş O, Taşkıran D (2020) The investigation of therapeutic potential of oxytocin and liraglutide on vincristine-induced neuropathy in rats. J Biochem Mol Toxicol 34:e22415. https://doi.org/10.1002/jbt.22415

Ertas M, Sagduyu A, Arac N, Uludag B, Ertekin C (1998) Use of levodopa to relieve pain from painful symmetrical diabetic polyneuropathy. Pain 75:257–259. https://doi.org/10.1016/s0304-3959(98)00003-7

Everett JR (2017) NMR-based pharmacometabonomics: a new paradigm for personalised or precision medicine. Prog Nucl Magn Reson Spectrosc 102-103:1–14. https://doi.org/10.1016/j.pnmrs.2017.04.003

Everett JR (2019) Pharmacometabonomics: the prediction of drug effects using metabolic profiling. Handb Exp Pharmacol 260:263–299. https://doi.org/10.1007/164_2019_316

Farshchian N, Alavi A, Heydarheydari S, Moradian N (2018) Comparative study of the effects of venlafaxine and duloxetine on chemotherapy-induced peripheral neuropathy. Cancer Chemother Pharmacol 82:787–793. https://doi.org/10.1007/s00280-018-3664-y

Feng RN, Niu YC, Sun XW, Li Q, Zhao C, Wang C, Guo FC, Sun CH, Li Y (2013) Histidine supplementation improves insulin resistance through suppressed inflammation in obese women with the metabolic syndrome: a randomised controlled trial. Diabetologia 56:985–994. https://doi.org/10.1007/s00125-013-2839-7

Ferrándiz ML, Alcaraz MJ (1991) Anti-inflammatory activity and inhibition of arachidonic acid metabolism by flavonoids. Agents Actions 32:283–288. https://doi.org/10.1007/bf01980887

Ferrier J, Bayet-Robert M, Dalmann R, El Guerrab A, Aissouni Y, Graveron-Demilly D, Chalus M, Pinguet J, Eschalier A, Richard D, Daulhac L, Marchand F, Balayssac D (2015) Cholinergic neurotransmission in the posterior insular cortex is altered in preclinical models of neuropathic pain: key role of muscarinic M2 receptors in donepezil-induced antinociception. J Neurosci 35:16418–16430. https://doi.org/10.1523/jneurosci.1537-15.2015

Ferrier J, Bayet-Robert M, Pereira B, Daulhac L, Eschalier A, Pezet D, Moulinoux JP, Balayssac D (2013) A polyamine-deficient diet prevents oxaliplatin-induced acute cold and mechanical hypersensitivity in rats. PLoS One 8:e77828. https://doi.org/10.1371/journal.pone.0077828

Filla LA, Edwards JL (2016) Metabolomics in diabetic complications. Mol Biosyst 12:1090–1105. https://doi.org/10.1039/c6mb00014b

Finsterer J, Wanschitz J, Quasthoff S, Iglseder S, Löscher W, Grisold W (2017) Causally treatable, hereditary neuropathies in Fabry's disease, transthyretin-related familial amyloidosis, and Pompe's disease. Acta Neurol Scand 136:558–569. https://doi.org/10.1111/ane.12758

Floyd JC Jr, Fajans SS, Conn JW, Knopf RF, Rull J (1966) Stimulation of insulin secretion by amino acids. J Clin Invest 45:1487–1502. https://doi.org/10.1172/JCI105456

Forbes JM, Cooper ME (2013) Mechanisms of diabetic complications. Physiol Rev 93:137–188. https://doi.org/10.1152/physrev.00045.2011

Fraser GE, Jaceldo-Siegl K, Henning SM, Fan J, Knutsen SF, Haddad EH, Sabaté J, Beeson WL, Bennett H (2016) Biomarkers of dietary intake are correlated with corresponding measures from repeated dietary recalls and food-frequency questionnaires in the adventist health study-2. J Nutr 146:586–594. https://doi.org/10.3945/jn.115.225508

Freeman OJ, Unwin RD, Dowsey AW, Begley P, Ali S, Hollywood KA, Rustogi N, Petersen RS, Dunn WB, Cooper GJ, Gardiner NJ (2016) Metabolic dysfunction is restricted to the sciatic nerve in experimental diabetic neuropathy. Diabetes 65:228–238. https://doi.org/10.2337/db15-0835

Frias B, Merighi A (2016) Capsaicin, nociception and pain. Molecules 21. https://doi.org/10.3390/molecules21060797

Fujita S, Hirota T, Sakiyama R, Baba M, Ieiri I (2019) Identification of drug transporters contributing to oxaliplatin-induced peripheral neuropathy. J Neurochem 148:373–385. https://doi.org/10.1111/jnc.14607

Fujita S, Ushio S, Ozawa N, Masuguchi K, Kawashiri T, Oishi R, Egashira N (2015) Exenatide facilitates recovery from oxaliplatin-induced peripheral neuropathy in rats. PLoS One 10: e0141921. https://doi.org/10.1371/journal.pone.0141921

Ganesh Yerra V, Negi G, Sharma SS, Kumar A (2013) Potential therapeutic effects of the simultaneous targeting of the Nrf2 and NF-κB pathways in diabetic neuropathy. Redox Biol 1:394–397. https://doi.org/10.1016/j.redox.2013.07.005

Ganz T, Wainstein J, Gilad S, Limor R, Boaz M, Stern N (2017) Serum asymmetric dimethylarginine and arginine levels predict microvascular and macrovascular complications in type 2 diabetes mellitus. Diabetes Metab Res Rev 33. https://doi.org/10.1002/dmrr.2836

Gao W, Zan Y, Wang ZJ, Hu XY, Huang F (2016) Quercetin ameliorates paclitaxel-induced neuropathic pain by stabilizing mast cells, and subsequently blocking PKCε-dependent activation of TRPV1. Acta Pharmacol Sin 37:1166–1177. https://doi.org/10.1038/aps.2016.58

Gonzalez-Quevedo A, Obregon F, Fernandez R, Santiesteban R, Serrano C, Lima L (2001) Amino acid levels and ratios in serum and cerebrospinal fluid of patients with optic neuropathy in Cuba. Nutr Neurosci 4:51–62. https://doi.org/10.1080/1028415x.2001.11747350

Gouin O, L'Herondelle K, Lebonvallet N, Le Gall-Ianotto C, Sakka M, Buhé V, Plée-Gautier E, Carré JL, Lefeuvre L, Misery L, Le Garrec R (2017) TRPV1 and TRPA1 in cutaneous neurogenic and chronic inflammation: pro-inflammatory response induced by their activation and their sensitization. Protein Cell 8:644–661. https://doi.org/10.1007/s13238-017-0395-5

Greenlee H, Hershman DL, Shi Z, Kwan ML, Ergas IJ, Roh JM, Kushi LH (2017) BMI, lifestyle factors and Taxane-induced neuropathy in breast cancer patients: the pathways study. J Natl Cancer Inst 109. https://doi.org/10.1093/jnci/djw206

Grim J, Ticha A, Hyspler R, Valis M, Zadak Z (2017) Selected risk nutritional factors for chemotherapy-induced polyneuropathy. Nutrients 9:535. https://doi.org/10.3390/nu9060535

Gründemann D (2012) The ergothioneine transporter controls and indicates ergothioneine activity--a review. Prev Med 54 Suppl: S71–S74. doi: https://doi.org/10.1016/j.ypmed.2011.12.001

Gu J, Lu H, Chen C, Gu Z, Hu M, Liu L, Yu J, Wei G, Huo J (2021) Diabetes mellitus as a risk factor for chemotherapy-induced peripheral neuropathy: a meta-analysis. Support Care Cancer. https://doi.org/10.1007/s00520-021-06321-7

Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J, Hu FB (2016) Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. Diabetes Care 39:833–846. https://doi.org/10.2337/dc15-2251

Guo K, Savelieff MG, Rumora AE, Alakwaa FM, Callaghan BC, Hur J, Feldman EL (2021) Plasma metabolomics and lipidomics differentiate obese individuals by peripheral neuropathy status. J Clin Endocrinol Metab. https://doi.org/10.1210/clinem/dgab844

Gupta A, Behl T, Sehgal A, Bhatia S, Jaglan D, Bungau S (2021) Therapeutic potential of Nrf-2 pathway in the treatment of diabetic neuropathy and nephropathy. Mol Biol Rep 48:2761–2774. https://doi.org/10.1007/s11033-021-06257-5

Haim N, Epelbaum R, Ben-Shahar M, Yarnitsky D, Simri W, Robinson E (1994) Full dose vincristine (without 2-mg dose limit) in the treatment of lymphomas. Cancer 73:2515–2519. https://doi.org/10.1002/1097-0142(19940515)73:10<2515::aid-cncr2820731011>3.0.co;2-g
Halliwell B, Cheah IK, Drum CL (2016) Ergothioneine, an adaptive antioxidant for the protection of injured tissues? A hypothesis. Biochem Biophys Res Commun 470:245–250. https://doi.org/10.1016/j.bbrc.2015.12.124
Harada S, Hirayama A, Chan Q, Kurihara A, Fukai K, Iida M, Kato S, Sugiyama D, Kuwabara K, Takeuchi A, Akiyama M, Okamura T, Ebbels TMD, Elliott P, Tomita M, Sato A, Suzuki C, Sugimoto M, Soga T, Takebayashi T (2018) Reliability of plasma polar metabolite concentrations in a large-scale cohort study using capillary electrophoresis-mass spectrometry. PLoS One 13:e0191230. https://doi.org/10.1371/journal.pone.0191230
Hardie DG, Ross FA, Hawley SA (2012) AMPK: a nutrient and energy sensor that maintains energy homeostasis. Nat Rev Mol Cell Biol 13:251–262. https://doi.org/10.1038/nrm3311
Hershman DL, Till C, Wright JD, Awad D, Ramsey SD, Barlow WE, Minasian LM, Unger J (2016) Comorbidities and risk of chemotherapy-induced peripheral neuropathy among participants 65 years or older in southwest oncology group clinical trials. J Clin Oncol 34:3014–3022. https://doi.org/10.1200/jco.2015.66.2346
Hershman DL, Unger JM, Crew KD, Till C, Greenlee H, Minasian LM, Moinpour CM, Lew DL, Fehrenbacher L, Wade JL 3rd, Wong SF, Fisch MJ, Lynn Henry N, Albain KS (2018) Two-year trends of taxane-induced neuropathy in women enrolled in a randomized trial of acetyl-L-carnitine (SWOG S0715). J Natl Cancer Inst 110:669–676. https://doi.org/10.1093/jnci/djx259
Hertz DL (2019) Concerns regarding use of patient-reported outcomes in biomarker studies of chemotherapy-induced peripheral neuropathy. Pharmacogenomics J 19:411–416. https://doi.org/10.1038/s41397-019-0093-1
Hertz DL, Childs DS, Park SB, Faithfull S, Ke Y, Ali NT, McGlown SM, Chan A, Grech L, Loprinzi CL (2021) Patient-centric decision framework for treatment alterations in patients with chemotherapy-induced peripheral neuropathy (CIPN). Cancer Treat Rev
Janes K, Little JW, Li C, Bryant L, Chen C, Chen Z, Kamocki K, Doyle T, Snider A, Esposito E, Cuzzocrea S, Bieberich E, Obeid L, Petrache I, Nicol G, Neumann WL, Salvemini D (2014) The development and maintenance of paclitaxel-induced neuropathic pain require activation of the sphingosine 1-phosphate receptor subtype 1. J Biol Chem 289:21082–21097. https://doi.org/10.1074/jbc.M114.569574
Jennaro TS, Fang F, Kidwell KM, Smith EML, Vangipuram K, Burness ML, Griggs JJ, Van Poznak C, Hayes DF, Henry NL, Hertz DL (2020) Vitamin D deficiency increases severity of paclitaxel-induced peripheral neuropathy. Breast Cancer Res Treat 180:707–714. https://doi.org/10.1007/s10549-020-05584-8
Jong NN, Nakanishi T, Liu JJ, Tamai I, McKeage MJ (2011) Oxaliplatin transport mediated by organic cation/carnitine transporters OCTN1 and OCTN2 in overexpressing human embryonic kidney 293 cells and rat dorsal root ganglion neurons. J Pharmacol Exp Ther 338:537–547. https://doi.org/10.1124/jpet.111.181297
Jude EB, Dang C, Boulton AJ (2010) Effect of L-arginine on the microcirculation in the neuropathic diabetic foot in type 2 diabetes mellitus: a double-blind, placebo-controlled study. Diabet Med 27:113–116. https://doi.org/10.1111/j.1464-5491.2009.02876.x
Kadiroglu AK, Sit D, Kayabasi H, Tuzcu AK, Tasdemir N, Yilmaz ME (2008) The effect of venlafaxine HCl on painful peripheral diabetic neuropathy in patients with type 2 diabetes mellitus. J Diabetes Complications 22:241–245. https://doi.org/10.1016/j.jdiacomp.2007.03.010
Kandula T, Park SB, Cohn RJ, Krishnan AV, Farrar MA (2016) Pediatric chemotherapy induced peripheral neuropathy: a systematic review of current knowledge. Cancer Treat Rev 50:118–128. https://doi.org/10.1016/j.ctrv.2016.09.005
Kantae V, Krekels EHJ, Esdonk MJV, Lindenburg P, Harms AC, Knibbe CAJ, Van der Graaf PH, Hankemeier T (2017) Integration of pharmacometabolomics with pharmacokinetics and

pharmacodynamics: towards personalized drug therapy. Metabolomics 13:9. https://doi.org/10.1007/s11306-016-1143-1
Kanzawa-Lee GA, Larson JL, Resnicow K, Smith EML (2020) Exercise effects on chemotherapy-induced peripheral neuropathy: a comprehensive integrative review. Cancer Nurs 43:E172–e185. https://doi.org/10.1097/ncc.0000000000000801
Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. Nat Rev Genet 19:299–310. https://doi.org/10.1038/nrg.2018.4
Kenkhuis MF, van der Linden BWA, Breedveld-Peters JJL, Koole JL, van Roekel EH, Breukink SO, Mols F, Weijenberg MP, Bours MJL (2021) Associations of the dietary World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) recommendations with patient-reported outcomes in colorectal cancer survivors 2-10 years post-diagnosis: a cross-sectional analysis. Br J Nutr 125:1188–1200. https://doi.org/10.1017/s0007114520003487
Kerckhove N, Collin A, Condé S, Chaleteix C, Pezet D, Balayssac D (2017) Long-term effects, pathophysiological mechanisms, and risk factors of chemotherapy-induced peripheral neuropathies: a comprehensive literature review. Front Pharmacol 8:86. https://doi.org/10.3389/fphar.2017.00086
Khodorova NV, Rutledge DN, Oberli M, Mathiron D, Marcelo P, Benamouzig R, Tomé D, Gaudichon C, Pilard S (2019) Urinary metabolomics profiles associated to bovine meat ingestion in humans. Mol Nutr Food Res 63:e1700834. https://doi.org/10.1002/mnfr.201700834
Kim S, Jang WJ, Yu H, Kim J, Lee SK, Jeong CH, Lee S (2020) Revealing metabolic perturbation following heavy methamphetamine abuse by human hair metabolomics and network analysis. Int J Mol Sci 21. https://doi.org/10.3390/ijms21176041
Kimata M, Shichijo M, Miura T, Serizawa I, Inagaki N, Nagai H (2000) Effects of luteolin, quercetin and baicalein on immunoglobulin E-mediated mediator release from human cultured mast cells. Clin Exp Allergy 30:501–508. https://doi.org/10.1046/j.1365-2222.2000.00768.x
Kirwan JA, Brennan L, Broadhurst D, Fiehn O, Cascante M, Dunn WB, Schmidt MA, Velagapudi V (2018) Preanalytical processing and biobanking procedures of biological samples for metabolomics research: a white paper, community perspective (for "Precision Medicine and Pharmacometabolomics Task Group" – the metabolomics society initiative). Clin Chem 64:1158–1182. https://doi.org/10.1373/clinchem.2018.287045
Kochlik B, Gerbracht C, Grune T, Weber D (2018) The influence of dietary habits and meat consumption on plasma 3-methylhistidine – a potential marker for muscle protein turnover. Mol Nutr Food Res 62:e1701062. https://doi.org/10.1002/mnfr.201701062
Koenig RJ, Peterson CM, Jones RL, Saudek C, Lehrman M, Cerami A (1976) Correlation of glucose regulation and hemoglobin AIc in diabetes mellitus. N Engl J Med 295:417–420. https://doi.org/10.1056/nejm197608192950804
Kolb NA, Smith AG, Singleton JR, Beck SL, Stoddard GJ, Brown S, Mooney K (2016) The association of chemotherapy-induced peripheral neuropathy symptoms and the risk of falling. JAMA Neurol 73:860–866. https://doi.org/10.1001/jamaneurol.2016.0383
Kölker S (2018) Metabolism of amino acid neurotransmitters: the synaptic disorder underlying inherited metabolic diseases. J Inherit Metab Dis 41:1055–1063. https://doi.org/10.1007/s10545-018-0201-4
Kramer R, Bielawski J, Kistner-Griffin E, Othman A, Alecu I, Ernst D, Kornhauser D, Hornemann T, Spassieva S (2015) Neurotoxic 1-deoxysphingolipids and paclitaxel-induced peripheral neuropathy. FASEB J 29:4461–4472. https://doi.org/10.1096/fj.15-272567
Kristoff CJ, Bwanali L, Veltri LM, Gautam GP, Rutto PK, Newton EO, Holland LA (2020) Challenging bioanalyses with capillary electrophoresis. Anal Chem 92:49–66. https://doi.org/10.1021/acs.analchem.9b04718
Kumar A, Mittal R (2017) Nrf2: a potential therapeutic target for diabetic neuropathy. Inflammopharmacology 25:393–402. https://doi.org/10.1007/s10787-017-0339-y
Lee JJ, Swain SM (2006) Peripheral neuropathy induced by microtubule-stabilizing agents. J Clin Oncol 24:1633–1642. https://doi.org/10.1200/jco.2005.04.0543

Li Y, Zhang H, Kosturakis AK, Cassidy RM, Zhang H, Kennamer-Chapman RM, Jawad AB, Colomand CM, Harrison DS, Dougherty PM (2015) MAPK signaling downstream to TLR4 contributes to paclitaxel-induced peripheral neuropathy. Brain Behav Immun 49:255–266. https://doi.org/10.1016/j.bbi.2015.06.003

Lin HT, Cheng ML, Lo CJ, Lin G, Lin SF, Yeh JT, Ho HY, Lin JR, Liu FC (2019) (1)H nuclear magnetic resonance (NMR)-based cerebrospinal fluid and plasma metabolomic analysis in type 2 diabetic patients and risk prediction for diabetic microangiopathy. J Clin Med 8. https://doi.org/10.3390/jcm8060874

Lloyd AJ, Favé G, Beckmann M, Lin W, Tailliart K, Xie L, Mathers JC, Draper J (2011) Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. Am J Clin Nutr 94:981–991. https://doi.org/10.3945/ajcn.111.017921

Loprinzi CL, Lacchetti C, Bleeker J, Cavaletti G, Chauhan C, Hertz DL, Kelley MR, Lavino A, Lustberg MB, Paice JA, Schneider BP, Lavoie Smith EM, Smith ML, Smith TJ, Wagner-Johnston N, Hershman DL (2020) Prevention and management of chemotherapy-induced peripheral neuropathy in survivors of adult cancers: ASCO guideline update. J Clin Oncol 38: 3325–3348. https://doi.org/10.1200/jco.20.01399

Ludolph AC, Ullrich K, Nedjat S, Masur H, Bick U (1992) Neurological outcome in 22 treated adolescents with hyperphenylalaninemia. A clinical and electrophysiological study. Acta Neurol Scand 85:243–248. https://doi.org/10.1111/j.1600-0404.1992.tb04039.x

Madhavi YV, Gaikwad N, Yerra VG, Kalvala AK, Nanduri S, Kumar A (2019) Targeting AMPK in diabetes and diabetic complications: energy homeostasis, autophagy and mitochondrial health. Curr Med Chem 26:5207–5229. https://doi.org/10.2174/0929867325666180406120051

Maekawa K, Ri M, Nakajima M, Sekine A, Ueda R, Tohkin M, Miyata N, Saito Y, Iida S (2019) Serum lipidomics for exploring biomarkers of bortezomib therapy in patients with multiple myeloma. Cancer Sci 110:3267–3274. https://doi.org/10.1111/cas.14178

Mao-Ying QL, Kavelaars A, Krukowski K, Huo XJ, Zhou W, Price TJ, Cleeland C, Heijnen CJ (2014) The anti-diabetic drug metformin protects against chemotherapy-induced peripheral neuropathy in a mouse model. PLoS One 9:e100701. https://doi.org/10.1371/journal.pone.0100701

Mathew AV, Jaiswal M, Ang L, Michailidis G, Pennathur S, Pop-Busui R (2019) Impaired amino acid and TCA metabolism and cardiovascular autonomic neuropathy progression in type 1 diabetes. Diabetes 68: 2035–2044. doi: https://doi.org/10.2337/db19-0145

McGale P, Taylor C, Correa C, Cutter D, Duane F, Ewertz M, Gray R, Mannu G, Peto R, Whelan T, Wang Y, Wang Z, Darby S (2014) Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. Lancet 383:2127–2135. https://doi.org/10.1016/s0140-6736(14)60488-8

Miao H, Xu J, Xu D, Ma X, Zhao X, Liu L (2019) Nociceptive behavior induced by chemotherapeutic paclitaxel and beneficial role of antioxidative pathways. Physiol Res 68:491–500. https://doi.org/10.33549/physiolres.933939

Miltenburg NC, Boogerd W (2014) Chemotherapy-induced neuropathy: a comprehensive survey. Cancer Treat Rev 40:872–882. https://doi.org/10.1016/j.ctrv.2014.04.004

Mitry P, Wawro N, Rohrmann S, Giesbertz P, Daniel H, Linseisen J (2019) Plasma concentrations of anserine, carnosine and pi-methylhistidine as biomarkers of habitual meat consumption. Eur J Clin Nutr 73:692–702. https://doi.org/10.1038/s41430-018-0248-1

Mols F, Beijers AJ, Vreugdenhil G, Verhulst A, Schep G, Husson O (2015) Chemotherapy-induced peripheral neuropathy, physical activity and health-related quality of life among colorectal cancer survivors from the PROFILES registry. J Cancer Surviv 9:512–522. https://doi.org/10.1007/s11764-015-0427-1

Mols F, Beijers T, Vreugdenhil G, van de Poll-Franse L (2014) Chemotherapy-induced peripheral neuropathy and its association with quality of life: a systematic review. Support Care Cancer 22: 2261–2269. https://doi.org/10.1007/s00520-014-2255-7

Mongiovi JM, Zirpoli GR, Cannioto R, Sucheston-Campbell LE, Hershman DL, Unger JM, Moore HCF, Stewart JA, Isaacs C, Hobday TJ, Salim M, Hortobagyi GN, Gralow JR, Thomas Budd G, Albain KS, Ambrosone CB, McCann SE (2018) Associations between self-reported diet during treatment and chemotherapy-induced peripheral neuropathy in a cooperative group trial (S0221). Breast Cancer Res 20:146. https://doi.org/10.1186/s13058-018-1077-9

Morvan D, Demidem A (2018) NMR metabolomics of fibroblasts with inherited mitochondrial complex I mutation reveals treatment-reversible lipid and amino acid metabolism alterations. Metabolomics 14:55. https://doi.org/10.1007/s11306-018-1345-9

Mustafa Ali M, Moeller M, Rybicki L, Moore HCF (2017) Long-term peripheral neuropathy symptoms in breast cancer survivors. Breast Cancer Res Treat 166:519–526. https://doi.org/10.1007/s10549-017-4437-8

Nagano H, Sanai H, Muraoka M, Takagi K (2012) Efficacy of lafutidine, a histamine H2-receptor antagonist, for taxane-induced peripheral neuropathy in patients with gynecological malignancies. Gynecol Oncol 127:172–174. https://doi.org/10.1016/j.ygyno.2012.06.029

Napoli E, Schneider A, Wang JY, Trivedi A, Carrillo NR, Tassone F, Rogawski M, Hagerman RJ, Giulivi C (2019) Allopregnanolone treatment improves plasma metabolomic profile associated with GABA metabolism in fragile X-associated tremor/ataxia syndrome: a pilot study. Mol Neurobiol 56:3702–3713. https://doi.org/10.1007/s12035-018-1330-3

Neuenschwander M, Ballon A, Weber KS, Norat T, Aune D, Schwingshackl L, Schlesinger S (2019) Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. BMJ 366:l2368. https://doi.org/10.1136/bmj.l2368

Ng T, Chan M, Khor CC, Ho HK, Chan A (2014) The genetic variants underlying breast cancer treatment-induced chronic and late toxicities: a systematic review. Cancer Treat Rev 40:1199–1214. https://doi.org/10.1016/j.ctrv.2014.10.001

Nishida K, Takeuchi K, Hosoda A, Sugano S, Morisaki E, Ohishi A, Nagasawa K (2018) Ergothioneine ameliorates oxaliplatin-induced peripheral neuropathy in rats. Life Sci 207: 516–524. https://doi.org/10.1016/j.lfs.2018.07.006

Nordlie RC, Foster JD, Lange AJ (1999) Regulation of glucose production by the liver. Annu Rev Nutr 19:379–406. https://doi.org/10.1146/annurev.nutr.19.1.379

O'Gorman A, Brennan L (2017) The role of metabolomics in determination of new dietary biomarkers. Proc Nutr Soc 76:295–302. https://doi.org/10.1017/s0029665116002974

Onodera S, Shibata M, Tanaka M, Inaba N, Arai Y, Aoyama M, Lee B, Yamaura T (1999) Gastroprotective mechanism of lafutidine, a novel anti-ulcer drug with histamine H2-receptor antagonistic activity. Arzneimittelforschung 49:519–526. https://doi.org/10.1055/s-0031-1300454

Onodera S, Shibata M, Tanaka M, Inaba N, Yamaura T, Ohnishi H (1995) Gastroprotective activity of FRG-8813, a novel histamine H2-receptor antagonist, in rats. Jpn J Pharmacol 68:161–173. https://doi.org/10.1254/jjp.68.161

Oza MJ, Laddha AP, Gaikwad AB, Mulay SR, Kulkarni YA (2021) Role of dietary modifications in the management of type 2 diabetic complications. Pharmacol Res 168:105602. https://doi.org/10.1016/j.phrs.2021.105602

Pallister T, Jennings A, Mohney RP, Yarand D, Mangino M, Cassidy A, MacGregor A, Spector TD, Menni C (2016) Characterizing blood metabolomics profiles associated with self-reported food intakes in female twins. PLoS One 11:e0158568. https://doi.org/10.1371/journal.pone.0158568

Park JE, Lim HR, Kim JW, Shin KH (2018) Metabolite changes in risk of type 2 diabetes mellitus in cohort studies: a systematic review and meta-analysis. Diabetes Res Clin Pract 140:216–227. https://doi.org/10.1016/j.diabres.2018.03.045

Paschen W (1992) Polyamine metabolism in different pathological states of the brain. Mol Chem Neuropathol 16:241–271. https://doi.org/10.1007/bf03159973

Pearce FL, Befus AD, Bienenstock J (1984) Mucosal mast cells. III. Effect of quercetin and other flavonoids on antigen-induced histamine secretion from rat intestinal mast cells. J Allergy Clin Immunol 73:819–823. https://doi.org/10.1016/0091-6749(84)90453-6

Peterson JW, Boldogh I, Popov VL, Saini SS, Chopra AK (1998) Anti-inflammatory and antisecretory potential of histidine in salmonella-challenged mouse small intestine. Lab Invest 78:523–534

Petrovchich I, Kober KM, Wagner L, Paul SM, Abrams G, Chesney MA, Topp K, Smoot B, Schumacher M, Conley YP, Hammer M, Levine JD, Miaskowski C (2019) Deleterious effects of higher body mass index on subjective and objective measures of chemotherapy-induced peripheral neuropathy in cancer survivors. J Pain Symptom Manage 58:252–263. https://doi.org/10.1016/j.jpainsymman.2019.04.029

Pini A, Obara I, Battell E, Chazot PL, Rosa AC (2016) Histamine in diabetes: is it time to reconsider? Pharmacol Res 111:316–324. https://doi.org/10.1016/j.phrs.2016.06.021

Plasmati R, Pastorelli F, Cavo M, Petracci E, Zamagni E, Tosi P, Cangini D, Tacchetti P, Salvi F, Bartolomei I, Michelucci R, Tassinari CA (2007) Neuropathy in multiple myeloma treated with thalidomide: a prospective study. Neurology 69:573–581. https://doi.org/10.1212/01.wnl.0000267271.18475.fe

Pop-Busui R, Lu J, Lopes N, Jones TL (2009) Prevalence of diabetic peripheral neuropathy and relation to glycemic control therapies at baseline in the BARI 2D cohort. J Peripher Nerv Syst 14:1–13. https://doi.org/10.1111/j.1529-8027.2009.00200.x

Powell HC, Rosoff J, Myers RR (1985) Microangiopathy in human diabetic neuropathy. Acta Neuropathol 68:295–305. https://doi.org/10.1007/bf00690832

Price TJ, Das V, Dussor G (2016) Adenosine monophosphate-activated protein kinase (AMPK) activators for the prevention, treatment and potential reversal of pathological pain. Curr Drug Targets 17:908–920. https://doi.org/10.2174/1389450116666151102095046

Qin XY, Akanuma H, Wei F, Nagano R, Zeng Q, Imanishi S, Ohsako S, Yoshinaga J, Yonemoto J, Tanokura M, Sone H (2012) Effect of low-dose thalidomide on dopaminergic neuronal differentiation of human neural progenitor cells: a combined study of metabolomics and morphological analysis. Neurotoxicology 33:1375–1380. https://doi.org/10.1016/j.neuro.2012.08.016

Quasthoff S, Hartung HP (2002) Chemotherapy-induced peripheral neuropathy. J Neurol 249:9–17. https://doi.org/10.1007/pl00007853

Raskin J, Smith TR, Wong K, Pritchett YL, D'Souza DN, Iyengar S, Wernicke JF (2006) Duloxetine versus routine care in the long-term management of diabetic peripheral neuropathic pain. J Palliat Med 9:29–40. https://doi.org/10.1089/jpm.2006.9.29

Rattray NJW, Deziel NC, Wallach JD, Khan SA, Vasiliou V, Ioannidis JPA, Johnson CH (2018) Beyond genomics: understanding exposotypes through metabolomics. Hum Genomics 12:4. https://doi.org/10.1186/s40246-018-0134-x

Razazian N, Baziyar M, Moradian N, Afshari D, Bostani A, Mahmoodi M (2014) Evaluation of the efficacy and safety of pregabalin, venlafaxine, and carbamazepine in patients with painful diabetic peripheral neuropathy. A randomized, double-blind trial. Neurosciences (Riyadh) 19: 192–198

Remmington T, Smith S (2021) Tyrosine supplementation for phenylketonuria. Cochrane Database Syst Rev 1:Cd001507. https://doi.org/10.1002/14651858.CD001507.pub4

Rivat C, Richebé P, Laboureyras E, Laulin JP, Havouis R, Noble F, Moulinoux JP, Simonnet G (2008) Polyamine deficient diet to relieve pain hypersensitivity. Pain 137:125–137. https://doi.org/10.1016/j.pain.2007.08.021

Roberts LD, Souza AL, Gerszten RE, Clish CB (2012) Targeted metabolomics. Curr Protoc Mol Biol. Chapter 30: Unit 30.2.1-24. https://doi.org/10.1002/0471142727.mb3002s98

Romanovsky AA, Almeida MC, Garami A, Steiner AA, Norman MH, Morrison SF, Nakamura K, Burmeister JJ, Nucci TB (2009) The transient receptor potential vanilloid-1 channel in thermoregulation: a thermosensor it is not. Pharmacol Rev 61:228–261. https://doi.org/10.1124/pr.109.001263

Rondón LJ, Farges MC, Davin N, Sion B, Privat AM, Vasson MP, Eschalier A, Courteix C (2018) L-arginine supplementation prevents allodynia and hyperalgesia in painful diabetic neuropathic rats by normalizing plasma nitric oxide concentration and increasing plasma agmatine concentration. Eur J Nutr 57:2353–2363. https://doi.org/10.1007/s00394-017-1508-x

Rosa AC, Fantozzi R (2013) The role of histamine in neurogenic inflammation. Br J Pharmacol 170:38–45. https://doi.org/10.1111/bph.12266

Rowbotham MC, Goli V, Kunz NR, Lei D (2004) Venlafaxine extended release in the treatment of painful diabetic neuropathy: a double-blind, placebo-controlled study. Pain 110:697–706. https://doi.org/10.1016/j.pain.2004.05.010

Roy Chowdhury SK, Smith DR, Saleh A, Schapansky J, Marquez A, Gomes S, Akude E, Morrow D, Calcutt NA, Fernyhough P (2012) Impaired adenosine monophosphate-activated protein kinase signalling in dorsal root ganglia neurons is linked to mitochondrial dysfunction and peripheral neuropathy in diabetes. Brain 135:1751–1766. https://doi.org/10.1093/brain/aws097

Schneider BP, Shen F, Jiang G, O'Neill A, Radovich M, Li L, Gardner L, Lai D, Foroud T, Sparano JA, Sledge GW Jr, Miller KD (2017) Impact of genetic ancestry on outcomes in ECOG-ACRIN-E5103. JCO Precis Oncol 2017. https://doi.org/10.1200/po.17.00059

Schwingshackl L, Hoffmann G, Lampousi AM, Knüppel S, Iqbal K, Schwedhelm C, Bechthold A, Schlesinger S, Boeing H (2017) Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. Eur J Epidemiol 32:363–375. https://doi.org/10.1007/s10654-017-0246-y

Seretny M, Currie GL, Sena ES, Ramnarine S, Grant R, MacLeod MR, Colvin LA, Fallon M (2014) Incidence, prevalence, and predictors of chemotherapy-induced peripheral neuropathy: a systematic review and meta-analysis. Pain 155:2461–2470. https://doi.org/10.1016/j.pain.2014.09.020

Shaw S, Uniyal A, Gadepalli A, Tiwari V, Belinskaia DA, Shestakova NN, Venugopala KN, Deb PK, Tiwari V (2020) Adenosine receptor signalling: probing the potential pathways for the ministration of neuropathic pain. Eur J Pharmacol 889:173619. https://doi.org/10.1016/j.ejphar.2020.173619

Shigematsu N, Kawashiri T, Kobayashi D, Shimizu S, Mine K, Hiromoto S, Uchida M, Egashira N, Shimazoe T (2020) Neuroprotective effect of alogliptin on oxaliplatin-induced peripheral neuropathy in vivo and in vitro. Sci Rep 10:6734. https://doi.org/10.1038/s41598-020-62738-w

Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, Arnold M, Erte I, Forgetta V, Yang TP, Walter K, Menni C, Chen L, Vasquez L, Valdes AM, Hyde CL, Wang V, Ziemek D, Roberts P, Xi L, Grundberg E, Waldenberger M, Richards JB, Mohney RP, Milburn MV, John SL, Trimmer J, Theis FJ, Overington JP, Suhre K, Brosnan MJ, Gieger C, Kastenmüller G, Spector TD, Soranzo N (2014) An atlas of genetic influences on human blood metabolites. Nat Genet 46:543–550. https://doi.org/10.1038/ng.2982

Shrikanth CB, Nandini CD (2020) AMPK in microvascular complications of diabetes and the beneficial effects of AMPK activators from plants. Phytomedicine 73:152808. https://doi.org/10.1016/j.phymed.2018.12.031

Sjölin J, Hjort G, Friman G, Hambraeus L (1987) Urinary excretion of 1-methylhistidine: a qualitative indicator of exogenous 3-methylhistidine and intake of meats from various sources. Metabolism 36:1175–1184. https://doi.org/10.1016/0026-0495(87)90245-9

Smith AG, Russell J, Feldman EL, Goldstein J, Peltier A, Smith S, Hamwi J, Pollari D, Bixby B, Howard J, Singleton JR (2006) Lifestyle intervention for pre-diabetic neuropathy. Diabetes Care 29:1294–1299. https://doi.org/10.2337/dc06-0224

Smith AG, Singleton JR (2013) Obesity and hyperlipidemia are risk factors for early diabetic neuropathy. J Diabetes Complications 27:436–442. https://doi.org/10.1016/j.jdiacomp.2013.04.003

Smith EM, Pang H, Cirrincione C, Fleishman S, Paskett ED, Ahles T, Bressler LR, Fadul CE, Knox C, Le-Lindqwister N, Gilman PB, Shapiro CL (2013) Effect of duloxetine on pain, function, and quality of life among patients with chemotherapy-induced painful peripheral neuropathy: a randomized clinical trial. JAMA 309:1359–1367. https://doi.org/10.1001/jama.2013.2813

Song TY, Chen CL, Liao JW, Ou HC, Tsai MS (2010) Ergothioneine protects against neuronal injury induced by cisplatin both in vitro and in vivo. Food Chem Toxicol 48:3492–3499. https://doi.org/10.1016/j.fct.2010.09.030

Stankovic JSK, Selakovic D, Mihailovic V, Rosic G (2020) Antioxidant supplementation in the treatment of neurotoxicity induced by platinum-based chemotherapeutics – a review. Int J Mol Sci 21. https://doi.org/10.3390/ijms21207753

Subramaniam S, O'Connor MJ, Masukawa LM, McGonigle P (1994) Polyamine effects on the NMDA receptor in human brain. Exp Neurol 130:323–330. https://doi.org/10.1006/exnr.1994.1210

Sucheston-Campbell LE, Clay-Gilmour AI, Barlow WE, Budd GT, Stram DO, Haiman CA, Sheng X, Yan L, Zirpoli G, Yao S, Jiang C, Owzar K, Hershman D, Albain KS, Hayes DF, Moore HC, Hobday TJ, Stewart JA, Rizvi A, Isaacs C, Salim M, Gralow JR, Hortobagyi GN, Livingston RB, Kroetz DL, Ambrosone CB (2018) Genome-wide meta-analyses identifies novel taxane-induced peripheral neuropathy-associated loci. Pharmacogenet Genomics 28:49–55. https://doi.org/10.1097/fpc.0000000000000318

Sun Y, Gao HY, Fan ZY, He Y, Yan YX (2020) Metabolomics signatures in type 2 diabetes: a systematic review and integrative analysis. J Clin Endocrinol Metab 105. https://doi.org/10.1210/clinem/dgz240

Sun Y, Kim JH, Vangipuram K, Hayes DF, Smith EML, Yeomans L, Henry NL, Stringer KA, Hertz DL (2018) Pharmacometabolomics reveals a role for histidine, phenylalanine, and threonine in the development of paclitaxel-induced peripheral neuropathy. Breast Cancer Res Treat 171:657–666. https://doi.org/10.1007/s10549-018-4862-3

Terrazzino S, Argyriou AA, Cargnin S, Antonacopoulou AG, Briani C, Bruna J, Velasco R, Alberti P, Campagnolo M, Lonardi S, Cortinovis D, Cazzaniga M, Santos C, Kalofonos HP, Canonico PL, Genazzani AA, Cavaletti G (2015) Genetic determinants of chronic oxaliplatin-induced peripheral neurotoxicity: a genome-wide study replication and meta-analysis. J Peripher Nerv Syst 20:15–23. https://doi.org/10.1111/jns.12110

Timmins HC, Mizrahi D, Li T, Kiernan MC, Goldstein D, Park SB (2021) Metabolic and lifestyle risk factors for chemotherapy-induced peripheral neuropathy in taxane and platinum-treated patients: a systematic review. J Cancer Surviv. https://doi.org/10.1007/s11764-021-00988-x

Traynelis SF, Hartley M, Heinemann SF (1995) Control of proton sensitivity of the NMDA receptor by RNA splicing and polyamines. Science 268:873–876. https://doi.org/10.1126/science.7754371

Tsukaguchi M, Shibano M, Matsuura A, Mukai S (2013) The protective effects of lafutidine for bortezomib induced peripheral neuropathy. J Blood Med 4:81–85. https://doi.org/10.2147/jbm.S44127

Umeda M, Fujita A, Nishiwaki H, Takeuchi K (1999) Effect of lafutidine, a novel histamine H2-receptor antagonist, on monochloramine-induced gastric lesions in rats: role of capsaicin-sensitive sensory neurons. J Gastroenterol Hepatol 14:859–865. https://doi.org/10.1046/j.1440-1746.1999.01976.x

Urpi-Sarda M, Almanza-Aguilera E, Llorach R, Vázquez-Fresno R, Estruch R, Corella D, Sorli JV, Carmona F, Sanchez-Pla A, Salas-Salvadó J, Andres-Lacueva C (2019) Non-targeted metabolomic biomarkers and metabotypes of type 2 diabetes: a cross-sectional study of PREDIMED trial participants. Diabetes Metab 45:167–174. https://doi.org/10.1016/j.diabet.2018.02.006

Vázquez-Fresno R, Llorach R, Urpi-Sarda M, Lupianez-Barbero A, Estruch R, Corella D, Fitó M, Arós F, Ruiz-Canela M, Salas-Salvadó J, Andres-Lacueva C (2015) Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: a 1- and 3-year follow-up in the PREDIMED study. J Proteome Res 14:531–540. https://doi.org/10.1021/pr5007894

Verma P, Devaraj J, Skiles JL, Sajdyk T, Ho RH, Hutchinson R, Wells E, Li L, Renbarger J, Cooper B, Ramkrishna D (2020) A metabolomics approach for early prediction of vincristine-induced peripheral neuropathy. Sci Rep 10:9659. https://doi.org/10.1038/s41598-020-66815-y

Wallace MS, Rowbotham MC, Katz NP, Dworkin RH, Dotson RM, Galer BS, Rauck RL, Backonja MM, Quessy SN, Meisner PD (2002) A randomized, double-blind, placebo-controlled trial of a glycine antagonist in neuropathic pain. Neurology 59:1694–1700. https://doi.org/10.1212/01.wnl.0000036273.98213.34

Wang J, Zhang XS, Tao R, Zhang J, Liu L, Jiang YH, Ma SH, Song LX, Xia LJ (2017) Upregulation of CX3CL1 mediated by NF-κB activation in dorsal root ganglion contributes to peripheral sensitization and chronic pain induced by oxaliplatin administration. Mol Pain 13: 1744806917726256. https://doi.org/10.1177/1744806917726256

Wang S, Kobayashi K, Kogure Y, Yamanaka H, Yamamoto S, Yagi H, Noguchi K, Dai Y (2018) Negative regulation of TRPA1 by AMPK in primary sensory neurons as a potential mechanism of painful diabetic neuropathy. Diabetes 67:98–109. https://doi.org/10.2337/db17-0503

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE (2011) Metabolite profiles and the risk of developing diabetes. Nat Med 17:448–453. https://doi.org/10.1038/nm.2307

Wei H, Viisanen H, You HJ, Pertovaara A (2016) Spinal histamine in attenuation of mechanical hypersensitivity in the spinal nerve ligation-induced model of experimental neuropathy. Eur J Pharmacol 772:1–10. https://doi.org/10.1016/j.ejphar.2015.12.039

Wiggin TD, Sullivan KA, Pop-Busui R, Amato A, Sima AA, Feldman EL (2009) Elevated triglycerides correlate with progression of diabetic neuropathy. Diabetes 58:1634–1640. https://doi.org/10.2337/db08-1771

Wilkins J, Sakrikar D, Petterson XM, Lanza IR, Trushina E (2019) A comprehensive protocol for multiplatform metabolomics analysis in patient-derived skin fibroblasts. Metabolomics 15:83. https://doi.org/10.1007/s11306-019-1544-z

Winkels RM, van Brakel L, van Baar H, Beelman RB, van Duijnhoven FJB, Geijsen A, van Halteren HK, Hansson BME, Richie JP, Sun D, Wesselink E, van Zutphen M, Kampman E, Kok DE (2020) Are ergothioneine levels in blood associated with chronic peripheral neuropathy in colorectal cancer patients who underwent chemotherapy? Nutr Cancer 72:451–459. https://doi.org/10.1080/01635581.2019.1637005

Winters-Stone KM, Horak F, Jacobs PG, Trubowitz P, Dieckmann NF, Stoyles S, Faithfull S (2017) Falls, functioning, and disability among women with persistent symptoms of chemotherapy-induced peripheral neuropathy. J Clin Oncol 35:2604–2612. https://doi.org/10.1200/jco.2016.71.3552

Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15:473–484. https://doi.org/10.1038/nrd.2016.32

Wishart DS (2019) Metabolomics for investigating physiological and pathophysiological processes. Physiol Rev 99:1819–1875. https://doi.org/10.1152/physrev.00035.2018

Wu FZ, Xu WJ, Deng B, Liu SD, Deng C, Wu MY, Gao Y, Jia LQ (2018) Wen-Luo-Tong decoction attenuates paclitaxel-induced peripheral neuropathy by regulating linoleic acid and glycerophospholipid metabolism pathways. Front Pharmacol 9:956. https://doi.org/10.3389/fphar.2018.00956

Xu X, Luo P, Wang Y, Cui Y, Miao L (2013) Nuclear factor (erythroid-derived 2)-like 2 (NFE2L2) is a novel therapeutic target for diabetic complications. J Int Med Res 41:13–19. https://doi.org/10.1177/0300060513477004

Yagihashi S, Matsunaga M (1979) Ultrastructural pathology of peripheral nerves in patients with diabetic neuropathy. Tohoku J Exp Med 129:357–366. https://doi.org/10.1620/tjem.129.357

Yamamoto H, Horie S, Uchida M, Tsuchiya S, Murayama T, Watanabe K (2001) Effects of vanilloid receptor agonists and antagonists on gastric antral ulcers in rats. Eur J Pharmacol 432:203–210. https://doi.org/10.1016/s0014-2999(01)01481-9

Yang Y, Luo L, Cai X, Fang Y, Wang J, Chen G, Yang J, Zhou Q, Sun X, Cheng X, Yan H, Lu W, Hu C, Cao P (2018) Nrf2 inhibits oxaliplatin-induced peripheral neuropathy via protection of mitochondrial function. Free Radic Biol Med 120:13–24. https://doi.org/10.1016/j.freeradbiomed.2018.03.007

Yehia R, Saleh S, El Abhar H, Saad AS, Schaalan M (2019) L-carnosine protects against oxaliplatin-induced peripheral neuropathy in colorectal cancer patients: a perspective on targeting Nrf-2 and NF-κB pathways. Toxicol Appl Pharmacol 365:41–50. https://doi.org/10.1016/j.taap.2018.12.015
Yerra VG, Areti A, Kumar A (2017) Adenosine monophosphate-activated protein kinase abates hyperglycaemia-induced neuronal injury in experimental models of diabetic neuropathy: effects on mitochondrial biogenesis, autophagy and neuroinflammation. Mol Neurobiol 54:2301–2312. https://doi.org/10.1007/s12035-016-9824-3
Yin X, Gibbons H, Rundle M, Frost G, McNulty BA, Nugent AP, Walton J, Flynn A, Gibney MJ, Brennan L (2017) Estimation of chicken intake by adults using metabolomics-derived markers. J Nutr 147:1850–1857. https://doi.org/10.3945/jn.117.252197
Yu J, Tang YY, Wang RR, Lou GD, Hu TT, Hou WW, Yue JX, Ohtsu H, Shi LY, Zhang SH, Chen Z (2016) A critical time window for the analgesic effect of central histamine in the partial sciatic ligation model of neuropathic pain. J Neuroinflammation 13:163. https://doi.org/10.1186/s12974-016-0637-0
Zajączkowska R, Kocot-Kępska M, Leppert W, Wrzosek A, Mika J, Wordliczek J (2019) Mechanisms of chemotherapy-induced peripheral neuropathy. Int J Mol Sci 20:1451. https://doi.org/10.3390/ijms20061451
Zhang W, Ramautar R (2021) CE-MS for metabolomics: developments and applications in the period 2018-2020. Electrophoresis 42:381–401. https://doi.org/10.1002/elps.202000203
Zhao XH, Wen ZM, Meredith CN, Matthews DE, Bier DM, Young VR (1986) Threonine kinetics at graded threonine intakes in young men. Am J Clin Nutr 43:795–802. https://doi.org/10.1093/ajcn/43.5.795
Zimmerman C, Atherton PJ, Pachman D, Seisler D, Wagner-Johnston N, Dakhil S, Lafky JM, Qin R, Grothey A, Loprinzi CL (2016) MC11C4: a pilot randomized, placebo-controlled, double-blind study of venlafaxine to prevent oxaliplatin-induced neuropathy. Support Care Cancer 24:1071–1078. https://doi.org/10.1007/s00520-015-2876-5
Zirpoli GR, McCann SE, Sucheston-Campbell LE, Hershman DL, Ciupak G, Davis W, Unger JM, Moore HCF, Stewart JA, Isaacs C, Hobday TJ, Salim M, Hortobagyi GN, Gralow JR, Budd GT, Albain KS, Ambrosone CB (2017) Supplement use and chemotherapy-induced peripheral neuropathy in a cooperative group trial (S0221): the DELCaP study. J Natl Cancer Inst 109. https://doi.org/10.1093/jnci/djx098

Metabolomics of Respiratory Diseases

Subhabrata Moitra, Arghya Bandyopadhyay, and Paige Lacy

Contents

1 Introduction 340
2 Respiratory Diseases with Metabolomic Signatures 343
2.1 Asthma 343
2.2 Chronic Obstructive Pulmonary Disease (COPD) 349
2.3 Pneumonia 352
2.4 Acute Lung Injury/Acute Respiratory Distress Syndrome (ARDS) 352
2.5 Occupational and Environmental Lung Diseases 355
3 Metabolomics of Lung Microbiome in Respiratory Diseases 357
3.1 Asthma 357
3.2 Chronic Obstructive Pulmonary Disease (COPD) 358
4 Conclusive Remarks 359
References 360

Abstract

Metabolomics is an expanding field of systems biology that is gaining significant attention in respiratory research. As a unique approach to understanding and diagnosing diseases, metabolomics provides a snapshot of all metabolites present in biological samples such as exhaled breath condensate, bronchoalveolar lavage, plasma, serum, urine, and other specimens that may be obtained from patients with respiratory diseases. In this article, we review the rapidly expanding field of metabolomics in its application to respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), pneumonia, and acute lung injury, along with its more severe form, adult respiratory disease syndrome. We also discuss

S. Moitra · A. Bandyopadhyay · P. Lacy (✉)
Department of Medicine, Alberta Respiratory Centre (ARC), University of Alberta, Edmonton, AB, Canada
e-mail: moitra@ualberta.ca; arghya1@ualberta.ca; placy@ualberta.ca

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_614

the potential applications of metabolomics for monitoring exposure to aerosolized occupational and environmental materials. With the latest advances in our understanding of the microbiome, we discuss microbiome-derived metabolites that arise from the gut and lung in asthma and COPD that have mechanistic implications for these diseases. Recent literature has suggested that metabolomics analysis using nuclear magnetic resonance (NMR) and mass spectrometry (MS) approaches may provide clinicians with the opportunity to identify new biomarkers that may predict progression to more severe diseases which may be fatal for many patients each year.

Keywords

Acute lung injury · Acute respiratory distress syndrome (ARDS) · Asthma · Chronic obstructive pulmonary disease (COPD) · Pneumonia

1 Introduction

Metabolomics is an established field of systems biology that has generated substantial new findings in respiratory research. The ability of metabolomics to produce a "snapshot" of small molecules within a given sample from the body provides a powerful tool for temporal analyses to follow the distribution and concentration of these molecules (Patti et al. 2012). Small molecules of interest include chemicals (such as drugs) and metabolites (including waste products of metabolism). These small molecules are a distinct group of compounds from the larger proteins and nucleic acids (RNA, DNA), and their measurement provides a valuable complement to other fields of systems biology (transcriptomics, genomics, proteomics, and others). Further, metabolomics informs other areas of systems biology as it lies downstream of proteins, RNA, and genes. Because of its ability to detect small molecules, metabolomics has the potential to discover novel biomarkers of disease as well as environmental and occupational exposure (Madsen et al. 2010; Robertson et al. 2011). To understand the relevance of metabolomics in respiratory diseases, it is important to establish how the metabolome is defined and how this aligns with other approaches in systems biology.

The metabolome of an organism reflects events that occur in the proteome, transcriptome, and genome. Changes in proteins, RNA, and genes result in alterations of metabolite concentrations in biological fluids and tissues. Perhaps unsurprisingly, measurement of metabolites in human samples is not a new procedure, since metabolites have been used for millennia to aid in the diagnosis of disease. For example, diabetes mellitus has been diagnosed since ancient times based on the taste of glucose, a small molecule metabolite, in urine from patients with type I diabetes. While this may be an unappetizing and perhaps unsafe practice these days, this was essentially the first diagnostic test for a metabolite in urine samples. The recognition that urine contains important biomarkers of disease led to the development of analytical tools to measure these in a variety of samples from the

body today. Today, the measurement of small molecules in human samples forms the basis of clinical chemistry, established to assist health professionals in diagnosis of illnesses.

Most clinical chemistry tests rely on the measurement of handful of metabolites, and often these are only qualitative (positive or negative) rather than quantitative tests. Because most metabolites measured using clinical tests are abundant and not specific to any one disease, their detection must always be taken into consideration with other clinical descriptors. Thus, the focus of clinical chemistry on such a small group of metabolites is a significant limitation that prevents the applicability of metabolite detection in the specific diagnosis of many diseases.

The limitations of traditional clinical chemistry highlight the advantages of metabolomics. Recent improvements in the sensitivity and specificity of metabolite detection using metabolomics have allowed the characterization and quantification of complex metabolic profiles resulting in concurrent analysis of hundreds of metabolites in a single sample. Metabolomics seeks to quantitatively assess complex metabolic patterns in patient samples and is coupled with computational technologies to allow the interpretation of data in the context of known metabolic pathways. The complexity of the metabolome in a patient sample is further augmented by the presence of metabolites that derive from the microbiome, which is present in almost all samples obtained from the human body. The microbiome generates metabolites that are unique to prokaryotic organisms and may be distinguished from the host's own metabolome, thus providing another possible approach for enhancing the diagnosis and prognosis of disease.

Despite substantial investments in genome analysis in diseases, genetic mutations that result in the manifestation of disease are rare. Only 1–2% of disease risk for a spectrum of conditions including asthma, chronic obstructive pulmonary disease (COPD), and acute respiratory distress syndrome (ARDS) can be explained by genetic mutations. Transcriptomic and proteomic analysis has generated more insight into their potential as biomarkers, but these too have not developed into standard disease indicators. In contrast, metabolomics and clinical chemistry reproducibly demonstrate that metabolites are highly predictive for a large proportion of complex diseases (Xia et al. 2013). Samples may be used from a broad range of sources including saliva, nasal lavage, exhaled breath condensate (EBC), bronchial washings, sweat, blood (plasma and serum), urine, feces, among others. Examples of established metabolic biomarkers include glucose for diabetes, as mentioned above, creatinine to detect kidney disease, cholesterol and triglycerides to evaluate the risk of developing cardiovascular disease, uric acid for gout detection, and thyroxine for hypo/hyperthyroidism. There are undoubtedly other metabolites that may be used to serve as biomarkers in a range of diseases.

These findings indicate that the metabolome is a much more dynamic group of analytes than the proteome, transcriptome, or genome, as it can change immediately in response to environmental or physiological changes (Fig. 1) (Wishart 2005). To appreciate the contribution that metabolomics may make to diagnosis of disease, it is useful to compare the impact of environmental and physiological impact on proteins, RNA, and genes. Environmental and physiological changes have negligible impact

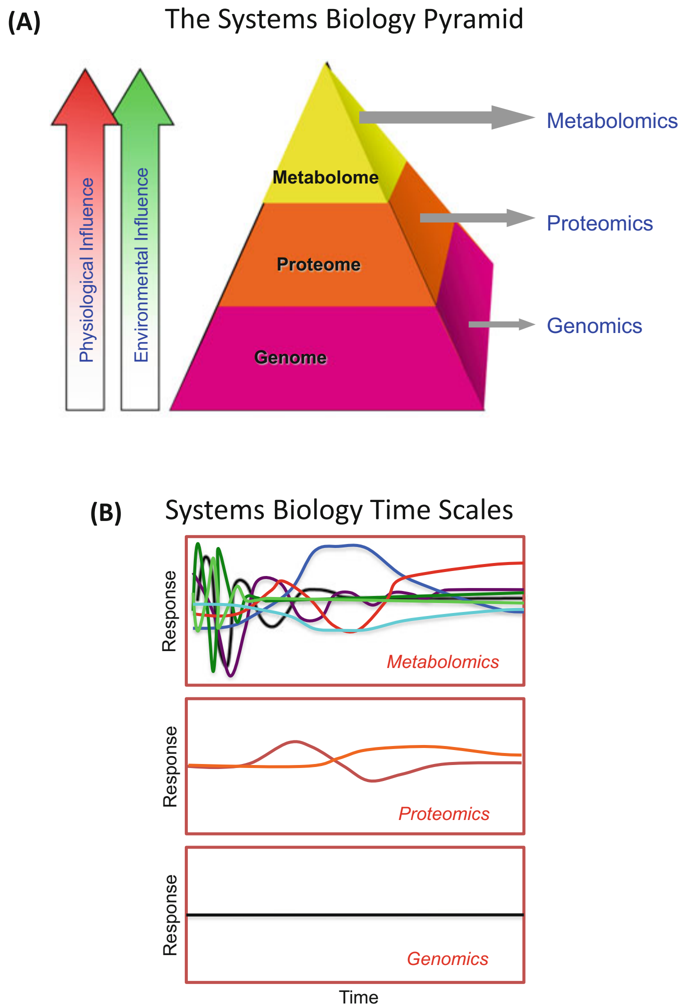


Fig. 1 The systems biology pyramid and time scales of responses to environmental influences. (**a**) Over 25,000 genes have been identified in human genomics, compared with a smaller number of enzymes and even smaller number of metabolites. The responsiveness to physiological and environmental insults of each of these components increases as we go from genomics to metabolomics. (**b**) While metabolomics shows rapid changes in multiple metabolites in a short period of time, proteomics shows smaller changes in abundance while genomics shows negligible changes over the same period

on somatic gene expression, while some transcriptomic and proteomic changes have been detected. In contrast, metabolomic changes in response to environmental and physiological factors closely correlate with these events and can be altered within seconds of exposure (Fig. 1b). Therefore, significant changes in metabolites may be measurable in samples over far shorter time scales than by other systems biology

approaches. This allows for a powerful approach for detection of changes in biomarkers in real time and provides an opportunity to use metabolomics as a biomonitoring tool in health and disease. Historically we have adhered to the concept of a single biomarker for each disease, but this limits the accuracy, precision, and sensitivity of the assay. New and developing metabolomics approaches suggest that we may use a pattern of metabolites to describe a given disease. However, by using multiple biomarkers for each disease, techniques become more sophisticated, and the computing power used to analyze the data becomes much more complicated.

In this article, we review the expanding field of metabolomics in its application to respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), pneumonia, acute lung injury/acute respiratory distress syndrome, and occupational and environmental lung diseases. We also discuss the metabolomics associated with the lung microbiome in asthma and COPD. These findings show that there remains a considerable amount of experimental work to be done to understand the role of the metabolome in respiratory diseases, and how this may be applied to the diagnosis and/or prognosis of illness. Recent findings have shone some light onto the relationship between the gut and lung microbiome metabolites in generating metabolic signatures that may provide mechanistic insights into various lung diseases, as well as deliver potential biomarkers associated with specific lung conditions.

2 Respiratory Diseases with Metabolomic Signatures

2.1 Asthma

Asthma is an inflammatory disease of the airways that is often triggered by exogenous perturbations. The recent Global Burden of Disease report stated that an estimated 262 million people were affected by asthma in 2019, and 461,000 people died from this disease in that year (Vos et al. 2020). Asthma is a highly heterogeneous disease with different phenotypic variations, as well as multiple causative agents, etiology, and complex inflammatory and pathophysiological features. Thus, it is proposed that significant metabolic changes are associated with different phenotypes of disease.

Metabolic profiling has demonstrated significant variations in serological and urinary metabolomic pathways that are distinct in various phenotypes of asthma and provides valuable information about the accuracy and precision of asthma diagnosis, disease progression, and response to treatment (Fig. 2) (Kelly, “Pharmacometabolomics of Asthma as a road map to Precision Medicine”). Using technologies such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS), several metabolomic studies have demonstrated comprehensive evidence of metabolic alterations in asthma. In Table 1, we have summarized some of the recent reports of metabolomic studies in asthma.

Studies conducted on asthma patients with varying degrees of disease severity, ages, or obesity have been reported. Interestingly, all these studies showed a high

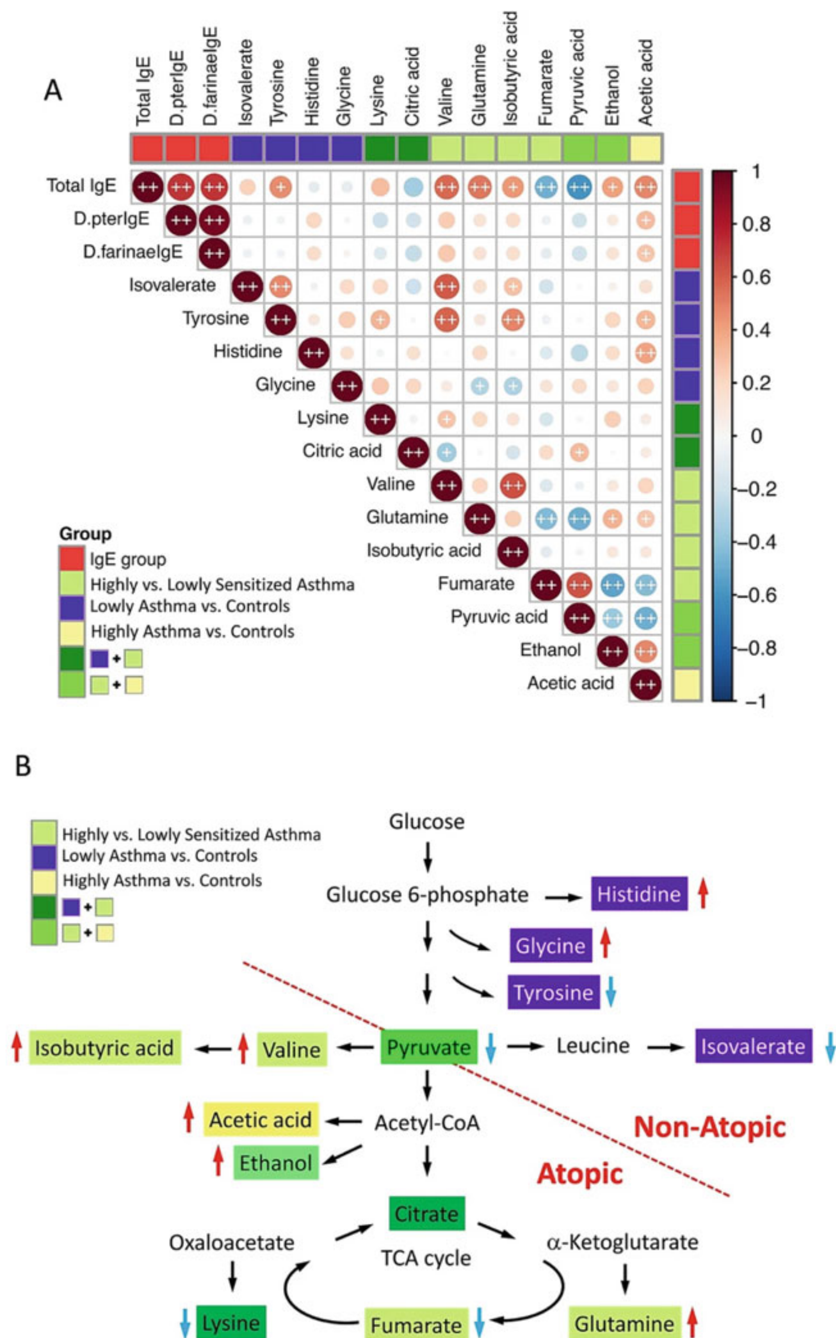


Fig. 2 Metabolites and metabolic pathways in childhood asthma. Heatmap of Spearman's rank correlation coefficients between metabolites associated with lowly and highly sensitized asthma (**a**) and metabolic pathways of metabolites associated with atopic and non-atopic asthma (**b**). Red color

correlation among specific metabolites with disease. Park et al. (2017) showed that in severe asthma in children (≤15 years of age) exhibiting corticosteroid resistance, tyrosine metabolism, degradation of aromatic compounds, and glutathione metabolism were suggested to be significant pathways related to corticosteroid resistance based on urine metabolites. A targeted LC-MS-based study for the presence of a unique biomarker in asthmatic children showed a combination of 2-isopropylmalic acid and betaine can classify children with asthma and controls. It was also shown in the same study that asthmatics had lower relative concentrations of serum ascorbic acid, 2-isopropylmalic acid, shikimate-3-phosphate, 6-phospho-D-gluconate, and reduced glutathione. In the case of overweight children, niacin concentrations were elevated in serum samples (Checkley et al. 2016). Loureiro et al. showed that lipid peroxidation-related metabolites in urine samples are associated with asthma severity and lung function, along with eosinophilic inflammation in nonobese asthmatic patients (Loureiro et al. 2016). In other studies, it was shown that metabolic pathways and pathway components like arginine, proline, taurine, hypotaurine, glyoxylate, and dicarboxylate in serum and urine samples were closely related to acute exacerbations of asthma as well as the choice of corticosteroid treatment (Quan-Jun et al. 2017). It was also found that a set of 15 volatile carbon compounds may discriminate between controlled and uncontrolled asthma and that 7 of these compounds detected in exhaled breath samples could predict exacerbation within the next 14 days with 88% sensitivity and 75% specificity (van Vliet et al. 2017).

Lipids have also been correlated with the diagnosis of asthma. Kang et al. showed that certain metabolites, primarily lipid biomolecules in bronchoalveolar lavage (BAL) fluid, could be markedly elevated in asthma compared to non-asthmatic healthy individuals (Kang et al. 2014). This observation, supported by other studies (Loureiro et al. 2016; Ghosh et al. 2020), indicates that lipid metabolism is altered in asthma, potentially as a result of increased oxidative stress. Such altered lipid metabolism was also associated with asthma severity, reduced lung function, and higher eosinophilic inflammation in asthmatic individuals (Loureiro et al. 2016). While asthma and obesity are known to share common systemic manifestations, Maniscalco et al. showed that methane, pyruvate, and glyoxylate and dicarboxylate metabolic pathways in EBC also greatly vary between obese and nonobese asthma patients (Maniscalco et al. 2017), which indicates more complex crosstalk between asthma and obesity than previously recognized.

Other recent studies have reported intriguing results of altered profiles of structural lipid molecules in asthma compared to healthy individuals (Kang et al. 2014; Ghosh et al. 2020; Reinke et al. 2017; Pang et al. 2018; Jiang et al. 2021). Bian et al. reported that some derivatives of serum arachidonic acid that serve as potential mediators for allergic responses were significantly elevated in asthma (Bian et al.

Fig. 2 (continued) represents positive correlations; blue color represents negative correlations; red arrow represents increase; blue arrow represents decrease. + symbol means a P-value < 0.05; ++ symbol means a P-value < 0.01. [Reproduced from Chiu et al. (2021)]

Table 1 Metabolomics studies – asthma

Study	Study population (adult/children)	Sample/method	Summary of results
Kang et al. (2014)	Adults (38 asthmatics and 13 healthy)	Quadrupole time-of-flight (QTOF) MS of bronchoalveolar lavage fluid (BALF)	• ↑ Lysophosphatidylcholine (LPC), triglyceride (TG), phosphatidylcholine (PC), phosphatidylglycerol (PG), phosphatidylserine (PS), and sphingomyelin (SM) in non-steroidal bronchial asthma NSBA) compared to healthy subjects • No difference was observed between steroid-treated bronchial asthma (SBA) and healthy subjects
Loureiro et al. (2016)	Adults (57 asthmatics)	Targeted solid phase microextraction (SPME) with two-dimensional gas chromatography and time-of-flight MS (GCxGC-TOF/MS) of urine	Metabolites related to lipid peroxidation were associated with ↑ asthma severity, ↓ lung function, and ↑ eosinophilic inflammation in nonobese patients with asthma
Ghosh et al. (2020)	Adults (34 asthmatics, 30 COPD, 35 ACO, 33 healthy)	GC-MS of serum	↑ 2-Palmitoylglycerol, cholesterol, serine, threonine, ethanolamine, glucose, stearic acid, linoleic acid, D-mannose, succinic acid in asthma than healthy ↓ Lactic acid, 2-palmitoylglycerol in asthma than healthy
Maniscalco et al. (2017)	Adults (25 obese patients with asthma, 30 obese patients without asthma, 30 lean patients with asthma and 72 adults in the external validation set)	Untargeted LC-MS of EBC	Participants with asthma, obesity, and obesity + asthma showed distinct variations in respiratory metabolic fingerprint
Reinke et al. (2017)	Adults (54 asthmatics, 22 healthy)	Untargeted LC-MS of serum	↑ Ceramide (C16:0, C18:0, C20:0, C22:0, C24:0, C24:1), sphingomyelin (C18:0, C18:1), hexosylceramide (C18:0, C24:1), and cysteinyl leukotriene E_4 (LTE_4) in asthma than healthy ↓ 14,15-Dihydroxyeicosatetraenoic acid (DiHETE), 19,20-Dihydroxydocosapentaenoic acid (DiHDPA) in asthmatics than healthy

(continued)

Table 1 (continued)

Study	Study population (adult/children)	Sample/method	Summary of results
Pang et al. (2018)	Adults (29 asthmatics, 15 healthy)	Ultra-performance liquid chromatography-tandem MS (UPLC-MS) of serum	↑ Monosaccharides, LysoPC(o-18:0, 18:1), Retinyl ester, PC(18:1/2:0), PC(16:0/18:1), arachidonic acid, PE(18:3/14:0) in asthma ↓ Glycerophosphocholine, PS(18:0/22:5), cholesterol glucuronide, Phytosphingosine, Sphinganine, LysoPC(p-18:1), retinols, PC(20:4/16:1)
Jiang et al. (2021)	Adults (33 asthmatics, 28 healthy)	LC-MS/MS of plasma	↑ Phosphatidylethanolamine (PE) (18:1p/22:6), PE (20:0/18:1), PE (38:1), sphingomyelin (SM) (d18:1/18:1), triglyceride (TG) (16:0/16:0/18:1) in asthmatics than healthy ↓ Phosphatidylinositol (PI) (16:0/20:4), TG (17:0/18:1/18:1), phosphatidylglycerol (PG) (44:0), ceramide (d16:0/27:2), lysoPC (22:4) in asthma
Chiu et al. (2020, 2018)	Adults (30 asthmatics, 30 healthy)	NMR of urine	↑ Guanidoacetate ↓ 1-Methylnicotinamide, allantoin
Chiu et al. (2021)	Children (28 asthmatics, 25 healthy)	NMR of plasma	↑ Lysine, isovalerate, histidine, tyrosine, glycine, citric acid, ethanol, acetic acid, pyruvic acid in asthma
Chang-Chien et al. (2021)	Adults (92 asthmatics, 73 healthy)	NMR of EBC	↑ Lactate, formate, butyric acid, isobutyrate in asthma
Bian et al. (2017)	Adults (15 asthmatics and 15 healthy)	Ultra-high performance liquid chromatography quadrupole time-of-flight (UHPLC)-Q-TOF- MS of serum	↑ 5(S)-Hydroxyeicosatetraenoic acid (HETE), 8(S)-HETE, 11(S)-HETE, 12(S)-HETE, 15(S)-HETE, 15(S)-Hydroxyeicosapentaenoic acid (HEPE), prostaglandin (PG)A2, PGB2, PGF1a, PGF2a, PGJ2, 15-keto-PGF2a in asthma compared to healthy ↓ Palmitic acid, Lauric acid in asthma than healthy
Checkley et al. (2016)	Children (50 asthmatics and 49 healthy between 9 and 19 years)	Targeted liquid chromatography-MS (LC-MS) of serum	↓ Relative concentrations of serum ascorbic acid, 2-isopropylmalic acid, shikimate-3-phosphate, 6-phospho-D-gluconate, and reduced glutathione in asthmatics than healthy

(continued)

Table 1 (continued)

Study	Study population (adult/children)	Sample/method	Summary of results
Kelly et al. (2017)	Children (380 asthmatics)	Targeted LC-MS (complementary methods) of plasma	• Metabolites (primarily glycerophospholipid, linoleic acid, and pyrimidine) were associated with airway hyperreactivity, and pre- and postbronchodilator FEV_1/FVC • Distinct metabolites showed moderate but important signatures between disease severity
Tao et al. (2019)	Children (80 asthmatics, 29 healthy)	GC-MS of urine	↑ Aspartic acid, Xanthosine, hypoxanthine, N-acetylgalactosamine ↓ Stearic acid, Heptadecanoic acid, uric acid, D-threitol
Li et al. (2020)	Children (30 asthmatics, 30 healthy)	GC-MS of urine	↑ Azelaic acid, citraconic acid 4, D-altrose 1, D-erythro-sphingosine 1, gentiobiose 2, 2-hydroxybutanoic acid, L-allothreonine 1, leucine, stearic acid, succinic acid, tyramine in asthmatics than healthy ↓ 3,4-dihydroxycinnamic acid, methionine 1, purine riboside, malonic acid 1, cysteine, erythrose 1, lactamide 1, uric acid, valine in asthma
Matysiak et al. (2020)	Children (13 asthmatics, 17 healthy)	LC-MS/MS of blood	↑ L-arginine, B-alanine, Ɣ-amino-N-butyric acid, L-histidine, Hydroxy-L-proline in asthma ↓ D,L-B-Aminoisobutyric acid, taurine, L-tryptophan, L-valine in asthma
Ferraro et al. (2020)	Children (26 asthmatics, 16 healthy)	UPLC-MS of EBC	↑ 9-amino-nonanoic acid, 12-amino-dodecanoic acid, lactone of PGF-MUM, N-linoleoyl taurine, 17-phenoxy trinor PGF2α ethyl amide, lysoPC (18:2(9Z,12Z)) in asthma
Van Vliet et al. (2017, 2016)	Children (96 asthmatics)	Targeted GC-TOF/MS for VOCs in EBC	• 7 VOCs (3 aldehydes, 1 hydrocarbon, 1 ketone, 1 aromatic compound, and 1 unidentified VOC) in exhaled breath could predict asthma exacerbations

2017). They described that some saturated fatty acids such as palmitic acid and lauric acid levels were decreased in asthma. In addition, metabolites derived from protein or carbohydrate metabolisms were found altered in asthma compared to non-asthmatic healthy individuals in EBC, plasma, and urine (Chang-Chien et al. 2021; Chiu et al. 2018, 2020). However, in adults, the asthmatic response can be caused, triggered, or aggravated by different risk factors such as allergy, environmental exposures, active or passive smoking, and workplace conditions. Therefore, more studies on different adult asthma phenotypes are required to better understand those metabolic alterations.

On the other hand, asthma in children is mostly caused by allergic conditions or genetic predisposition (such as parental atopy or asthma) and, to some extent, pregnancy-related issues such as gestational smoking. In a group of asthmatic and non-asthmatic children, Checkley et al. showed lowered relative concentrations of serum ascorbic acid, reduced glutathione (GSH), and some carbohydrate derivatives in asthma (Checkley et al. 2016). Kelly et al. further showed association between certain plasma metabolites (glycerophospholipid, linoleic acid, and pyrimidine) and airway hyperreactivity in asthmatic children (Kelly et al. 2017). They were able to demonstrate moderate but clinically important signatures of distinct metabolites in accordance with the disease severity (Kelly et al. 2017). Several other reports have demonstrated distinct metabolomic profiles in asthmatic children compared to non-asthmatic healthy individuals (Tao et al. 2019; Li et al. 2020; Matysiak et al. 2020; Chiu et al. 2021).

Recently, breath analysis has suggested some intriguing metabolic alterations in asthma, particularly related to volatile organic compounds (VOCs) in the EBC that could predict asthma exacerbations in children (van Vliet et al. 2017; Ferraro et al. 2020; Van Vliet et al. 2016). However, most of those analyses did not consider potential risk factors or confounding factors as mentioned earlier. Therefore, clinical correlations between metabolites and symptoms/severity are important to consider while inferring those results into clinical practice.

2.2 Chronic Obstructive Pulmonary Disease (COPD)

COPD is a major lung disease worldwide that causes significant morbidity and mortality and is among the top causes of death in many countries (Keogh and Mark 2021). COPD is a chronic inflammatory disease of the lungs that is progressive and irreversible in nature (Devine 2008). Although cigarette smoking is the most common major risk factor for COPD, occupational or environmental insults are also known to be prominent triggers for the onset and progression of this debilitating lung condition.

Recent studies have demonstrated that COPD is a variable condition with multimodal phenotypic variants, particularly because of the differences in causal agents, course, and progression of the disease. Although many metabolic alterations of COPD were unknown until the beginning of the twenty-first century, these have

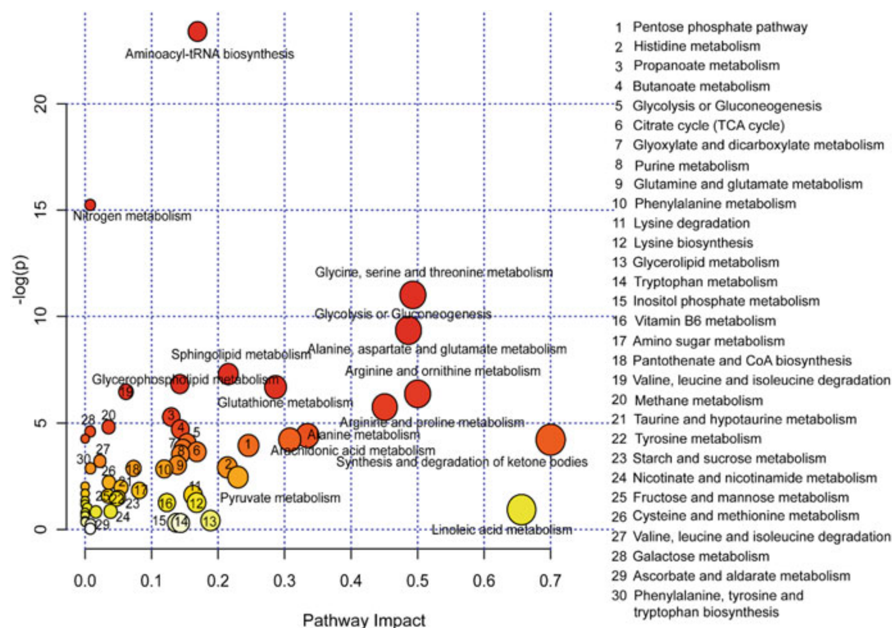


Fig. 3 Distinct metabolites identified in COPD-associated metabolomics studies. Metabolic pathways analysis based on distinct metabolites published in chronic obstructive pulmonary disease (COPD)-associated metabolomics studies performed by applying the Metabo-Analyst 4.0 platform. The names of 44 disturbed metabolic pathways were marked in the pathway figure, which mainly involved dysfunctions of amino acid metabolism, lipid metabolism, energy production pathways, and imbalance of oxidation and antioxidation. [Reproduced from Ran et al. (2019)]

so far exhibited an intriguing panorama based on what has been discovered to date (Fig. 3). In this section, we briefly describe some of the important findings from metabolomic research in COPD.

Several studies have identified metabolites that are distinctive in COPD (Turano, “NMR-based metabolomics to evaluate individual response to treatments”). Novotna et al. (2018) examined 10 COPD patients and 10 healthy individuals and observed that two amino acids, alanine and phenylalanine, were significantly lower in the peripheral blood of COPD patients than healthy individuals, while pyroglutamate level was higher in COPD patients. They also observed that the free carnitine to acylcarnitine ratio was significantly lower in COPD patients than the healthy individuals. Another report by Diao et al. (2019) further demonstrated that COPD patients had reduced serum levels of creatine, glycine, histidine, and threonine compared to non-COPD smokers. Although these findings indicate a possible subclinical malnutrition in the context of respiratory disease, results are still inconclusive regarding the association of these specific metabolites with COPD.

Body composition is greatly affected in COPD as the disease progresses (Schols et al. 2005). Chronic bronchitis and emphysema are the two distinct phenotype

variations of COPD and patients with these diseases have different body silhouettes, presumably due to difference in lipid metabolic pathways. Some reports have suggested that perturbation of lipid metabolism occurs in COPD (Chen et al. 2019; Rafie et al. 2018). In the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS) cohort, Halper-Stromberg et al. (2019) observed that phosphatidylethanolamine, phosphatidylcholine, phosphatidylinositol, leucine, and lysine from BAL fluid in COPD patients were associated with higher odds of developing emphysema. Liang et al. (2019) identified that serum metabolites such as glutamine, glycine, histidine, hypoxanthine, α-N-phenylacetyl-L-glutamine, L-pipecolic acid, P-chlorophenylalanine, pseudouridine, and L-citrulline levels were markedly different between asthma and COPD.

There is an increasing body of evidence suggesting that sphingolipids, which that play crucial roles in the structure and function of plasma membranes and signal transduction, also have roles in the pathogenesis of COPD, asthma, and other lung conditions (Vlahos 2020). Lipidomic studies have shown that COPD patients have higher plasma concentrations of very low density lipoprotein (VLDL) compared to healthy individuals, which strongly correlates with higher central and peripheral airway resistance (Rafie et al. 2018). Nambiar et al. (2021) found that blood palmitoleic acid, linoleic acid, and dihydrotestosterone were lower in COPD patients than healthy controls. Similarly, another study showed that the levels of serum lysophosphatidylcholine (LPC) 18:3, lysophosphatidylethanolamine (LPE) 16:1, and phosphatidylinositol (PI) 32:1 were markedly reduced in acute exacerbations in COPD, thus highlighting the role of glycerophospholipids in the pathophysiology of COPD (Gai et al. 2021).

Another recent report reiterated these findings in the context of disease onset and stages in COPD where the authors observed that phosphatidylcholine and LPC were key indicators of COPD onset and that phosphatidylserine and diacylglycerol could potentially indicate the various COPD stages (Zhou et al. 2020). In line with these observations, polyunsaturated acid metabolites were found to be associated with reduced lung function and disease severity in COPD (Ran et al. 2019; Yu et al. 2019; Xue et al. 2020). Pinto-Plata et al. identified plasma lipid metabolites that may predict survival differences in COPD patients (Pinto-Plata et al. 2019). Using the Karolinska COSMIC cohort, Naz et al. (2017) found that the autotaxin-lysophosphatidic acid axis may be dysregulated due to oxidative stress in COPD and that sex-regulated phenotypes are influential in the pathophysiology of disease. However, despite several reports demonstrating associations between metabolites with disease progression and severity in COPD, it is still not clear whether these metabolites may influence pathophysiological mechanisms. Furthermore, there may be several residual confounders that influence the dysregulation of metabolic pathways in disease conditions. Therefore, any interpretation and conclusions made from these metabolic outcomes should be made cautiously (Kilk et al. 2018).

2.3 Pneumonia

Community-acquired pneumonia remains a major cause of morbidity and mortality around the world, with over a million hospitalizations each year in the USA prior to COVID-19 (Griffin et al. 2013). Among the most common bacterial strains involved in community-acquired pneumonia are *Streptococcus pneumoniae* and *Staphylococcus aureus*, which are also found as commensal bacteria in healthy humans. The challenge in controlling the incidence of pneumonia is to determine the etiological process by which it occurs in individual patients. Using systems biology approaches, it is hoped that diagnosis and monitoring of disease may be enhanced to allow for more accurate prescription of drugs in pneumonia and similar inflammatory lung diseases (Wheelock et al. 2013).

Application of NMR analysis of pneumonia patient urine suggests that definitive metabolic profiles could be applied to infection with *S. pneumoniae* (Fig. 4). The pattern of urinary metabolites detected in pneumococcal pneumonia could be distinguished from pneumonia associated with viruses and other bacterial strains (Slupsky et al. 2009a). An animal model of pneumonia also demonstrated that distinct metabolic profiles could be detected in the urine of mice infected with *S. pneumoniae* or methicillin-resistant *S. aureus*, a major cause of antibiotic-resistant pneumonia (Fig. 5) (Slupsky et al. 2009b). These studies indicate that metabolomics has potential for the diagnosis, monitoring, and clinical management of pneumococcal diseases.

2.4 Acute Lung Injury/Acute Respiratory Distress Syndrome (ARDS)

Acute lung injury and its more severe form, ARDS, is characterized by infiltration of an inflammatory, fibrin-rich exudate into the pulmonary interstitium and alveolar spaces (Gattinoni et al. 2014; Martin and Matute-Bello 2011; Ware and Matthay

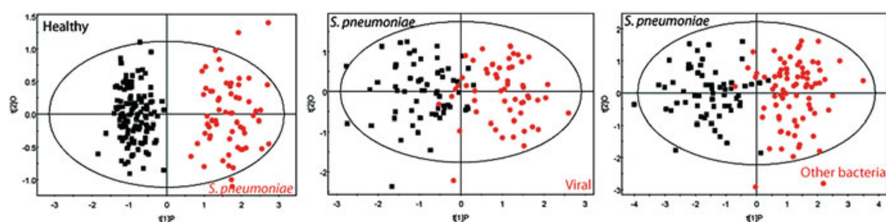


Fig. 4 Differentiating between different types of pneumonia in human patients. Urinary metabolites were found to be distinct in pneumonia caused by *S. pneumoniae* and other pathogens. These graphs show OPLS-DA models based on 61 measured metabolites found in the urine from *S. pneumoniae* patients compared with those found in viral pneumonia and other bacteria (including *Mycoplasma tuberculosis*, *Legionella pneumophila*, *S. aureus*, and others). Reprinted with permission from Slupsky et al. (2009b) J. Proteome Res. 8:5550–5558. Copyright 2009 American Chemical Society

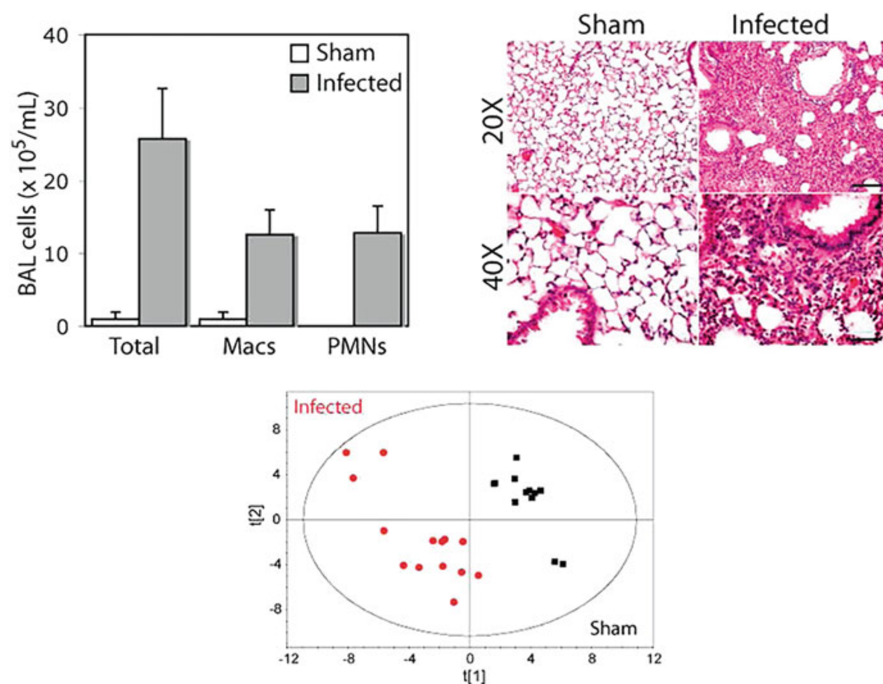


Fig. 5 Metabolic profiles in mice infected with *S. pneumoniae*. An inbred strain of mice (C57BL/6), maintained in specific virus antigen-free housing with autoclaved bedding and identical dietary supplies, was infected intratracheally with a clinical isolate of *S. pneumoniae*, serotype 14. After 24 h of infection, bronchoalveolar lavage (BAL) samples were analyzed for cell counts (**a**) and histology was carried out on lung sections (**b**) to confirm inflammation arising from infection. At the same time, urine samples were collected from animals that were subjected to NMR analysis and a PCA model of urinary metabolite concentrations was generated (**c**). Reprinted with permission from Slupsky et al. (2009a) J. Proteome Res. 8(6):3029–3036. Copyright 2009 American Chemical Society

2000; Li et al. 2011). This influx leads to impaired lung function and diminished gas exchange (Ware and Matthay 2000). First described in 1967 by Ashbaugh et al. (1967), ARDS is precipitated by many different causes, with the most common being sepsis, pneumonia, severe trauma, and more recently, severe COVID-19 (Huang et al. 2020). ARDS is accompanied by an extraordinarily high mortality rate (approximately 30% of patients die upon diagnosis of ARDS), and to date there have been few effective pharmacotherapies for its treatment that mainly serve to shorten the duration of illness rather than reverse it entirely. In addition, no effective predictive or prognostic biomarkers are available to indicate the likelihood of a patient developing ARDS. This has prompted a search for biomarkers of ARDS, which has been led by genomics and proteomics, although neither field has yielded suitable markers, and no candidate has progressed beyond the initial discovery phase (Serkova et al. 2011; Rogers and Matthay 2014; Meyer 2013, 2014). This is likely

due to the heterogeneity of disease, and much of the variation could lie beyond the proteome or genome, possibly in the metabolome (Serkova et al. 2011; Rogers and Matthay 2014). Thus, metabolomics presents itself as a potentially valuable tool for analysis in ARDS.

A challenge with understanding mechanisms associated with ARDS is that there are no translational animal models that accurately mimic human disease (Martin and Matute-Bello 2011; Matute-Bello et al. 2011; Matute-Bello and Downey 2013). Despite this limitation, there have been several metabolomic studies carried out in rodent models that demonstrate changes in metabolites (Stringer et al. 2016). In early experimental models, mechanical ventilation-induced ARDS in rodents generated metabolic profiles in serum, lung tissue, and BAL samples (Izquierdo-Garcia et al. 2014). Putative metabolites of ARDS were reported to be increased lactate and decreased glucose and glycine in lung tissues, together with increased glucose, lactate, acetate, 3-hydroxybutyrate, and creatine in BAL samples. NMR-detected metabolites in lung samples were associated with markers of ARDS phenotype (peak inspiratory pressure, PaO_2, and lung histology), but there was no association between these ARDS indices and serum metabolites. In one of the first studies examining the metabolomics of experimental ARDS, a cytokine-induced lung injury model was tested to determine the temporal association between inflammation in the lungs and changes in lung metabolome (Serkova et al. 2008). Cytokine-induced lung injury resulted in decreased ATP, energy balance, and energy charge levels, suggesting a decreased energy state. Together with this there was a significant increase in glycolytic activity, measured as elevated lactate-to-glucose levels that normalized 24 h after the induction of injury. Collectively these findings indicate that a shift in cell energy metabolism occurs in lung tissues in ARDS. The benefit of this study was that it demonstrated an association between phenotypic and metabolic changes, an important first step in biomarker discovery. To date, biomarkers have not been found that can differentiate between the two extremes of mild interstitial edema and extensive cellular injury in the spectrum of acute lung injury. However, continued analysis by magnetic resonance imaging and metabolic NMR spectroscopy may enhance the development of more robust and predictive longitudinal processes of experimental lung injury. Other animal models have shown significant metabolic shifts in ARDS induced by a variety of stimuli, reviewed in detail in Stringer et al. (2016).

Few clinical studies have reported metabolomics analysis of patients with ARDS. Several studies suggest that the use of BAL samples could provide insight into the metabolomic profile associated with ARDS. In one study, at least 26 and 18 endogenous metabolites, respectively, could be used to differentiate ARDS from healthy BAL samples using liquid chromatography-MS analysis (Evans et al. 2014). These included lactate and other energy metabolism-associated metabolites such as citrate, creatine, and creatinine which are increased in the plasma of patients with ARDS (Stringer et al. 2011). These findings demonstrate the utility of BAL as a biofluid for metabolomics analysis.

In addition, some reports have demonstrated the utility of exhaled breath as a vehicle for metabolomics analysis (Schubert et al. 1998; Bos 2018). For example, Schubert et al. demonstrated the utility of exhaled breath as a sample for metabolomics analysis (Schubert et al. 1998). This was furthered in a study by Bos et al. (2014) which found that three metabolites, octane, acetaldehyde, and 3-methylheptane, were able to discriminate ARDS from non-ARDS patients. Octane is an end-product of lipid peroxidation, one of the degenerative processes caused by oxidative stress (Riely et al. 1974; Horvat et al. 1964).

Interestingly, a recent study examining EBC from patients on mechanical ventilation due to severe COVID-19 or non-COVID-19 ARDS showed a characteristic "breathprint" for COVID-19 (Grassin-Delyle et al. 2021) that could be distinguished from non-COVID-19 ARDS. In this study, the four most prominent volatile compounds in COVID-19 patients were methylpent-2-enal, 2,4-octadiene, 1-chlorohelptane, and nonanal, suggesting that real-time metabolomics analysis of exhaled breath may identify patients with COVID-19. Nonanal is a sub-product of oxidative stress-mediated destruction of the cell membrane (Rahman 2003).

In summary, the metabolomics data generated from experimental and clinical studies of ARDS demonstrate that a disturbance in oxidative stress metabolism and energy levels occur in this disease, which is consistent with what has been described for the pathology of ARDS. To date, there appears to be no multi-center prospective studies done for metabolomics analysis of ARDS. Our understanding of ARDS metabolomics has been based on small studies that demonstrate feasibility for evaluation of ARDS phenotypes and for determining lung injury severity. Going forward, we will need to establish clinical trials aimed at testing prevention and treatment strategies in ARDS patients by applying metabolomics analysis to the spectrum of disease that presents in this population.

2.5 Occupational and Environmental Lung Diseases

Occupational exposure is one of the major risk factors associated with respiratory illnesses, and the incidence of occupational lung diseases is increasing due to expanding populations and consumer needs (Moitra et al. 2015). According to the report of the International Labour Organization, nearly two million people die each year due to workplace accidents, of which over 30% die due to lung cancer or other lung diseases as a result of workplace exposure (Cullinan et al. 2017). In many cases, occupational lung diseases are improperly recorded or detected, often due to a lack of causal evidence, all of which contributes to a significant underestimation of the true burden of these diseases. Although several biomarker-related reports have been published in the context of occupational exposure, metabolomic studies have been very limited to date. We discuss some of the few studies below on occupational and environmental lung diseases.

Among a group of workers employed in carbon-coating friction systems, Maniscalco et al. (2018) found that the concentrations of VOCs, including

1,2-propanediol, phenylalanine, 3-hydroxybutyrate, and isopropanol, were significantly elevated in the EBC of the workers who did not wear a mask at the work, compared to those who routinely wore masks.

Other markers such as polycyclic aromatic hydrocarbons (PAH) have been found to be associated with occupational exposure. Wei et al. studied the joint effects of arsenic exposure, smoking, and physical exercise on lung function changes among a group of coke-oven workers and found that urinary concentrations of PAH were significantly higher in coke-oven workers than office workers in the same industry who were not directly exposed to the ovens (Wei et al. 2021). Using a nationwide biomonitoring survey of the Korean National Environmental Health Survey, Koh et al. collected measurements of urinary 1-hydroxypyrene (1-OHP) as a metabolite of interest for PAH exposure at workplace. They found that the level of urinary 1-OHP was highest among people engaged in construction and mine-related occupations. Although that study did not explicitly study associations between urinary metabolites and respiratory health, the effect of PAH on respiratory health is already well known and therefore, urinary 1-OHP could potentially be a marker of PAH-associated respiratory dysfunctions.

The collapse of the World Trade Center (WTC) on 9/11 introduced a novel and unprecedented exposure scenario in which hundreds of thousands of New Yorkers were affected in the ensuing years. Firefighters and all first responders were exposed to huge amounts of dust containing numerous fibrous, chemical, and hazardous substances. To date, several reports have been published on the respiratory health of the workers who were exposed to WTC dust, resulting in a condition known as WTC lung disease (also known as WTC sarcoid-like granulomatous pulmonary disease). For example, a recent study provided novel insights into metabolic syndrome as a risk factor for lung function decline in a cohort of firefighters exposed to materials arising from the collapse of the WTC (Kwon et al. 2021). They also proposed that regulating metabolic syndrome, particularly dyslipidemia, could also help to decrease the risk of developing WTC lung disease. This group also showed previously that the serum metabolome, particularly the sphingolipid cluster containing sphingosine-1-phosphate, a pleiotropic inflammatory mediator, was low in WTC lung diseases, suggesting decreased bioavailability and increased risk of compromised vascular integrity in WTC lung disease (Crowley et al. 2018). A mouse model of WTC particulate matter exposure was also investigated and showed that several prominent metabolic pathways were affected, including advanced glycation end-products and lipids (including sphingolipids), that correlated with inflammatory changes and attenuation of antioxidant potential (Veerappan et al. 2020). However, despite these interesting outcomes, correlations between metabolomics and clinical evaluation in occupational lung diseases remain limited, and therefore more studies are required to elucidate the crosstalk between these two aspects.

3 Metabolomics of Lung Microbiome in Respiratory Diseases

3.1 Asthma

The lung microbiota and metabolome are likely to play a pivotal role in the onset of disease in the case of asthma (Barcik et al. 2020). It is now emerging that metabolically active microbiota that reside in the lung under normal conditions maintain a complex network of crosstalk with the host in a symbiotic manner. In disease conditions, however, this symbiosis is transformed into dysbiosis that can alter the host immune response, which influences the overall lung health (Loverdos et al. 2019). The composition of normal lung microbiota consists of *Bacteroidetes* and *Firmicutes* (the most abundant two genera), and apart from these two, *Proteobacteria*, *Actinobacteria*, and *Fusobacteria* have also been found by 16S rRNA sequencing in endobronchial brushing samples (Charlson et al. 2011; Bassis et al. 2015). Although normal lung microbiota consists of a relatively small bacterial population, estimated to be around 10^3 to $10^5/cm^2$, their intensely intricate crosstalk is thought to be primarily responsible for the conduct of most of the host-microbiome interplay (Charlson et al. 2011; Bassis et al. 2015; Hilty et al. 2010; Mathieu et al. 2018; Denner et al. 2016; Goleva et al. 2013).

In asthma, the bacterial pattern of the pulmonary microbiome has been characterized in several studies. It is evident that some bacterial species become elevated in nasopharyngeal swabs from asthmatics, such as *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Moraxella catarrhalis*, compared to healthy controls. These bacteria are well-known pathogens that can cause infectious exacerbations (Dickson et al. 2016). Interestingly, Huang et al. showed in patients with severe asthma, *Actinobacteria* is present at high abundance in correlation with elevated sputum leukocytes and eosinophils in bronchial biopsies (Huang et al. 2015). It has been also shown that elevated eosinophil numbers in lavage, along with reduced FEV_1, correlate with bacterial α-diversity (based on comparison of different species present in same sample) in endobronchial brushings of asthmatic subjects. Bacterial species associated with lower airway obstruction show distinctive features associated with FEV_1 levels. For example, patients with asthma exhibiting $FEV_1 < 60\%$ had low α-diversity but high β-diversity compared to asthma patients with $FEV_1 > 80\%$ (Denner et al. 2016).

Interestingly, the gut microbiome is an important component of asthma pathophysiology which has not been explored in detail. The human gut possesses a surface area of 150–200 m^2, which harbors 100,000 to 100 billion bacteria per mL of sample, depending on the region of sample collection (Sender et al. 2016). A relationship between the gut and lung was discovered upon the observation that different lung diseases may be influenced by changes in the gut microflora and vice versa. The microbiota in these two sites is therefore connected by a gut-lung axis that is important in relation to asthma (Marsland et al. 2015). Among many different metabolites produced by the gut microbiome, short-chain fatty acids (SCFAs), such as acetate, propionate, and butyrate, have been found to regulate physiological and immunological responses in humans. It is well known that not only do SCFAs

provide a source of energy, but they also function as signaling molecules. SCFAs have been shown to have multiple signaling effects: they inhibit histone deacetylases (HDACs) that increase cytokine gene expression by promoting an anti-inflammatory cell phenotype to maintain homeostasis, suppress transcription factors (nuclear factor κ-light-chain-enhancer of activated B cells (NF-κB)), and reduce tumor necrosis factor-α (TNF-α) production (Durack et al. 2017; Chambers et al. 2018; Tan et al. 2014). Depletion of SCFA-producing bacteria as a mechanistic link between the microbiome and asthma susceptibility or severity has been suggested by Cait et al. (2018). Hence, SCFAs derived from the metabolic activity of gut microbiota inhibit proinflammatory responses in the lungs. Although the mechanism underlying this pathway is unclear, the most likely mechanistic explanation is that the hepatic system may weaken innate immune responses by SCFAs binding to G protein-coupled receptors and/or inhibition of the mevalonate/isoprenoid pathway through HMGCoA reductase (Young et al. 2016). The specific mediators that make up the communication between gut and lung is still unclear, but it has been speculated that gut epithelial cells and immune cells absorb signals from the endothelium to form local cytokine microenvironments, and eventually this alters the immune response in distal sites such as the lung (Budden et al. 2017).

Overall, these studies demonstrate that the gut and lung microbiome, and its associated metabolome, have an enormous impact on patient outcomes in asthma. Findings from these reports could contribute to the discovery of mechanisms and novel biomarkers for asthma and its associated exacerbations.

3.2 Chronic Obstructive Pulmonary Disease (COPD)

Recent evidence suggests an association between the lung microbiome and COPD, suggesting a contribution of the lung bacterial community to disease progression in the form of dysbiosis (Hilty et al. 2010; Erb-Downward et al. 2011; Zakharkina et al. 2013; Pragman et al. 2012). Phylogenetic analysis of microbial populations in samples collected from the oropharynx and bronchial brushings from COPD patients and healthy controls showed increased populations of pathogenic *Proteobacteria* (*Haemophilus* spp.) over *Bacteroidetes* (*Prevotella* spp.), with the latter being especially reduced (Hilty et al. 2010). Other studies also demonstrated that healthy individuals commonly exhibit higher populations of *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Fusobacteria*, and *Actinobacteria*, in contrast to pathogenic *Haemophilus*, *Streptococcus*, *Klebsiella*, *Pseudomonas*, and *Moraxella* in COPD patients (Wu et al. 2014; Murphy et al. 2005). In addition, several reports have described that exposure to tobacco smoke can modify bacterial populations in the mouth and lungs. Though studies are limited in the context of COPD, numerous reports indicate an alteration of the oral and respiratory bacterial microbiome as an effect of tobacco smoking (Morris et al. 2013; Zhang et al. 2018; Huang and Shi 2019). In COPD patients, commensal colonization of *H. influenzae*, *S. pneumoniae*,

Pseudomonas, and *Moraxella* are frequently observed in the lungs (Simpson et al. 2016).

The gut-lung axis also features prominently in COPD (Young et al. 2016). Fecal microbiota derived from COPD patients have been demonstrated to contribute to the development of COPD in a mouse model (Li et al. 2021). The fecal microbiota of COPD patients were found to contain lower levels of SCFAs, which could contribute to the manifestation of COPD.

In another recent study comparing the metabolomic profiles of COPD patients with healthy humans (Bowerman et al. 2020), it was found that COPD patients and healthy individuals manifest significantly different sets of microbial and metabolic signatures in fecal samples. As many as 146 bacterial species differ in between these two groups, along with a group of the top 50 indicator metabolites that distinguished between healthy and COPD individuals, consisting of mostly lipids (46%), amino acids (20%), and xenobiotic compounds (20%). Hence, it can be deduced that the intricate mechanisms associated with the gut-lung axis and the host's microbial community play a crucial role in the manifestation and progression of COPD.

4 Conclusive Remarks

Taken together, we have reviewed some of the literature associated with metabolomics analysis of biological fluids obtained from patients and experimental animal models with a range of respiratory diseases. Metabolomics is a fundamental part of systems biology analysis that has enormous clinical potential in discovering novel biomarkers as well as understanding disease pathophysiology. Because of its rapidly changing properties in health as well as disease, metabolomics has the power to generate snapshots of metabolites from a given sample that can be followed over time with repeated sampling. Several high-throughput systems have the ability to capture the identities and qualities of metabolites in a rapid manner using NMR or MS-based techniques. Challenges remain with the application of NMR in complex biological samples, which is less sensitive to small amounts of metabolites in many cases than MS. An important distinction to make is that analysis of metabolites in lung-specific samples is predicted to provide greater sensitivity to the tissue-specific metabolome over that of blood-derived (plasma, serum, or whole blood) or urinary metabolites. This is especially evident in the case of analysis of the metabolomics of the lung microbiome. Variability of NMR-measured metabolites is also an issue, with differing results found within a single facility as well as multiple locations (Lacy et al. 2014). In addition, a substantial number of unknown metabolites have been detected by MS that await more detailed identification in biological samples. We look forward to a future where we can implement increasingly sophisticated analyses of biological samples using systems biology approaches in respiratory diseases.

References

Ashbaugh DG, Bigelow DB, Petty TL, Levine BE (1967) Acute respiratory distress in adults. Lancet 2(7511):319–323

Barcik W, Boutin RCT, Sokolowska M, Finlay BB (2020) The role of lung and gut microbiota in the pathology of asthma. Immunity 52(2):241–255

Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, Beck JM, Curtis JL, Huffnagle GB (2015) Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. MBio 6(2):e00037

Bian X, Sun B, Zheng P, Li N, Wu JL (2017) Derivatization enhanced separation and sensitivity of long chain-free fatty acids: application to asthma using targeted and non-targeted liquid chromatography-mass spectrometry approach. Anal Chim Acta 989:59–70

Bos LDJ (2018) Diagnosis of acute respiratory distress syndrome by exhaled breath analysis. Ann Transl Med 6(2):33

Bos LD, Weda H, Wang Y, Knobel HH, Nijsen TM, Vink TJ, Zwinderman AH, Sterk PJ, Schultz MJ (2014) Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome. Eur Respir J 44(1):188–197

Bowerman KL, Rehman SF, Vaughan A, Lachner N, Budden KF, Kim RY, Wood DLA, Gellatly SL, Shukla SD, Wood LG, Yang IA et al (2020) Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. Nat Commun 11(1):5886

Budden KF, Gellatly SL, Wood DL, Cooper MA, Morrison M, Hugenholtz P, Hansbro PM (2017) Emerging pathogenic links between microbiota and the gut-lung axis. Nat Rev Microbiol 15(1): 55–63

Cait A, Hughes MR, Antignano F, Cait J, Dimitriu PA, Maas KR, Reynolds LA, Hacker L, Mohr J, Finlay BB, Zaph C et al (2018) Microbiome-driven allergic lung inflammation is ameliorated by short-chain fatty acids. Mucosal Immunol 11(3):785–795

Chambers ES, Preston T, Frost G, Morrison DJ (2018) Role of gut microbiota-generated short-chain fatty acids in metabolic and cardiovascular health. Curr Nutr Rep 7(4):198–206

Chang-Chien J, Huang HY, Tsai HJ, Lo CJ, Lin WC, Tseng YL, Wang SL, Ho HY, Cheng ML, Yao TC (2021) Metabolomic differences of exhaled breath condensate among children with and without asthma. Pediatr Allergy Immunol 32(2):264–272

Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG (2011) Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med 184(8):957–963

Checkley W, Deza MP, Klawitter J, Romero KM, Klawitter J, Pollard SL, Wise RA, Christians U, Hansel NN (2016) Identifying biomarkers for asthma diagnosis using targeted metabolomics approaches. Respir Med 121(1532–3064 (Electronic)):59–66

Chen H, Li Z, Dong L, Wu Y, Shen H, Chen Z (2019) Lipid metabolism in chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 14:1009–1018

Chiu CY, Lin G, Cheng ML, Chiang MH, Tsai MH, Su KW, Hua MC, Liao SL, Lai SH, Yao TC, Yeh KW et al (2018) Longitudinal urinary metabolomic profiling reveals metabolites for asthma development in early childhood. Pediatr Allergy Immunol 29(5):496–503

Chiu CY, Cheng ML, Chiang MH, Wang CJ, Tsai MH, Lin G (2020) Metabolomic analysis reveals distinct profiles in the plasma and urine associated with IgE reactions in childhood asthma. J Clin Med 9(3):887

Chiu CY, Cheng ML, Chiang MH, Wang CJ, Tsai MH, Lin G (2021) Integrated metabolic and microbial analysis reveals host–microbial interactions in IgE-mediated childhood asthma. Sci Rep 11(1):23407

Crowley G, Kwon S, Haider SH, Caraher EJ, Lam R, St-Jules DE, Liu M, Prezant DJ, Nolan A (2018) Metabolomics of world trade center-lung injury: a machine learning approach. BMJ Open Respir Res 5(1):e000274

Cullinan P, Munoz X, Suojalehto H, Agius R, Jindal S, Sigsgaard T, Blomberg A, Charpin D, Annesi-Maesano I, Gulati M, Kim Y et al (2017) Occupational lung diseases: from old and novel exposures to effective preventive strategies. Lancet Respir Med 5(5):445–455

Denner DR, Sangwan N, Becker JB, Hogarth DK, Oldham J, Castillo J, Sperling AI, Solway J, Naureckas ET, Gilbert JA, White SR (2016) Corticosteroid therapy and airflow obstruction influence the bronchial microbiome, which is distinct from that of bronchoalveolar lavage in asthmatic airways. J Allergy Clin Immunol 137(5):1398–1405

Devine JF (2008) Chronic obstructive pulmonary disease: an overview. Am Health Drug Benefits 1(7):34–42

Diao W, Labaki WW, Han MK, Yeomans L, Sun Y, Smiley Z, Kim JH, McHugh C, Xiang P, Shen N, Sun X et al (2019) Disruption of histidine and energy homeostasis in chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 14:2015–2025

Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB (2016) The microbiome and the respiratory tract. Annu Rev Physiol 78(1):481–504

Durack J, Lynch SV, Nariya S, Bhakta NR, Beigelman A, Castro M, Dyer AM, Israel E, Kraft M, Martin RJ, Mauger DT et al (2017) Features of the bronchial bacterial microbiome associated with atopy, asthma, and responsiveness to inhaled corticosteroid treatment. J Allergy Clin Immunol 140(1):63–75

Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, Martinez FJ et al (2011) Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One 6(2):e16384

Evans CR, Karnovsky A, Kovach MA, Standiford TJ, Burant CF, Stringer KA (2014) Untargeted LC-MS metabolomics of bronchoalveolar lavage fluid differentiates acute respiratory distress syndrome from health. J Proteome Res 13(2):640–649

Ferraro VA, Carraro S, Pirillo P, Gucciardi A, Poloniato G, Stocchero M, Giordano G, Zanconato S, Baraldi E (2020) Breathomics in asthmatic children treated with inhaled corticosteroids. Meta 10(10):390

Gai X, Guo C, Zhang L, Zhang L, Abulikemu M, Wang J, Zhou Q, Chen Y, Sun Y, Chang C (2021) Serum glycerophospholipid profile in acute exacerbation of chronic obstructive pulmonary disease. Front Physiol 12:646010

Gattinoni L, Cressoni M, Brazzi L (2014) Fluids in ARDS: from onset through recovery. Curr Opin Crit Care 20(4):373–377

Ghosh N, Choudhury P, Kaushik SR, Arya R, Nanda R, Bhattacharyya P, Roychowdhury S, Banerjee R, Chaudhury K (2020) Metabolomic fingerprinting and systemic inflammatory profiling of asthma COPD overlap (ACO). Respir Res 21(1):126

Goleva E, Jackson LP, Harris JK, Robertson CE, Sutherland ER, Hall CF, Good JT Jr, Gelfand EW, Martin RJ, Leung DY (2013) The effects of airway microbiome on corticosteroid responsiveness in asthma. Am J Respir Crit Care Med 188(10):1193–1201

Grassin-Delyle S, Roquencourt C, Moine P, Saffroy G, Carn S, Heming N, Fleuriet J, Salvator H, Naline E, Couderc LJ, Devillier P et al (2021) Metabolomics of exhaled breath in critically ill COVID-19 patients: a pilot study. EBioMedicine 63:103154

Griffin MR, Zhu Y, Moore MR, Whitney CG, Grijalva CG (2013) U.S. Hospitalizations for pneumonia after a decade of pneumococcal vaccination. N Engl J Med 369(2):155–163

Halper-Stromberg E, Gillenwater L, Cruickshank-Quinn C, O'Neal WK, Reisdorph N, Petrache I, Zhuang Y, Labaki WW, Curtis JL, Wells J, Rennard S et al (2019) Bronchoalveolar lavage fluid from COPD patients reveals more compounds associated with disease than matched plasma. Metabolites 9(8):157

Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF et al (2010) Disordered microbial communities in asthmatic airways. PLoS One 5(1):e8578

Horvat RJ, Lane WG, Ng H, Shepherd AD (1964) Saturated hydrocarbons from autoxidizing methyl linoleate. Nature 203:523–524

Huang C, Shi G (2019) Smoking and microbiome in oral, airway, gut and some systemic diseases. J Transl Med 17(1):225
Huang YJ, Nariya S, Harris JM, Lynch SV, Choy DF, Arron JR, Boushey H (2015) The airway microbiome in patients with severe asthma: associations with disease features and severity. J Allergy Clin Immunol 136(4):874–884
Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506
Izquierdo-Garcia JL, Naz S, Nin N, Rojas Y, Erazo M, Martinez-Caro L, Garcia A, de Paula M, Fernandez-Segoviano P, Casals C, Esteban A et al (2014) A Metabolomic approach to the pathogenesis of ventilator-induced lung injury. Anesthesiology 120(3):694–702
Jiang T, Dai L, Li P, Zhao J, Wang X, An L, Liu M, Wu S, Wang Y, Peng Y, Sun D et al (2021) Lipid metabolism and identification of biomarkers in asthma by lipidomic analysis. Biochim Biophys Acta Mol Cell Biol Lipids 1866(2):158853
Kang YP, Lee WJ, Hong JY, Lee SB, Park JH, Kim D, Park S, Park CS, Park SW, Kwon SW (2014) Novel approach for analysis of bronchoalveolar lavage fluid (BALF) using HPLC-QTOF-MS-based lipidomics: lipid levels in asthmatics and corticosteroid-treated asthmatic patients. J Proteome Res 13(9):3919–3929
Kelly RS, Virkud Y, Giorgio R, Celedon JC, Weiss ST, Lasky-Su J (2017) Metabolomic profiling of lung function in Costa-Rican children with asthma. Biochim Biophys Acta Mol basis Dis 1863(6):1590–1595
Keogh E, Mark WE (2021) Managing malnutrition in COPD: a review. Respir Med 176:106248
Kilk K, Aug A, Ottas A, Soomets U, Altraja S, Altraja A (2018) Phenotyping of chronic obstructive pulmonary disease based on the integration of metabolomes and clinical characteristics. Int J Mol Sci 19(3):666
Kwon S, Lee M, Crowley G, Schwartz T, Zeig-Owens R, Prezant DJ, Liu M, Nolan A (2021) Dynamic metabolic risk profiling of world trade center lung disease: a longitudinal cohort study. Am J Respir Crit Care Med 204(9):1035–1047
Lacy P, McKay RT, Finkel M, Karnovsky A, Woehler S, Lewis MJ, Chang D, Stringer KA (2014) Signal intensities derived from different NMR probes and parameters contribute to variations in quantification of metabolites. PLoS One 9(1):e85732
Li G, Malinchoc M, Cartin-Ceba R, Venkata CV, Kor DJ, Peters SG, Hubmayr RD, Gajic O (2011) Eight-year trend of acute respiratory distress syndrome: a population-based study in Olmsted County, Minnesota. Am J Respir Crit Care Med 183(1):59–66
Li S, Liu J, Zhou J, Wang Y, Jin F, Chen X, Yang J, Chen Z (2020) Urinary Metabolomic profiling reveals biological pathways and predictive signatures associated with childhood asthma. J Asthma Allergy 13:713–724
Li N, Dai Z, Wang Z, Deng Z, Zhang J, Pu J, Cao W, Pan T, Zhou Y, Yang Z, Li J et al (2021) Gut microbiota dysbiosis contributes to the development of chronic obstructive pulmonary disease. Respir Res 22(1):274
Liang Y, Gai XY, Chang C, Zhang X, Wang J, Li TT (2019) Metabolomic profiling differences among asthma, COPD, and healthy subjects: a LC-MS-based metabolomic analysis. Biomed Environ Sci 32(9):659–672
Loureiro CC, Oliveira AS, Santos M, Rudnitskaya A, Todo-Bom A, Bousquet J, Rocha SM (2016) Urinary metabolomic profiling of asthmatics can be related to clinical characteristics. Allergy 71(9):1362–1365
Loverdos K, Bellos G, Kokolatou L, Vasileiadis I, Giamarellos E, Pecchiari M, Koulouris N, Koutsoukou A, Rovina N (2019) Lung microbiome in asthma: current perspectives. J Clin Med 8(11):1967
Madsen R, Lundstedt T, Trygg J (2010) Chemometrics in metabolomics – a review in human disease diagnosis. Anal Chim Acta 659(1–2):23–33
Maniscalco M, Paris D, Melck DJ, D'Amato M, Zedda A, Sofia M, Stellato C, Motta A (2017) Coexistence of obesity and asthma determines a distinct respiratory metabolic phenotype. J Allergy Clin Immunol 139(5):1536–1547

Maniscalco M, Paris D, Melck D, Chiariello N, Di Napoli F, Manno M, Iavicoli I, Motta A (2018) Biomonitoring of workers using nuclear magnetic resonance-based metabolomics of exhaled breath condensate: a pilot study. Toxicol Lett 298:4–12
Marsland BJ, Trompette A, Gollwitzer ES (2015) The gut-lung axis in respiratory disease. Ann Am Thorac Soc 12(Suppl 2):S150–S156
Martin TR, Matute-Bello G (2011) Experimental models and emerging hypotheses for acute lung injury. Crit Care Clin 27(3):735–752
Mathieu E, Escribano-Vazquez U, Descamps D, Cherbuy C, Langella P, Riffault S, Remot A, Thomas M (2018) Paradigms of lung microbiota functions in health and disease, particularly, in asthma. Front Physiol 9(1168):1168
Matute-Bello G, Downey GP (2013) Reply: defining lung injury in animals. Am J Respir Cell Mol Biol 48(2):267–268
Matute-Bello G, Downey G, Moore BB, Groshong SD, Matthay MA, Slutsky AS, Kuebler WM (2011) Acute Lung Injury in Animals Study G. An official American Thoracic Society workshop report: features and measurements of experimental acute lung injury in animals. Am J Respir Cell Mol Biol 44(5):725–738
Matysiak J, Klupczynska A, Packi K, Mackowiak-Jakubowska A, Breborowicz A, Pawlicka O, Olejniczak K, Kokot ZJ, Matysiak J (2020) Alterations in serum-free amino acid profiles in childhood asthma. Int J Environ Res Public Health 17(13):4758
Meyer NJ (2013) Future clinical applications of genomics for acute respiratory distress syndrome. Lancet Respir Med 1(10):793–803
Meyer NJ (2014) Beyond single-nucleotide polymorphisms: genetics, genomics, and other 'omic approaches to acute respiratory distress syndrome. Clin Chest Med 35(4):673–684
Moitra S, Puri R, Paul D, Huang YC (2015) Global perspectives of emerging occupational and environmental lung diseases. Curr Opin Pulm Med 21(2):114–120
Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, Flores SC, Fontenot AP, Ghedin E, Huang L, Jablonski K et al (2013) Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med 187(10):1067–1075
Murphy TF, Brauer AL, Grant BJ, Sethi S (2005) Moraxella catarrhalis in chronic obstructive pulmonary disease: burden of disease and immune response. Am J Respir Crit Care Med 172(2): 195–199
Nambiar S, Tan DBA, Clynick B, Bong SH, Rawlinson C, Gummer J, Corte TJ, Glaspole I, Moodley YP, Trengove R (2021) Untargeted metabolomics of human plasma reveal lipid markers unique to chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. Proteomics Clin Appl 15(2–3):e2000039
Naz S, Kolmert J, Yang M, Reinke SN, Kamleh MA, Snowden S, Heyder T, Levanen B, Erle DJ, Skold CM, Wheelock AM et al (2017) Metabolomics analysis identifies sex-associated metabotypes of oxidative stress and the autotaxin-lysoPA axis in COPD. Eur Respir J 49(6): 1602322
Novotna B, Abdel-Hamid M, Koblizek V, Svoboda M, Hejduk K, Rehacek V, Bis J, Salajka F (2018) A pilot data analysis of a metabolomic HPLC-MS/MS study of patients with COPD. Adv Clin Exp Med 27(4):531–539
Pang Z, Wang G, Wang C, Zhang W, Liu J, Wang F (2018) Serum metabolomics analysis of asthma in different inflammatory phenotypes: a cross-sectional study in Northeast China. Biomed Res Int 2018:2860521
Park YH, Fitzpatrick AM, Medriano CA, Jones DP (2017) High-resolution metabolomics to identify urine biomarkers in corticosteroid-resistant asthmatic children. J Allergy Clin Immunol 139(5):1518–1524
Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 13(4):263–269
Pinto-Plata V, Casanova C, Divo M, Tesfaigzi Y, Calhoun V, Sui J, Polverino F, Priolo C, Petersen H, de Torres JP, Marin JM et al (2019) Plasma metabolomics and clinical predictors of survival differences in COPD patients. Respir Res 20(1):219

Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE (2012) The lung microbiome in moderate and severe chronic obstructive pulmonary disease. PLoS One 7(10):e47305

Quan-Jun Y, Jian-Ping Z, Jian-Hua Z, Yong-Long H, Bo X, Jing-Xian Z, Bona D, Yuan Z, Cheng G (2017) Distinct metabolic profile of inhaled budesonide and salbutamol in asthmatic children during acute exacerbation. Basic Clin Pharmacol Toxicol 120(3):303–311

Rafie S, Moitra S, Brashier BB (2018) Association between the serum metabolic profile and lung function in chronic obstructive pulmonary disease. Turk Thorac J 19(1):13–18

Rahman I (2003) Oxidative stress, chromatin remodeling and gene transcription in inflammation and chronic lung diseases. J Biochem Mol Biol 36(1):95–109

Ran N, Pang Z, Gu Y, Pan H, Zuo X, Guan X, Yuan Y, Wang Z, Guo Y, Cui Z, Wang F (2019) An updated overview of Metabolomic profile changes in chronic obstructive pulmonary disease. Meta 9(6):111

Reinke SN, Gallart-Ayala H, Gomez C, Checa A, Fauland A, Naz S, Kamleh MA, Djukanovic R, Hinks TS, Wheelock CE (2017) Metabolomics analysis identifies different metabotypes of asthma severity. Eur Respir J 49(3):1601740

Riely CA, Cohen G, Lieberman M (1974) Ethane evolution: a new index of lipid peroxidation. Science 183(4121):208–210

Robertson DG, Watkins PB, Reily MD (2011) Metabolomics in toxicology: preclinical and clinical applications. Toxicol Sci 120:S146–S170

Rogers AJ, Matthay MA (2014) Applying metabolomics to uncover novel biology in ARDS. Am J Physiol Lung Cell Mol Physiol 306(11):L957–L961

Schols AM, Broekhuizen R, Weling-Scheepers CA, Wouters EF (2005) Body composition and mortality in chronic obstructive pulmonary disease. Am J Clin Nutr 82(1):53–59

Schubert JK, Muller WP, Benzing A, Geiger K (1998) Application of a new method for analysis of exhaled gas in critically ill patients. Intensive Care Med 24(5):415–421

Sender R, Fuchs S, Milo R (2016) Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell 164(3):337–340

Serkova NJ, Van Rheen Z, Tobias M, Pitzer JE, Wilkinson JE, Stringer KA (2008) Utility of magnetic resonance imaging and nuclear magnetic resonance-based metabolomics for quantification of inflammatory lung injury. Am J Physiol Lung Cell Mol Physiol 295(1):L152–L161

Serkova NJ, Standiford TJ, Stringer KA (2011) The emerging field of quantitative blood metabolomics for biomarker discovery in critical illnesses. Am J Respir Crit Care Med 184(6):647–655

Simpson JL, Baines KJ, Horvat JC, Essilfie AT, Brown AC, Tooze M, McDonald VM, Gibson PG, Hansbro PM (2016) COPD is characterized by increased detection of Haemophilus influenzae, Streptococcus pneumoniae and a deficiency of Bacillus species. Respirology 21(4):697–704

Slupsky CM, Cheypesh A, Chao DV, Fu H, Rankin KN, Marrie TJ, Lacy P (2009a) Streptococcus pneumoniae and Staphylococcus aureus pneumonia induce distinct metabolic responses. J Proteome Res 8(6):3029–3036

Slupsky CM, Rankin KN, Fu H, Chang D, Rowe BH, Charles PG, McGeer A, Low D, Long R, Kunimoto D, Sawyer MB et al (2009b) Pneumococcal pneumonia: potential for diagnosis through a urinary metabolic profile. J Proteome Res 8(12):5550–5558

Stringer KA, Serkova NJ, Karnovsky A, Guire K, Paine R 3rd, Standiford TJ (2011) Metabolic consequences of sepsis-induced acute lung injury revealed by plasma (1)H-nuclear magnetic resonance quantitative metabolomics and computational analysis. Am J Physiol Lung Cell Mol Physiol 300(1):L4–L11

Stringer KA, McKay RT, Karnovsky A, Quemerais B, Lacy P (2016) Metabolomics and its application to acute lung diseases. Front Immunol 7:44

Tan J, McKenzie C, Potamitis M, Thorburn AN, Mackay CR, Macia L (2014) Chapter three – the role of short-chain fatty acids in health and disease. In: Alt FW (ed) Advances in immunology, vol 121. Academic Press, pp 91–119

Tao JL, Chen YZ, Dai QG, Tian M, Wang SC, Shan JJ, Ji JJ, Lin LL, Li WW, Yuan B (2019) Urine metabolic profiles in paediatric asthma. Respirology 24(6):572–581

Van Vliet D, Smolinska A, Jobsis Q, Rosias PP, Muris JW, Dallinga JW, van Schooten FJ, Dompeling E (2016) Association between exhaled inflammatory markers and asthma control in children. J Breath Res 10(1):016014
van Vliet D, Smolinska A, Jobsis Q, Rosias P, Muris J, Dallinga J, Dompeling E, van Schooten FJ (2017) Can exhaled volatile organic compounds predict asthma exacerbations in children? J Breath Res 11(1):016016
Veerappan A, Oskuei A, Crowley G, Mikhail M, Ostrofsky D, Gironda Z, Vaidyanathan S, Wadghiri YZ, Liu M, Kwon S, Nolan A (2020) World trade center-cardiorespiratory and vascular dysfunction: assessing the phenotype and metabolome of a murine particulate matter exposure model. Sci Rep 10(1):3130
Vlahos R (2020) Lipids in chronic obstructive pulmonary disease: a target for future therapy? Am J Respir Cell Mol Biol 62(3):273–274
Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, Abbasi-Kangevari M, Abbastabar H, Abd-Allah F, Abdelalim A, Abdollahi M et al (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet 396(10258):1204–1222
Ware LB, Matthay MA (2000) The acute respiratory distress syndrome. N Engl J Med 342(18): 1334–1349
Wei W, Wu X, Bai Y, Li G, Meng H, Feng Y, Li H, Li M, Guan X, Fu M, Wang C et al (2021) Arsenic exposure and its joint effects with cigarette smoking and physical exercise on lung function impairment: evidence from an occupational cohort study. Environ Res 196:110419
Wheelock CE, Goss VM, Balgoma D, Nicholas B, Brandsma J, Skipp PJ, Snowden S, Burg D, D'Amico A, Horvath I, Chaiboonchoe A et al (2013) Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. Eur Respir J 42(3):802–825
Wishart DS (2005) Metabolomics: the principles and potential applications to transplantation. Am J Transplant 5(12):2814–2820
Wu D, Hou C, Li Y, Zhao Z, Liu J, Lu X, Shang X, Xin Y (2014) Analysis of the bacterial community in chronic obstructive pulmonary disease sputum samples by denaturing gradient gel electrophoresis and real-time PCR. BMC Pulm Med 14(1):179
Xia J, Broadhurst DI, Wilson M, Wishart DS (2013) Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics 9(2):280–299
Xue M, Cai C, Guan L, Xu Y, Lin J, Zeng Y, Hu H, Chen R, Wang H, Zhou L, Sun B (2020) Exploration of n-6 and n-3 polyunsaturated fatty acids metabolites associated with nutritional levels in patients with severe stable chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 15:1633–1642
Young RP, Hopkins RJ, Marsland B (2016) The gut-liver-lung axis. Modulation of the innate immune response and its possible role in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol 54(2):161–169
Yu B, Flexeder C, McGarrah RW 3rd, Wyss A, Morrison AC, North KE, Boerwinkle E, Kastenmuller G, Gieger C, Suhre K, Karrasch S et al (2019) Metabolomics identifies novel blood biomarkers of pulmonary function and COPD in the general population. Meta 9(4):61
Zakharkina T, Heinzel E, Koczulla RA, Greulich T, Rentz K, Pauling JK, Baumbach J, Herrmann M, Grunewald C, Dienemann H, von Muller L et al (2013) Analysis of the airway microbiota of healthy individuals and patients with chronic obstructive pulmonary disease by T-RFLP and clone sequencing. PLoS One 8(7):e68302
Zhang R, Chen L, Cao L, Li KJ, Huang Y, Luan XQ, Li G (2018) Effects of smoking on the lower respiratory tract microbiome in mice. Respir Res 19(1):253
Zhou J, Li Q, Liu C, Pang R, Yin Y (2020) Plasma metabolomics and Lipidomics reveal perturbed metabolites in different disease stages of chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 15:553–565

The Metabolomics of Critical Illness

Ana E. Pacheco-Navarro and Angela J. Rogers

Contents

1 Introduction: A Potential Role for Metabolomics in Understanding Critical Illness 368
2 Metabolic Alterations in Critical Illness 369
3 Endocrine Drivers of Altered Metabolism 369
4 Adrenergic Mediators and Their Changes in Critical Illness 371
5 Inflammatory Mediators (Cytokines) and Their Contribution to Metabolism 371
6 Metabolic Phases During Critical Illness 371
7 Autophagy and Mitophagy 372
8 Nutrient Changes During Critical Illness 373
8.1 Lactate: The Original Metabolomic Biomarker 374
9 Metabolomics in Sepsis 375
9.1 Diagnosis of Sepsis 375
9.2 Prognostication in Sepsis 376
9.3 Future Directions: Using Metabolomics to Guide Treatment 377
10 Metabolomics Studies in ARDS 377
10.1 Blood Metabolomics of ARDS 378
10.2 Using Blood Metabolomics to Identify High-Risk Subsets of ARDS Cohorts 379
10.3 Pulmonary Specific Samples for Metabolomic Profiling 379
11 Metabolomics of Critical Illness: Future Directions 380
References 381

Abstract

Critical illness is associated with dramatic changes in metabolism driven by immune, endocrine, and adrenergic mediators. These changes involve early activation of catabolic processes leading to increased energetic substrate availability; later on, they are followed by a hypometabolic phase characterized by

A. E. Pacheco-Navarro · A. J. Rogers (✉)
Division of Pulmonary, Allergy and Critical Care Medicine, Stanford University School of Medicine, Stanford, CA, USA
e-mail: ajrogers@stanford.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
V. Ghini et al. (eds.), *Metabolomics and Its Impact on Health and Diseases*,
Handbook of Experimental Pharmacology 277, https://doi.org/10.1007/164_2022_622

deranged mitochondrial function. In sepsis and ARDS, these rapid clinical changes are reflected in metabolomic profiles of plasma and other fluids, suggesting that metabolomics could one day be used to assist in the diagnosis and prognostication of these syndromes. Some metabolites, such as lactate, are already in clinical use and define patients with septic shock, a high-mortality subtype of sepsis. Larger-scale metabolomic profiling may ultimately offer a tool to identify subgroups of critically ill patients who may respond to therapy, but further work is needed before this type of precision medicine is readily employed in the clinical setting.

Keywords

Acute respiratory distress syndrome (ARDS) · Metabolism · Metabolomics · Prognostication · Sepsis

1 Introduction: A Potential Role for Metabolomics in Understanding Critical Illness

Each year, over four million critically ill Americans are admitted into intensive care units (ICUs) around the country, a number that has only increased since the start of the COVID-19 pandemic (ICU Outcomes 2022). Despite advances in medicine, ICU mortality remains high, ranging between 8 and 20%, with approximately 500,000 American deaths per year. In particular, sepsis – a systemic inflammatory illness caused by infection with a dysregulated host immune response – and the acute respiratory distress syndrome (ARDS) – a syndrome of respiratory failure with a diverse range of etiologies – are particularly challenging illnesses in the ICU, with mortality upward of 40% (Fleischmann et al. 2016; Gaieski et al. 2013; Rudd et al. 2020).

Part of the challenge in treating these syndromes lies in their significant biological heterogeneity. The majority of clinical trials in sepsis and ARDS have been negative, and available treatment remains largely supportive and directed at treating underlying infections, while not specifically addressing the inflammatory state. Recent re-analyses of a number of negative trials have revealed that subgroups of patients appear to respond to the intended therapy (Calfee et al. 2014; Puskarich et al. 2021), underscoring the heterogeneity of these syndromes and the need for better diagnostics in identifying high-risk patients and matching them with potential treatments.

Given the dramatic and rapid metabolic alterations that are observed during critical illness, metabolomics offers a powerful potential tool for better identifying subphenotypes of critically ill patients who may differentially respond to treatments and also identifying novel treatment pathways.

2 Metabolic Alterations in Critical Illness

In health, energetic needs are largely met by nutrient intake through food, rather than catabolism of the body's macromolecules. ATP, the main source of cellular energy, is derived by aerobic metabolism, and anabolic and catabolic processes are in balance.

In critical illness, however, dramatic changes in the endocrine and autonomic nervous system alter these metabolic processes, shifting them toward a catabolic state in which the body's carbohydrates, lipids, and protein stores are used to meet increased energetic needs. This imbalance is often further driven by reduced caloric intake, as critically ill patients frequently suffer from symptoms like encephalopathy and anorexia that curb eating. Clinically, such catabolism leads to skeletal muscle wasting, weakness, and prolonged recovery (Casaer and Van den Berghe 2014).

3 Endocrine Drivers of Altered Metabolism

Initially, acute stress such as that caused by sepsis and other forms of critical illness leads to an inflammatory cascade which includes the release of cytokines, including IL-1, IL-2, and IL-6 that can activate the hypothalamic-pituitary-adrenal axis. This can occur through direct cytokine production by glial cells in the brain (Khardori and Castillo 2012; Englert and Rogers 2016), and through free diffusion into the pituitary, which is outside the blood brain barrier. This cytokine-derived activation results in the release of corticotropin (ACTH) from the pituitary, which stimulates production and secretion of cortisol from the adrenal glands (Englert and Rogers 2016; Cooper 2003). In addition to its increased production by the adrenal gland, decreases in corticosteroid binding globulin during critical illness and impaired hepatic metabolism of the hormone contribute to increased levels of free, biologically active cortisol (Wasyluk et al. 2021; Marik et al. 2008).

Cortisol acts as the primary endocrine mediator in sepsis, leading to a variety of changes in metabolism, immune function, and cardiovascular function critical to maintaining homeostasis (Fig. 1). Metabolically, it leads to increased blood glucose through a variety of mechanisms: by activation of enzymes involved in gluconeogenesis, increasing insulin resistance, and decreasing glucose uptake by skeletal muscle (Marik et al. 2008). It activates lipolysis in adipose tissue, leading the release of free fatty acids (Marik et al. 2008).

In addition to increasing energy availability by increasing glucose and fatty acid concentrations, cortisol has a major role in increasing blood pressure. In vascular smooth muscle, it increases transcription of surface receptors for catecholamines and angiotensin II, resulting in increased sensitivity to these vasoactive hormones. Cortisol also plays a complex role in regulating and abrogating the immune response, acting in a negative feedback loop that both reduces the number of circulating immune cells and decreases their production of cytokines.

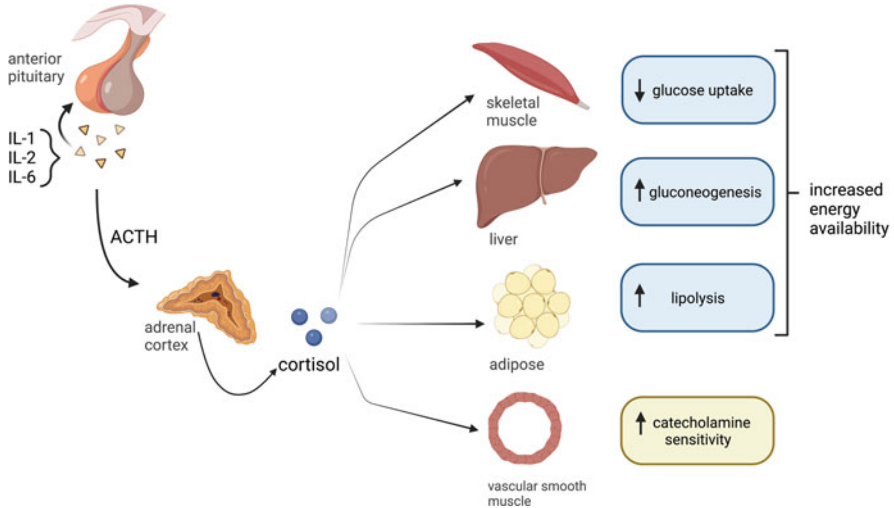


Fig. 1 Direct metabolic effects of cortisol. Figure created with biorender.com

While increased cortisol synthesis and secretion is part of the body's typical response to stress, this response can be attenuated or halted in a variety of conditions in the setting of critical illness. Head injuries, central nervous system depressants, and pituitary infarction can inhibit the CRH and corticotropin driven portion of the hypothalamic-pituitary-adrenal axis (Cooper 2003). At the level of the adrenal glands, drugs such as the induction agent etomidate – frequently used during intubation – can impair cortisol synthesis, and adrenal hemorrhage, though rare, can occur in septic patients with coagulopathies (Cooper 2003). More commonly, chronic intake of exogenous corticosteroids – e.g., in patients with rheumatologic disease – can lead to adrenal atrophy and a sluggish or absent response to corticotropin.

While absolute adrenal insufficiency is rare in critically ill patients, studies have shown that relative adrenal insufficiency – as defined by inadequate increase in cortisol level in response to a corticotropin challenge – is common in patients with septic shock (Annane et al. 2002). In 2007, a consensus statement between the Society of Critical Care Medicine and the European Society of Intensive Care Medicine first coined the term "CIRCI" (critical illness-related corticosteroid insufficiency) to recognize that many critically ill patients have an impaired hypothalamic-pituitary-adrenal response.

Despite this, it is unclear whether such patients benefit from hormone replacement (Annane et al. 2002; Sprung et al. 2008), as a number of large randomized controlled trials studying the use of steroids in sepsis have had conflicting results (Annane et al. 2002, 2018; Sprung et al. 2008; Venkatesh et al. 2018). At the

minimum, however, these studies suggest that steroid administration in critically ill patients is safe and should be considered, particularly for patients with refractory shock. As in the sepsis literature, the benefits of steroids are unclear in ARDS, but recent high-quality studies (Villar et al. 2020) – as well as the evidence in severe COVID-19 (RECOVERY Collaborative Group et al. 2021; Tomazini et al. 2020) – suggest that they may be beneficial.

4 Adrenergic Mediators and Their Changes in Critical Illness

The sympathetic nervous system is rapidly activated in the setting of critical illness. These stress signals can occur at the level of the peripheral nerves (e.g., in the case of trauma), by chemoreceptors (in the case of hypoxemia or hypercapnia), or by baroreceptors (in the setting of shock). These events all result in the release of norepinephrine from post-ganglionic neurons, and secretion of epinephrine and norepinephrine from chromaffin cells in the adrenal medulla (Preiser et al. 2014).

Along with exogenously delivered catecholamines that may be administered therapeutically, endogenous dopamine, norepinephrine, and epinephrine play major roles in activating catabolic pathways and increasing nutrient availability. Like cortisol, these hormones work to increase plasma glucose concentrations by upregulating hepatic gluconeogenesis and glycogenolysis and increasing insulin resistance.

5 Inflammatory Mediators (Cytokines) and Their Contribution to Metabolism

In addition to the endocrine and sympathetic nervous systems, cytokines produced by immune cells play important roles in mediating metabolic changes during critical illness. Most heavily implicated are tumor necrosis factor α – previously known in the cancer literature as "cachectin" for its upregulation of lipolysis – interleukin-1β (IL-1β), and interleukin-6 (Il-6). TNFα is a potent inducer of the other two, and together, they work to increase insulin resistance, proteolysis, and lipolysis.

6 Metabolic Phases During Critical Illness

Sepsis and other forms of critical illness are characterized by an early stress response, as described above, associated with increased catabolism and energy availability. This early phase is often followed by a period of decreased metabolic activity and organ dysfunction (Singer et al. 2004; Pool et al. 2018). Initially, the organ dysfunction and elevations in lactate that are frequently seen in sepsis were thought to result from impaired tissue perfusion, but studies have since shown that although malperfusion may play a role in the early phase of critical illness, tissue oxygen

delivery is adequate in the later phases of sepsis (Sair et al. 2001). Instead, cellular changes in metabolism – decreased oxidative phosphorylation and greater reliance on glycolytic pathways – termed *cytopathic hypoxia* – likely underly this organ dysfunction. These changes are hypothesized to be adaptive mechanisms to preserve energy for critical functions, prevent DNA damage, and limit additional injury.

In the heart, for example, sepsis and critical illness can result in reduced cardiac myocyte contractility through a process called myocardial hibernation, which preserves myocyte ATP levels at the cost of reduced function. In animal models, this change in contractility has been linked to inhibition of the oxidative phosphorylation enzyme cytochrome oxygenase, and this phenotype can be rescued by delivery of exogenous cytochrome c (Piel et al. 2007). Similar mitochondrial changes have also been implicated in animal models of sepsis-induced dysfunction of the lung, kidney, and liver (Pool et al. 2018). In the kidney, the presence of pro-inflammatory cytokines or lipopolysaccharide from bacteria leads to downregulation of energy-intensive sodium and chloride channels (Schmidt et al. 2007). Despite these observations, the full spectrum of metabolic changes during the phases of critical illness is not known. Metabolomics provides a potential avenue through which changes in multiple metabolic pathways can be better described.

7 Autophagy and Mitophagy

The mitochondrial dysfunction and electron transport chain inhibition described above likely result in increased *autophagy*, a cellular housekeeping mechanism that results in the lysosomal degradation of damaged and dysfunctional organelles and proteins. In brief, autophagy is upregulated by cellular/organelle damage – e.g., the mitochondrial membrane depolarization that might occur due to uncoupled respiration from an impaired electron transport chain – and inhibited by increased nutrients (Vanhorebeek et al. 2011).

Despite being a mechanism for intracellular protein catabolism, autophagy has been shown to play an important role in maintaining the health of myocytes in skeletal muscle, and inhibition of autophagy has been linked to accelerated muscle loss during fasting and muscle denervation (Masiero et al. 2009). For critically ill patients, who are in a persistent catabolic state and are at very high risk of muscle wasting and prolonged weakness (termed *critical illness myopathy*), the relationship between fasting and autophagy has been of increasing interest. In the 4,600+ patient EPaNIC trial (Early Parenteral Nutrition Completing Enteral Nutrition in Adult Critically Ill Patients), the effect of initiating parenteral nutrition (PN) on an early time scale (within 48 h of ICU admission) versus a late one (within 8 days) was compared. Late parenteral nutrition initiation was associated with faster recovery (as measured by ICU and hospital discharge) and fewer complications (Casaer et al. 2011). Significantly, in a prospectively planned subanalysis of the EPaNIC trial, in which a subset of patients underwent skeletal muscle biopsies 8 days after randomization and sequential strength testing through 100 days, patients randomized to the late PN group were found to have faster recovery of weakness and greater evidence

of autophagosome formation (Hermans et al. 2013). Taken together, this finding suggests that a tempered approach to early nutrition in the critically ill patient may allow for more autophagy and faster recovery of skeletal muscle strength.

8 Nutrient Changes During Critical Illness

Critical illness leads to significant changes in the metabolism of carbohydrates, protein, and lipids. Most notably, increased glycogenolysis, gluconeogenesis, and reduced sensitivity to insulin lead to increased plasma glucose concentrations through the endocrine, adrenergic, and inflammatory mechanisms described above (Table 1).

Catabolism of proteins and lipids is also upregulated, particularly in early sepsis. Amino acids generated in proteolytic processes are shuttled to the liver for synthesis of acute phase reactants. Protein breakdown and a net negative nitrogen balance lead to the wasting of skeletal muscles, which will contribute to deconditioning and prolong patients' recoveries. In adipose tissue, increased lipolysis leads to increased serum triglycerides and free fatty acids and decreased circulating lipoproteins.

Micronutrient levels also fluctuate during critical illness. Of these, changes in selenium and zinc levels in sepsis have been particularly well-described. Selenium and zinc are both trace elements with important anti-inflammatory and antioxidant properties, and levels of both minerals have been found to be lower in septic patients than in healthy controls (Allingstrup and Afshari 2015; Besecker et al. 2011). Supplementation of these micronutrients in the critically ill has been investigated with promising initial results (Manzanares et al. 2012), but due to small study sizes, further work is needed before their use in the ICU becomes widespread.

Table 1 Summary of metabolic changes in sepsis

Physiologic change in sepsis	Metabolic impact
↑ Gluconeogenesis, glycolysis, glycogenolysis	Hyperglycemia
↑ Protein catabolism	Altered circulating amino acids
↑ Lipolysis	↑ Triglycerides, ↓ lipoproteins
↓ Micronutrients	↑ Oxidative stress
↑ Neuroendocrine activation	↑ Catecholamines, ↑ counterregulatory hormones
↑ Cortisol	Hyperglycemia
↑ Catecholamine release	↑ Gluconeogenesis, ↑ glycolysis
↑ Cytokine release	Hyperglycemia, insulin resistance
Impaired oxygen utilization	↑ Reactive oxygen species

Adapted from Englert and Rogers (Englert and Rogers 2016) with their permission

8.1 Lactate: The Original Metabolomic Biomarker

Plasma lactate is one of the most validated and frequently used biomarkers in the intensive care unit. Lactate is a by-product of anerobic metabolic pathways, in which glucose is broken down into two pyruvate molecules and then fermented to 2 lactate molecules in a net reaction that produces 2 molecules of adenosine triphosphate (ATP) and regenerates the NAD+ cofactor used in glycolysis (Fig. 2). Accordingly, rising blood lactate levels can indicate inadequate tissue perfusion and oxygenation or a perfusion-independent shift toward anerobic pathways as can occur in malignancy (the Warburg effect) or in the later phases of sepsis.

In the ICU, lactate can be used for prognostication and also to guide resuscitation during critical illness. From a risk stratification perspective, hospitalized patients with infection-related admissions had a significantly higher risk of death if their serum lactate was >4 mmol/L (as compared to those with normal concentrations), and even modestly elevated lactates of 2.5–4 mmol were associated with increased mortality (Shapiro et al. 2005). Similarly, in a study of septic patients in the emergency department, intermediate and high lactate levels (2.5–3.9 mmol/L and >4 mmol/L) were associated with increased mortality even in patients with initially normal blood pressures and without evidence of organ dysfunction (Mikkelsen et al. 2009).

Lactate levels have also been applied to the goal-directed paradigm of sepsis management and have been used to guide resuscitation, though its role in this context has been less well-established. In the EMShockNet trial which compared the use of ScvO2 to serial lactates in the goal-directed resuscitation of patients

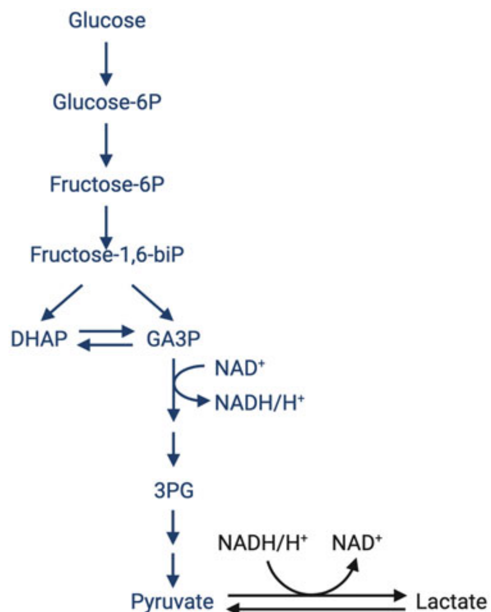


Fig. 2 Glycolysis and lactate fermentation. Adapted from "Warburg Effect," from Biorender.com (2022). Retrieved from https://app.biorender.com/biorender-templates and created in biorender.com

(in which both groups also had MAP and CVP goals), there was a trend toward reduced mortality in the lactate group, but this difference was not statistically significant (Jones et al. 2010). Similarly, in another trial comparing serial lactate-guided therapy to traditional early goal-directed therapy as described by Rivers et al., mortality trended lower in the lactate group, but was just outside of the predefined alpha for statistical significance ($p = 0.067$) (Jansen et al. 2010).

9 Metabolomics in Sepsis

Sepsis, or life-threatening organ dysfunction in response to infection (Singer et al. 2016), is one of the leading causes of death worldwide, accounting for an estimated five million deaths each year and many more hospitalizations (Fleischmann et al. 2016). The current treatment strategy is based on the delivery of early, broad-spectrum antibiotics and is otherwise supportive. Because of the disease's heterogeneity and the limited treatments available, precision medicine approaches – in which subgroups of patients with different mechanistic drivers are identified, leading to different treatment approaches – are appealing; these strategies have been effectively applied to a range of diseases, from cancer to asthma. While broader plasma metabolomic profiling studies have yet to impact care of septic patients to date, we will discuss overarching trends in the field with selected examples from larger studies and their clinical implications below.

9.1 Diagnosis of Sepsis

Noninfectious inflammatory responses – as might occur in pancreatitis or trauma – can resemble sepsis, delaying the diagnosis until culture data are finalized and potentially contributing to the administration of unnecessary antibiotics and increasing the development of drug-resistant bacteria. Metabolomics has been used to differentiate sepsis from sterile inflammation (Mickiewicz et al. 2013, 2014; Langley et al. 2014). In these studies, plasma samples from septic patients (or in one study, infected primates) were compared to those of critically ill uninfected controls with evidence of systemic inflammation, key metabolic differences were identified, then validated in an independent cohort of critically ill patients.

In one such study, Mickiewicz et al. (2013) used supervised analysis to investigate 58 metabolites in the serum of 60 children with early septic shock and compared profiles to those of 40 noninfected pediatric ICU patients with systemic inflammatory response syndrome (SIRS) and 40 healthy children. They used principal component analysis and partial least squares-discriminant analysis to identify 18 metabolites that differed in those who survived from those who did not. Interestingly, the responses of infants and toddlers differed from those of school-aged children, suggesting that the metabolic changes associated with sepsis may differ depending on life stage. Three metabolites (2-hydroxybutyrate, 2-hydroxyisovalerate, and lactate) involved in fat breakdown and anerobic

metabolism were elevated in septic patients as compared to controls regardless of age. Other studies reinforced these differences in lipid metabolism (Langley et al. 2013; Rogers et al. 2021) and also found differences in amino acid metabolism, mitochondrial metabolism, and the citric acid cycle and pentose phosphate pathway in septic patients versus noninfected inflamed controls (Eckerle et al. 2017).

More recently, Rydzak et al. used a metabolomic approach for rapid identification of blood stream infection pathogens and their antimicrobial susceptibility patterns (Rydzak et al. 2022), a highly valuable tool in sepsis, where blood stream infections contribute to significant morbidity and mortality. They used a metabolic preference assay that used the patterns of consumed and excreted metabolites of ex vivo microbial cultures to identify pathogens and their antibiotic susceptibilities, monitoring for changes in metabolite production in the presence of toxic antibiotics. They found that just seven metabolites were sufficient to differentiate between the seven most prevalent pathogens responsible for blood stream infections and identify their antimicrobial susceptibilities in half the time of traditional microbial identification and antimicrobial susceptibility testing protocols.

9.2 Prognostication in Sepsis

Sepsis has a wide range of outcomes, from full recovery to long-term disability or death (Singer et al. 2016). As such, it is not surprising that many studies in critical care metabolomics are focused on prognostication. In these studies, septic patients with worse outcomes (death or multiorgan failure) are compared to survivors or individuals with less organ failure.

Using targeted metabolomic profiling of more than 300 metabolites, Langley et al. evaluated the plasma metabolomes and proteomes in two independent cohorts (Langley et al. 2014). The group used 150 adult patients from the Community Acquired Pneumonia and Sepsis Outcome Diagnostics (CAPSOD) study, which enrolled patients who presented to the emergency department of three US hospitals with symptoms of sepsis (Langley et al. 2013) as a discovery cohort, and used patients with sepsis from the Brigham and Women's Hospital Registry of Critical Illness (RoCI) cohort as a validation set. They found that the metabolic signatures of survivors differed markedly from those that died. In particular, levels of proteins involved in fatty acid transport and β-oxidation, gluconeogenesis, and the citric acid cycle were significantly higher in those that died. The group used Support Vector Machine analysis to create an algorithm using two clinical features and five metabolite levels to predict survival in sepsis, using the RoCI data as a validation set. Though the RoCI cohort had higher mortality and more cancer as compared to the discovery cohort, the metabolic changes identified and the sepsis prognostication model were validated in the independent cohort.

Rogers et al. also used the CAPSOD and RoCI cohorts to perform targeted metabolic profiling on sepsis survivors and nonsurvivors (Rogers et al. 2014). Although the Langley and Rogers groups identified many of the same metabolites that differed in survivors and nonsurvivors, the Bayesian model developed by the

Rogers group using RoCI as the training cohort was very different, using seven completely different metabolites than those used in the Langley algorithm. This may reflect a redundancy of metabolites found in the same metabolic pathways, but also illustrates the evolving nature of analytic methodology in the field, in which the optimal statistical methods for dealing with highly correlated data are not yet known.

9.3 Future Directions: Using Metabolomics to Guide Treatment

One of the most promising applications for the use of metabolomics in critical care is in the identification of patients who might differentially respond to therapies. Sepsis treatment is currently focused on early antimicrobial administration and is otherwise supportive, as no mechanism-targeted therapies are approved for use in the syndrome. The failure of many studies to identify effective, mechanism-based treatments has been attributed to sepsis's significant heterogeneity, which may mask individual or subgroup responses to therapies in clinical trials.

Along these lines, Puskarich et al. used metabolomics to further phenotype patients in a reanalysis of a phase II study of L-carnitine infusion for the treatment of septic shock (Puskarich et al. 2015). The group performed metabolomics on existing serum samples from septic patients in the study and found that three metabolites involved in the synthesis and depredation of ketones differed in the pre-intervention samples of L-carnitine-treated survivors and nonsurvivors. In later work, they went on to identify acetylcarnitine as a particularly powerful biomarker for predicting L-carnitine mortality benefit (Puskarich et al. 2021). While the findings of retrospective reanalysis of a randomized controlled trials are likely insufficient to change clinical practice, the hope is that metabolomics can be used in the future to identify potential drug responders prior to their inclusion into prospective clinical trials.

10 Metabolomics Studies in ARDS

Like sepsis, ARDS is a highly heterogeneous and clinically important disease, defined by acute respiratory failure in the setting of a risk factor (such as sepsis, pneumonia, trauma, aspiration), bilateral infiltrates on imaging and low levels of blood oxygen (Ferguson et al. 2012; Thompson et al. 2017). It is estimated to affect >100,000 Americans per year with a mortality rate over 25% despite clinical improvements in recent decades (Bellani et al. 2016; Rubenfeld et al. 2005); those numbers predate the COVID epidemic, with ARDS and hypoxemic respiratory failure leading to the deaths of millions worldwide.

The heterogeneity of ARDS has no doubt contributed to a lack of effective therapies. ARDS mortality has fallen because of improvements in approaches to mechanical ventilation which decrease the pressure and volumes delivered to the lung, preventing further lung injury. But despite more than 50 years of research and substantial improvements in our understanding of the pathogenic underpinnings of

lung injury and inflammation, not a single drug had been approved to treat ARDS prior to COVID-19 (Matthay et al. 2019). In COVID-19, numerous drug trials have proved effective for the syndrome, likely because of increased homogeneity in the patient population: all cases share an underlying etiology of SARS-CoV2 infection, a typical time course of disease, and a substantial inflammatory component. Multiple anti-inflammatory drug therapies are now available for COVID ARDS patients (RECOVERY Collaborative Group et al. 2021; Investigators et al. 2021; Kalil et al. 2021).

Metabolomics studies in ARDS are compelling because of the dynamic metabolomic changes in the blood and lungs, with the potential for providing a real-time snapshot of disease and identifying high-risk subjects who could be targeted for trials (Millet, "Prospective metabolomic studies in precision medicine. The AKRIBEA project."; Turano, "NMR-based metabolomics to evaluate individual response to treatments"). Multiple studies have examined metabolomics of ARDS; major studies are summarized below and in a thorough recent review (Metwaly and Winston 2020).

10.1 Blood Metabolomics of ARDS

Several groups have performed metabolomic profiling of blood in ARDS patients and controls. As noted, there is substantial variability between studies, including (1) size of cohort, (2) how the control cohort is defined (healthy vs critically ill but without ARDS), and (3) the metabolomic profiling methodology used. Given these disparate methodologies and relatively small cohort sample sizes, it is perhaps unsurprising that the metabolites and potential pathways identified vary substantially.

Metwaly et al. examined 108 patients with ARDS, matched by age and gender with 27 mechanically ventilated controls (Metwaly and Winston 2020). They identified numerous pathways that differed in ARDS vs controls that were significant in both a training and testing cohort, with a particular focus on serine-glycine metabolism. As in prior studies, the mechanically ventilated controls were systemically much less sick than subjects with ARDS (e.g., less frequently required vasopressors, less kidney dysfunction, and markedly lower mortality).

Rogers et al. examined a large cohort of 78 ARDS cases and compared them to 75 critically ill controls with sepsis (Rogers et al. 2021). While there were numerous individual metabolites that differed, partial least squares-discriminant analysis (PLS-DA) could not separate ARDS from critically ill controls, who frequently had concomitant organ dysfunction (e.g., renal failure, liver failure, or GI malperfusion). In fact, in this cohort, many of the previously identified "ARDS metabolites" were associated with the Simplified Acute Physiology (SAPS) score, which is predictive of general ICU mortality, and were not ARDS-specific. This work suggests that many of the metabolic changes previously identified in ARDS may reflect systemic illness and inflammation rather than ARDS per se. As these

studies demonstrate, how to adjust for concomitant systemic organ dysfunction is an active area of investigation in the metabolomics of ARDS.

10.2 Using Blood Metabolomics to Identify High-Risk Subsets of ARDS Cohorts

Some of the most high-profile advances in ARDS pathogenesis in the past decade have involved identifying high-risk subsets of patients who are at increased risk of bad outcomes. Most of these risk markers have involved cytokines and proteins, for example angiopoietin 2 (Calfee et al. 2012; Zinter et al. 2016), IL-18 (Rogers et al. 2019; Dolinay et al. 2012), and a hypoinflammatory/hyperinflammatory latent class analysis (LCA) designation which can be assigned using a parsimonious model of 3 blood biomarkers (IL-6, TNF receptor alpha, and bicarbonate) (Calfee et al. 2014; Sinha et al. 2020). The latent class hyperinflammatory subset has been particularly well-studied and shown to be present in more than six ARDS populations (both clinical trials and clinical cohorts) and associated with differential treatment response in numerous ARDS clinical trials, including approach to positive end-expiratory pressure (PEEP) level, fluid management, and statin therapy (Calfee et al. 2014, 2018; Famous et al. 2017).

Fewer studies have focused on using metabolomics to identify high-risk subsets of ARDS. Metwaly et al. tested whether orthogonal projections to latent structures discriminant analysis (OPLS-DA) could differentiate previously identified subsets of ARDS, including direct pulmonary risk factor for developing ARDS (e.g., pneumonia) vs indirect risk factors (e.g., pancreatitis or non-pulmonary sepsis), and hyperinflammatory subphenotype using latent class (Metwaly et al. 2021). Indeed, using metabolites identified via high-VIP score for OPLS-DA, these subsets could be separated in both testing and training sets; metabolomics outperformed protein analyses in separating the direct and indirect classes.

The Rogers et al. study of 197 patients included 75 patients with ARDS (Rogers et al. 2021). As discussed above, PLS-DA was not able to differentiate patients with sepsis from ARDS. However, unbiased metabolic clustering of the entire cohort did reveal three subsets of patients, with differential metabolites marked by lipid classes driving the subclasses, with markedly different risks of mortality, regardless of ARDS status.

10.3 Pulmonary Specific Samples for Metabolomic Profiling

Perhaps one of the most promising aspects of metabolomic profiling is the ability to examine trace amounts of a lung-specific sample which could be captured without need for a tissue biopsy to use for metabolic profiling (e.g., bronchoalveolar lavage fluid, expired breath, or free-flowing pulmonary edema fluid). Such non-blood samples may overcome the non-specific metabolic changes that occur with critical illness and allow for a more lung injury-specific signature to emerge.

To date, lung-specific metabolomics studies have been limited by similar challenges that have been seen in blood cohorts: (1) fairly small sample size, particularly given the inherent heterogeneity of ARDS, (2) varying control population, and (3) differing lung sample fluids. Not surprisingly, given such disparities in sampling, there are no common metabolites emerging that consistently differentiate ARDS from non-ARDS fluid.

Only two studies to date examined the unbiased identification of high-risk subsets within ARDS using non-blood samples. In a very small study by Rogers et al. investigating the metabolome of pulmonary edema fluid in ARDS versus congestive heart failure, ARDS edema could not be reliably distinguished from cardiogenic edema via PLS-DA, but hierarchical clustering did separate a metabolically distinct subset of one-third of the sample which was at markedly increased risk of mortality (Rogers et al. 2017).

Viswan and colleagues performed a large study, examining both blood and BAL fluid in >150 ARDS patients to examine metabolites associated with three definitions of high-risk subsets of ARDS: those defined by (1) ARDS severity by PaO2:FIO2 ratio, (2) direct vs indirect ARDS, and (3) survivors and nonsurvivors (Viswan et al. 2019). They did identify numerous associations between subsets, including overlap between the BAL and blood metabolites identified in high-risk subsets, though some methodology is unclear (Rogers et al. 2021; Investigators et al. 2021).

ARDS is a clinically-defined syndrome which carries substantial risk of mortality. It is marked by both clinical and temporal heterogeneity, suggesting that a focus on metabolomics of blood and lung biomarkers could be fruitful. Though numerous metabolomics studies have been published in the last 2 decades, more work in large cohorts is needed to identify reproducible signatures that provide targetable metabolic subgroups of ARDS. Similarly, to what extent readily available plasma samples provide an adequate metabolic sample for lung biology, versus reflect systemic illness regardless of ARDS status requires ongoing study.

11 Metabolomics of Critical Illness: Future Directions

Both sepsis and ARDS are heterogenous and life-threatening clinical syndromes in which a dysregulated immune response triggers dramatic changes in metabolism. Identifying targeted and effective treatments for these syndromes remains a challenge. The current therapeutic paradigm hinges on treating underlying infections and otherwise providing supportive care; no effective therapies have been found to treat the metabolic and inflammatory derangements that drive these syndromes.

Metabolomics offers a potential tool for better identifying the biological mechanisms underlying these syndromes and has already shown promise in prognostication of both sepsis and ARDS. Further work will be needed to more cleanly differentiate the lung-specific injury that occurs in ARDS from the frequently overlapping changes seen in sepsis. Ultimately, metabolomics offers the opportunity to uncover important biologic pathways that may serve as therapeutic targets and

identify subgroups of patients in both sepsis and ARDS who may respond differentially to current and future therapeutics.

References

Allingstrup M, Afshari A (2015) Selenium supplementation for critically ill adults. Cochrane Database Syst Rev (7):CD003703. https://doi.org/10.1002/14651858.CD003703.pub3

Annane D, Sébille V, Charpentier C et al (2002) Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. JAMA 288(7):862–871. https://doi.org/10.1001/jama.288.7.862

Annane D, Renault A, Brun-Buisson C et al (2018) Hydrocortisone plus fludrocortisone for adults with septic shock. N Engl J Med 378(9):809–818. https://doi.org/10.1056/NEJMoa1705716

Bellani G, Laffey JG, Pham T et al (2016) Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. JAMA 315(8): 788–800. https://doi.org/10.1001/jama.2016.0291

Besecker BY, Exline MC, Hollyfield J et al (2011) A comparison of zinc metabolism, inflammation, and disease severity in critically ill infected and noninfected adults early after intensive care unit admission. Am J Clin Nutr 93(6):1356–1364. https://doi.org/10.3945/ajcn.110.008417

Calfee CS, Gallagher D, Abbott J, Thompson BT, Matthay MA, Network NA (2012) Plasma angiopoietin-2 in clinical acute lung injury: prognostic and pathogenetic significance. Crit Care Med 40(6):1731–1737. https://doi.org/10.1097/CCM.0b013e3182451c87

Calfee CS, Delucchi K, Parsons PE et al (2014) Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. Lancet Respir Med 2(8):611–620. https://doi.org/10.1016/S2213-2600(14)70097-9

Calfee CS, Delucchi KL, Sinha P et al (2018) Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. Lancet Respir Med 6(9):691–698. https://doi.org/10.1016/S2213-2600(18)30177-2

Casaer MP, Van den Berghe G (2014) Nutrition in the acute phase of critical illness. N Engl J Med 370(25):2450–2451. https://doi.org/10.1056/NEJMc1404896

Casaer MP, Mesotten D, Hermans G et al (2011) Early versus late parenteral nutrition in critically ill adults. N Engl J Med 365(6):506–517. https://doi.org/10.1056/NEJMoa1102662

Cooper MS (2003) Corticosteroid Insufficiency in acutely Ill patients. N Engl J Med:8

Dolinay T, Kim YS, Howrylak J et al (2012) Inflammasome-regulated cytokines are critical mediators of acute lung injury. Am J Respir Crit Care Med 185(11):1225–1234. https://doi.org/10.1164/rccm.201201-0003OC

Eckerle M, Ambroggio L, Puskarich MA et al (2017) Metabolomics as a driver in advancing precision medicine in sepsis. Pharmacotherapy 37(9):1023–1032. https://doi.org/10.1002/phar.1974

Englert JA, Rogers AJ (2016) Metabolism, metabolomics, and nutritional support of patients with sepsis. Clin Chest Med 37(2):321–331. https://doi.org/10.1016/j.ccm.2016.01.011

Famous KR, Delucchi K, Ware LB et al (2017) Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. Am J Respir Crit Care Med 195(3):331–338. https://doi.org/10.1164/rccm.201603-0645OC

Ferguson ND, Fan E, Camporota L et al (2012) The Berlin definition of ARDS: an expanded rationale, justification, and supplementary material. Intensive Care Med 38(10):1573–1582. https://doi.org/10.1007/s00134-012-2682-1

Fleischmann C, Scherag A, Adhikari NKJ et al (2016) Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. Am J Respir Crit Care Med 193(3):259–272. https://doi.org/10.1164/rccm.201504-0781OC

Gaieski DF, Edwards JM, Kallan MJ, Carr BG (2013) Benchmarking the incidence and mortality of severe sepsis in the United States. Crit Care Med 41(5):1167–1174. https://doi.org/10.1097/CCM.0b013e31827c09f8

Hermans G, Casaer MP, Clerckx B et al (2013) Effect of tolerating macronutrient deficit on the development of intensive-care unit acquired weakness: a subanalysis of the EPaNIC trial. Lancet Respir Med 1(8):621–629. https://doi.org/10.1016/S2213-2600(13)70183-8

ICU Outcomes. Philip R. Lee Institute for Health Policy Studies Accessed 24 Apr 2022. https://healthpolicy.ucsf.edu/icu-outcomes

Investigators RC, Gordon AC, Mouncey PR et al (2021) Interleukin-6 receptor antagonists in critically ill patients with Covid-19. N Engl J Med 384(16):1491–1502. https://doi.org/10.1056/NEJMoa2100433

Jansen TC, van Bommel J, Schoonderbeek FJ et al (2010) Early lactate-guided therapy in intensive care unit patients: a multicenter, open-label, randomized controlled trial. Am J Respir Crit Care Med 182(6):752–761. https://doi.org/10.1164/rccm.200912-1918OC

Jones AE, Shapiro NI, Trzeciak S et al (2010) Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial. JAMA 303(8):739–746. https://doi.org/10.1001/jama.2010.158

Kalil AC, Patterson TF, Mehta AK et al (2021) Baricitinib plus Remdesivir for hospitalized adults with Covid-19. N Engl J Med 384(9):795–807. https://doi.org/10.1056/NEJMoa2031994

Khardori R, Castillo D (2012) Endocrine and metabolic changes during sepsis. Med Clin North Am 96(6):1095–1105. https://doi.org/10.1016/j.mcna.2012.09.005

Langley RJ, Tsalik EL, van Velkinburgh JC et al (2013) An integrated clinico-metabolomic model improves prediction of death in sepsis. Sci Transl Med 5(195):195ra95. https://doi.org/10.1126/scitranslmed.3005893

Langley RJ, Tipper JL, Bruse S et al (2014) Integrative "omic" analysis of experimental bacteremia identifies a metabolic signature that distinguishes human sepsis from systemic inflammatory response syndromes. Am J Respir Crit Care Med 190(4):445–455. https://doi.org/10.1164/rccm.201404-0624OC

Manzanares W, Dhaliwal R, Jiang X, Murch L, Heyland DK (2012) Antioxidant micronutrients in the critically ill: a systematic review and meta-analysis. Crit Care Lond Engl 16(2):R66. https://doi.org/10.1186/cc11316

Marik PE, Pastores SM, Annane D et al (2008) Recommendations for the diagnosis and management of corticosteroid insufficiency in critically ill adult patients: consensus statements from an international task force by the American College of Critical Care Medicine. Crit Care Med 36(6):1937–1949. https://doi.org/10.1097/CCM.0b013e31817603ba

Masiero E, Agatea L, Mammucari C et al (2009) Autophagy is required to maintain muscle mass. Cell Metab 10(6):507–515. https://doi.org/10.1016/j.cmet.2009.10.008

Matthay MA, Zemans RL, Zimmerman GA et al (2019) Acute respiratory distress syndrome. Nat Rev Primer 5(1):18. https://doi.org/10.1038/s41572-019-0069-0

Metwaly SM, Winston BW (2020) Systems biology ARDS research with a focus on metabolomics. Metabolites 10(5). https://doi.org/10.3390/metabo10050207

Metwaly S, Cote A, Donnelly SJ et al (2021) ARDS metabolic fingerprints: characterization, benchmarking, and potential mechanistic interpretation. Am J Physiol Lung Cell Mol Physiol 321(1):L79–L90. https://doi.org/10.1152/ajplung.00077.2021

Mickiewicz B, Vogel HJ, Wong HR, Winston BW (2013) Metabolomics as a novel approach for early diagnosis of pediatric septic shock and its mortality. Am J Respir Crit Care Med 187(9): 967–976. https://doi.org/10.1164/rccm.201209-1726OC

Mickiewicz B, Duggan GE, Winston BW et al (2014) Metabolic profiling of serum samples by 1H nuclear magnetic resonance spectroscopy as a potential diagnostic approach for septic shock. Crit Care Med 42(5):1140–1149. https://doi.org/10.1097/CCM.0000000000000142

Mikkelsen ME, Miltiades AN, Gaieski DF et al (2009) Serum lactate is associated with mortality in severe sepsis independent of organ failure and shock. Crit Care Med 37(5):1670–1677. https://doi.org/10.1097/CCM.0b013e31819fcf68

Piel DA, Gruber PJ, Weinheimer CJ et al (2007) Mitochondrial resuscitation with exogenous cytochrome c in the septic heart. Crit Care Med 35(9):2120–2127. https://doi.org/10.1097/01.ccm.0000278914.85340.fe

Pool R, Gomez H, Kellum JA (2018) Mechanisms of organ dysfunction in sepsis. Crit Care Clin 34(1):63–80. https://doi.org/10.1016/j.ccc.2017.08.003

Preiser JC, Ichai C, Orban JC, Groeneveld ABJ (2014) Metabolic response to the stress of critical illness. Br J Anaesth 113(6):945–954. https://doi.org/10.1093/bja/aeu187

Puskarich MA, Finkel MA, Karnovsky A et al (2015) Pharmacometabolomics of l-carnitine treatment response phenotypes in patients with septic shock. Ann Am Thorac Soc 12(1):46–56. https://doi.org/10.1513/AnnalsATS.201409-415OC

Puskarich MA, Jennaro TS, Gillies CE et al (2021) Pharmacometabolomics identifies candidate predictor metabolites of an L-carnitine treatment mortality benefit in septic shock. Clin Transl Sci. https://doi.org/10.1111/cts.13088

RECOVERY Collaborative Group, Horby P, Lim WS et al (2021) Dexamethasone in hospitalized patients with Covid-19. N Engl J Med 384(8):693–704. https://doi.org/10.1056/NEJMoa2021436

Rogers AJ, McGeachie M, Baron RM et al (2014) Metabolomic derangements are associated with mortality in critically ill adult patients. PLoS One 9(1):e87538. https://doi.org/10.1371/journal.pone.0087538

Rogers AJ, Contrepois K, Wu M et al (2017) Profiling of ARDS pulmonary edema fluid identifies a metabolically distinct subset. Am J Physiol Lung Cell Mol Physiol 312(5):L703–L709. https://doi.org/10.1152/ajplung.00438.2016

Rogers AJ, Guan J, Trtchounian A et al (2019) Association of elevated plasma interleukin-18 level with increased mortality in a clinical trial of statin treatment for acute respiratory distress syndrome. Crit Care Med 47(8):1089–1096. https://doi.org/10.1097/CCM.0000000000003816

Rogers AJ, Leligdowicz A, Contrepois K et al (2021) Plasma metabolites in early sepsis identify distinct clusters defined by plasma lipids. Crit Care Explor 3(8):e0478. https://doi.org/10.1097/CCE.0000000000000478

Rubenfeld GD, Caldwell E, Peabody E et al (2005) Incidence and outcomes of acute lung injury. N Engl J Med 353(16):1685–1693. https://doi.org/10.1056/NEJMoa050333

Rudd KE, Johnson SC, Agesa KM et al (2020) Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study. Lancet Lond Engl 395(10219):200–211. https://doi.org/10.1016/S0140-6736(19)32989-7

Rydzak T, Groves RA, Zhang R et al (2022) Metabolic preference assay for rapid diagnosis of bloodstream infections. Nat Commun 13:2332. https://doi.org/10.1038/s41467-022-30048-6

Sair M, Etherington PJ, Peter Winlove C, Evans TW (2001) Tissue oxygenation and perfusion in patients with systemic sepsis. Crit Care Med 29(7):1343–1349. https://doi.org/10.1097/00003246-200107000-00008

Schmidt C, Höcherl K, Schweda F, Bucher M (2007) Proinflammatory cytokines cause down-regulation of renal chloride entry pathways during sepsis. Crit Care Med 35(9):2110–2119. https://doi.org/10.1097/01.ccm.0000281447.22966.8b

Shapiro NI, Howell MD, Talmor D et al (2005) Serum lactate as a predictor of mortality in emergency department patients with infection. Ann Emerg Med 45(5):524–528. https://doi.org/10.1016/j.annemergmed.2004.12.006

Singer M, De Santis V, Vitale D, Jeffcoate W (2004) Multiorgan failure is an adaptive, endocrine-mediated, metabolic response to overwhelming systemic inflammation. Lancet 364(9433):545–548. https://doi.org/10.1016/S0140-6736(04)16815-3

Singer M, Deutschman CS, Seymour CW et al (2016) The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 315(8):801–810. https://doi.org/10.1001/jama.2016.0287

Sinha P, Delucchi KL, McAuley DF, O'Kane CM, Matthay MA, Calfee CS (2020) Development and validation of parsimonious algorithms to classify acute respiratory distress syndrome

phenotypes: a secondary analysis of randomised controlled trials. Lancet Respir Med 8(3): 247–257. https://doi.org/10.1016/S2213-2600(19)30369-8

Sprung CL, Annane D, Keh D et al (2008) Hydrocortisone therapy for patients with septic shock. N Engl J Med 358(2):111–124. https://doi.org/10.1056/NEJMoa071366

Thompson BT, Chambers RC, Liu KD (2017) Acute respiratory distress syndrome. N Engl J Med 377(6):562–572. https://doi.org/10.1056/NEJMra1608077

Tomazini BM, Maia IS, Cavalcanti AB et al (2020) Effect of dexamethasone on days alive and ventilator-free in patients with moderate or severe acute respiratory distress syndrome and COVID-19: the CoDEX randomized clinical trial. JAMA 324(13):1307–1316. https://doi.org/10.1001/jama.2020.17021

Vanhorebeek I, Gunst J, Derde S et al (2011) Insufficient activation of autophagy allows cellular damage to accumulate in critically ill patients. J Clin Endocrinol Metab 96(4):E633–E645. https://doi.org/10.1210/jc.2010-2563

Venkatesh B, Finfer S, Cohen J et al (2018) Adjunctive glucocorticoid therapy in patients with septic shock. N Engl J Med 378:797. https://doi.org/10.1056/NEJMoa1705835

Villar J, Ferrando C, Martínez D et al (2020) Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. Lancet Respir Med 8(3):267–276. https://doi.org/10.1016/S2213-2600(19)30417-5

Viswan A, Ghosh P, Gupta D, Azim A, Sinha N (2019) Distinct metabolic endotype mirroring acute respiratory distress syndrome (ARDS) subphenotype and its heterogeneous biology. Sci Rep 9(1):2108. https://doi.org/10.1038/s41598-019-39017-4

Wasyluk W, Wasyluk M, Zwolak A (2021) Sepsis as a pan-endocrine illness-endocrine disorders in septic patients. J Clin Med 10(10):2075. https://doi.org/10.3390/jcm10102075

Zinter MS, Spicer A, Orwoll BO et al (2016) Plasma angiopoietin-2 outperforms other markers of endothelial injury in prognosticating pediatric ARDS mortality. Am J Physiol Lung Cell Mol Physiol 310(3):L224–L231. https://doi.org/10.1152/ajplung.00336.2015