Clinical Metabolomics Applications in Genetic Diseases

Anas M. Abdel Rahman
Editor

Clinical Metabolomics Applications
in Genetic Diseases

Anas M. Abdel Rahman
Editor

Clinical Metabolomics Applications in Genetic Diseases

Editor
Anas M. Abdel Rahman
Metabolomics Section, Department of Clinical Genomics, Center for Genome Medicine
King Faisal Specialist Hospital & Research Center (KFSHRC)
Riyadh, Saudi Arabia

ISBN 978-981-99-5161-1   ISBN 978-981-99-5162-8 (eBook)
https://doi.org/10.1007/978-981-99-5162-8

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 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 Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

The Advanced Technology and Clinical Application in Metabolomics . . . 1
Anas M. Abdel Rahman

Mass Spectrometry-Based Metabolomics for the Clinical Laboratory . . . 17
Joshua A. Dubland

Metabolomics: A Pipeline for Biomarker Discovery in Genetic Diseases . 43
Lina A. Dahabiyeh and Refat M. Nimer

Bioinformatic Tools for Clinical Metabolomics 71
David S. Wishart

Untargeted Metabolomics in Newborn Screening 97
Joshua Manor and Sarah H. Elsea

Untargeted Metabolomics, Targeted Care: The Clinical Utilities of Bedside Metabolomics 117
Joshua Manor and Sarah H. Elsea

Metabolomics in the Study of Human Mitochondrial Diseases 147
Rajaa Sebaa, Mary-Ellen Harper, Ruqaiah Al-Tassan, Mohammed Al-Owain, and Anas M. Abdel Rahman

Metabolomics of Rare Endocrine, Genetic Disease: A Focus on the Pituitary Gland 173
Afshan Masood, Abeer Malkawi, Anas M. Abdel Rahman, and Mohamed Siaj

Metabolomics and Genetics of Rare Endocrine Disease: Adrenal, Parathyroid Glands, and Cystic Fibrosis 189
Afshan Masood, Abeer Malkawi, Mohamed Siaj, and Anas M. Abdel Rahman

Metabolomic Role in Personalized Medicine: An Update 207
Minnie Jacob and Anas M. Abdel Rahman

Lipidomic Profiling in Clinical Practice Using LC-MS 225
Núria Amigó Grau and Pablo Ortiz Betes

Bringing Human Serum Lipidomics to the Forefront of Clinical Practice: Two Clinical Diagnosis Success Stories 239
Núria Amigó Grau and Pablo Ortiz Betes

LC-MS-Based Population Metabolomics: A Mini-Review of Recent Studies and Challenges from Sample Collection to Data Processing 269
Myriam Mireault and Lekha Sleno

Metabolomics and Transcriptomic Approach to Understand the Pathophysiology of Interstitial Lung Disease 301
Sanjukta Dasgupta, Anindita Bhattacharya, Priyanka Choudhury, Nilanjana Ghosh, Tanisha Das, Sushmita Roychowdhury, Riddhiman Dhar, and Koel Chaudhury

Transferring Metabolomics to Portable Diagnostic Devices: Trending in Biosensors 327
Shimaa Eissa

The Advanced Technology and Clinical Application in Metabolomics

Anas M. Abdel Rahman

Abstract Metabolomics identifies and quantifies small molecules (metabolites) using high-throughput techniques. The biological system metabolome integrates metabolomics data in conjugation with metabolic pathways, including other omics datasets, to produce a network of endogenous metabolites (metabotype) associated with the phenotypes. Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry are the main analytical techniques in combination with some separation techniques such as capillary electrophoresis, ultra-high-pressure liquid chromatography, and gas chromatography. The drastic improvement in the detection sensitivity and accuracy of the analytical techniques has widened the covered metabolomics. The comprehensive coverage of metabolomics becomes more integrated with other omics datasets to understand the system-level phenotypic changes and provide insight into the mechanisms that underlie various physiological conditions and diseases. This chapter highlights analytical methods for clinical metabolomics research and personalized medicine. Several innovative clinical metabolomics projects have reached up to patient services are discussed in this chapter.

Keywords Metabolomics · Chromatography · Lipidomics · Biomarker discovery · Mass spectrometry (MS) · Nuclear magnetic resonance (NMR)

Abbreviations

4MOP	4-Methyl-2-oxopentanoic acid
BHBA	β-hydroxybutyric acid
BSTFA	N,O-bis(trimethylsilyl)trifluoroacetamide
CE	Capillary electrophoresis

A. M. Abdel Rahman (✉)
Metabolomics Section, Department of Clinical Genomics, Center for Genome Medicine, King Faisal Specialist Hospital & Research Center (KFSHRC), Riyadh, Saudi Arabia
e-mail: aabdelrahman46@kfshrc.edu.sa

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_1

CI	Chemical ionization
EHMN	Edinburgh human metabolic network
ESI	Electrospray ionization
FFPE	Formalin-fixed paraffin-embedded
GSEA	Gene set enrichment analysis
HIES	Hyper-IgE syndromes
HILIC	Hydrophilic interaction liquid chromatography
HRM	High-resolution metabolomics
ICR	Ion cyclotron resonance
LC-MS	Liquid chromatography-mass spectrometry
LGPC	Linoleoylglycerophosphocholine
LIT	Linear quadrupole ion trap
NAFLD	Nonalcoholic fatty liver disease
NMR	Nuclear magnetic resonance
QIT	Quadrupole ion trap
ROC	Receiver operating characteristic
SMPDB	Small Molecule Pathway Database
TOF	Time of flight

1 Introduction

The history of metabolomics started from the essential biochemical genetics' techniques used routinely for patient diagnostics. The first metabolomics study was conducted in 1984 by Jermey Nicholson using nuclear magnetic resonance (NMR) and in 1995 by liquid chromatography-mass spectrometry (LC-MS) by Gary Siuzdak at Scripps Research Institute [1, 2]. Metabolomics and lipidomics are the techniques for studying the end product of the genetic makeup of living cells. Both techniques cover small molecules with a size below 1500 Da. This class of molecules covers endogenous and exogenous molecules, including drugs, food additives, microbiome secretome, and environmental exposome, with more than 217,920 compounds based on the human Metabolome Database (HMDB 5.0) [3].

In the last couple of decades, along with the advancement of MS and NMR, metabolomics and lipidomics have drastically populated different fields, including clinical research and drug development. Multiple disease models have been used to study the unique metabolomics profiles such as primary cell lines, mouse, and human biological materials (e.g., serum, plasma, tissue) [4, 5]. The use of metabolomics in medicine extends chronic diseases such as chronic kidney disease (CKD) [6], diabetes [7, 8], and rare syndromes with single gene deficiencies such as DOCK8 deficiency causing hyper-IgE syndromes (HIES) [9, 10].

In clinical research, metabolomics has shown to be a great choice for studying the disease mechanism, diagnostic and prognostic biomarkers, and therapeutic targets. More than 95% of the available clinical tests in the medical laboratory, 85% of the known drugs, and 50% of the genetic diseases are based on small molecules,

making metabolomics a great translational tool to improve the patient's quality of life. The technology of studying metabolomics has drastically improved in the last couple of decades alongside the cheminformatics and bioinformatic tools. For instance, mass spectrometry has been developed to increase the analyzers' mass-resolving power and quantitative sensitivities by advancing the hardware, such as using ion funnels in quadrupole time-of-flight (QTOF) analyzers [11]. However, the pre-analytical, analytical, and post-analytical limitations associated with the instinct of metabolites still need to be considered in any metabolomics research to reproduce the findings.

This chapter introduces metabolomics and lipidomics as major analytical technologies and their applications in clinical research and personalized medicine.

2 Mass Spectrometry

Mass spectrometry separates molecules based on their mass-to-charge ratio (m/z) in the gas phase under low pressure. At a given energy, molecules with the lowest m/z pass the analyzer the fastest. The MS performance can be evaluated based on resolution, precision, accuracy, and sensitivity. The MS resolution, the ability of the analyzer to separate ions and calculate as $m/\Delta m$, is based mainly on the complexity of the mixture and the length of the analysis path. The mass precision is based on the isotope abundance measurement reproducibility, represented as the coefficient of variation (CV%) for multiple measurements of the same sample. The MS accuracy is more challenging than precision evaluation, where the analysis has to be throughout interlaboratory standards. The accuracy of MS is known as the proximity of the experimental measurement to the true and exact mass (measurement error). The minimum sample size to obtain the optimal mass accuracy and precision is known as the sensitivity of MS.

The separation techniques enhanced the analytical performance of the MS and the accuracy of molecular annotation by introducing analytes in groups based on their physicochemical properties and reducing the matrix effect. Liquid chromatography (LC), gas chromatography (GC), and capillary electrophoresis (CE) are the main separation techniques hyphenated to MS used in metabolomics and lipidomics. Each technique has performance capabilities based on target molecules and the separation approach, such as in LC reversed-phase (RP) chromatography [12–15]. LC and CE analyze polar nonvolatile molecules without pretreatment or derivatization. RP LC is a dominating technique for metabolomics profiling. However, for the highly polar compounds, ion-pairing reagents are used for the stationary phase hydrophobicity and enhance their column retention for better analysis. In high-resolution and untargeted metabolomics analysis, ion-pairing reagents are used in negative mode ionization. Once switched to positive, ion-pairing reagent will drastically develop ion suppression and ion source contamination. Hydrophilic interaction liquid chromatography (HILIC) was found as an alternative to ion-pairing chromatography, where the polar stationary phase combines organic and aqueous mobile phases [16–18].

Capillary electrophoresis has been used widely to analyze polar and charged metabolites, where the molecular separation is based on electrophoretic mobility (charge-to-size ratio). Although both CE and HILIC are suitable for polar and charged molecules, HILIC is more sensitive due to the column capacity, while CE's peak is more efficient. The potential use of CE-MS in metabolomics has been reviewed [13]. The limitations of using CE are attributed to the lack of standardization of the small specimen load (poor sensitivity) and migration time variability.

The metabolomics throughput and depth of coverage have drastically increased after adding another dimension of post-ionization gas-phase separation using ion mobility (IM) to MS analyzers. The IM separation is based on molecular size, where the confidence of metabolites annotation is increased by adding the collision cross-section (CCS) as a molecular descriptive [19, 20]. In addition, IM plays a crucial role in lipidomics by enhancing the lipids' complex separation, improving isomer resolution, and increasing confidence in molecular identification and characterization [21].

Combining multiple chromatographic and IM techniques with high-resolution mass spectrometry and a data extraction algorithm is known as high-resolution metabolomics (HRM) [22]. The physiochemical properties and abundance heterogeneity of the cellular metabolome make the analytical tools used to study their expression quite challenging. Any research group aims to expand the metabolomics coverage to reach the low abundant molecules by combining multiple chromatographic approaches in different detection modes. The author's laboratory experience uses an alternating strategy: two chromatographic systems (RP and HILIC) and two ionization polarities (positive and negative) to cover the maximum number of metabolites. Several classes of molecules can be overlapped but in different sensitivities (Fig. 1).

Electrospray ionization (ESI) is the most common technique in both LC-MS and CE-MS use in metabolomics research. Moreover, the ESI ionization is considered soft to prevent uncontrolled in-source fragmentation, which is ideal for biomolecular analysis. The retention variations to using unbuffered mobile phase and ion suppression are the main restraints in using ESI in omics applications, mainly metabolomics (more details are covered in chapter "Metabolomics: A Pipeline for Biomarker Discovery in Genetic Diseases").

GC-MS is one of the most efficient and reproducible metabolomics platforms to match the standard of commercial and "in-house" established libraries and databases. As an analytical tool, GC-MS is considered robust, excellent separation capable, selective, sensitive, and reproducible to cover a large group of metabolites. Electron ionization (EI) and chemical ionization (CI) are GC-MS's main ionization techniques, producing fragments and molecular ion spectra for the target molecule. However, EI is the most informative ionization technique compatible with metabolomics' most available libraries and databases. Low molecular weight (50–600 Da) and volatile compounds are the most likely compounds that can be analyzed using GC-MS. The polar, thermolabile, and nonvolatile compounds can be analyzed using GC-MS after using some derivatization reagents to collect an informative analytical

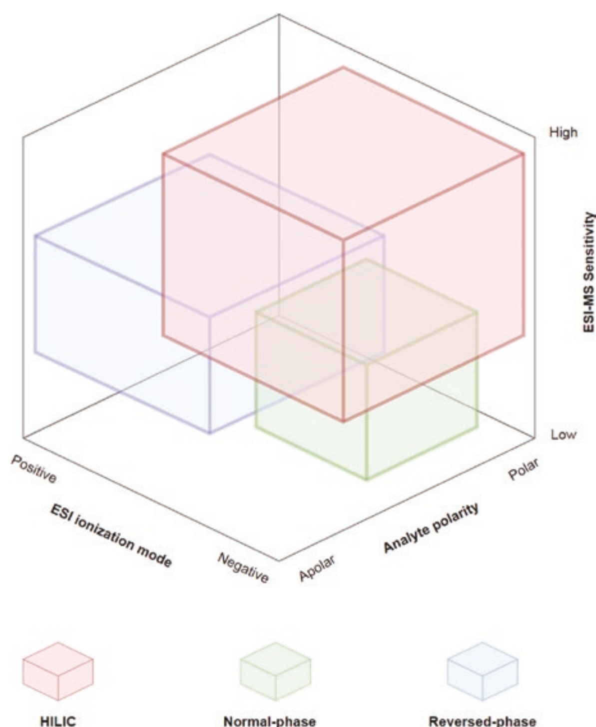

Fig. 1 The chromatographic option to increase the detection coverage of metabolites using LC-ESI-HRMS

signal [12]. The volatome or volatilome, a comprehensive and untargeted study of the expression of volatile compounds, was used interchangeably using GC-MS. For instance, the nonvolatile compounds can be analyzed on GC-MS after derivatization based on the functional groups (e.g., carboxylic, alcohol, amines, and thiol). Alkylation, acylation, and silylation are the common derivatization groups, whereas trimethylsilylation (TMS) is the most comprehensive to cover many functional groups. N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) is the most popular metabolomics and produces minimal "artifacts" compared to other silylation reagents [12].

Linear quadrupole ion trap (LIT), three-dimensional quadrupole ion trap (QIT), orbitrap, time of flight (TOF), and ion cyclotron resonance (ICR), all of these use the static or dynamic magnetic/electric field that are the main mass analyzers used in targeted and untargeted metabolomics and lipidomics studies. The proper selection of the mass analyzer depends on the resolution, mass range, scan rate, and detection limit (his part has been covered extensively in chapter "Metabolomics: A Pipeline for Biomarker Discovery in Genetic Diseases"). The choice of analytical technique depends on the research objectives, experimental design, and the biological sample queued for investigation. For unsupervised and discovery projects, it is highly recommended to use two or more independent or hyphenated techniques in multiple ionization polarities and separation modes to achieve a wide-ranging profile of metabolites.

3 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a spectroscopic analytical technique used to identify and quantify organic molecules based on hydrogen, carbon, nitrogen, flour, and phosphorus nuclei. The nuclei of the elements carry a charge that creates a magnetic dipole when the proton and neutron spins are not paired. The resulting magnetic dipole generates a spin axis along with the nucleus' spin axis, where the magnitude of this dipole is a fundamental nuclear property (nuclear magnetic moment, μ). The charge distribution is a function of the internal structure of the nuclei. The transitions between nuclear spins should be observable once a magnetic field is applied perpendicular to a nuclear magnetic energy level. Hydrogen NMR (^{1}H-NMR) is the most commonly targeted nucleus in analyzing samples of biological origin due to its natural abundance. NMR also targets other atoms, such as carbon and phosphorus, for additional information on a specific class of metabolites.

Initially, NMR has played a crucial role in metabolomics for molecular characterization and structural elucidation, followed by expression profiling and dynamic determination. The sensitivity and element-selective detection of the NMR to the nuclear spin made it one of the major tools in metabolomics studies. Coupling NMR with mass spectrometry enhances metabolomics profiling and identifications. NMR has proven its role in medical diagnosis and clinical research that targets the protein [23], lipid [24], and metabolites biomarkers in addition to the whole microbiological species (reviewed in chapter "Bringing Human Serum Lipidomics to the Forefront of Clinical Practice: Two Clinical Diagnosis Success Stories") [24]. Despite the analytical sensitivity, NMR is considered one of the highly robust techniques to quantify the naturally abundant metabolites in biological matrices for diagnostic purposes.

The NMR analytical workflow starts by extracting metabolites from the biological matrices and dissolving the dry extracts with NMR technique-compatible solvents. Post-acquisition, the NMR spectral data generated for each metabolite are univariate and multivariate analyzed. Then, the significant features are annotated using public libraries such as the Human Metabolome Database (HMDB) [3]. This part can be further explored in chapter "Metabolomics in the Study of Human Mitochondrial Diseases" by David Wishart.

In conclusion, NMR complements the MS by analyzing the isobaric, hard-to-be ionized, or structurally unknown compounds [25]. The metabolic transformation's dynamic and mechanism can be depicted using stable isotope labels-NMR (reviewed extensively by Markley et al. 2017) [26].

4 Metabolomics Data Analysis

The association of metabolic dysregulation with the clinical phenotype and the feasibility of being a diagnostic biomarker are explored further in multiple packages by developing a model. The validation approach considers

predictive accuracy, sensitivity, and specificity. For instance, the receiver operating characteristic (ROC)'s area under the curve (AUC) represents the probability of the classifier ranking a randomly chosen positive sample higher than a randomly chosen negative one. ROC curve avoids systematic bias and is the most used performance assessment method, where the ROC's perfect classifier (AUC = 1) while a random classifier will obtain AUC close to 0.5. An AUC > 0.7 is often considered the minimal performance for a biomarker test to be clinically useful [27]. The classification models are followed with a validation process to estimate the model's performance to a new set of samples, particularly when a small set of samples are used in the discovery cohort. Permutation analysis and cross-validation testing are the two main approaches for validation [28].

The connection between the discovered metabolomics panel in a discovery cohort with the disease etiology is usually achieved through the pathway and network analyses. Several computational platforms such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [29–31], Small Molecule Pathway Database (SMPDB) [32], Edinburgh Human Metabolic Network (EHMN) [33], Wiki Pathways [34], and MetaCyc [35] provide extensive information of a large number of metabolic pathways. The metabolic pathways-based methods are known as metabolite set enrichment analysis (MSEA). They are based on the gene set enrichment analysis (GSEA) approach, which was designed for pathway analysis of gene-expression data [36, 37].

5 Advancement of Metabolomics Application in Clinical Research

Decades of advancement in clinical chemistry are mainly based on biochemical analyses, where separation science and molecular spectroscopy techniques play the primary role. Metabolomics is considered an advanced version of the biochemical genetics and clinical biochemistry platforms, where newborn screening, and urinary organic acid profile using the GC-MS platform, is the first clinical application of metabolomics. The advancement of the MS techniques and the data analysis pipelines enhanced the discovery rate of metabolic biomarkers and translational potential for clinical usage. Several chapters in this book will cover some clinical applications of metabolomics, such as endocrinology (reviewed in detail in chapters "Metabolomics and Genetics of Rare Endocrine Disease: Adrenal, Parathyroid Glands, and Cystic Fibrosis" and "Metabolomics Role in Personalized Medicine: An Update"), mitochondrial diseases (reviewed in detail in chapter "Lipidomic Profiling in Clinical Practice Using LC-MS"), the inborn error of metabolism (reviewed in detail in chapters "Bioinformatics Tools for Clinical Metabolomics" and "Untargeted Metabolomics in Newborn Screening"), etc. Some applications are highlighted in this section.

Clinical metabolomics research is based mainly on study design, replication, clinical data collection, specimen collection, storage, metabolism quenching, metabolic extraction, and instrumental acquisition (reviewed in detail in chapters "Metabolomics: A Pipeline for Biomarker Discovery in Genetic Diseases" and "Metabolomics of Rare Endocrine, Genetic Disease: A Focus on the Pituitary Gland"). These factors have shown in multiple studies a drastic role in generating a reproducible and sensitive metabolic biomarker for disease diagnosis and treatment management. In 2019, a quality assurance (QA) and quality control (QC) consortium (mQACC) was established to engage the community in addressing the quality challenges in untargeted metabolomics [38].

In this decade, metabolomics has been extensively applied in oncology, where many studies have focused on potential cancer biomarkers discovery for diagnostic, prognostic, therapy, and prevention.

Depending upon the genetic alterations, patient-tailored therapy by cancer's genomic and epigenomic characteristics has become possible. Many studies that have been conducted on cancer metabolome, estrogen receptor, and Her2/neu status in breast cancer are some known examples in oncology [39–42].

Metabolomics has also unfolded different branches of medicine, including perinatology, prenatal care, and maternal-fetal medicine. Prenatal medicine pregnancy is a critical state with various aspects to consider, including the mother's and child's well-being. Perinatal pathologies include chromosomal aberration, preterm delivery (PTD), congenital heart defects, spina bifida, chorioamnionitis, and low birth weight. Many studies have been conducted for biomarker discovery in aneuploidy screening, preeclampsia, fetal growth restriction, preterm labor, and delivery [43–45].

5.1 Oncology

The Warburg effect was the first to report the metabolic arm of cancer by converting glucose to lactate at a higher rate than normal cells [46]. This effect moves the main source of cellular energy from the mitochondria to the cytoplasm, which shifts the cell from an inert entity producing ATP to rapidly dividing cells. The metabolic nature of cancer progression produces many amino acids, lipids, and nucleotides, to produce the cellular biomass for the highly proliferating cells. Metabolomics has a great role in cancer personalized medicine which has been extensively reviewed [47]. As the frozen tissue is quite limited, an alternative formalin-fixed paraffin-embedded (FFPE) option is still viable for cancer metabolic biomarkers, mainly for tumor classification [48]. This protocol combines sensitive targeted metabolomics that covers cancer's most important pathways, such as glycolysis, TCA, and pentose phosphate pathways. For instance, more than 30 endogenous metabolites in breast cancer have been reported, including cholate levels (resulting from increased phosphocholine), low glycerophosphocholine, and low glucose, as potential biomarkers. These biomarkers improve the sensitivity and sensibility up to 83–100% in

detecting malignancy more than tumor size, lymph node hormonal status, and histology [49].

Elevated levels of glutamine, glycine, cysteine, and threonine were reported in ovarian cancer studies, and tryptophan, histidine, and phenylalanine were degraded significantly (reviewed by Turkoglu et al.) [50]. The profile based on the next-generation metabolomics in lung cancer was reviewed extensively as biomarkers for diagnosis, pathogenesis, classifications, and precision medicine. Most LC metabolomics markers are involved in the upregulation and reprogramming of the TCA and glycolysis pathways and the upregulation of phospholipid metabolic pathways and fatty syntheses (reviewed by Turkoglu et al.) [51]. Serum lactic acid, progesterone, homocysteine, 3-hydroxybutyrate, linoleic acid, stearic acid, myristic acid, threonine, and valine levels were reported to be significantly dysregulated in endometrial cancer (EC) [52, 53]. A group of metabolites has shown great concordance with colorectal cancer (CC) recurrence, prognosis, and survival [54]. Most of these metabolites play roles in perturbing cellular respiration and carbohydrate metabolism (i.e., TCA cycle and anaerobic respiration), lipid metabolism (i.e., fatty acid oxidation), amino acid metabolism (i.e., histidine, methionine, and tryptophan), and nucleotide metabolites (i.e., uracil, *p*-cresol, etc.) [54]. Cancer as a metabolic disease has been reviewed recently by Wishart [55].

The cancer metabolic biomarkers (oncometabolomics) open the venue for utilizing mass spectrometry in translational medicine. Removing tumor tissue accurately with minimally invasive, especially in brain tumors, is now closely using iKnife technology. These surgical cutters connected to desorption electrospray ionization (DESI) mass spectrometry were developed by Graham Cooks groups at Purdue University and Nathalie Agar at Harvard Medical School. This ambient ionization technique mediated a medical cutter and mass spectrometry collecting analytical signals for oncometabolite to accurately diagnose tissue removed during brain surgery [56]. This approach reduces postsurgical complications, speeds up surgery, and minimizes the subjectivity of conventional pathology. Zoltán Takáts at Imperial College London capitalized on this technology and started to build an oncometabolite-based library from multiple laboratories using rapid evaporative ionization mass spectrometry (REIMS). In BC, 63 phospholipids and 6 triglyceride species were responsible for 24 spectral differences between normal and tumor tissues, increasing the diagnostic sensitivity and specificity to 93.4% and 94.9%, respectively, using REIMS [57].

5.2 *Prenatal Medicine*

Pregnancy is a very natural process and complicated from a research standpoint to consider the mother and the baby's well-being. A wide range of physiological changes is required throughout the gestation in maternal, which requires circulating some metabolites such as triglyceride, cholesterol, and lipids to satisfy in utero fetal development from energy and catabolic needs. Multiple complications are known to be

associated with pregnancy, e.g., bleeding, ectopic pregnancy, miscarriage or fetal loss, placental complication, and preeclampsia or eclampsia. These complications are associated with overlapped risk factors such as the mother's age, body mass index, and medical history. Preeclampsia is the leading pregnancy complication that affects about 5–8% of pregnant women worldwide. Gestational high blood pressure is the main preeclampsia characterization, and it is combined with protein in the urine, occasionally fluid retention, and in some severe cases, seizures, coma, and death.

Oxidative stress is the common metabolic pathway in most prenatal disorders, except preeclampsia and congenital anatomic defects. Acylcarnitine and amino acid metabolism are altered in both preeclampsia and CAD, while taurine metabolic dysregulation is unique in patients with preeclampsia. Preterm labor and deliveries alter bile acid and inflammation metabolism. Some energy metabolism (e.g., pentose phosphate pathways, ketone body production) is dysregulated in other pregnancy complications, including single gene disorders.

Throughout the pregnancy, noninvasive (e.g., history analysis, ultrasound, maternal serum analysis) and invasive (e.g., fetal samples of the placenta) assessments might be needed. Developing noninvasive tests with lower false-positive rates is fundamental in prenatal medicine, where metabolomics is potentially useful. Metabolomics has great potential for evaluating several biomarkers in a single, rapid, relatively low-cost, controlled experiment. This technology identifies the targeted pathway altered through multiple intermediates unsupervised, which can identify the fetal or pregnancy disorders using simultaneously different compartments (e.g., maternal, placental, and fetal). However, overinterpretation and high false discovery rates, with unestablished cutoff values for the target analytes, are the biggest limitation of metabolomics in prenatal medicine. Multiple metabolomics studies have been conducted on prenatal cases covering normal pregnancy in different trimesters [43–45]. Aneuploidy screening, preeclampsia [58], fetal, preterm labor and delivery [45], congenital anatomic defects (CAD), and single gene disorders [43].

The cause of preeclampsia is not known, and mothers with a history should be under continuous monitoring for any potential risk. Several metabolic biomarkers were reported in the literature to predict the early onset of preeclampsia from control, such as taurine and asparagine [58]. Metabolomics Diagnostics (Metabolomic Diagnostics), an Ireland-based, deep-tech medical diagnostics company, specializes in developing novel biomarker-based diagnostic solutions. This company established a preeclampsia screening program after screening more than 1000 pregnant women to validate a panel of biomarkers for preeclampsia prediction. Their panel of metabolites (PrePsia) is now executed as a prototype screening test to identify women at increased risk of developing preeclampsia in early pregnancy.

6 Frontiers in Metabolomics

Metabolomics platforms have been drastically advanced in the past decade, with great translational potential in medicine. The biological fluids and materials are used between the study groups in the standard clinical metabolomics pipeline.

Single-cell metabolomics has shown promising results that explain the cellular heterogeneity based on the phenotypical differences to help understand the cause of cellular biochemical activity with implications for health and disease [59]. Single-cell metabolomics experiments with low sample volume are based on a shotgun-like approach. This motivates the developments of the metabolomics pipelines such as the nano-DESI, ambient ionization technique, droplet-based microextraction, single probe with duel bore capillary, etc. Single-cell metabolomics has great potential for studying the cellular biochemical changes in a specific genetic disease that is hard to notice in biological fluids, such as mitochondrial disorders (reviewed in detail in chapter "Lipidomic Profiling in Clinical Practice Using LC-MS").

Molecular imprinting polymer (MIP)-based electrochemical sensor is greatly applied in targeted metabolomics toward developing point-of-care tests (POCT) [60]. Having this line of research and technique advancement encourages the application scientist to continue in biomarker discovery at multiple levels, including monitoring health statuses such as metabolic disorders and those with critical metabolic risk. This part of the research is covered extensively in chapter "Transferring Metabolomics to Portable Diagnostic Devices: Trending in Biosensors".

Together, the advancement of the biosensor of targeted metabolomics and the metabolomics biomarker discovery opens the venue toward real-time health monitoring, mainly for metabotyping individuals for well-being, nutritional, and personalized medicine. Urine metabolomics in multiple studies has shown a real-time behavior regardless of the analytical platform, from sample collection to analysis limitations [61]. By enabling patients to access continuous measurement and wearable devices to diagnose and control their illnesses precisely, metabolomics will have a great opportunity in this field by multiplexing the monitoring to avoid the confounding effects in each case.

Metabolomics is gaining significant interest in medicine and clinical diagnosis, where multiple novel assays have been licensed to provide metabolomics-based diagnostic services. Metabolon and Baylor College of Medicine have worked in the last decades in developing Global MAPSTM, a semiquantitative metabolomics profiling, to determine the disruption related to specific biochemical abnormalities. More details regarding their experience in detecting some metabolic diseases are detailed in chapters "Bioinformatics Tools for Clinical Metabolomics" and "Untargeted Metabolomics in Newborn Screening". Quantose®IR and Quantose®IGT are metabolomics-based clinical assays for insulin resistance (IR) and impaired glucose tolerance (IGT) identification in patients with multiple conditions, such as type 2 diabetes. In addition to insulin, Quantose®IR is based on a panel of biomarkers comprised of a small organic acid (α-hydroxybutyric acid (AHB)) and two lipids (oleic acid and linoleoylglycerophosphocholine (LGPC)) [62]. The algorithm scoring was developed based on a nondiabetic cohort from 13 European countries, where the score cutoff is 63 between insulin sensitivity and resistance. This test is performed using LC-MSMS for treatment monitoring, such as pioglitazone [62]. Quantose®IGT identifies the prediabetes risk based on the IGT. The Quantose®IGT scores were developed based on α-hydroxybutyric acid (AHB), 4-methyl-2-oxopentanoic acid (4MOP), oleic acid, linoleoylglycerophosphocholine (LGPC), β-hydroxybutyric acid (BHBA), serine, and pantothenic acid (vitamin B5) levels in the sample [63].

Nonalcoholic fatty liver disease (NAFLD) is a highly prevalent progressive chronic disease in which the liver displays histological features like those induced by excessive alcohol intake but in the absence of alcohol consumption. NAFLD is very common in people with diabetes and the obese. Its early identification is important due to the burden of the disease and the fact that NAFLD often presents with only mild or no symptoms. OWLiver® Care and OWLiver® are noninvasive assays for fatty liver screening and NASH diagnosis. These two tests use highly sensitive laboratory processes to determine the form of disease present in the patient's liver. These tests determine the risk of developing NASH, reflected in the patient's lifestyle and medical management. This part is covered extensively in chapter "Bringing Human Serum Lipidomics to the Forefront of Clinical Practice: Two Clinical Diagnosis Success Stories".

References

1. Holmes E, Antti H. Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra. Analyst. 2002;127(12):1549–57.
2. Cravatt BF, Prospero-Garcia O, Siuzdak G, Gilula NB, Henriksen SJ, Boger DL, et al. Chemical characterization of a family of brain lipids that induce sleep. Science (New York, NY). 1995;268(5216):1506–9.
3. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: the human metabolome database for 2022. Nucleic Acids Res. 2022;50(D1):D622–31.
4. Kennedy AD, Wittmann BM, Evans AM, Miller LAD, Toal DR, Lonergan S, et al. Metabolomics in the clinic: a review of the shared and unique features of untargeted metabolomics for clinical research and clinical testing. J Mass Spectrom. 2018;53(11):1143–54.
5. Abdel Rahman AM, Pawling J, Ryczko M, Caudy AA, Dennis JW. Targeted metabolomics in cultured cells and tissues by mass spectrometry: method development and validation. Anal Chim Acta. 2014;845:53–61.
6. Jacob M, Nimer RM, Alabdaljabar MS, Sabi EM, Al-Ansari MM, Housien M, et al. Metabolomics profiling of nephrotic syndrome towards biomarker discovery. Int J Mol Sci. 2022;23(20):12614.
7. Aleidi SM, Dahabiyeh LA, Gu X, Al Dubayee M, Alshahrani A, Benabdelkamel H, et al. Obesity connected metabolic changes in type 2 diabetic patients treated with metformin. Front Pharmacol. 2020;11:616157.
8. Hocher B, Adamski J. Metabolomics for clinical use and research in chronic kidney disease. Nat Rev Nephrol. 2017;13(5):269–84.
9. Jacob M, Gu X, Luo X, Al-Mousa H, Arnaout R, Al-Saud B, et al. Metabolomics distinguishes DOCK8 deficiency from atopic dermatitis: towards a biomarker discovery. Metabolites. 2019;9(11):274.
10. Jans JJM, Broeks MH, Verhoeven-Duif NM. Metabolomics in diagnostics of inborn metabolic disorders. Curr Opin Syst Biol. 2022;29:100409.
11. Heinemann J. Machine learning in untargeted metabolomics experiments. Methods Mol Biol. 2019;1859:287–99.
12. Beale DJ, Pinu FR, Kouremenos KA, Poojary MM, Narayana VK, Boughton BA, et al. Review of recent developments in GC-MS approaches to metabolomics-based research. Metabolomics. 2018;14(11):152.

13. Ramautar R. Capillary electrophoresis-mass spectrometry for clinical metabolomics. Adv Clin Chem. 2016;74:1–34.
14. Zhou J, Yin Y. Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. Analyst. 2016;141(23):6362–73.
15. Gordillo R. Supercritical fluid chromatography hyphenated to mass spectrometry for metabolomics applications. J Sep Sci. 2021;44(1):448–63.
16. Spagou K, Tsoukali H, Raikos N, Gika H, Wilson ID, Theodoridis G. Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. J Sep Sci. 2010;33(6–7):716–27.
17. Virgiliou C, Gika HG, Theodoridis GA. HILIC-MS/MS multi-targeted method for metabolomics applications. Methods Mol Biol. 2018;1738:65–81.
18. Tang DQ, Zou L, Yin XX, Ong CN. HILIC-MS for metabolomics: an attractive and complementary approach to RPLC-MS. Mass Spectrom Rev. 2016;35(5):574–600.
19. Zhang X, Quinn K, Cruickshank-Quinn C, Reisdorph R, Reisdorph N. The application of ion mobility mass spectrometry to metabolomics. Curr Opin Chem Biol. 2018;42:60–6.
20. Mairinger T, Causon TJ, Hann S. The potential of ion mobility-mass spectrometry for non-targeted metabolomics. Curr Opin Chem Biol. 2018;42:9–15.
21. Zandkarimi F, Brown LM. Application of ion mobility mass spectrometry in lipidomics. Adv Exp Med Biol. 2019;1140:317–26.
22. Uppal K. Models of metabolomic networks. In: Wolkenhauer O, editor. Systems medicine. Oxford: Academic Press; 2021. p. 134–42.
23. Hu Y, Cheng K, He L, Zhang X, Jiang B, Jiang L, et al. NMR-based methods for protein analysis. Anal Chem. 2021;93(4):1866–79.
24. Marchand J, Martineau E, Guitton Y, Le Bizec B, Dervilly-Pinel G, Giraudeau P. A multidimensional (1)H NMR lipidomics workflow to address chemical food safety issues. Metabolomics. 2018;14(5):60.
25. Wishart DS. NMR metabolomics: a look ahead. J Magn Reson. 1997;2019(306):155–61.
26. Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, et al. The future of NMR-based metabolomics. Curr Opin Biotechnol. 2017;43:34–40.
27. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics. 2013;9(2):280–99.
28. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proc Natl Acad Sci U S A. 2009;106(10):3698–703.
29. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
30. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353–61.
31. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–51.
32. Jewison T, Su Y, Disfany FM, Liang Y, Knox C, Maciejewski A, et al. SMPDB 2.0: big improvements to the small molecule pathway database. Nucleic Acids Res. 2014;42(Database issue):D478–84.
33. Hao T, Ma HW, Zhao XM, Goryanin I. Compartmentalization of the Edinburgh Human Metabolic Network. BMC Bioinform. 2010;11:393.
34. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al. WikiPathways: building research communities on biological pathways. Nucleic Acids Res. 2012;40(Database issue):D1301–7.
35. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 2008;36(Database issue):D623–31.
36. Tilford CA, Siemers NO. Gene set enrichment analysis. Methods Mol Biol. 2009;563:99–121.

37. Deng L, Ma L, Cheng KK, Xu X, Raftery D, Dong J. Sparse PLS-based method for overlapping metabolite set enrichment analysis. J Proteome Res. 2021;20(6):3204–13.
38. Beger RD, Dunn WB, Bandukwala A, Bethan B, Broadhurst D, Clish CB, et al. Towards quality assurance and quality control in untargeted metabolomics studies. Metabolomics. 2019;15(1):4.
39. Zhao X, Modur V, Carayannopoulos LN, Laterza OF. Biomarkers in pharmaceutical research. Clin Chem. 2015;61(11):1343–53.
40. Ni Y, Xie G, Jia W. Metabonomics of human colorectal cancer: new approaches for early diagnosis and biomarker discovery. J Proteome Res. 2014;13(9):3857–70.
41. Liesenfeld DB, Habermann N, Owen RW, Scalbert A, Ulrich CM. Review of mass spectrometry-based metabolomics in cancer research. Cancer Epidemiol Biomarkers Prev. 2013;22(12):2182–201.
42. Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. J Pharm Biomed Anal. 2014;87:1–11.
43. Monni G, Atzori L, Corda V, Dessolis F, Iuculano A, Hurt KJ, et al. Metabolomics in prenatal medicine: a review. Front Med. 2021;8:645118.
44. McKeating DR, Fisher JJ, Perkins AV. Elemental metabolomics and pregnancy outcomes. Nutrients. 2019;11(1):73.
45. Carter RA, Pan K, Harville EW, McRitchie S, Sumner S. Metabolomics to reveal biomarkers and pathways of preterm birth: a systematic review and epidemiologic perspective. Metabolomics. 2019;15(9):124.
46. Warburg O. On respiratory impairment in cancer cells. Science. 1956;124(3215):269–70.
47. Cacciatore S, Loda M. Innovation in metabolomics to improve personalized healthcare. Ann N Y Acad Sci. 2015;1346(1):57–62.
48. Kelly AD, Breitkopf SB, Yuan M, Goldsmith J, Spentzos D, Asara JM. Metabolomic profiling from formalin-fixed, paraffin-embedded tumor tissue using targeted LC/MS/MS: application in sarcoma. PLoS One. 2011;6(10):e25357.
49. Bathen TF, Geurts B, Sitter B, Fjøsne HE, Lundgren S, Buydens LM, et al. Feasibility of MR metabolomics for immediate analysis of resection margins during breast cancer surgery. PLoS One. 2013;8(4):e61578.
50. Turkoglu O, Zeb A, Graham S, Szyperski T, Szender JB, Odunsi K, et al. Metabolomics of biomarker discovery in ovarian cancer: a systematic review of the current literature. Metabolomics. 2016;12(4):60.
51. Yu L, Li K, Zhang X. Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: mini review. Oncotarget. 2017;8(70):115774–86.
52. Troisi J, Sarno L, Landolfi A, Scala G, Martinelli P, Venturella R, et al. Metabolomic signature of endometrial cancer. J Proteome Res. 2018;17(2):804–12.
53. Njoku K, Sutton CJ, Whetton AD, Crosbie EJ. Metabolomic biomarkers for detection, prognosis and identifying recurrence in endometrial cancer. Metabolites. 2020;10(8):314.
54. Zhang F, Zhang Y, Zhao W, Deng K, Wang Z, Yang C, et al. Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review. Oncotarget. 2017;8(21):35460–72.
55. Wishart D. Metabolomics and the multi-omics view of cancer. Metabolites. 2022;12(2):154.
56. Eberlin LS, Norton I, Orringer D, Dunn IF, Liu X, Ide JL, et al. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. Proc Natl Acad Sci U S A. 2013;110(5):1611–6.
57. St John ER, Balog J, McKenzie JS, Rossi M, Covington A, Muirhead L, et al. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. Breast Cancer Res. 2017;19(1):59.
58. Kuc S, Koster MP, Pennings JL, Hankemeier T, Berger R, Harms AC, et al. Metabolomics profiling for identification of novel potential markers in early prediction of preeclampsia. PLoS One. 2014;9(5):e98540.

59. Duncan KD, Fyrestam J, Lanekoff I. Advances in mass spectrometry based single-cell metabolomics. Analyst. 2019;144(3):782–93.
60. Ozcelikay G, Kaya SI, Ozkan E, Cetinkaya A, Nemutlu E, Kır S, et al. Sensor-based MIP technologies for targeted metabolomics analysis. TrAC Trends Anal Chem. 2022;146:116487.
61. Miller IJ, Peters SR, Overmyer KA, Paulson BR, Westphall MS, Coon JJ. Real-time health monitoring through urine metabolomics. NPJ Digit Med. 2019;2:109.
62. Cobb J, Gall W, Adam KP, Nakhle P, Button E, Hathorn J, et al. A novel fasting blood test for insulin resistance and prediabetes. J Diabetes Sci Technol. 2013;7(1):100–10.
63. Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhle PJ, et al. alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a non-diabetic population. PLoS One. 2010;5(5):e10883.

Mass Spectrometry-Based Metabolomics for the Clinical Laboratory

Joshua A. Dubland

Abstract In the clinical laboratory, analysis of small molecules using mass spectrometry (MS) primarily encompasses the targeted and quantitative determination of known biomarkers for disease diagnosis and monitoring, general health status evaluation, toxicology, and therapeutic drug monitoring. Although there are exceptions, MS-based assays in the clinical laboratory typically involve the utilization of analyte-specific calibration curves for quantitation, and stable isotope-labeled internal standards to correct for any sample preparation and instrument-related variability. A clinical MS-based assay usually consists of a relatively small panel of biomarkers in a certain diagnostic context that are compatible with the same sample preparation protocol. Alternatively, the term metabolomics generally refers to the comprehensive and systematic large-scale profiling of small molecules within a biological system. Targeted and quantitative MS-based assays containing relatively large panels of small molecules (metabolites) are routinely utilized for newborn screening (NBS), biochemical genetics testing, and toxicology. Broad nontargeted metabolomics investigations have found some utility in the aforementioned testing areas, but are not currently commonplace in the clinical laboratory. This chapter discusses current state-of-the-art MS instrumentation, describes several applications, and provides implementation considerations for MS-based metabolomics in the clinical laboratory.

Keywords Mass spectrometry · Metabolomics · Clinical laboratory · Instrumentation · Applications · Validation

J. A. Dubland (✉)
Department of Pathology and Laboratory Medicine, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada
e-mail: joshua.dubland@ubc.ca

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_2

Abbreviations

AC	Alternating current
APCI	Atmospheric chemical ionization
APPI	Atmospheric pressure photoionization
CCS	Collisional cross section
CI	Chemical ionization
CID	Collision-induced dissociation
Da	Dalton
DC	Direct current
EI	Electron ionization
ESI	Electrospray ionization
eV	Electron volt
FTICR	Fourier-transform ion cyclotron resonance
FWHM	Full width at half maximum
GC	Gas chromatography
HILIC	Hydrophilic interaction liquid chromatography
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry
ICP	Inductively coupled plasma
IEMs	Inborn errors of metabolism
IMS	Ion mobility spectrometry
kV	Kilovolt
LC	Liquid chromatography
m/*z*	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MRM	Multiple-reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
QTOF	Quadrupole time-of-flight
RF	Radiofrequency
TOF	Time-of-flight
UPLC	Ultra-high-performance liquid chromatography

1 Introduction

Mass spectrometry (MS) is a powerful tool for investigating biological processes and has been extensively utilized for metabolomics, which is the comprehensive and systematic large-scale analysis of small molecules within biological systems. MS-based metabolomics can provide insights into the biochemical status and flux in both the healthy and diseased states. Metabolomics investigations can be either targeted or nontargeted (or a mixture of these approaches).

Targeted metabolomics has been used in both research and clinical diagnostic laboratories, and focuses on identifying and accurately quantifying small metabolite

panels (typically <10 analytes). For targeted metabolomics, analytes have been preselected and the analytical approach has gone through some form of validation to ensure the accuracy of the results. Targeted and quantitative analysis of relatively large small molecule panels (>10 analytes) using MS has found applications in newborn screening (NBS), biochemical genetics testing, and toxicology.

Alternatively, nontargeted (global) metabolomics aims to detect as many specific metabolites or metabolomic features as possible within a given sample. The term "metabolomic feature" refers to a certain mass-to-charge ratio (m/z) at a unique chromatographic retention time. Nontargeted metabolomics can involve the following or combinations of the following: (1) m/z pattern recognition and relative signal comparisons between samples, (2) quantitative or semiquantitative analyses using calibration curves and/or surrogate calibrants, and (3) spectral comparisons with MS libraries to try and identify some or all of the metabolomic features. Nontargeted metabolomics analyses can lead to the discovery of new metabolites and potential biomarkers. Confirmatory analyses are still needed, including chromatographic retention time and m/z comparisons with synthesized or commercially available compounds, and/or nuclear magnetic resonance (NMR) spectroscopy to elucidate the chemical structure of isolated metabolite chromatographic fractions. Large-scale nontargeted (global) MS-based metabolomics investigations have been utilized primarily for research purposes and have not been widely utilized in the clinical diagnostic laboratory.

This chapter starts with a description of the basic concepts of a mass spectrometer and instrumental interfacing with several chromatographic techniques. Next, some current examples of MS-based metabolomics in the clinical diagnostic laboratory are described. A brief discussion on guidance documentation and validation considerations for small molecule MS assays in the regulated clinical setting is then provided.

2 Components of a Mass Spectrometer

On a simplistic level, a mass spectrometer consists of three components: an ion source, a mass analyzer, and a detector. The ion source first generates charged molecules (ions) in the gas phase from the sample being analyzed. The mass analyzer component separates and isolates the ions based on their m/z. The abundance of isolated ions at specific m/z's are then recorded by a detector. A mass spectrometer requires low pressure (high vacuum) to direct ions through the instrument and remove contaminants [1]. Specialized software is used to control the mass spectrometer and analyze data.

2.1 Sample Ionization

A variety of ionization techniques are available for introducing charged molecules into a mass spectrometer. Currently, the most utilized methods of ionization include electron ionization (EI), chemical ionization (CI), electrospray ionization (ESI),

atmospheric chemical ionization (APCI), and matrix-assisted laser desorption/ionization (MALDI). A mixture of ionization approaches is needed to improve metabolite coverage for global metabolomics analysis of biological specimens (extensively reviewed in chapter "The Advanced Technology and Clinical Application in Metabolomics") [2].

EI is frequently utilized for ionization and fragmentation of thermally stable volatile small molecules. It is a process where a beam of electrons (typically having a kinetic energy of 70 eV) is emitted from a heated filament and collides with gas-phase neutral molecules to generate charged ions and fragments. This process must occur in a vacuum in order to avoid oxidation and beam scattering. The electron beam abstracts one electron from a neutral molecule, generating a radical cation, which then breaks into smaller, either charged or neutral, fragments. Neutral fragments are eliminated or removed via a vacuum on the EI chamber walls. Positively charged ions are drawn out of the source using an electric field, and the beam of ions is focused [3]. The generated EI fragment spectra (ions are typically separated and isolated by a quadrupole mass analyzer prior to detection) are useful for quantitation, structural characterization, and library matching. EI is a very common ionization and fragmentation method used for coupling gas chromatography (GC) with a mass spectrometer (GC-MS).

CI is a soft ionization technique (lower energy than EI) where a proton is added or removed from the neutral molecule using a reagent gas such as methane or ammonia. An electron beam produces a number of ionized and reactive reagent gas species, which transfer a proton to the neutral molecule of interest and generate a product ion. Generally, there is minimal fragmentation, and a protonated molecular ion is abundantly present in the mass spectra. Typically, CI is operated in positive ion mode (addition of a proton), as most neutral molecules can form positive ions through reactive intermediates. CI can also be conducted in negative ion mode if the molecules being ionized can stabilize a negative charge from electron capture, such as carboxylic acid groups or highly electronegative halogens [4, 5]. A CI source can be used in GC-MS, but is less routinely utilized than EI.

ESI is also a soft ionization technique and operates at atmospheric pressure. It is the standard ionization source utilized when liquid chromatography (LC) is coupled to a mass spectrometer (LC-MS). The ESI source exists in many variations, but consists of the same general components. The liquid sample first passes through a capillary inside a sample probe. An electrical voltage of less than 5 kV is applied to the capillary, leading to the generation of charged droplets as the liquid exits the capillary [1]. A heated nebulization gas (usually nitrogen) moving parallel to the sample capillary is used to help facilitate liquid droplet formation, directionality, and evaporation. Charged liquid droplets generate smaller liquid droplets as they move through the atmospheric region, which then completely evaporate (desolvate), resulting in the formation of gas-phase ions. Singly or multiply charged ions can be generated. After moving through the atmospheric region, the ions enter the mass spectrometer via a small orifice in a sampling cone that has a voltage applied to it. The capillary that generates charged droplets can be directly pointed toward the sampling cone orifice or positioned orthogonally to it. One or more cones (skimmers) in a series are sometimes

utilized. A neutral gas running perpendicular to the direction of ions moving through the cone orifice is used to prevent contamination of the MS. ESI works well with polar compounds that can hold a charge or exist in a charged state [4, 6].

In the APCI source, no voltage is applied to the capillary needle. Instead, the heated nebulization gas alone, flowing parallel to the capillary, generates solvent droplets as the solvent exits the tip of the capillary. The nebulization gas helps with the desolvation of the liquid droplets and is typically heated to higher temperatures than in ESI. The solvent in the vapor becomes charged from being in close proximity to a corona discharge needle that has a current (microamps typically) applied to it. The charge on the solvent (e.g., water, methanol, acetonitrile) is transferred to molecules in the vapor to produce ions. Like ESI, APCI operates at atmospheric pressure. APCI can provide more efficient ionization than ESI for nonpolar molecules. Both ESI and APCI are considered soft ionization techniques and generate minimal analyte fragmentation compared to EI. In-source fragmentation can occur for some molecules with ESI and APCI as a result of the high heat (between 300 and 650 °C) and/or the electric charge applied. LC is most often coupled to MS using ESI or APCI sources [1, 4, 7]. The solvents (mobile phases) carrying analytes from the LC into the ESI or APCI source are frequently made basic or acidic using modifiers such as formic acid, ammonium hydroxide, or a pH buffer such as ammonium formate to enhance the addition or removal of a proton (charge). Solvent modification can increase the MS signal intensity as more ions may be generated in the source and then enter the MS.

A variation of APCI much less commonly utilized is atmospheric pressure photoionization (APPI). In APPI, a photon-emitting lamp generates ions instead of a corona discharge needle. An ionizable dopant such as toluene or acetone is commonly infused in parallel to the nebulization gas for APPI. The dopant increases charge exchange or proton transfer to the molecule(s) being analyzed [8, 9]. Like APCI, APPI is generally used for the analysis of nonpolar analytes. In some cases, using APPI can lead to increased analyte ionization and higher signal intensities than APCI.

Matrix-assisted laser desorption/ionization (MALDI) is an ionization technique where a pulsed laser generates ions from a sample on which a UV-absorbing matrix has been applied. The matrix often acts as a proton donor to generate charged analytes. MALDI is a soft ionization technique that produces mostly singly charged ions [5]. MALDI is often used to generate ions for imaging MS of proteins and small molecules in tissue samples [10, 11]. Each MALDI laser shot at a certain location on a tissue section provides a mass spectrum comparable to a pixel in digital imaging. The combined analysis of many pixels provides spatial distribution information about ions across a tissue. An alternative surface ionization technique to MALDI is desorption electrospray ionization (DESI). With DESI, a charged solvent is directed toward the surface of a sample instead of a laser, and the charged solvent desorbs ions directly from the sample [5]. It is, therefore, a combination of ESI and surface desorption approaches. Interestingly, DESI has potential applications for real-time surgical tissue analysis as it is performed at atmospheric pressure and has simplified sample preparation relative to MALDI [12].

Direct analysis in real time (DART) is a rapid analysis approach that also produces ions directly from the surface of a sample without the need for the addition of any matrix or time-consuming sample preparation. With DART, an electrical current is applied to a heated inert gas to generate excited neutral species, which are directed at the sample of interest. The electronically excited inert gas (e.g., helium) interacts with water vapor and other gases at atmospheric pressure to generate reagent ions, which then cause chemical ionization of the analytes on the sample's surface. Gas-phase ions generated from the surface of the specimen then enter the MS for analysis [13]. DART can be performed in an open environment and is well suited for trace analysis screening purposes (e.g., drugs of abuse, plant material, explosives) [14] and has also been utilized for metabolite profiling [15].

2.2 Mass Analyzers

A variety of mass analyzers are commercially available, the most common of which are the quadrupole, time-of-flight (TOF), ion trap, orbitrap, and Fourier-transform ion cyclotron resonance (FTICR) mass analyzers. Mass analyzers vary in their m/z range, mass accuracy and resolution, and acquisition speed. At a very basic level, mass analyzers can be classified as either beam or trapping-type instruments. In beam instruments, ions make a single pass through the mass analyzer and are recorded using a detector. In trapping instruments, ions are confined using combinations of electric and magnetic fields, which are manipulated in order to measure selected ions. Beam instruments generally operate on the timescale of microseconds to milliseconds, whereas trapping instruments can operate from milliseconds to minutes (or even longer). Quadrupole and TOF mass analyzers are beam instruments, whereas ion trap, orbitrap, and FTICR instruments are trapping instruments. Although the development of the first mass spectrometers date back over 100 years [16–18], the basic utilization of electric and magnetic fields for separating and isolating ions remain the same. It is common for several types of mass analyzers to be combined within the same instrument and incorporate collision cell(s). These technical arrangements allow for the analysis of intact charged metabolites (precursors) and charged fragments of metabolites (products), which is critical for targeted analyses as well as nontargeted investigations.

2.2.1 Quadrupole

The quadrupole mass analyzer consists of four symmetrically arranged cylindrical metal rods with an ion flight path in the center that is parallel to the rods' direction (the *Z*-axis). Both alternating current (AC) and direct current (DC) are applied to pairs of rods directly opposing each other in the configuration. One set of opposing rods has a positive AC potential applied (*X*-*Z* plane), while at the same time, the

other set of opposing rods has a negative potential applied (*Y*-*Z* plane). This means that the AC or radio frequency (RF) waveform applied to one set of rods is 180° out of phase with the waveform applied to the other set of rods. At the same time, the DC potential of one set of opposing electrodes is positive (*X*-*Z* plane), while the other is negative (*Y*-*Z* plane). The applied voltages affect the flight path of ions between the rods as they traverse the device. When the AC potential is positive, the ions are accelerated onto the central *Z*-axis. In contrast, when the AC potential is negative, the ions are defocused and accelerated toward the rods. Depending on the voltage applied, ions either travel along the length of the quadrupole and pass through or are eliminated by colliding with the rods [19, 20].

In the *X*-*Z* plane, heavier ions will be mostly influenced by the positive DC potential and not affected by the high-frequency AC potential. Lower mass ions, on the other hand, will be significantly affected in the *X*-*Z* plane by the AC potential and, if they are light enough, may collide with the rods and be eliminated. In the *Y*-*Z* plane, heavier ions will be defocused from the central *Z*-axis. They may be eliminated on the *Y*-*Z* rods as they will mainly feel the destabilizing effect of the negative DC potential and not the high-frequency AC potential. However, lighter ions will be focused onto the central *Z*-axis in the *Y*-*Z* plane by the high-frequency AC potential and pass through the quadrupole.

In essence, the opposing rods in the *X*-*Z* plane act as a high-pass mass filter as only higher *m*/*z* ions will be transmitted and lower *m*/*z* ions will strike the *X*-rods and then be removed as neutral species by the turbo pumps. Conversely, the opposing rods in the *Y*-*Z* plane act as a low-pass mass filter transmitting only lower *m*/*z* ions and eliminating higher *m*/*z* ions [19]. Mathieu's equation describes the stability or instability of certain *m*/*z* ions in the *X*- and *Y*-coordinates in relation to AC and DC voltages on electrodes of opposite potential [21, 22]. The combination of the high-pass and low-pass mass filters creates a narrow band pass window (Δ *m*/*z*) of ions transmitted through the quadrupole. Changing the amplitude of the RF voltage can select (tune) for different masses to be transmitted through the quadrupole. A complete mass spectrum can be obtained by simultaneously varying the amplitude of the AC and DC voltages applied to the quadrupole rods but keeping the DC/RF ratio fixed. The mass resolution can be varied by changing the DC/RF ratio [19]. Generally, parameters are set such that the band pass window (Δ *m*/*z*) is constant across the entire *m*/*z* range, which provides sufficient mass resolution to separate isotopes of small molecules that are singly charged. Mass resolution is calculated as [(*m*/*z*)/(Δ *m*/*z*)]. The quadrupole mass analyzer can act as either a mass filter (i.e., isolate single Δ *m*/*z*'s as set by the user) or as a scanning instrument (i.e., scan Δ *m*/*z* across a range as set by the user) [4].

The benefits of quadrupole mass analyzers include high sensitivity, a large dynamic range, fast positive and negative mode polarity switching, small size, robustness, and low cost. Limiting factors are a mass range typically less than 2000 *m*/*z*, low mass accuracy (greater than 100 ppm), and a mass resolution of 0.5 to 1 Dalton (Da) full width at half the maximum (FWHM) [1, 4]. One Da is equivalent to 1/12 of the mass of carbon 12 in its lowest energy state. Some quadrupole analyzers can reach resolutions of ≤0.1 Da at FWHM with specially designed

quadrupole rods [23]. A quadrupole mass analyzer is commonly combined in series with another quadrupole mass analyzer, but is also frequently combined with ion trap (termed a QTRAP) or TOF (termed a QTOF) mass analyzers.

2.2.2 Triple Quadrupole

Triple quadrupole or tandem mass spectrometers (MS/MS) are a very common instrument found in the clinical laboratory for quantitative analysis as a result of their high specificity and sensitivity. The development of the triple quadrupole mass spectrometer was initially reported in 1978 by Yost and Enke [24]. A triple quadrupole mass spectrometer consists of the following: a quadrupole mass filter (Q1), a low energy collision cell usually consisting of an RF-only quadrupole (Q2), another quadrupole mass filter (Q3), and then an electron multiplier to detect transmitted ions.

The tandem mass spectrometer has several modes of operation: multiple-reaction monitoring (MRM), product ion scan, precursor ion scan, and neutral loss scan. The most utilized mode of operation for quantitative analysis is MRM. In MRM mode, a single mass is selected in Q1, fragmented in the Q2 collision cell, and then a product ion is selected in Q3 and detected. The instrument can acquire one MRM at a time for a set amount of time, which is termed the dwell time. MRM dwell times are on the order of milliseconds (typically between 10 and 300 ms). A product ion scan selects a certain *m*/*z* in Q1, fragments the ion in Q2, and obtains a product ion spectrum by scanning an *m*/*z* range in Q3. A precursor ion scan is the reverse of a product ion scan. A fragment ion *m*/*z* is selected in Q3, and Q1 is scanned for all precursor *m*/*z*'s that give that Q3 fragment ion. This mode is useful for analyzing compounds with a known common fragment ion. In a neutral loss scan, both Q1 and Q3 are scanned, looking for a common neutral mass difference (loss) between precursor and product ions [4, 25].

The Q2 collision cell is usually a low-energy collision-induced dissociation (CID) device that contains an inert gas such as argon or nitrogen. Ions transmitted from Q1 are given kinetic energy as they enter the CID device, collide with the inert gas, and are fragmented. Collision energies are typically between 1 and 100 eV. The Q2 quadrupole is operated in an RF-only mode (i.e., all ions are transmitted) by removal of the DC potential on the quadrupole rods [19, 26]. Although CID is the most common fragmentation method utilized in a triple quadrupole MS, other methods such as electron capture dissociation or surface-induced dissociation fragmentation methods are possible [27].

2.2.3 TOF

A TOF is a beam type mass analyzer where *m*/*z* is determined from the time it takes for an ion with a certain kinetic energy to travel through a long tube under a vacuum. An electrostatic field first accelerates packets of ions entering the TOF. The ions all obtain the same kinetic energy and then are separated over a drift path and

registered by a detector. Lower mass ions arrive at the detector first. The ion flight path length limits a TOF device's resolution. To increase the flight path and not the overall size of the instrument flight tube, the flight path can be reflected multiple times within the tube [28, 29]. The TOF mass analyzer is a pulsed instrument by nature but is frequently paired with a continuous incoming ion beam from ionization sources such as ESI, APCI, or EI. In order to extract packets of pulsed ions from the continuous incoming ion beam, the TOF drift tube is placed orthogonally to it. Ion packets are then accelerated and injected into the TOF low-pressure drift tube in pulses [30]. A TOF mass analyzer is also often paired with MALDI, which is already a pulsed incoming ion beam.

The TOF instrument is a high-resolution mass spectrometer (HRMS) that can measure the exact mass of an ion, typically within a mass error of five parts per million (ppm). This high-resolution mass measurement allows the TOF mass analyzer to be useful in identifying unknown compounds by matching the measured exact mass to a library of compounds of known molecular weight. Often, a quadrupole mass analyzer followed by a collision cell is placed between the ionization source and the TOF mass analyzer in order that high resolution fragment ion spectra can also be collected. This instrumental arrangement is called a QTOF [31]. A TOF mass analyzer has a very fast acquisition rate of microseconds, making it compatible with being placed after a quadrupole mass analyzer that has an acquisition rate of milliseconds. The quadrupole allows for initial low-resolution selection of specific ions or a range of ions, followed by fragmentation of those ions in the collision cell and then high-resolution measurement of those fragments by TOF. Metabolomics studies are often performed using a QTOF. The combination of measuring both the exact mass of intact precursor ions and the exact mass of the fragmentation (product ions) spectra greatly increases the spectral library matching performance. In addition, LC or GC is usually performed prior to molecules entering the MS in order to separate compounds and thereby improve spectral quality. The LC or GC retention time data also can be informative for identifying unknown molecules in combination with the exact mass spectra [4, 20].

Benefits of a TOF mass analyzer include high-resolution mass measurement capabilities, high spectral acquisition rates, very large mass measurement ranges, relative simplicity, durability, and having relatively reasonable cost. Disadvantages include a lower analytical sensitivity range than quadrupole mass analyzers, the requirement for a highly controlled instrument temperature environment, and lower resolution capabilities than other HRMS systems such as orbitrap or FTICR instruments.

2.2.4 Ion Traps

An ion trap is a mass analyzer that traps and stores ions using electric or magnetic fields. Several configurations of ion traps exist and are discussed briefly in the following section. These include the 3D-type quadrupole ion trap, 2D-type linearity ion trap, orbitrap, and ion cyclotron resonance mass analyzer.

The 3D-type quadrupole ion trap (also called a Paul ion trap) functions similarly to a quadrupole mass analyzer as oscillating RF fields and DC voltages are utilized. Quadrupole ion traps differ in their configuration though, as they consist of a hyperbolic ring electrode situated between two symmetrical hyperbolic entrance and exit end cap electrodes. Ions enter and exit the device through holes in the end caps. Unlike a quadrupole, which is a mass filter in 2D space, the ion trap accumulates ions confined in a circular 3D space between the electrodes. Storage of selected ion species or a certain mass range can be set in the ion trap. The incoming ion beam is first trapped, and then, ions are subsequently scanned out based on their m/z by manipulating the electric fields [22, 25]. Ion traps have high sensitivity but are limited in the overall number of ions that can accumulate due to space charging effects, which restricts the dynamic signal range. Since ions are accumulated in the ion trap spectral skewing from chromatographic peak elution does not affect the mass analysis. Notably, a 3D-type quadrupole ion trap is able to perform multiple-stage fragmentation (MS^n) experiments. All ions except the ion to be fragmented are first ejected from the trap. Then, the ion of interest is fragmented by collisional activation, and the products are subsequently detected. The fragmentation and analysis process can then be repeated at higher orders. MS^n experiments are useful for structural investigations of molecules, but higher fragmentation orders lead to a loss of signal intensity [20, 22, 32].

A 2D-type linear ion trap consists of a quadrupole with the addition of electrostatic plates on both ends of the device that generate a stopping potential to trap ions. The linear ion trap is unique because it can function as a stand-alone quadrupole mass analyzer or an ion trap. Relative to 3D-type quadrupole (Paul type) ion traps, the linear ion trap has a higher capacity for storing ions [33, 34]. Linear ion traps have been paired in series following a standard quadruple mass analyzer (a hybrid instrumental arrangement called a QTRAP). In this configuration, the hybrid instrument can operate as a regular quadrupole mass analyzer (i.e., generate standard MRM data) or utilize the trapping features of the linear ion trap such as enhanced MS scan, enhanced product ion scan, enhanced resolution scan, enhanced multiply charged scan, and generation of MS^3 ion fragmentation data [35–38].

The orbitrap is a high-resolution mass analyzer where ions are trapped in orbit around a central spindle-like electrode in electrostatic fields. The general principles of orbitraps in use today are based on the ion trap device called the Kingdon trap that was reported in 1923 [39]. Commercial orbitrap instruments were developed by Makarov et al. in the early 2000s [40, 41]. An orbitrap mass analyzer is made up of a central spindle-like electrode surrounded by outer cup-shaped electrodes. An ion packet from an orthogonally positioned curved linear ion trap (called a C-trap) is injected into the orbitrap through a small slit in the outer cup-shaped electrodes. An incoming ion beam initially fills the C-trap and then the ion packet is pulsed orthogonally out of the C-trap into the orbitrap mass analyzer. The ions that enter the orbitrap are bent around the central axial electrode using a radial electric field and with the correct choice of parameters ions continue to orbit the central electrode. At the same time, an axial electric field induces harmonic axial ion oscillations. An image current signal (meaning a current induced by ions

passing by a conductor) from the frequency of the harmonic axial ion oscillations is then recorded by outer electrodes on the orbitrap device acting as receiver plates. The frequency of the oscillations is mathematically related to the *m/z* of the ion. After Fourier transformation of the image current signal, a mass spectrum is obtained [40–42]. The resolution of the orbitrap mass analyzer decreases as the scan speed increases and also as the *m/z* increases [43]. In some incidences, resolution may need to be reduced in order to increase the orbitrap scan speed. In comparison, a TOF mass analyzer obtains relatively lower mass resolution than an orbitrap, but the resolution is independent of detection time and *m/z*.

Orbitrap mass analyzers have also been coupled with other devices, such as quadrupoles, linear ion traps, and collision cells in order to increase selectivity and perform MS^n fragmentation for structural analysis [41, 44]. The coupling of different devices is facilitated via the C-trap, which can also serve as a T-device. Instrument configurations can vary, but generally, the ion beam first passes through a quadrupole mass filter or a linear ion trap (having fragmentation capabilities) and then enters the C-trap. From the C-trap, ions can be sent orthogonally to the orbitrap mass analyzer or pass straight through the C-trap to a high collision-induced dissociation (HCID) cell or linear ion trap where fragmentation is induced. The ions are then returned to the C-trap, where they are orthogonally injected into the orbitrap to generate high-resolution mass spectra [41]. In this way, both high-resolution precursor and product ion spectra can be obtained with various experiments. Other molecular fragmentation methods have been utilized with orbitrap mass analyzers, such as electron transfer dissociation (ETD), which utilizes reagent anions to interact with peptide cations. ETD has helped facilitate in-depth analysis of peptides and post-translational protein modifications [41, 45].

The FTICR mass analyzer currently offers the highest resolution and mass accuracy of commercially available mass spectrometers and works on the basis that ions within a magnetic field undergo cyclical motion (cyclotron motion). FTICR mass analyzers consist of a Penning trap, where a strong uniform magnetic field is applied to induce cyclotron motion of ions in the plane perpendicular to the magnetic field lines (the radial plane). A ring and endcap electrodes are used to generate a weak quadrupolar electric field, which traps ions in the axial plane. Within the Penning trap, ions undergo three independent motions: cyclotron, magnetron, and axial. Cyclotron motion is the large circular motion of ions in the plane perpendicular to the magnetic field lines (the radial plane). Magnetron motion is an additional slow circular drifting motion of ions in the radial plane (a drifting of the cyclotron motion centre). Axial motion is the harmonic oscillation of ions along the magnetic field lines. Excitation electrodes generate a sweeping RF potential that excites ions to larger cyclotron orbits in the radial plane so that they pass in close proximity to a pair of detection electrodes. Similar to an orbitrap mass analyzer, an image current is recorded by the pair of detection electrodes and then Fourier transformation is performed to generate mass spectra. The *m/z* of an ion is inversely proportional to the ion cyclotron frequency. Resolution can be improved by increasing the magnetic field strength or by increasing the scan time. FTICR instruments can also perform MS^n experiments [46–51].

3 Chromatographic Separations Interfacing with MS

Physical separation of small molecules for MS analysis is important for metabolomic investigations in order to reduce matrix effects, separate interferences, and resolve both isomeric and isobaric compounds. Standard chromatographic techniques, as mentioned earlier, include liquid chromatography (LC) and gas chromatography (GC). Capillary electrophoresis (CE) has also been utilized. A somewhat newer separation tool for MS-based metabolomics is ion mobility, where ions are separated in the gas phase by interactions with a neutral gas in the presence of an electric field. A brief discussion of these separation approaches is provided here.

3.1 Liquid Chromatography

LC is an important laboratory technique that separates a mixture into individual components based on the interaction between a liquid and a solid stationary phase. Generally, LC refers to a separation that is performed in a column packed with a stationary phase. Thin layer chromatography (TLC), though, is also a type of LC where the stationary phase is coated on a sheet of inert material (such as aluminum or glass), and the liquid is drawn up the plate based on capillary action. LC columns use pressurized liquid to speed up chromatographic separations. Two general types of instruments are currently available, the high-performance liquid chromatography (HPLC) system and the ultra-high-performance liquid chromatography (UPLC) system. HPLC operates in a system pressure range generally less than 6000 psi, whereas UPLC operates in a pressure range of less than 18,000 psi. Due to the higher system pressures for UPLC, the use of stainless steel tubing is required.

Sample introduction for HPLC and UPLC is done using an autosampler. A sampling needle is first used to take up a small amount (typically 1–20 μL) of sample from a prepared liquid specimen in a vial or plate within the autosampler. Pressurized mobile phases (MPs) then transfer the sample via a fix-loop or flow-through injector onto an LC column. Analytes are separated on the column based on their interaction with the MPs and the column's stationary phase. Once analytes have eluted from the column they are transferred via tubing to the MS source for ionization (typically ESI or APCI, as described earlier in this chapter).

HPLC and UPLC columns are small, typically on the order of 30–150 mm in length and 1–4.6 mm in diameter. The adsorbent particle sizes for HPLC are typically between 3 and 5 μm in diameter. UPLC columns utilize particle diameters below 2 μm, which improves chromatographic efficiency (peak dispersion). Chromatographic efficiency is inversely proportional to the particle size. Smaller particle sizes increase system pressure, as the particle size is also inversely proportional to the column back pressure. The use of UPLC columns, which have high efficiency, can increase chromatographic resolution (sharper and narrower peaks) and simultaneously reduce sample analysis time. Higher flow rates can be utilized

since the system can tolerate higher back pressures. UPLC paired with MS is a preferred approach for metabolomics investigations [52].

Generally, synthetic silica polymers are used as the adsorbent packing material in LC columns. Pore sizes of the silica polymer particles are typically in the 100–300 Å range. The smaller the pore size, the larger the surface area available on the particle to interact with analytes. Smaller pore sizes are generally utilized for small molecules, whereas larger pore sizes are preferable for larger molecules such as proteins. The silica polymers packed into an LC column are usually covalently modified using alkyl chains (e.g., C8 or C18 chain lengths) or other functional groups (e.g., phenyl, cyano, amide, amine, fluorophenyl) to improve selectivity. In reversed-phase chromatography, a nonpolar stationary phase (e.g., C18 modified silica) is utilized. The mobile phase composition is ramped from high aqueous (polar) to high organic (nonpolar) solvent content. The organic solvent competes with nonpolar hydrophobic analytes for interaction with the nonpolar adsorbent stationary phase. Analytes generally elute in order of decreasing polarity as the mobile phase composition is increased from high aqueous to high organic solvent content. In normal-phase chromatography, the stationary phase is polar (e.g., bare silica or amide-modified silica), and the mobile phase composition is nonpolar. The mobile phase can be ramped to higher polarity content in order to elute polar analytes. Nonpolar compounds generally elute first in normal-phase chromatography. Normal-phase chromatography is not frequently utilized with HPLC or UPLC. If the stationary phase is bare silica, organic solvents nonmiscible with water are generally utilized as water can become highly adsorbed to the silica and cause chromatographic variability. Polar-modified silica (e.g., amide modification) is a normal-phase chromatography approach where water can be tolerated in the mobile phase. Another normal-phase type approach that has also been developed is called hydrophilic interaction liquid chromatography (HILIC). In HILIC separations, polar hydrophilic compounds are retained more than nonpolar hydrophobic compounds by the stationary at high nonpolar solvent content. The stationary phase is polar and hydrophilic. As the water (polar) content in the mobile phase increases, polar hydrophilic compounds elute off the column. The mobile phases used for HILIC generally require buffers such as ammonium formate or ammonium acetate to improve chromatographic peak shape. Additionally, acid or base modifiers such as formic acid or ammonium hydroxide are often added to mobile phases to enhance either positive or negative mode ionization in the MS source [53].

3.2 Gas Chromatography

GC is a gas-phase separation technique where a gaseous sample mixture is separated into components using a narrow hollow metal tube filled with a porous silica-based stationary phase. A neutral gas such as helium or hydrogen is used to move the volatile analytes through the column as a temperate ramp or constant temperature is applied to the column using a temperature-controlled oven. GC oven temperatures

can reach up to 350 °C, and columns must be able to withstand these high temperatures. MS is frequently used for detecting eluting analytes (typically using an EI source at 70 eV, as discussed earlier) when a mass spectrum is required for analyte confirmation or library searching. Other methods include flame ionization detection and thermal conductivity detection. GC columns are usually very long, being on the order of 10 m or more in length [54]. The first reports of combining GC with MS date back to the late 1950s [55]. GC-MS is a robust and well-utilized method of analyzing specimens for metabolic investigations and allows for the identification of unknown analytes using well-developed 70 eV EI libraries, but can only analyze volatile compounds. To enhance the volatilization of molecules derivatization is frequently required. A common method is silylation to generate trimethylsilyl (TMS) derivatives from alcohols, carboxylic acids, and amines. Ketone groups form TMS derivatives following initial treatment with hydrazine. Sample preparation of biological specimens for GC-MS analysis typically involves liquid-liquid extraction protocols that are generally longer than sample preparation techniques required for LC-MS analysis of the same specimens. Urine organic acid profiling by GC-MS is a standard targeted and nontargeted metabolomics assay utilized in biochemical genetics laboratories to investigate inborn errors of metabolism (IEMs) [56].

3.3 Capillary Electrophoresis

CE can be interfaced with MS and is an emerging method of separating analytes prior to MS analysis. It is a technique based on the movement of ions in a high electric field. Analytes move from one end of the capillary to the other based on their size and charge. CE is particularly useful for the separation of polar and charged molecules. ESI is a common mode of interfacing CE with MS [57].

3.4 Ion Mobility Spectrometry

Ion mobility spectrometry (IMS) is an analysis technique where molecules are separated based on their mobility through an inert buffer gas such as nitrogen or helium in the presence of an electric field. IMS generates an analytical output called a collisional cross section (CCS) value that provides additional complementary data to *m*/*z* determined by MS. The CCS value is a measure, in square Ångströms ($Å^2$), of the interaction of a molecule with the buffer gas. CCS values depend on the molecules' specific size, shape, and charge and also vary based on the buffer gas utilized in the device. Several different types of IMS devices are commercially available. These include drift tube ion mobility spectrometry (DTIMS), traveling wave ion mobility spectrometry (TWIMS), structures for lossless ion manipulations (SLIM), field asymmetric waveform ion mobility spectrometry (FAIMS), and trapped ion mobility spectrometry (TIMS) [58–61].

IMS devices operate on the millisecond timescale and are, therefore, well suited to being incorporated into an MS instrument (termed IMS-MS), which are often additionally paired with either LC or GC front-end separations. IMS is a useful analytical technique for increasing the separation of isomers and interferences in addition to LC and GC, or it can also operate as a stand-alone separation method. An IMS-MS instrument paired with initial chromatographic separation can provide retention time, *m*/*z*, and CCS data that is useful for metabolomics investigations [59, 62]. Ion mobility spectrometry has not been utilized in the clinical laboratory to any great extent thus far, but in the future it may offer ways to increase selectivity, shorten analysis times, and potentially reduce reliance on traditional LC or GC separations [63].

4 Current Clinical Metabolomic Applications

MS has been utilized for diagnostic testing in several areas of clinical pathology including clinical biochemistry, microbiology, anatomical pathology, toxicology, and newborn screening and biochemical genetics testing (Fig. 1) [64–67]. Generally, clinical quantitative diagnostic assays that utilize MS consist of targeted small panels of analytes where sample preparation and instrument parameters can be fairly easily optimized. Applications of MS for clinical biochemistry include quantitative LC-MS/MS analysis for steroids, vitamins, therapeutic drug monitoring of small molecules and antibodies, and protein biomarkers for various disorders [64, 66]. These types of quantitative clinical assays typically consist of a single analyte or a small panel of analytes. Larger profiling panel assays have also been developed to improve patient diagnosis. For example, a 26 analyte panel urine steroid profiling assay by LC-HRMS has been reported [68]. Trace metal analysis is also done by inductively coupled plasma MS (ICP-MS) [69]. MS is utilized in microbiology for the qualitative identification of microorganisms by MALDI-QTOF MS via spectral matching of ionizable proteins and peptides present in a bacterial culture to a validated MS library [70]. In anatomical pathology, the use of MALDI-QTOF MS for spatial imaging analysis of small molecules, proteins, and peptides in cancer and other tissue biopsies is an emerging area [71]. MS has also been used during surgeries to guide the removal of tumor tissues [72].

In toxicology, LC-MS/MS, GC-MS, and LC-HRMS are routinely used for drug screening and investigation of toxic chemical exposures. ICP-MS is also utilized to investigate exposure to toxic levels of metals [73]. Detecting known and unknown drug substances in the body involves a combination of targeted and nontargeted approaches. LC-HRMS and GC-MS assays utilize analytical standards and spectral library matching to identify known and unknown drugs and toxic substances. LC-MS/MS is utilized for targeted quantitative analysis. Additionally, the use of large panel targeted and nontargeted MS assays to detect drug metabolites and determine the effect of drugs on biochemical pathway metabolite formation may find applications in clinical toxicology to improve workflows [74–77]. Global

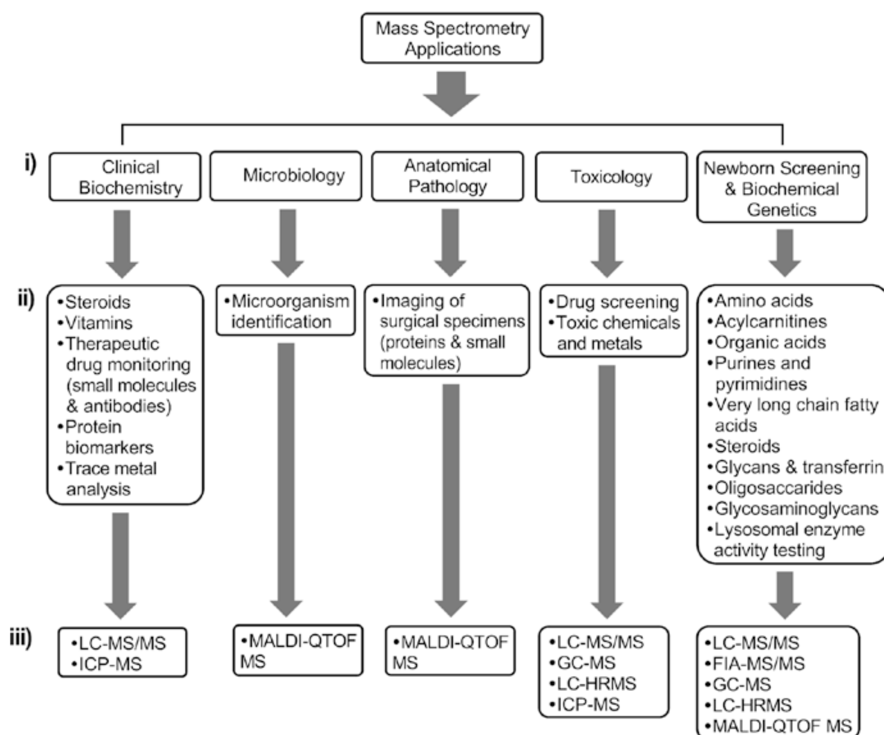

Fig. 1 Mass spectrometry applications to diagnostic testing by clinical pathology division. (i) Divisions of clinical pathology that MS has been applied to. (ii) Applications of MS in each division. (iii) MS platforms typically utilized in each division. Abbreviations: *FIA-MS/MS* flow injection analysis-tandem mass spectrometry, *GC-MS* gas chromatography-mass spectrometry, *ICP-MS* inductively coupled plasma-mass spectrometry, *LC-HRMS* liquid chromatography-high-resolution mass spectrometry, *LC-MS/MS* liquid chromatography-tandem mass spectrometry, *MALDI-QTOF* MS, matrix-assisted laser desorption/ionization-quadrupole time-of-flight mass spectrometry.

analysis looking at exogenous xenobiotic metabolites formed in the body, in addition to the effect of drug exposure on metabolites normally found in the body, is part of multidisciplinary metabolomics. Metabolomics being the large-scale systematic analysis of small molecules (metabolites) in a biological system.

Newborn screening and biochemical genetics is the clinical pathology division that currently heavily utilizes large-scale MS-based metabolomics investigations of biochemical pathway metabolites (both targeted large panels and nontargeted analyses) for diagnosing and monitoring IEMs. Targeted large panel assays primarily analyze for amino acids, acylcarnitines, organic acids, and purines and pyrimidines. Organic acid analysis also typically involves spectral library searching for nontargeted identification of unknowns. Other tests utilizing MS include very long chain fatty acids for analysis of peroxisomal disorders [78, 79], steroids for analysis of sterol biosynthesis and steroidogenesis disorders [80–83], glycans and intact transferrin for analysis of congenital disorders of glycosylation [84], oligosaccharide and

glycosaminoglycan analysis [85, 86], and lysosomal enzyme activity testing for lysosomal storage disorders [87].

Amino acids are the building blocks of life and are utilized to synthesize peptides and proteins in the body, such as enzymes, hormones, and neurotransmitters. They are also key energy sources and metabolic intermediates in biological pathways, and can be recycled when required. There are 21 amino acids that are utilized to generate proteins, and 9 of these are called “essential” amino acids as the body cannot synthesize them. Essential amino acids must be obtained in the diet [88]. Mutations in genes involved in amino acid catabolism can cause metabolic pathway blockages in the body, leading to toxic elevations in certain amino acids and metabolites [88, 89]. Deficiency in the synthesis pathways of amino acids caused by gene mutations can also occur [90]. As such, the measurement of amino acids in blood, plasma, urine, and cerebral spinal fluid is of essential importance for both diagnosing IEMs and subsequent patient monitoring [91]. Amino acid analysis for NBS is discussed later on. As part of the diagnostic workup, gene sequencing for patients suspected of an IEM is frequently performed in addition to the biochemical assessment.

Traditionally, deproteinized amino acids have been analyzed by ion exchange chromatography (IEC) with post-column ninhydrin derivatization for colorimetric detection [92, 93]. IEC has been considered the gold standard for amino acid analysis for over half a century. This approach involves lengthy sample analysis time (more than 1 h of analysis time per sample), lacks specificity and analytical resolution, and has minimal calibration. Many clinical laboratories have, therefore, moved away from IEC in favor of LC-MS/MS amino acid analysis approaches. LC-MS/MS approaches include pre-column or post-column amino acid derivatization, and underivatized analyses using ion pairing reagents in the LC mobile phases [94–97]. Assays must be able to cover large analytical ranges as amino acid concentrations span three orders of magnitude. LC-MS/MS provides a much faster sample analysis time relative to IEC, which can facilitate urgent patient care needs and improve laboratory workflows.

Acylcarnitine analysis is done to help diagnose mitochondrial β-oxidation and organic acid metabolism disorders [98, 99]. Typically, this is performed by flow injection analysis paired with MS/MS (FIA-MS/MS) for urine and serum specimens. Either derivatized or nonderivatized approaches can be used. Acylcarnitine analysis for NBS is discussed later on. Acylcarnitine isomers have also been separated using LC-MS/MS to increase specificity [100].

Urine organic acid analysis is traditionally done by GC-MS and covers a broad range of IEMs arising from enzyme or transporter deficiencies [56]. Disorders can affect many biochemical pathways, including the metabolism of organic acids, amino acids, carbohydrates, fatty acids, sterols, and purines and pyrimidines. It provides targeted analysis of known biomarkers and also nontargeted screening capabilities through spectral library searching (fragmentation energy of 70 keV is standard for GC-MS library generation and searching). Elevated levels of excreted metabolites and observed metabolite patterns are used to both identify and monitor disorders. Normalization of values to urine creatinine levels is required for

diagnostic evaluation. Analysis of urinary organic acids by GC-MS generally requires lengthy liquid-liquid extraction, derivatization to enhance organic acid volatilization, and long chromatography run times to separate analytes and interferences (typically greater than 30 min per sample). Recently, an LC-QTOF MS method was reported, allowing for simplified sample preparation without the need for sample derivatization [101].

Purines and pyrimidines are important molecules that help form DNA and RNA, act as metabolic regulators and intermediates, and provide an energy store [102]. Metabolic disorders of purine and pyrimidine metabolism are usually screened in a panel assay by LC-MS/MS [103, 104], identifying disorders by abnormal urinary excretion levels of these molecules. This analysis typically does not require sample derivatization as LC, and not GC, is utilized. Similar to urine organic acids, normalization to creatinine levels is required for urinary purine and pyrimidine analysis.

NBS programs aim to identify infants in the first few days after birth that are at high risk of having an IEM that may not be easily identifiable otherwise (i.e., having a lack of initial signs and symptoms of the disorder without a blood test). Early detection of certain disorders, such as phenylketonuria, allows for prompt treatment intervention in order to prevent death and other health complications. Dried blood spot (DBS) cards are typically utilized for NBS as they are relatively easy to collect via a heal prick from a newborn and provide sufficient analyte stability for the card to be transported at room temperature to a testing location. DBS cards are punched using an automated hole punch to provide a small sample for testing. Acylcarnitines and amino acids from DBS cards are analyzed simultaneously for NBS by FIA-MS/MS quantitation [105]. Derivatization with butanol hydrochloride is frequently utilized for NBS FIA-MS/MS. The initial (first-tier) FIA-MS/MS analysis lacks sufficient diagnostic specificity for some disorders and a follow-up (second-tier) test is needed to increase the overall positive predictive value without reducing screening sensitivity. Second-tier tests provide more specific biomarkers and/or separate interferences from the analysis. Typically, second-tier tests involve LC separations followed by MS/MS detection. The current necessity for lack of chromatographic separation for first-tier NBS prior to introduction into the MS is the sheer daily volume of specimens that must be analyzed [106, 107]. FIA-MS/MS analysis is less than 2 minutes a sample, whereas second-tier analyses can range anywhere from 2 to 15 minutes a sample. From the first-tier testing a smaller cohort of specimens is flagged for second-tier analysis, which allows for a more manageable screening workflow. Of note, a recent publication has outlined a strategy to combine first-tier and second-tier NBS using microfluidic capillary electrophoresis paired with HRMS analysis [108].

It is possible to develop large-scale targeted LC-HRMS or low-resolution LC-MS/MS clinical metabolomics panels for evaluations of IEMs (reviewed comprehensively in chapters "Bioinformatics Tools for Clinical Metabolomics" and "Untargeted Metabolomics in Newborn Screening") [56, 97, 101, 103, 109, 110]. Nontargeted metabolomics has limited clinical application to date outside urine organic acids analysis, which is typically performed by GC-MS. The advantage of large panel analyses is the efficient utilization of instrument time and a broader snapshot of the biological phenotype. A disadvantage of using very large metabolite

panels is that quality control issues arising from both the sample preparation and the instrumental analysis can occur more frequently. Analyte recovery can vary greatly when using a universal sample preparation approach. Obtaining optimal chromatographic peak shape may not be possible as columns are designed to work optimally with different classes of molecules. Matrix effects leading to analyte ion suppression or enhancement need to be considered. Using a large number of internal standards to correct matrix effects can be problematic from a technical perspective. In some situations, benefits and disadvantages need to be taken into consideration regarding either condensing or expanding existing small molecule test panels. Ultimately the approach must be beneficial to patient care and clinical laboratory workflows.

5 Guidance Documents and Validation Hurdles

Several guidance documents exist for validating small molecule MS assays [111, 112]. These documents describe analytical criteria that should be demonstrated in a method validation to prove that the assay is acceptable for use in the clinical laboratory. Assay parameters to be included in a method validation are sensitivity, specificity, linearity, analytical range, the limit of quantitation, the limit of detection, recovery, matrix effects, accuracy, precision, and sample stability. Evaluation of normal ranges, clinical diagnostic decision-making points, and abnormal ranges must be performed. Frequently, a method comparison of the new assay vs. another assay in use at a different laboratory is also conducted using real specimens [113, 114]. Targeted MS-based assays almost always involve the use of stable-labeled internal standards. Post-validation assay performance is monitored using a system suitability sample that is run prior to starting a batch of samples, in addition to QC samples analyzed during the run. Involvement in available external proficiency programs is also often done for assay performance monitoring.

Guidance strategies are currently available and work well, for single analyte or small panel assays in the clinical setting. However, for larger panel targeted metabolomic assays (such as some mentioned in the prior section), sample and batch acceptance criteria are not as straightforward. A fit-for-purpose approach, in some cases, is required. The clinical importance of certain biomarkers in the context of assay performance must be taken into consideration.

Nontargeted global metabolomics investigations are not generally utilized outside urine organic acids analysis in the routine clinical laboratory [56]. A major goal of many nontargeted metabolomics investigations is to identify differentially present biomarkers or biomarker signatures that may be informative of a disease or health state. Accurate quantitation in nontargeted metabolomics investigations is not generally a main goal [115]. The incorporation of HRMS nontargeted metabolomics investigations with clinically validated targeted small molecule panels may provide further health insights. Any potential biomarkers identified in nontargeted analyses should be validated in a fit-for-purpose manner and not by themselves drive clinical decisions. Harmonization of nontargeted metabolomics approaches

and the use of proficiency testing specimens [116, 117] will help improve results and data consistency from different laboratories [118–120]. A guidance document has also been published regarding combining and reporting ion mobility data from different instruments and buffer gases utilized [121].

6 Summary and Future Outlook

MS is an important analysis technique for metabolomics investigations. Targeted and nontargeted metabolomics analyses provide valuable insights into phenotype changes. These biochemical investigations provide complementary data to genomic analyses in relation to disease evaluation and overall health status. Typically, screening and diagnostic workflows in the clinical laboratory involve biochemical analyses first, followed by genome investigations when a genetic disorder is suspected from the biochemical profile. The order of investigations may change in the future as the cost of genomic testing continues to decrease. At some point in the future genomic analyses may be performed first, screening for known gene mutations and variants of unknown significance (VUS) associated with various metabolic disorders. Biochemical testing would then follow the genetic workup to evaluate for the presence of known disease biomarkers and provide further comprehensive (targeted and/or nontargeted) metabolic profiling as needed. Utilizing internal standards, quality controls, and external proficiency monitoring for biochemical MS assays is important to ensure the quality of the clinical laboratory data.

Further guidance documents are needed regarding approaches to acceptance criteria for large panel targeted MS-based metabolomics analyses, as current documents are aimed at small panels of analytes. Nontargeted metabolomics investigations have yet to be widely implemented in the routine clinical laboratory. Still, they may provide an investigative means in scenarios where existing targeted diagnostic testing has not provided clear answers. Harmonization of methodologies, validation criteria, and proficiency testing specimens for nontargeted MS-based metabolomics investigations is needed in the future. Incorporating nontargeted analyses into the clinical setting may facilitate the discovery of new biomarkers that can be validated and included in targeted analyte panels.

References

1. Kaklamanos G, Aprea E, Theodoridis G. Mass spectrometry: principles and instrumentation. In: Caballero B, Finglas PM, Toldrá F, editors. Encyclopedia of food and health. Oxford: Academic Press; 2016. p. 661–8.
2. Nordström A, Want E, Northen T, Lehtiö J, Siuzdak G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. Anal Chem. 2008;80(2):421–9.
3. Gross JH. Electron ionization. In: Gross JH, editor. Mass spectrometry: a textbook. Berlin: Springer; 2004. p. 193–222.

4. Rockwood AL, Kushnir MM, Clarke NJ. Chapter 2 - Mass spectrometry. In: Rifai N, Horvath AR, Wittwer CT, editors. Principles and applications of clinical mass spectrometry. Amsterdam: Elsevier; 2018. p. 33–65.
5. Bhardwaj C, Hanley L. Ion sources for mass spectrometric identification and imaging of molecular species. Nat Prod Rep. 2014;31(6):756–67.
6. Konermann L, Ahadi E, Rodriguez AD, Vahidi S. Unraveling the mechanism of electrospray ionization. Anal Chem. 2013;85(1):2–9.
7. Covey TR, Thomson BA, Schneider BB. Atmospheric pressure ion sources. Mass Spectrom Rev. 2009;28(6):870–97.
8. Robb DB, Covey TR, Bruins AP. Atmospheric pressure photoionization: an ionization method for liquid chromatography−mass spectrometry. Anal Chem. 2000;72(15):3653–9.
9. Raffaelli A, Saba A. Atmospheric pressure photoionization mass spectrometry. Mass Spectrom Rev. 2003;22(5):318–31.
10. Seeley EH, Caprioli RM. MALDI imaging mass spectrometry of human tissue: method challenges and clinical perspectives. Trends Biotechnol. 2011;29(3):136–43.
11. Reyzer ML, Caprioli RM. MALDI-MS-based imaging of small molecules and proteins in tissues. Curr Opin Chem Biol. 2007;11(1):29–35.
12. Calligaris D, Norton I, Feldman DR, Ide JL, Dunn IF, Eberlin LS, et al. Mass spectrometry imaging as a tool for surgical decision-making. J Mass Spectrom. 2013;48(11):1178–87.
13. Gross JH. Direct analysis in real time—a critical review on DART-MS. Anal Bioanal Chem. 2014;406(1):63–80.
14. Sisco E, Forbes TP. Forensic applications of DART-MS: a review of recent literature. Forensic Chem. 2021;22:100294.
15. Chernetsova ES, Morlock GE. Ambient desorption ionization mass spectrometry (DART, DESI) and its bioanalytical applications. Bioanal Rev. 2011;3(1):1–9.
16. Münzenberg G. Development of mass spectrometers from Thomson and Aston to present. Int J Mass Spectrom. 2013;349–350:9–18.
17. Nier KA, Yergey AL, Jane GP. A general chronicle of mass spectrometry and guide to the scope of the volume. In: Gross ML, Caprioli RM, editors. The encyclopedia of mass spectrometry. Boston: Elsevier; 2016. p. 7–12.
18. Nier KA. Dempster's descendants—the core of the development of mass spectrometry. J Mass Spectrom. 2020;55(8):e4353.
19. Miller PE, Denton MB. The quadrupole mass filter: basic operating concepts. J Chem Educ. 1986;63(7):617.
20. Li C, Chu S, Tan S, Yin X, Jiang Y, Dai X, et al. Towards higher sensitivity of mass spectrometry: a perspective from the mass analyzers. Front Chem. 2021;9:813359.
21. Stafford GC, Kelley PE, Syka JEP, Reynolds WE, Todd JFJ. Recent improvements in and analytical applications of advanced ion trap technology. Int J Mass Spectrom Ion Process. 1984;60(1):85–98.
22. Paul W. Electromagnetic traps for charged and neutral particles (Nobel Lecture). Angew Chem Int Ed Engl. 1990;29(7):739–48.
23. Yang L, Amad M, Winnik WM, Schoen AE, Schweingruber H, Mylchreest I, et al. Investigation of an enhanced resolution triple quadrupole mass spectrometer for high-throughput liquid chromatography/tandem mass spectrometry assays. Rapid Commun Mass Spectrom. 2002;16(21):2060–6.
24. Yost RA, Enke CG. Selected ion fragmentation with a tandem quadrupole mass spectrometer. J Am Chem Soc. 1978;100(7):2274–5.
25. Thomas SN. Chapter 10 - Mass spectrometry. In: Clarke W, Marzinke MA, editors. Contemporary practice in clinical chemistry. 4th ed. New York: Academic Press; 2019. p. 171–85.
26. Yost RA, Enke CG, McGilvery DC, Smith D, Morrison JD. High efficiency collision-induced dissociation in an RF-only quadrupole. Int J Mass Spectrom. 1979;30(2):127–36.

27. Sleno L, Volmer DA. Ion activation methods for tandem mass spectrometry. J Mass Spectrom. 2004;39(10):1091–112.
28. Plaß WR, Dickel T, Scheidenberger C. Multiple-reflection time-of-flight mass spectrometry. Int J Mass Spectrom. 2013;349–350:134–44.
29. Mamyrin BA. Time-of-flight mass spectrometry (concepts, achievements, and prospects). Int J Mass Spectrom. 2001;206(3):251–66.
30. Coles J, Guilhaus M. Orthogonal acceleration — a new direction for time-of-flight mass spectrometry: fast, sensitive mass analysis for continuous ion sources. TrAC Trends Anal Chem. 1993;12(5):203–13.
31. Allen DR, McWhinney BC. Quadrupole time-of-flight mass spectrometry: a paradigm shift in toxicology screening applications. Clin Biochem Rev. 2019;40(3):135–46.
32. Johnson JV, Yost RA, Kelley PE, Bradford DC. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. Anal Chem. 1990;62(20):2162–72.
33. Clarke W. Chapter 1 - Mass spectrometry in the clinical laboratory: determining the need and avoiding pitfalls. In: Nair H, Clarke W, editors. Mass spectrometry for the clinical laboratory. San Diego: Academic Press; 2017. p. 1–15.
34. Douglas DJ, Frank AJ, Mao D. Linear ion traps in mass spectrometry. Mass Spectrom Rev. 2005;24(1):1–29.
35. Cabruja M, Priotti J, Domizi P, Papsdorf K, Kroetz DL, Brunet A, et al. In-depth triacylglycerol profiling using MS3 Q-trap mass spectrometry. Anal Chim Acta. 2021;1184:339023.
36. Collings BA. Fragmentation of ions in a low pressure linear ion trap. J Am Soc Mass Spectrom. 2007;18(8):1459–66.
37. Lv S, Wang H, Yan Y, Ge M, Guan J. Quantification and confirmation of four aflatoxins using a LC–MS/MS QTRAP system in multiple reaction monitoring, enhanced product ion scan, and MS3 modes. Eur J Mass Spectrom. 2020;26(1):63–77.
38. Hopfgartner G, Varesio E, Tschäppät V, Grivet C, Bourgogne E, Leuthold LA. Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. J Mass Spectrom. 2004;39(8):845–55.
39. Kingdon KH. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. Phys Rev. 1923;21(4):408–18.
40. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem. 2000;72(6):1156–62.
41. Hecht ES, Scigelova M, Eliuk S, Makarov A. Fundamentals and advances of orbitrap mass spectrometry. In: Encyclopedia of analytical chemistry. New York: Wiley. p. 1–40.
42. Zubarev RA, Makarov A. Orbitrap mass spectrometry. Anal Chem. 2013;85(11):5288–96.
43. Crutchfield CA, Clarke W. Chapter 12 - High resolution accurate mass (HRAM) mass spectrometry. In: Nair H, Clarke W, editors. Mass spectrometry for the clinical laboratory. San Diego: Academic Press; 2017. p. 247–59.
44. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal Chem. 2006;78(7):2113–20.
45. Kim MS, Pandey A. Electron transfer dissociation mass spectrometry in proteomics. Proteomics. 2012;12(4–5):530–42.
46. Guan S, Marshall AG. Ion traps for Fourier transform ion cyclotron resonance mass spectrometry: principles and design of geometric and electric configurations. Int J Mass Spectrom Ion Process. 1995;146–147:261–96.
47. Nikolaev EN, Kostyukevich YI, Vladimirov GN. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: theory and simulations. Mass Spectrom Rev. 2016;35(2):219–58.
48. Blaum K, Eliseev S, Sturm S. Perspectives on testing fundamental physics with highly charged ions in penning traps. Quantum Sci Technol. 2021;6(1):014002.
49. Ostrander CM, Arkin CR, Laude D. Central ring electrode for trapping and excitation/detection in Fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectrom. 2001;12(1):30–7.

50. Marshall AG, Hendrickson CL. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. Int J Mass Spectrom. 2002;215(1):59–75.
51. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom Rev. 1998;17(1):1–35.
52. Perez de Souza L, Alseekh S, Scossa F, Fernie AR. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. Nat Methods. 2021;18(7):733–46.
53. Žuvela P, Skoczylas M, Jay Liu J, Baṃczek T, Kaliszan R, Wong MW, et al. Column characterization and selection systems in reversed-phase high-performance liquid chromatography. Chem Rev. 2019;119(6):3674–729.
54. Bartle KD, Myers P. History of gas chromatography. TrAC Trends Anal Chem. 2002;21(9):547–57.
55. Gohlke RS. Time-of-flight mass spectrometry and gas-liquid partition chromatography. Anal Chem. 1959;31(4):535–41.
56. Gallagher RC, Pollard L, Scott AI, Huguenin S, Goodman S, Sun Q. Laboratory analysis of organic acids, 2018 update: a technical standard of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2018;20(7):683–91.
57. Ramautar R. Capillary electrophoresis-mass spectrometry for clinical metabolomics. Adv Clin Chem. 2016;74:1–34.
58. Chouinard CD, Wei MS, Beekman CR, Kemperman RHJ, Yost RA. Ion mobility in clinical analysis: current progress and future perspectives. Clin Chem. 2016;62(1):124–33.
59. Delvaux A, Rathahao-Paris E, Alves S. Different ion mobility-mass spectrometry coupling techniques to promote metabolomics. Mass Spectrom Rev. 2022;41(5):695.
60. Dodds JN, Baker ES. Ion mobility spectrometry: fundamental concepts, instrumentation, applications, and the road ahead. J Am Soc Mass Spectrom. 2019;30(11):2185–95.
61. Ibrahim YM, Hamid AM, Deng L, Garimella SVB, Webb IK, Baker ES, et al. New frontiers for mass spectrometry based upon structures for lossless ion manipulations. Analyst. 2017;142(7):1010–21.
62. Paglia G, Astarita G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. Nat Protoc. 2017;12(4):797–813.
63. Dubland JA. Lipid analysis by ion mobility spectrometry combined with mass spectrometry: a brief update with a perspective on applications in the clinical laboratory. J Mass Spectrom Adv Clin Lab. 2022;23:7–13.
64. Fung AWS, Sugumar V, Ren AH, Kulasingam V. Emerging role of clinical mass spectrometry in pathology. J Clin Pathol. 2020;73(2):61.
65. Strathmann FG, Hoofnagle AN. Current and future applications of mass spectrometry to the clinical laboratory. Am J Clin Pathol. 2011;136(4):609–16.
66. Adaway JE, Keevil BG, Owen LJ. Liquid chromatography tandem mass spectrometry in the clinical laboratory. Ann Clin Biochem. 2015;52(1):18–38.
67. Jannetto PJ, Fitzgerald RL. Effective use of mass spectrometry in the clinical laboratory. Clin Chem. 2016;62(1):92–8.
68. Hines JM, Bancos I, Bancos C, Singh RD, Avula AV, Young WF, et al. High-resolution, accurate-mass (HRAM) mass spectrometry urine steroid profiling in the diagnosis of adrenal disorders. Clin Chem. 2017;63(12):1824–35.
69. Nuttall KL, Gordon WH, Ash KO. Inductively coupled plasma mass spectrometry for trace element analysis in the clinical laboratory. Ann Clin Lab Sci. 1995;25(3):264–71.
70. Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012;36(2):380–407.
71. Arentz G, Mittal P, Zhang C, Ho YY, Briggs M, Winderbaum L, et al. Chapter 2 - Applications of mass spectrometry imaging to cancer. In: Drake RR, McDonnell LA, editors. Advances in cancer research, vol. 134. New York: Academic Press; 2017. p. 27–66.

72. Basu SS, Stopka SA, Abdelmoula WM, Randall EC, Gimenez-Cassina Lopez B, Regan MS, et al. Interim clinical trial analysis of intraoperative mass spectrometry for breast cancer surgery. NPJ Breast Cancer. 2021;7(1):116.
73. Mbughuni MM, Jannetto PJ, Langman LJ. Mass spectrometry applications for toxicology. EJIFCC. 2016;27(4):272–87.
74. Szeremeta M, Pietrowska K, Niemcunowicz-Janica A, Kretowski A, Ciborowski M. Applications of metabolomics in forensic toxicology and forensic medicine. Int J Mol Sci. 2021;22(6):3010.
75. Plumb RS, Stumpf CL, Granger JH, Castro-Perez J, Haselden JN, Dear GJ. Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. Rapid Commun Mass Spectrom. 2003;17(23):2632–8.
76. Steuer AE, Brockbals L, Kraemer T. Untargeted metabolomics approaches to improve casework in clinical and forensic toxicology—"where are we standing and where are we heading?". WIREs Forensic Sci. 2022;4(4):e1449.
77. Keen B, Cawley A, Reedy B, Fu S. Metabolomics in clinical and forensic toxicology, sports anti-doping and veterinary residues. Drug Test Anal. 2022;14(5):794–807.
78. Al-Dirbashi OY, Santa T, Rashed MS, Al-Hassnan Z, Shimozawa N, Chedrawi A, et al. Rapid UPLC-MS/MS method for routine analysis of plasma pristanic, phytanic, and very long chain fatty acid markers of peroxisomal disorders. J Lipid Res. 2008;49(8):1855–62.
79. Valianpour F, Selhorst JJ, van Lint LE, van Gennip AH, Wanders RJ, Kemp S. Analysis of very long-chain fatty acids using electrospray ionization mass spectrometry. Mol Genet Metab. 2003;79(3):189–96.
80. Schwarz E, Liu A, Randall H, Haslip C, Keune F, Murray M, et al. Use of steroid profiling by UPLC-MS/MS as a second tier test in newborn screening for congenital adrenal hyperplasia: the Utah experience. Pediatr Res. 2009;66(2):230–5.
81. Janzen N, Sander S, Terhardt M, Steuerwald U, Peter M, Das AM, et al. Rapid steroid hormone quantification for congenital adrenal hyperplasia (CAH) in dried blood spots using UPLC liquid chromatography-tandem mass spectrometry. Steroids. 2011;76(13):1437–42.
82. Herman GE, Kratz L. Disorders of sterol synthesis: beyond Smith-Lemli-Opitz syndrome. Am J Med Genet C Semin Med Genet. 2012;160c(4):301–21.
83. Storbeck K-H, Schiffer L, Baranowski ES, Chortis V, Prete A, Barnard L, et al. Steroid metabolome analysis in disorders of adrenal steroid biosynthesis and metabolism. Endocr Rev. 2019;40(6):1605–25.
84. Abu Bakar N, Lefeber DJ, van Scherpenzeel M. Clinical glycomics for the diagnosis of congenital disorders of glycosylation. J Inherit Metab Dis. 2018;41(3):499–513.
85. Forni G, Malvagia S, Funghini S, Scolamiero E, Mura M, Della Bona M, et al. LC-MS/MS method for simultaneous quantification of heparan sulfate and dermatan sulfate in urine by butanolysis derivatization. Clin Chim Acta. 2019;488:98–103.
86. Huang R, Cathey S, Pollard L, Wood T. UPLC-MS/MS analysis of urinary free oligosaccharides for lysosomal storage diseases: diagnosis and potential treatment monitoring. Clin Chem. 2018;64(12):1772–9.
87. Gelb MH, Lukacs Z, Ranieri E, Schielen P. Newborn screening for lysosomal storage disorders: methodologies for measurement of enzymatic activities in dried blood spots. Int J Neonatal Screen. 2019;5(1):1.
88. Aliu E, Kanungo S, Arnold GL. Amino acid disorders. Ann Transl Med. 2018;6(24):471.
89. Ferreira CR, Rahman S, Keller M, Zschocke J, Group IA. An international classification of inherited metabolic disorders (ICIMD). J Inherit Metab Dis. 2021;44(1):164–77.
90. de Koning TJ. Amino acid synthesis deficiencies. J Inherit Metab Dis. 2017;40(4):609–20.
91. Phipps WS, Jones PM, Patel K. Chapter 2 - Amino and organic acid analysis: essential tools in the diagnosis of inborn errors of metabolism. In: Makowski GS, editor. Advances in clinical chemistry, vol. 92. Amsterdam: Elsevier; 2019. p. 59–103.

92. Berridge BJ Jr, Chao WR, Peters JH. Analysis of plasma and urinary amino acids by ion-exchange column chromatography. Am J Clin Nutr. 1971;24(8):934–9.
93. Spackman DH, Stein WH, Moore S. Automatic recording apparatus for use in chromatography of amino acids. Anal Chem. 1958;30(7):1190–206.
94. Carling RS, McDonald BA, Austin D, Burden D, Correia J, Leung J, et al. Challenging the status quo: a comparison of ion exchange chromatography with liquid chromatography-mass spectrometry and liquid chromatography-tandem mass spectrometry methods for the measurement of amino acids in human plasma. Ann Clin Biochem. 2020;57(4):277–90.
95. Waterval WAH, Scheijen JLJM, Ortmans-Ploemen MMJC, Habets-van der Poel CD, Bierau J. Quantitative UPLC-MS/MS analysis of underivatised amino acids in body fluids is a reliable tool for the diagnosis and follow-up of patients with inborn errors of metabolism. Clin Chim Acta. 2009;407(1):36–42.
96. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens JP, Bouchu D. Ion-pairing reversed-phase liquid chromatography/electrospray ionization mass spectrometric analysis of 76 underivatized amino acids of biological interest: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Rapid Commun Mass Spectrom. 2005;19(12):1587–602.
97. Le A, Ng A, Kwan T, Cusmano-Ozog K, Cowan TM. A rapid, sensitive method for quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry (LC-MS/MS). J Chromatogr B Analyt Technol Biomed Life Sci. 2014;944:166–74.
98. Miller MJ, Cusmano-Ozog K, Oglesbee D, Young S, Committee ALQA. Laboratory analysis of acylcarnitines, 2020 update: a technical standard of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2021;23(2):249–58.
99. Millington DS, Stevens RD. Acylcarnitines: analysis in plasma and whole blood using tandem mass spectrometry. Methods Mol Biol. 2011;708:55–72.
100. Peng M, Fang X, Huang Y, Cai Y, Liang C, Lin R, et al. Separation and identification of underivatized plasma acylcarnitine isomers using liquid chromatography–tandem mass spectrometry for the differential diagnosis of organic acidemias and fatty acid oxidation defects. J Chromatogr A. 2013;1319:97–106.
101. Körver-Keularts IMLW, Wang P, Waterval HWAH, Kluijtmans LAJ, Wevers RA, Langhans C-D, et al. Fast and accurate quantitative organic acid analysis with LC-QTOF/MS facilitates screening of patients for inborn errors of metabolism. J Inherit Metab Dis. 2018;41(3):415–24.
102. Nyhan WL. Disorders of purine and pyrimidine metabolism. Mol Genet Metab. 2005;86(1):25–33.
103. Monostori P, Klinke G, Hauke J, Richter S, Bierau J, Garbade SF, et al. Extended diagnosis of purine and pyrimidine disorders from urine: LC MS/MS assay development and clinical validation. PLoS One. 2019;14(2):e0212458.
104. Hartmann S, JrG O, Schmidt C, Langhans C-D, Garbade SF, Burgard P, et al. Comprehensive detection of disorders of purine and pyrimidine metabolism by HPLC with electrospray ionization tandem mass spectrometry. Clin Chem. 2006;52(6):1127–37.
105. Chace DH, Kalas TA, Naylor EW. Use of tandem mass spectrometry for Multianalyte screening of dried blood specimens from newborns. Clin Chem. 2003;49(11):1797–817.
106. Chace DH, Hannon WH. Impact of second-tier testing on the effectiveness of newborn screening. Clin Chem. 2010;56(11):1653–5.
107. Gramer G, Hoffmann GF. Second-tier strategies in newborn screening – potential and limitations. Med Genet. 2022;34(1):21–8.
108. Austin Pickens C, Isenberg SL, Cuthbert C, Petritis K. Combining first and second-tier newborn screening in a single assay using high-throughput Chip-based capillary electrophoresis coupled to high-resolution mass spectrometry. Clin Chem. 2021;67(12):1709–20.
109. Jacob M, Malkawi A, Albast N, Al Bougha S, Lopata A, Dasouki M, et al. A targeted metabolomics approach for clinical diagnosis of inborn errors of metabolism. Anal Chim Acta. 2018;1025:141–53.

110. Miller MJ, Kennedy AD, Eckhart AD, Burrage LC, Wulff JE, Miller LA, et al. Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. J Inherit Metab Dis. 2015;38(6):1029–39.
111. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM), editor. Bioanalytical method validation: guidance for industry; 2018.
112. CLSI. Liquid chromatography-mass spectrometry methods; approved guideline. CLSI document C62-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.
113. Wu AHB, French D. Implementation of liquid chromatography/mass spectrometry into the clinical laboratory. Clin Chim Acta. 2013;420:4–10.
114. Rappold BA. Review of the use of liquid chromatography-tandem mass spectrometry in clinical laboratories: part I-development. Ann Lab Med. 2022;42(2):121–40.
115. Naz S, Vallejo M, García A, Barbas C. Method validation strategies involved in non-targeted metabolomics. J Chromatogr A. 2014;1353:99–105.
116. Simón-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, et al. 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. 2013;85(24):11725–31.
117. Bowden JA, Heckert A, Ulmer CZ, Jones CM, Koelmel JP, Abdullah L, et al. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950–metabolites in frozen human plasma. J Lipid Res. 2017;58(12):2275–88.
118. Beger RD. Interest is high in improving quality control for clinical metabolomics: setting the path forward for community harmonization of quality control standards. Metabolomics. 2018;15(1):1.
119. Liu KH, Nellis M, Uppal K, Ma C, Tran V, Liang Y, et al. Reference standardization for quantification and harmonization of large-scale metabolomics. Anal Chem. 2020;92(13):8836–44.
120. Ulmer CZ, Maus A, Hines J, Singh R. Challenges in translating clinical metabolomics data sets from the bench to the bedside. Clin Chem. 2021;67(12):1581–3.
121. Gabelica V, Shvartsburg AA, Afonso C, Barran P, Benesch JLP, Bleiholder C, et al. Recommendations for reporting ion mobility mass spectrometry measurements. Mass Spectrom Rev. 2019;38(3):291–320.

Metabolomics: A Pipeline for Biomarker Discovery in Genetic Diseases

Lina A. Dahabiyeh and Refat M. Nimer

Abstract Genetic disorders (GD) affect hundreds of millions of lives worldwide, leading to specific molecular perturbations in the metabolic profile within a certain biological matrix. Metabolomic studies use advanced technologies (nuclear magnetic resonance (NMR) and mass spectrometry (MS)) with bioinformatics to identify and quantify a set of small molecules (such as carbohydrates, nucleic acids, amino acids, and lipids) present in a biological system. As metabolites represent the downstream product of gene and protein activity, they most likely reflect the phenotype of an organism at a specific time. Metabolomic studies provide novel insights into the underlying disease pathophysiological mechanisms, evaluate the progress of the disease, and identify unique biomarkers for the prediction of disease and therapeutic outcomes. This chapter highlights the applications of metabolomic techniques for biomarker discovery in GD. It discusses the workflow followed, the methods used for sample analysis and data interpretation, and the major challenges and limitations in applying the metabolomic approach for biomarker discovery in GD. Moreover, the chapter provides an overview of the biomarkers identified in four GD: cystic fibrosis, Down syndrome, sickle cell anemia, and glycogen storage disorders, highlighting the promising role of metabolomics in clinical applications.

Keywords Biomarker · Genetic diseases · Cystic fibrosis · Metabolomics · Mass spectrometry · NMR · Down syndrome

L. A. Dahabiyeh (✉)
Department of Pharmaceutical Sciences, School of Pharmacy, The University of Jordan, Amman, Jordan
e-mail: L.Dahabiyeh@ju.edu.jo

R. M. Nimer (✉)
Department of Medical Laboratory Sciences, Jordan University of Science and Technology, Irbid, Jordan
e-mail: rmnimer@just.edu.jo

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_3

Abbreviations

BALF	Bronchoalveolar lavage fluid
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CID	Collision-induced dissociation
DBS	Dried blood spot
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
EBC	Exhaled breath condensate
GC	Gas chromatography
GD	Genetic disorders
GSDs	Glycogen storage disorders
LOD	Limit of detection
LOQ	Limit of quantification
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
OPLS-DA	Orthogonal partial least squares-discriminant analysis
PCA	Principal component analysis
PLS-DA	Partial least squares-discriminant analysis
SCD	Sickle cell disease
VIP	Variable importance in projection

1 Introduction

Hundreds of millions of lives are affected by an estimated 10,000 unique genetically determined diseases [1]. Genetic disorders (GD) refer to any disease caused by mutations in one or more genes. Although GDs are individually rare, they account for approximately 80% of rare disorders [2]. Additionally, genetic material is considered a risk factor for several frequent complex multifactorial disorders, including cancer, asthma, heart disease, and diabetes [2, 3]. As with any other illness, GD will result in specific alterations in the profile of the biological molecules within a certain biological matrix, such as biofluids, cells, and tissues. Measurements of these biomolecules can be used to identify disease biomarkers and aid in understanding the underlying molecular mechanisms of the disease.

The last few decades have witnessed increasing interest in omics studies: genomics, transcriptomics, proteomics, and metabolomics (Fig. 1). They represent advanced and promising measurement approaches for disease biomarker discovery. Metabolomics is the global identification and quantification of a set of small molecules, less than 1500 Da (such as carbohydrates, nucleic acids, amino acids, and lipids) present in a biological system [4]. As metabolites represent the downstream product of gene and protein activity, they most likely reflect the phenotype of an organism at a specific time [5]. Metabolomic studies provide novel insights into the

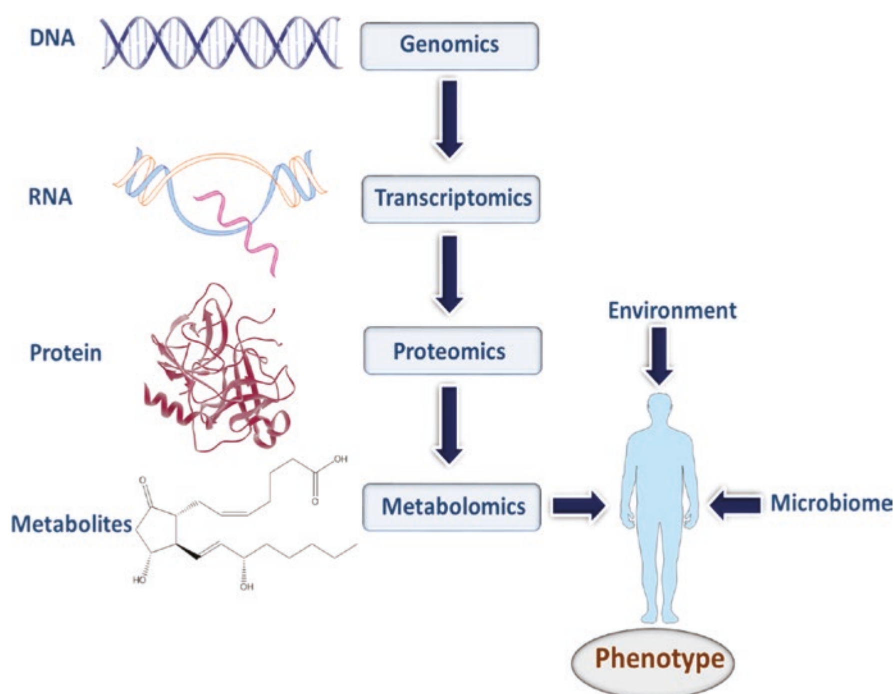

Fig. 1 The four omics sciences genomics, transcriptomics, proteomics, and metabolomics in systems biology approach. Metabolomics represents the downstream output of the genome, and together with environmental factors and the microbiome, it can reflect the phenotype of an organism

underlying disease pathophysiological mechanisms, evaluate the progress of the disease, and identify unique biomarkers for the prediction of disease and therapeutic outcomes [6–10].

Global metabolomic studies use advanced technologies with bioinformatics. Two main analytical technologies are employed: nuclear magnetic resonance (NMR) and mass spectrometry (MS) [8, 11]. Advances in MS coupled with chromatography (LC-MS) have improved the efficiency and reliability of metabolite profiling. Over the years, LC-MS-based metabolomics has witnessed tremendous improvements in sensitivity, mass resolution, metabolome coverage, and data processing, enabling reliable identification of diverse metabolites even within complex biological samples [12]. The application of metabolomics in biomedical research has identified biochemical pathways and has discovered diverse sets of potential biomarkers that have enhanced the understanding of the molecular mechanisms in several diseases and aided in their diagnosis, including diabetes [13, 14], cancer [15, 16], and GD [17]. A biomarker is a biomolecule (i.e., gene, protein, or metabolite-based substance) that indicates an abnormal condition within a subject. Biomarkers can be used in clinical and medical settings to determine specific disorders (known as diagnostic biomarkers), monitor disease progression (prognostic biomarkers), indicate probable response to therapy (predictive biomarkers), and indicate the risk

of developing a disease (predisposition biomarkers) [18]. Metabolomics allows the identification of tens of potential biomarkers that need to undergo analytical (with other analytical platforms such as polymerase chain reactions (PCR) or immune affinity-based assays) and clinical validations before being approved to be used in the clinic. Metabolome profiling of biofluids can be combined with whole-genome profiling of single nucleotide polymorphisms to identify many gene-metabolite associations simultaneously in metabolomic genome-wide association studies [19].

This chapter highlights the applications of metabolomic techniques for biomarker discovery in GD. The chapter discusses the workflow followed and the methods used for sample analysis and data interpretation in the metabolomic approach. Moreover, the chapter provides an overview of the most interesting recent biomarker discovered in the most frequent GD, such as cystic fibrosis, Down syndrome, and sickle cell anemia, highlighting the promising role of metabolomics in clinical applications. The current challenges and limitations in biomarker discovery in GD using the metabolomic approach will also be covered.

2 Sample Preparation in Metabolomics

Careful consideration of different aspects of metabolomic studies, such as study design, is vital to ensure high-quality data and reproducible biomarker generation. Following standardized protocols during sample collection and processing and using validated analytical methods to analyze samples will result in robust and reproducible analyses [20].

Experimental design in metabolomic studies may differ depending on the study's objective, available resources, and the type of biological sample. However, the general experimental workflow is the same. The typical workflow in metabolomics includes sample preparation and metabolite extraction, metabolite separation using chromatography (liquid (LC) or gas chromatography (GC)), metabolite analysis using MS or NMR spectroscopy, metabolite identification, and data processing using uni- and multivariate analyses for the discovery of potential biomarkers (Fig. 2).

2.1 Sample Collection

In metabolomic experiments, at least two samples (two conditions) are typically compared, one being the control (or reference) group. Biospecimens used for biomarker discovery are preferably collected from large case-control studies or cohorts with defined inclusion and exclusion criteria and complete information on the clinical and demographic characteristics, as much as possible.

Poorly defined groups, non-matched confounding factors, or heterogeneous samples are potential pitfalls in the study design and may negatively affect the study

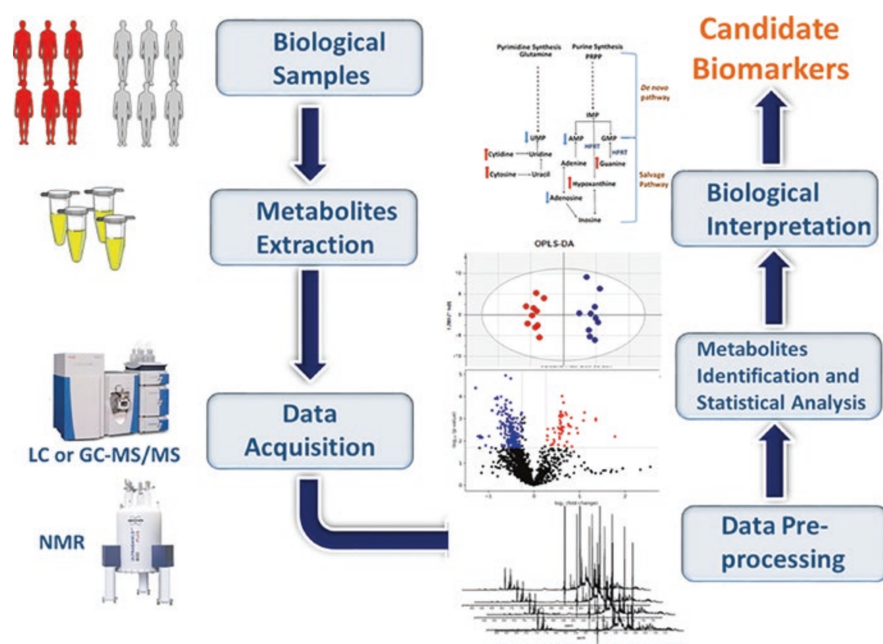

Fig. 2 Workflow followed in untargeted metabolomic analysis

outcomes. Poor study design during sample collection is one of the reasons that might explain the poor progression of discovered biomarkers toward clinical applicability [18].

2.2 Sample Type

Biomarker research aims to discover a biomarker using a noninvasive (e.g., urine, tears, saliva) or minimally invasive (e.g., blood and serum or plasma) sample and, if possible, try to avoid invasive samples (e.g., tissue). Blood is the most commonly used biofluid in metabolomic studies since it is easy to collect (compared to tissues). Its composition is stable and reflects the state of the body at the time of collection [21]. Blood is considered the most commonly used sample type in many GD, including cystic fibrosis (CF) and sickle cell disease [22–24]. Serum, plasma, sweat, dried blood spots, and epithelial cell cultures from cohorts of patients with CF [24–28] were used to identify novel metabolic abnormalities associated with CF and discover a panel of potential biomarkers for the disease. Metabolomic studies involving Down syndrome, the most common human chromosomal aberration, mainly used amniotic fluid from fetuses with Down syndrome [29, 30]. Cultured fibroblasts and liver tissues were used in metabolomic studies conducted on glycogen storage disease (GSD) and glucose-6-phosphatase deficiency mouse models, respectively [31, 32]. Inherited genetic mutations play a major role in about 5–10%

of all cancers. Many metabolomic studies on cancer patients mainly used tissue specimens to monitor disease progression, identify different disease stages, and discover biomarker candidates in breast and colon cancers [33–35].

2.3 Metabolite Extraction

The matrix of the biological sample has a certain level of interferences (such as salts and proteins) that can affect the operation of the analytical instrument. Therefore, sample preparation is crucial in metabolomic analyses and might be a major source of variability [36]. To ensure informative and accurate metabolite profile outcomes in global metabolomics, the sample preparation method should be (a) unselective to maximize metabolite coverage, (b) sensitive and compatible with the analytical approach followed, (c) simple and fast to avoid metabolite degradation and minimize variability, (d) reproducible, and (e) include a metabolism quenching step to stop any biochemical reactions after the time of sampling [37].

In the metabolomic study of serum or plasma, proteins will be precipitated first with organic solvent (such as methanol, acetonitrile, and acetone or a combination), followed by metabolite extraction. Among the most commonly used extraction methods are (a) liquid-liquid separation (with specified ratios of solvents), where metabolites of interest are separated into an immiscible solvent, (b) using a column or solid-phase extraction (SPE) approach to trap the metabolites, and (c) selective solubilization [38, 39]. Analysis involving GC-MS will typically include a step of derivatization before metabolic profiling. Among the solvents used in extraction, methanol was the best in metabolite coverage and method reproducibility [39]. On the other hand, sample preparation for NMR analysis is much simpler than MS-based approaches and requires a deuterated solvent and a chemical shift reference. Metabolite extraction during metabolite profiling using the LC-MS approach in CF, Down syndrome, and sickle cell disease mainly employed a solvent extraction method [23, 29, 40, 41]. It offers a simple and rapid extraction approach with excellent metabolome coverage. However, the latter highly depends on the solvents used [37]. To ensure system stability and aid in metabolite identification, stable isotope-labeled quality control internal standards are added during sample preparation, as in the metabolite profiling of sickle cell disease patients [22].

3 Technologies Used in Metabolomics

Current technologies used in metabolomic research involve MS and NMR spectroscopy.

3.1 Mass Spectrometry (MS)

Mass spectrometry (MS) is an indispensable analytical tool and a vital technology in metabolomic studies. It is widely applied due to its high sensitivity (typically at the pg level) and fast data acquisition speed. During the last decades, MS has moved into the front line of metabolomic research. Several well-established methods have identified potential biomarkers in GD from blood plasma-derived samples, cells, and tissues [29, 32, 34, 42, 43]. Advances in the MS analysis were achieved due to enhancements in the sensitivity, ionization, mass accuracy, and the high-throughput capabilities of different mass analyzers and the hyphenation of MS to different separation techniques, such as LC and GC, to allow for the analysis of complex biological samples. Common MS analyzers include quadrupole (Q; acts as a mass filter), high-resolution accurate mass analyzers, time-of-flight (TOF), and Orbitraps. To provide quality assurance for MS analysis and data acquisition, quality control (QC) samples, prepared from pooling aliquots of all samples to be analyzed, are randomized for analysis among the study samples [44]. Another approach used to compensate for the effect of ion suppression and increase the reliability of the generated data is to spike the samples with stable isotope-labeled internal standard structural analogs as internal standards, which is mainly applied in the targeted metabolomics [45, 46].

3.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy, including ^{1}H NMR, is used to profile metabolites for biomarker discovery and develop metabolic fingerprints for disease diagnosis and monitoring response to treatment due to its fundamental quantitative nature. NMR-based metabolomics has been applied to profile plasma and urine samples from subjects with Down syndrome and normal controls [47, 48] and plasma samples from sickle cell disease patients with normal albuminuria and patients with moderately or severely increased albuminuria [49]. The same approach was used to identify plasma metabolic phenotypes of children with autism spectrum disorder, idiopathic developmental delay, and Down syndrome compared to typically developed controls [50]. Unlike MS, NMR allows structural verification/identification of known and unknown metabolites and quantitation from the same measurement, requiring minimal sample preparation. However, NMR methods have lower sensitivity, dynamic range, and metabolite coverage when compared to LC-MS-based methods [12].

4 Metabolite Identification and Statistical Analysis

Advanced software combined with rich databases highly impacted advancement in metabolomic research. Typically, acquired data will be cleaned up to remove artifact peaks or peaks with poor repeatabilities, such as peaks detected in less than 50% QC samples or peaks with high variability (e.g., CV > 30% in QC). Normally, metabolite identification is performed by matching accurate masses of the detected peaks with metabolite in specialized databases and libraries, the retention times (RT) of authentic standards, and/or the MS/MS fragmentation database [41]. The confidence in metabolite identification in GD follows the general recommendations based on the Chemical Analysis Working Group Metabolomics Standards Initiative recommendation, where the confidence is assigned as level 1–4 (L1–4) [51, 52].

Different metabolomic studies in GD used multivariate analysis and univariate analysis to identify potential biomarkers of the disease [28–30, 47, 49]. For multivariate analysis, imported datasets are Pareto scaled, and principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and orthogonal partial least squares-discriminant analysis (OPLS-DA) are used for modeling the differences between the study groups. The robustness of the generated models is monitored by the fitness of the model (R2X) for PCA (R2Y) for PLS-DA and OPLS-DA and predictive ability (Q2) values. The model yielding large R2X and R2Y (close to 1) and Q2 values of ~0.5 indicates a robust model [53]. Mass ions are responsible for the class separation between diseased and control groups specified using variable importance in projection (VIP) of the generated PLS-DA or OPLS-DA model. Metabolites with VIP scores above 1.0 are considered important for the model and responsible for differentiating between samples [54]. An unpaired two-tailed Student's *t*-test is used for univariate analysis to identify significantly altered metabolites between the two compared groups. False discovery rate (FDR)-corrected *p*-value of less than 0.05 is considered significant.

5 Approaches Followed in Metabolomics for Biomarker Discovery

The key to discovering new biomarkers depends on the quantification, relative or absolute, of the changes in the levels of the metabolites in the study versus control samples. Two main approaches can be applied in metabolomic analyses: untargeted metabolomics (sometimes referred to as "shotgun metabolomics") and targeted [55, 56].

5.1 Untargeted Approach

Global untargeted metabolomics aims to capture all metabolites in a specific biological sample. In MS-based metabolomics, data can be acquired either in data-dependent acquisition (DDA) or data-independent acquisition (DIA). DDA is the classic LC-MS/MS method for scanning metabolites in metabolomic profiling experiments and generating further MS information for identification. In this approach, MS measures ions in duty cycles. A quick MS1 (the first mass analyzer) survey scan for all the detectable ions will be performed in each duty cycle. Second, the top ions with the highest intensities will be selected for MS2 (the second mass analyzer) analyses. Third, a series of MS2 scans will be done for fragment ions of each preselected target ion [57]. DIA has gained increased interest in recent years due to the advancement in computer algorithms and speed. This data acquisition does not rely on the MS1 survey scans to detect top precursor ions for fragmentation and subsequent product ion scan. Although DIA can increase MS/MS information, it cannot generate high-quality precursor-specific MS2 spectra [58]. For untargeted global metabolomics, high-resolution accurate-mass MS instruments, such as TOF and Orbitrap coupled to quadrupole (Q) mass filter, are preferred [44]. Knowledge of accurate masses facilitates ion identification across different samples using online or commercially available databases.

Most of the metabolomic studies applied in GD used an untargeted approach for biomarker discovery. LC coupled with a QTOF mass spectrometer with DDA or DIA was used to profile metabolites of the amniotic fluid of fetuses with Down syndrome [30], the amniotic fluid of women carrying a fetus with Down syndrome [29], and blood from 292 participants with Down syndrome [40]. On the other hand, a hybrid Q-Orbitrap mass spectrometer was used to identify plasma metabolites implicated in sickle cell disease clinical heterogeneity [22]. Metabolomic discovery untargeted screening was also used to identify metabolites associated with neutrophilic inflammation in bronchoalveolar lavage fluid (BALF) supernatant from preschool children with CF [59] and to identify perturbed metabolites in human sickle erythrocytes compared with human non-sickle erythrocytes [60, 61].

Although identifying specific biomarker candidates is the ultimate goal of the untargeted metabolomic approach, this approach will need better reproducibility and the detection of a high number of false positives in the presence of deficiencies in the experimental design. The latter will have a negative impact on the reproducibility of the method. Therefore, it is essential to have quality control over the analytical method by (a) randomization of the sequence of the samples, (b) analysis of blank and pooled QC samples, (c) monitoring mass accuracy of internal standards during the run, and (d) checking retention time and peak intensity for spiked internal standards [62, 63].

5.2 Targeted Approach

The targeted analysis will focus on a specific set of defined metabolites. Hence, it offers an improved limit of detection (LOD) and limit of quantification (LOQ) compared to global metabolomic methods. The approach can provide absolute quantification (when analyzed with standards) of tens-hundreds of metabolites simultaneously in a highly specific and sensitive manner. Triple quadruple MS instruments are the preferred analyzers in targeted quantification as they offer high sensitivity, particularly when running multiple reaction monitoring (MRM) modes [44, 64]. MRM requires the selection of precursor-to-fragment transitions; MS1 (Q1) will selectively filter ions of a particular mass over charge *m/z* (within a certain resolution) corresponding to the intact targeted metabolite. The precursor ions are then subjected to fragmentation using collision-induced dissociation (CID) in a collision cell (q2). Finally, the fragment ions of the target analyte with high ion intensity and specificity are filtered through MS2 (Q3), generating a signal of the MRM transition [65].

Due to its high specificity and sensitivity, MS-based targeted metabolomics can be used to quantify low-abundance metabolites. It can also be applied for verifying and validating candidate biomarkers, especially in easily accessible blood samples and body fluids. It is an essential step for translating disease biomarkers into clinical practice. Targeted MS metabolomics was employed to quantify extra- and intracellular metabolites in cultured fibroblasts from healthy controls and patients with the GSD Ia, GSD Ib, and GSD III [32], relevant metabolites in BALF and blood samples (spiked with isotopically labeled internal standard) from children with CF [27, 59, 66, 67], and plasma metabolites of normal individuals and patients with sickle cell disease [23].

6 Biomarker Discovery and Validation

A biomarker is a characteristic that can be objectively evaluated and measured as an indicator of a normal biological process, a pathogenic process, or a pharmacologic response to treatment [68]. Biomarkers are essential to early disease detection and provide useful predictive and monitoring information for more precise diagnoses and successful patient treatment and management. With the pivotal roles of biomarkers in clinical applications and the recent development of new technologies, biomarker discovery has been a topic of intensive research [69].

Metabolomics is the closest biomolecule to the phenotype. Therefore, metabolome analysis may be useful for identifying diagnostic and predictive biomarkers. With the growing availability of metabolome analysis, researchers' attention has shifted toward analyzing biofluids and tissues to find alterations in metabolites that can potentially be novel biomarkers of diseases. The general premise was relatively straightforward: identify as many metabolites as possible in a particular kind of

biofluid obtained from diseased patients and compare them to those found in healthy individuals matched for age, sex, and other factors. Once potential candidates are identified; the key step is validation, which produces disease biomarkers for eventual clinical applications.

Metabolomic biomarker discovery has led to identifying numerous potential candidates for disease diagnosis, prognosis, and prediction of response to therapy. The field of metabolomics has significantly developed over the last 10–15 years. However, very few potential biomarkers discovered have been validated clinically and are regularly utilized in clinical practice [70]. One challenge that has delayed the translation of the most promising biomarker candidates into a clinical setting is validation [71, 72]. Due to the large physicochemical varieties of the metabolome, validating metabolite characteristics is still regarded as a bottleneck [4]. Potential biomarkers should be further validated with large-scale studies [69]. However, one of the obstacles in biomarker validation is the limited access to large and independent cohort samples [72].

At the discovery level, metabolomic technologies permit only a few (20–100) samples to be analyzed in each class sampled from two independent populations. Then, the study is repeated at the internal validation stage to validate the previous findings from the discovery study. This step is needed to see if the proposed biomarker can distinguish disease states in a cohort similar to the discovery cohort. In the external validation stage, a larger sample size in the hundreds to thousands is analyzed; however, only a small number (e.g., one to ten) of metabolites are measured [73, 74]. One crucial issue often overlooked in biomarker development and validation is proper patient recruitment, simply because more samples need to be analyzed in the biomarker validation stage than in the biomarker discovery stage. During external validation, the biomarker must be evaluated in a population typical of the population for whom the test is designed; thus, ethnicity and place of origin should be considered. In addition, a suitable control group for the biomarker purpose must also be established [70]. Notably, the test and control groups must have precisely the same collection conditions with similar demographics. Confounding variables, such as sex and age, should always be matched; other confounders, such as experimental conditions, diet, and drug interactions, should also be controlled during the discovery and validation of a metabolomic biomarker [75, 76].

After the discovery phase, targeted studies can be carried out to validate the results. However, in patients with nonspecific phenotypes, untargeted metabolomics shows promise as a validation tool for variants of unknown significance in candidate genes for genetic disorders [77]. Moreover, biomarker validation studies can be performed using affinity multiplexing assays [78]. In a second independent study, NMR spectroscopy was used to validate metabolites from exhaled breath condensate of patients with unstable CF, stable CF, and healthy subjects [79]. Purine metabolism and protein catabolism pathways have been found as biomarkers for neutrophilic airway inflammation in CF patients.

Furthermore, these multiple metabolic pathways were validated by an independent cohort [80]. Calprotectin and urinary glucose tetrasaccharide for hepatic GSDs and Pompe disease are examples of biomarkers approved for clinical use in

GD. Nevertheless, most of the reported biomarkers for GSD were performed with few patients and needed clinical validation to transfer to routine use [81].

7 Application of Metabolomics in Disease Biomarker Discovery in Genetic Disorders

One of the greatest and ongoing challenges in genetics is the ability to predict the phenotypes of individuals from their genotypes. In GD, the genetic variants may alter an individual's metabolome; this link between genetic variants with changes in the metabolome may help predict novel phenotypes [82]. Metabolomics differs from other "omics" sciences in linking gene and environmental interactions. As a result, scientists may investigate gene-environment interactions by studying metabolites and metabolism [83, 84].

Our focus in this chapter will be on cystic fibrosis, Down syndrome, sickle cell anemia, and glycogen storage disorders.

7.1 Cystic Fibrosis

Cystic fibrosis (CF) is an autosomal recessive disease caused by mutations in the gene that codes for the cystic fibrosis transmembrane conductance regulator (CFTR) [85]. More than 2000 mutations have been identified and reported in the CFTR1 (CF Mutation Database) [86]. CF is a disease that affects various organ systems ranging from lung illness to liver disease and poor response to current treatments [87]. Although CF-causing mutations in the CFTR gene are known, there is significant clinical variability in phenotype level, and the etiology of many symptoms is unknown. Therefore, metabolic investigations were carried out to provide a deep understanding of how CFTR mutations cause disease consequences and aid in finding novel biomarkers [17]. A summary of selected studies and their main findings is presented in Table 1.

In a study, Quinn et al. [88] compared metabolite levels in the sputum of participants known to have CF with exacerbation during treatment and posttreatment and in stable groups [88]. They utilized LC-MS/MS to evaluate the impact of exacerbations on the sputum metabolome. Platelet-activating factor (PAF) and a related monacylglycerophosphocholine lipid were identified as possible biomarkers of cystic fibrosis pulmonary exacerbations (CFPE) [88]. Targeted metabolomic profiling on 39 dried blood spot (DBS) samples identified sorbitol as an important indicator for the mucoviscidosis seen in patients with CF. Esther et al. [59] conducted a study with 20 BALF samples analyzed by LC-MS/MS and GS-MS and then validated 34 extra BALF samples with targeted MS [59]. They verified that metabolites associated with adenyl purine metabolism, dipeptides, cellular energy, and lipids

Table 1 A summary of the literature on applying metabolomics for biomarker discovery in CF

Sample	Method	Sample size	Main findings	References
Sputum	LC-MS/MS	11	Platelet-activating factor (PAF) and a related monacylglycerophosphocholine lipid are elevated in CF patients with pulmonary exacerbations	Quinn et al. [88]
DBS	GC/MS and LC-MSMS	69	26 metabolites were significantly differentially expressed and characterized by amino acid, glycolysis, mitochondrial, peroxisomal metabolism, and sorbitol pathways	Al-Qahtani et al. [27]
			CF patients have a significantly lower level of osmolyte (sorbitol) than healthy individuals, indicating that their sorbitol pathway is disturbed, which may explain the mucoviscidosis found in those with CF	
BALF	GC/MS and LC-MSMS	20/ 34 for validation	Early structural lung disease was predicted by findings that involved the catabolism of proteins, oxidative stress, and the metabolism of adenosine products. The enzymes associated with adenosine metabolism were also elevated in those with early disease samples. Metabolites and pathways altered with neutrophilic inflammation and destructive lung disease	Esther et al. [59]
Serum	A chemical isotope labeling liquid chromatography-mass spectrometry	69	Metabolites changed between CF mutational classes II–VI and III–IV. The highly sensitive biomarkers for CF, 3,4-dihydroxymandelate-3-*O*-sulfate and 5-aminopentanoic acid, were found in all three analyses	Masood et al. [66]
Serum	GC/MS and LC-MSMS	62	Several pathways were found to be different in CF, indicating decreased activity in the oxidation of fatty acids CF has lower ketone bodies, lower medium chain carnitines, higher dicarboxylic acids, and lower 2-hydroxybutyrate from amino acid metabolism	Joseloff et al. [26]

(continued)

Table 1 (continued)

Sample	Method	Sample size	Main findings	References
Primary human airway epithelial cell cultures	Untargeted metabolomic/ UHLC/MS/MS	Three separate cohorts	The levels of purine nucleotides were significantly reduced in CF cells. Decreased glucose metabolism in CF cells may exacerbate oxidative stress and limit epithelial cell response to environmental stress	Wetmore et al. [25]
DBS	Nontargeted capillary electrophoresis-mass spectrometry (CE-MS)	152	N-glycated amino acids, oxidized glutathione disulfide, and nicotinamide were differentially expressed in normal birth weight CF neonates without meconium ileus compared to the control group	DiBattista et al. [89]
Sputum	Untargeted LC/ HRMS	34	*Pseudomonas aeruginosa* “metabotypes,” antibiotic resistance and virulence phenotypes, and clinical exacerbations were significantly associated	Moyne et al. [90]
Sweat	Nontargeted capillary electrophoresis-mass spectrometry (CE-MS)	68	Several metabolites associated with asymptomatic CF infants were identified in sweat, including asparagine and glutamine. Both pilocarpic acid, a synthetic sweat stimulant, and mono(2-ethylhexyl) phthalic acid, a natural sweat stimulant, were secreted in significantly lower concentrations in CF infants than in unaffected CF screen positive controls	Macedo et al. [28]
Exhaled breath condensate	UPLC-MS	35	4-hydroxycyclohexylcarboxylic acid and pyroglutamic acid were used to distinguish acute pulmonary exacerbation (APE) samples from stable CF samples. Lactic acid and pyroglutamic acid accurately distinguished pre-APE samples from stable CF samples and matched the APE signature when projected onto the APE vs. stable CF model. Post-APE samples had a metabolomic signature more similar to stable CF samples	Zang et al. [92]
Serum	UPLC-MS	30	Tryptophan-kynurenine, nitric oxide, bile acids, and microbial-derived amino acid metabolites were altered in serum to distinguish the pre- from post-exacerbation state	Muhlebach et al. [41]

Table 1 (continued)

Sample	Method	Sample size	Main findings	References
BALF	NMR	11	Alteration in amino acids and lactate may help distinguish the high- from the low-inflammation groups	Wolak et al. [91]
Plasma	UPLC-MS	52	Fatty acid, amino acid, and carbohydrate metabolism differed between baseline vitamin D and placebo-treated CF patients	Alvarez et al. [24]
			The amino acid pathways in CF patients treated with vitamin D3 versus placebo were altered. Several tricarboxylic acid cycle intermediates increased, while amino acid-related metabolites decreased in the placebo group but not in the vitamin D3 group	
Exhaled breath	GC-TOF-MS	105	Volatile organic compounds (VOCs) such as C16 polyunsaturated hydrocarbon in the exhaled breath of CF patients can distinguish between CF and non-CF patients and between CF patients with and without *Pseudomonas*	Robroeks et al. [93]

pathways are altered with neutrophilic inflammation and destructive lung disease [59]. Metabolites related to amino acids, di-, and tri-peptides, glutathione, glutamine, glutamate, and arginine metabolism pathways have also been described as potential biomarkers among the CF mutational classes (II–VI) and between the class III and IV through analyses of serum samples by chemical isotope-labeled MS-based metabolomic approach [66]. Joseloff et al. [26] applied LC-MS/MS and GC-MS in two groups of children with CF and non-CF lung disease. They reported an alteration in cellular energy metabolism in CF, potentially reflecting mitochondrial dysfunction, which may be useful in differentiating CF from non-CF lung diseases [26]. Metabolomic profiling showed differences in CF and non-CF primary lung epithelial cells [25]. The levels of glucose, sorbitol, glycerophosphocholine, and various glycolytic intermediates, including glucose 6-phosphate, fructose 6-phosphate, and lactate, were significantly reduced in CF cells compared with the non-CF cells [25].

Many studies have revealed that different biomarkers, based on metabolite profiles, could be detected in a range of biological samples, including blood [89], sputum [88, 90], BALF [80, 91], exhaled breath condensate (EBC) [92, 93], cell culture [25], sweat [28], plasma [24], and serum [26, 41, 66]. Biomarkers in the polyamine (e.g., putrescine, spermidine) metabolism highlighted significant associations between *Pseudomonas aeruginosa* “metabotypes,” expression of antibiotic resistance and virulence phenotypes, and frequency of clinical exacerbations in CF patients [90]. The level of ophthalmic acid was found to be downregulated, along

with amino acids (serine, threonine, proline, and glycine) present in neonatal DBS [89]. Mono(2-ethylhexyl) phthalic acid in sweat samples from positive CF infants was differentially downregulated compared to non-CF [28]. Zang et al. [92] reported a distinctive profile for CF in exhaled breath condensate, including two metabolic discriminant characteristics, 4-hydroxycyclohexylcarboxylic acid and pyroglutamic acid, between APE and stable CF samples. Moreover, lactic acid and pyroglutamic acid were found to differentiate stable CF from pre-APE [92].

7.2 Down Syndrome

Down syndrome (DS) is the most common chromosomal-related GD. The presence of an extra-human copy of chromosome 21 characterizes DS. DS patients are characterized by the facial appearance and various complications, including intellectual disabilities, mental and growth retardation, vision problems, hearing loss, infections, hypothyroidism, blood disorders, and cardiovascular abnormalities [94, 95]. Even though the anatomical and physiological abnormalities in DS are well known, and the genetic etiology of DS has been identified, understanding the exact cellular mechanisms linking genotype to phenotype is the major challenge. With the presence of well-established procedures for testing DS, such as ultrasonographic methods and amniocentesis, which looks for chromosome disorders, the diagnosis of DS based adequately on reliable biomarkers is underdeveloped [30, 96].

DS genotype affects the functional phenotype leading to changes in metabolomic profiles. Therefore, metabolomics seems to be a promising tool for identifying disease-related biomarkers. To date, few metabolomic studies have been conducted to study and discover novel biomarkers in DS disease. The urinary metabolome of 122 maternal urines has been analyzed using LC-MS and revealed that dihydrouracil was significantly elevated in the urine of women with a DS-affected pregnancy. In contrast, progesterone was decreased in a DS-affected pregnancy compared with normal pregnancies [97].

Parfieniuk et al. [29] reported significant differences in the level of methylhistidine, hexanoylcarnitine, diacetylspermine, and *p*-cresol sulfate when comparing amniotic fluid of 13 women with fetal DS with 13 healthy fetuses [29]. Diaz et al. [98] conducted a study on urine levels of 2-ketoglutarate, 1-methylhistidine, 3-hydroxybutyrate, 4-OH-hippurate, and dimethylamine and were able to distinguish the pregnant women carrying a baby with a chromosomal abnormality from the control group using NMR method [98]. Furthermore, alterations of cortisol levels, free amino acids (arginine, histidine, and glutamate), and pregnenolone sulfate in amniotic fluid from fetuses with DS were observed and validated in a study conducted by Haung et al. [30] using UPLC-MS. NMR-based metabolomic profiles of plasma samples from 129 people with DS and 46 healthy controls showed a dramatic difference in the 7 metabolites that may be used to distinguish between the DS and healthy groups [30]. However, the metabolomic patterns examined cannot be linked to the degree of intellectual disability (ID) [48]. Caracausi et al. [47]

analyzed the plasma and urine metabolome from 67 DS patients and 29 healthy controls using NMR and reported distinguished metabolic patterns in DS attributable mainly to mitochondrial dysmetabolism [47]. Moreover, a study using untargeted metabolomic analysis of amniotic fluid samples from women having normal and DS fetuses identified alterations in several steroid hormones and their derivatives, glutathione catabolites coupled, gamma-glutamyl amino acids, phospholipid catabolites, sugars, and dicarboxylic acids [99].

7.3 Sickle Cell Disease

Sickle cell disease (SCD) is a set of red blood cell hereditary disorders caused by structural hemoglobin (Hb) abnormalities known as sickle hemoglobin (HbS). A single nucleotide change in the gene-producing β-globin is the cause of SCD. The sixth position of the β-globin in hemoglobin S (HbS) is replaced by valine instead of glutamic acid. SCD is caused by a homozygous HbS condition (HbSS) or is the result of having inherited HbS with additional hemoglobin mutations such as beta0 thalassemia (HbS-beta0 that), which refers to the absence of production of beta-globin, HbC (HbSC), or beta+ thalassemia mutations (HbS-beta+thal) where the beta gene makes low levels of globin [100]. Sickle cell anemia (SCA) is a condition where a mutation in the β-globin gene on chromosome 11 may occur. This is an autosomal recessive condition. The sickle cell trait is characterized by the appearance of long polymers of deoxygenated HbS (deoxyHbS), resulting in sickle-shaped erythrocytes and the eventual vascular hemolysis [100]. SCD symptoms include anemia, acute chest syndrome, stroke, transient ischemia events, severe vaso-occlusive discomfort, severe pain, and splenic sequestration. In addition, SCD may cause problems in the central nervous system, lungs, kidneys, and gastrointestinal system [100, 101].

Currently, metabolomics is aiding scientists in precisely measuring functional phenotypes that arise from changes in genomic, transcriptomic, and proteomic levels. Therefore, this approach can be applied to identify new potential candidates' diagnostic, prognostic, and therapeutic biomarkers for SCD. In this context, several studies concerning the application of metabolomics in SCD have been conducted. Dembélé et al. [102] presented the results of metabolomic profiling of patients experiencing vaso-occlusive crises compared to their sickle cell disease baselines. To determine the differences between these two disease states, a standardized targeted metabolomic method was used for samples from 40 individuals, including plasma and erythrocyte fractions. They found that metabolic signatures in the plasma were especially notable for their differences in nitric oxide metabolism, which hints at connections with pain. In addition, during the crisis, red blood cells had extensive alterations in phospholipids, indicating significant membrane remodeling [102].

Several metabolomic studies were conducted on the transgenic mouse model of SCD. A list of 251 metabolites associated with 8 pathways, including glycolysis,

the pentose phosphate pathway (PPP), amino acids, nucleotides, xenobiotics, lipids, fatty acids, and carbohydrates, were significantly changed in SCD mouse blood [23]. In addition, the elevation of several metabolites was reported in SCD blood mice, for instance, nucleosides (including adenosine), lipids (such as sphingosine-1-phosphate [S1P], lysophospholipids, and free acyl fatty acids), and glycolytic intermediates (such as 2,3-BPG) [23, 103, 104].

Analysis of the red blood cell metabolome from 28 adult patients with the HbSS (hemoglobin SS) genotype in a steady state and comparing it to 24 healthy adults (HbAA) showed that 31 metabolites in key metabolic pathways (e.g., glycolysis, pentose phosphate, glutathione, ascorbate, polyamines, carnitine, creatine, and other amino acids) were significantly altered [60].

7.4 Glycogen Storage Disorders

Glycogen storage disorders (GSDs) are a group of metabolic abnormalities in glycogen metabolism resulting from deficiencies in glycogen production or degradation enzymes (*Inborn Metabolic Diseases: Diagnosis and Treatment*). GSDs are rare GD that usually primarily influence the liver, muscles, or the two.

Until now, studies conducted to find clinical biomarkers for GSDs have been limited. Among the 19 types of GSDs classified based on enzyme deficiency and affected tissue, glycogen storage disease type Ia (GSDIa, von Gierke disease) is considered the most common type that occurs in about 1 out of every 100,000 live births [105]. Tamara et al. [106] noted alterations in energy production pathways, such as the tricarboxylic acid cycle, creatine metabolism, urea cycle, amino acid, purine/pyrimidine metabolism, and enzyme cofactors, such as biotin. This study was conducted on 14 plasma samples from adult GSDI patients compared to 31 healthy controls utilizing LC-MS/MS [106]. Similar findings were reported by Farah et al. [31] using cell culture and a mouse model of GSDIa. The previous study reported oxidative phosphorylation dysfunction, abnormalities in TCA cycle metabolites, reduced mitochondrial membrane potential, and dysfunctional mitochondrial ultrastructure [31].

To differentiate between metabolite profiles from GSDI, GSD III subtypes, and healthy controls, cultured fibroblasts from the three groups were analyzed by targeted LCMS/MS [25]. Malfunctions in energy production pathways (glycolysis, Krebs cycle, succinate) and decreased creatinine and antioxidant defense of the cysteine and glutathione systems in GSDIa and GSD IIII have been detected [25]. For GSDII (Pompe disease), urine glucose tetrasaccharide has been authorized for clinical usage [107]. However, most of the discovered biomarkers for GSDs lack clinical validation.

A summary of selected studies and their main findings for DS, SCD, and GSDs are presented in Table 2.

Table 2 A summary of the literature on applying metabolomics for biomarker discovery in Down syndrome (DS), sickle cell disease (SCD,) and glycogen storage disorders (GSDs)

Sample	Method	Sample size	Main findings	References
1. Down syndrome (DS)				
Urine	ZIC-HILIC MS	122	Dihydrouracil is elevated in the urine of women with a DS-affected pregnancy, while progesterone is decreased	Trivedi et al. [97]
Amniotic fluid	LC-MSMS	26	There were significant differences in the levels of four metabolites: methylhistidine, hexanoylcarnitine, diacetylspermine, and *p*-cresol sulfate, which may be linked to the improper nervous system and muscle development	Parfieniuk et al. [29]
Urine	NMR spectroscopy	300	Urine levels of 2-ketoglutarate, 1-methylhistidine, 3-hydroxybutyrate, 4-OH-hippurate, and dimethylamine have been shown to differentiate pregnant women carrying a baby with a chromosomal abnormality from the control group	Diaz et al. [98]
Amniotic fluid	LC-QTOF-MS	50	Significant differences between DS fetuses and controls in porphyrin, bile acids, amino acids, and hormone metabolites such as taurochenodeoxycholate, L-arginine, and taurocholate. They also found differences in L-histidine and glycocholic acid metabolites	Huang et al. [30]
Plasma	NMR spectroscopy	175	Alterations in the level of 7 metabolites may be used to distinguish between the DS and healthy groups. No association between the differences in metabolites and the degree of ID	Antonaros et al. [48]
Plasma and urine	NMR spectroscopy	96	Several significantly altered metabolites of Down syndrome correlate with alteration of mitochondrial metabolism	Caracausi et al. [47]
Amniotic fluid	UPLC-MS	42	There were many different metabolites found in the amniotic fluid of DS pregnancies compared to normal pregnancies, including lower levels of several steroid hormones and their derivatives, higher levels of glutathione catabolites, and lower levels of gamma-glutamyl amino acids	Liu et al. [99]

(continued)

Table 2 (continued)

Sample	Method	Sample size	Main findings	References
2. Sickle cell disease (SCD)				
Red blood cells and plasma	Targeted quantitative using LC-MS/MS	40	The involvement of metabolites not previously identified in sickle cell disease, such as hexoses, acyl-carnitines, and -aminoadipate all help explain how sickle cell disease affects energetic metabolism	Dembélé et al. [102]
Red blood cells and plasma	LC/GC-MS	44 human and 6 mice	Erythrocyte sphingosine kinase 1 SPHK1-mediated elevation of sphingosine-1-phosphate (S1P) contributes to sickling and disease progression	Zhang et al. [23]
Whole blood and plasma	LC/GC-MS	37 human and 12 mice	Increased erythrocyte cytosolic phospholipase A2 (cPLA2) is a key factor in the imbalanced lands cycle seen in SCD erythrocytes, as well as the increased erythrocyte LysoPC and circulating arachidonic acid seen in SCD mice	Wu et al. [103]
Whole blood	UHPLC-MS	52	The concentrations of reduced glutathione (GSH) and oxidized glutathione (GSSG) were reduced in HbS cells, while their precursors (glutamine, glutamate, and glycine) were increased. *N*-acetylglutathione (NAG) was found to be decreased in HbS cells, as were two ascorbate metabolism metabolites, diketogulonic acid and threonolactone	Darghouth et al. [60]
3. Glycogen storage disorders (GSDs)				
Plasma	LC-MS/MS	45	Changes in metabolites in various metabolic pathways, such as the tricarboxylic acid cycle, creatine metabolism, urea cycle, amino acid, and purine/pyrimidine metabolism, as well as enzyme cofactors such as biotin these metabolic changes were seen in both GSD subtypes (Ia and Ib)	Tamara et al. [106]
Skin fibroblast cell lines	Targeted LCMS/MS	Skin fibroblasts	Energy production pathways (glycolysis, Krebs cycle, succinate) were dysfunctional in GSDIa and GSD III	Hannibal et al. [32]
Urine	UPLC-MS/MS	79	Urinary glucose tetrasaccharide (Glc4) determination may be useful in the follow-up of a positive newborn Pompe disease screening result. A high Glc4 indicates an infantile phenotype	Chien et al. [107]

8 Current Trends and Future Perspective

The field of metabolomics has recently seen a rapid expansion in clinical research, with goals ranging from elucidating disease pathogenesis to discovering clinical biomarkers. As a result, a new technology utilizing NMR and MS to assess metabolite profiles in clinical samples has been introduced to detect the disease's biomarker molecules and metabolic effects or its treatment [108].

It is difficult to determine GD's pathogenic mechanisms because of their complexity and effects on multiorgan systems. Furthermore, it appears that GD does not have a single biomarker that can distinguish between patients and accurately reflect the pathology of the disease. In this context, metabolomic analysis of clinical samples may be an excellent tool for distinguishing biomarkers from therapeutic targets [109].

Recently, the field of metabolomics has made significant contributions to studying GD. This is due to the ability of metabolomics, in contrast to other "omics" sciences, to link gene and environmental interactions. It represents the genome's downstream output and the upstream input of the surrounding environment [110]. As a result of its ability to detect subtle changes in large datasets through comprehensive metabolic measurements, metabolomics is a practical application in GD for clinical biomarker discovery [111].

Technological advancements in MS have made it possible to detect tens of thousands of signals in complex biological systems, allowing the detection of a wide range of metabolites in a single test and opening up new opportunities in the precision and personalized medicine [112]. Computing tools, databases, and big data analytics are evolving quickly, allowing a complete annotation and identification of these signals. Thus, it will be possible to discover new molecules, classes of compounds, or metabolism pathway configurations that have not previously been considered in studies of GD. Public databases for metabolomics are being established, and they will apply to a wide range of GD and therapeutic areas shortly.

Current biomarker discovery strategies on GD frequently rely on identifying alterations in metabolites and the association of these altered metabolites with a specific disease. When it comes to GD, the exact mechanism of these metabolites and the functional role of metabolite biomarkers are frequently unknown and understudied. An integrated biomarker discovery platform must be supplemented with genomic, transcriptomic, and proteomic data to be effective.

Despite significant progress in metabolomic research over the past decade, most identified biomarkers still need to replace existing clinical tests. A potential biomarker must be confirmed and validated using hundreds of specimens before being used in a clinical setting. It must also be reproducible, specific, and sensitive to be approved. Robust experimental designs, data acquisition, data mining, and biomarker validation must be performed to successfully translate the results of a metabolomic experiment to clinical use in GD. In addition, future methodology development in sample preparation and analysis is still necessary to achieve comprehensive metabolome studies.

9 Conclusions

Genetic diseases have widely varied in pathophysiological consequences and severity of the disease. Currently, the available treatments and markers for properly managing GD complications are limited. Additionally, the diagnosis and follow-up of GD still depend on examining chromosomes or DNA and routine laboratory tests, which are invasive or nonspecific. Hence, there is a high demand to identify and validate novel biomarkers that may have diagnostic, prognostic, and therapeutic values, improving the management and treatment of GD.

Metabolomic research provides insight into the human metabolome under specific disease states. Recently, metabolomics has shown promise as a readout of genes and the environment at a particular time. Developing a combination of disease biomarkers by metabolomic study is a promising approach, particularly because many of these diseases are heterogeneous and multifactorial. However, identified potential biomarkers should be validated (analytical and clinical validation) and confirmed using an independent sample set to ensure that they are specific to the disease state and are detected due to variability within the biological sample of patients. The ability of metabolomic studies to produce a detailed disease characterization allows personalized disease management and treatment, which are the basis of precision medicine.

References

1. Boycott KM, Hartley T, Biesecker LG, Gibbs RA, Innes AM, Riess O, et al. A diagnosis for all rare genetic diseases: the horizon and the next frontiers. Cell. 2019;177(1):32–7.
2. Jackson M, Marks L, May GHW, Wilson JB. The genetic basis of disease. Essays Biochem. 2018;62(5):643–723.
3. Ropers HH. New perspectives for the elucidation of genetic disorders. Am J Hum Genet. 2007;81(2):199–207.
4. Zhou B, Xiao JF, Tuli L, Ressom HW. LC-MS-based metabolomics. Mol BioSyst. 2012;8(2):470–81.
5. Rinschen MM, Ivanisevic J, Giera M, Siuzdak G. Identification of bioactive metabolites using activity metabolomics. Nat Rev Mol Cell Biol. 2019;20(6):353–67.
6. Tolstikov V, Moser AJ, Sarangarajan R, Narain NR, Kiebish MA. Current status of metabolomic biomarker discovery: impact of study design and demographic characteristics. Metabolites. 2020;10(6):224.
7. Zhang AH, Sun H, Wang XJ. Saliva metabolomics opens the door to biomarker discovery, disease diagnosis, and treatment. Appl Biochem Biotechnol. 2012;168(6):1718–27.
8. Dahabiyeh LA, Malkawi AK, Wang XH, Colak D, Mujamammi AH, Sabi EM, et al. Dexamethasone-induced perturbations in tissue metabolomics revealed by chemical isotope labeling LC-MS analysis. Metabolites. 2020;10(2):42.
9. Dahabiyeh LA, Mahmoud NN, Al-Natour MA, Safo L, Kim DH, Khalil EA, et al. Phospholipid-gold nanorods induce energy crisis in MCF-7 cells: cytotoxicity evaluation using LC-MS-based metabolomics approach. Biomolecules. 2021;11(3):364.

10. Dahabiyeh LA, Mujammami M, Arafat T, Benabdelkamel H, Alfadda AA, Rahman AMA. A metabolic pattern in healthy subjects given a single dose of metformin: a metabolomics approach. Front Pharmacol. 2021;12:705932.
11. Aranibar N, Borys M, Mackin NA, Ly V, Abu-Absi N, Abu-Absi S, et al. NMR-based metabolomics of mammalian cell and tissue cultures. J Biomol NMR. 2011;49(3–4):195–206.
12. Ren JL, Zhang AH, Kong L, Wang XJ. Advances in mass spectrometry-based metabolomics for investigation of metabolites. RSC Adv. 2018;8(40):22335–50.
13. Pallares-Mendez R, Aguilar-Salinas CA, Cruz-Bautista I, del Bosque-Plata L. Metabolomics in diabetes, a review. Ann Med. 2016;48(1–2):89–102.
14. Razquin C, Toledo E, Clish CB, Ruiz-Canela M, Dennis C, Corella D, et al. Plasma lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. Diabetes Care. 2018;41(12):2617–24.
15. Chen ZH, Li ZH, Li HR, Jiang Y. Metabolomics: a promising diagnostic and therapeutic implement for breast cancer. Onco Targets Ther. 2019;12:6797–811.
16. Chen XL, Zhu Y, Jijiwa M, Nasu M, Ai JM, Dai SM, et al. Identification of plasma lipid species as promising diagnostic markers for prostate cancer. BMC Med Inform Decis Mak. 2020;20:223.
17. Liessi N, Pedemonte N, Armirotti A, Braccia C. Proteomics and metabolomics for cystic fibrosis research. Int J Mol Sci. 2020;21(15):5439.
18. Quezada H, Guzman-Ortiz AL, Diaz-Sanchez H, Valle-Rios R, Aguirre-Hernandez J. Omics-based biomarkers: current status and potential use in the clinic. Bol Med Hosp Infant Mex. 2017;74(3):219–26.
19. Robinette SL, Holmes E, Nicholson JK, Dumas ME. Genetic determinants of metabolism in health and disease: from biochemical genetics to genome-wide associations. Genome Med. 2012;4:30.
20. Goossens N, Nakagawa S, Sun XC, Hoshida Y. Cancer biomarker discovery and validation. Transl Cancer Res. 2015;4(3):256–69.
21. Wishart DS. Metabolomics for investigating physiological and pathophysiological processes. Physiol Rev. 2019;99(4):1819–75.
22. Ilboudo Y, Garrett ME, Bartolucci P, Brugnara C, Clish CB, Hirschhorn JN, et al. Potential causal role of L-glutamine in sickle cell disease painful crises: a Mendelian randomization analysis. Blood Cells Mol Dis. 2021;86:102504.
23. Zhang YJ, Berka V, Song AR, Sun KQ, Wang W, Zhang WR, et al. Elevated sphingosine-1-phosphate promotes sickling and sickle cell disease progression. J Clin Investig. 2014;124(6):2750–61.
24. Alvarez JA, Chong EY, Walker DI, Chandler JD, Michalski ES, Grossmann RE, et al. Plasma metabolomics in adults with cystic fibrosis during a pulmonary exacerbation: a pilot randomized study of high-dose vitamin D-3 administration. Metabolism. 2017;70:31–41.
25. Wetmore DR, Joseloff E, Pilewski J, Lee DP, Lawton KA, Mitchell MW, et al. Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells. J Biol Chem. 2010;285(40):30516–22.
26. Joseloff E, Sha W, Bell SC, Wetmore DR, Lawton KA, Milburn MV, et al. Serum metabolomics indicate altered cellular energy metabolism in children with cystic fibrosis. Pediatr Pulmonol. 2014;49(5):463–72.
27. Al-Qahtani W, Jabar MA, Masood A, Jacob M, Nizami I, Dasouki M, et al. Dried blood spot-based metabolomic profiling in adults with cystic fibrosis. J Proteome Res. 2020;19(6):2346–57.
28. Macedo AN, Mathiaparanam S, Brick L, Keenan K, Gonska T, Pedder L, et al. The sweat metabolome of screen-positive cystic fibrosis infants: revealing mechanisms beyond impaired chloride transport. ACS Cent Sci. 2017;3(8):904–13.
29. Parfieniuk E, Pietrowska K, Samczuk P, Kretowski A, Ciborowski M, Zbucka-Kretowska M. Amniotic fluid metabolic fingerprinting indicated metabolites which may play a

role in the pathogenesis of foetal Down syndrome - a preliminary report. Ginekol Pol. 2021;92(3):188–94.
30. Huang J, Mo JH, Zhao GL, Lin QY, Wei GH, Deng WN, et al. Application of the amniotic fluid metabolome to the study of fetal malformations, using Down syndrome as a specific model. Mol Med Rep. 2017;16(5):7405–15.
31. Farah BL, Sinha RA, Wu YJ, Singh BK, Lim A, Hirayama M, et al. Hepatic mitochondrial dysfunction is a feature of Glycogen Storage Disease Type Ia (GSDIa). Sci Rep. 2017;7:44408.
32. Hannibal L, Theimer J, Wingert V, Klotz K, Bierschenk I, Nitschke R, et al. Metabolic profiling in human fibroblasts enables subtype clustering in glycogen storage disease. Front Endocrinol. 2020;11:579981.
33. Tang XH, Lin CC, Spasojevic I, Iversen ES, Chi JT, Marks JR. A joint analysis of metabolomics and genetics of breast cancer. Breast Cancer Res. 2014;16(4):415.
34. Long ZP, Zhou JD, Xie K, Wu Z, Yin HH, Daria V, et al. Metabolomic markers of colorectal tumor with different clinicopathological features. Front Oncol. 2020;10:981.
35. Yusof HM, Ab-Rahim S, Suddin LS, Saman MSA, Mazlan M. Metabolomics profiling on different stages of colorectal cancer: a systematic review. Malays J Med Sci. 2018;25(5):16–34.
36. Alvarez-Sanchez B, Priego-Capote F, de Castro MDL. Metabolomics analysis II. Preparation of biological samples prior to detection. Trends Analyt Chem. 2010;29(2):120–7.
37. Vuckovic D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. Anal Bioanal Chem. 2012;403(6):1523–48.
38. Pinu FR, Villas-Boas SG. Extracellular microbial metabolomics: the state of the art. Metabolites. 2017;7(3):43.
39. Brauer R, Leichtle AB, Fiedler GM, Thiery J, Ceglarek U. Preanalytical standardization of amino acid and acylcarnitine metabolite profiling in human blood using tandem mass spectrometry. Metabolomics. 2011;7(3):344–52.
40. Mapstone M, Gross T, Macciardi F, Cheema A, Petersen M, Head E, et al. Metabolic correlates of prevalent mild cognitive impairment and Alzheimer's disease in adults with Down syndrome. Alzheimers Dement (Amst). 2020;12(1):e12028. https://doi.org/10.1002/dad2.12028.
41. Muhlebach MS, Sha W, MacIntosh B, Kelley TJ, Muenzer J. Metabonomics reveals altered metabolites related to inflammation and energy utilization at recovery of cystic fibrosis lung exacerbation. Metabol Open. 2019;3:100010. https://doi.org/10.1016/j.metop.2019.100010.
42. Dahabiyeh LA. The discovery of protein biomarkers in pre-eclampsia: the promising role of mass spectrometry. Biomarkers. 2018;23(7):609–21.
43. Rahman AMA, Pawling J, Ryczko M, Caudy AA, Dennis JW. Targeted metabolomics in cultured cells and tissues by mass spectrometry: method development and validation. Anal Chim Acta. 2014;845:53–61.
44. Steuer AE, Brockbals L, Kraemer T. Metabolomic strategies in biomarker research-new approach for indirect identification of drug consumption and sample manipulation in clinical and forensic toxicology? Front Chem. 2019;7:319.
45. Wei R, Li GD, Seymour AB. High-throughput and multiplexed LC/MS/MRM method for targeted metabolomics. Anal Chem. 2010;82(13):5527–33.
46. Jacob M, Malkawi A, Albast N, Al Bougha S, Lopata A, Dasouki M, et al. A targeted metabolomics approach for clinical diagnosis of inborn errors of metabolism. Anal Chim Acta. 2018;1025:141–53.
47. Caracausi M, Ghini V, Locatelli C, Mericio M, Piovesan A, Antonaros F, et al. Plasma and urinary metabolomic profiles of Down syndrome correlate with alteration of mitochondrial metabolism. Sci Rep. 2018;8:2977.
48. Antonaros F, Ghini V, Pulina F, Ramacieri G, Cicchini E, Mannini E, et al. Plasma metabolome and cognitive skills in Down syndrome. Sci Rep. 2020;10(1):10491.
49. Elsherif L, Pathmasiri W, McRitchie S, Archer DR, Ataga KI. Plasma metabolomics analysis in sickle cell disease patients with albuminuria - an exploratory study. Br J Haematol. 2019;185(3):620–3.

50. Orozco JS, Hertz-Picciotto I, Abbeduto L, Slupsky CM. Metabolomics analysis of children with autism, idiopathic-developmental delays, and Down syndrome. Transl Psychiatry. 2019;9:243.
51. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics. 2007;3(3):211–21.
52. Sumner LW, Lei ZT, Nikolau BJ, Saito K, Roessner U, Trengove R. Proposed quantitative and alphanumeric metabolite identification metrics. Metabolomics. 2014;10(6):1047–9.
53. Eriksson L, Andersson PL, Johansson E, Tysklind M. Megavariate analysis of environmental QSAR data. Part I - a basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). Mol Divers. 2006;10(2):169–86.
54. Yin PY, Wan DF, Zhao CX, Chen J, Zhao XJ, Wang WZ, et al. A metabonomic study of hepatitis B-induced liver cirrhosis and hepatocellular carcinoma by using RP-LC and HILIC coupled with mass spectrometry. Mol BioSyst. 2009;5(8):868–76.
55. Kind T, Tsugawa H, Cajka T, Ma Y, Lai ZJ, Mehta SS, et al. Identification of small molecules using accurate mass MS/MS search. Mass Spectrom Rev. 2018;37(4):513–32.
56. Ghanbari R, Sumner S. Using metabolomics to investigate biomarkers of drug addiction. Trends Mol Med. 2018;24(2):197–205.
57. Michalski A, Cox J, Mann M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. J Proteome Res. 2011;10(4):1785–93.
58. Boxler MI, Schneider TD, Kraemer T, Steuer AE. Analytical considerations for (un)-targeted metabolomic studies with special focus on forensic applications. Drug Test Anal. 2019;11(5):678–96.
59. Esther CR, Turkovic L, Rosenow T, Muhlebach MS, Boucher RC, Ranganathan S, et al. Metabolomic biomarkers predictive of early structural lung disease in cystic fibrosis. Eur Respir J. 2016;48(6):1612–21.
60. Darghouth D, Koehl B, Madalinski G, Heilier JF, Bovee P, Xu Y, et al. Pathophysiology of sickle cell disease is mirrored by the red blood cell metabolome. Blood. 2011;117(6):E57–66.
61. Adebiyi MG, Manalo JM, Xia Y. Metabolomic and molecular insights into sickle cell disease and innovative therapies. Blood Adv. 2019;3(8):1347–55.
62. Valdes A, Lucio-Cazana FJ, Castro-Puyana M, Garcia-Pastor C, Fiehn O, Marina ML. Comprehensive metabolomic study of the response of HK-2 cells to hyperglycemic hypoxic diabetic-like milieu. Sci Rep. 2021;11(1):5058.
63. Barupal DK, Zhang Y, Shen T, Fan SL, Roberts BS, Fitzgerald P, et al. A comprehensive plasma metabolomics dataset for a cohort of mouse knockouts within the International Mouse Phenotyping Consortium. Metabolites. 2019;9(5):101.
64. Boja ES, Fehniger TE, Baker MS, Marko-Varga G, Rodriguez H. Analytical validation considerations of multiplex mass-spectrometry-based proteomic platforms for measuring protein biomarkers. J Proteome Res. 2014;13(12):5325–32.
65. Yao XD, Bajrami B, Shi Y. Ultrathroughput multiple reaction monitoring mass spectrometry. Anal Chem. 2010;82(3):794–7.
66. Masood A, Jacob M, Gu X, Abdel Jabar M, Benabdelkamel H, Nizami I, et al. Distinctive metabolic profiles between cystic fibrosis mutational subclasses and lung function. Metabolomics. 2021;17(1):4.
67. Zardini Buzatto A, Abdel Jabar M, Nizami I, Dasouki M, Li L, Abdel Rahman AM. Lipidome alterations induced by cystic fibrosis, CFTR mutation, and lung function. J Proteome Res. 2020;20(1):549–64.
68. Fidock M, DeSilva B. Bioanalysis of biomarkers for drug development. Bioanalysis. 2012;4(20):2425–6.
69. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. Biomed Res Int. 2015;2015:354671.

70. Marchand CR, Farshidfar F, Rattner J, Bathe OF. A framework for development of useful metabolomic biomarkers and their effective knowledge translation. Metabolites. 2018;8(4):59.
71. López-López A, López-Gonzálvez A, Barker-Tejeda TC, Barbas C. A review of validated biomarkers obtained through metabolomics. Expert Rev Mol Diagn. 2018;18(6):557–75.
72. Nagana Gowda A, G, Raftery D. Biomarker discovery and translation in metabolomics. Curr Metabolomics. 2013;1(3):227–40.
73. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chem Soc Rev. 2011;40(1):387–426.
74. Usher-Smith J, Harshfield A, Saunders C, Sharp S, Emery J, Walter F, et al. External validation of risk prediction models for incident colorectal cancer using UK Biobank. Br J Cancer. 2018;118(5):750–9.
75. Putri SP, Yamamoto S, Tsugawa H, Fukusaki E. Current metabolomics: technological advances. J Biosci Bioeng. 2013;116(1):9–16.
76. Emwas A-HM, Salek RM, Griffin JL, Merzaban J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. Metabolomics. 2013;9(5):1048–72.
77. Almontashiri NAM, Zha L, Young K, Law T, Kellogg MD, Bodamer OA, et al. Clinical validation of targeted and untargeted metabolomics testing for genetic disorders: a 3 year comparative study. Sci Rep. 2020;10(1):9382.
78. Ali SE, Farag MA, Holvoet P, Hanafi RS, Gad MZ. A comparative metabolomics approach reveals early biomarkers for metabolic response to acute myocardial infarction. Sci Rep. 2016;6(1):36359.
79. Montuschi P, Paris D, Melck D, Lucidi V, Ciabattoni G, Raia V, et al. NMR spectroscopy metabolomic profiling of exhaled breath condensate in patients with stable and unstable cystic fibrosis. Thorax. 2012;67(3):222–8.
80. Esther CR Jr, Coakley RD, Henderson AG, Zhou Y-H, Wright FA, Boucher RC. Metabolomic evaluation of neutrophilic airway inflammation in cystic fibrosis. Chest. 2015;148(2):507–15.
81. Molares-Vila A, Corbalán-Rivas A, Carnero-Gregorio M, González-Cespón JL, Rodríguez-Cerdeira C. Biomarkers in glycogen storage diseases: an update. Int J Mol Sci. 2021;22(9):4381.
82. Swain-Lenz D, Nikolskiy I, Cheng J, Sudarsanam P, Nayler D, Staller MV, et al. Causal genetic variation underlying metabolome differences. Genetics. 2017;206(4):2199–206.
83. Balashova EE, Maslov DL, Lokhov PG. A metabolomics approach to pharmacotherapy personalization. J Pers Med. 2018;8(3):28.
84. Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. Mass Spectrom Rev. 2019;38(3):221–38.
85. Tsui L-C, Dorfman R. The cystic fibrosis gene: a molecular genetic perspective. Cold Spring Harb Perspect Med. 2013;3(2):a009472.
86. Cystic Fibrosis Mutation Database [updated April 25, 2011; cited 2021, available from: http://www.genet.sickkids.on.ca/].
87. Ciuca IM, Marian P, Monica M. Biomarkers in cystic fibrosis lung disease–a review. Rom J Anaesth Intensive Care. 2020;27(2):34.
88. Quinn RA, Lim YW, Mak TD, Whiteson K, Furlan M, Conrad D, et al. Metabolomics of pulmonary exacerbations reveals the personalized nature of cystic fibrosis disease. PeerJ. 2016;4:e2174.
89. DiBattista A, McIntosh N, Lamoureux M, Al-Dirbashi OY, Chakraborty P, Britz-McKibbin P. Metabolic signatures of cystic fibrosis identified in dried blood spots for newborn screening without carrier identification. J Proteome Res. 2019;18(3):841–54.
90. Moyne O, Castelli F, Bicout DJ, Boccard J, Camara B, Cournoyer B, et al. Metabotypes of Pseudomonas aeruginosa correlate with antibiotic resistance, virulence and clinical outcome in cystic fibrosis chronic infections. Metabolites. 2021;11(2):63.

91. Wolak JE, Esther CR Jr, O'Connell TM. Metabolomic analysis of bronchoalveolar lavage fluid from cystic fibrosis patients. Biomarkers. 2009;14(1):55–60.
92. Zang X, Monge ME, McCarty NA, Stecenko AA, Fernández FM. Feasibility of early detection of cystic fibrosis acute pulmonary exacerbations by exhaled breath condensate metabolomics: a pilot study. J Proteome Res. 2017;16(2):550–8.
93. Robroeks CM, van Berkel JJ, Dallinga JW, Jöbsis Q, Zimmermann LJ, Hendriks HJ, et al. Metabolomics of volatile organic compounds in cystic fibrosis patients and controls. Pediatr Res. 2010;68(1):75–80.
94. Gardiner K, Herault Y, Lott IT, Antonarakis SE, Reeves RH, Dierssen M. Down syndrome: from understanding the neurobiology to therapy. J Neurosci. 2010;30(45):14943–5.
95. Bull MJ. Health supervision for children with down syndrome. Am Acad Pediatr. 2011;128:393.
96. Kolialexi A, Tsangaris GT, Papantoniou N, Anagnostopoulos AK, Vougas K, Bagiokos V, et al. Application of proteomics for the identification of differentially expressed protein markers for Down syndrome in maternal plasma. Prenat Diagn. 2008;28(8):691–8.
97. Trivedi DK, Iles RK. Shotgun metabolomic profiles in maternal urine identify potential mass spectral markers of abnormal fetal biochemistry–dihydrouracil and progesterone in the metabolism of Down syndrome. Biomed Chromatogr. 2015;29(8):1173–83.
98. Diaz SO, Barros AS, Goodfellow BJ, Duarte IF, Galhano E, Pita C, et al. Second trimester maternal urine for the diagnosis of trisomy 21 and prediction of poor pregnancy outcomes. J Proteome Res. 2013;12(6):2946–57.
99. Liu X, Quan S, Fu Y, Wang W, Zhang W, Wang X, et al. Study on amniotic fluid metabolism in the second trimester of trisomy 21. J Clin Lab Anal. 2020;34(3):e23089.
100. Bakshi N, Morris CR. The role of the arginine metabolome in pain: implications for sickle cell disease. J Pain Res. 2016;9:167.
101. Kassim AA, Galadanci NA, Pruthi S, DeBaun MR. How I treat and manage strokes in sickle cell disease. Blood. 2015;125(22):3401–10.
102. Dembélé KC, Veyrat-Durebex C, Guindo A, Chupin S, Tessier L, Goïta Y, et al. Sickle cell disease: metabolomic profiles of vaso-occlusive crisis in plasma and erythrocytes. J Clin Med. 2020;9(4):1092.
103. Wu H, Bogdanov M, Zhang Y, Sun K, Zhao S, Song A, et al. Hypoxia-mediated impaired erythrocyte lands' cycle is pathogenic for sickle cell disease. Sci Rep. 2016;6(1):29637.
104. Zhang Y, Dai Y, Wen J, Zhang W, Grenz A, Sun H, et al. Detrimental effects of adenosine signaling in sickle cell disease. Nat Med. 2011;17(1):79–86.
105. Özen H. Glycogen storage diseases: new perspectives. World J Gastroenterol. 2007;13(18):2541.
106. Tamara M, Martin P, Matthias G, Barbara P, Matthias B, Hochuli M, editors. Untargeted plasma metabolomics identifies broad metabolic perturbations in glycogen storage disease type I. 22nd European Congress of Endocrinology, 2020: BioScientifica.
107. Chien Y-H, Goldstein JL, Hwu W-L, Smith PB, Lee N-C, Chiang S-C, et al. Baseline urinary glucose tetrasaccharide concentrations in patients with infantile-and late-onset Pompe disease identified by newborn screening. JIMD Rep. 2014;19:67–73.
108. Kang J, Zhu L, Lu J, Zhang X. Application of metabolomics in autoimmune diseases: insight into biomarkers and pathology. J Neuroimmunol. 2015;279:25–32.
109. Jafari A, Babajani A, Rezaei-Tavirani M. Multiple sclerosis biomarker discoveries by proteomics and metabolomics approaches. Biomark Insights. 2021;16:11772719211013352.
110. Sasaki K, Sagawa H, Suzuki M, Yamamoto H, Tomita M, Soga T, et al. Metabolomics platform with capillary electrophoresis coupled with high-resolution mass spectrometry for plasma analysis. Anal Chem. 2018;91(2):1295–301.
111. Zhang X-w, Li Q-h, Dou J-j. Mass spectrometry-based metabolomics in health and medical science: a systematic review. RSC Adv. 2020;10(6):3092–104.
112. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007;26(1):51–78.

Bioinformatic Tools for Clinical Metabolomics

David S. Wishart

Abstract Clinical metabolomics may be used for the discovery of novel disease biomarkers, the diagnosis of known diseases, and the understanding or rationalization of disease mechanisms. This chapter introduces readers to some of the bioinformatic tools that can be used to facilitate clinical metabolomics and provide key insights into disease processes. In particular, readers will be introduced to several important bioinformatic tools and resources in clinical metabolomics that are widely used for metabolite identification, for biomarker discovery, for disease diagnosis, and for understanding disease mechanisms. These will include discussions on the Metabolomics Standards Initiative (MSI), software tools for metabolite identification and quantification (such as Bayesil and XCMS), data resources for metabolite annotation such as the Human Metabolome Database (HMDB) and MarkerDB (a biomarker database), data analysis and biomarker discovery tools such as MetaboAnalyst, and resources for interpreting or characterizing disease mechanisms, such as the Small Molecule Pathway Database (SMPDB). Using these and other well-known bioinformatic resources, clinicians are now able to make more precise diagnoses, integrate multiple types of omics data, and provide the framework for making metabolomics and integral part of precision medicine.

Keywords Metabolomics · Software · Review · Bioinformatics · Clinical applications

D. S. Wishart (✉)
Department of Computing Science, University of Alberta, Edmonton, AB, Canada

Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada
e-mail: dwishart@ualberta.ca

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_4

1 Introduction

Metabolomics is a branch of "omics" science that is focused on the comprehensive characterization of the small molecule metabolites in the metabolome. Clinical metabolomics involves the application of metabolomic techniques toward discovery, diagnosis, and monitoring of human diseases. Clinical metabolomics can also be used to explore and understand the molecular basis to disease and disease mechanisms. As highlighted throughout this book, metabolomics is being successfully used in many areas of medicine and medical genetics, including the diagnosis and monitoring of inborn errors of metabolism (IEMs), the detection of various endocrine disorders, the characterization of neurodegenerative diseases, and the diagnosis of cancers and cardiovascular diseases. The success of metabolomics in these many diverse areas of medicine and medical genetics lies in the fact that metabolites represent the downstream products of upstream events occurring within the genome, the transcriptome, and the proteome. Indeed, metabolites are sometimes called the "canaries of the genome." Just as canaries were used by miners in the 1800s to serve as sensitive indicators of toxic gases in coal mines, metabolites can serve as remarkably 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 [1]. This exceptional sensitivity is the basis to newborn screening, in which metabolite tests have been used to detect IEMs (such as phenylketonuria) for many decades [2]. Metabolite concentrations are not only very sensitive to what goes on in the genome, but they are also very sensitive to what goes on in the environment. Indeed, metabolite concentrations are heavily influenced by nutrition, physical activity, exposure to environmental chemicals, the time of day, or the even the outside temperature [3, 4].

Because metabolites are the end products of complex interactions happening inside the cell (the genome and the transcriptome) and events happening outside the cell (the environment), metabolomics is ideal for assessing the interactions between genes and the environment (i.e., measuring the phenotype). Therefore, metabolomics offers clinicians and medical researchers an ideal route to measure and monitor both human phenotypes and human diseases in "real time." This gives metabolomics an important advantage over genomics. While the genome can tell you what *might* happen, the metabolome actually tells you what *is* happening.

As highlighted throughout this book, there are two distinct "flavors" of metabolomics: (1) targeted metabolomics and (2) untargeted metabolomics. Targeted metabolomics is focused on the identification (and often absolute quantification) of a specific, predefined collection or category of metabolites in a tissue, biofluid, or biological matrix. Because of its ability to achieve absolute quantification, targeted metabolomics is widely used in clinical medicine, biomarker testing/discovery, and disease monitoring. On the other hand, untargeted metabolomics involves the broad, unbiased identification of the maximum number metabolites or metabolic features in a tissue, biofluid, or biological matrix. Untargeted metabolomics is not reliably quantitative, and so it is more widely used in early-stage biomarker discovery or

hypothesis generation applications. Both targeted and untargeted metabolomics can be, and are currently, used in clinical metabolomics. Both can be used to discover biomarkers or biomarker profiles, and both methods can be performed with standard metabolomic platforms such as liquid chromatography mass spectrometry (LC-MS), gas chromatography mass spectrometry (GC-MS), and nuclear magnetic resonance (NMR) systems. All three of these analytical platforms are capable of separating, detecting, and characterizing hundreds, even thousands of chemicals in complex chemical mixtures. In almost all cases, when NMR, GC-MS, or LC-MS instruments are used to analyze clinical materials, they produce spectra or chromatograms consisting of many hundreds to thousands of peaks.

The quantity of data generated by these platforms necessitates the use of computers and a wide variety of bioinformatic software tools. The main bioinformatic challenges in metabolomics, and clinical metabolomics in particular, are (1) determining which peaks in these spectra match to which chemical compounds (metabolite identification); (2) detecting which compounds are significantly altered in concentration or abundance (determining metabolite significance); (3) determining which metabolites or combination of metabolites can serve as robust disease biomarkers (biomarker discovery); and (4) understanding the biological and genetic context for the observed metabolic changes (finding disease mechanisms or causes).

This chapter is intended to provide an overview to some of the bioinformatic tools that can be used to facilitate clinical metabolomics and provide key insights into disease processes. In particular, readers will be introduced to a number of software tools, data resources, and data standards for facilitating compound identification (for both targeted and untargeted metabolomics), for detecting which compounds are significantly altered in abundance, for identifying and assessing metabolite biomarkers or biomarker panels, and for understanding biological and genetic context of the observed metabolite changes. These will include discussions on the Metabolomics Standards Initiative (MSI), software tools for metabolite identification and quantification (such as AMIX, Bayesil, AMDIS, and XCMS), data resources for metabolite annotation such as the Human Metabolome Database (HMDB) and MarkerDB (a biomarker database), data analysis and biomarker discovery tools such as MetaboAnalyst, and resources for interpreting or characterizing disease mechanisms, such as the Small Molecule Pathway Database (SMPDB).

2 Bioinformatic Tools for Metabolite Identification

Depending on the method used (targeted vs. untargeted) and the type of compounds being measured (lipids vs. non-lipids), different levels of metabolite identification can be achieved. According to the Metabolomics Standards Initiative (MSI) [5], there are actually four levels of metabolite identification: (1) positively identified compounds, (2) putatively identified compounds, (3) compounds putatively identified to be part of a compound class, and (4) unknown compounds. Positively identified compounds correspond to those chemicals that have a name, a known structure,

a CAS (Chemical Abstracts Services) number, or an InChI (International Chemical Identifier) string. To fall into this category, a compound must be identified using a purified, authentic standard collected under identical or near-identical data collection conditions. For targeted metabolomic studies, most metabolites are identified at a Level 1 standard. On the other hand, for most untargeted studies, achieving Level 1 identification is rare. Putatively identified compounds (Level 2) correspond to those where the compound is identified based on a spectral match (i.e., MS/MS) or an alternative parameter (i.e., retention time and exact mass) match to a reference database value. In these cases, an authentic standard is not available, meaning that there is some ambiguity about the compound's true identity. Certainly, if the compound is known to exist in a human biofluid as indicated by numerous literature reports or data resources such as the HMDB, these Level 2 compound identifications are much stronger and may be considered "near positive."

The third level of compound identification is typical of many lipids, where the exact structure of the compound cannot be determined, but it is known to be a specific class of lipid (a phospholipid or triglyceride) or a lipid where the total mass of the acyl chains is known, but the type of position of the individual acyl chains is not known (i.e. the phosphatidyl choline PC(38:3)). Level 3 identification is common for untargeted MS-based lipidomic studies. The fourth level of compound identification is the "unknown" category. Again, this is usually only seen with untargeted metabolomic studies. In many cases, a compound is labeled as an "unknown" simply because the investigator has not been very thorough in their analyses or because their software/database being used for compound identification is inadequate, incomplete, or too small.

2.1 Bioinformatic Tools for Metabolite Identification Via NMR

NMR-based metabolomics has been used in clinical metabolomics for a number of years. These include applications in diagnosing IEMs [6], detecting novel genetic disorders [7], and for lipoprotein profiling [8, 9]. There are three methods for performing metabolite identification by NMR. One is to manually spike the suspected compound into the sample, collect the NMR spectrum of the spiked biofluid, and confirm that the spiked compound changes the observed NMR spectrum in the expected manner. This method is obviously slow and less than ideal for analyzing large numbers of samples or identifying large numbers of metabolites. The second is to use manual chemical shift "lookup" tables to match observed NMR chemical shifts with known chemical shifts. Many NMR labs compile their own chemical shift reference tables to perform manual assignments. However, there are now several online NMR spectral databases that contain both experimentally measured and (accurately) predicted NMR spectra for thousands of human metabolites. These include the Human Metabolome Database or HMDB [10], the Natural Products Magnetic Resonance Database or NP-MRD [11], and the BioMagResBank or BMRB [12]. The HMDB is a particularly important metabolomic resource for

clinical metabolomics. It currently contains the largest collection (>250,000) of known human metabolites along with detailed information on their structures, descriptions, names and synonyms, biological pathways, biofluid concentrations, and disease associations along with their corresponding MS and NMR spectra (to aid in compound identification). In particular, the HMDB has experimentally measured and predicted ^{1}H and ^{13}C NMR spectra for more than 220,000 human metabolites at NMR spectrometer frequencies ranging from 300 MHz to 1000 MHz. While somewhat smaller and less clinically relevant, the NP-MRD has experimentally measured and computationally predicted ^{1}H and ^{13}C NMR spectra for nearly 90,000 metabolites and natural products at NMR spectrometer frequencies ranging from 300 MHz to 1000 MHz. The BMRB has nearly 1000 commonly occurring metabolites with experimentally measured ^{1}H and ^{13}C NMR spectra at 400 MHz and 600 MHz. All of these databases support web-based, automated metabolite identification and NMR spectral searches and/or peak matching.

The third approach to perform metabolite identification via NMR spectroscopy is to use spectral deconvolution. Spectral deconvolution is a computational approach that involves taking a complex spectrum consisting of a mixture with many chemicals and simplifying it into individual spectra of its "pure" chemical components (Fig. 1).

This process requires a specially constructed spectral database as well as carefully developed spectral fitting software. The spectral database used in NMR spectral deconvolution typically consists of reference 1D ^{1}H or ^{13}C NMR spectra of the pure compound(s) that are known or expected to be in the biological sample of

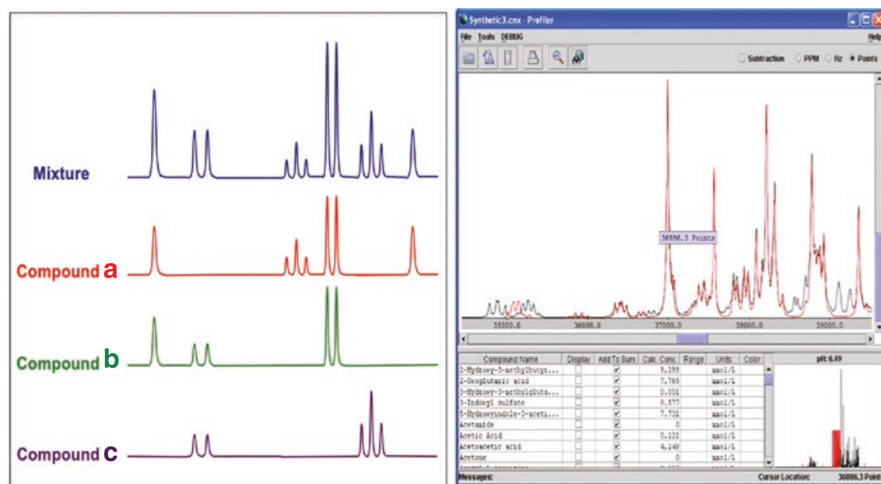

Fig. 1 An image illustrating spectral deconvolution. On the left side is a simplified depiction of how an NMR mixture spectrum could be decomposed into three separate "pure" spectra from three separate compounds (A, B, and C). Summing the three spectra together produces the spectrum of the mixture at the top. On the right side of this image is a spectral deconvolution performed on a real NMR spectrum with the list of identified compounds and their concentrations

interest. These reference NMR spectra must be collected under exactly the same conditions (same temperature, same solvent, same salt, same pH) under which the biological sample was analyzed.

The fact that most metabolites have distinct, almost invariant chemical shift "fingerprints" made up of several compound-specific peaks is one reason why spectral deconvolution works so well for NMR. Furthermore, given that most metabolites have many NMR peaks helps to alleviate the problem of spectral redundancy. To put it in another way, it is highly unlikely that any two distinct compounds will have the same number of peaks, chemical shifts, peak intensities, spin couplings, or line shapes in their NMR spectra. Another reason why spectral deconvolution works so well with NMR is because NMR peak intensities provide precise information about compound concentrations. In other words, accurate spectral matches not only lead to exact compound IDs, but they also lead to exact or near-exact concentration determinations.

There are several commercial programs that support NMR spectral deconvolution for metabolite identification, including AMIX (Bruker) and NMR Suite (Chenomx) (for small molecule metabolites). These software packages have large NMR spectral reference libraries consisting of hundreds of metabolites. Newer versions of these packages such as Bruker B.I. QUANT and Bruker FoodScreener as well as NMR Suite (Version 7 and above) now support semiautomatic deconvolution for higher-throughput analysis. In addition to small molecule metabolite analysis for clinical applications, several other companies, including LipoScience, Bruker, and Nightingale, have begun to offer lipoprotein analyses of serum or plasma samples through an automated or semiautomated spectral deconvolution. These programs or services generate quantitative clinically useful data on high-, medium-, and low-density lipoprotein particles (HDL, MDL, and LDL) from human plasma samples.

In addition to these commercial programs or commercial services for clinically based NMR metabolomics, several freely available academic programs have been developed to perform semiautomatic compound identification or quantification via NMR. These include Batman [13], AQuA [14], ASICS 2.0 [15], and rDolphin [16]. Unfortunately, these programs do not support automated data processing, which means a separate software package such as NMRPipe [17] or NMRFx [18] must be used to manually process the data prior to analysis. This requirement for manual processing significantly slows the analysis (hours per spectrum) and adds an element of human error to the analysis pipeline.

Recently, two new NMR spectral deconvolution programs called Bayesil [19] and MagMet [20] have been introduced. These programs support fully automated NMR metabolite identification and quantification. Both Bayesil and MagMet can perform fully automated data processing and spectral deconvolution of one-dimensional (1D) ^{1}H NMR spectra to identify and quantify upward of 50 to 60 compounds in 3–4 min. Just as with the Chenomx NMR Suite, Bayesil and MagMet work with most NMR instrument models and field strengths but are limited to analyzing specific biofluids such as serum, plasma, or fecal water. Both Bayesil (http://

bayesil.ca/) and MagMet (https://magmet.ca/) are freely accessible through web servers.

2.2 Bioinformatic Tools for Metabolite Identification Via GC-MS

GC-MS has been used in clinical chemistry for more than 50 years [21]. Indeed, many clinical labs continue to routinely use GC-MS as their go-to platform for metabolite profiling. GC-MS can be used to identify and quantify amino acids, organic acids, hormones, and other important clinical biomarkers. As a result, GC-MS can be used to diagnose and monitor many IEMs or other genetic disorders. While most clinical chemists are reluctant to call GC-MS analysis a form metabolomics, the simple fact is that clinical GC-MS is one of the most robust and most widely used metabolomic platforms in clinical chemistry.

Just as with NMR-based metabolomics, compound identification via GC-MS is best done through a form of spectral deconvolution. A typical GC-MS spectrum or total ion chromatogram (TIC) from a metabolite mixture will consist of dozens of sharp peaks (corresponding to ion counts) covering an elution time of about 30–45 min. Each peak often consists of one or more EI (electron ionization) mass spectra arising from one or more compounds. A variety of commercial GC-MS data analysis tools such as AMDIS, which stands for Automated Mass Spectral Deconvolution and Identification System [22], MassHunter (Agilent), ChromaTOF (Leco), and AnalyzerPro (SpectralWorks) can be used to identify and quantify metabolites. Once the EI-MS spectra are extracted, metabolite identification is performed in a similar manner to what is done for NMR. Namely, the extracted EI-MS spectra from the biofluid are compared, one at a time, to EI-MS spectral reference libraries containing the EI-MS spectra of thousands of pure, derivatized, and authenticated compounds. This process is done semiautomatically with users making metabolite identification calls based on the information and spectral image overlays that the computer programs provide.

There are three key factors that ultimately determine the quality of a compound identification by GC-MS: (1) the quality of the extracted query spectrum, (2) the quality of the spectral matching algorithm, and (3) the quality and comprehensiveness of the reference spectral database. Unlike NMR, where "false-positive" peaks are extremely rare, GC-MS is frequently plagued with an abundance of false-positive peaks. In some cases, up to 50% of features seen in GC-MS spectra are fragments, adducts, or derivatives of either the column matrix, the derivatization reagents, or of the metabolites themselves. Different software packages tend to handle these spectral artifacts differently. A study by Lu et al. [23] compared three commonly used GC-MS deconvolution packages (AMDIS, ChromaTOF, and AnalyzerPro) using a defined mixture of 35 compounds with widely varying concentrations. It was found that both the AMDIS and ChromaTOF packages produced

unusually high numbers of false positives or false/impure spectra, while the AnalyzerPro package generally performed best.

Ultimately, the main factor driving the success in compound identification by GC-MS is the size and quality of the EI-MS spectral reference database. The most common and widely used EI-MS resource is the NIST (National Institute of Standards and Technology) EI-MS spectral database. The latest release contains EI-MS spectra for more than 300,000 compounds or derivatized compounds along with retention index (RI) values for another 140,000 compounds. However, most of the NIST compounds are not metabolites nor are they derived from biological materials. The paucity of real metabolites in the NIST library can lead to a number of false-positive identifications, especially if authentic standards are not used to verify the identity of a given compound. On the other hand, the HMDB [10] provides a much larger collection of human metabolites (>250,000) along with a much larger library of corresponding GC-MS data, including almost 2.3 million predicted and experimental EI-MS spectra and nearly seven million retention indices. In particular, the latest version of the HMDB currently contains 6,696,000 accurately predicted RI values for 26,880 parent compounds (and 2.1 million TMS and TBDMS derivatives of those parent compounds) and 2,282,000 predicted and experimental EI-MS spectra. These spectra and RI data can be readily searched, separately or together. However, it is important to note that the quality of most predicted EI-MS spectra is not yet sufficiently high to achieve 100% identification accuracy. Therefore, identifications made using predicted EI-MS data or predicted RI data should always be viewed as putative identifications (Level 2).

2.3 Bioinformatic Tools for Metabolite Identification Via LC-MS

Over the past three decades, LC-MS has become the most popular analytical platform for clinical chemistry and clinical metabolomics [2, 24]. Indeed, most newborn screening activities in the developed world are done using triple quadrupole (QQQ) tandem mass (MS/MS) spectrometers [2]. Relative to NMR or GC-MS, LC-MS methods offer much greater sensitivity, more comprehensive compound detection, and generally higher throughput. These advantages largely explain its growing popularity. Most LC-MS methods adopted in clinical chemistry laboratories employ targeted approaches that use authentic, isotopically labelled standards to simultaneously identify and quantify a small (15–25) number of high-priority metabolites. Most of these targeted LC-MS methods use defined tables of specific metabolite retention times (for their given liquid chromatography system), metabolite-specific multiple reaction monitoring (MRM) or single reaction monitoring (SRM) peak lists, and multipoint calibration curves for metabolite identification and quantification. A number of different software packages from various LC-MS vendors are available to facilitate targeted LC-MS analysis. These include

Analyst (from Sciex), MassHunter (from Agilent), Progenesis QI (from Waters), and TraceFinder/LCQUAN (from Thermo Fisher). In addition to these vendor-specific packages, Biocrates Life Sciences provides a vendor-independent software package, called MetIDQ with its targeted metabolomic kits to perform semiautomated MRM-based metabolite identification and quantification. Modern, targeted LC-MS-based metabolomic software and methods typically allow the identification and quantification of up to 700 metabolites in a given sample in less than 30 min.

Untargeted LC-MS-based metabolomics is normally reserved for biomarker discovery as opposed to biomarker testing. A typical LC-MS spectrum from an untargeted metabolomic study will consist of many sharp peaks (corresponding to ion counts) covering an elution time of about 10–35 min. Each peak may consist of one or more ESI (electrospray ionization) m/z values arising from one or more compounds. As a result, untargeted LC-MS metabolomic studies can easily generate a huge number of spectral features or putative compounds (>10,000). This is many times more than what is seen by NMR or GC-MS. Many of these LC-MS features turn out to be noise peaks, column contaminants, insource fragments, adducts, and isotopic variants. As a result, untargeted LC-MS data typically requires a considerable amount of post-processing and peak consolidation to reduce the number of peaks to a reliable, countable number (preferably <2000 putative compounds).

Untargeted LC-MS data is often further complicated by the fact that liquid chromatographic data is substantially more variable from run to run than NMR or GC-MS data. As a result, metabolomic data acquired via untargeted LC-MS techniques typically requires additional de-noising, spectral alignment, and spectral averaging to ensure that the correct peaks are being picked and compared. This kind of spectral processing requires sophisticated software that either comes bundled with the LC-MS instrument or which is designed, written, and distributed by highly specialized MS laboratories. Examples of some of the instrument-specific tools include Mass Frontier (Thermo Fisher), MassHunter (Agilent), XCMS-Plus (Sciex), Profile Analysis (Bruker), and Progenesis (Waters). There are also a number of platform-independent freeware systems including XCMS [25], MS-DIAL [26], and MzMine2 [27]. All of these software packages support chromatographic and MS spectral alignment, peak finding, multivariate statistics (for data reduction), parent ion mass matching, molecular formula calculation, and MS/MS spectral matching.

Metabolite identification via accurate parent ion mass (or more correctly the mass-to-charge, *m*/*z*) measurement requires the use of very high-resolution MS instruments such as quadrupole time-of-flight (QTOFs), Orbitraps, or Fourier-Transform Ion Cyclotron Resonance (FT-ICR) spectrometers. If a parent ion mass is measured to 4–5 decimal places, which corresponds to a mass accuracy of <5 ppm, it is usually possible to determine the ion's molecular formula and its putative identity (Level 3 identification) through a chemical formula calculator. Several commercial MS chemical formula calculators exist such as SigmaFit (Bruker), Formula Predictor (Shimadzu), and MassHunter (Agilent) as well as a number of freeware packages including 7-Golden-Rules [28] and SIRIUS [29]. By including restrictions on the types of elements typically found in metabolites as well as requirements on hydrogen/carbon ratios and isotopic abundances, it is often

possible to reduce the number of feasible chemical formulas even further [30]. Unfortunately, even with these improvements, parent ion-based metabolite identification is still very risky as there are often many masses or molecular formulae that can still match dozens of metabolites in existing compound databases.

The preferred route of metabolite identification for most untargeted LC-MS metabolomic studies is to use both parent ion (or formula) matching and MS/MS spectral matching. MS/MS spectra, with their characteristic fragmentation patterns, provide very useful structural information about molecules. Successful LC-MS/MS spectral matching is critically dependent on having instrument-specific or condition-specific MS/MS product ion fragment libraries. Many of these libraries are bundled with the instrument-specific software packages mentioned earlier. On the other hand, commercial MS/MS databases, such as the NIST20 database and METLIN [31], as well as public MS/MS databases such as MassBank of North America (MoNA), [32], and HMDB [10] are normally used by the freeware packages (XCMS, MS-DIAL, and MzMine) to perform MS/MS spectral matching. Table 1 provides list of MS/MS databases with experimentally acquired (and predicted) MS/MS spectral data and their relative size.

The challenge with using experimentally acquired MS/MS spectra from these MS databases is that each compound is often represented by dozens of different MS/MS spectra collected on different MS instruments under different ionization conditions or at different collision energies. So, while the number of *experimentally* collected MS spectra is large, the actual number of unique (parent) compounds represented by this diverse collection is quite small. Indeed, it is thought that the current experimental MS/MS spectral collection represents <20% of known or expected human metabolites. Given the striking shortage of experimentally collected MS/MS spectra, a number of investigators have started to use computational tools to predict MS/MS spectra for individual compounds where no experimental MS/MS spectra exist [33, 34]. Many of these in silico predicted MS/MS spectral libraries are now available through the HMDB [10] and the CFM-ID database [34]. These data resources support direct MS/MS spectral searches (to find specific compounds) as well as neutral loss searches (to find related compounds). Other

Table 1 A list of MS/MS spectral databases with the reported numbers of compounds and MS/MS spectra

Database name	Number of compounds	Number of MS/MS spectra
Metlin	870,000	2,510,000 (experimental)
CFM-ID	216,890	1,771,460 (predicted and experimental)
MassBank of North America (MoNA)	223,614	658,790 (experimental and predicted)
NIST20 MS/MS	31,808	1,300,000 (experimental)
MassBankEU	15,055	89,769 (experimental)
GNPS	11,947	584,567 (experimental)
mzCloud	11,495	3,243,574 (experimental)

computational MS/MS spectral interpretation tools, including CSI:FingerID [35] and SIRIUS4 [36], allow users to input an experimental MS/MS spectrum and will generate a structure match or a structure class match without the need to match against any predicted MS/MS spectra. These programs use a technique called chemical or spectral fingerprint analysis rather than spectral prediction [34].

Even with the best MS/MS spectral databases (experimentally acquired or computationally predicted) and the best chemical/spectral fingerprint analysis tools, it is still quite difficult to confidently identify (MSI Level 2) more than 400–500 metabolites via untargeted LC-MS-based metabolomics. While the instrumental times for most untargeted metabolomic LC-MS assays are relatively quick (15–20 min per sample), the data analysis times are often quite slow. Indeed, they often run for several hours per sample as most workflows require considerable manual inspection and manual intervention.

3 Bioinformatic Tools for Detecting Metabolite Differences

Regardless of whether targeted or untargeted approaches are used, one of the central goals of any clinical metabolomic study is to determine which peaks or which metabolites are significantly different for those individuals with a disease or condition relative to healthy controls. This comparison between healthy concentration values versus diseased concentration values is how most known disease biomarkers are measured and how new disease biomarkers are discovered. In clinical metabolomics, there are three routes for determining which metabolites are significantly different or significantly differentiating. These include (1) reference-based metabolite differentiation, (2) multivariate metabolite differentiation, and (3) multivariate peak differentiation.

Reference-based metabolite differentiation involves comparing quantitatively measured metabolite concentrations in a given biofluid for a diseased individual (measured via targeted metabolomic methods) with healthy, age-specific and sex-specific reference metabolite concentrations for the same biofluid, as reported in a database or a reference textbook. Reference-based metabolite differentiation is ideal for diagnosing individuals or for identifying biomarkers in individuals afflicted with rare diseases, such as IEMs, or other genetic disorders. This approach is the most widely used method for detecting significant metabolite differences in clinical metabolomics.

The second approach, called multivariate metabolite differentiation, detects metabolite differences by conducting case-versus-control studies using targeted metabolomic methods. This involves quantitatively measuring specific "named" metabolites from biofluid samples collected from multiple individuals with the disease of interest and biofluid samples from multiple healthy individuals (age and sex matched) without the disease. This approach, which requires the use of multivariate statistical techniques, is ideal for discovering multiple metabolite biomarkers for a given disease and for creating robust multi-marker profiles or multi-marker models.

The third approach to detecting metabolite differences is specific to untargeted metabolomics. Like the second approach, it uses a case-versus-control experimental design and multivariate statistics, but the goal is to use multivariate statistics in combination with relative peak intensity differences (as opposed to absolute concentrations of fully identified metabolites) to identify which peaks or features in the untargeted dataset are differentially abundant. The goal is to reduce the initial list of thousands of unidentified features or peaks to a more manageable list of a few dozen features that exhibit strong differences between cases and controls. From this smaller list of yet-to-be-identified features, it is possible to use various techniques (spectral matching, spike-in experiments, etc.) to identify the actual metabolites showing the most significant concentration changes. Multivariate peak differentiation is well suited for novel biomarker discovery and early-stage or putative biomarker identification. However, these putative markers must ultimately be validated using a targeted metabolomic technique that quantitatively measures the presumptive metabolites.

3.1 Bioinformatic Tools for Reference-Based Metabolite Differentiation

Reference-based metabolite differentiation requires a large and reliable set of human reference metabolite concentrations for different ages, sexes and biofluids. Currently the most complete set of healthy metabolite reference values for different biofluids for different ages and sexes is the Human Metabolome Database or HMDB [10]. The HMDB is widely regarded as the most complete open-access database on human metabolites and their disease associations. Currently the HMDB contains reference concentration values for 3073 metabolites in serum/plasma, 1757 metabolites in urine, 447 metabolites in cerebrospinal fluid, 883 metabolites in saliva and 1805 metabolites in feces. These values include the literature sources, age group or age range of the measured cohort and sex (if available). In many cases, multiple values are provided as different analytical methods can lead to slight differences in the reported concentration values. Abnormal concentrations are also reported in the HMDB, along with the associated conditions and the corresponding literature sources. Currently, the HMDB contains abnormal metabolite concentration data for more than 660 IEMs and other genetic disorders. Users can easily query the HMDB via its web interface with a specific metabolite name, metabolite structure or metabolite InChI identifier to obtain the corresponding biofluid concentrations and explore what else is known about the queried metabolite. Likewise, users may also query or browse the HMDB by disease or condition names and the results will provide the lists of altered metabolites associated with that condition.

Another useful source of reference metabolite concentrations and reference-based metabolite differentiation is MarkerDB [37]. MarkerDB is the world's most comprehensive open-access biomarker database. It contains more than 26,600

genetic, protein, and metabolite biomarkers for 670 human disorders or conditions. MarkerDB not only provides metabolite concentration data for healthy individuals (with information on age- and sex-specific values), but it also provides data on the unhealthy metabolite concentrations (in different biofluids), descriptions of the associated disorders, detailed descriptions of the metabolites, and even biomarker performance indications, such as biomarker sensitivity and specificity. In this regard, the disease and biomarker data in MarkerDB is probably more complete than the data in HMDB.

Both HMDB and MarkerDB are primarily designed for querying or browsing a small number (<3) of metabolites and assessing their concentration differences relative to healthy normal values. If multiple (10 or more) metabolites need to be queried to determine if the observed metabolite concentrations are significantly different from normal, it is possible to use an online metabolomic web server called MetaboAnalyst [38]. In particular, the Single Sample Profiling (SSP) option within MetaboAnalyst's enrichment analysis (EA) module allows users to enter long lists of metabolite names and concentrations which are then compared against those values reported in the HMDB for a wide variety of biofluids. Metabolites that are higher (H) or lower (L) than the normal reference values are flagged in the resulting output. Hyperlinks to the HMDB compound database entries are also provided. The SSP option with MetaboAnalyst provides users a fast and convenient route to perform reference-based metabolite differentiation from targeted metabolomic data. It also allows important metabolite features (i.e., those marked with H or L) to generate a biomarker profile for a given disease.

3.2 Bioinformatic Tools for Multivariate Metabolite/ Peak Differentiation

When conducting biomarker discovery or biomarker validation studies in clinical metabolomics, it is standard practice to perform well-powered case-versus-control studies. Case-control studies often involve hundreds of subjects, thousands of biofluid samples, and the collection of very large metabolomic datasets. As highlighted earlier, any single targeted metabolomic assay can easily generate hundreds of named metabolites and metabolite concentrations. Likewise, an untargeted metabolomic assay can easily generate thousands of un-named "features" or peaks along with their corresponding relative peak intensities. Because the number of variables in these types of clinical studies is so large, special statistical methods must be used to help manage the data, differentiate up- or downregulated metabolites, and reduce the problems of overlap, false positives, and significance. In particular, the techniques that must be used are called multivariate statistics. In multivariate (short for multiple variable) statistics, the variables are called "dimensions." One of the primary objectives of multivariate statistics is to reduce the number of variables or dimensions so that the problem can be tackled more simply using traditional

univariate statistics, such as Student's t-tests or ANOVA techniques. Multivariate statistics uses a class of mathematical techniques called dimensional reduction methods to make multivariate data look more like univariate (single variable) data. Dimensional reduction allows one to identify the key components in a large multivariate dataset that contain the maximum amount of information or maximize the differences among groups. As a result, dimensional reduction reduces a long list of metabolites to a shorter list of the most significant metabolites. This is the essence of multivariate feature/metabolite differentiation. The most common form of dimensional reduction is called principal component analysis or PCA.

3.3 Principal Component Analysis

Principal component analysis (PCA) is an unsupervised clustering technique. Clustering is the process of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. Clustering helps distinguish groups, such as cases and controls, from one another based on their metabolic parameters. In a more formal "mathematical" sense, PCA determines an optimal linear transformation for a collection of data points such that the properties of that set of data points are most clearly displayed along a small number of coordinate (or principal) axes. Simply put, PCA allows metabolomic researchers to easily plot, visualize, and cluster multiple lists of metabolites and their concentrations based on linear combinations of their shared features. PCA is most commonly used in clinical metabolomics to determine whether one or more samples are different from another. It also allows one to identify which variables or metabolites contribute most to this difference and whether those metabolites contribute in the same way (i.e., are correlated) or independently (i.e., uncorrelated) from each other. PCA is particularly appealing because it allows one to visually detect sample clusters or groupings. In particular, the results of a PCA are usually discussed in terms of **scores** and **loadings**. The scores represent the original data in the new coordinate system, and the loadings are the weights applied to the original data during the projection process. Plotting out the data using two sets of scores (one for the X axis and one for the Y axis) will produce a "scores" plot. The "weightings" of the individual components correspond to a PCA "loadings" plot. With untargeted metabolomic data, the loadings plot can be used to narrow down the list of features or peaks to just a few important ones that need to be identified. This makes PCA ideal for reducing the number of features in untargeted metabolomic data from 1000s to just a few dozen or less. It can also help reduce the list of metabolites in targeted metabolomic studies from 100 s to just a dozen or fewer. Furthermore, PCA can be used to identify the most important or most informative metabolites required to generate a biomarker profile for a given disease.

PCA can be easily conducted using a variety of free or nearly free software programs such as MatLab or the R project (http://www.r-project.org) using R's *prcomp* or *princomp* commands. PCA can also be performed using freely available,

downloadable software packages such as XCMS [25], MS-DIAL [26], MAVEN [39], and GALAXY-M [40], which are frequently used for processing LC-MS data. Freely available web servers are also available that support PCA and other common multivariate statistical techniques. The most widely used web server for multivariate statistical analysis in metabolomics is MetaboAnalyst [38]. MetaboAnalyst, which is freely available, provides an easy-to-use graphical interface that allow users to simply point and click to perform complex multivariate statistical operations or to generate colorful, interactive graphs or tables. Nearly one-third of all published metabolomic papers use MetaboAnalyst in the metabolomic data analysis pipeline.

3.4 Partial Least Squares Discriminant Analysis

PCA is not the only multivariate statistical approach that can be used to identify important metabolites or reduce the number of spectral features. Another type of multivariate statistical method that can be used for this purpose is known as supervised classification. Supervised classifiers are programs or algorithms that require that information about the class identities must be provided in advance of running the analysis. In other words, prior knowledge about which samples belong to the "cases" and which samples belong to the "controls" is used to label each of the samples. Examples of supervised classifiers include SIMCA (soft independent modeling by class analogy), PLS-DA (partial least squares discriminant analysis), and OPLS-DA (orthogonal projections to latent structures discriminant analysis). All of these techniques can be used to help convert extensive NMR, LC-MS/MS, and GC-MS metabolite lists (for targeted metabolomics) or their corresponding spectral features (for untargeted metabolomics) into much shorter lists of highly significant metabolites and/or features.

PLS-DA or partial least squares discriminant analysis is often used when PCA techniques do not generate sufficiently distinct clusters or sufficiently distinct metabolite sets. In particular, PLS-DA can be used to enhance the separation between data points in a PCA "scores" plot by essentially rotating the PCA components such that a maximum separation among classes is obtained. This enhanced separation allows one to better understand which variables are most responsible for separating the observed (or apparent) classes. Care must be taken in using PLS-DA methods because these classification techniques can be overtrained. That is, PLS-DA can create convincing clusters or classes that have no statistical meaning (i.e., they over-fit the data). The best way of avoiding these problems is to use permutation (random relabeling) approaches to ensure that the data clusters derived by PLS-DA are real and robust. A number of freely available metabolomic software packages and web servers, such as MetaboAnalyst, are able to perform these permutation tests. Another way of quantitatively assessing a PLS-DA model is to report R^2 and/or Q^2 values. Both R^2 and Q^2 are typically reported by metabolomic web servers and software packages such as MetaboAnalyst. R^2 is the correlation index and refers to the goodness of fit or the explained variation, while Q^2 refers to the predicted

variation or quality of prediction. A poorly fit model will have an R^2 of 0.2 or 0.3, while a nicely fit model will have an R^2 of 0.7 or 0.8. In practice, Q^2 typically tracks very closely to R^2. However, if the PLS-DA model becomes over-fit, Q^2 reaches a maximum value and then begins to fall. Generally, a $Q^2 > 0.5$ if considered good, while a Q^2 of 0.9 is outstanding.

If a robust PLS-DA model can be generated, the set of important metabolites (generated via targeted metabolomics) or features (generated via untargeted metabolomics) arising from the variable importance plot (VIP) can be more easily interpreted than those determined via a PCA loading plot. PLS-DA is generally among the most powerful and useful methods for reducing the number of features in untargeted metabolomic data from 1000s to just a few dozen or less. PLS-DA is also very effective in reducing the list of important or differential metabolites in targeted metabolomic studies from 100 s to just a dozen or fewer. Furthermore, PLS-DA can be used to identify the most important or most informative metabolites required to generate a biomarker profile for a given disease. The utility of PLS-DA in biomarker development and discovery is discussed in the next section.

4 Bioinformatic Tools for Biomarker Discovery

One of the principal goals of clinical metabolomics is to discover and/or measure metabolite biomarkers of human disease. Biomarkers are typically defined as objectively measurable biological characteristics that can be used to diagnose, monitor, or predict the risk of disease [41]. For example, blood glucose is a standard chemical biomarker for monitoring diabetes, while serum creatinine is a chemical marker for kidney function. Many traditional clinical chemistry biomarkers consist of just a single measured entity. However, metabolomics has allowed clinicians to measure multiple chemicals at once. This means it is now possible to measure multiple biomarkers or develop multi-biomarker panels to predict or diagnose diseases with greatly improved sensitivity and specificity. Indeed, it has long been common practice among physicians to combine multiple physiological biomarkers (age + BMI + triglyceride level + cholesterol level = cardiac disease risk) to improve biomarker sensitivity and specificity. Now, with metabolomics, it is possible to create diagnostic or predictive models from multiple metabolites, which can be used to classify individuals into specific groups (i.e., healthy vs. diseased) with much improved sensitivity and specificity.

Sensitivity and specificity have very formal definitions in biomarker studies and the biomarker literature. In standard case vs. control studies, sensitivity (Sn) is mathematically defined as Sn = TP/(TP + FN), and specificity (Sp) is mathematically defined as Sp = TN/(TN + FP), where TP is the number of true positives, TN is the number of true negatives, FN is the number of false negatives, and FP is the number of false positives. Sensitivity (also known as the true positive rate) can be considered as the probability of a positive test result given that a subject has an actual positive outcome. Specificity (also known as the true negative rate) can be

considered as the probability of a negative test result given that a subject has an actual negative outcome. For instance, if a biomarker or biomarker panel has a sensitivity of 0.95 and a specificity of 0.60, this indicates that if a patient has a test score that is above the decision boundary there is a 95% chance that the patient is correctly diagnosed with the disease/condition; but if the test score is below the decision boundary, then there is only a 60% chance that the patient is correctly classified as being healthy. A promising biomarker must have both high sensitivity (i.e., to give a positive test result when the disease is actually present) and high specificity (i.e., to give a negative test result when the disease is absent).

One of the best ways to observe how a decision boundary affects sensitivity and specificity is through a receiver operator characteristic (ROC) curve. A ROC curve shows how the sensitivity and specificity change as the classification decision boundary is varied across the range of available biomarker scores. Because an ROC curve depicts the performance of a biomarker test over the complete range of possible decision boundaries, it allows the optimal specificity and associated sensitivity to be determined by visual inspection. When one evaluates a biomarker using a ROC curve, there is no need to be worried about the "data normality" of either the predicted positive or negative score distributions nor whether the two distributions have equal numbers of subjects or equal variance. As a result, ROC curve analysis is widely considered to be the most objective and statistically valid method for biomarker performance evaluation [42].

ROC curves are often summarized into a single metric known as the "Area Under the Curve" (AUC or AUROC). The AUROC indicates a biomarker model's ability to discriminate between cases (positive examples) and non-cases (negative examples.). If all positive cases are ranked before negative ones (i.e., a perfect classifier), the AUC is 1.0. An AUC of 0.5 is equivalent to randomly classifying subjects as either sick or healthy (i.e., the classifier is of no practical utility). A rough guide for assessing the utility of a biomarker based on its AUROC is as follows: 0.9–1.0 = excellent; 0.8–0.9 = good; 0.7–0.8 = fair; 0.6–0.7 = poor; and 0.5–0.6 = fail (see Fig. 2).

Currently, the most useful tool for biomarker discovery, biomarker selection, and for performing sensitivity/specificity analysis (via ROC curve analysis) with metabolomic data is MetaboAnalyst [38]. In particular, the MetaboAnalyst biomarker module supports three common ROC-based analysis modes: (1) classical univariate ROC curve analysis, (2) multivariate ROC curve exploration, and (3) manual biomarker model creation and evaluation. The most popular and useful option is the multivariate ROC curve exploration which supports automated multi-biomarker selection and optimization using Monte Carlo cross validation (MCCV). This allows the biomarker panel's AUROC to be maximized while minimizing the number of biomarkers being used. MetaboAnalyst will typically generate several biomarker models with different numbers of metabolites and different AUROCs to allow users some choice over what biomarker panel matches their biomarker requirements or performance expectations.

Four different biomarker modeling options are currently offered with MetaboAnalyst's Biomarker module: (1) partial least squares discriminant analysis (PLS-DA), (2) support vector machine (SVM), (3) random forests, and (4) logistic

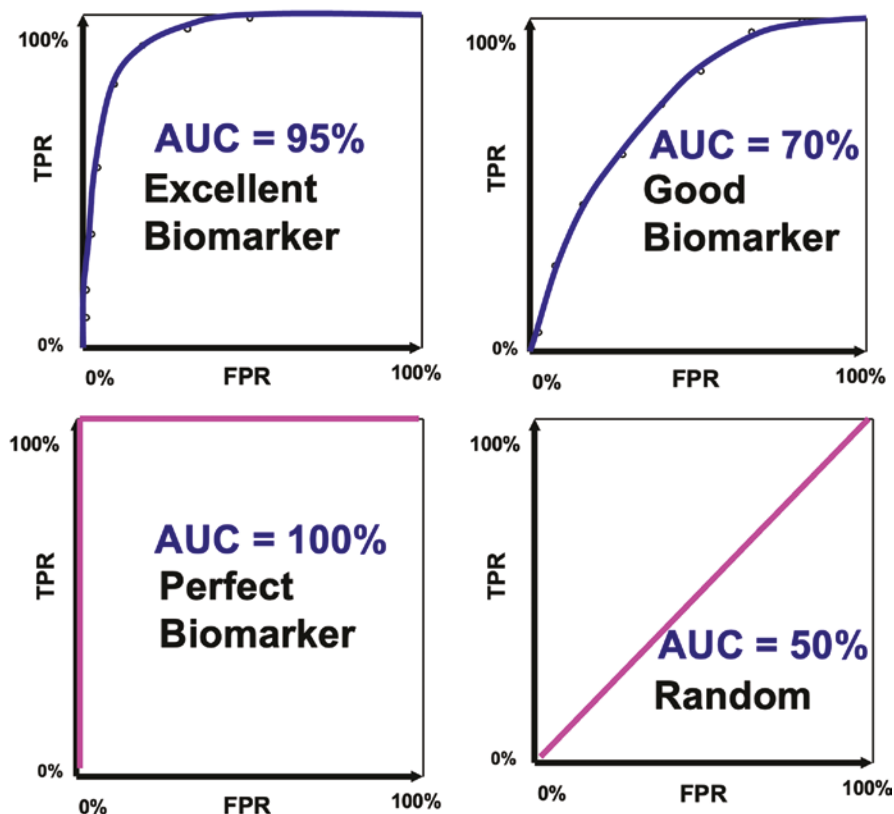

Fig. 2 A depiction of several different ROC curves for different biomarker tests with the area under the ROC curves indicated. On the bottom left is an example of a perfect biomarker with a perfect ROC curve having an AUROC of 1.0. On the top left is an example of an excellent biomarker profile with an AUROC of 0.9. On the top right is an example of a moderately good biomarker profile with an AUROC of 0.7. On the bottom left is an example of a random biomarker with no predictive or diagnostics capability

regression. The most useful of these four options is the logistic regression model as it provides an equation, or set of equations, incorporating metabolite concentrations that can be universally used for calculating cutoff thresholds or decision boundaries. MetaboAnalyst also generates a number of useful graphs, ROC curves, confidence intervals, and charts to help users assess the selected biomarkers and biomarker models. The simplicity with which biomarker models can be developed (mostly via point-and-click operations) and rich graphical support in MetaboAnalyst within its Biomarker module makes it the ideal tool for biomarker discovery in clinical metabolomics.

It is important to note that the metabolomic data being uploaded into MetaboAnalyst's Biomarker module should be absolutely quantitative. As with most analytical methods supported by MetaboAnalyst, the metabolite data uploaded

into the biomarker module must be properly normalized, scaled, and transformed so that metabolite values are comparable and therefore more robustly analyzable.

5 Bioinformatic Tools for Biomedical Interpretation and Data Integration

Biomarker identification is usually limited to disease diagnosis or disease prediction [37]. Biomarkers are not necessarily intended to help uncover the underlying cause of the disease or explain specific disease mechanisms. Of course, metabolic biomarkers may be strongly associated with disease mechanisms, but without some kind of biological interpretation or some kind of biomedical context, association does not imply causation, nor does it lead to mechanistic insights. To properly determine disease causes or disease mechanisms from clinical metabolomic data, it is often necessary to turn to metabolic pathways or to integrate both genomic and metabolomic data together. Metabolite interpretation via pathway analysis often involves determining whether the identified metabolites belong to a single pathway or a smaller set of related pathways. In many cases, this requires searching or reading carefully through various online metabolic pathway databases.

Metabolic pathway databases provide a centralized collection of schematic pathways that depict the current state of the knowledge regarding metabolic (catabolic, anabolic, or signaling) processes that occur within a cell, tissue, or organism. Pathway databases combine large collections of carefully curated metabolite data, with large amounts of carefully collected protein and/or genetic data through a series of illustrated enzyme-mediated reactions, receptor-mediated signaling processes, or protein-aided transport activities. These represent the key molecular and cellular activities that underlie all physiological processes. Because pathway databases combine multi-omic (metabolomic, proteomic, genomic) data together along with general information about physiological or biological consequences, these databases can play a key role in the biological analysis or biomedical interpretation of metabolomic data.

Some of the most popular small molecule pathway databases include KEGG [43], the Reactome database [44], the “Cyc” databases [45], WikiPathways [46], the Small Molecule Pathway Database or SMPDB [47], and PathBank [48]. A number of commercial pathway databases also exist such as BioCarta, TransPath (from BioBase Inc.), and Ingenuity Pathway Analysis (from Ingenuity Systems Inc.). The most useful pathway database for clinical metabolomics is SMPDB as it offers that largest number and most diverse pathways specific to human biology and human diseases. In particular, SMPDB contains 150 signaling pathways, 20,250 disease pathways (covering many IEMs and genetic disorders), 468 drug pathways, and 27,800 metabolic (catabolic/anabolic) pathways.

Most pathway databases support interactive image mapping with hyperlinked information content that allows users to view chemical information (if a compound

is clicked) or brief summaries of genes and/or proteins (if a protein or enzyme is clicked). Almost all pathway databases support some kind of limited text search, and a few, such as Reactome, SMPDB, and the "Cyc" databases, support the mapping of gene, protein, and/or metabolite expression data onto pathway diagrams. Only a few pathway databases, such as SMPDB, provide their pathway data in common, machine-readable data exchange formats such as BioPAX [49], SBML [Systems Biology Markup Language] [50], or SBGN-ML [Systems Biology Graphical Notation Markup Language] [51].

Nearly all of the major pathway databases used in metabolomics today (KEGG, the Reactome database, the "Cyc" databases, WikiPathways, and SMPDB) permit users to upload metabolite data and generate highlighted pathway plots indicating the location of key metabolites in a given pathway. Unfortunately, most metabolite/metabolism databases (such as KEGG, the Cyc databases, WikiPathways, Reactome) only contain anabolic or catabolic pathways associated with endogenous metabolites. Almost no information is provided on metabolite signaling pathways, disease pathways, metabolic diseases (such as phenylketonuria), or drug action pathways (how aspirin works). As a result, many metabolomic pathway analyses are limited to interpreting complex metabolite data in only the simplest of terms. An important exception to this is SMPDB. SMPDB resource contains hundreds of human-specific pathways including dozens of signaling pathways as well as hundreds of disease and drug pathways. Currently, SMPDB is the only open-access database that covers such a broad diversity of human disease or disease mechanism pathways – especially for small molecules. This makes SMPDB one of the most popular tools for interpreting and integrating clinical metabolomic data.

While pathway visualization can provide some important qualitative insight into the biological roles for metabolites detected in a clinical metabolomic study, it is also important to remember that more quantitative tools for pathway analysis also exist. In particular, pathway enrichment and pathway topological analysis are two quantitative methods that can be quite helpful. MetaboAnalyst offers several advanced pathway enrichment analysis procedures along with pathway topological analysis to help identify the most relevant metabolic pathways involved in a given clinical metabolomic study. The pathway analysis module in MetaboAnalyst uses simple point and click operations to support three types of analyses: (1) pathway enrichment analysis, (2) pathway topological analysis, and (3) pathway impact analysis. Pathway enrichment analysis can be done using either overrepresentation analysis or via metabolite set enrichment analysis using Fishers' exact test, the hypergeometric test, and GlobalAncova [52]. Pathway topological analysis is based on the centrality measures of a metabolite in a given metabolic network. Centrality is a quantitative measure of the position of a metabolite relative to the other metabolites in a pathway. Centrality can be used to estimate a metabolite's relative importance or role in a pathway or network diagram. MetaboAnalyst uses relative "betweenness" centrality and "out-degree" centrality to calculate the relative importance of a metabolite. Centrality means that metabolites located on the periphery of a pathway or those that are involved in side reactions have little consequence and are not particularly "central." On the other hand, metabolites that are in pathway

bottlenecks or those that serve as hubs or precursors for many reactions are more "central." By calculating the topological importance of different metabolites in a given pathway, as well as the enrichment of certain metabolites in a pathway, it is possible to calculate a pathway impact score.

By plotting the pathway impact score against the number of significant metabolites appearing that pathway, it is possible to generate a plot that illustrates the most important pathways detected from a set of significantly altered metabolites in a given metabolomic experiment (Fig. 3).

In this example, the *X*-axis displays the pathway impact score, while the *Y*-axis displays the level of enrichment. The size of the colored circles represents the number of metabolites in the illustrated pathway, and the color of the circle indicates its overall significance (with red being most significant and pale yellow being least significant). By clicking on the colored circles, it is possible to see more details about the pathway name, the pathway components, and their topological relationships. Within MetaboAnalyst, each detected metabolite is also "clickable" so that a box-and-whisker plot can be generated that illustrates the metabolite concentrations and range between the "case" and "control" samples.

In addition to pathway analysis, there are also a number of other approaches that can be used to interpret, visualize, or explore clinical metabolomic data. One particularly useful approach involves using a technique called metabolite set

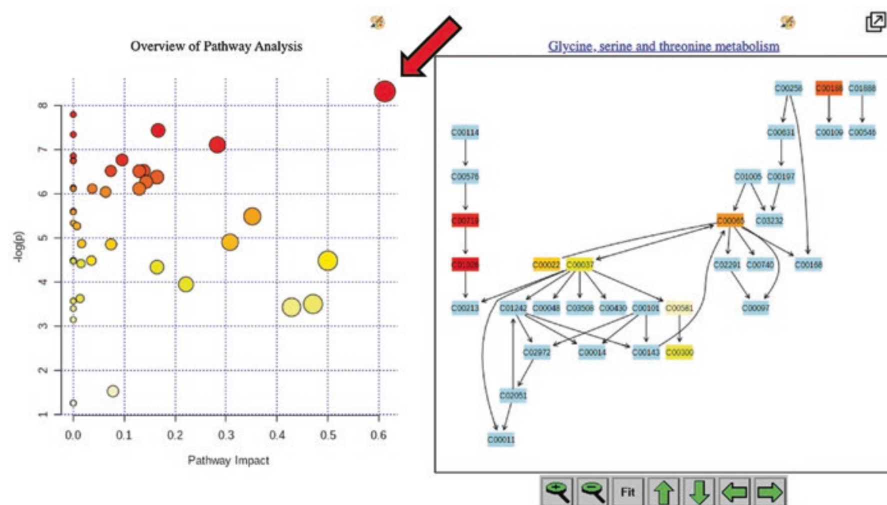

Fig. 3 An example of a pathway impact diagram from MetaboAnalyst. For the graph on the left side of the image, the *X*-axis displays the pathway impact score, while the *Y*-axis displays the level of enrichment. The size of the colored circles represents the number of metabolites in the illustrated pathway, and the color of the circle indicates its overall significance (with red being most significant and pale yellow being least significant). By clicking on the colored circles, it is possible to see more details about the pathway name, the pathway components, and their topological relationships. This expanded view is shown on the right side of the image with the pathway diagram being taken from KEGG and the individual metabolites being identified with KEGG identifiers

enrichment or MSEA [53]. MSEA is a form of functional enrichment analysis similar to gene set enrichment analysis (GSEA). For metabolite set enrichment to be effective, one usually needs a comprehensive database of metabolic pathways, a database of healthy/diseased metabolite concentrations, or a database with associations between metabolites and SNPs or metabolites and gene expression levels. Ideally, a good MSEA system should have all of these databases and support all of these functional analyses. In this regard, the MSEA module in MetaboAnalyst actually has all of these databases and functional tools, making it particularly useful for clinical interpretation. Another approach to interpreting clinical metabolomic data is to combine it with gene expression or protein expression data [54]. There are a number of bioinformatic tools that support this kind of integration. One example is MetScape [55]. MetScape is a plugin for the widely used open-source network analysis and visualization tool called Cytoscape. MetScape supports the interactive, network-based exploration and visualization of both metabolite and gene expression data by integrating both the KEGG and EHMN (Edinburgh human metabolic network) databases. MetScape allows users to identify enriched pathways from gene/metabolite expression profiling data, build and analyze gene/metabolite networks, and interactively visualize changes in gene/metabolite data. Another integrated "omics" approach that offers similar capabilities is called Integrated Metabolomic and Expression Analysis or INMEX [54]. This web-based tool is now available through MetaboAnalyst. Like MetScape, INMEX makes use of the KEGG pathway database as well as a number of pathways from SMPDB.

How these software tools and resources are used and how the data is eventually interpreted depends somewhat on the knowledge of the user. Naïve analyses performed by a naïve individual will lead to naïve interpretation. Taking the time to read the literature and to discover what else is known (genetically or metabolically) about a given disease or condition will allow for a much more efficient use of the software and a much more intelligent interpretation of the data. In this regard, it is always important to remember that bioinformatics should always be used as an aid to support and extend one's own biochemical and biological knowledge.

6 Summary

This chapter has provided a high-level overview of the bioinformatic resources needed to analyze clinical metabolomic data. As highlighted at the beginning of this chapter, the main bioinformatic challenges in clinical metabolomics are (1) metabolite identification, (2) determining metabolite significance, (3) biomarker discovery, and (4) finding disease mechanisms or causes. To address these challenges, we introduced and discussed a number of software tools, data resources, and data standards for facilitating compound identification, for detecting which compounds are significantly altered in abundance, for identifying and assessing metabolite biomarkers or biomarker panels, and for understanding biological and genetic context of the observed metabolite changes. In particular, we discussed the Metabolomics

Standards Initiative (MSI) for compound identification and introduced software tools for metabolite identification and quantification for NMR, GC-MS, and LC-MS/MS (such as AMIX, Bayesil, AMDIS, and XCMS). We also discussed data resources for metabolite annotation such as the Human Metabolome Database (HMDB) and MarkerDB, as well as data analysis and biomarker discovery tools such as MetaboAnalyst. Finally, we closed the chapter with a discussion on different bioinformatic resources for interpreting or characterizing disease mechanisms, such as the Small Molecule Pathway Database (SMPDB).

The field of clinical metabolomics has grown considerably over the past 10 years, and detailed descriptions of all the bioinformatic tools and resources that have been developed for clinical metabolomics could easily fill several books. This chapter is only intended to serve as an introduction so that individuals who are interested in pursuing clinical metabolomics and using bioinformatic tools for clinical metabolomics can better appreciate what is available, what is possible, and what still needs to be done.

References

1. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, et al. HMDB: the human metabolome database. Nucleic Acids Res. 2007;35(Database issue):D521–6.
2. Levy PA. An overview of newborn screening. J Dev Behav Pediatr. 2010;31:622–31.
3. Bassini A, Cameron LC. Sportomics: building a new concept in metabolic studies and exercise science. Biochem Biophys Res Commun. 2014;445:708–16.
4. Brown SA. Circadian metabolism: from mechanisms to metabolomics and medicine. Trends Endocrinol Metab. 2016;27:415–26.
5. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics. 2007;3:211–21.
6. Moolenaar SH, Engelke UF, Wevers RA. Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. Ann Clin Biochem. 2003;40(Pt 1):16–24.
7. Verrips A, Hoefsloot LH, Steenbergen GC, Theelen JP, Wevers RA, Gabreëls FJ, van Engelen BG, van den Heuvel LP. Clinical and molecular genetic characteristics of patients with cerebrotendinous xanthomatosis. Brain. 2000;123(Pt 5):908–19.
8. Otvos JD, Jeyarajah EJ, Bennett DW, Krauss RM. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. Clin Chem. 1992;38(9):1632–8.
9. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet. 2015;8:192–206.
10. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, et al. HMDB 5.0: the human metabolome database for 2022. Nucleic Acids Res. 2022a;50(D1):D622–31.
11. Wishart DS, Sayeeda Z, Budinski Z, Guo A, Lee BL, Berjanskii M, Rout M, Peters H, Dizon R, Mah R, et al. NP-MRD: the natural products magnetic resonance database. Nucleic Acids Res. 2022b;50(D1):D665–77.

12. Markley JL, Ulrich EL, Berman HM, Henrick K, Nakamura H, Akutsu H. BioMagResBank (BMRB) as a partner in the worldwide protein data Bank (wwPDB): new policies affecting biomolecular NMR depositions. J Biomol NMR. 2008;40:153–5.
13. Hao J, Liebeke M, Astle W, De Lorio M, Bundy JG, Ebbels TMD. Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. Nat Protoc. 2014;9:1416–27.
14. Röhnisch HE, Eriksson J, Mullner E, Agback P, Sandstrom C, Moazzami AA. AQuA: an automated quantification algorithm for high-throughput NMR-based metabolomics and its application in human plasma. Anal Chem. 2018;90:2095–102.
15. Lefort G, Liaubet L, Marty-Gasset N, Canlet C, Vialaneix N, Servien R. Joint automatic metabolite identification and quantification of a set of (1)H NMR spectra. Anal Chem. 2021;93:2861–70.
16. Cañueto D, Gómez J, Salek RM, Correig X, Cañellas N. rDolphin: a GUI R package for proficient automatic profiling of 1D (1)H-NMR spectra of study datasets. Metabolomics. 2018;14:24.
17. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6:277.
18. Norris M, Fetler B, Marchant J, Johnson BA. NMRFx processor: a cross-platform NMR data processing program. J Biomol NMR. 2016;65:205–16.
19. Ravanbakhsh S, Liu P, Bjorndahl TC, Mandal R, Grant JR, Wilson M, Eisner R, Sinelnikov I, Hu X, Luchinat C, Greiner R, Wishart DS. Accurate, fully-automated NMR spectral profiling for metabolomics. PloS One. 2015;10:e0124219.
20. Foroutan A, Fitzsimmons C, Mandal R, Berjanskii MV, Wishart DS. Serum metabolite biomarkers for predicting residual feed intake (RFI) of Young Angus bulls. Metabolites. 2020;10:491.
21. Jellum E, Stokke O, Eldjarn L. Combined use of gas chromatography, mass spectrometry, and computer in diagnosis and studies of metabolic disorders. Clin Chem. 1972;18:800–9.
22. Stein SE. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J Am Soc Mass Spectrom. 1999;10:770–81.
23. Lu H, Liang Y, Dunn WB, Shen H, Kell DB. Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. Trends Anal Chem. 2008;27:215–27.
24. Ismail IT, Showalter MR, Fiehn O. Inborn errors of metabolism in the era of untargeted metabolomics and lipidomics. Metabolites. 2019;9:242.
25. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006;78:779–87.
26. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, VanderGheynst J, Fiehn O, Arita M. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods. 2015;12:523–6.
27. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11:395.
28. Kind T, Fiehn O. Seven Golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics. 2007;8:105.
29. Dührkop K, Scheubert K, Böcker S. Molecular Formula Identification with SIRIUS. Metabolites. 2013;3:506–16.
30. Kind T, Fiehn O. Advances in structure elucidation of small molecules using mass spectrometry. Bioanal Rev. 2010;2:23–60.
31. Tautenhahn R, Cho K, Uritboonthai W, Zhu Z, Patti GJ, Siuzdak G. An accelerated workflow for untargeted metabolomics using the METLIN database. Nat Biotechnol. 2012;30:826–8.
32. Kind T, Tsugawa H, Cajka T, Ma Y, Lai Z, Mehta SS, Wohlgemuth G, Barupal DK, Showalter MR, Arita M, Fiehn O. Identification of small molecules using accurate mass MS/MS search. Mass Spectrom Rev. 2017;37(4):513–32.

33. Allen F, Greiner R, Wishart DS. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. Metabolomics. 2015;11:98–110.
34. Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart DS. CFM-ID 4.0: more accurate ESI-MS/MS spectral prediction and compound identification. Anal Chem. 2021;93:11692–700.
35. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci U S A. 2015;112:12580–5.
36. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, Dorrestein PC, Rousu J, Böcker S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods. 2019;16:299–302.
37. Wishart DS, Bartok B, Oler E, Liang KYH, Budinski Z, Berjanskii M, Guo A, Cao X, Wilson M. MarkerDB: an online database of molecular biomarkers. Nucleic Acids Res. 2021;49(D1):D1259–67.
38. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46(W1):W486–94.
39. Melamud E, Vastag L, Rabinowitz JD. Metabolomic analysis and visualization engine for LC-MS data. Anal Chem. 2010;82:9818–26.
40. Davidson RL, Weber RJ, Liu H, Sharma-Oates A, Viant MR. Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. Gigascience. 2016;5:10.
41. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89–95.
42. Soreide K. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. J Clin Pathol. 2009;62:1–5.
43. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue):D199–205.
44. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011;39:D691–7.
45. Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, Pellegrini-Toole A. The EcoCyc and MetaCyc databases. Nucleic Acids Res. 2000;28:56–9.
46. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR. WikiPathways: building research communities on biological pathways. Nucleic Acids Res. 2012;40(Database issue):D1301–7.
47. Jewison T, Su Y, Disfany FM, Liang Y, Knox C, Maciejewski A, Poelzer J, Huynh J, Zhou Y, Arndt D, et al. SMPDB 2.0: big improvements to the small molecule pathway database. Nucleic Acids Res. 2014;42(Database issue):D478–84.
48. Wishart DS, Li C, Marcu A, Badran H, Pon A, Budinski Z, Patron J, Lipton D, Cao X, Oler E, et al. PathBank: a comprehensive pathway database for model organisms. Nucleic Acids Res. 2020;48(D1):D470–8.
49. Strömbäck L, Lambrix P. Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. Bioinformatics. 2005;21:4401–7.
50. Gillespie CS, Wilkinson DJ, Proctor CJ, Shanley DP, Boys RJ, Kirkwood TB. Tools for the SBML community. Bioinformatics. 2006;22:628–9.
51. van Iersel MP, Villéger AC, Czauderna T, Boyd SE, Bergmann FT, Luna A, Demir E, Sorokin A, Dogrusoz U, Matsuoka Y, et al. Software support for SBGN maps: SBGN-ML and LibSBGN. Bioinformatics. 2012;28:2016–21.
52. Xia J, Wishart DS. MetPA: a web-based metabolomics tool for pathway analysis and visualization. Bioinformatics. 2010a;26:2342–4.
53. Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res. 2010b;38(Web Server issue):W71–7.

54. Xia J, Fjell CD, Mayer ML, Pena OM, Wishart DS, Hancock RE. INMEX--a web-based tool for integrative meta-analysis of expression data. Nucleic Acids Res. 2013;41(Web Server issue):W63–70.
55. Karnovsky A, Weymouth T, Hull T, Tarcea VG, Scardoni G, Laudanna C, Sartor MA, Stringer KA, Jagadish HV, Burant C, Athey B, Omenn GS. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. Bioinformatics. 2012;28:373–80.

Untargeted Metabolomics in Newborn Screening

Joshua Manor and Sarah H. Elsea

Abstract Since its inception six decades ago, newborn screening has been lauded as a highly successful and cost-effective public health program by identifying disorders at the presymptomatic stage, enabling early disease-modifying intervention that otherwise invariably leads to death or permanent damage if treated at the symptomatic stage. The advent of multiplex high-throughput assays involving chromatography coupled with mass spectroscopy enabled the analysis of multiple disorders in a single run, vastly increasing the repertoire of screened disorders while keeping the cost nearly the same. Industrialized countries provide unified screening for more than 50 conditions, compared to about a dozen, a mere decade ago. Inevitably, we now screen, in essence, more than we know how to treat. Nonetheless, as a constant flow of new therapies breaks ground, providing accurate diagnostic data is vital for patient outcomes. Breaking the diagnostic barrier can mean new research, new drugs, and ultimately increased survival. In this chapter, we overview the concept of untargeted metabolomics as applied to newborn screening, how it fares compared to the well-standardized tests of the targeted screening, and its ability to screen for more disorders that are currently "unscreenable."

Keywords Untargeted metabolomics · Inborn error of metabolism (IEM) · Mass spectrometry (MS) · Dried blood spot (DBS) · Newborn screening (NBS) · Biomarker discovery

---

J. Manor
Metabolic Disease Unit, Haim Sheba Medical Center, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Israel

S. H. Elsea (✉)
Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
e-mail: Sarah.Elsea@bcm.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_5

1 Population-Wide Untargeted Screening

Variant reclassification, as demonstrated in Chap. 3, is not only important for accurate genetic counseling but also to guide management in treatable hereditary disorders, which is best exemplified in IEMs. IEMs constitute a large group of disorders of inherited defects in metabolic pathways and, if left untreated, can lead to significant morbidity and mortality [1]. In a global effort to reduce the IEM morbidity burden and to improve patients' outcomes, a more than half-century-long effort of newborn screening bore fruition in many countries. What started in the 1960s as a single disorder screening assay has evolved into an extended 50^{+} IEM screening machinery for each newborn [2]. This expansion was made possible primarily with the advent of MS-based targeted metabolomics (TM), allowing multiplexed screening in a single sample and facilitating the overall screening process. This is in contrast to the ad hoc methodology—each for a single disorder per test—e.g., electrophoresis isoelectric focusing for hemoglobinopathies and enzyme-linked immunosorbent assay (ELISA) for thyroxine (T4) quantification. The multiplexed assays have also altered the traditional view of NBS. From screening for disorders that (a) are occurring at significant frequency, (b) can be screened inexpensively, (c) include effective treatment that can avert significant morbidity, and (d) cannot be screened based on other signs and symptoms presented at birth [2], screening is concentrating now on only the second principal, and many disorders (some are exceedingly rare) are also caught in the wider net, despite having inadequate treatment options. One can conceive an "expanded" NBS not simply as a longer list of disorders that are screened at birth but actually as a screening procedure for disorders that are now screened for presymptomatically, merely because technology allows it. With the rapid growth of precision medicine, improved drug delivery, improved blood-brain barrier penetration, and the constant deluge of clinical trials, the need for expanded screening may prove beneficial quicker than previously imagined. To that end, a NBS expansion can mean moving forward from TM to untargeted metabolomics (UM).

So, our next question is how does a qualitative UM compared with quantitative TM when it comes to disease screening?

UM, by probing for not only a relevant subset of predefined metabolites (e.g., tyrosine levels in tyrosinemia patients), can identify up to 1000s [3] to 10,000s [4, 5] unique metabolites in each sample. Coupled with high-throughput separation techniques like high-resolution liquid chromatography (HPLC) or gas chromatography (GC), successful identification of disease-defining metabolites is made by demonstrating a statistically significant alteration (elevations or reductions), exceeding a $-2 \leq \log_2 \geq 2$ ratio, meaning a four-fold change from the normal range.

2 Screening for IEM Using UM

Miller et al. employed both GC and LC coupled to single-quadruple MS on 120 patients diagnosed with 21 different IEMs and positively identified 20 IEMs in 112 patients by showing perturbation of 2–20 metabolites in relevant pathways for each IEM [6]. Coene et al. applied a HPLC in tandem to quadrupole time-of-flight MS as a single platform on a cohort of 46 IEMs, identifying the key metabolites in 42 [4]. Bonte et al. utilized pentafluorophenylpropyl phase-based HPLC with Orbitrap MS in a cohort of 33 IEMs, identifying successfully 31 [5]. Haijes et al. applied direct infusion MS without chromatography-based separation, circumventing the need to create an experimental library for analytes' retention times [7]. Screens were performed on both dried blood spots (DBS) and plasma. Two additional cohorts were constructed for urine samples [8, 9] with biomarkers from 41 IEMs successfully detected using liquid chromatography coupled to either quadruple time-of-flight (QTOF) or high-resolution MS. Correct identification was reported for 90–95% of the IEMs. Results are detailed together in Table 1.

Importantly, UM is performed well on the "can't miss" diagnoses. Propionic acidemia (MIM 606054), an autosomal recessive inborn error of metabolism of propionic acid (PA), is due to a defective propionyl-CoA carboxylase. It results in the toxic accumulation of PA derived from the catabolism of methionine, valine, isoleucine, threonine, and odd-chain fatty acids. NBS reveals elevated propionylcarnitine (also referred to as C3), and confirmatory testing will show elevations of 2-methylcitrate and 3-OH-propionate. When utilizing UM, researchers found elevations in both primary and secondary metabolites while noticing a decrease in 2-methylmalonyl carnitine, an analyte associated with the product of propionyl-CoA carboxylase. Similar picture was obtained for the other two main organic acidemias, methylmalonic aciduria (MIM 251000) and isovaleric aciduria (MIM 243500). For the aminoacidopathy, maple syrup urine disease (MSUD), in which the keto acids of deaminated leucine, isoleucine, and valine cannot be further catabolized resulting in a severe encephalopathic disorder of hyperleucinosis, plasma was assessed from seven individuals. In these patients, isoleucine, the disease-pathognomonic biomarker alloisoleucine, and the corresponding analytes of the keto acids were found to be elevated, while downstream products were decreased: 3-hydroxyisobutyrate (valine catabolite), isovalerylcarnitine (isoleucine catabolite), and hydroxyisovalerylcarnitine (leucine catabolite), demonstrating a diagnostic biochemical profile for this disorder (a more complete profile that cannot be obtained by TM alone). For another can't miss diagnosis, medium-chain acyl-CoA dehydrogenase deficiency (MCADD, MIM 201450), an autosomal recessive disorder of β-oxidation of medium-chain fatty acids, not all known analytes were discovered, but a near pathognomonic picture with elevations of hexanoylcarnitine (C6) and octanoylcarnitine (C8) and the dicarboxylic acid suberic and sebacic acid was obtained, pointing out the diagnosis rather easily. The same was true for phenylketonuria, tyrosinemia type 1 (MIM 276700), and histidinemia (MIM 235800), while the finding of elevated tetradecenoylcarnitine was detected in four cases of very

Table 1 Successful identification of inborn errors of metabolism (IEM) based on untargeted metabolomics (UM) [4–9, 11, 12, 26–28]

Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Disorders of branched-chain amino acid metabolism	Propionic acidemia	606054	**Propionylcarnitine** (P), propionylglycine (U), 3-hydroxypropionic acid (U), methylcitric acid (P, U)	+	+		CR
	Methylmalonyl-CoA mutase deficiency	251000	**Propionylcarnitine** (P), methylmalonic acid (P, U), 3-hydroxypropionic acid (U), methylcitric acid (P, U)	+		+	CR
	Isovaleric acidemia	243500	**Isovalerylcarnitine** (P), isovaleric acid (P), isovalerylglycine (U), 3-hydroxyisovaleric acid (U)	+	+	+	CR
	3-methylcrotonyl-CoA carboxylase deficiency type 1 or 2	210200 and 210210	**Hydroxyisovalerylcarnitine** (P), 3-methylcrotonylglycine (U)	+	+	+	CR
	2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency[A]	300256	**2-methyl-3-hydroxybutyrylcarnitine (P)**, tiglylglycine (P, U)	+			SE
	2-methylbutyryl-CoA dehydrogenase deficiency	610006	**2-methylbutyrylcarnitine** (P, U)	+	+		SE
	3-hydroxy-3-methylglutaryl-CoA lyase deficiency	246450	**Hydroxyisovalerylcarnitine** (P), methylglutarylcarnitine (P), 3-hydroxy-3-methylglutaric acid (U), 3-hydroxyisovaleric acid (P, U), 3-methylglutaconic acid (P, U)	+	+		CR
	Methylmalonyl-CoA epimerase deficiency[B]	251120	Methylmalonic acid (U)	+			NO
	Malonyl-CoA decarboxylase deficiency	248360	Malonylcarnitine (P), malonic acid (U)	+	+		SE
	Combined methylmalonic and malonic aciduria (ACSF3 deficiency)[C]	614265	**Propionylcarnitine** (P), malonylcarnitine (P), methylmalonic acid (P, U), malonic acid (U)	+			NO
	Methylmalonate semialdehyde dehydrogenase deficiency[D]	614105	3-hydroxyisobutyric acid (U), 3-aminoisobutyric acid (U)		+		NO
	Maple syrup urine disease	248600	**Leucine** (P), isoleucine (P), valine (P), alloisoleucine (P), 2-oxoisocaproic acid (U), 2-oxo-3-methylvaleric acid (U), 2-oxoisovaleric acid (U)	+	+[E]	+	CR

Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Disorders of fatty acid oxidation and transport	Short-chain acyl-CoA dehydrogenase deficiency[F]	201470	**Butyrylcarnitine** (P), methylsuccinic acid (P, U), ethylmalonic acid (P, U)		+		NO
	Medium-chain acyl-CoA dehydrogenase deficiency	201450	**Octanoylcarnitine** (P), **hexanoylcarnitine** (P)**,** **decenoylcarnitine** (P)**,** decanoylcarnitine (P), hexanoylglycine (U), suberylglycine (U), 3-phenylpropionylglycine (U)	+	+	+	CR
	Very-long-chain acyl-CoA dehydrogenase deficiency	201475	**Tetradecenoylcarnitine** (P), C6-C10 dicarboxylic acids (U) [e.g., sebacic acid, adipic acid]	+		+	CR
	Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency	609016	**Hydroxyhexadecanoylcarnitine** (P)**, 3-hydroxy octadecenoylcarnosine** (P), C6-C14 dicarboxylic acids (U)	+		+	CR
	Carnitine palmitoyltransferase I deficiency	255120	**Carnitine** (P), reduction in other acylcarnitines, C6-C10 dicarboxylic acids (U) [e.g., sebacic, suberic, and adipic acids]	0	+[G]	+	SE
	Carnitine palmitoyltransferase II deficiency[H]	600649 608836 255110	**Hexadecanoylcarnitine** (P), stearoylcarnitine (P), oleoylcarnitine (P), Linoleoylcarnitine (P)	+		+	SE

(continued)

Table 1 (continued)

Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Disorders of ammonia detoxification	Carbamoyl phosphate synthetase I deficiency	237300	Ammonia (P), glutamine (P), arginine (P)	+[I]			NO
	Ornithine transcarbamylase deficiency	300461	Ammonia (P), glutamine (P), arginine (P), orotic acid (U)	+		0	NO
	Citrullinemia type 1	215700	**Citrulline** (P, U), ammonia (P), glutamine (P), orotic acid (U)	+	+		CR
	Argininosuccinic acid lyase deficiency	207900	**Citrulline** (P), argininosuccinate (P, U), ammonia (P), glutamine (P), orotic acid (P, U)	+			CR
	Arginase deficiency	207800	Arginine (P), ammonia (P) glutamine (P)	+			SE
	Hyperammonemia-hyperornithinemia-homocitrullinuria syndrome	238970	**Ornithine** (P), homocitrulline (U)	+	+/–		NO
Disorders of phenylalanine metabolism	Phenylketonuria, hyperphenylalaninemia	261600	**Phenylalanine** (P), phenylpyruvic acid (U)	+	+	+	CR
Disorders of tyrosine metabolism	Tyrosinemia type 1	276700	**Tyrosine** (P), **succinylacetone** (P, U)	+	+	J	CR
	Alkaptonuria	203500	Homogentisate (U)	0	+		NO
	Hawkinsinuria	140350	4-hydroxycyclohexylacetate (U), hawkinsin (U)			+	NO
Disorders of sulfur amino acid and sulfide metabolism	Homocystinuria	236200	**Methionine** (P), homocysteine (P, U)	+		+	CR
	Methionine adenosyltransferase I/III deficiency	250850	Methionine (P), S-adenosylmethionine (P), methionine sulfoxide (U)	+		+	NO
	Ethylmalonic encephalopathy[F]	602473	C4-butyrylcarnitine (P), 2-methylbutyrylcarnitine (P), ethylmalonic acid (P, U), methylsuccinic acid (P, U)	+			NO

Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Disorders of lysine metabolism	Glutaric acidemia type 1	231670	Glutarylcarnitine (P), glutaric acid (U, P), 3-hydroxyglutaric acid (U)	+	+	+/−[K]	CR
	Pyridoxine-dependent epilepsy	266100	L-α-aminoadipic semialdehyde (P,U), pipecolic acid (P)	+			NO
	Hyperlysinemia type 1 or 2[L]	268700 238700	Lysine (P, U), saccharopine (P, U)	+			NO
Disorders of proline and ornithine metabolism	Hyperprolinemia type 2	239510	Proline (P, U), hydroxyproline (P, U), pyrroline-5-carboxylate, (P, U)	+			NO
	Ornithine aminotransferase deficiency	613349	**Ornithine** (P, U), lysine (U), arginine (U)	+		+	NO
Disorders of serine and glycine metabolism	Phosphoglycerate dehydrogenase deficiency	601815	Low CSF serine, low fasting plasma serine	+		0[M]	NO
	Glycine encephalopathy[N]	605899	Glycine (P)	+		+	NO
Disorder of histidine metabolism	Histidinemia	235800	Histidine (P, U)	+			NO
Disorders of purine metabolism	Adenylosuccinate lyase deficiency	103050	Succinyladenosine (P, U), succinylaminoimidazolecarboxamide riboside (SAICA riboside) (P, U)	+	+		NO
	Hypoxanthine guanine phosphoribosyltransferase deficiency	300322 and 300323	Uric acid (P, U), hypoxanthine (P, U)	+			NO
Disorders of pyrimidine metabolism	Uridine monophosphate synthase deficiency	258900	Orotic acid (P, U)	+	+		NO
	β-Ureidopropionase deficiency[O]	613161	N-carbamoyl-β-alanine (P, U), N-carbamoyl-β-aminoisobutyric acid (P)	+			NO
Disorders of carnitine metabolism	Primary carnitine deficiency	212140	**Low free carnitine** (P, U)	+		+	CR
	Trimethyllysine hydroxylase epsilon deficiency	300872	Low free carnitine (P)	+			NO

(continued)

Table 1 (continued)

Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Disorders of ketone body metabolism	β-Ketothiolase T2 deficiency[A]	203750	**2-methyl-3-hydroxybutyrylcarnitine** (P, U), tiglylglycine (P, U), 2-methylacetoacetic acid (U)	+	+		CR
	3-hydroxy-3-methylglutaryl-CoA synthase 2 deficiency	605911	4-hydroxy-6-methyl-2-pyrone (U), 3-OH-dicarboxylic acids (U)		+		NO
Disorders of biotin metabolism	Holocarboxylase synthetase deficiency	253270	**Hydroxyisovalerylcarnitine** (P), 3-methylcrotonylglycine (U), 3-hydroxypropionic acid (U), lactate (P)	+			CR
Disorders of cobalamin metabolism	Cobalamin-related disorder[P]		See comment[P]	+			CR
Disorders of folate metabolism	Glutamate formiminotransferase deficiency	229100	Formiminoglutamate (P, U), hydantoin-5-propionate (U)	+			NO
	Methylenetetrahydrofolate reductase deficiency[Q]	236250	Homocysteine (P)	0		+	NO
Disorders of riboflavin metabolism	Multiple acyl-CoA dehydrogenase deficiency	231680	**Glutarylcarnitine** (P), **isovalerylcarnitine** (P), C6-C18 esters (P), ethylmalonic acid (U), isovaleric acid (U), glutaric acid (U), isovalerylglycine (U), hexanoylglycine (U)	+	+		SE
Disorders of molybdenum metabolism	Molybdenum cofactor deficiency[R]	252150, 252160, 615501	S-sulfocysteine (P, U), taurine (P, U), xanthine (P, U), low uric acid (P)			+/−	NO
	Xanthinuria type 2[R]	603529	Hypoxanthine (P, U), xanthine (P, U)	+			NO

Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Disorders of neurotransmitters	Aromatic L-amino acid decarboxylase deficiency	608643	**High 3-O-methyldopa (CSF, P)** and 5-hydroxytryptophan (CSF); **low** homovanillic acid (CSF, P), 5-hydroxyindoleacetic acid (CSF, P), 3-methoxy-4-hydroxyphenyl glycol (CSF), and dopamine-3-O-sulfate (P)	S	+		NO
	Succinic semialdehyde dehydrogenase deficiency	271980	4-hydroxybutyric acid (U), γ-aminobutyric acid (P)		+		NO
Disorders of creatine metabolism	Guanidinoacetate methyltransferase deficiency	612736	Guanidinoacetate (P, U), low creatine (P, U)	+[T]		+	NO
Disorders of amino acid transport	Lysinuric protein intolerance	222700	Lysine (U), ornithine (U), arginine (P, U), citrulline (P), orotic acid (U), glutamine (P), ammonia (P)	+[U]	+	+/−	NO
Disorders of the pentose phosphate pathway and glutathione synthesis	Transaldolase deficiency	606003	Sedoheptulose (U), sedoheptulose-7-P (U), erythronic acid (U), erythritol (U), ribitol (U), arabitol (U)	+[V]	+		NO
	Transketolase deficiency	617044	Arabitol (P,U), erythritol (P, U), ribitol (P, U)	+	+		NO
	Glutathione synthetase deficiency	266130	5-oxoproline (U)		+		NO
Disorders of cholesterol and bile acid synthesis	Mevalonic aciduria	610377	Mevalonic acid (U)	0	+		NO
	Cerebrotendinous xanthomatosis	213700	Cholestanol (P)	+			NO
	α-Methylacyl-CoA racemase deficiency	604489	(25R)-3a,7a,12a-trihydroxy-5b-cholestan-26-oic acid (THCA) (P), (25R)-3a,7a-dihydroxy-5b-cholestan-26-oic acid (DHCA) (P), pristanic acid (S)	+			NO
Disorders of galactose and fructose metabolism	Galactosemia	230400	**Galactose-1-phosphate** (P), **galactose** (P), galactitol (P, U)	+[W]			CR
	Glycerate kinase deficiency	610516	D-glycerate (P, U)	+			NO

(continued)

Table 1 (continued)

Group	Disease	MIM	Common diagnostic markers (in bold—when analytes used for NBS-DBS)	P	U	DBS	RUSP
Aminoacylase deficiency	Aminoacylase I deficiency	609924	N-acetylalanine (U), N-acetylglutamic acid (U), N-acetylglycine (U)	+	+		N
	Canavan disease (aspartoacylase deficiency)	271900	N-acetylaspartic acid (P, U)	+	+		N
Peroxisomal disorders	Refsum disease	266500	Phytanic acid (P), pristanic acid (P), pipecolic acid (P, U)	+[X]			N
	Primary hyperoxaluria type 1	259900	Oxalic acid (P, U), glycolic acid (P, U)		+[Y]		N
Oligosaccharidosis	β-Mannosidosis	248510	–	+			N
Disorders of the Krebs cycle	Fumarate hydratase deficiency	606812	2-ketoglutarate (U), lactate (P), succinic acid (U), fumaric acid (U)		+		N
	GTP-specific succinyl-CoA ligase α subunit deficiency	245400	**Propionylcarnitine** (P), methylmalonic acid (P, U)	+/–[Z]	+		N
Disorders of mitochondrial nucleotide pool maintenance	Mitochondrial neurogastrointestinal encephalopathy (MNGIE)	603041	Thymidine (P, U), deoxyuridine (P, U)	+			N
Congenital disorders of glycosylation	N-acetylneuraminic acid phosphate synthase deficiency	610442	N-acetyl-D-mannosamine (P, U)	+			N

MIM accession number of the (online) Mendelian inheritance in man, *P* plasma (sample), *U* urine, *S* serum, *DBS* dried blood spot, *RUSP* Secretary of the Department of Health and Human Services (HHS) Recommended Uniform Screening Panel, + successful identification, 0 failed to identify, +/− identification of only some diagnostic metabolites

For RUSP, CR denotes a core recommendation, SE, secondary recommendation, and NO, does not appear in the list

A—2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency and β-ketothiolase deficiency cannot be differentiated based on elevated analytes: tiglylglycine, tiglylcarnitine, and 2-methyl-3-hydroxybutyric acid

B—Altered analytes for methylmalonyl-CoA epimerase deficiency are propionylcarnitine and 3-OH-propionic acid, which are also elevated in other disorders branched-chain amino acid metabolism, e.g., the more prevalent propionic academia

C—Combined methylmalonic and malonic aciduria was identified by the elevation of methylmalonic acid without elevation of malonic acid, making this metabolic profile nonspecific

D—Elevation of 3-hydroxy-isobutyric acid is suggestive of methylmalonate semialdehyde dehydrogenase deficiency as it does not accumulate in 3-hydroxy-isobutyryl-CoA hydrolase deficiency (HIBCH, MIM 611283)

Table 1 (continued)

E—In urine, MSUD was identified by elevations of 2-hydroxy-3-methlypentanoic acid, 2-hydroxy-3-methylbutyric acid, and 2-oxoisocaproic acid, which are different analytes than the "traditional" MSUD analytes of alloisoleucine and leucine (and in addition, valine and isoleucine)

F—Currently, the benign short-chain acyl-CoA dehydrogenase deficiency and the severe ethylmalonic encephalopathy are identified similarly with elevations in ethylmalonic acid, isobutyrylglycine, and methylsuccinic acid

G—Identification of CPT1 deficiency was achieved based on elevations of medium-chain dicarboxylic acids which are the detectable metabolite in urine in this condition; however, these findings are nonspecific for this condition and can also indicate medium-chain acyl-CoA dehydrogenase deficiency. In addition, pyrrole-2-carboxylic acid was significantly elevated in the urine, a finding not known to be associated with CPT1 but rather with hyperprolinemia type 2 [29, 30]

H—Carnitine palmitoyltransferase II deficiency cannot be distinguished based on metabolic profiling from carnitine acylcarnitine translocase deficiency (MIM 212138)

I—Altered analytes for carbamoyl phosphate synthetase I deficiency include elevation of glutamine and glutamate; both are nonspecific indicators for compensated hyperammonemic status

J—Tyrosinemia of uncertain type was correctly identified by DBS with elevation of 4-hydroxyphenyllactic acid, 4-hydroxyphenyllacetic acid, 4-hydroxyphenylpyruvic acid, and tyrosine but without succinylacetone, fitting for a diagnosis of tyrosinemia type 2 or type 3

K—DBS detected elevations in glutarylcarnitine, similar to traditional targeted screening, but failed to detect the pathognomonic 3-hydroxyglutaric acid (which was detected readily in plasma and urine)

L—Elevations in lysine, N-acetyl-lysine, and L-pipecolic acid cannot distinguish between the two forms of hyperlysinemia

M—Failure of DBS to detect reduction in serine and glycine is attributed to supplementation. However, one should remember that low levels are more difficult to detect on DBS as the lower limit of normal can approach 0

N—Only 50% of cases of glycine encephalopathy (nonketotic hyperglycinemia) were detected by the elevation of glycine, due to only mild elevations in the samples

O—Identification of β-ureidopropionase deficiency was made based on elevations of dihydrouracil and dihydrothymine. This profile can also fit dihydropyrimidinase deficiency (MIM 222784)

P—Several disorders of intracellular cobalamin metabolism can fit the screening results of elevation of 2-methylmalonylcarnitine, propionylcarnitine, and 2-methylcitrate, a metabolic profile of methylmalonic acidemia (MMA). No homocysteine was detected in plasma, which is not detected in many assays as it requires an additional unique processing. According to HHS RUSP, screening for cobalamin defects should be performed by screening for MMA. Disorders of intracellular cobalamin metabolism resulting in isolated MMA elevations (without elevations in homocysteine) are cobalamin A (MMAA, MIM 251100), cobalamin B (MMAB, MIM 251110), and several cases of cobalamin D (MIM 277410)

Q—Identification of methylenetetrahydrofolate reductase (MTHFR) deficiency was achieved by elevation of homocysteine derivatives (homocysteine thiolactone and homocysteine). This is not specific for MTHFR deficiency and can be found also in other re-methylation defects, such as cobalamin G (MIM 250940), cobalamin E (MIM 236270), and some cobalamin D (MIM 277410) defects. See also comment P above for homocysteine detection

Table 1 (continued)

R—Consistently found altered in MoCD in the cohorts above is elevation of xanthine. Only one plasma sample presented, in addition, statistically significant reduction in uric acid and elevation of S-sulfocysteine [7]. Other samples did not show statistically significant reduction in uric acid including samples from DBS. Elevation of xanthine and reduction of uric acid can also be seen in xanthinuria type 1 (MIM 278300) and type 2 (MIM 603592)

S—Successful identification of this disorder was previously reported [31] by identification of elevated 3-methoxytyrosine (3-*O*-methyldopa) in plasma, a methylation product of L-DOPA by catechol-*O*-methyltransferase (COMT). 3-methoxytyrosine is elevated when L-DOPA is not further metabolized to dopamine due to defective aromatic L-amino acid decarboxylase (AADC)

T—Guanidinoacetate failed detection in two cohorts but was identified in a third

U—Lysinuric protein intolerance was identified successfully in two studies while two others failed. LPI diagnostic rates increase when combined with a urine sample

V—In one case, only sedoheptulose was detected in the sample. This finding is not specific for transaldolase deficiency and can be seen in isolated sedoheptulose kinase deficiency (MIM 617213), a benign condition but also due to a common copy number variant found in patients with cystinosis (MIM 219800)

W—In plasma, only galactitol was identified in a single case of galactosemia. Galactitol will be elevated also in galactokinase deficiency (MIM 230200) and galactose epimerase deficiency (MIM 230350). In urine, elevations in galactose were also seen, giving a similar nonspecific picture

X—Identification of Refsum disease (phytanoyl-CoA hydroxylase deficiency) was made based on elevation of phytanic acid. Elevation of pipecolic acid was not seen. Elevation in phytanic acid can also be seen in peroxisomal biogenesis defects (Zellweger spectrum disorder) and rhizomelic chondrodysplasia punctata (MIM 215100)

Y—Identification of primary hyperoxaluria type 1 was made based on elevation of glycolic acid. These elevations can also be seen in fumarate hydratase deficiency (in which fumaric acid will be elevated) or D-2-hydroxyglutaric acid (in which 2-OH-glutaric acid will be elevated)

Z—Elevated MMA was detected in plasma, indicating a disorder of elevated MMA. In urine, the additional detection of elevated lactate creates a more specific diagnosis of this disorder of mitochondrial DNA depletion

Table 2 Examples of diagnostic metabolites detected by untargeted metabolomics leading to non-benign IEMs that are not routinely screened by NBS

Metabolite(s)	Condition(s)	MIM	Specimen
N-acetylalanine, N-acetylglutamate, N-acetylmethionine, N-acetylleucine	Aminoacylase I deficiency	609924	Plasma, urine
N-acetylaspartate	Canavan disease	271900	Plasma
Orotic acid	Uridine monophosphate synthase deficiency (primary), urea cycle defects (secondary)	613891 (primary)	Plasma, urine
Mevalonate	Mevalonic aciduria	610377	Urine
Homocysteine thiolactone/ homocysteine[a]	Methylenetetrahydrofolate reductase deficiency, cobalamin-related disorders	236250 236270 250940 277380 277400 277410 309541	Plasma, DBS
Xanthine, S-sulfocysteine	Molybdenum cofactor deficiency types A, B and C, isolated sulfite oxidase deficiency	252150 252160 615501 272300	Plasma
Succinyladenosine	Adenylosuccinate lyase deficiency	103050	Plasma
3-methoxytyrosine (3-*O*-methyldopa)	Aromatic L-amino acid decarboxylase deficiency	608643	Plasma, urine
Guanidinoacetate↑, +/− ↓creatine	Guanidinoacetate methyltransferase deficiency	612736	Plasma, DBS
Proline and ornithine	Ornithine aminotransferase deficiency[b]	258870	Plasma, DBS
↓Lysine, ↓ornithine, ↓arginine	Lysinuric protein intolerance	222700	Plasma
Glycine	Nonketotic hyperglycinemia	605899	Plasma, DBS
Sedoheptulose	Transaldolase deficiency, rare cases of cystinosis	606003	Plasma
↓Serine, ↓glycine	Phosphoglycerate dehydrogenase deficiency[c]	256520 601815	Plasma
Cholestane-3,7,12,25,25-pentol	Cerebrotendinous xanthomatosis	213700	Plasma
Di- and tri-hydroxy-5b-cholestan-26-oic acids[d]	α-Methylacyl-CoA racemase deficiency	614307	Plasma
N-acetyl-D-mannosamine	N-acetylneuraminic acid phosphate synthase deficiency	610442	Plasma
Phytanic acid[e]	Peroxisomal biogenesis defects, Refsum disease	266500 (Refsum)	Plasma
Fumaric acid	Fumarate hydratase deficiency	606812	Urine

Table 2 (continued)

Metabolite(s)	Condition(s)	MIM	Specimen
4-Hydroxy-6-methyl-2--pyrone[d]	3-hydroxy-3-methylglutaryl-CoA synthase 2 deficiency	605911	Urine
2-Pyrrolidinone, succinamic acid	GABA transaminase deficiency	613163	Plasma

UM untargeted metabolomics, *IEM* inborn error of metabolism, *NBS* newborn screening, ↓ reduction in metabolite; otherwise, alteration is elevation

[a]With normal levels of methionine. Reported only by some laboratories as the detection of homocysteine requires special preparatory steps

[b]Elevations of proline and ornithine may not be seen in ornithine aminotransferase deficiency during the neonatal period due to substrate stoichiometry

[c]Similar metabolomics picture can be seen in phosphoserine aminotransferase 1 deficiency (PSAT1D, MIM 616038), which can present similarly to phosphoglycerate dehydrogenase deficiency

[d]Not routinely reported by all laboratories

[e]Alterations may also include the elevation of pipecolic acid

long-chain acyl-CoA dehydrogenase deficiency (VLCADD, MIM 201475) in one study [4], although undetected in two cases from another study [6]. Homocysteine and homocysteine thiolactone are both accumulating in cystathionine β-synthase deficiency (homocystinuria, MIM 236200) and 5,10-methylenetetrahydrofolate reductase deficiency (MTHFR, MIM 607093) [1, 10]. These two disorders were successfully diagnosed in DBS by demonstration of significant elevation in one of the two analytes in conjunction with elevation of methionine (a “can’t miss” diagnosis) in the former but not the latter. Homocysteine was only elevated in DBS, as in plasma metabolomics homocysteine is not detected using high-throughput methodology for sample prep, which does not allow for the pretreatment required for proper analysis.

Patients with β-ketothiolase deficiency (T2 deficiency, MIM 203750), a disorder of isoleucine breakdown and ketone utilization deficiency, were found to have elevation of tiglylglycine and 2-methyl-3-hydroxybutyric acid with no elevation of 2-methylacetoacetic acid, making this metabolic profile indistinguishable from 2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency (HSD17B10 deficiency, MIM 300438). Contrarily, for the single patient with 2-methyl-3-hydroxybutyric deficiency in the cohorts, tiglylglycine was elevated, but 2-methyl-3-hydroxybutyric was not. This suggests that tiglylglycine is a sensitive marker for both of these two distal isoleucine breakdown disorders; however, specific biomarkers for either disorder are lacking.

An important group of disorders that are only partially screened for in NBS are urea cycle defects. Most notably is the severe ornithine transcarbamylase deficiency, an X-linked urea cycle defect (OTCD, MIM 300461) for which early detection is crucial for improved outcomes, yet this condition is difficult to detect due to limitation on measuring low quantities (of citrulline in the case of OTCD) and extraction of orotic acids in high-throughput methods. In one case of OTCD, significant elevations of uridine were found, which is a secondary disease biomarker. The elevation

in uridine stems from shunting of carbamoyl phosphate from the urea cycle into the pyrimidine biosynthesis pathway via the orotic acid intermediate; however, orotic acid was not elevated in the two patients with OTCD. For citrin deficiency (citrullinemia type 2, MIM 603471 and 605814), citrulline was elevated along with amino acid markers of liver injury (tyrosine and its degradation products 4-hydroxyphenyllactic acid 4-hydroxyphenylpruvic acid, lysine, arginine, and 5′-S-methyl-5′-thioadenosine), a common finding in neonatal presentation of this disorder. In a patient with hyperornithinemia-hyperammonemia-homocitrullinuria syndrome (HHH, MIM 238970), a secondary UCD caused by mitochondrial ornithine transporter defect, only general markers for urea cycle dysfunction were identified (elevations in uracil and orotic acid), yet this finding should merit a focused investigation for UCDs. Homocitrulline and ornithine were both elevated in plasma UM of a 30-year-old patient with HHH, [11], while it failed to detect elevations in citrulline in a patient with citrin deficiency in another study [12]. For OTCD, DBS samples showed orotic acid elevation in one of two cases; however, no significant elevations in uridine or uracil were found, in contrast to plasma, where both were elevated. Nonetheless, these results are somewhat encouraging giving hope for expanded DBS screening for the most common urea cycle defect [2], with severe presentation in males and no current screening protocols.

A few important examples demonstrate the increased diagnostic repertoire of UM (see Table 2). Lysinuric protein intolerance (LPI, 222700) is a disorder caused by defective cationic amino acid (CAA) transport at the basolateral membrane of epithelial cells in the kidney and intestine leading to increased renal excretion of CAA leading to the depletion of mainly lysine, arginine, and ornithine, the latter leading to secondary urea cycle defect (UCD) which can result in hyperammonemia. UM in two patients identified excess glutamine, a marker for compensated hyperammonemia and reduction of plasma CAA (although mild in one study). In urine, low glutarylcarnitine and N^6-trimethyllysine were recorded, with mildly elevated N^6-acetyllysine [9]; the latter two are products of lysine catabolism. Urine levels of CAA did not differ significantly from the control.

Successful identifications of four severe nucleotide degradation IEMs have been reported, two pyrimidine degradation disorders, β-ureidopropionase deficiency (MIM 606673), in plasma, and dihydropyrimidine dehydrogenase deficiency (MIM 274270), in urine, and two purine metabolism syndromes: X-linked hypoxanthine-guanine phosphoribosyltransferase deficiency (causing either Lesch-Nyhan syndrome (MIM 300322) or HPRT-related hyperuricemia (MIM 300323)) in two patients (one each in plasma and in urine) and adenylosuccinate lyase deficiency (MIM 608222) in the other two (also one each in plasma and in urine). These results are promising, as currently purine and pyrimidine metabolism defects can cause a severe neurological phenotype which may be amenable for treatment but currently are not routinely screened among presymptomatic newborns. Moreover, elevation of xanthine and xanthosine (purine catabolites) and low uric acid (the degradation product of xanthine) were found in five patients, three with molybdenum cofactor deficiency (MoCD, MIM 252150) and two with xanthinuria type 2 (MIM 603592). The former disorder is a severe, rapidly progressive encephalopathy due to

deficiency in the molybdenum and biopterin cofactors, the latter a benign disorder of accumulation of xanthine due to defect in xanthine oxidase and aldehyde oxidase (AO, affecting mainly N^1-methylnicotinamide degradation). Of note, elevations in the specific MoCD marker S-sulfocysteine was seen in only one sample, while thiosulfate and taurine were not detected. On DBS, increased levels of α-amino adipic semialdehyde (AASA) were also seen, which are also observed in sulfite oxidase deficiency (MIM 272300) and AASA deficiency (antiquitin deficiency, also called pyridoxine-dependent epilepsy, MIM 266100) [13], both are disorders bearing importance of early screening. For the MoCD samples without elevated S-sulfocysteine, both patients were under treatment; however, a profile containing elevations in xanthine and xanthosine and reduction of uric acid is expected to include sulfa-containing analytes, as well. Identification of these metabolites can facilitate the diagnosis of MoCD, which for its most common genetic cause a therapeutic option is available, and avoidance of permanent damage can be achieved if therapy is initiated within days after birth [14]. 3-phosphoglycerate dehydrogenase (3-PGDH) deficiency (Neu-Laxova syndrome, MIM 606879), a severe neurodevelopmental due to serine biosynthesis defect, is characterized by low plasma levels of serine and glycine, and while this profile was correctly captured in the plasma, it was not observed in DBS, presumably due to treatment.

Cerebrotendinous xanthomatosis (CTX, MIM 213700), an autosomal recessive bile acid synthesis disorder resulting in a neurodegeneration and premature atherosclerosis due to diffuse xanthomata, was successfully identified in five patients due to elevations in cholestane-3,7,12,24,25-pentol, a specific marker for CTX [15]. These results would allow early treatment with chenodeoxycholic acid early in the course of the disease and can prevent neurodegeneration in a disorder otherwise diagnosed well into the development of systemic symptoms. Elevations of phytanic acids were seen in one patient with adult Refsum disease (MIM 266500), a disease of defective α-oxidation of phytanoyl esters into pristanic esters, resulting in vision decline, hearing loss, polyneuropathy, ichthyosis, and cardiac conduction defects. Early recognition can initiate low phytanic acid diet can improve significantly the overall prognosis of this disorder that is usually diagnosed after the appearance of symptoms [1].

Unlike organic acidemias, amino acidopathies, and urea cycle disorders, the screening for organelle-based diseases, such as storage diseases, lags behind with currently only two types of mucopolysaccharidoses being screened [16]. UM offers hope as a highly sensitive screening tool for additional metabolites that can indicate abnormal intra-organelle catabolism, which are pathognomonic for certain disorders. For lysosomes, β-mannosidosis, a neurodegenerative lysosomal storage disorder (LSD) due to a defect in removal of β-D-mannose from glycoproteins (MIM 248510), was successfully identified by elevation of the pathognomonic mannosyl-β1,4-N-acetylglucosamine (GlcNAc-Man). In the mitochondria, mitochondrial neurogastrointestinal encephalopathy (MNGIE, MIM 603041) is a multisystem disorder of mitochondrial DNA depletion causing severe gastrointestinal (GI) dysmotility with peripheral neuropathy, hearing and vision impairment, and early death in the fourth decade [17]. On average, diagnosis is delayed by 12 years [18] due to

the initial nonspecific GI symptoms due to deficiency of thymidine phosphorylase. Here, UM screening assists in early diagnosis, with elevations of thymidine and deoxyuridine, and, thus, can reduce morbidity with the emergence of new therapies [17].

In the cohorts of urine samples [8, 9], two patients with aromatic L-amino acid decarboxylase deficiency (AADCD, MIM 608643), a monoamine neurotransmitter synthesis defect, had significant elevations of vanillactic acid, which is expected to increase with the accumulation of L-DOPA, a substrate of AADC. In addition, 4-hydroxybutyric acid was elevated in two patients with succinic semialdehyde dehydrogenase deficiency (SSADHD, MIM 271980), a γ-aminobutyric acid (GABA) catabolism defect. Both conditions are severe and are not typically screened, with emerging gene therapy options for the former [19, 20]. Additionally, adenosine deaminase deficiency, a purine degradation disorder that is the fourth most common etiology for severe combined immunodeficiency (SCID, MIM 102700) [21], was successfully screened by the identification of 2-deoxyadenosine and 2-deoxyinosine [9].

Few notable limitations of UM are worth mentioning. For GAMT deficiency, which converts guanidinoacetate (GAA) and S-adenosylmethionine (SAM) into creatine and S-adenosylhomocysteine (SAH), only one study out of the three successfully identified GAA accumulation and reduction of creatine in plasma. However, more recently, an increased sensitivity to GAA is reported by laboratories, which may lead to improved identification of GAMT deficiency. In plasma, homogentisic acid was not detected in the case of alkaptonuria, a tyrosine breakdown disorder causing a degenerative joint and cartilage disorder (MIM 203500), and similarly, mevalonic acid was not detected for mevalonic aciduria, a heterogeneous multisystemic disorder of cholesterol synthesis (MIM 610377 and 260920). In urine samples, however, both homogentisic acid and mevalonic acid were detected readily [8]. For LPI, not all cohorts detected low levels of CAA, leaving the nonspecific hyperglutaminemia in plasma the only consistent finding in these cases. In urine, alteration of lysine metabolites was noted (low N^6-trimethyllysine and high N^6-acetyllysine), as mentioned above.

Importantly, despite all methods grouped under “UM,” different preparation techniques yielded different metabolic profiles and detection rates of the IEMs. Examples are LPI, VLCADD, GAMT deficiency, and alkaptonuria, in which the primary metabolites CAA, C14:1, GAA, and homogentisic acid were found only in some studies but not in others. Additionally, while UM detected up to 10,000 analytes per sample [4], only about 300–400 analytes were matched to known library of reference and analyzed for perturbations to create metabolic profiling [4, 5]. Large population screening, such as newborn screening, must maintain consistency in the selection of derivatives to be detected; the same small group of representative molecules must be detected in each and every sample. To be effectively used for large population screening, steps for streamlining sample collection and preparation must be undertaken, as well efforts to reduce unreliable peak integration in the analysis process, supporting semiautomated result interpretation [22]. One approach is

the implementation of automated interpretation pipeline [23] or utilization of machine learning to more clearly define biochemical phenotypes [24, 25].

Glossary

Dried blood spot (DBS) A method of whole blood sample collection, in which a small amount of fresh blood is blotted onto an absorbent filter paper, followed by drying. This method provides a convenient storage and shipment platform and is widely used for newborn screening. Typically, a small punch from the DBS paper is eluted with phosphate-buffered saline, availing the sample for testing.

Multiplex assay An assay measuring simultaneously multiple analytes in a single testing. These tests are becoming more popular in the metabolic sciences where several similar analytes are tested for alterations from the normal range, e.g., urine polyols for evaluation of the pentose phosphate pathway, urine glycosaminoglycans for the diagnosis of mucopolysaccharidoses, or carbohydrate moieties for congenital disorders of glycosylation.

References

1. Saudubray JM, Baumgartner MR, Walter JH. Inborn metabolic diseases: diagnosis and treatment. Berlin, Heidelberg: Springer; 2016.
2. Lee B, Scaglia F. Inborn errors of metabolism : from neonatal screening to metabolic pathways. New York, NY: Oxford University Press; 2015.
3. Wright Muelas M, Roberts I, Mughal F, O'Hagan S, Day PJ, Kell DB. An untargeted metabolomics strategy to measure differences in metabolite uptake and excretion by mammalian cell lines. Metabolomics. 2020;16(10):107.
4. Coene KLM, Kluijtmans LAJ, van der Heeft E, Engelke UFH, de Boer S, Hoegen B, et al. Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients. J Inherit Metab Dis. 2018;41(3):337–53.
5. Bonte R, Bongaerts M, Demirdas S, Langendonk JG, Huidekoper HH, Williams M, et al. Untargeted metabolomics-based screening method for inborn errors of metabolism using semi-automatic sample preparation with an UHPLC-Orbitrap-MS Platform. Metabolites. 2019;9(12):289.
6. Miller MJ, Kennedy AD, Eckhart AD, Burrage LC, Wulff JE, Miller LA, et al. Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. J Inherit Metab Dis. 2015;38(6):1029–39.
7. Haijes HA, Willemsen M, Van der Ham M, Gerrits J, Pras-Raves ML, Prinsen H, et al. Direct infusion based metabolomics identifies metabolic disease in patients' dried blood spots and plasma. Metabolites. 2019;9(1):12.
8. Körver-Keularts I, Wang P, Waterval H, Kluijtmans LAJ, Wevers RA, Langhans CD, et al. Fast and accurate quantitative organic acid analysis with LC-QTOF/MS facilitates screening of patients for inborn errors of metabolism. J Inherit Metab Dis. 2018;41(3):415–24.
9. Kennedy AD, Miller MJ, Beebe K, Wulff JE, Evans AM, Miller LA, et al. Metabolomic profiling of human urine as a screen for multiple inborn errors of metabolism. Genet Test Mol Biomarkers. 2016;20(9):485–95.

10. Jakubowski H, Boers GH, Strauss KA. Mutations in cystathionine beta-synthase or methylenetetrahydrofolate reductase gene increase N-homocysteinylated protein levels in humans. FASEB J. 2008;22(12):4071–6.
11. Liu N, Xiao J, Gijavanekar C, Pappan KL, Glinton KE, Shayota BJ, et al. Comparison of untargeted metabolomic profiling vs traditional metabolic screening to identify inborn errors of metabolism. JAMA Netw Open. 2021;4(7):e2114155.
12. Almontashiri NAM, Zha L, Young K, Law T, Kellogg MD, Bodamer OA, et al. Clinical validation of targeted and untargeted metabolomics testing for genetic disorders: a 3 year comparative study. Sci Rep. 2020;10(1):9382.
13. Mills PB, Footitt EJ, Ceyhan S, Waters PJ, Jakobs C, Clayton PT, et al. Urinary AASA excretion is elevated in patients with molybdenum cofactor deficiency and isolated sulphite oxidase deficiency. J Inherit Metab Dis. 2012;35(6):1031–6.
14. Schwahn BC, Van Spronsen FJ, Belaidi AA, Bowhay S, Christodoulou J, Derks TG, et al. Efficacy and safety of cyclic pyranopterin monophosphate substitution in severe molybdenum cofactor deficiency type a: a prospective cohort study. Lancet. 2015;386(10007):1955–63.
15. Shefer S, Dayal B, Tint GS, Salen G, Mosbach EH. Identification of pentahydroxy bile alcohols in cerebrotendinous xanthomatosis: characterization of 5beta-cholestane-3alpha, 7alpha, 12alpha, 24xi, 25-pentol and 5beta-cholestane-3alpha, 7alpha, 12alpha, 23xi, 25-pentol. J Lipid Res. 1975;16(4):280–6.
16. Jalal K, Carter RL, Barczykowski A, Tomatsu S, Langan TJ. A roadmap for potential improvement of newborn screening for inherited metabolic diseases following recent developments and successful applications of bivariate Normal limits for pre-symptomatic detection of MPS I, Pompe disease, and Krabbe disease. Int J Neonatal Screen. 2022;8(4):61.
17. Bax BE. Mitochondrial neurogastrointestinal encephalomyopathy: approaches to diagnosis and treatment. J Transl Genet Genom. 2020;4:1–16.
18. Danjou M, Guardia D, Geoffroy PA, Seguy D, Cottencin O. Mitochondrial neuro-gastro-intestinal encephalopathy (MNGIE): when and how to suspect it in front of an atypical anorexia nervosa? Encéphale. 2016;42(6):574–9.
19. Tai CH, Lee NC, Chien YH, Byrne BJ, Muramatsu SI, Tseng SH, et al. Long-term efficacy and safety of eladocageneexuparvovec in patients with AADC deficiency. Mol Ther. 2021;30:509.
20. Merola A, Kobayashi N, Romagnolo A, Wright BA, Artusi CA, Imbalzano G, et al. Gene therapy in movement disorders: a systematic review of ongoing and completed clinical trials. Front Neurol. 2021;12:648532.
21. Routes J, Verbsky J. Newborn screening for severe combined immunodeficiency. Curr Allergy Asthma Rep. 2018;18(6):34.
22. Chetnik K, Petrick L, Pandey G. MetaClean: a machine learning-based classifier for reduced false positive peak detection in untargeted LC-MS metabolomics data. Metabolomics. 2020;16(11):117.
23. Hoegen B, Zammit A, Gerritsen A, Engelke UFH, Castelein S, van de Vorst M, et al. Metabolomics-based screening of inborn errors of metabolism: enhancing clinical application with a robust computational pipeline. Metabolites. 2021;11(9):568.
24. Heinemann J. Machine learning in untargeted metabolomics experiments. Methods Mol Biol. 2019;1859:287–99.
25. García-Pérez P, Zhang L, Miras-Moreno B, Lozano-Milo E, Landin M, Lucini L, et al. The combination of untargeted metabolomics and machine learning predicts the biosynthesis of phenolic compounds in Bryophyllum medicinal plants (genus Kalanchoe). Plants (Basel). 2021;10(11):2430.
26. Odom JD, Sutton VR. Metabolomics in clinical practice: improving diagnosis and informing management. Clin Chem. 2021;67(12):1606–17.
27. Alaimo JT, Glinton KE, Liu N, Xiao J, Yang Y, Reid Sutton V, et al. Integrated analysis of metabolomic profiling and exome data supplements sequence variant interpretation, classification, and diagnosis. Genet Med. 2020;22(9):1560–6.

28. Shayota BJ, Donti TR, Xiao J, Gijavanekar C, Kennedy AD, Hubert L, et al. Untargeted metabolomics as an unbiased approach to the diagnosis of inborn errors of metabolism of the non-oxidative branch of the pentose phosphate pathway. Mol Genet Metab. 2020;131(1–2):147–54.
29. Heacock AM, Adams E. Formation and excretion of pyrrole-2-carboxylate in man. J Clin Invest. 1974;54(4):810–8.
30. Wajner M, Wannmacher CM, Purkiss P. High urinary excretion of N-(pyrrole-2-carboxyl) glycine in type II hyperprolinemia. Clin Genet. 1990;37(6):485–9.
31. Atwal PS, Donti TR, Cardon AL, Bacino CA, Sun Q, Emrick L, et al. Aromatic L-amino acid decarboxylase deficiency diagnosed by clinical metabolomic profiling of plasma. Mol Genet Metab. 2015;115(2–3):91–4.

Untargeted Metabolomics, Targeted Care: The Clinical Utilities of Bedside Metabolomics

Joshua Manor and Sarah H. Elsea

Abstract The second decade of the twenty-first century saw a quiet revolution in the field of inborn errors of metabolism. Decades of extensive research into metabolic pathways of physiologically active cells and tissues, along with an improved resolution of high-throughput screening capabilities, brought forth the clinical metabolome. Clinicians can now take a metabolic snapshot while assessing their patients and receive invaluable information on pathological processes, rule in or rule out a proposed diagnosis, highlight early signs of decompensation, assess response to treatment, explore new disease biomarkers, and even suggest novel treatment options. In this chapter, we review the major strengths of clinical metabolomics as a diagnostic aid and its capabilities in promoting novel biomarker discovery. We also provide an outlook for how next-gen interpretation modalities (such as machine learning) are expected to revolutionize this field further to benefit patients worldwide.

Keywords Clinical metabolomics · Inborn error of metabolism (IEM) · Mass spectrometry (MS) · Dried blood spot (DBS) · Newborn screening (NBS) · Biomarker discovery

J. Manor
Metabolic Unit, Haim Sheba Medical Center, Edmond and Lily Safra Children's Hospital, Tel Hashomer, Israel

S. H. Elsea (✉)
Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
e-mail: Sarah.Elsea@bcm.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_6

1 Introduction

In recent years, clinical practice has seen a dramatic increase in the utilization of targeted and untargeted metabolomics (TM and UM, respectively) [1, 2]. Metabolomics has been termed "the stethoscope of the twenty-first century" with broad applications in many fields of medicine: In oncology, it is utilized for biomarkers; in neurology, for severity stratification of neurodegenerative disorders; in endocrinology for diabetes modulators; in rheumatology and cardiology where certain metabolites may improve prognostication of chronic diseases (e.g., osteoarthritis and atherosclerosis); and in gastroenterology, where metabolomics may assist in differentiating between Crohn's disease and ulcerative colitis [3]. UM has been used in multifactorial disorders to assist in risk assessment and early diagnosis, such as the observation that increased plasma levels in three out of five aromatic and branched-chain amino acids (isoleucine, leucine, valine, tyrosine, and phenylalanine) confers a fivefold increased likelihood for developing type 2 diabetes [4] or that increased plasma levels of glycocholate, taurocholate, and glycochenodeoxycholate are associated with nonalcoholic fatty liver disease (NAFLD). In contrast, decreased plasma levels of free carnitine, butyrylcarnitine, methyl-butyryl carnitine, and cysteine-glutathione are seen in nonalcoholic steatohepatitis (NASH) [5]. Similarly, breast cancer showed a pattern of increased total choline-containing substances and decreased glycerophosphocholine in the plasma [6], correlating malignancy to a glycerophosphocholine-to-phosphocholine ratio switch [7].

Naturally, UM can highlight alterations in complex "metabolic disorders" and refine our understanding of metabolic flux in health and disease. The untargeted global assessment can further provide a high-resolution cellular homeostasis map and multivariant perturbation metrics that can assist in prognostication, the need for intervention, and the effect of a therapeutic modality. From a clinical perspective, a rare disease is a primary focus, and areas of influence for untargeted metabolomics include improving diagnostic rates, allowing affordable high-throughput disease screening, and identifying novel disease biomarkers that can be translated to the clinic. UM can contribute significantly toward a facilitated diagnosis and direct targeted treatment, thus enabling a reduction of the traditionally high morbidity and mortality associated with the under-recognition and undertreatment of metabolic disorders.

2 Metabolomics Joins the Diagnostic Front Seat

Clinicians justifiably consider exome sequencing (ES) to be the panacea of all diagnostic dilemmas, particularly when routine laboratory studies and tissue biopsies fail to establish a diagnosis. Indeed, ES has ushered in a new diagnostic era. Thanks to the reduced cost of sequencing, massive utilization of ES is now widely available, turning it into first-tier clinical testing for diagnostic evaluation of developmental

delay and other congenital anomalies [8–10]. ES is perhaps most heavily relied upon in the neonatal intensive care unit (ICU), in which a third of admitted patients are due to genetic causes [11]. ES is also heavily relied upon for primary mitochondrial disorders (PMD) [12], for which there is a notorious lack of biomarkers with adequate specificity [13]. This is further exemplified by the staggering finding of normal muscle respiratory chain enzymes in 10–20% of cases with mitochondrial myopathy undergoing invasive diagnostic procedures [14].

Nonetheless, as intriguing as it may appear, a diagnosis relying solely on ES without additional laboratory results lowers the test's pretest probability and increases the likelihood of false-positive results [15]. The American College of Medical Genetics and Genomics (ACMG) recommends that effort should be made to avoid using the pathogenicity of a variant as the *sole evidence* of a Mendelian disease but rather should be used in conjunction with other clinical information [16]. Perhaps, the best demonstration of the chasm between the diagnostic promise of ES and reality is variants of uncertain significance (VUS) [17], perceived as a "challenge" in ~1/3 of the articles reporting ES results [18]. With inconclusive molecular data, generating a snapshot of metabolites during illness creates a functional bioassay for candidate metabolic pathways. It provides an independent tool for ruling in or ruling out suspected diagnoses. Following the ACMG variant interpretation algorithm [16], metabolomic data fall into the functional data category by providing "well-established functional studies to show deleterious [or non-deleterious] effect," which is considered strong supportive information for the pathogenic or benign nature of a variant. Metabolomic data, therefore, can push the needle from the neutral zone (a VUS) to the actionable zone [19, 20]. It is, nevertheless, of utmost importance to report only significant variations in metabolites to avoid reporting fluctuations stemming merely from dietary changes, environmental factors, drug exposure, and normal daily changes in metabolism. However, the definition of significance is still laboratory dependent [21]. It is also of equal importance to provide detailed clinical information and the nutritional and therapeutic status of the patient to the performing laboratory to optimize the data analysis. Given that a VUS is a common finding in ES, estimated to occur in 30–80% of clinically indicated ES tests [22, 23], and comprises ~30% of variants found by targeted sequencing for suspected inborn errors of metabolism (IEMs) [24], a variant validation and classification tool alongside molecular diagnostics is critical. UM serves as a valuable biochemical, functional validation instrument that can be integrated into clinical care.

2.1 Monoamine Synthesis

The utilization of UM for VUS reclassification was demonstrated by Atwal et al. in a diagnostic odyssey of an 11-month-old boy presenting with intellectual disability and hypotonia with episodes of generalized stiffening [25]. Initially diagnosed with cerebral palsy, exome sequencing showed two missense VUS in *DCC*, which

encodes aromatic-L-amino acid decarboxylase (AADC), a key enzyme in monoamine synthesis. AADC deficiency (AADCD) results in a severe neurometabolic disorder due to the combined deficiency of serotonin, dopamine, norepinephrine, and epinephrine. The onset of the disease is typically in the first months of life. However, the phenotypic spectrum of disease is broad and includes hypotonia, oculogyric crises, ptosis, dystonia, hypokinesia, developmental delays, and autonomic dysfunction [26]. Diagnosis is typically made by abnormal monoamine metabolites in CSF, followed by molecular confirmation or enzyme analysis in plasma. Elevation of 3-O-methyldopa (3-OMD), also termed 3-methoxytyrosine (3-MT), a catabolite of dihydroxyphenylalanine (L-DOPA), indicating a blockage in the conversion of L-DOPA into dopamine. For the patient described in Atwal et al., UM showed a plasma level of 3-MT 69 times higher than the control population.

Further confirmation of the UM findings was achieved by demonstrating absolute levels of 3-MT in CSF > three-fold the upper limit of normal, along with undetectable levels of the dopamine and serotonin derivatives, homovanillic acid (HVA), and 5-hydroxy indoleacetic acid (5-HIAA), respectively. Neurotransmitter analysis in CSF has been considered the gold standard for diagnosing AADC deficiency. Yet, this work presents that UM can attain similar results in plasma without requiring an invasive procedure. At times of atypical presentation and a wide differential diagnosis, UM can cast a wide net in a single test and may render specific diagnostic tests and procedures, each for a single suspected disease, inessential. The presentation can be atypical for the late-onset subgroup of patients with AADCD, including milder symptoms of hypotonia, dystonia, and fatigue [27]. As symptoms of AADC uniformly present in the first years of life [28, 29], a differential diagnosis may include "dopa-responsive dystonia," with a trial of L-DOPA initiated as part of the diagnostic workup [30]. In light of this treatment approach [31], an additional work examined the ability of UM to differentiate an AADCD metabolic profile from other L-DOPA-treated conditions, as 3-MT will be elevated in both cases [32]. Two patients with AADCD before initiation of L-DOPA showed similar elevations of 3-MT as five patients with non-AADCD pathology (*Z*-scores of +5.88 and +7.65, respectively). However, reduced levels of dopamine 3-O-sulfate (D3OS) and vanillylmandelic acid (VMA) downstream of the AADC enzyme were seen in AADCD patients. In contrast, non-AADCD patients had elevated levels of D3OS and 3-methoxytyramine sulfate (3-MTS), showing a surplus of dopamine following treatment with L-DOPA. As these results were obtained from plasma and not from CSF, these data again demonstrate the robustness of the metabolic profile of AADCD obtained noninvasively.

2.2 Ornithine Metabolism

Diagnosis by variant reclassification was also made in a 7-year-old male with global developmental delay (GDD), ADHD, epilepsy, and ectodermal abnormalities [33]. The boy was found to have a novel, presumed splice-site VUS in *ODC1*. This gene

encodes ornithine decarboxylase, converting ornithine into putrescine, the first step of the spermidine pathway. UM showed excessive N-acetylputrescine, a metabolite of putrescine, confirming a gain-of-function variant in this gene, consistent with a diagnosis of Bachmann-Bupp syndrome (MIM 619075), an autosomal dominant disorder of neurodevelopment characterized by GDD, macrocephaly, white matter and callosal abnormalities, spasticity, seizures, and ectodermal defects (alopecia, cutaneous vascular malformations) [34, 35].

2.3 NAD(P)HX Repair System

VUS reclassification can also direct treatment for metabolic disorders amenable to the intervention. The NAD(P)HX repair system is a highly conserved two-enzyme system that restores damaged NAD(P)H [36]. In an acidic and hyperthermic environment, NADH and NADPH can both undergo hydration into NAD(P)HX nonenzymatically [37]. Without the repair system, the inactive NAD(P)HX accumulates and depletes the NAD^+ pool under cellular stress. Biallelic pathogenic variants in either of the genes coding for the two enzymes, *NAXD* and *NAXE*, cause a rapidly progressive neurometabolic disorder triggered by inflammatory stress (mostly febrile illness), bearing high mortality in the first decade of life (MIM 618321 and 617186, respectively) [38–40]. A case of a fever-triggered encephalomyopathy crisis in a 16-year-old adolescent demonstrated a small deletion encompassing the first two exons of *NAXD*, *in trans* to a missense VUS in the exon 1-intron 1 splice donor, suspected to alter both splicing and the mitochondrial localization of NAXD, which contains the mitochondrial targeting sequence in its first exon [41]. Plasma UM during metabolic crisis demonstrated NAD^+ depletion and led the medical team to initiate niacin therapy. Under therapy, clinical status improved, and baseline metabolomics demonstrated repletion of NAD^+. Not only did UM biochemically support the pathogenicity of the variant, but it also provided monitoring data for the targeted treatment. These results were further validated in a case of NAXE deficiency, for which evidence of depletion of NAD^+ derivatives was presented during a crisis. Niacin treatment appeared to prevent metabolic decompensation during a subsequent febrile illness while on niacin supplementation (Fig. 1).

2.4 Riboflavin Metabolism

Diagnosis and treatment guidelines were also provided in a case of a 2-year-old male diagnosed with hypoplastic macrocytic anemia (hemoglobin of 5.5 g/dL and mean corpuscular volume of 107.1 fL, range 10.5–14.0 g/dL and 76–90 fL, respectively) who later developed ataxia and nystagmus in the context of respiratory syncytial virus (RSV) infection [42]. Brain magnetic resonance imaging (MRI) showed enhancement of cranial nerves III and V, diffuse intramedullary T2 hyperintensity of

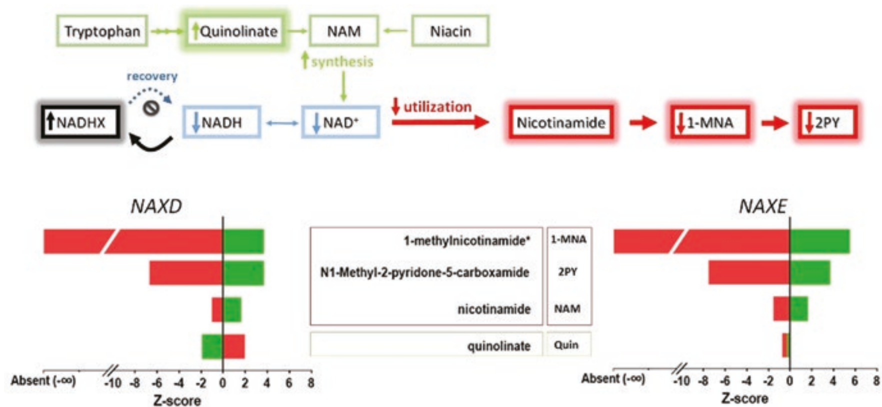

Fig. 1 NAD(P)HX repair system deficiency and plasma untargeted metabolomics. **Top:** NAD$^+$ utilization (red) and repletion by biosynthesis (green) pathways are shown. When the repair system is defective (left, black), correction of spontaneously converted NADH to NADHX back to NADH (blue) is impaired, lowering available NAD$^+$ and, thus, lowering utilization products while increasing upstream biosynthesis markers (quinolinate). **Bottom:** Plasma metabolomic profiles of deficiency in the NAD(P)HX repair system when patients are under inflammatory stress. Two patient profiles are shown, one for each repair system enzyme (NAXD, NAXE), with the effect of NAD$^+$ depletion on selected analytes (red bars) and after supplementation with niacin (green bars). Taken together, in both NAXD and NAXE deficiencies, these results point to NAD$^+$ depletion during inflammatory stress that is amenable to correction with substrate repletion. **Bottom: Left panel**—At the time of acute inflammatory stress, the red bars show the metabolomic profile from a patient with NAD(P)HX dehydratase deficiency (NAXD, EC 4.2.1.93), demonstrating the absence of 1-methylnicotinamide, marked deficiency in N1-methyl-2-pyridine-5-carboxamide, and increased quinolinate (red bars). At 11 months' post-crisis, the patient underwent another UM profiling (green bars), showing a reversal of these alterations while under niacin treatment. **Bottom: Right panel**—Shown are the same key molecules in plasma from a patient with NAD(P)HX epimerase deficiency (NAXE, EC 5.1.99.6) demonstrating similar trends during an acute inflammatory crisis (red bars). Repeat UM profiling at 9 months' post-crisis while on niacin supplementation showed a reversal of these alterations (green bars)

the entire cord, and cauda equina nerve root thickening and enhancement. ES showed a pathogenic variant and a novel VUS predicted to cause an in-frame single amino acid deletion (p.Phe153del) in *SLC52A2*, an intestinal basolateral and a blood-brain barrier riboflavin transporter (MIM 607882). UM showed elevated C6 (hexanoylcarnitine), C8 (octanoylcarnitine), C10 (decanoylcarnitine), and C10:1 (decenoylcarnitine), supporting the pathogenicity of the in-frame deletion variant, in addition to elevations in 2-hydroxyglutarate, methyl succinate, and ethylmalonate, common secondary alterations in short- and medium-chain fatty acid oxidation defects. These perturbations can also be detected on TM, yet the value of UM was nicely demonstrated by showing additional perturbations in metabolic pathways that are related to riboflavin deficiency. Riboflavin is the precursor of flavin adenine dinucleotide (FAD), a necessary cofactor for kynurenine-3-monooxygenase, the *de novo* NAD$^+$ biosynthesis pathway member; in the patient, the proximal kynurenine was increased, and the distal picolinate was decreased. Reduction of

5,10-methylene tetrahydrofolate into 5-methyltetrahydrofolate is also dependent on FAD, leading to inhibition of the one-carbon cycle and recycling of sulfur-containing amino acid, ultimately resulting in elevations in sarcosine, dimethylglycine, methionine, and glycine. These values normalized upon initiation of high-dose riboflavin (70 mg/kg/day), with a resolution of the anemia and nystagmus and improvement in ataxia. While ataxia and nystagmus are common findings, the highly suggestive upper body proximal muscle weakness with prominent neck flexion and the early symptom of dysphagia were not seen; however, macrocytic anemia is a rare finding [43], making a diagnosis based on clinical suspicion very difficult. Indeed, SLC52A2 deficiency, also called Brown-Vialetto-Van Laere syndrome 2 (MIM 614707), is a difficult diagnosis to make, with an average time to diagnosis of more than 2 years, a significant delay for a disorder manifesting in the first decade of life [43]. For example, normocytic anemia was misdiagnosed in a toddler as pure red cell aplasia overlooking an insidious weakness and areflexia and delaying the diagnosis of SLC52A2 deficiency for 3 years [44].

2.5 Histidine Metabolism

Not only for VUS can interpretation by UM assist in diagnostic dilemmas. Deficiency of urocanase (also called urocanate hydratase), an enzyme participating in the histidine deamination breakdown pathway, was previously considered to be associated with intellectual disability [45, 46]. Intellectual disability (ID) is a relatively common clinical finding, with a 0.8–3.7% prevalence in the pediatric population [47, 48]. Common disorders can manifest independently in rare diseases. A connection can be falsely established due to (a) small cohorts (an inherent problem of rare disorders), (b) consanguinity increasing independently the prevalence of IEM [49, 50] and ID [51, 52], and (c) publication bias, resulting from a top-down sequencing of patients presenting with a multitude of symptoms. In a recent work, plasma and urine UM showed a significant increase in both cis- and trans-urocanate and imidazole propionate in two asymptomatic patients with biallelic pathogenic variants in *UROC1* and normal intellect and with no other significant metabolic alterations [53] (Fig. 2). UM helped expand our understanding of the benign nature of urocanase deficiency (MIM 276880) and the need to pursue a genetic diagnosis for ID in patients with urocanase deficiency, per standard guidelines [10].

2.6 The Diagnostic Rate Among Inborn Error of Metabolism

A direct comparison of TM (plasma amino acids [PAA], acylcarnitine profile [ACP], and urine organic acids [UOA]) to UM showed that the latter has an overall ~sixfold higher diagnostic rate for IEM (7.1% vs. 1.3%) [19]. That result may not be unexpected, yet the power of UM was demonstrated by the variety of diagnoses it allows

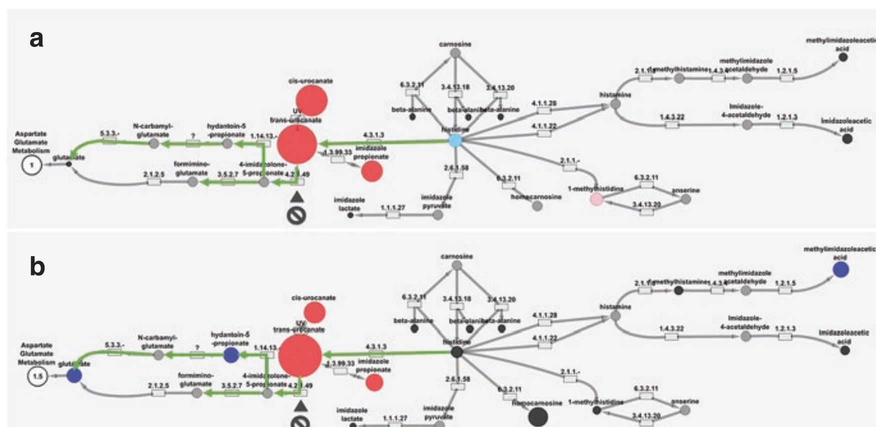

Fig. 2 Metabolomic map of the histidine catabolism pathway in a patient with urocanate hydratase (EC 4.2.1.49, urocanase) deficiency in (**a**) plasma and (**b**) urine. The affected arm of the pathway, highlighted in green arrows in both plasma and urine, shows the accumulation of analytes immediately upstream to the enzymatic block, trans-urocanate and cis-urocanate. These accumulations then result in the subsequent accumulation of imidazole propionate (in equilibrium with trans-urocanate) and, to a lesser degree, 1-methylhistidine. In urine, the deficiency of the downstream analytes, hydantoin-5-propionate and glutamate, are also shown (blue circles). The color, diameter, and shading of each circle are proportional to the *Z*-score. Red circles indicate analytes in excess (*Z*-score > +2); blue circles represent deficient analytes (*Z*-score < −2); pink circles indicate analyte excess with a *Z*-score > +1.5 < +2. Black circles indicate analytes within $-1.5 \leq Z$-score $\leq +1.5$. Gray circles indicate analytes that are not measured in this assay. (Adapted from Glinton et al., 2018. Urocanate hydratase (urocanase) is indicated by the black "no entry" sign)

over TM. Diagnoses included disorders of synthesis of neurotransmitters, cholesterol, and peroxisomal biogenesis. γ-aminobutyric acid-(GABA)-transaminase deficiency (MIM 613163). This early infantile epileptic encephalopathy with a movement disorder and hypersomnolence due to GABA catabolism defect, was diagnosed in a 1-year-old patient with hypotonia and movement disorder and compound heterozygous VUS in *ABAT*, which encodes GABA-transaminase. UM showed 2-pyrrolidinone, succinamic acid, and succinimide elevations that were not seen on UOA. These three metabolites are reliably detected in both CSF and plasma, in contrast to CSF GABA alone, the gold standard for diagnosis but prone to false-negative results if not handled appropriately [27]. Specifically, 2-pyrrolidinone is a butyrolactam spontaneously converted from GABA when the latter is not broken down to succinic semialdehyde (SSA) by GABA-transaminase. 2-pyrrolidinone is converted to succinimide and succinamic acid [54], making these three metabolites alternative biomarkers. The same metabolic profile was seen in a 6-year-old with global developmental delay (GDD) and movement disorder who was also homozygous for a VUS and two other patients presenting with encephalopathy, seizures, cortical blindness, motor developmental delay, hypotonia, strabismus, ataxia, and intellectual disability [19, 54].

A diagnosis of Smith-Lemli-Opitz (MIM 270400), a disorder of cholesterol biosynthesis, was made in a typically presenting 10-year-old male with microcephaly, hypotonia, developmental delays, and a congenital cardiac defect. In that patient, 7-dehydrocholesterol was elevated, while cholesterol was decreased. Initially, a diagnosis could be ascertained by comparing this profile to the near identical profile of a molecularly confirmed 3-year-old boy [19] and later confirmed molecularly. In a different study, such a diagnosis was excluded in a 28-year-old patient with developmental delays by demonstrating normal 8-dehydrocholesterol and 7-dehydrocholesterol, downgrading a VUS to likely benign [55].

The latter study presented 16 cases of either homozygous or compound heterozygous VUS in which metabolomics assisted in variant reclassification after ES was nondiagnostic. Noteworthy in this study are (1) a case of a patient presenting with early-onset recurrent nephrolithiasis, urosepsis, and transfusion-dependent anemia, showing an excess of orotic acid in the urine (a metabolite not easily detected in UOA); thus, the two missense VUS were reclassified to pathogenic and likely pathogenic, and treatment with uridine monophosphate was initiated; and (2) a case of psychomotor retardation and retinitis pigmentosa with increased urinary 5-oxoproline excretion, and despite negative ES, the finding that prompted targeted sequencing of the *GSS* gene, demonstrating homozygosity for a deep intronic variant for this highly heterogeneous condition, 5-oxoprolinase deficiency (MIM 260005).

UM performed on plasma from a 17-year-old with agenesis of the corpus callosum, autism spectrum disorder, lactic acidosis, hyperammonemia, and electrolyte abnormalities revealed reduced pantothenate, carnitine, and carnitine derivatives due to *SLC5A6* deficiency, confirmed by biallelic VUS identified by subsequent exome sequencing [19]. SLC5A6 is a cellular cotransporter of pantothenate, biotin, and α-lipoic acid in the intestine and blood-brain barrier [56]. SLC5A6 deficiency leads to early-onset neurodegenerative disorder and also includes failure to thrive, acquired microcephaly, movement disorder, immunodeficiency, gastrointestinal dysfunction, and osteopenia. Early treatment with high-dose pantothenate, biotin, and α-lipoic acid seems to improve outcomes [57, 58]. UM also confirmed the diagnosis of medium-chain acyl-CoA dehydrogenase deficiency, argininemia, and X-linked glycerol kinase deficiency (GKD, MIM 307030), in which molecular diagnoses were inconclusive or missing, among other conditions [19].

To directly analyze the contribution of UM to the interpretation of variants identified in ES, Alaimo et al. examined clinical samples with both ES and UM testing that were sent for diagnostic purposes in a cohort of 170 patients [20]. This cohort was similarly primarily pediatric, with>90% presenting with neurological symptoms. The researchers identified 145 variants in 74 patients in their cohort in 73 genes associated with an IEM. Based on the metabolomic data, the 12.3% diagnostic rate in this cohort facilitated the reclassification of 27 variants (19%). Of the reclassified variants, 24 VUS were reclassified as either likely pathogenic (n = 15) or likely benign (n = 9), while an additional three variants were upgraded from likely pathogenic to pathogenic. Classifications were done according to the ACMG guidelines [16]. For 17 additional pathogenic variants, the study presented

confirmatory metabolic perturbations. One case upgraded a homozygous VUS to a likely pathogenic variant, making the diagnosis of guanidinoacetate methyltransferase (GAMT) deficiency likely in the patient and allowing the clinician to focus on the treatment for this disorder of cerebral creatine deficiency. Moreover, for autosomal recessive conditions in which sequencing revealed only heterozygous variants, UM ruled out the suspected condition in ~60% of cases by showing a normal metabolite profile for the metabolic pathway in question, thus excluding the possibility of a biochemically symptomatic carrier possessing a second disease-causing allele not detected due to the constraints of ES.

2.7 Automation of Data Interpretation

Similarly to ES diagnostic capabilities, no discussion of UM clinical diagnostic capabilities will be complete without discussing the interpretation of the (untargeted) data. The ability to correctly diagnose a condition based on UM heavily depends on the art of data interpretation. Rather than the convoluted manual analysis of metabolomic data (with or without genomic information), automated bioinformatic tools provide means for pattern recognition and thus hold a great promise of an improved matching between datasets and a particular disease or pathway. One approach applies the clique problem in a metabolomic graph. A clique is defined as an interconnected group of metabolites (nodes), and small highly connected cliques are extracted based on computational analysis bounding cliques' p values [59]. Across 539 plasma samples, this connect-the-dots (CTD) approach reproduced accurate diagnosis of 16 different IEMs [60]. The top-down bioinformatic approach seeks to identify causative genes by providing a likelihood scoring system based on heuristic algorithms predicting the effects they expect to exert on the -*omic* dataset. Several metrics achieved prioritization of candidate genes based on UM. In *cross-omics*, a metabolomic study performed using dried blood spots (DBS) [61], candidate genes were prioritized based on their distance from the perturbed metabolite (where each reaction accounts for 1 step), with a narrowing process of including metabolites that participate in only limited amount of reactions ("uniqueness") and limiting metabolites to only those with a significant Z-score ("significance") to generate gene-specific metabolite sets. Successful prioritization was achieved by considering metabolites up to four reactions away from the primary reaction, uniqueness of up to 15 reactions, and significance of $>+3$ or <-3. In *metPropagate* [62], each protein is assigned a rank based on its associated metabolite enrichment, and dynamic ranking is propagated with the protein's functional linkage network. This approach successfully prioritized causative genes (within the top 20th centile) in 20% of IEMs in the study group and 82% (9/11) in the test group of neurometabolic disorders. This algorithm outperformed Exomiser [63], a causative variant

prioritization tool based on human phenotype ontology data and standard pathogenicity prediction (variant prevalence, conservation, and information from model organisms and inheritance patterns). In another approach, Reafect [64] assigned a score based on cumulative pathway perturbations, including metabolites with sub-significant *Z*-scores, correctly predicting the causative gene in the top fifth centile in 80% of the 76 patients harboring 36 different IEMs. When combined with the deleteriousness predictive score, Combined Annotation Dependent Depletion (CADD) [65], specificity further increased. Another promising approach is the implementation of a siamese neural network, weighing in the tandem computational metabolic network (for predicting metabolite flux) and machine learning (ML, for matching metabolic networks to diseases) trained by the ML algorithm [66]. The model used a single simulated profile for each disease with real data points from only 2% of the diseases to outperform a generic algorithm that prioritizes causative genes based on distance from real data profiles (using the L1 Manhattan metrics). Limitations to ML include the quality of the training set and the generalizability of metabolic profiles to novel diseases. Nonetheless, implementing predictive ML-based algorithms promises to reduce the number of patient-derived samples required for disease discovery (smaller cohorts), a vital prerequisite in the world of rare diseases such as IEMs [59].

3 Metabolomic Fingerprinting: Identifying Diseases' Biometrics and Finding New Disease Biomarkers

Biomarkers are crucial for the effective screening, designation, and diagnostic confirmation or exclusion of diseases by UM. Biomarkers are also critical in monitoring an affected patient's metabolic status and treatment efficacy. Clinical biomarkers for IEMs are abundant [67], although most are derived from TM. Many biomarkers are assessed by routine biochemical tests, such as ammonia, lactate, uric acid, and cholesterol, or from TM, such as UOA (e.g., trimethylamine in fish odor syndrome) or PAA (high phenylalanine in PKU). To increase specificity, either identification of disease-specific biomarkers can assist in the diagnosis (or ruling out) of a disease, such as allo-isoleucine in MSUD or argininosuccinic acid in argininosuccinate lyase (ASL) deficiency, or the identification of several non-pathognomonic metabolites, such as low levels of lysine, ornithine, and arginine in plasma amino acids suggestive of LPI, or the elevation of propionylcarnitine on ACP, methylmalonic acid in UOA, low methionine on PAA, and homocysteine in the blood, indicative of an intracellular cobalamin utilization defect. UM can instigate both strategies. By its untargeted nature, UM has the potential to uncover biomarkers that were not known to be associated with an IEM or not visible by targeted testing.

3.1 Peroxisome Biogenesis

Peroxisome biogenesis disorders in the Zellweger spectrum (PBD-ZSD) are a heterogeneous group of genetic disorders caused by mutations in genes responsible for normal peroxisome assembly and functions [68]. The majority of cases are due to biallelic pathogenic variants in *PEX1* (61%), followed by *PEX6* (15%) and *PEX12* (8%), with additional contribution from at least another 10 *PEX* genes [69], an important group of proteins essential for the assembly of peroxisomes and the recognition and transport of cytoplasmic proteins that contain peroxisomal targeting sequence [70]. The severe end of PBD-ZSD includes neuronal migration defects with leukodystrophy, neonatal onset seizures, hypotonia, failure to thrive, liver dysfunction, bony stippling respiratory insufficiency, cataracts, sensorineural hearing loss, and renal cortical microcysts; the mild end of the spectrum includes developmental delays and intellectual disability that can be mild and slowly progressive retinopathy and sensorineural hearing loss [68, 69, 71]. Severe cases can be screened by elevation of the very-long-chain fatty acid (VLCFA), C26:0-lysophosphatidylcholine (DBS, plasma), phytanic and pristanic acids (plasma); reduction in plasmalogens (plasma, erythrocyte membranes); increase in pipecolic acid (plasma, urine); and increase in the bile acids, dihydroxycholestanoate, and trihydroxycholestanoate (plasma, urine) [69]. Intermediate and mild cases may show only subtle alterations, and in combination with subtle clinical symptoms, screening for PBD-ZSD may not be performed. In a cohort of 19 mild to intermediate PBD-ZSD pediatric patients, Wangler et al. showed a distinct PBD-ZSD plasma metabolome (Fig. 3), with elevated levels of long-chain dicarboxylic acids, pipecolic acid, and the bile acid derivative 7α-hydroxy-3-oxo-4-cholestenoic acid and with reductions in phosphatidylcholines, phosphatidylethanolamines, and plasmalogens [71]. Pipecolic acid and the lysophospholipid 1-lignoceroyl-GPC (24:0) were most strikingly elevated with a *Z*-score of +3.7. Less anticipated changes included dicarboxylic acids of 16–22 carbons. UM also proposed novel biomarkers, with a reduction in nine sphingomyelin species. Reduction of several sphingomyelins in the clinical UM database, excluding PBD-ZSD, was observed in only 2% (interestingly, one of which was diagnosed with bifunctional protein deficiency, which can mimic PBD-ZSD). Moreover, the plasma elevations in pipecolic acid and reductions in the sphingomyelins attenuated with age, an expected finding correlating milder phenotypes among older surviving patients.

3.2 Urea Cycle

The urea cycle is the principal mechanism for the clearance of waste nitrogen resulting from protein turnover, the sole source of endogenous production of arginine, ornithine, and citrulline, and a principal component of the nitric oxide (NO) production pathway [72]. The urea cycle is also connected with TCA anaplerosis

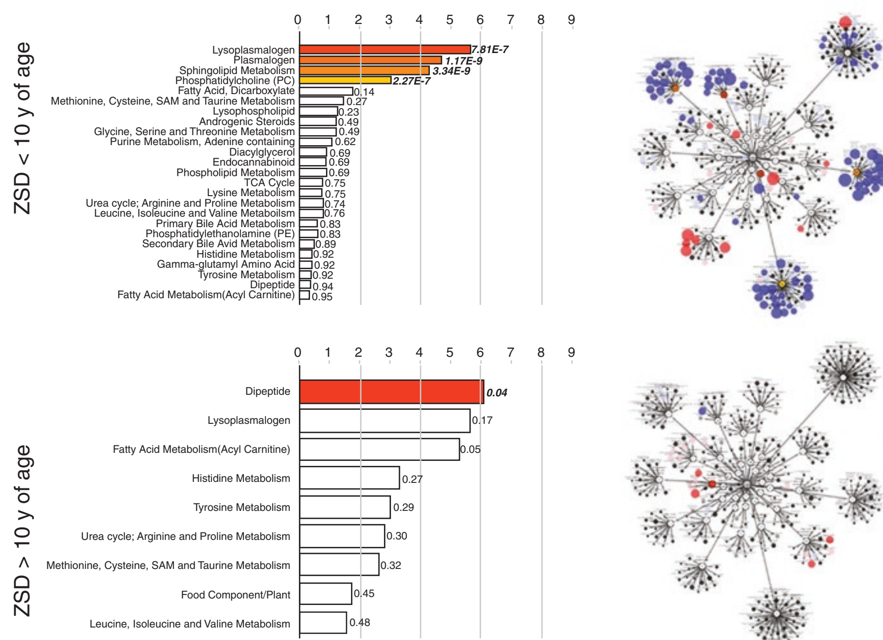

Fig. 3 Plasma metabolomic footprint of peroxisome biogenesis disorder-Zellweger spectrum disorder (PBD-ZSD). **Top: Left panel**. Early diagnosis (age <10 years) of PBD-ZSD shows a more significant metabolic phenotype, with enrichment of lysoplasmalogens, plasmalogens, sphingolipid metabolism, and phosphatidylcholines, reaching statistical significance (colored bars) for patients when compared to control population. **Top: Right panel.** The relative deficiencies of most of these lipids in PBD-ZSD in children under the age of 10 years are represented by the blue circles in the metabolomic tree of lipid analytes (lipidomics). **Bottom: Left panel**. Later diagnosis of PBD-ZSD (age >10 years) exhibits a more subtle metabolic phenotype. **Bottom: Right panel**. The metabolomic fingerprint of PBD-ZSD in individuals >10 years of age showed mild perturbations in which reductions of the plasmalogens and lysoplasmalogens did not reach statistical significance. The lipidomic tree shows considerably fewer perturbations when compared to plasma obtained from a younger patient. Trees were rendered using Cytoscape (https://www.cytoscape.org). (Figure adapted from Wangler et al. 2018)

via the alternative synthesis of fumarate. UCDs include eight different IEMs resulting from defects in any one of the six enzymes or two transporters involved in the hepatic removal of ammonia as waste nitrogen by its conversion to urea and excretion by the kidneys [73]. Mortality and morbidity primarily contribute to neurological damage resulting from hyperammonemia (HA) and the elevation of other neurotoxic intermediates of metabolism [74]. A typical presentation of a UCD includes neonatal hyperammonemic crisis; however, nontypical presentations of later-onset HA, acute liver dysfunction, intellectual disability, or insidious pyramidal signs of the lower extremities with minimal HA crises (in arginase deficiency) have been described [27]. It has been suggested that arginase deficiency, with its unique presentation compared to other UCDs, exerts a neurotoxic

effect by the guanylation of glycine, with the excess arginine serving as a guanidine donor rather than acute HA crises [75]. In a cohort of 13 patients with argininemia, Burrage et al. demonstrated an increase in additional guanidine compounds, namely, N-acetylarginine, homoarginine, argininate, and 2-oxoarginine, to a greater degree than guanidinoacetate (GA) [76]. Guanidine compound (GC) toxicity is believed to be derived from the observed in vitro neurotoxicity: diminished response to the inhibitory GABA and glycine neurotransmitters [77], promotion of non-apoptotic cell death and axonal hypersprouting [78], and inhibition of Na^+/K^+-ATPase activity and glutamate uptake, and decrease in antioxidant defense in the rat brain [79]. Other toxic effects of GC include ethanol-induced liver injury, stimulated osteoclastogenesis, generation of reactive oxygen species (ROS), and modulation of cerebral cortex potentials [79]. In GAMT deficiency, GA is the main GC accumulating and phenotypically exerting a greater degree of intellectual disability, refractory epilepsy, and dystonia. At the same time, pyramidal signs are less dominant as compared to arginase deficiency. While these results may indicate an important role of GA in the developing brain, another important modifier between the two disorders is the creatine level, which is normal in arginase deficiency and low in GAMT deficiency [80]. Peripheral administration of polyethylene glycol amalgamated to (PEGylated) arginase, now in advanced stages of development, will help elucidate further pathomechanistic insights by the peripheral reduction of arginine excess without restoring urea cycle function in the liver.

The same study by Burrage and colleagues also examined a UM profile of OTC deficiency, a severe X-linked disorder bearing high morbidity and mortality among males and affected females. In this UCD, morbidity is attributed to HA crises, and biomarkers are scarce, making this disorder “unscreenable.” In this study, 83% of patients (10/12) with a history of hyperammonemia (excluding females with no such history) showed a significant elevation (*Z*-score >+2) of either orotate or uridine. Other biomarkers of increased pyrimidine metabolism (due to the shunting of the cabamoylphosphate from the dysfunctional urea cycle), β-ureidopropionate, and uracil were not significantly elevated. For screening purposes, the sensitivity of these markers is ~60% (10/17), and the specificity is also low, given shared pathway perturbation with other UCD and pyrimidine metabolism defects; however, these results point to an important consideration for biochemical testing, indicating such abnormalities in cases of uncertain diagnosis given the sensitive nature of this disorder. The authors are aware of a case of a 1.5-year-old female presenting with fulminant liver failure and nonspecific liver biopsy for which only uracil was elevated in urine on traditional metabolic screening. Based on that result, a presumptive diagnosis of OTC deficiency was made, and targeted treatment was promptly provided; molecular testing later confirmed the suspected diagnosis. Both citrullinemia and ASL deficiency did not show unique metabolic fingerprints beyond the accumulation of citrulline and argininosuccinic acid, respectively [76].

3.3 Pyruvate Kinase

Pyruvate kinase deficiency (PKD, MIM 266200) is the most common form of inherited anemia due to glycolytic defects. It results in a spectrum of hemolytic anemia that can result in infantile-onset transfusion-dependent anemia or a milder form of compensated anemia [81]. A diagnostic gap exists for PKD due to the unsatisfactory performance of activity assays, a genetic composition complicating molecular diagnosis (variants in regulatory elements or effector genes such as *KLF1*), and the confounding effect of frequent blood transfusion on these methods [82]. In a cohort of 16 patients with PKD (against 32 controls), van Dooijeweert et al. showed three groups of metabolites in DBS that differed between the two cohorts: glycolytic products phosphoenolpyruvate and 2- and 3-phosphoglycerate, as one might expect, along with polyamines, such as spermine, spermidine, N^1-acetylspermidine, and putrescine, which are associated with red blood cell (RBC) membrane integrity, and acylcarnitines such as methylmalonylcarnitine and propionylcarnitine which are involved in turnover and repair of the RBC membrane [83]. Principal core analysis showed a separation of metabolic profiles between the two groups. Interestingly, the mildly affected patients with no history of transfusion dependence or splenectomy more closely resembled the control group, followed by transfusion dependence (severe phenotype), probably due to the frequent retrieval of donor RBC, and the splenectomized patients (moderate phenotype) were furthest away from control. Based on these findings, a machine learning algorithm trained on a subset of the cohort could predict the disease in 94% of cases.

3.4 Glucose Transporter 1

UM can also reveal a metabolic fingerprinting of disease-modifying treatments, aiding the monitoring of both therapeutic efficacy and disease progression. The brain glucose transporter GLUT1 facilitates the diffusion of glucose to brain tissue to compensate for insufficient passive diffusion, given the elevated requirement for glucogenic energy production by that tissue. Heterozygous pathogenic variants in *SLC2A1*, which encodes GLUT1, result in a brain energy failure syndrome (GLUT1DS) caused by impaired glucose transport and result in a spectrum of phenotypes including epileptic encephalopathy, intellectual disability, acquired microcephaly, ataxia, action limb dystonia, chorea, and tremor in its severe form and paroxysmal ataxia in its mild end. Diagnosis is made by demonstrating a low CSF concentration of glucose in the presence of plasma normoglycemia and/or identification of a pathogenic variant in *SLC2A1* [84]. Aside from probable energy depletion, it was suggested that depletion of glycolysis intermediates might play a role in the disorder's pathogenesis; however, it is unknown to what extent, as in vitro

models include near-complete abrogation of the transporter [85]. The sole therapeutic option available is a classic ketogenic diet (CKD). By altering the brain energy fuel into mainly ketone bodies, this therapy dramatically affects seizures and can improve cognitive outcomes. However, diet discontinuation was reported in up to 10% of patients due to side effects [86], and low compliance led experts to recommend alternative low carbohydrate diets (such as the modified Atkins diet) for adolescents and adults [87] and even exploring amylopectin-based diet of the low glycemic index [88]. Monitoring CKD is based on serum β-hydroxybutyrate levels, as measuring urine ketones—a convenient and qualitative measure of acetoacetate—has a limited role in diet monitoring [87]. A more comprehensive picture created by UM performed on six treated patients with GLUT1DS [89] showed elevations in β-hydroxybutyrate, β-hydroxybutyrylcarnitine, β-methyladipate, and N-acetylglycine. Other elevated derivatives were α-ketobutyrate, β-hydroxylaurate, 10-nonadecenoate, margarate, 15-methylpalmitate, and α-aminoheptanoate. Furthermore, pathway analysis using Kruskal-Wallis analysis (comparing pathway metabolite perturbations vs. non-pathway metabolite perturbations) showed involvement of long-chain fatty acids, phospholipids, acylcarnitines, and, to a lesser degree, sphingolipids, mono-hydroxy fatty acids, and polyunsaturated fatty acids. In accordance with fatty acid utilization, free carnitine levels were low, while carnitine-bound metabolites were elevated. CSF UM of three patients prior to CKD onset revealed low levels of glycerol 3-phosphate, an intermediary metabolite in lipid metabolism, and an increased level of isocitrate, which can indicate a TCA dysfunction [89]. These results supply a CKD metabolic profile with a more complete ketosis map, which can assist in diet fine-tuning, e.g., increasing the fat-to-carbohydrate ratio to increase overall ketosis or examining the effect of decreasing the ratio in reduced diet tolerability. The results can also guide the need for supplemental carnitine due to increased secondary excretion in ketosis, which, although it is considered a benign supplementary agent, can also exacerbate GI-related symptoms in patients under this GI-unfriendly diet and can also independently elevate plasma trimethylamine N-oxide (TMAO) levels, an atherosclerotic agent [90, 91].

3.5 Serine Metabolism

Serine biosynthesis defects occur due to deficiency in either of the three enzymes converting 3-phosphoglycerate into serine, phosphoglycerate dehydrogenase (PGDH), phosphoserine aminotransferase (PSAT), and phosphoserine phosphatase (PSP), and result in a severe neurometabolic disorder including severe intellectual disability, ataxia, nystagmus, epilepsy, hypertonia/spasticity, microcephaly, and poor growth. Deficiencies in enzymes of the serine de novo biosynthesis pathway result in low plasma and CSF levels of serine. Low serine impedes the synthesis of sphingolipids (from serine and palmitoyl-CoA), phosphatidylserine, D-serine (an agonist to the ionotropic glutamate receptor N-methyl-D-aspartate (NMDA)), and

5,10-methylenetetrahydrofolate (by converting serine to glycine) [92]. This multi-pathway effect could explain neurodevelopmental abnormalities associated with this disorder; however, a more direct association has not been clinically demonstrated. In four children with two serine biosynthesis defects, PGDH deficiency (three children) and PSAT deficiency (one child), Glinton et al. demonstrated an abnormal phospholipid profile at the time of diagnosis and, upon treatment, normalization thereof. Before initiation of treatment, serine and glycine were both decreased in all patients, and multiple phospholipids were reduced, including species of the mono- and di-unsaturated 18-carbon phosphatidylcholine and phosphatidylethanolamine in three or four of the patients [93]. Multiple sphingolipid species were reduced in two patients, most notably sphingomyelin. Except for phospholipids, no other compounds were found to be altered (either reduced or elevated) consistently in all patients evaluated [93]. The majority of these abnormalities normalized under supplementation. These results emphasize the need for early treatment. De novo synthesis of phosphatidylcholine (PC) and phosphatidylethanolamine (PE) from choline, ATP, cytidine triphosphate (CTP), and diacylglycerol by choline kinase, phosphocholine cytidylyltransferase, and choline transferase (also referred to as the Kennedy pathway) may, in fact, be inadequate at the time of rapid growth or neuronal differentiation in utero and require supplementation of PE and PC from phosphatidylserine pool (by mitochondrial phosphatidylserine decarboxylase). Another possibility is serine palmitoyltransferase (SPT) substrate promiscuity, leading to the condensation of alanine and palmitoyl-CoA into 1-deoxy-sphinganine (instead of 1-dehydro-sphinganine when serine is sufficient). Indeed, decreased sphingomyelin and increased 1-deoxy-sphingomyelin have been reported in targeted metabolomics applied for primary serine biosynthesis defects and secondary serine deficiency due to mitochondrial disorders [94]. Interestingly, a DBS sample from one patient at 38 hours of life showed markedly reduced serine, which could have served as abnormal NBS, as discussed above [93].

3.6 Pentose Phosphate Pathway and Polyol Metabolism

The pentose phosphate pathway is an important cytosolic pathway that converts glucose into ribulose-5-phosphate and produces reduced nicotinamide adenine dinucleotide phosphate (NADPH) for restoration of the antioxidant glutathione. Its oxidation product, ribulose-5-phosphate, can serve as a nucleotide building block or be further converted into phosphorylated mono-sugars that can serve as glycolytic intermediates [95]. The nonoxidative portion includes four enzymes, ribulose 5-phosphate isomerase and reductase, transaldolase (TALDO), and transketolase (TKT). Common features of TALDO deficiency (MIM 606003) are hepatosplenomegaly, anemia, thrombocytopenia, renal tubulopathy, heart abnormalities, and cholestatic liver dysfunction that can develop into cirrhosis [96]. For TKT deficiency (MIM 617044), manifestations include short stature, developmental delays, and congenital heart defects [27]. These manifestations are nonspecific; further,

considering the potential reversibility of these two disorders (despite no currently available treatment), high-yield screening can facilitate diagnosis and improve outcomes. While TALDO deficiency was indicated by UM screening by identification of sedoheptulose, a single polyol that can point out an abnormality within the nonoxidative portion of PPP, Shayota et al. successfully demonstrated high-specificity multiple polyol alterations in plasma and urine metabolomics for the diagnosis of these two enzyme deficiencies (see Table 1) [97]. Two novel biomarkers were shown for both disorders (erythronate and ribonate). Secondary alterations in purine and pyrimidine metabolites, as seen in this work, are most probably due to defect in ribulose-5-phosphate processing.

Table 1 TM versus UM findings in two disorders of the nonoxidative portion of the pentose phosphate pathway [97]

	Metabolic pathway	Targeted metabolomics		Untargeted metabolomics	
		Plasma	Urine	Plasma	Urine
TALDO deficiency	Polyol	Arabitol Erythritol Ribitol Xylitol Sedoheptulose	Arabitol Sedoheptulose	Arabitol/xylitol[a] Ribitol Erythritol Sedoheptulose Erythronate Ribonate	Arabitol/xylitol[a] Ribitol Erythronate Ribonate
	Tryptophan	–	–	Kynurenate Xanthurenate	Quinolinate Xanthurenate
	Purine	–	–		Xanthosine
	Pyrimidine				N-carbamoylaspartate
	TCA	–	–	α-Ketoglutarate	Succinate Fumarate Malate
TKT deficiency	Polyol	Arabitol Erythritol Ribitol Xylitol Sedoheptulose	None	Arabitol/xylitol Ribitol Erythritol Ribose Erythronate Ribonate	Arabitol/xylitol Ribitol Erythritol Erythronate Ribonate
	Tryptophan	–	–	Kynurenine Xanthurenate Quinolinate Indolelactate	Kynurenine 3-hydroxykynurenine
	Purine	–	–	Inosine	Xanthosine Guanosine
	Pyrimidine				N-carbamoylaspartate

Underline—a newly discovered disease biomarker
[a]Arabitol could not be easily distinguished from xylitol or arabonate from xylonate by this UM platform, given the similar structure and molecular weight

3.7 Metabolomics of Muscular Diseases

The muscular disease has a broad differential diagnosis and multiple pathological mechanisms. An important etiological group is the mitochondrial myopathies, with can have early- or late-onset and acute or subacute course and are progressive in nature [14]. Buzkova et al. compared the metabolomic profiles of mitochondrial myopathies and ataxias (lumping the sporadic inflammatory disorder inclusion body myositis, which also affects mitochondrial functioning) in comparison to non-mitochondrial neuromuscular diseases [98]. This study utilized TM of 94 metabolites but is included herein, given its broad application and interest in disease fingerprinting. The first group demonstrated alterations in the transsulfuration pathway, including elevated cystathionine (1.9–4.1-fold increase) and a less consistent reduction in taurine. In contrast, the second group was characterized by normal levels of cystathionine, depletion of nicotinamide (−1.7-fold change), and increased creatine (2.1-fold change). Alterations were also found in carbohydrate metabolism leading the authors to propose a quad-biomarker set of elevations in sorbitol, alanine, myoinositol, and cystathionine, producing an area under the curve similar to fibroblast growth factor-21 (FGF-21), lactate, and pyruvate, to distinguish the mitochondrial origin of myopathy (with an overall 76% sensitivity, 95% specificity). A particular mitochondrial myopathy, mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS, MIM 540000) showed elevations in carbohydrate derivatives (sorbitol, glucuronate, myoinositol, and sucrose), decreased arginine, and an increase in transsulfuration intermediates (cystathionine, γ-glutamyl-cysteine S-adenosylmethionine, and glutamate) with a decrease in adenosine, guanidinoacetate, and betaine. In another UM study of MELAS patients, Sharma et al. demonstrated novel amino acid, acylcarnitine, and fatty acid biomarkers [99]: N-lactoyl attached to the branched-chain amino acids leucine, isoleucine, or valine or to the aromatic amino acids phenylalanine and tyrosine; β-hydroxy acylcarnitines of even-length C10:0 to C16:0 (C10:0, C12:0, C14:0, C16:0); and β-hydroxy fatty acids of even-length C8:0 to C14:0. These biomarkers were associated with the degree of diseased severity (per Karnofsky performance score). Interestingly, β-OH-C16:0 carnitine showed a severity correlation similar to the well-accepted growth/differentiation factor-15 (GDF-15) [100], while β-hydroxy acylcarnitines and β-hydroxy fatty acids correlated with ventricular lactate levels, and the N-lactoyl-amino acids correlated with urine heteroplasmy. Table 2 summarizes selected metabolic pathways in which more than 10% of metabolites are significantly altered in the individual disorders (two-sample *t*-test for significance).

Table 2 Plasma metabolic pathways in which ≥10% of metabolites are altered compared to control in selected mitochondrial myopathies and ataxias, compared to non-mitochondrial myopathies [98]

Group	Family	Disorder	OMIM	Gene	Altered metabolic pathways
Mitochondrial	mtDNA depletion	IOSCA	271245	*TWNK*	Primary bile acids; nitrogen; alanine, aspartate, and glutamine; cysteine and methionine; histidine; butanoate; D-glutamine and D-glutamate; cyanoamino acids; glutathione
		MIRAS	607459	*POLG*	Cysteine and methionine; alanine aspartate and glutamine; taurine and hypotaurine; histidine; β-alanine; nitrogen; aminoacyl-tRNA; cyanoamino acids; pantothenate and CoA
	Single mtDNA deletion	PEO	530000 258450 157640 609286	Sporadic mtDNA deletion *TWNK* *POLG*	**Plasma:** Alanine, aspartate, and glutamine; cysteine and methionine; nitrogen; glycine serine and threonine; D-glutamine and D-glutamate; arginine and proline; aminoacyl-tRNA
					Muscle tissue (invasive): Cyanoamino acids; branched-chain amino acids; nitrogen; pantothenate and CoA; alanine aspartate and glutamine; cysteine and methionine; aminoacyl-tRNA; histidine
	Mitochondrial tRNA defects	MELAS[a]	540000	m.3243A>G	D-arginine and D-ornithine; taurine and hypotaurine; cysteine and methionine; alanine, aspartate, and glutamine; arginine and proline; aminoacyl-tRNA; glutathione
	Immune dysregulation	Inclusion body myositis	147421	Sporadic	Nitrogen; glycine serine and threonine; cysteine and methionine; pyrimidine; arginine and proline; glutathione; primary bile acid; histidine
Non-mitochondrial	Neuromuscular diseases			AD or AR *DMPK* *ZNF9* *CAPN3* *PABPN1* *SMN1* *TIA1* Or unknown	Arginine and proline; folate (one-carbon pool); glutathione; glycine, serine, and threonine; histidine; butanoate; β-alanine

Pathways are presented in decreasing order of alteration

mtDNA mitochondrial DNA, *IOSCA* infantile-onset spinocerebellar ataxia, *MIRAS* mitochondrial recessive ataxia syndrome, *MELAS* mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes

[a]**Sharma et al.**

4 Future Perspectives

We shall end our discussion of bedside UM with a futuristic perspective of UM in the service of autism spectrum disorder (ASD). The autism spectrum is a range of neurodevelopmental conditions exhibiting persistent deficits in reciprocal social interaction and restricted, repetitive patterns of behavior, interests, or activities. Historically, ASD gained an independent psychiatric status from schizophrenia only in the second half of the twentieth century (the word autism was initially coined to describe severe schizophrenia) [101] and became a spectrum only at the end of that century. We now appreciate a 60–80% heritability of ASD [102, 103], with an estimated genetic or genomic etiology in 30–40% of cases [104]. Within the group of ASD, due to a known genetic variant, an estimated 3–5% are due to an IEM [105]. Examples include [106] amino acidopathies (PKU, homocystinuria, S-adenosylhomocysteine hydrolase deficiency, MSUD, UCD), organic acidurias (PA, L-2-hydroxyglutaric aciduria), cholesterol biosynthesis defects (Smith-Lemli-Opitz syndrome), disorders of neurotransmitter synthesis or degradation (SSA deficiency; SSADH deficiency), disorders of purine metabolism (ADSLD, Lesch-Nyhan syndrome), cerebral creatine deficiency syndromes (GAMT deficiency, creatine transporter defect), disorders of folate transport and metabolism (cerebral folate deficiency, MTHFR deficiency), lysosomal storage disorders (mucopolysaccharidosis type III, neuronal ceroid lipofuscinoses (NCL), Niemann-Pick disease type C), CTX, MELAS, Wilson disease, and several types of neurodegeneration with brain iron accumulation, among others. Moreover, multiple lines of evidence have suggested biochemical alterations in children with ASD compared to peers. Alterations can derive from the large intestine microbiome, showing differential fecal content for isopropanol, p-cresol, and short-chain fatty acids. Mitochondrial dysfunction can result in complex I, IV, or V deficiencies in children with ASD (up to 7%). Alterations were also demonstrated in some cases of glutathione reduction in the CNS [107]. While attempts to find unifying markers of autism demonstrated interesting candidates, they failed to show consistency across multiple UM platforms and ASD cohorts. Glutaric acid, arginine, histidine (and its catabolites), taurine, β-alanine, and succinic acid were most consistently elevated, along with a reduction of creatine and creatinine [107]. In a cohort of 52 pregnant women whose children developed ASD, compared to 62 control pregnant women in an NMR-based UM, differences were found in glycosphingolipid metabolism, N-glycan and pyrimidine metabolism, bile acid pathways, and C21-steroid hormone biosynthesis but with only a mild perturbation in each metabolite [108]. The latter two candidates are of interest due to the involvement of cholesterol metabolism, as ASD is highly prevalent among patients with Smith-Lemli-Opitz syndrome, a disorder of cholesterol synthesis. Abnormal bile acid synthesis can cause perturbations in taurine, and pyrimidine synthesis defects can alter β-alanine. However, a lesson learned from these works is that ASD may represent shared neurodevelopmental outcomes of quite different neurometabolic processes, complicating the search for biomarkers, as the case separation from control may be shadowed by intragroup variability

[109]. Instead, a machine learning approach to recognize a pattern of altered seemingly unrelated metabolites in a study group could allow the discovery of heterogeneous biomarkers, indicating a high-risk status for ASD, Alzheimer's disease, Parkinson's disease, or other conditions, and, once defined, may also be then targeted for treatment and/or monitoring. Such a learning process was performed on UM data from a cohort of 500 children with ASD (against 200 controls) [110]. The group was halved into study and discovery groups. The study group yielded 34 groups sharing metabolic phenotypes with a specificity of ≥95%. Those groups were then clustered into six metabolic groups, each with an abnormal ratio of either α-ketoglutarate, 4-hydroxyproline, glycine, lactate/pyruvate, ornithine, or succinic acid compared to other metabolites. These clusters were then validated in the discovery group. When screening for the clusters together, this assay had 53% sensitivity and 91% specificity for indicating "high risk" for ASD. These outcomes demonstrate the potential for a higher level of complexity in data interpretation. Identification of such abnormalities by the physician ordering an "ASD risk stratification" test may be a difficult task. Still, computer-assisted data analysis could flag such abnormal results and promote a revolution in the field of developmental neuroscience.

While UM offers an opportunity for discovery both inside and outside clinical settings, its use in the clinical lab supports broad screening for inborn errors of metabolism well beyond the newborn screen, supports the development of metabolic profiles for a disease that may be monitored during treatment, and, when integrated with genomics, provides a precision medicine approach to the diagnosis of rare disease.

Glossary

Dried blood spot (DBS) A method of whole blood sample collection in which a small amount of fresh blood is blotted onto an absorbent filter paper, followed by drying. This method provides a convenient storage and shipment platform and is widely used for newborn screening. Typically, a small punch from the DBS paper is eluted with phosphate-buffered saline, availing the sample for testing.

Elevation/reduction (of a metabolite) In the context of this chapter, a metabolite is considered *reduced* (insufficient) or *elevated* (in excess) when UM reveals a *Z*-score ≥+2 or ≤−2. The *Z*-score is the number of standard deviations that a data point differs from the population means, representing the relative level of a given metabolite. Raw values for individual metabolites are $\log_2$-transformed, and the relative *Z*-score is calculated compared to a lab-specific reference population [91, 111–113].

Inborn error of metabolism (IEM) A heterogeneous group of mostly inherited disorders involving a failure of the certain metabolic pathway(s) to break down or store biomolecules (typically carbohydrates, lipids, or amino acids) in the cell.

Although any given inborn error of metabolism is rare, taken as a group, inborn errors of metabolism occur in 1 in 2000 births [114].

Molecular confirmation A suspected diagnosis, as suggested by biochemical testing such as UM, is said to be molecularly confirmed when genomic sequencing reveals pathogenic variants, either monoallelic for autosomal dominant or X-linked disorders or biallelic for recessive disorders in the gene associated with the metabolic abnormality. Sequencing can be targeted for the specific gene(s) or untargeted as exome or genome sequencing. In the latter case, further confirmation by Sanger sequencing may be performed to validate the variants identified.

Reference population A lab-specific reference population created by performing UM on samples received in the clinical laboratory, with careful inclusion and exclusion of clinical samples to ensure pathways and analytes are covered for comprehensive clinical assessment. Raw data for each metabolite are median scaled, $\log_2$ transformed, extreme outliers removed, and *Z*-scores generated based on the mean and standard deviation in this reference population [91, 111–113].

Traditional screening methods Screening tools for certain common abnormalities indicative of diseases. While there is no definition per se for traditional screening methods, they usually include organic acids measured in urine (urine organic acids, UOA); measurements of standard amino acids in plasma (PAA); and carnitine conjugates of fatty acids (acylcarnitine profile, ACP). ACP and PAA are examples of TM where predefined biochemicals are quantitatively measured compared to known standards. UOA is a semiquantitative, untargeted analysis in which analytes are qualitatively compared against a few laboratory-specific internal standards. UOA and PAA are typically performed by liquid chromatography (LC) and/or gas chromatography (GC) coupled with mass spectrometry (MS). In contrast, ACP is performed by tandem MS/MS, where indicators are aimed at detecting carnitine daughter ions.

References

1. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. Biomed Res Int. 2015;2015:354671.
2. Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harb Mol Case Stud. 2015;1(1):a000588.
3. Ashrafian H, Sounderajah V, Glen R, Ebbels T, Blaise BJ, Kalra D, et al. Metabolomics: the stethoscope for the twenty-first century. Med Princ Pract. 2021;30(4):301–10.
4. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448–53.
5. Kalhan SC, Guo L, Edmison J, Dasarathy S, McCullough AJ, Hanson RW, et al. Plasma metabolomic profile in nonalcoholic fatty liver disease. Metabolism. 2011;60(3):404–13.
6. Mimmi MC, Finato N, Pizzolato G, Beltrami CA, Fogolari F, Corazza A, et al. Absolute quantification of choline-related biomarkers in breast cancer biopsies by liquid chromatography electrospray ionization mass spectrometry. Anal Cell Pathol (Amst). 2013;36(3–4):71–83.
7. Glunde K, Jacobs MA, Bhujwalla ZM. Choline metabolism in cancer: implications for diagnosis and therapy. Expert Rev Mol Diagn. 2006;6(6):821–9.

8. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet Med. 2018;20(4):435–43.
9. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet Med. 2019;21(11):2413–21.
10. Manickam K, McClain MR, Demmer LA, Biswas S, Kearney HM, Malinowski J, et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2021;23:2029.
11. Baxter SK, King MC. A time to sequence. JAMA Pediatr. 2017;171(12):e173435.
12. Theunissen TEJ, Nguyen M, Kamps R, Hendrickx AT, Sallevelt S, Gottschalk RWH, et al. Whole exome sequencing is the preferred strategy to identify the genetic defect in patients with a probable or possible mitochondrial cause. Front Genet. 2018;9:400.
13. Lehtonen JM, Auranen M, Darin N, Sofou K, Bindoff L, Hikmat O, et al. Diagnostic value of serum biomarkers FGF21 and GDF15 compared to muscle sample in mitochondrial disease. J Inherit Metab Dis. 2021;44(2):469–80.
14. Mancuso M, Klopstock T. Diagnosis and management of mitochondrial disorders. Berlin: Springer International; 2019.
15. Friedman JM, Jones KL, Carey JC. Exome sequencing and clinical diagnosis, vol. 324. JAMA; 2020. p. 627.
16. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24.
17. Cooper GM. Parlez-vous VUS? Genome Res. 2015;25(10):1423–6.
18. Bertier G, Hétu M, Joly Y. Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. BMC Med Genet. 2016;9(1):52.
19. Liu N, Xiao J, Gijavanekar C, Pappan KL, Glinton KE, Shayota BJ, et al. Comparison of untargeted metabolomic profiling vs traditional metabolic screening to identify inborn errors of metabolism. JAMA Netw Open. 2021;4(7):e2114155.
20. Alaimo JT, Glinton KE, Liu N, Xiao J, Yang Y, Reid Sutton V, et al. Integrated analysis of metabolomic profiling and exome data supplements sequence variant interpretation, classification, and diagnosis. Genet Med. 2020;22(9):1560–6.
21. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, et al. Performance of ACMG-AMP variant-interpretation guidelines among nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. Am J Hum Genet. 2016;98(6):1067–76.
22. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Köster J, Marais A, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet. 2017;25(2):176–82.
23. Shashi V, McConkie-Rosell A, Schoch K, Kasturi V, Rehder C, Jiang YH, et al. Practical considerations in the clinical application of whole-exome sequencing. Clin Genet. 2016;89(2):173–81.
24. Park KJ, Park S, Lee E, Park JH, Park JH, Park HD, et al. A population-based genomic study of inherited metabolic diseases detected through newborn screening. Ann Lab Med. 2016;36(6):561–72.
25. Atwal PS, Donti TR, Cardon AL, Bacino CA, Sun Q, Emrick L, et al. Aromatic L-amino acid decarboxylase deficiency diagnosed by clinical metabolomic profiling of plasma. Mol Genet Metab. 2015;115(2–3):91–4.
26. Wassenberg T, Molero-Luis M, Jeltsch K, Hoffmann GF, Assmann B, Blau N, et al. Consensus guideline for the diagnosis and treatment of aromatic l-amino acid decarboxylase (AADC) deficiency. Orphanet J Rare Dis. 2017;12(1):12.

27. Saudubray JM, Baumgartner MR, Walter JH. Inborn metabolic diseases: diagnosis and treatment. Berlin, Heidelberg: Springer; 2016.
28. Pearson TS, Gilbert L, Opladen T, Garcia-Cazorla A, Mastrangelo M, Leuzzi V, et al. AADC deficiency from infancy to adulthood: symptoms and developmental outcome in an international cohort of 63 patients. J Inherit Metab Dis. 2020;43(5):1121–30.
29. Fusco C, Leuzzi V, Striano P, Battini R, Burlina A, Spagnoli C. Aromatic L-amino acid decarboxylase (AADC) deficiency: results from an Italian modified Delphi consensus. Ital J Pediatr. 2021;47(1):13.
30. Luc QN, Querubin J. Clinical management of dystonia in childhood. Paediatr Drugs. 2017;19(5):447–61.
31. Kim R, Jeon B, Lee WW. A systematic review of treatment outcome in patients with Dopa-responsive dystonia (DRD) and DRD-plus. Mov Disord Clin Pract. 2016;3(5):435–42.
32. Pappan KL, Kennedy AD, Magoulas PL, Hanchard NA, Sun Q, Elsea SH. Clinical metabolomics to segregate aromatic amino acid decarboxylase deficiency from drug-induced metabolite elevations. Pediatr Neurol. 2017;75:66–72.
33. Almontashiri NAM, Zha L, Young K, Law T, Kellogg MD, Bodamer OA, et al. Clinical validation of targeted and untargeted metabolomics testing for genetic disorders: a 3 year comparative study. Sci Rep. 2020;10(1):9382.
34. Rodan LH, Anyane-Yeboa K, Chong K, Klein Wassink-Ruiter JS, Wilson A, Smith L, et al. Gain-of-function variants in the ODC1 gene cause a syndromic neurodevelopmental disorder associated with macrocephaly, alopecia, dysmorphic features, and neuroimaging abnormalities. Am J Med Genet A. 2018;176(12):2554–60.
35. Bupp CP, Schultz CR, Uhl KL, Rajasekaran S, Bachmann AS. Novel de novo pathogenic variant in the ODC1 gene in a girl with developmental delay, alopecia, and dysmorphic features. Am J Med Genet A. 2018;176(12):2548–53.
36. Marbaix AY, Noël G, Detroux AM, Vertommen D, Van Schaftingen E, Linster CL. Extremely conserved ATP- or ADP-dependent enzymatic system for nicotinamide nucleotide repair. J Biol Chem. 2011;286(48):41246–52.
37. Marbaix AY, Tyteca D, Niehaus TD, Hanson AD, Linster CL, Van Schaftingen E. Occurrence and subcellular distribution of the NADPHX repair system in mammals. Biochem J. 2014;460(1):49–58.
38. Van Bergen NJ, Guo Y, Rankin J, Paczia N, Becker-Kettern J, Kremer LS, et al. NAD(P)HX dehydratase (NAXD) deficiency: a novel neurodegenerative disorder exacerbated by febrile illnesses. Brain. 2019;142(1):50–8.
39. Kremer LS, Danhauser K, Herebian D, Petkovic Ramadža D, Piekutowska-Abramczuk D, Seibt A, et al. NAXE mutations disrupt the cellular NAD(P)HX repair system and cause a lethal Neurometabolic disorder of early childhood. Am J Hum Genet. 2016;99(4):894–902.
40. Van Bergen NJ, Walvekar AS, Patraskaki M, Sikora T, Linster CL, Christodoulou J. Clinical and biochemical distinctions for a metabolite repair disorder caused by NAXD or NAXE deficiency. J Inherit Metab Dis. 2022;45(6):1028–38.
41. Manor J, Calame DG, Gijavanekar C, Tran A, Fatih JM, Lalani SR, et al. Niacin therapy improves outcome and normalizes metabolic abnormalities in an NAXD-deficient patient. Brain. 2022;145(5):e36–40.
42. Pillai NR, Amin H, Gijavanekar C, Liu N, Issaq N, Broniowska KA, et al. Hematologic presentation and the role of untargeted metabolomics analysis in monitoring treatment for riboflavin transporter deficiency. Am J Med Genet A. 2020;182(11):2781–7.
43. Amir F, Atzinger C, Massey K, Greinwald J, Hunter LL, Ulm E, et al. The clinical journey of patients with riboflavin transporter deficiency type 2. J Child Neurol. 2020;35(4):283–90.
44. Liu Z, Peng Q, Li J, Rao C, Lu X. BVVLS2 overlooked for 3 years in a pediatric patient caused by novel compound heterozygous mutations in *SLC52A2* gene. Clin Chim Acta. 2021;523:402–6.
45. Kalafatic Z, Lipovac K, Jezerinac Z, Juretic D, Dumic M, Zurga B, et al. A liver urocanase deficiency. Metabolism. 1980;29(11):1013–9.

46. Espinós C, Pineda M, Martínez-Rubio D, Lupo V, Ormazabal A, Vilaseca MA, et al. Mutations in the urocanase gene UROC1 are associated with urocanic aciduria. J Med Genet. 2009;46(6):407–11.
47. Camp BW, Broman SH, Nichols PL, Leff M. Maternal and neonatal risk factors for mental retardation: defining the ‘at-risk’ child. Early Hum Dev. 1998;50(2):159–73.
48. Boyle CA, Yeargin-Allsopp M, Doernberg NS, Holmgreen P, Murphy CC, Schendel DE. Prevalence of selected developmental disabilities in children 3-10 years of age: the metropolitan Atlanta developmental disabilities surveillance program, 1991. MMWR CDC Surveill Summ. 1996;45(2):1–14.
49. Keyfi F, Nasseri M, Nayerabadi S, Alaei A, Mokhtariye A, Varasteh A. Frequency of inborn errors of metabolism in a northeastern Iranian sample with high consanguinity rates. Hum Hered. 2018;83(2):71–8.
50. Afzal RM, Lund AM, Skovby F. The impact of consanguinity on the frequency of inborn errors of metabolism. Mol Genet Metab Rep. 2018;15:6–10.
51. Saad HA, Elbedour S, Hallaq E, Merrick J, Tenenbaum A. Consanguineous marriage and intellectual and developmental disabilities among Arab Bedouins children of the Negev region in southern Israel: a pilot study. Front Public Health. 2014;2:3.
52. Jamra R. Genetics of autosomal recessive intellectual disability. Med Genet. 2018;30(3):323–7.
53. Glinton KE, Levy HL, Kennedy AD, Pappan KL, Elsea SH. Untargeted metabolomics identifies unique though benign biochemical changes in patients with pathogenic variants in UROC1. Mol Genet Metab Rep. 2019;18:14–8.
54. Kennedy AD, Pappan KL, Donti T, Delgado MR, Shinawi M, Pearson TS, et al. 2-Pyrrolidinone and Succinimide as clinical screening biomarkers for GABA-transaminase deficiency: anti-seizure medications impact accurate diagnosis. Front Neurosci. 2019;13:394.
55. Ferreira EA, Veenvliet ARJ, Engelke UFH, Kluijtmans LAJ, Huigen M, Hoegen B, et al. Diagnosing, discarding, or de-VUSsing: a practical guide to (un)targeted metabolomics as variant-transcending functional tests. Genet Med. 2023;25(1):125–34.
56. Subramanian VS, Constantinescu AR, Benke PJ, Said HM. Mutations in SLC5A6 associated with brain, immune, bone, and intestinal dysfunction in a young child. Hum Genet. 2017;136(2):253–61.
57. Byrne AB, Arts P, Polyak SW, Feng J, Schreiber AW, Kassahn KS, et al. Identification and targeted management of a neurodegenerative disorder caused by biallelic mutations in SLC5A6. NPJ Genom Med. 2019;4:28.
58. Holling T, Nampoothiri S, Tarhan B, Schneeberger PE, Vinayan KP, Yesodharan D, et al. Novel biallelic variants expand the SLC5A6-related phenotypic spectrum. Eur J Hum Genet. 2022;1-11:439.
59. Thistlethwaite LR, Petrosyan V, Li X, Miller MJ, Elsea SH, Milosavljevic A. CTD: an information-theoretic algorithm to interpret sets of metabolomic and transcriptomic perturbations in the context of graphical models. PLoS Comput Biol. 2021;17(1):e1008550.
60. Thistlethwaite LR, Li X, Burrage LC, Riehle K, Hacia JG, Braverman N, et al. Clinical diagnosis of metabolic disorders using untargeted metabolomic profiling and disease-specific networks learned from profiling data. Sci Rep. 2022;12(1):6556.
61. Kerkhofs M, Haijes HA, Willemsen AM, van Gassen KLI, van der Ham M, Gerrits J, et al. Cross-omics: integrating genomics with metabolomics in clinical diagnostics. Metabolites. 2020;10(5):206.
62. Graham Linck EJ, Richmond PA, Tarailo-Graovac M, Engelke U, Kluijtmans LAJ, Coene KLM, et al. metPropagate: network-guided propagation of metabolomic information for prioritization of metabolic disease genes. NPJ Genom Med. 2020;5:25.
63. Smedley D, Jacobsen JO, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the exomiser. Nat Protoc. 2015;10(12):2004–15.
64. Bongaerts M, Bonte R, Demirdas S, Huidekoper HH, Langendonk J, Wilke M, et al. Integration of metabolomics with genomics: metabolic gene prioritization using metabolomics data and genomic variant (CADD) scores. Mol Genet Metab. 2022;136(3):199–218.

65. Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. Genome Med. 2021;13(1):31.
66. Messa GM, Napolitano F, Elsea SH, di Bernardo D, Gao X. A Siamese neural network model for the prioritization of metabolic disorders by integrating real and simulated data. Bioinformatics. 2020;36(Suppl_2):i787–i94.
67. Garg U, Smith LD. Biomarkers in inborn errors of metabolism: clinical aspects and laboratory determination. Amsterdam: Elsevier Science; 2017.
68. Braverman NE, Raymond GV, Rizzo WB, Moser AB, Wilkinson ME, Stone EM, et al. Peroxisome biogenesis disorders in the Zellweger spectrum: an overview of current diagnosis, clinical manifestations, and treatment guidelines. Mol Genet Metab. 2016;117(3):313–21.
69. Steinberg SJ, Raymond GV, Braverman NE, Moser AB. Zellweger spectrum disorder. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, et al., editors. GeneReviews(®). Seattle, WA: University of Washington; 1993.
70. Kim PK, Hettema EH. Multiple pathways for protein transport to peroxisomes. J Mol Biol. 2015;427(6 Pt A):1176–90.
71. Wangler MF, Hubert L, Donti TR, Ventura MJ, Miller MJ, Braverman N, et al. A metabolomic map of Zellweger spectrum disorders reveals novel disease biomarkers. Genet Med. 2018;20(10):1274–83.
72. Ah Mew N, Simpson KL, Gropman AL, Lanpher BC, Chapman KA, Summar ML. Urea cycle disorders overview. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, et al., editors. GeneReviews(®). Seattle, WA: University of Washington; 1993.
73. Stone WL, Basit H, Jaishankar GB. Urea cycle disorders. In: StatPearls. Treasure Island, FL: StatPearls; 2022.
74. Sen K, Whitehead M, Castillo Pinto C, Caldovic L, Gropman A. Fifteen years of urea cycle disorders brain research: looking back, looking forward. Anal Biochem. 2022;636:114343.
75. Amayreh W, Meyer U, Das AM. Treatment of arginase deficiency revisited: guanidinoacetate as a therapeutic target and biomarker for therapeutic monitoring. Dev Med Child Neurol. 2014;56(10):1021–4.
76. Burrage LC, Thistlethwaite L, Stroup BM, Sun Q, Miller MJ, Nagamani SCS, et al. Untargeted metabolomic profiling reveals multiple pathway perturbations and new clinical biomarkers in urea cycle disorders. Genet Med. 2019;21(9):1977–86.
77. De Deyn PP, Marescau B, Macdonald RL. Guanidino compounds that are increased in hyperargininemia inhibit GABA and glycine responses on mouse neurons in cell culture. Epilepsy Res. 1991;8(2):134–41.
78. Hanna-El-Daher L, Béard E, Henry H, Tenenbaum L, Braissant O. Mild guanidinoacetate increase under partial guanidinoacetate methyltransferase deficiency strongly affects brain cell development. Neurobiol Dis. 2015;79:14–27.
79. Ostojic SM. Safety of dietary Guanidinoacetic acid: a villain of a good guy? Nutrients. 2021;14(1):75.
80. Ingoglia F, Chong JL, Pasquali M, Longo N. Creatine metabolism in patients with urea cycle disorders. Mol Genet Metab Rep. 2021;29:100791.
81. Grace RF, Bianchi P, van Beers EJ, Eber SW, Glader B, Yaish HM, et al. Clinical spectrum of pyruvate kinase deficiency: data from the pyruvate kinase deficiency natural history study. Blood. 2018;131(20):2183–92.
82. Bianchi P, Fermo E, Glader B, Kanno H, Agarwal A, Barcellini W, et al. Addressing the diagnostic gaps in pyruvate kinase deficiency: consensus recommendations on the diagnosis of pyruvate kinase deficiency. Am J Hematol. 2019;94(1):149–61.
83. Van Dooijeweert B, Broeks MH, Verhoeven-Duif NM, Van Beers EJ, Nieuwenhuis EES, Van Solinge WW, et al. Untargeted metabolic profiling in dried blood spots identifies disease fingerprint for pyruvate kinase deficiency. Haematologica. 2021;106(10):2720–5.
84. Wang D, Pascual JM, De Vivo D. Glucose transporter type 1 deficiency syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, et al., editors. GeneReviews(®). Seattle, WA: University of Washington; 1993.

85. Tang M, Monani UR. Glut1 deficiency syndrome: new and emerging insights into a prototypical brain energy failure disorder. Neurosci Insights. 2021;16:26331055211011507.
86. Wibisono C, Rowe N, Beavis E, Kepreotes H, Mackie FE, Lawson JA, et al. Ten-year single-center experience of the ketogenic diet: factors influencing efficacy, tolerability, and compliance. J Pediatr. 2015;166(4):1030–6.e1.
87. Klepper J, Akman C, Armeno M, Auvin S, Cervenka M, Cross HJ, et al. Glut1 deficiency syndrome (Glut1DS): state of the art in 2020 and recommendations of the international Glut1DS study group. Epilepsia Open. 2020;5(3):354–65.
88. Almuqbil M, Go C, Nagy LL, Pai N, Mamak E, Mercimek-Mahmutoglu S. New paradigm for the treatment of glucose transporter 1 deficiency syndrome: low glycemic index diet and modified high amylopectin cornstarch. Pediatr Neurol. 2015;53(3):243–6.
89. Cappuccio G, Pinelli M, Alagia M, Donti T, Day-Salvatore DL, Veggiotti P, et al. Biochemical phenotyping unravels novel metabolic abnormalities and potential biomarkers associated with treatment of GLUT1 deficiency with ketogenic diet. PloS One. 2017;12(9):e0184022.
90. Miller MJ, Bostwick BL, Kennedy AD, Donti TR, Sun Q, Sutton VR, Elsea SH. Chronic Oral L-Carnitine Supplementation Drives Marked Plasma TMAO Elevations in Patients with Organic Acidemias Despite Dietary Meat Restrictions. JIMD Rep. 2016;30:39–44. https://doi.org/10.1007/8904_2016_539. Epub 2016 Mar 3. PMID: 26936850; PMCID: PMC5110437.
91. Kennedy AD, Miller MJ, Beebe K, Wulff JE, Evans AM, Miller LA, et al. Metabolomic profiling of human urine as a screen for multiple inborn errors of metabolism. Genet Test Mol Biomarkers. 2016;20(9):485–95.
92. El-Hattab AW. Serine biosynthesis and transport defects. Mol Genet Metab. 2016;118(3):153–9.
93. Glinton KE, Benke PJ, Lines MA, Geraghty MT, Chakraborty P, Al-Dirbashi OY, et al. Disturbed phospholipid metabolism in serine biosynthesis defects revealed by metabolomic profiling. Mol Genet Metab. 2018;123(3):309–16.
94. Ferreira CR, Goorden SMI, Soldatos A, Byers HM, Ghauharali-van der Vlugt JMM, Beers-Stet FS, et al. Deoxysphingolipid precursors indicate abnormal sphingolipid metabolism in individuals with primary and secondary disturbances of serine availability. Mol Genet Metab. 2018;124(3):204–9.
95. Litwack G. Chapter 7 - Glycogen and Glycogenolysis. In: Litwack G, editor. Human biochemistry. Boston: Academic Press; 2018. p. 161–81.
96. Banne E, Meiner V, Shaag A, Katz-Brull R, Gamliel A, Korman S, et al. Transaldolase deficiency: a new case expands the phenotypic Spectrum. JIMD Rep. 2016;26:31–6.
97. Shayota BJ, Donti TR, Xiao J, Gijavanekar C, Kennedy AD, Hubert L, et al. Untargeted metabolomics as an unbiased approach to the diagnosis of inborn errors of metabolism of the non-oxidative branch of the pentose phosphate pathway. Mol Genet Metab. 2020;131(1–2):147–54.
98. Buzkova J, Nikkanen J, Ahola S, Hakonen AH, Sevastianova K, Hovinen T, et al. Metabolomes of mitochondrial diseases and inclusion body myositis patients: treatment targets and biomarkers. EMBO Mol Med. 2018;10(12):e9091.
99. Sharma R, Reinstadler B, Engelstad K, Skinner OS, Stackowitz E, Haller RG, et al. Circulating markers of NADH-reductive stress correlate with mitochondrial disease severity. J Clin Invest. 2021;131(2):e136055.
100. Yatsuga S, Fujita Y, Ishii A, Fukumoto Y, Arahata H, Kakuma T, et al. Growth differentiation factor 15 as a useful biomarker for mitochondrial disorders. Ann Neurol. 2015;78(5):814–23.
101. Evans B. How autism became autism: the radical transformation of a central concept of child development in Britain. Hist Human Sci. 2013;26(3):3–31.
102. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The heritability of autism Spectrum disorder. JAMA. 2017;318(12):1182–4.
103. Colvert E, Tick B, McEwen F, Stewart C, Curran SR, Woodhouse E, et al. Heritability of autism Spectrum disorder in a UK population-based twin sample. JAMA Psychiatry. 2015;72(5):415–23.

104. Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. Genet Med. 2013;15(5):399–407.
105. Ghaziuddin M, Al-Owain M. Autism spectrum disorders and inborn errors of metabolism: an update. Pediatr Neurol. 2013;49(4):232–6.
106. Žigman T, Petković Ramadža D, Šimić G, Barić I. Inborn errors of metabolism associated with autism Spectrum disorders: approaches to intervention. Front Neurosci. 2021;15:673600.
107. Glinton KE, Elsea SH. Untargeted metabolomics for autism spectrum disorders: current status and future directions. Front Psych. 2019;10:647.
108. Ritz B, Yan Q, Uppal K, Liew Z, Cui X, Ling C, et al. Untargeted metabolomics screen of mid-pregnancy maternal serum and autism in offspring. Autism Res. 2020;13(8):1258–69.
109. Courraud J, Ernst M, Svane Laursen S, Hougaard DM, Cohen AS. Studying autism using untargeted metabolomics in newborn screening samples. J Mol Neurosci. 2021;71(7):1378–93.
110. Smith AM, Natowicz MR, Braas D, Ludwig MA, Ney DM, Donley ELR, et al. A metabolomics approach to screening for autism risk in the Children's autism metabolome project. Autism Res. 2020;13(8):1270–85.
111. Ford L, Kennedy AD, Goodman KD, Pappan KL, Evans AM, Miller LAD, et al. Precision of a clinical metabolomics profiling platform for use in the identification of inborn errors of metabolism. J Appl Lab Med. 2020;5(2):342–56.
112. Kennedy AD, Pappan KL, Donti TR, Evans AM, Wulff JE, Miller LAD, et al. Elucidation of the complex metabolic profile of cerebrospinal fluid using an untargeted biochemical profiling assay. Mol Genet Metab. 2017;121(2):83–90.
113. Kennedy AD, Wittmann BM, Evans AM, Miller LAD, Toal DR, Lonergan S, et al. Metabolomics in the clinic: a review of the shared and unique features of untargeted metabolomics for clinical research and clinical testing. J Mass Spectrom. 2018;53(11):1143–54.
114. Waters D, Adeloye D, Woolham D, Wastnedge E, Patel S, Rudan I. Global birth prevalence and mortality from inborn errors of metabolism: a systematic analysis of the evidence. J Glob Health. 2018;8(2):021102.

Metabolomics in the Study of Human Mitochondrial Diseases

Rajaa Sebaa, Mary-Ellen Harper, Ruqaiah Al-Tassan, Mohammed Al-Owain, and Anas M. Abdel Rahman

Abstract Mitochondria are dynamic cellular organelles playing many biological roles that are fundamentally required for cellular functions. The primary role of mitochondria is ATP production through oxidative phosphorylation (OXPHOS). Mitochondria are found in nearly all cell types, and their number within cells varies in a tissue−/organ-dependent manner. Tissues/organs characterized by high-energy demands contain abundant mitochondria, and these tissues/organs are most frequently affected when their mitochondria are dysfunctional. The resulting pathologies can be generally referred to as mitochondrial diseases (MDs). MDs can be caused by nuclear or mitochondrial DNA mutations in genes encoding mitochondrial proteins, including OXPHOS proteins. Also, MDs can be developed through nongenetic mechanisms such as those involving environmental factors, mitotoxicity drugs, oxidative stress, and aging. MDs can appear over the entire life span. Patients with particular MDs present a wide range of heterogeneous phenotypes with differ-

R. Sebaa
Department of Medical Laboratories, College of Applied Medical Sciences, Shaqra University, Al-Dawadmi, Saudi Arabia

M.-E. Harper
Department of Biochemistry, Microbiology, and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

R. Al-Tassan
Department of Internal Medicine, King Abdullah Hospital, Bisha, Saudi Arabia

Department of Medical Genomics, Centre for Genomic Medicine, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

M. Al-Owain
Department of Medical Genomics, Centre for Genomic Medicine, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

A. M. Abdel Rahman (✉)
Metabolomics Section, Department of Clinical Genomics, Center for Genome Medicine, King Faisal Specialist Hospital & Research Center (KFSHRC), Riyadh, Saudi Arabia
e-mail: AabdelRahman46@kfshrc.edu.sa

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_7

ent levels of disease severity. The wide variety of leading causes and heterogenous phenotypes of MDs make diagnosing MDs notoriously challenging. Despite these challenges, multiple diagnostic examinations and tools, including family history, phenotypic examinations, neurological imaging, biochemical tests, and genetic analyses, have collectively enhanced the diagnosis of MDs. As a result of the diagnostic limitations and drawbacks, there have been demands for developing new diagnostic approaches capable of detecting metabolic perturbations of MDs used as metabolic biosignatures. For that reason, the metabolomic approach, the study of small metabolites ≤1500 daltons, has recently garnered attention. While metabolomics offers significant advances, it is recommended that data sets be integrated with other diagnostic approaches. This chapter reviews the application of metabolomic analyses in studying human MDs.

Keywords Mitochondria · Mitochondrial diseases · Mass spectrometry · Nuclear Magnetic Resonance (NMR) Spectroscopy · Untargeted metabolomics · Targeted metabolomics · Metabolic biosignatures

Abbreviations

ANT	Adenine nucleotide translocase
ATP	Adenosine triphosphate
BAT	Brown adipose tissue
CAT	Catalase
CE-MS	Capillary electrophoresis-coupled mass spectrometry
CIL	Chemical isotope labeling
CPEO	Chronic progressive external ophthalmoplegia
DRP1	Dynamin-related protein-1
ETF	Electron transfer flavoprotein
FIS1	Fission protein-1
GC-MS	Gas chromatography-coupled mass spectrometry
GPxs	Glutathione peroxidases
GSH	Glutathione
IMM	Inner mitochondrial membrane
IMS	Intermembrane space
KSS	Kearns-Sayre syndrome
LC-MS	Liquid chromatography-coupled mass spectrometry
LHON	Leber hereditary optic neuropathy
MDs	Mitochondrial diseases
MELAS	Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes
MERRF	Myoclonic epilepsy with ragged-red fibers
MFF	Mitochondrial fission factor
MFN1	Mitofusins-1
MFN2	Mitofusins-2
MiD49	Mitochondrial dynamic proteins of 49 kDa

MiD51	Mitochondrial dynamic proteins of 51 kDa
MNGIE	Mitochondrial neurogastrointestinal encephalopathy
MS	Mass spectrometry
mtDNA	Mitochondrial DNA
NAC	N-acetylcysteine
NARP	Neuropathy, ataxia, and retinitis pigmentosa
nDNA	Nuclear DNA
NMR	Nuclear magnetic resonance
OMM	Outer mitochondrial membrane
OPA1	Optic atrophy-1
OXPHOS	Oxidative phosphorylation
PMF	Proton motive force
PMS	Pearson marrow pancreas syndrome
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
SOD	Superoxide dismutase
SPG7	Hereditary spastic paraplegia 7
TCA	Tricarboxylic acid cycle
TFAM	Transcription factor A of mitochondria
TFB2M	Mitochondrial transcription factor B2
tRNA	Transfer RNA
Trx	Thioredoxin
UCP1	Uncoupling protein-1
VUS	Variants of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing

1 Introduction

Mitochondria are organelles most commonly referred to as the powerhouses of cells due to their fundamental function of ATP production to fuel the energy-demanding process in cells. Mitochondria transduce energy substrates into adenosine triphosphate (ATP) in key metabolic tissues [1]. Mitochondria are found in nearly all cell types, although their numbers vary in a tissue−/organ-dependent manner. Tissues characterized by high energy demands have many mitochondria, but those with less energy demand have fewer mitochondria to match their cellular energy requirements [2]. Despite the variations in mitochondrial content across different tissues, certain factors can increase or decrease the mitochondrial number in adaptive processes when cells undergo physiological changes such as exercise, dietary energy surfeit or deficit, tissue growth or atrophy, and aging [3, 4]. The role of mitochondria in energy transduction is widely understood. However, mitochondria are also involved in other vital biological functions, including reactive oxygen species (ROS) production, redox signaling, Ca^{+2} hemostasis, and cellular apoptosis [5, 6].

Moreover, mitochondrial functions are influenced by certain properties such as mitochondrial morphology, ultrastructure, and dynamics [7, 8].

Depending on cell type and physiological and energetic status, mitochondria present diverse lengths from 0.5 to 10 μm and shapes, but they share the main structural components [9]. Mitochondria have five main compartments, including the outer mitochondrial membrane (OMM), inner mitochondrial membrane (IMM), intermembrane space (IMS), cristae, and matrix. The OMM and IMM are mainly composed of phospholipids and proteins to support the function of these membranes, with the IMM having a far greater membrane protein density than the OMM. The specific compositions of phospholipids and proteins in the OMM and IMM determine the degree of the integrity and permeability of these membranes. The OMM is relatively permeable, transporting low molecular metabolites, solutes, and ions from the cytoplasm into the IMS. The IMM is highly selectively impermeable to most solutes and metabolites as this structural feature of the IMM is important to allow OXPHOS to occur.

Furthermore, the IMM is associated with a wide range of transporters and protein shuttles to support the many mitochondrial metabolic and bioenergetic pathways [10, 11]. Aqueous regions defined by the membranes are the IMS and matrix. While IMS is present between OMM and IMM, the matrix is defined and enveloped by the IMM. Both the IMS and matrix are important for metabolic events and pathways. The matrix contains many enzymes involved in metabolic pathways such as the tricarboxylic acid cycle, fatty acid β-oxidation, ketogenesis, amino acid metabolism, urea cycle, hormone synthesis, etc. Also, the matrix houses the circular mitochondrial DNA (mtDNA), which exclusively encodes mitochondrial components and its genetic machinery elements such as ribosomes and RNA (more details mentioned below). The fifth mitochondrial compartment is the cristae, defined as the folds of IMM into the matrix to increase the surface area of IMM for enhancing mitochondrial metabolic activity and ATP production and to allow important protein-protein interactions [12]. The mitochondrial structure is illustrated in (Fig. 1).

Concerning mitochondrial morphology, mitochondria exist in many morphologies, such as short oval or spherical tubules, long elongated tubules, or reticular networks. Mitochondrial morphology is distinctively different across cell/tissue types. Moreover, multiple morphologies of mitochondria can be seen with one cell

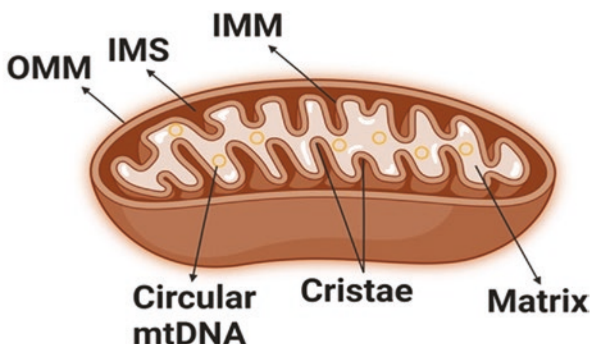

Fig. 1 Mitochondrial structure consists of outer mitochondrial membrane (OMM), inner mitochondrial membrane (IMM), intermembrane space (IMS), matrix, cristae, and circular mitochondrial genome (mtDNA)

type [13–15]. Under various physiological and pathological conditions, mitochondria continuously remodel their morphology to allow the cells to adapt [16, 17].

Mitochondria are highly dynamic and continuously undergo events of fission and fusion. Mitochondrial dynamics maintain mitochondrial health, morphology, size, and numbers [18, 19]. The morphological changes are induced when cells are triggered by certain stressors or undergo energetic changes and are required during cellular adaptation [9, 19]. Mitochondrial dynamics is under the control of GTPase proteins that mediate the events of fission and fusion. Mitochondrial fission is the fragmentation of one mitochondrion into two or more mitochondria. It requires the recruitment of a cytosolic protein called dynamin-related protein-1 (DRP1) through the action of dynamic proteins, including mitochondrial fission protein-1 (FIS1), mitochondrial fission factor (MFF), and mitochondrial dynamic proteins of 49 and 51 kDa (MiD49 and MiD51). Mitochondrial fission is important for quality control as the fragmented, damaged mitochondria are removed. This is referred to as mitophagy, the selective removal of damaged mitochondria by autophagy [20–22]. In contrast to mitochondrial fission, mitochondrial fusion promotes the joining of two separate mitochondria by merging OMM, IMM, and matrix to make one elongated organelle. Mitochondrial fusion needs the action of several GTPase proteins, including mitofusin (MFN1) and mitofusin-2 (MFN2), to promote OMM fusion and optic atrophy-1 (OPA1) protein for IMM fusion [23, 24]. Fission and fusion should be balanced to maintain a healthy mitochondrial reticulum in cells; otherwise, mitochondrial dysfunction occurs and contributes to the development of mitochondrial diseases (MDs).

From the genetic aspect, mitochondria have the exceptional feature of being the only cellular organelle possessing DNA beyond that found in the nucleus. mtDNA is maternally inherited since paternal sperm mitochondria are targeted and destroyed after egg fertilization during embryogenesis. mtDNA is circular, and there are 100–10,000 copies per cell. mtDNA contains 37 genes encoding only 13 mitochondrial protein subunits involved in OXPHOS, 2 ribosomal RNA (rRNA), and 22 transfer RNA (tRNA) for intra-mitochondrial protein synthesis [25, 26]. In contrast, approximately 1500 mitochondrial proteins are encoded by nuclear DNA (nDNA), translated into the cytoplasm, and imported to the mitochondria [27]. As mitochondria have their transcriptional machinery, mtDNA is colocalized with transcriptional factors, which are nuclear-encoded, such as transcription factor A of mitochondria (TFAM) and mitochondrial transcription factor B2 (TFB2M), to initiate and regulate mtDNA transcription and eventually synthesize OXPHOS subunits [28]. mtDNA is particularly susceptible to mutations compared to nDNA due to the limited repair mechanisms and the high exposure to ROS emission in mitochondria. Mutations in mitochondrial protein-encoding genes lead to dysfunctional mitochondria and various MDs (mentioned below).

2 Mitochondrial Metabolic Pathways

Mitochondria are hubs of cellular metabolism. They are involved in many crucial metabolic pathways, including the tricarboxylic acid cycle (TCA), fatty acid β-oxidation, acyl-carnitine metabolism, urea cycle, amino acid degradation, ketogenesis/ketolysis, steroidogenesis, and OXPHOS as shown in (Fig. 2). These pathways produce small molecular weight intermediate molecules called metabolites, which are also important for cellular proliferation, signaling, survival, and function. Defects in mitochondrial metabolism are associated with alterations in the level of metabolites, which can directly or indirectly affect the physiology of tissues/organs in the body [12, 29].

The OXPHOS system in the IMM consists of NADH dehydrogenase (complex I), succinate dehydrogenase (complex II), cytochrome *c* reductase (complex III), cytochrome *c* oxidase (complex IV), and ATP synthase (complex V) [30]. In detail, in the presence of oxygen, mitochondria convert chemical energy stored in energy substrates, such as pyruvate, acyl-CoA, ketone bodies, etc., into ATP through coupled OXPHOS systems mediated via complex V. These energetic metabolites undergo oxidative reactions producing reducing agents such as nicotinamide adenine dinucleotide hydrogen (NADH + H) and flavin adenine dinucleotide dihydrogen (FADH2). These reduced coenzymes then provide electrons either to certain OXPHOS complexes, including complex I and complex II or other IMM-bounded

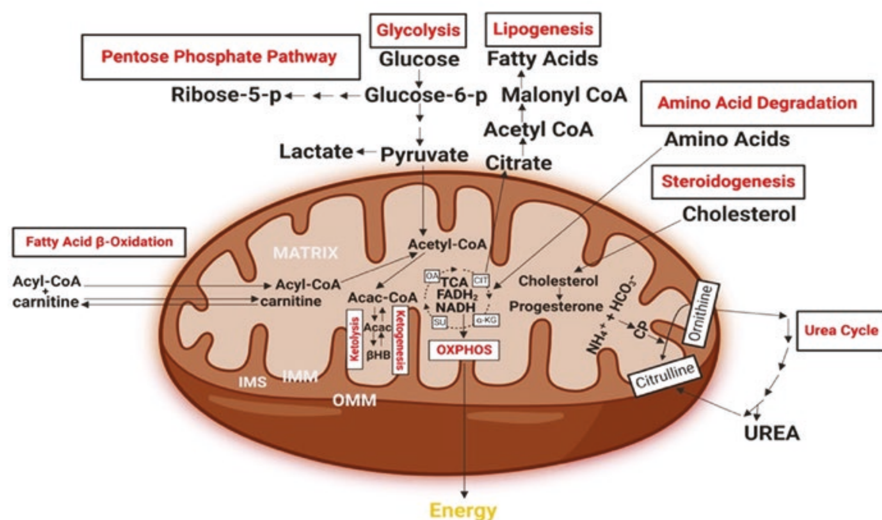

Fig. 2 Simplified illustration of various metabolic pathways taking place in mitochondria

proteins in conjunction with OXPHOS, such as electron transfer flavoprotein (ETF). When electrons are accepted by complex I, complex II, or ETF, they are subsequently passed to coenzyme Q, a mobile fat-soluble electron carrier found in the IMM. Then, coenzyme Q transfers the electrons to complex III, which are then transferred to complex IV through cytochrome C. Ultimately, the electrons reduce oxygen to produce water. While electron flow happens along complexes I–IV, protons are pumped from the matrix to IMS through complex I, complex III, and complex IV, contributing to the generation of an electrochemical gradient called proton motive force (PMF) located across the IMM. Subsequently, free energy stored in PMF drives protons back to the mitochondria matrix through complex V, resulting in the production of ATP, as shown in (Fig. 3) [30–33].

The coupling of electron flow (and oxygen consumption) from ATP production is far from perfect. A certain amount of uncoupling occurs through proton leaks in all cell types [34]. However, one highly metabolic tissue called brown adipose tissue (BAT) that uncoupled OXPHOS has the physiological function of heat production (thermogenesis) instead of ATP production. BAT has low levels of ATP synthase but high amounts of uncoupling protein 1 (UCP1). UCP1 belongs to a gene family of UCPs, and in BAT, it mediates proton leak for thermoregulatory functions. When UCP1 is activated, oxidative reactions upstream of UCP1 support a high influx of reducing equivalents (i.e., electrons) into the electron transfer system in complexes I–IV. This thereby results in proton pumping from the matrix into IMS, causing the formation of PMF, which is rapidly dissipated through UCP1-mediated proton leak

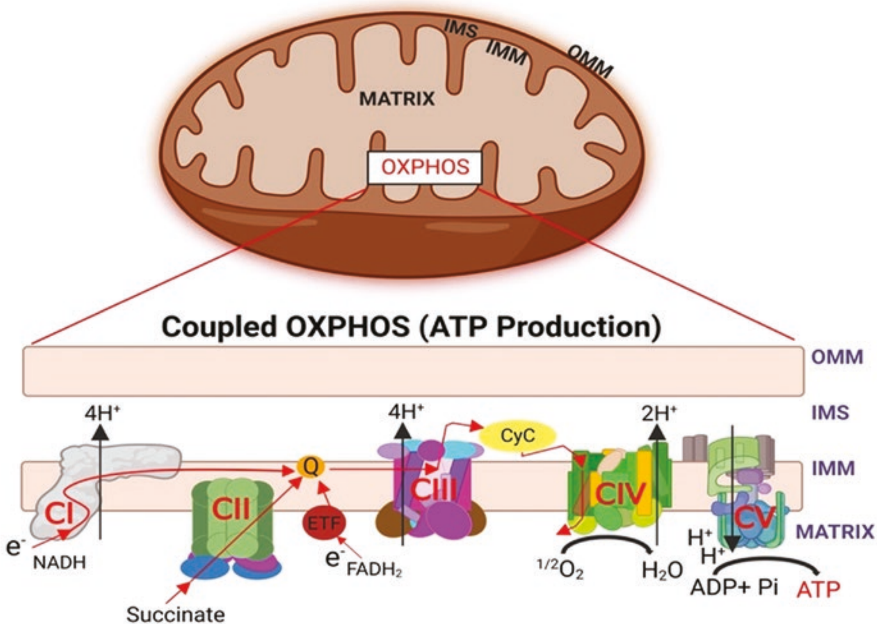

Fig. 3 Mitochondrial-coupled OXPHOS pathways

activity. Thermogenesis results from the upstream oxidative reactions [35]. This is depicted in (Fig. 4). Interestingly, there are other uncoupling proteins named UCP2 and UCP3, which protect against oxidative damage and facilitate fatty acid oxidation in the tissues/cells [36]. UCP2 is found in most tissue types, while UCP3 is predominantly found in the skeletal muscle and BAT, but their levels of expression are approximately two orders of magnitude lower than the expression of UCP1 in BAT [37, 38]. Adenine nucleotide translocase (ANT), which normally exchanges cytosolic ADP for mitochondrial ATP, has been shown to cause proton leak across the IMM and thereby possibly protect against oxidative damage [38–42].

During coupled or uncoupled OXPHOS pathways, ROS, which are byproducts of aerobic metabolism, can be formed. ROS include superoxide anion (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radicals (OH·). The last are reactive molecules and free radicals derived from molecular oxygen produced in mitochondria. Although they are needed at certain physiological levels to support cellular signaling and transduction, increased and uncontrolled levels of ROS are detrimental, causing pathological effects on cells/tissues such as proteins, lipids, and DNA [43–45]. Therefore, cells/tissues have antioxidant systems that neutralize excessive amounts of ROS. The antioxidant systems can be enzymatic and nonenzymatic based. The enzymatic antioxidant system contains several enzymes, including superoxide dismutase (SOD), catalase (CAT), glutathione peroxidases (GPxs), and

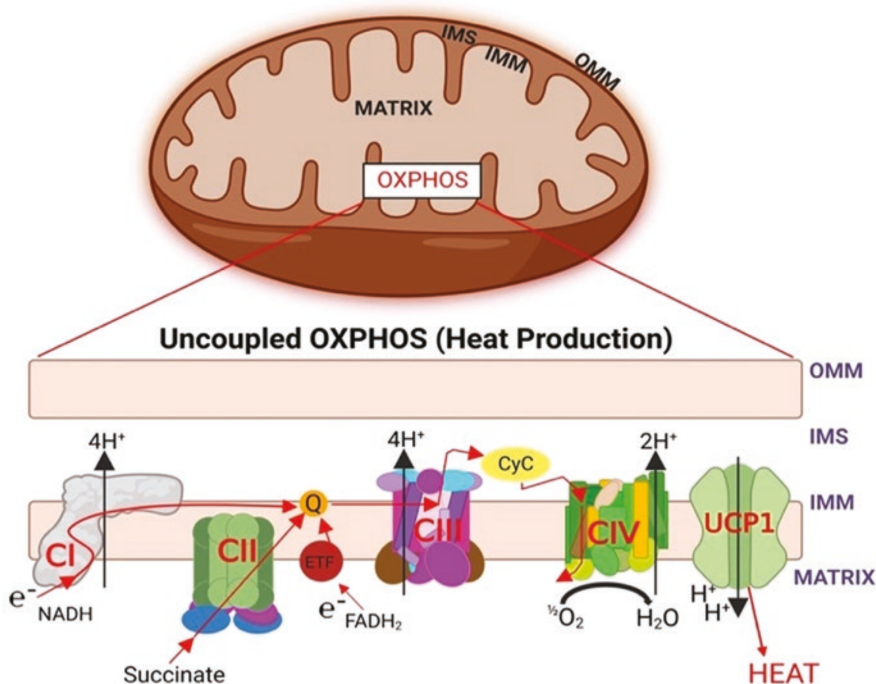

Fig. 4 Mitochondrial uncoupled OXPHOS pathways

thioredoxin (Trx). Nonenzymatic-based antioxidant systems include reduced glutathione (GSH), GSH precursors and reducers, such as N-acetylcysteine (NAC), and vitamins such as vitamins C, A, and E and their derivatives. Accordingly, the antioxidant systems are fundamental as a protective cellular mechanism against oxidative damage. Thus, impairments in the antioxidant mechanisms can lead to oxidative stress due to the imbalance between ROS and antioxidant systems. Oxidative stress can lead to pathological conditions, including MDs [46–49].

3 Mitochondrial Diseases

As described above, MDs can be developed because of defects in mitochondrial properties, including mitochondrial genome, metabolism, structure, and dynamics. MDs are heterogeneous and complex diseases sharing the common feature of mitochondrial dysfunction. MDs are mainly caused directly by inherited mutations in genes encoding mitochondrial OXPHOS proteins, regardless of whether genes in nDNA or mtDNA encode the proteins. MDs can also develop because of inherited alterations in non-OXPHOS mitochondrial proteins, which are essentially required for mitochondrial function. In addition, nongenetic factors such as environmental stressors, myotoxicity drugs, oxidative stress, and aging can progressively cause dysfunctional mitochondria leading to eventually MDs during a lifetime [50].

Examples of MDs developed by either genetic or nongenetic causes are mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS), chronic progressive external ophthalmoplegia (CPEO), neuropathy, ataxia, and retinitis pigmentosa (NARP), myoclonic epilepsy with ragged-red fibers (MERRF), Leber hereditary optic neuropathy (LHON), Kern-Sayre syndrome (KSS), Pearson marrow pancreas syndrome (PMS), mitochondrial neurogastrointestinal encephalopathy (MNGIE), Leigh syndrome, Alpers-Huttenlocher Syndrome, Barth Syndrome, Parkinson's disease, Alzheimer's disease, autism, Huntington's disease, amyotrophic lateral sclerosis, Wilson's disease, Charcot-Marie-Tooth type 2 K, hereditary spastic paraplegia 7 (SPG7), Friedreich's ataxia, schizophrenia, sepsis, cardiovascular diseases, cancers, diabetes, and metabolic syndromes [51–55]. Expectedly, new MDs are about to develop and continue to be discovered. The prevalence of MDs is not accurately recorded. It has been challenging to precisely determine the prevalence of MDs due to the following reasons: First, the identification of MDs accounted only for diagnostic patients who have undergone molecular testing for the monogenic MDs, which does not account for other unidentified and/or suspected individuals of MDs, especially for those who had mitochondrial defect without developing obvious symptoms of MD during their lifetimes. Second, one particular MD can show heterogenous phenotypes and symptoms in different patients, which makes the symptoms of that particular MD not firm. Third, MDs can develop from different factors, including genetic and nongenetic factors, and for nongenetic factors, there is a possibility that new factors might arise with time.

Thus, the current prevalence of MDs is probably underestimated. A better estimate of the true prevalence of MDs will lead to increased efforts to diagnose and effectively treat the diseases accurately. To do so, there have been increasing global efforts from the research and clinical community to develop clinical approaches for diagnosing patients with MDs, contributing to the correction of underestimated MD prevalence.

Currently, studies have reported that MDs observed in children account for approximately 5 individuals per 100,000 population [56]. Another study reported that the prevalence of MDs in adults is estimated to be one individual per 5000 [57]. MDs are currently more common than previously thought, requiring an early diagnosis to enhance the health outcomes of patients at an early stage and provide them with proper therapeutic interventions.

4 Diagnostic Tools for MDs

Various clinical approaches are utilized for diagnosing MDs, ranging from simple to sophisticated methods. Suspected MD patients undergo a series of examinations and tests. Examinations are extensive and include investigations of a patient's family medical history, clinical phenotypes, biochemical parameters, molecular genetic tests, tissue biopsies, and neurological abnormalities. All these examinations have usually been considered in diagnosing MDs because they give a comprehensive picture of pathological status at distinct levels. Consensus-based recommendations written by the Mitochondrial Medicine Society mention the optimal diagnostic tools that can be used for MDs involving biochemical tests, genetic analyses, tissue biopsy examinations, and neuroimaging as they explained the purposes of each approach to verify the diagnosis of MDs [58].

Following the initial investigation of the medical family history and clinical phenotypes of the suspected MD patients, biochemical tests are usually performed on biological fluids such as blood, urine, and CFS samples to measure the levels of certain metabolites commonly disrupted in MDs, including pyruvate, lactate, acylcarnitines, ketone bodies, amino acids, and organic acids [58]. Lactate levels and the ratio of lactate/pyruvate are commonly increased when mitochondria are defective as pyruvate utilization by the mitochondria decreases. Moreover, ketone bodies are often perturbed in MD patients. In addition, there are alterations in the level of amino acids and organic acids in MD patients acting as an indication of mitochondria dysfunction [59]. All the biochemical analyses mentioned above are not specific and exclusive to MDs because these alterations are a common feature in most MDs and also could be seen in non-MDs.

Thus, supportive and specific analyses are needed to perform in addition to the biochemical tests to improve the diagnostic investigation of MDs, involving neurological imaging of the central nervous system and collecting tissue biopsies from MD patients for enzymatic analyses of OXPHOS and other mitochondrial proteins. These methods show their great ability to inform about alterations in mitochondria

in MDs cases; however, these methods have certain pitfalls as neurological imaging detect abnormalities in the brain and nerve system developed only in particular, but not all, MD patients [58]. In addition, enzymatic assays provide valuable functional analyses of mitochondria, although they require an invasive tissue biopsy procedure.

In addition, genetic testing has been extensively included in the diagnostic criteria of MDs since most MDs are inherited diseases. Genetic testing discovers the genetic mutations leading to MDs. The implication of genetic testing has been done for MDs through testing for common pathogenic variants in nDNA or mtDNA genes associated with MDs by sequencing the whole mtDNA or nDNA via whole genome sequencing (WGS) and/or whole exome sequencing (WES). MDs that were identified by multiple genetic testing approaches include those MDs developed by various pathogenic variants in genes coding for the following proteins, NADH dehydrogenase, pyruvate dehydrogenase complex component X (PDHX), ethylmalonic encephalopathy 1 protein (ETHE1), mitochondrial inner membrane protein (MPV17), mitochondrial carnitine/acylcarnitine carrier protein (SLC25A20), mitochondrial fission factor (MFF), F-Box and leucine-rich repeat protein 4 (FBXL4), elaC ribonuclease Z 2 (ELAC2), protein PET100 (PET100), iron-sulfur cluster assembly 2 (ISCA2), mitochondrial-processing peptidase subunit alpha (PMPCA), metal cation symporter ZIP8 (SLC39A8), mitochondrial coenzyme A transporter (SLC25A42), ATP-dependent zinc metalloprotease (YME1L1), mitochondrial intermediate peptidase (MIPEP), mitochondrial calcium uptake protein 2 (MICU2), cytochrome c oxidase subunit 5A (COX5A), ubiquinone biosynthesis methyltransferase COQ5 (COQ5), and nundid hydrolase 2 (NUDT2) [60–74].

Regardless of the substantial implications of genetic approaches in identifying MDs, the genetic approaches have certain diagnostic downsides. Specifically, these genetic tests require reading out an unlimited number of genes, and the roles of many genes are still not fully understood. Finding mutations in metabolism-related genes that are functionally unknown, called variants of uncertain significance (VUS), could be misleading in the context of MDs because it is not fully understood if these unknown mutations affect health. Furthermore, mutations identified by genetic tests might be secondary findings of other diseases not discovered or developed yet and unrelated to MDs of interest, which is considered a false-positive discovery. In addition, genetic testing cannot be used for those MDs that are developed by nongenetic factors over a patient's lifetime [75, 76].

Consequently, based on the previous issues associated with the genetic approaches, there is still a definitive need to perform alternative functional analyses of suspected VUS found in metabolism-related genes to ensure whether these genetic mutations truly cause MDs. Also, these alternative functional analyses could help examine the pathogenic effects of nongenetic factors leading to MDs. Since mitochondria work as a hub of metabolism and as metabolite producers, metabolomics, which detect small metabolites ≤1500 Daltons through high-throughput mass spectrometry, can predict mitochondrial status in health and disease. Metabolomics has recently gained great attention because of its high potential to be used as a diagnostic tool for all types of MDs, either genetic or nongenetic MDs. The strengths of

using metabolomics as a diagnostic tool are as follows: First, metabolomics measures intermediates or products of ongoing metabolic pathways, and their measurements directly reflect the snapshot readouts of physiological conditions, which make metabolomics more real functional analyses of metabolism compared to other omic approaches. Thus, if there are any metabolic perturbations, as seen in MDs, they could be detected by the metabolomic approach. Second, metabolomics shows high sensitivity with extraordinary capabilities for identifying distinct metabolic biosignatures/profiles of MDs, reflecting the disrupted metabolic pathways in MD conditions. Third, metabolomic data analysis is easier than other omic data because of the limited number of metabolites that currently can be accurately measured compared to the huge numbers of genes or proteins that can be accurately measured. Fourth, compared to other approaches, metabolomics is not expensive, requires relatively little time, and provides comprehensive biological measurements. All these reasons encourage researchers and scientists to focus on applying metabolomics to MDs.

5 Metabolomics of MDs

Several metabolomic studies have been conducted to ultimately identify diagnostic biomarkers/biosignatures of MDs using different biological samples with various mass spectrometry (MS) machines. Herein, we mentioned examples of these metabolomic studies conducted for MDs (Table 1). For instance, a study performed gas chromatography-coupled mass spectrometry (GC-MS)-based metabolomic analyses on plasma samples and skeletal muscle fibers collected from patients diagnosed with mitochondrial myopathy/progressive external ophthalmoplegia disease showing elevated levels of certain metabolites such as cystathionine, glutamic acid, serine, and arachidyl carnitine compared to the samples collected from the controls [77]. Also, GC-MS-based metabolomic analyses of urine samples collected from children diagnosed with deficiencies in OXPHOS proteins revealed increased levels of organic acids, including fumaric acid, glutaric acid, lactic acid, malic acid, and others, compared to the control children [78]. Another study used a metabolomic approach as a pre-screening tool to identify metabolic biosignatures of MD patients who need to undergo tissue biopsies. In this way, MD patients who do not need a tissue biopsy do not need to undergo the invasive procedure [79]. Their study would help to select the proper MD patient who needs to undergo tissue biopsies. Smuts et al. identified biosignatures in urine samples collected from patients with deficient OXPHOS to be used as indicators before the tissue biopsies. They performed untargeted nuclear magnetic resonance (NMR). They targeted GC-MS on OXPHOS-deficient patients' urine samples. They found six organic acids (lactic, succinic, 2-hydroxybutyric, 3-hydroxybutyric, 3-hydroxyisovaleric, and 3-hydroxy-3-methylglutaric acids), six amino acids (alanine, glycine, glutamic acid, serine, tyrosine, and α-aminoadipic acid), and creatine as biosignatures of these patients that need to undergo tissue biopsies for the validation of MDs [79].

Table 1 Examples of metabolomic studies done to screen metabolomic biosignatures of MDs by using various biological samples and metabolomic approaches

MD name	Biological specimen	Major findings	Technique	Ref.
Mitochondrial myopathy/ progressive external ophthalmoplegia disease	Plasma samples and skeletal muscle fibers	Elevated levels of certain metabolites such as cystathionine, glutamic acid, serine, and arachidoyl-carnitine	GC-MS	Nikkanen et al. [77]
Deficiencies in OXPHOS proteins	Urine samples	Increased levels of organic acids, including fumaric acid, glutaric acid, lactic acid, malic acid, and others	GC-MS	Reinecke et al. [78]
Deficient OXPHOS in muscles	Urine samples	Altered six organic acids (lactic, succinic, 2-hydroxybutyric, 3-hydroxyisobutyric, 3-hydroxyisovaleric and 3-hydroxy-3-methylglutaric acids), six amino acids (alanine, glycine, glutamic acid, serine, tyrosine, and α-aminoadipic acid), and creatine	Untargeted NMR and targeted GC-MS	Smuts et al. [79]
French-Canadian Leigh syndrome	Urine and plasma samples	Altered 45 metabolite markers, including alanine, asparagine, ketones, acylcarnitines, succinate, kynurenine, lactate, and pyruvates	Targeted GC-MS and LC-MS	Thompson Legault et al. [80]
MELAS	Plasma samples	Elevated levels of pyruvate, lactate, malate, alanine, α-hydroxybutyrate, N-lactoyl-amino acids, β-hydroxy acylcarnitines, and β-hydroxy fatty acids	Targeted and untargeted MS	Sharma et al. [81]
MELAS and MIDD with renal dysfunction	Urine samples	Lower levels of 4-cresyl sulfate, S-methyl-cysteine-sulfoxide, N-methylnicotinamide, and hippuric acid	NMR spectroscopy	Hall et al. [82]
LHON	Fibroblasts	Decreases in amino acids, spermidine, putrescine, isovaleryl-carnitine, propionyl-carnitine, and five sphingomyelin species but increases in ten phosphatidylcholine species	Targeted LC-MS	Chao de la Barca et al. [83]

(continued)

Table 1 (continued)

MD name	Biological specimen	Major findings	Technique	Ref.
Barth syndrome	Plasma samples	Perturbations in creatinine, fatty acids, methionine, and proline	NMR spectroscopy	Sandlers et al. [84]
		Alterations in acylcarnitines, bigeneric amines, PC/lysoPC, and amino acids	Targeted LC-MS	
KSS	Urine samples	Increased levels of pyruvate, fumarate, and 3-hydroxybutyrate	Targeted GC-MS	Semeraro et al. [85]
PMS	Urine samples	Increased levels of lactate, 3-hydroxybutyrate, 3-hydroxyisobutyrate, fumarate, pyruvate, 2-hydroxybutyrate, 2-methyl-2,3-dihydroxybutyrate, 3-methylglutarate, 2-ethyl-3-hydroxypropionate, 3-methylglutaconate, malate, and tiglylglycine		
MELAS	Urine samples	Increases in caproic/caprylic acid, 2-hydroxyglutaric acid, butyric/valeric/2-hydroxybutyric/3-methyl-2-oxovaleric acid, 4-pentenoic acid, acetylcarnitine, propionylcarnitine, taurine, acetic acid but decreased metabolites pyruvic acid, glycerol, carbamate, 2,5-furandicarboxylic acid, fumaric acid, pseudouridine, glycolic acid, and arabinose	Targeted LC-MS/MS and untargeted GC–MS and NMR spectroscopy	Esterhuizen et al. [86]
MIDD	Urine samples	Increases in myoinositol, 2-hydroxyglutaric acid, 4-pentenoic acid, glucuronic acid, 2-hydroxyisovaleric acid, glucose, 2-ethylhydracrylic acid, and 3-hydroxyisobutyric acid. Metabolites' decrease includes glycolic acid, sarcosine, 1,2-ethandiol, 3-methylphenol, 2,5-furandicarboxylic acid, homocysteine, arabinose, and pseudouridine		
Myopathy	Urine samples	Increases in 2-hydroxyglutaric acid, 4-pentenoic acid, 3-methylphenol, 2-ethylhydracrylic acid, and creatine but decreased glycolic acid		

Table 1 (continued)

MD name	Biological specimen	Major findings	Technique	Ref.
Parkinson's disease	Plasma and CSF samples	Altered pattern of metabolites involved in the metabolism of glycerophospholipid, sphingolipid, acylcarnitine, and amino acids	Untargeted LC-MS and MS/MS	Stoessel et al. [87]
Parkinson's disease	Urine samples	Altered metabolites related to tryptophan and tyrosine metabolism	Untargeted LC-MS	Luan et al. [88]
Alzheimer's disease	CSF samples	Altered patterns of choline, dimethylarginine, arginine, valine, proline, serine, histidine, creatine, carnitine, and suberylglycine	Untargeted capillary electrophoresis mass spectrometry (CE-MS)	Ibáñez et al. [89]
Medium-chain acyl-coenzyme A dehydrogenase deficiency disease	Dried blood spots	Altered levels of certain amino acids and acylcarnitines	Targeted LC-MS	Scolamiero et al. [90]
Long-chain acyl-coenzyme A dehydrogenase deficiency disease	Dried blood spots and serum samples	Decreases in lysine, valine, glycerol, and niacinamide but increases in glutamine succinic acid and guanosine	Targeted LC-MS/MS	Jacob et al. [91]
Type 2 diabetes (T2D)	Urine samples	Decreases in t3-hydroxyundecanoyl-carnitine	Targeted LC-MS	Salihovic et al. [92]
Insulin resistance	Serum samples	Insulin resistance altered amino acids such as amino acids (Asn, Gln, and his), methionine (met) sulfoxide, 2-methyl-3-hydroxy-5--formylpyridine-4-carboxylate, serotonin, L-2-amino-3-oxobutanoic acid, and 4,6-dihydroxyquinoline	Chemical isotope labeling (CIL) liquid chromatography-mass spectrometry (LC-MS)	Gu et al. [93]
Type 2 diabetes		Distinct amino acids, amino acid metabolites, and dipeptides for T2D		

A cohort of French-Canadian Leigh syndrome patients with mutations in the LRPPRC gene causing defects in one of the OXPHOS proteins was involved in a metabolomic study. Urine and plasma samples were collected from the Leigh syndrome patients and were analyzed using targeted GC-MS and liquid chromatography-coupled mass spectrometry (LC-MS). As a result, they found 45 outstanding metabolic biomarkers, including alanine, asparagine, ketones, acylcarnitines, succinate, kynurenine, lactate, and pyruvates, altered compared to controls (Thompson [80]). Additionally, another research group recruited mitochondrial encephalomyopathy lactic acidosis and stroke-like episode (MELAS) syndrome and maternally inherited diabetes and deafness (MIDD) patients with renal dysfunction for identification of urinary metabolic markers via NMR spectroscopy-based metabolomics and found that these patients distinguished from controls by having lower levels of 4-cresyl sulfate, S-methyl-cysteine-sulfoxide, N-methylnicotinamide, and hippuric acid [82]. Recently, Sharma et al. performed metabolomic analyses on plasma samples collected from MELAS and controls by using targeted and untargeted MS, revealing that elevated levels of pyruvate, lactate, malate, alanine, α-hydroxybutyrate, N-lactoyl-amino acids, β-hydroxy acylcarnitines, and β-hydroxy fatty acids [81]. Another example of metabolomic study is those conducted for Barth syndrome. They identified dysregulated metabolic markers and pathways underlying Barth syndrome using human plasma samples showing perturbations in creatinine, fatty acids, methionine, and proline detected by NMR. At the same time, LC-MS revealed metabolic alterations found in acylcarnitines, bigeneric amines, PC/lysoPC, and amino acids in patients compared to controls [84].

Furthermore, a targeted LC-MS-based metabolomic approach was applied to fibroblasts taken from LHON patients to uncover metabolites affected in this disease used as biomarkers, including decreases in amino acids, spermidine, putrescine, isovaleryl-carnitine, propionyl-carnitine, and five sphingomyelin species but increases in ten phosphatidylcholine species [83]. Also, metabolomic studies of PMS and KSS were performed recently using a targeted GC-MS approach on urine samples collected from PMS and KSS patients. Semeraro et al.'s study revealed that abnormal alterations in urinary organic acids were detected in both PMS and KSS. Still, the alterations are more pronounced in PMS patients than in KSS. They found that urine samples from KSS had increased levels of pyruvate, fumarate, and 3-hydroxybutyrate, while urinary metabolites detected in PMS patients indicated elevated levels in lactate, 3-hydroxybutyrate, 3-hydroxyisobutyrate, fumarate, pyruvate, 2-hydroxybutyrate, 2-methyl-2,3-dihydroxybutyrate, 3-methylglutarate, 2-ethyl-3-hydroxypropionate, 3-methylglutaconate, malate, and tiglylglycine [85].

Very recently, Esterhuizen et al. attempted to clinically distinguish between three types of MDs, including MELAS, MIDD, and myopathy, by identifying disease-specific metabolic profiles. They used combined multi-metabolomic approaches of targeted LC-MS/MS, untargeted GC-MS, and NMR spectroscopy to comprehensively investigate metabolites in urine samples collected from MELAS, MIDD, and myopathy patients compared to individual disease-matched controls. Although they found some metabolic similarities in identified metabolites shared in the three diseases, their metabolomic data shows that each of the three diseases has certain

unique metabolites compared to their controls. Compared to controls, MELAS patients showed increased caproic/caprylic acid, 2-hydroxyglutaric acid, butyric/valeric/2-hydroxybutyric/3-methyl-2-oxovaleric acid, 4-pentenoic acid, acetylcarnitine, propionylcarnitine, taurine, and acetic acid but decreased metabolites pyruvic acid, glycerol, carbamate, 2,5-furandicarboxylic acid, fumaric acid, pseudouridine, glycolic acid, and arabinose. Furthermore, MIDD patients showed increased myoinositol, 2-hydroxyglutaric acid, 4-pentenoic acid, glucuronic acid, 2-hydroxyisovaleric acid, glucose, 2-ethylhydracrylic acid, 3-hydroxyisobutyric acid although decreased metabolites were shown including glycolic acid, sarcosine, 1,2-ethandiol, 3-methylphenol, 2,5-furandicarboxylic acid, homocysteine, arabinose, pseudouridine. Lastly, they showed that myopathy patients revealed increased 2-hydroxyglutaric acid, 4-pentenoic acid, 3-methylphenol, 2-ethylhydracrylic acid, and creatine but decreased glycolic acid [86].

In addition to the previous studies, Stoessel et al. focused on the metabolic changes in Parkinson's disease, in which they performed untargeted LC/MS and MS/MS analysis to profile metabolic changes in plasma and CSF samples collected from Parkinson's disease patients. They uncovered a perturbed pattern of metabolites involved in the metabolism of glycerophospholipid, sphingolipid, acylcarnitine, and amino acids as Parkinson's disease biosignatures [87]. In concordance with the previous Parkinson's disease study, another research group studied Parkinson's disease and discovered alterations in metabolites related to tryptophan and tyrosine metabolism and other pathways using untargeted LC-MS approaches applied to urine samples from Parkinson's disease patients [88]. Moreover, Ibáñez et al. used untargeted metabolomic profiling based on capillary electrophoresis-coupled mass spectrometry (CE-MS) to identify Alzheimer's disease biosignatures in CSF samples collected from Alzheimer's disease patients for the prediction of the disease and its progression at distinct stages. They found choline, dimethylarginine, arginine, valine, proline, serine, histidine, creatine, carnitine, and suberylglycine as potential Alzheimer's disease progression markers [89]. Also, Scolamiero et al. analyzed dried blood spots (DBS) from patients diagnosed with medium chain acyl-coenzyme A dehydrogenase deficiency disease via targeted LC-MS to identify metabolic profiling associated with the disease condition. Their results showed that the level of certain acylcarnitine (C2, C6, C8, and C10) was perturbed distinctively in these affected patients compared to the controls, potentially used as diagnostic markers [90]. In addition, another research group performed a metabolomic study on patients diagnosed with long-chain acyl-coenzyme A dehydrogenase deficiency disease by performing targeted LC-MS analyses on dried blood spots and serum samples collected from the patients showing that there were alterations in metabolites in patients' samples compared to the controls. Examples of altered metabolites are decreases in lysine, valine, glycerol, and niacinamide but increases in glutamine succinic acid and guanosine [91].

Furthermore, diabetes was extensively studied in different biological samples with different MS approaches; thus, example studies are mentioned. A recent study conducted nontargeted LC-MS-based metabolomics to profile urine metabolites in patients diagnosed with type 2 diabetes (T2D), resulting in the discovery of a

promising biomarker t3-hydroxyundecanoyl-carnitine, which was decreased with T2D [92]. In addition, another study done by another research group revealed metabolite biosignatures related to insulin resistance and T2D. This study collected serum samples from normal, obese T2D participants for metabolic profiling using chemical isotope labeling (CIL) and LC-MS. Their results showed that certain biomarkers associated with insulin resistance, such as amino acids (asparagine, glutamine, histidine), methionine sulfoxide, 2-methyl-3-hydroxy-5-formylpyridine-4-carboxylate, serotonin, L-2-amino-3-oxobutanoic acid, and 4,6-dihydroxyquinoline and other biomarkers for T2D, including amino acids, amino acid metabolites, and dipeptides [93].

Last but not least, there is a growing body of recent studies demonstrating the application of metabolomic technologies in sepsis [94–96], cardiovascular diseases [97, 98], and cancers [99–102].

6 Conclusion

The application of metabolomics can be promising for the diagnosis and prognosis of MDs in humans. However, it has been reported that metabolomic analyses have certain challenges and limitations that need to be considered. For instance, metabolomic analyses can be conducted by multiple mass spectrometry approaches, including NMR, GC-MS, and LC-MS. All of them are used for various purposes due to their different functionalities, metabolite coverage, and experimental design. Each approach has advantages and disadvantages. For example, NMR is suitable for significant volume samples, is nondestructive to samples, does not require sample preparations, and gives reproducible data. However, NMR is not as sensitive as other techniques, such as GC-MS and LC-MS, which can detect a large range of metabolites with very tiny amounts of sample. GC-MS and LC-MS function based on the polarity of metabolites, as GC-MS is mainly used more for nonpolar metabolites. Thus LC-MS is the preferable approach for polar metabolites.

Nevertheless, MS approaches have certain drawbacks, such as lower reproducibility, higher-cost expensive techniques, and time-consuming sample preparation steps. Thus, no single approach can comprehensively detect all metabolites with various chemical and physical properties in biological samples within one analytical run. Other limitations are related to metabolomic data preprocessing and spectral library databases. Examples of issues during metabolomic data processing missing data can be seen, and that is due to several reasons, such as the concentration of metabolites being either lower or higher than the detection limit of the machine. Another limitation of metabolomics is the availability of metabolite spectral libraries for data analyses and interpretation. NMR and LC-MS approaches have limited libraries compared to GC-MS libraries.

Many of the challenges associated with metabolomics can be overcome by combining various approaches to detect and cover a comprehensive range of metabolites. Furthermore, metabolomics shows the powerful capability and potential in

diagnosing MDs as it represents real perturbations associated with MDs. Nonetheless, it is best to integrate metabolomic diagnostic data with other diagnostic approaches (i.e., patient and family histories, physical examinations, clinical biochemical and histological tests, and enzymatic assays). Finally, research in this area and the corresponding field development will improve with the involvement of larger patient cohorts. Increased global collaborations between many clinical communities will strengthen and validate the finding of metabolic biosignatures of human MDs and improve diagnostic platforms.

References

1. Elston T, Wang H, Oster G. Energy transduction in ATP synthase. Nature. 1998;391(6666):510–3. https://doi.org/10.1038/35185.
2. Veltri KL, Espiritu M, Singh G. Distinct genomic copy number in mitochondria of different mammalian organs. J Cell Physiol. 1990;143(1):160–4. https://doi.org/10.1002/jcp.1041430122.
3. Spiegelman BM. Transcriptional control of mitochondrial energy metabolism through the PGC1 coactivators. In: Mitochondrial biology: new perspectives. Chichester: John Wiley & Sons; 2007. p. 60–9. https://doi.org/10.1002/9780470725207.ch5.
4. Tzameli I. The evolving role of mitochondria in metabolism. Trends Endocrinol Metab. 2012;23(9):417–9. https://doi.org/10.1016/j.tem.2012.07.008.
5. Friedman JR, Nunnari J. Mitochondrial form and function. Nature. 2014;505(7483):335–43. https://doi.org/10.1038/nature12985.
6. Herst PM, Rowe MR, Carson GM, Berridge MV. Functional mitochondria in health and disease. Front Endocrinol. 2017;8:296. https://doi.org/10.3389/fendo.2017.00296.
7. McBride HM, Neuspiel M, Wasiak S. Mitochondria: more than just a powerhouse. Curr Biol. 2006;16(14):R551–60. https://doi.org/10.1016/j.cub.2006.06.054.
8. Shaw JM, Nunnari J. Mitochondrial dynamics and division in budding yeast. Trends Cell Biol. 2002;12(4):178–84. https://doi.org/10.1016/S0962-8924(01)02246-2.
9. Trushina E. A shape shifting organelle: unusual mitochondrial phenotype determined with three-dimensional electron microscopy reconstruction. Neural Regen Res. 2016;11(6):900–1. https://doi.org/10.4103/1673-5374.184477.
10. Gohil VM, Greenberg ML. Mitochondrial membrane biogenesis: phospholipids and proteins go hand in hand. J Cell Biol. 2009;184(4):469–72. https://doi.org/10.1083/jcb.200901127.
11. Kühlbrandt W. Structure and function of mitochondrial membrane protein complexes. BMC Biol. 2015;13(1):89. https://doi.org/10.1186/s12915-015-0201-x.
12. Bowsher CG, Tobin AK. Compartmentation of metabolism within mitochondria and plastids. J Exp Bot. 2001;52(356):513–27. https://doi.org/10.1093/jexbot/52.356.513.
13. Glancy B, Hartnell LM, Malide D, Yu Z-X, Combs CA, Connelly PS, Subramaniam S, Balaban RS. Mitochondrial reticulum for cellular energy distribution in muscle. Nature. 2015;523(7562):617–20. https://doi.org/10.1038/nature14614.
14. Glancy B, Kim Y, Katti P, Willingham TB. The functional impact of mitochondrial structure across subcellular scales. Front Physiol. 2020;11:541040. https://doi.org/10.3389/fphys.2020.541040.
15. Yoon Y. Sharpening the scissors. Cell Biochem Biophys. 2004;41(2):193–205. https://doi.org/10.1385/CBB:41:2:193.
16. Ding W-X, Li M, Biazik JM, Morgan DG, Guo F, Ni H-M, Goheen M, Eskelinen E-L, Yin X-M. Electron microscopic analysis of a spherical mitochondrial structure*. J Biol Chem. 2012;287(50):42373–8. https://doi.org/10.1074/jbc.M112.413674.

17. Navaratnarajah T, Anand R, Reichert AS, Distelmaier F. The relevance of mitochondrial morphology for human disease. Int J Biochem Cell Biol. 2021;134:105951. https://doi.org/10.1016/j.biocel.2021.105951.
18. Busch KB, Kowald A, Spelbrink JN. Quality matters: how does mitochondrial network dynamics and quality control impact on mtDNA integrity? Philos Trans R Soc Lond B Biol Sci. 2014;369(1646):20130442. https://doi.org/10.1098/rstb.2013.0442.
19. Mishra P, Chan DC. Metabolic regulation of mitochondrial dynamics. J Cell Biol. 2016;212(4):379–87. https://doi.org/10.1083/jcb.201511036.
20. Ashrafi G, Schwarz TL. The pathways of mitophagy for quality control and clearance of mitochondria. Cell Death Differ. 2013;20(1):31–42. https://doi.org/10.1038/cdd.2012.81.
21. Losón OC, Song Z, Chen H, Chan DC. Fis1, Mff, MiD49, and MiD51 mediate Drp1 recruitment in mitochondrial fission. Mol Biol Cell. 2013;24(5):659–67. https://doi.org/10.1091/mbc.E12-10-0721.
22. Otsuga D, Keegan BR, Brisch E, Thatcher JW, Hermann GJ, Bleazard W, Shaw JM. The dynamin-related GTPase, Dnm1p, controls mitochondrial morphology in yeast. J Cell Biol. 1998;143(2):333–49.
23. Chen H, Detmer SA, Ewald AJ, Griffin EE, Fraser SE, Chan DC. Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. J Cell Biol. 2003;160(2):189–200. https://doi.org/10.1083/jcb.200211046.
24. Song Z, Ghochani M, McCaffery JM, Frey TG, Chan DC. Mitofusins and OPA1 mediate sequential steps in mitochondrial membrane fusion. Mol Biol Cell. 2009;20(15):3525–32. https://doi.org/10.1091/mbc.e09-03-0252.
25. Gustafsson CM, Falkenberg M, Larsson N-G. Maintenance and expression of mammalian mitochondrial DNA. Annu Rev Biochem. 2016;85:133. https://doi.org/10.1146/annurev-biochem-060815-014402.
26. Wai T, Ao A, Zhang X, Cyr D, Dufort D, Shoubridge EA. The role of mitochondrial DNA copy number in mammalian Fertility1. Biol Reprod. 2010;83(1):52–62. https://doi.org/10.1095/biolreprod.109.080887.
27. Rath S, Sharma R, Gupta R, Ast T, Chan C, Durham TJ, Goodman RP, Grabarek Z, Haas ME, Hung WHW, Joshi PR, Jourdain AA, Kim SH, Kotrys AV, Lam SS, McCoy JG, Meisel JD, Miranda M, Panda A, Mootha VK. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res. 2021;49(D1):D1541–7. https://doi.org/10.1093/nar/gkaa1011.
28. D'Souza AR, Minczuk M. Mitochondrial transcription and translation: overview. Essays Biochem. 2018;62(3):309–20. https://doi.org/10.1042/EBC20170102.
29. Demine S, Reddy N, Renard P, Raes M, Arnould T. Unraveling biochemical pathways affected by mitochondrial dysfunctions using metabolomic approaches. Meta. 2014;4(3):831–78. https://doi.org/10.3390/metabo4030831.
30. Wallace DC, Fan W, Procaccio V. Mitochondrial energetics and therapeutics. Annu Rev Pathol. 2010;5:297–348. https://doi.org/10.1146/annurev.pathol.4.110807.092314.
31. Dimroth P, Kaim G, Matthey U. Crucial role of the membrane potential for ATP synthesis by F(1)F(o) ATP synthases. J Exp Biol. 2000;203(1):51–9. https://doi.org/10.1242/jeb.203.1.51.
32. Mitchell P, Moyle J. Chemiosmotic hypothesis of oxidative phosphorylation. Nature. 1967;213(5072):137. https://doi.org/10.1038/213137a0.
33. Watmough NJ, Frerman FE. The electron transfer flavoprotein: ubiquinone oxidoreductases. Biochim Biophys Acta. 2010;1797(12):1910–6. https://doi.org/10.1016/j.bbabio.2010.10.007.
34. Divakaruni AS, Brand MD. The regulation and physiology of mitochondrial proton leak. Physiology (Bethesda). 2011;26(3):192–205. https://doi.org/10.1152/physiol.00046.2010.
35. Cannon B, Nedergaard J. Brown adipose tissue: function and physiological significance. Physiol Rev. 2004;84(1):277–359. https://doi.org/10.1152/physrev.00015.2003.

36. Mailloux RJ, Harper M-E. Uncoupling proteins and the control of mitochondrial reactive oxygen species production. Free Radic Biol Med. 2011;51(6):1106–15. https://doi.org/10.1016/j.freeradbiomed.2011.06.022.
37. Brand MD, Esteves TC. Physiological functions of the mitochondrial uncoupling proteins UCP2 and UCP3. Cell Metab. 2005;2(2):85–93. https://doi.org/10.1016/j.cmet.2005.06.002.
38. Nicholls DG. Mitochondrial proton leaks and uncoupling proteins. Biochim Biophys Acta. 2021;1862(7):148428. https://doi.org/10.1016/j.bbabio.2021.148428.
39. Bertholet AM, Chouchani ET, Kazak L, Angelin A, Fedorenko A, Long JZ, Vidoni S, Garrity R, Cho J, Terada N, Wallace DC, Spiegelman BM, Kirichok Y. H+ transport is an integral function of the mitochondrial ADP/ATP carrier. Nature. 2019;571(7766):515–20. https://doi.org/10.1038/s41586-019-1400-3.
40. Bevilacqua L, Seifert EL, Estey C, Gerrits MF, Harper M-E. Absence of uncoupling protein-3 leads to greater activation of an adenine nucleotide translocase-mediated proton conductance in skeletal muscle mitochondria from calorie restricted mice. Biochim Biophys Acta. 2010;1797(8):1389–97. https://doi.org/10.1016/j.bbabio.2010.02.018.
41. Echtay KS, Esteves TC, Pakay JL, Jekabsons MB, Lambert AJ, Portero-Otín M, Pamplona R, Vidal-Puig AJ, Wang S, Roebuck SJ, Brand MD. A signalling role for 4-hydroxy-2-nonenal in regulation of mitochondrial uncoupling. EMBO J. 2003;22(16):4103–10. https://doi.org/10.1093/emboj/cdg412.
42. Klumpe I, Savvatis K, Westermann D, Tschöpe C, Rauch U, Landmesser U, Schultheiss H-P, Dörner A. Transgenic overexpression of adenine nucleotide translocase 1 protects ischemic hearts against oxidative stress. J Mol Med (Berl). 2016;94(6):645–53. https://doi.org/10.1007/s00109-016-1413-4.
43. Ayala A, Muñoz MF, Argüelles S. Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-Hydroxy-2-Nonenal. Oxidative Med Cell Longev. 2014;2014:e360438. https://doi.org/10.1155/2014/360438.
44. Cadet J, Davies KJA, Medeiros MH, Di Mascio P, Wagner JR. Formation and repair of oxidatively generated damage in cellular DNA. Free Radic Biol Med. 2017;107:13–34. https://doi.org/10.1016/j.freeradbiomed.2016.12.049.
45. Davies MJ. Singlet oxygen-mediated damage to proteins and its consequences. Biochem Biophys Res Commun. 2003;305(3):761–70. https://doi.org/10.1016/S0006-291X(03)00817-9.
46. Christofidou-Solomidou M, Muzykantov VR. Antioxidant strategies in respiratory medicine. Treat Respir Med. 2006;5(1):47–78. https://doi.org/10.2165/00151829-200605010-00004.
47. Holzerová E, Prokisch H. Mitochondria: much ado about nothing? How dangerous is reactive oxygen species production? Int J Biochem Cell Biol. 2015;63:16–20. https://doi.org/10.1016/j.biocel.2015.01.021.
48. Murphy MP. How mitochondria produce reactive oxygen species. Biochem J. 2009;417(Pt 1):1–13. https://doi.org/10.1042/BJ20081386.
49. Shan Y, Schoenfeld RA, Hayashi G, Napoli E, Akiyama T, Iodi Carstens M, Carstens EE, Pook MA, Cortopassi GA. Frataxin deficiency leads to defects in expression of antioxidants and Nrf2 expression in dorsal root ganglia of the Friedreich's ataxia YG8R mouse model. Antioxid Redox Signal. 2013;19(13):1481–93. https://doi.org/10.1089/ars.2012.4537.
50. Sulaiman SA, Mohd Rani Z, Mohd Radin FZ, Abdul Murad NA. Advancement in the diagnosis of mitochondrial diseases. J Transl Genet Genom. 2020;4(3):159–87. https://doi.org/10.20517/jtgg.2020.27.
51. Chinnery PF. Primary mitochondrial disorders overview. In: Adam MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJ, Gripp KW, Amemiya A, editors. GeneReviews®. Seattle: University of Washington; 1993; http://www.ncbi.nlm.nih.gov/books/NBK1224/.
52. Khajuria K, Khajuria V, Sawhney V. SECONDARY MITOCHONDRIAL DYSFUNCTION. Int J Pharm Pharm Sci. 2021;14:14–9. https://doi.org/10.22159/ijpps.2021v13i3.40335.

53. Niyazov DM, Kahler SG, Frye RE. Primary mitochondrial disease and secondary mitochondrial dysfunction: importance of distinction for diagnosis and treatment. Mol Syndromol. 2016;7(3):122–37. https://doi.org/10.1159/000446586.
54. Rötig A, Cormier V, Blanche S, Bonnefont JP, Ledeist F, Romero N, Schmitz J, Rustin P, Fischer A, Saudubray JM. Pearson's marrow-pancreas syndrome. A multisystem mitochondrial disorder in infancy. J Clin Invest. 1990;86(5):1601–8. https://doi.org/10.1172/JCI114881.
55. Schlieben LD, Prokisch H. The dimensions of primary mitochondrial disorders. Front Cell Dev Biol. 2020;8:1441. https://doi.org/10.3389/fcell.2020.600079.
56. Skladal D, Halliday J, Thorburn DR. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. Brain. 2003;126(8):1905–12. https://doi.org/10.1093/brain/awg170.
57. Gorman GS, Schaefer AM, Ng Y, Gomez N, Blakely EL, Alston CL, Feeney C, Horvath R, Yu-Wai-Man P, Chinnery PF, Taylor RW, Turnbull DM, McFarland R. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. Ann Neurol. 2015;77(5):753–9. https://doi.org/10.1002/ana.24362.
58. Parikh S, Goldstein A, Koenig MK, Scaglia F, Enns GM, Saneto R, Anselm I, Cohen BH, Falk MJ, Greene C, Gropman AL, Haas R, Hirano M, Morgan P, Sims K, Tarnopolsky M, Van Hove JLK, Wolfe L, DiMauro S. Diagnosis and management of mitochondrial disease: a consensus statement from the mitochondrial medicine society. Genet Med. 2015;17(9):689–701. https://doi.org/10.1038/gim.2014.177.
59. Rodenburg RJT. Biochemical diagnosis of mitochondrial disorders. J Inherit Metab Dis. 2011;34(2):283–92. https://doi.org/10.1007/s10545-010-9081-y.
60. Al Aqeel AI, Rashid MS, Pn Ruiter J, Ijlst L, Ja Wanders R. A novel molecular defect of the carnitine acylcarnitine translocase gene in a Saudi patient. Clin Genet. 2003;64(2):163–5. https://doi.org/10.1034/j.1399-0004.2003.00117.x.
61. Al-Hassnan ZN, Al-Dosary M, Alfadhel M, Faqeih EA, Alsagob M, Kenana R, Almass R, Al-Harazi OS, Al-Hindi H, Malibari OI, Almutari FB, Tulbah S, Alhadeq F, Al-Sheddi T, Alamro R, AlAsmari A, Almuntashri M, Alshaalan H, Al-Mohanna FA, Kaya N. ISCA2 mutation causes the infantile neurodegenerative mitochondrial disorder. J Med Genet. 2015;52(3):186–94. https://doi.org/10.1136/jmedgenet-2014-102592.
62. Baertling F, Al-Murshedi F, Sánchez-Caballero L, Al-Senaidi K, Joshi NP, Venselaar H, van den Brand MA, Nijtmans LG, Rodenburg RJ. Mutation in mitochondrial complex IV subunit COX5A causes pulmonary arterial hypertension, lactic acidemia, and failure to thrive. Hum Mutat. 2017;38(6):692–703. https://doi.org/10.1002/humu.23210.
63. Boycott KM, Beaulieu CL, Kernohan KD, Gebril OH, Mhanni A, Chudley AE, Redl D, Qin W, Hampson S, Küry S, Tetreault M, Puffenberger EG, Scott JN, Bezieau S, Reis A, Uebe S, Schumacher J, Hegele RA, McLeod DR, Abou Jamra R. Autosomal-recessive intellectual disability with cerebellar atrophy syndrome caused by mutation of the manganese and zinc transporter gene SLC39A8. Am J Hum Genet. 2015;6(97):886–93. https://doi.org/10.1016/j.ajhg.2015.11.002.
64. Eldomery MK, Akdemir ZC, Vögtle F-N, Charng W-L, Mulica P, Rosenfeld JA, Gambin T, Gu S, Burrage LC, Al Shamsi A, Penney S, Jhangiani SN, Zimmerman HH, Muzny DM, Wang X, Tang J, Medikonda R, Ramachandran PV, Wong L-J, Sutton VR. MIPEP recessive variants cause a syndrome of left ventricular non-compaction, hypotonia, and infantile death. Genome Med. 2016;8(1):106. https://doi.org/10.1186/s13073-016-0360-6.
65. Haack TB, Kopajtich R, Freisinger P, Wieland T, Rorbach J, Nicholls TJ, Baruffini E, Walther A, Danhauser K, Zimmermann FA, Husain RA, Schum J, Mundy H, Ferrero I, Strom TM, Meitinger T, Taylor RW, Minczuk M, Mayr JA, Prokisch H. ELAC2 mutations cause a mitochondrial RNA processing defect associated with hypertrophic cardiomyopathy. Am J Hum Genet. 2013;93(2):211–23. https://doi.org/10.1016/j.ajhg.2013.06.006.
66. Hartmann B, Wai T, Hu H, MacVicar T, Musante L, Fischer-Zirnsak B, Stenzel W, Gräf R, van den Heuvel L, Ropers H-H, Wienker TF, Hübner C, Langer T, Kaindl AM. Homozygous

YME1L1 mutation causes mitochondriopathy with optic atrophy and mitochondrial network fragmentation. elife. n.d.;5:e16078. https://doi.org/10.7554/eLife.16078.
67. Heberle LC, Tawari AAA, Ramadan DG, Ibrahim JK. Ethylmalonic encephalopathy—report of two cases. Brain Dev. 2006;28(5):329–31. https://doi.org/10.1016/j.braindev.2005.10.005.
68. Jobling RK, Assoum M, Gakh O, Blaser S, Raiman JA, Mignot C, Roze E, Dürr A, Brice A, Lévy N, Prasad C, Paton T, Paterson AD, Roslin NM, Marshall CR, Desvignes J-P, Roëckel-Trevisiol N, Scherer SW, Rouleau GA, Yoon G. PMPCA mutations cause abnormal mitochondrial protein processing in patients with non-progressive cerebellar ataxia. Brain. 2015;138(6):1505–17. https://doi.org/10.1093/brain/awv057.
69. Malicdan MCV, Vilboux T, Ben-Zeev B, Guo J, Eliyahu A, Pode-Shakked B, Dori A, Kakani S, Chandrasekharappa SC, Ferreira C, Shelestovich N, Marek-Yagel D, Pri-Chen H, Blatt I, Niederhuber JE, He L, Toro C, Taylor RW, Deeken J, Anikster Y. A novel inborn error of the coenzyme Q10 biosynthesis pathway: cerebellar ataxia and static encephalomyopathy due to COQ5 C-methyltransferase deficiency. Hum Mutat. 2018;39(1):69–79. https://doi.org/10.1002/humu.23345.
70. Ramadan DG, Head RA, Al-Tawari A, Habeeb Y, Zaki M, Al-Ruqum F, Besley GTN, Wraith JE, Brown RM, Brown GK. Lactic acidosis and developmental delay due to deficiency of E3 binding protein (protein X) of the pyruvate dehydrogenase complex. J Inherit Metab Dis. 2004;27(4):477–85. https://doi.org/10.1023/B:BOLI.0000037336.91549.44.
71. Shamseldin HE, Alasmari A, Salih MA, Samman MM, Mian SA, Alshidi T, Ibrahim N, Hashem M, Faqeih E, Al-Mohanna F, Alkuraya FS. A null mutation in MICU2 causes abnormal mitochondrial calcium homeostasis and a severe neurodevelopmental disorder. Brain. 2017;140(11):2806–13. https://doi.org/10.1093/brain/awx237.
72. Shamseldin HE, Smith LL, Kentab A, Alkhalidi H, Summers B, Alsedairy H, Xiong Y, Gupta VA, Alkuraya FS. Mutation of the mitochondrial carrier SLC25A42 causes a novel form of mitochondrial myopathy in humans. Hum Genet. 2016;135(1):21–30. https://doi.org/10.1007/s00439-015-1608-8.
73. Spinazzola A, Santer R, Akman OH, Tsiakas K, Schaefer H, Ding X, Karadimas CL, Shanske S, Ganesh J, Di Mauro S, Zeviani M. Hepatocerebral form of mitochondrial DNA depletion syndrome: novel MPV17 mutations. Arch Neurol. 2008;65(8):1108–13. https://doi.org/10.1001/archneur.65.8.1108.
74. Yavuz H, Bertoli-Avella AM, Alfadhel M, Al-Sannaa N, Kandaswamy KK, Al-Tuwaijri W, Rolfs A, Brandau O, Bauer P. A founder nonsense variant in NUDT2 causes a recessive neurodevelopmental disorder in Saudi Arab children. Clin Genet. 2018;94(3–4):393–5. https://doi.org/10.1111/cge.13386.
75. Ferguson E. The strengths and weaknesses of whole-genome sequencing; 2020. p. 3.
76. Petersen B-S, Fredrich B, Hoeppner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. BMC Genet. 2017;18(1):14. https://doi.org/10.1186/s12863-017-0479-5.
77. Nikkanen J, Forsström S, Euro L, Paetau I, Kohnz RA, Wang L, Chilov D, Viinamäki J, Roivainen A, Marjamäki P, Liljenbäck H, Ahola S, Buzkova J, Terzioglu M, Khan NA, Pirnes-Karhu S, Paetau A, Lönnqvist T, Sajantila A, Suomalainen A. Mitochondrial DNA replication defects disturb cellular dNTP pools and remodel one-carbon metabolism. Cell Metab. 2016;23(4):635–48. https://doi.org/10.1016/j.cmet.2016.01.019.
78. Reinecke CJ, Koekemoer G, van der Westhuizen FH, Louw R, Lindeque JZ, Mienie LJ, Smuts I. Metabolomics of urinary organic acids in respiratory chain deficiencies in children. Metabolomics. 2012;8(2):264–83. https://doi.org/10.1007/s11306-011-0309-0.
79. Smuts I, van der Westhuizen FH, Louw R, Mienie LJ, Engelke UFH, Wevers RA, Mason S, Koekemoer G, Reinecke CJ. Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach. Metabolomics. 2013;9(2):379–91. https://doi.org/10.1007/s11306-012-0455-z.
80. Legault JT, Strittmatter L, Tardif J, Sharma R, Tremblay-Vaillancourt V, Aubut C, Boucher G, Clish CB, Cyr D, Daneault C, Waters PJ. A metabolic signature of mitochondrial dysfunction

revealed through a monogenic form of Leigh syndrome. Cell Rep. 2015;13(5):981–9. https://doi.org/10.1016/j.celrep.2015.09.054.

81. Sharma R, Reinstadler B, Engelstad K, Skinner OS, Stackowitz E, Haller RG, Clish CB, Pierce K, Walker MA, Fryer R, Oglesbee D, Mao X, Shungu DC, Khatri A, Hirano M, De Vivo DC, Mootha VK. Circulating markers of NADH-reductive stress correlate with mitochondrial disease severity. J Clin Invest. 2021;131(2):136055. https://doi.org/10.1172/JCI136055.
82. Hall AM, Vilasi A, Garcia-Perez I, Lapsley M, Alston CL, Pitceathly RDS, McFarland R, Schaefer AM, Turnbull DM, Beaumont NJ, Hsuan JJ, Cutillas PR, Lindon JC, Holmes E, Unwin RJ, Taylor RW, Gorman GS, Rahman S, Hanna MG. The urinary proteome and metabonome differ from normal in adults with mitochondrial disease. Kidney Int. 2015;87(3):610–22. https://doi.org/10.1038/ki.2014.297.
83. 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, Reynier P. The metabolomic signature of Leber's hereditary optic neuropathy reveals endoplasmic reticulum stress. Brain. 2016;139(11):2864–76. https://doi.org/10.1093/brain/aww222.
84. Sandlers Y, Mercier K, Pathmasiri W, Carlson J, McRitchie S, Sumner S, Vernon HJ. Metabolomics reveals new mechanisms for pathogenesis in Barth syndrome and introduces novel roles for Cardiolipin in cellular function. PLoS One. 2016;11(3):e0151802. https://doi.org/10.1371/journal.pone.0151802.
85. Semeraro M, Boenzi S, Carrozzo R, Diodato D, Martinelli D, Olivieri G, Antonetti G, Sacchetti E, Catesini G, Rizzo C, Dionisi-Vici C. The urinary organic acids profile in single large-scale mitochondrial DNA deletion disorders. Clin Chim Acta. 2018;481:156–60. https://doi.org/10.1016/j.cca.2018.03.002.
86. Esterhuizen K, Lindeque JZ, Mason S, van der Westhuizen FH, Rodenburg RJ, de Laat P, Smeitink JAM, Janssen MCH, Louw R. One mutation, three phenotypes: novel metabolic insights on MELAS, MIDD and myopathy caused by the m.3243A>G mutation. Metabolomics. 2021;17(1):10. https://doi.org/10.1007/s11306-020-01769-w.
87. Stoessel D, Schulte C, Teixeira dos Santos MC, Scheller D, Rebollo-Mesa I, Deuschle C, Walther D, Schauer N, Berg D, Nogueira da Costa A, Maetzler W. Promising metabolite profiles in the plasma and CSF of early clinical Parkinson's disease. Front Aging Neurosci. 2018;10:51. https://doi.org/10.3389/fnagi.2018.00051.
88. Luan H, Liu L-F, Meng N, Tang Z, Chua K-K, Chen L-L, Song J-X, Mok VCT, Xie L-X, Li M, Cai Z. LC–MS-based urinary metabolite signatures in idiopathic Parkinson's disease. J Proteome Res. 2015;14(1):467–78. https://doi.org/10.1021/pr500807t.
89. Ibáñez C, Simó C, Martín-Álvarez PJ, Kivipelto M, Winblad B, Cedazo-Mínguez A, Cifuentes A. Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis–mass spectrometry metabolomics. Anal Chem. 2012;84(20):8532–40. https://doi.org/10.1021/ac301243k.
90. Scolamiero E, Cozzolino C, Albano L, Ansalone A, Caterino M, Corbo G, di Girolamo MG, Di Stefano C, Durante A, Franzese G, Franzese I, Gallo G, Giliberti P, Ingenito L, Ippolito G, Malamisura B, Mazzeo P, Norma A, Ombrone D, Ruoppolo M. Targeted metabolomics in the expanded newborn screening for inborn errors of metabolism. Mol BioSyst. 2015;11(6):1525–35. https://doi.org/10.1039/c4mb00729h.
91. Jacob M, Malkawi A, Albast N, Al Bougha S, Lopata A, Dasouki M, Abdel Rahman AM. A targeted metabolomics approach for clinical diagnosis of inborn errors of metabolism. Anal Chim Acta. 2018;1025:141–53. https://doi.org/10.1016/j.aca.2018.03.058.
92. Salihovic S, Broeckling CD, Ganna A, Prenni JE, Sundström J, Berne C, Lind L, Ingelsson E, Fall T, Ärnlöv J, Nowak C. Non-targeted urine metabolomics and associations with prevalent and incident type 2 diabetes. Sci Rep. 2020;10(1):16474. https://doi.org/10.1038/s41598-020-72456-y.

93. Gu X, Al Dubayee M, Alshahrani A, Masood A, Benabdelkamel H, Zahra M, Li L, Abdel Rahman AM, Aljada A. Distinctive metabolomics patterns associated with insulin resistance and type 2 diabetes mellitus. Front Mol Biosci. 2020;7:411. https://doi.org/10.3389/fmolb.2020.609806.
94. Bu Y, Wang H, Ma X, Han C, Jia X, Zhang J, Liu Y, Peng Y, Yang M, Yu K, Wang C. Untargeted Metabolomic profiling of the correlation between prognosis differences and PD-1 expression in sepsis: a preliminary study. Front Immunol. 2021;12:740. https://doi.org/10.3389/fimmu.2021.594270.
95. Lin S-H, Fan J, Zhu J, Zhao Y-S, Wang C-J, Zhang M, Xu F. Exploring plasma metabolomic changes in sepsis: a clinical matching study based on gas chromatography–mass spectrometry. Ann Transl Med. 2020;8(23):1568. https://doi.org/10.21037/atm-20-3562.
96. Wang J, Sun Y, Teng S, Li K. Prediction of sepsis mortality using metabolite biomarkers in the blood: a meta-analysis of death-related pathways and prospective validation. BMC Med. 2020;18(1):83. https://doi.org/10.1186/s12916-020-01546-5.
97. Cavus E, Karakas M, Ojeda FM, Kontto J, Veronesi G, Ferrario MM, Linneberg A, Jørgensen T, Meisinger C, Thorand B, Iacoviello L, Börnigen D, Woodward M, Schnabel R, Costanzo S, Tunstall-Pedoe H, Koenig W, Kuulasmaa K, Salomaa V, BiomarCaRE consortium. Association of circulating metabolites with risk of coronary heart disease in a European population: results from the biomarkers for cardiovascular risk assessment in Europe (BiomarCaRE) consortium. JAMA Cardiol. 2019;4(12):1270–9. https://doi.org/10.1001/jamacardio.2019.4130.
98. Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. Nat Rev Cardiol. 2011;8(11):630–43. https://doi.org/10.1038/nrcardio.2011.138.
99. Debik J, Euceda LR, Lundgren S, Gythfeldt HVDL, Garred Ø, Borgen E, Engebraaten O, Bathen TF, Giskeødegård GF. Assessing treatment response and prognosis by serum and tissue metabolomics in breast cancer patients. J Proteome Res. 2019;18(10):3649–60. https://doi.org/10.1021/acs.jproteome.9b00316.
100. Feng LR, Barb JJ, Regan J, Saligan LN. Plasma metabolomic profile associated with fatigue in cancer patients. Cancer Med. 2021;10(5):1623–33. https://doi.org/10.1002/cam4.3749.
101. Guida F, Tan VY, Corbin LJ, Smith-Byrne K, Alcala K, Langenberg C, Stewart ID, Butterworth AS, Surendran P, Achaintre D, Adamski J, Amiano P, Bergmann MM, Bull CJ, Dahm CC, Gicquiau A, Giles GG, Gunter MJ, Haller T, Johansson M. The blood metabolome of incident kidney cancer: a case–control study nested within the MetKid consortium. PLoS Med. 2021;18(9):e1003786. https://doi.org/10.1371/journal.pmed.1003786.
102. Wojakowska A, Chekan M, Widlak P, Pietrowska M. Application of metabolomics in thyroid cancer research. Int J Endocrinol. 2015;2015:e258763. https://doi.org/10.1155/2015/258763.

Metabolomics of Rare Endocrine, Genetic Disease: A Focus on the Pituitary Gland

Afshan Masood, Abeer Malkawi, Anas M. Abdel Rahman, and Mohamed Siaj

Abstract Endocrine diseases and disorders are influenced by individual genetic and environmental factors that directly influence metabolite levels revealing unique metabolomic disease profiles. In the last few years, metabolomic-derived biomarkers and quantitative traits have helped identify the underlying mechanisms for many rare endocrine diseases and demonstrated a high potential for use in precision medicine. Using the metabolomic platform has also provided disease-specific biomarkers for diagnosis and monitoring treatment beyond the use of conventional biochemical immunoassays. Although the development of metabolomic profiling in pituitary disorders is at an early stage, recent advances have shown it to be a promising approach for identifying specific disease biomarkers in cases of growth hormone disorders (acromegaly, short stature) and Cushing's syndrome. Implementing high-performance metabolomic analysis techniques in pituitary disease will be a helpful clinical tool for significantly improving diagnosis and, potentially, the therapeutic approach by identifying highly specific disease biomarkers and novel molecular pathogenic mechanisms.

Keywords Rare endocrine disease · Rare genetic disease · Metabolomics · Biomarkers · Pituitary adenomas · Growth hormone · Acromegaly · Short stature · Pituitary dwarfism · Cushing's syndrome

A. Masood · A. Malkawi · M. Siaj (✉)
Department of Chemistry and Biochemistry, Université du Québec à Montréal (UQAM), Montreal, QC, Canada
e-mail: siaj.mohamed@uqam.ca

A. M. Abdel Rahman (✉)
Metabolomics Section, Department of Clinical Genomics, Center for Genome Medicine, King Faisal Specialist Hospital & Research Center (KFSHRC), Riyadh, Saudi Arabia
e-mail: AabdelRahman46@kfshrc.edu.sa

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_8

Abbreviations

CS	Cushing's syndrome
DMA	Dimethylamine
GH	Growth hormone
GPR101	G protein-coupled receptor 101
GWAS	Genome-wide association
IR	Insulin resistance
LC-MS	Liquid chromatography-mass spectrometry
NMR	Nuclear magnetic resonance
PA	Pituitary adenomas
rhGH	Recombinant human growth hormone
UPLC Q-TOF	Ultra-performance liquid chromatography quadrupole time of flight

1 Introduction

Endocrinology is a branch of medicine that deals with studying the endocrine system. The endocrine system comprises a system of specialized ductless glands responsible for synthesizing, storing, and releasing their secretions, that is, hormones, directly into circulation. Classically, hormone action occurs at different peripheral target tissues having receptors for the specific hormones. In addition to the classical endocrine signaling, hormones also act on neighboring cells (paracrine effect), on the cell secreting the hormone itself (autocrine effect), or locally within the cell without actually getting released from it (intracranial effect).

Hormonal action can be broadly described as important in controlling and coordinating whole-body metabolism and maintaining body homeostasis. Hormones are responsible for maintaining energy balance, regulating metabolic pathways, reproduction, growth and development, and response to injury, stress, and environmental factors. The hormonal synthesis, production, and secretion are tightly regulated processes, broadly, through complex regulatory feedback control loops or axes between different endocrine glands. The feedback mechanisms control the circulating concentrations of the hormones within a tight physiological range necessary for optimal hormone action at the cellular level. The most well known among these are those under hypothalamic-pituitary control, such as the hypothalamo-pituitary-thyroid axis and hypothalamo-pituitary-adrenal axis that control and regulate thyroid and adrenal hormone action via the regulatory and organ-stimulating hormones. Endocrine dysfunction arises as a result of either subnormal (e.g., hypothyroid) or excess production (e.g., hyperthyroid) of hormones or due to peripheral resistance to hormone action (e.g., insulin resistance (IR)), which leads to different disease states.

Clinical evaluation of endocrine dysfunction is generally based on measuring single effector hormones using immunoassays in the clinical laboratory. Although these methods are well established and widely used in different hospital laboratories, single hormone measurements, as commonly perceived, do not completely

represent disease or hormonal defects [1]. The complexity in endocrine disease diagnosis arises as virtually all complex processes are regulated at any point with one hormone whose functions are closely intertwined. Therefore, measurement of single hormone levels may not be sufficient in identifying disease. The immunoassays are further limited in their detection capabilities due to their narrow specificity, sensitivity, and a high coefficient of variation between the different immunoassays for the same hormone [2, 3]. With the advent of deeper genome sequencing, an increasing number of endocrine diseases are grouped as a single disease with a heterogeneous presentation but are now identified as distinct disorders based on their genetic makeup. In the present chapter, we will look at the metabolomics of diseases of the pituitary gland.

2 Endocrine Diseases and Genetics

Although the paradigm of studying hormonal action and related diseases focused on the concept of one anatomical gland and its associated hormone, this view is changing with many interrelated complex interactions being identified between the different hormones. Hormones do not act in isolation; a single hormone affects multiple organs and functions, and several hormones control each function. Identification of various genetic polymorphisms conveying an increased risk of developing endocrine disease has been identified through genome-wide association (GWAS) studies. These studies have greatly improved knowledge and genetic testing, becoming a significantly important component for the thorough clinical diagnostic workflow in endocrine diagnostics and the routinely performed biochemical laboratory analysis [4, 5].

Recent studies using GWAS have identified many genetic mutations (such as single nucleotide polymorphisms, allelic loss or gene amplification, and epigenetic changes, usually by promoter methylation chromosomal defects) and their variants that are known to lead to endocrine disorders and pathogenic or benign neoplastic endocrine disease conditions [5–9]. This has been facilitated through clinical genetic studies in patients carrying mutations with well-characterized presentations and manifestations. These genetic mutations have been identified in genes involved in regulating the cell cycle, growth factors, signaling pathways, and hormone receptors. The genetic mutations may either result in gain of function or loss of function of the specific genes and their downstream protein products. These defects in proteins translate into biochemical enzymatic defects resulting in alterations in the hormonal levels due to disorders of hormone synthesis or receptor uptake, to name a few, resulting in endocrine disease [9–13].

Additionally, advances in cellular biology and transcriptomic studies have shown the significant role that hormones play in regulating gene expression. Mutations in the germline result in familial diseases, while somatic mutations may present with expression only within specific tissues. Evaluation of rare endocrine genes besides the characterization of the hormonal defects requires a comprehensive genetic workup [4, 5].

Genetic diagnostic testing in rare endocrine conditions has become important for achieving an early molecular diagnosis and identifying presymptomatic individuals. Although significant, genetic evaluation is not without disadvantages as it is highly specialized and not easily available routinely. Recent advances in OMIC technologies, specifically in metabolomics, have shown that changes in the metabolite pattern reflect the associated disease conditions [14]. The disease is often seen to exceed the influence of clinical factors. The variations in many metabolites have been heritable and linked to distinct loci identified by GWAS [15]. Genetic variants, particularly of the enzyme coding genes, are associated with changes in the proteins regulating the homeostasis of key lipids, carbohydrates, or amino acid metabolisms and metabolites.

Over the last decade, the understanding of physiological and pathological processes underpinning endocrine and endocrine-related disease has significantly expanded, aided by advances in mass spectrometry (MS) approaches and novel molecular biological and computational tools. The metabolic compound dysregulation is associated with changes in physiology which are consequences of a unique expression of the genes. A small variation in protein expression can significantly affect the metabolic pathway activity with alterations in concentrations of metabolites related to the relevant proteins. Thus, compared to both the proteome and transcriptome, metabolomic profile is far more sensitive. The development of the high-throughput omic platforms has helped identify the specific metabolites altered with endocrine diseases and more so in the cases of rare genetic, endocrine diseases. Determining metabolomic features as causes of endocrine disorders provides a better holistic opportunity to understand the related pathophysiology and develop approaches for personalized therapy [16].

3 Metabolomics of Endocrine Disease

Endocrine diseases deeply impact the human metabolome as hormonal defects affect the functions of enzymes, causing alterations in the normal metabolic activity of tissues or glands, leading to disease. Metabolomics has contributed significantly to endocrinology by providing means to understand better the inner workings of individual cells. It also helped to understand the interactions between different organ systems and providing insight into physiological and pathological processes. The main application of metabolomics in endocrine disease has been clinical research dealing with disease metabolomic profiling and biomarker discovery for diagnostic purposes. Plasma profiling of the disease and control states has allowed the characterization of a comprehensive set of metabolites in biological samples. Fluctuations in concentration of these small molecules, in response to different stimuli, serve as markers for monitoring of disease. Two principal analytical platforms have been most used to analyze metabolites in body fluids, primarily derivatives of blood (serum, plasma), tissues, and urine. These include nuclear magnetic resonance (NMR) spectroscopy and liquid/gas chromatography coupled with mass

spectrometry (LC-MS). The analysis of lipid metabolites was enhanced by using ultra-performance liquid chromatography quadrupole time-of-flight (UPLC Q-TOF) MS and GC-MS. TMS major complementary approaches have been applied. One involves the targeted analysis of metabolites, which measures the concentrations of known disease-associated metabolites. The second is an untargeted hypothesis-driven approach that provides a global profile of a large set of metabolites associated with the disease semiquantitatively.

Alterations in metabolite levels or a specific group of metabolites reflect the intricate relationships between genes and their products, as well as the activity of related enzymes within metabolic pathways either physiologically or in a disease state. In combination with genetics, untargeted metabolomics has provided additional information to determine the association between genetic variants and disease. Combining the clinical (phenotype) with the different molecular datasets (OMIC technologies) in a bioinformatic-based clinical decision-making system also has the advantage of advancing personalized medicine. These advances are in understanding the pathophysiology, diagnosis, stratification, management, and assessment of treatment efficacy in endocrinology. Metabolomics adds to this milieu by including metabolic information, furthering the accuracy of diagnosis in individuals and discovering novel metabolites as reported in the human metabolome database with the disease. For each metabolic gene, prediction models can then be developed for predicting changes in concentrations of metabolites in different biological fluids (serum, plasma, urine, saliva, etc.) or tissue specimens with various genetic mutations (Table 1).

3.1 Metabolomics of Hypothalamus and Pituitary Gland Dysfunction

The hypothalamus along with pituitary gland together constitutes a functional unit responsible for secreting regulatory hormones that exert control over the functions of nearly all the endocrine glands. The pituitary is the master regulatory gland of the body, acting through several hypothalamic-pituitary-target organ regulatory axes. Any defects arising within these regulatory systems, synthesis, storage, or release of the hormones impact the human body's overall physiological functions, leading to disorders and diseases.

Primary dysfunction of the pituitary gland is uncommon. The majority of the cases of hypo-, hyper-, or panhypopituitarism arise secondary to the presence of either neoplastic or benign tumors such as adenomas. Pituitary adenomas (PA) are benign neoplasms that are highly heterogeneous with varying subtypes and accounts for nearly 15% of primary tumors arising in the brain. They are either secretory based on the increase or decrease in the hormones secreted by the affected cell types or nonsecretory. Inherited genetic mutations, although rare, have been identified by GWAS and are known to lead to the development of benign PA, with several genetic

Table 1 The different endocrine glands with the associated disease, gene locus, and major metabolites altered with the disease

Endocrine gland	Hormones	Disease	Genes	Metabolomic platform	Major significant metabolites	References
Pituitary	GH defects	GH acromegaly	*AIP*, *GNAS*, *GPR101*, *USP8*, *USP48*, and *BRAF. MEN1*, *PRKAR1A*, *GPR101*, *SDHx*, *CDKN1B*, and *MAX*	GC-MS (untargeted), LC-MS (targeted), and ^{1}H-NMR (untargeted)	Serine, dihydrocoumarin, glyceric acid, L-dithiothreitol, 5-aminovaleric acid, N-acetyl-L-glutamic acid, gluconic acid, mono-olein, and dimethylamine (↑) BCAA, lysine, D-erythronolactone, taurine, carbamoyl-aspartic acid, mucic acid, and lactate (↓)	[24–26]
		Short stature	*PIT-1*, *HESX13*, *SOX3*, *FGFR3*, and *PAPP-A2*	NMR (untargeted)	Citrate, lactate, mannitol, alanine, and DMA (↑). Creatine, creatinine, TMAO, and acetoacetate (↓)	[32]
				GC-MS (untargeted)	Threonic acid, cysteine, cystine, and palmitoleic acid (↓). Glutamic acid, aspartic acid, hypoxanthine-like, uridine, and glyceric acid (↑)	[31]
				NMR (untargeted)	Citrate, phenylalanine, creatinine, and tyrosine (↑) Glucose, serine, betaine, inositol, lysine, glycerol, and glutamine (↓)	[33]
	ACTH increase	Cushing's disease	*USP8*	LC-MS/MS (targeted)	11-Deoxycortisol, cortisol, cortisone, corticosterone, 11-deoxycorticosterone, androstenedione, 18-oxocortisol, DHEA, DHEA-SO4, and aldosterone (↓)	[38, 40]
Adrenal gland	Cortisol	Cushing's syndrome	PRKACA and ARMC5	LC-MS/MS (targeted)	2-Hydroxybutyric acid, aminoadipic acid, L-aspartic acid, 3-hydroxyphenyl acetic acid, hypoxanthine, 4-pyridoxic acid, quinolinic acid, sucrose, xanthine, glucose 6-phosphate, deoxycholic acid, and 3-methyladipate (direction of change unspecified)	[39]

mutations identified as single gene mutations. These have been identified for single genes encoding for guanine nucleotide-binding protein G(s) subunit alpha (*GNAS*) and G protein-coupled receptor 101 (*GPR101*), aryl hydrocarbon receptor-interacting protein (*AIP*), and deubiquitinase genes: *USP8*, *USP48*, and *BRAF*. PA can also occur in familial syndromes, including *MEN1*, *SDHx* (mutated in the PA, pheochromocytomas, and paragangliomas), von Hippel-Lindau, DICER1, *PRKAR1A* (mutated in Carney complex), succinate dehydrogenase-related familial PA, *GPR101* (involved in X-linked acrogigantism), neurofibromatosis type 1 (NF1), and Lynch syndromes [17]. The phenotype of the secretory PA depends on the affected cell types and the hormones secreted; growth hormone (GH)-secreting somatotroph adenomas result in acromegaly, corticotroph cell adenomas-secreting corticotropin hormone result in Cushing's disease, lactotroph cell adenomas-secreting prolactin result in hyperprolactinemia, and thyrotropin-secreting adenomas result in hyperthyroidism. The nonsecreting adenomas, on the other hand, manifest as incidental pituitary sellar masses and lead to hypogonadism. The clinical diagnosis of pituitary disease, irrespective of the cell type affected, is challenging. This is because of the wide range of presenting symptoms, nonspecific changes in single-hormone measurements, the need for adopting stimulation or suppression testing, and the requirement for imaging studies. Metabolomic approaches using GC/LC-MS have profiled pituitary-related diseases and identified biomarkers differentiating between patients with PA and healthy controls. Plasma secretory plasma metabolomic profiling revealed significant alterations in the metabolism of amino acid, specifically in alanine, glutamate (metabolites of the glutamate oxidationcycle), and aspartate (metabolite of the urea cycle involved in gluconeogenesis), and increase in homocysteine levels. Increased homocysteine levels in patients with PA are considered to be the cause for increased cytotoxicity and oxidative stress [18].

In addition to the findings in plasma, pituitary tissue metabolomic profiling was also done on tissue sections removed postoperatively from patients with gonadotropin- and PRL-secreting PA. The metabolomic profile showed differential regulation of phosphoethanolamine, glutamate, glutamine, N-acetyl aspartate, aspartate, and myoinositol, which are known metabolites involved in the regulation of the central nervous system. Levels of these metabolites distinctly differentiated between the two adenomas with PRL-secreting PA showing a decrease in levels of phosphoethanolamine, N-acetyl aspartate, and myoinositol while aspartate, glutamate, and glutamine levels were increased. Women with PRL-secreting PA showed higher estrogen metabolites and 17-ketosteroids in the urine. The increase in these metabolites was most likely due to reduced enzymatic actions of 3-beta-hydroxysteroid dehydrogenase and 5α-reductase enzymes among these patients. These patients also showed characteristic increase in the ratios of levels between 5-beta and 5-alpha-hydrogensteroids and delta 5 and delta 4 steroid ratios which served as novel markers for disease detection [19, 20].

3.2 Growth Hormone Secretion Defects

The primary cause of GH defects is pituitary adenomas that secrete aberrant GH. They can appear as single (isolated) tumors or as component of other systemic conditions as multiple endocrine neoplasia type 1 or type 4, Carney complex, McCune-Albright syndrome, or in association with pheochromocytoma and paraganglioma. GH acts via the somatotropic axis consisting of GH, GH receptors, insulin-like growth factors (IGF) 1 and 2, their associated carrier proteins, receptors, and releasing factors, which regulate growth and body composition. GH secretion defects can manifest clinically as an increase in its circulating levels leading to acromegaly or a decrease that leads to short stature.

3.2.1 Growth Hormone Excess: Acromegaly

An excess of GH or its chronic hypersecretion results in acromegaly or gigantism, a rare endocrine disease that is debilitating and leads to multiple comorbidities with reduction in life expectancy. The uncontrolled production of GH stimulates the liver to increasingly synthesize and secrete insulin-like growth factor-1 (IGF-1) which in turn influences metabolic changes in various organ systems and stimulates somatic (internal organs, tissues, bones, and muscles) overgrowth and development of comorbidities, including cardiovascular and malignant diseases. In skeletal muscles, GH exerts an anabolic impact that increases amino acid intake and protein synthesis and decreases protein oxidation. This action shifts the normal metabolism from mainly utilizing glucose and protein substrates to oxidation of lipid oxidation, altering the linear body growth and organ growth during childhood. The resulting clinical phenotype of these patients presents with growth acceleration, enlarged feet and hands, increased appetite, and broadening and coarsening of the facial features.

The primary cause of pituitary gigantism in nearly 50% of individuals is an underlying genetic variant or mutation [21, 22]. In 95% of cases, acromegaly occurs sporadically because of adenomas of somatotroph cells that are characterized by excessive secretion of GH and IGF-1. Fifty percent of these tumors are attributable to familial germline mutations in aryl hydrocarbon receptor-interacting protein (AIP) and probable G-protein-coupled receptor 101 (GPR101) genes. On the other hand, cases with pituitary hyperplasia are a rarity and are usually encountered as part of syndromes such as the McCune-Albright syndrome, Carney complex disease, or X-linked acrogigantism. Besides the mutations seen in AIP, germline mutations have also been documented in other genes such as GPR101, PRKAR1A, CDKN1B, GNAS, MAX, SDHx, and MEN1. Somatic mutations have been shown to be related to mutations in GNAS gene and account for 40% of tumors [23].

Presently the diagnosis of active acromegaly is made by determining elevations in GH levels, after a standard oral glucose tolerance testing, and in levels of IGF-1 when compared to age and gender matched controls. Confirmation of the diagnosis is through radiological MRI evaluation that validates the findings of an adenoma.

Biochemical evaluations of GH and IGF-1 levels are sometimes inconsistent, and an increase in GH levels is not paralleled with that of IGF-1. The heterogeneous clinical presentation of the disease lacks early warning signs, and the inconsistencies in the measurements of GH and IGF-1 levels lead to delay or misdiagnosis. Molecular genetic testing using chromosomal microarray analysis or single gene duplication studies has helped identify cases with germline or somatic mutations [22].

In addition to genetic testing, metabolomic analysis has aided the diagnosis of acromegaly by creating a metabolomic fingerprint of the disease and identifying characteristic metabolites that differentiate patients with active disease from controls. Elevations were noted in gluconeogenic substrates (glycerol, lactate, propionate, and glucogenic amino acids such as valine and isoleucine); serine, 5-aminovaleric acid, dihydrocoumarin, mono-olein, N-acetyl-L-glutamic acid, glyceric acid, L-dithiothreitol, and gluconic acid were observed in the metabolomic profile in active acromegaly compared with normal controls. In contrast, serum levels of D-erythronolactone, taurine, carbamoyl-aspartic acid, and mucic acid were decreased. Among the different metabolites, glyceric acid and taurine had the highest sensitivities to discriminate between patients with acromegaly from normal controls [24, 25].

Patients with active disease having increased GH but normal IGF-1 levels showed significantly lower levels of essential branched-chain amino acids (BCAAs) (valine, isoleucine), lysine, and lactate, whereas levels of the metabolite dimethylamine were higher. The inverse correlation between valine and isoleucine was seen with higher GH levels and not with IGF-1 indicating that lower BCAA levels, which represent the main metabolic fingerprint of acromegaly, were influenced by GH rather than IGF-1 as the primary mediator [26]. Metabolomic analysis in patients with acromegaly also highlighted significant alterations in major metabolic pathways in these groups of patients. The pathways affected as a result of active disease included glycerolipid, glyoxylate, taurine, hypotaurine, dicarboxylate, and pyrimidine pentose phosphate pathway. Dysregulation of these pathways was proposed to be the underlying reason that supported hyperplasia of tissues observed in different organs, due to excess GH secretion [19]. An increase in gluconeogenic metabolites along with decreased BCAA levels detected was suggested to be a result of an accelerated consumption of BCAAs. Assessing levels of BCAAs, besides identifying the active disease, were also beneficial in monitoring the response to therapy.

Metabolomic profiling in patients with acromegaly, with associated cardiovascular dysfunction, identified distinct changes in amino acids and in plasma lipids. Specific perturbations were noted in the glycerophospholipid, sphingolipid, and metabolites involved in the linoleic acid metabolic pathways. Levels of phosphatidyl ethanolamine (PE) (22:6/16:0) positively correlated with changes in left ventricular mass, while lysophosphatidyl choline (LysoPC) (16:0) was positively correlated with alterations in fractional shortening and left ventricle ejection fraction [27]. Patients with active acromegaly also present with increased insulin resistance (IR) accompanied with lowered hepatocellular lipids and cholesterol. This is in stark contrast to other IR-associated metabolic diseases which are commonly seen associated with fatty liver disease. The lower hepatic cholesterol levels reflect

the decreased liver fat content, a lower unsaturated-to-saturated lipid ratio, and decreased levels of different carnitine species, namely, plasma butyryl carnitine hexanoyl carnitine [25]. Among the various FFA species, levels of cholesteryl esters (CEs, 18:3) were lower. In contrast, within phosphocholine (PC) lipids, increased levels of LPC (18:0), as well as decreased PC (36:5), ether PC (38:6), PC (40:7), and PC (42:5), were noted, and in the sphingomyelin class, SM (36:0) was significantly lower. The hepatic lipid content decreased from increased hepatic ATP synthesis due to excess GH.

3.2.2 Growth Hormone Deficiency

Insufficient production of GH leads to GH deficiency or pituitary dwarfism, a rare endocrine disease condition. The causes of GHD can be inherited, congenital, or acquired. The usual manifestation of the disease is during childhood, with children clinically presenting with abnormally short stature but normal body proportions. Lately, the importance of GH has also been observed in adult patients, especially concerning its effect on lipid profile, body composition, bone mass, and cognition [28]. Adult growth hormone deficiency is a well-defined clinical condition that is characterized by abnormal body composition, poor quality of life, dyslipidemia, and increased risk of cardiovascular disease and death [29]. Genetic analysis by whole genome sequencing has identified multiple genes that affect the normal development of the pituitary, hypothalamic signaling or influence the actions of GH and IGF-1 in target tissues by altering binding to its receptors. Isolated GH deficiency can arise due to mutations or defects in the growth hormone (GH1) gene itself or the GH-releasing hormone receptor which lead to either classical GH deficiency or produce physiologically inactive GH. More rarely, growth hormone deficiency can result from mutations in HESX13, SOX3, FGFR3 gene, and pregnancy-associated plasma protein (PAPP)-A2 [30].

The diagnosis of GHD is based on laboratory evaluation, through radiological imaging examinations and the insulin hypoglycemia test, which is the gold standard test in adults. The biochemical evaluation is done through measuring levels of GH, IGF-1, and IGFBP-3 (insulin-like growth factor binding protein 3) as indirect markers of growth hormone action. Although these markers are used routinely, they lack sensitivity, and among them, IGF-1 is a least sensitive marker and a poor diagnostic indicator of GHD.

Untargeted metabolomic profiling using GC-MS identified numerous metabolites and potential biomarkers for diagnosing GHD and was also used for monitoring recombinant human growth hormone (rhGH) replacement therapy in these patients. Levels of 13 metabolites including threonic acid, cystine, cysteine, palmitoleic acid, glutamic acid, glyceric acid, aspartic acid, uridine, and hypoxanthine-like were reported to be significantly dysregulated in adult patients with GHD. On the other hand, successful treatment with rhGH showed a reversal in glutamic acid, glyceric acid, hexadecanoic acid, and palmitoleic acid to normal control levels [31]. The effectiveness of growth hormone treatment was similarly assessed in a patient

with a rare pituitary-specific positive transcription factor (PIT-1) genetic defect that caused short stature. PIT-1 is responsible for the normal development of the anterior pituitary gland. Urine metabolomic profile from this patient using NMR-based metabonomics showed that higher levels of creatine, creatinine, lactate, and trimethylamine-N-oxide and lower levels of urinary citrate, dimethylamine (DMA), and alanine observed before GH treatment were reversed after treatment and returned to normal values [32]. It was also possible to determine the metabolic differences between children with short stature (SS) and healthy controls, using NMR-based blood metabolomic analysis, to identify metabolites which can serve as potential biomarkers for the diagnosis of SS. Serum levels of creatinine, citrate, phenylalanine, and tyrosine were increased, and levels of inositol, lysine, glycerol, glucose, betaine, serine, and glutamine were lower in comparison to normal controls. These dysregulated metabolites were associated with disturbances in metabolic pathways related to carbohydrate and amino acid metabolism [33]. Metabolomic profiling was also used to monitor therapy and assess the metabolic effects of IGF-1 treatment in PAPPA2-deficient patients. The metabolomic analysis highlighted profound changes in lipid and protein metabolism after replacement. An increase in BCAA, hydroxyproline, glutamic acid, glutamine, and asparagine was noted along with a decrease in levels of glycerol and FFA, palmitate, oleate, arachidonate, linoleate, palmitoleate, and stearate [34]. Metabolomics in GHD extends beyond identifying metabolites as biomarkers for disease diagnosis as it also helped to assess the risk of patients for developing cardiovascular disease. The identified metabolites have been linked in adults with untreated GHD to a number of CV risk factors, including morphological changes in the heart, metabolic alterations, and visceral obesity [26]. More recently, epigenetics and metabolomics have been increasingly developed to detect distinctive fingerprints, which could predict an increased risk of CVD in patients with GHD.

Because of its anabolic effects on protein metabolism and muscle growth, athletes have used GH to improve their performance. Another aspect of metabolomics has been its extension toward analyzing biological fluids, urine, and plasma samples for evidence of doping. In the treated groups, application of direct discriminant analysis distinguished the treatments and was also used to classify them accordingly [35].

3.3 Hypercortisolism (Cushing's Syndrome and Cushing's Disease)

Hypercortisolism is a term that encompasses an increase in cortisol production and is a feature of both Cushing's syndrome (CS) and Cushing's disease. It can be classified into three types: ACTH-dependent (due to pathology of the pituitary gland), ectopic ACTH secretion, and ACTH-independent (due to adrenal pathology). The most common pituitary pathology is adenoma of the pituitary gland affecting the

corticotrophic cells resulting in glucocorticoid (cortisol) excess production from the adrenal glands. The uncontrolled ACTH secretion from a pituitary adrenocorticotrophic adenoma is a rare pathology and is termed as endogenous Cushing's disease.

Clinically cortisol excess manifests by a well-characterized group of signs and symptoms that include centripetal obesity, hypertension, muscle weakness, moon face, fatigue, diabetes mellitus, cognitive difficulties skin changes, osteoporotic fractures, headaches, and changes in mood. In the female patients, menstrual irregularities and hirsutism (female) are also additionally seen [36]. Diagnosis of Cushing's syndrome (CS) (and Cushing's disease) is a tedious multistep process requiring verification of hypercortisolism as the first step, followed by identifying the cause of hyperfunctioning of adrenocortical axis. This is further hampered by the poor diagnostic utility of the glucocorticoid metabolic pathway indices such as plasma cortisol levels, urinary-free cortisol, and other clinically measurable biochemical parameters, such as 11-deoxycortisol. Assessing the cause of hypercortisolism is challenging as each factor in the differential diagnosis of the disease needs to be ruled out.

Metabolomic evaluation of patients suspected of CS, with clinical features of hypercortisolism but indeterminate pituitary imaging, identified many affected metabolic pathways. These mainly involved metabolic pathways related to fatty acids, amino sugars, carbohydrates (glycolysis/gluconeogenesis), purines, glycated nucleotides, amino acids (glutamate, alanine, aspartate, and lysine), vitamin B metabolism, aminoacyl-tRNA biosynthesis, and starch and sucrose metabolism [37]. Patients with ACTH-secreting pituitary adenomas showed distinct changes in plasma metabolite levels identified using LC-MS/MS compared to the control group. These included 2-hydroxybutyric acid, 3-hydroxyphenyl acetic acid, hypoxanthine, L-aspartic acid, aminoadipic acid, 4-pyridoxic acid, quinolinic acid, deoxycholic acid, xanthine, 3-methyladipate, and sucrose and glucose 6-phosphate. Besides the plasma and urine metabolomics, tissue metabolomic analysis from ACTH-secreting pituitary adenoma showed increased short-chain fatty acids (hexanoic, capric, heptanoic, octanoic, and nonanoic fatty acids) [38]. At the same time, glucose-6-phosphate was decreased compared to the control group [19]. Short- and medium-chain acylcarnitines, branched-chain and aromatic amino acids, and polyamine levels were also lower in patients with hypercortisolism. Independently, the severity of hypercortisolism is linked to changes in intermediate metabolism, which are consistent with skewed protein synthesis and catabolism and incomplete β-oxidation, indicating the presence of metabolic inflexibility exists [19, 39].

Differentiating between the different subtypes of hypercortisolism is also clinically challenging. Using metabolomic approaches through serum and urine steroid profiling has helped identify sensitive markers in patients with CS. Steroid metabolite profiling has shown promise for accurately subtyping patients with CS. The profiling revealed that patients with ACTH-dependent CS (caused by ACTH-secreting pituitary adenomas or ectopic malignancies) have elevated androgens and related metabolites, whereas patients with CS having glucocorticoid hypersecretion from the adrenals which is autonomous or independent from ACTH have lower levels [38, 40]. Targeted metabolomic analysis using a combination panel of ten

plasma steroids was shown to provide a sensitive panel of markers for diagnosis of CS with a high discriminatory capacity between the three subtypes of CS. The various steroids in the panel (11-deoxycortisol, cortisol, cortisone, corticosterone, 11-deoxycorticosterone, androstenedione, 18-oxocortisol, DHEA, DHEA-SO4, and aldosterone) demonstrated significantly distinct profile patterns among patients with ACTH-independent and ACTH-dependent forms of CS. From the metabolites mentioned earlier, increases in cortisol levels, its precursor 11-deoxycortisol and derivative 21-deoxycortisol, cortisone, corticosterone, 11-deoxycorticosterone, and decrease in plasma aldosterone and 18-oxocortisol differentiated between patients with and without CS. Higher levels of 11-deoxycorticosterone and 11-deoxycortisol were higher in patients with pituitary and adrenal CS, while patients with ectopic CS disease showed lower plasma levels of 18-oxocortisol and aldosterone. The plasma levels of adrenal androgens, androstenedione, DHEA, and DHEAS were higher in pituitary and ectopic CS compared to adrenal CS. On the other hand, ectopic CS had highest increase in glucocorticoids.

The urine metabolomic profile was also evaluated in patients with CS and showed a higher urinary excretion of DHEA and DHEAS metabolites in pituitary disease compared to adrenal CS [19, 37]. The ratio of cortisol to metabolites of cortisone was higher, indicating the suppression of activity of the cortisol-inactivating enzyme HSD11B2. The unrestricted action of cortisol on the mineralocorticoid receptor explains the clinical symptoms of hypertension and hypokalemia commonly seen in these patients.

Cushing's syndrome leads to various metabolic dysfunctions, including cardiovascular disease due to a proatherogenic shift in the circulating lipids. Metabolomic analysis identified increased urinary ceramides, glycerophospholipids, and sphingolipid metabolites, independently associated with higher urinary-free cortisol. In the study by Vega-Beyhart, patients with CS showed significant alterations in 93 metabolites showing an increase in sulfur containing amino acids, ceramides, triacylglycerols, cholesteryl esters, and glycerophospholipids. On the other hand, a decrease was seen in essential and nonessential AA, polyunsaturated fatty acids, conjugated bile acids, and second messenger glycerolipid concentrations. Twenty-four-hour urinary-free cortisol was associated with alterations in the concentration of many lipids and amino acid metabolites. The identified metabolites comprising of ten amino acids and ten lipid metabolites demonstrated an AUC-ROC of 95% and served as a metabolic signature for the classification of CS. The identified amino acids, acylcarnitines, ceramides, and glycerophospholipid markers were suggested as potential biomarkers of cardiovascular risk in patients with CS [41].

4 Conclusion

Metabolomic profiling of endocrine diseases with its high-power discrimination is a powerful tool for the clinical investigation and diagnosis of rare pituitary endocrine diseases. The realm of metabolomics, in managing pituitary diseases, will in

the future extend beyond identifying metabolites as biomarkers for disease diagnosis and towards personalized medicine. This approach will aim to deliver targeted individualized therapy that optimizes the effectiveness of treatments and avoid unwanted adverse effect reactions or procedures. Establishing disease-specific metabolite panels on a chip could further revolutionize the diagnostic industry. Achieving this milestone would require further studies with larger populations, meta-analysis, and a uniform standardization of analytical and statistical steps to yield reproducible results.

References

1. Wallaschofski H. What will metabolomics studies mean to endocrinology? J Endocrinol. 2012;215(1):1–2.
2. Barth JH. An evidence-base for laboratory endocrinology? Clin Biochem Rev. 2008;29(3):97–101.
3. Hillebrand JJ, Wickenhagen WV, Heijboer AC. Improving science by overcoming laboratory pitfalls with hormone measurements. J Clin Endocrinol Metab. 2021;106(4):e1504–12.
4. De Sousa SM, et al. Genetic testing in endocrinology. Clin Biochem Rev. 2018;39(1):17–28.
5. Eggermann T, et al. Genetic testing in inherited endocrine disorders: joint position paper of the European reference network on rare endocrine conditions (Endo-ERN). Orphanet J Rare Dis. 2020;15(1):144.
6. Goodarzi MO. Genetics of common endocrine disease: the present and the future. J Clin Endocrinol Metab. 2016;101(3):787–94.
7. Hirschhorn JN, et al. A comprehensive review of genetic association studies. Genet Med. 2002;4(2):45–61.
8. Song ZJ, et al. The genome-wide mutational landscape of pituitary adenomas. Cell Res. 2016;26(11):1255–9.
9. Fang Q, et al. Genetics of combined pituitary hormone deficiency: roadmap into the genome era. Endocr Rev. 2016;37(6):636–75.
10. Hwangbo Y, Park YJ. Genome-wide association studies of autoimmune thyroid diseases, thyroid function, and thyroid cancer. Endocrinol Metab (Seoul). 2018;33(2):175–84.
11. Rahmioglu N, et al. Variability of genome-wide DNA methylation and mRNA expression profiles in reproductive and endocrine disease related tissues. Epigenetics. 2017;12(10):897–908.
12. Persani L, et al. GENETICS IN ENDOCRINOLOGY: genetic diagnosis of endocrine diseases by NGS: novel scenarios and unpredictable results and risks. Eur J Endocrinol. 2018;179(3):R111–23.
13. Eriksson D, et al. GWAS for autoimmune Addison's disease identifies multiple risk loci and highlights AIRE in disease susceptibility. Nat Commun. 2021;12(1):959.
14. Jacob M, et al. Metabolomics toward personalized medicine. Mass Spectrom Rev. 2019;38(3):221–38.
15. Rhee EP, et al. A genome-wide association study of the human metabolome in a community-based cohort. Cell Metab. 2013;18(1):130–43.
16. Macel M, Van Dam NM, Keurentjes JJ. Metabolomics: the chemistry between ecology and genetics. Mol Ecol Resour. 2010;10(4):583–93.
17. Gadelha MR, Kasuki L, Korbonits M. The genetic background of acromegaly. Pituitary. 2017;20(1):10–21.
18. Bicikova M, et al. Aminothiols in human brain tumors. Clin Chem Lab Med. 2006;44(8):978–82.
19. Pinzariu O, Georgescu B, Georgescu CE. Metabolomics-a promising approach to pituitary adenomas. Front Endocrinol (Lausanne). 2018;9:814.

20. Lee SH, Nam SY, Chung BC. Altered profile of endogenous steroids in the urine of patients with prolactinoma. Clin Biochem. 1998;31(7):529–35.
21. Rostomyan L, et al. Clinical and genetic characterization of pituitary gigantism: an international collaborative study in 208 patients. Endocr Relat Cancer. 2015;22(5):745–57.
22. Iacovazzo D, et al. Germline or somatic GPR101 duplication leads to X-linked acrogigantism: a clinico-pathological and genetic study. Acta Neuropathol Commun. 2016;4(1):56.
23. Boguslawska A, Korbonits M. Genetics of acromegaly and gigantism. J Clin Med. 2021;10(7):1377.
24. Yu H, et al. Metabolic profiling of acromegaly using a GC-MS-based nontargeted metabolomic approach. Endocrine. 2020;67(2):433–41.
25. Fellinger P, et al. Increased ATP synthesis might counteract hepatic lipid accumulation in acromegaly. JCI Insight. 2020;5(5):e134638.
26. Biagetti B, et al. Metabolic fingerprint of acromegaly and its potential usefulness in clinical practice. J Clin Med. 2019;8(10):1549.
27. Wang M, et al. High-performance liquid chromatography-mass spectrometry-based lipid metabolite profiling of acromegaly. J Clin Endocrinol Metab. 2020;105(3):e1075.
28. De San-Martin BS, et al. Metabolomics as a potential tool for the diagnosis of growth hormone deficiency (GHD): a review. Arch Endocrinol Metab. 2021;64(6):654–63.
29. Molitch ME, et al. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2011;96(6):1587–609.
30. Argente J, et al. Genetics of growth disorders-which patients require genetic testing? Front Endocrinol (Lausanne). 2019;10:602.
31. Hoybye C, et al. Metabolomics: a tool for the diagnosis of GH deficiency and for monitoring GH replacement? Endocr Connect. 2014;3(4):200–6.
32. Abd Rahman S, et al. Urine metabonomic profiling of a female adolescent with PIT-1 mutation before and during growth hormone therapy: insights into the metabolic effects of growth hormone. Growth Hormon IGF Res. 2013;23(1–2):29–36.
33. Xu R, et al. Metabolomic analysis reveals metabolic characteristics of children with short stature caused by growth hormone deficiency. Clin Sci (Lond). 2019;133(6):777–88.
34. Mastrangelo A, et al. Metabolomics changes in patients with PAPP-A2 deficiency in response to rhIGF1 treatment. Growth Hormon IGF Res. 2018;42-43:28–31.
35. Narduzzi L, et al. Applying metabolomics to detect growth hormone administration in athletes: proof of concept. Drug Test Anal. 2020;12(7):887–99.
36. Nieman LK, et al. The diagnosis of Cushing's syndrome: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2008;93(5):1526–40.
37. Oklu R, et al. Identification of small compound biomarkers of pituitary adenoma: a bilateral inferior petrosal sinus sampling study. J Neurointerv Surg. 2014;6(7):541–6.
38. Arlt W, et al. Steroid metabolome analysis reveals prevalent glucocorticoid excess in primary aldosteronism. JCI Insight. 2017;2(8):e93136.
39. Di Dalmazi G, et al. Cortisol-related metabolic alterations assessed by mass spectrometry assay in patients with Cushing's syndrome. Eur J Endocrinol. 2017;177(2):227–37.
40. Eisenhofer G, et al. Plasma steroid metabolome profiling for diagnosis and subtyping patients with Cushing syndrome. Clin Chem. 2018;64(3):586–96.
41. Vega-Beyhart A, et al. Endogenous cortisol excess confers a unique lipid signature and metabolic network. J Mol Med (Berl). 2021;99(8):1085–99.

Metabolomics and Genetics of Rare Endocrine Disease: Adrenal, Parathyroid Glands, and Cystic Fibrosis

Afshan Masood, Abeer Malkawi, Mohamed Siaj, and Anas M. Abdel Rahman

Abstract Recent advances in metabolomic technologies and methodologies have identified significant metabolites related to rare endocrine disease conditions of the adrenal gland (hyperaldosteronism, primary adrenal insufficiency), parathyroid (hypoparathyroidism), and cystic fibrosis. Metabolomic profiling combined with genomics is increasingly being employed for improving understanding, clinical diagnosis, and management of these clinically challenging conditions. Advances in gas and liquid chromatography combined with tandem mass spectrometry (GC/LC–MS/MS) techniques have improved the profiling of steroid metabolites. Significant alterations in levels of these metabolites demonstrate the potential to serve as specific markers of disease, help in their stratification, and contribute toward moving to personalized medicine.

Keywords Rare endocrine disease · Genetic disease · Metabolomics · Primary aldosteronism · Primary adrenal insufficiency hyperaldosteronism · Parathyroidism Pheochromocytoma · Paragangliomas · Cystic fibrosis

A. Masood · A. Malkawi · M. Siaj (✉)
Department of Chemistry and Biochemistry, Université du Québec à Montréal (UQAM), Montreal, QC, Canada
e-mail: siaj.mohamed@uqam.ca

A. M. Abdel Rahman (✉)
Metabolomics Section, Department of Clinical Genomics, Center for Genome Medicine, King Faisal Specialist Hospital & Research Center (KFSHRC), Riyadh, Saudi Arabia
e-mail: AabdelRahman46@kfshrc.edu.sa

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_9

Abbreviations

ATP1A1	ATPase Na+/K + -transporting subunit alpha 1
ATP2B3	ATPase plasma membrane Ca2+ transporting 3
CACNA1H	Calcium voltage-gated channel subunit alpha 1 H
CACNA1H	Calcium voltage-gated channel subunit alpha 1 H
CASR	G protein-coupled calcium-sensing receptor
CCND1/PRAD1	Cyclin D1
CDKN1C	Cyclin-dependent kinase inhibitor 1C
CFTR	CF transmembrane conductance regulator
CHD7	Chromodomain helicase DNA binding protein 7
CLCN2	Chloride voltage-gated channel 2
CSDE1	Cold shock domain-containing E1
CTNNB1	Catenin beta 1
CYP11B2	Cytochrome P450 family 11 subfamily B member 2
DAX-1 (NR0B1) SF-1	Nuclear receptor subfamily 0 group B member 1
DLST	Dihydrolipoamide S-succinyltransferase
FH	Fumarate hydratase
GATA3	GATA binding protein 3
GCM2	Glial cells missing transcription factor 2
GNA11	G protein subunit alpha 11
H3F3A	H3.3 histone A
HIF2A	Hypoxia-inducible factor 1 subunit alpha
HRAS	HRas proto-oncogene, GTPase
IDH	Isocitrate dehydrogenase (NADP(+)) 1
IRP1	Iron regulatory protein
KCNJ5	Potassium inwardly rectifying channel subfamily J member 5
MAML3	Mastermind-like transcriptional coactivator 3
MDH2	Malate dehydrogenase 2
NF1	Neurofibromin 1
NR5A1	Nuclear receptor subfamily 5 group A member 1
P450scc/CYP11A1	Cytochrome P450 family 11 subfamily A member 1
PHD1	Prolyl hydroxylase 1
POLE1	DNA polymerase epsilon, catalytic subunit
PTH	Parathyroid hormone
RET	Ret proto-oncogene
SAMD9	Sterile alpha motif domain containing 9
SDHx	Succinate dehydrogenase complex iron-sulfur subunit B
SEMA3E	Semaphorin 3E
SGPL1	Sphingosine-1-phosphate lyase 1
SLC25A11d	Solute carrier family 25 member 13
SOX3	SRY-Box transcription factor 3
TMEM127	Transmembrane protein 127
VHL/EPAS	Von Hippel–Lindau tumor suppressor

1 Introduction

Over the course of the past 10 years, research based on metabolomics has grown significantly and emerged as a promising instrument for clinical diagnostics as well as for improving our comprehension of the physiological and pathological processes that are the foundation of study for endocrine-related and rare diseases. In this chapter, we looked at how metabolomics helped diagnose, treat, and follow up rare endocrine disease of the pituitary gland. Beyond the pituitary, metabolomics has also been applied in disease stratification and management and in identifying biomarkers with applications in disease prediction, diagnosis, prognosis, and therapy.

2 Metabolomics of Adrenal Dysfunction

The adrenal glands play an important role in regulation of body homeostasis. Anatomically, they are made up of the cortex and medulla that secrete hormones involved in maintaining electrolyte and mineral balance, control metabolic pathways, provide response to stress (cortisol production in the adrenal cortex and catecholamines in adrenal medulla), and are crucial for sexual differentiation (through producing steroid hormones in the adrenal cortex). Diseases of the adrenal glands result in the resistive synthesis of glucocorticoids, sex hormones, and catecholamines (epinephrine and norepinephrine). Excessive circulating glucocorticoid (cortisol) levels, independent of ACTH, primarily arise from the adrenal gland disease (CS), while increased secretion of corticotropin (ACTH)-dependent cortisol is primarily due to disease of the pituitary and in some cases due to other glands. CS accounts for 15% of the cases while a majority is 70% due to CD and other causes including ectopic production is 15% [1]. The associated genetic defects and metabolite changes related to endogenous and exogenous hypercortisolism were covered in the previous chapter. Adrenal gland dysfunction also results in disorders of aldosterone synthesis (hyperaldosteronism or Conn's syndrome and adrenal insufficiency or Addison's disease), steroidogenesis, and the synthesis of sex hormones. Rare forms of these conditions arise due to germline mutations resulting in benign adrenal tumors or adrenocortical adenomas having an overall incidence in the general population of 1–2 cases per million [2]. Clinical evaluation of these disorders of adrenal steroidogenesis and disease requires measurement of specific hormonal levels, radiological imaging, and histopathology of the biopsy specimens. Advances in GC–MS and LC–MS/MS techniques have greatly improved the diagnosis of these diseases by metabolomic identification and quantification of the steroid metabolome that includes steroid hormones along with metabolic derivatives in bodily fluids (e.g., serum, urine) for clinical (diagnostics and treatment monitoring) as well as research purposes.

2.1 Hyperaldosteronism (Conn's Syndrome)

Primary aldosteronism (PA) is characterized by inappropriate and excessive secretion of the adrenal steroid hormone and aldosterone (hyperaldosteronism), the main mineralocorticoid hormone responsible for salt and water reabsorption, as well as increased potassium and proton secretion from the kidneys. Among the primary causes of PA is aldosterone-producing adenomas that present clinically as secondary endocrine hypertension. The nonneoplastic rare causes of PA (5%) are due to familial hyperaldosteronism (FH I–IV) and bilateral adrenal hyperplasia (BAH) [3, 4]. The advent of next-generation sequencing (NGS) technology and its wider application determined a largely genetic basis for the rare (5%) causes of PA through the identification of overlapping set of genes carrying numerous disease-causing germline mutations. These genome-wide association studies (GWAS) identified germline variants in CACNA1H (encoding a subunit of T-type voltage-gated calcium channel, CaV3.2), KCNJ5 (potassium inwardly rectifying channel subfamily J member 5), CYP11B2 (encoding aldosterone synthase), CACNA1D (calcium channel voltage-dependent L-type alpha-1D subunit), and CLCN2 (encoding voltage-gated chloride channel ClC-2) [5–8]. Metabolomic approaches and metabolome profiling using gas chromatography–mass spectrometry (GC–MS) and ultra-HPLC–tandem mass spectrometry (UHPLC–MS/MS) have comprehensively profiled the steroid metabolite profiling in sera and urine of patients and have found applications in the areas of personalized medicine for diagnostic purpose and prediction of prognosis. Coupling metabolomics together with genomics using GWAS provides a platform for employing integrated OMICS toward understanding and identifying the clinical phenotype. These genetic variations with their resulting metabolic alterations have created the metabolic phenotypes termed "genetically determined metabotypes" that is now being considered as a diagnostic feature [9]. Clinically, the characteristic presentation of PA is an increase in blood pressure, that is, hypertension along with disturbances in the electrolyte levels. When compared to patients with primary hypertension, these patients are at a higher risk of developing cardiovascular and kidney disease [10, 11], making an early diagnosis vital. Aside from these complications, there is also the need to differentiate between the different PA subtypes and unilateral or bilateral disease as clinical management of both conditions differs; the former is managed surgically while the latter is managed medically.

A metabolomic approach using liquid chromatography with tandem mass spectrometry (LC–MS/MS) successfully quantified adrenal steroids. The multi-steroid signatures associated with steroid biosynthesis and metabolism disorders showed a high level of diagnostic accuracy to differentiate between PA and adrenal hyperplasia that have similar presentations. Levels of cortisol derivatives (18-hydroxycortisol and 18-oxocortisol) were elevated in patients with PAs, carrying KCNJ5 mutations, in comparison to those with adrenal hyperplasia. Urinary 18-hydroxycortisol showed a high accuracy in distinguishing between the two conditions [12]. On the other hand, patients with adrenal hyperplasia showed elevated levels of plasma dehydroepiandrosterone (DHEA), DHEA-S, cortisol, and corticosterone [13].

Targeted metabolomic analysis by LC-MS/MS using a 32-metabolite steroid panel that included 11-deoxycorticosterone, aldosterone, cortisol, 11-deoxycortisol, 21-deoxycortisol, corticosterone, progesterone, 17-hydroxyprogesterone, 18-oxocortisol, 18-hydroxycortisol, cortisone, pregnenolone, androstenedione, DHEA, and DHEA-S was also used. Levels of 18-oxocortisol and 18-hydroxycortisol were found to have distinctively higher levels in cases of FH I and III compared to controls and also showed a high level of correlation with histological features of adenoma. In addition, mutations in *CTNNB1*, coding for β-catenin, have been identified in 2–5% of cases with aldosterone-producing benign adenoma. A somatic mutation in *CLCN2*, coding for the chloride channel ClC-2 (chloride channel protein 2) mutated in familial hyperaldosteronism type II (FH-II) and early-onset PA, has recently been identified. Previous studies have shown that CYP11B2 can convert 11-deoxycortisol efficiently to 18-hydroxycortisol and 18-oxocortisol, while CYP11B1 can synthesize only 18-hydroxycortisol [5, 14, 15] (Table 1).

Hyperaldosteronism is suggested to cause inflammation and metabolic dysregulation and contribute to development of cardiovascular disease. Metabolomic profiling in patients with Conn's disease revealed significant alterations in levels of triglyceride concentrations, large VLDL particles with urate concentrations, and derivatives of the linoleic acid metabolism pathway [16]. Steroid profiling has also revealed high production of the "hybrid" cortisol metabolites 18-hydroxycortisol and 18-oxocortisol in patients with rare, familial forms of PA associated with specific genetic errors (*CYP11B1*/*CYP11B2* hybrid gene, *KCNJ5* mutations) [8, 17, 18]. The levels of these hybrid metabolites also served as markers to differentiate between PA and BAH. Specifically, the 18-oxocortisol/cortisol ratio in adrenal vein samples and urinary 18-hydroxycortisol levels showed sufficient diagnostic accuracy to distinguish APAs from BAHs of patients [19, 20]. In addition to the clear elevation of plasma 18-oxocortisol in PAs, increased levels of plasma cortisol, corticosterone, DHEA, and DHEA-S were documented in patients with BAH [13]. Differences between the metabolite levels among the varying subtypes were also demonstrated using in situ metabolomics and demonstrated that levels of 18-oxocortisol and 18-hydroxycortisol negatively correlate with the CYP11B1. The steroid profiles were also correlated with their respective genotypes. The PAs carrying KCNJ5 mutations presented significantly higher levels of 18-oxocortisol in both adrenal vein and peripheral plasma samples than all other PAs, while PAs harboring ATPase mutations displayed the highest peripheral concentrations of aldosterone, cortisol, 11-deoxycorticosterone, and corticosterone. At the same time, patients with CACNA1D-mutated PAs had lower aldosterone and corticosterone concentrations than all other groups [21]. Distinct molecular signatures between KCNJ5- and CACNA1D-mutated PAs involving metabolites of steroidogenesis as well as purine metabolism KCNJ5 carriers displayed significantly higher levels of 18-hydroxycortisol and 18-oxocortisol when compared to CACNA1D carriers. Activation of purine metabolism was observed in KCNJ5 mutant APAs, with a significant increase in adenosine monophosphate and diphosphate [14].

Table 1 The different endocrine glands with the associated disease, the gene locus, and the major metabolites altered with the disease

Endocrine gland	Hormones	Disease	Genes and chromosomal locus	Metabolomic platform and approach	Major significant metabolites	References
Adrenal gland	Aldosterone	Primary hyperaldosteronism	*KCNJ5*, *CACNA1D*, *ATP1A1*, *ATP2B3*, *CTNNB1*, *CYP11B2*, *CLCN2*, and *CACNA1H*	*LC–MS/MS (targeted) and steroid profiling using GC-MS and LC–MS/MS (untargeted)*	*18-Hydroxycortisol and 18-oxocortisol*, *tetrahydro-18-oxocortisol*, *aldosterone*, *cortisol*, *11-deoxycorticosterone, and corticosterone* (↑) *Long-chain acylcarnitines C18:1, C18:2, ornithine, and spermidine*	[13, 17]
	ACTH decrease	Adrenal insufficiency	*DAX-1 (NR0B1) SF-1, NR5A1, CDKN1C, SAMD9, POLE1, P450scc/ CYP11A1, and SGPL1*	*GC-MS, LC-MS/MS, 1HNMR, and steroid profiling (targeted)*	*Total cortisol metabolites, 11β-HSD1 11β-HSD2, and 5α- and 5β-reductase* (↓)	[17, 24, 27]
		Pheochromocytoma/ paraganglioma	*NF1, RET, TMEM127, HRAS, SDHx, FH, IDH, H3F3A, HIF2A, MDH2, PHD1, IRP1, SLC25A11d, DLST, VHL/EPAS, and CSDE1 MAML3*	ESI–LC–MS/MS (targeted), *LC–MS*	*Succinate, citrate, isocitrate, cis-aconitate* (↑), *fumarate, and 2-hydroxyglutarate* (↓) *2-Oxoglutarate/malate carrier and glutamic-oxaloacetic transaminase 2* (↓)	[29, 38]
Parathyroid	PTH	Hypoparathyroidism	*GCM2, SOX3, CHD7, SEMA3E, GATA3, PTH, CASR, and GNA11*	*UHPLC–MS (untargeted)*	*Adenosine, inosine, hypoxanthine, guanosine, and xanthine* (↓) *N-acetyl aspartate*	[45]

Pancreas		Cystic fibrosis	*CFTR*	*GC/LC–MS (untargeted)* *UHLC–MS/MS and GC–MS (untargeted)*	*Sorbitol and cholesteryl esters* *Purine nucleotides, adenosine, inosine, hypoxanthine, and guanosine (↑)* *Oxidized glutathione, S-lactoylglutathione, S-nitrosoglutathione, and ophthalmate (↓)*	[21, 49, 50, 55–57]

ATP1A1 ATPase Na+/K + -transporting subunit alpha 1, *ATP2B3* ATPase plasma membrane Ca2+ transporting 3, *CACNA1D* calcium voltage-gated channel subunit alpha 1 D, *CACNA1H* calcium voltage-gated channel subunit alpha 1 H, *CACNA1H* calcium voltage-gated channel subunit alpha 1 H, *CASR* G protein-coupled calcium-sensing receptor, *CDKN1C* cyclin-dependent kinase inhibitor 1C, *CFTR* CF transmembrane conductance regulator, *CHD7* chromodomain helicase DNA binding protein 7, *CLCN2* chloride voltage-gated channel 2, *CSDE1* cold shock domain containing E1, *CTNNB1* catenin beta 1, *CYP11B2* cytochrome P450 family 11 subfamily B member 2, *DAX-1 (NR0B1) SF-1* nuclear receptor subfamily 0 group B member 1, *DLST* dihydrolipoamide S-succinyltransferase, *FH* fumarate hydratase, *GATA3* GATA binding protein 3, *GCM2* glial cells missing transcription factor 2, *GNA11* G protein subunit alpha 11, *H3F3A* H3.3 histone A, *HIF2A* hypoxia-inducible factor 1 subunit alpha, *HRAS* HRas proto-oncogene, *GTPase, IDH* isocitrate dehydrogenase (NADP(+)), *1IRP1* iron regulatory protein, *KCNJ5* potassium inwardly rectifying channel subfamily J member 5, *MAML3* mastermind-like transcriptional coactivator 3, *MDH2* malate dehydrogenase 2, *NF1* neurofibromin 1, *NR5A1* nuclear receptor subfamily 5 group A member 1, *P450scc/CYP11A1* cytochrome P450 family 11 subfamily A member 1, *PHD1* prolyl hydroxylase, *POLE1* DNA polymerase epsilon, catalytic subunit, *PTH* parathyroid hormone, *PTPN22* protein tyrosine phosphatase non-receptor type 22, *RET* Ret proto-oncogene, *SAMD9* sterile alpha motif domain containing 9, *SDHx* succinate dehydrogenase complex iron sulfur subunit B, *SEMA3E* semaphorin 3E, *SGPL1* sphingosine-1-phosphate lyase 1, *SLC25A11d* solute carrier family 25 member 13, *SOX3* SRY-Box transcription factor 3, *TMEM127* transmembrane protein 127, *UPLC–MS/MS* ultra-PLC–tandem mass spectrometry, *VHL/EPAS* Von Hippel-Lindau tumor suppressor

2.2 Primary Adrenal Insufficiency

Primary adrenal insufficiency (PAI), a deficiency of glucocorticoid and mineralocorticoid production, is a relatively rare life-threatening condition due to autoimmune disorders and enzymatic defects. The patients clinically present with skin and mucous membrane hyperpigmentation, craving for salt, failure to thrive, depression, and fatigue. PAI is caused due to pathology within the adrenal glands which results in stimulation of the hypothalamo–pituitary axis and the renin–angiotensin–aldosterone system regulatory feedback loop. PAI is typically diagnosed by measuring levels of ACTH and proopiomelanocortin peptides, which are elevated in addition to inappropriately low cortisol secretion. A delayed diagnosis of PAI is linked with an adverse quality of life and raises the patient's risk of an adrenal crisis that might be fatal. Recent studies have implicated several genetic mutations in the pathophysiology of the disease. These include defects in the nuclear receptors DAX-1 (NR0B1), steroidogenic factor-1 (SF-1/NR5A1), CDKN1C and SAMD9 or loss of POLE1, P450scc/CYP11A1 insufficiency, and sphingosine-1-phosphate lyase-1 (SGPL1) defects [22]. Treatment of PAI conventionally is modeled around corticosteroid replacement therapy that is conventionally administered three times a day [23].

Metabolomic studies in PAI are limited in the literature. In a study, metabolite profiling was carried out in the sera or urine for disease identification and stratification and for evaluating optimal replacement therapy. The natural circadian rhythm of cortisol cycle cannot be entirely replicated by current glucocorticoid replacement regimens, which leads to either over- or under-replacement. The urinary cortisol metabolome was assessed to determine optimal cortisol replacement in patients with PAI. The metabolic profile of patients using two hydrocortisone replacement therapies were compared, namely, the once-a-day dual-release hydrocortisone (DHC) and three-times-a-day hydrocortisone (TID-HC) therapy. In the 24-h urine samples, total cortisol metabolites decreased after DHC therapy compared to TID-HC and were more in line with the usual control levels. 11-β-Hydroxysteroid dehydrogenase (11β-HSD) type 1 activity dropped with DHC compared to TID-HC therapy, whereas 11-β HSD2 activity fell with TID-HC but returned to normal with DR-HC. Moreover, 5α- and 5β-reduced metabolites were decreased with DR-HC compared to TID-HC. Patients undergoing traditional TID-HC replacement treatment with enhanced 11β-HSD1 activity exhibits significant alterations in the urine cortisol metabolome, which may explain the adverse metabolic profile in patients with PAI. Its shift toward normalcy with DHC therapy might serve as an indicator for more favorable metabolic outcomes [17, 24]. Aside from providing means for optimizing therapeutic dosage, the metabolomic approach was also used to measure the effects of glucocorticoid therapy and identify biomarkers related to its action. Serum metabolic profiling was also undertaken in PAI patients, using GC–MS and LC–MS, during glucocorticoid therapy and after its withdrawal to assess response to therapy. The differentially expressed metabolites identified were amino acids (tyrosine, tryptophan, asparagine), malic acid, lactic acid, and uracil. The metabolism of tryptophan, which modulates mood and energy homeostasis, is regulated by

glucocorticoids through the kynurenine pathway. Metabolomic analysis was able to identify that administering high doses of glucocorticoid, especially after a treatment of 10-week treatment, resulted in decreased tryptophan levels by influencing the kynurenine pathway [25, 26]. Hence, metabolomics assisted in assessing the therapeutic effects, allowing for the individualization of approaches and optimization of glucocorticoid therapy.

2.3 Metabolomics of Pheochromocytoma

Pheochromocytomas (PCC) and paragangliomas (PPGLs) are a group of rare heterogeneous neuroendocrine tumors that arise from either the adrenal medullary chromaffin cells or from outside the adrenal gland in the neural crest cells (sympathetic and parasympathetic paraganglia). Most PPGLs are benign tumors, with an incidence of approximately of one per million population per year [27]. The characteristic clinical phenotype of these patients is associated with features of excess circulating catecholamine levels due to increased synthesis or release. The presenting signs and symptoms classically range from a triad consisting of sweating, headaches, and palpitations to nonspecific symptoms such as weight loss, nausea, tiredness, or flushing [27, 28]. The diverseness and nonspecificity of the clinical manifestations, heterogeneity of these tumors regarding the age of presentation, and differences in their location make an early clinical diagnosis of PCC difficult [29]. PCC and PPGLs have the highest degree of heritability, where PPGLs carrying a germline mutation account for 30–40%. More than 20 susceptibility genes with varying mutations have been identified as predisposing factors to this condition, placing it among the rare genetic endocrine conditions. The metabolic phenotypes in PCC are based on the affected specific gene/protein [30] that determines the secretory profile, molecular features, metabolic changes, clinical outcomes, and potential for malignancy [31].

PCC and PPGLs have been classified based on their inheritance, multiple endocrine neoplasia type 2 (MEN2), familial Von Hippel–Lindau (VHL) syndrome and less commonly neurofibromatosis type 1 (NF1)) or sporadic [30], or by molecular pathway subtypes, kinase signaling subtype (RET, transmembrane protein 127 (TMEM127), mutations in the NF1, and HRAS genes), pseudohypoxia (Von Hippel–Lindau (VHL/EPAS)-related and tricarboxylic acid cycle (TCA)-related mutations, and Wnt-altered subtype (CSDE1 somatic mutations and mastermind-like transcriptional coactivator 3 (MAML3) fusion genes)). The TCA-related PPGL subtype consists of tumors having mutations in the succinate dehydrogenase subunits A–D (SDHx), fumarate hydratase (FH), and isocitrate dehydrogenase (IDH). In addition, other genetic mutations in the H3F3A, malate dehydrogenase 2 (MDH2), PHD1, IRP1, SLC25A11, and DLST were identified with a lower frequency and have not yet been included in the Cancer Genome Atlas. PCC and PPGLs are also classified into clusters based on their secretory profile as adrenergic and noradrenergic clusters. The noradrenergic pseudohypoxic phenotype (secreting

norepinephrine and normetanephrine (NMN)) constitutes cluster 1. It includes tumors with SDHx mutations, along with VHL, FH, (MDH2), hypoxia-induced factor (HIF2α), and IDH mutations and the newly identified SLC25A11 [32]. Tumors with the adrenergic phenotype (secreting epinephrine and metanephrine (MN)), which are associated with abnormal kinase signaling pathways and include mutations in the genes rearranged during transfection (RET), NF1, TMEM127, kinesin family member 1B (KIF1B), and MYC-associated factor X (MAX), make up cluster 2. Cluster 3 is associated with the Wnt signaling pathway; it includes somatic mutations of cold shock domain-containing E1 (CSDE1) and mastermind-like transcriptional coactivator 3 (MAML3) fusion genes [28, 32].

The gold standard for diagnosis relies on biochemical measurements of urinary or plasma products of the catecholamine degradation, noradrenaline (MN), adrenaline (NMN), and dopamine (methoxytyramine (MTY)) [33]. Measurements of plasma-free MN have been proven in several independent investigations to have diagnostic sensitivity surpassing 96% and specificity between 85 and 100%. An alternate method with a comparable degree of diagnostic sensitivity is provided by urinary-fractionated MN [28]. The measurement of plasma and urine MN by LC–MS/MS is presently widely accepted in the USA and many other laboratories as the gold standard approach [28, 34]. In addition to laboratory measurement, all patients with documented PCC and PPGLs should have genetic determination of PPGL phenotype as part of the diagnostic panel. It is very common to find mixed phenotypes of both adrenergic and dopaminergic secreting tumors in comparison to either adrenergic or dopaminergic ones. Each of these phenotypes has been linked to mutations in different genes. The genetic mutations in TMEM127 gene have been associated with only the adrenergic tumors while mutations in the KIF1B, MAX, RET, and NF1 genes have been associated with tumors with adrenergic mixed phenotype. On the other hand, it is known that extra-adrenal PPGLs having noradrenergic and dopaminergic phenotypes have mutations in PHD1/PHD2, HIF2A, SDHx, SDHAF2, FH, and IDH genes. The majority of PPGL tumors with HIF2A and VHL-mutated are typically noradrenergic while predominantly dopaminergic secreting tumors are known to be commonly associated with SDHx mutations [33]. Although confirmatory, genetic testing can be complex and, in many cases, unavailable at all centers. This potentially leads to delayed or inconclusive diagnosis [30, 35, 36]. Due to the probable increased risk of metastatic illness in these patients, it becomes crucial for an effective clinical management to distinguish early on between tumors with underlying germline mutations and those that are sporadic [37]. In these instances, quantifying metabolites can help verify functionality and identify underlying mechanisms and factors for germline or somatic mutations in patients with unresolved genetic testing results.

Metabolomic studies have helped to bridge this gap by identifying metabolites that have not only helped in detailing the metabolic pathways affected by the disease but also to differentiate between the phenotypes, stratify the disease, and propose metabolites that are amenable to diagnostic applications. The tumor metabolomic profile distinguishes these different subtypes of tumors to classify patients with PGLs as sporadic or hereditary. Untargeted metabolomic approaches

aided in profiling the disease pathology and associating the changes with the different variants of genetic mutations. On the other hand, targeted metabolomic approaches using the identified metabolites have also been studied to determine their impact on metabolism and utilize them as diagnostic markers in clinical laboratories [29] for monitoring therapeutic response potential conversion metastases [38]. Surgical resection of the PPGL with normalization of catecholamine levels was associated with significant changes in the metabolites. Following surgery levels of glycerophospholipids (phosphatidylcholine diacyl (PC aa) 42:0, phosphatidylcholine acyl-alkyl (PCae) 42:5, PCae (44:5), and PCae (44:6) and hexoses were lower, while levels of amino acids (biogenic amines), namely, histidine and creatinine, were demonstrated to be higher [29]. The metabolomic profile in each of the different genetic variants of these tumors was also deciphered. Around 15–25% of all PCC/PPGLs were linked to defects in the Krebs cycle enzymes, SDH, FH, MDH, and IDH, with SDH faults, being the most frequent. The 2-oxoglutarate/malate carrier, glutamic-oxaloacetic transaminase 2, and others have more recently been linked to hereditary PPGL as regulators of mitochondrial metabolites [36]. The PCC/PGLs associated with mutations in the pseudo hypoxic cluster (cluster 1) were associated with the hypoxia-inducible factor (HIF) signaling pathway and involved mutations in genes encoding the HIF2A, succinate dehydrogenase subunits or their assembly factors (SDHx [SDHA, SDHB, SDHC, SDHD]), succinate dehydrogenase complex assembly factor 2 (SDHAF2), Von Hippel–Lindau tumor suppressor (VHL), and egl-9 prolyl hydroxylases 1 and 2 (EGLN1/EGLN2). The pathogenic mutations in these genes lead to an accumulation of their related metabolites, that is, succinate, fumarate, or 2-hydroxyglutarate, which in turn were responsible for tumor development.

Distinct differences were noted in metabolites and pathways related to oxygen sensing, hypermethylation, DNA repair, and overexpression of certain transporters and receptors; notably Krebs cycle enzymes have been through the genetic investigations in PCC/PGL tumors [39–41]. Metabolomic profiling identified differential regulation of metabolites between the various genetic causes of PCC/PGL. Metabolomic analysis of PCC/PGL arising from mutations in SDHx revealed a decrease in activity of SDH (mitochondrial electron transport chain complex II) enzyme and other TCA cycle metabolites, including fumarate glutamate and aspartate with elevated succinate levels [37]. The ratio of two metabolites, succinate to fumarate, was determined as a novel metabolic marker to detect paraganglioma with underlying SDHB/D mutations [38, 42]. Moreover, in tumors linked to SDHx mutations, glutamate levels and ATP/ADP/AMP values were shown to be lower with modest but significant changes in levels of histidine, threonine, and lysoPC (C28:0) [32, 42]. Significant correlations were also noted between plasma MN and total urine catecholamine levels with the sum of detected hexoses (reflecting glucose) which were found [33], along with the increase in levels of glutamine [40]. Significant alterations were noted in isocitrate, cis-aconitate, and citrate levels in patients with mutations in FH. In contrast, IDHx mutations were characterized by higher citrate, isocitrate, and cis-aconitate levels [37]. These differences could serve as potential biomarkers for early diagnosis of disease.

3 Metabolomics of Parathyroid Dysfunction

The parathyroid glands are four small pea-sized glands behind the thyroid gland secreting parathyroid hormone (PTH). The primary endocrine glands maintain calcium and phosphorus homeostasis with other hormones, including vitamin D and fibroblast growth factor (FGF23). PTH regulation occurs mainly between three organs, the intestine, kidney, and bone. A complex interplay occurs between PTH, active vitamin D (1,25(OH)2D), and calcium sensor receptors (CaSRs) that maintain serum calcium concentration within a narrow physiological range to maintain mineral homeostasis. Dysfunction of the parathyroid glands occurs as a primary disease of the gland or secondary to other diseases such as chronic kidney disease. Parathyroid dysfunction results in inappropriate parathyroid hormone (PTH) production, resulting in abnormal calcium homeostasis. Phenotypically, it can manifest as either an increase, hyperparathyroidism, or a decrease, hypoparathyroidism, in circulating PTH levels. Primary parathyroid dysfunction is relatively rare compared to secondary causes and can be seen as an isolated condition or component of a complex endocrine syndrome. Hypocalcemia and hyperphosphatemia are the characteristic hallmarks of primary hypoparathyroidism, which is caused by inadequate quantities of circulating parathyroid hormone.

3.1 Hypoparathyroidism

Incidental hypoparathyroidism is a rare disorder with an estimated prevalence of 0.25 per 1000 individuals. Hypoparathyroidism is clinically characterized by decreased parathyroid hormone levels resulting in hypocalcemia that directly impacts calcium and phosphorus homeostasis and the bone. It can occur as an isolated condition or as part of a complex endocrine syndrome. The most common cause of hypoparathyroidism is transient postsurgical hypoparathyroidism resulting from a functional impairment or surgical resection of parathyroid glands after acute manipulation during neck surgery. Other causes include impairment of PTH action, pseudo-hypoparathyroidism, and genetic causes. Numerous somatic or germline mutations have been identified, leading to dysgenesis of the parathyroid gland or an inability of the parathyroid glands to secrete PTH. These include mutation in the PTH gene, GCM2, SOX3, CASR, GNA11, TBX1, CHD7, GATA3, and TBCE [43, 44]. Rare genetic defects involving the transient receptor potential ion channel (TRMP6) and tight-junction gene claudins 16 and 19 have been identified, resulting in the abnormal homeostasis of magnesium leading to hypoparathyroidism (23). The cause of low magnesium levels is attributed to nutritional deficiencies or chronic diseases, including T2DM, hypertension, and renal conditions, which either decrease secretion of PTH or increase resistance to the actions of PTH in the bone and kidneys. Metabolomic studies evaluating changes in the metabolites

within hypoparathyroidism are limited. A single study by Paprocka et al. in children with hypoparathyroidism identified alterations in N-acetyl aspartate by 1H magnetic resonance spectroscopy [45]. Further metabolomic studies are needed to identify the different metabolites altered with hypoparathyroidism.

4 Exocrine Pancreatic Dysfunction: Metabolomics of Cystic Fibrosis

Metabolomics has also been important in studying the pathology of cystic fibrosis, which leads to endocrine-related complications of the pancreas. Cystic fibrosis is a lethal autosomal recessive disorder arising from mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene expressed in the apical membranes of various epithelial cells. It is a cAMP-regulated channel that conducts ATP and regulates several apical membrane-associated channels, including the sodium, chloride, and potassium channel along with regulating release of bicarbonate. The disease represents an example of a monogenic defect with over 2000 mutations that leads to characteristic multisystemic disease. Besides the characteristic pulmonary manifestations, patients with CF show endocrine defects in the pancreas and the reproductive system. The most common cause of the pathology is blockage of endocrine ducts due to the thickened secretions. CF mutations can be grouped as those causing severe or mild disease and are further categorized as one of six classes; classes I–III represent severe disease, while mild mutations are classes IV–VI.

Recently, metabolomics has been utilized as an invaluable tool to study the changes in the metabolic profiles in CF, understand the pathophysiology, and elucidate the different metabolic pathways altered with CF [46, 47]. The gold standard for initial newborn screening is a measurement of immunoreactive trypsinogen (IRT) in dry blood spots (DBSs), followed by targeted CFTR mutation analysis and confirmation with abnormally elevated sweat chloride. Our laboratory identified significant differences in 26 metabolites involved in peroxisomal, amino acids, sorbitol, glycolysis, and mitochondrial metabolic pathways. A distinct and interesting finding was the decrease in the osmolyte sorbitol in adult patients with CF patients compared to healthy controls. In order to maintain correct cellular activities and cell survival, organic osmolytes are crucial for regulating cell volume and fluid balance. The perturbation in the sorbitol pathway was identified as a causative factor for the mucoviscidosis [48]. A reduction in the sorbitol levels, and glycerol phosphorylcholine, another osmolyte, was noted in an untargeted metabolomic analysis of primary human airway epithelial cell culture in CF patients. Additionally, significant alterations were noted in the purine nucleotides, adenosine, inosine, hypoxanthine, and guanosine, which may regulate cellular responses via purinergic signaling. Reductions were also seen in metabolites related to glutamate, including oxidized glutathione levels, in *S*-lactoylglutathione, *S*-nitrosoglutathione, and ophthalmate [49].

Metabolomic profiling identified differences in patients with different grades and severity of the disease and between the functional classes of CF. Distinct metabolites were identified that related to clinical phenotype and lung function. We identified specific metabolites between the different CF functional classes using chemical isotope-labeled LC–MS-based metabolomics. The metabolomic profile was assessed between CF and controls, between the different mutation classes of CF, and specifically among classes III and IV. Significant alterations were seen in glutathione, glutamine, glutamate, and arginine metabolism, amino acids, and di- and tripeptides. The significant metabolites include gamma-glutamylglutamic acid, 1-aminopropan-2-ol, cystathionine, ophthalmate, and serotonin. An above-average FEV1% level of lung function was associated with decreased glutamic acid and increased guanosine levels. Metabolomic profiling, between the three analyses, demonstrated alterations in several amino acids and dipeptides governing glutathione metabolism and identified two metabolites in common between the analyses. These metabolites, namely, 3,4-dihydroxymandelate-3-O-sulfate and 5-aminopentanoic acid, could serve as biomarkers for CF [50].

Moreover, serum metabolomics was employed to evaluate CF bacterial lung illness in the preform post-exacerbation stage and identify which systemically measurably connected pathways were impacted throughout recovery. Bile acids, amino acid metabolites generated from microorganisms, increases in the lipid classes of glycerophospholipid, glycerolipids, cholesterol, phospholipids, and the class of sphingolipids were among the compounds and pathways affected. The resolution of the exacerbation was characterized by alterations of the tryptophan–kynurenine pathway, decreased polyamines, a reduction in lipid markers such as fatty acids (n6/n3), and increased in nitric oxide pathway metabolites [51]. On the other hand, metabolites altered with acute pulmonary exacerbation in CF patients demonstrated lower essential amino acids, L-arginine, and oxoproline levels than healthy controls. This decrease was mainly attributed to the skeletal muscle wasting, poor protein intake, increased amino acid utilization, and decreased intestinal absorption of proteins leading to an overall protein-deficient state [52, 53].

In addition to the derangements in the amino acids, patients with CF also showed abnormal lipid metabolism for most lipid subclasses, with significant plasma elevations in odd-chain and polyunsaturated fatty acyl lipids and a decrease in the plasma levels of several species of lysophosphatidylcholine (18:0, 18:2, 20:3, and 20:5) and phosphatidylcholine (36:5, O-38:0, 38:4, 38:5, 38:6, and P-40:1). Plasma phospholipid signatures were found to discriminate between mild and severe forms of CF. In contrast, levels of phosphatidic acids and diacylglycerols were particularly affected by different genotypic mutation classes. A biomarker panel of five oxidized lipids successfully differentiated patients with reduced lung function. Four species of PC (36:3, 36:5, 38:5, and 38:6) were consistently downregulated in severe vs. mild patients, while sphingolipid SM(d18:0) was significantly increased in all patients [54]. The lung function of CF patients is often assessed through forced vital expiratory capacity (FEV1) measurements using FEV1% or FEV1/FVC ratio. Lipid

fractions of the PUFA (C20:3n-9, C20:5n-3, C22:5n-3, and C22:6n-3) positively correlated with FEV1, along with PC (32:2) and PC (36:4), and oleoyl ethanolamide was negatively correlated with FEV1 progression. Lower PC(32:2), PC(38:5), and C18:3n-3, triacylglycerols higher cholesterol, and cholesterol esters were noted in chronically infected patients [55–57].

Metabolomic analysis was also carried out in other body fluids, including sputum, saliva, sweat, urine, bronchoalveolar lavage fluid, and exhaled breath analysis. These studies were generally aimed at identifying the differences in metabolite patterns to unravel the underlying pathophysiological mechanisms of CF and evaluate the effectiveness of treatment modalities. A recent study identified the changes in the lung microbial composition through untargeted metabolomic analysis of the sputum and exhaled breath. Patients with homozygous Phe508del genotype usually receive treatment with combination therapy lumacaftor and ivacaftor. Lumacaftor targets CFTR class II mutations specifically, while Ivacaftor improves the gating (class III) or conduction (class IV) defect in the mutant channels.

CFTR modulators improve CFTR function significantly by partly restoring the function of the chloride channel and improving transport of epithelial fluid in the airways. Besides improving lung function, treatment with CFTR modulators alters the pulmonary microbiome by reducing the abundance of the bacteria, for example, *Pseudomonas aeruginosa*. Metabolomic analysis by GC-TOF/MS showed changes in concentrations of the metabolite phenyl pyruvate in the sputum. On the other hand, the breath metabolome showed alterations in volatile organic compounds such as 4-ethylbenzanoic acid 2-pentyl ester, suggesting a strong link between oxidative stress and inflammation [58]. Untargeted metabolomic profiling of sweat between carriers and cases showed significant alterations in purine derivatives, organic acids, dipeptides, amino acids, and amino acid derivatives, in affected patients, and alterations in levels of asparagine and glutamine, in asymptomatic patients [59]. Patients with CF also present with lung disease characterized by bronchial inflammation due to chronic bacterial infection. The resulting inflammatory response is predominantly dominated by neutrophils. Metabolomic studies were used to identify and quantify the metabolites in the bronchoalveolar lavage fluid samples from these patients. A targeted metabolomic approach identified and quantified metabolites related to proteins, metabolism of purines, polyamines, and nicotinamide which correlated strongly with the clinical markers and neutrophil counts [60]. In addition to these body fluids, the urine metabolomic profile was also studied. The urinary metabolome in CF although heterogeneous showed metabolic alteration that were distinct when compared to non-CF groups. A targeted metabolomic study in the urine revealed an altered methyl status and oxidative stress in children with CF using NMR. Additionally, a subgroup of these children with pancreatic insufficiency showed a considerable rise of phthalate chemicals in their urine NMR spectra in comparison to children with CF who did not have pancreatic insufficiency [47, 61, 62].

5 Conclusion and Future Perspectives

Metabolomics has slowly made inroads into many aspects of patient care and has shown its relevance in understanding disease pathophysiology, diagnosis, and therapeutic monitoring. It provides a bridge between knowledge accumulated from basic science to clinical research as it considers the individual's metabolic characteristics. Combining the clinical (phenotype) with the metabolomic and genomics data will aid the clinical decision-making process by providing more sensitive and specific analyte panels for diagnostic testing. The potential of this omic approach is to further advance in bringing an era of personalized medicine in endocrinology.

References

1. Nieman LK, Ilias I. Evaluation and treatment of Cushing's syndrome. Am J Med. 2005;118(12):1340–6.
2. Mourtzi N, et al. Unravelling the genetic basis of primary aldosteronism. Nutrients. 2021;13(3):875.
3. Young WF Jr. Diagnosis and treatment of primary aldosteronism: practical clinical perspectives. J Intern Med. 2019;285(2):126–48.
4. Monticone S, et al. GENETICS IN ENDOCRINOLOGY: the expanding genetic horizon of primary aldosteronism. Eur J Endocrinol. 2018;178(3):R101–11.
5. Azizan EA, et al. Somatic mutations in ATP1A1 and CACNA1D underlie a common subtype of adrenal hypertension. Nat Genet. 2013;45(9):1055–60.
6. Seidel E, Schewe J, Scholl UI. Genetic causes of primary aldosteronism. Exp Mol Med. 2019;51(11):1–12.
7. Funder JW, et al. The management of primary aldosteronism: case detection, diagnosis, and treatment: an endocrine society clinical practice guideline. J Clin Endocrinol Metab. 2016;101(5):1889–916.
8. Storbeck KH, et al. Steroid metabolome analysis in disorders of adrenal steroid biosynthesis and metabolism. Endocr Rev. 2019;40(6):1605–25.
9. Tokarz J, et al. Endocrinology meets metabolomics: achievements, pitfalls, and challenges. Trends Endocrinol Metab. 2017;28(10):705–21.
10. Hundemer GL, et al. Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. Lancet Diabetes Endocrinol. 2018;6(1):51–9.
11. Hundemer GL, et al. Renal outcomes in medically and surgically treated primary aldosteronism. Hypertension. 2018;72(3):658–66.
12. Spyroglou A, et al. Transcriptomics, epigenetics, and metabolomics of primary aldosteronism. Cancers (Basel). 2021;13(21):5582.
13. Eisenhofer G, et al. Mass spectrometry-based adrenal and peripheral venous steroid profiling for subtyping primary aldosteronism. Clin Chem. 2016;62(3):514–24.
14. Murakami M, et al. In situ metabolomics of aldosterone-producing adenomas. JCI Insight. 2019;4(17):e130356.
15. De Sousa K, et al. Genetic, cellular, and molecular heterogeneity in adrenals with aldosterone-producing adenoma. Hypertension. 2020;75(4):1034–44.
16. van der Heijden C, et al. Vasculometabolic and inflammatory effects of aldosterone in obesity. J Clin Endocrinol Metab. 2020;105(8):2719.
17. Papathomas TG, Nose V. New and emerging biomarkers in endocrine pathology. Adv Anat Pathol. 2019;26(3):198–209.

18. Ulick S, Chu MD. Hypersecretion of a new corticosteroid, 18-hydroxycortisol in two types of adrenocortical hypertension. Clin Exp Hypertens A. 1982;4(9–10):1771–7.
19. Nakamura Y, et al. 18-oxocortisol measurement in adrenal vein sampling as a biomarker for subclassifying primary aldosteronism. J Clin Endocrinol Metab. 2011;96(8):E1272–8.
20. Mulatero P, et al. 18-hydroxycorticosterone, 18-hydroxycortisol, and 18-oxocortisol in the diagnosis of primary aldosteronism and its subtypes. J Clin Endocrinol Metab. 2012;97(3):881–9.
21. Williams TA, et al. Genotype-specific steroid profiles associated with aldosterone-producing adenomas. Hypertension. 2016;67(1):139–45.
22. Fluck CE. MECHANISMS IN ENDOCRINOLOGY: update on pathogenesis of primary adrenal insufficiency: beyond steroid enzyme deficiency and autoimmune adrenal destruction. Eur J Endocrinol. 2017;177(3):R99–R111.
23. Nowotny H, et al. Therapy options for adrenal insufficiency and recommendations for the management of adrenal crisis. Endocrine. 2021;71(3):586–94.
24. Espiard S, et al. Improved urinary cortisol metabolome in addison disease: a prospective trial of dual-release hydrocortisone. J Clin Endocrinol Metab. 2021;106(3):814–25.
25. Chantzichristos D, et al. Identification of human glucocorticoid response markers using integrated multi-omic analysis from a randomized crossover trial. elife. 2021;10:10.
26. Sorgdrager FJH, et al. Hydrocortisone affects fatigue and physical functioning through metabolism of tryptophan: a randomized controlled trial. J Clin Endocrinol Metab. 2018;103(9):3411–9.
27. Hescot S, et al. Prognosis of malignant pheochromocytoma and paraganglioma (MAPP-Prono study): a European network for the study of adrenal tumors retrospective study. J Clin Endocrinol Metab. 2019;104(6):2367–74.
28. Darr R, et al. Pheochromocytoma—update on disease management. Ther Adv Endocrinol Metab. 2012;3(1):11–26.
29. Erlic Z, et al. Metabolic impact of pheochromocytoma/paraganglioma: targeted metabolomics in patients before and after tumor removal. Eur J Endocrinol. 2019;181(6):647–57.
30. Neumann HPH, Young WF Jr, Eng C. Pheochromocytoma and paraganglioma. N Engl J Med. 2019;381(6):552–65.
31. Fishbein L, et al. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. Cancer Cell. 2017;31(2):181–93.
32. Eijkelenkamp K, et al. Clinical implications of the oncometabolite succinate in SDHx-mutation carriers. Clin Genet. 2020;97(1):39–53.
33. Mercado-Asis LB, et al. Pheochromocytoma: a genetic and diagnostic update. Endocr Pract. 2018;24(1):78–90.
34. Lenders JW, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. J Clin Endocrinol Metab. 2014;99(6):1915–42.
35. Richter S, et al. Krebs cycle metabolite profiling for identification and stratification of pheochromocytomas/paragangliomas due to succinate dehydrogenase deficiency. J Clin Endocrinol Metab. 2014;99(10):3903–11.
36. Dwight T, et al. Metabolomics in the diagnosis of pheochromocytoma and paraganglioma. Horm Metab Res. 2019;51(7):443–50.
37. Richter S, et al. Metabolome-guided genomics to identify pathogenic variants in isocitrate dehydrogenase, fumarate hydratase, and succinate dehydrogenase genes in pheochromocytoma and paraganglioma. Genet Med. 2019;21(3):705–17.
38. Lendvai N, et al. Succinate-to-fumarate ratio as a new metabolic marker to detect the presence of SDHB/D-related paraganglioma: initial experimental and ex vivo findings. Endocrinology. 2014;155(1):27–32.
39. Chae YC, et al. Landscape of the mitochondrial Hsp90 metabolome in tumours. Nat Commun. 2013;4:2139.
40. Jochmanova I, Pacak K. Pheochromocytoma: the first metabolic endocrine cancer. Clin Cancer Res. 2016;22(20):5001–11.

41. de Cubas AA, et al. DNA methylation profiling in pheochromocytoma and paraganglioma reveals diagnostic and prognostic markers. Clin Cancer Res. 2015;21(13):3020–30.
42. Imperiale A, et al. A new specific succinate-glutamate metabolomic hallmark in SDHx-related paragangliomas. PLoS One. 2013;8(11):e80539.
43. Li D, et al. Heterozygous mutations in TBX1 as a cause of isolated hypoparathyroidism. J Clin Endocrinol Metab. 2018;103(11):4023–32.
44. Gordon RJ, Levine MA. Genetic disorders of parathyroid development and function. Endocrinol Metab Clin N Am. 2018;47(4):809–23.
45. Paprocka J, et al. Neurological picture and 1H MRS in 4 children with hypoparathyroidism. Przegl Lek. 2005;62(7):680–4.
46. Liessi N, et al. Proteomics and metabolomics for cystic fibrosis research. Int J Mol Sci. 2020;21(15):5439.
47. Muhlebach MS, Sha W. Lessons learned from metabolomics in cystic fibrosis. Mol Cell Pediatr. 2015;2(1):9.
48. Al-Qahtani W, et al. Dried blood spot-based metabolomic profiling in adults with cystic fibrosis. J Proteome Res. 2020;19(6):2346–57.
49. Wetmore DR, et al. Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells. J Biol Chem. 2010;285(40):30516–22.
50. Masood A, et al. Distinctive metabolic profiles between cystic fibrosis mutational subclasses and lung function. Metabolomics. 2021;17(1):4.
51. Muhlebach MS, et al. Metabonomics reveals altered metabolites related to inflammation and energy utilization at recovery of cystic fibrosis lung exacerbation. Metabol Open. 2019;3:100010.
52. Alvarez JA, et al. Plasma metabolomics in adults with cystic fibrosis during a pulmonary exacerbation: a pilot randomized study of high-dose vitamin D3 administration. Metabolism. 2017;70:31–41.
53. Grasemann H, et al. Decreased systemic bioavailability of L-arginine in patients with cystic fibrosis. Respir Res. 2006;7:87.
54. Zardini Buzatto A, et al. Lipidome alterations induced by cystic fibrosis, CFTR mutation, and lung function. J Proteome Res. 2021;20(1):549–64.
55. Guerrera IC, et al. A novel lipidomic strategy reveals plasma phospholipid signatures associated with respiratory disease severity in cystic fibrosis patients. PLoS One. 2009;4(11):e7735.
56. Ollero M, et al. Plasma lipidomics reveals potential prognostic signatures within a cohort of cystic fibrosis patients. J Lipid Res. 2011;52(5):1011–22.
57. Grothe J, et al. Plasma phosphatidylcholine alterations in cystic fibrosis patients: impaired metabolism and correlation with lung function and inflammation. Cell Physiol Biochem. 2015;35(4):1437–53.
58. Neerincx AH, et al. Lumacaftor/ivacaftor changes the lung microbiome and metabolome in cystic fibrosis patients. ERJ Open Res. 2021;7(2):00731.
59. Macedo AN, et al. The sweat metabolome of screen-positive cystic fibrosis infants: revealing mechanisms beyond impaired chloride transport. ACS Cent Sci. 2017;3(8):904–13.
60. Esther CR Jr, et al. Metabolomic evaluation of neutrophilic airway inflammation in cystic fibrosis. Chest. 2015;148(2):507–15.
61. Innis SM, et al. Choline-related supplements improve abnormal plasma methionine-homocysteine metabolites and glutathione status in children with cystic fibrosis. Am J Clin Nutr. 2007;85(3):702–8.
62. Keller BO, Davidson AG, Innis SM. Phthalate metabolites in urine of CF patients are associated with use of enteric-coated pancreatic enzymes. Environ Toxicol Pharmacol. 2009;27(3):424–7.

Metabolomic Role in Personalized Medicine: An Update

Minnie Jacob and Anas M. Abdel Rahman

Abstract Metabolomics is a rapidly evolving omic technology in personalized medicine and has been extensively valued because it involves prescribing the right medicine to the right patient. The breathtaking boost in metabolomic technology has paved the huge potential for its application in personalized medicine. Correlating the metabolic phenotype of individuals into subgroups that respond differently is also becoming a reality through metabolomics. The perception of the metabotype has emerged and played a crucial role in developing a personalized healthcare system. Metabotypes are groups of individuals defined based on their similarities in metabolic profiles. Metabolomics has been utilized in the therapeutic outcomes of drugs, thereby mapping the metabolic profiles of the patients with their responses.

In contrast, the efficacy and toxicity of drugs can be predicted in the pharmacometabolomic method to provide the theoretical basis for individualized medical treatment. This chapter overviews clinical metabotyping, disease biomarker discovery, and pharmacometabolomics toward personalized medicine, improving drug efficacy. These three approaches enhance the understanding of the disease's pathophysiological mechanisms and the metabolic side effects of drugs on human bodies.

Keywords Personalized medicine · Metabolomics · Biomarker discovery · Pharmacometabolomics · Metabotyping

M. Jacob · A. M. Abdel Rahman (✉)
Metabolomics Section, Department of Clinical Genomics, Center for Genome Medicine, King Faisal Specialist Hospital and Research Center (KFSHRC), Riyadh, Saudi Arabia
e-mail: aabdelrahman46@kfshrc.edu.sa

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_10

Abbreviations

3-OHKY	3-hydroxykyenurinine
CAR	Chimeric antigen receptor
CIL-LC/MS	Chemical isotope labeling liquid mass spectrometry
DBS	Dried blood spots
DILI	Idiosyncratic drug-induced liver injury
DIPP	Diabetes prediction and prevention
DOCK8	Dedicator of cytokinesis 8
EPA	Environment Protection Agency
FDA	Food and Drug Administration
IEM	Inborn errors of metabolism
LC-MSMS	Liquid chromatography-tandem mass spectrometry
MSI-CE-MS	Multisegment injection-capillary electrophoresis-mass spectrometry
NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance
NUDT15	Nudix hydrolase 15
PC	Pancreatic cancer
REIMS	Rapid evaporative ionization mass spectrometry
SRM	Selected reaction monitoring
SSRIs	Selective serotonin reuptake inhibitors
ToF	Time of flight
TPMT	Thiopurine methyltransferase

1 Introduction

The swift growth in metabolomics leads to a recharged enthusiasm for cellular metabolism and the role of small molecules in many biological processes. Metabolomics' advanced analysis combined with sophisticated computational techniques is used for molecular characterization and relative expression [1]. Metabolome, a collection of small molecules (<15,000 Da), is the most sensitive biomolecule as regards to physiological, biological, and environmental changes [2]. Integrating metabolomics' relative expression with the pathway analyses helps understand disease pathophysiology and mechanisms. Metabolite expression is associated with the genetic blueprint, sequence variation, RNA transcription, and protein translational processes interacting with environmental exposure and reflecting genotype, especially for multifactorial diseases. Among genomics and proteomics, metabolomics perhaps is closely linked to the phenotype. Hence, it provides

valuable information on healthy and pathophysiological conditions and the response to an external stimulus, such as treatment and environmental exposures [2]. The metabolomics' relative expression helps to know the pathophysiology of the disease at the molecular level, identify the biomarkers of disease prediction and diagnosis, assess disease progression, interpret the influence of the environment and lifestyle, and lastly assess the disease toxicity and related adverse reactions [3–5].

The research community accepted using more than one omic technique to read a disease phenotype's complexity functionally. Healthcare systems worldwide are shifting from the traditional "one-size-fits-all" approach to proactive medical models. Such models are becoming predictive, preventive, personalized, and participatory, known as "Proactive P4 Medicine" [6]. However, integrating multiple omic datasets is promising in P4 personalized medicine, including developing and tailoring effective therapies for individual patients [7, 8].

Considering multiple layers of biochemical reactions between genotypes and phenotypes, different degrees of sensitivity, stability, and physiological and environmental influences are unique for each omic molecule. Accordingly, metabolites are the most functional biomolecules to use as biomarkers, given that the biological connection between the phenotype and genotype has been proven. More than 95% of clinical assays in medical laboratories are based on measuring small molecules. Integrating multi-omics is an approach that allows small molecules through upstream biology to eventually be used for clinical services.

Multiple international projects have invested in their population and genomic profiling after developing next-generation sequencing (NGS), for example, the Genome Canada, Genome England, and Saudi Human Genome Project. However, the widespread use of this advanced technology in routine clinical diagnoses is still encumbered by multiple obstacles, including cost and data analysis. Combining genomic data with a more functional layer of data, such as epigenomics, transcriptomics, proteomics, and metabolomics, will overcome most of the limitations associated with the exclusive use of genomic profiling.

The terms "personalized medicine" and "precision medicine" are used interchangeably, and in addition to a comprehensive clinical profile (phenotype), they refer to the long-term collection of multiple layers of data in various fields, including genomics (genotype), transcriptomics, proteomics, metabolomics, microbiomics, and exosomes. This approach can help healthcare providers customize care management, decisions, and recommendations based on a highly accurate estimation of individual patients' risks and potential outcomes. Personalized or stratified medicine enables clinicians to efficiently prescribe the right medicine to the right patient at the right time, thereby improving healthcare quality and reducing unnecessary diagnostic testing and therapies. Precision medicine can be defined as "treatments targeted to the individual patients based on genetic biomarker, phenotype or physiological characteristics of a given patient from other patients with similar clinical presentations" [9, 10].

Irrespective of the diagnoses' complexities, significant advancements are still required for proper diagnosis, prognosis, therapeutic monitoring, and clinical management. The sensitivity and specificity of the available biomarkers for the particular disease determine the diagnostic efficiency. Typically, liquid chromatography-tandem mass spectrometry (LC-MS/MS) has been used for many decades for several inborn errors of metabolism (IEM) diagnosis, where population-based screening programs usually rely on high-quality biomarkers [11]. However, there are no biomarkers with perfect diagnostic performance, so a combination of multiple biomarkers (including ratios) is used for disease screening and diagnostic purposes. For example, in the newborn screening for phenylketonuria (PKU), both phenylalanine (Phe) levels and the ratio of phenylalanine-to-tyrosine (Phe/Tyr) are quite remarkable, and using them together reduces the false favorable/negative detection rates drastically [12]. Screening for medium-chain acyl Co-A dehydrogenase (MCAD) deficiency relies on the combined profile of elevated acylcarnitines (C6, C8, C10, C10:1), although they have different weights. Regardless of the natural expression in different matrices, metabolites vary quantitatively based on the tumors' types and locations, which determines the suitable metabolic biomarker particularly. Accordingly, personalized medicine can provide a complete and integrated picture of the most related metabolites to the phenotype.

This chapter discusses the updates on metabolomics' role in personalized medicine's main applications: disease metabotyping, biomarker discovery, and pharmacometabolomics.

2 Disease Metabolic Profiling "Metabotyping"

The grouping of individuals based on their metabolic and phenotype characteristics into coherent subgroups is referred to as metabolic phenotypes or metabotypes. The distinctive metabolic profile of the physiological changes during disease progression and medical intervention enables an understanding of the connections of metabotypes to individual factors. Modifying the metabotypes and responding individually under a given intervention is the ultimate goal of developing the personalized diagnostic profile [13, 14]. These physiological-based assays display distinctive metabolic profiles associated with the disease's main cause, clinical manifestation, treatment and management, and health outcomes.

Metabolomics is a rapidly emerging tool that helps analyze these metabotypes using the newly developed analytical strategies that pave the way to annotate several metabolites. For example, in type I diabetes and many diseases, physiological systems play a pivotal role in the progression of a disease. The pathogenesis and etiology require an integrative system biology approach, which requires intensive data acquisition, refinement, and validation of mathematical models that include the contributing factors [15].

The BABYDIAB study in Germany represents metabolic profiles related to age at the onset of islet autoimmunity. Islet autoantibodies appear before developing type I diabetes in children. Still, those developing before 2 years of age are not the same as those developing later in childhood [16]. Metabolic profiling of these two subgroups showed significant differences in the profiles relative to age and islet autoantibody status. A twofold lower concentration of methionine was observed in children who developed autoantibodies by 2 years, compared with those who developed autoantibodies in their late childhood or those found to be autoantibody negative. The critical role of methionine is highlighted in the pathways related to the development of islet autoantibodies in early infancy [17]. This research was in agreement with the Diabetes Prediction and Prevention (DIPP) study [18], where most of the children progressed to islet autoimmunity before 2 years of age. In these two studies, it would be interesting to know if the early detected metabotypes that precede autoimmunity are specific to children who later progress to type I diabetes or were mainly found in children who progress to one or more islet autoantibodies. Another example was a study on overweight obesity, where obesity could be predicted in children by studying the urolithin metabotypes, which led to early biomarkers related to obesity [19].

Metabotype profiling of patients diagnosed with cystic fibrosis (CF) using dried blood spot (DBS) samples reported intermediate byproducts associated with patients' genotypes and phenotypes [20, 21]. The identified metabolic profile included 26 significantly differentially expressed metabolites involving the amino acids, glycolysis, mitochondrial and peroxisomal metabolism, and sorbitol pathways. Specifically, the osmolyte (sorbitol) was remarkably downregulated in CF patients compared to healthy controls indicating perturbation in the sorbitol pathway, which may be responsible for the mucoviscidosis seen in patients with CF [22]. These findings may be supported by the clinical utility of inhaled mannitol and hypertonic saline in patients with CF. A few examples of metabotypes in various genetic disorders are described in Table 1.

Table 1 Metabotypes in genetic disorders

Genetic disorders	Class	Disease/gene	Metabotypes	Perturbed key metabolites	Reference
Single gene disease	Autosomal dominant	Huntington's disease/HTT	Huntington's disease patients and controls	Arginine, citrulline, glycine, d-serine, cholesterol esters, diacylglycerides, triacylglycerides, phosphatidylcholines, phosphatidylethanolamines, sphingomyelin, and NAD+	[23]
		Acute intermittent porphyria/*HMBS*	Patients with asymptomatic acute intermittent porphyria (aAIP) and familial or sporadic porphyria cutanea tarda	Glycolysis and energy-conversion pathways, notably acetate, citrate, and pyruvate	[24]
		Tuberous sclerosis complex 2/ TCS2	Before, during, and after 6 months of treatment with sirolimus and hydroxychloroquine	5′-Methylthioadenosine and arginine and adenosylmethionine decarboxylase 1	[25]
	Autosomal recessive	Hyper-IgE syndrome_DOCK8 deficiency/*DOCK8*	DOCK8 and AD	3-Anthranilic acid, aspartic acid, hypotaurine, guanosine, leucyl-phenylalanine, and glycyl-phenylalanine	[26]
		Cystic fibrosis/*CFTR*	Cystic fibrosis and healthy controls	Sorbitol, glutamic acid, pipecolic acid, fructose-1,6-bisphosphate, hypoxanthine, and lactate	[21, 22]
		Sickle cell disease/*HBB*	Sickle cell patients and controls	2,3-Bisphosphoglycerate, glutathione, carnitines, choline, and increased expression of metabolic precursors involved in polyamine synthesis	[27]
		Phenylketonuria/*PAH*	Phenylketonuria patients and healthy controls	LDL2 Apo-B, LDL2 particle number, LDL2 phospholipids, LDL2 cholesterol, VLDL5-free cholesterol, VLDL5 cholesterol, VLDL5 phospholipids, and VLDL4-free cholesterol, tyrosine, glutamine, creatinine, and citric and glutamic acids	[28]
		Mucopolysaccharidosis-type VI (MPS VI)/*ARSB*	MPS VI patients and healthy control	Dermatan sulfate, amino acids, carnitine, arginine-proline, histidine, and glutathione metabolism	[29]
		Hereditary hemochromatosis/*HFE*	Hereditary hemochromatosis patients and 20 healthy subjects	Serum phosphatidylcholine (PC) and phosphatidylethanolamine	[30]
	Mitochondrial disease	Myoclonic epilepsy with ragged red fibers (MERRF)	MERRF patients and controls	Sodium, phosphate, and magnesemia levels, N-methyl nicotinamide, and hippurate	[31]
		Mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS)	MELAS patients and controls	Glutathione GSH, glutathione disulfide, and cysteine	[32]

Chromosomal abnormalities		Down syndrome	Down syndrome patients and healthy controls	Pyruvate, citrate, succinate, lactate, leucine, phenylalanine, and formate	[33]
		Turner syndrome	Girls with turner syndrome and ge-matched girls with obesity	BCAAs such as methionine, phenylalanine, lysine, tryptophan, histidine, tyrosine, alanine and ornithine, arginine, tyrosine, glutamic acid, citrulline, and alanine	[34]
Multifactorial diseases		Type 1 diabetes	Age and islet autoantibody status	Triglycerides and polyunsaturated fatty acids containing phospholipids	[17]
		Obesity	Children and adolescents	Urolithin metabotypes (UM-A, UM-B, and UM-0), the top-ten contributing SNPs being rs1801253-ADRB1, rs4343-ACE, rs8061518-FTO, rs1130864-CRP, rs659366-UCP2, rs6131-SELP, rs12535708-LEP, rs1501299-ADIPOQ, rs708272-CETP, and rs2241766-ADIPOQ	[19]
		Celiac disease	Celiac disease patients and healthy controls	Triacylglycerols and phosphatidylcholines	[35]
		Rheumatoid arthritis	Rheumatoid arthritis and healthy controls	Phospholipids, diol and its derivatives, arsonoacetate, oleanolic acid acetate, docosahexaenoic acid methyl ester, linolenic acid, and eicosatrienoic acid derivatives	[36]
		Cancer	Pre-cachectic patients and weight-stable cancer patients	α-Fetoprotein, carcinoembryonic antigen, carbohydrate antigen 15–3, carbohydrate antigen 242, carbohydrate antigen 50, cytokeratin 19 fragments, procalcitonin, and interleukin-6. Altered levels of direct bilirubin, total bilirubin, total cholesterol, total bile acid, alkaline phosphatase, γ-glutamyltransferase alanine aminotransferase, aspartate aminotransferase, α-hydroxybutyric dehydrogenase, and lactate dehydrogenase. Biochemical biomarkers include direct bilirubin, total bilirubin, total bile acid, alkaline phosphatase, γ-glutamyltransferase, aspartate aminotransferase, α-hydroxybutyric dehydrogenase, and lactate dehydrogenase	[37]

3 Personalized Medicine and Biomarker Discovery

The eventual goal of personalized medicine is to enable clinicians to prescribe the right medicine with maximum efficacy and minimal toxicity by predicting disease onset among populations, thereby improving healthcare quality and reducing unnecessary diagnostic testing and therapies. The well-tested medicine may lead to the right treatment at the right time but not necessarily the right one for that individual [9]. President Barack Obama introduced a precision medicine initiative in January 2015 called “All of US” (www.whitehouse.gov/precisionmedicine) that enrolled over one million Americans, and they were expected to share the data generated over 10 years from sequencing, electronic medical records, personal reported information, and digital health technologies.

Biomarker discovery uses a combination of technologies to capture the data, which is then translated to select biomarkers that most reliably detect the disease. Firstly, a biospecimen has to be analyzed using a robust method ensuring consistent performance. The emerging biomarkers need further evaluation and validation against a larger sample size to ensure the highest data quality, depending on the clinical model. The validated biomarkers should be sensitive and reliable enough to distinguish patients from healthy subjects.

A metabolomic biomarker-based device called an intelligent knife (Iknife) was introduced recently, which is based on rapid evaporative ionization mass spectrometry (REIMS) technology. Iknife can discriminate cancer from normal tissue in different tumor sites, including the brain, breast, ovaries, and colon [38].

Sensor technology is evolving at a fast rate. It can be well translated into clinical biomarker discovery, wherein targeted biomarkers for a condition can be monitored or detected by wearable sensors by the end-user and then uploaded the data for the researcher to build a model for further prediction and response to the therapeutic intervention. An example is the mPower study for Parkinson’s disease [36], wherein the aspects were investigated through surveys and frequent sensor-based recordings from participants with and without Parkinson’s disease. Similar metabolomic data can be streamlined to look at a panel of biomarkers detectable by sensor chips capable of reporting and interpreting the real-time data. Yet another metabolomic approach published by Blasco et al. (2017) characterized plasma levels in phenylketonuria on a multiplatform, where the commonly dysregulated metabolites were glutamine, arginine, succinate, and alpha aminobutyric acid, in addition to the pathophysiological mechanisms like protein synthesis, energetic metabolism, and oxidative stress, thereby confirming specific metabolic signature related to tyrosine and phenylalanine concentrations (Table 2) [43].

Table 2 Application of metabolomics in different disease conditions

Disease/condition	Metabolomic platform		Biomarkers/major findings	Matrix	Platform	Reference
	Targeted	Untargeted				
Propionic and methylmalonic acidemia		√	Propionyl-carnitine and gamma-butyrobetaine	Plasma	ToF-MS	[39]
Diabetes mellitus	√	√	1,5-Anhydroglucitol	Saliva	LC-MSMS	[40]
Gestational diabetes mellitus			Glycine, serine, and threonine metabolism, steroid hormone biosynthesis, tyrosine metabolism, glycerophospholipid metabolism, and fatty acid metabolism	Serum	LC-MSMS QToF-MS	[41]
Diabetes mellitus			3-Indoxy sulfate, glycerophospholipids, free fatty acids, and bile acids	Blood	LC-MS and NMR	[42]
Phenylketonuria	√		Tyrosine and phenylalanine	Urine	NMR/GC-MS/ amino acid analyzer	[43]
Cystic fibrosis	√		Sorbitol	DBS	LC-MSMS	[22]
	√		3,4-Dihydroxymandelate-3-*O*-sulfate and 5-aminopentanoic acid. Ophthalmic acid, nicotinamide, deoxycarnitine, and a few amino acids	Serum	CIL-LC/MS MSI-CE-MS	[20, 21]
	√		Lactic acid and 4-hydroxy cyclohexyl carboxylic acid	Exhaled breath condensate	UPLC-MS/MS	[44]
HIES	√		Hypotaurine, 3-hydroxy anthranilic acid, and glycyl-phenylalanine	Serum	CIL-LC/MS	[26]

CIL-LC/MS chemical isotope labeling liquid mass spectrometry, *NMR* nuclear magnetic resonance, *ToF* time of flight, *DOCK8* dedicator of cytokinesis 8. *MSI-CE-MS* multisegment injection-capillary electrophoresis-mass spectrometry

Recently, a clinical validation study was reported on 12 disease groups belonging to the study of IEM groups which included disorders in the metabolism of amino acids, fatty acids, ketones, purines, pyrimidines, carbohydrates, porphyrias, neurotransmitters, vitamins, cofactors, and creatine. Remarkably, even mild metabolite patterns are seen in mild multiple acyl-CoA dehydrogenase deficiencies (GA-II), and maple syrup urinary disease (MSUD) could be differentiated easily in this study [45].

Jacob et al. (2019) identified seven positively identified metabolites, distinguishing DOCK8 deficiency from atopic dermatitis (AD) patients. Aspartic acid and 3-hydroxy anthranilic acid (3HAA, a tryptophan degradation pathway intermediate) were upregulated in DOCK8 deficiency, whereas hypotaurine, leucyl-phenylalanine, glycyl-phenylalanine, and guanosine were downregulated. Hypotaurine, 3-hydroxy anthranilic acid, and glycyl-phenylalanine were identified as potential biomarkers specific to DOCK8 deficiency [26] (Table 2).

Currently, personalized medicine faces a few challenges in obtaining approval for routine use from regulatory agencies and healthcare stakeholders before using them routinely in clinics. Moreover, tailored or personalized therapies can be expensive, such as autologous chimeric antigen receptor (CAR) T cell transplant therapies for certain types of cancer [46] and mutation-specific medicines to treat CF patients [47, 48]. Before the biomarkers can be utilized as novel therapy, there is a need for rigorous, standardized protocols and pipelines for biomarker discovery, analysis, validation, and reporting. The pipeline shows that a metabolomic study can be performed either in a targeted or untargeted manner, depending on the metabolomic platform. Untargeted analysis covers and attempts to identify all detected peaks above the noise threshold using scan mode and utilizes both annotated and unannotated peak information for statistical analysis. The targeted analysis seeks only known metabolites detected by selected reaction monitoring (SRM). The identified metabolite biomarkers associated with disease development and progression may pave the way for discovering predictive, diagnostic, and prognostic biomarkers and monitoring therapeutic outcomes after validation and prototype assay development. These results demonstrate that robust metabolomics has the potential as a noninvasive strategy and is a promising screening tool to evaluate the potential of these metabolites in the early diagnosis of patients and provide new insight into pathophysiological mechanisms (Fig. 1).

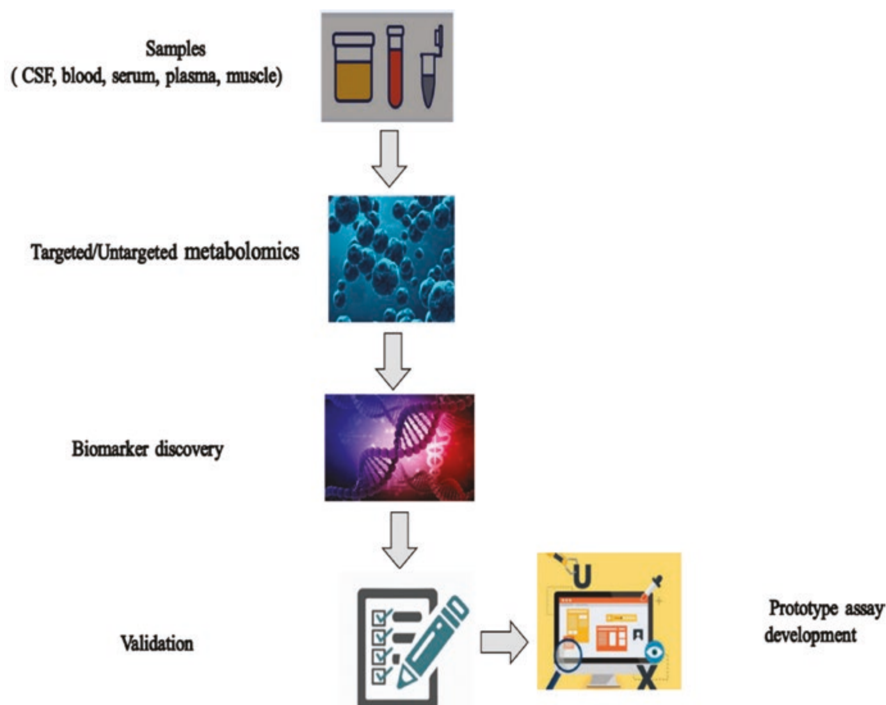

Fig. 1 Biomarker development pipeline workflow

4 Pharmacometabolomics

Pharmacometabolomics was introduced by Clayton et al. in 2006 [49] as an evolving research field aiming to achieve a tailored therapeutic regime and focuses on predicting or evaluating responses to drug treatments based on their metabolic fingerprints. Pharmacometabolomics can be used interchangeably with pharmacometabolomics, which predicts drug effects based on a mathematical model of predose metabolite profiles [49] and monitoring drug metabolic pathway alteration. It is based on metabolic phenotypes or metabotypes, which are the ultimate result of genetic, physiological, chemical, and environmental influences [50]. In

pharmacometabolomics, gut microbiome and environment exposures reveal information about the metabotypes and treatment outcomes, thereby creating metabolic signatures to find potential biomarkers that might inform treatment outcomes. It also provides tools for mapping drug effects and revealing pathways contributing to drug response phenotypes.

4.1 Prediction of Treatment Outcomes

New pathways can be identified for therapeutic discovery by comparing the metabolomes of patients with different clinical parameters like slow vs. fast progressors. For example, in the case of depression, comparing metabolic signatures of fast-acting drugs like ketamine vs. slow-acting drugs like SSRIs (selective serotonin reuptake inhibitors) can define more effective therapies [51]. Pharmacometabolomic studies have identified biomarkers associated with drug metabolism responders, nonresponders, or patients with adverse drug responses. Models can be constructed with baseline samples or treatment samples, as sample collection during or after drug exposure is possible. Such treatment samples may be useful for finding mechanisms associated with drugs that do not cause a positive or negative effect for several months after the initial dosing period. For instance, increased cardiotoxicity risk is seen for many years after treating cancer patients with chemotherapy drugs [52]. Similarly, prolonged latency is observed in a few patients who suffer from idiosyncratic drug-induced liver injury (DILI) [53].

Sixty-eight thousand chemicals out of a total of 81,000 parent chemicals were registered under the Toxic Substances Control Act as per the report from the Environment Protection Agency (EPA). Most of these entities are registered without including the biotic and abiotic transformation products, estimated to be around a million exposure in a lifetime [54]. Although an exogenous biomarker can be measured using specific and sensitive methods, a comprehensive approach covering hundreds to thousands of molecules is quite challenging without using high-throughput techniques such as metabolomics. The endogenous metabolites, nutrients and lipids, phytochemicals, pharmaceuticals, and environmental exposures (xenobiotics) can be covered using exposome-based metabolomics, the only tool to explore the environment and genetic integration to understand the disease risk and development. The xenobiotics and their transformed products are contributed to the host's phenotype through micro- and macroscale interactions with endogenous processes. On the other hand, genetic polymorphisms reflect the xenobiotic clearance and bioactivity across the study population, which is worth to be evaluated and integrating.

4.2 Integrating Drug Metabolism Pathway Alteration

Pharmaceuticals are exogenous molecules that drastically perturb endogenous metabolism, instantly reflecting the phenotype. Pharmacometabolomics plays an important role in our lives because the response to drugs in humans and animals is different in all individuals depending on their predose phenotypes, which are influenced by their genomes, environment, and microbiome. Pharmacometabolomic experiments mainly establish a correlation between variations in an individual's predose biofluid metabolite profile and their post-drug dose responses, using different matrices like blood, urine, or plasma.

In patients, a drug interaction is expected to return the metabolic profile to a healthy state in case of a positive therapeutic response. However, ineffective treatment in patients might shift the metabolomic profile to a toxic response, such as the accumulation of thiopurine drugs. These toxic metabolites result from variant detection on thiopurine methyltransferase (TPMT) and nudix hydrolase 15 (NUDT15) genes. Thus, metabolomics significantly can be involved in all the drug discovery phases, from identifying the therapeutic target to therapeutic monitoring. The drug exposome, including the factors of drug administration route, frequency, formulation, and ADME (absorption, distribution, metabolism, and excretion), is essential for data mining in any environmental exposome studies [55].

Numerous patients experience little or no efficacy, or sometimes even toxicity, due to their prescribed drugs. As per the study of Lazarou et al., over two million people in American hospitals face serious adverse drug reactions requiring hospitalization or leading to permanent disability [56]. Gebregiworgis et al.'s studies focused on pancreatic cancer (PC) cells that respond or develop resistance to gemcitabine treatment. This study compared the wild type and resistant type of PC cell lines before and after treatment with gemcitabine, which revealed unique metabolic changes differentiating the response or the acquired resistance to the drug. The resistant type was combined with stable-isotope labeling experiments using ^{13}C6-glucose, primarily derived for nucleotide synthesis to compensate for gemcitabine activity. In the wild type, glucose is directed toward glycolysis after treatment [57].

The first pharmacometabolomic study on patients demonstrated that the predose serum levels of lactate, alanine, and percentage body fat could predict weight gain in a group of 21 breast cancer patients undergoing 5-fluorouracil, cyclophosphamide, or epirubicin-based chemotherapy [58]. Weight gain is the risk factor for reoccurrence. Predictive tests could be very useful in these patients. The levels of 3-hydroxykyenurinine (3-OHKY) could predict the severity of clinical symptoms during the early stages in schizophrenic patients [59]. Low predose levels of 3-OHKY were predictive of greater responsibility in the fourth week with one or more antipsychotic drugs.

In a study by Clayton et al. (2009), a group of volunteers was administered acetaminophen and showed increased levels of urinary p-cresol sulphonate acetaminophen (a metabolite of acetaminophen) [60]. Before taking the drug, individuals with high urine p-cresol sulphonate showed a low concentration of sulfonated acetaminophen in the subsequent urine sample, suggesting that p-cresol sulphonate acts as a competitive inhibitor of acetaminophen. This confirms that each individual is unique, with certain metabolic pathways involved in the catabolism of the drug, hence focusing on the importance of decoding individual metabotypes before the first dose.

Pharmacometabolomics can play a pivotal role in improving personalized medicine or stratified healthcare for clinicians to help choose the optimal treatment for the subsets of patients. Interestingly, the Critical Path Opportunities Report published by the FDA (Food and Drug Administration) considered pharmacometabolomics a fundamental part of the early phases of drug development [61].

Rapidly improving analytical technologies have taken metabolisms to a great level, but the clinical use of pharmacometabolomics is quite slow. To better understand or predict individual patient responses to the drug, the combination of pharmacogenomics and pharmacometabolomics provides genetic, drug metabolite, systemic metabolite, and environmental information [51, 62, 63]. Pharmacometabolomics has been used to discover provisional and/or safety biomarkers, which can help in patient selection during clinical trials. In this decade, the need of the hour is to understand how environmental factors influence drug responses. Pharmacometabolomics is important in providing information on xenobiotic, endogenous, and gut-microbe metabolites present in a patient before, during, and after drug exposure.

5 Conclusion

In summary, identifying the metabotypes in a particular state of health will likely benefit the patient's health. Metabotyping helps identify metabolically similar subpopulations or patient subgroups responding differently to nutritional or drug interventions, which can help, for example, tailor a dietary recommendation for a particular subgroup of the obese population. Since the metabolome is complex, it requires complementary analytical methodologies to explain the underlying biological processes, as global lifestyle and genetic components play a decisive role. Metabolomics is expected to revolutionize the pattern of biochemical information used to assess health and disease states, as it promises a better prognosis for the disease. Metabolomics-based personalized medicine helps detect altered metabolic pathways correlated with a given health condition and its progression over time.

Pharmacometabolomics has the potential to accelerate drug development by identifying clinical development processes at an early stage, and it is the integrated outcome of genomics, proteomics, and environmental influences on the organism which help in providing vital information on drug responses not easily expressed by

other omics. Early planning and identification of potential metabolic signatures, ethics consideration, education, and sample processing are vital inclusions in pharmacometabolomic principles in clinical development, which can shorten clinical development timelines and lower the overall developmental costs. The integrated analysis of data obtained from different platforms (metabolomics, genomics, and proteomics) can help characterize the treatment effects and benefit the personalized medicine field, thereby improving treatment selection for the patients. Further in vivo and in vitro studies are required to understand better the biological mechanism underlying metabolic changes that ultimately lead to the discovery of sensitive and specific biomarkers. Specimen stability and pre-analytical issues are crucial in accomplishing a successful biomarker discovery.

In the near future, significant improvements in imaging technologies and prediction algorithms will provide immense knowledge and differentiate between healthy and disease conditions. Medical databases must be developed and populated with large-scale metabotyping data for various diseases. Pharmacometabolomics plays a crucial role in shortening the clinical development timelines, bringing down the overall cost, and benefiting the healthcare system through overall development and translational effectiveness. The interplay of pharmacogenomics, metabolomics, and pharmacometabolomics is quite fascinating, as the identification of the genetic components of the metabolome seems to be the key to the discovery of many biomarkers. "Pharmacometabolomics-informed pharmacogenomics" is expected to give new insight and contribute toward finding the right metabotypes leading to personalized medicine.

References

1. German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. Metabolomics. 2005;1(1):3–9.
2. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the human metabolome database. Nucleic Acids Res. 2007;35(Database issue):D521–6.
3. Kohler I, Verhoeven A, Derks RJ, Giera M. Analytical pitfalls and challenges in clinical metabolomics. Bioanalysis. 2016;8(14):1509–32.
4. Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. Mass Spectrom Rev. 2017;38:221.
5. Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z, Lindon JC. Metabolic phenotyping in clinical and surgical environments. Nature. 2012;491(7424):384–92.
6. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. New Biotechnol. 2012;29(6):613–24.
7. Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. Mass Spectrom Rev. 2019;38(3):221–38.
8. Wishart DS. Metabolomics: the principles and potential applications to transplantation. Am J Transpl. 2005;5(12):2814–20.
9. Jameson JL, Longo DL. Precision medicine--personalized, problematic, and promising. N Engl J Med. 2015;372(23):2229–34.

10. Abdel Rahman AM, Lopata AL, Randell EW, Helleur RJ. Absolute quantification method and validation of airborne snow crab allergen tropomyosin using tandem mass spectrometry. Anal Chim Acta. 2010;681(1–2):49–55.
11. Villoria JG, Pajares S, López RM, Marin JL, Ribes A. Neonatal screening for inherited metabolic diseases in 2016. Semin Pediatr Neurol. 2016;23(4):257–72.
12. Bhattacharya K, Wotton T, Wiley V. The evolution of blood-spot newborn screening. Transl Pediatr. 2014;3(2):63–70.
13. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181–9.
14. Holmes E, Loo RL, Stamler J, Bictash M, Yap IK, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008;453(7193):396–400.
15. Marinković T, Sysi-Aho M, Orešič M. Integrated model of metabolism and autoimmune response in β-cell death and progression to type 1 diabetes. PLoS One. 2012;7(12):e51909.
16. Ziegler AG, Nepom GT. Prediction and pathogenesis in type 1 diabetes. Immunity. 2010;32(4):468–78.
17. Pflueger M, Seppänen-Laakso T, Suortti T, Hyötyläinen T, Achenbach P, Bonifacio E, et al. Age- and islet autoimmunity-associated differences in amino acid and lipid metabolites in children at risk for type 1 diabetes. Diabetes. 2011;60(11):2740–7.
18. Oresic M, Simell S, Sysi-Aho M, Näntö-Salonen K, Seppänen-Laakso T, Parikka V, et al. Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. J Exp Med. 2008;205(13):2975–84.
19. Cortés-Martín A, Colmenarejo G, Selma MV, Espín JC. Genetic polymorphisms, mediterranean diet and microbiota-associated urolithin metabotypes can predict obesity in childhood-adolescence. Sci Rep. 2020;10(1):7850.
20. Masood A, Jacob M, Gu X, Abdel Jabar M, Benabdelkamel H, Nizami I, et al. Distinctive metabolic profiles between Cystic Fibrosis mutational subclasses and lung function. Metabolomics. 2021;17(1):4.
21. DiBattista A, McIntosh N, Lamoureux M, Al-Dirbashi OY, Chakraborty P, Britz-McKibbin P. Metabolic signatures of cystic fibrosis identified in dried blood spots for newborn screening without carrier identification. J Proteome Res. 2019;18(3):841–54.
22. Al-Qahtani W, Abdel Jabar M, Masood A, Jacob M, Nizami I, Dasouki M, et al. Dried blood spot-based metabolomic profiling in adults with cystic fibrosis. J Proteome Res. 2020;19(6):2346–57.
23. McGarry A, Gaughan J, Hackmyer C, Lovett J, Khadeer M, Shaikh H, et al. Cross-sectional analysis of plasma and CSF metabolomic markers in Huntington's disease for participants of varying functional disability: a pilot study. Sci Rep. 2020;10(1):20490.
24. Luck M, Schmitt C, Talbi N, Gouya L, Caradeuc C, Puy H, et al. Correction to: urinary metabolic profiling of asymptomatic acute intermittent porphyria using a rule-mining-based algorithm. Metabolomics. 2018;14(3):21.
25. Tang Y, El-Chemaly S, Taveira-Dasilva A, Goldberg HJ, Bagwe S, Rosas IO, et al. Alterations in polyamine metabolism in patients with lymphangioleiomyomatosis and tuberous sclerosis complex 2-deficient cells. Chest. 2019;156(6):1137–48.
26. Jacob M, Gu X, Luo X, Al-Mousa H, Arnaout R, Al-Saud B, et al. Metabolomics distinguishes DOCK8 deficiency from atopic dermatitis: towards a biomarker discovery. Meta. 2019;9(11):274.
27. Adebiyi MG, Manalo JM, Xia Y. Metabolomic and molecular insights into sickle cell disease and innovative therapies. Blood Adv. 2019;3(8):1347–55.
28. Cannet C, Pilotto A, Rocha JC, Schäfer H, Spraul M, Berg D, et al. Lower plasma cholesterol, LDL-cholesterol and LDL-lipoprotein subclasses in adult phenylketonuria (PKU) patients compared to healthy controls: results of NMR metabolomics investigation. Orphanet J Rare Dis. 2020;15(1):61.
29. Tebani A, Abily-Donval L, Schmitz-Afonso I, Piraud M, Ausseil J, Zerimech F, et al. Analysis of mucopolysaccharidosis type VI through integrative functional metabolomics. Int J Mol Sci. 2019;20(2):446.

30. Seeßle J, Gan-Schreier H, Kirchner M, Stremmel W, Chamulitrat W, Merle U. Plasma Lipidome, PNPLA3 polymorphism and hepatic steatosis in hereditary hemochromatosis. BMC Gastroenterol. 2020;20(1):230.
31. Hall AM, Vilasi A, Garcia-Perez I, Lapsley M, Alston CL, Pitceathly RD, et al. The urinary proteome and metabonome differ from normal in adults with mitochondrial disease. Kidney Int. 2015;87(3):610–22.
32. Homma K, Toda E, Osada H, Nagai N, Era T, Tsubota K, et al. Taurine rescues mitochondria-related metabolic impairments in the patient-derived induced pluripotent stem cells and epithelial-mesenchymal transition in the retinal pigment epithelium. Redox Biol. 2021;41:101921.
33. Antonaros F, Ghini V, Pulina F, Ramacieri G, Cicchini E, Mannini E, et al. Plasma metabolome and cognitive skills in Down syndrome. Sci Rep. 2020;10(1):10491.
34. Bugajska J, Berska J, Wójcik M, Starzyk JB, Sztefko K. Metabolic fingerprint of turner syndrome. J Clin Med. 2020;9(3):664.
35. Sen P, Carlsson C, Virtanen SM, Simell S, Hyöty H, Ilonen J, et al. Persistent alterations in plasma lipid profiles before introduction of gluten in the diet associated with progression to celiac disease. Clin Transl Gastroenterol. 2019;10(5):1–10.
36. Carlson AK, Rawle RA, Wallace CW, Adams E, Greenwood MC, Bothner B, et al. Global metabolomic profiling of human synovial fluid for rheumatoid arthritis biomarkers. Clin Exp Rheumatol. 2019;37(3):393–9.
37. Yang QJ, Zhao JR, Hao J, Li B, Huo Y, Han YL, et al. Serum and urine metabolomics study reveals a distinct diagnostic model for cancer cachexia. J Cachexia Sarcopenia Muscle. 2018;9(1):71–85.
38. Tzafetas M, Mitra A, Paraskevaidi M, Bodai Z, Kalliala I, Bowden S, et al. The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease. Proc Natl Acad Sci U S A. 2020;117(13):7338–46.
39. Wikoff WR, Gangoiti JA, Barshop BA, Siuzdak G. Metabolomics identifies perturbations in human disorders of propionate metabolism. Clin Chem. 2007;53(12):2169–76.
40. Halama A, Kulinski M, Kader SA, Satheesh NJ, Abou-Samra AB, Suhre K, et al. Measurement of 1,5-anhydroglucitol in blood and saliva: from non-targeted metabolomics to biochemical assay. J Transl Med. 2016;14(1):140.
41. Liu T, Li J, Xu F, Wang M, Ding S, Xu H, et al. Comprehensive analysis of serum metabolites in gestational diabetes mellitus by UPLC/Q-TOF-MS. Anal Bioanal Chem. 2016;408(4):1125–35.
42. Suhre K, Meisinger C, Döring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS One. 2010;5(11):e13953.
43. Blasco H, Veyrat-Durebex C, Bertrand M, Patin F, Labarthe F, Henique H, et al. A multiplatform metabolomics approach to characterize plasma levels of phenylalanine and tyrosine in phenylketonuria. JIMD reports. 2017;32:69–79.
44. Zang X, Monge ME, Gaul DA, McCarty NA, Stecenko A, Fernández FM. Early detection of cystic fibrosis acute pulmonary exacerbations by exhaled breath condensate metabolomics. J Proteome Res. 2020;19(1):144–52.
45. Steinbusch LKM, Wang P, Waterval H, Stassen F, Coene KLM, Engelke UFH, et al. Targeted urine metabolomics with a graphical reporting tool for rapid diagnosis of inborn errors of metabolism. J Inherit Metab Dis. 2021;44(5):1113–23.
46. Fry TJ, Shah NN, Orentas RJ, Stetler-Stevenson M, Yuan CM, Ramakrishna S, et al. CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. Nat Med. 2018;24(1):20–8.
47. Davis PB, Yasothan U, Kirkpatrick P. Ivacaftor. Nat Rev Drug Discov. 2012;11(5):349–50.
48. Gulland A. Cystic fibrosis drug is not cost effective, says NICE. BMJ. 2016;353:i3409.
49. Clayton TA, Lindon JC, Cloarec O, Antti H, Charuel C, Hanton G, et al. Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature. 2006;440(7087):1073–7.
50. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008;134(5):714–7.

51. Kaddurah-Daouk R, Weinshilboum R. Metabolomic signatures for drug response phenotypes: pharmacometabolomics enables precision medicine. Clin Pharmacol Ther. 2015;98(1):71–5.
52. Jain D, Ahmad T, Cairo M, Aronow W. Cardiotoxicity of cancer chemotherapy: identification, prevention and treatment. Ann Transl Med. 2017;5(17):348.
53. Mosedale M, Watkins PB. Drug-induced liver injury: advances in mechanistic understanding that will inform risk management. Clin Pharmacol Ther. 2017;101(4):469–80.
54. Niedzwiecki MM, Walker DI, Vermeulen R, Chadeau-Hyam M, Jones DP, Miller GW. The exposome: molecules to populations. Annu Rev Pharmacol Toxicol. 2019;59:107–27.
55. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods. 2000;44(1):235–49.
56. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA. 1998;279(15):1200–5.
57. Gebregiworgis T, Bhinderwala F, Purohit V, Chaika NV, Singh PK, Powers R. Insights into gemcitabine resistance and the potential for therapeutic monitoring. Metabolomics. 2018;14(12):156.
58. Keun HC, Sidhu J, Pchejetski D, Lewis JS, Marconell H, Patterson M, et al. Serum molecular signatures of weight change during early breast cancer chemotherapy. Clin Cancer Res. 2009;15(21):6716–23.
59. Condray R, Dougherty GG Jr, Keshavan MS, Reddy RD, Haas GL, Montrose DM, et al. 3-Hydroxykynurenine and clinical symptoms in first-episode neuroleptic-naive patients with schizophrenia. Int J Neuropsychopharmacol. 2011;14(6):756–67.
60. Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc Natl Acad Sci U S A. 2009;106(34):14728–33.
61. Woodcock J, Woosley R. The FDA critical path initiative and its influence on new drug development. Annu Rev Med. 2008;59:1–12.
62. Abo R, Hebbring S, Ji Y, Zhu H, Zeng ZB, Batzler A, et al. Merging pharmacometabolomics with pharmacogenomics using '1000 Genomes' single-nucleotide polymorphism imputation: selective serotonin reuptake inhibitor response pharmacogenomics. Pharmacogenet Genomics. 2012;22(4):247–53.
63. Oh J, Yi S, Gu N, Shin D, Yu KS, Yoon SH, et al. Utility of integrated analysis of pharmacogenomics and pharmacometabolomics in early phase clinical trial: a case study of a new molecular entity. Genomics Informa. 2018;16(3):52–8.

Lipidomic Profiling in Clinical Practice Using LC-MS

Núria Amigó Grau and Pablo Ortiz Betes

Abstract The advances in lipidomic profiling techniques in the last two decades have significantly improved our understanding of the biological processes involved in health and disease. Currently, many lipidomic profiling applications are moving from research laboratories to clinical application, as lipidomics has the potential to be implemented into the clinical routine, complementing traditional clinical factors to improve the diagnosis, to stratify risk post-diagnosis, and to support treatment monitoring of both pharmaceutical and lifestyle interventions.

In this chapter, we describe how clinical environment is opening the door to the implementation of lipidomic-based applications for predicting and monitoring a wide range of metabolic diseases. We present a use case based on LC-MS lipidomics already operating in hospital clinical workflows for Non-Alcoholic Fatty Liver Disease assessment, exemplifying the potential of precision medicine approaches for personalized diagnostics.

Keywords Lipidomics · LC-MS lipidomics · NMR metabolomics · Advanced lipoprotein profiling · Nonalcoholic fatty liver disease · Cardiovascular risk assessment · Cardiovascular diseases · Precision medicine

---

N. Amigó Grau (✉)
Biosfer Teslab, Av. Universitat 1, Reus, Spain

Department of Basic Medical Sciences, Universitat Rovira i Virgili (URV), Institut d'Investigació Sanitària Pere Virgili (IISPV), Av. Universitat 1, Reus, Spain

Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III (ISCIII), Madrid, Spain
e-mail: namigo@biosferteslab.com

P. Ortiz Betes
Biosfer Teslab, Av. Universitat 1, Reus, Spain

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_11

Abbreviations

ApoB	Apoprotein B
AUC	Area under the curve
AUROC	Area under the receiver operating characteristic curve
BMI	Body mass index
CIBERDEM	Centro de Investigación Biomédicas en Red de Diabetes y Enfermedades Metabólicas
CLIA	Clinical Laboratory Improvement Amendments
CMD	Cardiometabolic diseases
CVD	Cardiovascular disease
DAG	Diacylglycerol
DG	Diglycerides
EAS	European Atherosclerosis Society
ESC	European Society of Cardiology
FFAs	Free fatty acids
HCC	Hepatocarcinoma
HDL	High-density lipoprotein
IDL	Intermediate-density lipoproteins
IISPV	Institut d'Investigació Sanitària Pere Virgili
IMT	Intima–media thickness
IVD	In vitro diagnostics
LC-MS	Liquid chromatography-mass spectrometry
LDL	Low-density lipoprotein
LITMUS	Liver Investigation Testing Marker Utility in Steatohepatitis
LMWM	Low-molecular-weight metabolites
NAFLD	Nonalcoholic fatty liver disease
NASH	Nonalcoholic steatohepatitis
NMR	Nuclear magnetic resonance
NPV	Negative predictive value
OWL	One-way liver
PC	Phosphatidylcholine
PPV	Positive predictive value
PUFAs	Polyunsaturated fatty acids
RCT	Reverse cholesterol transport
ROC	Receiver operating characteristics
VLDL	Very-low-density lipoprotein
TAG	Triacylglycerol
TG	Triglycerides
US FDA	US Food and Drug Administration
WHO	World Health Organization

1 Introduction

Metabolomics is an emerging approach in the systems biology field especially in clinical trials and is considered the closest biology system to the phenotypes.

Metabolites are the products of complex molecular pathways (including genomics, transcriptomics, and proteomics). Therefore, metabolomic analysis is a promising strategy for identifying disease-associated biomarkers: metabolites, small molecules, and end products produced as a result of interactions between genes and environmental factors.

There are two possible approaches for discovering new biomarkers. The first is a classic approach, whereby a biomarker is sought that explains the pathophysiology of metabolic diseases, such as glucose quantification to determine whether a person has diabetes or the quantification of LDL-C for assessing someone's risk of cardiovascular diseases (CVD). The second approach involves searching for metabolic patterns that are characteristic of the physiology of a metabolic disease [1]. In this second approach, high-performance techniques such as liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) are used to carry out metabolomic studies without prior knowledge of the metabolites' involvement or role in the disease mechanism or physiology [2], since the metabolites are the final product of the interactions between gene and protein expression and environmental exposure and consequently of the biochemical activity that characterize part of the patient's metabolism.

The metabolomic approach compares the metabolomic profiles of fluids or tissues of a patient with those of a healthy subject to identify which metabolites are expressed differently [3].

Metabolomics generally includes all kinds of molecules constituting a biological matrix, and lipidomics is the study of specific lipids. In particular, lipidomics involves the study of the structure, function, and metabolism of lipids in living organisms [4]. Lipidomics is a broad field that encompasses analysis of the complete set of lipids present in a biological sample and the pathways and enzymes involved in their synthesis, degradation, and modification.

Various analytical techniques are used in lipidomics, such as mass spectrometry, NMR spectroscopy, and chromatography, to identify and quantify the different lipid species present in a sample. These techniques allow researchers to obtain a detailed picture of the lipid content of a sample and to study the changes in lipid levels in response to different conditions or treatments [5].

Lipidomics has a wide range of applications in various fields, including medicine, nutrition, and environmental science. It is used to study the role of lipids in health and disease, and the mechanisms through which lipids affect cellular processes and signaling pathways, and to identify potential therapeutic targets for treating lipid-related disorders [6].

Compared to other omic techniques, such as proteome or genome analysis, the lipidome is much more complex because of the enormous structural diversity comprising both linear and coupled macromolecules [5]. This diversity increases the number of lipid molecules and the complexity of their carriers, lipoproteins. For this reason, analyzing the lipidome and lipoproteins is very challenging from a technical standpoint. In recent years, LC-MS and NMR have proved to be important analytical tools for metabolomic studies in biological fluids. Unlike other techniques, LC-MS and NMR can quantify many molecular entities simultaneously and effectively [7, 8]. These techniques are widely used in both routine clinical laboratories and large epidemiological studies.

This chapter discusses lipidomic platforms for clinical applications using LC-MS. It explores their benefits, challenges, difficulties, and future opportunities. Application of these technologies in the clinical environment has opened the door to their future implementation for predicting and monitoring numerous diseases. The technologies analyzed in the present chapter are already being applied in clinics, exemplifying the potential of precision medicine approaches for personalized diagnostics.

2 Lipids and Lipoproteins: A Biochemical Approach

2.1 General Concepts

Lipids are a diverse group of organic molecules that are important for many biological functions and essential to life. They include fats, oils, waxes, and other water-insoluble compounds [9].

Lipids are synthesized by living cells through a process called lipid biosynthesis. This process involves condensing fatty acids with glycerol or other alcohols to form triglycerides, the primary component of fats. Alternatively, lipids can also be synthesized by modifying existing lipids, such as by adding a phosphate group to form a phospholipid or adding a carbohydrate group to form a glycolipid.

The structure of lipids is characterized by their hydrophobic nature, which arises from the presence of long, nonpolar hydrocarbon chains. This nonpolarity allows lipids to interact with other molecules and form aggregates, such as micelles or lipid bilayers, which can serve as structural components of cell membranes [10].

In general, the structure of lipids in plasma is that of a submicroscopic oil droplet containing an outer layer of phospholipids, unesterified cholesterol, and proteins, with a core of neutral lipids, predominately cholesterol esters and triglycerides, named lipoproteins [11]. The lipoprotein classes differ in their lipid composition (being either cholesterol [C]-rich or triglyceride [TG]-rich), their protein composition (apoprotein), and their density (which is proportional to their protein/lipid ratio). Based on protein composition, there are two types of lipoproteins, those that

contain ApoB and those that contain ApoA. Based on lipid composition, there are also two types of lipoproteins, those that are C-rich and those that are TG-rich.

Lipids have many functions in living organisms, from energy reserve to insulation, as well as being a structural component of every cell and tissue. They also play important roles in cellular signaling and regulation. This structural diversity is mirrored by the enormous variation in their physiological function. The abundance of individual lipid molecular species in plasma may indicate the variety of specific human diseases.

Around 4500 metabolites have been detected/identified in the human serum metabolome to date: half are phospholipids and over a thousand are glycerolipids (triglycerides TG, diglycerides DG, and monoacylglycerols) [12]. In other words, lipids make up approximately three-quarters of the known human serum metabolome.

To date, thousands of distinct molecular species have been quantified, covering the six main mammalian lipid categories: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, and prenol lipids. The number of metabolites keeps increasing, as each of these types of lipid can exist in multiple forms [13]. In addition, lipids can undergo chemical modifications, such as the addition of a phosphate group or a carbohydrate group, which can result in the creation of new lipid species. This structural diversity is particularly relevant within the sphingolipid and glycerophospholipid categories, principally determined by variations in fatty acid content and head groups.

2.2 Lipids Mirror Present and Future Metabolic Health: Two Sides of the Same Problem

Abnormal plasma lipids and lipoproteins are important risk factors for metabolic and cardiovascular diseases. Metabolic diseases are increasing exponentially worldwide, and their complications, in cardiovascular and liver diseases, are the leading cause of mortality worldwide [14]. They have a common characteristic: their etiology is linked to excess fat and associated inflammation. Nonalcoholic fatty liver disease (NAFLD), associated with obesity and excess fat in the liver, is increasingly common around the world, especially in Western countries, and affects approximately a quarter of the population in countries such as the USA [15]. Symmetrically, arterial health is compromised by excess fat, which accelerates the progression of arteriosclerosis, the underlying cause of myocardial infraction or stroke, the leading cause of death in all developed countries [14]. Both liver disease and atherosclerosis can develop and progress more rapidly when accompanied by several well-known risk factors and comorbidities, such as obesity and dyslipidemia.

Lipids circulate in the blood in the form of lipoprotein particles, including chylomicrons, very-low-density lipoproteins (VLDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). Some parameters, such as cholesterol and

triglyceride levels, are assessed routinely, and physicians regularly prescribe lipid-lowering drugs to patients with dyslipidemia [16].

The duration of exposure to high lipid levels is also a crucial risk factor in cardiometabolic health, especially for metabolic disorders that start early in childhood. Evidence indicates that atherosclerosis begins in childhood with the accumulation of lipids in the intima of arteries to form fatty streaks [17]. Nearly all children have at least some degree of aortic fatty streaks by 3 years of age [18]. These fatty streaks increase after 8 years of age, with atherosclerotic plaques being found in the coronary arteries during adolescence [19]. Clusters of risk factors in childhood predict the presence of risk factors in adults [20].

Carotid ultrasonography screening for subclinical arteriosclerosis has been validated in observational, longitudinal, and randomized clinical studies. Those results were significantly correlated with intravascular coronary ultrasonography, coronary angiography, and pathologic findings of arterial lesions in healthy and CVD patients, and longitudinal studies have demonstrated that increased lipid alterations and intima–media thickness in young adults are associated with cardiovascular risk factors in childhood [21, 22].

The contribution of LDL-associated cholesterol to the development of CVD has been well described. While LDL particle accumulation increases the atherosclerotic process, HDL helps remove excess cholesterol by reverse cholesterol transport. Low levels of HDL are associated with high cardiovascular risk [23].

Conversely, in circulating chylomicrons and VLDL, triglycerides undergo hydrolysis to generate a pool of free fatty acids (FFAs), which are used as an energy source in tissues. Excess FFAs are stored in adipocytes, favoring the expansion and dysfunction of adipose tissue, increasing insulin resistance and diabetes. This process is associated with abnormally high plasma levels of saturated FFAs, allowing increasing their uptake into hepatocytes to exceed metabolic requirements, which leads to hepatic steatosis and inflammation [24].

3 Clinical Relevance of the Lipidome

Lipids participate in many biological processes, so it is not surprising that defects in lipoprotein homeostasis are the direct cause of many diseases and therefore be considered markers of the disease. Detailed knowledge of the composition and concentration of plasma lipidome is expected to expand our diagnostic capabilities and improve the pharmacological evaluation and efficacy of prescribed therapy. A deep analysis of the lipidome might reflex altered synthesis of specific lipid species or identify abnormal underlying pathological lipoprotein patterns. Knowledge about the role of lipids and lipoproteins in disease mechanisms is constantly growing as more information on lipids and lipoprotein physiology becomes clear.

3.1 Lipids and NAFLD: Introduction Through LC-MS

Nonalcoholic fatty liver disease (NAFLD) includes a wide spectrum of disorders ranging from benign lipid accumulation in the liver (steatosis) to a more complicated clinical stage when fat induces hepatic inflammation and hepatocyte necrosis producing a new stage called nonalcoholic steatohepatitis (NASH) also named as steatohepatitis and, eventually, when fibrosis progression is added to fat and inflammation ("NASH at risk") [25]. The final stage of disease progression is cirrhosis and/or hepatocarcinoma (HCC) [26]. Interestingly, another potential evolution from NASH directly to HCC without any significant fibrosis contribution has also been described [27].

The exact cause of NAFLD is not known. Many factors contribute to this condition, such as excessive food intake, obesity, type 2 diabetes, and dyslipidemia, but not all patients develop NAFLD/NASH, and not all patients with NAFLD/nonalcoholic steatohepatitis (NASH) suffer from these conditions [28].

The pathophysiology of NAFLD is quite complex, and the progression from hepatic steatosis to the different stages of this condition is not completely understood. However, lipid metabolic changes, including the production of lipotoxic species in the liver, could be responsible for disease progression in NAFLD [27].

LC-MS allows the tracking of more than 400 different lipid species in the liver and serum. Changes in many of these metabolites were followed in several transverse cohort studies [29–35], giving robust data about the lipidomic signature of the different clinical stages of NAFLD, from liver steatosis to steatohepatitis and advanced fibrosis. In the past 20 years, research studies in serum lipidomics have successfully identified biomarkers to differentiate the stages of NAFLD [29].

The first attempt to diagnose fatty liver through lipidomics in humans was reported by Puri et al. in 2007. They analyzed the hepatic lipid profiles of subjects with normal liver histology, steatosis, and NASH [30]. The study showed that there was no difference in the FFAs between the three groups. The triacylglycerol (TAG) and diacylglycerol (DAG) levels were increased, and the phosphatidylcholine (PC) content was decreased in NAFLD, which suggests that PC hydrolysis may contribute to DAG and TAG accumulation in fatty liver and thereby increased lipogenesis in NAFLD.

These findings were confirmed by Kotronen and García-Cañaveras through semi-quantification of the full range of lipids using LC-MS [31, 32]. Total lysophospholipids, DAG, and TAG were elevated in NAFLD. The stearic-to-oleic acid ratio was decreased in NAFLD, indicative of increased TAG biosynthesis, and NAFLD is also characterized by an increase in DAG and a reduction in polyunsaturated fatty acids (PUFAs).

An additional pilot study (42 biopsy samples) and a pivotal validation study (467 biopsy samples) were published by Barr et al. in 2010 and 2012, respectively, whereby LC-MS was used to identify lipidomic signatures associated with NAFLD

progression [33]. A variety of lipid biomarkers were identified that correlated with NAFLD progression. In the second study, which included 467 biopsy samples with normal liver ($n = 90$) or diagnosed with NAFLD (steatosis, $n = 246$; NASH, $n = 131$), approximately 540 circulating metabolites were analyzed, including amino acids, FFA, DAG, TG, PC, PE, PI, ceramides, SM, cholesteryl esters, and bile acids. Analysis of the lipidomic data allowed the definition of a robust BMI-dependent lipidomic signature that reliably and accurately differentiated liver steatosis from NASH. The area under the curve (AUC) was 0.84 for lean/pre-obese, 0.85 for obese, and 0.87 for morbidly obese patients. More recently, using this same cohort of patients, a set of 25 BMI-dependent lipid profiles was established that could differentiate between steatosis and NASH with AUC values of 0.99, 0.90, and 0.91 for lean/pre-obese, obese, and morbidly obese patients, respectively [34].

Distinguishing between simple steatosis and NASH is relevant for differentiating between a generally benign condition and one with increased morbidity and mortality [35]. Therefore, there is an unmet need for noninvasive biomarkers that are robust, reliable, and cost-effective for patients with NAFLD.

Mayo et al. reanalyzed the lipidomic data from Barr's pivotal trial and a new cohort of 192 biopsy samples from NAFLD patients, culminating in the validation of a BMI-dependent algorithm with 20 TGs. The area under the receiver operating characteristic curve (AUROC) versus biopsy for the discrimination between NASH and NAFLD was 0.95 with sensitivity, specificity, positive predictive value, and negative predictive value of 0.83, 0.94, 0.89, and 0.90, respectively [34].

Bril et al. evaluated the lipidomic signature in 220 patients with type 2 diabetes mellitus to differentiate between steatosis and NASH. They found an AUROC of 0.79 (95% CI 0.68–0.90) in patients with adequate glycemic control. However, this differentiation was quite poor in patients with high insulin resistance or poor glycemic control [36].

Validation of the lipidomic signatures in such important cohorts indicates that lipidomic markers play a significant role in elucidating the phenotypic stages of NAFLD. Therefore, LC-MS is not just for basic research: it is a very efficient tool for providing clinically relevant information for patient management. The contribution of lipidomics to understand NAFLD complexity was well evaluated in a review by Masoodi et al. [29].

4 Applications of LC-MS-Based Lipidomics in Clinical Practice

4.1 The OWLiver® Test

The one-way liver (OWLiver®) panel is an in vitro diagnostic test based on the serum lipidomic signature of patients with NAFLD. This panel can stratify this condition into four clinically relevant stages: normal liver, steatosis, NASH, and

"NASH at risk" (NASH + F2 or more). The panel was developed by OWL metabolomics in Spain as a CE-certified IVD for over 8 years.

The company collects serum samples and biopsies from different cohorts of NAFLD patients in Spain, the Czech Republic, Chile, Mexico, Israel, and the USA (Florida and New York). Together with the biopsy readings and clinical and analytical information, they have built up an impressive database consisting of more than 1200 NAFLD patients.

Furthermore, the panel was selected for the two major international consortia: Noninvasive Biomarkers of Metabolic Liver Diseases (NIMBLE) and Liver Investigation: Testing Marker Utility in Steatohepatitis (LITMUS) (US and EU consortia created for the validation of noninvasive diagnosis of NAFLD spectrum) [37].

Currently, the OWLiver panel is available to most Spanish hospitals, and its use is fully reimbursed for the Health Service of the Basque Country. As the panel obtained self-certificate CE marked, it is available for most EU countries. CLIA test is expected to be launched in the USA in late 2023.

The panel applies three different algorithms to the lipidomic signatures obtained after the injection of a small volume (10 μl) into an LC-MS. All three algorithms are BMI-dependent, while two of them are ALT- and AST-dependent. The final output of the panel is the classification of the patient into one of the four mentioned diagnostic categories.

The first algorithm, based on BMI, ALT, AST, and 12 complex lipids, identifies patients with "NASH at risk." Accuracy in comparison with biopsy results in an AUC close to 0.8, which compares very well with the other noninvasive alternatives in development. The selection of this group of patients is clinically very relevant since they would benefit greatly from early treatment.

The second algorithm, based on BMI, ALT, AST, and 16 complex lipids, identifies patients with NASH, which correlates with an AUC versus biopsy of close to 0.8. This classification is also clinically important for patients' prognosis since the morbidity–mortality for this category is higher than for patients with normal liver and simple steatosis. Only NIS 4 has shown comparable accuracy in diagnosing NASH [38]. To date, no other noninvasive procedures with the capability to recognize the inflammatory component of this condition have been reported.

Based on BMI and 11 triglycerides, the third algorithm differentiates NAFLD from normal liver with an AUC close to 0.9 versus biopsy [34]. The accuracy is very close to that observed with Fibroscan. Although this category is not very clinically relevant since a simple echography can provide the same information, it could be useful for epidemiological or population studies because it only requires a blood test.

The first results from NIMBLE were published in 2022 after blinded evaluation of more than 1000 samples from NAFLD patients and exhibited the same accuracy as observed by Mayo et al. [34].

The OWLiver panel, combined with the Fibrosis-4 score (FIB4), is widely in hepatology units used to select "NASH at-risk" patients. This combination reduces false negatives from FIB 4, identifying most of the patients suffering from NASH and fibrosis stage 2 or above [34].

Endocrinology units are also very interested in identifying "NASH at-risk" and NASH patients because the technology allows them to carefully monitor overweight type 2 diabetes patients during their potential long-term progression.

Finally, knowing whether these metabolite changes are reversed when patients improve or even return to normality is vital. Some longitudinal studies are currently being carried out by two major consortia, NIMBLE and LITMUS, to confirm if the original lipidomic signature is recovered when patients improve.

In conclusion, the accuracy shown by the different lipidomic signatures in the OWLiver panel was very competitive compared with biopsy, and the panel is complementary to the other noninvasive tools developed in parallel for the same purpose, such as NIS 4 (based on microRNA), Fibroscan, and MRI [38].

4.2 *Other Diseases*

LC-MS lipidomics has enormous potential for improving the diagnosis and treatment/understanding of complex disease, as was already shown with NAFLD. Several other conditions have been explored using LC-MS lipidomics in the last decade with promising results. Selected studies are presented in Table 1 to show the heterogeneity of the conditions and the potential of the diagnostic tools under development.

Table 1 Potential clinical applications of LC-MS

Biomarkers for disease diagnosis	
Idiopathic noncirrhotic portal hypertension	Seijo et al. [39] showed that a subset of five metabolites differentiates patients with idiopathic noncirrhotic portal hypertension from patients with liver cirrhosis and healthy volunteers (AUROC = 0.8871 [0.838–0.924]). Using high and low cut-off values, the model can diagnose or exclude idiopathic non-cirrhosis portal hypertension, respectively
Multiple sclerosis	Villoslada et al. [40] identified metabolomic signatures for classifying patients versus controls with high accuracy, as well as for classifying patients with a medium to high disability (EDSS 3.0). Among them, sphingomyelin and lysophosphatidylethanolamine were the metabolites that showed a more robust pattern in the time series analysis for discriminating between patients and controls
Alzheimer's disease	Olarazan et al. [41] developed a panel with seven metabolites that could differentiate Alzheimer's disease patients from those with amnestic mild cognitive impairment. The final panel consisted of seven metabolites: Three amino acids (glutamic acid, alanine, and aspartic acid), one non-esterified fatty acid (22:6n-3, DHA), one bile acid (deoxycholic acid), one phosphatidylethanolamine [PE (36:4)], and one sphingomyelin [SM (39:1)]
Colorectal cancer	Cubiella et al. [42] evaluated the fecal levels of 105 metabolites and found 18 that were significantly altered in patients with advanced neoplasia compared to controls. The combinations of seven metabolites, ChoE (18:1), ChoE (18:2), ChoE (20:4), PE (16:0/18:1), SM (d18:1/23:0), SM (42:3), and TG (54:1), discriminated advanced neoplasia patients from healthy controls. These seven metabolites were employed to construct a predictive model that provides an AUC value of 0.821 for cancer diagnosis

5 Summary and Future Outlook

The advances in lipidomic profiling techniques in the last decades have significantly improved our understanding of the biological processes involved in health and disease. Currently, many lipidomic profiling applications are moving from research laboratories to clinical application. However, lipidomics has the potential to widely assist in the routine diagnosis of disease, to stratify risk post-diagnosis, and to monitor the efficacy of both pharmaceutical and lifestyle interventions.

The global objective is to implement these precise and personalized technologies in an impactful way, to scale and make them accessible and effective to help healthcare providers around the world to be time and cost-effective. After technical development and validation, clinical lipidomic platforms will have to be approved by the relevant regulatory body in the country of use, as has already been done for the first platforms in the EU and USA.

Clinical and Implementation Risks Lipidomic applications are based on highly innovative technologies for bio-screening purposes. Although metabolomics has repeatedly been demonstrated to be cost-effective for CVD management, the price of the technology surpasses that of the gold standard, which is a significant barrier to implementation in stressed healthcare systems.

Moreover, lipidomic applications are primarily focused on cardiometabolic diseases (CMD). Lifestyle interventions can address obesity and concomitant metabolic alterations. Innovative solutions may be addressed after adherence to therapeutic improvement and lifestyle interventions, and complementary technologies can appear before the present solutions are broadly applied. However, personalized systems could still be applied to create patient awareness about personalized health status, favoring the adoption of a healthy lifestyle routine.

Regulatory Difficulties Although regulatory bodies currently include novel technologies, large amounts of time and resources are required for approval by a notification body. Particularly for US FDA certification, if there is no precedent, novel technologies may experience serious delays before clinical introduction, even when promoted by previous IVD-CE marketing equivalent technology, threatening the economic sustainability of the whole technological development. Based on solutions presented in this chapter, lipidomics has already crossed the regulatory barriers under the risk classification ISO 13485:2016, requiring technical file preparation, CE declaration, and registry with the European Competent Authority. For future lipidomic applications, the previous regulatory strategy might be useful. The technical documentation must provide evidence of conformance with the essential requirements of 98/79/EC and the imminent regulation (EU) 2017/746 on in vitro medical devices.

Despite the difficulties, the use of lipidomic technologies as diagnostic, prognostic, and evaluation tools is expected to expand enormously as a result of the developments in modern medicine.

Lipidomic-based applications are expected to increase awareness regarding the need for diagnostic and prognostic technologies for other diseases. Lipidomics will help with the clinical assessment of different pathologies, spread the importance of metabolomic-based approaches, and ensure uptake of the already developed applications by clinicians, hospital decision-makers, and regulators.

The introduction of advanced molecular profiling is aligned with the global strategy to incorporate personalized screening tools to define better monitoring and therapeutic strategies that will improve healthcare systems and empower clinicians with more detailed information, allowing an early therapeutic response and thus reducing the development of life-threatening symptoms.

References

1. Monteiro MS, Carvalho M, Bastos ML, Guedes de Pinho P. Metabolomics analysis for biomarker discovery: advances and challenges. Curr Med Chem. 2013;20:257–71. https://doi.org/10.2174/092986713804806621.
2. Rhee EP, Gerszten RE. Metabolomics and cardiovascular biomarker discovery. Clin Chem. 2012;58:139–47. https://doi.org/10.1373/clinchem.2011.169573.
3. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. Biomed Res Int. 2015;2015:354671. https://doi.org/10.1155/2015/354671.
4. Dennis EA. Lipidomics joins the omics evolution. Proc Natl Acad Sci U S A. 2009;106:2089–90. https://doi.org/10.1073/pnas.0812636106.
5. Quehenberger O, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res. 2010;51:3299–305. https://doi.org/10.1194/jlr.M009449.
6. Meikle TG, Huynh K, Giles C, Meikle PJ. Clinical lipidomics: realizing the potential of lipid profiling. J Lipid Res. 2021;62:100127. https://doi.org/10.1016/j.jlr.2021.100127.
7. Burla B, et al. MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines. J Lipid Res. 2018;59:2001–17. https://doi.org/10.1194/jlr.S087163.
8. Li J, Vosegaard T, Guo Z. Applications of nuclear magnetic resonance in lipid analyses: an emerging powerful tool for lipidomics studies. Prog Lipid Res. 2017;68:37–56. https://doi.org/10.1016/j.plipres.2017.09.003.
9. Havel RJ, Eder HA, Bragdon JH. The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. J Clin Invest. 1955;34:1345–53. https://doi.org/10.1172/jci103182.
10. Magos L. C. Lentner (ed.). Geigy Scientific Tables, 8th edition. Vol. 1. Units of Measurement. Body Fluids. Composition of the Body. Nutrition. 1981, 298 pp. Vol. 2. Introduction to Statistics. Statistical Tables. Mathematical Formulae. 1982, 241 pp. Vol. 3. Physical Chemistry. Composition of the Blood. Haematology. Human Somatometric Data. 1984, 359 pp. Vol. 4. Biochemistry. Metabolism of Xenobiotics. Inborn Error of Metabolism. Pharmacogenetics and Ecogenetics. 1986, 330 pp. Ciba-Geigy, Basel, £12.50 each volume. Distributed in U.K. by Farrand Press. J Appl Toxicol. 1987;7:413. https://doi.org/10.1002/jat.2550070617.
11. Fahy E, et al. A comprehensive classification system for lipids. J Lipid Res. 2005;46:839–61. https://doi.org/10.1194/jlr.E400004-JLR200.
12. Psychogios N, et al. The human serum metabolome. PloS one. 2011;6:e16957. https://doi.org/10.1371/journal.pone.0016957.
13. Quehenberger O, Dennis EA. The human plasma lipidome. N Engl J Med. 2011;365:1812–23. https://doi.org/10.1056/NEJMra1104901.

14. Roth GA, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol. 2020;76:2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010.
15. Riazi K, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2022;7:851–61. https://doi.org/10.1016/s2468-1253(22)00165-0.
16. Langlois MR, et al. Quantifying atherogenic lipoproteins for lipid-lowering strategies: consensus-based recommendations from EAS and EFLM. Clin Chem Lab Med. 2020;58:496–517. https://doi.org/10.1515/cclm-2019-1253.
17. McGill HC Jr. The pathogenesis of atherosclerosis. Clin Chem. 1988;34:B33–9.
18. Napoli C, Pignalosa O, de Nigris F, Sica V. Childhood infection and endothelial dysfunction: a potential link in atherosclerosis? Circulation. 2005;111:1568–70. https://doi.org/10.1161/01.Cir.0000161816.52136.66.
19. Stary HC, et al. A definition of the intima of human arteries and of its atherosclerosis-prone regions. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation. 1992;85:391–405. https://doi.org/10.1161/01.cir.85.1.391.
20. Bao W, Srinivasan SR, Wattigney WA, Berenson GS. Persistence of multiple cardiovascular risk clustering related to syndrome X from childhood to young adulthood. The Bogalusa Heart Study. Arch Intern Med. 1994;154:1842–7.
21. Peters SA, et al. Extensive or restricted ultrasound protocols to measure carotid intima-media thickness: analysis of completeness rates and impact on observed rates of change over time. J Am Soc Echocardiogr. 2012;25:91–100. https://doi.org/10.1016/j.echo.2011.09.009.
22. Li S, et al. Childhood cardiovascular risk factors and carotid vascular changes in adulthood: the Bogalusa Heart Study. JAMA. 2003;290:2271–6. https://doi.org/10.1001/jama.290.17.2271.
23. Ouimet M, Barrett TJ, Fisher EA. HDL and reverse cholesterol transport. Circ Res. 2019;124:1505–18. https://doi.org/10.1161/circresaha.119.312617.
24. Sandesara PB, Virani SS, Fazio S, Shapiro MD. The forgotten lipids: triglycerides, remnant cholesterol, and atherosclerotic cardiovascular disease risk. Endocr Rev. 2019;40:537–57. https://doi.org/10.1210/er.2018-00184.
25. Choi SS, Diehl AM. Hepatic triglyceride synthesis and nonalcoholic fatty liver disease. Curr Opin Lipidol. 2008;19:295–300. https://doi.org/10.1097/MOL.0b013e3282ff5e55.
26. Page JM, Harrison SA. NASH and HCC. Clin Liver Dis. 2009;13:631–47. https://doi.org/10.1016/j.cld.2009.07.007.
27. Day CP, James OF. Steatohepatitis: a tale of two "hits"? Gastroenterology. 1998;114:842–5. https://doi.org/10.1016/s0016-5085(98)70599-2.
28. Bellentani S, Scaglioni F, Marino M, Bedogni G. Epidemiology of non-alcoholic fatty liver disease. Dig Dis. 2010;28:155–61. https://doi.org/10.1159/000282080.
29. Masoodi M, et al. Metabolomics and lipidomics in NAFLD: biomarkers and non-invasive diagnostic tests. Nat Rev Gastroenterol Hepatol. 2021;18:835–56. https://doi.org/10.1038/s41575-021-00502-9.
30. Puri P, et al. A lipidomic analysis of nonalcoholic fatty liver disease. Hepatology (BaltimoreMd.). 2007;46:1081–90. https://doi.org/10.1002/hep.21763.
31. Kotronen A, et al. Hepatic stearoyl-CoA desaturase (SCD)-1 activity and diacylglycerol but not ceramide concentrations are increased in the nonalcoholic human fatty liver. Diabetes. 2009;58:203–8. https://doi.org/10.2337/db08-1074.
32. García-Cañaveras JC, Donato MT, Castell JV, Lahoz A. A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations. J Proteome Res. 2011;10:4825–34. https://doi.org/10.1021/pr200629p.
33. Barr J, et al. Obesity-dependent metabolic signatures associated with nonalcoholic fatty liver disease progression. J Proteome Res. 2012;11:2521–32. https://doi.org/10.1021/pr201223p.

34. Mayo R, et al. Metabolomic-based noninvasive serum test to diagnose nonalcoholic steatohepatitis: results from discovery and validation cohorts. Hepatol Commun. 2018;2:807–20. https://doi.org/10.1002/hep4.1188.
35. Mishra A, Younossi ZM. Epidemiology and natural history of non-alcoholic fatty liver disease. J Clin Exp Hepatol. 2012;2:135–44. https://doi.org/10.1016/s0973-6883(12)60102-9.
36. Bril F, et al. Use of a metabolomic approach to non-invasively diagnose non-alcoholic fatty liver disease in patients with type 2 diabetes mellitus. Diabetes Obes Metab. 2018;20:1702–9. https://doi.org/10.1111/dom.13285.
37. Rasmussen DGK, et al. NAFLD and NASH biomarker qualification in the LITMUS consortium - lessons learned. J Hepatol. 2022;78:852. https://doi.org/10.1016/j.jhep.2022.11.028.
38. Sanyal AJ, Shankar S, Yates KP, Daly E, Dehn CA, Bolognese JA, Neuschwander-Tetri BA, Kowdley KV, Vuppalanchi RK, Guy CA, Tonascia JA, Samir AE, Sirlin CB, Sherlock SP, Fowler KJ, Heymann H, Kamphaus TN, Loomba R, Calle RA, for the NIMBLE Project Team. AASLD meeting. 2021.
39. Seijo S, Lozano JJ, Alonso C, Miquel R, Berzigotti A, Reverter E, García-Pagán JC. Metabolomics as a diagnostic tool for idiopathic non-cirrhotic portal hypertension. Liver Int. 2016;36(7):1051–8.
40. Villoslada P, et al. Metabolomic signatures associated with disease severity in multiple sclerosis. Neurol Neuroimmunol Neuroinflamm. 2017;4:e321. https://doi.org/10.1212/nxi.0000000000000321.
41. Olazarán J, et al. A blood-based, 7-metabolite signature for the early diagnosis of Alzheimer's disease. J Alzheimers Dis. 2015;45:1157–73. https://doi.org/10.3233/jad-142925.
42. Cubiella J, et al. Targeted UPLC-MS metabolic analysis of human faeces reveals novel low-invasive candidate markers for colorectal cancer. Cancers (Basel). 2018;10:300.

Bringing Human Serum Lipidomics to the Forefront of Clinical Practice: Two Clinical Diagnosis Success Stories

Núria Amigó Grau and Pablo Ortiz Betes

Abstract The present chapter describes two clinical applications based on LC-MS and NMR lipidomics that have already been introduced into clinical workflows to better stratify metabolic health, including staging nonalcoholic fatty liver disease according to a specific lipid signature for the disease progression and improving the cardiovascular disease risk based on advanced lipoprotein profiling.

The chapter includes a list of potential applications based on the same technologies and details the envisaged risks and limitations.

The implications of developing advanced high-throughput technologies for clinical applications go much further, such as accelerating the deployment of lipidomic-based assessments in the healthcare system, favoring true disruption through precise and personalized medicine based on global bio-screening approaches.

Keywords Nuclear magnetic resonance (NMR) · Metabolomics · Clinical diagnosis · Cardiovascular diseases (CVD) · Lipidomics · Personalized medicine

Abbreviations

ALT	Alanine transaminase
ApoA/B	Apolipoprotein A/B
AST	Aspartate transaminase

N. Amigó Grau (✉)
Biosfer Teslab, Reus, Spain

Department of Basic Medical Sciences, Institut d'Investigació Sanitària Pere Virgili (IISPV), Av. Universitat, Reus, Spain

Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III (ISCIII), Madrid, Spain
e-mail: namigo@biosferteslab.com

P. Ortiz Betes
Biosfer Teslab, Reus, Spain

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_12

AUC/ AUROC	Area under the ROC curve
BMI	Body mass index
CIBERDEM	*Centro de Investigación Biomédicas en Red de Diabetes y Enfermedades Metabólicas*
CLIA	Clinical Laboratory Improvement Amendments
CMD	Cardiometabolic disease
CMR	Cardiometabolic risk
EAS	European Atherosclerosis Society
ESC	European Society of Cardiology
FFA	Free fatty acids
HCC	Hepatocellular carcinoma
HDL	High-density lipoproteins
IDL	Intermediate-density lipoproteins
IISPV	*Institut d'Investigació Sanitària Pere Virgili*
IMT	Intima-media thickness
IRAS	Insulin resistance atherosclerosis
IVD	In vitro diagnostic
LC	Liquid chromatography-mass
LDL	Low-density lipoproteins
LITMUS	Liver Investigation: Testing Marker Utility in Steatohepatitis
LMWM	Low-molecular-weight metabolites
MAFLD	Metabolic-associated fatty liver disease
MESA	Multi-Ethnic Study of Atherosclerosis
ML	Machine learning
MUFAs	Monounsaturated fatty acids
NASH	Nonalcoholic steatohepatitis
NIMBLE	Noninvasive biomarkers of metabolic liver disease
NPV	Negative predictive value
PPV	Positive predictive value
RUO	Research use only
VLDL	Very-low-density lipoproteins
WHO	World Health Organization

1 Introduction

Metabolomics is an omic approach and is closest to phenotypes because metabolites are the end products of complex and transverse molecular pathways (including genomics, transcriptomics, and proteomics). Therefore, metabolomic analysis is a promising strategy for identifying disease-associated biomarkers: metabolites, small molecules, and end products of the interaction between genes and environmental factors [1].

In the search for new biomarkers, there are two main possible approaches that can be taken. The first is a classic approach based on searching for a biomarker that

explains the pathophysiology of metabolic diseases, such as glucose quantification to determine whether a person has diabetes or the quantification of LDL-C for assessing the risk of cardiovascular diseases (CVDs). The second approach involves searching for metabolic patterns that are characteristic of the physiology of a metabolic disease [2]. In this second approach, high-performance analytical techniques, especially liquid chromatography-mass spectrometry (LC-MS) or nuclear magnetic resonance (NMR), are used to analyze the metabolome without prior knowledge of the metabolites' involvement or role in the disease mechanism or physiology [3]. Since the metabolites are the final products of the interactions between the expression of genes and proteins and environmental exposure, and consequently of the biochemical activity that gives rise to the phenotype of the measured sample, they thus characterize part of the patient's metabolism. The metabolomic approach compares the metabolites in fluids or tissues of a patient with those of a healthy subject to see which metabolites are expressed differently [4].

Metabolomics generally includes a wide range of molecules within a biological matrix. The term "lipidomics" should be adopted when the studied metabolites are specifically lipids. In particular, lipidomics is the study of the structure, function, and metabolism of lipids in living organisms [5]. It is a broad field that encompasses the measurement of all of the lipids present in a biological sample as well as the pathways and enzymes involved in their synthesis, degradation, and modification.

Various analytical techniques are applied in lipidomics to identify and quantify the different lipid species present in a sample, especially mass spectrometry, NMR spectroscopy, and gas or liquid chromatography. These techniques allow researchers to get a detailed picture of the lipids that are present in a sample and to study changes in lipid levels in response to different conditions or treatments [6].

Applications of lipidomics have been reported in a range of different areas, including medicine, nutrition, and environmental science. It is used to study the roles of lipids in health and disease, understand the mechanisms through which lipids affect cellular processes and signaling pathways, and identify potential therapeutic targets for treating lipid-related disorders [7].

Lipidome analysis is much more complex than other omic techniques, such as proteome or genome analysis, because of the enormous structural diversity of lipids, which can differ in both their linear and coupled macromolecular compositions [6]. This diversity increases the number of lipid molecules and the complexity of their carriers, lipoproteins. Thus, lipidome and lipoprotein analysis is technically very challenging. LC-MS and NMR are indispensable analytical tools for metabolomic studies in biological fluids. LC-MS and NMR have the advantage of being able to quantify and identify a large number of compounds simultaneously, reproducibly, and effectively [8, 9]. These techniques are often used in large epidemiological studies and are starting to be applied routinely in clinical laboratories.

This chapter is focused on LC-MS- and 1H-NMR-based lipidomic platforms for clinical applications, exploring their benefits, challenges, difficulties, and future opportunities. The application of these technologies in the clinical environment has opened the doors to the future use of such advanced technologies for predicting and monitoring several diseases. The technologies analyzed in the present chapter are

already applied in clinics, exemplifying the potential of precision medicine approaches for personalized diagnostics.

2 Lipids and Lipoproteins: A Biochemical Approach

2.1 General Concepts

Lipids are a diverse group of organic molecules that are essential to life and many biological functions. They include fats, oils, waxes, and other water-insoluble compounds [10].

Lipids are synthesized by living cells through a process called lipid biosynthesis. This process involves condensing fatty acids with glycerol or other alcohols to form triglycerides, the primary component of fats. Lipids can also be synthesized by modifying existing lipids, such as by adding a phosphate group to form a phospholipid or adding a carbohydrate group to form a glycolipid [11].

The structure of lipids is characterized by their hydrophobic nature, which arises from the presence of long, nonpolar hydrocarbon chains. This nonpolarity allows lipids to interact with each other's molecules and form aggregates, such as micelles or lipid bilayers, which can serve as structural components of cell membranes [12].

Lipids are generally found in plasma as lipoproteins, which are macromolecular complexes with a hydrophobic core of neutral lipids, mainly cholesterol esters and triglycerides, surrounded by a hydrophilic layer of phospholipids, unesterified cholesterol, and proteins [13]. Lipoproteins can be classified based on their density, protein composition (apoprotein), or lipid composition (rich in cholesterol or triglycerides). Considering the protein composition, there are two types of lipoproteins, ApoA and ApoB.

Lipids have many functions in living organisms: from energy reserve to insulation, as well as being a structural component of every cell and tissue. They are also important for cellular signaling and regulation [14]. The diversity of lipid structures is reflected by a wide range of physiological functions they perform. The levels of particular lipids in plasma can be used to diagnose diseases.

So far, approximately 4500 metabolites have been identified in human serum. Around half of these metabolites are phospholipids while over a thousand are glycerolipids (triglycerides [TG], diglycerides [DG], and monoacylglycerols) [15]. Thus, lipids make up approximately 75% of the known human serum metabolome.

Analysts have already quantified thousands of distinct species of lipids, including fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, and prenol lipids, representing the six main categories of lipids found in mammals. Moreover, the number is continually increasing as each of these types of lipids can exist in multiple forms [16]. In addition, lipids can undergo chemical modifications, such as the addition of a phosphate group or a carbohydrate group, which can result in the creation of new lipid species. Sphingolipids and glycerophospholipids are

particularly diverse in structure, due to variations in fatty acid content and head groups.

2.2 Lipids Mirror Present and Future Metabolic Health: Two Sides of the Same Problem

Abnormal levels of lipids and lipoproteins in the blood are major risk factors for metabolic and cardiovascular diseases. Metabolic diseases are increasing exponentially worldwide, and their complications, at cardiovascular and liver levels, are the leading cause of mortality worldwide [17]. They have a common characteristic: their etiology is linked to excess fat and associated inflammation. Nonalcoholic fatty liver disease (NAFLD) is associated with obesity and excess fat in the liver, and its prevalence is increasing around the world, especially in Western countries, currently affecting approximately a quarter of people in countries such as the USA [18]. Symmetrically, arterial health is compromised by an excess of fat, which accelerates the progression of arteriosclerosis, the underlying disease of heart disease and stroke and the leading cause of death in all developed countries [17]. Both liver disease and atherosclerosis can develop and progress more rapidly when accompanied by several well-known risk factors and comorbidities, such as obesity and dyslipidemia.

Lipids circulate in the blood in the form of lipoprotein particles, ranging from chylomicrons, as the largest lipoprotein, to very-low-density lipoproteins (VLDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL), which are the smallest. Health services routinely check cholesterol and triglyceride, and doctors frequently prescribe lipid-lowering drugs to treat patients with dyslipidemia [19].

The duration of exposure to high lipid levels is also a crucial risk factor in cardiometabolic health, dramatically increasing the risk of major cardiovascular events, especially for metabolic disorders that start early in childhood. Atherosclerosis typically begins in childhood [20]. Fatty streaks, buildups of lipids in the intima of arteries, are present in nearly all children by 3 years of age [21]. After 8 years of age, these fatty streaks increase, and atherosclerotic plaques form the coronary arteries during adolescence [22]. Children with clusters of risk factors for cardiovascular diseases are likely to have those same risk factors as adults [23].

Carotid ultrasonography screening for subclinical arteriosclerosis has been validated in observational, longitudinal, and randomized clinical assays. Those results were significantly correlated with intravascular coronary ultrasonography, coronary angiography, and pathologic findings of arterial lesions in healthy and CVD patients, and longitudinal studies have demonstrated that increased lipid alterations and intima–media thickness (IMT) in young adults are linked to cardiovascular risk factors in childhood [24, 25].

The contribution of LDL-associated cholesterol to the development of CVD has been well described. While LDL particle accumulation increases the atherosclerotic process, HDL helps eliminate excess cholesterol by reverse cholesterol transport. Low levels of HDL are associated with high cardiovascular risk [26].

Conversely, in circulating chylomicrons and VLDL, triglycerides undergo hydrolysis, producing a pool of free fatty acids (FFAs), which are used by tissues as an energy source. Adipocytes store excess FFAs, favoring the expansion and dysfunction of adipose tissue, increasing insulin resistance and diabetes. This process is associated with excessive levels of saturated FFAs in plasma, increasing their uptake in hepatocytes to exceed metabolic requirements, which leads to hepatic steatosis and inflammation [27].

3 Clinical Relevance of the Lipidome

Lipids play important roles in many biological processes, so it is not surprising that many diseases can be caused by defects in lipoprotein homeostasis, but it does mean that lipids can be used as markers of the disease. Expanding our knowledge of the composition and concentration of lipid metabolites in the plasma lipidome will lead to improved diagnostic capabilities, as well as enhancing pharmacological evaluation and the efficacy of prescribed treatments [28]. A deep analysis of the lipidome might reflex altered synthesis of specific lipid species or identify abnormal underlying pathological lipoprotein patterns. Our understanding of the roles played by lipids and lipoproteins in disease mechanisms is constantly growing as more information on lipids and lipoprotein physiology becomes available, further highlighting the physiological importance of the lipidome composition and transport, which is strictly regulated and interrelated to cellular response.

3.1 Lipids and NAFLD: Introduction to LC-MS

Nonalcoholic fatty liver disease (NAFLD) covers a wide range of disorders, from benign lipid accumulation in the liver (steatosis) to a more complicated clinical stage when fat induces hepatic inflammation and hepatocyte necrosis producing a new stage called nonalcoholic steatohepatitis (NASH) also named as steatohepatitis, when fibrosis progression is added to fat and inflammation. The final stage of disease progression is cirrhosis and/or hepatocarcinoma (HCC) [29]. Interestingly another potential evolution from NASH directly to HCC without any significant fibrosis contribution has also been described [30].

The exact cause of NAFLD is not known. Many factors contribute to this condition, such as excessive food intake, obesity, type 2 diabetes, and dyslipidemia, but not all patients develop NAFLD/NASH, and not all patients with NAFLD/NASH suffer from one of these conditions [30, 31].

The pathophysiology of NAFLD is quite complex, and the progression from hepatic steatosis to the different stages of this condition is not completely understood. However, lipid metabolic changes, including the production of lipotoxic species in the liver, could be responsible for disease progression in NAFLD [30].

LC-MS allows the tracking of more than 400 different lipid species in the liver and serum. Changes in many of these metabolites were followed in several transverse cohorts giving robust data about the lipidomic signature of the different clinical stages of NAFLD, from liver steatosis to steatohepatitis and advanced fibrosis.

Lipids are highly likely to be involved in both the origin and the progression of the disease, so serum lipidomics has, in the last 20 years, been one of the most successful research lines in identifying markers to differentiate different stages of NAFLD [32].

The first attempt to diagnose NAFLD in humans using lipidomics was reported by Puri et al. in 2007 [33]. The researchers analyzed the lipid levels in the livers of people with normal liver tissue, fatty liver disease, and nonalcoholic steatohepatitis (NASH) and found no differences in the FFA contents in the three groups. However, in NAFLD both the TG and DG levels were higher, while the PC level was decreased, which suggests that PC hydrolysis may contribute to DG and TG accumulation in the fatty liver. The SFA-to-MUFA ratio generally decreased across multiple classes of lipids, providing evidence for increased lipogenesis in NAFLD.

These findings were confirmed in two small series by Kotronen and García-Cañaveras through semiquantitative analysis of the full range of lipids by LC-MS [34, 35]. Total lysophospholipids, DG, and TG were found to be elevated in NAFLD, while the stearic-to-oleic acid ratio was lower, indicating increased TG biosynthesis. Increased DG and reduced PUFA are also characteristic of NAFLD.

An additional pilot study (42 biopsied patients) and a pivotal validation one (467 biopsied patients) were published by Barr et al. in 2010 and 2012, respectively, that clearly identified lipidomic signatures associated with NAFLD progression using LC-MS [36, 37].

Especially relevant is Barr's 2012 study that included 467 biopsied individuals, comprising 90 with normal liver and 377 diagnosed with NAFLD (steatosis, $n = 246$; NASH, $n = 131$). In this study, they analyzed 540 circulating metabolites, including amino acids, FA, DG, TG, PC, PE, PI, ceramides, SM, cholesteryl esters, and bile acids. After analyzing the lipidomic data, the authors established a robust "body mass index-dependent lipidomic signature" for reliably and accurately distinguishing liver steatosis from NASH. The areas under the curve (AUC) for lean/pre-obese, obese, and morbidly obese patients were 0.84, 0.85, and 0.87, respectively. Subsequently, a set of 25 BMI-dependent lipids was established using this same cohort of patients, allowing differentiation between steatosis and NASH with AUC of 0.99, 0.90, and 0.91 for lean/pre-obese, obese, and morbidly obese patients, respectively [38].

It is important to distinguish between simple steatosis and NASH because simple steatosis is a generally benign condition, while NASH is a more serious condition that can lead to increased morbidity and mortality [39].

Mayo et al. reanalyzed the lipidomic data from the cohort from Barr's pivotal trial [36] and a new cohort of 192 biopsied NAFLD patients, which allowed them to establish and validate a BMI-dependent algorithm with 20 TGs. The algorithm was able to distinguish between NASH and NAFLD with a high degree of accuracy, with an AUROC of 0.95 versus biopsy. The sensitivity, specificity, positive predictive value, and negative predictive value of the test were 0.83, 0.94, 0.89, and 0.90, respectively [38].

Using the lipidomic signature in 220 patients with type 2 diabetes mellitus, Bril et al. found an AUROC of 0.79 (95% CI 0.68–0.90) in patients with adequate glycemic control for discrimination between steatosis and NASH. However, the discrimination ability was poor in patients with high insulin resistance or poor glycemic control [40].

Validation of the lipidomic signatures in such important cohorts indicates that lipidomic markers will have a major impact in elucidating the phenotypic stages of NAFLD. Therefore, LC-MS is not just for use in basic research: it has also become a very efficient tool for providing clinically relevant information for patient management. The review by Masoodi et al. gives an in-depth discussion of the contribution of LC-MS to lipidomics and the complexity of NAFLD [32].

3.2 Lipoproteins and CVD: Introduction to NMR

According to data from the World Health Organization (WHO), CVDs are the leading cause of mortality and morbidity in developed countries. They are a major burden on society, leading to a significant reduction in the quality of life and health, as well as healthcare costs [17]. These diseases can be present for many years before becoming clinically apparent, making clinical management difficult. Thus, early CVD risk identification is important to delay and prevent its onset.

Traditionally, CVDs have been diagnosed based on the analysis of risk factors such as smoking, high (total and LDL) cholesterol, high blood pressure, obesity, sedentary lifestyles, or type 2 diabetes. However, it has not been possible to accurately identify all individuals at risk for cerebrovascular accidents or complications using such risk factors. Unfortunately, unexpected acute ischemic events still occur at high rates, both in patients known to have arteriosclerosis and subjects thought to be healthy [41].

LDL cholesterol (LDL-C) is the most important lipid factor for assessing an individual's cardiovascular risk. However, many individuals with CVD have normal LDL-C levels [42]. LDL-C concentrations are often normal or only slightly elevated in people with metabolic disorders such as diabetes, obesity, or metabolic syndrome. At the same time, the level of LDL particles (LDL-P) is raised owing to the presence of smaller particles, lower cholesterol levels, and higher atherogenicity, so they are able to access the arterial wall easily. Among the LDL particles, the smallest and densest can easily infiltrate the arterial wall and stick to the extracellular matrix proteoglycans [43]. Small LDL-P

significantly increase the risk of CVD but are not considered by the traditional risk estimation methods.

Beyond LDL-P, lipoproteins rich in triglycerides, including VLDL, intermediate-density lipoproteins (IDL), and cholesterol remnants, favor the development of atherosclerosis as a complementary source of cholesterol and promote inflammation. In particular, lipoproteins below 70 nm in diameter can cross the vascular endothelial barrier and interact with macrophages exacerbating inflammatory mechanisms [44]. Remnant cholesterol transported by triglyceride-rich particles shows a CVD predictive value even higher than that of LDL cholesterol (LDL-C) [45].

The basic and generally used lipid panel for CV risk assessment does not offer an approach to the real protagonists of atherogenicity, the lipoprotein particles [46]. The proatherogenic particle concentration can be indirectly characterized by determining the plasma apolipoprotein B (ApoB) levels [47]. ApoB-100 is the major apolipoprotein of atherogenic lipoproteins. Each lipoprotein particle has only one ApoB molecule, from chylomicrons and VLDL to LDL. For the same amount of LDL-C, a higher concentration of ApoB in plasma ApoB indicates the presence of more atherogenic particles and smaller lipoproteins [48]. The latest guidelines on dyslipidemia and cardiovascular prevention from the European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS) recognize that the LDL-C level is a surrogate for the level of atherogenic particles and that lipid-lowering drugs achieve their beneficial effects by reducing the number of atherogenic particles [49]. These guidelines propose the determination of the ApoB level as a surrogate marker for CVD closer to the number of LDL particles and set out certain concentrations of ApoB as secondary therapeutic goals. Despite this, the concentration of ApoB does not distinguish between triglyceride-rich particles and LDL particles.

For this reason, advanced lipoprotein testing is being developed into a new diagnostic system, helping health specialists improve the control of cardiovascular risk, providing an accurate picture of the complete lipid profile, and establishing personalized therapy for patients [50].

Advanced lipoprotein profiling using NMR allows the determination in one NMR run of the basic lipid profile, including total cholesterol, LDL, HDL, non-HDL, and triglycerides, as well as a more advanced lipoprotein profile that includes the lipid composition, particle size, and concentration of the main lipoprotein classes (VLDL, LDL, and HDL). The complete characterization of the blood lipoprotein profile contributes to personalized preventive and therapeutic decisions in estimating and addressing the CVD risk [51]. This comprehensive characterization of the lipoprotein profile facilitates the detection of individuals with an increased CVD risk [52].

The NMR technique for lipoprotein profiling was set in 1992 and has repeatedly proved helpful measuring a wide spectrum of cardiovascular risk factors in a high-throughput operational mode [53]. NMR allows us to obtain advanced metabolic profiling that includes the detailed lipoprotein profile and a complementary set of information related to the cardiometabolic status (glycoprotein profile and the concentrations of low-molecular-weight metabolites [LMWM], among others).

In particular, NMR biomarker profiling has repeatedly demonstrated cost-effectiveness and high performance and has been used for both preclinical screening and in vitro diagnostics (IVD)-by-NMR discovery and validation [54, 55]. Except for some applications, NMR is broadly considered a research use only (RUO) device. However, the development of the technology over the last decade has facilitated the implementation of NMR metabolomics as an IVD system in the clinical workflow. The specific application of NMR lipoprotein profiling for cardiovascular risk assessment has recently been introduced in advanced lipid units in some countries, favoring true disruption to a preventive, predictive, and precise medicine approach to help with diagnosis and the development of effective, safe medications and doses that are tailored to patients' individual lipidomic profiles [42].

4 Measurement Techniques for Characterization of Lipid Species and Lipoproteins

4.1 Main Techniques Used to Measure Lipid Species for Cardiometabolic Health Assessment

Various MS-based methods can be used to analyze plasma lipids: LC-MS, direct flow injection, and direct-infusion/shotgun MS (DIMS) are the most common approaches. LC-MS/MS is typically applied for the targeted analysis of very low abundancy (nanomoles per liter range) lipid mediators (e.g., eicosanoids, specialized pro-resolving mediators, oxysterols). Meanwhile, DIMS as well as LC-MS and LC-MS/MS can be used for analyzing lipid classes with higher abundancy, in the high micromoles per liter to millimoles per liter range (e.g., glycerolipids, glycerophospholipids, cholesterol esters, and ceramides).

As mentioned, in plasma, lipids are often found bound to soluble carrier proteins (e.g., albumin) or associated with multiprotein assemblies (lipoproteins). Therefore, a single lipid recovery protocol is unlikely to be effective for all analytical approaches. This is because different lipid families require different extraction and analysis protocols. For example, some lipids are more stable than others and can be extracted using milder methods. Additionally, different lipids require different analytical techniques. Furthermore, some analytical approaches require that the lipids be kept in their native state, which means that they cannot be disrupted by the extraction or analysis process. This is important because the structure of the lipids can provide important information about their function.

Therefore, truly comprehensive lipidome analysis requires multiple analytical platforms (each suited for a subset of lipid classes) to be applied in parallel, particular sample preparation techniques and non-destructive approaches for lipoprotein structure profiling [56]. Frequently, plasma lipidome analysis only includes lipid classes that can be determined in one run using the only mass spectrometer that the researchers have available and supplemented by structural lipoprotein analysis using NMR spectroscopy [57].

It is important, however, to achieve cost-effectiveness through high-throughput approaches. MS is applied for more detailed characterization of the lipid family. MS characterization is based on mass difference, while NMR characterization of lipids is based on the spectroscopically different response of the lipids to an external magnetic field according to the size of the lipoprotein particle carrying them.

In this section, we will report two novel clinical applications based on MS and NMR methods for quantifying circulating lipids and lipoproteins as robust and reliable tools for biomedical diagnosis of liver and cardiometabolic diseases.

The two analytical techniques included in the present chapter for performing high-throughput lipid analysis in cardiometabolic health assessment are liquid chromatography-mass spectrometry (LC-MS) and NMR spectroscopy [58]. Both techniques have advantages and disadvantages [59]. MS can resolve more compounds than NMR, has a higher sensitivity, and requires a smaller sample volume [60]. However, MS requires standards and quality control samples for absolute quantification and reproducibility. NMR is highly reproducible and intrinsically quantitative and can be scaled for use in different laboratories [61, 62]. NMR can immediately provide qualitative and quantitative information, including lipoproteins.

Historically, it has been challenging to perform wide-scale lipid profiling. This is because lipid metabolites have a variety of physical properties, which require the use of different purification systems and complex technical procedures. However, the evolution of lipidomics has led to the development of new analytical platforms, particularly in mass spectrometry. These new platforms have streamlined the procedures involved in lipid profiling, allowing for the analysis of many more lipid molecules in greater detail.

4.2 The Complementarity of Serum/Plasma LC/MS and NMR for Lipoprotein Analysis

Mass spectrometry coupled with chromatography (LC-MS) is the most common technique used for analyzing the lipidome due to its sensitivity and selectivity. When LC is combined with a high-resolution accurate mass instrument (e.g., Orbitrap or time-of-flight instruments), large numbers of analytes can be analyzed at the same time. Typically, reversed-phase chromatography is used to separate the analytes before reaching the mass spectrometer, which determines their structure and concentration. Various chromatographic columns are available for separating lipids depending on their chemical structure, such as their lipid class (e.g., phosphatidylcholine or phosphatidylethanolamine head groups) and fatty acyl composition (e.g., chain length and degree of unsaturation) [58, 63–66].

On the other side, NMR is an excellent technique for profiling biofluids under physiological temperature conditions and is especially adept at characterizing complex solutions. Minimal sample preparation is required for profiling serum/plasma samples by NMR. Large numbers of plasma/serum samples can be collected from patients at a high frequency in a standard clinical routine, allowing the detailed characterization of

Table 1 Strengths and weaknesses of LC-MS and NMR for lipidomic analysis

	LC-MS	NMR
High throughput	+	+++
Sensitivity	+++	+
Equipment cost	+	++
Implementation level	+	++
Versatility	++	+
Robustness	+	+
Scalability	+	++
Absolute concentration	+[a]	+++
Labor intensity	+	+++
Specific application field	NAFLD	CVD

[a]Only when the standard is available

dynamic metabolic events. One of the advantages of NMR metabolomics is that it can be used to generate quantitative profiles of solute-state fluids with minimal sample preparation. This allows for a naturalistic, largely unbiased view of their composition that closely represents the in vivo state. In metabolic research, NMR spectroscopy allows comprehensive metabolic profiling associated with different metabolic conditions and inflammation: from small molecules (known as the aqueous metabolome) to large macromolecular complexes (advanced lipoprotein testing based on NMR technology and NMR glycoprotein profiling) from intact plasma or serum, as well as the characterization of different lipid species from lipid plasma or serum extracts.

The clinically relevant information that can be found in 1H NMR spectra of blood plasma goes far beyond standard lipid panels, including a set of parameters associated with lipoprotein profiling for cardiometabolic health (size, composition, and particle concentration) and other parameters, including glycoproteins, LMWM, and some lipid species. Since the vascular wall releases molecules into the bloodstream that reflect the patient pathological processes, the concentrations of these molecules participating in pathological processes could be potential biomarkers for the future appearance of diseases with which to establish predictive mathematical models that can be used in clinics.

As mentioned in the previous section, measuring lipids and lipoproteins using LC-MS and NMR can be complementary, presenting different strengths and weaknesses (Table 1).

The sensitivity of LC-MS exceeds that of NMR by an order or magnitude and hence the higher number of visible metabolites and potential versatility. However, lipoprotein profiling should be performed using NMR. Moreover, NMR exhibits low experimental variability between laboratories. The main source of variability in NMR is intra-operator and intra-day. NMR can measure a quantitative lipoprotein profile over time for individual longitudinal studies.

4.3 LC-MS Lipidomics and 1H NMR Lipoprotein Profiling Technical Aspects (Table 2)

Table 2 LC-MS Lipidomics and 1H NMR Lipoprotein Profiling Specifications

	LC-MS	NMR
Sample handling	Serum and plasma samples derived from blood should be frozen at the LC-MS lab and immediately aliquoted following a platform-specific protocol depending on the lipid class(es) that will be analyzed	Different biochemical protocols are applied to serum and plasma fractions derived from blood samples after collection The coagulation factors (i.e., fibrinogen) and blood cells are removed from serum by centrifugation Plasma is typically obtained from blood samples that have had an anticoagulant agent added (i.e., heparin or EDTA), which produce high-intensity peaks (EDTA) or overlapping signals (heparin) in the NMR spectra, so serum can be preferable to plasma for some NMR applications
Storage	Samples should be stored at −20 °C for less than 10 days or −70 °C or −80 °C if stored for longer periods [67]	Storage must be considered when analyzing lipoproteins and other plasma/serum metabolites by NMR Samples can be stored at 2–4 °C for up to 7 days or at −20 °C for up to 1–2 months However, as some enzymes, for example, plasma esterase, are still active at −20 °C, samples should be stored at −70 °C or −80 °C for longer periods. Good lipoprotein stability in frozen samples stored for more than 10 years has been reported in some cases [68]

5 LC-MS Lipidomics and 1H NMR Lipoprotein Analysis in Clinical Practice

5.1 Clinical Application of LC-MS: The OWLiver® Test

The OWLiver® panel is an in vitro diagnostic test based on the serum lipidomic signature of patients with NAFLD. This panel can stratify this condition into four clinically relevant stages: normal liver, steatosis, NASH, and “NASH at risk” (NASH + F2 or more). The panel was developed by OWL metabolomics in Spain and has been applied as a CE-certified IVD for over 8 years.

The company collects serum samples and biopsies from different cohorts of NAFLD patients in Spain, the Czech Republic, Chile, Mexico, Israel, and the USA (Florida and New York). Combining the biopsy readings with the clinical and analytical data, they have established an impressive database including data for more than 1200 NAFLD patients.

Furthermore, the panel was selected for the two major international consortia: Non-invasive Biomarker of Metabolic Liver Diseases (NIMBLE) and Liver Investigation Testing Marker Utility in Steatohepatitis (LITMUS) [69] (US and EU consortia created for the validation of noninvasive diagnostic of NAFLD stages).

Currently, the OWLiver panel is available to most Spanish hospitals and its use is fully reimbursed for the Health Service of the Basque Country. A CLIA lab test is

also available for most EU countries interested in the panel, and the CLIA test is expected to launch in the USA in 2023.

The panel applies three different algorithms to the lipidomic signatures obtained after the injection into an LC-MS of a very low amount of serum (~10 μl). All three algorithms are BMI-dependent, and two of them also need the patient's alanine amino transferase (ALT) and aspartate amino transferase (AST). The final output of the panel is the classification of the patient into one of the four mentioned diagnostic categories.

The first algorithm, based on BMI, ALT, AST, and 12 complex lipids, identified patients suffering from "NASH at risk." Accuracy in comparison with biopsy results in an AUC close to 0.8, which compares very well with the other noninvasive alternatives in development. The diagnosis of this group of patients is clinically very relevant since they are the ones that would benefit the most from earlier treatment.

The second algorithm, based on BMI, ALT, AST, and 16 complex lipids, identifies patients with NASH, with an AUC versus biopsy of close to 0.8. This classification is also clinically important for the patients' prognosis since the morbidity–mortality of this category is higher than for patients with normal liver and simple steatosis. Only NIS 4 has shown comparable accuracy in diagnosing NASH [70]. To date, no other noninvasive procedure has been able to recognize the inflammatory component of this condition.

Based on BMI and 11 triglycerides, the third algorithm differentiates NAFLD from normal liver patients with an AUC close to 0.9 versus biopsy [38]. The accuracy is very close to that observed with Fibroscan. Although this category is not very clinically relevant since a simple echography can provide the same information, it could be useful for epidemiological or population studies because it only requires a blood test.

The first results from NIMBLE were published in 2022 after evaluation in blind more than 1000 samples from NAFLD patients and exhibited the same accuracy as observed by Mayo et al. [38, 70].

The main use of the OWLiver panel in hepatology units is in combination with FIB4 to select "NASH at-risk" patients. This combination reduces false negatives from FIB 4, identifying most of the patients suffering from NASH and fibrosis stage 2 or more [38, 71].

Endocrinology units are also very interested in identifying "NASH at-risk" and NASH patients because the technology allows them to carefully monitor overweight type 2 diabetes patients during their potential long-term progression.

Finally, knowing if these metabolite changes are reversed when patients improve or even return to normality is vital. Some longitudinal studies are currently being carried out by two major consortia, NIMBLE and LITMUS, to confirm if the original lipidomic signature is recovered once the patient recovers.

In conclusion, the accuracy shown by the different lipidomic signatures in the OWLiver panel was very competitive compared with biopsy, and the panel is complementary to the other noninvasive tools developed in parallel with the same purpose, such as NIS 4 (based on microRNA, Fibroscan, and MRI) [70].

5.2 *Clinical Application of 1H NMR: The Liposcale Test*

Although lipoprotein particle size and number of particles are not routinely measured in clinical practice, they can be efficiently and simultaneously assessed using NMR spectroscopy. Lipoprotein profiling by using NMR was set two decades before and has proved helpful concerning a wide spectrum of metabolic risk factors [72].

Lipoprotein determination by NMR has been established for some years, in both basic research and using commercially available tests; most approaches (one-dimensional approaches) are limited to analysis of lipids and lipoproteins using empirical models based on correlations between the raw NMR data and laboratory biochemical measurements.

The analysis of lipoproteins by 1D 1H NMR spectroscopy is based on particle size. The lipid methyl groups that are transported within the lipoproteins resonate at slightly different frequencies depending on the lipoprotein that transports them, with smaller particles resonating at lower frequencies. Therefore, it is possible to quantify lipoproteins by decomposing the NMR signal of the methyl group of the lipids in individual signals. This method, commercialized by Liposcience (recently acquired by LabCorp), provides the concentrations of particles of the major lipoprotein classes and subclasses from indirectly estimated size, as it is based on a library of one-dimensional NMR spectra of previously isolated lipoprotein classes and an algorithm, which adjusts the NMR signal depending on samples being analyzed.

Therefore, an advanced lipoprotein profile based on NMR allows for the determination of both the basic lipid profile, including total cholesterol, LDL, HDL, non-HDL, and triglycerides, as well as a more advanced lipoprotein profile, which includes lipid composition, particle size, and concentration of the main lipoprotein classes (VLDL, LDL, and HDL) and the concentrations of nine subclasses of lipid particles. The complete characterization of the blood lipoprotein profile contributes to personalized preventive and therapeutic decisions in estimating and addressing the CVD risk. This comprehensive characterization of the lipoprotein profile facilitates the identification of individuals with an increased CVD risk.

The application of an advanced lipoprotein profile to identify the situations described in the following section is particularly relevant.

5.2.1 Individuals with Discordant Levels of LDL-C and LDL-P

LDL particles are highly heterogeneous in size and lipid content. This heterogeneity has led to the definition of two phenotypes or patterns associated with a greater or lesser risk of CVD, depending on the relationship between the levels of LDL-C and LDL-P:

LDL-C > LDL-P or those individuals with particularly large LDL particles with a higher cholesterol content and lower CVD risk.

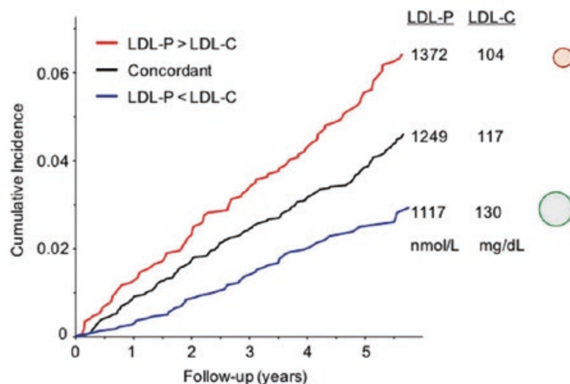

Fig. 1 Incidence of cardiovascular events according to the stratification of the MESA study population concerning LDL-C/LDL-P levels. (Adapted from Otvos et al. [73])

LDL-C < LDL-P or those individuals with especially small LDL particles with a lower cholesterol content and increased risk.

A relevant report by Otvos and colleagues published in the *Journal of Clinical Lipidology*, including approximately 6000 individuals from the prospective Multi-Ethnic Study for Atherosclerosis (MESA), showed that the number of cardiovascular and cerebrovascular accidents accumulated over the years is significantly associated with the level of LDL-P but not with the level of LDL-C if these two magnitudes show discrepancies [73]. As shown in Fig. 1, the incidence of CVD in individuals with especially large LDL particles (and therefore levels of LDL-C > LDL-P) was observed to be significantly lower than in the group of individuals with small particles.

For individuals with a low LDL-P concentration, the CVD risk is overestimated if the traditional LDL-C factor is used. In contrast, for individuals with high LDL-P concentrations but normal LDL-C levels, the CVD risk is underestimated [73].

5.2.2 Lipoprotein Profiles Associated with the Future Development of Type 2 Diabetes and Insulin Resistance

Type 2 diabetes mellitus has been classified as the epidemic of the twenty-first century both for its growing magnitude and its impact on CVD. Determining the factors associated with the onset of diabetes before it starts (prediabetic signs) is a challenge representing a breakthrough in CVD prevention.

In this sense, Dr. Samia Mora [74] and Dr. Rafael Carmena [75] argue that advanced lipoprotein characterization helps to identify individuals at higher risk of developing cardiometabolic diseases by allowing the identification of a characteristic lipoprotein profile years before the onset of clinical manifestations.

This concept has been reinforced in other relevant works, such as the report published in *Circulation* of the IRAS (Insulin Resistance Atherosclerosis Study) [76].

The study showed that NMR can be used to identify lipoprotein profiles that are associated with an increased risk of developing cardiometabolic diseases, such as type 2 diabetes and insulin resistance. These profiles show high concentrations of large VDL particles and small HDL particles. This abnormal distribution between lipoprotein subclasses is reflected in an increased VLDL size and a decreased size of HDL. Early detection of these particular patterns can help prevent the future development of chronic hyperglycemia.

In any case, the measurement of lipoprotein profiles is of great interest to those patients who are exhibiting [77]:

- Family history of atherosclerotic cardiovascular risk
- Elevated triglycerides
- Low levels of HDL-C
- Metabolic syndrome
- Diabetes mellitus
- Secondary prevention: recurrent events due to atherosclerotic cardiovascular risk despite intervention in lifestyle changes and/or administration of lipid-lowering therapy

A recent study by Puig-Jove et al. [78] in the journal *Revista Española de Cardiología* provides a clinical overview of the use of 1H NMR serum lipidomics to directly assess the number, composition, and size of the different lipoprotein particles in other areas of cardiovascular research, such as premature cardiovascular disease (CVD) or heart failure.

Several methodological approaches have been commercially developed to quantify blood lipoproteins from NMR spectra. The first commercially available was based on a deconvolution method (LipoProfile, LabCorp Inc., USA) and includes decomposition of the NMR signal into individual signals obtained from a library of NMR spectra for isolated lipoproteins. Another alternative based on linear regression modeling for lipid prediction was developed by Mika Ala-Korpela and is currently commercialized by the Nightingale Health Ltd., Finland [79, 80]. These approaches do not allow direct quantification of the particle size and provide an indirect measurement of the concentration of particles and estimation of the size.

As an alternative to the abovementioned one-dimensional 1H NMR methods, two-dimensional 1H NMR appears as a second-generation approach for lipoprotein characterization. Diffusion Ordered Nuclear Magnetic Resonance Spectroscopy (DOSY-NMR) is a type of NMR experiment that can be used to measure the hydrodynamic characteristics of molecules. This includes the diffusion coefficient, which is a measure of how fast the molecules move through a solution. The diffusion coefficient of each subclass of lipoprotein can be measured using DOSY-NMR. The sizes of the different lipoprotein subclasses can then be calculated directly from the diffusion coefficients using the Stokes-Einstein equation. However, it is important to directly measure the size of the lipoprotein particles since this is used to calculate the number of lipoprotein particles.

Therefore, the two-dimensional 1H NMR approach allows direct calculation of the sizes and gives more precise particle concentrations than can be obtained using

one-dimensional NMR methods by showing better correlations between [1] the LDL-P and ApoB content of isolated LDL fractions, [2] the VLDL-P and ApoB content of isolated VLDL fractions, and [3] the HDL-P and ApoA-I content of HDL fractions [81].

This methodology was developed recently (Liposcale®, CE marked) and is currently commercialized in Spain (Biosferteslab.com). The technology is also available as a CLIA lab test for most EU countries. The launch in the USA is also expected for 2023. This advanced lipoprotein profiling method determines the number of lipoprotein particles from each of the three main classes of lipoproteins (VLDL, LDL, and HDL) and their three subclasses (large, medium, and small). The size of each lipoprotein class is also determined, as well as the cholesterol and triglyceride content in each fraction, including remnant cholesterol.

5.2.3 Other Diseases and Applications

LC-MS-based and NMR lipidomics has enormous potential to contribute to improving the treatment, understanding and diagnosis of complex disease, as was already shown for NAFLD. Several other conditions have been explored in the last decade with promising results. A short selection of studies is discussed now to show the heterogeneity of the conditions and the potential of the diagnostic tools under development (Table 3).

The earlier examples show the diversity and dimension of the lipidomic and metabolomic fields for clinical applications. Still, they need to follow a long pathway to clinical use. As shown in the section discussing the OWLiver panel, these candidates should be validated in big cohorts from a wide range of patients, and "revalidation" in independent cohorts will be required for international recognition.

The field of potential applications for NMR metabolomics is certainly broad: beyond advanced lipoprotein profiling, 1H NMR technology is capable of simultaneously detecting the presence of a set of blood metabolites (and/or combinations thereof), such as glycoproteins, LMWM (amino acids, sugars), and lipid species that are associated with chronic inflammatory processes and the risk of development of other metabolic diseases and that can therefore be used as early biomarkers in different diseases complementing lipoprotein profiling.

Protein Glycation A change in the protein glycosylation pattern has been identified as a major event that occurs during the transition from healthy to diseased tissue, with significant changes being observed during chronic inflammatory processes, such as obesity, metabolic syndrome, polycystic ovaries syndrome, or rheumatoid arthritis [72, 84–86].

Recently, glyc-A, an NMR-derived biomarker, has been shown to be an independent risk factor for CVD inflammatory diseases, as well as being linked with the residual risk of CVD and death in patients treated with low LDL cholesterol levels [87–90].

Table 3 Potential clinical applications of LC-MS

Biomarkers for disease diagnosis	
Idiopathic noncirrhotic portal hypertension	Seijo et al. showed that a subset of five metabolites can differentiate patients with idiopathic noncirrhotic portal hypertension from patients with liver cirrhosis and healthy volunteers (AUROC = 0.8871 [0.838–0.924]). The model can diagnose or exclude idiopathic non-cirrhosis portal hypertension using high or low cut-off values, respectively
Multiple sclerosis	Villoslada et al. [82] identified metabolomic signatures for distinguishing patients from controls with high accuracy and for classifying patients with a medium to high disability (EDSS 3.0). From their time series analysis, sphingomyelin and lysophosphatidylethanolamine were reported to be the most robust metabolites for discriminating between patients and controls
Alzheimer's disease	Olarazan et al. developed a panel with seven metabolites that could differentiate Alzheimer's disease patients from those with amnestic mild cognitive impairment. Seven metabolites were used in the final panel: Three amino acids (glutamic acid, alanine, and aspartic acid), one non-esterified fatty acid (22:6n-3, DHA), one bile acid (deoxycholic acid), one phosphatidylethanolamine [PE (36:4)], and one sphingomyelin [SM (39:1)]
Colorectal cancer	Cubiella et al. [83] measured the levels of 105 metabolites in feces, and found that 18 were significantly different in patients with advanced neoplasia compared to controls. Seven metabolites, namely, ChoE (18:1), ChoE (18:2), ChoE (20:4), PE (16:0/18:1), SM (d18:1/23:0), SM (42:3), and TG (54:1), could be used to discriminate advanced neoplasia patients from healthy controls. A predictive model establishing using these metabolites exhibited an AUC value of 0.821 for cancer diagnosis
Alcoholic hepatitis	In a study with 90 patients with alcoholic hepatitis and alcoholic cirrhosis, Michelena et al. were able to build an algorithm using only 4 metabolites to diagnose alcoholic hepatitis with an AUC value of 0.932. With another four metabolites, it was possible to select alcoholic hepatitis patients with poor prognosis

LMWM and Specific Lipid Families Dysregulated lipid metabolism is an important and well-known risk factor in cardiovascular diseases. Understanding the whole lipidome signature in vascular pathophysiology is a current challenge in CVD research [91]. On the other hand, aqueous metabolites, including branched-chain and aromatic amino acids, FFAs, and some low-molecular-weight metabolite products of energetic and nitrogenate metabolism, like glycerol, are predictors of insulin resistance and the development of hyperglycemia and type 2 diabetes. It was reported that insulin resistance is linked to higher plasma glutamate levels but lower plasma glutamine levels and glutamine/glutamate ratios. The study also found that an excess of glutamine in blood relative to glutamate was associated with a reduced risk of incidence for T2DM [92].

The clinical development of high-throughput NMR for biomedical screening has recently experienced a significant expansion. Now, it is possible to obtain advanced metabolic profiling that includes the detailed lipoprotein profile and a complementary set of information related to the cardiometabolic status and inflammation, including a systemic glycoprotein profile and the concentrations of LMWM simultaneously. The whole analysis is compatible with the clinical requirements

regarding time, reproducibility, and scalability. It is expected to open a new horizon in evaluating metabolism and position lipidomics toward precision medicine.

In that sense, the present and future of NMR clinical applications—complementing the CE-IVD Liposcale® test for lipoprotein profiling—seems to be a winning strategy shortly approaching affordable refined bioscreening tools to the patients (Table 4).

Table 4 Potential clinical applications of NMR

Biomarkers for disease events and risk prediction	
CVD	Phenylalanine and MUFAs are predictive for a high risk of CVD events. Omega-6 fatty acids and docosahexaenoic acid levels are inversely associated with the risk of a CVD event [93]
The 5-year risk of death	Glycoprotein acetylation, albumin, and VLDL particle size [94]
DM2	Eight amino acids with glycemia, branched-chained and aromatic amino acids, alanine, and glutamine were predictive of diabetes risk. Ketones acetoacetate and β-hydroxybutyrate, lipids, and lipoprotein subclass measures are associated with glycemia and type 2 diabetes risk [95–97]
Inborn errors	NMR-based screening of newborn urine is currently the optimum method for detecting inherited errors of metabolism [98, 99]
Prognosis for HIV patients	A baseline NMR serum metabolomic signature is associated with immunological CD4(+) T-cell recovery in HIV-infected patients [100]
Chronic kidney disease	Lipoprotein alterations [101, 102]
Clinical oncology and cancer research	Branched-chain amino acids, lipoprotein alterations, and glycoprotein profile alteration [103–105]
Parkinson's cognitive impairment	Glycoprotein alterations and HDL composition [106, 107]
Drug interventions	
Statin therapy	Lower levels of small VLDL particles and remnant cholesterol, in addition to the LDL-lowering effects [108]
Hormone therapy	Changes in many fatty acids and amino acids [109]
PPAR-α/PPAR-γ agonist for treatment of NAFLD	Lower proatherogenic profile and remnant cholesterol, in addition to the TG-lowering effects and inflammatory GlycA markers profile [110]
Metabolic risk factor characterization	
Adiposity	Causal effects on numerous metabolic measures: Branched-chain and aromatic amino acids, omega-6 fatty acids, and glycoprotein acetylation, as well as multiple sizes and lipid classes of lipoprotein [111]
Insulin resistance	Lipoprotein subclass profiling [112]
Birth weight	Metabolic signature as the metabolite association pattern with higher adiposity or future cardiometabolic risk [113–115]
Menopause/aging	Glutamine, tyrosine, and isoleucine, in addition to the atherogenic lipoprotein pattern [116]
Alcohol consumption	Biomarkers for alcohol intake beyond routine lipids, including adverse associations with omega-6 fatty acids, MUFAs, glutamine, and citrate [117]
Vitamin D	Large VLDL and small LDL subclasses and related measures, for example, serum triglycerides [118]

6 Concluding Remarks and Future Perspective

The advances in lipidomic profiling techniques in the last two decades have opened up a new era of research into the role of lipids in health and disease, significantly improving our understanding of the underlying biological processes involved. More recently, the related technologies in specific fields, such as NMR lipoprotein profiling and LC-MS liver health evaluation, reached the clinical environment after high-throughput sample preparation, and screening became feasible. These technologies have been shown to have clinical utility in several applications and applied to the analysis of large cohort clinical trials and are cost-effective in the diagnosis of subclinical CVD and metabolic-related conditions, such as metabolic-associated fatty liver disease (MAFLD) evaluation and stratification.

Currently, many lipidomic profiling applications are crossing the barrier from research to clinical application. The technology is now being applied in some advanced clinical settings, and it has the potential to be widely used in the future to diagnose diseases, stratify risk, and monitor the efficacy of all kinds of treatments. Although the technology and techniques are still under development, lipidomic profiling has the potential to revolutionize the way we diagnose and treat diseases. For example, it could be used to identify biomarkers for diseases that are currently difficult to diagnose, such as cancer and Alzheimer's disease. It could also be used to stratify risk for diseases, so that we can better target treatments to those who are most likely to benefit. Additionally, it could be used to monitor the efficacy of treatments, so that clinicians can check that they are working and that patients are not experiencing any adverse side effects. The widespread use of this technology has the potential to improve the health of millions of people around the world.

The global objective is to implement NMR and LC-MS lipidomics as precise and personalized technologies in an impactful way, to scale and make them accessible and effective to help improve the time-efficiency and cost-effectiveness of healthcare providers around the world. Once clinical lipidomic platforms have been developed and validated, they will need to be approved by the relevant regulatory body in the country where they will be used, as has already been done for the first platforms in the EU and USA. This is an important procedure, and it helps to ensure that the platforms are safe and effective.

Specific applications based on MS and NMR are already accessible around the world. However, the cost of the equipment is high, and the potential of the technology is greater than the currently developed applications. The next step is to integrate lipidomic and lipoprotein profiling into clinical practice. For this to be achieved, the standardization of sample preparation is essential to ensure the accuracy and reproducibility of results. Moreover, the streamlining of analytical procedures will make lipidomic profiling more accessible to a wider range of laboratories, and the establishment of metabolite databases will allow for the rapid identification of lipids and the comparison of results between different laboratories.

6.1 Envisaged Risks and Limitations of Clinical Lipidomics

Despite the technological advances, the implementation of new technologies into the clinical routine is not high-throughput. We have identified the principal risks for introducing lipidomics into clinical workflows into either clinical and implementation risks related to the adoption of current LC-MS and NMR approaches and technical and regulatory risks.

Clinical and Implementation Risks Lipidomic applications are based on highly innovative technologies for bioscreening purposes. Although lipidomics has repeatedly been demonstrated to be cost-effective for CVD management, the price of the technology surpasses that of the gold standard, which is a significant barrier in stressed healthcare systems.

Moreover, lipidomic applications are primarily focused on cardiometabolic diseases (CMD). However, lifestyle interventions and lifestyle changes can easily address obesity and concomitant metabolic alterations; the more expensive applications are unlikely to be easily implemented. Innovative solutions may be addressed after adherence to therapeutic improvement and lifestyle interventions, and complementary technologies can appear before the present solutions are broadly applied. However, personalized systems could still be applied to create patient awareness about personalized health status, favoring the adoption of a healthy lifestyle.

Technological Risks Molecular signatures based on LC-MS-NMR are not sufficient to explain the variability associated with cardiometabolic diseases and fail to identify individuals at higher risk of cardiometabolic risk (CMR). Alternatively, developments may produce faster diagnostic tests based on other technologies. However, based on previous evidence-based literature, lipidomic approaches have demonstrated discriminatory capacity among several diseases. They can be used for advanced lipoprotein testing (relevant for CVD events), which has already been recommended in the clinical guidelines for risk management in specific patients in countries such as the USA, Canada, and Spain. Moreover, these technologies could still be applied to future metabolomic profiling in pharmacological interventions, clinical trials, or other clinical cost-effective applications (CVD, inborn errors).

Regulatory Difficulties Although regulatory bodies currently encourage novel technologies, large amounts of time and resources are required for approval by a regulatory body. Particularly for USFDA certification, if there is no precedent, novel technologies may experience serious delays before clinical introduction, even when promoted by previous IVD-CE marketing equivalent technology, threatening the economic sustainability of the whole technological development. Based on solutions presented in this chapter, lipidomics has already crossed the regulatory barriers under the risk classification ISO 13485:2016, requiring technical file preparation, CE declaration, and registry with the European Competent Authority. For future lipidomic applications, the established regulatory strategy might be useful. The technical documentation must provide evidence of conformance with the essential

requirements of 98/79/EC and the imminent regulation (EU) 2017/746 on in vitro diagnostic medical devices.

Despite the difficulties, the use of lipidomic technologies as diagnostic, prognostic, and evaluation tools is expected to expand enormously as a result of the development of modern medicine.

Lipidomic-based applications are expected to increase awareness regarding the need for diagnostic and prognostic technologies for other diseases. Lipidomics will help clinical assessment of different pathologies, spread the importance of metabolomic-based approaches, and ensure uptake of the already developed applications by clinicians, hospital decision-makers, and regulators.

The introduction of advanced molecular profiling is aligned with the global strategy to incorporate personalized screening tools to define better monitoring and treatment strategies that will improve healthcare systems and empower clinicians with more detailed information, allowing an early therapeutic response and thus reducing the development of life-threatening symptoms.

The presented lipidomic-based technologies for liver disease and lipoprotein characterization exemplify the possibility of deploying advanced molecular screening tools in the clinical workflow. Beyond these two examples, using “omic” approaches to unravel the complexity of diseases will accelerate the understanding of diseases, help identify altered mechanisms that have not been described previously, discover prognostic biomarkers, monitor health status, and select better treatments and therapies facilitating the healthcare transformation toward personalized medicine.

References

1. Idle JR, Gonzalez FJ. Metabolomics. Cell Metab. 2007;6(5):348–51.
2. Monteiro MS, Carvalho M, Bastos ML, Guedes de Pinho P. Metabolomics analysis for biomarker discovery: advances and challenges. Curr Med Chem. 2013;20(2):257–71.
3. Rhee EP, Gerszten RE. Metabolomics and cardiovascular biomarker discovery. Clin Chem. 2012;58(1):139–47.
4. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. Biomed Res Int. 2015;2015:354671.
5. Dennis EA. Lipidomics joins the omics evolution. Proc Natl Acad Sci U S A. 2009;106(7):2089–90.
6. Quehenberger O, Armando AM, Brown AH, Milne SB, Myers DS, Merrill AH, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res. 2010;51(11):3299–305.
7. Meikle TG, Huynh K, Giles C, Meikle PJ. Clinical lipidomics: realizing the potential of lipid profiling. J Lipid Res. 2021;62:100127.
8. Burla B, Arita M, Arita M, Bendt AK, Cazenave-Gassiot A, Dennis EA, et al. MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines. J Lipid Res. 2018;59(10):2001–17.
9. Li J, Vosegaard T, Guo Z. Applications of nuclear magnetic resonance in lipid analyses: an emerging powerful tool for lipidomics studies. Prog Lipid Res. 2017;68:37–56.

10. Havel RJ, Eder HA, Bragdon JH. The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. J Clin Invest. 1955;34(9):1345–53.
11. Yeagle PL. Lipid regulation of cell membrane structure and function. FASEB J. 1989;3(7):1833–42.
12. Magos L. C. Lentner (ed.). Geigy Scientific Tables, 8th edition. Vol. 1. Units of Measurement. Body Fluids. Composition of the Body. Nutrition. 1981, 298 pp. Vol. 2. Introduction to Statistics. Statistical Tables. Mathematical Formulae. 1982, 241 pp. Vol. 3. Physical Chemistry. Composition of the Blood. Haematology. Human Somatometric Data. 1984, 359 pp. Vol. 4. Biochemistry. Metabolism of Xenobiotics. Inborn Error of Metabolism. Pharmacogenetics and Ecogenetics. 1986, 330 pp. Ciba-Geigy, Basel, £12.50 each volume. Distributed in U.K. by Farrand Press. J Appl Toxicol. 1987;7(6):413.
13. Fahy E, Subramaniam S, Brown HA, Glass CK, Merrill AH Jr, Murphy RC, et al. A comprehensive classification system for lipids. J Lipid Res. 2005;46(5):839–61.
14. Sunshine H, Iruela-Arispe ML. Membrane lipids and cell signaling. Curr Opin Lipidol. 2017;28(5):408–13.
15. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. PloS One. 2011;6(2):e16957.
16. Quehenberger O, Dennis EA. The human plasma lipidome. N Engl J Med. 2011;365(19):1812–23.
17. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol. 2020;76(25):2982–3021.
18. Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2022;7(9):851–61.
19. Langlois MR, Nordestgaard BG, Langsted A, Chapman MJ, Aakre KM, Baum H, et al. Quantifying atherogenic lipoproteins for lipid-lowering strategies: consensus-based recommendations from EAS and EFLM. Clin Chem Lab Med. 2020;58(4):496–517.
20. McGill HC Jr. The pathogenesis of atherosclerosis. Clin Chem. 1988;34(8b):B33–9.
21. Napoli C, Pignalosa O, de Nigris F, Sica V. Childhood infection and endothelial dysfunction: a potential link in atherosclerosis? Circulation. 2005;111(13):1568–70.
22. Stary HC, Blankenhorn DH, Chandler AB, Glagov S, Insull W Jr, Richardson M, et al. A definition of the intima of human arteries and of its atherosclerosis-prone regions. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation. 1992;85(1):391–405.
23. Bao W, Srinivasan SR, Wattigney WA, Berenson GS. Persistence of multiple cardiovascular risk clustering related to syndrome X from childhood to young adulthood. The Bogalusa Heart Study. Arch Intern Med. 1994;154(16):1842–7.
24. Peters SA, den Ruijter HM, Palmer MK, Grobbee DE, Crouse JR 3rd, O'Leary DH, et al. Extensive or restricted ultrasound protocols to measure carotid intima-media thickness: analysis of completeness rates and impact on observed rates of change over time. J Am Soc Echocardiogr. 2012;25(1):91–100.
25. Li S, Chen W, Srinivasan SR, Bond MG, Tang R, Urbina EM, et al. Childhood cardiovascular risk factors and carotid vascular changes in adulthood: the Bogalusa Heart Study. JAMA. 2003;290(17):2271–6.
26. Ouimet M, Barrett TJ, Fisher EA. HDL and reverse cholesterol transport. Circ Res. 2019;124(10):1505–18.
27. Sandesara PB, Virani SS, Fazio S, Shapiro MD. The forgotten lipids: triglycerides, remnant cholesterol, and atherosclerotic cardiovascular disease risk. Endocr Rev. 2019;40(2):537–57.
28. Choi SS, Diehl AM. Hepatic triglyceride synthesis and nonalcoholic fatty liver disease. Curr Opin Lipidol. 2008;19(3):295–300.
29. Page JM, Harrison SA. NASH and HCC. Clin Liver Dis. 2009;13(4):631–47.
30. Day CP, James OF. Steatohepatitis: a tale of two "hits"? Gastroenterology. 1998;114(4):842–5.

31. Bellentani S, Scaglioni F, Marino M, Bedogni G. Epidemiology of non-alcoholic fatty liver disease. Dig Dis. 2010;28(1):155–61.
32. Masoodi M, Gastaldelli A, Hyötyläinen T, Arretxe E, Alonso C, Gaggini M, et al. Metabolomics and lipidomics in NAFLD: biomarkers and non-invasive diagnostic tests. Nat Rev Gastroenterol Hepatol. 2021;18(12):835–56.
33. Puri P, Baillie RA, Wiest MM, Mirshahi F, Choudhury J, Cheung O, et al. A lipidomic analysis of nonalcoholic fatty liver disease. Hepatology. 2007;46(4):1081–90.
34. García-Cañaveras JC, Donato MT, Castell JV, Lahoz A. A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations. J Proteome Res. 2011;10(10):4825–34.
35. Kotronen A, Seppänen-Laakso T, Westerbacka J, Kiviluoto T, Arola J, Ruskeepää AL, et al. Hepatic stearoyl-CoA desaturase (SCD)-1 activity and diacylglycerol but not ceramide concentrations are increased in the nonalcoholic human fatty liver. Diabetes. 2009;58(1):203–8.
36. Barr J, Caballería J, Martínez-Arranz I, Domínguez-Díez A, Alonso C, Muntané J, et al. Obesity-dependent metabolic signatures associated with nonalcoholic fatty liver disease progression. J Proteome Res. 2012;11(4):2521–32.
37. Barr J, Vázquez-Chantada M, Alonso C, Pérez-Cormenzana M, Mayo R, Galán A, Caballería J, Martín-Duce A, Tran A, Wagner C, Luka Z, Lu SC, Castro A, Le Marchand-Brustel Y, Martínez-Chantar ML, Veyrie N, Clément K, Tordjman J, Gual P, Mato JM. Liquid chromatography-mass spectrometry-based parallel metabolic profiling of human and mouse model serum reveals putative biomarkers associated with the progression of nonalcoholic fatty liver disease. J Proteome Res. 2010;9(9):4501–12.
38. Mayo R, Crespo J, Martínez-Arranz I, Banales JM, Arias M, Mincholé I, et al. Metabolomic-based noninvasive serum test to diagnose nonalcoholic steatohepatitis: results from discovery and validation cohorts. Hepatol Commun. 2018;2(7):807–20.
39. Mishra A, Younossi ZM. Epidemiology and natural history of non-alcoholic fatty liver disease. J Clin Exp Hepatol. 2012;2(2):135–44.
40. Bril F, Millán L, Kalavalapalli S, McPhaul MJ, Caulfield MP, Martinez-Arranz I, et al. Use of a metabolomic approach to non-invasively diagnose non-alcoholic fatty liver disease in patients with type 2 diabetes mellitus. Diabetes Obes Metab. 2018;20(7):1702–9.
41. Sachdeva A, Cannon CP, Deedwania PC, Labresh KA, Smith SC Jr, Dai D, et al. Lipid levels in patients hospitalized with coronary artery disease: an analysis of 136,905 hospitalizations in get with the guidelines. Am Heart J. 2009;157(1):111–7.e2.
42. Langlois MR, Chapman MJ, Cobbaert C, Mora S, Remaley AT, Ros E, et al. Quantifying atherogenic lipoproteins: current and future challenges in the era of personalized medicine and very low concentrations of LDL cholesterol. A consensus statement from EAS and EFLM. Clin Chem. 2018;64(7):1006–33.
43. Tehrani DM, Zhao Y, Blaha MJ, Mora S, Mackey RH, Michos ED, et al. Discordance of low-density lipoprotein and high-density lipoprotein cholesterol particle versus cholesterol concentration for the prediction of cardiovascular disease in patients with metabolic syndrome and diabetes mellitus (from the multi-ethnic study of atherosclerosis [MESA]). Am J Cardiol. 2016;117(12):1921–7.
44. Borén J, Williams KJ. The central role of arterial retention of cholesterol-rich apolipoprotein-B-containing lipoproteins in the pathogenesis of atherosclerosis: a triumph of simplicity. Curr Opin Lipidol. 2016;27(5):473–83.
45. Castañer O, Pintó X, Subirana I, Amor AJ, Ros E, Hernáez Á, et al. Remnant cholesterol, not LDL cholesterol, is associated with incident cardiovascular disease. J Am Coll Cardiol. 2020;76(23):2712–24.
46. Ference BA, Graham I, Tokgozoglu L, Catapano AL. Impact of lipids on cardiovascular health: JACC health promotion series. J Am Coll Cardiol. 2018;72(10):1141–56.
47. Cole TG, Contois JH, Csako G, McConnell JP, Remaley AT, Devaraj S, et al. Association of apolipoprotein B and nuclear magnetic resonance spectroscopy-derived LDL particle num-

ber with outcomes in 25 clinical studies: assessment by the AACC Lipoprotein and Vascular Diseases Division Working Group on Best Practices. Clin Chem. 2013;59(5):752–70.
48. Pichler G, Amigo N, Tellez-Plaza M, Pardo-Cea MA, Dominguez-Lucas A, Marrachelli VG, et al. LDL particle size and composition and incident cardiovascular disease in a South-European population: the Hortega-Liposcale Follow-up Study. Int J Cardiol. 2018;264:172–8.
49. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, et al. 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. Eur Heart J. 2020;41(1):111–88.
50. Masana L, Ibarretxe D. Magnetic resonance-assessed lipoprotein profile. The time has come for its clinical use. Rev Esp Cardiol (Engl ed). 2022;75(1):5–8.
51. Aday AW, Lawler PR, Cook NR, Ridker PM, Mora S, Pradhan AD. Lipoprotein particle profiles, standard lipids, and peripheral artery disease incidence. Circulation. 2018;138(21):2330–41.
52. Mora S, Buring JE, Ridker PM. Discordance of low-density lipoprotein (LDL) cholesterol with alternative LDL-related measures and future coronary events. Circulation. 2014;129(5):553–61.
53. Otvos JD, et al. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. Clin Chem. 1992;38:1632.
54. Folse HJ, Goswami D, Rengarajan B, Budoff M, Kahn R. Clinical-and cost-effectiveness of LDL particle-guided statin therapy: a simulation study. Atherosclerosis. 2014;236(1):154–61.
55. Allaire J, Vors C, Couture P, Lamarche B. LDL particle number and size and cardiovascular risk: anything new under the sun? Curr Opin Lipidol. 2017;28(3):261–6.
56. Li M, Yang L, Bai Y, Liu H. Analytical methods in lipidomics and their applications. Anal Chem. 2014;86(1):161–75.
57. Wishart DS. Quantitative metabolomics using NMR. TrAC Trends Anal Chem. 2008;27(3):228–37.
58. Cajka T, Fiehn O. Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. Trends Analyt Chem. 2014;61:192–206.
59. Jeremy K, Nicholson JCL. Systems biology: metabonomics. Nature. 2008;455(7216):1054–6.
60. DeFilippis AP, Trainor PJ, Hill BG, Amraotkar AR, Rai SN, Hirsch GA, Rouchka EC, Bhatnagar A. Identification of a plasma metabolomic signature of thrombotic myocardial infarction that is distinct from non-thrombotic myocardial infarction and stable coronary artery disease. PLoS One. 2017;12(4):e0175591.
61. Gallo V, Intini N, Mastrorilli P, Latronico M, Scapicchio P, Triggiani M, et al. Performance assessment in fingerprinting and multi component quantitative NMR analyses. Anal Chem. 2015;87:6709–17.
62. Eckhart AD, Beebe K, Milburn M. Metabolomics as a key integrator for "omic" advancement of personalized medicine and future therapies. Clin Trans Sci. 2012;5:285–8.
63. Cajka T, Fiehn O. Increasing lipidomic coverage by selecting optimal mobile-phase modifiers in LC–MS of blood plasma. Metabolomics. 2016;12(34).
64. Isaac G, McDonald S, Astarita G. Lipid separation using UPLC with charged surface hybrid technology. Milford, MA: Waters Corp.
65. Ulmer CZ, Patterson RE, Koelmel JP, Garrett TJ, Yost RAA. Robust lipidomics workflow for mammalian cells, plasma, and tissue using liquid-chromatography high- resolution tandem mass spectrometry. Methods Mol Biol. 2017;1609:91–106.
66. Haider A, Wei Y-C, Lim K, Barbosa AD, Liu C-H, Weber U, Mlodzik M, Oras K, Collier S, Hussain MM, et al. PCYT1A regulates phosphatidylcholine homeostasis from the inner nuclear membrane in response to membrane stored curvature elastic stress. Dev Cell. 2018;45:481–95.
67. Pawula M, Hawthorne G, Smith GT, Hill HM. Best practice in biological sample collection, processing, and storage for LC-MS in bioanalysis of drugs. In: Handbook of LC-MS bioanalysis: best practices, experimental protocols, and regulations; 2013. p. 139–64.

68. Loo RL, Lodge S, Kimhofer T, Bong SH, Begum S, Whiley L, Holmes E. 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. 2020;19(11):4428–41.
69. Rasmussen DGK, Anstee QM, Torstenson R, Golding B, Patterson SD, Brass C, Karsdal MA. NAFLD and NASH Biomarker Qualification in the LITMUS consortium–Lessons learned. J Hepatol. 2022. 2023;78(4):852–65. https://doi.org/10.1016/j.jhep.2022.11.028.
70. Sanyal AJ, Yates KP, Daly E, Dehn CA, Neuschwander-Tetri B, Kowdley KV, Vuppalanchi RK, Tonascia JA, Samir AE, Sirlin CB, Fowler KJ, Heymann H, Kamphaus TN, Loomba R, Calle RA; the NIMBLE Project Team. Primary results of the nimble stage 1-NASH CRN study of circulating biomarkers for nonalcoholic steatohepatitis and its activity and fibrosis stage. AASLD meeting. 2021.
71. Noureddin MM, et al. Si llega a tiempo la cita ponemos esa, si no referenciamos el abstract poblicado en J of Hepatology por los mismos autores. Hepatology. 2023.
72. Connelly MA, Gruppen EG, Wolak-Dinsmore J, et al. GlycA, a marker of acute phase glycoproteins, and the risk of incident type 2 diabetes mellitus: PREVEND study. Clin Chim Acta. 2016;452:10–7.
73. Otvos JD, Mora S, Shalaurova I, Greenland P, Mackey RH, Goff DC Jr. Clinical implications of discordance between LDL cholesterol and LDL particle number. J Clin Lipidol. 2011;5(2):105–13.
74. Mora S, Otvos JD, Rifai N, Rosenson RS, Buring JE, Ridker PM. Lipoprotein particle profiles by nuclear magnetic resonance compared with standard lipids and apolipoproteins in predicting incident cardiovascular disease in women. Circulation. 2009;119:931–9.
75. Carmena R, Duriez P, Fruchart JC. Atherogenic lipoprotein particles in atherosclerosis. Circulation. 2004;109(23 Suppl 1):III2–7.
76. Festa A, Williams K, Hanley AJ, et al. Nuclear magnetic resonance lipoprotein abnormalities in prediabetic subjects in the Insulin Resistance Atherosclerosis Study. Circulation. 2005;111(25):3465–72.
77. Pintó X, Masana L, Civeira F, et al. Consensus document of an expert group from the Spanish Society of Arteriosclerosis (SEA) on the clinical use of nuclear magnetic resonance to assess lipoprotein metabolism (Liposcale®). Clin Investig Arterioscler. 2020;32(5):219–29.
78. Puig-Jové C, Castelblanco E, Falguera M, et al. Advanced lipoprotein profile in individuals with normal and impaired glucose metabolismPerfil lipoproteico avanzado en individuos con metabolismo glucémico normal y alterado. Rev Esp Cardiol (Engl Ed). 2022;75(1):22–30.
79. Aru V, Lam C, Khakimov B, Hoefsloot HCJ, Zwanenburg G, Lind MV, Schäfer H, van Duynhoven J, Jacobs DM, Smilde AK, Engelsen SB. Quantification of lipoprotein profiles by nuclear magnetic resonance spectroscopy and multivariate data analysis. Trends Analyt Chem. 2017;94:210–9.
80. Ala-Korpela M, Korhonen A, Keisala J, et al. 1H NMR-based absolute quantitation of human lipoproteins and their lipid contents directly from plasma. J Lipid Res. 1994;35(12):2292–304.
81. Mallol R, Amigó N, Rodríguez MA, et al. Liposcale: a novel advanced lipoprotein test based on 2D diffusion-ordered 1H NMR spectroscopy. J Lipid Res. 2015;56(3):737–46.
82. Villoslada P, Alonso C, Agirrezabal I, Kotelnikova E, Zubizarreta I, Pulido-Valdeolivas I, et al. Metabolomic signatures associated with disease severity in multiple sclerosis. Neurol Neuroimmunol Neuroinflamm. 2017;4(2):e321.
83. Cubiella J, Clos-Garcia M, Alonso C, Martinez-Arranz I, Perez-Cormenzana M, Barrenetxea Z, et al. Targeted UPLC-MS metabolic analysis of human faeces reveals novel low-invasive candidate markers for colorectal cancer. Cancers (Basel). 2018;10(9):300.
84. Carmona-Maurici J, Amigó N, Cuello E, et al. Bariatric surgery decreases oxidative stress and protein glycosylation in patients with morbid obesity. Eur J Clin Investig. 2020;50(11):e13320.
85. Moncayo S, Insenser M, Martínez-García MÁ, et al. Acute-phase glycoprotein profile responses to different oral macronutrient challenges: influence of sex, functional hyperandrogenism and obesity. Clin Nutr. 2021;40(3):1241–6.

86. Fuertes-Martín R, Taverner D, Vallvé JC, et al. Characterization of 1H NMR plasma glycoproteins as a new strategy to identify inflammatory patterns in rheumatoid arthritis. J Proteome Res. 2018;17(11):3730–9.
87. Fuertes-Martín R, Correig X, Vallvé JC, Amigó N. Human serum/plasma glycoprotein analysis by 1H-NMR, an emerging method of inflammatory assessment. J Clin Med. 2020;9(2):354.
88. McGarrah RW, Kelly JP, Craig DM, et al. A novel protein glycan-derived inflammation biomarker independently predicts cardiovascular disease and modifies the association of HDL subclasses with mortality. Clin Chem. 2017;63(1):288.
89. Holmes MV, Millwood IY, Kartsonaki C, et al. Lipids, lipoproteins, and metabolites and risk of myocardial infarction and stroke. J Am Coll Cardiol. 2018;71(6):620–32.
90. Otvos JD, Guyton JR, Connelly MA, et al. Relations of GlycA and lipoprotein particle subspecies with cardiovascular events and mortality: a post hoc analysis of the AIM-HIGH trial. J Clin Lipidol. 2018;12(2):348–55.
91. Kohno S, Keenan AL, Ntambi JM, Miyazaki M. Lipidomic insight into cardiovascular diseases. Biochem Biophys Res Commun. 2018;504(3):590–5.
92. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation. 2012;125(18):2222–31.
93. Würtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation. 2015;131(9):774–85.
94. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on-omic technologies. Am J Epidemiol. 2017;186(9):1084–96.
95. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448–53.
96. Ahola-Olli AV, Mustelin L, Kalimeri M, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia. 2019;62(12):2298–309.
97. Hameed A, Mojsak P, Buczynska A, Suleria HAR, Kretowski A, Ciborowski M. Altered metabolome of lipids and amino acids species: a source of early signature biomarkers of T2DM. J Clin Med. 2020;9(7):2257.
98. Mussap M, Zaffanello M, Fanos V. Metabolomics: a challenge for detecting and monitoring inborn errors of metabolism. Ann Transl Med. 2018;6(17):338.
99. Embade N, Cannet C, Diercks T, et al. NMR-based newborn urine screening for optimized detection of inherited errors of metabolism. Sci Rep. 2019;9:13607.
100. Malo AI, Rull A, Girona J, et al. Glycoprotein profile assessed by 1H-NMR as a global inflammation marker in patients with HIV infection. A prospective study. J Clin Med. 2020;9(5):1344.
101. Zhang Y, Zhang S, Wang G. Metabolomic biomarkers in diabetic kidney diseases—a systematic review. J Diabetes Complicat. 2015;29(8):1345–51.
102. Bermudez-Lopez M, Perpiñan H, Amigo N, et al. Advanced lipoprotein parameters could better explain atheromatosis in non-diabetic chronic kidney disease patients. Clin Kidney J. 2021;14(12):2591–9.
103. Elebo N, Omoshoro-Jones J, Fru PN, et al. Serum metabolomic and lipoprotein profiling of pancreatic ductal adenocarcinoma patients of African ancestry. Meta. 2021;11(10):663.
104. Mayers JR, Wu C, Clish CB, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. Nat Med. 2014;20(10):1193–8.
105. Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG, Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. Cancer J Clin. 2021;71(4):333–58.
106. Laguna A, Xicoy H, Tolosa E, et al. Serum metabolic biomarkers for synucleinopathy conversion in isolated REM sleep behavior disorder. npj Parkinsons Dis. 2021;7:40.

107. Rademacher TD, Meuth SG, Wiendl H, Johnen A, Landmeyer NC. Molecular biomarkers and cognitive impairment in multiple sclerosis: state of the field, limitations, and future direction–a systematic review and meta-analysis. Neurosci Biobehav Rev. 2023;146:105035.
108. Langlois MR, Chapman MJ, Cobbaert C, et al. Quantifying atherogenic lipoproteins: current and future challenges in the era of personalized medicine and very low concentrations of LDL cholesterol. A consensus statement from EAS and EFLM. Clin Chem. 2018;64(7):1006–33.
109. Karppinen JE, Törmäkangas T, Kujala UM, et al. Menopause modulates the circulating metabolome: evidence from a prospective cohort study. Eur J Prev Cardiol. 2022;29(10):1448–59.
110. Gawrieh S, Noureddin M, Loo N, et al. Saroglitazar, a PPAR-α/γ agonist, for treatment of NAFLD: a randomized controlled double-blind phase 2 trial. Hepatology. 2021;74(4):1809–924.
111. O'Keeffe LM, Bell JA, O'Neill KN, et al. Sex-specific associations of adiposity with cardiometabolic traits in the UK: a multi–life stage cohort study with repeat metabolomics. PLoS Med. 2022;19(1):e1003636.
112. Iida M, Harada S, Takebayashi T. Application of metabolomics to epidemiological studies of atherosclerosis and cardiovascular disease. J Atheroscler Thromb. 2019;26(9):747–57.
113. Miranda J, Simões RV, Paules C, et al. Metabolic profiling and targeted lipidomics in small for gestational age and foetal growth restriction. Sci Rep. 2018;9:13614.
114. Youssef L, Simões RV, Miranda J, et al. Paired maternal and fetal metabolomics reveal a differential fingerprint in preeclampsia versus fetal growth restriction. Sci Rep. 2021;11(1):14422.
115. Algaba-Chueca F, Maymó-Masip E, Ballesteros M, et al. Cord blood advanced lipoprotein testing reveals an interaction between gestational diabetes and birth-weight and suggests a new early biomarker of infant obesity. Biomedicine. 2022;10(5):1033.
116. Panyard DJ, Yu B, Snyder MP. The metabolomics of human aging: advances, challenges, and opportunities. Sci Adv. 2022;8(42):eadd6155.
117. Silva RA, Pereira TCS, Souza AR, Ribeiro PR. 1H NMR-based metabolite profiling for biomarker identification. Clin Chim Acta. 2020;502:269–79.
118. Bashir NA, Bashir AAM, Bashir HA. Effect of vitamin D deficiency on lipid profile. Am J Lab Med. 2019;4(1):11–8.

LC-MS-Based Population Metabolomics: A Mini-Review of Recent Studies and Challenges from Sample Collection to Data Processing

Myriam Mireault and Lekha Sleno

Abstract Metabolomics aims to identify and quantify metabolites in biological samples to understand better biological changes resulting from lifestyle, environment, or disease. This is challenging due to the structural diversity of the metabolites and the complexity of samples of interest, such as blood and urine, useful in population studies to study biological changes in large cohorts. The limited number of commercially available standards and incomplete metabolite spectral databases impedes the identification of many metabolites. Furthermore, the need for more standardization in sample preparation, analysis, and interpretation of data is an important issue that can influence results in large cohort studies. Variations or errors occurring during the pre-analytical stage can highly affect levels of metabolites. In this mini-review, we outline the challenges associated with population metabolomic studies and show an overview of current practices in the field with some case studies.

Keywords Population studies · Metabolomics · Sample collection · Analytical considerations · Data processing · Mini-review · Liquid chromatography · Mass spectrometry

Abbreviations

APCI	Atmospheric pressure chemical ionization
ASQ	Age and stage questionnaire
BMI	Body mass index
CI	Chemical ionization
COMETS	COnsortium of METabolomic Studies
CV	Coefficient of variation

M. Mireault · L. Sleno (✉)
Chemistry Department/CERMO-FC, Université du Québec à Montréal (UQAM), Montreal, QC, Canada
e-mail: Sleno.lekha@uqam.ca

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_13

DNA	Deoxyribonucleic acid
EDIH	Empirical dietary index of hyperinsulinemia
EDTA	Ethylenediaminetetraacetic acid
EI	Electron ionization (electron impact)
ESI	Electrospray ionization
FDR	False discovery rate
GC-MS	Gas chromatography-mass spectrometry
HDL	High-density lipoprotein
HMDB	Human metabolome database
IS	Internal standard
LC-MS	Liquid chromatography-mass spectrometry
LOD	Limit of detection
LTR	Long-term reference
MALDI	Matrix-assisted laser desorption/ionization
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random
MS	Mass spectrometry
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
OPLS-DA	Orthogonal partial least-squares discriminant analysis
PBS	Phosphate-buffered saline
PCA	Principal component analysis
QA	Quality assurance
QC	Quality control
QqQ	Triple quadrupole
QqTOF	Quadrupole time of flight
QRILC	Quantile regression imputation of left-censored data
RNA	Ribonucleic acid
ROC	Receiver operating characteristic curve
SRM	Standard reference material
VLDL	Very-low-density lipoprotein
WGCNA	Weighted gene co-expression network analysis

1 Introduction

Metabolomics is one of the most recent branches of "omic" science, studying the molecules involved in the structure, function, or dynamics of a cell, tissue, or organism [1, 2]. It encompasses the qualitative and quantitative analysis of metabolites present in a biological sample to interrogate metabolic pathways. It is possible with metabolomics to investigate potential therapeutic targets and identify biomarkers for a disease's early detection and progression [3, 4]. Small molecule metabolites (less than 1500 Da) are intermediates or end products of metabolism and comprise

a large variety of chemical structures [2, 5]. They are also considered to correlate more with the phenotype of a tissue or organism than proteins and genes, thereby allowing fast detection of biological changes [5].

The metabolome's structural diversity increases the analyses' complexity compared to the proteome, genome, and transcriptome (Fig. 1). Proteins are combinations of 20 amino acids, while DNA and RNA are each composed of 4 nucleotides. Metabolites are formed from different endogenous or exogenous compounds and do not have predefined structures [6]. Isomeric metabolites have different structures but identical exact mass, increasing the difficulty of metabolomic analyses [5].

The choice of a biological system can influence the metabolome. Urine and blood are commonly used in metabolomics because of their wide range of metabolites and their accessibility, as well as having the ability to represent an overall picture of metabolism from different organs and tissues [7]. Measuring polar metabolites in urine can inform on the major elimination pathways from food, drugs, and environmental contaminants [8]. Blood is composed of polar and non-polar molecules, excreted or secreted by tissues [9]. Thus, these two biofluids offer a global view of a person's exposure and health status [7].

In metabolomics, two approaches can be employed; targeted, and untargeted analysis. The targeted approach quantifies a limited number of known metabolites in a sample. However, the small number of known metabolites and commercially available pure reference standards limits this approach in epidemiological studies. The untargeted analysis provides a more comprehensive view by identifying known and unknown metabolites. It requires a nonselective and reproducible preparation to

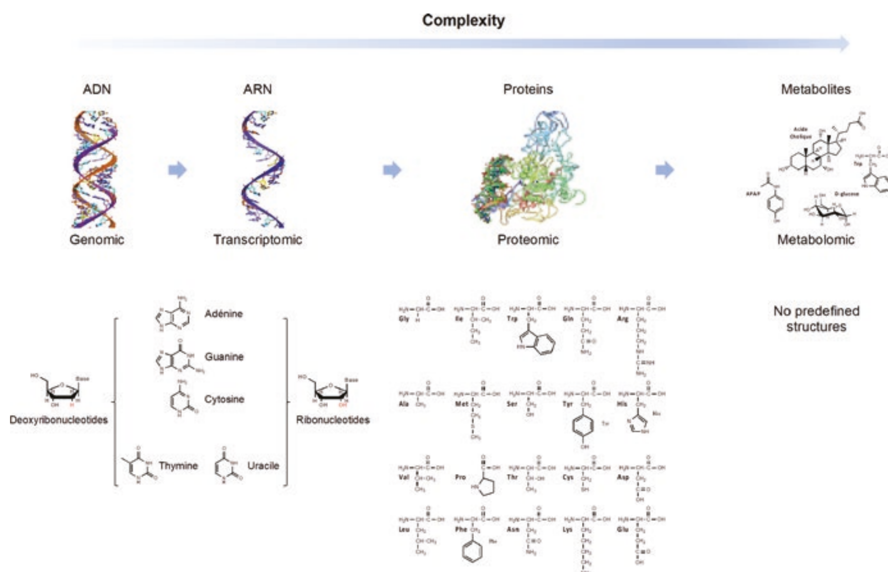

Fig. 1 The complexity of the different omic sciences, showing structural diversity involved in metabolomic studies

avoid any change in the metabolic profile [10]. Due to a lack of standardization of protocols, several pre-analytical, analytical, and post-analytical approaches may influence the results of such studies. Common practices and challenges related to epidemiological studies using urine and blood will be discussed in this chapter.

2 Pre-Analytical Factors

The pre-analytical steps (Figs. 2 and 3) greatly impact the results, corresponding to 60–80% of laboratory errors [11, 12]. Due to the presence of enzymes, metabolism continues even after biological samples are collected unless properly quenched, resulting in varying levels of some compounds [7]. Sample degradation is also an important factor to consider when analyzing metabolites with limited stability. Therefore, sample collection, preparation, and storage must be considered crucial steps in metabolomic studies [13].

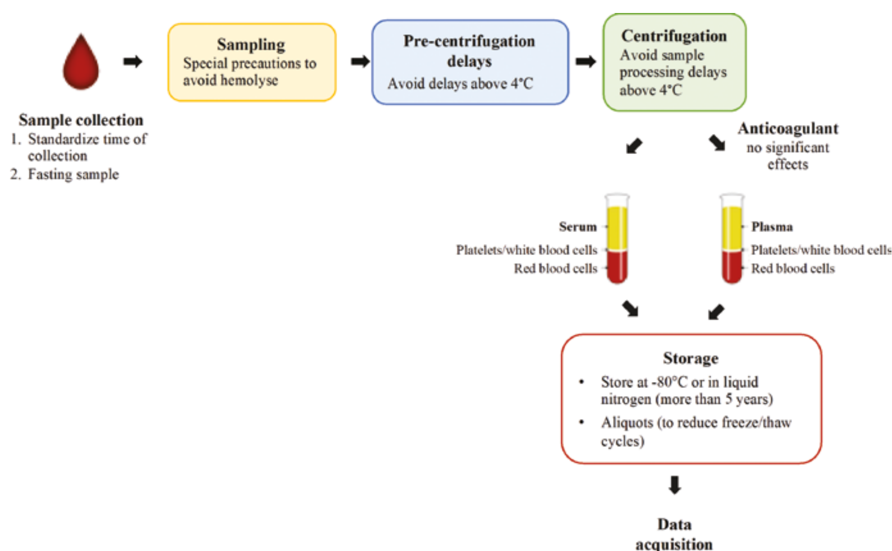

Fig. 2 Pre-analytical steps involved in metabolomic analyses of blood samples

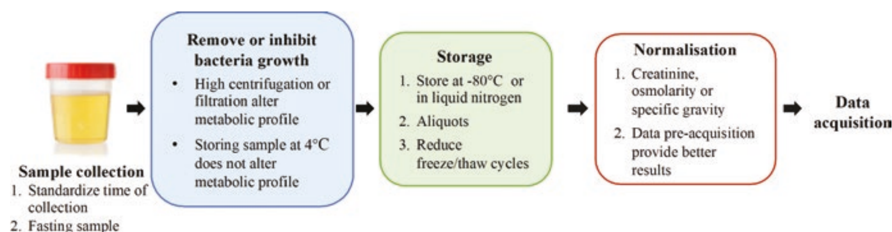

Fig. 3 Pre-analytical steps to consider for urine sample

2.1 Blood Samples

Whole blood, plasma, or serum is usually prepared for metabolomic studies. When plasma samples are collected, anticoagulants, such as ethylenediaminetetraacetic acid (EDTA), sodium citrate, and heparin, are added to inhibit the coagulation cascade [7, 14]. EDTA and sodium citrate are chelators of calcium which acts as a cofactor in the cascade, while heparin activates antithrombin [14]. The choice of one anticoagulant over another does not appear to significantly affect the levels of metabolites [13]. Nevertheless, citrate and EDTA can be detected by mass spectrometry and interfere with the analysis of metabolites of interest. Citrate can also alter the sample pH and reduce the yield of metabolite extraction. Unlike the other two anticoagulants, heparin is not detectable with MS, but it is often added as lithium or sodium salts that can form adducts during electrospray ionization [7].

Special precautions must be taken when collecting blood samples to avoid hemolysis. Erythrocytes can release intracellular components, including hemoglobin, and alter the metabolic profile of plasma and serum [7]. Other pre-analytical factors influencing metabolite levels include sample collection time, delays, and ambient temperature [7, 13].

Samples are often stored for a long period of time in an epidemiological study. However, storage for over 5 years at −80 °C can significantly change metabolite concentrations. It is preferable to store samples in liquid nitrogen as aliquots to avoid repeated freeze-thaw cycles that can alter the metabolic profile [7, 13].

2.2 Urine Samples

Urine differs from other biofluids in that it is easily collected in large quantities [7, 8]. However, collection at different times of the day and with or without fasting may interfere with the urinary metabolic profile [7, 13]. Metabolites from the diet are also more likely to be detected in urine than in other biofluids [15]. Therefore, it is preferable to standardize the time of collection and obtain samples after fasting [13].

Bacteria can also alter urinary metabolite levels in samples, which grow rapidly even after samples are collected [7]. Centrifugation and filtration can remove bacteria and cellular debris. However, high-speed pre-centrifugation can break the cells and cause the release of cellular components, which can alter the metabolic profile. Mild centrifugation at 1000–3000 g is preferred to remove contaminants without damaging cells. Filtration can sometimes also lead to a loss of metabolites [16]. Storing samples at 4 °C or lower immediately after sample collection inhibits bacterial growth [7]. In the long term, it is preferable to store samples in liquid nitrogen or at −80 °C and in smaller aliquots to limit freeze/thaw cycles and reduce the potential for metabolite degradation [7, 13].

The concentration of metabolites in urine can also vary significantly between samples depending on hydration status and other variables [17]. Creatinine concentrations or osmolarity can be used to normalize urine samples. However, creatinine

levels can vary depending on different factors, including age, gender, ethnicity, diet, muscle mass, exercise level, time of day, and disease states. Osmolarity would provide better separation of biological groups compared with creatinine but is less available in practice. Specific gravity, generally used as an alternative to osmolarity, corresponds to the ratio between the density of urine and water measured at constant temperature by refractometry [17]. Pre-acquisition normalization has been shown to provide better results than post-acquisition normalization, with a combination of both being even better [13], for instance, normalizing urine dilution steps during sample preparation based on creatinine concentration followed by a second normalization of raw data using either endogenous signals in the urine or average ion intensity for the chromatographic run.

3 Quality Assurance and Quality Control

Like many aspects of metabolomics, quality standards are not well defined. Poor data quality management leads to biased results, wasted resources, and can damage the credibility of this field of study [18]. Quality assurance (QA) establishes the prerequisites (equipment maintenance, personnel training, etc.) to ensure the quality and reproducibility of results, while quality control (QC) is important to make sure that no bias occurs during sample preparation and analysis [18–20]. The QC procedure includes using several control samples, including blanks, pooled QC or intra-study QC samples, long-term reference (LTR) or intra-laboratory QC samples, and standard reference material (SRM)/interlaboratory QC samples. Internal standards, technical replicates, and random sample analysis are also often used [18]. However, each laboratory uses different methods, which makes it difficult to harmonize results [18, 20]. Long et al. developed a list of recommendations for good laboratory practices (QA and QC) to ensure quality results [19]. They are classified into five steps (pre-pre-analytical, pre-analytical, analytical, post-analytical, and post-post-analytical), and each is accompanied by commonly made errors.

During collection, the most common errors are due to the misidentification of samples or improper collection, resulting in sample hemolysis [12]. Centrifugation conditions (time and speed), handling (sorting, pipetting/aliquoting), storage, and inappropriate transport of samples are also errors often observed in the medical environment and can have an impact on the metabolic profile [11, 12, 19]. Although less frequent, analytical errors can also occur [12]. They are caused by equipment malfunction, interference from an endogenous or exogenous compound, or a failure not detected by quality control. Therefore, it is important to perform regular maintenance on the equipment and use various QC samples to ensure the quality of the data generated [19]. Following data acquisition, there may also be erroneous validation, incorrect data entry, or misinterpretation of results [19]. Thus, errors can occur at different stages of the analysis process. However, many of them are human errors, and it is important to prioritize staff training to avoid them.

4 Identification of Metabolites

4.1 Confidence Levels

Confidence levels ensure the quality of metabolite identification. They classify from 0 to 4, with 0 being the highest confidence level. This level requires determining the three-dimensioanl structure and the complete stereochemistry of the metabolites of interest. Level 1 uses a reference standard to identify metabolites from two orthogonal techniques, such as MS/MS spectrum and retention time. Level 2 allows a possible identification of metabolites by comparing information from two orthogonal techniques with the literature or a database. Level 3 determines a possible class of metabolites based on at least one technique. Finally, level 4 does not allow the identification, but the compound remains present in the sample after extraction and can be quantified [19, 21].

To ensure the identification of metabolites present in samples, it is always preferable to use reference standards (level 1) [5, 21]. However, obtaining a complete set of metabolite standards can be an arduous and expensive task. Initiatives for creating metabolite standard mixes are an interesting concept to reduce the cost of individual standards. A comparison of MS/MS spectra and retention times to those present in databases allows, at best, a level 2 confidence [22].

4.2 Databases for Metabolite Identification

Several databases, such as Metlin, HMDB, MassBank, and NIST, are available to support metabolite identification. These can contain endogenous metabolites, exogenous compounds, and transformation products from food, microbiome, drugs, plants, and pollutants [23]. Nevertheless, metabolite identification remains complex, and it is estimated that less than 25% of MS/MS spectra can be identified due to a low number of metabolites listed in these databases and the limited number of commercially available pure reference standards [5, 23].

Some isomeric metabolites have a distinct structure but identical molecular formulae and, therefore, the same exact mass. The presence of these isomers can lead to false discovery rates (FDRs) due to the similarity between many MS/MS spectra present in databases [5]. A low FDR allows reliable annotation of a small number of metabolites, while a high FDR allows annotation of a larger number of metabolites whose quality may be poor [24]. Thus, FDRs are generally higher in large databases because they have many MS/MS spectra for the same compound. Each device provides a different spectrum, and the diversity of the methods used leads to many redundancies [5]. It is generally accepted that the best scenario would be to compare spectra obtained under the exact conditions as the metabolomic data was acquired. Different platforms can show variable fragmentation, especially if the collision

energies, for instance, are not the same. Another important concept would be the difference between instruments having varying mass resolutions. A low-resolution system, such as a triple quadrupole, allows only unit resolution to be achieved; therefore, molecular formulae of fragment ions detected cannot be confirmed. MassBank and NIST contain MS/MS spectra from different instruments (QqTOF, Orbitrap, QqQ, ion-trap) and different ionization techniques (ESI, EI, CI, APCI, and MALDI) [23]. HMDB also contains spectra from different devices but uses only electrospray ionization (ESI) [25]. Metlin is a database whose spectra are acquired on the electrospray-quadrupole-time-of-flight (QqTOF) platform from the Agilent Technologies. The spectra have been generated in positive and negative modes at three distinct collision energies (10, 20, and 40 V) [23, 26].

4.3 Data Processing Tools

Data processing is an essential and usually very time-consuming step in untargeted metabolomics. For this purpose, many tools are available, both commercially or freely online, using script platforms (such as R and python) [24, 27]. XCMS, MZmine, and MS-DIAL are commonly used in metabolomics. They allow filtering, feature detection, alignment, annotation, and identification of metabolites [28–30]. XCMS and MZmine also include statistical tests, while only XCMS allows the analysis of metabolic pathways and data integration with proteomics and genomics [22, 28, 29]. Other metabolomic analysis tools are available and have been discussed in various reviews [21, 22, 24]. However, all these softwares use different algorithms, resulting in a lack of reproducibility and consistency. Studies have shown differences in the number of features detected and in statistical analysis between tools, resulting in the identification of false-positive biomarkers [31–34]. Hohrenk et al. [33] tested four different softwares (MZmine2, enviMass, Compound Discoverer, and XCMS online) and found that only 10% of identified features were common.

5 Analytical Challenges

5.1 Variation of Metabolites

Metabolite variability is an important issue in epidemiological studies. It can arise from inter- and intra-subject variability or lack of technical reproducibility [4]. Metabolites can vary according to different factors, including age, gender, BMI, pregnancy, diet, ethnicity, smoking, environment, and physical activity [15, 35, 36]. Therefore, obtaining a large cohort that could correct all these criteria is difficult. Several authors report having an insufficient number of participants in the limitations of their study, despite a very large number of participants [37–42].

5.2 Missing Values

Missing values are a key factor to consider when analyzing the results. They result from metabolite signals in some samples and are absent in others. This may be due to biological variation between two groups of samples (control vs. disease) or external factors, such as pre-analytical conditions or chemical instability [4]. Missing values from technical factors are divided into three classes, missing not at random (MNAR) values, missing at random (MAR) values, and missing completely at random (MCAR) values, also called abundance-dependent missing values [43, 44]. MCAR values result from random errors acquired during data acquisition, such as reduced ionization efficiency or ion suppression [43]. MAR values are more general and are difficult to distinguish from MCAR, as they are often combined [44]. They may result from poor peak detection or deconvolution of two co-eluting peaks [45]. Finally, the MNAR values correspond to metabolites with concentrations below the limit of detection [43].

In epidemiological studies, missing values are often replaced by half or a fraction of the lowest detected value [4]. However, the different types of missing values may influence the choice of approach to adopt. Random forest is considered the best approach for MCAR/MAR, while quantile regression imputation of left-censored data (QRILC) is preferred for MNAR [45]. When the missing value is from a biological factor, a change may distort the results; therefore, it would be better, in this case, to keep these values as null values [4]. Therefore, the choice of method considers the authors' opinion on the nature of the missing values, which can be verified by comparing the different results obtained using these methods [46].

Following the missing data processing, the metabolites usually undergo a log transformation to normalize the results [4]. This method allows us to obtain a Gaussian-like distribution in the metabolite variables and to compare the statistical analyses [46]. However, this approach cannot be applied to all metabolites, and a comparison of the distribution of values must be performed before and after the transformation. When there is no change, it is better to keep the initial values [46].

6 Statistical Analysis

Several statistical analyses are used in epidemiological studies. Nevertheless, some tests seem to be used more than others. According to a survey of laboratories participating in the COnsortium of METabolomics Studies (COMETS), univariate regression is mostly used in metabolomic analysis [4]. Specifically, it is the linear regression that associates a metabolite with an outcome. It is combined with multiple correction tests, such as Benjamini-Hochberg, to determine the false discovery rate or Bonferroni correction to control the false positive or type 1 error [4, 46]. However, the Bonferroni correction is a conservative approach and can increase the risk of false negatives or type 2 error [47]. Multivariable analysis of metabolites

with covariates and principal component analysis (PCA) are also widely used [4]. Unlike linear regression, PCA does not identify a specific biomarker. However, it allows the correlation between a set of metabolites and an outcome [46].

The choice of statistical tests will depend on the objective of the study. Many researchers use the area under the Receiver Operating Characteristic Curve (ROC) to determine the quality of a biomarker, while partial correlation is the most common method for analyzing intercorrelations between different metabolites. Laboratories that have performed network analysis [4] used the weighted gene co-expression network analysis (WGCNA) and MetaboAnalyst, for a freely available and user-friendly data analysis online platform. Recently, a recent version of MetaboAnalyst (5.0) has been released online and provides better support for statistical analysis (univariate and multivariate) and functional analysis (enrichment, pathway, and functional meta-analysis) [48, 49].

7 Selected Case Studies

Since Oliver et al. [50] introduced the metabolome concept in 1998, metabolomic studies related to disease and new biomarkers have continued to increase [5]. Cancers (breast, colon, lung) [39, 51–55], autism spectrum disorder [56], hyperinsulinemia [38], eating disorders [41], chronic kidney disease [57], depression [58], osteoporosis [59], metabolic syndrome [60], ovarian reserve dysfunction [61], and obesity [62] are few examples of diseases that are the focus of population-based metabolomic studies. They perform metabolomic profiling to predict disease incidence or progression. Kelly et al. [56] showed that plasma metabolite levels could predict autism in children aged 8 years with good sensitivity and specificity. Tryptophan and tyrosine metabolic pathways would be associated with better ages and stages questionnaire (ASQ) communication scores. Tabung et al. [38] evaluated the correlation between plasma metabolites and the empirical dietary index of hyperinsulinemia (EDIH). They showed increased levels of diacylglycerol, triacylglycerols, C10:2 carnitine, and C18:2 sphingomyelins and decreased phospholipid levels of trigonelline, and eicosapentaenoate would correlate with elevated EDIH [38]. Another study performed metabolic profiling of plasma in 7-year-old children. It showed that high levels of VLDL, triglycerides, apolipoprotein-B/apolipoprotein-A, and monounsaturated fatty acids ratio could decrease the probability of developing anorexia nervosa at the age of 18, while high levels of HDL, docosahexaenoic acid, polyunsaturated fatty acid ratio, and fatty acid unsaturation could increase this risk [41].

Other population-based studies focus more on factors leading to metabolite variability in different biological matrices to characterize them better. Wang et al. performed metabolic profiling in plasma [63] and urine [64] to identify diet-related biomarkers. They associated 238 plasma metabolites with 74 food groups and 513 urinary metabolites with 79 food groups. Darst et al. [65] instead studied the effect

of age and sex on the metabolic profile of plasma. They found that 623 metabolites were associated with age, of which 29 steroids decreased with age while levels of most fatty acids and sphingolipids increased. Furthermore, 695 metabolites were associated with sex, 55% of which decreased in women (mainly steroids and amino acids).

Nevertheless, for each study, information regarding the pre-analytical procedures (collection, preparation, or storage sample) and methods to ensure quality control or data processing is often lacking. This increases the difficulty of comparing two studies with each other. Thus, the experimental approach used in some metabolomic-related population studies is summarized in Tables 1, 2 and 3.

7.1 Pre-Analytical Procedures

For sample collection, it was common for patients to be fasting (2 h overnight). Plasma samples were collected in tubes containing EDTA. Only the study by Larkin et al. [53] used lithium heparin, whereas three studies did not specify the anticoagulant used. Several studies did not describe the steps involved in the preparation of plasma and serum samples. In general, centrifugation parameters, time delays, or any other information that could influence the levels of metabolites when obtaining either of these two matrices were not specified.

Conversely, some studies shared not only all of this information but also the average total time required for sample preparation [40, 52, 65, 66], an important factor for the variability of blood metabolites. The studies did not use pre-centrifugation and filtration for urine samples to remove bacteria. Some of them kept the samples at 4 °C and at a lower temperature, or in some cases, it is simply not defined. Two of five studies normalized metabolite concentrations to osmolality. Only one performed a creatinine normalization, which was not indicated in the other two studies. For all biofluids, samples were generally aliquoted before being stored. In the long term, samples were stored at −80 °C to avoid changes in metabolite levels. Some studies opted for storage in liquid nitrogen or at −70 °C.

7.2 Data Acquisition

Liquid chromatography-mass spectrometry (LC-MS) is the most used analytical method. Nuclear magnetic resonance (NMR) is also widely used, while only three studies used gas chromatography-mass spectrometry (GC-MS). To ensure the quality of the results, the most common method was to use a pool of samples injected several times. The coefficient of variation (CV) was then determined from this pool. Duplicates or triplicates were also used to assess inter- or intra-batch variability. Lau et al. [66] used several techniques to ensure the quality of their results. Internal

Table 1 Cancer-related metabolomic population studies

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analyses	Outcomes of study	Reference
Breast cancer	Serum	174 women (aged 35–64)	– Fasting blood sample – Aliquoted – Stored in liquid N2	NMR	– MetaboAnalyst – Missing values half of the LOQ – > 20% under LOQ excluded	– Log-transformation – Wilcoxon rank-sum tests – Frequencies – Pearson's chi-squared tests – Conditional logistic regression models – Benjamini-Hochberg – PCA – OPLS-DA	– Alanine, leucine, tyrosine, valine, lactic acid, pyruvic acid, triglycerides, lipid main fraction, and 11 VLDL lipid subfractions were inversely associated with high MBD BC cases – Acetic acid was directly associated with high MBD cases – Phenylalanine, tyrosine, and tryptophan pathway emerged MBD BC cases	[51]

Breast cancer	Serum	62 Asian females	Collection – Incubated at r.t. for 30 min – Centrifuged at 3000 r/min, 5 min – Stored at −80 °C, Metabolite extraction – Protein precipitation by MeOH – Centrifuged at 12000 rpm, 15 min at 4 °C – Samples kept at 4 °C throughout analysis	UHPLC-QTOF MS – Leucine encephalin used as lock mass for internal mass calibration – Pooled QC sample injected every eight samples	– Agilent Mass Profiler software – In-house library, METLIN and HMDB MetScape – Correlation-based metabolic networking analysis – MetaboAnalyst 4.0 for metabolic pathway analysis (KEGG)	– Log-transformation – Autoscaling SIMCA-P – PCA – OPLS-DA MetaboAnalyst 4.0 – Univariate statistical analysis MetScape – Pearson's correlation coefficient – Debiased squared partial correlation SPSS – ROC curve	– All altered amino acids were upregulated, while all cardiolipin (CL) species are downregulated in the TNBC samples – Glycerophospholipid metabolism, aminoacyl-tRNA biosynthesis, and valine, leucine, and isoleucine biosynthesis were significantly altered – dUMP, L-octanoylcarnitine, L-proline, lysoPC (22:1), PS (22:0/0:0), and uric acid correlated with a five-year survival rate in TNBC patients	[54]

(continued)

Table 1 (continued)

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analyses	Outcomes of study	Reference
Cancer	Plasma	293 participants (age $\geq$ 40)	– Fasting blood (2 h) – Lithium-heparin tubes – Left to stand at room temperature – Centrifugation at 2200 × g, 10 min – Stored at −80 ° C – Defrosted on ice – Mixed with NMR buffer – Centrifugation at 16,000 × g, 3 min	NMR – Pooled samples	– Assignments by reference to literature values – HMDB	R package – OPLS-DA – Two-sided Kolmogorov-Smirnov test – ROC curves – Pearson's correlation	– OPLS-DA models separated unwell patients with solid tumor diagnoses from those with noncancer diagnoses with an AUC of 0.91, sensitivity of 94%, and a specificity of 82% – OPLS-DA models also separated patients with a metastatic cancer from those with nonmetastatic cancer with an AUC of 0.91, sensitivity of 94%, and specificity of 88%	[53]

Colorectal cancer (cachexia)	Urine	52 participants	– Aliquoted – Stored at −80 °C	– Creatinine normalization – IS GC-MS – Pooled QC samples (every batch) – Kovats retention mixtures (every batch) – LOESS function – %RSD – Analyzed paired samples in same analytical batches NMR – QC samples were run at the beginning and end of each analytical batch – Citrate was measured in duplets	– MetaboAnalyst software 4.0 GC-MS – MZMine 2.0 – HMDB and PubChem – Authentic reference standard NMR – Bruker TopSpin software with zero filling – DataChord spectrum miner software – Chenomx library/HMDB – Authentic reference standard	– Log-transformation – Autoscaled – Benjamini-Hochberg – One-way ANOVA – Pearson's chi-squared test – Pearson's partial correlation coefficients – OPLS-DA	– Levels of acetone, arginine, 2.3.-butanediol, and 2.3.-dihydroxybutyrate decreased in cachectic compared to non-cachectic patients – Glycerol phosphate shuttle metabolism and glycine, ketone body metabolism, and serine metabolism were the top 3 pathways	[39]

(continued)

Table 1 (continued)

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analyses	Outcomes of study	Reference
Lung cancer	Urine	564 never-smoking females (aged 52–66)	– Sterilized cup containing ascorbic acid (antioxidant) – Kept at 0 °C to −4 °C – Processed within 6 h – Storage at −70 °C	LC-MS and NMR – Pooled QC samples (every five patient samples)	LC-MS – Progenesis (nonlinear dynamics) – XCMS software – CV >30%, features were excluded – In-house databases, HMDB and METLIN NMR – MATLAB R2012b – Normalized by probabilistic quotient Missing data – Metabolites were excluded from all analyses Mummichog analysis – Pathway enrichment	R package, SAS, and SIMCA-P+ software – Log- transformation – PCA – OPLS-DA – Wilcoxon rank sum test – Fisher exact test – Separate unconditional logistic regression models – Multivariable linear regression model – Mediation analysis – Spearman correlation coefficients – Benjamini-Hochberg	– 5-Methyl-2-furoic acid was significantly associated with lower lung cancer risk – 25% of the association between soy consumption and higher lung cancer risk was significantly mediated via 5- methyl-2-furoic acid – 1-carbon metabolism, nucleotide metabolism, oxidative stress, and inflammation were associated with lung cancer risk	[55]

Table 2 Population metabolomic studies related to other diseases

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analysis	Outcomes of study	Reference
Autism spectrum disorder	Plasma	806 children (ages 1 and 3 years)	– Separated – Stored at – 80 °C	LC-MS – Two batches sent 6 months apart – Normalized to sample mass	– In-house library Missing value – 50% or more metabolites were excluded	R package – Log-transformation – PCA – PLS-DA – Bonferroni correction – ROC	– Tryptophan and tyrosine metabolism were associated with better communication score – Plasma metabolite level provides a high sensitivity (88.9%) and specificity (84,5%) for predicting autism	[56]
	Stool	806 children (age 3 years)	– Stored at −80 °C – Precipitation with MeOH – Five aliquots	LC-MS – One batch – Normalized to sample mass	– In-house library Missing value – S/N < 10 or with missing levels >10% were excluded			
Hyperinsulinemia	Plasma	1919 women (aged 50–79)	– Fasting blood sample (12 h) – EDTA tube – Stored at −70 °C – Shipment on dry ice	LC-MS – Pooled plasma reference samples – Running every 20 samples	– Authentic reference standards or reference samples – Coefficients of variation (CVs) Missing value – Signal-to-noise ratio < 10 – Half the lowest observed value	– Multivariable-adjusted linear regression models	– Mainly phospholipids as well as trigonelline and eicosapentaenoate decreased with increasing EDIH scores – Mainly diacylglycerol and triacylglycerols as well as C10:2 carnitine and C18:2 sphingomyelins increased with EDIH scores	[38]

(continued)

Table 2 (continued)

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analysis	Outcomes of study	Reference
Eating disorders	Plasma	2929 children (7 years old)	– Non-fasting – EDTA tube	NMR – Sample standard deviation (SD)		R package – Logistic regression	– Elevated VLDL, triglycerides, Apo-B/A, and monounsaturated fatty acids ratio at age 7 were associated with lower odds of anorexia nervosa at age 18 – Elevated HDL, docosahexaenoic acid, and fatty acid unsaturation at age 7 were associated with higher risk for anorexia nervosa at 18 years	[41]
Chronic kidney disease (CKD)	Serum or plasma	454 participants (mean age 68 ± 12)	Trained personal under strict quality control Collection – Fasting blood (overnight) – Prepared – Stored in liquid N_2 – Serum samples thawed on ice Extraction plasma metabolites (diabetes) – Solvent mixture – Spiked with IS – Centrifuged at 15.800 × g, 15 min at 4°C	GC-MS – Three technical replicates – Blank: Deionized water – QC (serum pool) samples everyday	– Agilent Fiehn GC/MS metabolomics RTL library – NIST library 11 – Agilent MassHunter workstation quantitative analysis Agilent's MassHunter software – IS d27-myristic acid using the RTL system MetaboAnalyst – Pathway analysis – KEGG homo sapiens library Missing value >50% below LOD were excluded	R package and SPSS – Cox regression models – Multivariable models – Unadjusted and adjusted fine and gray models – ROC curves – Partial correlations MetaboAnalyst Hypergeometric test adjusted	– D-malic acid, acetohydroxamic acid, and butanoic acid were independently associated with death – Docosahexaenoic acid was inversely associated to mortality – Lactose and 2-O-glycerol-α-D-galactopyranoside were associated to end stage renal disease (ESRD) risk – Tyrosine showing an inverse relationship with ESRD	[57]

Depression	Plasma	10,145 control 5283 depressed persons	– Fasting blood (overnight) – EDTA tubes Centrifugation and stored at −80 °C or Centrifugation at 4 °C and stored at −20 °C	NMR – Sample handling to data processing is highly standardized and fully automated – Unaware of depression cases vs. control status – Bayesian modeling	Missing values: Metabolite values in subjects with outlying concentrations (6 5 SD) were additionally set as missing	R package – Log transformation – Random effects meta-analyses – I^2 with 95% confidence intervals – Wald tests – FDR method	– Higher levels of Apo B, VLDL, triglycerides, diglycerides, monounsaturated fatty acids, fatty acid chain length, glycoprotein acetyls, tyrosine, and isoleucine were associated with increased odds of depression – Lower levels of HDL, acetate, and Apo A1 were associated with increased odds of depression	[58]
Osteoporosis and bone mineral density (BMD)	Serum	320 participants	– Fasting blood – EDTA – Centrifugation – Frozen at −80 °C – AbsoluteIDQ p180 kit	LC-MS	– Biocrates proprietary MetIQ TM software – Semi- quantitation was applied to the lipid quantitation based on isotopic lipid IS Missing value – >20% below LOD were excluded	SAS – Log-transformation – PCA – PLS-DA – Wilcoxon test – $\chi 2$ test, – Random forest classifier method – ROC curves	– In males, glutarylcarnitine, hydroxy sphingomyelin C16:1, sphingomyelin C18:0, lysine, and serine were associated with osteoporosis – In postmenopausal females, acetylcarnitine, phosphatidylcholine diacyl, phosphatidylcholine acyl-alkyl, and hydroxyproline were associated with osteoporosis – ROC curve for BTMs and serum metabolites increased significantly compared to BTMs only	[59]

(continued)

Table 2 (continued)

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analysis	Outcomes of study	Reference
Metabolic syndrome	Plasma	115 participants	– Fasting plasma – EDTA Metabolite extraction – H2O: MeOH (1:9) – Incubated for 2 min – Stored on ice for 2 h – Centrifuged at 18620 RCF, 10 min at 4 °C Lipid extraction – $CHCl_3$: MeOH (2:1) – Stored at room temperature for 60 min – Centrifuged at 18620 RCF, 3 min at 4 °C	LC-MS/MS Reference interface – Purine m/z 121.05 and m/z 119.03632 – HP-0921 m/z 922.0098 and m/z 966.000725 – Positive and negative modes, respectively GC-MS/MS	LC-MS – Profinder™ software package – In-house libraries – Metabolite and lipid signals were normalized to the total peak area GC-MS – MATLAB – In-house library – Normalization to IS	SIMCA software – PCA – OPLS – 95% confidence interval using the jackknife method – CV-ANOVA – Permutation analyses R package – Univariate linear regression – Log-transformation	– In the univariate analyses, the metabolic syndrome (score) was associated with multiple individual metabolites, including valeryl carnitine, pyruvic acid, lactic acid, alanine, and lipids (diglyceride)	[60]
Functional ovarian reserve	Serum	398 women (ages 18–45)	– Non-fasted blood – Centrifuged – Frozen at −80 °C – Assays within 1 year of storage – No freeze/thaw cycles	NMR	– Quantification is achieved through three molecular windows from each sample – PERCH NMR software – Serum extract metabolites are scaled via the total cholesterol	R package – Multivariable linear regression – Scaled to standard deviation (SD) units – Bonferroni	– Anti-Müllerian hormone (AMH) showed positive associations with HDL, omega-6, and polyunsaturated fatty acids and the amino acids isoleucine, leucine, and tyrosine and negatively with acetate – Antral follicle count (AFC) was positively associated with alanine, glutamine, and glycine	[61]

Obesity	Plasma	1020 children (ages 5.5, 8, and 10)	CHOP study – Fasted blood – Centrifugation – Frozen at −70 °C UBCS – Fasted blood (>10 h) – Aliquoted – Frozen at −80 °C GINIplus/LISA study – Aliquoted – Frozen at −80 °C For all, sent on dry ice and re-stored at −80 °C Metabolite extraction – Protein precipitation with MeOH, containing IS – Centrifuged at 4000 rpm, 10 min at room temperature	LC-MS CHOP study and GINIplus/LISA study – 6 QC samples per batch – Intra-batch CV < 0.2, the batch was included – QC > 50% of the batches, metabolite was included UBCS – Six QC samples measured twice – QC > 35% CV, metabolites were excluded	– Quantification of SM 32:2 was achieved by comparison to commercially available standard SM (d18:1/18:0) – Metabolite were excluded if >1.5 x SD from second highest value	R package – Linear models regressing – Bivariate linear models – Multiple linear models – Partial R2 – Cochran's Q31 – Bonferroni	– Only SM 32:2 was significantly associated with BMI z-score in all four populations – Alanine showed the strongest positive association with HOMA, while acylcarnitines and nonesterified fatty acids were negatively associated with HOMA – SM d18:2/14:0 is a powerful marker for molecular changes in childhood obesity	[62]

(continued)

Table 3 Population metabolomic studies related to age, sex, BMI, pregnancy, and lifestyle habits

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analysis	Outcomes of study	Reference
Age and sex	Plasma	1212 participants (mean age 61)	Plasma metabolomics – Fasting blood in EDTA tubes – Centrifuged at 3000 rpm, 15 min at RT – Aliquoted and frozen −80 °C, 30 min – Shipped on dry ice – Stored at −80 °C Metabolite extraction – Recovery standards – Protein precipitation with MeOH – Centrifugation – Stored under N_2 Genomics – DNA was extracted using PUREGENE® DNA isolation kit	LC-MS/MS – Pooled matrix sample – Blank: Extracted water samples – Cocktail of QC – Instrument performance monitoring – Aided chromatographic alignment – Median relative standard deviation (RSD) Study samples were randomized QC samples spaced evenly among the injections	– Metabolon's hardware and software – In-house library – Identified by future acquisition of a matching purified standard or by classical structural analysis – Metabolites with interquartile range of zero excluded from analyses Missing metabolite values – Imputed to the lowest level of detection for each metabolite	Metabolomics – Median-scaled – Log-transformation – Pearson r SAS – Linear mixed effects regression models – Benjamini-Hochberg Genomics – PCA – Pearson r GCTA h2 of each metabolite	– 623 were associated with age; 29 steroid lipids significantly decreased while higher levels of most fatty acid, sphingolipids and amino acids increased with age – 695 were associated with sex; most steroid lipids and amino acids were in lower levels while most fatty acids were higher in women – The heritabilities of metabolites ranged dramatically (0.2–99.2%)	[65]

Sex-specific associations of BMI and body fat	Urine	369 participants (aged 16–18)	– Stored at −80 °C – Protein precipitation by MeOH	LC-MS/MS – Pools of sample were used as quality controls	– In-house library – Normalization with block correction – Normalization by urine osmolality Missing value – >20% below LOD were excluded – Random forest method built into “mice” package	SAS software and R software – Log-transformation – ICA – Linear regression model – enjamini-Hochberg	– Ten metabolites associated with both BMI and BF – Eleven metabolites associated with only BF and nine with only BMI – None of these associations was in females – Strong sexual dimorphism in the relationship between body composition and the urine metabolome	[67]

(continued)

Table 3 (continued)

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analysis	Outcomes of study	Reference
Age, sex, BMI and dietary habits	Serum	1192 children (aged 6–11)	– Collection in silica plastic tubes – Inverted gently for 6–7 times – Spun at 2500 g, 15 min at 4 °C – Median processing time (collection to freezing): 1.8 h – Median time between last meal and collection: 3.3 h	LC-MS/MS – AbsoluteIDQ p180 kit with IS – Blank PBS samples (three replicates) – Randomized samples – NIST SRM 1950 plasma reference material (4 replicates) – QC material (two replicates, SeraLab, S-123-M-27485) – QCs from manufacturer in three concentrations CVs: NIST SRM 1950	LC-MS/MS (serum) – MetIDQ™ software Missing value: – CV > 30% > 30% below LOD	R package – Log-transformation – Metabolome-wide association study (MWAS) – Multiple linear regression models – Bonferroni correction – Partial R2 approach – PCA – Pearson's correlation coefficients – Cytoscape software and the MetScape plugin application	Sex: – Urinary isoleucine was at lower while 5-oxoproline and tyrosine were higher in males – Neurotransmitter serotonin in serum was higher in males while serine, lysine, ornithine, acylcarnitines, and three sphingolipids were higher in females Age: – Creatinine BMI z-score: – Urinary 4-deoxyerythronic acid, urinary valine, serum carnitine, acylcarnitines, glutamate, BCAA valine, lysoPC, and sphingolipids Dietary metabolite: – Urinary creatine and serum PC with meat – Serum PC with fish – Urinary hippurate with vegetables and urinary proline betaine – Hippurate with fruit Population-specific variance was better captured in the serum than in the urine profile	[66]
	Urine		– Collection: Evening and morning – Kept in fridge overnight – Transported in temperature-controlled environment – Aliquoted and frozen within 3 h – Centrifuged at 13,000 g, 10 min at 4 °C	NMR – Randomised – Pooled urine samples from 20 individuals CVs: Pooled QC	– MATLAB – Probabilistic quotient normalisation – HMDB – ChenomxNMRsuite 7.1 profiler Missing value: – CV > 30% > 30% below LOD			

BMI	Serum	304 postmenopausal women (aged 50–74)	– Fasting blood (≥10 h) – Processed – Stored at −86 °C within 12 h – Several recovery standards – Protein precipitation by MeOH	LC-MS/MS – Pooled QC standard – CVs calculated for 38 samples using single pooled QC sample	– In-house library – Normalization to the run day Missing value – >90% below LOD were excluded – Minimum observed value of that specific metabolite	– Log-transformation SAS – Partial Pearson correlations Bonferroni correction	– 50 BMI-correlated metabolites – BMI-related metabolites were more strongly correlated with fat mass than lean mass – Eight metabolites that were correlated with fat mass or lean mass but not BMI	[68]
Diet	Urine	648 participants (mean age 52.2 ± 9.4)	– 24 h urine collections – Refrigerate or with cooler packs – Aliquoted – Frozen and shipped on dry ice – Storage in liquid N_2 – Precipitate proteins with MeOH	LC-MS/MS – Triplicates of 44 participant samples – Interclass correlation coefficients (ICCs)	– In-house library – Normalization by osmolality Missing value: – ICC < 0.5 were excluded – >90% below LOD were excluded – Imputed with the minimum	– Log-transformation – Autoscaled – Pearson's partial correlation – Bonferroni – ROC curve using R package	– 708 diet-metabolite associations were identified – 513 unique metabolites correlated with 79 food groups/items	[64]
Diet	Plasma	671 participants (mean age 52.3 ± 9.5)	– Fasting blood (8 h) – EDTA – Refrigerated – Centrifugation – Aliquoted – Shipped on dry ice – Storage in liquid N_2 – Protein precipitation by MeOH	LC-MS/MS – Duplicates of 60 participant samples – Interclass correlation coefficients (ICCs)	– In-house library – Each metabolite was divided by its daily median – Excluded – Missing values were assigned the minimum detection value	– Log-transformation – Autoscaled – Pearson's partial correlation – Bonferroni – ROC curve using R package	– A total of 677 diet-metabolite associations – 238 plasma metabolites were associated with 74 food groups/items	[63]

(continued)

Table 3 (continued)

Subject	Sample	Cohort	Data pre-acquisition	Data acquisition	Data post-acquisition	Statistical analysis	Outcomes of study	Reference
Pregnancy	Serum	8774 women	– Fasting serum – Processed within 2.5 h – Centrifuged at 3500 rpm, 10 min at room temperature – Stored at −80 °C	NMR	– PERCH NMR software – Repeated analyses including only complete case data to test whether any missing data were altering the results	R package – Multivariable linear regression	– White European women (WEs) had higher levels of most lipoprotein, cholesterol, glycerides, phospholipids, monosaturated fatty acids, pyruvate, glycerol, and creatinine – South Asian women had higher levels of glucose, linoleic acid, omega-6 and polyunsaturated fatty acids, and most amino acids – Higher BMI and having GD had higher levels of lipoprotein, triglycerides, mostly with stronger associations in WEs	[40]

Breast cancer preventive lifestyle behaviors	Plasma	1319 participants (aged 30–74)	– EDTA tubes – Stored at 4 °C – Centrifuged at 1300 g,10 min at 4 °C with brake on – Stored at −80 °C (19–24 h after blood collection) – Free of hemolysis and lipidemia	NMR	**Missing values:** – Half minimum value of that metabolic measure in total population – CV: Duplicate samples (5%)	R package and Stata – Log-transformation – Multivariate imputations by chained equation (MICE) – Linear regression – Pearson pairwise correlations MetScape – Conditional correlation networks	Fruit and vegetable consumption – Negative associations between glycoprotein acetyls, an inflammation-related metabolite with lower BMI – Positive association between polyunsaturated fatty acids BMI – Positive associations between HDL with lower BMI Alcohol consumption – Positive association of ApoA1 BMI and fruit and vegetable consumption were generally consistent but were opposite with alcohol consumption	[52]

standards (IS), blanks with PBS, different control samples (NIST SRM 1950, commercially available serum QCs and QCs provided by the manufacturer), and randomized batches were used for serum samples.

7.3 Post-Acquisition Data Processing

Missing values are due to the absence of metabolites in some samples. When a metabolite is not detected in several samples (20–90%), it is usually excluded from the analysis. In some cases, the missing value is replaced by a fraction of the lowest detected value or the minimum value. However, it is not always specified whether this is the limit of detection (LOD) or the value detected in a sample set. The study by McClain et al. [68] used the random forest method built into the "mice" package to handle missing data. Several studies use an in-house library to identify metabolites, which allows for reducing the FDR and obtaining MS/MS spectra specific to the method used. Online databases are also used. The most common was HMDB, followed by METLIN and NIST. Studies have combined in-house libraries and online databases to increase the number of identified metabolites. Some studies confirm with commercially available standards to achieve a level 1 confidence. Prior to statistical analyses, most studies perform a log transformation to obtain a Gaussian distribution. Subsequently, the most used statistical tests were the principal component analysis (PCA), the area under the receiver operating characteristic (ROC) curve, and the multivariate linear regression model. Other analyses, such as orthogonal partial least-squares discriminant analysis (OPLS-DA), Pearson correlation coefficient, and Pearson partial correlation coefficient, were also present in several studies. Many studies have performed multiple correction tests, with a slight preference for the Bonferroni correction. Thus, despite the diversity of methods used, some approaches seem to be more common than others.

8 Conclusion

Recently, population-based metabolomic studies have increased significantly due to their potential to predict, diagnose, and monitor disease progression [5, 69]. However, due to the lack of standardization of protocols, these studies use different approaches, which may lead to limited reproducibility and consistency between results. Therefore, it is essential to establish standards for a specific workflow for untargeted metabolic analysis. Furthermore, to get a complete view of the biological mechanisms involved, it is preferable to integrate metabolomics with other "omic" sciences, such as proteomics, transcriptomics, and exposomics.

References

1. Vailati-Riboni M, Palombo V, Loor JJ. What are omics sciences? In: Periparturient diseases of dairy cows; 2017. p. 1–7.
2. Aderemi AV, et al. Metabolomics: a scoping review of its role as a tool for disease biomarker discovery in selected non-communicable diseases. Meta. 2021;11(7):418.
3. Gika HG, et al. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. J Pharm Biomed Anal. 2014;87:12–25.
4. Playdon MC, et al. Metabolomics analytics workflow for epidemiological research: perspectives from the consortium of metabolomics studies (COMETS). Meta. 2019;9(7):145.
5. Cui L, Lu H, Lee YH. Challenges and emergent solutions for LC-MS/MS-based untargeted metabolomics in diseases. Mass Spectrom Rev. 2018;37(6):772–92.
6. Villas-Boas SG, et al. Mass spectrometry in metabolome analysis. Mass Spectrom Rev. 2005;24(5):613–46.
7. Kirwan JA, et al. 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. 2018;64(8):1158–82.
8. Bouatra S, et al. The human urine metabolome. PLoS One. 2013;8(9):e73076.
9. Psychogios N, et al. The human serum metabolome. PLoS One. 2011;6(2):e16957.
10. Chen Y, et al. Methods used to increase the comprehensive coverage of urinary and plasma metabolomes by MS. Bioanalysis. 2016;8(9):981–97.
11. Lippi G, et al. Preanalytical variability: the dark side of the moon in laboratory testing. Clin Chem Lab Med. 2006;44(4):358–65.
12. Szecsi PB, Odum L. Error tracking in a clinical biochemistry laboratory. Clin Chem Lab Med. 2009;47(10):1253–7.
13. Stevens VL, et al. Pre-analytical factors that affect metabolite stability in human urine, plasma, and serum: a review. Metabolites. 2019;9(8):156.
14. Denery JR, Nunes AA, Dickerson TJ. Characterization of differences between blood sample matrices in untargeted metabolomics. Anal Chem. 2011;83(3):1040–7.
15. Wu J, Gao Y. Physiological conditions can be reflected in human urine proteome and metabolome. Expert Rev Proteomics. 2015;12(6):623–36.
16. Bernini P, et al. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. J Biomol NMR. 2011;49(3–4):231–43.
17. Wu Y, Li L. Sample normalization methods in quantitative metabolomics. J Chromatogr A. 2016;1430:80–95.
18. Evans AM, et al. Dissemination and analysis of the quality assurance (QA) and quality control (QC) practices of LC-MS based untargeted metabolomics practitioners. Metabolomics. 2020;16(10):113.
19. Long NP, et al. Toward a standardized strategy of clinical metabolomics for the advancement of precision medicine. Meta. 2020;10(2):51.
20. Beger RD, et al. Towards quality assurance and quality control in untargeted metabolomics studies. Metabolomics. 2019;15(1):4.
21. Blazenovic I, et al. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Meta. 2018;8(2):31.
22. Spicer R, et al. Navigating freely-available software tools for metabolomics analysis. Metabolomics. 2017;13(9):106.
23. Vinaixa M, et al. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects. TrAC Trends Anal Chem. 2016;78:23–35.
24. Misra BB. New software tools, databases, and resources in metabolomics: updates from 2020. Metabolomics. 2021;17(5):49.
25. Wishart DS, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 2018;46(D1):D608–17.

26. Kind T, et al. Identification of small molecules using accurate mass MS/MS search. Mass Spectrom Rev. 2018;37(4):513–32.
27. Cajka T, Fiehn O. Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. Anal Chem. 2016;88(1):524–45.
28. Tautenhahn R, et al. XCMS online: a web-based platform to process untargeted metabolomic data. Anal Chem. 2012;84(11):5035–9.
29. Pluskal T, et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinform. 2010;11(1):1–11.
30. Tsugawa H, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods. 2015;12(6):523–6.
31. Gurdeniz G, et al. The effect of LC-MS data preprocessing methods on the selection of plasma biomarkers in fed vs. fasted rats. Metabolites. 2012;2(1):77–99.
32. Chen Y, et al. Assessment of data pre-processing methods for LC-MS/MS-based metabolomics of uterine cervix cancer. Analyst. 2013;138(9):2669–77.
33. Hohrenk LL, et al. Comparison of software tools for liquid chromatography-high-resolution mass spectrometry data processing in nontarget screening of environmental samples. Anal Chem. 2020;92(2):1898–907.
34. Coble JB, Fraga CG. Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. J Chromatogr A. 2014;1358:155–64.
35. Thevenot EA, et al. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. J Proteome Res. 2015;14(8):3322–35.
36. Lawton KA, et al. Analysis of the adult human plasma metabolome. Pharmacogenomics. 2008;9(4):383–97.
37. Perng W, et al. Urate and nonanoate mark the relationship between sugar-sweetened beverage intake and blood pressure in adolescent girls: a metabolomics analysis in the ELEMENT cohort. Meta. 2019;9(5):100.
38. Tabung FK, et al. Identifying metabolomic profiles of Insulinemic dietary patterns. Meta. 2019;9(6):120.
39. Ose J, et al. Multiplatform urinary metabolomics profiling to discriminate cachectic from non-cachectic colorectal cancer patients: pilot results from the ColoCare study. Meta. 2019;9(9):178.
40. Taylor K, et al. Differences in pregnancy metabolic profiles and their determinants between white European and South Asian women: findings from the born in Bradford cohort. Meta. 2019;9(9):190.
41. Santos Ferreira DL, et al. Associations between blood metabolic profile at 7 years old and eating disorders in adolescence: findings from the Avon longitudinal study of parents and children. Meta. 2019;9(9):191.
42. Ellul S, et al. Metabolomics: population epidemiology and concordance in Australian children aged 11-12 years and their parents. BMJ Open. 2019;9(Suppl 3):106–17.
43. Karpievitch YV, Dabney AR, Smith RD. Normalization and missing value imputation for label-free LC-MS analysis. BMC Bioinform. 2012;13(16):1–9.
44. Lazar C, et al. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. J Proteome Res. 2016;15(4):1116–25.
45. Wei R, et al. Missing value imputation approach for mass spectrometry-based metabolomics data. Sci Rep. 2018;8(1):663.
46. Antonelli J, et al. Statistical workflow for feature selection in human metabolomics data. Meta. 2019;9(7):143.
47. Andrade C. Multiple testing and protection against a type 1 (false positive) error using the Bonferroni and Hochberg corrections. Indian J Psychol Med. 2019;41(1):99–100.
48. Pang Z, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res. 2021;49(W1):W388–96.

49. Xia J, et al. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res. 2009;37(Web Server issue):W652–60.
50. Oliver SG, et al. Systematic functional analysis of the yeast genome. Trends Biotechnol. 1998;16(9):373–8.
51. Bendinelli B, et al. Prediagnostic circulating metabolites in female breast cancer cases with low and high mammographic breast density. Sci Rep. 2021;11(1):13025.
52. Qi J, et al. Metabolomics and cancer preventive behaviors in the BC generations project. Sci Rep. 2021;11(1):12094.
53. Larkin JR, et al. Metabolomic biomarkers in blood samples identify cancers in a mixed population of patients with nonspecific symptoms. Clin Cancer Res. 2022;28:1651.
54. Li L, et al. Metabolomics-based discovery of molecular signatures for triple negative breast cancer in Asian female population. Sci Rep. 2020;10(1):370.
55. Seow WJ, et al. Association of untargeted urinary metabolomics and lung cancer risk among never-smoking women in China. JAMA Netw Open. 2019;2(9):e1911970.
56. Kelly RS, et al. Metabolomics and communication skills development in children; evidence from the ages and stages questionnaire. Meta. 2019;9(3):42.
57. Titan SM, et al. Metabolomics biomarkers and the risk of overall mortality and ESRD in CKD: results from the Progredir cohort. PLoS One. 2019;14(3):e0213764.
58. Bot M, et al. Metabolomics profile in depression: a pooled analysis of 230 metabolic markers in 5283 cases with depression and 10,145 controls. Biol Psychiatry. 2020;87(5):409–18.
59. Wang J, et al. Discovery of potential biomarkers for osteoporosis using LC-MS/MS metabolomic methods. Osteoporos Int. 2019;30(7):1491–9.
60. Surowiec I, et al. Metabolomic and lipidomic assessment of the metabolic syndrome in Dutch middle-aged individuals reveals novel biological signatures separating health and disease. Metabolomics. 2019;15(2):23.
61. Al Rashid K, et al. Association of the functional ovarian reserve with serum metabolomic profiling by nuclear magnetic resonance spectroscopy: a cross-sectional study of ~ 400 women. BMC Med. 2020;18(1):247.
62. Hellmuth C, et al. An individual participant data meta-analysis on metabolomics profiles for obesity and insulin resistance in European children. Sci Rep. 2019;9(1):5053.
63. Wang Y, et al. Identification and reproducibility of plasma metabolomic biomarkers of habitual food intake in a US diet validation study. Meta. 2020;10(10):382.
64. Wang Y, et al. Identification and reproducibility of urinary Metabolomic biomarkers of habitual food intake in a cross-sectional analysis of the cancer prevention Study-3 diet assessment sub-study. Meta. 2021;11(4):1.
65. Darst BF, et al. Longitudinal plasma metabolomics of aging and sex. Aging (Albany NY). 2019;11(4):1262.
66. Lau CE, et al. Determinants of the urinary and serum metabolome in children from six European populations. BMC Med. 2018;16(1):202.
67. Brachem C, et al, Oluwagbemigun K. Associations of BMI and body fat with urine metabolome in adolescents are sex-specific: a cross-sectional study. Metabolites. 2020;10(8).
68. McClain KM, et al. Body composition and metabolomics in the Alberta physical activity and breast cancer prevention trial. J Nutr. 2022;152(2):419–28.
69. Zanetti KA. The future of metabolomic profiling in population-based research: opportunities and challenges. J Anal Bioanal Tech. 2014;5(4):2.

Metabolomics and Transcriptomic Approach to Understand the Pathophysiology of Interstitial Lung Disease

Sanjukta Dasgupta, Anindita Bhattacharya, Priyanka Choudhury, Nilanjana Ghosh, Tanisha Das, Sushmita Roychowdhury, Riddhiman Dhar, and Koel Chaudhury

Abstract Interstitial lung diseases (ILD) are a heterogeneous group of parenchymal pulmonary disorders that result from varying degrees of inflammation or fibrosis in the lung interstitium, that is, the septum between alveoli and the blood capillaries. The clinical presentation of ILD is complex and the diagnosis is often challenging. Therefore, the need to establish disease-specific molecular fingerprints to better understand the underlying pathogenesis is well realized. "Omics" is a powerful tool that collectively depicts and quantifies biomolecules, including key genomic, transcriptomic, proteomic, and metabolomic signatures, and discloses their dynamic interactions within an organism. Metabolomics is a branch of omics that identifies numerous small molecules from body fluids or tissues and holds immense potential for early diagnosis, therapeutic monitoring, and understanding of disease pathophysiology. Another evolving popular omic field is transcriptomics, which identifies key genetic regulations and posttranscriptional modifications triggering diseases. The findings of 17 original articles on metabolomics and 63 on transcriptomics of ILD reported are discussed. Though each omic dataset provides valuable information, integrating these platforms offers an overall snapshot of the interplay between the candidate molecules and genes, thereby paving the path for highlighting the genotype-to-phenotype relationship and assisting in making more effective treatment decisions for complex diseases.

S. Dasgupta · A. Bhattacharya · P. Choudhury · N. Ghosh · T. Das · K. Chaudhury (✉)
School of Medical Science and Technology, Indian Institute of Technology Kharagpur, Kharagpur, India
e-mail: koel@smst.iitkgp.ac.in

S. Roychowdhury
Fortis Hospital, Kolkata, India

R. Dhar
Department of Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_14

Keywords Interstitial lung diseases · Metabolomics · Transcriptomics · Systems biology

1 Introduction

Interstitial lung disease (ILD) is an umbrella term that encompasses about 300 parenchymal pulmonary disorders, resulting from varying degrees of inflammation or fibrosis in the lung interstitium, that is, the septum between alveoli and the blood capillaries. A schematic diagram of a healthy vs. ILD lung is shown in Fig. 1.

Numerous studies across the globe have reported the incidence, prevalence, and relative frequency of ILD. The annual incidence of ILD varies between 1 and 31.5 per 100,000 [1]. The incidence and prevalence vary among populations, likely due to differences in study design, data collection, and incorrect recognition of the disease subtypes [2]. ILD is classified based on clinical, radiological, and histopathological features. The latest classification focuses on recognizing the underlying etiology since this often impacts both prognostication and management decisions. ILD mainly consists of disorders of known causes [collagen vascular disease, hypersensitivity pneumonitis (HP)] as well as disorders of unknown/idiopathic causes [idiopathic interstitial pneumonia (IIP), sarcoidosis] [3]. ILD registries comprising patients from Western countries suggest that idiopathic pulmonary fibrosis (IPF) and sarcoidosis are the most common phenotypes. However, the ILD registry of India indicates HP to be the most common, which accounts for nearly 50% of all ILD cases [4].

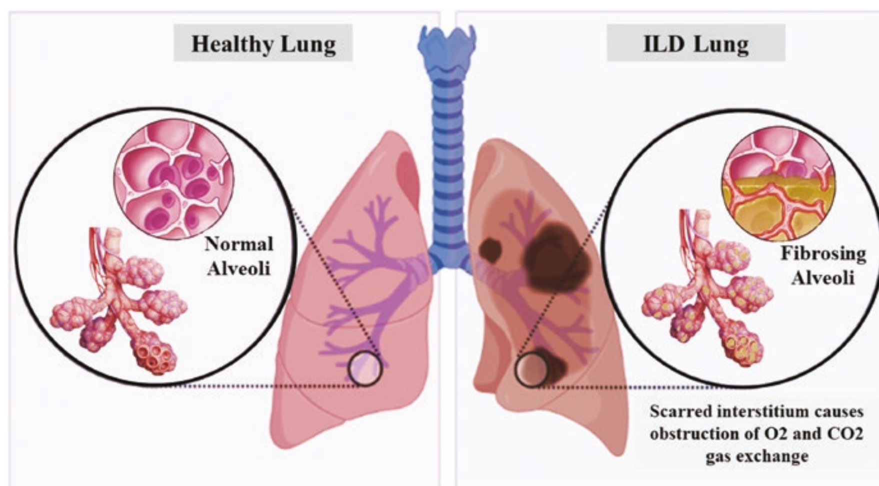

Fig. 1 Healthy lung vs. interstitial lung disease (created using BioRender.com)

The emerging field of metabolomics, in which many small molecules from body fluids or tissues can be identified, holds immense potential for early diagnosis, therapeutic monitoring, and understanding of disease pathophysiology. Over the past two decades, nuclear magnetic resonance (NMR) spectroscopy and gas chromatography (GC)/liquid chromatography (LC) coupled with mass spectrometry (MS) combined with chemometric analysis have emerged as principal analytical techniques for use in metabolomics. Several biofluids including cerebrospinal fluid (CSF), bronchoalveolar lavage fluid (BALF), bile, seminal fluid, amniotic fluid, synovial fluid, gut aspirate, serum/plasma, saliva, exhaled breath condensate (EBC), and urine contain hundreds to thousands of detectable metabolites which have been extensively studied so far [5]. More recently, metabolic profiling of intact tissue and extracts of lipid and aqueous metabolites are gaining increasing importance for detection of biomarkers.

Another branch of popular omic science is transcriptomics, which provides detailed information about gene regulation in normal and diseased conditions. Two key contemporary techniques commonly used for transcriptomic analysis are hybridization-based microarray techniques, which quantify a set of predetermined sequences, and next-generation sequencing (NGS), which uses high-throughput sequencing to capture all sequences [6]. In the last decade, these two transcriptomic approaches have been utilized most widely to understand the underlying disease pathogenesis at both molecular and genetic levels and also for molecular diagnosis and clinical therapy. Human biofluids including amniotic fluid, aqueous humor, ascites, bile, BALF, breast milk, CSF, colostrum, gastric fluid, pancreatic cyst fluid, plasma, saliva, seminal fluid, serum, sputum, stool, synovial fluid, sweat, tears, urine, and tissues are widely used for transcriptomic studies to identify biomarkers of several diseases [7–9].

2 Types of ILD

ILD, as mentioned earlier, refers to a group of lung diseases ranging from occasional self-limited inflammatory processes to severe debilitating fibrosis of the lung parenchyma. There are varied causes of ILD, which generally result from a range of environmental, occupational, recreational, or drug-related exposures or could arise from the various systemic autoimmune or connective tissue diseases (CTD) [10]. Classification of different types of ILD is shown in Fig. 2. A few of the common ILD subtypes are described in the present section.

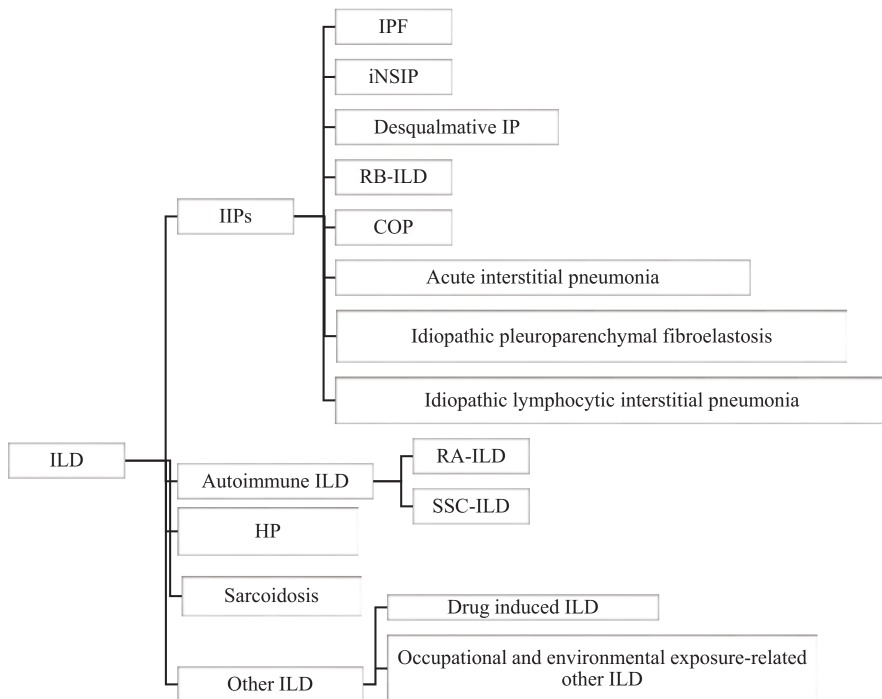

Fig. 2 Classification of different types of interstitial lung disease (Cottin et al. 2018) [3]. *ILD* interstitial lung disease, *IIP* idiopathic interstitial pneumonia, *IPF* idiopathic pulmonary fibrosis, *iNSIP* idiopathic nonspecific interstitial pneumonia, *RB-ILD* respiratory bronchiolitis-associated ILD, *COP* cryptogenic organizing pneumonia, *RA-ILD* rheumatoid arthritis-associated ILD, *SSC-ILD* systemic sclerosis-associated ILD, *HP* hypersensitivity pneumonitis

2.1 Idiopathic Interstitial Pneumonia (IIP)

The cause of IIP, comprising of diffuse parenchymal lung diseases, remains unknown. IIP is characterized by varying degrees of inflammation and fibrosis in the lung interstitium. These characteristics split IIP into eight clinicopathologic entities, that is, IPF, nonspecific interstitial pneumonia (NSIP), cryptogenic organizing pneumonia (COP), acute interstitial pneumonia, respiratory bronchiolitis-associated interstitial lung disease, desquamative interstitial pneumonia, lymphoid interstitial pneumonia, and idiopathic pleuroparenchymal fibroelastosis [11]. Among all IIPs, IPF is the most common phenotype characterized by fibroblastic foci and the presence of inflammation and honeycombing in the lung parenchyma.

2.2 Autoimmune ILD

Autoimmune ILD is caused specifically by autoimmune disorders, which involve the body's immune system attacking the lungs. This ILD group gradually develops and emerges over a long period of time. The symptoms of this ILD include difficulty in breathing, dry cough, and shortness of breath. Connective tissue disease-related ILD (CTD-ILD), rheumatoid arthritis-associated ILD (RA-ILD), and systemic sclerosis-associated ILD (SSC-ILD) are the common types of autoimmune ILD [12, 13].

2.3 Hypersensitivity Pneumonitis (HP)

HP, also referred to as extrinsic alveolar alveolitis, is a complex subtype of ILD arising from repeated exposure to certain antigens, most commonly avian, microbial (especially molds), or chemical. HP is the third most prevalent ILD after IPF and CTD-ILD. The inhaled antigen triggers type III and type IV hypersensitivity reactions, which causes the damage of alveolar epithelial cells. An impaired repair mechanism may result in fibroblast activation, deposition of collagen by the destruction of extracellular matrix, and parenchymal architecture [14]. The major forms of HP are acute, subacute, and chronic. Acute and subacute HP is mainly characterized by influenza-like symptoms, such as cough, dyspnea, and fever, developing after 2–9 h of antigen exposure. The chronic form of HP arises from repetitive, low-level exposure to the causative agent. Still, the identity of the causative antigen may remain unknown in more than half the cases. Chronic HP patients slowly develop fibrosis in the lung interstitium and are associated with a significantly high mortality rate [15].

2.4 Sarcoidosis

Sarcoidosis is a systemic, inflammatory disease resulting from an unknown origin. Chronic immune response to an idiopathic antigen may lead to sarcoidosis in genetically susceptible subjects. Almost 90% of sarcoidosis patients have pulmonary involvement. Dry cough, chest tightness, chronic dyspnea on exertion, shortness of breath, wheezing, hypoxemia, and decline in pulmonary function are the common signs and symptoms of sarcoidosis. Near about 20% of sarcoidosis patients develop pulmonary fibrosis, that is, stage IV sarcoidosis which is associated with high mortality [16].

2.5 *Occupational and Environmental Exposure-Related Other ILDs*

Long-term exposure to occupational or environmental antigens could cause certain types of ILD via pulmonary and systemic inflammation and oxidative stress. Many different types of mineral dust, such as silica, asbestos, beryllium, coal mine dust, metal, and organic dust, including mold spores, can also affect the lung airways, either by a direct wound or through reactive oxygen molecules. Common conditions include asbestosis, which is associated with asbestos fibers, and silicosis, which is caused by free crystalline silicon dioxide or silica particles [17, 18].

3 Metabolomics: An Emerging Tool in Clinical Research

Metabolomics, one of the newest omics science, is an evolving field in clinical research. Metabolomics is the scientific study of metabolic fingerprints that all cellular processes leave behind in a biological sample [19]. It provides a snapshot of the metabolic state of an individual at a given point in time. On the other hand, "metabonomics," a term first coined by Jeremy Nicholson, refers to "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification" [20, 21]. The terms "metabolomics" and "metabonomics" are often used interchangeably. Among the different omic approaches, metabolomics is considered to modulate best and depict the molecular phenotype of health and disease [22]. Thus, it is increasingly becoming a useful and powerful tool for the investigation of complex diseases with unclear etiology, enabling the discovery of novel biomarkers, which, in turn, aid in the prevention and early diagnosis of diseases. Metabolomics can also monitor the effect of pharmacotherapy, allowing clinicians to choose the best treatment option for patients suffering from potentially devastating disorders. The two analytical techniques popularly used have their own advantages and disadvantages. While mass spectrometry can analyze a wider range of metabolites and is more sensitive, it results in the destruction of the analyzed sample. NMR spectroscopy, on the other hand, is highly reproducible and does not destroy the sample; however, sensitivity is limited [23, 24]. Over the years, application of metabolomics in diseases is rapidly growing, and recent studies exploring the metabolic profiles of various human samples, including but not limited to plasma, serum, urine, BALF, exhaled breath, saliva, and tissues, bring this technology closer to the patients' bedside, thereby enhancing its clinical utility. A schematic representation of the metabolomic workflow is shown in Fig. 3.

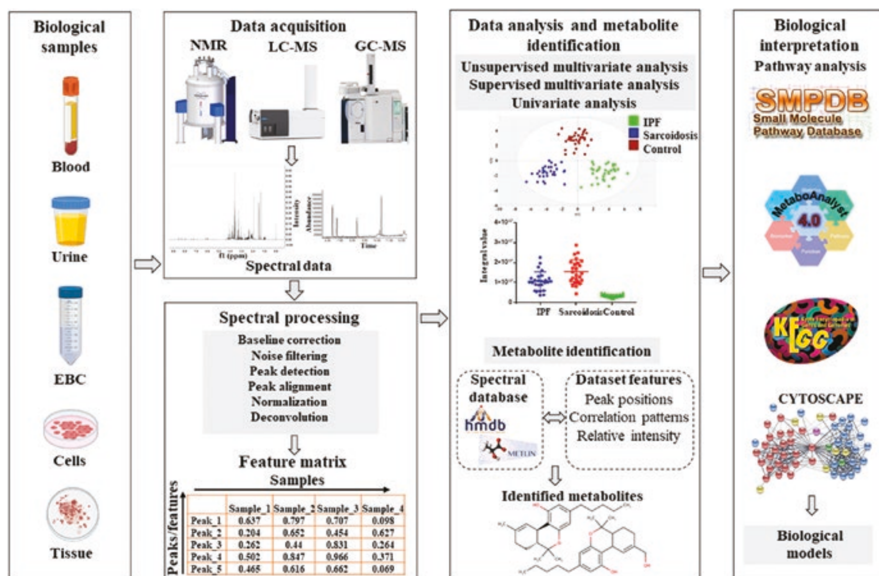

Fig. 3 Schematic representation of the metabolomic workflow (created using BioRender.com)

3.1 Metabolomics in ILD

Several attempts have been made to understand the metabolic status of ILD patients and identify prospective biomarkers in lung tissues and various body fluids using a nontargeted and targeted metabolomic approach. Studies utilizing the metabolomic approaches to investigate ILD are summarized in Table 1.

Table 1 A summary of studies exploring different types of ILD in humans using metabolomic approach

IPF			
Biological sample	Technique	Main findings	References
Tissue	NMR (untargeted)	Lactic acid levels significantly elevated in IPF lung tissue, suggested to be a key driver of myofibroblast differentiation, as well as onset and progression of fibrotic disorders	[25]
Tissue	MS (untargeted)	Alterations in glycolytic, adenosine triphosphate degradation, glutathione biosynthesis, and ornithine aminotransferase pathways indicated in lung tissues of IPF patients	[26]
Tissue	MS (targeted)	Free fatty acid dysregulation in IPF lungs; stearic acid suggested exhibiting antifibrotic effect in IPF	[27]
Tissue	MS (untargeted)	Dysregulation in sphingolipid metabolic pathway, arginine pathway, glycolysis, TCA cycle, and mitochondrial β-oxidation; dysregulated haem, bile acid, and glutamate/aspartate metabolism suggested to play a crucial role in IPF pathogenesis	[28]
Exhaled breath	MS (untargeted)	Distinct metabolic profile with 58 discriminatory metabolites identified in EBC of IPF patients	[29]
Exhaled breath	MS (targeted)	Significantly increased expression levels of proline, 4-hydroxyproline, alanine, valine, leucine/isoleucine, and allysine were detected in exhaled breath of IPF patients	[30]
Plasma	MS (targeted)	62 altered lipids, including 24 types of glycerophospholipids, 30 types of glycerolipids, 3 types of sterol lipids, 4 types of sphingolipids, and 1 type of fatty acid identified in the plasma of IPF patients	[31]
Plasma	MS (untargeted)	Lysophosphatidylcholine (lysoPC) and several fatty acids, including palmitoleic acid, oleic acid, and linoleic acid, significantly upregulated, whereas dihydrotestosterone significantly downregulated in IPF patients	[32]
Plasma	MS (targeted)	Discrimination between stable and progressive IPF patients based on differences in plasma levels of triglycerides and phosphatidylcholine; this difference further confirmed in lung tissue of IPF	[33]
Serum	MS (untargeted)	LysoPC was found to be significantly dysregulated in IPF patients, indicating its potential as a biomarker for diagnosis and monitoring of IPF	[34]
HP			
Serum, EBC, and BALF	NMR (untargeted)	Three metabolites, including lactate, pyruvate, and proline, significantly altered in all three biofluids	[35]

Table 1 (continued)

IPF			
Biological sample	Technique	Main findings	References
Sarcoidosis			
Serum	NMR (untargeted)	Three major pathways, including fatty acid metabolism, glycolysis/TCA cycle, and homocysteine/methylamine, altered in sarcoidosis	[36]
Plasma	NMR (untargeted) and MS (targeted)	Distinct metabolomic and metallomic profiles were observed in veterans with sarcoidosis as compared to civilians, with levels of magnesium, calcium, aluminium, titanium, and iron increased in sarcoidosis	[37]
RA-ILD			
Serum	MS (untargeted)	Four serum metabolites (mannosamine, alliin, kynurenine, and 2-hydroxybutyric acid) exhibit better performance in distinguishing types of RA patients with acute-onset diffuse ILD (AoDILD) as compared to existing AoDILD markers, KL-6 and SP-D	[38]
Serum	MS (untargeted)	Significantly altered expression of decanoic acid, glycerol, and morpholine was observed on comparing RA-ILD (usual interstitial pneumonia-associated RA and NSIP-associated RA) and RA patients without any chronic lung disease	[39]
Silicosis			
Plasma	MS (untargeted and targeted)	L-arginine and kynurenine associated with severity of silicosis with a predictive role in disease monitoring	[40]
Lymphangioleiomyomatosis			
Cell line	MS (untargeted)	Targeting E_2-dependent cellular metabolic pathways may have favorable therapeutic effects on lymphangioleiomyomatosis patients	[41]

4 Transcriptomics: A Promising Omic Approach

Transcriptome analysis utilizes high-throughput methods to study the complete set of RNA transcripts produced by the genome under specific circumstances. It covers all types of transcripts, including mRNAs, miRNAs, and different types of long noncoding RNAs (lncRNAs). Transcriptome analysis gives us an overview of all genes' expression levels and enables us to understand the physiology of the cell. More precisely, it also discloses key regulations of biological processes triggering diseases. While microarrays are generally less complex and easier to use than NGS, the latter is associated with greater flexibility, high throughput, and high discovery potential. A schematic representation of the transcriptomic workflow is shown in Fig. 4.

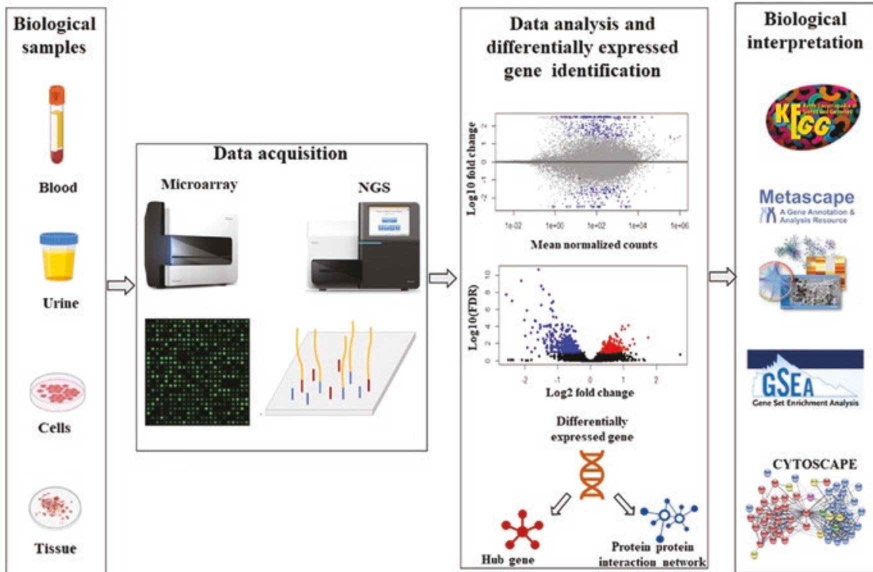

Fig. 4 Schematic representation of the transcriptomic workflow (created using BioRender.com)

4.1 Transcriptomics in ILD

Various studies have been performed to understand the transcriptomic signatures of ILD patients and identify prospective biomarkers in lung tissues and various biofluids using NGS and microarray techniques. Despite increasing interest and effort invested by clinicians and scientists during the last decade, the etiology of ILD remains elusive and controversial.

As mentioned earlier, IPF is characterized by remodeling or scarring of the airway epithelium. The activated extracellular matrix (ECM)-produced myofibroblasts play a key role in the process of fibrotic tissue remodeling. Advances in transcriptomic techniques have allowed high-throughput analysis and discovery of gene deregulation in IPF. Several studies using lung tissues have reported that IPF is associated with variances in the expression levels of genes such as *CCL8* [42], *CXCL14* [43], *CXCL4* and *CXCL12* [44], *NOTCH2* [45], *TGF-β1* and *RhoA* kinase [46], *REVERBα* [47], *IL-1β* [48], *FLIL33* and *POU2AF1* [49], *FOXL1* [50], *COL6A3*, and *POSTN* [51]. Microarray analysis of peripheral blood by Abe et al. (2020) has shown dysregulated *PDGF B*, *VEGF B*, and *FGF 2*. The authors confirmed their findings using ELISA, western blot, immunofluorescence, and ^{3}H thymine uptake assays. Xia and co-workers (2021) recently utilized weighted gene co-expression network analysis (WGCNA) of BALF samples and could associate four genes, *TLR2*, *CCR2*, *HTRA1*, and *SFN*, with disease prognosis.

Pathway enrichment analysis based on dysregulated genes highlights the associated biological pathways, molecular functions, and cellular components. This

method identifies all biological pathways enriched in a gene list more than would be expected by chance. The KEGG pathway tool maps the pathways associated with dysregulated genes in a specific disease. Pathway enrichment analysis of IPF patients revealed that the differentially expressed genes were majorly associated with myofibroblast differentiation and massive ECM deposition. The transcriptomic signatures of fibroblasts suggest that characterization of lung proteins, specifically lung fibrotic ECM, helps determine its composition and define targetable molecules for advanced stages of fibrosis. Boesch and his team (2020) isolated fibrosis-specific mesenchymal stem cell-like cells from lung tissue of IPF subjects and observed that the differentially expressed genes were enriched with hypoxia, fibrosis, and bacterial colonization factors which are the typical hallmarks of pulmonary fibrosis. They found that the cells isolated from IPF patients express genes associated with activating canonical TGF-β, HIPPO/YAP, PI3K/AKT, p53, and WNT signaling cascades, which are activated in an integrated network. Another interesting study by Hsu and co-authors (2011) suggested that IPF lungs enriched in fibrosis-related genes, insulin-like growth factor signaling, and caveolin-mediated endocytosis. This microarray analysis also highlighted the common molecular signatures between lung tissue and fibroblasts of these patients.

Like IPF, HP is associated with matrix remodeling and formation of fibrosis. There exist only two studies where transcriptomics has been used to explore genetic alterations in HP. Sarcoidosis, as mentioned earlier, is an immune-mediated multisystem disease characterized by the formation of non-caseating granuloma. Multiple pro-inflammatory signaling pathways, including *IFN-γ/STAT-1*, *IL-6/STAT-3*, and *NF-κB*, have been implicated in mediating macrophage activation and granuloma formation in sarcoidosis. Utilizing RT-PCR, Christophi et al. (2014) have demonstrated that *IL-6*, *COX-2*, *MCP-1*, *IFN-γ*, T-bet, *IRF-1*, *Nox2*, *IL-33*, and eotaxin-1 hold potential for differential diagnosis between sarcoidosis, suture, and fungal granulomas. In another recent study, Lepzien and co-workers (2021) have shown that allogeneic T cell proliferation increased after coculture with monocytes and dendritic cells of sarcoidosis patients. The authors also found that mainly T-bet and RORγt-expressing T cells produce IFN-γ. Monocytes from sarcoidosis patients can activate and polarize T cells towards Th1 and Th17.1 cells. In a comparative study between sarcoidosis and IPF, cluster analysis of BALF cells showed elevated mRNA expression of genes associated with ribosome biogenesis in sarcoidosis patients. Clusters formed by genes with altered mRNA expression in patients with IPF could be implicated in cell migration and adhesion processes, metalloproteinase expression, and negative regulation of cell proliferation. Various studies highlighting the transcriptome fingerprints and associated pathways in different ILD subtypes are summarized in Table 2.

Table 2 A summary of studies exploring different types of ILD in humans using transcriptomic approach

IPF			
Biofluid	Technique	Findings	Reference
Plasma, BALF, And tissues	Microarray	*CCL8* is a key molecule for differential diagnosis of IPF and can also predict survival	[42]
Tissue	NGS	Differentially expressed genes in IPF are associated with fibrosis, hypoxia, bacterial colonization, and pulmonary fibrosis metabolism	[43]
Tissue	Microarray	TGF-β1, RhoA kinase, and the TSC2/RHEB axis form major signaling clusters associated with collagen gene expression in IPF	[44]
Tissue	NGS	Specific connective tissue-related genes including alpha-smooth muscle actin, fibrillin, fibronectin, tenascin C, osteopontin, chains of highly abundant structural collagens and other collagens, multiple matrix metalloproteinases, and Wilms tumor protein are elevated in IPF	[45]
Tissue	Microarray	*TGF-β1* increases the risk of developing IPF in smokers	[46]
Tissue	NGS	Notch signaling regulates the maintenance of an expanded pool of secretory primed basal cells in the distal lung of IPF patients	[47]
Tissue	Microarray	IPF lungs are enriched with fibrosis-related gene, insulin-like growth factor signaling, and caveolin-mediated endocytosis	[48]
Tissue	Microarray	Lower expression of cell migration-inducing and hyaluronan-binding protein in pirfenidone-treated IPF patients	[49]
Tissue	Microarray	IPF lungs are enriched with cell adhesion, molecule binding, chemical homeostasis, surfactant homeostasis, and receptor binding genes	[50]
Tissue	NGS	Elevated expression of numerous immune, inflammation, and extracellular matrix-related mRNAs observed in IPF	[45]
Tissue	NGS	Alternative splicing COL6A3 and POSTN may be involved in the pathogenesis of IPF	[51]
Tissue	Microarray	Twist1 as a regulator of noncanonical NF-κB signaling through CXCL12 may have a profibrotic effect in IPF	[52]
Tissue	Microarray	*CXCL14* and *CXCL4* may be involved in the activation of fibroblasts within IPF lungs and are involved in disease pathogenesis	[53]
Tissue	Microarray	A significant upregulation of EGFR, both at protein and mRNA level, was observed in IPF, fibrotic NSIP, and COP compared with controls	[54]
Tissue	NGS	*MMP7* is differentially expressed in IPF patients	[55]

Table 2 (continued)

IPF			
Biofluid	Technique	Findings	Reference
Tissue	NGS	Hypoxia and TGF-β1 synergistically increase myofibroblast marker expression in IPF	[56]
Tissue	NGS	Discrete types of macrophages expressing (1) monocyte markers and (2) higher levels of FABP4, INHBA, SPP1, and MERTK present in IPF lungs	[57]
Tissue	NGS	FOXL1 can control a wide array of genes that potentiate fibroblast function, including *TAZ*/*YAP* signature genes and PDGF receptor-α in IPF	[58]
Tissue	NGS	*POU2AF1* regulates fibrosis in IPF	[59]
Tissue	NGS	*FLIL33* overexpression and stimulation with TGF-β differentially regulates the fibroblast transcriptome in IPF	[60]
Tissue	Microarray	Genes associated with cell adhesion, molecule binding, chemical homeostasis, surfactant homeostasis, and receptor binding are dysregulated in lungs of IPF patients	[50]
Tissue	Microarray	LncRNAs are crucial regulators of proliferation and inflammation in human lung fibroblasts, suggesting their possible involvement in the lower inflammatory response in IPF	[61]
Tissue	NGS	Increased CD44 is a characteristic of IPF mesenchymal progenitor cells	[62]
Tissue	NGS	Following TGF-β1 stimulation, collagen secretion is elevated in IPF patients	[63]
Tissue and plasma	NGS	The expression of *GDF15* is increased in IPF and is associated with the progression of the disease	[64]
Tissue	NGS	Altered basaloid cells that express basal epithelial, mesenchymal, senescence, and developmental markers are located at the myofibroblast foci edge. Ectopically expanded cell populations are observed in vascular endothelial cells	[65]
Tissue	Microarray	*CXCL12*, collagen 3A1, *MMP2*, and *MMP14* are upregulated in fibrotic ILD, including IPF, NSIP, organizing pneumonia, and alveolar fibroelastosis as compared with controls	[66]
Tissue	NGS	IPF fibroblast transcriptional signatures indicate enrichment of *WNT*, *TGF-β*, and ECM genes and downregulation of miR-29b-3p, miR-138-5p, and miR-146b-5p	[67]
Tissue	Microarray	Pathways associated with vascular proliferation, WNT signaling, and apoptosis are dysregulated in IPF arterioles	[68]

(continued)

Table 2 (continued)

IPF			
Biofluid	Technique	Findings	Reference
Tissue	NGS	Alveolar type 1 (AT1), AT2, and conducting airway selective markers are frequently co-expressed by IPF cells, and aberrant activation of canonical signaling via TGF-β, HIPPO/YAP, p53, WNT, and AKT/PI3K is predicted via pathway analysis	[69]
Tissue	Microarray	Expression of cilium genes appears to identify two unique molecular phenotypes Of IPF/UIP, which may affect therapeutic responsiveness	[70]
BALF	Microarray	IPF is associated with cell migration, cell adhesion, metalloproteinase expression, and negative regulation of cell proliferation	[71]
BALF	Microarray	*TLR2*, *CCR2*, *HTRA1*, and *SFN* are involved in the prognosis of IPF	[72]
Peripheral blood	Microarray	*PDGF B*, *VEGF B*, and *FGF 2* genes are associated with IPF	[73]
Peripheral blood	Microarray	*YBX3*, *UTRN*, hsa_circ_0001924, and *FENDR* could be potential diagnostic biomarkers of IPF	[74]
Peripheral blood	Microarray	Increased circulating *FUT3* level is associated with reduced risk of IPF	[75]
PBMC, monocytes, and serum	NGS	Type I IFN pathway is the key regulator for driving chronic inflammation and fibrosis in IPF	[76]
Nasal biopsy	NGS	Pathways related to immune response and inflammatory signaling are elevated in IPF patients	[77]
Tissue	Fluorescence-based RNA quantitation assay	IGF-1 signaling, ERK/MAPK signaling, protein ubiquitination, PI13/AKT signaling, cardiac b-adrenergic signaling, actin-cytoskeleton signaling, integrin signaling, and NRF2-mediated oxidative stress response pathways are associated with IPF	[78]
HP			
Tissue	NGS	HP is associated with specific genes, including *CXCL9*, an IFN-γ-inducible chemokine, and ligand for CXCR3	[79]
Tissue	NGS	Antigen presentation and extracellular matrix-associated transcriptomic signatures are present in mild HP cases, whereas B cells are predominant in fibrotic HP	[80]
Sarcoidosis			
Tissue	Microarray	Multiple pro-inflammatory signaling pathways mediate macrophage activation and granuloma formation in sarcoidosis	[81]

(continued)

Table 2 (continued)

IPF			
Biofluid	Technique	Findings	Reference
Tissue	NGS	*STAB1*, *HBEGF*, and *NOTCH4* genes are associated with sarcoidosis pathogenesis	[82]
BALF cells	Microarray	Increased mRNA gene expression associated with ribosome biogenesis and proteasome apparatus observed in sarcoidosis patients	[71]
BAL	Microarray	Cathepsin S is significantly upregulated in sarcoidosis	[83]
BAL	NGS	In four sarcoidosis endotypes (hilar lymphadenopathy, extraocular involvement, chronic stage, and multiorgan involvement condition), elevated acute T-cell response, PI3K pathways, increased immune response pathways, and increased IL-1 and IL-18 immune and inflammatory responses are observed	[84]
Blood and BAL	NGS	Monocytes of sarcoidosis patients can activate and polarize T cells toward Th1 and Th17.1	[85]
Blood and BAL	NGS	Monocytes/monocyte-derived cells increased in blood and BAL of sarcoidosis compared to healthy controls	[86]
Blood	Microarray	Interferon-inducible neutrophil-driven blood transcriptional signature observed in sarcoidosis	[87]
PBMC and BAL cells	Microarray	Alterations in TLR2 signaling pathway and downstream of NF-κB apoptosis and proliferation evidenced in sarcoidosis	[88]
PBMC	NGS	Dysfunctional p53, cell death, and TNFR2 signaling associated with sarcoidosis	[89]
PBMC, in vitro granuloma, and tissue	Microarray	Molecular pathways, regulated by IL-13, which helps in activated M2 macrophage polarization, is associated with the pathogenesis of sarcoidosis	[90]
SSC-ILD			
Tissue	NGS	Mesenchymal cell population including SPINT2hi, MFAP5hi, few WIF1hi fibroblasts, and a new large myofibroblast population may be actively involved in the regulation of disease pathogenesis	[91]
Tissue	NGS	Cellular stress pathways are upregulated in SSC-ILD, a population of KRT5-/KRT17+ aberrant basaloid cells representing markers of epithelial-mesenchymal transition and cellular senescence identified in the disease for the first time	[92]
Tissue	Microarray	Increased expression of TGF-β response signature is the key regulator of fibrosis formation in fibrotic SSC-ILD	[93]

(continued)

Table 2 (continued)

IPF			
Biofluid	Technique	Findings	Reference
Tissue	Microarray	Targeting *IL-6* trans-signaling, *IGFBP2*, *IGFL2*, and the coagulation cascade represent potential therapeutic strategies against the disease	[94]
Skin biopsy	Microarray	*SELP*, *MMP 3*, and *CCL2* which are involved in the adhesion and extravasation of inflammatory cells are associated with SSC-ILD	[95]
Serum	Microarray	Hepatic fibrosis, granulocyte and agranulocyte adhesion, and diapedesis are associated with SSC-ILD	[96]
Silicosis			
Tissue	NGS	Several critical genes, including *MUC5AC* and *FGF10*, serve as potential drug targets in silicosis	[97]
Cell line	NGS	Transcription factors, *EGR2* and *BHLHE40*, are upregulated while *TBX2*, *NR1H3*, *NR2F1*, *PPAR-γ*, and *EPAS1* are downregulated, which may play a crucial regulatory role in disease pathogenesis	[98]
Dermatomyositis-associated ILD			
Blood	NGS	*PLAUR* may play an important role in disease pathogenesis by regulating the neutrophil-associated immune response	[99]

5 Integration of Metabolomic and Transcriptomic Fingerprints

As mentioned earlier, clinical metabolomics is primarily used to identify low molecular weight compounds differentially expressed in a particular disease. In contrast, transcriptomics identifies the complete set of dysregulated RNAs associated with a disease. Integration of metabolomic and transcriptomic signatures has emerged as a popular application-driven method for investigating underlying disease mechanisms, monitoring disease progression, and identifying potential biomarkers [100–102]. The omic tools highlight alterations in genotype and phenotype and provide complementary information about genetic alterations, protein synthesis, metabolism, and cellular function. Pathways and network connections further reflect the association between key metabolites and candidate transcripts.

Biological pathway networks reveal hidden patterns in unstructured data by converting them into logically structured and visually evident representations, with nodes representing genes and metabolites and edges suggesting relationships between nodes and clusters with similar chemical activities. VANTED [103], VisAnt

[104], Impala [105], and Metscape2 [106] are some of the network-based visualization tools that interface with public databases. In addition, Arena3D allows users to envision three-dimensional biological networks [107]. Interactive editing is frequently performed for small biological networks. However, for major networks, automated layout web tools, that is, Cytoscape [108], NAViGaTOR [109], and Cerebral [110], are more convenient. Alternatively, pathway visualization tools highlight the biochemical activities and different interactive pathways in experimental datasets. Pathguide offers an overview of nearly 190 web-usable network databases and biological pathways [110]. Arakawa and his team have developed a pathway visualization tool for KEGG-based pathways. Users can capture systematic features of biological activity by visualizing pathways at the level of different omic data representations [111]. Paintomics, another software program, analyzes the expression of genes and concentration of metabolite data and displays it on KEGG pathway maps [112]. ProMeTra can display dynamic data and accept annotated images in SVG format [113]. In plants, KaPPa-View and MapMan show the number of metabolites and transcripts for preset route blocks [114, 115]. Other tools like MAYDAY enable viewing expression data in a genomic context with any metadata [116], and PaVESy creates personalized pathways using proteins and metabolites provided by the user [117]. A schematic representation of integrated metabolomic and transcriptomic workflow is shown in Fig. 5.

In a recent study, our group used NMR coupled with chemometric analysis to identify the unique metabolic fingerprints in BALF of HP subjects. A total of six metabolites were found to be significantly altered in HP compared to non-HP controls [35]. Next, we considered NGS data of lung tissues from HP patients and controls, reported in the NCBI-GEO public database by Furusawa et al., and

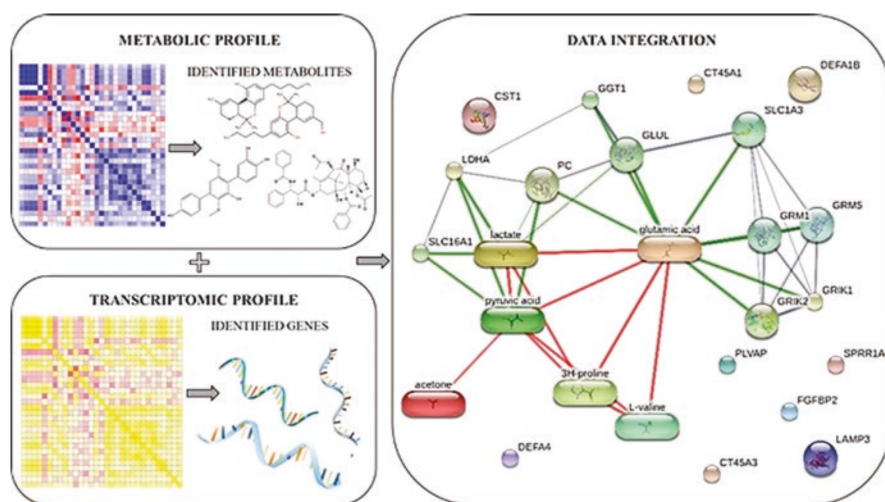

Fig. 5 Schematic representation of integrated metabolomic and transcriptomi data (created using BioRender.com, STITCH database, and Graph pad prism version 7)

performed bioinformatic analysis. A total of 555 genes were dysregulated (373 upregulated and 182 downregulated) in HP cases. An interaction network between the six candidate metabolites and most significantly altered genes (five upregulated and five downregulated) was established utilizing the Search Tool for Interactions of Chemicals (STITCH) database. The metabolite-gene interaction by STITCH demonstrated 19 nodes connected via 16 edges. The clustering coefficient of the network was found to be 0.768 (protein-protein interaction enrichment p-value: 0.0838). Overall pathway overrepresentation analysis was performed by integrating the candidate metabolites and transcripts utilizing IMPaLA version 12. Glycolysis and phosphatidylinositol 3-kinase-protein kinase B (PI3K-AKT) signaling pathways emerged to be most significantly associated with the pathogenesis of HP. These findings are encouraging, since association of these pathways in chronic HP is well established. Since glycolysis is the key energy driving force for myofibroblast differentiation and formation of fibrosis, perturbation of glycolysis seems likely [118]. The involvement of PI3K-AKT pathway is also evidenced in bleomycin-induced pulmonary fibrosis. It is hypothesized that PI3K-AKT plays a central role in fibrosis development [119, 120]. A novel insight into the pathogenesis of HP is envisioned by integrating the findings of the two omic platforms.

6 Challenges and Future Scope

Most of the omic-driven studies conducted on ILD so far have included a small number of patients, which is quite understandable considering that ILD is a severe condition with a short average life expectancy. Power and sample size estimation, however difficult, would be useful because the low sample size is connected with statistical errors and risks of overfitting and misleading calculations. Since omic output is highly dynamic, clinical variables such as physiological status, age, gender, and treatment may influence the findings. Hence, baseline characteristics of recruited ILD subjects need to be closely matched. Lack of a rigorous subject selection approach could also result in discovering markers that are not exclusive to ILD subtypes. It is observed that only a few groups have included healthy controls in their omic-driven research on ILD. Also, nonuniformity in including smokers and nonsmokers is frequently observed while comparing disease populations with healthy controls. This makes unbiased comparisons and conclusions impossible. A few groups were also unable to validate ILD candidate markers, which is crucial for biomarker identification. In fact, one of the main reasons why most of the omic-based disease markers identified so far have not made it to clinical practice is due to a lack of adequate validation trials. Another observation that warrants attention while using omics is that different research groups identify different biomarkers in the same biofluid for a particular disease. This is not surprising given the fact that factors such as sampling methods, sample collection, handling and preparation, instrumentation, and data mining protocols tend to vary from one setup to another. To generate robust and reproducible data, the practices and procedures should be

standardized and rigorously followed across all clinics and research laboratories. Metabolic flux analysis is crucial to obtain insight into dysregulated cellular metabolism caused by disease perturbations. It is expected that stratifying ILD patients based on disease severity and subtypes will significantly improve metabolome and transcriptome coverage. Assessment of sensitivity, specificity, and clinical relevance of the differentially expressed molecules is also recommended. For a reliable and unbiased diagnosis of this severe pulmonary disease, large-scale, well-designed, multicentric clinical studies and recruitment of suitable controls are recommended.

The ultimate focus of metabolomic and transcriptomic data integration is identifying key metabolic and genetic factors that contribute significantly to disease etiology. Integrated omics is more than a collection of tools; it is a comprehensive paradigm for interpreting multi-omic datasets in a way that can provide new insights into basic biology, as well as health and disease. Machine learning approaches for multi-omic data analyses is an emerging trend for exploring molecular pathways in detail and drawing a holistic representation of a given phenotype using all biological and clinical information of an individual. One of the major advantages is incorporating biological domain knowledge into the machine learning models as inductive biases to reduce data overfitting. Additionally, as omic tools evolve, they need to be user-friendly, interoperable, and effective for computationally intensive analyses. Machine learning methods offer novel techniques to integrate such omic datasets. With the emerging precision medicine initiative, where disease prevention and management take into account the variability in genes, environment, and lifestyle of each individual in contrast to the conventional one-size-fits-all approach, integration of clinical data with the patients' metabolome and genetic makeup will provide an in-depth understanding of disease pathophysiology and facilitate designing of targeted therapies for individuals, thereby revolutionizing precision medicine-based decision-making in the clinic.

Acknowledgments Figures were created with BioRender.com.

References

1. Rivera-Ortega P, Molina-Molina M. Interstitial lung diseases in developing countries. Ann Glob Health. 2019;85(1):4.
2. ERS. Interstitial lung diseases. In: ERS European lung White-book. Chapter 22. ERS; 2015. http://www.erswhitebook.org/chapters/interstitial-lung-diseases/. Accessed 15 Sep 2016.
3. Cottin V, Hirani NA, Hotchkin DL, Nambiar AM, Ogura T, Otaola M, Skowasch D, Park JS, Poonyagariyagorn HK, Wuyts W, Wells AU. Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. Eur Respir Rev. 2018;27(150):180076.
4. Singh S, Collins BF, Sharma BB, Joshi JM, Talwar D, Katiyar S, Singh N, Ho L, Samaria JK, Bhattacharya P, Gupta R. Interstitial lung disease in India. Results of a prospective registry. Am J Respir Crit Care Med. 2017;195(6):801–13.
5. Gowda GN, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. Expert Rev Mol Diagn. 2008;8(5):617–33.

6. Hulstaert E, Morlion A, Cobos FA, Verniers K, Nuytens J, Eynde EV, Yigit N, Anckaert J, Geerts A, Hindryckx P, Jacques P. Charting extracellular transcriptomes in the human biofluid RNA atlas. Cell Rep. 2020;33(13):108552.
7. El-Mogy M, Lam B, Haj-Ahmad TA, McGowan S, Yu D, Nosal L, Rghei N, Roberts P, Haj-Ahmad Y. Diversity and signature of small RNA in different bodily fluids using next generation sequencing. BMC Genomics. 2018;19(1):1–24.
8. Ferrero G, Cordero F, Tarallo S, Arigoni M, Riccardo F, Gallo G, Ronco G, Allasia M, Kulkarni N, Matullo G, Vineis P. Small non-coding RNA profiling in human biofluids and surrogate tissues from healthy individuals: description of the diverse and most represented species. Oncotarget. 2018;9(3):3097.
9. Godoy PM, Bhakta NR, Barczak AJ, Cakmak H, Fisher S, MacKenzie TC, Patel T, Price RW, Smith JF, Woodruff PG, Erle DJ. Large differences in small RNA composition between human biofluids. Cell Rep. 2018;25(5):1346–58.
10. Guler SA, Corte TJ. Interstitial lung disease in 2020: a history of progress. Clin Chest Med. 2021;42(2):229–39.
11. Oliveira DS, Araújo JDA, Paiva AFL, Ikari ES, Chate RC, Nomura CH. Idiopathic interstitial pneumonias: review of the latest American Thoracic Society/European Respiratory Society classification. Radiol Bras. 2018;51(5):321–7.
12. Antoniou KM, Margaritopoulos G, Economidou F, Siafakas NM. Pivotal clinical dilemmas in collagen vascular diseases associated with interstitial lung involvement. Eur Respir J. 2009;33(4):882–96.
13. Mathai SC, Danoff SK. Management of interstitial lung disease associated with connective tissue disease. BMJ. 2016;352:h6819.
14. Takemura T, Akashi T, Ohtani Y, Inase N, Yoshizawa Y. Pathology of hypersensitivity pneumonitis. Curr Opin Pulm Med. 2008;14(5):440–54.
15. Sforza GGR, Marinou A. Hypersensitivity pneumonitis: a complex lung disease. Clin Mol Allergy. 2017;15(1):1–8.
16. Nardi A, Brillet PY, Letoumelin P, Girard F, Brauner M, Uzunhan Y, Naccache JM, Valeyre D, Nunes H. Stage IV sarcoidosis: comparison of survival with the general population and causes of death. Eur Respir J. 2011;38(6):1368–73.
17. Gulati M, Redlich CA. Asbestosis and environmental causes of usual interstitial pneumonia. Curr Opin Pulm Med. 2015;21(2):193.
18. Leung CC, Yu ITS, Chen W. Silicosis. Lancet. 2012;379(9830):2008–18.
19. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends Biotechnol. 1998;16(9):373–8.
20. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181–9.
21. Faber JH, Malmodin D, Toft H, Maher AD, Crockford D, Holmes E, Nicholson JK, Dumas ME, Baunsgaard D. Metabonomics in diabetes research. J Diabetes Sci Technol. 2007;1(4):549–57.
22. Guijas C, Montenegro-Burke JR, Warth B, Spilker ME, Siuzdak G. Metabolomics activity screening for identifying metabolites that modulate phenotype. Nat Biotechnol. 2018;36(4):316–20.
23. Li S, Todor A, Luo R. Blood transcriptomics and metabolomics for personalized medicine. Comput Struct Biotechnol J. 2016;14:1–7.
24. Schreier C, Kremer W, Huber F, Neumann S, Pagel P, Lienemann K, Pestel S. Reproducibility of NMR analysis of urine samples: impact of sample preparation, storage conditions, and animal health status. Biomed Res Int. 2013;2013:878374.
25. Kottmann RM, Kulkarni AA, Smolnycki KA, Lyda E, Dahanayake T, Salibi R, Honnons S, Jones C, Isern NG, Hu JZ, Nathan SD. Lactic acid is elevated in idiopathic pulmonary fibrosis and induces myofibroblast differentiation via pH-dependent activation of transforming growth factor-β. Am J Respir Crit Care Med. 2012;186(8):740–51.

26. Kang YP, Lee SB, Lee JM, Kim HM, Hong JY, Lee WJ, Choi CW, Shin HK, Kim DJ, Koh ES, Park CS. Metabolic profiling regarding pathogenesis of idiopathic pulmonary fibrosis. J Proteome Res. 2016;15(5):1717–24.
27. Kim HS, Yoo HJ, Lee KM, Song HE, Kim SJ, Lee JO, Hwang JJ, Song JWW. Stearic acid attenuates profibrotic signalling in idiopathic pulmonary fibrosis. Respirology. 2021;26(3):255–63.
28. Zhao YD, Yin L, Archer S, Lu C, Zhao G, Yao Y, Wu L, Hsin M, Waddell TK, Keshavjee S, Granton J. Metabolic heterogeneity of idiopathic pulmonary fibrosis: a metabolomic study. BMJ Open Respir Res. 2017;4(1):e000183.
29. Rindlisbacher B, Strebel C, Guler S, Kollár A, Geiser T, Fiedler GM, Leichtle AB, Bovet C, Funke-Chambour M. Exhaled breath condensate as a potential biomarker tool for idiopathic pulmonary fibrosis—a pilot study. J Breath Res. 2017;12(1):016003.
30. Gaugg MT, Engler A, Bregy L, Nussbaumer-Ochsner Y, Eiffert L, Bruderer T, Zenobi R, Sinues P, Kohler M. Molecular breath analysis supports altered amino acid metabolism in idiopathic pulmonary fibrosis. Respirology. 2019;24(5):437–44.
31. Yan F, Wen Z, Wang R, Luo W, Du Y, Wang W, Chen X. Identification of the lipid biomarkers from plasma in idiopathic pulmonary fibrosis by lipidomics. BMC Pulm Med. 2017;17(1):1–12.
32. Nambiar S, Tan DBA, Clynick B, Bong SH, Rawlinson C, Gummer J, Corte TJ, Glaspole I, Moodley YPP, Trengove R. Untargeted metabolomics of human plasma reveal lipid markers unique to chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. Proteom Clin Appl. 2021;15(2–3):e2000039.
33. Nambiar S, Clynick B, How BS, King A, Walters EH, Goh NS, Corte TJ, Trengove R, Tan D, Moodley Y. There is detectable variation in the lipidomic profile between stable and progressive patients with idiopathic pulmonary fibrosis (IPF). Respir Res. 2021;22(1):1–8.
34. Rindlisbacher B, Schmid C, Geiser T, Bovet C, Funke-Chambour M. Serum metabolic profiling identified a distinct metabolic signature in patients with idiopathic pulmonary fibrosis–a potential biomarker role for LysoPC. Respir Res. 2018;19(1):1–12.
35. Dasgupta S, Ghosh N, Choudhury P, Joshi M, Chowdhury SR, Bhattacharyya P, Chaudhury K. NMR metabolomic and microarray-based transcriptomic data integration identifies unique molecular signatures of hypersensitivity pneumonitis. Mol Omics. 2022;18(2):101–11.
36. Geamanu A, Gupta SV, Bauerfeld C, Samavati L. Metabolomics connects aberrant bioenergetic, transmethylation, and gut microbiota in sarcoidosis. Metabolomics. 2016;12(2):35.
37. Banoei MM, Iupe I, Bazaz RD, Campos M, Vogel HJ, Winston BW, Mirsaeidi M. Metabolomic and metallomic profile differences between veterans and civilians with pulmonary sarcoidosis. Sci Rep. 2019;9(1):1–12.
38. Furukawa H, Oka S, Shimada K, Hashimoto A, Komiya A, Matsui T, Fukui N, Tohma S. Serum metabolomic profiles of rheumatoid arthritis patients with acute-onset diffuse interstitial lung disease. Biomark Insights. 2019;14:1177271919870472.
39. Furukawa H, Oka S, Shimada K, Okamoto A, Hashimoto A, Komiya A, Saisho K, Yoshikawa N, Katayama M, Matsui T, Fukui N. Serum metabolomic profiling in rheumatoid arthritis patients with interstitial lung disease: a case–control study. Front Med. 2020;7:599794.
40. Xue C, Wu N, Fan Y, Ma J, Ye Q. Distinct metabolic features in the plasma of patients with silicosis and dust-exposed workers in China: a case–control study. BMC Pulm Med. 2021;21(1):1–10.
41. Sun Y, Gu X, Zhang E, Park MA, Pereira AM, Wang S, Morrison T, Li C, Blenis J, Gerbaudo VH, Henske EP. Estradiol promotes pentose phosphate pathway addiction and cell survival via reactivation of Akt in mTORC1 hyperactive cells. Cell Death Dis. 2014;5(5):e1231.
42. Lee JU, Cheong HS, Shim EY, Bae DJ, Chang HS, Uh ST, Kim YH, Park JS, Lee B, Shin HD, Park CS. Gene profile of fibroblasts identify relation of CCL8 with idiopathic pulmonary fibrosis. Respir Res. 2017;18(1):1–12.

43. Boesch M, Baty F, Brutsche MH, Tamm M, Roux J, Knudsen L, Gazdhar A, Geiser T, Khan P, Hostettler KE. Transcriptomic profiling reveals disease-specific characteristics of epithelial cells in idiopathic pulmonary fibrosis. Respir Res. 2020;21(1):1–9.
44. Guillotin D, Taylor AR, Platé M, Mercer PF, Edwards LM, Haggart R, Miele G, McAnulty RJ, Maher TM, Hynds RE, Jamal-Hanjani M. Transcriptome analysis of IPF fibroblastic foci identifies key pathways involved in fibrogenesis. Thorax. 2021;76(1):73–82.
45. Luzina IG, Salcedo MV, Rojas-Peña ML, Wyman AE, Galvin JR, Sachdeva A, Clerman A, Kim J, Franks TJ, Britt EJ, Hasday JD. Transcriptomic evidence of immune activation in macroscopically normal-appearing and scarred lung tissues in idiopathic pulmonary fibrosis. Cell Immunol. 2018;325:1–13.
46. Checa M, Hagood JS, Velazquez-Cruz R, Ruiz V, Garcia-De-Alba C, Rangel-Escareño C, Urrea F, Becerril C, Montaño M, García-Trejo S, Lira JC. Cigarette smoke enhances the expression of profibrotic molecules in alveolar epithelial cells. PLoS One. 2016;11(3):e0150383.
47. Carraro G, Mulay A, Yao C, Mizuno T, Konda B, Petrov M, Lafkas D, Arron JR, Hogaboam CM, Chen P, Jiang D. Single-cell reconstruction of human basal cell diversity in normal and idiopathic pulmonary fibrosis lungs. Am J Respir Crit Care Med. 2020;202(11):1540–50.
48. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. Arthritis Rheum. 2011;63(3):783–94.
49. Kwapiszewska G, Gungl A, Wilhelm J, Marsh LM, Puthenparampil HT, Sinn K, Didiasova M, Klepetko W, Kosanovic D, Schermuly RT, Wujak L. Transcriptome profiling reveals the complexity of pirfenidone effects in idiopathic pulmonary fibrosis. Eur Respir J. 2018;52(5):1800564.
50. Lee JU, Son JH, Shim EY, Cheong HS, Shin SW, Shin HD, Baek AR, Ryu S, Park CS, Chang HS, Park JS. Global DNA methylation pattern of fibroblasts in idiopathic pulmonary fibrosis. DNA Cell Biol. 2019;38(9):905–14.
51. Nance T, Smith KS, Anaya V, Richardson R, Ho L, Pala M, Mostafavi S, Battle A, Feghali-Bostwick C, Rosen G, Montgomery SB. Transcriptome analysis reveals differential splicing events in IPF lung tissue. PLoS One. 2014;9(3):e92111.
52. Tan J, Tedrow JR, Nouraie M, Dutta JA, Miller DT, Li X, Yu S, Chu Y, Juan-Guardela B, Kaminski N, Ramani K. Loss of Twist1 in the mesenchymal compartment promotes increased fibrosis in experimental lung injury by enhanced expression of CXCL12. J Immunol. 2017;198(6):2269–85.
53. Rodriguez LR, Emblom-Callahan M, Chhina M, Bui S, Aljeburry B, Tran LH, Novak R, Lemma M, Nathan SD, Grant GM. Global gene expression analysis in an in vitro fibroblast model of idiopathic pulmonary fibrosis reveals potential role for CXCL14/CXCR4. Sci Rep. 2018;8(1):1–12.
54. Tzouvelekis A, Ntolios P, Karameris A, Vilaras G, Boglou P, Koulelidis A, Archontogeorgis K, Kaltsas K, Zacharis G, Sarikloglou E, Steiropoulos P. Increased expression of epidermal growth factor receptor (EGF-R) in patients with different forms of lung fibrosis. Biomed Res Int. 2013;2013:654354.
55. Vukmirovic M, Herazo-Maya JD, Blackmon J, Skodric-Trifunovic V, Jovanovic D, Pavlovic S, Stojsic J, Zeljkovic V, Yan X, Homer R, Stefanovic B. Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with idiopathic pulmonary fibrosis. BMC Pulm Med. 2017;17(1):1–12.
56. Senavirathna LK, Huang C, Pushparaj S, Xu D, Liu L. Hypoxia and transforming growth factor β1 regulation of long non-coding RNA transcriptomes in human pulmonary fibroblasts. Physiol Rep. 2020;8(1):e14343.
57. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, Jiang Y, Kass DJ, Gibson K, Chen W, Mora A. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. Eur Respir J. 2019;54(2):1802441.

58. Miyashita N, Horie M, Suzuki HI, Saito M, Mikami Y, Okuda K, Boucher RC, Suzukawa M, Hebisawa A, Saito A, Nagase T. FOXL1 regulates lung fibroblast function via multiple mechanisms. Am J Respir Cell Mol Biol. 2020;63(6):831–42.
59. McDonough JE, Ahangari F, Li Q, Jain S, Verleden SE, Herazo-Maya J, Vukmirovic M, DeIuliis G, Tzouvelekis A, Tanabe N, Chu F. Transcriptional regulatory model of fibrosis progression in the human lung. JCI Insight. 2019;4(22):e131597.
60. Luzina IG, Fishelevich R, Hampton BS, Courneya JP, Parisella FR, Lugkey KN, Baleno FX, Choi D, Kopach P, Lockatell V, Todd NW. Full-length IL-33 regulates Smad3 phosphorylation and gene transcription in a distinctive AP2-dependent manner. Cell Immunol. 2020;357:104203.
61. Hadjicharalambous MR, Roux BT, Csomor E, Feghali-Bostwick CA, Murray LA, Clarke DL, Lindsay MA. Long intergenic non-coding RNAs regulate human lung fibroblast function: implications for idiopathic pulmonary fibrosis. Sci Rep. 2019;9(1):1–15.
62. Beisang DJ, Smith K, Yang L, Benyumov A, Gilbertsen A, Herrera J, Lock E, Racila E, Forster C, Sandri BJ, Henke CA. Single-cell RNA sequencing reveals that lung mesenchymal progenitor cells in IPF exhibit pathological features early in their differentiation trajectory. Sci Rep. 2020;10(1):1–12.
63. Roach KM, Sutcliffe A, Matthews L, Elliott G, Newby C, Amrani Y, Bradding P. A model of human lung fibrogenesis for the assessment of antifibrotic strategies in idiopathic pulmonary fibrosis. Sci Rep. 2018;8(1):1–15.
64. Zhang Y, Jiang M, Nouraie M, Roth MG, Tabib T, Winters S, Chen X, Sembrat J, Chu Y, Cardenes N, Tuder RMM. GDF15 is an epithelial-derived biomarker of idiopathic pulmonary fibrosis. Am J Phys Lung Cell Mol Phys. 2019;317(4):L510–21.
65. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, Chu SG, Raby BA, DeIuliis G, Januszyk M, Duan Q. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Sci Adv. 2020;6(28):eaba1983.
66. Jonigk D, Stark H, Braubach P, Neubert L, Shin HO, Izykowski N, Welte T, Janciauskiene S, Warnecke G, Haverich A, Kuehnel M. Morphological and molecular motifs of fibrosing pulmonary injury patterns. J Pathol Clin Res. 2019;5(4):256–71.
67. Mullenbrock S, Liu F, Szak S, Hronowski X, Gao B, Juhasz P, Sun C, Liu M, McLaughlin H, Xiao Q, Feghali-Bostwick C. Systems analysis of transcriptomic and proteomic profiles identifies novel regulation of fibrotic programs by miRNAs in pulmonary fibrosis fibroblasts. Genes. 2018;9(12):588.
68. Patel NM, Kawut SM, Jelic S, Arcasoy SM, Lederer DJ, Borczuk AC. Pulmonary arteriole gene expression signature in idiopathic pulmonary fibrosis. Eur Respir J. 2013;41(6):1324–30.
69. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, Wikenheiser-Brokamp KA, Perl AKT, Funari VA, Gokey JJ, Stripp BR. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. JCI Insight. 2016;1(20):e90558.
70. Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, Rosen R, Neidermyer AJ, McKean DF, Groshong SD, Cool C. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. Thorax. 2013;68(12):1114–21.
71. Paplińska-Goryca M, Goryca K, Misiukiewicz-Stępień P, Nejman-Gryz P, Proboszcz M, Górska K, Maskey-Warzęchowska M, Krenke R. mRNA expression profile of bronchoalveolar lavage fluid cells from patients with idiopathic pulmonary fibrosis and sarcoidosis. Eur J Clin Investig. 2019;49(9):e13153.
72. Xia Y, Lei C, Yang D, Luo H. Construction and validation of a bronchoalveolar lavage cell-associated gene signature for prognosis prediction in idiopathic pulmonary fibrosis. Int Immunopharmacol. 2021;92:107369.
73. Abe S, Sato S, Aono Y, Azuma M, Kishi M, Koyama K, Takahashi N, Kagawa K, Kawano H, Nishioka Y. Functional analysis of human fibrocytes derived from monocytes reveals their profibrotic phenotype through paracrine effects. J Med Investig. 2020;67(1.2):102–12.

74. Di Mauro S, Scamporrino A, Fruciano M, Filippello A, Fagone E, Gili E, Scionti F, Purrazzo G, Di Pino A, Scicali R, Di Martino MT. Circulating coding and long non-coding RNAs as potential biomarkers of idiopathic pulmonary fibrosis. Int J Mol Sci. 2020;21(22):8812.
75. Nakanishi T, Cerani A, Forgetta V, Zhou S, Allen RJ, Leavy OC, Koido M, Assayag D, Jenkins RG, Wain LV, Yang IV. Genetically increased circulating FUT3 level leads to reduced risk of idiopathic pulmonary fibrosis: a Mendelian randomisation study. Eur Respir J. 2021;59:2003979.
76. Fraser E, Denney L, Antanaviciute A, Blirando K, Vuppusetty C, Zheng Y, Repapi E, Iotchkova V, Taylor S, Ashley N, St Noble V. Multi-modal characterization of monocytes in idiopathic pulmonary fibrosis reveals a primed type I interferon immune phenotype. Front Immunol. 2021;12:226.
77. Sala MA, Balderas-Martínez YI, Buendía-Roldan I, Abdala-Valencia H, Nam K, Jain M, Bhorade S, Bharat A, Reyfman PA, Ridge KM, Pardo A. Inflammatory pathways are upregulated in the nasal epithelium in patients with idiopathic pulmonary fibrosis. Respir Res. 2018;19(1):1–10.
78. Boon K, Bailey NW, Yang J, Steel MP, Groshong S, Kervitsky D, Brown KK, Schwarz MI, Schwartz DA. Molecular phenotypes distinguish patients with relatively stable from progressive idiopathic pulmonary fibrosis (IPF). PLoS One. 2009;4(4):e5134.
79. Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS, Bang TJ, Schwarz MI, Brown KK, Kropski JA, Rojas M. Chronic hypersensitivity pneumonitis, an interstitial lung disease with distinct molecular signatures. Am J Respir Crit Care Med. 2020;202(10):1430–44.
80. De Sadeleer LJ, McDonough JE, Schupp JC, Yan X, Vanstapel A, Van Herck A, Everaerts S, Geudens V, Sacreas A, Goos T, Aelbrecht C. Lung microenvironments and disease progression in fibrotic hypersensitivity pneumonitis. Am J Respir Crit Care Med. 2021;205(1):60–74.
81. Christophi GP, Caza T, Curtiss C, Gumber D, Massa PT, Landas SK. Gene expression profiles in granuloma tissue reveal novel diagnostic markers in sarcoidosis. Exp Mol Pathol. 2014;96(3):393–9.
82. Casanova NG, Gonzalez-Garay ML, Sun B, Bime C, Sun X, Knox KS, Crouser ED, Sammani N, Gonzales T, Natt B, Chaudhary S. Differential transcriptomics in sarcoidosis lung and lymph node granulomas with comparisons to pathogen-specific granulomas. Respir Res. 2020;21(1):1–12.
83. Tanaka H, Yamaguchi E, Asai N, Yokoi T, Nishimura M, Nakao H, Yoneda M, Ohtsuka Y, Konno S, Yamada N. Cathepsin S, a new serum biomarker of sarcoidosis discovered by transcriptome analysis of alveolar macrophages. Sarcoidosis Vasc Diffuse Lung Dis. 2019;36(2):141.
84. Vukmirovic M, Yan X, Gibson KF, Gulati M, Schupp JC, DeIuliis G, Adams TS, Hu B, Mihaljinec A, Woolard TN, Lynn H. Transcriptomics of bronchoalveolar lavage cells identifies new molecular endotypes of sarcoidosis. Eur Respir J. 2021;58(6):2002950.
85. Lepzien R, Nie M, Czarnewski P, Liu S, Yu M, Ravindran A, Kullberg S, Eklund A, Grunewald J, Smed-Sörensen A. Pulmonary and blood dendritic cells from sarcoidosis patients more potently induce IFNγ-producing Th1 cells compared with monocytes. J Leukoc Biol. 2021;2021:1–10.
86. Lepzien R, Liu S, Czarnewski P, Nie M, Österberg B, Baharom F, Pourazar J, Rankin G, Eklund A, Bottai M, Kullberg S. Monocytes in sarcoidosis are potent TNF producers and predict disease outcome. Eur Respir J. 2021;58(1):2003468.
87. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y, Xu Z, Wilkinson KA, Wilkinson RJ, Kendrick Y, Devouassoux G. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. PLoS One. 2013;8(8):e70630.
88. Kachamakova-Trojanowska N, Jazwa-Kusior A, Szade K, Kasper L, Soja J, Andrychiewicz A, Jakiela B, Plutecka H, Sanak M, Jozkowicz A, Sladek K. Molecular profiling of regulatory T cells in pulmonary sarcoidosis. J Autoimmun. 2018;94:56–69.

89. Garman L, Pelikan RC, Rasmussen A, Lareau CA, Savoy KA, Deshmukh US, Bagavant H, Levin AM, Daouk S, Drake WP, Montgomery CG. Single cell transcriptomics implicate novel monocyte and T cell immune dysregulation in sarcoidosis. Front Immunol. 2020;11:567342.
90. Locke LW, Crouser ED, White P, Julian MW, Caceres EG, Papp AC, Le VT, Sadee W, Schlesinger LS. IL-13–regulated macrophage polarization during granuloma formation in an in vitro human sarcoidosis model. Am J Respir Cell Mol Biol. 2019;60(1):84–95.
91. Jung SM, Park KS, Kim KJ. Integrative analysis of lung molecular signatures reveals key drivers of systemic sclerosis-associated interstitial lung disease. Ann Rheum Dis. 2021;81(1):108–16.
92. Valenzi E, Bulik M, Tabib T, Morse C, Sembrat J, Bittar HT, Rojas M, Lafyatis R. Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. Ann Rheum Dis. 2019;78(10):1379–87.
93. Valenzi E, Tabib T, Papazoglou A, Sembrat J, Trejo Bittar HE, Rojas M, Lafyatis R. Disparate interferon signaling and shared aberrant basaloid cells in single-cell profiling of idiopathic pulmonary fibrosis and systemic sclerosis-associated interstitial lung disease. Front Immunol. 2021;12:960.
94. Lindahl GE, Stock CJ, Shi-Wen X, Leoni P, Sestini P, Howat SL, Bou-Gharios G, Nicholson AG, Denton CP, Grutters JC, Maher TM. Microarray profiling reveals suppressed interferon stimulated gene program in fibroblasts from scleroderma-associated interstitial lung disease. Respir Res. 2013;14(1):1–14.
95. Renaud L, da Silveira WA, Takamura N, Hardiman G, Feghali-Bostwick C. Prominence of IL6, IGF, TLR, and bioenergetics pathway perturbation in lung tissues of scleroderma patients with pulmonary fibrosis. Front Immunol. 2020;11:383.
96. Assassi S, Wu M, Tan FK, Chang J, Graham TA, Furst DE, Khanna D, Charles J, Ferguson EC, Feghali-Bostwick C, Mayes MD. Skin gene expression correlates of severity of interstitial lung disease in systemic sclerosis. Arthritis Rheum. 2013;65(11):2917–27.
97. Hutter C, Kauer M, Simonitsch-Klupp I, Jug G, Schwentner R, Leitner J, Bock P, Steinberger P, Bauer W, Carlesso N, Minkov M. Notch is active in Langerhans cell histiocytosis and confers pathognomonic features on dendritic cells. Blood. 2012;120(26):5199–208.
98. Pang J, Luo Y, Wei D, Cao Z, Qi X, Song M, Liu Y, Li Z, Zhang J, Li B, Chen J. Comparative transcriptome analyses reveal a transcriptional landscape of human silicosis lungs and provide potential strategies for silicosis treatment. Front Genet. 2021;12:652901.
99. Yao W, Yang P, Qi Y, Jin L, Zhao A, Ding M, Wang D, Li Y, Hao C. Transcriptome analysis reveals a protective role of liver X receptor alpha against silica particle-induced experimental silicosis. Sci Total Environ. 2020;747:141531.
100. Chen J, Zhang R, Xie M, Luan C, Li X. Transcriptome sequencing identifies PLAUR as an important player in patients with dermatomycitis-associated interstitial lung disease. Front Genet. 2021;12:784215.
101. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17(7):451–9.
102. Chambers DC, Carew AM, Lukowski SW, Powell JE. Transcriptomics and single-cell RNA-sequencing. Respirology. 2019;24(1):29–36.
103. Junker BH, Klukas C, Schreiber F. VANTED: a system for advanced data analysis and visualization in the context of biological networks. BMC Bioinform. 2006;7(1):1–13.
104. Kamburov A, Cavill R, Ebbels TM, Herwig R, Keun HC. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. Bioinformatics. 2011;27(20):2917–8.
105. Hu Z, Hung JH, Wang Y, Chang YC, Huang CL, Huyck M, DeLisi C. VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. Nucleic Acids Res. 2009;37(Web Server issue):W115–21.
106. Karnovsky A, Weymouth T, Hull T, Tarcea VG, Scardoni G, Laudanna C, Sartor MA, Stringer KA, Jagadish HV, Burant C, Athey B. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. Bioinformatics. 2012;28(3):373–80.

107. Pavlopoulos GA, O'Donoghue SI, Satagopam VP, Soldatos TG, Pafilis E, Schneider R. Arena3D: visualization of biological networks in 3D. BMC Syst Biol. 2008;2(1):1–7.
108. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
109. McGuffin MJ, Jurisica I. Interaction techniques for selecting and manipulating subgraphs in network visualizations. IEEE Trans Vis Comput Graph. 2009;15(6):937–44.
110. Barsky A, Munzner T, Gardy J, Kincaid R. Cerebral: visualizing multiple experimental conditions on a graph with biological context. IEEE Trans Vis Comput Graph. 2008;14(6):1253–60.
111. Arakawa K, Kono N, Yamada Y, Mori H, Tomita M. KEGG-based pathway visualization tool for complex omics data. In Silico Biol. 2005;5(4):419–23.
112. García-Alcalde F, García-López F, Dopazo J, Conesa A. Paintomics: a web based tool for the joint visualization of transcriptomics and metabolomics data. Bioinformatics. 2011;27(1):137–9.
113. Neuweger H, Persicke M, Albaum SP, Bekel T, Dondrup M, Hüser AT, Winnebald J, Schneider J, Kalinowski J, Goesmann A. Visualizing post genomics datasets on customized pathway maps by ProMeTra–aeration-dependent gene expression and metabolism of Corynebacterium glutamicum as an example. BMC Syst Biol. 2009;3(1):1–14.
114. Tokimatsu T, Sakurai N, Suzuki H, Ohta H, Nishitani K, Koyama T, Umezawa T, Misawa N, Saito K, Shibata D. KaPPA-view. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. Plant Physiol. 2005;138(3):1289–300.
115. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 2004;37(6):914–39.
116. Symonsy S, Zipplies C, Battke F, Nieselt K. Integrative systems biology visualization with MAYDAY. J Integr Bioinform. 2010;7(3):1–14.
117. Lüdemann A, Weicht D, Selbig J, Kopka J. PaVESy: pathway visualization and editing system. Bioinformatics. 2004;20(16):2841–4.
118. Xie N, Tan Z, Banerjee S, Cui H, Ge J, Liu RM, Bernard K, Thannickal VJ, Liu G. Glycolytic reprogramming in myofibroblast differentiation and lung fibrosis. Am J Respir Crit Care Med. 2015;192(12):1462–74.
119. Hsu HS, Liu CC, Lin JH, Hsu TW, Hsu JW, Su K, Hung SC. Involvement of ER stress, PI3K/AKT activation, and lung fibroblast proliferation in bleomycin-induced pulmonary fibrosis. Sci Rep. 2017;7(1):1–11.
120. Zhao H, Li C, Li L, Liu J, Gao Y, Mu K, Chen D, Lu A, Ren Y, Li Z. Baicalin alleviates bleomycin induced pulmonary fibrosis and fibroblast proliferation in rats via the PI3K/AKT signaling pathway. Mol Med Rep. 2020;21(6):2321–34.

Transferring Metabolomics to Portable Diagnostic Devices: Trending in Biosensors

Shimaa Eissa

Abstract Metabolic biomarkers are very popular in clinical laboratories, where more than 95% of the clinical assays are based on small molecules. Biosensors can offer a faster, simpler, and cheaper alternative to conventional analytical assays as they can make the metabolic biomarkers more accessible in a high-throughput fashion for point-of-care testing. As research continues in different aspects of the biosensor design, more sensitive and selective biosensors are being developed for various metabolic biomarkers for diagnostic applications. In this chapter, we provide a brief overview of the current biosensor designs and formats for applications in metabolomics. Various biorecognition receptors and transducers used in the development of biosensors for metabolic biomarkers are discussed. Major advances in the biosensors for metabolites such as the use of aptamers as new recognition receptors as well as the utilization of nanomaterials as transducers are highlighted. The developments of multiplexed array biosensors for the simultaneous detection of multiple biomarkers and wearable biosensors are discussed as emerging diagnostic tools in metabolomics. The challenges and future perspectives in the use of biosensors in metabolomics are discussed.

Keywords Biosensing platforms · Metabolic biomarkers · Point-of-care biosensors · Metabolomics · Electrochemical biosensors · Optical biosensors · Multiplexed array biosensors · Detection

---

S. Eissa (✉)
Department of Chemistry, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Advanced Materials Chemistry Center (AMCC), Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates
e-mail: shimaa.eissa@ku.ac.ae

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. M. Abdel Rahman (ed.), *Clinical Metabolomics Applications in Genetic Diseases*, https://doi.org/10.1007/978-981-99-5162-8_15

Abbreviations

Aptasensors	Aptamer-based biosensors
HQ	Hydroquinone
MS	Mass spectrometry
NMR	High-resolution nuclear magnetic resonance spectroscopy
POCT	Point-of-care testing
SELEX	Systematic evolution of ligand by exponential enrichment
TMB/H_2O_2	Tetramethylbenzidine/hydrogen peroxide

1 Introduction

The development of fast, accurate, and reliable diagnostic methods is crucial to improving the clinical course of any disease. Metabolomics offers new opportunities to discover biomarkers associated with complex diseases [1]. Detecting these low molecular weight metabolic biomarkers in biological samples can provide useful information about the pathology and the progress of the disease [2]. Moreover, it can provide insight into the response of the patients to specific treatment and helps to understand the mechanism of the disease. Therefore, metabolomics has become a powerful tool in clinical research for discovering biomarkers in disease diagnosis [1]. With the recent advancements in the bionanotechnology field, various promising metabolomic technologies are being developed to identify new therapeutic targets and improve the disease prognosis and diagnosis [3]. Advances in bioinformatics, mass spectrometry (MS), high-resolution nuclear magnetic resonance spectroscopy (NMR), and ultra-performance liquid chromatography methods have led to a drastic improvement in the reliability and efficiency of metabolic profiling [4]. Moreover, the coupling of mass spectrometry with chromatography has enabled more comprehensive coverage of metabolic biomarkers [4].

Biosensors are analytical devices that gained significant interest over the last decades as it offers a lower cost, simpler, faster, and potentially more versatile alternative to traditional analytical techniques. Biosensing devices are of high importance in metabolomic applications due to their potential to provide simple and noninvasive measurements of various biomarkers in biofluids such as saliva, blood, sweat, tears, and urine. There are various types of biosensors which can be used for different applications. A typical biosensor contains three main units: a biorecognition receptor such as antibody, enzyme, or aptamer which is responsible for the specific binding with the target analyte; a transducer such as optical, electrochemical, or mass-based detection techniques which can transform the binding event between the recognition receptor and the analyte into a measurable signal; and a signal detector which reports and displays the biosensor signal. The potential utility of biosensors in metabolomics is evident from the rapidly increasing reported biosensing platforms for various metabolic biomarkers for point-of-care testing. In the next section, the different recognition receptors and transducers used in the biosensors for metabolites are discussed. Moreover, the integration of microfluidics with

array multiplexed biosensors as well as flexible materials for wearable biosensors as promising improvements in the biosensors for metabolites is highlighted.

2 Biorecognition Receptors Used in the Biosensors for Metabolic Biomarkers

The selection of the proper bioreceptor is a major element in the biosensor design. The choice of the biorecognition receptor used to fabricate the biosensor depends mainly on the type of target analyte under investigation. Antibodies are mainly used for the detection of proteins, whereas ssDNA can be used for the detection of genomic sequences by hybridization. However, some antibodies have been also produced against some metabolites. Enzymes are widely used as bioreceptors in biosensors for some small molecules where the enzyme is used to catalyze specific biochemical reaction. However, since low molecular weight metabolites can have various structural and chemical properties, the choice of specific recognition receptors for each individual metabolite can be challenging. Various suitable bioreceptors are used for different types of metabolites. Specific enzymes for some metabolites like glucose, lactate, and uric acid have been used for the development of biosensors. Antibodies for other metabolites have been also developed and used for the fabrication of immunosensors such as hormones. Aptamers have appeared in the recent years as potential alternative to antibodies for the biosensor's development, particularly for small molecules. Therefore, aptamers hold considerable promise in the metabolomic field because of their low cost, easier in vitro synthesis, and high stability. In the next subsections, we will discuss various bioreceptors, which have been used for the development of biosensing platforms for the detection of different metabolites.

2.1 Enzyme-Based Biosensors

Enzymes are selective bioreceptors that are mainly used to develop catalytic biosensors for certain target analytes. The enzyme catalyzes a reaction leading to the formation or disappearance of an electroactive product which can be detected using an electrochemical technique such as amperometry [5, 6]. Enzyme-based biosensors are easy to construct and can provide sensitive and rapid analysis. However, there are no available specific enzymes for many target molecules. Moreover, the enzyme stability is limited as it gradually loses activity over time, and thus, the shelf lifetime of the enzyme-based biosensors is short.

Monitoring blood glucose, one of the most common metabolic biomarkers, was the main driving force for the research work to develop enzymatic biosensors for medical applications [5, 7, 8]. Blood glucose has been established as a biomarker for diagnosis of diabetes. The development of biosensors to detect glucose level has started since almost 60 years [8]. Significant effort has been made to improve the

glucose biosensor technology in terms of sensitivity, reliability, stability, and portability. The historical development of glucose biosensors has been described in detail in several reviews [5, 7–9]. In summary, glucose biosensors are divided into three generations. The first-generation enzyme biosensors were oxygen-based, whereas the second-generation are mediator-based. However, the third-generation glucose biosensors are the directly coupled enzyme electrodes. The continuous emergences of new nanostructures and nanocomposites have led to the developments of various glucose biosensors with improved electron transfer efficiency and electrocatalytic activity [10].

Similarly, other enzyme-based biosensors for the detection of different metabolites have been later developed such as lactate [11, 12], xanthine [13, 14], caffeine [13], glycolic acid [15], bile acids [16, 17], L-arginine [18], choline [19], and biosensors. Lactate is produced when glucose is broken down and is commonly used as avital biomarker in medical monitoring [20]. The concentration of lactate in blood often rises during exercises such as running; however, it can be also altered due to hemorrhage, trauma, and ischemia. Lactate is also used as a biomarker for conditions such as bacterial meningitis or acidosis [21]. Various enzymes have been used as biorecognition receptors in lactate biosensors such as lactate monooxidase, lactate oxidase, lactate dehydrogenase, and cytochrome b [6]. Mediators such as NAD+/NADH and ferricyanide are sometimes used in these enzymatic electrochemical biosensors leading to the production of a current after applying certain potential that is measured amperometrically [6]. The uric acid is a metabolite used as indicator of gout which is produced by the breaking down of purine nucleotides with xanthine oxidase enzyme. The detection of xanthine is also of increasing medical interest. Thus, several amperometric biosensors using xanthine oxidase have been developed for the detection of xanthine [14].

The detection of choline is important in clinical practice, especially in the early diagnosis of some brain disorders such as Parkinson's and Alzheimer's diseases [22]. Many enzyme-based amperometric biosensors for choline have been developed. In these biosensors, choline oxidase was used, and the detection was based on amperometric monitoring of hydrogen peroxide produced when choline oxidase catalyzes the reaction. Hydrogen peroxide reacts either directly with the redox mediator [23–26], or its reaction is catalyzed by a second enzyme such as horseradish peroxidase to enhance the current signal [27, 28]. Different nanomaterials have been integrated in these enzyme-based biosensors for choline such as carbon nanotubes [19, 26, 29], gold nanoparticles, and graphene [27].

2.2 *Antibody-Based Biosensors*

The production of antibodies against low molecular weight compounds is usually challenging. Because of their small size, most metabolites are not usually immunogenic and need to be conjugated with larger carrier protein before using in immunizing the animals. Moreover, most of the antibodies for small molecules suffer from low specificity issues. However, some successful examples of antibodies produced

for metabolites have been reported. Antibodies against various hormones such as progesterone, testosterone, cortisol, 11-deoxycortisol, and diethylstilbestrol have been produced. Table 1 shows the reported biosensors for metabolic biomarker which used antibody as recognition receptor. Serafín et al. [30] have reported the development of a biosensor based on an anti-progesterone antibody for the detection of progesterone in saliva (Fig. 1). A competitive amperometric biosensor was fabricated on a low-cost disposable electrode which allowed a fast (45 min) and sensitive detection of progesterone with a LOD of 5 pg/mL. Good selectivity against other hormones such as testosterone, corticosterone, cortisol, and 17-β-ethynylestradiol was shown with comparable results to commercial ELISA. Other studies have shown the use of anti-progesterone antibody for the development of electrochemical immunosensors utilizing a thionine/graphene oxide composite achieving LOD of 6.3 pg/mL [31] and gold nanoparticles yielding LODs of 430 pg/mL [32] and 80 pg/mL [33].

Several biosensors have utilized antibody against cortisol for the development of biosensors to measure the concentration of cortisol in buffer [34–37], artificial saliva [38], real human saliva [39, 40], and interstitial fluid [39, 41]. Different approaches were used in these immunosensors to minimize the matrix effect especially when cortisol was detected in human saliva such as dilution [39] or using fluid control system [42].

Kämäräinen et al. [40] have utilized anti-cortisol antibody to develop a competitive electrochemical disposable immunosensor for the detection of cortisol in human saliva using cortisol-alkaline phosphatase conjugate and screen-printed electrodes showing good sensitivity, reproducibility, and repeatability. Moreover, the results of the cortisol immunosensor were comparable with ultra-high pressure liquid chromatography-tandem mass spectrometry.

Many studies have shown the integration of antibody against estradiol in several immunosensors for the detection of estradiol [43, 44]. The reported immunosensors have mainly utilized a competitive assay where a protein-estradiol conjugate was employed to compete with the free estradiol molecules on the sample for the antibody immobilized on the sensor surface. Enzymes such as alkaline phosphatase [43, 44] and horse radish peroxidase [45, 46] or bovine serum albumin [47] were conjugates with estradiol; in some of these studies, a competitive immunosensor was developed. A non-labeled competitive immunosensor for estradiol has been also reported using hydroquinone as redox marker and differential pulse voltammetry for the detection [48]. Ojeda et al. [46] have described the integration of an antibody in a competitive electrochemical immunosensor for estradiol hormone. The immunosensor was fabricated on a screen-printed electrode modified with streptavidin on which a biotinylated anti-estradiol was immobilized. A competitive assay was performed using horse radish peroxidase-labeled estradiol, and the detection was achieved amperometrically employing hydroquinone as redox mediator. The immunosensor exhibited good sensitivity with a LOD of 0.77 pg/mL as well as good selectivity against other hormones. Moreover, this estradiol immunosensor showed good applicability in spiked serum and urine samples.

Antibody for testosterone was produced and utilized in the fabrication of some immunosensors. Eguílaz et al. [49] have reported the development of an amperometry immunosensor based on disposable screen-printed carbon electrodes and

Table 1 Immunosensors for different metabolites

Target metabolite	Material	Transducer	Limit of detection	Reference
Progesterone	Magnetic microbeads	Amperometry	5 pg/mL	[30]
Progesterone	Thionine-graphene Oxide composites	Amperometry	6.3 pg/mL	[31]
Progesterone	Colloidal gold-graphite-Teflon	Amperometry	430 pg/mL	[32]
Progesterone	Gold nanoparticles	Amperometry	80 pg/mL	[33]
Cortisol	Dithiobis (succinimidyl propionate)-modified gold microarray	Electrochemical impedance spectroscopy	1 pM	[34]
Cortisol	Polyaniline-protected gold nanoparticles	Cyclic voltammetry	1 pM	[37]
Cortisol	Low temperature co-fired ceramic (LTCC)-based microfluidic system	Cyclic voltammetry	10 pM	[36]
Cortisol	Single-walled, carbon nanotube	Chemiresistor	1 pg/mL	[38]
Cortisol	Graphite	Square wave voltammetry	1.7 ng/mL	[40]
Estradiol	Carbon	Amperometry	50 pg/mL	[44]
Estradiol	Gold nanoparticles	Amperometry	6 pg/mL	[45]
Estradiol	Carbon	Amperometry	0.77 pg/mL	[46]
Estradiol	Gold nanoparticle thiolated protein G-scaffold	Square wave voltammetry and electrochemical impedance spectroscopy	18 pg/mL and 26 pg/mL	[47]
Testosterone	Magnetic beads	Amperometry	1.7 pg/mL	[49]
Testosterone	Carbon	Amperometry	26 pg/mL and 1.8 pg/mL in buffer and urine	[50]
Vitamin D_3	Cellulose acetate fibers	Amperometry	10 ng/mL	[54]
Vitamin D_3	Magnetite nanoparticles incorporated into electrospun polyacrylonitrile nanofibers	Differential pulse voltammetry	0.12 ng/mL	[55]
Vitamin D_3	Gold-platinum bimetallic nanoparticle-coated 3 -(aminopropyl) triethoxysilane	Differential pulse voltammetry	4.9 pg/mL	[56]
Vitamin D_3	Gold nanoparticles	Surface plasmon Resonance	1 μg/mL	[57]
Vitamin D_3	Nanostructured cerium (IV) oxide (nCeO2)	Differential pulse voltammetry	4.63 ng/mL	[58]

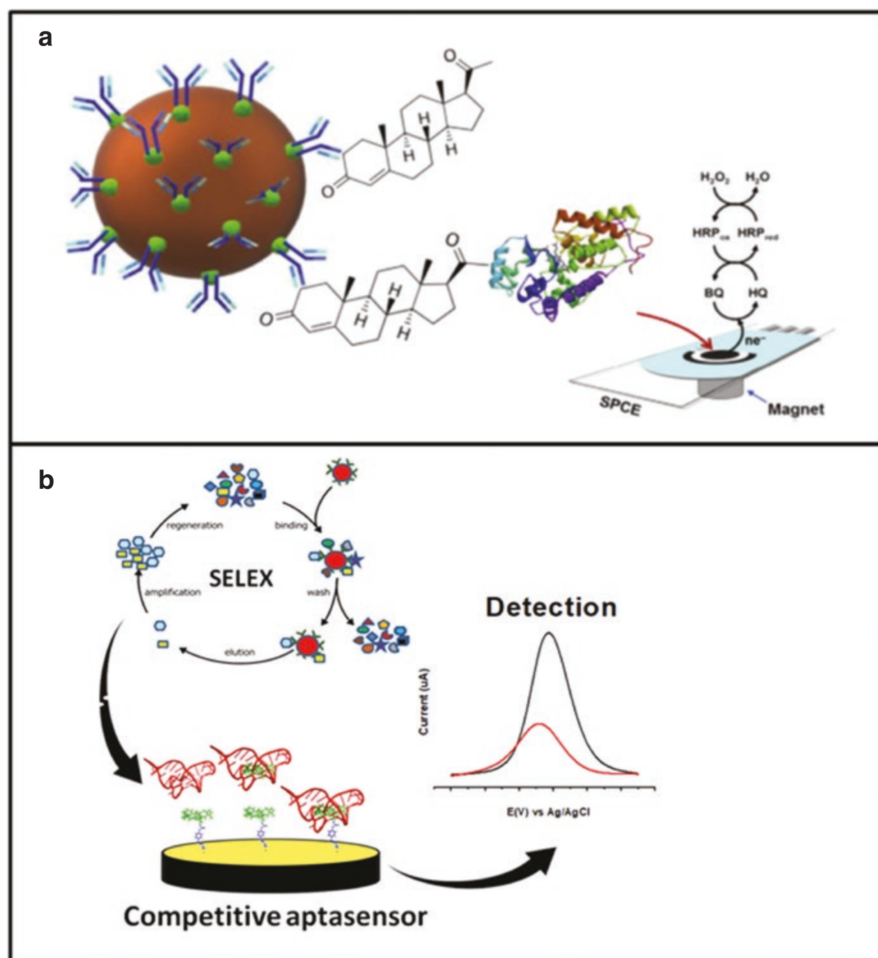

Fig. 1 (**a**) Scheme of the design of progesterone amperometric immunosensor. (Reprinted with permission from Ref. [30]). (**b**) Schematic illustration of the selection process of sepiapterin aptamer and the working principle of the competitive biosensor using square wave voltammetry detection. (Reprinted with permission from Ref. [87])

protein A-functionalized magnetic beads. The antibody was immobilized onto the magnetic beads, and competitive assay was performed using HRP-testosterone conjugate. The amperometric detection was achieved after the addition of H_2O_2 using hydroquinone (HQ) as redox mediator showing a LOD of 1.7 pg/mL. The immunosensor showed good selectivity against other steroid hormones and was successfully applied for the detection of testosterone in spiked human serum samples. A recombinant Fab fragment was also used for the fabrication on another electrochemical immunosensor using screen printed electrode and was applied for the detection of testosterone in bovine urine [50]. A competitive assay was utilized using

HRP-labeled testosterone and tetramethylbenzidine/hydrogen peroxide (TMB/H_2O_2) substrate for the signal development. The detection was realized using chronoamperometry at +100 mV.

Vitamin D_3 is one of the metabolic biomarkers that has significant physiological functions. Vitamin D_3 deficiency is considered a global health issue because of its serious health impact and correlation with various diseases [51] such as cardiovascular diseases, bone disorder, diabetes, infections, tuberculosis, osteoarthritis, hypertension, cancer, and even COVID-19 [52, 53]. Thus, antibodies for vitamin D_3 have been produced and employed for the fabrication of various immunosensors to monitor vitamin D_3 level [54–58]. Chauhan et al. [54] have described the fabrication of a low-cost and eco-friendly immunosensor for the detection of 25-hydroxy vitamin-D_3 using disposable conducting paper substrate decorated with electrospun cellulose acetate fibers. The detection was realized using a chronoamperometric technique showing a LOD of 10.0 ng/mL. Moreover, the immunosensor was applied for the detection of vitamin D3 in serum samples exhibiting good agreement with the results obtained from ELISA. The same research group has also reported that the development of another electrochemical immunosensor for vitamin D3 detection using magnetite nanoparticles incorporated electrospun polyacrylonitrile nanofibers. The immunosensor has shown a LOD of 0.12 ng/mL [55]. Kaur et al. [56] have also developed an electrochemical immunosensor for the detection of vitamin D3 using gold-platinum bimetallic nanoparticle-coated 3-(aminopropyl)triethoxysilane on fluorine tin oxide glass electrode. The vitamin D3 antibody was attached to the electrode covalently via glutaraldehyde as a cross linker. The immunosensor showed good sensitivity with a LOD of 0.49 pg/mL. Carlucci et al. [57] have reported the detection of vitamin D3 using both electrochemical and surface plasmon resonance immunosensors showing LODs of 10 ng/mL and 45 ng/mL, respectively. Recently, Chauhan et al. [58] have described the fabrication of carbon cloth-based immunosensor for detection of 25-hydroxy vitamin D_3 using a label-free format. The immunosensor was prepared by depositing nanostructured cerium (IV) oxide on carbon cloth followed by the immobilization of anti-vitamin D_3 antibodies. The immunosensor exhibited good sensitivity with a LOD of 4.63 ng/mL and a fast response time of 15 min. The immunosensor was successfully applied in real serum samples and demonstrated good agreement with the conventional ELISA [58].

2.3 Aptamer-Based Biosensors

Aptamers are single-stranded DNA or RNA which were firstly reported in 1990 [59] and widely considered as promising alternative to the gold standard antibodies in biosensing applications [60–62]. Many aptamers have been identified and tested against various analytes in the last two decades. However, only few aptamers have reached the commercialization stage for therapeutic and diagnostic applications. Aptamers have been widely identified and successfully applied for several

diagnostic applications showing great promise for point-of-care testing (POCT) [63–67]. Aptamers offers several advantages over antibodies such as their high stability, ease of in vitro production, and simplicity of synthesis at very low cost. These remarkable advantages, along with the ease of their immobilization and regeneration as well as their small size, have made them excellent candidates for the development of aptamer-based biosensors (aptasensors). Since metabolites are small molecule health markers, their detection by conventional antibody-based methods can be challenging. Thus, the identification of aptamers against metabolites for health-related and diagnostic applications has recently received considerable attention. Aptamers are usually identified through a process called systematic evolution of ligand by exponential enrichment (SELEX) [68]. The selection is realized by exposing nucleic acid library which consists of random sequences with the target analyte. Then, a partitioning step is performed to separate the bound from the unbound sequences followed by amplification of the bound sequences using a polymerase chain reaction. The amplified DNA is then purified and used to start a new SELEX cycle. The selection cycles are repeated from 10 to 20 times until the DNA pool is enriched with the highest affinity binders to the target.

Several aptamers have been selected against hormones such as estradiol [69, 70], progesterone [71, 72], cortisol [73, 74], thyroxine [75], 11-deoxycortisol [76], testosterone [77], and vasopressin [78, 79]. These aptamers have been exploited for the development of several biosensors for point-of-care diagnostic applications. The increase of progesterone levels can lead to several health issues. Therefore, the detection of progesterone in clinical samples is very important to protect the public health. Jiménez et al. [71] have described the selection, identification, and characterization of high binding affinity DNA aptamers against progesterone using in vitro selection. Electrochemical impedance spectroscopy and fluorometric assays were utilized to determine the dissociation constants. The highest affinity aptamer has shown dissociation constant of 17 nM without any significant cross-reactivity to similar analogues such as 11-norethisterone and 17β-estradiol. The aptamer was then immobilized on gold electrode, and a complementary short sequence was hybridized to the aptamer at different sites to optimize the signal gain of the aptasensor. The detection relied on the conformational change of the aptamer upon binding with the analyte as confirmed by circular dichroism spectroscopy. The aptasensor has shown excellent sensitivity with a LOD of 0.90 ng/mL. This aptamer has been optimized in another study [72] and used to fabricate a fluorescence-based biosensor for the detection of progesterone. The aptamer was truncated in this study to eliminate the nonbinding region which had negative impact on the binding affinity of the aptamer to progesterone. Fluorescence mapping was performed to enhance the affinity and the specificity of the aptamer with 16-fold increase in the dissociation constant compared to the original aptamer. The truncated aptamer was then used in a displacement fluorescence assay where it was hybridized at different sites to fluorescein- and quencher-labelled complementary DNA sequences to form duplex structures. The detection of progesterone was realized via displacement of the complementary sequence which causes enhancement of the fluoresce signal.

Despite of the numerous advantages of aptamers as bioreceptors, the selection of aptamers against various metabolites is still limited mainly due to the tediousness and high cost of the traditional SELEX process. Eissa et al. [76] have reported a new method for the selection of aptamers against 11-deoxycortisol hormone. The selection was based on an electrochemical method which enabled a cost-effective, efficient, and rapid enrichment process. In this method, the hormone molecules were immobilized on a gold electrode which was then used as the solid support for the SELEX method. The enrichment of the DNA during the selection process was monitored using square wave voltammetry in a label-free format unlike the conventional SELEX protocol for small molecules which often requires the use of fluorescently labeled DNA. High-affinity aptamers against 11-deoxycortisol hormone were successfully selected after eight cycles showing dissociation constants at the subnanomolar level. The selected aptamer was utilized to fabricate an electrochemical biosensor for the detection of 11-deoxycortisol showing very high sensitivity. This aptasensor has shown good selectivity and successful supplication in spiked serum samples.

Akki et al. [70] have reported the selection of DNA aptamers against the endocrine-disrupting compounds: 17β-estradiol and 17α-ethynylestradiol using SELEX. The selected aptamers have demonstrated good affinity with dissociation constants of 0.6 and 0.5 μM for 17β-estradiol and 17α-ethynylestradiol, respectively. The authors found that the selected aptamer against 17β-estradiol has shown good specificity against 17α-ethynylestradiol. Similarly, one of the selected aptamers against 17α-ethynylestradiol showed good specificity against 17β-estradiol and the similar analogue, estrone. It is very important to study the selectivity when selecting aptamers against small molecules and to evaluate to which extent the aptamer binds to other structurally similar compounds.

The stress hormone, cortisol, is one of the most important metabolites synthesized by the stimulation of adrenal cortex upon stress. Cortisol has immunosuppressive and anti-inflammatory effects as well as a strong impact on blood pressure, heart rate, and reproductive and digestive activities. A selection method for an aptamer against the stress biomarker cortisol based on tunable stringency magnetic bead was described by Martin et al. [74]. After 15 rounds of selection, the enriched pool showed a single sequence with high copy number. A next-generation sequencing analysis indicated a correlation between the number of aptamer copies and enhanced affinity to the target under certain conditions. Two methods were used for the estimation. The highest-affinity aptamer has shown dissociation constant of 6.9 and 16.1 μM by equilibrium and microscale thermophoresis methods, respectively. A gold nanoparticle assay incorporating the selected aptamer was performed showing good discrimination between cortisol and other structurally related biomarkers: epinephrine, norepinephrine, and cholic acid.

In another study, an in silico approach of molecular docking was used to investigate the interactions between aptamers and cortisol [73]. The tertiary conformational structures of ten aptamers were studied against cortisol and other related hormones. It was found that the hydrophobic interactions of cortisol with the

aptamer have the major impact on the binding compared to the hydrogen bonding according to the docking results.

RNA has the capability to bind to a wide range of small molecules which can be exploited in many applications in molecular biology. Riboswitches are examples of RNA domains which can serve as bioreceptors that bind to specific metabolite. Despite that most of the aptamers selected against metabolic biomarkers are DNA based, some RNA aptamers have been also reported. LEVESQUE et al. [75] have reported the identification of thyroxine-specific aptamers using SELEX showing good affinity and selectivity for thyroxine over its inactive derivative, thyronine. Transcripts with site-specific modified nucleotides, mutational studies, circular dichroism, and binding shift assays were used to investigate the binding of the aptamer to thyroxine. This study suggested that the iodine moiety in the thyroxine molecule is the main binding site to the aptamer.

The selection of an aptamer against the main sex hormone, testosterone, has been reported [77]. Testosterone is responsible for the regulation of various physiological processes in males such as growth of skeletal muscles and bones and other male sex characteristics. Thus, the accurate detection of testosterone levels in biological fluids is highly important. The selection of testosterone aptamer has been conducted using classical SELEX via immobilizing the target on magnetic beads and performing counterselections against other steroids with similar chemical structures [77]. Ten aptamer sequences were identified using next-generation sequencing showing dissociation constants in the nanomolar range. The conformational change of the aptamers upon binding with testosterone was studied using circular dichroism.

An aptamer has been selected against the nine-amino acid peptide hormone, vasopressin [78–80]. This hormone is considered a biomarker in patients with hemorrhagic shocks as it plays an important role in enhancing peripheral vascular resistance which leads to an increase in arterial blood pressure. The applicability of aptamers in biological fluids especially blood is often limited by their instability due to the presence of nucleases. However, the selection of single-stranded DNA stable and nuclease-resistant aptamer against vasopressin has been reported [78–80]. The aptamer has been also used to develop biosensors for the detection of vasopressin using different designs and transducers [78, 81]. Williams et al. [79] have reported the selection of an aptamer for the enantiomer of vasopressin to identify a mirror-image DNA aptamer (enantiomer) that binds with vasopressin with high stability to nucleases. The enantiomer of the aptamer was synthesized and showed high stability and good bioactivity as vasopressin antagonist in cell culture.

Graphene oxide-based SELEX has been used to select aptamer against 25-hydroxy vitamin D_3 [82]. This immobilization-free method has led to the identification of high-affinity and specificity aptamers. The affinity of the aptamers was investigated using both isothermal titration and gold nanoparticle-based colorimetric assay. From the selected aptamer pool, 9 sequences showed good affinity out of 16 aptamer candidates. The aptamer which showed the highest binding affinity to 25-hydroxy vitamin D_3 with a dissociation constant of 11 nM has been utilized to develop gold nanoparticle-based colorimetric biosensor showing a LOD of 1 μM. The conformation change of the aptamer upon binding with the target was

also confirmed using circular dichroism analysis as well as by utilizing a displacement assay with both magnetic beads and streptavidin-coated 96-well plates.

Many metabolic biomarkers are used for the diagnosis of genetic diseases. Among them, sepiapterin level has been used as indicator of a rare inborn genetic error of neurotransmitter metabolism called sepiapterin reductase deficiency [83–85]. This genetic disease is characterized by cognitive and motor abnormalities. The early diagnosis of such neurotransmitter diseases is crucial to enhance the treatment and avoid the progression of the disease [86]. The detection of sepiapterin is challenging because there is no available specific antibody for sepiapterin in market until now. We have reported recently the selection and identification of DNA aptamers which binds specifically to sepiapterin using SELEX [87] (Fig. 1). Few aptamers have been identified exhibiting high affinity and specificity with dissociation constants in the nanomolar range. The aptamer with the highest affinity to sepiapterin was used to fabricate a competitive electrochemical biosensor. The detection was achieved via competition of the free sepiapterin in the sample with immobilized analyte on gold electrode for binding to the free aptamer. Square wave voltammetry technique was used for the detection showing high sensitivity and selectivity against closely related molecules.

Metabolomics has also led to the discovery of biomarkers associated with asthma pathogenesis in serum. Among these biomarkers, it was reported that a decreased level of arginine can be used as indicator of asthma [88]. Yuan et al. [89] have reported the development of an aptamer-based biosensor for the chiral recognition of arginine enantiomers. The biosensor was based on fluorescence detection using gold nanoparticles on which fluorophore-labeled aptamers were immobilized

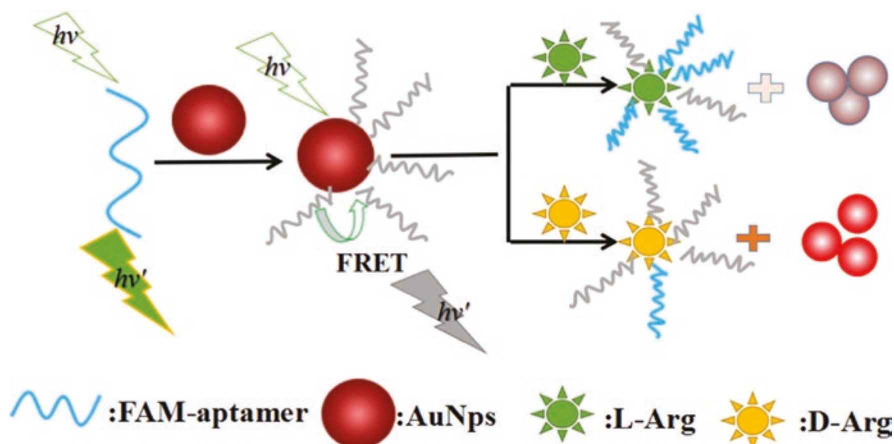

Fig. 2 A schematic of fluorescence-based biosensor for the detection of D, L- arginine. The fluorescence of the labeled aptamers was quenched when the aptamer adsorbs on gold nanoparticles due to the fluorescence resonance energy transfer. The binding of the aptamer with D, L- arginine leads to desorption of the aptamers from the gold surface and thus enhancement in the fluorescence intensity. (Reprinted with permission from [89])

(Fig. 2). The detection was achieved by following the increase in the fluorescence intensity when the aptamers bind to the targets leading to the release of the fluorescence label from the gold nanoparticles.

3 Transducers for Detection of Metabolic Biomarkers

Many different transducers are used for the development of various biosensors. The most popular transducers used in most of the biosensor's designs are the optical, electrochemical, thermal, and mass-based transducers. Each of these detection techniques has its own advantages and disadvantages and can be better suited for the detection of different target analytes.

Several biosensors have been reported for the detection of metabolic biomarkers for the point-of-care diagnosis of diseases [16, 35, 41, 67, 76, 90]. Considerable research effort has been particularly devoted toward the fabrication of electrochemical and optical biosensors for the detection of different types of metabolites. The integration of various nanomaterials, such as carbon nanotubes, graphene, carbon nanofibers, magnetic nanoparticles, and gold nanoparticles into these biosensors, has led to significant improvements in their analytical performance because of their large surface area, high electrical conductivity, and chemical stability [91]. In the next subsections, we will focus mainly on the electrochemical and optical detection techniques used in the biosensor designs for different metabolites.

3.1 Electrochemical Detection

Electrochemical detection is the main detection approach used in most of the biosensors as transducer via monitoring the electrochemical signal generated when the analyte binds to the recognition receptor. These electrochemical signals can be a measurable charge accumulation or potential (potentiometry), current (amperometry/voltammetry), conductivity (conductometry), or resistance and capacitance (Electrochemical impedance spectroscopy). Electrochemical detection methods in biosensors offer several advantages such as their high sensitivity, low cost, ease of use, and capability of miniaturization. These advantages have led to significant applications of electrochemical biosensors in point-of-care testing as they are easier to implement in integrated biosensors and operate than optical techniques. Moreover, the advances in the screen-printing technology have led to a wide range of applications of screen-printed electrodes in point-of-care diagnostic biosensors. Electrochemical detection techniques used in the biosensors for metabolites are mostly voltammetry/amperometry and electrochemical impedance spectroscopy. Thus, these techniques will be discussed in the next sections.

3.1.1 Amperometry/Voltammetry

Voltammetric and amperometric techniques are used to measure the current resulting from an electrochemical oxidation or reduction processes when a potential is applied to a working electrode versus a reference electrode. In the voltammetric techniques, the potential is scanned over a set potential range resulting in a current response in the form of a peak, whereas in the amperometry, a constant potential is maintained, and the generated current is monitored directly with time. The most used voltammetric methods in biosensors are the cyclic voltammetry, differential pulse voltammetry, linear sweep voltammetry, and square wave voltammetry. Amperometric methods are primarily used for biocatalytic/enzymatic biosensors, whereas voltammetric-based detection is often used on affinity biosensors (immunosensors and aptasensors).

Several metabolites have been detected using amperometric- [15–19, 92] or voltammetric-based biosensors [76, 87] utilizing various electrodes and nanomaterials. Since most of the amperometric−/voltammetric-based detection methods are mainly used in enzyme-based biosensors, they were described in more detail in Sect. 2.1.

3.1.2 Electrochemical Impedance Spectroscopy

Electrochemical impedance spectroscopy is an electrochemical technique which used to measure the capacitive and resistive properties of an electrode upon perturbation of a system with a small amplitude sinusoidal ac excitation signal typically of 2–10 mV. The current response is then determined with changing frequency over a wide range, and the result is displayed in the form of an impedance spectrum. Impedance-based detection is mainly used in affinity biosensor showing good advantages in terms of sensitivity and nondestructive nature. Several impedance-based biosensors for different metabolites have been reported [71, 93, 94]. For instance, impedance-based biosensor for the detection of parathyroid hormone for the diagnosis of thyroid cancer, hypoparathyroidism, and hyperparathyroidism has been described [93]. The biosensor was fabricated by immobilizing antibody for parathyroid on poly amidoamine dendrimer which was attached to gold electrode. The detection was achieved via monitoring the change in the charge transfer resistance of the electrode upon binding of the hormone with the immunosensor. The linear range of the biosensor was from 10 to 60 fg/mL. Immunosensor was used to detect parathyroid levels in artificial serum samples.

3.2 Optical Detection

The detection of the binding event between the recognition receptor and the target in the biosensors through monitoring the change in the optical signals is very popular. The optical detection methods are usually based on the use of an optically active label conjugated to the target molecule. These labels can be a simple fluorophore or quencher molecules for fluorescence detection or nanoparticle tags for colorimetric or surface plasmon resonance-based biosensors.

A fluorescence-based aptasensor for arginine has been reported [89]. The detection is based on the quenching of the carboxyfluorescein-labeled aptamer when it is adsorbed on the negatively charged gold nanoparticles because of the fluorescence resonance energy transfer effect. However, when the D- or L-arginine exist in the sample, the aptamer binds to the target causing a change in the conformation of the aptamer. This prevents the fluorescence label from being adsorbed on the gold nanoparticles leading to a recovery in the fluorescence intensity. In this study, it was shown that the increase in the fluorescence signal was stronger when L-arginine was used compared to D-arginine. An aptamer-based surface-enhanced Raman spectroscopy biosensor for the detection of vasopressin was reported [81]. Densely packed metal nanotube arrays prepared using an anodized alumina nanoporous membrane were used to create the active substrate. The integration of the membranes with a polydimethylsiloxane microfluidic device has demonstrated good sensitivity with a LOD of 5.2 μU/mL. Colorimetric-based detection of cortisol is simple and low cost and does not require the use of sophisticated equipment as the results can be seen by the naked eye. Four different chromogens (sulfuric acid, Porter-Silber reagent, Prussian blue, and blue tetrazolium) have been used for the detection of cortisol in artificial saliva and human sweat. This method has shown comparable sensitivity to the electrochemical biosensors. The simplicity of the colorimetric detection makes them suitable for point-of-care testing of different metabolites [95]. Colorimetric-based aptasensor was reported for the detection of progesterone in human serum and urine [96]. This method was based on the alteration of the aggregating properties of the gold nanoparticles upon addition of progesterone, aptamer, and the cationic surfactant, hexadecyltrimethylammonium bromide. When the aptamer binds to progesterone, the surfactant causes aggregation of the gold nanoparticles leading to a change of the color of the solution from red to blue. Progesterone has been also detected colorimetrically using lateral flow assay [97]. A specific aptamer for progesterone was immobilized on gold nanoparticles, and a biotin-labeled complementary DNA sequence was then hybridized with the aptamer. The test line was coated with streptavidin which allows the capture of the biotinylated gold nanoparticles – aptamer duplex. However, upon binding of the aptamer with the target progesterone, a displacement of the biotinylated complementary DNA has occurred which prevented the capture of the gold particles on the test line. This aptamer-based lateral flow assay method has led to sensitive detection of progesterone in the nanomolar range as well as good selectivity against other hormones.

4 Advances in the Development of Biosensors for Metabolic Biomarkers

Recent advances in nanotechnology have accelerated the improvements in the development of biosensors for metabolic biomarkers in point-of-care diagnosis. Multiplexed biosensors offered high-throughput simultaneous screening of various targets which is highly important in metabolomics. Particularly, the ease of fabrication of multiple individually addressable arrays of working electrodes on small chips has led to the development of many multiplexed electrochemical biosensors [98]. Microfluidics enables the use of very small sample volumes for detection, and thus, it is perfectly suited for the diagnostic point-of-care biosensors where a small blood sample is often used [99–101]. Zhao et al. [102] have reported the development of a microfluidic paper-based multiplexed electrochemical biosensor array for the simultaneous detection of three metabolic biomarkers, glucose, lactate, and uric acid. Channels were fabricated on chromatographic paper via solid wax printing, and the silver connections and carbon electrodes were printed on the surface of the paper using screen printing. An array of eight biosensors were developed, and the signals were detected using a portable handheld custom-made potentiostat. Simultaneous measurements for multiple analytes were successfully recorded, and the analytical performance of the device was comparable with the commercial platforms.

We have reported the development of a multiplexed electrochemical immunosensor for the simultaneous detection of the metabolites: morphine, benzoylecgonine, and tetrahydrocannabinol in urine [103]. Gold nanoparticle-modified screen-printed carbon array electrodes were utilized on which specific antibodies for the three metabolites were immobilized. A competitive assay was used for the detection by using bovine serum albumin-conjugated analytes. The multiplexed biosensor showed fast response and high sensitivity and selectivity.

Wearable biosensors are gaining significant interest because of their potential to provide noninvasive and real-time measurements of biomarkers of different metabolites in biofluids, such as saliva, sweat, and tears [104]. Wearable biosensors combine the microfluidic sampling, multiplexed biosensing, and transport systems with flexible materials to form miniaturized and easily operating detection tool (Fig. 3). Recent wearable biosensors for healthcare monitoring have been recently reviewed by Kim et al. [104].

Recently, a wearable lactate electrochemical biosensor for sweat analysis has been reported [11]. An outer plasticized polymeric layer containing the tetradodecylammonium tetrakis (4-chlorophenyl) borate salt has been used to prevent the lactate oxidase enzyme from being in direct contact with the sample which significantly reduced the effect of temperature and pH. The authors reported that their biosensor showed higher sensitivity than other reported biosensors with good selectivity and reproducibility. The analytical performance of this wearable biosensor was comparable with ion chromatography method. The biosensor was also applied for the detection of lactate in sweat on three different body locations (forehead, back, and

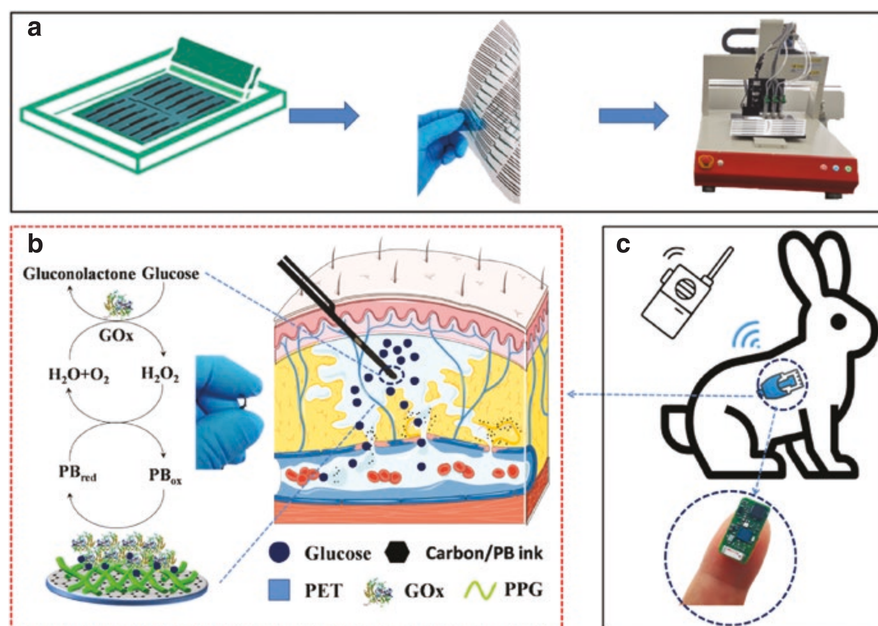

Fig. 3 (**a**) flexible screen-printed electrode array. (**b**) The enzymatic biosensing mechanism of glucose detection and implantation of wearable biosensor. (**c**) Schematic of the integrated wearable biosensor for the continuous monitoring of glucose on a rabbit. (Reprinted with permission from [105])

thigh) showing good performance indicating that it holds great promise toward other healthcare applications.

5 Conclusions and Future Perspectives

The development of biosensors for metabolic biomarkers offers a faster, simpler, and lower cost alternative to the conventional analytical assays. Several biosensors for various metabolites have been developed for point-of-care diagnosis. In this chapter, we reviewed the different biosensor designs used in metabolomics. Extensive research has been devoted for the development of enzyme-based biocatalytic sensors for some metabolites used in medical diagnosis such as glucose, lactate, and xanthene. Various types of nanoparticles have been integrated in these biosensors to enhance their analytical performance which resulted in high sensitivity and rapid analysis. However, there are no specific enzymes for many metabolites. Moreover, the enzyme stability is limited as it gradually loses activity over time which represents a major challenge in the enzymatic biosensors. Therefore, recent research has focused on the development of other biorecognition receptors

such as antibodies and aptamer to fabricate affinity biosensors for various metabolites. The continuous identification of new aptamers that binds specifically to various types of metabolites is highly needed and can open the door for the development of sensitive biosensing platforms for point-of-care diagnosis. Optical and electrochemical detections have been used in most of the reported biosensors for metabolic biomarkers. Particularly, electrochemical biosensors offer great promise because of their low cost, high sensitivity, ease of integration into portable biosensors, and capability of multiplexing which make them ideal for high-throughput screening in metabolomics. Advances in the development of wearable biosensors have been also highlighted in this chapter as they provide continuous and noninvasive measurements for different metabolites in body fluids. Yet, large cohort studies of these wearable biosensors are required for validation to reach clinical acceptance.

References

1. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. Biomed Res Int. 2015;2015:354671. https://doi.org/10.1155/2015/354671.
2. Song Q, Zhang A-h, Yan G-l, Liu L, Wang X-J. Technological advances in current metabolomics and its application in tradition Chinese medicine. RSC Adv. 2017;7(84):53516–24. https://doi.org/10.1039/C7RA02056B.
3. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner HP. Targeted metabolomics for biomarker discovery. Angew Chem Int Ed Engl. 2010;49(32):5426–45. https://doi.org/10.1002/anie.200905579.
4. Clarke CJ, Haselden JN. Metabolic profiling as a tool for understanding mechanisms of toxicity. Toxicol Pathol. 2008;36(1):140–7. https://doi.org/10.1177/0192623307310947.
5. Wang J. Glucose biosensors: 40 years of advances and challenges. Electroanalysis. 2001;13(12):983–8. https://doi.org/10.1002/1521-4109(200108)13:12<983::AID-ELAN983>3.0.CO;2-#.
6. Ronkainen NJ, Halsall HB, Heineman WR. Electrochemical biosensors. Chem Soc Rev. 2010;39(5):1747–63. https://doi.org/10.1039/B714449K.
7. Yoo EH, Lee SY. Glucose biosensors: an overview of use in clinical practice. Sensors (Basel). 2010;10(5):4558–76. https://doi.org/10.3390/s100504558.
8. Juska VB, Pemble ME. A critical review of electrochemical glucose sensing: evolution of biosensor platforms based on advanced nanosystems. Sensors (Basel). 2020;20(21):6013. https://doi.org/10.3390/s20216013.
9. Oliver NS, Toumazou C, Cass AEG, Johnston DG. Glucose sensors: a review of current and emerging technology. Diabet Med. 2009;26(3):197–210. https://doi.org/10.1111/j.1464-5491.2008.02642.x.
10. Taguchi M, Ptitsyn A, McLamore ES, Claussen JC. Nanomaterial-mediated biosensors for monitoring glucose. J Diabetes Sci Technol. 2014;8(2):403–11. https://doi.org/10.1177/1932296814522799.
11. Xuan X, Pérez-Ràfols C, Chen C, Cuartero M, Crespo GA. Lactate biosensing for reliable on-body sweat analysis. ACS Sens. 2021;6(7):2763–71. https://doi.org/10.1021/acssensors.1c01009.
12. Hiraka K, Tsugawa W, Asano R, Yokus MA, Ikebukuro K, Daniele MA, et al. Rational design of direct electron transfer type l-lactate dehydrogenase for the development of multiplexed biosensor. Biosens Bioelectron. 2021;176:112933. https://doi.org/10.1016/j.bios.2020.112933.

13. Raj MA, John SA. Simultaneous determination of uric acid, xanthine, hypoxanthine and caffeine in human blood serum and urine samples using electrochemically reduced graphene oxide modified electrode. Anal Chim Acta. 2013;771:14–20. https://doi.org/10.1016/j.aca.2013.02.017.
14. Pundir CS, Devi R. Biosensing methods for xanthine determination: a review. Enzym Microb Technol. 2014;57:55–62. https://doi.org/10.1016/j.enzmictec.2013.12.006.
15. Tsiafoulis CG, Prodromidis MI, Karayannis MI. Development of amperometric biosensors for the determination of glycolic acid in real samples. Anal Chem. 2002;74(1):132–9. https://doi.org/10.1021/ac0106896.
16. Tian G, Ding M, Xu B, He Y, Lyu W, Jin M, et al. A novel electrochemical biosensor for ultrasensitive detection of serum total bile acids based on enzymatic reaction combined with the double oxidation circular amplification strategy. Biosens Bioelectron. 2018;118:31–5. https://doi.org/10.1016/j.bios.2018.07.030.
17. Dong E, Zheng J, Liu H. Bile acid detection by biosensors-a review. Sheng Wu Gong Cheng Xue Bao. 2020;36(12):2779–90. https://doi.org/10.13345/j.cjb.200196.
18. Verma N, Singh AK, Singh M. L-arginine biosensors: a comprehensive review. Biochem Biophys Rep. 2017;12:228–39. https://doi.org/10.1016/j.bbrep.2017.10.006.
19. Magar HS, Ghica ME, Abbas MN, Brett CMA. A novel sensitive amperometric choline biosensor based on multiwalled carbon nanotubes and gold nanoparticles. Talanta. 2017;167:462–9. https://doi.org/10.1016/j.talanta.2017.02.048.
20. Weis BK, Balshaw D, Barr JR, Brown D, Ellisman M, Lioy P, et al. Personalized exposure assessment: promising approaches for human environmental health research. Environ Health Perspect. 2005;113(7):840–8. https://doi.org/10.1289/ehp.7651.
21. Boldt J. Clinical review: hemodynamic monitoring in the intensive care unit. Crit Care. 2002;6(1):52–9. https://doi.org/10.1186/cc1453.
22. Sánchez-Paniagua López M, Hervás Pérez JP, López-Cabarcos E, López-Ruiz B. Amperometric biosensors based on choline oxidase entrapped in polyacrylamide microgels. Electroanalysis. 2007;19(2–3):370–8. https://doi.org/10.1002/elan.200603742.
23. Zhang H, Yin Y, Wu P, Cai C. Indirect electrocatalytic determination of choline by monitoring hydrogen peroxide at the choline oxidase-prussian blue modified iron phosphate nanostructures. Biosens Bioelectron. 2012;31(1):244–50. https://doi.org/10.1016/j.bios.2011.10.026.
24. Ricci F, Amine A, Palleschi G, Moscone D. Prussian blue based screen printed biosensors with improved characteristics of long-term lifetime and pH stability. Biosens Bioelectron. 2003;18(2):165–74. https://doi.org/10.1016/S0956-5663(02)00169-0.
25. Shi H, Yang Y, Huang J, Zhao Z, Xu X, Anzai J, et al. Amperometric choline biosensors prepared by layer-by-layer deposition of choline oxidase on the Prussian blue-modified platinum electrode. Talanta. 2006;70(4):852–8. https://doi.org/10.1016/j.talanta.2006.02.012.
26. Keihan AH, Sajjadi S, Sheibani N, Moosavi-Movahedi AA. A highly sensitive choline biosensor based on bamboo-like multiwall carbon nanotubes/ionic liquid/Prussian blue nanocomposite. Sensors Actuators B Chem. 2014;204:694–703. https://doi.org/10.1016/j.snb.2014.08.039.
27. Deng K, Zhou J, Li X. Noncovalent nanohybrid of ferrocene with chemically reduced graphene oxide and its application to dual biosensor for hydrogen peroxide and choline. Electrochim Acta. 2013;95:18–23. https://doi.org/10.1016/j.electacta.2013.02.009.
28. Razola SS, Pochet S, Grosfils K, Kauffmann JM. Amperometric determination of choline released from rat submandibular gland acinar cells using a choline oxidase biosensor. Biosens Bioelectron. 2003;18(2–3):185–91. https://doi.org/10.1016/s0956-5663(02)00186-0.
29. Song Z, Huang J-D, Wu B-Y, Shi H-B, Anzai J-I, Chen Q. Amperometric aqueous sol–gel biosensor for low-potential stable choline detection at multi-wall carbon nanotube modified platinum electrode. Sensors Actuators B Chem. 2006;115(2):626–33. https://doi.org/10.1016/j.snb.2005.10.030.
30. Serafín V, Martínez-García G, Aznar-Poveda J, Lopez-Pastor JA, Garcia-Sanchez AJ, Garcia-Haro J, et al. Determination of progesterone in saliva using an electrochemical immunosensor

and a COTS-based portable potentiostat. Anal Chim Acta. 2019;1049:65–73. https://doi.org/10.1016/j.aca.2018.10.019.

31. Dong X-X, Yuan L-P, Liu Y-X, Wu M-F, Liu B, Sun Y-M, et al. Development of a progesterone immunosensor based on thionine-graphene oxide composites platforms: improvement by biotin-streptavidin-amplified system. Talanta. 2017;170:502–8. https://doi.org/10.1016/j.talanta.2017.04.054.
32. Carralero V, González-Cortés A, Yáñez-Sedeño P, Pingarrón JM. Nanostructured progesterone immunosensor using a tyrosinase–colloidal gold–graphite–teflon biosensor as amperometric transducer. Anal Chim Acta. 2007;596(1):86–91. https://doi.org/10.1016/j.aca.2007.05.046.
33. Monerris MJ, Arévalo FJ, Fernández H, Zon MA, Molina PG. Integrated electrochemical immunosensor with gold nanoparticles for the determination of progesterone. Sensors Actuators B Chem. 2012;166-167:586–92. https://doi.org/10.1016/j.snb.2012.03.015.
34. Arya SK, Chornokur G, Venugopal M, Bhansali S. Dithiobis(succinimidyl propionate) modified gold microarray electrode based electrochemical immunosensor for ultrasensitive detection of cortisol. Biosens Bioelectron. 2010;25(10):2296–301. https://doi.org/10.1016/j.bios.2010.03.016.
35. Kumar A, Aravamudhan S, Gordic M, Bhansali S, Mohapatra SS. Ultrasensitive detection of cortisol with enzyme fragment complementation technology using functionalized nanowire. Biosens Bioelectron. 2007;22(9):2138–44. https://doi.org/10.1016/j.bios.2006.09.035.
36. Vasudev A, Kaushik A, Tomizawa Y, Norena N, Bhansali S. An LTCC-based microfluidic system for label-free, electrochemical detection of cortisol. Sensors Actuators B Chem. 2013;182:139–46. https://doi.org/10.1016/j.snb.2013.02.096.
37. Arya SK, Dey A, Bhansali S. Polyaniline protected gold nanoparticles based mediator and label free electrochemical cortisol biosensor. Biosens Bioelectron. 2011;28(1):166–73. https://doi.org/10.1016/j.bios.2011.07.015.
38. Tlili C, Myung NV, Shetty V, Mulchandani A. Label-free, chemiresistor immunosensor for stress biomarker cortisol in saliva. Biosens Bioelectron. 2011;26(11):4382–6. https://doi.org/10.1016/j.bios.2011.04.045.
39. Arya SK, Chornokur G, Venugopal M, Bhansali S. Antibody modified gold micro array electrode based electrochemical immunosensor for ultrasensitive detection of cortisol in saliva and ISF. Procedia Eng. 2010;5:804–7. https://doi.org/10.1016/j.proeng.2010.09.230.
40. Kämäräinen S, Mäki M, Tolonen T, Palleschi G, Virtanen V, Micheli L, et al. Disposable electrochemical immunosensor for cortisol determination in human saliva. Talanta. 2018;188:50–7. https://doi.org/10.1016/j.talanta.2018.05.039.
41. Venugopal M, Arya SK, Chornokur G, Bhansali S. A realtime and continuous assessment of cortisol in ISF using electrochemical impedance spectroscopy. Sensors Actuators A Phys. 2011;172(1):154–60. https://doi.org/10.1016/j.sna.2011.04.028.
42. Yamaguchi M, Matsuda Y, Sasaki S, Sasaki M, Kadoma Y, Imai Y, et al. Immunosensor with fluid control mechanism for salivary cortisol analysis. Biosens Bioelectron. 2013;41:186–91. https://doi.org/10.1016/j.bios.2012.08.016.
43. Volpe G, Fares G, Delli Quadri F, Draisci R, Ferretti G, Marchiafava C, et al. A disposable immunosensor for detection of 17beta-estradiol in non-extracted bovine serum. Anal Chim Acta. 2006;572(1):11–6. https://doi.org/10.1016/j.aca.2006.05.008.
44. Pemberton RM, Mottram TT, Hart JP. Development of a screen-printed carbon electrochemical immunosensor for picomolar concentrations of estradiol in human serum extracts. J Biochem Biophys Methods. 2005;63(3):201–12. https://doi.org/10.1016/j.jbbm.2005.05.002.
45. Liu X, Wong DK. Picogram-detection of estradiol at an electrochemical immunosensor with a gold nanoparticle|protein G-(LC-SPDP)-scaffold. Talanta. 2009;77(4):1437–43. https://doi.org/10.1016/j.talanta.2008.09.027.
46. Ojeda I, López-Montero J, Moreno-Guzmán M, Janegitz BC, González-Cortés A, Yáñez-Sedeño P, et al. Electrochemical immunosensor for rapid and sensitive determination of estradiol. Anal Chim Acta. 2012;743:117–24. https://doi.org/10.1016/j.aca.2012.07.002.

47. Liu X, Duckworth PA, Wong DK. Square wave voltammetry versus electrochemical impedance spectroscopy as a rapid detection technique at electrochemical immunosensors. Biosens Bioelectron. 2010;25(6):1467–73. https://doi.org/10.1016/j.bios.2009.10.047.
48. Kuramitz H, Matsuda M, Thomas JH, Sugawara K, Tanaka S. Electrochemical immunoassay at a 17β-estradiol self-assembled monolayer electrode using a redox marker. Analyst. 2003;128(2):182–6. https://doi.org/10.1039/b209590b.
49. Eguílaz M, Moreno-Guzmán M, Campuzano S, González-Cortés A, Yáñez-Sedeño P, Pingarrón JM. An electrochemical immunosensor for testosterone using functionalized magnetic beads and screen-printed carbon electrodes. Biosens Bioelectron. 2010;26(2):517–22. https://doi.org/10.1016/j.bios.2010.07.060.
50. Conneely G, Aherne M, Lu H, Guilbault GG. Development of an immunosensor for the detection of testosterone in bovine urine. Anal Chim Acta. 2007;583(1):153–60. https://doi.org/10.1016/j.aca.2006.09.062.
51. Wang H, Chen W, Li D, Yin X, Zhang X, Olsen N, et al. Vitamin D and chronic diseases. Aging Dis. 2017;8(3):346–53. https://doi.org/10.14336/AD.2016.1021.
52. Demir M, Demir F, Aygun H. Vitamin D deficiency is associated with COVID-19 positivity and severity of the disease. J Med Virol. 2021;93(5):2992–9. https://doi.org/10.1002/jmv.26832.
53. Weir EK, Thenappan T, Bhargava M, Chen Y. Does vitamin D deficiency increase the severity of COVID-19? Clin Med (Lond). 2020;20(4):e107–e8. https://doi.org/10.7861/clinmed.2020-0301.
54. Chauhan D, Solanki PR. Hydrophilic and insoluble electrospun cellulose acetate fiber-based biosensing platform for 25-Hydroxy vitamin-D3 detection. ACS Appl Polym Mater. 2019;1(7):1613–23. https://doi.org/10.1021/acsapm.9b00179.
55. Chauhan D, Gupta PK, Solanki PR. Electrochemical immunosensor based on magnetite nanoparticles incorporated electrospun polyacrylonitrile nanofibers for vitamin-D(3) detection. Mater Sci Eng C Mater Biol Appl. 2018;93:145–56. https://doi.org/10.1016/j.msec.2018.07.036.
56. Kaur A, Kapoor S, Bharti A, Rana S, Chaudhary GR, Prabhakar N. Gold-platinum bimetallic nanoparticles coated 3-(aminopropyl)triethoxysilane (APTES) based electrochemical immunosensor for vitamin D estimation. J Electroanal Chem. 2020;873:114400. https://doi.org/10.1016/j.jelechem.2020.114400.
57. Carlucci L, Favero G, Tortolini C, Di Fusco M, Romagnoli E, Minisola S, et al. Several approaches for vitamin D determination by surface plasmon resonance and electrochemical affinity biosensors. Biosens Bioelectron. 2013;40(1):350–5. https://doi.org/10.1016/j.bios.2012.07.077.
58. Chauhan D, Yadav AK, Solanki PR. Carbon cloth-based immunosensor for detection of 25-hydroxy vitamin D3. Microchim Acta. 2021;188(4):145. https://doi.org/10.1007/s00604-021-04751-y.
59. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990;346(6287):818–22. https://doi.org/10.1038/346818a0.
60. Song S, Wang L, Li J, Fan C, Zhao J. Aptamer-based biosensors. TrAC Trends Anal Chem. 2008;27(2):108–17. https://doi.org/10.1016/j.trac.2007.12.004.
61. Marrazza G. Aptamer sensors. Biosensors (Basel). 2017;7(1):5. https://doi.org/10.3390/bios7010005.
62. Ruscito A, DeRosa MC. Small-molecule binding aptamers: selection strategies, characterization, and applications. Front Chem. 2016;4:14. https://doi.org/10.3389/fchem.2016.00014.
63. Zhou W, Jimmy Huang P-J, Ding J, Liu J. Aptamer-based biosensors for biomedical diagnostics. Analyst. 2014;139(11):2627–40. https://doi.org/10.1039/c4an00132j.
64. Eissa S, Siddiqua A, Chinnappan R, Zourob M. Electrochemical selection of a DNA aptamer, and an impedimetric method for determination of the dedicator of cytokinesis 8 by self-assembly of a thiolated aptamer on a gold electrode. Microchim Acta. 2019;186(12):828. https://doi.org/10.1007/s00604-019-3817-7.

65. Eissa S, Zourob M, et al. Sci Rep. 2017;7(1):1016. https://doi.org/10.1038/s41598-017-01226-0.
66. Almusharraf AY, Eissa S, Zourob M. Truncated aptamers for total and glycated hemoglobin, and their integration into a graphene oxide-based fluorometric method for high-throughput screening for diabetes. Microchim Acta. 2018;185(5):256. https://doi.org/10.1007/s00604-018-2789-3.
67. Prante M, Segal E, Scheper T, Bahnemann J, Walter J. Aptasensors for point-of-care detection of small molecules. Biosensors. 2020;10(9):108. https://doi.org/10.3390/bios10090108.
68. Musheev MU, Krylov SN. Selection of aptamers by systematic evolution of ligands by exponential enrichment: addressing the polymerase chain reaction issue. Anal Chim Acta. 2006;564(1):91–6. https://doi.org/10.1016/j.aca.2005.09.069.
69. Rather JA, Khudaish EA, Kannan P. Graphene-amplified femtosensitive aptasensing of estradiol, an endocrine disruptor. Analyst. 2018;143(8):1835–45. https://doi.org/10.1039/C7AN02092A.
70. Akki SU, Werth CJ, Silverman SK. Selective aptamers for detection of estradiol and ethynylestradiol in natural waters. Environ Sci Technol. 2015;49(16):9905–13. https://doi.org/10.1021/acs.est.5b02401.
71. Contreras Jiménez G, Eissa S, Ng A, Alhadrami H, Zourob M, Siaj M. Aptamer-based label-free impedimetric biosensor for detection of progesterone. Anal Chem. 2015;87(2):1075–82. https://doi.org/10.1021/ac503639s.
72. Alhadrami HA, Chinnappan R, Eissa S, Rahamn AA, Zourob M. High affinity truncated DNA aptamers for the development of fluorescence based progesterone biosensors. Anal Biochem. 2017;525:78–84. https://doi.org/10.1016/j.ab.2017.02.014.
73. Zulkeflee Sabri M, Abdul Hamid AA, Sayed Hitam SM, Abdul Rahim MZ. In-silico study of single-strand DNA aptamers against cortisol as a potential biomarker receptor in therapeutics and diagnostics. Mater Today Proc. 2020;31:A90–A7. https://doi.org/10.1016/j.matpr.2020.12.1080.
74. Martin JA, Chávez JL, Chushak Y, Chapleau RR, Hagen J, Kelley-Loughnane N. Tunable stringency aptamer selection and gold nanoparticle assay for detection of cortisol. Anal Bioanal Chem. 2014;406(19):4637–47. https://doi.org/10.1007/s00216-014-7883-8.
75. Lévesque D, Beaudoin J-D, Roy S, Perreault J-P. In vitro selection and characterization of RNA aptamers binding thyroxine hormone. Biochem J. 2007;403(1):129–38. https://doi.org/10.1042/BJ20061216.
76. Eissa S, Siddiqua A, Chinnappan R, Zourob M. Electrochemical SELEX technique for the selection of DNA aptamers against the small molecule 11-Deoxycortisol. ACS Appl Bio Mater. 2019;2(6):2624–32. https://doi.org/10.1021/acsabm.9b00294.
77. Skouridou V, Jauset-Rubio M, Ballester P, Bashammakh AS, El-Shahawi MS, Alyoubi AO, et al. Selection and characterization of DNA aptamers against the steroid testosterone. Microchim Acta. 2017;184(6):1631–9. https://doi.org/10.1007/s00604-017-2136-0.
78. He P, Oncescu V, Lee S, Choi I, Erickson D. Label-free electrochemical monitoring of vasopressin in aptamer-based microfluidic biosensors. Anal Chim Acta. 2013;759:74–80. https://doi.org/10.1016/j.aca.2012.10.038.
79. Williams KP, Liu XH, Schumacher TN, Lin HY, Ausiello DA, Kim PS, et al. Bioactive and nuclease-resistant L-DNA ligand of vasopressin. Proc Natl Acad Sci U S A. 1997;94(21):11285–90. https://doi.org/10.1073/pnas.94.21.11285.
80. Purschke WG, Eulberg D, Buchner K, Vonhoff S, Klussmann S. An L-RNA-based aquaretic agent that inhibits vasopressin in vivo. Proc Natl Acad Sci U S A. 2006;103(13):5173. https://doi.org/10.1073/pnas.0509663103.
81. Huh YS, Erickson D. Aptamer based surface enhanced raman scattering detection of vasopressin using multilayer nanotube arrays. Biosens Bioelectron. 2010;25(5):1240–3. https://doi.org/10.1016/j.bios.2009.09.040.

82. Lee BH, Nguyen VT, Gu MB. Highly sensitive detection of 25-HydroxyvitaminD3 by using a target-induced displacement of aptamer. Biosens Bioelectron. 2017;88:174–80. https://doi.org/10.1016/j.bios.2016.08.011.
83. Echenne B, Roubertie A, Assmann B, Lutz T, Penzien JM, Thöny B, et al. Sepiapterin reductase deficiency: clinical presentation and evaluation of long-term therapy. Pediatr Neurol. 2006;35(5):308–13. https://doi.org/10.1016/j.pediatrneurol.2006.05.006.
84. Blau N, Bonafé L, Thöny B. Tetrahydrobiopterin deficiencies without Hyperphenylalaninemia: diagnosis and genetics of DOPA-responsive dystonia and sepiapterin reductase deficiency. Mol Genet Metab. 2001;74(1):172–85. https://doi.org/10.1006/mgme.2001.3213.
85. Blau N. Disorder of tetrahydrobiopterin and related biogenic amines. The metabolic and molecular bases of inherited disease. 2000.
86. Assmann B, Surtees R, Hoffmann GF. Approach to the diagnosis of neurotransmitter diseases exemplified by the differential diagnosis of childhood-onset dystonia. Ann Neurol. 2003;54(S6):S18–24. https://doi.org/10.1002/ana.10628.
87. Eissa S, Alkhaldi S, Chinnappan R, Siddiqua A, Abduljabbar M, Abdel Rahman AM, et al. Selection, characterization, and electrochemical biosensing application of DNA aptamers for sepiapterin. Talanta. 2020;216:120951. https://doi.org/10.1016/j.talanta.2020.120951.
88. Jung J, Kim SH, Lee HS, Choi GS, Jung YS, Ryu DH, et al. Serum metabolomics reveals pathways and biomarkers associated with asthma pathogenesis. Clin Exp Allergy. 2013;43(4):425–33. https://doi.org/10.1111/cea.12089.
89. Yuan H, Huang Y, Yang J, Guo Y, Zeng X, Zhou S, et al. An aptamer-based fluorescence bio-sensor for chiral recognition of arginine enantiomers. Spectrochim Acta A Mol Biomol Spectrosc. 2018;200:330–8. https://doi.org/10.1016/j.saa.2018.04.038.
90. Imani S, Bandodkar AJ, Mohan AMV, Kumar R, Yu S, Wang J, et al. A wearable chemical–electrophysiological hybrid biosensing system for real-time health and fitness monitoring. Nat Commun. 2016;7(1):11650. https://doi.org/10.1038/ncomms11650.
91. Malhotra BD, Ali MA. Nanomaterials in biosensors: fundamentals and applications. Nanomat Biosens. 2018;1:1–74. https://doi.org/10.1016/B978-0-323-44923-6.00001-7.
92. He C, Xie M, Hong F, Chai X, Mi H, Zhou X, et al. A highly sensitive glucose biosensor based on gold nanoparticles/bovine serum albumin/Fe3O4 biocomposite nanoparticles. Electrochim Acta. 2016;222:1709–15. https://doi.org/10.1016/j.electacta.2016.11.162.
93. Özcan HM, Sezgintürk MK. Detection of parathyroid hormone using an electrochemical impedance biosensor based on PAMAM dendrimers. Biotechnol Prog. 2015;31(3):815–22. https://doi.org/10.1002/btpr.2060.
94. Saxena R, Srivastava S. A sensitive and one-step quantification of thyroid stimulating hormone using nanobiosensor. Mater Today Proc. 2019;18:1351–7. https://doi.org/10.1016/j.matpr.2019.06.600.
95. Tu E, Pearlmutter P, Tiangco M, Derose G, Begdache L, Koh A. Comparison of colorimetric analyses to determine cortisol in human sweat. ACS Omega. 2020;5(14):8211–8. https://doi.org/10.1021/acsomega.0c00498.
96. Du G, Wang L, Zhang D, Ni X, Zhou X, Xu H, et al. Colorimetric aptasensor for progesterone detection based on surfactant-induced aggregation of gold nanoparticles. Anal Biochem. 2016;514:2–7. https://doi.org/10.1016/j.ab.2016.09.006.
97. Alnajrani MN, Alsager OA. Lateral flow aptasensor for progesterone: competitive target recognition and displacement of short complementary sequences. Anal Biochem. 2019;587:113461. https://doi.org/10.1016/j.ab.2019.113461.
98. Eissa S, Zourob M. Ultrasensitive peptide-based multiplexed electrochemical biosensor for the simultaneous detection of listeria monocytogenes and Staphylococcus aureus. Microchim Acta. 2020;187(9):486. https://doi.org/10.1007/s00604-020-04423-3.
99. Jayamohan H, Sant HJ, Gale BK. Applications of microfluidics for molecular diagnostics. Methods Mol Biol. 2013;949:305–34. https://doi.org/10.1007/978-1-62703-134-9_20.

100. Nikoleli G-P, Siontorou CG, Nikolelis DP, Bratakou S, Karapetis S, Tzamtzis N. Chapter 13—biosensors based on microfluidic devices lab-on-a-chip and microfluidic technology. In: Nikolelis DP, Nikoleli G-P, editors. Nanotechnology and biosensors. Amsterdam: Elsevier; 2018. p. 375–94.
101. Zhao C, Liu X. A paper-based microfluidic device for multiplexed electrochemical detection of biomarkers. 2013 International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale. 2013. p. 162–5.
102. Zhao C, Thuo MM, Liu X. A microfluidic paper-based electrochemical biosensor array for multiplexed detection of metabolic biomarkers. Sci Technol Adv Mater. 2013;14(5):054402. https://doi.org/10.1088/1468-6996/14/5/054402.
103. Eissa S, Almthen RA, Zourob M. Disposable electrochemical immunosensor array for the multiplexed detection of the drug metabolites morphine, tetrahydrocannabinol and benzoylecgonine. Microchim Acta. 2019;186(8):523. https://doi.org/10.1007/s00604-019-3646-8.
104. Kim J, Campbell AS, de Ávila BE-F, Wang J. Wearable biosensors for healthcare monitoring. Nat Biotechnol. 2019;37(4):389–406. https://doi.org/10.1038/s41587-019-0045-y.
105. Jin X, Li G, Xu T, Su L, Yan D, Zhang X. Fully integrated flexible biosensor for wearable continuous glucose monitoring. Biosens Bioelectron. 2022;196:113760. https://doi.org/10.1016/j.bios.2021.113760.