

Establishing parameters for the characterization of rare neurometabolic and neurodegenerative diseases using mass-spectrometry based Metabolomics

Dissertation

for the award of the degree

“Doctor rerum naturalium”

of the Georg-August-Universität Göttingen

within the doctoral program International Max Planck Research School for Neuroscience
of the Georg-August University School of Science (GAUSS)

submitted by

Henry Gerd Klemp

from Hildesheim, Germany

Göttingen, 2021

Thesis Committee:

Prof. Dr. André Fischer, Epigenetik und Systemmedizin bei Neurodegenerativen Erkrankungen, Deutsches Zentrum für Neurodegenerative Erkrankungen Göttingen

Prof. Dr. Klaus-Armin Nave, Abteilung für Neurogenetik, Max Planck Institut für experimentelle Medizin

Prof. Dr. Jutta Gärtner, Klinik für Kinder- und Jugendmedizin, Universitätsmedizin Göttingen

Members of the Examination Board:

Referee: Prof. Dr. André Fischer, Epigenetik und Systemmedizin bei Neurodegenerativen Erkrankungen, Deutsches Zentrum für Neurodegenerative Erkrankungen Göttingen

2nd Referee: Prof. Dr. Klaus-Armin Nave, Abteilung für Neurogenetik, Max Planck Institut für experimentelle Medizin

Further members of the Examination Board:

Prof. Dr. Jutta Gärtner, Klinik für Kinder- und Jugendmedizin, Universitätsmedizin Göttingen

Prof. Dr. Christine Stadelmann-Nessler, Institut für Neuropathologie, Universitätsmedizin Göttingen

Prof. Dr. Wolfgang Brück, Sprecher des Vorstands, Forschung und Lehre, Universitätsmedizin Göttingen

Prof. Dr. Jürgen Brockmöller, Institut für Klinische Pharmakologie, Universitätsmedizin Göttingen

Date of oral examination: 23.09.2021

Table of contents

Foreword to the structure of this work	IV
List of Publications	V
Acknowledgements	VI
Abbreviations	VII
Abstract	VIII
1.0 General introduction: The diagnostic odyssey of rare disorders	1
1.1 Introduction to rare disorders and their statistics	2
1.1.1 Common diagnostic procedures in IMD and the role of newborn screening	2
1.1.2 The role of genetics, sequencing, and genomics in rare disorders	3
1.2 Introduction to metabolomics as a holistic perspective on metabolism	3
1.2.1 Planning a metabolomics study requires multiple steps	4
1.3 Introduction to liquid chromatography and mass spectrometry in metabolomics	11
1.3.1 Liquid chromatography	12
1.3.2 Mass spectrometry	14
1.4 Summary of the introduction and goal of this thesis project	17
2.0 General materials and methods	18
2.1 Materials	18
2.1.1 List of chemicals	18
2.1.2 List of bought buffers, other solutions and disposable materials	18
2.1.3 List of devices and other equipment	19
2.1.4 List of scientific programs and other software	21
2.1.5 Buffers and other solutions	21
2.2 Methods	22
2.2.1 Preparation of EDTA-plasma for metabolomics	22
2.2.2 Preparation of cell cultures for metabolomics analysis	23
2.2.3 Cell lysis of cell culture samples for metabolomics analysis	23
2.2.4 Standardization of cell culture lysates	23
2.2.5 Sample preparation for untargeted metabolomics	24
2.2.6 Processing of the Biocrates Absolute IDQ p180-kit	24
2.2.7 Cleaning the G2-S QToF mass spectrometer before use	25
2.2.8 Preparing the I-Class-LC and G2-S QToF mass spectrometer for a measurement	26
2.2.9 General conditions of the I-Class-LC and G2-S QToF mass spectrometer run	26
2.2.10 Final conditions of the ZIC-HILIC-MS method	26
2.2.11 Final conditions of the CortecsT3-RPLC-MS lipidomics method	27
2.2.12 General data analysis and data storage	28
3.0 Development of the metabolomics platform for rare neurometabolic disorders	32
3.1 Introduction to the development of a metabolomics platform	32
3.1.1 The AbsoluteIDQ p180-kit for targeted metabolomics	32
3.1.2 Introduction to the development of an untargeted metabolomics platform	34
3.2 Methods for the development of the untargeted LC-MS platform	43
3.2.1 A LC-MS metabolite library of authentic reference compounds	43
3.2.2 Generating an identification file for use in open-source metabolomics programs	44
3.3. The AbsoluteIDQ p180-kit, a general description of results	47
3.4. The in-house platform for untargeted metabolomics analysis provides broader metabolite information	49

Table of contents

3.4.1 A ZIC-HILIC column is an appropriate selection for a hydrophilic metabolite LC-MS/MS method	49
3.4.2 The human metabolite reference library for the ZIC-HILIC-MS method contains over 400 compounds	51
3.4.3 The RPLC-MS/MS platform for lipidomics analysis	54
3.4.4. Development of a sample preparation method for the untargeted metabolomics platform	56
3.4.5 Stability of the untargeted metabolomics platform and analytical validation	61
3.5 Discussion: Development of a platform to detect rare neuropediatric disorders with mass spectrometry-based metabolomics	81
3.5.1 We successfully applied the AbsoluteIDQ p180-kit to our institute	81
3.5.2 We established an untargeted metabolomics pipeline	83
3.6 Summary: Development of a LC-MS-based metabolomics platform for characterization of neuropediatric disorders.	87
4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders	88
4.1 Introduction to metabolomics for peroxisomal disorders	88
4.2 Specialized methods and cohorts: new LC-MS based platform for peroxisomal disorders	90
4.3 Results: a new LC-MS based platform for peroxisomal disorders	91
4.3.1 167 metabolites were found in plasma samples	91
4.3.2 ZSS patients present with decreased ether-linked phospholipids	91
4.3.3 VLCFA-linked lipids are increased in ZSS and X-ALD, but sphingomyelin is only decreased in ZSS patients	92
4.3.4 D-bifunctional protein deficiency type III patients have altered peroxisomal biomarkers	94
4.3.5 Summary of results in peroxisomal disorders	95
4.4 Discussion: Commercial metabolomics platform may simplify peroxisomal disorder research	96
5.0 Applications: Targeted metabolomics in X-ALD	99
5.1. Introduction to targeted metabolomics in X-ALD	99
5.2 Additional methods in X-ALD	101
5.2.1 Generation of mice samples	101
5.2.2 Human patient cohort	101
5.3. Results of targeted metabolomics in X-ALD	102
5.3.1 <i>Abcd1^{tm1Kds}</i> mice show changes in lipid metabolism of very long chain fatty acids	102
5.3.2 Metabolomic of <i>Abcd1^{tm1Kds}</i> mice reveals differences between symptomatic and asymptomatic EAE animals	104
5.3.3 Alterations of lysoPC 20:3 and lysoPC 20:4 in human X-ALD patients	108
5.4. Discussion of targeted metabolomics in X-ALD	109
6.0 Applications: A human patient with an ELOVL1-mutation	112
6.1 Introduction to a patient with a mutation in the ELOVL1 gene	112
6.2 Methods and cohorts	114
6.2.1 Cell culture methods	114
6.3. Results of a human patient with an ELOVL1 mutation	115
6.3.1 Analysis of ELOVL1 patient plasma shows decreased VLFCFA-SM	115
6.3.2 Analysis of ELOV1 patient plasma using the targeted metabolomics kit confirms reduction of membrane lipid species	117
6.3.3 ELOVL1 patient fibroblasts lipid profile reacts to external fatty acid withdrawal	119
6.4 Discussion: A mutation in the ELOVL1 gene leads to changes in the lipid profile of a human patient	122
7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1	126
7.1 Introduction to the analysis of species differences in the organic cation transporter 1	126
7.2 Specialized method for analyzing the interspecies difference in OCT1	127
7.3 Results: mOCT1 but not hOCT1 shows increased accumulation of acyl-carnitines in cell supernatant	128
7.4 Discussion: Untargeted metabolomics pointed to the transport acyl-carnitines by murine but not the human OCT1 isoform	131

Table of contents

8.0 General discussion, conclusion, and perspectives	133
8.1 Metabolomics can help characterize rare disorders together with genomics methods.....	133
8.2 We successfully created a metabolomics platform based on untargeted and targeted metabolomics	134
8.3 The metabolomics platform still faces challenges ahead.....	135
8.3.1 The access to appropriate control groups is challenging	135
8.3.2 The selection of a proper study matrix is challenging	135
8.3.3 The diversity and scarcity of rare disorders is challenging.....	136
8.3.4 The conception of bioinformatics pipelines is challenging	137
8.4 Summary of discussion.....	138
9.0 References.....	139
10.0 Supplemental material	154
10.1 Glossary of mass spectrometry terms.....	154
10.2 Lipid nomenclature	156
10.3 Supplementary Data.....	158
10.3.1 Human metabolite mass and retention time library.....	158
10.3.2 CortecsT3 fatty acid identification data.....	180
10.4 List of MetaSci metabolites for metabolite library	181
10.5 List of figures	185
10.6 List of tables	187

Foreword to the structure of this work

Dear reader,

As the title already suggests, this thesis work deals with rare neurometabolic disorders (a biological question) and liquid chromatography mass spectrometry (a more methodological aspect). Therefore, this work is structured into an introduction that leads from the general topic of rare neurometabolic disorders to a short introduction to basics in mass spectrometry. Consequently, this chapter-based thesis work further consists of multiple chapters dealing with singular aspects such as the development of the metabolomics pipeline and applications to different rare disorders. Each chapter may be read individually, however all are connected to solve the greater question of the use of mass spectrometry in rare disorders characterization. The thesis then concludes with a general discussion of the advancements of mass spectrometry-based metabolomics in rare neurometabolic disorders. In the supplementary chapters 10.1 (p. 154) and 10.2 (p.156) we have included a short glossary of important mass spectrometry terms and lipid nomenclature that may improve the understanding of this thesis.

Henry G. Klemp

List of Publications

1. Klemp, H.G., Kettwig, M., Streit, F., Gärtner, J., Rosewich, H., and Krätzner, R. (2021). LC-MS Based Platform Simplifies Access to Metabolomics for Peroxisomal Disorders. *Metabolites* 11, 347.
2. Kettwig, M., Klemp, H., Nessler, S., Streit, F., Krätzner, R., Rosewich, H., and Gärtner, J. (2021). Targeted metabolomics revealed changes in phospholipids during the development of neuroinflammation in *Abcd1tm1Kds* mice and X-linked adrenoleukodystrophy patients. *J Inherit Metab Dis*.
3. Jensen, O., Matthaei, J., Klemp, H.G., Meyer, M.J., Brockmöller, J., and Tzvetkov, M.V. (2021). Isobutyrylcarnitine as a Biomarker of OCT1 Activity and Interspecies Differences in its Membrane Transport. *Front Pharmacol* 12.

Acknowledgements

During the course of the last three years, many kind and talented researchers have contributed to me becoming a better scientist and person. Science is a field that lives from cooperation, with each person contributing to a bigger goal than ourselves. Therefore, I think acknowledging all the great people in this project should always stand in the beginning and not the end of a doctoral thesis.

Firstly, I would like to thank Prof. Dr. Jutta Gärtner for enabling me to work at her institute, for contributing to our funding and advising me, as well partaking in my thesis advisory committee. Then, I would like to thank my supervisor PD Dr. Ralph Krätzner for many great years, contributing to my career and development as a person, enabling me to follow my own ideas and teaching me.

Secondly, I would like to thank Prof. Dr. André Fischer and Prof Dr. Klaus-Armin Nave for being part of my thesis committee and many helpful hints and discussions. Similarly, I would very much like to thank Prof. Dr. Christine Stadelmann-Nessler, Prof. Dr. Jürgen Brockmöller and Prof. Dr. Wolfgang Brück for agreeing to participate in my extended examination board.

I would like to thank all my collaboration partners for all the amazing projects that we were able to achieve together which also offered many possibilities for me to learn. Here I am especially grateful for our collaboration with Dr. Matthias Kettwig, Ole Jensen, Dr. Keit Men Wong and Prof. Dr. Hupke, Dr. Susann Weissbach, as well as Dr. Vivek Venkatamarani. In addition, I would like to thank Dr. Lutz Binder of the institute of Clinical Chemistry, as well as Dr. Marcel Grapp for our collaboration in the use of the mass spectrometers and localities.

I am very thankful for Ina Martetschläger giving me the proper training in handling a mass spectrometer, for all her advice, her technical skills and her time. I would like to extend my gratitude to Antony Grüness, our former student assistant who now started his PhD in our lab. Thanks to both for all their support, making my experience in mass spectrometry so much nicer.

Every lab has a backbone, and, in our lab, this are our technical assistants. This applies to everything from bringing structure, providing advice and support and without them there would be no moving forward. Therefore, I would like to extend special gratitude towards Monika Schneider, Sabine Linnemann, Elisabeth Ehbrecht, Mark Ziegenbein, Kathrin Schreiber, Tanja Wilke and Corinna Dickel. Further thanks go to all the amazing people in the clinic for pediatrics, including all my colleagues and fellow PhD students. You made my time in the lab so much more worthwhile.

On a special note, I would like to think all those amazing people contributing to the open accessibility of science, here I would like to thank the great development teams that developed the software tools I was able to use: MetaboAnalyst and MSDial.

I would also like to thank and acknowledge for our funding which was provided partly by the Deutsche Forschungsgemeinschaft (DFG) under transregional collaborative research center (CRC) 274/1 2020-408885537, GA 354/14-1 project number 249473948 and Germany's Excellence Strategy (EXC 2067/1-390729940). I have also been very fortunate to be part of the International Max Planck Research School for Neuroscience (DFG grant GSC 226). Special thanks also go to my program coordinators Sandra Drube, Dr. Jonas Barth and Franziska Kühne for all their help and support.

Most importantly, but on a more personal note, I would like to thank my family and my friends. Every single one helped me growing up to become the person that I am. Above all, I am incredible grateful for everything my mother Sabine Klemp has done for me, my education and my life, and for my wife Melanie Nuesch, which is the best partner I could ever wish for. Thank you for everything!

Abbreviations

aa: acyl-acyl bonding in lipids
ae: acyl-ethyl bonding in lipids
BHT: butylated hydroxytoluene
cALD: cerebral X-linked Adrenoleukodystrophy
D-BPIII: D-bifunctional enzyme type 3
EDTA: Ethylaminetetraethylacetate
ELOVL: Very long chain fatty acid elongase
FIA: Flow injection analysis, also called direct injection analysis
HILIC: Hydrophilic interaction liquid chromatography
IMD: Inherited metabolic disorder
LC: Liquid chromatography
LOESS: Locally weighted scatter-plot smoother
lysoPC: Lysophosphatidylcholine
lysoPE: Lysophosphatidylethanolamine
MRI: Magnetic resonance imaging
MS: Mass spectrometry
OCT1: Organic cation transporter 1
PC: Phosphatidylcholine
PE: Phosphatidylethanolamine
Pex1: Peroxisomal biogenesis factor 1
PKU: Phenylketonuria
RPLC: Reversed-phase liquid chromatography
SM: Sphingomyelin
SPE: Solid-phase extraction
S/N: Signal to noise-ratio
TIC: Total ion chromatogram
TG/TAG: Triacylglyceride
 VLCFA: Very long chain fatty acid
VUS: Variants of unknown significance
X-ALD: X-linked Adrenoleukodystrophy
ZIC: Zwitterionic chromatography
ZSS: Zellweger syndrome spectrum disorders

Abstract

Rare diseases are a class of disorders where every single disease has a low prevalence, but due to their incredible variety, they affect more than 100 million people worldwide. A significant number of these disorders lead to neuropediatric, neurometabolic, and neurodegenerative disorders. The high diversity of disorders and widespread missing information hinder potentially life-saving diagnoses. While newborn screening successfully alleviated diagnostic problems for the most common rare disorders, the majority remain unstudied. Next-generation sequencing methods have led to a giant leap forward in rare disease detection, but undocumented variants of unknown significance can still hinder diagnosis. Additionally, the outbreak and severity of many of these disorders cannot be predicted from the genotype alone. As metabolism is the interphase between exogenous factors and endogenous factors, the holistic study of metabolism, metabolomics, may help characterize rare neurometabolic disorders working in concert with genomics techniques. Consequently, the main aim of this project was to establish a liquid chromatography mass spectrometry-based platform for the characterization of rare neurometabolic and neuropediatric disorders.

For this, we developed an in-house untargeted metabolomics platform, as well as applied the commercial AbsoluteIDQ p180-kit from Biocrates for targeted metabolomics applications. For our untargeted platform, we selected a combination of a chromatography method for hydrophilic analytes, as well as a technique for lipids. To alleviate identification problems common to untargeted metabolomics, we decided to generate a metabolite identification library for the hydrophilic method comprising 408 human disease-relevant metabolites. Our untargeted method compared well with other methods in the field and was able to show additional validation parameters commonly not studied by other metabolomics methods.

In a second part, we aimed to apply our complete metabolomics platform to several research projects in the rare disease field. Here we were able to discover new potential biomarkers for peroxisomal disorders, find potential prognostic biomarkers for developing the cerebral phenotype of X-Adrenoleukodystrophy, provide metabolic validation data to a single patient with an alteration in fatty acid elongation, as well as study the organic cation transporter 1 (OCT-1).

Using these applications, we could successfully fulfill the aims of this project and show the utility of metabolomics in characterizing neurometabolic conditions. Further research is needed to study the abilities of this metabolic platform in detail. However, this metabolomics platform may significantly contribute to the diagnosis and characterization of rare disorders, especially in association with other omics disciplines.

1.0 General introduction: The diagnostic odyssey of rare disorders

In the late 1920s, two parents made an alarming discovery regarding their children. Even though she had developed normally, the oldest child, a girl, failed to acquire speech as she was already three years old. Her younger brother grew normally for a couple of months but then lost interest in the environment and never learned to sit or walk. While physically, both developed normally, there seemed to be something affecting their mind. At the same time, the parents started recognizing a “strange musty odor” of their children’s urine. After severe worsening of the boy’s symptoms, the parents went from one medical professional to another. However, all standard laboratory work and clinical assessment did not result in a diagnosis. The parents even saw advice from self-proclaimed natural healers or prophets. But they did not give up. The father remembered that he had heard lectures from Prof. Fölling, a specialist in metabolic disorders in college. Thus, they started contacting Prof. Fölling, and he agreed to see the children. While classical laboratory work on the children remained normal initially, Prof. Fölling decided to do one last test for ketones in the urine. Ketonuria is usually caused by metabolic abnormalities such as those found in diabetes or starvation. It was imaginable, those might (when remaining for a longer time) lead to neurologic effects. The skilled physician then made an astounding discovery: Upon addition to the test solution (ferric chloride), the solution turned green instead of the classical red-brown color. This reaction has never been described before. In cooperation with the parents over multiple months, Prof. Fölling managed to isolate the substance causing this reaction from “about 20 liters of urine”. After extensive chemical analysis, he determined the substance to be phenyl pyruvic acid, a phenyl ketone. Later, he went to multiple psychiatric wards and found further patients with this disorder. In a research article, he then coined the term “imbecillitas phenylpyruvica” for this disorder, which was later renamed phenylketonuria (PKU). Partly adapted from Centerwall and Centerwall, 2000; Fölling, 1934

In 1947 Dr. Jervis uncovered that the accumulation of phenyl pyruvic acid resulted from defective phenylalanine catabolism (Jervis, 1947). Based on this, Dr. Bickel (and others) developed a phenylalanine-low diet that significantly reduced phenylketonuria and substantially improved symptoms (Bickel et al., 1954). Subsequently, it was proposed that early recognition of the disorder and early dietary changes may prevent symptoms from progressing and enable the children an everyday life. For this early detection, Dr. Guthrie developed a bacterial growth inhibition test that could be conducted from heel-prick blood spotted on a filter paper (Guthrie, 1992). The Guthrie test enabled massive newborn screening, which was formally introduced for the first time in Massachusetts in 1962 and was introduced to Germany in the 1970s. Since then, the newborn screens have been steadily improved, and more diseases included.

The history of phenylketonuria is an excellent example of stories of many neurometabolic disorders. Typical “hallmarks” include:

- normal development until a certain age
- the combination of a neurologic phenotype with blood or urine abnormalities
- multiple occurrences in the family (inheritance)
- first recognition by primary caregivers, but no abnormalities found in routine medical screening
- diagnostic odyssey
- discovery of biochemical changes by specialized tests
- discovery of genetic or metabolic causes
- sometimes: development of therapy options that can even prevent disease progression
- sometimes: inclusion into a screening program

In the above example, it might seem evident that one fundamental necessity for proper patient treatment is diagnosis. However, as seen in this example, the diagnosis of uncommon disorders can be cumbersome.

1.0 General introduction: The diagnostic odyssey of rare disorders

For these uncommon disorders, the legal term “rare disorder” was coined. The following sections will include further definitions of rare disorders, diagnostic opportunities, and their connection to the “omics” era.

1.1 Introduction to rare disorders and their statistics

Rare disorders are an umbrella term for a multitude of different disorders with strongly varying causes. Legal definitions vary from country to country; European legislation defines a rare disorder to have a prevalence of less than 5 in 10,000 or 1 in 2000 (Richter et al., 2015). However, the term rare disorder can be misleading, and there are also no clear definitions of what defines a disorder, as opposed to single patients with medical alterations. Thus, rare disorder numbers vary greatly and range from 5000 to 9000 entities, though rare disorders are highly diverse no matter the exact number (Ferreira, 2019). Consequently, the cumulated worldwide prevalence of all rare disorders is guessed to be in the range of multiple hundred million patients (Nguengang Wakap et al., 2020). The high diversity of this group and insufficient information about their causes might contribute to an extraordinarily high number of misdiagnoses. This was further analyzed by the EurordisCare studies 1-3 conducted by the European rare disease patient advocacy group EURORDIS from 2002-2008 in 6-16 exemplary rare disorders (<https://www.eurordis.org/publication/voice-12000-patients>, accessed 05.08.2021). An astonishing number of 25% had to wait longer than 5 years until reaching a diagnosis, in some cases, even up to 30 years. Also, around 40% of patients received initial misdiagnoses. Apart from the burden for patients, this creates an economic burden for society as well. The average cost of around 80,000 € yearly per patient in Germany has been described based on 8 major rare disorders (Linertová et al., 2017). For pediatric care, this has prime importance as it is guessed that over 50% of neonatal or infant death is caused by rare malformations or genetic conditions (Stevenson and Carey, 2004). Numbers vary greatly; nevertheless, most authors agree that the most significant causes for rare disorders are genetic (Ferreira, 2019). Rare genetic disorders also include the group of inherited disorders of metabolism, where most are also neurometabolic neuropediatric disorders (Filiano, 2006; Rath et al., 2010; Willemsen et al., 2016).

Diagnostic capabilities must be thus further extended. The international rare disease research consortium (IRDIRC) aims in their goals for 2017-2027 for a reduction of diagnosis time to less than one year (Austin et al., 2018). EURORDIS has further confirmed this in their 2030 goals (Kole, A., Hedley V., et al., 2021). For this, more efficient and holistic diagnostic possibilities are needed. Here we aim to delineate further diagnostic principles and opportunities for the group of inherited disorders of metabolism (IMD).

1.1.1 Common diagnostic procedures in IMD and the role of newborn screening

As seen in the introduction, the first significant achievement of rare disease diagnostics was the invention of the Guthrie method for the newborn screen of phenylketonuria (Guthrie, 1992). In the beginning, those tests were biochemical, with the Guthrie-test relying on bacterial growth inhibition, or enzymatic tests with photometric or fluorescence detection, for instance, in Galactosemia (Beutler and Baluda, 1966). With the upcoming of these methods that allowed for newborn screening, Wilson and Jungner defined criteria which diseases should be screened for, that were later widely adopted (Wilson and Jungner, 1968). Apart from other requirements, especially the treatability of the disorders played a big role, which in fact excludes many rare IMDs. In Europe, most nations have established national screening programs, although partly analyzing different diseases (Therrell et al., 2015).

1.0 General introduction: The diagnostic odyssey of rare disorders

The advances in newborn screening and their wide availability would not have been possible without immense technological achievements. One was the introduction of liquid chromatography mass spectrometry (LC-MS) into newborn screening, first developed to detect acyl-carnitines (la Marca, 2014; Millington et al., 1990). This massively increased detection speed, needed to process the 100-1000 samples that arrive at the laboratories every day (Rousseau et al., 2012). However, the typically up to 40 parameters (in most countries less) measured cannot diagnose all 1500 IMDs classified (Ferreira et al., 2021). Thus, most rare IMDs remain undiscovered and undiagnosed by the newborn screen, and other diagnostic procedures are necessary.

1.1.2 The role of genetics, sequencing, and genomics in rare disorders

In the previous sections, we established that many rare inherited disorders of metabolism are monogenetic. Clearly, genetic screens are essential for the diagnosis of these disorders. Simple genetic sequencing techniques, such as Sanger sequencing, have long been used to confirm biochemical results (Katsanis and Katsanis, 2013). Albeit recent advances in next-generation sequencing have helped rare disease diagnostics a giant leap forward (Turro et al., 2020; Vinkšiel et al., 2021; Wright et al., 2018). Techniques such as whole-exome sequencing or whole-genome sequencing have arrived as almost standard methods for examining undiagnosed rare disease patients. While indeed highly beneficial to patients and researchers alike, these techniques have pitfalls (Vinkšiel et al., 2021). One of the most significant pitfalls is the inability to judge the relevance of findings conclusively. In these screens, results of altered genes are rather common. Even after filtering using patient-parent trio sequencing (Zhu et al., 2015) and other advanced filtering techniques (Sevim Bayrak and Itan, 2020), a higher number of potential causes can remain (Wright et al., 2018). If these variants are not previously documented and cannot be judged for their pathogenicity using computational predictors (McLaren et al., 2016), a high number of those variants of unknown significance (VUS) remain (Federici and Soddu, 2020; Sullivan, 2021). These VUS require further testing; in the case of IMDs, studying the potentially affected enzymes and metabolic pathways is the logical next step. Still, those disorders can be complex, affecting multiple pathways, and can show interdependently changed metabolites. New techniques to further elucidate changes in the metabolism in a holistic manner being comparable to genomics are needed. One approach would be the use of metabolomics.

1.2 Introduction to metabolomics as a holistic perspective on metabolism

To our knowledge the first terms that used the suffix -ome, were “biome” and “genome” (Clements, 1917; Winkler, 1920). In more recent times, this suffix has then been further concretized by system biology. One possible definition is that -ome is the holistic perspective on a system that shows emergent properties (Mazzocchi, 2012). An emergent system is a system where its parts, based on their interaction effects, gain properties that they would not possess independently. Typically omic-disciplines study the -ome using high-throughput technologies (Micheel et al., 2012).

Thus, the metabolome is the holistic study of metabolites and their interaction effects (Fiehn, 2002; Lindon et al., 2000; Oliver et al., 1998). Most definitions for metabolism include small molecules of up to 1500-2000 Da (Wishart et al., 2013). While there are many deviating definitions, for this work, we define a metabolite as an endogenous molecule below 2000 Da that partakes in processes necessary to sustain life in that organism.

1.0 General introduction: The diagnostic odyssey of rare disorders

Metabolism itself provides interesting insights into diseases, as it is influenced by internal (genome, proteome), as well as external factors (e.g. nutrition) and thus provides a “link between phenotype and genotype” (**Figure 1**; Fiehn, 2002). It is to note that there are also other contributors, that do not follow the scheme below. These include effectors such as ribozymes (Walter and Engelke, 2002) or epigenetic modifications (Cavalli and Heard, 2019).

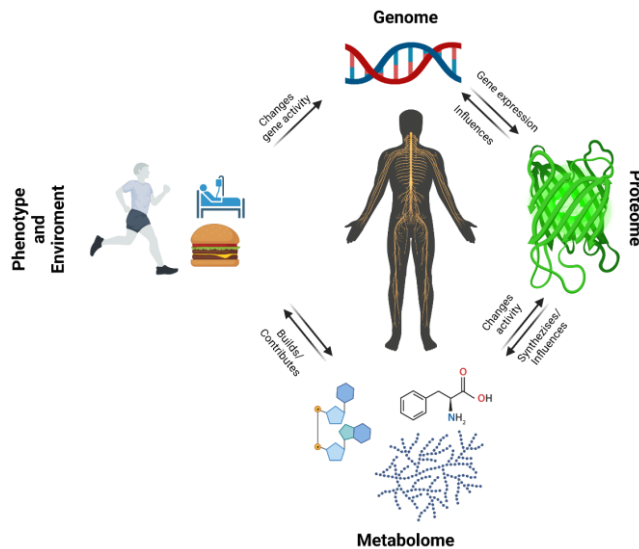


Figure 1 The metabolome is the interface between genotype and phenotype. The metabolome is influenced by endogenous factors such as the proteome and genome, but also by external factor such as nutrition and physical activity. In return, the metabolome can influence genome and proteome on its own. Created with BioRender.com

We can separate two main approaches to study metabolism with an omics-perspective: Untargeted and targeted metabolomics.

Untargeted metabolomics is an approach where the set of analyzed metabolites is not known and not defined beforehand. The approach allows an unbiased overview of the metabolome and is especially useful for the characterization of unknown phenotypes or the search for biomarkers (Fiehn, 2002; Schrimpe-Rutledge et al., 2016; Vinayavekhin and Saghatelian, 2010). In its most extreme form, as metabolic fingerprinting, single metabolites are not identified but rather interpreted on a feature level. A feature in this context can be a signal of any kind from the used detector, typically assigned identifiers specific to the used detector.

Targeted metabolomics is studying a predefined list of metabolites, enabling the techniques to be more sensitive and specific (Roberts et al., 2012). Due to different reasons (explained in chapter 1.3.2.6, p. 16), only targeted metabolomics can be fully quantitative.

1.2.1 Planning a metabolomics study requires multiple steps

As mentioned above, most -omics fields utilize high-throughput techniques to benefit from broader coverage. Many papers deal with the correct planning of a metabolomics analysis, and many specific approaches exist (Barnes et al., 2016; Jacyna et al., 2019; Shriver, 2016). In a synopsis of the literature, metabolomics analyses require the definition of (**Figure 2**):

- The purpose of the study
- The appropriate analytical technology
- The cohort
- Sample preparation guidelines
- Data analysis

1.0 General introduction: The diagnostic odyssey of rare disorders

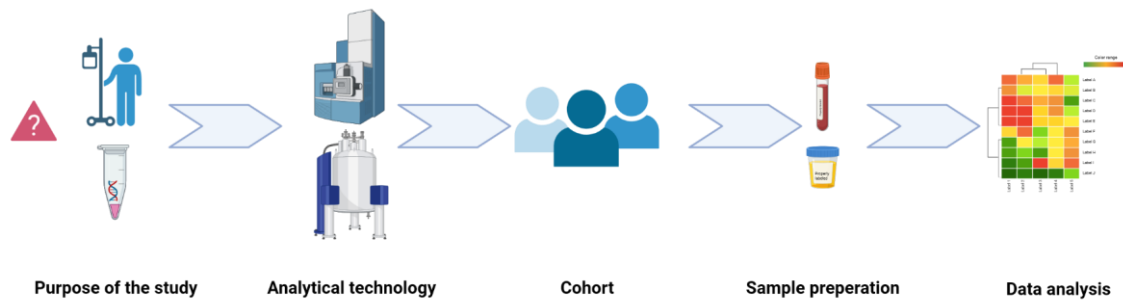


Figure 2 Planning a metabolomics study requires the definition of multiple steps. Planning of a metabolomics study requires the definition of its purpose and then selecting the appropriate technology, cohort, sample preparation and data analysis. Created with BioRender.com

1.2.1.1 Metabolomics requires definition of an analytical purpose

Independent of the specifications of an analytical platform, every project starts with a general project initiative cause or purpose. Frequently, initiative causes can be:

- a basic scientific research question (e.g., “Why is the brain so reliant on glucose consumption?”)
- or in the case of our clinically oriented neuropaediatric institute, a specific patient case (e.g., “An unconscious pediatric patient with honey-sweet urine is charged into intensive care, what is the cause for the disorder and how can we treat it?”).

The purpose defines all further steps, such as requirements for sensitivity or specificity. Noteworthy is also if the sample can be disintegrated or needs to remain intact. For example, blood samples may be completely disintegrated in a blood biomarker study, but if the dynamic metabolism in a human brain should be studied, disintegration is not recommended.

After the initiative cause is determined, the first experiments are performed to build a hypothesis. For this, very open screening methods are used. In the case of metabolomics, typically untargeted metabolomics methods are utilized. These methods are also commonly only semiquantitative. Based on the newly generated hypothesis, specific metabolites are picked and analyzed in confirmatory analyses. In a second step, targeted metabolomics analysis consists of ideally fully quantitative methods and can then confirm or disprove the initial hypothesis (Jacyna et al., 2019).

1.2.1.2 Metabolomic studies need defined cohorts

In metabolomic studies, designing an appropriate cohort is very important, as the metabolome reacts to many influences (Johnson and Gonzalez, 2012; Kochhar et al., 2006; Lenz et al., 2004). In studies from biofluids (to less extent also tissues), impacts can include:

- Age
- Sex
- Nutrition (including liquid intake)
- Bodyweight, body fat, and muscle mass
- Physical activity
- Sleep deprivation
- Iatrogenic influences, e.g., drugs
- Diurnal effects
- Ethnicity

1.0 General introduction: The diagnostic odyssey of rare disorders

Thus, the cohort should be carefully selected and should observe a specialized protocol.

In the case of the (typically) monogenetic inherited disorders of metabolism (IMD), metabolomic alterations are usually very fulminant (Miller et al., 2015). Changes in the range of 10-1000-fold are usual, and most of the discussed effects are typically smaller. Consequently, in IMD research and diagnosis, cohort selection is accepted to be less strict.

In IMD research and diagnosis, patient and control cohorts are usually very small, with most studies being single-patient, being lower than the recommended minimum of 5 biological replicates (Sumner et al., 2007). Additionally, invasive procedures such as blood draws cannot be performed in pediatric research without a medical indication. Therefore, specific control cohorts include patients that are not healthy in the original sense but are not assumed to have the same alterations as the patient of interest (Griggs et al., 2009). For example, in peroxisomal disease research (which mainly affects lipid metabolism), phenylketonuria patients who have been admitted for routine check-ups may be used. Unfortunately, there is often a large age-mismatch between target patients and controls. The use of the patients' parents as controls is also not uncommon. While this represents a challenge for research in rare neuropediatric disorders, it is alleviated by the above-mentioned strong changes expected in IMDs. To concretize, the deciding factor is the statistical power of the analysis. If the metabolic effect is large enough, smaller sample sizes can be acceptable (Blaise et al., 2016).

1.2.1.3 *The analytical technology needs to match the studies' purpose*

After defining the purpose of the analysis, requirements regarding sensitivity, specificity, and destructiveness are defined. The next step is to determine an analytical technique that can generate distinguishable features. In metabolomics, nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) are the two most common techniques (Barnes et al., 2016). The most significant advantage is that NMR works in situ and is not destructive. Therefore, even intact tissues or live organisms can be measured (Emwas, 2015). In vivo NMR is also known as magnetic resonance spectroscopy, MRS (Vingara et al., 2013). However, not all elements (or isotopes) generate NMR signals, and the signal intensities are comparatively low, reducing the sensitivity of detection.

The other technique is mass spectrometry (MS). While this technique is destructive, it typically performs with higher sensitivity, is amenable to all elements and a more comprehensive range of compounds, and has higher linearity and dynamic range (Aretz and Meierhofer, 2016; Dettmer et al., 2007). Most biomarker studies are performed on biofluids; thus, the destructiveness of the analysis does not play a significant role (Califf, 2018). Based on these vital points, most metabolomics studies are conducted using mass spectrometry.

Mass spectrometry can be combined with a wide range of other methods that increase its specificity and analytical power. The most predominant are chromatography techniques (Croley et al., 2012). There are different chromatography techniques, each dependent on the physical state of the analyte and separation setup. In metabolomics, the most common are liquid chromatography (LC) or gas chromatography (GC), with the name defining whether the analytes are in a liquid or gaseous state. GC needs analytes in the gaseous state; hence the analytes need to be volatile (Dettmer et al., 2007). However, a wide range of metabolites are not volatile, so that GC requires derivatization to change their physicochemical properties and make them amenable to analysis (Halket et al., 2005). The derivatization removes the metabolome from its native state, and many metabolites are excluded by derivatization.

1.0 General introduction: The diagnostic odyssey of rare disorders

Even though it has these shortcomings, it is a very reliable and stable technique that can analyze complicated analytes efficiently. For our analysis, we decided to use LC techniques commonly not requiring derivatization, allowing samples to remain in a more native state.

In summary, for this project, we decided to use mass spectrometry combined with liquid chromatography, LC-MS.

1.2.1.4 Sample preparation is demanded to be adapt to the analytical technology

Independent of the used analytical techniques, every method requires a clear definition of analytical targeted groups and principles for acquiring analytical material. Apart from selecting experimental groups that fit the pipeline's initiative cause, the appropriate material collection is key to a successful project.

The appropriate material collection is always dependent on the target analytes, the analytical matrix, and the analytical platform itself (Fan, 2012; Vuckovic, 2012, 2013).

Starting with the analytes, generally, small hydrophilic molecules are highly unstable (Hu et al., 2020). This especially applies to metabolites with high intrinsic energy, such as metabolites that participate in energy converting reactions of primary metabolism. Described here are ATP, NAD, FADH₂, hexose/pentose phosphates, and similar compounds (Fu et al., 2019; Mičová et al., 2017; Trammell and Brenner, 2013). Their biological purpose is to transfer energy, and they can do this even when not intended by the experimental protocol. Among those high energy metabolites, common reactions are the oxidation reactions leading to loss of protons and electrons (e.g., $NADH \rightarrow NAD + H^+ + 2 e^-$) or hydrolysis causing loss of high energy phosphates (e.g., $ATP + H_2O \rightarrow ADP + P_i$). Other reactions include deamination (e.g., $Gln \rightarrow Glu + NH_3$). Lipophilic molecules are generally more stable, mainly if not containing previously mentioned groups. Triacylglycerides containing 3 fatty acids bound by carboxy-esters to the polyol glycerol do not hydrolyze under normal conditions but require high temperatures and extreme pH (Giegel, 1974; Matthan et al., 2010). Aromatic groups also stabilize more lipophilic compounds due to their delocalized π -electrons. Only unsaturated lipids (double bonds), especially when they are occurring more than once in a molecule (e.g., polyunsaturated fatty acids), are more reactive. Double bonds are reactive, can quickly oxidize to form reactive carbonyl-radicals, and oxidize other molecules in the sample (Metherel and Stark, 2016). For example, the hardening of oil paint relies on the formation of lipid radicals in between α -linolenic acid molecules (3 double bonds) contained in the used linseed oil (Boelhouwer et al., 1967). The stability of all these metabolites can be affected by pure physical/chemical effects, but of course, also by enzymes. Hence, the analytical matrix changes the requirements for appropriate sample collection.

Biological materials such as cells and tissues can demand complex guidelines (Cuperlović-Culf et al., 2010). Both have a high content of working enzymes influencing the metabolites. Especially when natural membranes (cell and organelle membranes) are destroyed, a highly reactive environment forms. Enzymes responsible for degradation (e.g., the lysosome) or oxidation (e.g., the Peroxisome or the Mitochondria) are freely reacting in the suspension. Thus, one process with key importance is the quick deactivation of enzymatic reactions to keep the metabolites in their original state (metabolic quenching; (Vijay Kapoore et al., 2017; Wahrheit and Heinzle, 2014).

In cells and tissue, this requires another component: making the inside of cells and tissues accessible to extract analytes.

1.0 General introduction: The diagnostic odyssey of rare disorders

Most obviously, in tissue metabolomics, the targeted organs need to be harvested quickly. Their location (e.g., the location of the brain inside the skull) or their contamination with biofluids like blood (e.g., the liver needs to be drained of blood) can impede the process. Tissue samples also commonly require manual or automatic physical disintegration, different homogenizers based on beads or blenders are used. Frozen tissue can also be ground using a mortar (Römisch-Margl et al., 2012).

Handling cell culture samples is more facile, commonly only requiring removing cell culture supernatant and a washing step. All these procedures should be occurring as fast as possible, so the metabolic quenching can occur to preserve a pristine metabolic state.

For metabolic quenching, multiple methods are used. The most ubiquitous is applying low temperatures to slow enzyme activity by either applying precooled buffers or placing the sample in a low-temperature container (Pabst et al., 2010).

This method is mild, and it does not lead to denaturation of enzymes enabling precise analysis of these. However, it is not very effective. Cooling is slow, and enzymes still show residual activity, as temperatures lower than 0°C cannot be used due to the freezing of aqueous buffers. While the freezing-point depression with salts (e.g., sodium/potassium chloride in road antifreeze) or polyols (e.g., ethylene glycol or glycerol in car-anti-freeze) is possible, it typically affects further preparation steps negatively. In the case of tissue extraction from organs, advanced techniques have been developed, including the application of a focused microwave beam applied to the in-situ brain, leading to the denaturation of enzymes and metabolic quenching (Epstein et al., 2013).

Instead of thermal techniques for metabolic quenching (especially in cell culture), typically organic solvents are used, leading to instant denaturation and deactivation of enzymes (Bi et al., 2013; Dunn and Winder, 2011). Typically, methanol or acetonitrile are used alone or in combination with each other. Sometimes also ethanol or butanol is added (Löfgren et al., 2016). Still, denaturation also leads to more inhomogeneities of the samples; not all of the samples can be transferred by pipetting.

Metabolic quenching can also use strong acids such as sulfosalicylic acid and perchloric acid. Perchloric acid is often used to determine glucose or lactate (Astles et al., 1994), but it is not used in metabolomics due to the high risk of hydrolyzing other metabolites.

If not already occurred, for instance by applying organic solvents, cells need to be lysed. Unfortunately, the most efficient lysing agents, detergents such as sodium dodecyl sulfate (SDS), cannot be used due to their strong bonding to LC columns and MS signal depression. Instead, only physical techniques are used. Common is the use of hypotonic buffers that lead to osmotic stress and cell lysis. Repeated freezing and thawing steps to create expansion stress and cell lysis are utilized, too (Tansey, 2006).

Biofluids such as blood, urine, or CSF do not face this problem and are mostly stable. Only blood requires further workup, as separating cellular components from plasma is necessary. Cellular enzymes in EDTA full blood partly remain active (Jobard et al., 2016). Thus, removing the cells quickly is essential. Especially glucose and lactate are metabolites that are known to be affected, as the strongly glycolytic erythrocytes stay active in drawn blood.

1.0 General introduction: The diagnostic odyssey of rare disorders

Prolonged storage may also lead to lysis of erythrocytes (hemolysis), contaminating the plasma with intracellular compounds (Malm et al., 2016). Very importantly, full blood should never be frozen, as it leads to even more substantial hemolysis after thawing.

As previously mentioned, reducing sample extraction time, proper metabolic quenching, and storing the samples appropriately are essential factors for keeping metabolites stable. Additionally, other agents may be added to stabilize the compounds. One agent commonly used in metabolomics is butylated hydroxytoluene (BHT), an aromatic molecule that can scavenge free radicals, especially for lipid analysis (Burla et al., 2018; Metherel et al., 2013). Also, EDTA can help stabilize the metabolome, removing metal ions from reactions (Khadka et al., 2019). Most enzymes have a metallic cofactor, thus by chelation with EDTA, the enzymes are deactivated. Also, by means of the Fenton reaction, metallic ions (mainly Fe^{2+}) create free radicals, thus removing them also leads to lower oxidation (Winterbourn, 1995). Both agents are also relatively non-toxic and partially even used in food.

Tissue, cell, and urine metabolomics face a similar problem: samples can contain diverging amounts of metabolites. Urine is a biofluid that can have highly altered metabolite concentrations. Depending on the intake of fluids or an unrelated unphysiological reason (e.g. ethanol-intoxication or cardio-vascular medication), urine can be concentrated in the kidney extensively. As we are not interested in the osmotic effects, this effect has to be normalized. In clinical chemistry, mainly the molecule creatinine is used, as it shows stable glomerular excretion and is not reabsorbed (Sawant et al., 2018). Creatinine is formed in a nonenzymatic irreversible reaction from creatine, a compound used primarily in muscles and brain as creatine-phosphate for short-term energy storage (Wyss and Kaddurah-Daouk, 2000). Unfortunately, creatinine excretion is dependent predominantly on muscle mass and can also be influenced by muscle damage as for instance, in accidents (crush-syndrome) or in response to drugs (e.g., propofol (Rácz et al., 2012)). Consequently, in newer urine metabolomics studies, exceedingly other markers such as total osmolality are used as more stable markers (Khamis et al., 2018).

Tissue and cells show natural variations in their metabolite contents due to differences in cell numbers or cell sizes. Non-adherent cells can be counted in aliquots before lysis, but adherent cell numbers cannot be counted. Cell numbers can be partially controlled in cell culture, as the seeding cell number can be determined; however, cells in culture keep dividing (and/or dying) so that seeding cell number does not necessarily correlate with cell number before cell lysis (Muschet et al., 2016). Tissues can be accurately weighed, and buffer volume adjusted to match concentrations between different samples. However, the tissue homogenization and/or cell lysis procedures are never complete and can vary so that metabolite content should be primarily determined in the lysate. Mainly described as a proxy for metabolite content is the use of total protein determination (Dettmer et al., 2011). Protein is generally accepted to be correlated to the concentration of metabolites and can be determined by simple means such as the bicinchoninic acid reaction (BCA). However, when analyzing protein concentration, protein denaturation using organic solvent crushing presents a problem, creating inhomogeneities.

1.0 General introduction: The diagnostic odyssey of rare disorders

Especially for lipid analysis, the total lipid can be determined using the phosphovanillic acid reaction, albeit this reaction is not particularly sensitive and may not work in low-lipid samples such as cell culture samples (Zöllner and Kirsch, 1962). Quantification of cellular DNA was also described as a normalization method (Silva et al., 2013).

Each of the discussed factors (sample preparation, sample matrix, and respective metabolite stability) can also influence which metabolites can be detected by untargeted metabolomics methods. While metabolites may be not excluded actively, these factors can exclude metabolites passively just by their analytical design.

In summary, sample preparation is a crucial step in metabolomics that can heavily influence results. Sample preparation needs to observe metabolite stability and must be carefully optimized. Therefore optimizing sample preparation is an important part of this work. Sample preparation requirements specific to LC-MS and our metabolomics platform are found in the next chapters.

1.2.1.5 Adapt data analysis is a core interpretative step

Data analysis in metabolomics is complex, and many approaches exist (Bartel et al., 2013; Cambiaghi et al., 2017; De Livera et al., 2013). Very superficially, it can be separated into nine stages:

1. Metabolomic features need to be extracted, separating single entities from a continuous data flow, denoted using a unique identifier and an instrument signal intensity is associated with it.
2. In most studies, features need to be identified to be specific metabolites and redundant data is removed (one metabolite often forms multiple features)
3. Data needs to be checked for data integrity and often exported into other data formats for statistical exploration
4. Data needs to be filtered, meaningless features and outliers need to be removed according to predefined neutral guidelines to ensure meaningful interpretation. E.g., signals with a low signal to noise ratio are removed
5. Data needs to be transformed and normalized. In most statistical methods, metabolites are compared so that they need to have the same range and are centralized around a zero mean. Data normality is a further requirement for most statistical testing.
6. General chemometric methods such as principal component analysis (PCA) are used for assessing data quality and basic separation
7. Univariate (e.g., Student's T-test) and multivariate (PCA, neuronal networks) analysis methods are used for further analysis.
8. Especially relevant metabolites are selected and subjected to further analysis, including classical literature search or metabolite pathway analysis
9. The conception of further confirmatory studies (that may use targeted metabolomics)

In clinical research and especially in rare disease research, experimental groups are typically very small. Most statistical techniques were conceived for more extensive studies, with multiple hundreds of samples. In our research, most studies are single-patient studies, typically also with small control cohorts. Therefore, statistical evaluations should proceed with utmost care.

1.0 General introduction: The diagnostic odyssey of rare disorders

Predominant analysis types in rare disease research are two-group comparisons (“healthy”-vs.- “diseased”). Thus, multiple t-tests, ANOVAs, and volcano-plots (displaying logarithmic p-values depending on logarithmic fold changes) are common.

Heatmaps together with clustering techniques are used as well. For hypothesis generation, partial least squares discriminant analysis (PLS-DA) is used (Fonville et al., 2010). This is a biased technique that can be used to determine metabolites that lead to the separation of two groups under the assumption that there is a separation.

However, while these techniques are the easiest to interpret, all techniques rely on linear relationships between the metabolite and the phenotype or with each other. Many effects in metabolism and other omics fields are not linear and involve network effects. Therefore, techniques such as artificial neural networks are more adapted to analyze this data (Pomyen et al., 2020). As there are numerous pitfalls and optimization is needed, neural networks in clinical metabolomics are still in development (Mendez et al., 2019).

The key to successful metabolomics studies is then to select metabolites that should be analyzed further. Here literature search can be used or, more importantly, pathway analysis tools such as pathway enrichment (Karnovsky and Li, 2020). In multi-omics studies, sequencing data can massively help to associate gene and metabolite differences to diseases.

Typically, further confirmatory analysis follows to further delineate and prove the generated hypothesis.

1.3 Introduction to liquid chromatography and mass spectrometry in metabolomics

Liquid chromatography mass spectrometry (LC-MS) is the most prevalent analysis technique in metabolomics. It typically enables the study of the widest variety of intact metabolites, as opposed to NMR, while it has superior sensitivity and dynamic range (Dettmer et al., 2007). It consists of two parts acting together:

- a liquid chromatography system that separates molecules based on their chemical interactions
- a mass spectrometer that separates molecular ions based on physical interactions and mass/charge ratio (Gu, 2019).

These techniques are not kept separate, but the chemically separated molecules are directly transported into the mass spectrometer so that the chromatographic separation properties can now be associated with the mass spectrometry properties. These combined LC-MS data can then also be used for the identification of metabolites.

1.3.1 Liquid chromatography

(Adapted from "Introduction to liquid chromatography", Snyder and Dolan 2009)

Liquid chromatography (LC) is a technique that separates molecules based on their chemical properties. LC uses a tube (usually made from stainless steel or PEEK) filled with particles of a specific size and a defined chemical surface (LC-column, **Figure 3**). The particles are typically composed of porous silica, where the silica can also be modified by functional groups such as long alkanes or zwitterionic groups. Other materials exist; however, they are more rare (e.g. porous graphitic carbon; (Pereira, 2008). The columns have a large surface area that can interact with material moving through the column (stationary phase).

A fluid is pumped through the column that carries the analytes (mobile phase). Analytes then interact with the stationary phase and mobile phase in varying ways based on their chemical properties (retention).

Depending on the strength and number of interactions of the analyte with the stationary phase, the analytes are retained in the column for a time that is specific to their chemistry (retention time). Thus, the retention time can be used as a molecular identification parameter specific to the analyte's molecular interactions. The retention time and separation efficiency can be influenced by multiple parameters that need to be thoroughly optimized for designated purposes. Influences include the length of the column, the surface modifications, temperature, and mobile phase composition. The mobile phase composition is also a parameter that can be altered during the run to optimize separation (gradient). Different techniques for LC separation exist. In mass spectrometry-based metabolomics, the two most predominant are reversed-phase (RP) and hydrophilic liquid interaction chromatography (HILIC). RP is optimal for the separation of lipophilic compounds, separating them using a column modified with unpolar groups (e.g., an alkane chain such as octadecane, C18). HILIC is optimal for the separation of hydrophilic compounds and separates them on a chromatographic column that carries hydrophilic groups (zwitterionic groups, such as sulfobetaine). These hydrophilic groups form an aqueous layer that can interact with hydrophilic analytes (Alpert, 1990). Each of them requires different optimization strategies and different sample preparation strategies to make them amenable for analysis.

1.0 General introduction: The diagnostic odyssey of rare disorders

While there is an excellent separation of analytes using LC, there is still a large range of analytes with the same retention time (coeluting analytes) in every method. Thus, a different technique is needed for better identification, such as mass spectrometry.

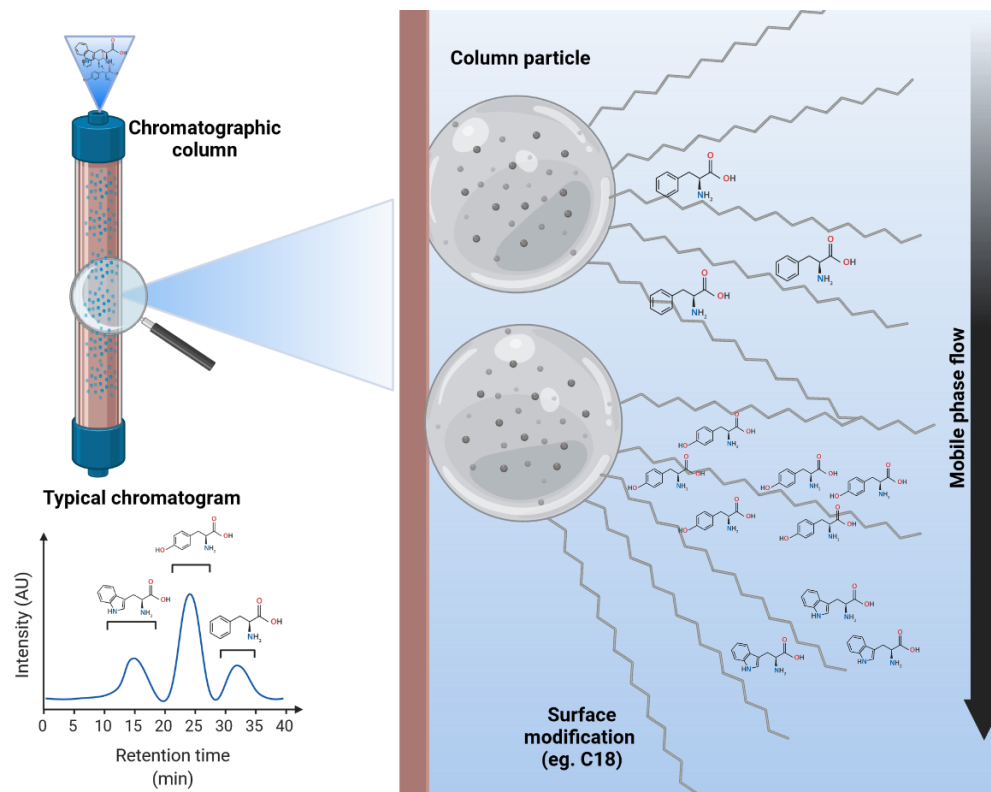


Figure 3 Liquid chromatography is a separation technique. Liquid chromatography (LC) requires a chromatographic column filled with a chemically modified particulate stationary phase. A liquid mobile phase solvent carries the analytes, which are separated by their interaction effects of the column and mobile phase. Thus, LC is a pure separation technique demanding a further detector for complete analysis. Created with BioRender.com

LC by itself also needs a way to record analytes that exit the column (detector), such as a mass spectrometer. Other techniques such as fluorescence detection, UV detection are available but scarcely used in metabolomics.

1.3.2 Mass spectrometry

(Adapted from "Massenspektrometrie", Gross, 2013)

Mass spectrometry is a technique to separate analytes based on their physical properties. The essential requirement is that the analyte must be charged/ionized. Ions can be manipulated using electromagnetic fields (Lorentz force), which is essential for the transport inside the spectrometer and the analysis itself. Most mass spectrometers consist of at least 3 parts: the ionization source, a mass analyzer, and the detector (**Figure 4**). Some mass spectrometers also include a fragmenting compartment (collision cell), which can fragment molecular ions, and an additional mass analyzer. To avoid unwanted collisions with air molecules, the instrument is evacuated.

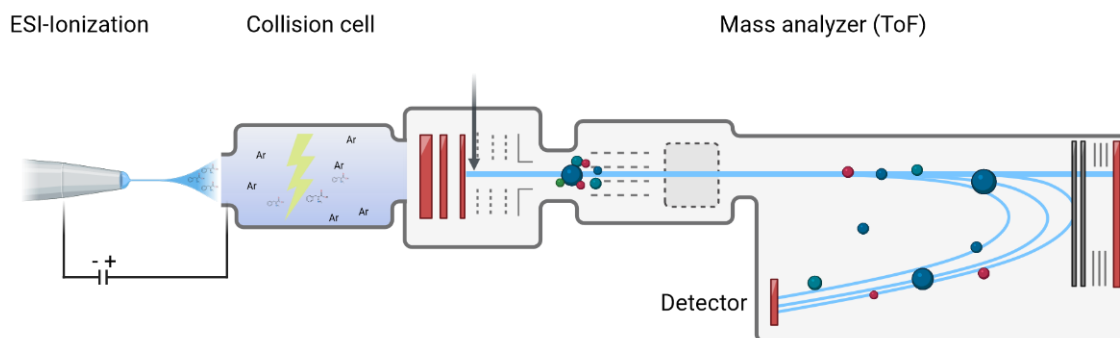


Figure 4 Mass spectrometers separate analyte ions based on their mass to charge ratio. The scheme represents a typical mass spectrometer, based on a time-of-flight mass (ToF) analyzer. The analytes enter the MS on the left side through the ESI capillary and become ionized. Molecular analyte ions can be fragmented in a collision cell and then get drawn into a mass analyzer. The mass analyzer separates ions based on their mass to charge-ratio using different physical principles. Ions are conclusively entering the detector and intensities and mass/charge information is then yielded. Created with BioRender.com.

1.3.2.1 Ionization is key to mass spectrometry analysis

The ion source is responsible for ionizing incoming analyte molecules to make them amenable for mass analysis. In LC-MS, most ion sources use electrospray ionization (ESI). Other sources exist but are typically limited to more specialized applications, such as atmospheric pressure chemical ionization (APCI; Kostianen et al., 2003). After leaving the LC, the mobile phase with analytes enters a narrow bore capillary with an open-end (ESI-capillary). The capillary is charged with a high voltage and surrounded by a (heated) nitrogen gas flow. First, by running through this needle, a cone forms (Taylor cone) at the tip from which a spray originates. The spray consists of analyte and solvent drops. As there is a high voltage applied, strong forces inside the drops lead to cycles of breaking apart by charge repulsion until only analyte ions remain. Unionized molecules (e.g., solvents) are removed by vacuum suction; ionized molecules enter the mass spectrometer. The applied voltage can be positive or negative (positive and negative ionization mode), leading to either positive or negative analyte ions. These ions can be complex and depend on the analyte's chemical properties. For example, acidic molecules predominantly form negative ions by proton donation; basic molecules accept protons and form positive ions. Ions that cluster the analyte molecule with inorganic ions (Na^+ , Cl^-) and molecules (NH_4^+) exist, too. While ESI is regarded as a mild ionization technique, ionization can induce fragmentation, and loss of H_2O or CO_2 are common. Ionization in ESI depends on the environment of the solvent, non-volatile solvents (or buffers), and other ions can reduce ionization efficiency (ion suppression).

1.3.2.2 Fragmentation of analyte can be used to increase identification specificity

The fragmentation compartment or collision cell is responsible for purposefully creating smaller fragments out of the molecular analyte ions, and the intensity of fragmentation can typically be altered by applying different fragmentation energies. Different strategies for fragmentation exist, however collision induced fragmentation (CID) is the most common in LC-MS (Sleno and Volmer, 2004). CID works by accelerating ions using different energies into a compartment filled with an inert gas, mostly argon (Morris et al., 1994). Upon impact of the ion into the gas atom (molecule), the kinetic impact energy leads to fragmentation. The combination of fragments can be very specific to respective analytes and present a vital identification possibility. The fragmentation follows typical reactions in the gas phase and can be partially predicted (Allen et al., 2015); albeit, the actual fragmentation is dependent on the surrounding environment and thus on the LC method and sometimes the specific sample.

1.3.2.3 Mass analyzers separate analyte ions based on their mass to charge ratio

The mass analyzer is responsible for the separation of analyte ions. However, different possibilities for the mass analyzer exist. In metabolomics, quadrupole, time of flight and ion trap mass spectrometers are the most common (Lei et al., 2011). Quadrupole analyzers consist of four parallel metal rods that generate an alternating electromagnetic field. Based on their mass to charge ratio, ions can either fly through the electromagnetic field or become unstable and get attracted to the metal rods. The electromagnetic field can be modified so only the analyte ion with a specific mass to charge ratio passes and reaches the detector. Simply put, one can imagine the quadrupole analyzer as a filter for a subset of ions. As at every time point, only one mass/charge ratio can pass, to analyze the whole spectrum, the analyzer must scan through all mass ranges. These devices typically have a low mass resolution (Zhou et al., 2012). Based on this relatively targeted nature, quadrupole-based mass spectrometers are mostly used for targeted metabolomics (Roberts et al., 2012). As only one targeted ion at each time can pass, it reduces surrounding noise, making them able to be highly sensitive and suitable for quantification.

The ion traps similarly use electromagnetic fields, but they are used to store packages of ions. By slowly alternating the electromagnetic field, ions with specific mass/charge-ratio can leave the ion trap and enter the detector. These devices are typically sensitive and have low noise; however, their resolution is strongly enhanced compared to quadrupole analyzers. These devices generally are slightly slower than quadrupole mass spectrometers (Zhou et al., 2012).

The third type are the time of flight (ToF) mass spectrometers. When a package of analyte ions enters the mass spectrometer, they are accelerated with the same energy into an evacuated flight tube using an electromagnetic field. As the kinetic energy ions gain is different depending on their mass/charge ratio, the resulting velocity of the ions is different. By accurately measuring the time the ion spends for the flight until it reaches the detector (time of flight), one can calculate the mass/charge ratio of the ion. This type of analyzer is faster than ion traps and shows better resolution than the quadrupole. However, the noise is higher, making them less able to quantify low abundant metabolites. Thus, ToF mass spectrometers are typically most useful for untargeted metabolomics (Freiburghaus et al., 2019).

Combinations of different mass analyzers exist as well. For quantification, the most common are arrays of three sequential quadrupoles (Triple-Quads or QqQ), where the middle quadrupole is a collision cell.

1.0 General introduction: The diagnostic odyssey of rare disorders

This enables very selective analyses by targeting a specific ion with the first quadrupole and then using the third quadrupole to target one specific fragment ion. Triple-Quads are exceptionally sensitive and have good specificity (Zhou et al., 2012). Also, quadrupole time of flight mass spectrometers (QToF) exist as well.

During the course of this work, we worked with a QToF and a Triple-Quad mass spectrometer in our institute, combining abilities for targeted and untargeted metabolomics.

1.3.2.4 Detectors in mass spectrometry often use electron amplifiers

After passing through the mass spectrometer, the analyte ions reach the detector. Most detectors use different types of secondary electron amplifiers; microchannel-plate detectors are common (Fraser, 2002; Geno, 1992). Here ions hit the side of a channel arranged in a plate, producing the exit of electrons. The plate is charged so that the electrons accelerate and hit the channel wall again, generating a higher number of electrons. This repeats multiple times, creating a strongly amplified electron avalanche that can then be detected.

1.3.2.5 Identification in mass spectrometry is based on mass to charge information

Mass spectrometers can collect information about the mass to charge ratio of the analyte molecular ion and the collected fragments. As the atomic mass of single elements is known, using combinatoric calculations elemental formulas can be calculated for each of the mass/charge ratios. Using different databases, the most likely molecule can be identified. The identification usually follows specific rules that reduce the number of possible combinations (Vijlder et al., 2018).

However, often there is a high number of analytes that share the same mass/charge ratio and cannot be uniquely identified (isobaric compounds). Potential fragments can be calculated from the structural formulas and help with the identification, although they often need to be predetermined. Even though measurement approaches that do not utilize LC exist (flow injection analysis, FIA; Fuhrer et al., 2011), separation by LC is beneficial for sensitivity and necessary for proper identification using retention time information (Xiao et al., 2012).

1.3.2.6 Quantification in mass spectrometry requires adept internal standards

Mass analyzers are a priori “quantitative” instruments that count analyte ions hitting the mass spectrometer detector. This “counts” are then translated into ion intensities, sometimes called ion abundances. However, ion production efficiency from analytes in the ionization source varies greatly and depends on the specific mass spectrometer. Also, the translation of ion counts to abundances varies significantly between manufacturers, and even slight changes in the configuration of the ion source (e.g., changing the ESI capillary) can change abundances. Thus, ion abundances are not a quantitative marker and cannot be compared between metabolites, different mass spectrometers, or even between runs, making ion intensities/abundances a semiquantitative feature. In order to translate ion abundances in quantitative values (e.g., concentration), a standard/calibration curve is necessary. There is another level of complexity, as the ionization efficiency varies in the surrounding chromatographic environment (ion suppression). For correcting these effects, the use of internal standards is necessary. Consequently, (absolutely) quantitative results can only be yielded from targeted metabolomics approaches, as only here the appropriate calibration curves and internal standards are used.

1.0 General introduction: The diagnostic odyssey of rare disorders

1.4 Summary of the introduction and goal of this thesis project

The term “rare diseases” is an umbrella term for a very diverse group of disorders, where the single disorders have a particularly low prevalence, but cumulatively patients in the hundred million range are affected. Due to missing widespread disease information and high diversity, most rare disorders go by undetected for a long time. This limits patient care as well as research. Among the rare disorders, a large group are the inherited disorders of metabolism, leading to neuropediatric neurometabolic, and neurodegenerative disorders. While some rare disorders are part of the newborn screens mandated by many countries, the great majority are still unrecognized. As these disorders affect metabolism, the holistic approach to study metabolism, metabolomics might further detect and study these diseases. Many strategies for metabolomics measurements exist. However, most metabolomics studies utilize liquid chromatography mass spectrometry (LC-MS).

Therefore, we propose the use of LC-MS-based metabolomics to characterize rare neurometabolic and neurodegenerative disease further. This thesis aims to develop an LC-MS-based metabolomics platform to characterize these disorders in an optimized fashion. In a second step, we aim to prove the utility of our platform on the example of four selected projects.

2.0 General materials and methods

2.1 Materials

2.1.1 List of chemicals

Acetonitrile for LC-MS $\geq 99.95\%$ (Art.-Nr.: 2697.2500, CHEMSOLUTE/Geyer, Höxter, Germany)

Ammonia solution 25% for LC-MS, LiChropur (Art.-Nr.: 5.33003.0050; Sigma-Aldrich/Merck, Darmstadt, Germany)

Ammonium acetate (Art.-Nr.: 431311-50G; Honeywell, Charlotte, USA)

Ammonium formate $\geq 99.995\%$ trace metal basis (Art.-Nr.: 516961-100G; Sigma-Aldrich, Darmstadt, Germany)

Butylated hydroxytoluene (BHT) $\geq 99\%$, FCC, FG (BHT; Art.-Nr.: W218405-SAMPLE-K; Sigma-Aldrich, Darmstadt, Germany)

Formic acid 99%, ULC/MS-CC/SFC (Art.-Nr.:00069141A8BS, Biosolve, Valkenswaard, The Netherlands)

Human metabolite reference standard library, self-modified (MetaSci, Toronto, Canada). A list of acquired metabolites can be found in **Table 22** (p. 181).

Methanol absolute ULC/MS-CC/SFC (Art.-Nr.:0013684101BS, Biosolve, Valkenswaard, The Netherlands)

Palmitic acid-d31 98 atom%D, 99% CP (Art.-Nr.: 366897-100MG, Sigma-Aldrich; Sigma-Aldrich/Merck, Darmstadt, Germany)

Potassium hydroxide, Emsure (Art.-Nr.:105033, Merck Millipore, Darmstadt, Germany)

Tetrahydrofuran, 99%, stabilized with 250-350 ppm BHT (Art.-Nr.: L13304, AlfaAesar, Haverhill, USA)

2-propanol for LC-MS $\geq 99.95\%$ (Art.-Nr.: 1178.1000, CHEMSOLUTE/Geyer, Höxter, Germany)

2.1.2 List of bought buffers, other solutions and disposable materials

AbsoluteIDQ p180-Kit (Art.-Nr.: p180WU-6334, Biocrates, Innsbruck, Austria)

Acquity UPLC 700 μ l Round 96-Well Sample-Plate (Art.-Nr.:186005837, Waters, Milford, USA)

BCA-Solution: BCA protein assay kit (Art.-Nr.: UP95424, Interchim, Montluçon, France), 1:50 solution B to A, prepared on use

Bond Elut 96 Square-well, C18, 100 mg (Art.-Nr.:A396011C, Agilent, Santa Clara, USA)

BSA-Standard: Uptima BSA-standard for protein assay 2 mg/ml in water (Art.-Nr.:UP36859A, Interchim, Montluçon, France), stored at 4°C

Buffer solution pH 7.0 (Art.-Nr.: A518.1, Roth, Karlsruhe, Germany)

Buffer solution pH 4.0 (Art.-Nr.: A517.1, Roth, Karlsruhe, Germany)

2.0 General materials and methods

Buffer solution pH 9.0 (Art.-Nr.: A519.1, Roth, Karlsruhe, Germany)

Chromabond HILIC, 45 µm, 3 ml, 500 mg (Art.-Nr.: 730593, Macherey-Nagel, Düren, Germany)

Combitips advanced 50 mL (Art.-Nr.: 0030089.480, Eppendorf, Hamburg, Germany)

Combitips advanced 10 mL (Art.-Nr.: 0030089.820, Eppendorf, Hamburg, Germany)

Conical tubes, PP, 15 mL, sterile (Art.-Nr.: 62.554.502, Sarstedt, Nümbrecht, Germany)

Conical tubes, PP, 50 mL, sterile (Art.-Nr.: 62.547.254, Sarstedt, Nümbrecht, Germany)

Cortecs UPLC T3 column (CortecsT3) 2.1 x 150 mm 1.6µm column (Art.-Nr.: 186008500, Waters, Milford, USA)

LC-MS glass screw neck vials with cap, PTFE/silicone septa pre-slit (Art.-Nr.:600000670cv, Waters, Milford, USA)

Microplate 96-well, PS, F-Bottom, Clear (Art.-Nr.: 655101, Greiner bio-one, Frickenhausen, Germany)

Pipette tips 10 µL (Art.-Nr.: 70.113, Sarstedt, Nümbrecht, Germany)

Pipette tips 200 µL (Art.-Nr.: 70.760.002, Sarstedt, Nümbrecht, Germany)

Pipette tips 1000 µL (Art.-Nr.: 70.3050.020, Sarstedt, Nümbrecht, Germany)

Polypropylene mat caps for round plates (Art.-Nr.:186002483, Waters, Milford, USA)

Reaction tubes, SafeSeal, 1.5 mL (Art.-Nr.: 72.706.200, Sarstedt, Nümbrecht, Germany)

Reaction tubes, SafeSeal, 2.0 mL (Art.-Nr.: 72.695.400, Sarstedt, Nümbrecht, Germany)

Reaction tubes, SafeSeal, 5 mL (Art.-Nr.: 72.701, Sarstedt, Nümbrecht, Germany)

Rotiprotect-Nitril green (Art.-Nr.: HXK2.1, Roth, Karlsruhe, Germany)

Serological pipettes, 5 mL (Art.-Nr.: 86.1253.001, Sarstedt, Nümbrecht, Germany)

Serological pipettes, 10 mL (Art.-Nr.: 86.1254.001, Sarstedt, Nümbrecht, Germany)

Serological pipettes, 25 mL (Art.-Nr.: 86.1685.001, Sarstedt, Nümbrecht, Germany)

Sequant ZIC-HILIC 3.5 µm, 250 x 2.1 mm 100 Å PEEK-coated HPLC-column (Art.-Nr.: 1.50443.0001); Sequant/Merck, Darmstadt, Germany)

Storage solution for pH and ORP electrodes (Art.-Nr.: HI70300, Hanna instruments, Woonsocket, USA)

Water for LC-MS: LC-MS water was used freshly from an Arium basic water purification system (Sartorius Stedim, Göttingen, Germany) with a maximum conductivity of 0.075 µS/cm.

2 ml Square Collection Plate (Art.-Nr.:186002482, Waters, Milford, USA)

2.1.3 List of devices and other equipment

Acquity I-Class liquid chromatography system with binary solvent manager and FTN Sample manager (Waters, Milford, USA)

2.0 General materials and methods

Bio-Freezer -80°C (Forma-Scientific/Thermo Fisher, Waltham, USA)

Centrifuge 5424 (Eppendorf, Hamburg, Germany)

EF103 easy-fit ice machine (Scotsman, Milano, Italy)

FCB table top balance (Kern, Balingen, Germany)

Gen5 1.11 (Biotek, Winooski, USA)

Measuring pipette, glas, Blaubrand, 10 mL (Brand, Wertheim, Germany)

Measuring pipette, glas, Blaubrand, 25 mL (Brand, Wertheim, Germany)

Measuring pipette, glas, Blaubrand, 50 mL (Brand, Wertheim, Germany)

MH 15 Magnetic stirrer with heater (Roth, Karlsruhe, Germany)

Mikro 220R centrifuge (Hettich, Tuttlingen, Germany)

OptiPlex 5040 with Intel Core i3-6100 3 GHz and 4 GB RAM and Windows 10 Enterprise (Dell, Round Rock, USA)

OptiPlex 5060 with Intel Core i5-8500 3 GHz and 16 GB RAM and Windows 10 Enterprise Version 1809 (Dell, Round Rock, USA)

Oven 400 HY (Bachofer, Reutlingen, Germany)

pH-Meter CG820 (Schott Geräte)

Pipette, manual, Eppendorf Research, 10 µL (Eppendorf, Hamburg, Germany)

Pipette, manual, Eppendorf Research, 200 µL (Eppendorf, Hamburg, Germany)

Pipette, manual, Eppendorf Research, 1000 µL (Eppendorf, Hamburg, Germany)

Predator PH315-53 with Intel Core i7-10750H 2.6 GHz and 32 GB RAM and Windows 10 Home Version (Acer, New-Taipeh, Taiwan)

Savant SpeedVac SPD140DDA Vacuum concentrator equipped with a Savant RVT5105 Refrigerated Vapor Trap and a OFP400 Vacuum Pump (ThermoFisher Scientific, Waltham, USA)

Synergy Mx Microplate reader (Biotek, Winooski, USA)

Thermal Shake lite (VWR, Vienna, Austria)

ThinkStation C30 with Intel Xeon CPU E5-2620 2.1 GHz and 16 GB RAM and Windows 10 Enterprise Version (Lenovo, Quarry Bay, Hongkong)

Universal 320 centrifuge (Hettich, Tuttlingen, Germany)

VF2 Vortex mixer (Janke&Kunkel IKA Labortechnik, Staufen im Breisgau, Germany)

Water bath (Mettler, Schwabach, Germany)

Xevo G2-S QToF mass spectrometer (Waters, Milford, USA)

Xevo TQ-S mass spectrometer (Waters, Milford, USA)

2.0 General materials and methods

2.1.4 List of scientific programs and other software

BioRender.com (BioRender, Toronto, Canada)

ClassyFire 1.0 (Djoumbou Feunang et al., 2016)

MassLynx V4.1 software (Waters, Milford, USA)

Matplotlib 3.0 (Hunter, 2007)

MetIDQ Carbon (Biocrates, Innsbruck, Austria)

MetaboAnalyst 5.0 (Pang et al., 2021; Xia et al., 2009)

Microsoft Office Professional Plus 2016 (Microsoft, Redmond, USA)

MS-Dial 4.60 (Tsugawa et al., 2015, 2020)

Prism 8.3 (GraphPad Software, San Diego, USA)

Progenesis QI 2.4 (Nonlinear Dynamics, Newcastle upon Tyne, UK)

Python 3.6 (The python software foundation, 2021)

UNIFI version (Waters, Milford, USA)

Zotero 5.0.96 (Roy Rosenzweig Center for History and New Media of the George Mason University, Fairfax, USA)

2.1.5 Buffers and other solutions

BHT solution: For a 1% (w/v) butylated hydroxytoluene (BHT) solution, 1 g BHT were dissolved in 100 ml methanol. The solution was stored at room temperature in the dark for up to 4 weeks.

CortecsT3 sample solvent: For 100 ml, 50 ml water, 25 ml methanol and 25 ml acetonitrile were combined. The solvent can be stored for at least 3 months at room temperature.

Leucine-Enkephaline lockspray solution: First, prepare 200 ml dilution solvent by combining 100 ml LC-MS grade water, 100 ml acetonitrile and 200 μ l formic acid. For the first dilution of 400 ng/ μ l dissolve 3 mg of leucine-enkephalin in 7.5 ml water. Then for the second dilution step of 1 ng/ μ l, dissolve 50 μ l of the 400 ng/ μ l dilutions in 20 ml dilution solution. This dilution can be stored in PTFE-coated containers at -14°C for 6 months. Before immediate use, dilute 20 ml of the 1 ng/ μ l dilution step in 80 ml dilution solution and sonicate for 15 minutes. The solutions can be stored at room temperature for at least one month.

Modified Bligh and Dyer lipid extraction solvent: Modified from (Bligh and Dyer, 1959). For 100 ml, 32.2 ml chloroform, 61.4 ml methanol, 3.2 ml BHT solution, and 3.2 ml 3M HCl were combined and stored at room temperature for up to 4 weeks and prechilled to -15°C before immediate use.

Phosphate buffer for cell lysis: For 1L of 10 mM buffer, 1.4 g sodium dihydrogen phosphate was dissolved in LC-MS-grade water, and pH was adjusted with HCl to pH 7.0. Buffer was stored at room temperature for up to 2 weeks and prechilled to 4°C before immediate use.

Protein precipitation buffer for RPLC-SPE: For 50 ml combine 47.5 ml of methanol with 3.5 ml 1% BHT-solution. The solution was stored at room temperature in the dark for up to 4 weeks.

2.0 General materials and methods

RPLC-MS CortecsT3 buffer A: For 1 L, weight 630 mg ammonium formate in a 2 ml reaction tube and dissolve in 1ml LC-MS 100% formic acid by vortexing vigorously.

Then combine 400 ml LC-MS water, 600 ml acetonitrile, and formic acid/ ammonium formate solution and sonicate for 10 minutes. Prepare one day in advance before use. Buffer was stored at room temperature for up to 2 weeks.

RPLC-MS CortecsT3 buffer B: For 1 L, weight 630 mg ammonium formate in a 2 ml reaction tube and dissolve in 1ml LC-MS 100% formic acid by vortexing vigorously. Then combine 900 ml LC-MS 2-propanol, 100 ml acetonitrile, and formic acid/ ammonium formate solution and sonicate for 10 minutes. Prepare one day in advance before use. Buffer was stored at room temperature for up to 2 weeks.

RPLC-MS CortecsT3 purge buffer: For 1 L, combine 400 ml LC-MS water and 600 ml acetonitrile. Buffer was stored at room temperature for up to 3 months.

RPLC-MS CortecsT3 wash buffer: For 1 L, combine 100 ml LC-MS water and 900 ml acetonitrile. Buffer was stored at room temperature for up to 3 months.

RPLC-SPE lipid elution buffer: For 100 ml, combine 25 ml tetrahydrofuran (with BHT) and 75 ml methanol. The solution was stored at room temperature in the dark for up to 4 weeks.

Sodium formate QTOF calibration solution: First, prepare 1 ml 10% formic acid solution by adding 100 μ l formic acid to 900 μ l water. Then dissolve 12.5 μ l 1 M sodium hydroxide solution in 22.5 ml 2-propanol and 2.5 ml water. Add 10 % formic acid and sonicate for 15 minutes and use immediately. The solutions can be stored at room temperature for at least one month.

ZIC-HILIC SPE buffer: For 1L, 700 ml tetrahydrofuran, 250 ml acetonitrile, 50 ml methanol were combined. Buffer was stored at room temperature for up to 3 months.

ZIC-HILIC-MS buffer stock: For 500 ml of a 30 mM ammonium acetate stock solution: 1.16 g LC-MS grade ammonium acetate was dissolved in 500 ml LC-MS-grade water was adjusted to pH 7.0 using 25% ammonia solution. Use immediately.

ZIC-HILIC-MS buffer A: For 500 ml, 450 ml ZIC-HILIC-MS running buffer stock were combined with 50 ml acetonitrile. Buffer was stored at room temperature for up to 2 weeks.

ZIC-HILIC-MS buffer B: For 500 ml, 50 ml ZIC-HILIC-MS running buffer stock were combined with 450 ml acetonitrile. Buffer was stored at room temperature for up to 2 weeks.

ZIC-HILIC-MS purge buffer: For 1 L, combine 200 ml LC-MS water and 800 ml acetonitrile. Buffer was stored at room temperature for up to 3 months.

ZIC-HILIC-MS wash buffer: For 1 L, combine 900 ml LC-MS grade water and 100 ml acetonitrile. Buffer was stored at room temperature for up to 3 months.

2.2 Methods

2.2.1 Preparation of EDTA-plasma for metabolomics

After participant approval according to the local ethics regulation, venous blood was drawn into Sarstedt EDTA-plasma vials by the responsible medical professional and shook lightly to distribute EDTA in the blood. Blood was filled to appropriate markings to ensure equal concentration of EDTA.

2.0 General materials and methods

Where possible, EDTA blood was delivered to the testing laboratory in less than 1 hour. Then EDTA-blood was centrifuged at 3000 rpm for 5 minutes and plasma (supernatant) was aliquoted in 550 μ l aliquots without disturbing the pellet.

Samples with visually recognizable hemolysis were excluded from the study. Aliquots were then stored at -80°C .

2.2.2 Preparation of cell cultures for metabolomics analysis

Cells were seeded preferably in 150 mm dishes or 6-well plates, with consistent seeding densities for the specific experiment. Seeding densities were dependent on the respective cell type. Typically, 1×10^6 - 3×10^6 cells were used for large cells (such as fibroblasts) and 10×10^6 cells for smaller cell types (HEK or HELA cells). Cells were then cultured in the appropriate cell culture medium for at least 3 days before cell lysis.

2.2.3 Cell lysis of cell culture samples for metabolomics analysis

Phosphate buffer for cell lysis was prechilled to 4°C and stored on ice. Then cells in 150 mm plates were removed from the incubator and washed quickly without force one time with PBS (room temperature) and all remaining liquid was removed. Then ice cold cell lysis solution was applied to the plate, and cells were mechanically rapidly brought into the solution with a cell scraper. Around 1000 μ l of the cell homogenate was aliquoted into Safe seal 2 ml reaction tubes and immediately snap-frozen in liquid nitrogen. As the short time between removing the cells from the incubator and snap-freezing is crucial, only up to 3 plates were handled simultaneously. Then cells were lysed using a repeated freeze-thaw protocol. Here cells were first defrosted for around 1-2 minutes at 80°C using a thermo-block with intermittent vortexing and, upon near-complete defrosting, snap-frozen immediately. The time in the thermo-block should be kept at less than 2 minutes to avoid high internal sample temperatures. After snap-freezing, this procedure was repeated an additional 3 times. Then the cell homogenate was sonicated for 15 mins in a cold ice-water bath. The lysate can be used immediately, stored on ice for up to 2 hours, or held at -80°C for at least 3 months.

2.2.4 Standardization of cell culture lysates

For standardization of different samples from the same cell type, total protein concentration in the cell lysate was used to determine an approximator for the number of cells in the sample and cell lysis efficiency to ensure reproducible results. After cell lysis, protein concentration was determined using bichinoic acid assay (BCA) reaction with the Interchim kit according to the guidelines of the manufacturer. Briefly, cell lysates were diluted in a transparent 96-round-well microplate with water. For large cell types (e.g., fibroblasts) with lower protein concentrations, 15 μ l lysate was diluted with 10 μ l water. For small cell types with higher protein concentrations, 5 μ l lysate and 20 μ l water were used. Then using the 2 mg/ml BSA standard, a calibration curve was prepared, comprising of 0, 5, 10, 15, 20, and 30 μ g of BSA protein. All measurements occurred in technical duplicates. The assay solution was prepared by combining reagent A with B in a ratio of 1:50, and 200 μ l were applied to each well. The microwell plate was incubated for 30 minutes at 37°C , and absorbance at 540 nm was determined using a microplate reader. A water blank was subtracted from each well. Then samples were standardized to the sample with the lowest protein concentration in the same experimental set, setting its volume to 1000 μ l. The volumes of other lysates were then adjusted negative proportionally to their protein concentration and filling their volume to 1000 μ l. The adjusted lysate can be used directly or stored at -80°C for at least 3 months.

2.2.5 Sample preparation for untargeted metabolomics

2.2.5.1 Lipid extraction using the modified Bligh and Dyer method

500 µl EDTA-plasma or 1000 µl standardized cell culture lysate was added to 1000 µl cold (-20°C) modified Bligh and Dyer (Bligh and Dyer, 1959) extraction buffer with 1 µl palmitic acid d31 (5 mg/ml in methanol) as internal standard. After intensive vortexing, samples were centrifuged at 14,000 rpm for 5 minutes in a cold centrifuge, and the resulting two-phase system aliquoted separately. 1000 µl of the upper, hydrophilic, phase was then used for extraction with the hydrophilic ZIC-SPE method. Either 150 µl for the plasma samples or 300 µl for cell culture samples of the lower lipid-phase were aliquoted and dried using a vacuum centrifuge at 35°C and below 100 mPa for 15 minutes. The dried extract was redissolved in 100 µl CortecsT3 sample solvent and after vortexing centrifuged for 2 minutes and max. speed. The extract was then transferred to the LC-MS measurement vial.

2.2.5.2 Hydrophilic metabolite extraction with the ZIC-SPE

1000 µl of the hydrophilic extract from the Bligh and Dyer extraction (chapter 2.2.5.1, p. 24) were mixed with 1500 µl ZIC-SPE solvent. The ZIC-SPE columns were initiated with 2 ml water and primed with 3 ml ZIC-SPE solvent. Then the samples were applied and eluted in a cold (4°C) environment. The eluate was removed and metabolites eluted with 2.5 ml water. The eluate was collected in two reaction tubes, snap-frozen, and dried in the frozen state in a vacuum centrifuge at 60°C, below 10 mPa for 2.5 hours. The dried extracts were dissolved in 100 µl ZIC-HILIC mobile phase B and combined. Particulates were removed using centrifugation for 2 minutes at maximum speed and aliquoted to a LC-MS sample 96-well plate.

2.2.5.3 Metabolite extraction using the RPLC-SPE method

For the sample extraction with the RPLC-SPE method, 500 µl of plasma sample or 1000 µl of standardized cell culture lysate was applied to a round-well 2 ml 96-well plate. Then 1000 µl BHT solution was applied to the samples and shaken for 15 mins in a cold (4°C) environment. Then samples were centrifuged for 5 minutes at max. speed and the supernatant used further. The 96-well RPLC SPE plate was once washed with 1 ml methanol, then 2 ml water, and then 1000 µl sample supernatant were applied to each well. The hydrophilic metabolites were collected in this eluate in a different 2 ml collection plate. The hydrophilic extract was dried in a vacuum centrifuge at 60°C and below 10 mPa for 3 hours. The dried hydrophilic extract was reconstituted in 150 µl ZIC-HILIC mobile phase B by agitating at 1500 rpm for 30 minutes. The lipophilic metabolites were eluted by applying 1000 µl RPLC-SPE extraction buffer. The lipophilic eluate was then dried at 45°C and below 10 mPa for 1.5 hours in a vacuum centrifuge and the extract reconstituted in 250 µl CortecsT3 lipid sample buffer by agitating for 15 minutes at 1500 rpm.

2.2.6 Processing of the Biocrates Absolute IDQ p180-kit

For the metabolome analysis using the AbsoluteIDQ p180-kit, EDTA-plasma or serum samples were processed according to the manufacturer's guideline and as described previously (Siskos et al., 2017). The kit was used on a BEH amide column with a Xevo TQ-S mass spectrometer.

The kit is conceived in a 96-well-plate format, with an extraction plate stacked onto a second collection plate. Each well of the extraction plate contains a small insert carrying a small round filter piece. At least 11 of these wells are reserved for the application of a 7-point calibration curve (for the quantification of amines), one well for a blank (mainly necessary to determine the limit of quantification for lipids),

2.0 General materials and methods

and 3 human plasma-based quality controls in different concentration levels. However, to check the quality of the run, 3 further wells are used by distributing the medium quality control level over the plate. The remaining 85 wells can be used for samples.

To determine the suitability of the LC-MS system, checkups are required before every kit run. For LC-MS, a test mix of biogenic amines and respective internal standards is provided that is run manually using the provided LC-MS method. Here, the retention times of each analyte in the mass spectrometry method are modified so that they match with times in the test mix. This is to ensure proper detection of all compounds, as retention times are subjected to variation. Before every FIA-Measurement the mass spectrometer and sample manager fluidics have to be cleaned extensively, as remaining lipids strongly interfere with FIA-MS lipid analysis. Following cleaning, the blank intensity is determined. If the blank total ion chromatogram is higher than 1×10^6 , additional cleaning is performed. Then an FIA-MS test mix, consisting of several proprietary compounds, is measured and evaluated by the kit.

Before the kit can be used, samples must be registered and assigned to specific well-positions in MetIDQ, a software delivered with the kit. This software then generates plate schemes for the wet-lab processing and sample lists for the mass spectrometry measurement using unique sample identifiers. As the evaluation of the FIA-MS data and validations of quality controls can only be done using the software, prior registration and names are necessary for correct kit measurements. After the kit is registered, it can be processed in the wet lab.

Briefly, a proprietary internal standard solution was first applied to all but one (blank) well for the wet lab processing. Then 10 μl of either sample, calibrator, or quality control were applied to the wells as assigned in MetIDQ software. The plate was then dried using a positive pressure manifold, and amines were derivatized using phenylisothiocyanate. After another drying step, metabolites were extracted from the filter plate using 5 mM ammonium acetate solution in methanol. Then the extract was diluted in separate plates. For FIA-MS, the extract was diluted with FIA-MS solvent. Water was used to dilute LC-MS samples. While the LC-MS-Plate should be measured directly after extraction, the FIA-MS plate may be stored for up to two weeks at 4°C.

The kit was then measured using the liquid chromatography and mass spectrometry methods provided with the kit. Solvents are defined by the kit manufacturer; the FIA-MS solvent contains a special proprietary solution that is delivered with the kit. After the kit run, LC-MS data were evaluated using the provided analysis method with MassLynx. To ensure proper determination of the metabolites, peak integration of every peak was controlled manually and adjusted if necessary. Then LC-MS Data were imported in MetIDQ. FIA-MS-Data were imported directly into MetIDQ and evaluated automatically. After the import of both analyses, MetIDQ performs automatized quality controls by comparing the results of the three different concentration levels and the sample distributed over the plate with internally saved values. If accuracy and precision are within the defined limits ($\pm 15\%$), an analyte is validated, and data of concentrations (e.g., in μM) can be exported for statistical evaluation.

2.2.7 Cleaning the G2-S QTOF mass spectrometry before use

To avoid signal deterioration based on contamination of the mass spectrometer by usage, the mass spectrometer sample inlet (sampling cone) and lock spray-deflection baffle (baffle) was cleaned before each run.

2.0 General materials and methods

Cleaning was performed by immersing the mass spectrometer parts in 50% methanol and 50% of a 10% formic acid in water solution and sonicating for 15 minutes in a sonic water bath. Then the solution was exchanged to 50% water and 50% methanol, and sonication was repeated before repeating the sonication a third time with 100% methanol. The parts were then air-dried and used.

2.2.8 Preparing the I-Class-LC and G2-S QToF mass spectrometer for a measurement.

Independent of the specific type of analysis, preparing the I-Class liquid chromatography system and G2-S QToF mass spectrometer follows the same general protocol. To ensure proper equilibration, the LC mobile phase was prepared one day before use and not older than 2 weeks for the highly aqueous mobile phases and not more than 4 weeks for higher organic mobile phases. Mobile phases were attached to the LC system and primed for 5 minutes. Then the column was equilibrated with 50% of each phase for 15 minutes and equilibrated at LC starting conditions. In parallel, the mass spectrometer is set into operation mode and equilibrated for 30 minutes, while the infusion of mobile phase avoids corrosion by corona discharge. After equilibration, the intensity and stability of the internal mass spectrometry compound ("Lockmass"/"Lockspray") were determined using the internal IntelliStart program. Then the calibration of the mass spectrometer was inspected by using a calibration mix containing sodium formate using IntelliStart. Sodium formate forms clusters of different mass to charge ratios covering the entire mass range with defined and known masses. The deviation of each measured mass to the expected is determined, and the square root of mean deviations is calculated. We defined a variation of less than 10 ppm as acceptable.

After equilibration of the LC system, one LC gradient without an injection of a sample is run.

2.2.9 General conditions of the I-Class-LC and G2-S QToF mass spectrometer run

Generally, samples are stored in LC-MS-grade glass vials with PTFE-membrane caps or in LC-MS grade 700µl 96-well-plates. Before the run, samples were stored at -80°C and during the run at 8°C. Each run consists of a blank sample made from the sample solvent and a quality control sample made by pooling an equal amount of each sample to reach an amount of at least 100 µl (pool). Blank and pool samples are repeated multiple times over the course of the run. Typically experiments with replicates were run by groups with one replicate of each sample group followed by a blank and a pooled sample, continuing with a second group of replicates. At least three blank/pool samples are included in each run. In order to avoid long-term intensity deterioration and retention time shifts, the maximum runtime was limited to 4 days for each sample set.

2.2.10 Final conditions of the ZIC-HILIC-MS method

First, the ZIC-HILIC-MS mobile phase was prepared as above: mobile phase A was 90 % 30mM ammonium acetate pH 7.0 in water, 10% acetonitrile; mobile phase B was 10% 30mM ammonium acetate pH 7.0 in water, 90% acetonitrile. 80% acetonitrile, 20% water was used as autosampler purge solvent; 10% acetonitrile, 90% water was used as autosampler wash solvent. We used a Sequant ZIC-HILIC in the dimensions 2.1x250mm at room temperature, and 5 µl sample were injected. The 30 min gradient used a flow of 0.2 ml/min, initially starting with 90% mobile phase B held for 2 min. This was followed by a linear gradient to 0% B in the next 18 minutes and held for 4 mins. In the next minute, the solvent returned to 90% B using a linear gradient and the column was equilibrated for 5 minutes before the next run. We used a Xevo G2-S QToF mass spectrometer with an ESI ionization source.

2.0 General materials and methods

We used the mass spectrometer in positive and negative ionization mode in sensitivity settings using the conditions displayed in **Table 1**. Data were acquired in continuous MSe mode.

Table 1 Mass spectrometer conditions for ZIC-HILIC-MS method.

Setting	Positive mode	Negative mode
ESI capillary voltage	2.5 kV	-1.5 kV
Cone voltage		50 V
Source offset		60 V
Source temperature		120 °C
Desolvation temperature		500 °C
Mass range		50-800 Da
Fragmentation		Ramp between 10V-40V

2.2.11 Final conditions of the CortecsT3-RPLC-MS lipidomics method

The CortecsT3 lipidomics method was adapted from a method widely used in the field (Narváez-Rivas and Zhang, 2016; Rampler et al., 2018). Here we shortened the run time to 30 mins by using a chromatographic column with a higher peak efficiency. Briefly, as described above, this method used 40% 10 mM ammonium formate in water with 0.1% formic acid and 60% acetonitrile as mobile phase A; 90% 2-propanol and 10 % acetonitrile with 0.1% formic acid and 10 mM ammonium formate were used as mobile phase B. 40% water and 60% acetonitrile were used as autosampler purge, and 90% acetonitrile and 10% water were used as autosampler wash solvent. We used a Cortecs UPLC T3 column in the dimensions 2.1x150mm at 45°C, and 5 µl samples were injected. The 35 min gradient used a flow of 0.3 ml/min and started with 35% mobile phase B increasing linearly to 40% B. In the next 5 mins, a linear gradient was used to reach 43% B before changing to 50% B in 0.1 mins. Afterward, a linear gradient in 12.40 mins to 54% B was used, followed by a jump to 70% B in the next minute. In the following 6 minutes, 95% B was applied and held for 3 mins. In the end, the column was equilibrated back to 35 % B in 2 minutes and held additional 3 minutes. We used a Xevo G2-S mass spectrometer with an ESI ionization source in continuous MSe mode. We used the mass spectrometer in positive ionization mode in resolution settings using the conditions displayed in **Table 2**.

Table 2 Mass spectrometer conditions for the RPLC-MS CortecsT3 lipidomics method.

Setting	Positive mode
ESI capillary voltage	0.8 kV
Cone voltage	25 V
Source offset	80 V
Source temperature	150 °C
Desolvation temperature	20 °C
Mass range	50-800 Da
Fragmentation	Ramp between 10V-40V

2.2.12 General data analysis and data storage

After measurement, data and sample lists were regularly secured on a high-fidelity server. For data analysis using MS-Dial, data were copied to the computer used for analysis.

Generally, mass spectrometry data were first imported into specialized mass spectrometry software (UNIFI, MetIDQ, or MS Dial). Here, metabolite features were extracted, abundances determined, and metabolites identified. Subsequently, data were exported to an excel-document and stored together with raw data. Extracted metabolite data were then further analyzed using MetaboAnalyst 4.0, and specific features examined using GraphPad prism. If not otherwise mentioned, data are shown as mean \pm SEM. The significance level of p-values after statistical testing are displayed as *P < .05, **P < .01, ***P < .001.

2.2.12.1 Analysis of mass spectrometry data using the open-source metabolomics software MS-Dial

For untargeted metabolomics analysis, we decided to use MS-Dial released by the RIKEN Center for Sustainable Resource Science: Metabolome Informatics Research Team under the lead development of Hiroshi Tsugawa (Tsugawa et al., 2015, 2020). Mass spectrometry data from untargeted metabolomics experiments were imported into MS-Dial and processed using optimized parameters based on the program's recommendation. Processing parameters were adapted to the specific analysis method (ZIC-HILIC-MS negative/positive, CortecsT3 lipidomics method). Common to all analysis methods was the alignment to a quality control sample (pool) and the signal normalization using either the identified metabolite total ion chromatogram (mTIC) or quality control-based local regression (LOESS).

For ZIC-HILIC-MS analyses, parameters in **Table 3** for positive ionization mode and **Table 4** for negative mode were used. Compared to the recommended settings, the mass range was adjusted to cover most small hydrophilic compounds while reducing data size (50-800 Da). Mass accuracies were adjusted based on typical mass deviations while selecting slightly broader ranges to encompass more metabolites characteristic for a screening method. Alignment was modified to higher retention time tolerances and broader peaks specific for a HILIC method with an HPLC column. Identification was carried out using our in-house human metabolite identification library combined with the mass/fragment information from the FiehnHILIC database (<https://mona.fiehnlab.ucdavis.edu/>, last accessed 04.06.2021). For further analysis, we included metabolites with matching retention time (difference < 0.5 min) as identified compounds but also features that were only determined by the mass and/or fragments from the FiehnHilic library. After processing with MS-Dial, we used the normalized aligned peaks for further statistical analysis. Features of special relevance to the experiment that our metabolite library did not identify were subjected to further identification studies.

2.0 General materials and methods

Table 3 MS-Dial processing parameters for ZIC-HILIC-MS method in positive ionization mode.

Data collection		Peak detection		Alignment	
MS1 tolerance:	0.01 Da	Minimum peak height:	300 A.U	Retention time tolerance:	0.4 min
MS2 tolerance:	0.05 Da	Mass slice width:	0.02 Da	MS1 tolerance:	0.03 Da
Retention time begin:	1.5 min	MS2Dec		Retention time factor:	0.4
Retention time end:	26 min	Sigma window value:	0.2	MS1 factor:	0.6
MS1 mass range begin:	50 Da	MS/MS abundance cut off:	500 A.U	N% detected in group:	50%
MS1 mass range end:	800 Da	Identification			
MS/MS mass range begin:	50 Da	Retention time tolerance:	1 min		
MS/MS mass range end:	800 Da	Accurate mass tolerance (MS1):	0.05 Da		
Maximum charged number:	2	Accurate mass tolerance (MS2):	0.1 Da		
Adduct		Identification score cutoff	10 %		
[M+H] ⁺ , [M+NH ₄] ⁺ , [M+Na] ⁺		Only report top hit	yes		

Table 4 MS-Dial processing parameters for ZIC-HILIC-MS method in negative ionization mode.

Data collection		Peak detection		Alignment	
MS1 tolerance:	0.01 Da	Minimum peak height:	300 A.U	Retention time tolerance:	0.4 min
MS2 tolerance:	0.05 Da	Mass slice width:	0.02 Da	MS1 tolerance:	0.03 Da
Retention time begin:	1.5 min	MS2Dec		Retention time factor:	0.4
Retention time end:	26 min	Sigma window value:	0.2	MS1 factor:	0.6
MS1 mass range begin:	50 Da	MS/MS abundance cut off:	500 A.U	N% detected in group:	50%
MS1 mass range end:	800 Da	Identification			
MS/MS mass range begin:	50 Da	Retention time tolerance:	1 min		
MS/MS mass range end:	800 Da	Accurate mass tolerance (MS1):	0.05 Da		
Maximum charged number:	2	Accurate mass tolerance (MS2):	0.1 Da		
Adduct		Identification score cutoff	10 %		
[M-H] ⁻ , [M-H ₂ O-H] ⁻ , [M-Cl] ⁻ , [M+Hac-H] ⁻		Only report top hit	yes		

For lipidomics analyses using the CortecsT3 method, we used the parameters in **Table 5**. Here mass ranges were adapted to typical ranges of lipids (100Da-1200Da). Also, retention time tolerances were adjusted to the higher retention time stability of RPLC methods. For the identification of lipid species, we used the identification pipeline integrated into MS-Dial. After analysis and selection of lipids of special interest, identification of lipid species was confirmed based on the typical retention times ranges of lipid classes and were possible, by acquisition of new reference standards.

2.0 General materials and methods

Table 5 MS-Dial processing parameter for the CortecsT3 method.

Data collection		Peak detection		Alignment	
MS1 tolerance:	0.01 Da	Minimum peak height:	300 A.U	Retention time tolerance:	0.3 min
MS2 tolerance:	0.1 Da	Mass slice width:	0.02 Da	MS1 tolerance:	0.03 Da
Retention time begin:	0.8 min	MS2Dec		Retention time factor:	0.5
Retention time end:	32 min	Sigma window value:	0.5	MS1 factor:	0.5
MS1 mass range begin:	100 Da	MS/MS abundance cut off:	800 A.U	N% detected in group:	33%
MS1 mass range end:	1200 Da	Identification			
MS/MS mass range begin:	100 Da	Retention time tolerance:	n.a		
MS/MS mass range end:	1200 Da	Accurate mass tolerance (MS1):	0.05 Da		
Maximum charged number:	1	Accurate mass tolerance (MS2):	0.1 Da		
Adduct		Identification score cutoff	10 %		
[M+H] ⁺ , [M+NH ₄] ⁺ , [M+Na] ⁺		Only report top hit	yes		

2.2.12.2 Statistical analysis of metabolomics data using MetaboAnalyst

After transforming metabolomics data (targeted and untargeted data) from the respective mass spectrometry software to a tabular, excel format, we proceeded with statistical data analysis. Here we used the R-based web tool MetaboAnalyst developed by McGill University under lead by Jianguo Xia (Pang et al., 2021; Xia et al., 2009). This tool is adapted for the purpose of metabolomics analysis and encompasses necessary steps from data integrity checks to data normalization, and provides functions for univariate analyses (e.g., Student's T-tests, volcano-plots), chemometric analyses (PCA, PLS-DA), and cluster analyses (cluster-heatmaps).

For general data analysis of our metabolomics data, we selected features from the previous mass spectrometry programs that led to identification at least by mass. Features that were not identified were removed to reduce data noise and enable better interpretability.

The intensities of different metabolites can vary over large magnitudes. Thus, for metabolites to be comparable in analysis (e.g., comparing fold changes), metabolic features need to be scaled (van den Berg et al., 2006). Here we used either auto-scaling or pareto-scaling. When defining x_{ij} as the the j -th value (measurement) from the metabolite with the i -th index, \bar{x}_i as the mean of the i -th metabolite, s_i as its standard deviation and \widetilde{x}_{ij} as the transformed value; autoscaling can be defined by $\widetilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$ and pareto scaling is defined by $\widetilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$. For most classical statistical testing procedures, data should be normally distributed and possess sufficiently low heteroscedasticity. To achieve better normality, we calculated the generalized logarithm transform \widehat{x}_{ij} of x_{ij} . First $\widetilde{x}_{ij} = \log_{10} x_{ij}$ is calculated and then the mean of each metabolite \widetilde{x}_{ij} is deducted from each value and the transformed value \widehat{x}_{ij} is yielded as $\widehat{x}_{ij} = \widetilde{x}_{ij} - \bar{\widetilde{x}}_i$.

For first data exploration, we commonly used principal component analysis to assess the global separation of experimental groups. Furthermore, we used cluster-heatmaps to determine the global extent of metabolites altered between groups and find patterns of metabolites that are altered together (higher correlation). We used Euclidean distance with Ward-Clustering on the normalized data for heatmap analysis, if not otherwise mentioned.

2.0 General materials and methods

The Euclidean distance between two points (e.g., different samples) is the squared root of the quadrated differences in all dimensions (e.g., the difference between metabolomics features)

$$D = \sqrt{(x_{1,j} - x_{1,k})^2 + \dots + (x_{i+n,j} - x_{i+n,k})^2}, \text{ where } x_{ij} \text{ is the intensity/concentration of the } j\text{-th}$$

sample in the i -th metabolite, x_{ik} is the intensity in the k -th sample, that this pair-wise distance is calculated with (Gower, 1982). Iterating over all metabolites, results in the Euclidean distance between two samples, analyzing the distances between all samples results in a distance matrix. This distance matrix can then be used for clustering by the ward algorithm (Ward, 1963). Ward algorithm is an agglomerative clustering algorithm, that operates by merging the clusters (or points) that have the lowest sum of squared distances. For the hypothesis-generation purpose and to simplify data interpretability, typically only 65 or 80 metabolites with the lowest p -value after Student's T-test or ANOVA were used.

For correlation analysis we used Pearson correlation coefficient (Pearson, 1895). Pearson correlation coefficient is defined as: $R_{x,y} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$, where x_i and y_i are two individual sample points (e.g., metabolite concentrations) to be correlated, n is the sample size, \bar{x} and \bar{y} the sample means for x and y respectively.

For the evaluation of linear regressions we used the coefficient of determination $R^2 = \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$, where y_i is the observed data point, \bar{y} the mean of all y_i and f_i the calculated data point (Ellis et al., 1968).

2.2.12.3 Specialized statistical data analysis with python

For specialized data analysis we used python with the Matplotlib, Seaborn, and Scikit-learn libraries (Hunter, 2007). Volcano plots for two-group comparisons were generated by first calculating Student's T-test p -values and foldchange by using Metaboanalyst. Then we plotted the negative \log_{10} transformed p -value versus the \log_2 -transformed fold change using the python library Matplotlib. For data sets of the untargeted lipidomics method or the AbsoluteIDQ p180-kit, we used different colors to indicate lipid classes of the presumptively identified compounds.

For an overview of lipidomics data with two experimental groups before identification, we calculated foldchange and Student's T-test p -value for each metabolite. Then we calculated the \log_{10} of the fold changes between the groups and assigning the color blue to metabolites with a negative \log_{10} fold change (decreased metabolites) and red to the positive fold changes (increased metabolites). The transparency of the colors represented the magnitude of the change, with metabolites with the most significant fold change having the lowest transparency (the highest color boldness). Here the highest 5 % of the foldchanges were set at minimum transparency, and the other metabolites scaled using min-max-scaling.

The $-\log_{10}$ p -values of the Student's T-tests of each metabolite were then calculated and scaled as above. Scaled p -values were then used to display the point size. Then each metabolite was plotted using the color, transparency, and point size to the respective retention time and mass to charge ratio. This plot was then used to determine global changes in the lipidome.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

3.1 Introduction to the development of a metabolomics platform

The holistic nature of metabolomics analyses can provide deeper insights into the metabolism in diseased states than classical biochemical studies may allow (D'Alessandro et al., 2012; German et al., 2005). Therefore, the holistic nature may offer an advanced diagnostic power for rare neurometabolic disorders (Coene et al., 2018; Piras et al., 2016; Willemsen et al., 2016). However, the development of a metabolic platform consists of many steps that need to be optimized and untargeted as well as targeted methods that need to be developed (**Figure 5**). In parallel, we decided to establish also a commercial targeted metabolomics kit to gain first insights into metabolomics. This Absolute IDQ p180-kit from Biocrates company will provide our platform with an easily comparable method between institutes facilitating data exchange and reproducibility. While this method is established, we will then focus on developing an untargeted metabolomics platform to cover the metabolome holistically.

To summarize, in this project, we aim to set up a metabolomics platform based on two pillars:

- 1) An in-house untargeted metabolomics platform.
- 2) A targeted metabolomics platform, based on multiple methods, starting with the commercial AbsoluteIDQ p-180-kit.

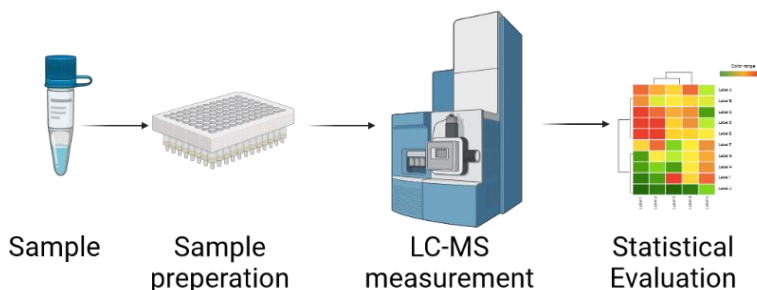


Figure 5 Metabolomics platforms need to include multiple steps in their pipelines. Metabolomics analyses require sample preparation, followed by LC-MS measurement and statistical evaluation. Created with BioRender.com.

3.1.1 The AbsoluteIDQ p180-kit for targeted metabolomics

As an complimentary approach to our in-house metabolomics pipeline, we applied the AbsoluteIDQ p180-kit from Biocrates. The p180-kit is a set of sample preparation and mass spectrometry methods as well as disposables and standards for the analysis of up to 187 metabolites from various sample matrices (<https://biocrates.com/absoluteidq-p180-kit/>, Accessed on 31.05.2021). It consists of a common sample preparation method for two different MS analysis methods (**Figure 6**):

- Firstly, a classical liquid chromatography mass spectrometry (LC-MS) measurement for biogenic amines.
- Secondly, a flow-injection mass spectrometry (FIA-MS) measurement without an LC separation to analyze lipids and carnitines.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

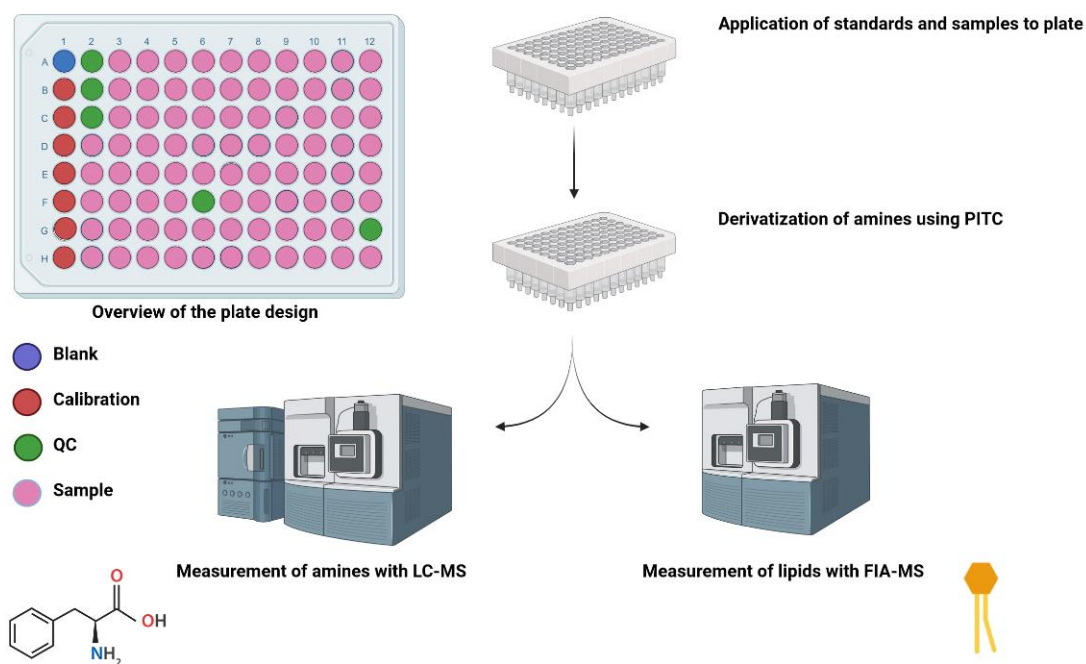


Figure 6 The AbsoluteIDQ p180-kit from Biocrates is a set of metabolomics method for the analysis of lipids and biogenic amines. The targeted metabolomics AbsoluteIDQ p180-kit from Biocrates is organized in a 96-well plate scheme, including the calibration curve, the quality controls (QC) and samples. After application to the plate, amines are derivatized with phenylthioisocyanate (PITC) and split into two dilutions for flow-injection analysis (FIA-MS) for lipids and LC-MS for biogenic amines. Created with BioRender.com

According to the manufacturer's application notes, the kit can quantify up to 145 metabolites in the FIA-MS mode (40 acyl-carnitines, 15 sphingolipids, 76 phosphatidylcholines, and 15 lysophosphatidylcholines). In the LC-MS mode, up to 42 metabolites (21 canonical amino acids and 21 other biogenic amines, e.g., neurotransmitters) can be quantified. The kit has been conceived based on recommendations of different regulatory bodies, such as the FDA and EMA (Bioanalytical method validation guide, EMA, 2012), and also it has been successfully validated in an international ring trial (Siskos et al., 2017). Additionally, the kit has already been applied in many different laboratories and to a myriad of various projects. Still, each kit application to another mass spectrometer or laboratory may pose certain insecurities that should be assessed. Thus, before deploying the kit, we first evaluated the general suitability of our devices and at least evaluated the quality and simplicity analysis qualitatively.

3.1.2 Introduction to the development of an untargeted metabolomics platform

Untargeted metabolomics is the metabolome analysis without actively excluding a set of compounds to generate an open overview of the phenotype (Schrimpe-Rutledge et al., 2016). Truthfully, a completely open method does not exist. Based on every method's specifics, there is always a certain bias to certain analytes due to their intrinsic chemical properties and analytes are excluded even in untargeted studies simply by analysis design (Christians et al., 2011). Thus, typically a combination of different techniques is used to acquire data from a more significant subset of metabolites. Most metabolomic platforms divide the analyses based on two general properties of a metabolite: the polarity of the analyte, separating hydrophilic analytes such as amino acids from unpolar analytes such as lipids; secondly, the ability to generate positively charged or negatively charged ions in mass spectrometry. Based on the different chemical properties, most methods for hydrophilic and unpolar analyses use different chromatographic protocols, while positive and negatively ionizable molecules often can use the same chromatography procedures (Masson et al., 2010). For a higher sample throughput, it is essential to reduce the number of chromatographic methods, as each requires a significant amount of time (typically 15-60 min per analysis).

Also, each different chromatography method increases the number of chromatographic columns and the solvent consumption, increasing the overall cost of the investigation.

Consequently, up to four mass spectrometry methods are used for most untargeted platforms, and between one to four chromatographic techniques are used. In the analysis of lipophilic compounds (for lipids: lipidomics), often analysis of negatively charged molecules is omitted due to the substantial prevalence of positively ionizable lipids (Naz et al., 2017; Wang et al., 2015). Because of the broader range in chemical space, hydrophilic methods typically analyze both. Opposing to this common platforms, there are also concepts that just use one analytical run (Schwaiger et al., 2019).

Based on this, we decide to develop one untargeted method in positive ionization mode for lipids and one chromatographic analysis method for hydrophilic compounds in positive and negative ionization mode.

3.1.2.1 Analytical LC-MS method development is needed for further steps

The development of a metabolomics method commonly starts reversed to the final procedures for the samples (Honour, 2011). First, the LC-MS analytical methods need to be developed, as the analytical method is the only way to assess features such as the efficiency of sample preparation (Gika et al., 2014). Generally, LC-MS development requires developing methods for the MS part and the LC part. Designing a metabolomics method often focuses on the chromatography method, but an MS method is necessary to evaluate this appropriately. However, the MS method is also dependent on the specific chromatographic environment (Furey et al., 2013).

Albeit mostly, the initial MS conditions are adjusted to notes of the manufacturer or literature. Then a reference mixture should be selected containing the target analytes in a sufficient amount, concentration, and purity (Bergeron et al., 2009). This reference is used for further optimization. For targeted analyses, simply the metabolites of interest are used. For untargeted analyses, this may present more complicated, as there is no definite target for optimization.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Some groups have used mixtures of metabolites with the highest possible chemical difference to assess all chemical space. Other groups have used biological material such as plasma from local donors or acquired standardized material such as the NIST standard plasma SRM 1950 (Ribbenstedt et al., 2018). The advantage of directly using plasma for optimization is the closeness to the final sample; however, many signals may overlay, and there is no definitive group for optimization, making the optimization analysis more difficult (Wang et al., 2019). The selection of specific metabolites to optimize simplifies analysis, but when analyzing actual samples after optimization, certain metabolites may be overlaid by other signals from the matrix.

After the selection of the optimization targets, optimization of the LC-MS method can occur. Optimization parameters include selecting the LC column, selecting the mobile phase, and lastly, the appropriate separation gradient. Each parameter can partially affect the other. Therefore, for complete optimization, all parameters need to be assessed in all possible combinations; however, this requires a high amount of time and material costs due to the high amount of different combinations. Thus, commonly, optimization follows an elimination pattern, also called a “one-factor-at-a-time” (OFAT) approach (Dolan and Snyder, 2017; Lakka and Kuppan, 2019; Peraman et al., 2015). At first chromatographic columns are compared using mobile phases as recommended by the manufacturer or literature. Then the best performing column is selected for further optimization. For reversed-phase analyses, the most common columns include columns with a saturated 18-carbon chain connected to the silica beads (C18 column), other columns such as C8 or even columns with perfluorinated aromatic groups are also available (pentafluorophenyl, PFP; (Criscuolo et al., 2019). For hydrophilic analytes, more modifications are available, including bare silica columns or silica modified with amine (NH₂), amide, or a variety of different zwitterionic groups (ZIC) (Buszewski and Noga, 2012).

Then the different mobile phases are optimized, changing the solvent, buffering salt, buffering salt concentration, and pH in each round of optimization. As buffers in MS need to be volatile, typically only ammonium acetate or ammonium formate are used. For most methods, a range from 5-30 mM salt is used. Conjugated acids such as acetic acid and formic acid (Monnin et al., 2018), or bases such as ammonium hydroxide (Tang et al., 2016) can be added to modify the pH. For specialized methods ion-pairing reagents, amphiphilic molecules that enable separation of polar metabolites on RPLC columns, exist. However, they are usually limited to peptide analysis (Åsberg et al., 2017; Guo et al., 1987). Even metal ions, such as cesium, have been described as additives for carbohydrate analytics (Rogatsky et al., 2005).

As metabolomics methods are often screening methods, the content of organic and aqueous components in mobile phases is often as high (or low) as possible to achieve a maximum range of separation. For the analysis of hydrophilic molecules, HILIC methods are often used. A high organic starting condition is therefore required, conditions higher than 90% are mostly described (Buszewski and Noga, 2012). Reversed-phase techniques require a starting mobile phase with almost no organic content, e.g., typically below 10%.

Elimination-wise optimization is a suitable mechanism for optimization in a fast and resource-saving manner. Though, it leads only to local optimization.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

A better method may be available but is not found, given that only a specific combination of parameters leads to it. One way to reach a global optimum is to test all possible combinations, which is not feasible. However, newer strategies have introduced automatized multiparameter optimization techniques, that by strategically altering multiple parameters at once, can create a more global overview of method optima. These techniques use polynomial equation solving techniques such as response-surface modeling for finding optimal method parameters (Tome et al., 2019). Multi-platform technologies are available, such as the “multi-platform unbiased optimization of spectrometry via closed-loop experimentation (MUSCLE)”-tool development by the University of Birmingham (Bradbury et al., 2015).

In typical LC-MS method optimizations for regulated environments, but also for targeted LC-MS analyses in research, specific method parameters are assessed. During the development of the LC-MS method itself, the linearity of the method and analytical stability (e.g., retention time stability) are determined.

3.1.2.2 Development of analytical sample preparation pipelines makes sample amenable to analysis

Following the optimization of the LC-MS method, the sample preparation pipeline needs to be developed. Optimization aims to decrease unwanted compounds while increasing compounds of interest and making the sample amendable for analysis (Vuckovic, 2012). In untargeted metabolomics, the extracted metabolites should be from the widest chemical range possible.

To minimize metabolite alterations and possible contamination risk, samples should be handled as fast and with the lowest treatment as possible. In metabolomics, simple techniques from basic protein removal to liquid-liquid extractions and elaborate solid-phase extractions are used.

To assess sample preparation efficiency, mainly important is the determination of signal abundances and reproducibility of targeted analytes, optimizing the methods for high signal intensity and high reproducibility. To generate more easily interpretable data, in untargeted metabolomics, this can be done on a metabolite class levels (e.g., “amino acids”) or based on global intensities (e.g., total ion chromatogram). Some groups have monitored protein removal efficiency as a marker (utilizing protein assays) for analytical cleanup (Polson et al., 2003). For hydrophilic analytes, especially the removal of phospholipids is measured (Bylda et al., 2014). Protein aggregates lead to clogging of instrument capillaries and columns, and phospholipids can lead to lower sensitivities by causing ion suppression.

Protein removal

Protein removal can be achieved utilizing flow molecular weight filters (typically below 3000 Da) or by precipitation using adding organic solvents. Ultrafiltration is commonly used for biofluids (e.g., urine) and removes proteins based on their size. It does not change the solvent composition of the analytes. However, certain metabolites are not liberated from their protein carriers, leading to unwanted removal. In high proteinaceous samples (cell lysate, blood), filters clog and are often not usable. Protein precipitation techniques are more commonly used and usually work by denaturation.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Strong pH changes by acids (Link and LaBaer, 2011), changing ionic strength by using chaotropic salts (Shih et al., 1992), flocculation by polyelectrolytes (Chen et al., 1992), nanomaterials (Zhang et al., 2016), or removing solvation shells by organic solvents (Merrill and Fleisher, 1932) can precipitate proteins. The first ones are uncommon to LC-MS as they can lead to interferences or reduces metabolite stability. Instead, organic solvent precipitation is used predominantly (Vuckovic, 2012).

Different protein precipitation solvents better extract different compounds and different sizes and precipitates' stabilities (Polson et al., 2003; Want et al., 2006). Acetonitrile, methanol, acetone, and butanol are typical organic solvents used and they are often mixed or modified using acids and salt buffers to regulate pH and ionic strength.

Metabolite extraction

Certain groups of metabolites can be extracted from the sample to reduce interference from the respective not extracted group (Yang et al., 2013b). For example, lipids can be extracted from hydrophilic metabolites to minimize ion suppression from phospholipids. Metabolite extraction primarily uses either liquid-liquid extraction or solid-phase extraction.

Two-phase liquid-liquid extraction employs the use of an aqueous sample buffer and an organic, water-immiscible solvent (Martin et al., 2014). Different solvents are used, with chloroform being the most common (Bligh and Dyer, 1959; Folch et al., 1951), but also ethyl-acetate or methyl-tert-butyl ether (MTBE) are used, with the latter being favored due its lower volatility and lesser toxicity (Matyash et al., 2008). Liquid-liquid extraction can be optimized by choice of organic solvent, pH change, salts to change ionic strength, and the volume ratios of extraction solvent and sample (Ulmer et al., 2018).

Liquid-liquid extraction cannot handle many samples at the same time and is not easily automatable. Based on the manual liquid handling, the inter-sample deviations are also high. While matrices such as diatomaceous earth have been used for an automatable solid-phase supported liquid-liquid extraction (Cheng and Jiang, 2019), they are optimized for only the recovery of one phase (primarily organic phase). Thus, these methods cannot be used when both extracts are relevant.

The most automatable sample preparation technique is the use of solid-phase extraction (Dettmer et al., 2007). Similar to liquid chromatography, this technique uses a silica stationary phase coated with functional groups and filled in a column structure, and compounds are separated from a liquid stationary phase. Typically, the silica particles have a bigger diameter enabling higher flow rates with lower backpressure. The elution of compounds from SPE also differs; generally, elution occurs step-wise, with different solvent volumes applied and not by continuous flow as in LC-MS (Poole, 2003). Other formats are widely available, including higher throughput able 96-well formats and even integrates systems that use SPE during an LC-MS measurement (online-SPE, (Calderón-Santiago et al., 2015)). Very similarly to liquid chromatography, the selected solid phase and the selected solvents need to be optimized.

Additionally, based on the stepwise nature of SPE volumes applied to the column need to be optimized. Optimizing follows similar principles as in LC optimization.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

3.1.2.3 Normalization of samples corrects for sample-wise and artificially introduced intensity variation

Loss of analytes during sample preparation is unavoidable (Ejigu et al., 2013; Misra, 2020). While this should be reduced by designing appropriate sample preparation pipelines and reducing the number of cleanup steps to a certain amount, it will persist. Clearly, analyte loss influences sensitivity, but more important is to monitor the homogeneity of analyte loss in every sample. This is important to achieve high reproducibility. Typically, analyte loss is very dependent on the specific metabolite due to individual properties (e.g., logP, pKa), and analyte losses do not linearly correlate with each other. The solubility of each analyte also depends on the different analytes in the sample, contributing to increasing solubility or decreasing it. Apart from the sample preparation, errors can also be introduced by the LC-MS measurement. Here samples can differ in small amounts, or analyte intensities can vary due to sample-specific ion suppression by matrix compounds, such as phospholipids (Annesley, 2003; King et al., 2000).

Therefore, in targeted analyses, every analyte of interest typically corresponds to a respective internal standard with the same chemical properties (Bansal and DeStefano, 2007). In most LC-MS analyses, the same metabolites are used as the targeted metabolites, only differing in isotopic composition; this is called isotopic dilution mass spectrometry (Ciccimaro and Blair, 2010). The stable hydrogen isotopes deuterium (H^2), but in fewer cases, the carbon isotope C^{13} are exchanged with their more abundant counterparts protium (H^1) or C^{12} . The isotopically modified internal standards retain most chemical properties (e.g., extraction efficiency, retention time, or ion suppression) while being identified as different substances by their varied masses in the mass spectrometer. However, these are expensive and not available for all compounds making the reduction of their use necessary (Stokvis et al., 2005). Especially in untargeted metabolomics, target analytes are unknown beforehand, and very varied, so specific internal standards cannot be added.

Instead, then internal standards are selected for distinct classes e.g., in lipidomics for each lipid class (Boysen et al., 2018; Sysi-Aho et al., 2007; Wang et al., 2017). Sometimes just single substances are added for normalization of extraction. This is common for untargeted screening methods as the standard deviations are accepted to be higher.

Other internal standard types exist, such as standards similar in chemical aspects, but not occurring naturally in the samples, e.g., tricarballic acid for organic acids or miltefosine for phospholipids (Godfrey et al., 2017). Those are commonly cheaper than isotopically modified standards, but they to be carefully validated.

Normalization can also occur on a data level after measurement. Used for normalization are the total ion chromatogram (TIC), the total ion chromatogram of identified metabolites (mTIC), or other multiparameter methods (Ejigu et al., 2013; Misra, 2020). Among the latter, the locally weighted scatter plot smoother (LOESS) algorithm is the most common (Cleveland, 1979; Dunn et al., 2011; van der Kloet et al., 2009).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

3.1.2.4 Analytical method validation is the last step in metabolomics method development

The expected last step of developing an analytical platform is method validation (Bansal and DeStefano, 2007; Honour, 2011). Different authorities document method validation guidelines for controlled environments such as good laboratory or good manufacturing practices (ICH, FDA, EMA). Validation of metabolomics methods for scientific use is, unfortunately, less well documented. Currently, there are efforts to standardize validation criteria, but typically there are some similar procedures followed as in EMA or FDA guidelines. Based on the wide and untargeted nature of potential analytes as well as the endogenous nature of analytes, these guidelines can only be followed in very open ways (Broadhurst et al., 2018; Coene et al., 2018; Klupczynska et al., 2020; Ribbenstedt et al., 2018). The assessment of the following parameters is generally recommended by FDA, EMA, ICH or others (FDA, EMA, ICH; Bansal and DeStefano, 2007):

- specificity/the selectivity
- sensitivity
- linearity
- accuracy and precision
- stability
- matrix effects, dilution integrity

Assessment of specificity and selectivity

The specificity and selectivity are typically the first parameters to be assessed as they are essential to determine the identity of a compound (Peris-Vicente et al., 2015). Specificity of a method is the ability of the method to react only to one compound with an appropriate response but not to others. Selectivity is the ability to differentiate the analytes from other endogenous analytes. In typical validations, authorized reference standards are used for each compound. In untargeted metabolomics analysis, this is usually not available. In LC-MS the important identification criteria for metabolites are retention time and the mass/charge-ratios of the compound as well its fragments (Kushnir et al., 2005; Rozet et al., 2011; Saah and Hoover, 1997).

In classical method validation, specificity and selectivity are assessed mainly by determining the aberration of these parameters in samples to the reference standards. Additionally, the reproducibility of retention time and mass parameters between different runs in a sample can be assessed (inter-run retention time stability and mass accuracy). The latter criteria can even be determined in untargeted metabolomics methods as shown e.g., in a publication by Stefanie Wernisch and others (Wernisch and Pennathur, 2016).

Assessment of sensitivity

The sensitivity of compounds is the ability to detect a compound from a matrix safely (Peris-Vicente et al., 2015). Conceptually it can be separated into two distinct parameters: the limit of detection (LOD) and the lower limit of quantification (LLOQ). The limit of detection is defined as the lowest possible concentration where the analyte is safely identified from the surrounding noise. The ICH accepts three ways to determine the LOD: visual evaluation, signal-to noise-ratio and standard deviation of the blank.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

The LLOQ is defined as the lowest possible concentration that can be quantified and depends on the signal/noise range and the linearity of the method. Commonly, the sensitivity is assessed by standard addition to the sample matrix. In endogenous analytes, the isotope standards are being used. If not otherwise possible, sample matrix substitutes are possible, containing typical matrix compounds but the target analyte.

The FDA and EMA recommend the LLOQ to be the lowest non-zero standard of the calibration curve. Its intensity should be at least five times higher than the zero calibrators.

In untargeted metabolomics, this presents as difficult. Metabolites expected in analyses are commonly unknown, of a high number and wide range of different analytes. A possible proxy marker for the sensitivity is the dilution integrity. By diluting the sample in various stages, at least a broad information can be yielded by which dilution factor becomes undetectable.

Calibration linearity

The calibration linearity is a marker for the dependency of the analytes concentration and instrument response caused by it (Shabir, 2003). Sufficient linearity is necessary for the quantification of an analyte. The linearity is commonly determined by adding a nonendogenous compound to the sample matrix and plotting the resulting signal intensity. For endogenous analytes, matrix substitutes or internal isotopic standards can be used. The calibration points should encompass the range of concentrations that should be determined later. Typically, apart from matters of linearity, the LLOQ should be at least 5 times higher than the blank. The regression coefficient is used for assessing linearity; in most quantitative methods, an $R^2 > 0.9$ is found (Karnes and March, 1991). In untargeted metabolomics, linearity cannot be determined usually due to the absence of standards and number, but similar to the sensitivity might be assessed by doing sequential dilutions.

Precision

Precision is the property of an analysis to determine the same value for the same sample between different analysis trails. Accuracy is the ability to produce analytical results as close as possible to a known reference.

Precision can be assessed by determining the covariance (CV) of analyte concentrations between different runs and accuracy to the known reference. The FDA recommends a maximum covariance of 15-20% for accuracy and precision. As there is no concentration reference in metabolomics, precision is mainly determined as intra-run signal stability (Lin et al., 2020). In untargeted metabolomics, CVs tend to be substantially increased, as they are semiquantitative methods.

Metabolite stability

The stability of an analyte can be determined by treating the sample with different conditions (Stevens et al., 2019). Here influences on the sample should be studied that are relevant to typical analyses. Typical is the effect of different time spans or temperatures on different steps of the analysis. Different matrices also can influence stability in different ways. For instance, occurrences such as hemolysis can influence multiple parameters (Denihan et al., 2015).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Specific analytes may be increased due to the higher intracellular concentration of metabolites in erythrocytes. Some other analytes may be decreased due to instability or due to problems in sample preparation.

Matrix effects

Matrix effects can be judged by analyzing the difference between the analyte signal intensity in a pure solvent and the intensity in an actual sample. Due to the different impacts such as, e.g., ion suppression, analytes present with lower intensities in real matrices such as blood (Chamberlain et al., 2019). For example, certain metabolites in the blood may not be detectable, but they are detectable in urine even though they had the same concentration. In untargeted metabolomics, this is mostly relevant when comparing different matrices. In general, this comparison is deprecated, even though the previously mentioned effect of hemolysis on the metabolome can still be studied.

3.1.2.5 Identification of analytes need to be optimized in metabolomics

In the strictest sense, in early biomarker development or early untargeted metabolomics, metabolomic features do not require identification (Dettmer et al., 2007). Each feature can be assigned a unique identifier (typically retention time and mass/charge ratio). This identifier can then be used for statistical analysis, and parts can be identified later after becoming relevant in the statistical analyses. However, it is crucial to identify analytes for doing pathway analyses and separating data from the noise. Mass spectrometry provides mass/charge-ratios as a basal identification criterium. Based on combinatoric calculations, a list of empirical formulas can be calculated and matched to online databases. In MS/MS also fragments can be matched with potential hits in databases (Kind and Fiehn, 2007; Schymanski et al., 2014). Lastly, isotopic distribution criteria can provide a secondary criterium for identification. There are a vast number of different libraries accessible and quickly provide preliminary identifications such as the human metabolite database (HMDB) by the lab Prof. Wishart (Wishart et al., 2013). Additionally a list of minimum identification criteria has been proposed by the metabolomics standards initiative (Sumner et al., 2007). However, this identification is typically of lower confidence, mainly due to isobaric compounds. These are structurally different compounds that still have the same (or similar enough) mass/charge-ratio to not be distinguishable by mass alone (e.g., leucine and isoleucine). Hundreds of isobaric analytes can be found in certain mass ranges, severely impacting identification confidence and increasing misidentifications (Opialla et al., 2020).

Retention time and fragmentation patterns are parameters with supreme importance for the identification of metabolites. These parameters cannot be calculated but have to be determined empirically (Xiao et al., 2012). Classical untargeted metabolomics methods can only identify metabolites by their calculated mass and by potential fragments, thus suffering from a high rate of misidentified analytes. After determining features of interest, their identity must then be confirmed in yet another measurement. This two-step procedure can take a long time due to the partly long times for reference compound delivery and interpreting data that is not useful for analysis.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

To address this issue, we decided to create a different set-up for our untargeted method platform. For identifying a metabolite, it does not need to be targeted. Every untargeted data set still contains the criteria necessary for identification: retention time and mass so that metabolites, where those are known, can be identified without changing the original data set (Coene et al., 2018; Naz et al., 2017).

3.1.2.6 Summary of the introduction to metabolomics method development

Rare neuropediatric disorders are typically complex disorders, affecting multiple metabolic pathways and interacting with the environment. To better understand and diagnose these disorders, we decided to build a metabolomics platform. We chose to concept a metabolomic platform based on an untargeted in-house metabolomics method and targeted metabolomics methods, including a commercially available metabolomics kit.

To enable faster and more confident identification in our untargeted method, we decided to build a metabolite reference library, including spectral data, fragment data, and retention times. In this library, the retrospective addition of metabolites and the retrospective identification of compounds should be possible, too.

As for most metabolomics platforms, there is not much validation information available, consequently we were also interested in how this method performs in different validation criteria. Additionally, we were also interested in seeing changes in metabolite stabilities under different circumstances and finding out about the limitations of the method.

To summarize, using targeted metabolomics with the AbsoluteIDQ p180-kit and untargeted metabolomics analysis with an in-house metabolite identification library, we created a metabolomics platform for the analysis of rare neuropediatric disorders.

3.2 Methods for the development of the untargeted LC-MS platform

3.2.1 A LC-MS metabolite library of authentic reference compounds

For accurate metabolite identification, LC methods heavily rely on retention time and fragments of analytes as prime identifiers. However, these parameters cannot be calculated (as, for instance, the molecular mass) and therefore have to be determined empirically. To increase identification confidence in time-saving manner, we decided to create an in-house metabolite identification database (**Figure 7**).

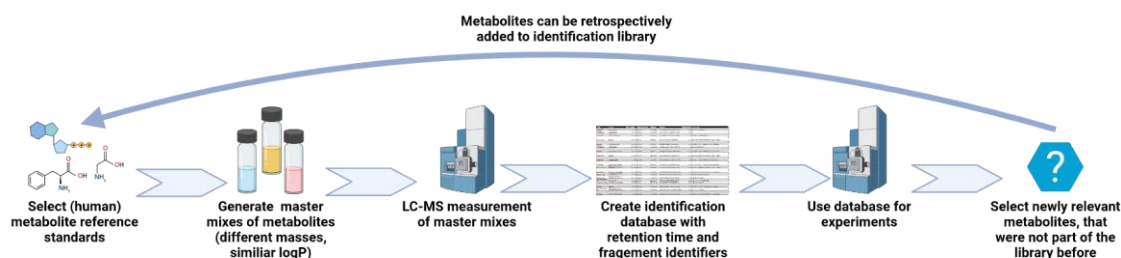


Figure 7 The metabolite identification library is generated from the measurements of metabolite reference standards. Metabolites for the library are selected and combined into master mixes to reduce measurement time and retention time and fragment data acquired using LC-MS. Identification data is then curated into the in-house database and then can be used to identify metabolites in experimental data. When metabolites relevant to an experiment are found, which were not part of the database before, they can be retrospectively included. Created with BioRender.com.

3.2.1.1 Human metabolite standards of important pathways are preferentially selected

As the purchase of a sufficient number of metabolites in singular purchases is highly laborious and expensive, we selected a company for the custom preparation of metabolite standards. Here we used MetaSci (Toronto, Canada) to produce a customized ordered library of undissolved reference standards placed in small glass vials. First, we manually selected the offered 1200 metabolites based on their relevance to our research. We primarily selected metabolites participating in endogenous human metabolism, with a focus on inherited disorders of metabolism.

A small number of xenobiotics were selected because of their importance in nutrition or their involvement in therapies common to neuropediatric practice (e.g., anti-convulsive drugs). Based on these constraints, we acquired 502 reference standards, a list of metabolite reference standards from MetaSci can be found in **Table 22** (p.181).

3.2.1.2 Master mixes are were created to shorten measurement time

Commonly, LC-MS measurements can last for 30 to 60 mins; thus, measuring 502 metabolites as single measurements of each compound is unfeasible. Hence, we combined multiple compounds into 39 different master mixes, reducing measurement time drastically.

For this, we first determined whether a compound is rather soluble in water or organic solvents by determining the logP from either experimental or predicted data. The logP is the decadic logarithm of the octanol-water-partition coefficient, which can be a measure for lipophilicity (positive logP) or hydrophilicity (negative logP) of a compound (Sangster, 1997). Then samples with a $\log P > 0$ and $\log P < 0$ were grouped separately. More hydrophilic compounds were dissolved in water, more hydrophobic compounds in acetonitrile. Compounds were diluted to have a concentration of 5 mg/ml or 10 mg/ml.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

In either group, we then determined the monoisotopic mass and grouped metabolites in a way that all metabolites in a master mix had a mass difference of more than 1 Da. This was done, to avoid misidentification by having peaks with the same mass/charge ratio in the same measurement. Then metabolites were added to their master mixes and diluted to a final concentration of 100 μ M in 50% acetonitrile/ 50% water. Before measurement, metabolites were stored at -80°C.

3.2.1.3 Measurement of metabolite standards and generation of library database

Based on the method screening studies results, we selected the ZIC-HILIC-MS method as the primary method for the generation of the library. Thus, we acquired LC-MS data in the positive as well as in the negative ionization mode of each master mix. Using UNIFI, we then processed each master mix and evaluated each metabolite that the mix contained. Peaks that had a sufficient signal intensity (>1000 A.U.) and sufficiently narrow peak widths (<1.5 minutes) were included in the library with their respective retention time. Based on their structural formula UNIFI predicted possible fragments for each metabolite. We evaluated the predicted structural formula whether the fragment chromatogram matched the main adduct chromatogram (coelution, similar profile). From this, only fragments that were expected to behave reproducibly in ESI-ionization were included. Principally excluded were carbon-carbon bond breaks (e.g., acyl-chains) and aromatic ring breaks due to their low repeatability under ESI conditions. Fragments of bonds between functional groups (e.g., amides fragmenting in carboxy- and amino-groups) and fragments leaving a highly stabilized group (such as tryptophan fragmenting into an aromatic indole) were preferred. Chromatograms, fragmentation spectra and from this corresponding retention time and fragmentation data were stored in a database system using UNIFI.

For use in other programs (such as open-source mass spectrometry software, e.g., MS-Dial), retention time and fragmentation information were exported to a tabular format.

3.2.2 Generating an identification file for use in open-source metabolomics programs

To facilitate the possibility of open data exchange and for use in open-source metabolomics software, we exported our retention time library into other formats. Commonly used are ASCII-text-based formats such as the MSP format developed by the national standards institute for science and technology (NIST). This is a standardized text format containing names, molecular ions, and fragment masses with corresponding relative intensities. Additionally, other groups have added retention times and molecular structure identifiers such as InChIKey or SMILES. To identify metabolites that were not part of our library (even in a less confident manner), we combined our library with a further library. Here we used the FiehnHilic library, which can also be openly accessed through the MassBank of North America (MoNA) (<https://mona.fiehnlab.ucdavis.edu>, last accessed on 10.06.2021). In our identification file, we replaced metabolites in the FiehnHilic-File with our identification data.

3.2.2.1 Processing of mass spectrometry data using UNIFI

Especially for the systematic analysis and organization of mass spectrometry data for the metabolite library, we selected the UNIFI mass spectrometry software from Waters. This software provides the ability to analyze mass spectrometry data and extract discrete metabolic features from the provided sample. For the identification of mass spectrometry features, the software accepts a predefined list of compounds and structural formulas.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Using the structural formula, UNIFI can predict theoretical fragments and, where applicable, points them out in the experimental fragmentation spectrum. The experimentally verified retention time and fragmentation data can then be saved with spectral data using a database functionality. Additionally, the software is applied to a server that provides regular backup functions.

The combination of data analysis and the possibilities for database curation make this software an ideal tool for the generation of a human metabolite identification database.

After completion of the mass spectrometry measurement, data was imported into UNIFI for processing. The parameters were based on the recommended settings and optimized to the type of liquid chromatography and mass spectrometer used. The optimized parameters can be found in **Table 6** and **Table 7**. Generally, we used “Accurate Mass Screening Analysis on MSe data”.

Table 6 UNIFI processing parameters for ZIC-HILIC positive mode. Processing parameters for the processing of ZIC-HILIC-MS data in positive mode for the metabolite identification library with UNIFI software. Parameters were modified based on recommendation of the software.

Smooth		3D Isotope clustering		Analysis Specific Settings	
Smoothing algorithm:	mean	Maximum considered charge for cluster:	1	Adducts:	+H, +NH4, +Na
Half width:	5	Maximum number of isotopes per cluster:	3	Lock Mass	
Iterations:	1	Target by mass		Combine width:	3
Find 3D Peaks		Match tolerance:	20 ppm	Mass window:	0.5
Apply lockmass correction:	yes	Fragment match tolerance:	5 mDa	Reference mass:	556.2766
Chromatographic peak width	1.5	Allow scores below	5		
High energy intensity threshold:	200	Look for in-source fragments:	True		
Low energy mass start:	55 Da	Maximum Candidates:	50000		
Low energy mass end:	800 Da	Maximum Unidentified Candidates:	10000		
Low energy intensity threshold	500	Maximum number of targets with the same m/z to keep	5		
Maximum peaks per channel	10000				

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Table 7 UNIFI processing parameters for ZIC-HILIC negative mode. Processing parameters for the processing of ZIC-HILIC-MS data in negative mode for the metabolite identification library with UNIFI software. Parameters were modified based on recommendation of the software.

Smooth		3D Isotope clustering		Analysis Specific Settings	
Smoothing algorithm:	mean	Maximum considered charge for cluster:	2	Adducts:	-H, +Cl
Half width:	5	Maximum number of isotopes per cluster:	3	Lock Mass	
Iterations:	1	Target by mass		Combine width:	3
Find 3D Peaks		Match tolerance:	25 ppm	Mass window:	0.5
Apply lockmass correction:	yes	Fragment match tolerance:	5 mDa	Reference mass:	554.2609
Chromatographic peak width	1.5	Allow scores below	5		
High energy intensity threshold:	50	Look for in-source fragments:	True		
Low energy mass start:	50 Da	Maximum Candidates:	50000		
Low energy mass end:	800 Da	Maximum Unidentified Candidates:	10000		
Low energy intensity threshold	250	Maximum number of targets with the same m/z to keep	5		
Maximum peaks per channel	10000				

3.3. The AbsoluteIDQ p180-kit, a general description of results

The AbsoluteIDQ p180-kit is a commercially available set of sample preparation methods, various standards, and LC-MS methods to enable mass spectrometry-experienced laboratories to conduct metabolomics experiments quickly (<https://biocrates.com/absoluteidq-p180-kit/>, accessed on 31.05.2021). Using one sample preparation, the kit measurement is separated into two parts: a classical LC-MS analysis for the determination of biogenic amines, and secondly a flow-injection mass spectrometry analysis for the determination of various lipid classes. While the kit has been validated in an international ring trial (Siskos et al., 2017) and therefore is shown to produce comparable results between many laboratories, each kit application to a lab has specific challenges. Thus, the first kit application is commonly accompanied by the in-house modification of parameters by a Biocrates application scientist. Additionally, prior to every kit measurement, the system must be checked for suitability using included tests. So, while the kit is a commercially, thoroughly validated kit, we deemed a short note on the application of the kit to our lab a necessary part of this work. While the kit has been widely used to study different research areas, to our knowledge, rare neuropediatric diseases have not been studied widely (Klemp et al., 2021).

3.3.1. *The AbsoluteIDQ p180-kit: System suitability testing shows good identification reliability*

Before each run of the commercial kit, the system suitability of the LC-MS system for the p180-kit needs to be examined. For this, the LC-MS method and FIA-MS method are tested separately. Due to the lower stability of the biogenic amines studied by the LC-MS method, the kit supplier recommends starting the kit measurement with the LC-MS part. We made the experience that the FIA-MS system suitability should only be checked after the LC-MS run, as the LC-MS run can lead to slight contaminations of the mass spectrometer (Keller et al., 2008; Yao et al., 2016a). These may not be noticed during LC-MS runs, but in the FIA-MS run with commonly very low sample injection flows, it becomes evident.

The system suitability tests of the LC-MS part require the most user interaction; due to the natural aging of the chromatographic columns, the retention time slightly varies between measurements (McMaster, 2007). Therefore, to guarantee optimal detection of the metabolites, during the system check, each metabolites' retention time has to be examined and the mass spectrometry method adjusted to it. We observed that different kit runs with the same column varied between around 0.02 minutes and about 0.04 minutes with different columns. During nine kit runs, with around 864 injections, we used three different LC-columns due to peak and intensity deterioration. In classical LC-MS methods, similar columns last around 1000 injections. Similarly, we observed that certain metabolites in the test-mix, such as the polyamines spermine, spermine, and the alanine-derivative sarcosine, often presented with peaks not applicable for analysis. Due to multiple amino groups, polyamines are strongly positively charged and are not easily separated in RPLC. Even when using advanced ion-pairing methods, peak tailing still occurs, making polyamines notoriously difficult to analyze (Häkkinen, 2011; Magnes et al., 2014).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

The system suitability testing of the FIA-MS part begins with the checkup of blank intensities to examine potential contamination by lipids in the analysis. Lipids are omnipresent in most sample preparations in the clinical setting (Keller et al., 2008; Yao et al., 2016a) and, due to their amphiphilic nature, often accumulate in mass spectrometry systems (Guo and Lankmayr, 2011). Therefore, if studying lipids samples with low intensities, it is important to reduce background noise caused by lipid contamination in the system.

Also, the lower limit of quantification of the FIA-MS part is determined by the blank intensity. Thus, an extensive cleaning protocol was necessary before every FIA-MS measurement. Then a system suitability test mix with a proprietary mixture of compounds was injected to assess signal intensity and stability. Due to the compounds' proprietary nature, the test mix assessment had to occur through the kit software MetIDQ. Here we occasionally saw unstable signals but generally sufficient intensities (**Figure 8**).

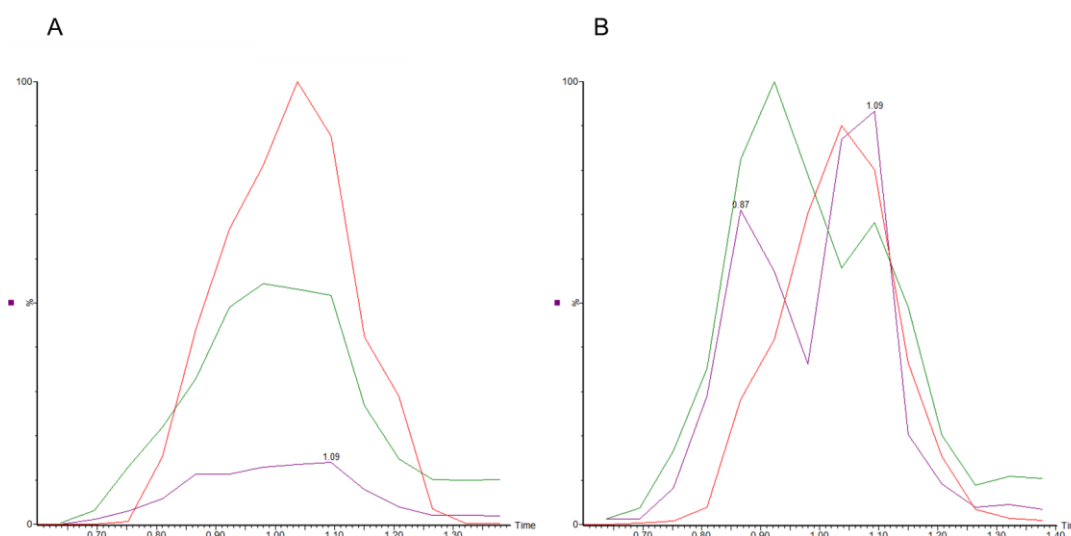


Figure 8 Exemplary chromatograms of acceptable and problematic FIA-MS measurements. A: Typical chromatogram of FIA-MS measurement using the Biocrates p180-kit, B: Signal deterioration and collapse in an FIA-MS measurement that leads to quantification issues. Three different exemplary mass/charge-ratios are shown in different colors in both chromatograms.

3.3.2 The AbsoluteIDQ p180-kit: Quality control samples show validity of most metabolites
After the kit measurement using LC-MS and FIA-MS, the data sets are imported into MetIDQ. Here the software also performs quality controls necessary for the validation of the kit. While the validation of most LC-MS parameters presented flawlessly, in some cases, the polyamines (spermine/spermidine) presented faulty validation, as already seen in the system check. Additionally, leucine and isoleucine often did not pass the quality check.

In the international ring trial, the spermine, spermidine, and carnosine also presented problems in the validation, exceeding their specifications in two points higher than 20% (Siskos et al., 2017).

In the FIA-MS part, certain runs had to be repealed entirely, as they could not be validated. Further assessment showed low signal intensities as well as high noise and deviating flow profiles. Only after repeated cleaning, signal intensities were restored and the kit validated. In no kit measurement, we were able to quantify the sum of hexoses. To our knowledge, the entire kit measurement is conducted in positive ionization mode; therefore, hexoses that are typically only visible in negative mode might be undetected.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

In the ring trial by Siskos et al., SM C18:0 and C12 acyl-carnitine were found to have poor inter-laboratory reproducibility (Siskos et al., 2017). Apart from the mentioned difficulties, the other metabolites were validated well.

3.3.3 *The AbsoluteIDQ p180kit: 167 metabolites are typically visible in human serum*

In human blood-derived biofluids such as plasma and serum, we commonly found around 167 of the 187 quantifiable kit metabolites. Thus the kit performed comparably to the ring trial of Siskos et al., where they excluded 27 compounds as they were below the LOD or showed high variability (Siskos et al., 2017).

They also documented that the lipids PC aa C30:2, PC aa C32:2, PC aa C38:1, SM C22:3 are in general not analyzable by mass spectrometers from Waters, our MS supplier. Certain metabolites such as the previously mentioned polyamines and the sum of hexoses were not found. Additionally, the number of metabolites varied slightly between measurements as the limit of quantification also changed between measurements. Apart from the mentioned limitations, we were successfully able to use the kit in multiple projects in serum, plasma, and tissue, which are being discussed in later chapters, but two are also published (Kettwig et al., 2021; Klemp et al., 2021).

3.4. The in-house platform for untargeted metabolomics analysis provides broader metabolite information

The AbsoluteIDQ p180-kit as a targeted assay is an excellent introduction to metabolomics analysis. However, our metabolomics platform's main goal is to screen a wide range of the metabolism in an untargeted manner to find even previously undocumented metabolic aberrations in rare neurometabolic disorders. Based on previous literature (Contrepolis et al., 2015; Naz et al., 2017; Ribbenstedt et al., 2018; Tang et al., 2016; Wernisch and Pennathur, 2016), we decided to divide our untargeted analysis into 2 parts: a method for relatively hydrophilic compounds with smaller masses common to primary metabolism; secondly a method for the study of lipid metabolism. Ideally, both methods should possess orthogonality, together covering most of the metabolism without too much redundancy. While there have been many methods for metabolomics proposed, the exact pipeline heavily depends on the infrastructure accessible (e.g., mass spectrometer type) and is strongly dependent on the focus of research institutes. Here we aimed to develop a platform suited for the needs of neuropediatric research and potentially diagnostics. Accurate identification using retention time information is of greater importance for hydrophilic analytes. We first began our untargeted platform development with the development of our LC-MS method, then continued generating the metabolite reference library, established the sample preparation, and concluded with an insight into essential validation parameters.

3.4.1 A ZIC-HILIC column is an appropriate selection for a hydrophilic metabolite LC-MS/MS method

Method development for a metabolomics method requires authentic reference compounds that the method should be optimized with (Bergeron et al., 2009). Using these reference compounds, general parameters such as retention, intensity, and peak shape can be judged more easily than when using a prepared sample that might introduce matrix effects. As a reference we decided to use a standard mixture of biogenic amino acids, commonly used in our institute.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Amino acids have a wide range of polarities and pKa (and therefore different retention), are commonly found and easily acquired, and have supreme importance in the field of inherited disorders of metabolism (Le et al., 2014).

Target of the optimization was to generate a method with a broad spectrum (increase the number of metabolites seen) while having high sensitivity (increase the signal intensity). Optimization of the peak shape played secondary roles.

Firstly, we conducted intensive literature research on methods that other groups have used to study metabolomics. Then we decided to first start with a comparison of different column chemistries. We decided to examine two different chromatography concepts: Reversed-phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (Buszewski and Noga, 2012; Tang et al., 2016). As there is a wealth in different column modifications for HILIC, we also determined the applicability of other HILIC-columns (Jandera, 2008). However, comparison of varying column chemistries is difficult, as each column chemistry requires different mobile phases and mass spectrometry conditions. For example, a reversed-phase column initially requires very low organic mobile phases, but HILIC chemistries require very high organic phases (Alpert, 1990). Thus to compare reasonably, we decided to use the recommendation of the manufacturer and/or further metabolomics literature for the mobile phases used for primary selection (Contrepolis et al., 2015; Ribbenstedt et al., 2018; Si-Hung et al., 2017; Sonnenberg et al., 2019; Vass et al., 2016; Wernisch and Pennathur, 2016). Here we found that HILIC-column chemistry with zwitterionic surface groups (ZIC-HILIC) was able to analyze the broadest range of amino acids, as well as had the highest overall signal intensities. The reversed-phase column (C18) had low intensities with small and/or positive amino acids, such as lysine, valine, serine, and proline. The other HILIC phases globally performed worse and showed lower intensities with arginine and lysine. Wernisch et al. found in their analysis that in their 764 metabolite library, the best total coverage was achieved by an RPLC method. However, the method with good retention and third-highest total coverage was achieved by a zwitterionic sulfobetaine column comparable to our method (Wernisch and Pennathur, 2016). Together the RPLC column and sulfobetaine column outperformed all other tested column combinations. Similarly, in most tested criteria, the ZIC column outperformed the other column chemistries, such as amide modifications. Similarly, in a comparative study of 46 compounds by Contrepolis, the ZIC-HILIC method was found to be superior compared to other studied column chemistries (Contrepolis et al., 2015).

Combining ZIC-HILIC columns and RPLC is common (Contrepolis et al., 2015; Schwaiger et al., 2019), especially using QToF mass spectrometers such as ours (Naz et al., 2017). Given our results and backup by other groups, we decided to optimize the chromatographic conditions with the ZIC-HILIC column further (**Figure 9**).

We then further modified mobile phase parameters such as pH and buffer concentration and the type of gradient used for analysis. We then selected a buffer containing 30 mM ammonium acetate and a pH of 7.0. Also, a steeper linear gradient was selected instead of multiple different steps of an isocratic gradient. The use of mobile phases widely varies in the literature, with some groups utilizing no buffer salts (0.1 % formic acid in water or acetonitrile; (Naz et al., 2017), some others have described 10 mM ammonium acetate.

3.0 Development of the metabolomics platform for rare neurometabolic disorders



Figure 9 Exemplary picture of the LC-MS method selection process. A reference mixture of amino acids was used to determine MS signal abundance (numbers) and peak shapes (color). Green colors represent optimal peaks (symmetric and narrow), yellow suboptimal (skewed and/or broad), and red non-admissible peaks (extremely broad or absent). Exemplary selection of a set of methods that lead to the development of the current pipeline.

Lastly, we optimized mass spectrometry parameters to increase the overall signal intensities, as they should be adapted to the used mass spectrometry mobile phases. The complete optimized method can be found in the method section (chapter 2.2.9, p. 26)

3.4.2 The human metabolite reference library for the ZIC-HILIC-MS method contains over 400 compounds

High confidence identification of metabolites by LC-MS requires at least valid mass and retention time identifiers (Schrimpe-Rutledge et al., 2016). While mass/charge ratio can be calculated from the elemental sum formula, the retention time has to be determined empirically. This holds true especially for ZIC-HILIC chromatography, as there is not always a straightforward prediction from the hydrophilicity possible, such as in RPLC (Chirita et al., 2011). In untargeted metabolomics, it is possible to work with data sets without identifying the features contained in them and identifying them after the measurement. A set of most likely reference metabolites is acquired, retention time determined, and retrospectively identified metabolic features. However, the selection of features deemed relevant for identification can be cumbersome when there is no knowledge about their biological function and interaction with other features/metabolites.

Another drawback is that there is no possibility to efficiently study the limitations of the method, as, for instance, the typical limits for detection of metabolites are undetermined. Also, retrospective purchase of new reference standards results in waiting times, decreasing the overall speed of project progress.

Therefore, we decided to avoid those obstacles by creating a metabolite identification library containing over 500 human metabolite reference standards to provide better identification information for the ZIC-HILIC-MS method (chapter 3.2.1., p. 43). To develop the human metabolite library of 502 compounds (list in **Table 22**, p. 181), we first decided to split the LC-MS acquisition into 39 sets of metabolites (master mixes) to increase the speed of analysis.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

We detected 402 of the 502 substances in the reference mixes across positive and negative ionization modes. In positive mode, 174 metabolites were detected with stable fragmentation patterns, which could be used as identification parameters for later studies (**Figure 10**). 100 metabolites did not produce relevant fragments. 29 metabolites had a retention time lower than 2.5 minutes, a timeframe lower than to be chromatographically resolved.

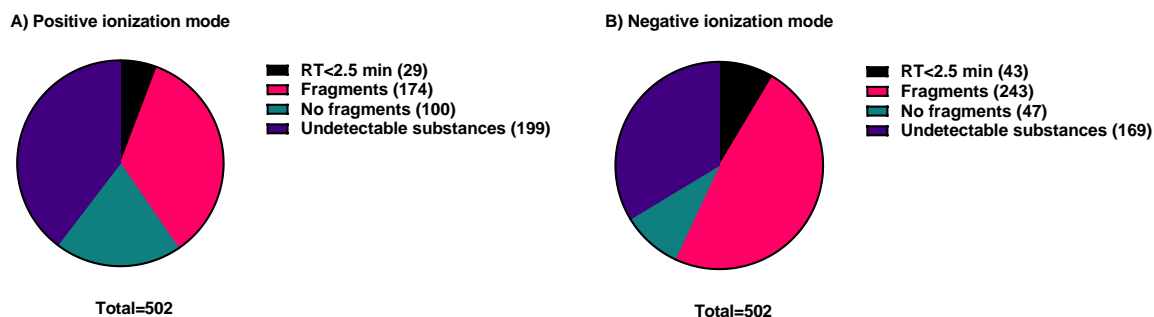


Figure 10 Coverage of the metabolite reference library: ZIC-HILIC-MS. A total of 502 human metabolite reference standards were acquired and divided into 39 measurement sets. Data of retention time, main adduct, and fragment mass were acquired using the ZIC-HILIC-MS method in positive A) and negative ionization mode B). Compounds were either not sufficiently detectable (“Undetectable substances”), delivered reproducible fragments (“Fragments”/“No Fragments”), or had a retention time equal to the flow time without chromatographic separation (dead volume, “RT < 2.5 min”). A complete overview of metabolite identification data can be found in **Table 19** and **Table 20**.

Partly overlapping with this, in negative mode, 243 metabolites were detected with fragments, 47 without fragments, and 43 compounds had a retention time lower than 2.5 minutes. Across both ionization modes, we can therefore detect around 80% from standards, which is very slightly higher than the around 70% in the ZIC-HILIC method discussed by Wernisch et al. (Wernisch and Pennathur, 2016).

We then decided to create a database of the retention time library using the commercial UNIFI software for better data integrity and more straightforward handling. This allows for the structured archiving of metabolite identifiers (such as retention time, mass, fragments) alongside the original chromatograms and fragment mass spectra of these compounds. This database allows for the identification of compounds even years after the initial creation of the database.

For better metabolomics data analysis using open-source software, we stored the main identification criteria in the form of a tabular format (excel) and a plain text text-file format (msp) that was created by the national institutes of science and technology (NIST). An example of what information is used for the identification of metabolites can be found in **Table 8**. The complete list of identification criteria can be found in the annex **Table 19** and **Table 20** for positive and negative ionization mode separately. The high number of identification information of metabolites are an important result, as they are a main necessity to detect rare disease using LC-MS.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Table 8 Exemplary data set from tryptophan in the metabolite identification library. Authentic metabolite reference standards were analyzed using the ZIC-HILC-MS method in positive ionization mode. Data was imported and assessed using UNIFI and retention time and correlating predicted fragments entered the database. For further analysis, important identification information was exported to excel. An exemplary data entry from Tryptophan in positive ionization mode is shown; the full table can be found in the annex.

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Tryptophan	Indolyl carboxylic acids and derivatives	203.082	C ₁₁ H ₁₂ N ₂ O ₂	QIVBCDJIAJPQS-VIFPVBQESA-N	7.55	159.09255,130.06607,74.02535

In order to avoid a high number of misidentifications, we only used the mass spectrometry adducts “-H” and “+H” as precursor types for the database. The metabolite ontology was created using ClassyFire “Subclass” after batch import of InChIkey (Djombou Feunang et al., 2016).

For identifying metabolites that were not part of our library, we combined our library with the publicly available FiehnHilic library (<https://mona.fiehnlab.ucdavis.edu>, last accessed on 10.06.2021).

3.4.3 The RPLC-MS/MS platform for lipidomics analysis

Based on previously mentioned literature (Contrepois et al., 2015; Schwaiger et al., 2019; Wernisch and Pennathur, 2016) and the interest in neuropediatric research, we found that using a method for lipidomics would complement our ZIC-HILIC-MS method for hydrophilic analytes.

The lipidomics reversed phased LC-MS/MS method was modified from an already established method, commonly used in the lipidomics community (Narváez-Rivas and Zhang, 2016; Rampler et al., 2018). However, the run time with 60 mins per sample was long and increased runtime related artifacts. In order to increase sample throughput, a UPLC column with solid-core particles was used due to their higher peak separation power with higher flow rates rate (higher peak efficiency; Kirkland et al., 2013). Subsequently, by increasing the flow rate and compressing the gradient we were able to shorten sample analysis to 30 minutes (the full gradient can be found in chapter 2.2.10, p. 27). A higher flow rate mostly leads to a faster elution of compounds. Based on the used chromatographic column, we refer to the method as the CortecsT3 lipidomics method.

3.4.3.1 Structured lipids allow for higher identification certainty even without retention time information

Lipids can be structured into a lipid-class defining head group and varying fatty acid acyl chains that define the specific lipid (**Figure 11**; Gyamfi et al., 2019). Further information to lipids and their nomenclature can be found in chapter 10.2, p. 156 in the annex.

Exemplary lipid (here: a phosphatidylcholine; PC 16:0_16:0, PC aa 32:0)

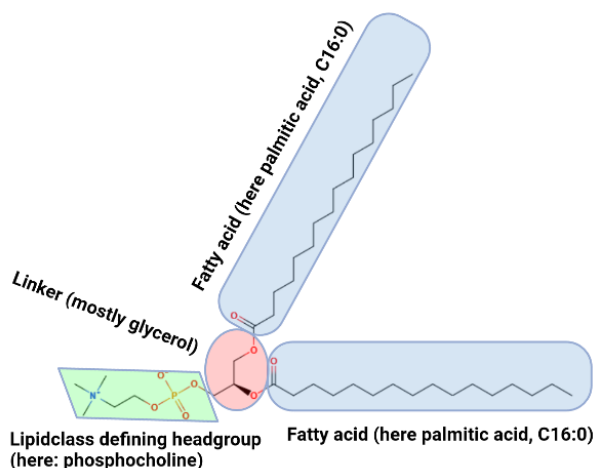


Figure 11 Scheme of a typical, structured lipid. Most lipids can be structured into lipid-class defining headgroup two variable fatty acids using a linking molecule. Created with BioRender.com

In RPLC-MS, there is a relationship between the polarity of the analyte (increasing retention time with decreasing polarity; (Knittelfelder et al., 2014). The polarity of a lipid is greatly determined by the polarity of the head group and, to a lesser extent, by the length of the attached fatty acid (the more carbon-atoms, the less polar). Thus, the retention time of lipids may be extrapolated by these criteria (Aicheler et al., 2015; Tsugawa et al., 2017). Here, we aimed to examine whether lipids of the same class indeed appear in grouped retention times and if they match the patterns from authentic lipid reference standards.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

We used EDTA plasma from a healthy human volunteer as a typical sample and acquired LC-MS data using the RPLC CortecsT3 lipidomics method. A typical graph of the total sum of all ions (from all mass ranges) dependent on time (total ion chromatogram, TIC) shows a high concentration of signals in discrete time points (**Figure 12**). A higher density of LC-MS signals can be found primarily in the beginning, middle, and end of the part chromatogram.

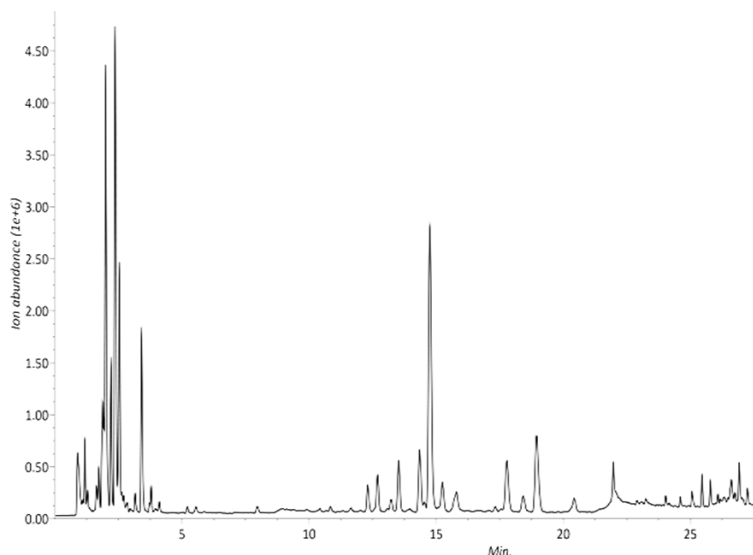


Figure 12 A typical total ion chromatogram of EDTA plasma measured with the lipidomics method shows different lipid signals. Healthy human EDTA plasma was prepared and analyzed using the CortecsT3 lipidomics method. Graph of total ion abundance (TIC, sum over all ions from different mass ranges) dependent on time is presented.

After identification using the adduct mass and fragments, a similar pattern of high signal concentrations may be found (**Figure 13**).

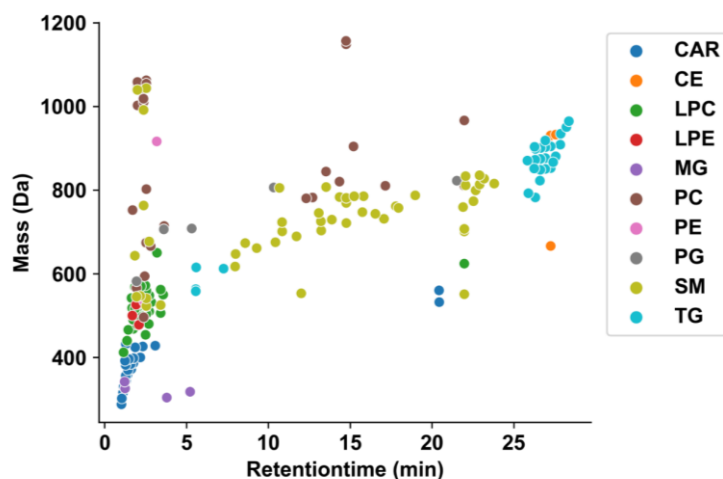


Figure 13 A plot of identified lipids and lipid classes from EDTA-plasma over different retention and mass ranges further shows a structure in lipid retention times. Healthy human EDTA-plasma was prepared and analyzed using the CortecsT3 lipidomics method. MS-features were filtered based on the presence of fragments and identified using adduct mass and fragment mass by the MS DIAL lipidomics library. MS-features were then graphed based on their respective identified mass and retention time; the color represents different lipid classes: CAR: acyl-carnitine; CE: cholesteryl-ester; LPC: lysophosphatidylcholine; LPE: lysophosphatidylethanolamine; MG: monoacylglycerol; PC: phosphatidylcholine; PE: phosphatidylethanolamine; PG: phosphatidylglycerol, SM: sphingomyelin; TG triacylglycerol;

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Using the presumptive identification of the lipid classes (note the color code), we recognized that especially polar lipid classes such as carnitines or lysolipids appear in concentrated regions at the beginning. Less polar lipids (e.g., triglycerides) appeared at the end of the chromatogram. Some lipid classes, such as sphingomyelins, can be found more distributed over the middle part of the chromatogram. On average, lipid classes, such as lysophosphatidylcholines, produced peaks of higher signal intensity, potentially because of their higher abundance in plasma or better MS ionization efficiency.

To confirm these observations and generalize them to other chromatograms, we then acquired retention time data of selected lipids with reference standards (**Figure 14**).

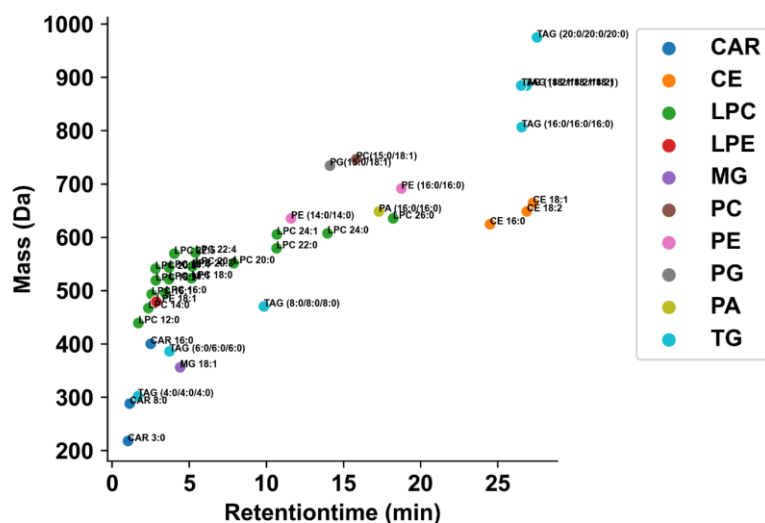


Figure 14 Reference lipids of different lipid classes similarly show patterning. Data were acquired from 38 lipid reference standards of 10 different lipid classes using the CortecsT3 lipidomics method. MS-features were filtered based on the presence of fragments then identified using adduct mass and fragments using the MS DIAL lipidomics library. MS-features were then graphed based on their respective mass and retention time; the color represents different lipid classes: CAR: acyl-carnitine; CE: cholesteryl-ester; LPC: lysophosphatidylcholine; LPE: lysophosphatidylethanolamine; MG: monoacylglycerol; PC: phosphatidylcholine; PE: phosphatidylethanolamine; PG: phosphatidylglycerol, SM: sphingomyelin; TG triacylglycerol;

Similarly, also in the authentic lipid reference standards we observed a clustering of different lipids according to their lipid class. Here is the dependence on the head group polarity, an increase of retention time with an increase of fatty acid chain length, and a decrease of retention with increasing unsaturation. These results match well with the commonly published results as for instance of Ovčačiková et al. (Ovčačiková et al., 2016).

3.4.4. Development of a sample preparation method for the untargeted metabolomics platform

Human biofluids (tissues and cells) are rarely directly admissible to LC-MS. Thus, sample preparation must occur. The main goals for sample preparation in LC-MS-based metabolomics are removing unwanted contaminants (e.g., proteins) while enriching the metabolites of interest and delivering the sample in a way that is admissible for the type of analysis (e.g., correct solvent). Essential criteria for assessing the preparation are the broadness of visible metabolites (variety of different metabolites classes, number of metabolites), intensity or signal-to-noise ratio, and reproducibility of the analysis (standard deviation between the same samples, prepared multiple times).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

As we selected two LC-MS methods to optimize the analysis of hydrophilic and lipophilic analytes separately, the sample preparation method should also extract these groups with maximum orthogonality. Because samples are often limited, ideally, both extracts should be prepared from one sample.

3.4.4.1 RPLC-SPE provides reliable lipid extraction than the Bligh and Dyer method

Generally, two methods are primarily used for the extraction of lipophilic metabolites: liquid-liquid and solid-phase extraction (Dettmer et al., 2007). While the first method also removes proteins by denaturation and accumulation in the interphase between the chloroform and water-phase, solid-phase extraction requires a previous removal of proteins. However, the solid-phase extraction is available in higher throughput, 96-well-plate formats.

Thus, we aimed to test whether the liquid-liquid extraction (modified Bligh and Dyer method; (Bligh and Dyer, 1959; Ulmer et al., 2018) or the solid-phase extraction (RPLC-SPE) is more admissible to our metabolomics pipeline. We analyzed the mean intensity and mean covariance of different lipid classes to overview extraction efficiency, as chemical properties (e.g., polarity) are most different between different lipid classes.

In most lipid classes, the reversed-phase SPE method (RPLC-SPE) showed superior mean group intensities of lipid classes than the liquid-liquid extraction method (**Figure 15**). This especially applies to rather polar lipid classes (e.g., sphingomyelins, phosphatidylethanolamines). On the other side, lipid classes that are very unpolar (such as triacylglycerol) show higher intensities with the Bligh and Dyer method. Thus, both methods appear to target different polarity ranges of metabolites.

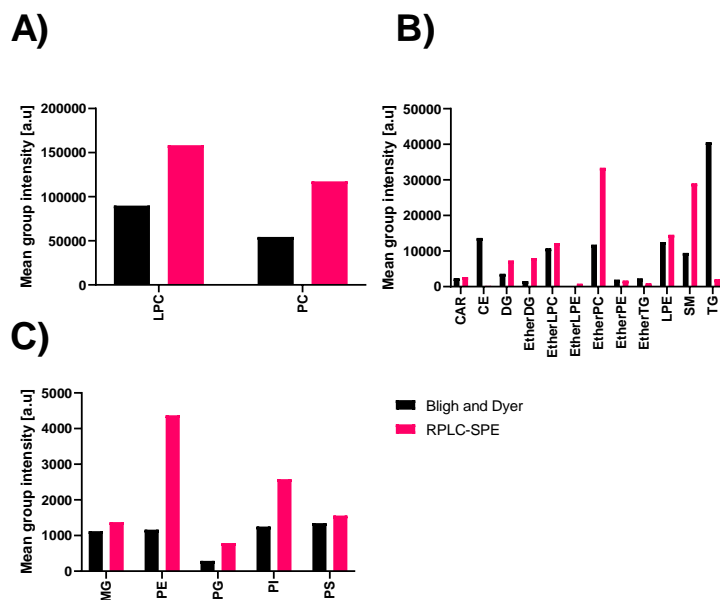


Figure 15 The mean intensity of different lipid classes is greater in the RPLC-SPE. Human EDTA-plasma was prepared using liquid-liquid extraction (modified “Bligh and Dyer” method) and the C18-reversed-phase solid phase (RPLC-SPE) extraction methods. Data were acquired using the Cortecst3 lipidomics method, and lipids were identified by mass and fragment masses. Mean intensity over different lipids from the specific lipid classes was calculated. n=5 extraction replicates with the same original sample were used. Subgraphs use different exes due to different intensities: A) LPC: lysophosphatidylcholine; PC: phosphatidylcholine; B) CAR: acyl-carnitine; CE: cholesteryl-ester; LPE: lysophosphatidylethanolamine; SM: sphingomyelin; TG triacylglycerol; C) MG: monoacylglycerol; PE: phosphatidylethanolamine; PG: phosphatidylglycerol.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Another important point to consider is the repeatability of the analysis. Sample preparation inaccuracies can increase standard deviation and obscure biologically significant effects. Similarly, we tested the mean variance of different lipid classes in multiple preparations of the same sample.

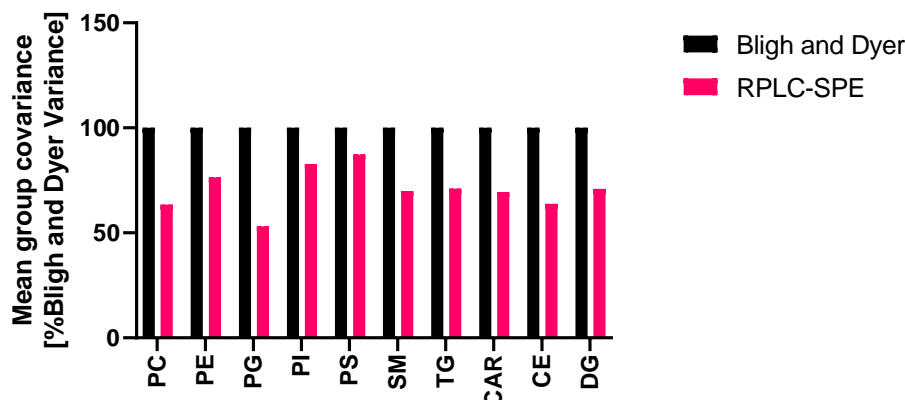


Figure 16 The mean group variance of different lipid classes is lower in the RPLC-SPE. Human EDTA-plasma was prepared with liquid-liquid extraction (modified “Bligh and Dyer” method) and the C18-reversed-phase solid phase extraction (RPLC-SPE) methods multiple times from the same sample. Data were acquired using the CortecsT3 lipidomics method, and lipids were identified by mass and fragment masses. Mean intensity and corresponding variance (in % mean) over different lipids from the specific lipid class were calculated. $n=5$ extraction replicates with the same original sample were used. PC: phosphatidylcholine; PE: phosphatidylethanolamine; PG: phosphatidylglycerol; PI: phosphatidylinositol; PS: phosphatidylserine; SM: sphingomyelin; TG: triacylglycerol; CAR: acyl-carnitine; CE: cholesteryl-ester; DG: diacylglycerol.

The manual Bligh and Dyer method has a higher mean group covariance in all tested lipid classes than the 96-well-plate scheme RPLC-SPE method (**Figure 16**). As a high mean group variance points to higher inaccuracies in sample preparation, the RPLC-SPE method appears to be superior.

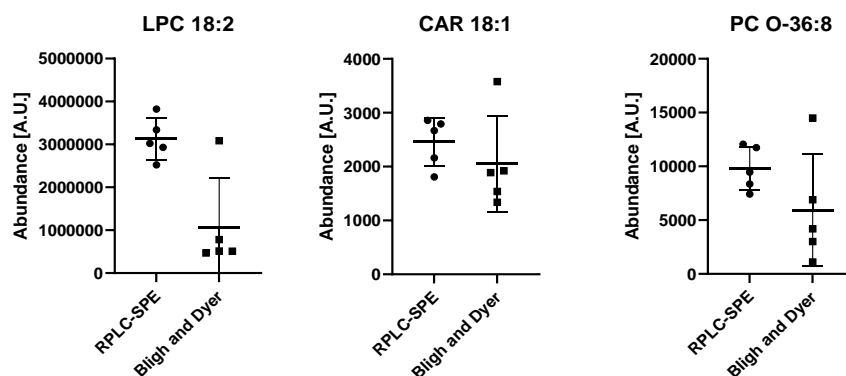


Figure 17 The RPLC-SPE has superior intensities and smaller deviations than the Bligh and Dyer method: Exemplary lipids. Human EDTA-plasma was prepared with liquid-liquid extraction (modified “Bligh and Dyer” method) and the C18-reversed-phase solid phase extraction (RPLC-SPE) methods multiple times from the same sample. Data were acquired using the CortecsT3 lipidomics method, and lipids were identified by mass and fragment masses. $n=5$ extraction replicates with the same original sample were used. PC: phosphatidylcholine; CAR: acyl-carnitine; LPC: lysophosphatidylcholine. Mean \pm SD.

However, the mean group variance should not be compared to variances in typical analytical methods. The ranges here are higher than in analytical method specifications (>15-20% for FDA validated methods). As the mean group average presents an average over all identified MS signals, lipids with a low signal-to-noise ratio in the respective preparation are included in the analysis. Thus, noisy signals can lead to higher variance, as they are nonetheless included in the mean group variance. Therefore, this marker should be used with caution only to analyze sample preparation methods and does not substitute for by-compound analysis. In **Figure 17** we extracted the raw abundances of three exemplary lipids, where the higher intensities and lower standard deviation of the RPLC-SPE is well noticeable.

Unfortunately, most research groups did not assess mean group intensities to compare between methods, but rather the number of uniquely identified lipid species, the intensity of exemplary lipids, or analyze recovery from artificial material (Matyash et al., 2008; Reis et al., 2013; Yang et al., 2013b). Interestingly we were not able to find further use of an RPLC-SPE method for a general extraction of lipids, instead, most SPE methods focused on specific lipid classes (Aldana et al., 2020; Züllig et al., 2020). However, a variant of SPE, solid-phase microextraction, SPME, was superior to the Bligh and Dyer method in terms of analytical variability and lipid diversity (Birjandi et al., 2017). This overall matched our results from the SPE method.

Based on lower variability in sample extraction, the high-throughput ability, the ability to create extracts for hydrophilic and hydrophobic analytes, and the higher lipid diversity, we decided to proceed with the RPLC method.

3.4.4.2 The RPLC-SPE can also be used for sample preparation of hydrophilic metabolites

In previous experiments, we identified the RPLC-SPE method to be superior for the reproducible analysis of lipids. The main goal of the sample preparation pipeline is to be able to extract lipophilic metabolites, as well as hydrophilic metabolites from the same sample. Consequently, we aimed to determine if the RPLC-SPE can also be used as sample preparation for hydrophilic metabolites. The alternative method for analyzing hydrophilic samples would be the specific extraction of hydrophilic metabolites, for example, by using a ZIC-HILIC-SPE column. This method has been developed before by the author and is mentioned elsewhere (Jensen et al., 2021). While this method may increase the extraction efficiency, it would not be directly compatible with the RPLC-SPE and is not readily (commercially) available in a 96-well plate format. Thus, we mainly aimed to determine if the ZIC-HILIC-SPE is necessary or the RPLC-SPE alone delivers acceptable intensities for hydrophilic metabolites.

As the metabolite classes of hydrophilic analytes include a wider variety of different classes than lipids, we decided against the use of the mean group intensity or covariance as a marker to evaluate the methods as we did in the previous section. Instead, we decided to group metabolites based on their respective signal-to-noise ratios (S/N).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Using the RPLC-SPE method and analyzing hydrophilic metabolites using the ZIC-HILIC-MS method in positive mode, we observed that the distribution of S/N-ratios was very similar (**Figure 18**).

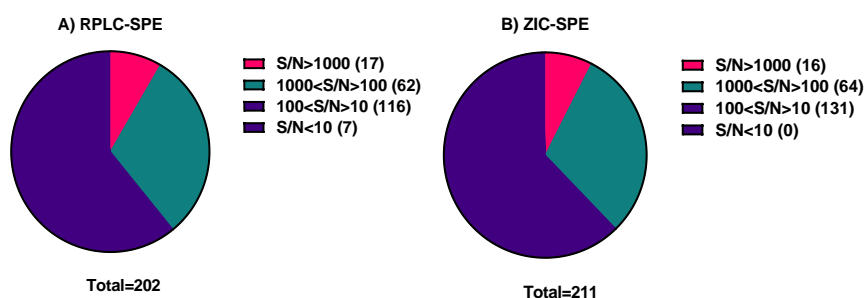


Figure 18 Signal-to-noise ratios of two SPE methods: ZIC-HILIC MS in positive ionization mode. Human EDTA-plasma was prepared using the A) RPLC-SPE or B) ZIC-SPE method and analyzed using the ZIC-HILIC-MS method in positive ionization mode. Metabolites were identified using MS-DIAL, and only metabolites that were identified by at least mass were used for the analysis. Signal to noise-ratios (S/N) were determined using MS-Dial in n=5 extraction replicates with the same original sample.

In negative ionization mode, the ZIC-SPE method appears to enrich more metabolites. However, slightly more metabolites have an S/N-ratio > 1000 in the RPLC-SPE method (**Figure 19**). Yet, this seems to be only a minor benefit, as the ZIC-SPE method has higher metabolite numbers in the following categories.

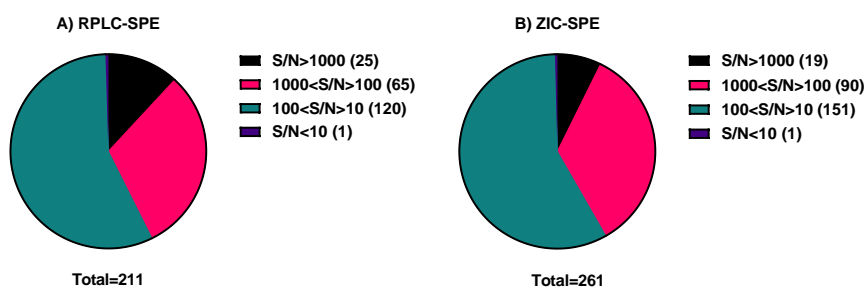


Figure 19 Signal-to-noise ratios of two SPE methods: ZIC-HILIC MS in negative ionization mode.

Human EDTA-plasma was prepared using the A) RPLC-SPE or B) ZIC-SPE method and analyzed using the ZIC-HILIC-MS method in positive ionization mode. Metabolites were identified using MS-DIAL, and only metabolites that were identified by at least mass were used for the analysis. Signal to noise-ratios (S/N) were determined using MS-Dial in n=5 extraction replicates with the same original sample.

While the ZIC-SPE method has increased signal-to-noise ratios for the hydrophilic metabolites, the RPLC-SPE method appears to have higher metabolite intensities and higher reproducibility for the lipophilic metabolites.

In **Figure 20** we extracted the raw signal abundance of three representative metabolites (tryptophan and phenylalanine in negative ionization mode; carnitine in positive mode). Here we can see that the RPLC-SPE extracted metabolites do not always have the highest signal intensities, but typically a lower variation.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

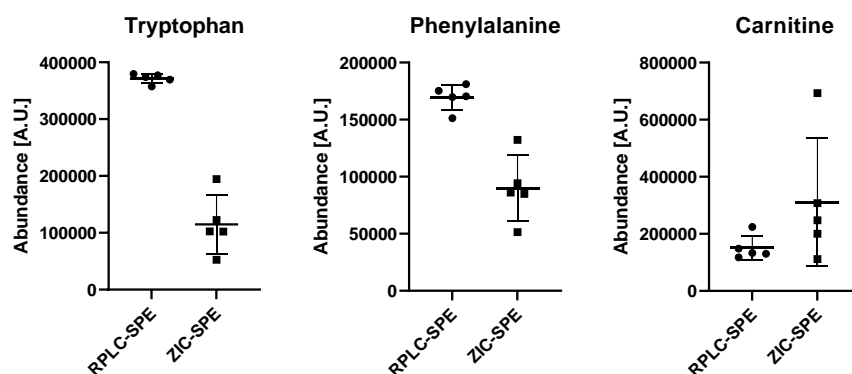


Figure 20 The RPLC-SPE method has lower deviations, but not higher mean intensities than the ZIC-SPE method in selected hydrophilic metabolites. Human EDTA-plasma was prepared using the RPLC-SPE or -SPE method and analyzed using the ZIC-HILIC-MS method in positive (Carnitine) or negative (Tryptophan and Phenylalanine) ionization mode. Metabolites were identified using MS-DIAL, and only metabolites that were identified by at least mass were used for the analysis. Raw signal abundance of $n=5$ extraction replicates with the same original sample are graphed with mean \pm SD.

Counterintuitively so far, we have not encountered use of RPLC-SPE in the way we proposed. Most groups in metabolomics analysis focus on either simple protein precipitation or liquid-liquid extraction techniques to yield hydrophilic and lipophilic extracts e.g., the group of Masson et al (Masson et al., 2010). Common for the analysis of hydrophilic metabolites, is also the removal of phospholipids by specialized Zirconia-SPEs (Tsakelidou et al., 2017). HILIC-SPE, in the context of metabolomics, was primarily described for phospholipid removal (Bylda et al., 2014; Van Damme et al., 2014). The combination of liquid-liquid extraction and the HILIC-SPE process is also not readily available in 96-well-plate formats. This limits sample throughput (only around 15 samples can be prepared at the same time) and is not easily automatable.

As the ZIC-SPE was not necessary to increase extraction efficiency, we selected the RPLC-SPE alone for further analyses. We decided to use the RPLC-SPE for the separation of hydrophilic and lipophilic metabolites.

3.4.5 Stability of the untargeted metabolomics platform and analytical validation

In the first part, we developed an untargeted metabolomics platform comprising a ZIC-HILIC-MS method in positive and negative ionization mode for hydrophilic analytes and a CortecsT3 RPLC-MS method for lipidomics. Naturally, we were interested in the analytical performance and especially the limitations of this method, as most untargeted metabolomics platforms do not comprehensively assess them (Masson et al., 2010; Naz et al., 2017; Ribbenstedt et al., 2018). This publication mainly focused only on a subset of known metabolites or discussed only reproducibility or the number of detected metabolites. Some others also include retention time stability or peak area variations of known metabolites (Bonte et al., 2019). However, they still lacked information about the influences of errors in the sample preparation, commonly required to be assessed in a regulated environment (FDA, EMA, ICH).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

The first part of this chapter (chapter 3.4.1 p. 49 to chapter 3.4.6 p. 59) focused mainly on detecting reference standards from solvents. To study the analytical validation, we now primarily concentrate on measurements from biomaterials and, as opposed to other studies, try to evaluate validation parameters for a broader range of compounds.

Different parameters are used to study the validity of analytical methods. According to regulatory bodies such as the ICH, FDA or EMA essential parameters include the specificity/selectivity of the method, the sensitivity of a method, and the linearity of a method. Due to the endogenous nature of the metabolites and the large number of different metabolites to be studied, classical methods to study validation parameters cannot easily be applied to untargeted metabolomics. To gain some insights into these parameters, we assessed proxy-markers that can give at least a particular estimation. For the specificity and selectivity in LC-MS analytical methods, retention time and mass/charge ratios have utmost significance, as they are the primary identifiers of specific metabolites. The sensitivity in metabolomics methods may be studied by the detectability of metabolites in different biomaterials or dilutions of biomaterials with solvents. The latter may also be used to study an approximation of the linear range. Due to the absence of accurate reference standards, we can only correctly assess the specificity of metabolites in the ZIC-HILIC-MS library, as standards for the lipid method are rarer.

3.4.5.1 Influences on specificity: Mass error and retention time stability

The main parameters to monitor specificity in LC-MS/MS methods are the mass deviation to a reference mass (mass error) and the retention time deviation to a reference retention time (retention time error). As peak detection and thus retention and mass errors may be different depending on different signal intensities, we analyzed the errors over different sample concentrations.

We first analyzed the distribution of mass errors and the stability of masses inside the run as the molecular ion mass is the key parameter for identification, especially of metabolites without retention time data.

Most identified metabolites had a mean mass difference to the reference library of less than 5 mDa, comprising 65 % of the metabolites, and 95 % of metabolites have a mass error of less than 38 mDa (**Figure 21**). The distribution is left-skewed, with an increased proportion of lower mass deviations, as expected.

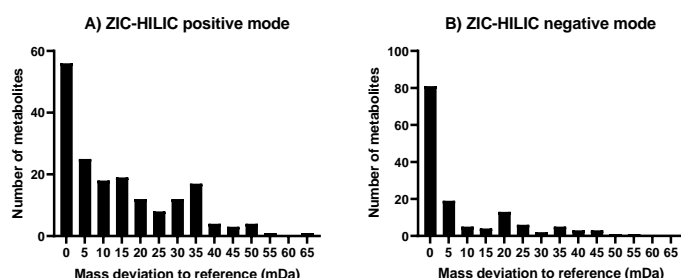


Figure 21 Stability of mass deviation to reference standards: ZIC-HILIC-MS. 10, 50, 50,100, 150, 300, 400, and 500 μ l human EDTA-plasma were prepared and measured using the ZIC-HILIC-MS method in positive A) and negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and mean mass deviation for each metabolite was calculated for $n=21$ measurements of 7 different concentrations in triplicates. A histogram with a bin width of 5 mDa was created.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

When examining the intra-run mass error, most of the error lies below 2 mDa, comprising 89% of the metabolites (**Figure 22**). 95% of the metabolites had a mass error of less than 6.8 mDa. Similarly, the distribution is also left-skewed.

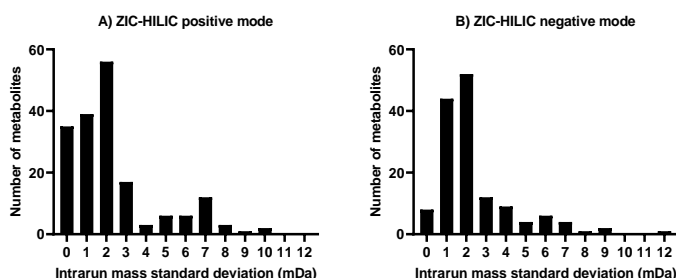


Figure 22 Stability of intrarun mass deviation: ZIC-HILIC-MS. 10, 50, 50,100, 150, 300, 400, and 500 μ l human EDTA-plasma were prepared and measured using the ZIC-HILIC-MS method in positive A) and negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and intrarun mass standard deviation for each metabolite were calculated for $n=21$ measurements of 7 different concentrations in triplicates. A histogram with a bin width of 1 mDa was created.

Based on calibration-checkup data of our mass spectrometer before every measurement, average mass errors to the calibration mix in a range of 2-5 mDa were reported, which is typical for a QToF mass spectrometer (Bristow and Webb, 2003; Köfeler and Gross, 2005).

The other important parameter for identification is the retention time; hence, a highly reproducible retention time is desirable. Therefore, we determined the mean retention time error to our reference library and inside a run (**Figure 23**).

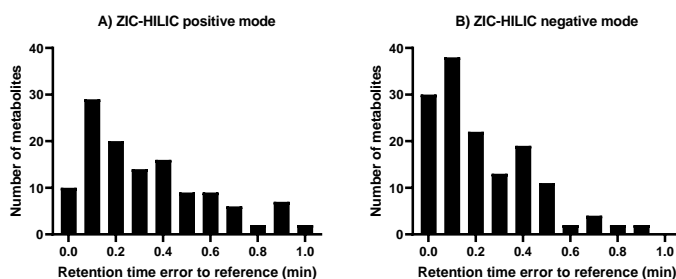


Figure 23 Stability of retention error to reference standards: ZIC-HILIC-MS. 10, 50, 50,100, 150, 300, 400, and 500 μ l human EDTA-plasma were prepared and measured using the ZIC-HILIC-MS method in positive A) and negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial and retention time deviation to reference standard for each metabolite was calculated for $n=21$ measurements of 7 different concentrations in triplicates. A histogram with a bin width of 0.1 min was created.

The retention time error to the reference library appears with a broader shape, without a clear maximum. 50% of metabolites occur with a retention time error of less than 0.16 min. 95% of metabolites have an error to the reference of fewer than 0.65 mins. Typically, most LC-MS methods have a retention time error of fewer than 0.2 mins in targeted methods. However, while our methods exceed this, this is not unexpected for an untargeted method that is not specific for an analyte. Additionally, some analytes may present with broader peaks, making the accurate determination of their maximum and, therefore, their retention time harder. In their metabolomics method, Bonte et al. have found a retention time CV of around 5% for most of their selected analytes, leading to a mean RT difference of around 0.25 min (Bonte et al., 2019).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

However, this was only estimated for a small, targeted set of analytes, not for an extensive range of different metabolites. Our ZIC-HILIC-MS method is also conceived as a screening method, where results frequently will be confirmed using other analytical techniques. Thus, intra-run stability is of greater importance than retention time errors to standards.

Most of the intrarun retention time errors appear below 0.04 min, comprising 52% of the metabolites. 95% of the metabolites have an intra-run retention time error of less than 0.1 mins (**Figure 24**). The intra-run retention time error is mainly influenced by shifts in the chromatographic analysis due to small variations on the column surface. Adsorption of matrix components such as lipids but even metabolites of interest can change the column selectivity (Dolan and Snyder, 2017). Also, a method can never be completely optimized for all analytes and this method was especially creating for screening of a wider range of compounds and not accurate detection.

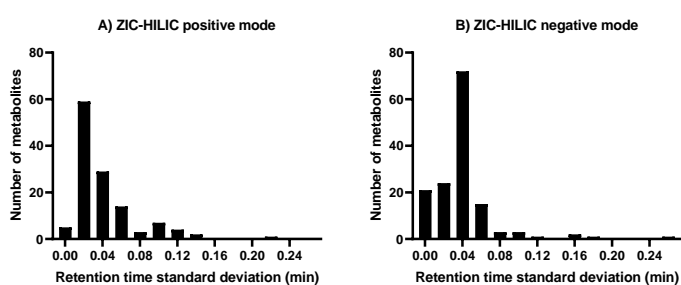


Figure 24 Stability of intra-run retention time error: ZIC-HILIC-MS. 10, 50, 50,100, 150, 300, 400, and 500 μ l human EDTA-plasma were prepared and measured using the ZIC-HILIC-MS method in positive A) and negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and intrarun retention time standard deviation for each metabolite was calculated for n= 21 measurements of 7 different concentrations in triplicates. A histogram with a bin width of 0.02 min was created.

The low intra-run retention time deviations point to a well reproducible method, at least inside the same run. This allows for longer run times required by large batches of a large number of samples. To summarize, while specific metabolites show a higher than typical retention time error, the intra-run retention time cv is very low, pointing out the high reproducibility of the method.

3.4.5.2 Influences on specificity: retention time distribution

For the identification of the metabolite signals apart from the retention time stability, the separation between different chromatographic peaks is important, as it decreases the chance of confusing different metabolites. A better chromatographic separation can also increase sensitivity by lower ion suppression effects.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Thus, ideally, the total amount of metabolite signals should be distributed equally over the retention time range.

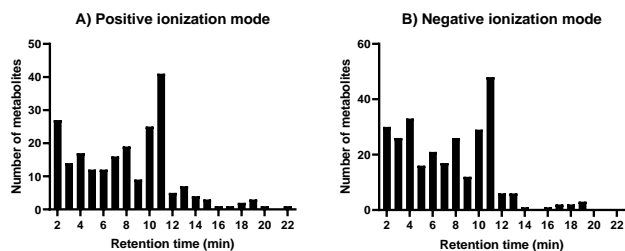


Figure 25 Retention time distribution of metabolites in ZIC-HILIC MS. Retention times of metabolites in the retention time reference library were analyzed using the ZIC-HILIC-MS metabolomics method in positive A) and negative B) ionization mode. A histogram was calculated with a bin of 1 minute.

In the ZIC-HILIC-MS method as well in the positive as in the negative mode, metabolite retention time appears to be distributed between mainly the first 10 minutes (**Figure 25**). After the first 10 minutes, only a minority of metabolites are eluting. In both modes, no single apparent maximum can be determined. An even more distributed metabolite profile would be desirable, but it is still very acceptable.

3.4.5.3 Sensitivity: Coverage of the metabolome in the healthy state

Classically, the sensitivity of an analytical method would be determined by adding a reference standard to the matrix in decreasing concentrations and assessment when a defined minimum signal to noise ratio is fallen below. As in metabolomics, the target metabolites are endogenous, an accurate limit of detection cannot be determined easily.

Instead, we first studied which metabolites can be determined from different matrices (EDTA-plasma, fibroblast cell lysate) in a healthy/control situation. This is a common marker used in metabolomics method development (Zhang et al., 2020). Using the sample preparation pipeline, we can detect around 192 metabolites from our reference library across positive and negative ionization mode in healthy human plasma and wild-type human fibroblast cell lysate (**Figure 26 A**).

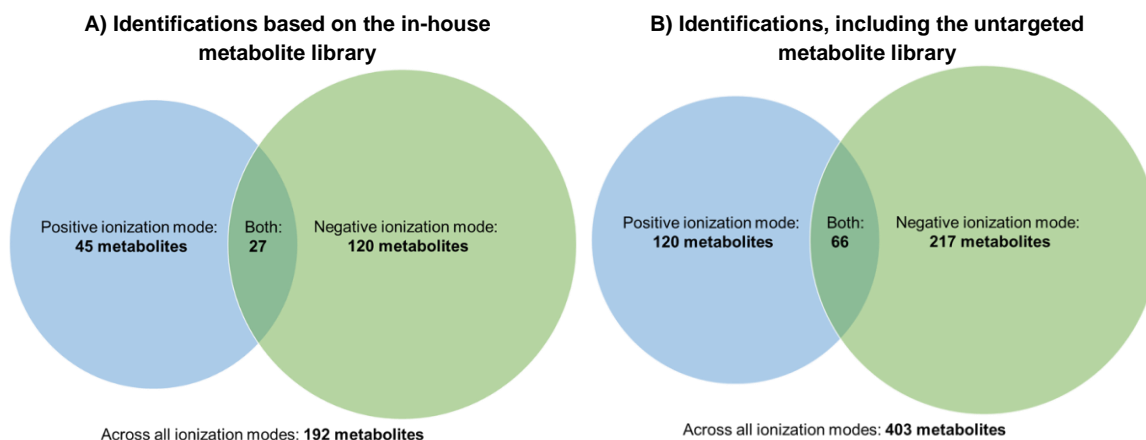


Figure 26 The ZIC-HILIC-MS method shows a high number of detectable metabolites. Healthy human EDTA-plasma and fibroblast cell culture lysate of 1×10^6 cells were measured using the ZIC-HILIC-MS method in positive and negative ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and only metabolites with matching mass and retention time are presented (A) or metabolites that were also identified in an untargeted manner using mass/charge-ratios were analyzed (B).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Both ionization modes show considerable orthogonality, with most metabolites that can only be analyzed in one ionization mode separately. If including metabolites that are identified presumptively using mass adduct and fragmentation information, up to 403 metabolites can be identified (**Figure 26 B**).

While some metabolites can be seen in both ionization modes as well as in EDTA-plasma and cell lysates, others can only be seen in one matrix or ionization mode (**Figure 27**).

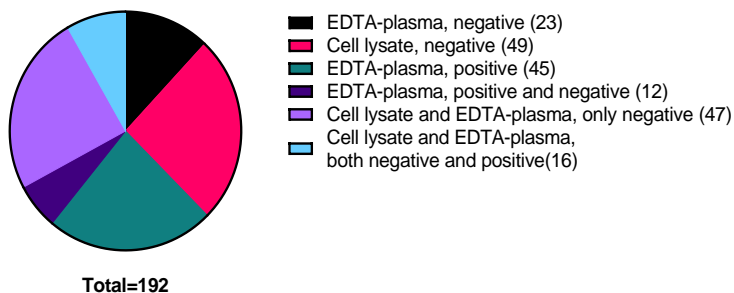


Figure 27 Different metabolites are detectable in different matrices. Healthy human EDTA-plasma and fibroblast cell culture lysate of 1×10^6 cells were measured using the ZIC-HILIC-MS method in positive and negative ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and only metabolites with matching mass and retention time are presented.

Using the 192 metabolites from our metabolite reference library, the metabolome can thus be covered in critical key pathways (**Figure 28**), pointing at the usability of our method.

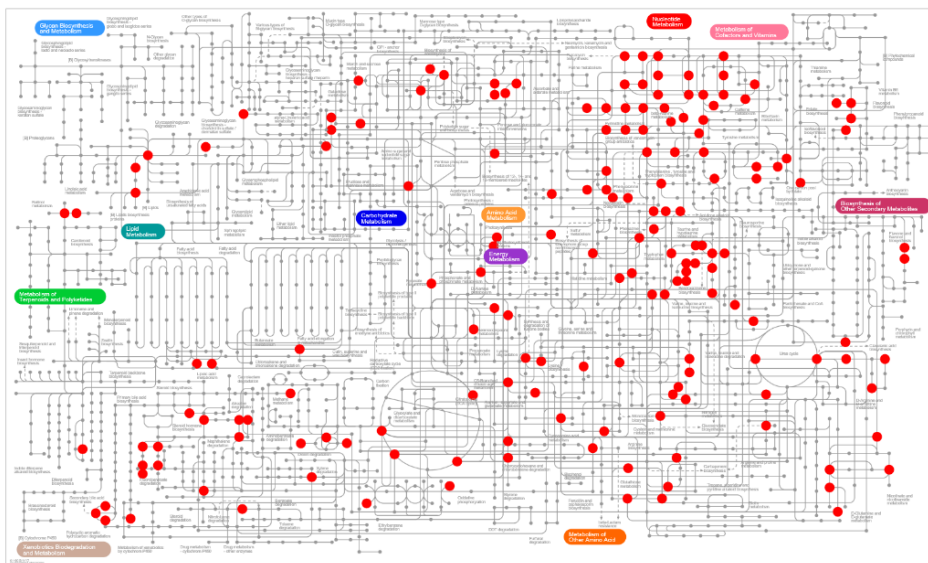


Figure 28 Using the ZIC-HILIC metabolomics pipeline, the metabolome can be covered in important checkpoints. Healthy human EDTA-plasma and wildtype fibroblast cell lysate was analyzed using the ZIC-HILIC-MS pipeline and identified using the metabolite reference library. 192 metabolites with matching retention time and masses were then plotted as red dots on the metabolome using IPath3 to illustrate metabolite coverage.

This included a number of metabolites from amino acid metabolism (esp. tryptophan metabolism) and nucleotide metabolism (esp. purine metabolism), but also a great number of other pathways.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Unfortunately, while our reference library contains a great number of metabolites from the glycolysis and citrate cycle, these were not detected. This may be due to the fast reaction times in glycolysis, the similar structure of hexose, and the high charge of tricarboxylic acids, hindering successful chromatography. Other essential metabolites such as biogenic amines can be detected well.

Our method performed comparatively well with the method of Bonte et al. AI that detected 178 annotated metabolites (Bonte et al., 2019), the method of Naz et al. AI detected around 112 metabolites (Naz et al., 2017), the method of Coene et al. AI identified around 222 metabolites in patient plasma (Coene et al., 2018). In all described methods, metabolite retention times were confirmed using authentic reference standards. Consequently, compared to other analytical methods in the field, our method performs well and provides a broad insight into different metabolic pathways.

3.4.5.4 Sensitivity: Most metabolites can be detected in only 100 µl of human EDTA-plasma

As metabolite sensitivity cannot be determined using standard addition, we used a dilution of healthy human EDTA-plasma with PBS to establish roughly how many-fold a metabolite can drop from its initial concentration before it is not detectable anymore. Apart from the method validation aspect, this is also helpful in case material is limited and only a smaller volume can be analyzed. By this analysis, we can extrapolate which analytes might be detectable. For this purpose, we defined a metabolite as observable when the signal abundance is at least three times higher than in an analytical blank (only PBS).

In the positive mode of the ZIC-HILIC-MS method, most of the metabolites can already be detected in 300 µl of EDTA-plasma, comprising around 50.0 % of metabolites (**Figure 29 A**). 21 metabolites were already detectable in 10 µl plasma. 55 metabolites showed very high abundances in the blank and should be analyzed only with caution. In negative mode, most of the metabolites can be detected in a minimum of 100 µl of EDTA-plasma, comprising of around 62.5 % of metabolites (**Figure 29 B**), 31 metabolites were already detectable in 10 µl plasma. 19 metabolites showed very high abundances in the blank.

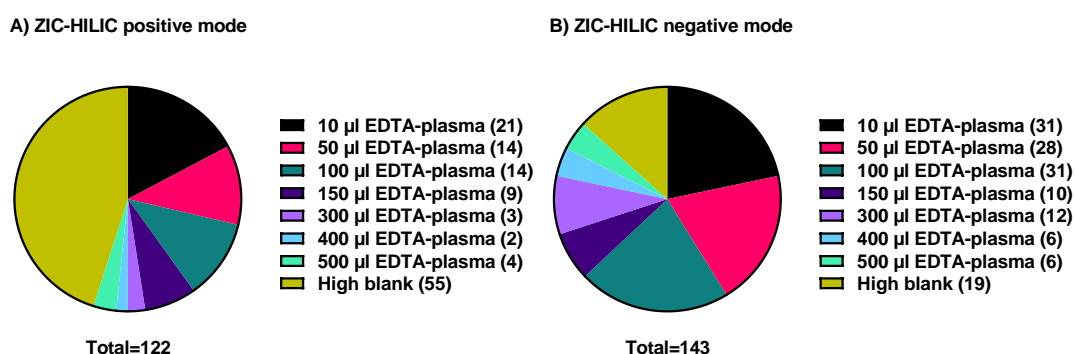


Figure 29 Detectability of metabolites in different plasma concentrations in ZIC-HILIC-MS. 10, 50, 50, 100, 150, 300, 400, and 500 µl human EDTA-plasma were prepared and measured using the ZIC-HILIC-MS method in positive A) and negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and mean signal abundance for each metabolite was calculated for n= 21 measurements of 7 different concentrations in triplicates. The limit of detection was defined as thrice the signal abundance as the PBS-blank.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

As mass spectrometry methods tend to be very sensitive, environmental contamination can increase analytical noise, as seen in the blank of this analysis. This may be originating from solvent production, used materials (e.g., vials), or even as reaction products on the column or the mass spectrometer ionization source. Unfortunately, this data cannot be easily compared with other analytical platforms. However, other relevant methods usually use 100 μ l of plasma in their respective methods (Coene et al., 2018; Naz et al., 2017).

3.4.5.5 Sensitivity: Limit of detection of metabolites that are not detectable in wildtype biological matrices

Some metabolites that have been detected in reference standards have not been detected in biological matrices. However, as in inherited disorders of metabolism, enzyme dysfunctions often can lead to increased metabolite concentrations to the 100-1000 fold; metabolites may then become well detectable. To estimate to which concentrations these metabolites need to reach to be observable, we selected 50 metabolites of particular importance due to their involvement in different disorders. We then added these metabolites to EDTA plasma and assessed their detectability in different plasma dilutions. The previous chapter defined a metabolite as detectable when the signal intensity was at least three times higher than the blank.

In the 50 metabolites that were previously not detectable, in this measurement, we found 7 metabolites in the positive method and 3 metabolites that were detectable even in the untreated plasma. The mass spectrometers' sensitivity varies slightly between measurements, depending on factors such as contamination from previous runs, but even exogenous factors such as room temperature can slightly influence sensitivity. Using the standard addition to EDTA-plasma, we found that 22% of the metabolites can be detected in ZIC-HILIC-MS positive mode when adding them in concentrations up to 100 μ M to plasma (**Figure 30 A**). In negative ionization mode 32% of metabolites can be detected, which are mostly not overlapping with the positive mode (**Figure 30 B**). Around half of the compounds can be determined when they reach concentrations of at least 100 μ M. The other half of metabolite cannot be detected from biomaterials after standard additions.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

This may be caused by improper sample preparation, degradation in plasma, or even absorption of proteins. This effect is expected, as we include many metabolites with very different chemical properties.

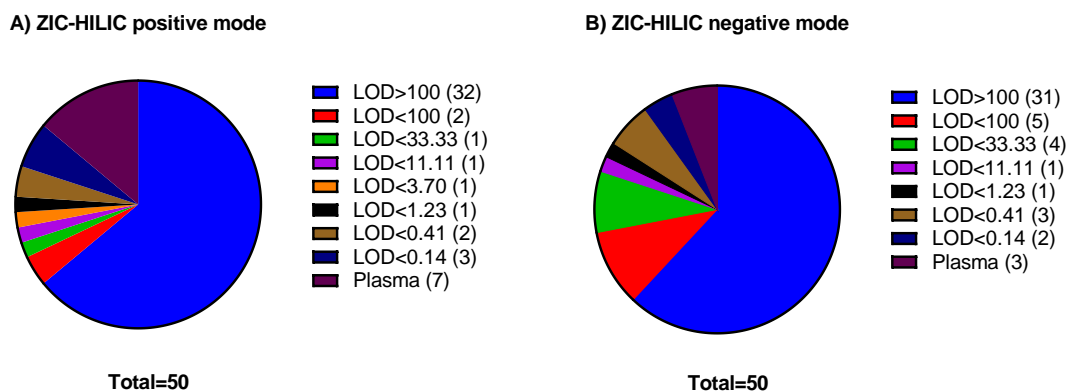


Figure 30 Limit of detection of 50 metabolites that were not detectable in healthy specimens. 50 metabolites that were not previously detectable in healthy human EDTA-plasma or wildtype fibroblasts with 1×10^6 cells were spiked into healthy human EDTA-Plasma and serially diluted to reach 8 different concentration levels. Then samples were measured using the ZIC-HILIC-MS method in positive A) and negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and mean signal abundance for each metabolite was calculated for $n=18$ measurements of 9 different sample concentrations in duplicates. The limit of detection was defined as thrice the signal abundance as the PBS-blank.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Adding metabolites to EDTA-plasma also allows us to determine the linearity of those metabolites in closer detail (**Figure 31**). In positive as in negative ionization mode, some of the 50 compounds show excellent linearities, such as nicotinic acid ($R^2=0.988$), thiamine ($R^2=0.967$), and biotine ($R^2=0.987$). Others such as creatine ($R^2=0.412$) did not correlate well. This is not unexpected, as this method was conceived mainly as a screening method.

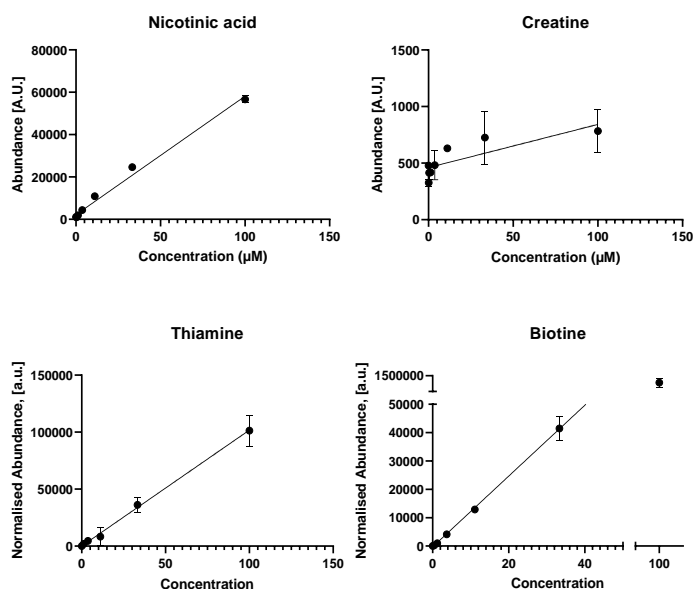


Figure 31 4 exemplary metabolites that were previously not detected in biospecimen showed high linearity in spiked samples: ZIC-HILIC-MS. 50 metabolites that were not detectable in healthy human EDTA-plasma or wildtype fibroblasts with our method were spiked into healthy human EDTA-plasma and serially diluted to reach 8 different concentration levels. Then samples were measured using the ZIC-HILIC-MS method in positive and negative ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and raw signal abundance plotted for $n=18$ measurements of 9 different sample concentrations in duplicates. 2 exemplary metabolites with high and low linearity are presented.

The linearity was consistently in the expected range and, together with the detected metabolite in standard addition, shows its sensitivity.

3.4.5.6 Linearity of the metabolomics method

Similar to the sensitivity, the linearity of the method cannot be determined classically due to the complexity of the metabolome and its endogenous nature. Thus, we can only assess the correlation of metabolite intensities to different EDTA-plasma dilutions.

Using plasma concentrations of 7 different dilutions, we find that over 50% of the analytes identified by ZIC-HILIC-MS in negative mode have a Pearson correlation coefficient of over 0.899 (**Figure 32 B**). 16 metabolites have a coefficient of less than 0.5. In positive ionization mode we find that over 50% of metabolites had a correlation greater than 0.35 with 73 metabolites that had a correlation of less than 0.5 (**Figure 32 A**). The negative ionization mode presented with a good amount of correlating metabolites that is comparable to other methods. The positive ionization mode has a greater amount of analytes that did not correlate well with the plasma concentration. It is possible that there is a higher amount of noise affecting correlation in the positive ionization mode. Mainly phospholipids may introduce the noise due to their ionization in positive mode. Phospholipids then may contribute to ion suppression, especially with increasing plasma concentrations.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

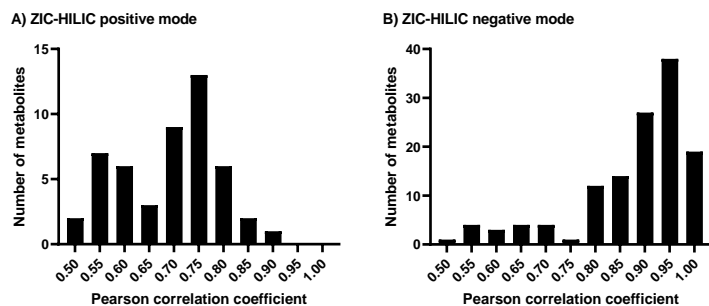


Figure 32 Distribution of Pearson correlation coefficient of identified metabolites from ZIC-HILIC-MS. 10, 50, 50,100, 150, 300, 400, and 500 μ l human EDTA-plasma were prepared and measured using the ZIC-HILIC-MS method in positive A) and negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial, and Pearson correlation coefficient for each metabolite was calculated for $n=21$ measurements of 7 different concentrations in triplicates. Metabolites with a coefficient greater than 0.5 were plotted as a histogram with a bin width of 0.05.

Previous validation parameters such as specificity could not be analyzed for the CortecsT3 lipidomics method due to the absence of reference standards for most metabolites. However, the sensitivity of the method using plasma dilution can be analyzed.

When analyzing the CortecsT3 lipidomics method, we find that over 50% of the 540 identified metabolites in the analysis have a Pearson correlation coefficient greater than 0.835 (**Figure 33**). 193 metabolites had a coefficient of less than 0.5.

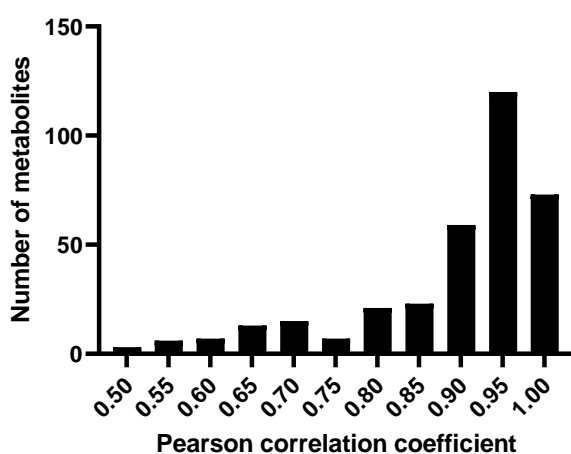


Figure 33 Distribution of Pearson correlation coefficient of identified metabolites from CortecsT3 lipidomics method. 10, 50, 50,100, 150, 300, 400, and 500 μ l human EDTA-plasma were prepared and measured using the CortecsT3 lipidomics method. Peaks were identified using the retention time reference library in MS-Dial, and the Pearson correlation coefficient for each metabolite was calculated for $n=21$ measurements of 7 different concentrations in triplicates. Metabolites with a coefficient greater than 0.5 were plotted as a histogram with a bin width of 0.05.

Additionally, to the distribution of correlation coefficients, we were also interested in whether there is a relationship to other parameters such as signal intensity.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

While the correlation of the analyzed metabolites does not seem to be in a relationship with the intensity of the metabolites alone, the correlation coefficients appear to be in a connection between the ratio between the mean abundance of the maximum plasma concentration to the mean abundance of the PBS-blank. This may be used to represent an alternative signal-to-noise ratio (**Figure 34, A**).

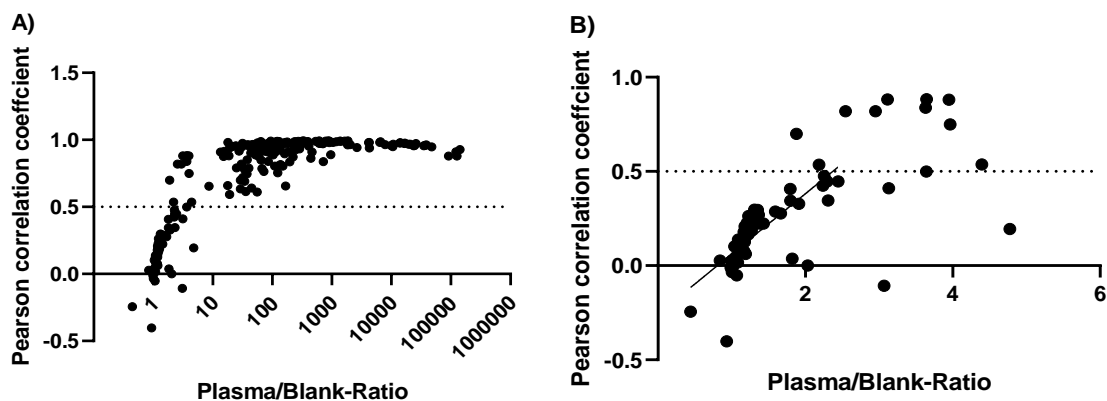


Figure 34 Graph of Pearson correlation coefficient in dependence on the maximum plasma/blank-ratio: A alternative signal to noise ratio. 10, 50, 50,100, 150, 300, 400, and 500 μ l human EDTA-plasma were prepared and measured using the CortecsT3 lipidomics method. Peaks were identified using the retention time reference library in MS-Dial and Pearson correlation coefficient for each metabolite was calculated for $n=21$ measurements of 7 different concentrations in triplicates. Then mean abundance of 500 μ l plasma to mean abundance of PBS-blank-ratio to Pearson correlation of plasma concentration was plotted. A) Plasma/Blank-Ratio was plotted logarithmically. B) A linear regression of ratios up to 2.5 was calculated (black line). A correlation coefficient may be used as measure for the linearity of the metabolite and metabolites with a high linearity may behave more reliably when comparing between different samples.

For Plasma/Blank-Ratios of up to 2.5, there is an acceptable linear fit ($R^2=0.559$) of the Plasma/Blank-Ratio to plasma correlation coefficient (**Figure 34, B**) in the CortecsT3 lipidomics method. The regression line meets the correlation coefficient of 0.5 at a ratio of approximately 2.4. Setting a threshold of 2.4 removes 91% of the metabolites with a correlation of less than 0.5, and 5.5% of these values have a correlation of more than 0.5.

Similarly, for ZIC-HILIC-MS, the minimum Plasma/Blank-Ratio can be calculated. For negative mode, the linear fit was close to the lipidomics method ($R^2=0.591$); the threshold appears to be at 1.8. Setting the threshold at 1.8, 75% of metabolites with a correlation of less than 0.5 were removed, and this value contained 10.4% values with a correlation of larger than 0.5. Due to larger scattering in the ZIC-HILIC positive mode, the threshold cannot be determined unequivocally. However, using the threshold from negative mode set at 1.8, removing 76.7% metabolites with a concentration of less than 0.5 and 3.1% have a correlation higher than 0.5.

The Plasma/Blank-Ratios very well reach the recommended level of a signal/noise ratio of 3.0. Even so, all values were slightly smaller, indication an above-average performance for most analytes.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

3.4.5.7 Selectivity: Certain metabolites are influenced by hemolysis

Human plasma samples are prepared by using centrifugation to separate the liquid plasma from cellular components. Of the cellular components, erythrocytes are the most abundant cell type. Due to various reasons, such as, e.g., prolonged or inappropriate storage, different disorders, or others, erythrocytes can lose their plasma integrity leading to a mixture of erythrocyte cell contents with blood plasma. This can lead to various shifts in the concentration of a significant number of metabolites and influence sample preparation (Denihan et al., 2015). Possible effects include the increase of typically intracellular metabolites and the decrease of metabolites due to problems in sample preparation.

Thus, we investigated which of the metabolites tested by our platform are affected by using four different concentrations of hemolyzed blood.

In ZIC-HILIC-MS with negative ionization mode, we found a strong increase in especially 2-octenoic acid and reduced glutathione and decreased metabolites such as bilirubin (**Table 9**). Reduced glutathione is a specific intracellular metabolite, protecting the erythrocytes from oxidative damage. Reduced glutathione is typically not found inside the plasma due to its oxidative environment. An increased concentration of reduced glutathione after hemolysis was also found by another group (Giustarini et al., 2004).

2-octenoic acid is a monounsaturated medium chain fatty acid, while an intermediate of β -oxidation, its occurrence has not been described in Erythrocytes so far. It is possible that this compound may be a product of reactions in the mass spectrometry source and originates from other compounds.

Table 9 TOP 5 metabolites altered by hemolyzed blood: ZIC-HILIC-MS in negative mode. Healthy human EDTA-plasma was spiked with 2, 10, 20, and 50% of lysed whole blood from the same blood draw to determine metabolite alterations caused by hemolysis. The samples were measured using the ZIC-HILIC-MS method in negative ionization mode, and peaks were identified using the retention time reference library in MS-Dial and Spearman correlation coefficient for each metabolite was calculated for n=10 samples of 5 different lysed blood concentrations, measured in duplicates.

Increased with hemolyzed blood		Decreased with hemolyzed blood	
Metabolite	Correlation	Metabolite	Correlation
2-Octenoic acid	0.995	O-Succinylhomoserine	-0.994
Quinaldic acid	0.986	N-Acetylserine	-0.987
2-Methylhippuric acid	0.980	Bilirubin	-0.973
Glutathione(reduced)	0.974	Tetradecanedioic acid	-0.956
Methylhistidine	0.962	Dihydroxybenzaldehyde	-0.949

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Both compounds correlate very well with the concentration of lysed blood over a great range of concentrations (**Figure 35**). These compounds may be used as estimators for estimating the degree of hemolysis. However, they may also be influenced by the diseased state.

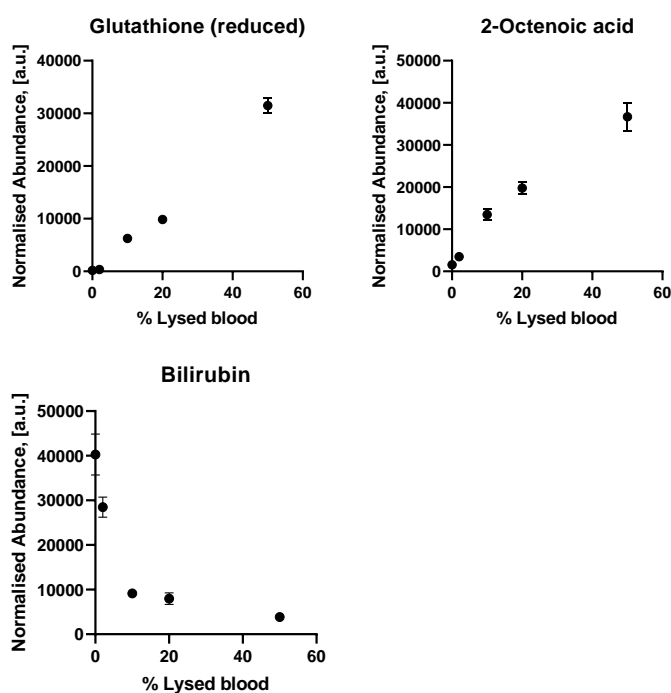


Figure 35 Exemplary metabolites that changed with the concentration of lysed human whole blood: ZIC-HILIC negative mode. Healthy human EDTA-plasma was spiked with 2, 10, 20, and 50% of lysed whole blood from the same blood draw to determine metabolite alterations caused by hemolysis. Samples were measured using the ZIC-HILIC-MS method in negative ionization mode, and peaks were identified using the retention time reference library in MS-Dial. n=10 examples of 5 different lysed blood concentrations, measured in duplicates.

Among the decreased metabolites in negative mode especially was bilirubin, a degradation product of hemoglobin. While the positive correlating metabolites increase linearly, most negative correlation metabolites rather decreased exponentially. Some intracellular erythrocyte components, such as e.g., hemoglobin, may decrease the retention of bilirubin on the solid phase extraction column, leading to a decrease. Using bilirubin, even small degrees of hemolysis may be detected. Interference of hemolysis in bilirubin analysis is known widely in clinical chemistry (Brady and O'Leary, 1998; Brunori et al., 2011). However, here it primarily originates from interference in indicator reaction and detection. To our knowledge, an influence in the extraction of bilirubin was not documented.

In ZIC-HILIC positive ionization, especially metabolites such as imidazole acetic acid and 2-hexenal, were increased in the lysed blood (**Table 10, Figure 36**). The imidazole-ring is a very prominent feature in the structure of heme, thus probably originating as a degradation product. However, to our knowledge there were no previous information published in this context. Imidazole acetic acid has also been described as degradation product of histamine (Tabor, 1954).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Similar to octanoic acid, 2-hexenal has not been described so far in lysed blood but also may be originating from degradation products. 2-hexenal is known as marker for oxidative stress (Frankel et al., 1989). It possibly is created by lipid peroxidation due to increased oxidative stress by hemolysis.

Table 10 TOP 5 metabolites altered by hemolyzed blood: ZIC-HILIC-MS in positive mode. Healthy human EDTA-plasma was spiked with 2, 10, 20, and 50% of lysed whole blood from the same blood draw to determine metabolite alterations caused by hemolysis. The samples were measured using the ZIC-HILIC-MS method in positive ionization mode and peaks were identified using the retention time reference library in MS-Dial and spearman correlation coefficient for each metabolite was calculated for n=10 samples of 5 different lysed blood concentrations, measured in duplicates.

Increased with hemolyzed blood		Decreased with hemolyzed blood	
Metabolite	Correlation	Metabolite	Correlation
Imidazole acetic acid	0.986	Bilirubin	-0.925
2-Hexenal	0.975	Arginine	-0.889
Gluconolactone	0.933	Glutamic acid	-0.873
Anserine	0.927	Citrulline	-0.872
Nicotinamide	0.922	Glycero-3-Phosphocholine	-0.853

The degree of hemolysis especially affected small positive molecules such as arginine and citrulline.

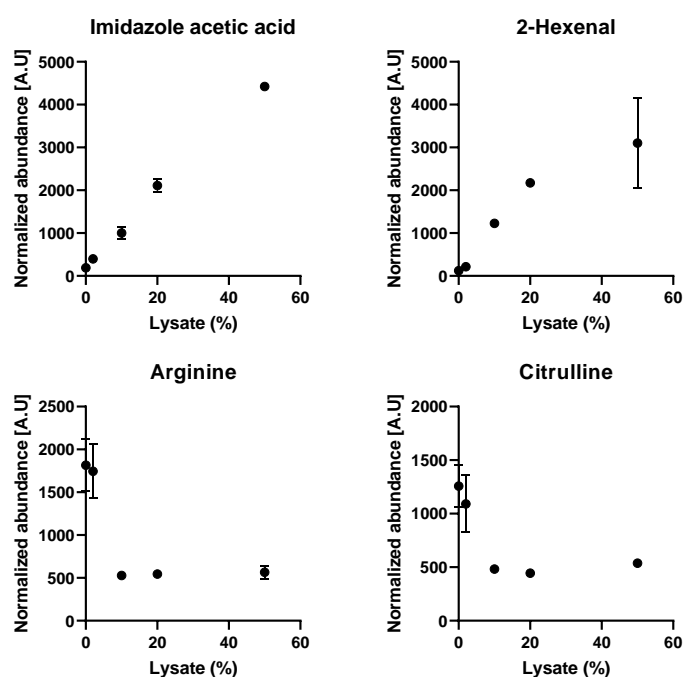


Figure 36 Exemplary metabolites that changed with the concentration of lysed human whole blood: ZIC-HILIC positive mode. Healthy human EDTA-plasma was spiked with 2, 10, 20, and 50% of lysed whole blood from the same blood draw to determine metabolite alterations caused by hemolysis. The samples were measured using the ZIC-HILIC-MS method in positive ionization mode, and peaks were identified using the retention time reference library in MS-Dial. n=10 samples of 5 different lysed blood concentrations, measured in duplicates.

Especially when studying urea cycle disorders, the degree of hemolysis should be controlled.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Interestingly we only found few and contradictory information on alterations of these metabolite in the context of hemolysis. Instead of a decrease, others have found a slight increase of arginine and citrulline during hemolysis (Kamlage et al., 2014). Thus, we concluded that the variation in our hands might originate from interfering with sample preparation.

In CortecsT3 lipidomics analysis, especially ether-phosphatidylethanolamines (Ether-PE) and cardiolipins were increased in lysed blood (**Table 11, Figure 37**). Ether-PEs are well known as essential constituents of cell membranes. Cardiolipins as well are important lipids in membranes. An increase is, therefore, only logical. Ether-linked triacylglycerides and monoacylglycerides show a negative correlation with lysed blood. However, the effect is less intense.

Table 11 TOP 5 metabolites altered by hemolyzed blood: CortecsT3 Lipidomics method. Healthy human EDTA-plasma was spiked with 2, 10, 20, and 50% of lysed whole blood from the same blood draw to determine metabolite alterations caused by hemolysis. The samples were measured using the CortecsT3 lipidomics method, and peaks were identified using MS-Dial. The total sum of a lipid class was calculated, and the Spearman correlation coefficient for each lipid class was calculated for n=10 samples of 5 different lysed blood concentrations, measured in duplicates. LPC: lysophosphatidylcholine; PC: phosphatidylcholine; CAR: acyl-carnitine; CE: cholesteryl-ester; LPE: lysophosphatidylethanolamine; SM: sphingomyelin; TG triacylglycerol; MG: monoacylglycerol; PE: phosphatidylethanolamine; PG: phosphatidylglycerol.

Increased with hemolyzed blood		Decreased with hemolyzed blood	
Metabolite	Correlation	Metabolite	Correlation
EtherPE	0.979	EtherTG	-0.854
CL	0.970	MG	-0.833
SHexCer	0.968	SM	-0.787
TG_EST	0.952	EtherPC	-0.772
OxTG	0.934	Cer_NDS	-0.662

Also, regarding ether-phosphatidylethanolamines and cardiolipins, we did not find further evidence in the literature regarding their influence by hemolysis.

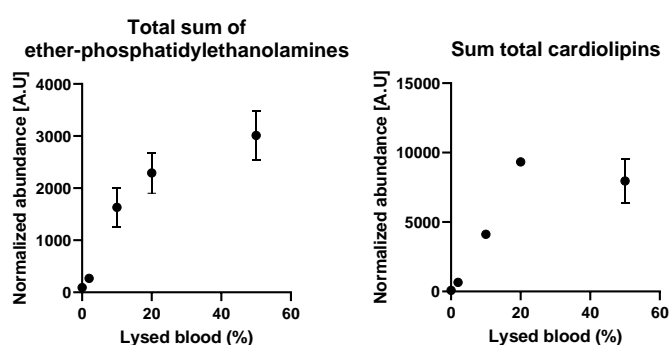


Figure 37 Exemplary metabolites that changed with the concentration of lysed human whole blood: CortecsT3 lipidomics method. Healthy human EDTA-plasma was spiked with 2, 10, 20, and 50% of lysed whole blood from the same blood draw to determine metabolite alterations caused by hemolysis. The samples were measured using the CortecsT3 lipidomics method, and peaks were identified using MS-Dial. The total sum of lipid classes were then calculated for n=10 samples of 5 different lysed blood concentrations, measured in duplicates.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

In a study by Denihan et al. they found ether-phosphatidylcholines, lysophosphatidylcholines and sphingomyelins decreased (Denihan et al., 2015). That indeed matches also our observations with reduced sphingomyelin and ether-phosphatidylcholine levels.

To summarize, we found increases in different metabolites ranging from typical intracellular metabolites over metabolites that may originate from their degradation and cell membrane lipids that may show hemolysis in plasma membranes. Also, we learned that especially parameters such as bilirubin and small positive amines such as arginine are severely affected by hemolysis. While there were no published details about most alterations for hydrophilic metabolites, we saw similar decreases of sphingomyelin and ether-phosphatidylcholines than another group.

3.4.5.8 Storage stability of different metabolites.

Metabolomic analyses require partly long sample handling times before analysis. The time from blood draw from the patient to analysis and variable conditions (such as temperature) during sample analysis strongly influence the sample stability (Breier et al., 2014; Haid et al., 2018; Stevens et al., 2019). While there is sufficient knowledge of the stability of some metabolites, the influences on stability may be strongly dependent on the analytical pipeline. Hence, we also aimed to assess which analytes are most affected by sample preparation and storage and which conditions have the greatest impact. For this, we treated a human healthy control sample with 8 different conditions. Apart from a control sample prepared freshly using the classical separation method, we also incubated EDTA whole blood for different times and different temperatures before plasma preparation. Also, plasma and the prepared metabolite extract were studied. Additionally, we tested for the effects of multiple freeze-thaw cycles on plasma. To get an overview of which conditions have the strongest influences on metabolites, we decided to compare the significantly ($p < 0.05$) changed metabolites compared to the control (**Figure 38**).

In ZIC-HILIC-MS positive mode, we found that storage of prepared EDTA-plasma at different temperatures and freeze cycles did not have a substantial effect on the metabolome. However, storing either EDTA-blood or prepared extract above 4°C leads to a high number of altered metabolites. In negative mode, we found strong alterations in all conditions; however, the freeze-thawed sample and the samples stored as EDTA-plasma had the least number of altered metabolites. In the publication by Haid et al. EDTA-plasma had a higher stability (compared to serum), with more metabolites changed at room temperature and a low number of alterations found for up to 3 hours (Haid et al., 2018)

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Among the strongly altered compounds in the negative mode were especially lactate, hypoxanthine, and glutamic acid (**Figure 39**). Lactate is specifically increased in the sample stored as EDTA blood, but very slightly increased also in plasma stored overnight at room temperature. The alteration of lactate in the blood may be caused by the continuous consumption of glucose and aerobic glycolysis to lactate. Temperature-dependent increases in full-blood are widely described (Seymour et al., 2011).

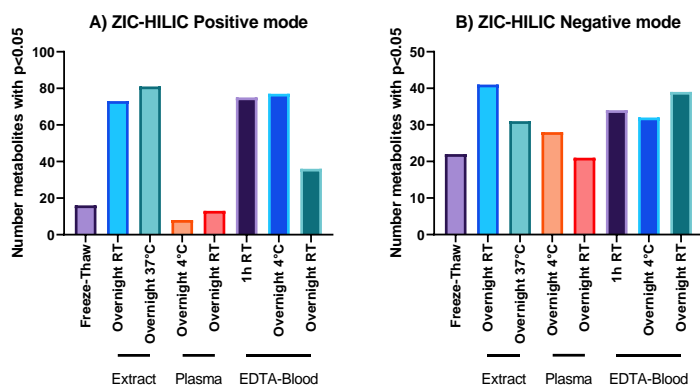


Figure 38 Significant alterations in metabolome in different extraction states under various external conditions: ZIC-HILIC-MS. Healthy human EDTA-plasma was stored under different extraction conditions (as LC-MS sample extract, EDTA-plasma, or EDTA-full-blood) and different storage times and temperatures (overnight, room temperature/RT or 4°C). Then samples were measured using the ZIC-HILIC-MS method in positive A) or negative B) ionization mode. Peaks were identified using the retention time reference library in MS-Dial and normalized using LOESS-Algorithm. Samples were measured in technical duplicates. Student's T-test p-values were calculated on the changes to normally treated plasma, and values with a p<0.05 were counted.

Hypoxanthine is a degradation product of purine bases, including adenine, and is a known factor increased in hypoxia and known to leak from erythrocytes (Bouliou et al., 1983; Saugstad, 1975). Thus, strongly increased Hypoxanthine may be related to the hypoxic condition after the blood draw.

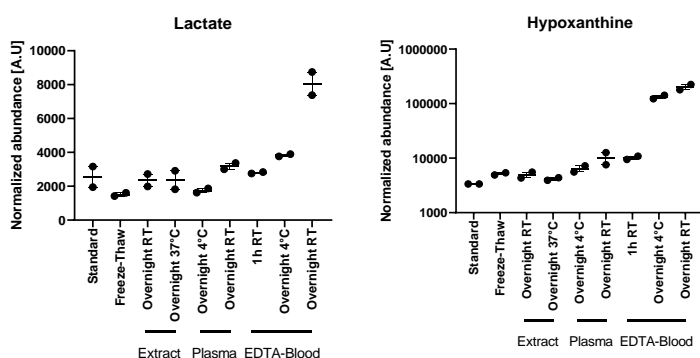


Figure 39 Representative metabolites changed by different extraction states under various external conditions: ZIC-HILIC-MS negative mode. Healthy human EDTA-plasma was stored under different extraction conditions (as LC-MS sample extract, EDTA-plasma, or EDTA whole blood) and different storage times and temperatures (overnight, room temperature/RT or 4°C). Then samples were measured using the ZIC-HILIC-MS method in negative ionization mode. Peaks were identified using the retention time reference library in MS-Dial and normalized using LOESS-Algorithm. Samples were measured in technical duplicates.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Glutamic acid is decreased significantly in the overnight stored samples (**Figure 40**). Glutamic acid may be deaminated either by enzymes (in plasma and EDTA whole blood) or nonenzymatically transformed (stored extract). Interestingly, so far, mostly degradation of glutamine to glutamate (with a concurrent increase of glutamate) was observed (Kornhuber et al., 1991; Rosenling et al., 2011)

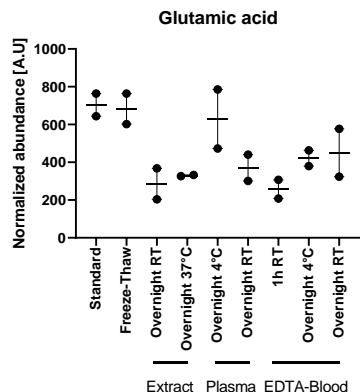


Figure 40 Representative metabolites changed by different extraction states under various external conditions: ZIC-HILIC-MS positive mode. Healthy human EDTA-plasma was stored under different extraction conditions (LC-MS sample extract, EDTA-plasma, or EDTA whole blood) and different storage times and temperatures (overnight, room temperature/RT, or 4°C). Then samples were measured using the ZIC-HILIC-MS method in positive ionization mode. Peaks were identified using the retention time reference library in MS-Dial and normalized using LOESS-Algorithm. Samples were measured in technical duplicates.

In the lipidome, we see the least changes under overnight storage as extract at room temperature and short time storage of one hour as EDTA-blood (**Figure 41**). The most significant changes can be seen in storing blood overnight or the extract at a high temperature.

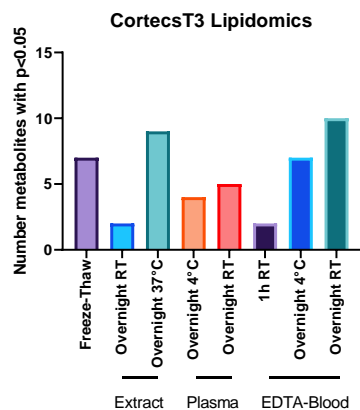


Figure 41 Significant changes in metabolome in different extraction states under various external conditions: CortecsT3 lipidomics method. Healthy human EDTA-plasma was stored under other extraction conditions (LC-MS sample extract, EDTA-plasma or EDTA-full-blood) and different storage times and temperatures (overnight, room temperature/RT or 4°C). Then samples were measured using the CortecsT3 lipidomics method. Peaks were identified using the retention time reference library in MS-Dial and normalized using LOESS-Algorithm. Samples were measured in technical duplicates. The total sum of lipid classes was then calculated. Student's t-test p-values were calculated on the changes to normal treated plasma, and values with a p < 0.05 were counted.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Especially total acyl-carnitines, as well as the total sum of ether-PE, was increased under different conditions (**Figure 42**). We found the highest increase of acyl-carnitines in blood stored overnight at 4°C, and it is decreased in extract stored at 37 °C. Acyl-carnitines are classical intracellular metabolites involved, for instance, in β -oxidation. Breier and Yang et al. showed similar decreases of different acyl-carnitines at 37°C (Breier et al., 2014; Yang et al., 2013a). However, they found most acyl-carnitines to be unchanged when stored at room temperature.

We also found an increase in ether-phosphatidylethanolamine in EDTA blood. As seen before, Ether-PE is increased in samples with higher hemolysis. Thus, the rise of compound classes may be attributed to the stronger hemolysis under those conditions.

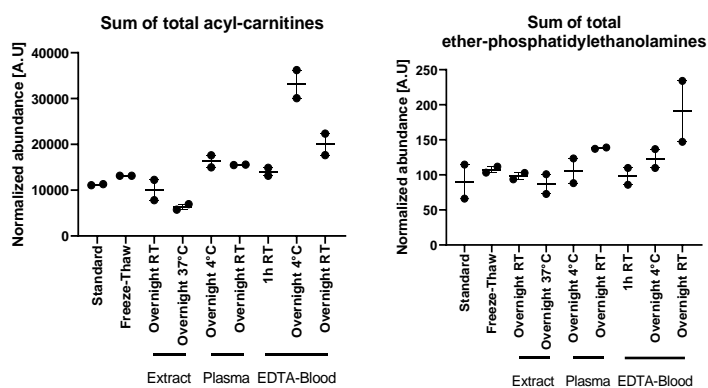


Figure 42 Representative metabolites changed by different extraction states under various external conditions: CortecsT3 lipidomics method. Healthy human EDTA-plasma was stored under different extraction conditions (as LC-MS sample extract, EDTA-plasma, or EDTA-full-blood) and different storage times and temperatures (overnight, room temperature/RT or 4°C). Then samples were measured using the CortecsT3 Lipidomics method. Peaks were identified using the retention time reference library in MS-Dial and normalized using LOESS-Algorithm. The total sums of lipid classes were then calculated. Samples were measured in technical duplicates.

Summarizing, we showed that the sample should ideally be stored as EDTA-plasma. Other conditions such as time after blood draw or storage as extract should be kept as small as possible. For most metabolites, the time to the extraction of up to one hour did not lead to strong alterations. Lactate and hypoxanthine were shown as very sensitive to storage as whole-blood.

3.5 Discussion: Development of a platform to detect rare neuropediatric disorders with mass spectrometry-based metabolomics

Rare neuropediatric disorders are often very complex disorders involving a multitude of different metabolomic pathways, even in monogenetic disorders (Willemsen et al., 2016). While next-generation sequencing techniques have ameliorated diagnostic complications, diagnosis and research on those remain challenging (Vinkšiel et al., 2021). This especially applies to cases where disease-causing mutations are not previously documented, and the clinical significance of these is undetermined (variants of unknown significance, VUS). Thus, a technique that can assess actual metabolic changes and even judge the relevance of VUS may significantly improve diagnostics of rare neurometabolic disorders. Metabolomics, the holistic study of metabolism, may provide such diagnostic abilities (D'Alessandro et al., 2012; German et al., 2005; Piras et al., 2016). Together with next generation sequencing studies, metabolomics might deliver highly complementary results.

Hence, we decided to establish a mass spectrometry-based metabolomics platform for the characterization of rare neurometabolic disorders. Metabolomics has been applied to study various rare disorders in predefined groups (to uncover biomarkers), but the diagnostic utility of metabolomics to our knowledge was only studied by few groups (Almontashiri et al., 2020; Bonte et al., 2019; Coene et al., 2018; Miller et al., 2015). Curiously, the analytical methods intensely varied, pointing out the need to develop in-house methods adapted to the local requirements and infrastructure. However, all those groups, Bonte et al., Coene et al., and Miller et al., used reversed-phase chromatography (RPLC) for their analyses. While certainly reducing measurement times and having high analytical stability, it has been reported that using hydrophilic interaction chromatography (HILIC) in combination with RPLC has greater orthogonality and provides a higher number of possible analytes (Contrepois et al., 2015; Naz et al., 2017; Tang et al., 2016; Wernisch and Pennathur, 2016). The previously proposed methods, for instance, do not provide insides into disorders such as lipid metabolism disorders (e.g., peroxisomal disorders).

Additionally, the publications of these authors mainly did not provide data about analytical limitations, such as sensitivities, specificities, or metabolites that are especially sensitive to interferences (e.g., of storage conditions, hemolysis).

Consequently, this project aimed to develop an untargeted metabolomics platform that would be primarily based on HILIC-MS but still be complemented using RPLC-MS, utilizing the same sample preparation platform. To ensure identification confidence, we created an in-house metabolite identification library. For further targeted and quantitative metabolomics analyses, we decided to apply the commercial AbsoluteIDQ p180-kit from Biocrates.

3.5.1 We successfully applied the AbsoluteIDQ p180-kit to our institute

To enable metabolomic analysis before establishing our untargeted metabolomics pipeline, we used the AbsoluteIDQ p80-kit by Biocrates (<https://biocrates.com/absoluteidq-p180-kit/>, Accessed on 31.05.2021). In the later stages of our metabolomics platform, we launched the kit as the second pillar of our platform for targeted metabolomics. Using a classical LC-MS and a fast, flow injection analysis MS (FIA-MS) method, the kit can analyze up to 187 metabolites in combined 17 minutes per sample. While this kit has been widely applied to many research areas and studied in an international ring trial (Siskos et al., 2017), every application to different laboratories is different.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

Additionally, not many have systematically studied the suitability over a longer time frame. Thus we wanted to get a small, qualitative overview of how the kit performed over the course of 4 years.

In general, we have produced well-received results using this method and successfully applied the method. System suitability tests, as well as quality controls, have primarily shown a stable running method. However, to ensure these checks' success, intensive cleaning of the flow paths in the devices necessary. This, applied mainly to the FIA-MS part, where even small (not detectable by other methods) lipid contaminations can severely increase the background noise and reduce the sensitivity of the analysis. In a few instances, we also observed strong intra-run signal breakdowns. Interestingly these were not reported by the ring-trial study of Siskos et. al (Siskos et al., 2017). High background may be caused by chemical noise in the mass spectrometer, where liquid chromatography buffer depositions lead to degradations of capillary material and signal breakdowns (Doff et al., 2012; Kang et al., 2017). Organic compounds such as phospholipids also accumulate and caused ion suppression (Annesley, 2003). Commonly found environmental contaminations of lipids, e.g., from gloves, hand creams, reaction tubes, and other materials, can further lead to worse detection limits due to higher noise (Keller et al., 2008). Of note is also that ESI-capillaries are optimized for higher flow methods and have a larger diameter than capillaries that should be used for lower flow methods (Gama et al., 2013). Lower flow methods, such as the FIA-MS method, when used with a non-optimized capillary, may lead to lower reproducibility of signals (Schmidt et al., 2003). So also, the p180-kit ring trial study observed interlaboratory reproducibility issues of C12:0 and C18:0-acylcarnitines, as well as the inability to analyze the sum of hexoses, PC aa C30:2, PC aa C32:2, PC aa C38:1, and SM C22:3 with our mass spectrometer supplier (Siskos et al., 2017). However, in most instances, the FIA-MS part of the p180-kit well-performed suitability checks and quality controls to fastly produce relevant results.

In regard to the system suitability checks and quality controls, we observed mainly well-validated signals with stable retention times. Curiously, the chromatographic columns had a reduced lifetime even with careful storage, compared to usual methods. For around 860 sample injections, three columns were used. Typically these columns last for about 500-1000 (Dong, 2006) injections each, using some methods even more. The reduced lifetime can have a multitude of reasons. One possibility would be the degradation of not completely removed derivatization reagents, which can be highly reactive or lead to the formation of particulates in the samples. Due to the used sample extraction method, typical other causes for low lifetimes, such as protein buildup, are improbable. Similarly observed by some laboratories in the ring trial (Siskos et al., 2017), we have seen problems in the standards as well as in the samples for the measurement of spermine, spermidine, and carnosine. These polyamines have a high polarity due to their typically multiple charges, and that may remain even after derivatization. This affects separation, leading to varying retention times. In fact, this was also observed by researchers using other methods (Häkkinen, 2011; Magnes et al., 2014). We, therefore, decided to exclude these metabolites in most analyses. For most other analytes, the p180-kit produced very stable results. Most measurements of serum or plasma were able to detect around 167 of the 187 metabolites theoretically measurable by the kit. This is congruent with the international ring trial.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

While, as discussed, certain analytes had to be excluded due to analytical reasons, certain lipids may just not be occurring in these sample types high enough for detection. Thus, in our institute, this kit provides the astonishing ability to quantify a very high number of metabolites using an inherently comparable method between different laboratories.

One note should be made to the identification limitations of the kit. Due to the used methods, the kit can only provide a total sum carbon count inside the esterified fatty acids chain and a total sum of double bonds without specifying the exact chain length or double bonds of the specific fatty acid chains. Additionally, there is no information on the position of the double bonds. This results in the absence of knowledge which fatty acid composition this feature now really has, and lipids with the same combination (same mass) interfere with the detection (isobaric analytes). As the fatty acid composition can significantly influence lipid function, follow-up studies are recommended. This could be potentially alleviated by using mass spectrometers that generate fragments using electron-induced dissociation (EID) techniques as opposed to the more common collision induced fragmentation. EID is able to reproducibly create fragments that allow for subdifferentiation of fatty acid conformation and double bond position (Jones et al., 2015). Albeit this approach is not even widely utilized in the untargeted lipidomics field. However, due to the fast and easy application of the FIA-MS p180-kit method, it merits potential identification caveats.

In summary, we were able to apply the AbsoluteIDQ p180-kit from Biocrates to our institute successfully. We were able to solve practical problems and were able to generate reproducible data. In the following chapters (chapter 4.0 p. 88 and chapter 5.0 p. 99), we then applied the kit to different disorders and gained relevant insights that were also published in peer-reviewed journals (Kettwig et al., 2021; Klemp et al., 2021).

3.5.2 We established an untargeted metabolomics pipeline

The AbsoluteIDQ p180-kit enabled us to quantitatively study up to 187 metabolites in a fast and reproducible manner. However, the coverage of the metabolome using only this method would be small. Thus, as a second pillar of our metabolomics platform, we developed untargeted methods to characterize a broader range of metabolites. This may allow us to detect metabolomic changes with higher sensitivities. Only a small set of platforms were described for the detection and research of previously not diagnosed rare disorders (Bonte et al., 2019; Coene et al., 2018; Miller et al., 2015). Curiously, these methods vary strongly; however, all used reversed-phase liquid chromatography (RPLC) and focused primarily on targets with medium polarity. Very polar metabolites or lipids were not targets of their screens. To increase the scope of analyzable metabolites, several authors have already proposed combining RPLC-methods with an orthogonal method, hydrophilic ion interaction chromatography (HILIC; Contrepolis et al., 2015; Naz et al., 2017; Tang et al., 2016; Wernisch and Pennathur, 2016). While the use of a single RPLC-method by the groups of Coene, Bonte, and Miller reduces the time of analysis and RPLC methods are also known to be more stable than HILIC methods, we decided to use an approach similar to the one by Naz et al. and combined a lipid-analysis RPLC method with HILIC-MS for hydrophilic metabolites. As discussed previously, for high identification confidence, especially hydrophilic metabolites, retention time and accurate fragment data are needed. Thus, we also decided to acquire 502 metabolites to generate a metabolite identification library and retrospectively analyze metabolites, even from an untargeted dataset.

3.0 Development of the metabolomics platform for rare neurometabolic disorders

3.5.2.1 We developed a LC-MS method for the untargeted metabolomics pipeline

We successfully developed a HILIC-MS method for the detection of up to 408 compounds. An overview of the most important metabolite identification parameters can be found in the annex (**Table 19, Table 20**).

To gain more information on the validity of our metabolite identification, we studied the retention time and mass error inside specific runs, as well as our metabolite reference library. Opposed to most other groups, we did not only explore a distinct set of internal standards but all detectable metabolites from our library. We found values well within the ranges of other methods, typical for this type of mass spectrometer regarding the mass error (Bristow and Webb, 2003; Köfeler and Gross, 2005). The retention time errors inside a run were low; however, the retention time errors to our metabolite library were partly very high. Most metabolites had a retention time error of less than 0.2 minutes. However, some others had an error of up to 0.65 minutes. Compared to the method by Bonte et al., as well as to the selected standards by Coene et al. and classical validation guidelines this is quite high (Bonte et al., 2019; Coene et al., 2018). The comparatively lower retention time stability of HILIC methods is widely described (Spagou et al., 2010; Tang et al., 2016; Wernisch and Pennathur, 2016). Instabilities are typically caused by insufficient (re-)equilibration times and complex interactions with sample matrix compounds, including salt and phospholipids (Alpert, 1990; Erngren et al., 2021; Naidong, 2003). However, as we mainly plan to use this method as a first-tier screening method and results are confirmed in secondary steps, a broader metabolite coverage is of greater importance than absolute identification confidence. Furthermore, in the lipidomics method we showed that lipid identification does not rely as strongly on retention time libraries due to the structured nature of lipids.

3.5.2.2 RPLC-SPE was introduced as superior method for sample preparation

Based on this, we continued developing a sample preparation method that allows us to extract samples for both, the RPLC and HILIC measurement from just one respective sample (**Figure 43**). Here we decide to use an RPLC solid-phase extraction (SPE) protocol. Interestingly, the use of an RPLC-SPE to split hydrophilic and lipophilic metabolites into two samples as described by us is uncommon to the metabolomics setting. Typically SPE methods in general are only used for the removal of unwanted compounds such as phospholipids (Bylda et al., 2014; Tsakelidou et al., 2017; Van Damme et al., 2014). In metabolomics, most platforms incorporate either simple protein precipitation or Liquid-liquid extraction techniques to yield hydrophilic and lipophilic extracts (Bonte et al., 2019; Coene et al., 2018; Dettmer et al., 2007).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

In our method, the aqueous extract has a high ion strength as ions are not removed by the SPE column and increased by sample evaporation. High ion strengths in LC-MS can be problematic as they lead to interferences/ion suppression as well as lead to degradation of columns as mass spectrometers (Doff et al., 2012; Erngren et al., 2021).

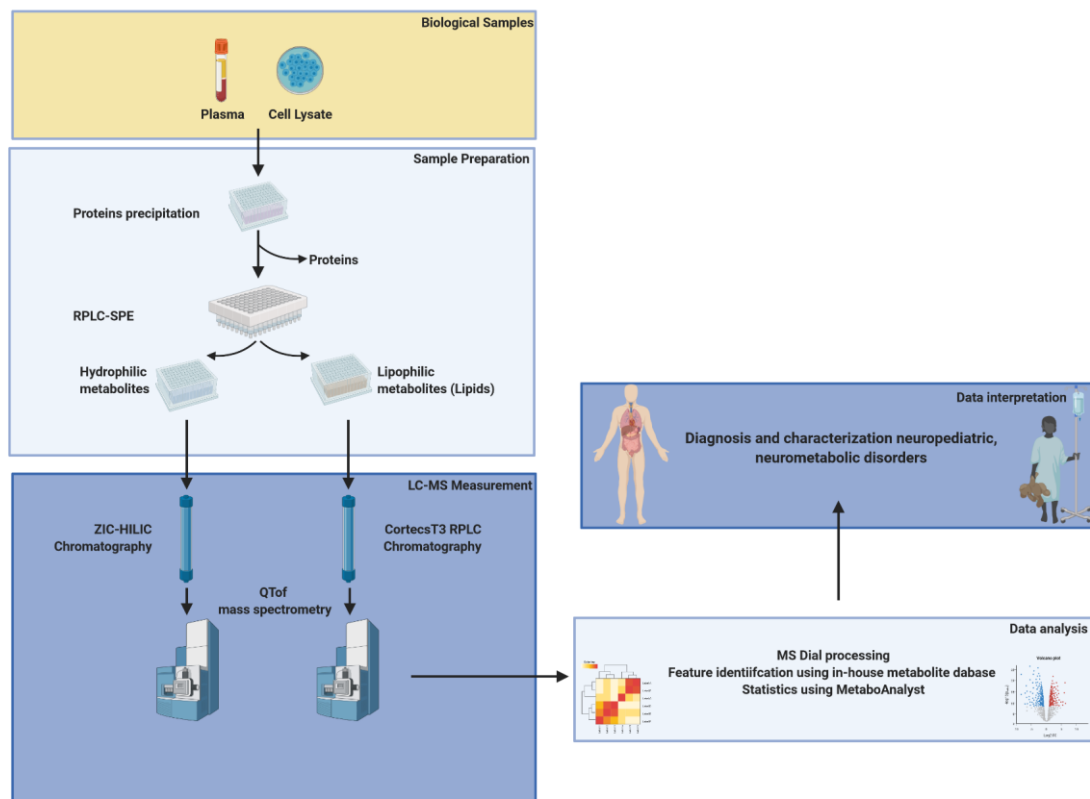


Figure 43 Our untargeted metabolomics pipeline uses RPLC-SPE and two LC-MS-methods for the analysis of lipids and hydrophilic metabolites. The untargeted metabolomics pipeline consists out of multiple steps. First the protein of biological samples is removed and then hydrophilic and lipophilic metabolites are extracted from the same sample using a reversed phase solid phase extraction (RPLC-SPE). The hydrophilic metabolites are measured using the ZIC-HILIC-MS method, the lipophilic metabolites are measured using the CortecsT3 RPLC-MS lipidomics method. The data are evaluated using MS-Dial and ZIC-HILIC measurements evaluated using the in-house identification database. Further statistical analysis with MetaboAnalyst and other tools allows interpretation and characterization of disorders.

Reduction of inorganic ions cannot simply be achieved, as standard techniques such as ion-exchange column also remove charged metabolites of interest. Most harmful ions leave the column either before or after most of the analytes. To protect at least the mass spectrometer, we therefore direct these into the waste using a fluidic valve. For the analysis of the compounds of interest, the valve then switches the LC flow back into the MS.

3.5.2.3 We were able to include 408 metabolites into the library and detect 192 in biological samples

Using our complete method platform for hydrophilic metabolites, we detected up to 192 metabolites from our retention time library in the healthy state. With these metabolites, we were able to cover an extensive range of the metabolome. Including metabolites that we could identify by using publicly available mass/fragment libraries, we could detect up to 408 compounds (with less confidence).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

We observed limitations of our method, especially for metabolites in the central carbon metabolism (primarily sugars and certain tricarboxylic acids), typically not distinguishable or showing very broad signals. Due to the polyol nature and their high similarity, the separation of sugars is not easy to achieve.

Therefore, sugar analysis is mainly done with GC-MS (Xia et al., 2018) or high specialized LC-MS methods, including derivatizations (Licea-Perez et al., 2016). Even though sugar analysis is not a strength of this method, we still might recognize changes in this metabolite class and reanalyze the sample with a targeted metabolomics method.

In comparison to platforms with similar aims, our method performed well. Bonte et al. detected 178 annotated metabolites (Bonte et al., 2019), the method of Coene et al. identified around 222 metabolites in patient plasma (Coene et al., 2018), each with verified retention times. Additionally, our pipeline also analyzes the lipid profile, significantly increasing the detectable metabolites, even without retention time information.

We determined that most metabolites were detectable even in plasma samples as low as 100 μ l. The information on which metabolites are typically visible in specific plasma volumes is highly beneficial for analyzing volume-restricted samples. In neuropediatric care, a great proportion of patients is very young, and blood cannot be extracted easily. Thus, it is important to know which metabolite will not be visible when reducing sample volume. With this data set, we also showed a good correlation between the applied sample amount and LC-MS signal for a great number of metabolites. A good correlation is indicative that also samples that have different concentrations (e.g., "healthy"- "diseased") can be distinguished. Very high intensity and very low-intensity metabolites showed less good correlation, and certain metabolites had high intensities in blanks. A lower change of signal intensity of high-intensity metabolites is a usual phenomenon and can be caused either by saturation of the extraction process, ionization source, or detector saturation (Lu et al., 2017). In our analysis, we showed some typically low intensity and/or high blank intensity metabolites. Increased background noise and smaller intensities (low signal/to noise) ultimately lead to larger intensity variations. As a result, this then leads to worse correlation and ultimately to lower distinguishing power in the samples. Hence, we determined the sample intensity/blank intensity as a proxy parameter for the signal/noise ratio and showed a relationship to the correlation coefficient. We then estimated optimal parameters for the cut-off to ensure reproducibility for our analyses.

3.5.2.3 Discussion: Interferences and metabolite stability

In the following steps, we determined the sensitivity of our pipeline against interferences and metabolite stability. Unfortunately, to our knowledge, there are no comprehensive untargeted metabolomics studies available. We found only one article that directly analyzed hemolysed blood using a targeted metabolomics method (Denihan et al., 2015). In our study, hemolysis mainly impacted levels of glutathione, bilirubin, arginine, citrulline, ether-phosphatidylethanolamines, as well as sphingomyelins and ether-phosphatidylcholines. The increase in glutathione (Giustarini et al., 2004), as well as the decrease of sphingomyelins and ether-phosphatidylcholines (Denihan et al., 2015), were previously mentioned. For the other metabolites, we could not find relevant documentation. Other substances such as hemoglobin and salts that are typically altered in routine clinical chemistry are not amendable to our analysis and were not studied (Koseoglu et al., 2011).

3.0 Development of the metabolomics platform for rare neurometabolic disorders

We also broadly studied which storage conditions affect metabolite stability and which metabolites are altered. We found that most metabolites tolerated storage of up to one hour as EDTA whole blood and storage of EDTA-plasma overnight at 4°C. This matched observations commonly found in literature (González-Domínguez et al., 2020; Kamlage et al., 2014).

In our analysis, we especially found that lactate and hypoxanthine and etherphosphatidylethanolamines increased in response to more extended storage as whole-blood. The increase of lactate is well described and the result of anaerobic glycolysis in erythrocytes (Jobard et al., 2016; Malm et al., 2016; Seymour et al., 2011). Hypoxanthine also has been found after whole blood storage and is probably caused by hypoxic conditions in the blood or leaking from erythrocytes (Boulieu et al., 1983; Saugstad, 1975). The increase of etherphosphatidylethanolamines has not been described, but we observed in the previous chapter as an increased metabolite in hemolysis.

3.6 Summary: Development of a LC-MS-based metabolomics platform for characterization of neuropediatric disorders.

To summarize, we successfully developed an LC-MS-based metabolomics platform for the characterization of rare disorders. The platform consists of untargeted methods as well as targeted methods, including the ring-trial validated AbsoluteIDQ p180-kit. We developed an untargeted metabolomics platform consisting of an RPLC and a HILIC method utilizing a quadrupole time of flight mass spectrometer. For the reliable identification of hydrophilic compounds, we then successfully established an identification library of 408 compounds that contains retention time and fragment data. This library can be increased with every project, and metabolites can even be identified retrospectively. Then we developed a sample preparation method to study 193 of those metabolites in the HILIC measurement and were also able to extract lipids for the RPLC method. We determined validation parameters and compared our method with the methods of other researchers. Lastly, we broadly analyzed which parameters are altered upon different storage conditions or affected by hemolysis.

In the following chapters, we want to present some of the successful applications comprehensively to demonstrate the utility of our platform.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

Parts of this work have been published as Klemp HG, Kettwig M, Streit F, Gärtner J, Rosewich H, Krätzner R. LC-MS Based Platform Simplifies Access to Metabolomics for Peroxisomal Disorders. Metabolites. 2021 May 29;11(6):347. doi: 10.3390/metabo11060347.

4.1 Introduction to metabolomics for peroxisomal disorders

Peroxisomes are key organelles for lipid metabolism inside cells, being central to the catabolism of very long chain fatty acids (VLCFA), phytanic acid, pristanic acid, and the synthesis of ether-phospholipids (Waterham et al., 2016). Alterations in the genes responsible for their biogenesis, the PEX genes, lead to devastating and complex disorders. These disorders, classified as Zellweger Spectrum disorders (ZSS), typically involve the brain, liver, and kidneys (“cerebro-hepato-renal syndrome”; Kelley, 1983; Klouwer et al., 2015). Neurological symptoms range from attention-deficits and behavioral abnormalities to dementia, poor development, and convulsions. Other peroxisomal disorders, such as disorders in single enzymes, may influence only a subset of the peroxisomal metabolism and partly lead to less severe symptoms (Wanders and Waterham, 2006). Belonging to this subgroup are diseases such as X-Adrenoleukodystrophy (X-ALD) and D-bifunctional protein deficiency type III (D-BP_{III}). X-ALD is caused by a mutation in the ABC-transporter D1 gene (*ABCD1*, X-ALD protein, ALDP; Blaw, ME, 1970; Mosser et al., 1994), most likely responsible for the import of very long chain fatty acids into peroxisomes. Mutations in the *HSD17B4* gene, a protein taking part in peroxisomal β -oxidation, result in D-BP_{III} (Ferdinandusse et al., 2006).

As mentioned, one central function of peroxisomes is the degradation of very long chain fatty acids (VLCFA). Hence this is intensively used in biochemical diagnostics of most peroxisomal disorders (Moser and Moser, 1996). The primary VLCFA markers are lignoceric acid (24-carbon fatty acid) and cerotic acid (26-carbon fatty acid), often used also as a ratio to a non-VLCFA such as behenic acid (22-carbon fatty acid) to provide more nutrition independent markers. For subdifferentiation of peroxisomal disorders, ether-phospholipids (e.g., plasmalogens) are used as diagnostic markers, as peroxisomes are necessary for introducing the ether-bond in those lipids (Hardeman and van den Bosch, 1989). Here conditions such as peroxisomal biogenesis disorders (e.g., ZSS) show a reduction, diseases affecting VLCFA-alone such as X-ALD do not affect ether-lipids. Additionally, other peroxisomal metabolites such as phytanic and pristanic acid can be determined (Verhoeven et al., 1998; Wanders et al., 2011). All analytes are typically determined using gas chromatography mass spectrometry (Verhoeven et al., 1995). While this technique provides very stable results, it often requires perilous and laborious derivatization procedures leading to low analytical speed. In the last years, several ZSS cases with milder or atypical classical peroxisomal laboratory biomarkers were reported (Bacino et al., 2015; Mignarri et al., 2012; Zeharia et al., 2007). As the biochemical biomarkers are still crucial for the assessment before the genetic screen is started (Braverman et al., 2016), this suggests that there are several wrongly characterized patients. Thus, it remains necessary to find additional new biomarkers for improved diagnostics and research purposes. The peroxisomal metabolism is complex and interacts with multiple other organelles (Sargsyan and Thoms, 2020), such that a holistic technique as metabolomics may be able to find further useful biomarkers.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

Previously there have been mainly untargeted metabolomics studies, especially in fibroblasts and serum (Herzog et al., 2016, 2018a, 2018b). These studies were able to find new biomarkers such as sphingomyelins and certain species of phosphatidylcholines. Untargeted metabolomics pipelines nevertheless face problems (Liebisch et al., 2017, 2019). These pipelines often require specialized knowledge in setting up analytical infrastructure, are only semiquantitative, and often produce results not easily comparable between different institutes. The incidences of peroxisomal disorders are low (Suzuki et al., 1996). To advance research, it is of supreme importance to easily compare the results of different groups to maximize knowledge in these disorders.

To alleviate this, we decided to apply the commercial AbsoluteIDQ p180-kit from Biocrates. As mentioned in the previous sections, this kit is a liquid-chromatography and flow injection mass spectrometry-based approach to quantify up to 187 metabolites. This kit can deliver quantitative results and was validated in international ring trials, showing its high comparability between different institutes. Here we decided to use ZSS and X-ALD as two significant peroxisomal disorders to find whether the AbsoluteIDQ p180-kit delivers results comparable to untargeted metabolomics methods and may provide additional biomarkers for this group of peroxisomal disorders. This is the first application of this metabolomics kit in our institute, making it very relevant for us to assess whether the kit integrates well in our metabolomics platform.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

4.2 Specialized methods and cohorts: new LC-MS based platform for peroxisomal disorders

A total of 17 ZSS patients with a genetically confirmed and known mutation in the Peroxisome biogenesis factor 1 (*PEX1*) gene with a mean age of 2.6 ± 4.9 years were analyzed. Patients had a ratio of lignoceric acid (C22:0) to behenic acid (C24:0) of 1.6 ± 0.3 as determined by total serum fatty acid analysis by GC-MS.

In addition to these patients with peroxisomal biogenesis defects, a total of 38 patients with peroxisomal single enzyme defects were examined with this kit. This group consisted of 36 X-ALD patients with genetically confirmed and known mutations in the *ABCD1* gene and two patients with D-bifunctional protein deficiency type III (D-BPIII) with mutations in the *HSD17B4* gene.

The control cohort consisted of 28 age-matched patients, which submitted serum for a routine check-up of amino acids in serum but did not display any abnormalities and did not have any prior history of peroxisomal disorders.

All procedures occurred following the ethical standards of the responsible committee on human experimentation and the Helsinki Declaration. The institutional Ethics Committee of University Medical Centre Göttingen approved this study (protocol code 4/11/16, date: 06.12.2016).

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

4.3 Results: a new LC-MS based platform for peroxisomal disorders

The AbsoluteIDQ p180-kit is a flow-injection and liquid chromatography mass spectrometry-based platform for the quantitative analysis of up to 187 metabolites (further information can be found in chapter 3.1.1, p. 32). We were interested if it can be used for the analysis of peroxisomal disorders, can reproduce key findings of other (untargeted) metabolomics studies, and may even be able to find new biomarkers for the detection of these disorders. Additionally, our group was interested in whether the commercial kit can be integrated as part of our metabolomics platform.

To assess the suitability of the commercial kit, we examined serum samples of 17 Zellweger spectrum patients (ZSS, *PEX1*-mutation), 36 X-adrenoleukodystrophy patients (X-ALD, *ABCD1*-mutation), as well as two D-bifunctional protein deficiency patients (*DPBIII*, *HSD17B4*-mutation).

4.3.1 167 metabolites were found in plasma samples.

Using the commercial kit, we found 167 metabolites with concentrations higher than the limit of quantification. These metabolites were used for further evaluations.

4.3.2 ZSS patients present with decreased ether-linked phospholipids

Among the altered metabolites in ZSS patients, especially acyl-ether-linked phospholipids were severely depleted, as visualized by the volcano plot (**Figure 44**). In previous untargeted metabolomics studies, acyl-ethyl-linked phospholipids such as PC ae 36:4, 40:6 and 44:5 were similarly found to be decreased (Herzog et al., 2018b).

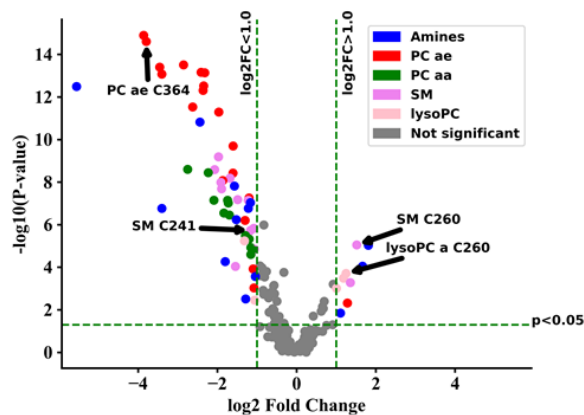


Figure 44 Volcano-plot showing metabolic changes of Zellweger syndrome spectrum patients

(ZSS/*PEX1*) in serum compared to the control group. Components with a raw p-value < 0.05 and log₂ fold change (log₂FC) > 1.0 are displayed in color. lysoPC: lysophosphatidylcholine; PC aa: phosphatidylcholines with two acyl-linkages; PC ae: phosphatidylcholine with one acyl and one ether linkage; SM: sphingomyelin

The determination of ether-lipids by GC-MS is a routine method for the biochemical diagnosis of ZSS. For this determination, two derivatization products of these lipids, commonly C16-diacetal and C18-diacetal are used as markers, as the lipids themselves are not stable in GC-MS.

Hence, we also examined C16 and C18-diacetal in our ZSS patients. Curiously, in four cases C18-diacetal and in 3 instances also C16-diacetal appeared normal (**Table 12**).

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

While the patients presented normally using the classical GC-Marker, the new FIA-MS ether-lipid marker PC ae 36:4 was clearly decreased.

Table 12 ZSS patients with irregularly high GC-MS plasmalogen markers can still be identified as plasmalogen deficient by the FIA-MS marker PC ae 36:4. Red blood cell membrane plasmalogen from PEX1 patients was analyzed using GC-MS, and serum from the same sample was analyzed using the Biocrates AbsoluteIDQ p180 kit. For the reference range, a 95% confidence interval from 28 controls was used.

Patient	C16:0-diacetal (ratio to 16:0 FA)		C18:0-diacetal (ratio to 18:0 FA)		PC ae 36:4 (µM)	
	Subject	Reference range	Subject	Reference-range	Subject	95% confidence
ZSS_1	5.2	6.8-11.9	11.2	10.6-24.9	0.801	7.24-11.32
ZSS_2	9.0	6.8-11.9	20.4	10.6-24.9	0.619	7.24-11.32
ZSS_3	6.9	6.8-11.9	15.4	10.6-24.9	3.69	7.24-11.32
ZSS_4	7.3	6.8-11.9	18.0	10.6-24.9	1.25	7.24-11.32

Typically, analysis of the ether-lipid markers C16- and C18-diacetal (with GC-MS) was only possible by isolation of erythrocyte membranes, as sensitivity and reproducibility from serum were not sufficient (Björkhem et al., 1986). However, these membranes cannot be stored for long periods and are also not stored in biobanks. Thus, analysis of larger cohorts or retrospective analysis of a patient is not easily possible. Our study and another study by Herzog et al. (Herzog et al., 2016) showed that analysis of ether-lipids (including plasmalogens) from serum is indeed possible, when directly analyzing specific ether-lipids. This is possible even in samples older than ten years. To our knowledge, however, our study is the first to characterize ether-lipids as ZSS biomarkers from serum using an FIA-MS-based method.

4.3.3 VLCFA-linked lipids are increased in ZSS and X-ALD, but sphingomyelin is only decreased in ZSS patients

In ZSS patients most sphingomyelin species (such as SM 16:0,16:1,18:0,18:1, 24:0,24:1) were decreased, but SM 26:0 with an even longer acyl-chain length was increased compared to controls (**Table 13**). While SM 26:0 was also increased in X-ALD patients, the shorter sphingomyelins were not reduced.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

Sphingomyelins are not currently used as biomarkers for peroxisomal disorders, but untargeted lipidomics studies in blood also have found the reduction of SM (SM 16:0,16:1,18:1,22:0, 24:1; (Wangler et al., 2018) and a study in fibroblasts the accumulation of SM 26:0 (Herzog et al., 2016).

Table 13 VLCFA sphingomyelin species are elevated in the serum of X-ALD and ZWS patients, non-VLCFA sphingomyelin species are reduced in ZSS patients compared to controls. Serum from patients analyzed using the Biocrates AbsoluteIDQ p180 kit. Concentrations of sphingomyelin species (SM) in μM , n= 36 (X-ALD/ABCD1), 28 (Controls), 17 (ZSS/ PEX1).

	Mean serum concentration (μM)			SEM (μM)		
	X-ALD	Control	ZWS	X-ALD	Control	ZWS
SM C16:0	113.98	131.31	62.73	4.44	14.44	5.68
SM C16:1	15.72	15.41	6.61	0.57	1.12	0.63
SM C18:0	14.70	29.81	7.93	0.72	4.22	1.19
SM C18:1	8.35	13.57	4.23	0.44	1.34	0.75
SM C20:2	0.48	0.56	0.19	0.04	0.13	0.03
SM C24:0	33.97	28.31	13.39	1.46	2.15	1.28
SM C24:1	59.63	68.97	30.85	2.79	6.11	3.14
SM C26:0	0.84	0.28	0.81	0.07	0.03	0.14
SM C26:1	1.27	0.56	1.45	0.09	0.06	0.29

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

Additionally, lysoPC 26:0 was increased in both peroxisomal disorders (**Figure 45**). The VLCFA-lysophosphatidylcholine lysoPC 26:0 is used as a marker for X-ALD in newborn screening and is commonly found in lipidomics studies (Herzog et al., 2018a, 2018b). Increases in phosphatidylcholines with VLCFA were found in other studies but were not measured using this kit.

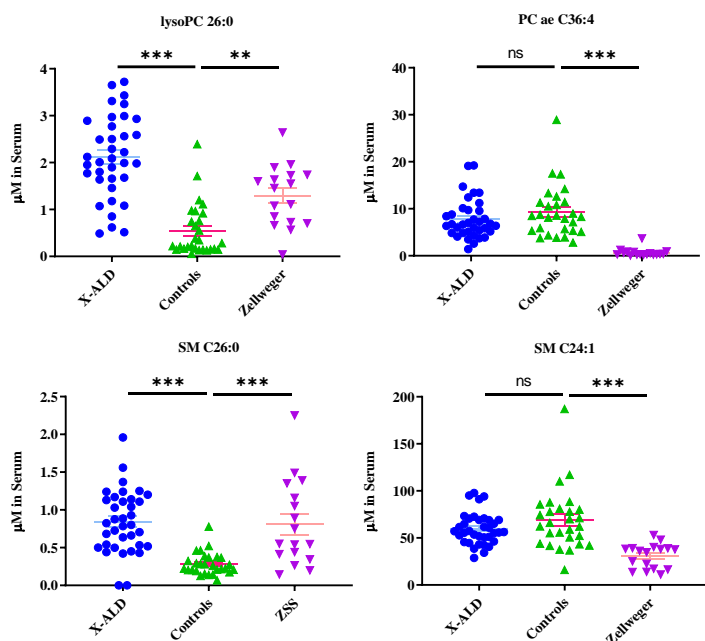


Figure 45 Membrane lipids indicate differential phenotype between Zellweger syndrome spectrum patients (ZSS) and X-linked adrenoleukodystrophy (X-ALD) patients, as well as controls. Mean±SEM, n= 36 (X-ALD/ABCD1), 28 (Controls), 17 (ZSS/ PEX1). *: p<0.05, **: p<0.01, ***:<0.001 after 1-way ANOVA multiple comparisons analysis, Tukey correction. lysoPC: lysophosphatidylcholine; PC aa: phosphatidylcholines with two acyl-linkages; PC ae: phosphatidylcholine with one acyl and one ether linkage; SM: sphingomyelin

4.3.4 D-bifunctional protein deficiency type III patients have altered peroxisomal biomarkers

As we were able to characterize two different peroxisomal disorders successfully, we were interested in getting an insight into another peroxisomal disorder, D-bifunctional protein deficiency type III (D-PBIII). We analyzed two patients with this disorder, where one of the patients provided two samples. Here we saw increased levels of lysoPC 26:0 (patient 1 with 1.33 μM and 5.86 μM; patient 2 with 3.97 μM) compared to controls (0.54±0.15 μM), similar to the other peroxisomal disorders. The ether lipid PC ae 36:4 was slightly decreased in the patients (5.66 μM and 5.32 μM in patient 1, 2.82 μM in patient 2 and 9.28±1.04 μM in controls). While a decrease of ether-lipids is unusual for DPBIII (Ferdinandusse et al., 2006) and points in the direction of ZSS, an even more substantial reduction was found in GC-MS plasmalogens analysis (**Table 14**). Even though a decrease of ether-lipids was also seen in PC ae 36:4 analysis, the separation to ZSS patients (0.67±0.2 μM) was considerably more significant than in GC-MS analysis. This may allow for superior identification of D-PBIII, especially from ZSS patients.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

Table 14 D-BPIII patient with irregularly low GC-MS plasmalogen markers can still be identified as different to ZSS patients, but also controls, by the FIA-MS marker PC ae 36:4. Red blood cell membrane plasmalogen from D-BPIII patient was analyzed using GC-MS, and serum from the same sample was analyzed using the Biocrates AbsoluteIDQ p180 kit. For reference range, a 95% confidence interval from 28 controls was used.

Patient	C16:0-diacetal (ratio to 16:0 FA)		C18:0-diacetal (ratio to 18:0 FA)		PC ae 36:4 (µM)	
	Subject	Reference range	Subject	Reference range	Subject	95% confidence
DBIII_1	4.3	6.8-11.9	6.3	10.6-24.9	5.32	7.24-11.32
DBIII_2	12.8	6.8-11.9	15.0	10.6-24.9	2.82	7.24-11.32

Similarly, to ZSS, SM C24:1 was reduced in D-BPIII patients (Patient 1: 21.2 µM and 35.5 µM; Patient 2: 32.6 µM) and SM 26:0 increased (Patient 1: 0.525 µM and 1.13 µM; Patient 2: 1.08 µM).

To summarize, we found an increase in VLCFA-lipids lysoPC 26:0 and SM26:0, as well as a reduction in SM 24:1, similar to ZSS, being in accordance with another lipidomics study (Herzog et al., 2018a) in fibroblasts. Additionally, we also found a slight decrease in the ether-lipid PC ae 36:4, which was not as strong as the decrease in ZSS.

4.3.5 Summary of results in peroxisomal disorders.

Using the commercial kit, we were able to generate metabolomics data in a fast, reproducible manner. We were able to show separate metabolomic profiles for X-ALD and ZSS while gaining the first insights into biomarkers for D-BPIII. Especially the ability to determine the ether-phospholipid marker PC ae C36:4 from plasma instead of erythrocyte membranes using FIA-MS may be beneficial for further research.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

4.4 Discussion: Commercial metabolomics platform may simplify peroxisomal disorder research

Recently, multiple milder or atypical cases of peroxisomal disorders with no or minor biochemical abnormalities using classical clinical laboratory markers were documented (Bacino et al., 2015; Mignarri et al., 2012; Zeharia et al., 2007). This underlines the need for additional biomarkers that also complement the use of modern genetics.

In the last years, untargeted metabolomics was used more commonly to study peroxisomal disorders in further detail. While successfully identifying potentially interesting biomarkers, resources such as equipment and human resources are limited in clinical laboratories. The establishment of untargeted metabolomics pipelines in those is thus rather difficult. To overcome this, we selected a commercially available metabolomics method called the AbsoluteIDQ p180-kit, enabling the user to quantify up to 187 metabolites. This kit requires knowledge of mass spectrometry but provides all methods and reagents so that in-depth knowledge is not necessary to generate targeted metabolomics data. We evaluated the kit for its suitability to peroxisomal disorders research, using samples from patients that were already genetically diagnosed to be affected with the peroxisomal disorders ZSS, X-ALD, and D-BP111. Here we successfully were able to separate patients with ZSS and patients with X-ALD from controls and between each other. Especially in the case of ZSS, it is unsurprising that the reduction of peroxisomes leads to severe alterations in many pathways.

When inspecting the deregulated metabolites more closely, we find that especially (polyunsaturated) ether-phosphatidylcholines (PC ae) were decreased in the ZSS patients compared to controls. A lack of ether-phospholipids was also found by another study by Herzog et. Al (Herzog et al., 2016). Ether-lipids and especially plasmalogens have been long used as vital markers for the differentiation of peroxisomal disorders. These can also be seen in comparison to X-ALD, another peroxisomal disorder, that did not show a reduction in ether-phospholipids. Ether-lipid biosynthesis starts in the peroxisome by the action of alkyldihydroxyacetonphosphate synthase (ADHAPS), generating an alkyldihydroxyacetonphosphate (alkyl-DHAP) from an acyldihydroxyacetonphosphate (acyl-DHAP). Thus, an absence or dysfunction of peroxisomes alters ether-phospholipid levels (Hajra and Das, 1996; Wanders, 2004).

Interestingly, due to their lower plasma levels, plasmalogens are generally determined with GC-MS only from erythrocyte membranes (Björkhem et al., 1986). However, whole blood cannot be stored for more than a few days and is not stored in biobanks. Erythrocyte membranes thus always need to be prepared freshly from non-hemolyzed blood. This significantly hinders the retrospective analysis of older samples, for diagnostics but also research purposes. As the incidences of peroxisomal disorders are low (Suzuki et al., 1996), samples need to be stored for long periods to gain sufficient sample sizes for research.

Our analysis also found that in four of the 17 ZSS patients, the classical GC-MS plasmalogen markers C18 diacetal were within the reference range. In 3 patients even the other plasmalogen marker C16 diacetal was also normal. This hints at a significant number of misdiagnoses using these GC-MS markers alone.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

Albeit the GC-MS marker were unchanged, we found that the ether-phospholipid marker PC ae 36:4 was able to differentiate patients still as affected. This marker can be analyzed using faster methods and without the perilous derivatization methods needed for GC-MS.

We also found a majority of sphingomyelins decreased in ZSS (SM 16:0, SM 16:1, SM 18:0, SM 18:1, SM 24:0 and SM 24:1) but not in X-ALD. This effect also has been studied by previous untargeted lipidomics studies (Herzog et al., 2016; Wangler et al., 2018). Even though these sphingomyelins are decreased in patients, the very long fatty acid chain SM 26:0 is increased in ZSS as well as X-ALD. These concur with the increase in lysoPC 26:0 that we found. The rise of lysoPC 26:0 in peroxisomal disorders is well documented in literature (Hubbard et al., 2006) and is even used in the newborn screen of X-ALD (Vogel et al., 2015). A possible explanation for both phenomena would be the relative overrepresentation of VLCFA in those cells. As the peroxisomal oxidation of cerotic acid (FA 26:0) decreases, cerotic acid becomes severely overrepresented in the cells. Enzymes attaching/exchanging fatty acids to lipids, may therefore face a changed equilibrium, forcing more cerotic acid into these lipids. However, not much is known about the specificity of these enzymes. Examples for enzymes partaking in this equilibrium shift may be glycerolphosphate-acyltransferases (GPAT), responsible for the attachment of the acyl-group to phospholipids in the ER (Takeuchi and Reue, 2009) or lysophospholipid acyltransferases (LPLAT) in Land's Cycle (Lands, 1958).

Additionally, using this targeted metabolomics method, we were able to gain first insights into another single enzyme disorder, D-bifunctional protein deficiency type III (D-BPIII). Due to its very low prevalence, we were only able to include two patients in the study. Using this targeted assay, we found similar changes in lipids relating to impaired VLCFA metabolism. Similar to X-ALD and ZSS, we saw increases in lysoPC 26:0 as well as SM 26:0. The alteration of VLCFA lipids was also reported by other groups and is consistent with the lower activity of a main enzyme in the peroxisomal VLCFA metabolism, such as the D-bifunctional protein. Counterintuitively, we also found a drop of the plasmalogen markers C16:0-diacetal and C18:0-diacetal using the routine GC-MS diagnostics, which would be only expected from diseases such as ZSS affecting the ether-lipid metabolism or the peroxisome as a whole, but not of D-BPIII. We also found a decrease of PC ae 36:4, an ether lipid from the targeted metabolomics kit. However, this alteration was minor and less potent compared to ZSS. Patients of D-BPIII may therefore be well separated from ZSS patients. Due to their low incidence, the sample size with only 2 patients is meager. We theorized that the drop of ether lipids in D-BPIII might originate from a general disturbance of peroxisomal metabolism. While not described in serum so far, this effect has been described at least in patient fibroblasts (Herzog et al., 2018a).

X-ALD showed a considerably less altered lipid profile than ZSS, relating to less affected pathways. When we use the proposed set of markers and especially lysoPC 26:0 and SM 26:0, we are well able to detect X-ALD patients and differentiate them from controls. SM 24:1 and PC ae 36:4 are normal, differentiating X-ALD patients from ZSS patients. When analyzing lysoPC 26:0, we see a considerable overlap between controls and patients, typically not seen in the LC-MS-based analysis of lysoPC 26:0 in newborn screening (Hubbard et al., 2006). The loss of discriminative power may originate from mild drawbacks in the flow injection mass spectrometry (FIA-MS) method used for phospholipid analysis.

4.0 Applications: LC-MS based platform simplifies access to metabolomics for peroxisomal disorders

It has been described that the actual concentrations are slightly overestimated, especially in measurements at the lower concentration end (Natarajan et al., 2019). This may lead to impaired discrimination of controls and patients with slightly decreased (but still abnormally high) lysoPC 26:0 levels. Factors influencing here may be residual (lipid-) contamination in the mass spectrometer or isobaric compounds (compounds with the same mass/charge ratio) with very low intensities playing a role. As we mainly established this commercial kit as a quantitative screening method and we plan specific measurements of metabolites using e.g., LC-MS, this does not display a significant problem for our purposes.

To summarize, using the AbsoluteIDQ p180-kit for targeted metabolomics, we were quickly able to reproduce findings of critical diagnostic metabolites in peroxisomal disorders and even found new biomarkers for these disorders. One of those was the ether-lipid PC ae 36:4, which can be analyzed quickly from serum without long, laborious, and perilous GC-MS derivatization. Due to its ability to be analyzed from serum, long-term stored biobanking samples may now be analyzed.

Hence, this kit may be applied for extended screens for patients with unclear peroxisomal disorders. In case of an atypical finding, an in-depth metabolomic and genetic study may be well justified.

5.0 Applications: Targeted metabolomics in X-ALD

This work originates from an intra-institute collaboration with Dr. Matthias Kettwig. Parts of this work have been published as Kettwig M, Klemp H, Nessler S, Streit F, Krätzner R, Rosewich H, Gärtner J. Targeted metabolomics revealed changes in phospholipids during the development of neuroinflammation in Abcd1tm1Kds mice and X-linked adrenoleukodystrophy patients. J Inherit Metab Dis. 2021 Apr 14. doi: 10.1002/jimd.12389.

5.1. Introduction to Targeted metabolomics in X-ALD

X-linked adrenoleukodystrophy (X-ALD) is the most common peroxisomal disorder (Bezman and Moser, 1998). X-ALD most commonly leads to neuroinflammatory demyelination in the central and peripheral nervous systems (cerebral ALD, cALD; Berger and Gärtner, 2006). Other tissues such as the testis and adrenal gland are also affected and lead to an Addison-like phenotype. Molecularly, X-ALD originates from a mutation in the peroxisomal ATP cassette transporter 1 (*ABCD1*), most likely responsible for importing very long chain fatty acids in this organelle (Mosser et al., 1994).

Curiously, while almost all affected children develop adrenal dysfunctions, far from all develop cerebral symptoms, the inflammatory lesions cause is unknown (Berger et al., 2014). Also, the onset age of cALD varies widely. Hematopoietic stem cell therapy is the only treatment option after the onset of cALD, but treatment success heavily relies on early diagnosis as destroyed neural tissue mostly continues affecting the patient throughout life (Raymond et al., 2019). Due to the invasiveness, strong side effects, and low availability of donors, stem cell therapy is not considered without indications of cALD development.

X-ALD is part of the biochemical newborn screen (acyl-carnitine 26:0 or lysoPC 26:0; Vogel et al., 2015) and is diagnosed genetically. However, the only reliable diagnostic options for an outbreak of the cerebral form are regular MRI scans to find leukodystrophic lesions (Melhem et al., 2000). Even though biochemical markers can have near-perfect sensitivity for detecting the genotype (especially lysoPC 26:0; Hubbard et al., 2006), there is no correlation between cALD and concentration in known biomarkers. MRI scans, however, can be traumatic, may require anesthesia, as well as lead to deposits of MRI contrast agents, and can thus only be done with broad timeframes between measurements (Guo et al., 2018). Hence new biomarkers are severely needed, and multiple studies have tried to find alterations using genomics and lipidomics analysis (Hama et al., 2018; Richmond et al., 2020). Very recently neurofilament light chain (Nfl), a marker for neuronal damage, has also been found elevated in cALD patients, however many questions remain (Weinhofer et al., 2021). Even though these studies provided more profound insights into the general pathophysiology, a promising predictive marker (or marker combination) was not found.

As shown in the previous analysis of different peroxisomal disorders (chapter 4.0, p. 88), the AbsoluteIDQ p180-kit is a commercial kit-based approach for the study of up to 187 metabolites in a reproducible, fast assay. As we showed previously, the kit can be used for the characterization of peroxisomal disorders. Thus, it may also be applied to study biomarkers for disease progression in X-ALD.

5.0 Applications: Targeted metabolomics in X-ALD

To understand X-ALD disease progression better, an *ABCD1*-mutant mouse line (*Abcd1^{tm1Kds}*) was kindly provided by Dr. Kirby Smith. As *Abcd1^{tm1Kds}* mice do not develop a neuroinflammatory phenotype, experimental autoimmune encephalitis (EAE) was induced in some mice to simulate lesions typical to cALD.

Then we studied biomarkers of the unstimulated, as well as the EAE dependent phenotype in *Abcd1^{tm1Kds}* and wild type (C57BL/6) mice serum. We also analyzed brain tissue of the corresponding mice. Based on these results, we tried to identify whether some of the alterations may be reproduced in human patients.

5.2 Additional methods in X-ALD

Experiments involving mice and collecting clinical data from patients were kindly provided by Dr. Matthias Kettwig and are described in detail in our published article (Kettwig et al., 2021).

5.2.1 Generation of mice samples

Briefly, C57BL/6 mice were obtained by Charles River laboratories, *Abcd1*^{tm1Kds} mice were kindly provided by Dr. Kirby Smith (Lu et al., 1997). Mice were housed under SPF conditions in an animal facility of the Georg-August University of Göttingen, and experiments were performed according to the German animal protection law and permission by the necessary authorities. Due to the localization on the X-chromosome and predominantly male affected human patients, only male mice were used for analysis.

For the generation of experimental autoimmune encephalitis, 20-30 weeks old male mice were immunized with a peptide of myelin oligodendrocyte glycoprotein (MOG₃₅₋₅₅) in emulsion with complete Freund's adjuvant, substituted with M. tuberculosis strain H37Ra, mice then also received pertussis toxin (Nessler et al., 2007).

After the onset of EAE, serum for metabolomics analysis was collected. The terminally anesthetized animals were perfused transcardially with PBS and brain tissue removed. Brains were dissected into smaller pieces, accurately weighed, and homogenized in ethanol/phosphate buffer (85%/15%) using 6 µl buffer for 1 mg tissue. Debris was removed by centrifugation and supernatant stored for metabolomics analysis.

5.2.2 Human patient cohort

Human serum samples were collected from genetically confirmed X-ALD patients during a biannual routine MRI monitoring session. Samples were collected after written, informed consent and according to the Helsinki declaration and the local ethics board. Cerebral adrenoleukodystrophy was diagnosed by assessing white matter lesions using MRI (Loes et al., 1994). 15 patients with an absence of neurological symptoms (Addison only), 4 patients that developed neurological symptoms within 12 months after diagnosis (prior cALD), and 21 patients with cALD symptoms during the time of blood draw (cALD) were used. Additionally, we were able to include samples of all stages from 3 patients. Further clinical characteristics can be found in **Table 15**.

Table 15 Clinical characteristics of X-ALD patients with confirmed ABCD1 mutation.

Compound	Addison only	Prior cALD	cALD
Number of samples	15	4	21
Age (years)	11.7 (5.1-18.3)	5.5 (4.4-7.3)	8.3 (5.1-20.6)
Time after MRI changes (months)			4.1 (0.2-11.0)
LOES-Score			12.4 (1-24)

5.3. Results of targeted metabolomics in X-ALD

5.3.1 *Abcd1*^{tm1Kds} mice show changes in lipid metabolism of very long chain fatty acids

Currently, there are no described biomarkers for the detection of the onset of cerebral symptoms in X-ALD. Thus, this project aims to gain first insights into possible biomarkers of cALD onset using a mice-based emulation of the cALD phenotype and assess potential candidates in human patients further.

Firstly, we assessed the basal changes between the *Abcd1*^{tm1Kds} and wild-type mice without experimental autoimmune encephalitis to gain information on whether the metabolic profile of mice matches the profile of humans as described in the previous chapter (chapter 4.0, p. 88).

Similarly, as in humans, we especially found accumulations of very long chain fatty acid (VLCFA) containing membrane lipids such as lysoPC 26:0 and 26:1, as well as SM 26:0 and PC aa C42:2 (**Figure 46, Table 16**). PC aa C42:2 may represent either of the VLCFA-lipids PC aa 18:2_24:0 or PC aa 16:2_26:0 but cannot be determined further due to analytical specifications of the kit.

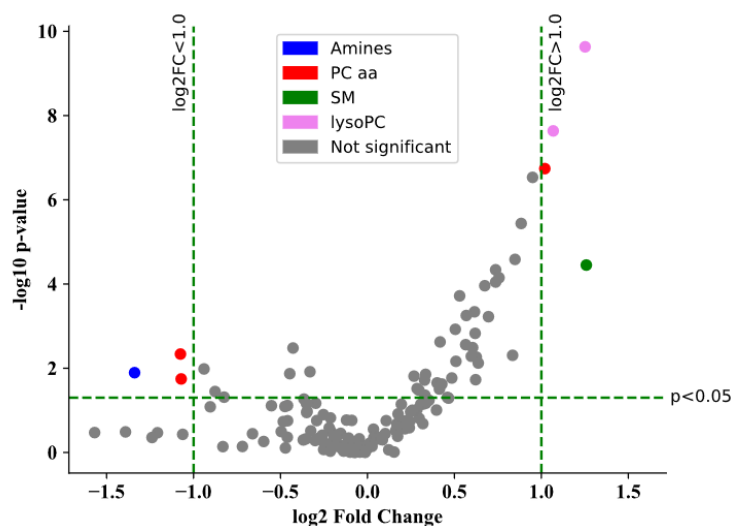


Figure 46 Volcano plot reveals accumulations of VLCFA lipids in *Abcd1*^{tm1Kds} mice compared to controls. Volcano plot showing the log₂ fold changes between n=19 *Abcd1*^{tm1Kds} and n=19 control mice serum of all 187 metabolites from the commercial kit, versus the -log₁₀ transformed adjusted p-value from Student's T-test.

5.0 Applications: Targeted metabolomics in X-ALD

As mentioned in previous sections, the alteration of VLCFA containing lipids is not unexpected due to their centrality of the *ABCD1* transporter in peroxisomal VLCFA degradation. We also found an alteration of the biogenic amine alpha-amino adipic acid (alpha-AAA) in mice serum, which was not detected in our human studies.

Table 16 Significant changes of serum metabolites in *Abcd1*^{tm1Kds} mice compared to controls. n=19 *Abcd1*^{tm1Kds} and n=19 control mice serum using metabolomics analysis of the commercial kit, versus the -log₁₀ transformed adjusted p-value from Student's T-test.

Compound	Raw fold change	Log ₂ fold change	Log ₁₀ p-value
lysoPC 26:0	2.381	1.252	9.632
lysoPC 26:1	2.097	1.068	7.636
PC aa C42:2	2.028	1.020	6.740
SM C26:0	2.393	1.259	4.450
PC aa C36:5	0.475	-1.075	2.337
alpha-AAA	0.395	-1.339	1.895
PC aa C34:4	0.476	-1.071	1.746

As there is insufficient information about the lipidome in brain tissue of *Abcd1*^{tm1Kds} mice, we analyzed brain tissue as well and found increases of lysoPC 26:0. However, SM C26:0 was not altered as for instance in the serum of mice or humans (**Table 17**). Additionally, we found lysoPC 24:0 and PC aa C42:1 to be altered. The slightly changed lipid profile may be caused by basal differences of brain and blood metabolome or by the brain lipid extraction process.

Table 17 Significant changes of metabolites in brain tissue of *Abcd1*^{tm1Kds} mice compared to controls. n=14 *Abcd1*^{tm1Kds} and n=14 control mice brain tissue using metabolomics analysis of the commercial kit, versus the -log₁₀ transformed adjusted p-value from Student's T-test.

Compound	Raw fold change	Log ₂ fold change	Log ₁₀ p-value
PC ae C36:3	0.655	-0.611	1.359
PC ae C38:3	0.661	-0.596	1.427
PC ae C38:2	0.662	-0.595	1.359
lysoPC a C24:0	1.444	0.530	1.476
lysoPC a C26:0	1.661	0.732	1.575
PC aa C42:1	1.932	0.950	1.779

The results of lysoPC 26:0 are especially relevant as this lipid is extensively used in newborn screening for X-ALD in humans (Vogel et al., 2015). In the previous chapter, we have also shown that lysoPC 26:0 can be analyzed by this metabolomics kit.

Even though the lysoPC 26:0 concentrations are significantly altered (**Figure 47**), contrary to previous studies in humans, we detected a slight overlap of wild type and *Abcd1*^{tm1Kds} mice. In brain tissue, this overlap is stronger than in serum, most likely due to minor errors introduced during the more extensive sample preparation.

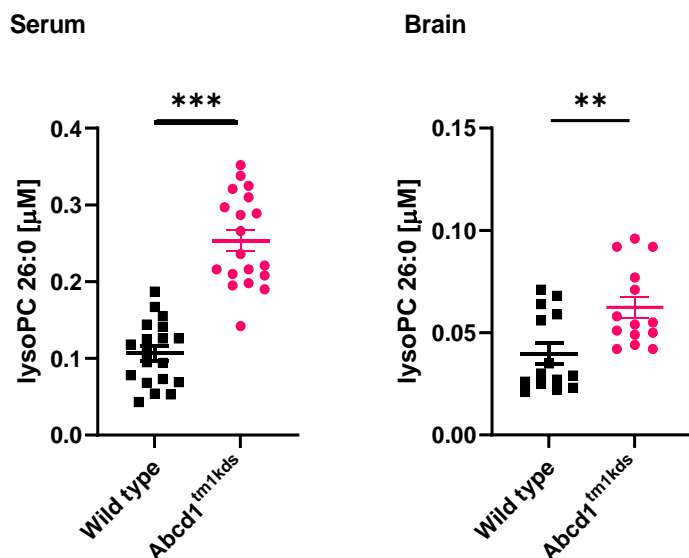


Figure 47 LysoPC 26:0 is increased in serum and brain tissue of *Abcd1*^{tm1Kds} mice. Analysis of serum or brain tissue using the commercial metabolomics kit. For serum $n=19$ *Abcd1*^{tm1Kds} and $n=19$ control mice and for brain tissue $n=14$ *Abcd1*^{tm1Kds} and $n=14$ control mice were used. P-values of Student's T-test are displayed as **: $p<0.01$ and ***: $p<0.001$.

5.3.2 Metabolomics of *Abcd1*^{tm1Kds} mice reveals differences between symptomatic and asymptomatic EAE animals.

Undesirably, *Abcd1*^{tm1Kds} mice do not develop spontaneous cALD-like white matter lesions. To introduce neuroinflammation, we used the experimental autoimmune encephalitis model (EAE). While there is no spontaneous onset of cALD in mice, we did find an earlier onset of EAE symptoms in *Abcd1*^{tm1Kds} mice (8 days versus 13 days after EAE) and a higher disease score (3.4 versus 2.1) compared to wild type mice. Moreover, *Abcd1*^{tm1Kds} showed encephalopathic symptoms and higher mortality uncommon for wild-type EAE mice. The results may indicate a more fulminant and more substantial inflammation in *Abcd1*^{tm1Kds} mice. Interestingly, in both groups, *Abcd1*^{tm1Kds} and wild-type mice, some did not develop EAE symptoms.

Next, we studied whether the *Abcd1^{tm1Kds}* mice show metabolome changes after stimulation by EAE and if animals that did not develop EAE symptoms are different from either group.

Using heatmap clustering analysis, we discovered that all three groups of *Abcd1^{tm1Kds}* mice (non-treated NT, treated without symptoms Tr-NS and treated with symptoms Tr-S) can be separated (**Figure 48**).

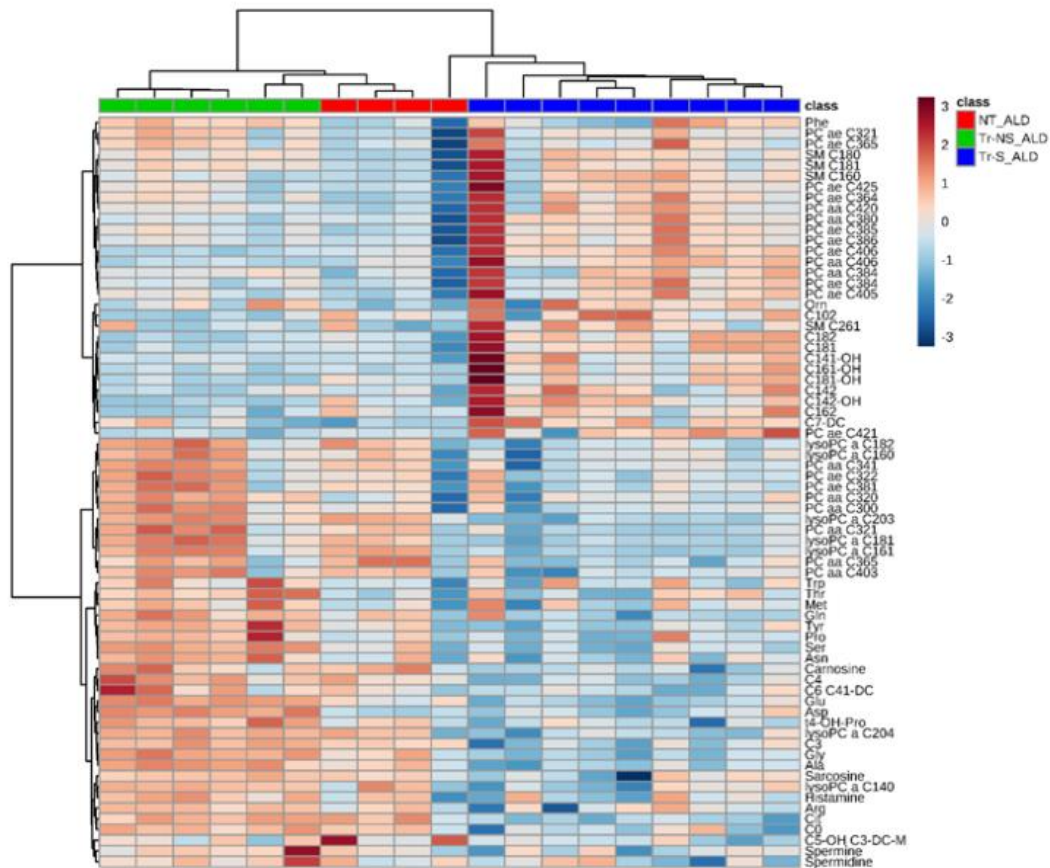


Figure 48 *Abcd1^{tm1Kds}* mice can be clustered based on different reactions to EAE. Heatmap clustering of the three different disease conditions in mice: not treated (NT, n=4), treated non- sick (Tr-NS, n=6), and treated sick (Tr-S, n=9) after determining the metabolome using the AbsoluteIDQ p180 kit. Using MetaboAnalyst, data were normalized with cube-root transform and auto-scaling. The 65 metabolites with the lowest ANOVA p-values were then clustered using Euclidian distance and ward algorithm.

Samples from animals without treatment and treatment but without symptoms clustered closely together, and symptomatic animals show a distinct grouping. A wide range of different metabolites, including lysoPC, carnitines, and sphingomyelins were dysregulated between these groups.

5.0 Applications: Targeted metabolomics in X-ALD

As some of the animals were treated but asymptomatic, we sought to look more closely at the difference between those and the symptomatic animals, as these might display metabolites valuable for monitoring disease progression. Thus, we selected a volcano plot to further assess markers in this 2-group comparison.

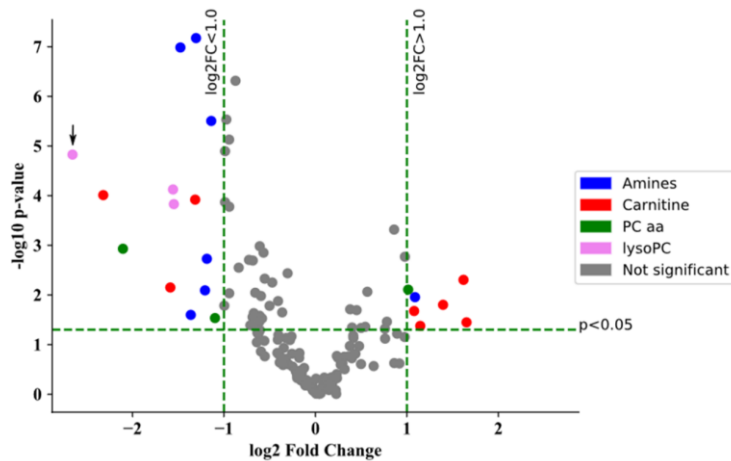


Figure 49 Volcano plot reveals metabolome differences in serum of symptomatic compared to asymptomatic *Abcd1^{tm1Kds}* EAE animals. Serum of treated non- sick (Tr-NS, n=6) and treated sick (Tr-S, n=9) were analyzed using the AbsoluteIDQ p180 kit. Using MetaboAnalyst, data was normalized with cube-root transform and auto-scaling, and \log_2 fold change, as well as \log_{10} p-value after Student's T-test, was plotted. Arrow marks the severely deregulated lipid lysoPC 20:3.

We found alterations of different metabolites, especially amines, carnitines, and (lyso-) phosphatidylcholines (**Figure 49**). While especially changes in amine concentration may reflect general health disturbances in affected mice, very curious was the decrease of the lysophosphatidylcholine lysoPC 20:3 (marked with an arrow in **Figure 49**).

The attached fatty acid 20:3 may represent mead acid or dihomogammalinolenic acid, both 20 carbon fatty acids, only differing positions of the three double bonds. Both fatty acids are essential metabolites in the eicosanoid metabolism, critical mediators of inflammation reactions (Sergeant et al., 2016). Given the changes in lysoPC 20:3, we were also interested if the lysoPC with the other important eicosanoid fatty acid, lysoPC 20:4, is also altered.

Indeed, we found both lysoPC 20:3 and 20:4 to be severely decreased in symptomatic EAE mice (Tr-S; **Figure 50**).

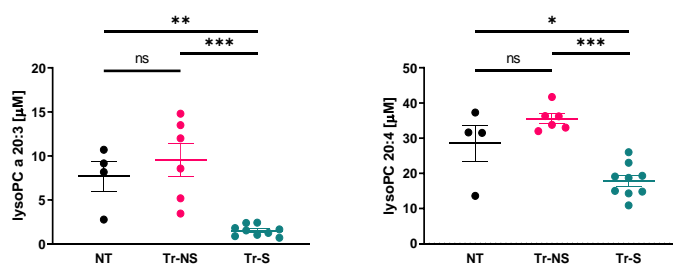


Figure 50 Two polyunsaturated fatty acid-containing lysophosphatidylcholines are reduced in symptomatic EAE *Abcd1^{tm1Kds}* mice. *Abcd1^{tm1Kds}* were either not treated (NT; n=4), treated with EAE and displayed no symptoms (Tr-NS; n=6) or treated with EAE and displayed symptoms (Tr-S; n=9). P-values of Student's T-test are displayed as **: $p < 0.01$ and ***: $p < 0.001$.

Interestingly, similar changes were also observed in wild type mice (**Figure 51**), which may point to the role of lysoPC 20:3 and 20:4 as markers for general (neuro-)inflammation, independent of the genotype, although the decrease in wild type mice was less severe as in symptomatic *Abcd1^{tm1Kds}* mice.

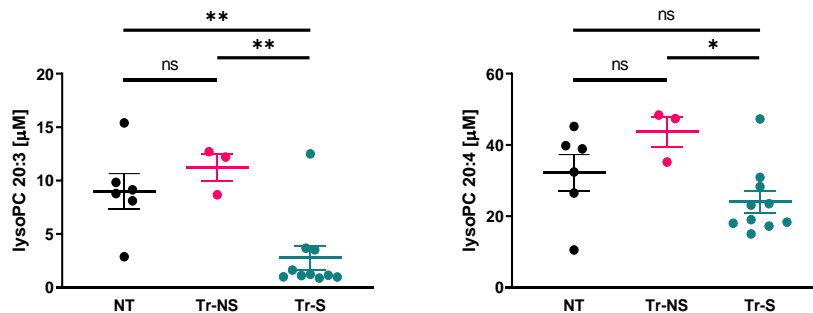


Figure 51 Two polyunsaturated fatty acid-containing lysophosphatidylcholines are reduced in symptomatic wild-type mice. Wild type was either not treated (NT; n=4), treated with EAE and displayed no symptoms (Tr-NS; n=6), or treated with EAE and displayed symptoms (Tr-S; n=9). P-values of Student's T-test are displayed as **: p<0.01 and ***: p<0.001.

Unfortunately, we were not able to show a correlation of lysoPC 20:3 with the EAE scores, but lysoPC 20:4 did correlate acceptably with the EAE scores, and even a linear regression was possible ($R^2=0.647$; **Figure 52**). A further evaluation in a study with a larger group size and a continuous and more quantitative marker than the EAE score should be done.

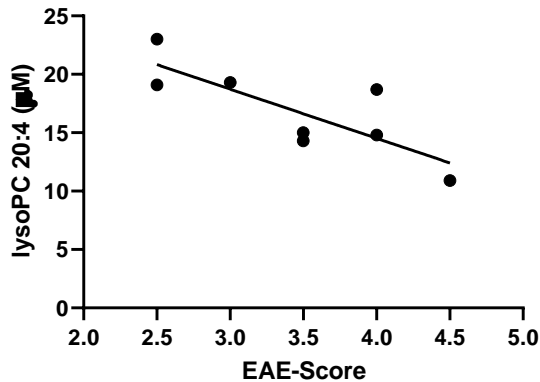


Figure 52 Linear regression of lysoPC 20:4 serum concentration to phenotypical EAE scores in EAE *Abcd1^{tm1Kds}* mice hints at a potential connection between lysoPC concentration and disease strenght. The serum concentration of lysoPC 20:4 was plotted against the phenotypical EAE score of symptomatic *Abcd1^{tm1Kds}* EAE mice (n=8). The EAE score is a phenotypical multiparameter score for assesing the strenght of EAE symptoms. Then linear regression was calculated: black line.

Even though the small group size prohibits further assumptions, this may further show that there might be a strong relationship between lysoPC 20:4 and cALD symptoms.

5.3.3 Alterations of lysoPC 20:3 and lysoPC 20:4 in human X-ALD patients

Based on the data of mice, we were interested in whether lysoPC 20:3 and 20:4 might also provide some insights into the inflammatory reactions in human cerebral X-ALD patients. Thus, we applied the metabolomics kit to three patient groups. Firstly, patients who did not display cALD symptoms and did not develop X-ALD in less than 12 months after the blood draw and only had mild adrenal/Addison-like deficiencies (Addison only). Secondly, patients that developed a cALD approximately up to 12 months after the blood draw (prior cALD). Thirdly, patients that presented with an active cALD at the time of blood draw (cALD).

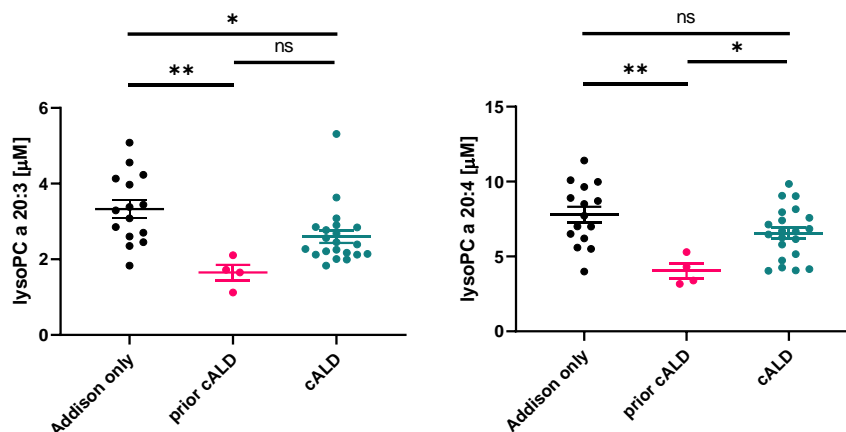


Figure 53 LysoPC 20:3 and 20:4 are decreased in human patients before and after cALD beginning.

Human X-ALD patient serum from patients without cALD (Addison only; n= 15), patients 6-12 months before begin of cALD (prior cALD;n= 4), and after beginning of cALD (cALD; n=21) was analyzed using the commercial kit. P-values of Student's T-test are displayed as *: p<0.05 and **: p<0.01.

In the human X-ALD patients, we saw the strongest decrease of both lipids in the patients prior to cALD (**Figure 53**). We were also able to observe a decrease between cALD patients and Addison-only patients in lysoPC 20:3. However, in 20:4, the effect was not significant. Interestingly, the concentration of those lipids seems to initially decrease in the patients before cALD and then increase again. In 3 cases, we were able to acquire data from each time point and saw a similar pattern of an initial decrease and long-term increase. LysoPC 20:3 or 20:4 do not appear to be linearly correlated to either the age of the patient (lysoPC 20:3 $R^2= 0.149$; lysoPC 20:4 $R^2=0.127$) or to the age of the sample (lysoPC 20:3 $R^2= 0.014$; lysoPC 20:4 $R^2=0.010$).

5.4. Discussion of targeted metabolomics in X-ALD

Unfortunately, despite multiple efforts with multi-omics techniques, no reliable prognostic biomarker for the onset of cerebral X-ALD (cALD) symptoms has been found (Hama et al., 2018; Richmond et al., 2020). The early detection of the onset of cALD is critical for starting the only therapy option (hematopoietic stem cell transplantation) before permanent neurological damage (Raymond et al., 2019).

Our project aimed to identify promising biomarker candidates in first-time use of the Biocrates AbsoluteIDQ p180-kit on a murine model of X-ALD before then transitioning to human patient samples.

Firstly, we decided to characterize the murine X-ALD model (*Abcd1^{tm1Kds}* mice). We tried to identify if the model recapitulated basal metabolomic patterns that we found in humans in our previous studies (chapter 4.0, p. 88). Here we chiefly found accumulations of very long chain fatty acids (VLCFA) containing lipids such as lysoPC 26:0. As established in previous sections, lysoPC 26:0 is an important marker, even used extensively in newborn screening. It possibly relates to the impaired VLCFA degradation by a peroxisomal import decrease to a defunct ABCD1 transporter (Mosser et al., 1994). The recapitulation of the metabolic changes speaks for at least some degree in the similarity of primary VLCFA metabolism in humans and the mouse model.

Unfortunately, the murine model does not show the same kind of white matter lesions as human patients with cALD. This curious phenotypic difference makes the study of cALD notoriously tricky. We decided to use the well-established model of experimental autoimmune encephalitis to at least partly simulate the reaction of X-ALD mice to a neuroimmunological stressor. Due to its artificialness, the EAE remains one of the most discussed protocols (Constantinescu et al., 2011). However, keeping the pitfalls in mind and with careful interpretation, it still may be used to study the cALD phenotype in mice partly. Interestingly in our project, *Abcd1^{tm1Kds}* mice were indeed more severely affected by EAE than wild-type mice, speaking for the validity of its use in X-ALD.

When comparing *Abcd1^{tm1Kds}* mice that were not affected and mice that were affected by EAE we curiously found the two 20-carbon lysophosphatidylcholines (lysoPC) with either 3 or 4 double bonds (lysoPC 20:3 and 20:4) severely decreased. In a preliminary analysis, the decrease of lysoPC 20:3 could even be linearly regressed to the EAE score, pointing to a potential connection to the disease strength. Later we proceeded to study whether these parameters are also altered in human X-ALD patients. Here, both lipids were especially decreased between the group without cALD and up to one year before cALD. The group of cALD patients displayed a slight increase of the lipids compared to the prior-cALD group but was lower than the non cALD group. This may indicate a change of the lipids with the start of pathomechanisms while later stabilizing after the strong onset of the symptoms. Commonly patients after cALD show symptom worsening compared to their previous state, yet they show an increase of both lipids. Their-increase may also point out that the 2 lipids are not simply altered by altered nutrition or health status.

However, some comparisons lack to be significant, and the sample size for the group before cALD is unfortunately very low, making definite conclusions difficult. Additionally, the timepoint of blood draw after symptom onset in mice and humans strongly deviated.

While in mice, symptom progression was observed daily, and three days after symptom appearance blood was directly taken, the sampling point in human strongly varied. Due to the long term toxicity of contrast agents and psychological stress, MRI was done biannually and blood was only taken in this session to relate to the MRI scans directly. To properly assess whether lysoPC 20:3 and 20:4 may be markers that allow for monitoring or even predict progressing symptoms, a shorter time interval between neurologic/neuroradiologic measurements is necessary.

Not much is known about the pathophysiological relevance and metabolism of lysoPC 20:3 and 20:4, including the decrease of these metabolites. In blood, the largest amount of lysoPC is formed by the action of lecithin cholesterol acyltransferase and phospholipase A2 from phosphatidylcholine (Steinman and Zamvil, 2005). The underrepresentation of fatty acids 20:3 and 20:4 may originate from their increased uptake by tissues/cells. Phospholipids containing those fatty acids are decreased and as a consequence also lysoPCs are reduced as well. The use of arachidonic acid from cellular phospholipids is widely known, and the remodeling from PC and lysoPC have been shown as well (Okuno et al., 2018; Weller, 2016). While different models for fatty acid transport in tissues exist (such as the flip/flop models or others (Dhopeshwarkar and Mead, 1973; Hamilton and Brunaldi, 2007; Pifferi et al., 2021) details especially regarding transport in metabolically restricted regions such as the brain, remain unclear. Interestingly transporters such as Mfsd2a were shown to contribute majorly to the transport of polyunsaturated fatty acids esterified as lysoPC (Chan et al., 2018; Wong and Silver, 2020).

No matter the exact process, the decrease of lysoPCs may relate strongly to the decline of their corresponding fatty acid. The reduction of the fatty acid may then be caused by either limited uptake/production or more substantial use of the fatty acid.

Unfortunately, the analytical specifications using flow injection mass spectrometry do not allow us to gain information on the specific position of the double bonds. In turn, this does not allow us directly to identify the corresponding fatty acid. However, there is a wealth of knowledge on both fatty acids 20:3 and 20:4. No matter the exact position, both are polyunsaturated fatty acids and eicosanoic acids, taking part in different reactions in the eicosanoid metabolism (Hanna and Hafez, 2018; Sergeant et al., 2016). Eicosanoids are potent cytokines that are primarily pro-inflammatory but can also be anti-inflammatory. Belonging to this are import mediators such as prostaglandins, thromboxanes and hydroxyeicosatetraenoic acids (HETE).

Potential isomers represented by 20:3 could be mead acid or dihomo- γ linolenic acid. Dihomo- γ linolenic acid is an elongation product of γ linolenic acid and generally leads to the production of anti-inflammatory mediators (Hsu et al., 2013; Johnson et al., 1997). Though dihomo- γ linolenic acid is usually very low in serum, its detection and relevance are improbable. More probable would be mead acid as an isomer of 20:3, being a fatty acid used to synthesize inflammatory mediators (Ichi et al., 2014). Interestingly mead acid is preferentially used when there is a relative insufficiency of arachidonic acid, a potential isomer of 20:4. There is another positional isomer of arachidonic acid. However, it is comparatively very low and does not partake in reactions. Arachidonic acid is the main crucial fatty acid in eicosanoid metabolism.

5.0 Applications: Targeted metabolomics in X-ALD

Curiously, the mainly described source of arachidonic acid in cells for eicosanoid production is phospholipase A2, and this enzyme is even involved in the regulation of eicosanoid production (Clark et al., 1991; Hanna and Hafez, 2018; Murphy and Folco, 2019).

In turn, these inflammatory mediators may then mediate the activity of immunologically active cells and steer neuroinflammation. This described process should not be unique to X-ALD but may be relevant to more neuroinflammatory diseases. Interestingly the connection of peroxisomal dysfunction increased PLA2, an increase of arachidonic acid in tissue has already been shown to lead to axonal loss and even neuroinflammation (Kassmann and Nave, 2008; Kassmann et al., 2007). Eicosanoids such as Leukotrienes have already been found to be increased in cerebrospinal fluid (Mayatepek et al., 2003). More compelling arguments are also provided by Ruiz et al. that discovered alteration of glycerophospholipids and PLA2 in *Abcd1* deficient mouse spinal cord that was associated with increases in the proinflammatory cytokines IL6 and IL8 (Ruiz et al., 2015)

Unfortunately, eicosanoids themselves are unstable molecules mostly acting as cytokines in the local environment (Hanna and Hafez, 2018). Thus, asserting long-term changes in the blood are not measurable. One previously not studied possibility may be the monitoring of eicosanoid metabolites (such as glucuronidated markers) in urine, which may provide a long-term overview of eicosanoid metabolism.

In summary, using a targeted metabolomics assay, we were successfully able to identify lysoPC 20:3 and 20:4 as neuroinflammatory serum biomarkers in an EAE X-ALD mouse model. We were also able to show that these biomarkers are altered in humans and may provide further insight into pathophysiology.

6.0 Applications: A human patient with an ELOVL1-mutation

This work originates from an intra-institute collaboration with Dr. Keit Men Wong and is part of a larger study of ELOVL1.

6.1 Introduction to a patient with a mutation in the ELOVL1 gene

One of the main reasons for developing a metabolomics platform was to understand further and optimize the diagnosis of rare neuropediatric disorders. Apart from characterizing our pipeline from a pure methodological point of view, we are intensely interested in which diseases can be studied with our untargeted metabolomics platform. After introducing metabolomics analysis using the AbsoluteIDQ p180-kit, we expanded our projects to use the untargeted lipidomics platform. In rare neurometabolic disorders, correctly diagnosing single-patient case studies is an integral part of the aims. Here we would like to present one specific patient case.

A pediatric patient was presented to Prof. Dr. med P. Huppke in the university medical center Göttingen with a neurologic and dermatologic phenotype including a primary developmental delay, mental retardation, extrapyramidal movement disorder, hypomyelination, and ichthyosis. Using whole-genome sequencing, the patient was found to have a mutation in introducing a premature stop-codon in exon 6 (p.Trp154*) in the very long fatty acid elongase gene 1 (ELOVL1), leading to losing a significant protein part. The only commercial antibody, which is directed against the c-terminal, does not detect the shortened protein in the western blot. Furthermore, the symptoms were similar to an ELOVL1 patient case study previously described by the group of Professor Płoski (Kutkowska-Kaźmierczak et al., 2018). In this study, both patients were found to have a heterozygous p.Ser165Phe mutation in ELOVL1. Using gas chromatography mass spectrometry in patient fibroblasts as well as a transfected HEK293 cell line with the mutation, ELOVL1 mutants were shown to have a strong reduction of the very long chain fatty acid (VLFCFA) cerotic acid (fatty acid, FA 26:0). Reductions of other VLFCAs FA 24:0-28:0 were found, however, to a milder extent. The serum VLFCFA of the patients were normal, but the 24:0/22:0-ratio was decreased. In an in vitro-assay, reduced ELOVL1 function was then shown as well (Mueller et al., 2019).

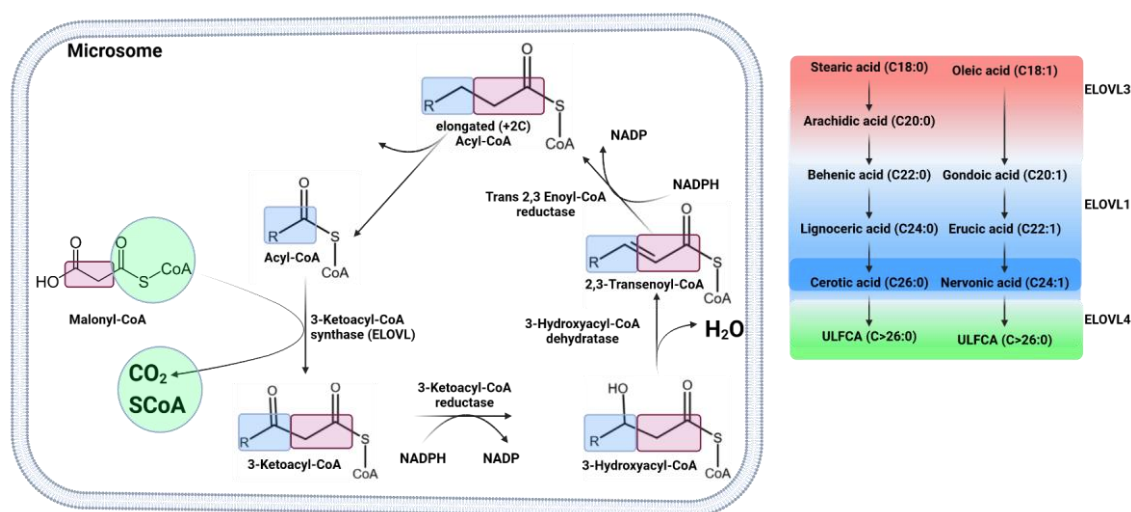


Figure 54 Very long chain fatty acid elongation cycle and ELOVL specificity. Created with BioRender.com

6.0 Applications: A human patient with an ELOVL1-mutation

The endproducts of de-novo synthesis of fatty acids using fatty acid synthase are primarily palmitate (16:0) and stearate (18:0) (Heil et al., 2019; Smith, 1994). However, longer fatty acids have to be synthesized using the fatty acid elongation cycle (**Figure 54**). The very long fatty acid elongase enzymes or 3-keto acyl-CoA synthases (ELOVL) are enzymes of the endoplasmatic reticulum/microsome (Jakobsson et al., 2006). Overall in every passing of the cycle 2 carbon-atoms are added by condensing one molecule of malonyl-CoA to an acyl-CoA (Leonard et al., 2004). The initial step of this reaction is catalyzed by the action of the ELOVL enzymes, leading to the generation of a 3-keto-acyl-CoA elongated by two carbon atoms and generation of CO₂. In the following reactions, the keto group is then removed by reduction, dehydration, and reduction of the formed enoyl-CoA. The process is similar to the fatty acid synthase. However, each reaction is carried out by separate enzymes (Ikeda et al., 2008; Moon and Horton, 2003). There are seven described ELOVL enzymes (ELOVL 1-7). Each of them shows separate specificities, but they are partly overlapping (Ohno et al., 2010; Wang et al., 2005). Not all enzymes are expressed in all tissues, and some are restricted to a specific tissue.

ELOVL1 is ubiquitously expressed, though, in most tissues, it is supplemented by ELOVL3 (Westerberg et al., 2004) or ELOVL4 (Hopiavuori et al., 2019; Vasireddy et al., 2007). The fatty acid elongation by ELOVL1 starts at C18:0, but mostly C22:0, C24:0, C26:0 are catalyzed (Ohno et al., 2010). The same work by Ohno et al. also showed that lipids containing VLFCA, especially 24:0-ceramides, were markedly reduced. An ELOVL1-mutant (ELOVL1-Mt) knockout mouse model showed increased epidermal permeability and a reduction of VLFCA (>26:0)-ceramides (Cer) and of VLFCA sphingomyelins (>24:0) in mice skin (Sassa et al., 2013). In the same mouse model, a reduction of brain VLFCA sphingolipids was found, including sphingomyelins, ceramide, galactosyl-ceramide, sulfatides, and others (Isokawa et al., 2019). These mice experienced difficulties in the balance beam test, speaking for motor dysfunction, mild hypomyelination was seen by MRI, and less reaction in the acoustic startle test, pointing to hearing defects.

The patient of Prof. Huppke indeed experienced a similar skin phenotype and motor alterations. Combined with the shortening of the ELOVL1 protein, there is compelling evidence that this may indeed be disease-causing. However, to ascertain a direct link, a biochemical profiling technique such as metabolomics should be used.

In order to further validate this and to further understand the phenotype of this disorder, we applied our untargeted lipidomics pipeline. While there have been some patient studies on ELOV1 already, using this project, we also hope to demonstrate the usefulness of our platform further.

6.2 Methods and cohorts

For analysis of the metabolome of an affected pediatric patient, we collected EDTA-plasma, as well as a skin biopsy for the generation of a primary fibroblast cell line. As a control group we collected EDTA-plasma from 3 family members without a metabolic phenotype. Written informed consent from patient and family members was confirmed, and all procedures were performed according to the Helsinki declaration and the local ethics board. Patient-derived samples were kindly provided by Prof. Dr. Huppke.

6.2.1 Cell culture methods

Dr. Keit Men Wong kindly provided cell culture experiments and samples. Briefly, primary fibroblast cell culture was obtained from a skin biopsy. Due to the invasiveness of the skin biopsy, for controls two established primary cell lines were used that did not display an metabolically altered phenotype. Cells were cultured at 37°C with 5% CO₂ in DMEM with 1 g/l glucose and 10% FCS. For experiments, 1x10⁶ cells were seeded in 150 mm plates and grown for at least 3 days. In FCS-withdrawal experiments, after seeding cells in 150 mm plates, cells were cultured 5 days in DMEM without FCS. Cells were lysed as described in the method section (chapter 2.2.3, p. 23) and all samples extracted and measured using the CortecsT3 lipidomics pipeline (chapter 2.2.10, p. 27).

6.3. Results of a human patient with an ELOVL1 mutation

6.3.1 Analysis of ELOVL1 patient plasma shows decreased VLFCA-SM

First, to get an overview of the alterations in the human patient, we analyzed EDTA-plasma using our untargeted lipidomics method. As an unbiased overview, we selected a presentation similar to the one seen in chapter 3.4.5 (p. 54). Here we found that if we present lipid signals as a plot of mass/charge-ratio to retention time, lipids can be structured based on their class, acyl-chain-length, and desaturation. ELOVL1 is an enzyme responsible for the elongation of very long chain fatty acids. Thus, a dysfunction should lead to a global reduction of lipids containing very long chain fatty acids (VLCFA). As, on average, longer chain lipids have a higher retention time and mass, reducing this lipid class should lead to a decrease in this specific region.

In patient plasma, we found a decrease of signals, especially in the range of 21-24 mins and a mass higher than 800 Da (**Figure 55**). However, also increases in a similar region are detectable. Given their retention time, they may also contain triglycerides that commonly also have a high mass and a late retention time given their three acyl groups. While certain effects of the ELOVL-1 mutation may be visible, they may be smaller due to compensatory impacts and nutrition. Unfortunately, to our knowledge previous patient studies have not analyzed the blood lipidome of ELOVL1-Mt patients so far, but only the fatty acid profile with GC-MS (Kutkowska-Kaźmierczak et al., 2018; Mueller et al., 2019). This group did not find any absolute alteration by GC-MS fatty acid profiling, only of the 24:0/22:0 ratio.

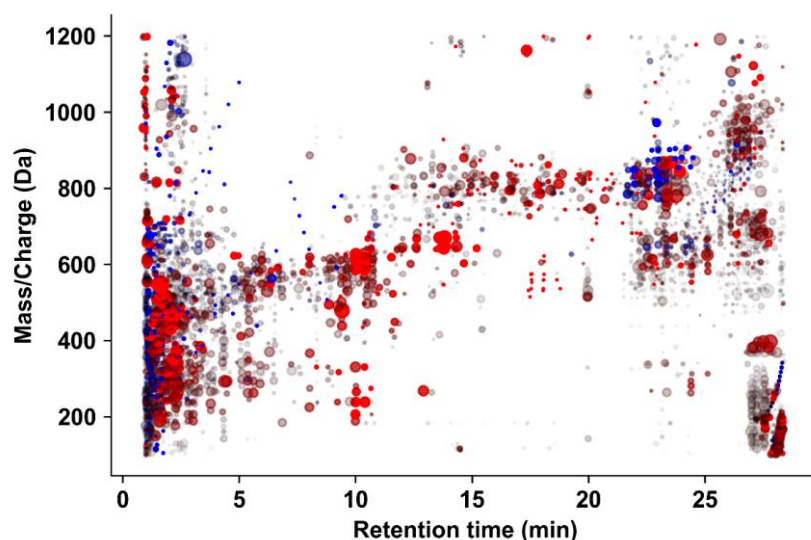


Figure 55 An overview of the plasma lipidome of a patient with an ELOVL1-mutation shows decreased signals in the high retention time and mass region. EDTA plasma of controls (3 family members without a metabolic disorder) and a patient with an ELOVL1 mutation (3 technical replicates) were analyzed using the CortecsT3 lipidomics method. Data were normalized using LOESS normalization and fold changes, and p-value by Student's T-test were calculated and corrected using Welsch correction. Then signals (features) were separated into groups of increased in patients (fold change > 1) and decreased (fold changes < 1) and normalized separately. Then metabolites are displayed at their specific retention time and mass/charge-ratio. Signals with an increased FC are represented in red, decreased signals in blue; the size of the circle depicts the negative decadic logarithm of the p-value (larger circles represent smaller p-values).

6.0 Applications: A human patient with an ELOVL1-mutation

We then identified corresponding lipids using fragmentation data and an untargeted lipidomics library (internal to MS Dial) to overview certain lipid classes that are more involved in the phenotype and if specific lipids are even more severely decreased.

Using a volcano-plot we were able to find certain lipids that were most aberrant regarding their log₂ fold change and their p-value (**Figure 56**). Among the reduced lipid classes were mostly sphingomyelins, ether-phosphatidylcholines, lysophosphatidylcholines, and diacylglycerols. Given their high importance in myelin and epithelia, a decrease in very long chain sphingomyelin may lead to severe complications. The reduction of VLFCFA sphingomyelins was found by the group of Prof. Kihara in tissue of (brain/skin-) tissue, but was not studied in blood (Isokawa et al., 2019; Sassa et al., 2013).

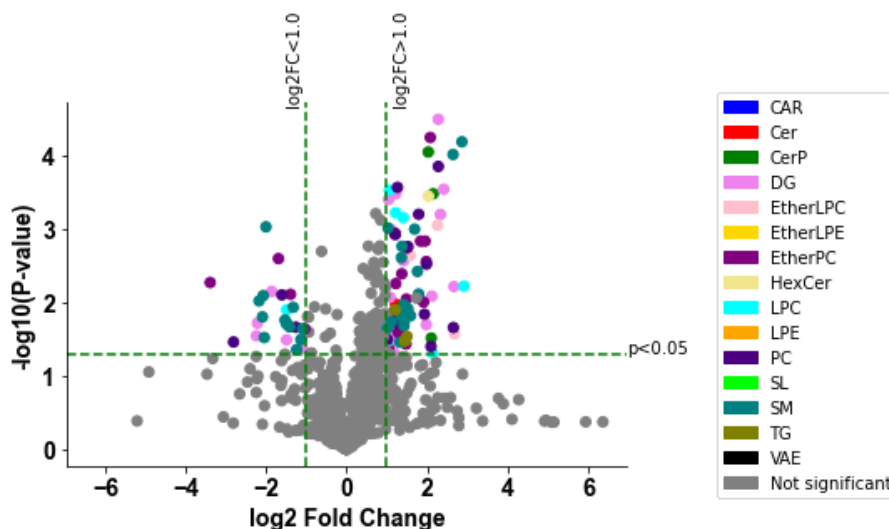


Figure 56 Volcano-Plot comparing ELOVL1-Mt patient to controls in plasma shows decreased membrane lipids. EDTA plasma of controls (3 family members without a metabolic disorder) and a patient with an ELOVL1 mutation (3 technical replicates) were analyzed using the CortecsT3 lipidomics method. Lipids were identified using mass and fragment mass and lipids without fragmentation information were excluded. Then fold change and p-value were calculated using Student's T-test with Welsch correction.

6.0 Applications: A human patient with an ELOVL1-mutation

One of the dysregulated sphingomyelins was SM C24:1 (**Figure 57**). The sphingomyelin that is associated with the 24-carbon fatty acid nervonic acid is of high interest, as nervonic acid is highly abundant among VLFCA-membrane lipids (Bourre et al., 1986). Therefore, alterations in this lipid may be a good marker for the ELOVL1 deficient phenotype.

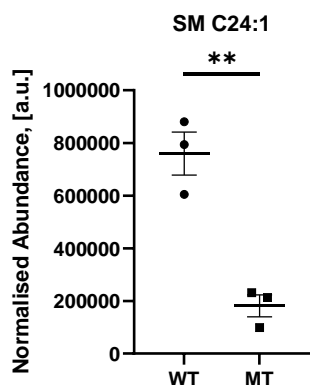


Figure 57 SM C24:1 is decreased in ELOVL1-Mt patient plasma. EDTA plasma of controls (WT; 3 family members without a metabolic disorder) and a patient with an ELOVL1 mutation (MT; 3 technical replicates) were analyzed using the CortecsT3 lipidomics method. Lipids were identified using mass and fragment mass and lipids without fragmentation information excluded. P-values of Student's T-test are displayed as **: $p < 0.01$.

6.3.2 Analysis of ELOV1 patient plasma using the targeted metabolomics kit confirms reduction of membrane lipid species

Our untargeted lipidomics method is a semiquantitative method and is expected to have less linearity than targeted mass spectrometry approaches. Thus, we aimed to compare our lipidomic method with the international ring-trial validated Absolute IDQ p180-kit. As mentioned in previous chapters, this kit can analyze up to 187 metabolites in a targeted and quantitative fashion. In this analysis, we used the same samples as in trials before.

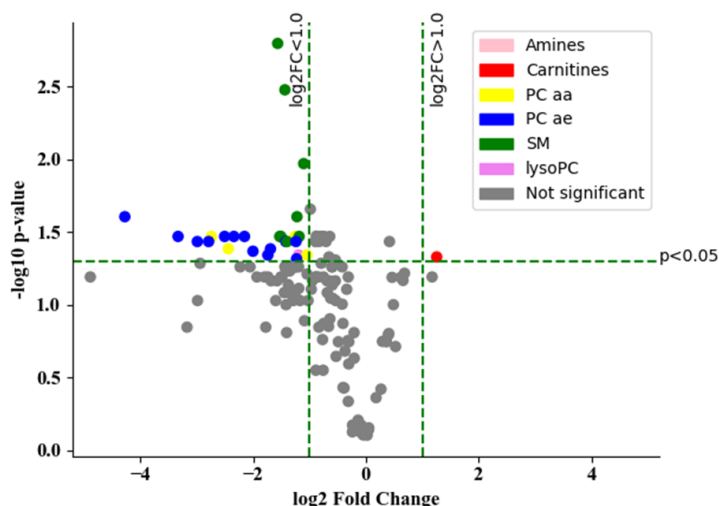


Figure 58 Volcano-Plot comparing ELOVL1-Mt patient plasma to controls using the targeted metabolomics kit reveals mostly reduction in tested membrane lipids. EDTA plasma of controls (3 family members without a metabolic disorder) and a patient with an ELOVL1 mutation (3 technical replicates) were analyzed using the Biocrates AbsoluteIDQ p180-kit. Then fold change and p-value were calculated using Student's T-test with Welsch correction.

6.0 Applications: A human patient with an ELOVL1-mutation

When using the kit, we similarly found alterations in various lipids, including phosphatidylcholines (PC aa), ether-phosphatidylcholines (PC ae), and chiefly a decrease in various sphingomyelins (SM; **Figure 58**). One of the strongly decreased sphingomyelins was SM C24:1 (**Figure 59**), confirming the results from the untargeted lipidomics study. In regard to SM C24:1 the fold changes with 0.24 (untargeted method) and 0.34 (targeted method) are quite similar. Only the variation (standard deviation in % mean) are higher in the untargeted method with 18.5 % (controls) than in the targeted method 8.4 % (controls).

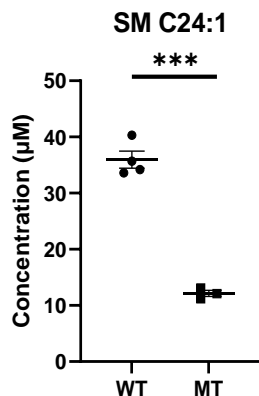


Figure 59 SM C24:1 is altered in patient ELOVL1-deficient patient plasma. EDTA plasma of controls (3 family members without a metabolic disorder) and a patient with an ELOVL1 mutation (3 technical replicates) were analysed using the Biocrates AbsoluteIDQ p180-kit. P-values of Student's t-test are displayed as and ***: $p < 0.001$

Additionally, to the long-chain sphingomyelins, several other sphingomyelins are also decreased in the patient, including SM 16:0 (**Figure 60**). We did not find those metabolites dysregulated using the untargeted method, possibly due to a lower overall sensitivity and higher variation. This is an unfortunate but well-known side effect of untargeted screening approaches (Roberts et al., 2012). However, the targeted method by Isokawa et al. also did not find this reduction in tissue. Tissue phenotype often does not match blood phenotype as seen, e.g., in the *Abcd1^{Kds}* mice brains in the previous chapter (chapter 5.3.1, p. 102).

Untargeted method: CortecsT3-RPLC-MS Targeted method: AbsoluteIDQ p-180-kit

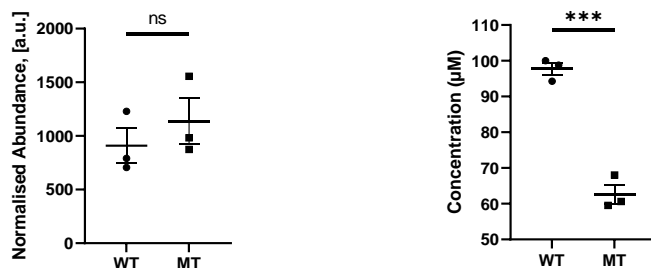


Figure 60 Sphingomyelin SM C16:0 is found decreased in ELOVL1-Mt plasma using the p180-kit but not using the untargeted lipidomics method. EDTA plasma of controls (WT; 3 family members without a metabolic disorder) and a patient with an ELOVL1 mutation (MT; 3 technical replicates) were analyzed using the Biocrates AbsoluteIDQ p180-kit and the CortecsT3 Lipidomics method. P-values of Student's T-test are displayed as and ***: $p < 0.001$.

6.3.3 ELOVL1 patient fibroblasts lipid profile reacts to external fatty acid withdrawal

To further analyze the patient phenotype in a controlled environment and study the effect of different conditions, including possible supplementations or treatments, we generated a primary fibroblast cell model of this patient. Like the blood plasma study, we studied the global lipidomic changes using the mass/charge to retention time graph (**Figure 61**).

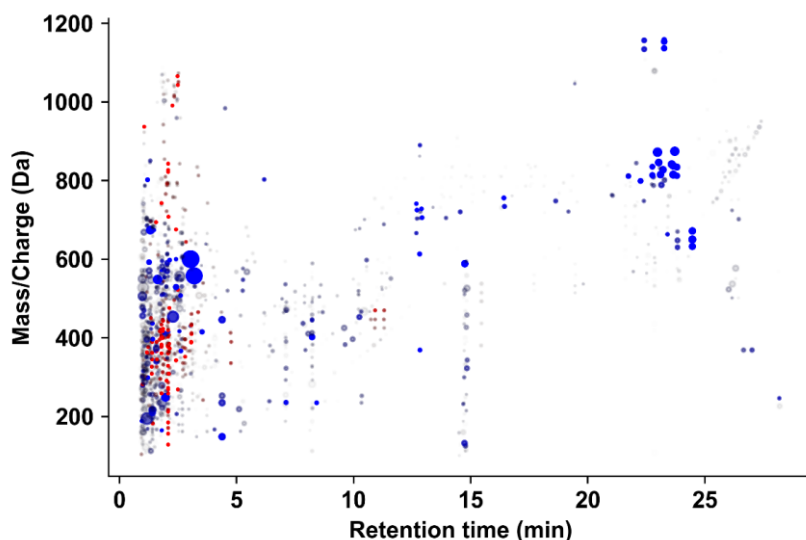


Figure 61 Overview of ELOVL1-Mt patient fibroblast lipidome compared to controls under standard culture conditions shows mildly decreased signals in the high retention time and high mass region.

Fibroblast cell lysate of controls (2 patients without a metabolic disorder with 3 cell culture replicates each) and a patient with an ELOVL1 mutation (3 cell culture replicates) were analyzed using the CortecsT3 lipidomics method. Data were normalized using LOESS normalization and fold changes, as well as p-value by Student's T test, were calculated and corrected using Welsch correction. Then signals (features) were separated into groups of increased in patients (fold change > 1) and decreased (fold change < 1) and normalized separately. Then metabolites are displayed at their specific retention time and mass/charge ratio. Signals with an increased FC are represented in red, decreased signals in blue, the size of the circle depicts the negative decadic logarithm of the p-value (larger circles represent smaller p-values).

In the patient-derived fibroblast cell lysates, we were able to show a more fulminant decrease of signals in the high mass and high retention time region.

6.0 Applications: A human patient with an ELOVL1-mutation

This can be increased even more when challenging the cell line with fetal calf serum (FCS) removal for three days (**Figure 62**).

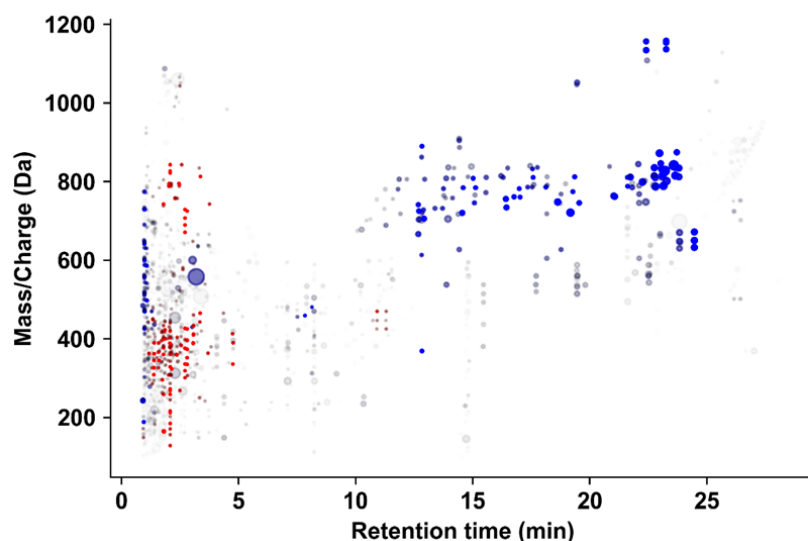


Figure 62 Overview of ELOVL1-Mt patient fibroblast lipidome compared to controls under FCS withdrawal reveals strongly decreased signals in the high retention time and high mass region

Fibroblast cell lysate of controls (2 patients without a metabolic disorder with 3 cell culture replicates each) and a patient with an ELOVL1 mutation (3 cell culture replicates) were treated by removal of FCS for 3 days and analyzed using the CortecsT3 lipidomics method. Data were normalized using LOESS normalization and fold changes, as well as p-values by Student's t test, were calculated and corrected using Welsch correction. Then signals were separated into groups of increased in patients (fold change > 1) and decreased (fold change < 1) and normalized separately. Then metabolites are displayed at their specific retention time and mass/charge ratio. Signals with an increased FC are represented in red, decreased signals in blue, the size of the circle depicts the negative decadic logarithm of the p-value (larger circles represent smaller p-values).

When removing the FCS for a short time period, we can see a clear decrease of signals in the high mass/ high retention time region compared to controls. Additionally, signals in the lower mass/ lower retention time region are increased. In cell culture, FCS is the source of (esterified) fatty acids (Bailey et al., 1972), including very long chain fatty acids. Thus, challenging the cells by removing their main source for fatty acids, they have to be relying on their own fatty acid synthesis and elongation (Yao et al., 2016b). Additionally, starving the cells of fatty acids also reduces the content of stored triglycerides that may have represented other signals in the high mass range.

6.0 Applications: A human patient with an ELOVL1-mutation

Given the results from the analysis of patient plasma, we were particularly interested, whether the very long chain sphingomyelin SM 24:1 was also altered in fibroblasts (**Figure 63**).

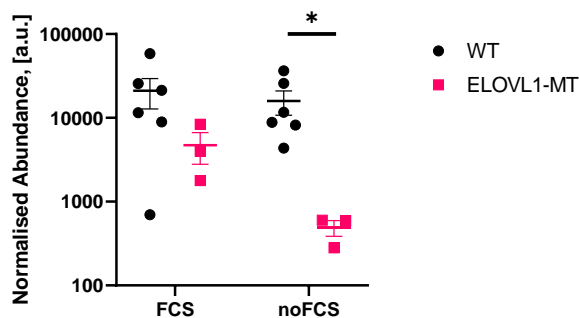


Figure 63 SM C24:1 deficient phenotype in ELOVL1-deficient cells can be exacerbated by FCS-challenge. Fibroblast cell lysate of controls (2 patients without a metabolic disorder with 3 cell culture replicates each) and a patient with an ELOVL1 mutation (3 cell culture replicates) were analyzed using the CortecsT3 Lipidomics method. Data were normalized using LOESS normalization and fold changes, as well as p-values by Student's T test, were calculated and corrected using Welsch correction. P-values of the Student's T-test are displayed as *: $p < 0.05$. For better display of data, normalized abundances are presented on a decadic logarithmic axis.

While under standard cell culture conditions, SM 24:1 in ELOVL1-Mt fibroblasts is slightly decreased compared to controls, the difference lacks to be significant. This may be related to some remaining esterified very long chain fatty acids inside the FCS. However, when challenging the cells by removing FCS, a dramatic decrease is visible.

In summary, using the CortecsT3 lipidomics method, we showed a decrease of high mass/high retention time signals in ELOVL1-deficient patient plasma and a patient-generated cell line. Withdrawing the patient-derived cells of a source of lipids (FCS) increased the effect strongly. Also, a reduction in VLFCA sphingomyelins was shown in patient plasma and fibroblasts. Altogether, these results point to an alteration of VLFCA metabolism as expected by a fatty acid elongation enzyme deficiency.

Concerning the metabolomics platform, we were able to show that the untargeted lipidomics platform can produce results matching with the results of other researchers. Additionally, the untargeted method can produce similar results as the quantitative, validated commercial kit approach.

6.4 Discussion: A mutation in the ELOVL1 gene leads to changes in the lipid profile of a human patient

The main goal of establishing this method platform is the ability to diagnose and understand neuropediatric diseases better. While we extensively characterized the methodological capacities of the method before (chapter 3.4.7, p. 61), a patient case enables us to show whether we achieved this goal.

In this project, we generated an untargeted metabolomic profile of a neuropediatric patient with a nonsense mutation in the very long fatty acid elongase 1 gene (Elov1). Elov1 is ubiquitously expressed and mainly responsible for synthesizing very long chain fatty acids such as 24:0, 24:1, and 26:0 (Ohno et al., 2010).

First, we analyzed the lipidome of patient plasma. In an unbiased overview, we first plotted liquid chromatography mass spectrometry (LC-MS) features based on their retention time and mass/charge ratio, as well as fold change and significance. Here we were able to identify a small but distinct cluster of downregulated metabolic features in the high mass and high retention time range. Lipids with very-long-chain fatty acids would be expected to be decreased in the patient due to the missing action of ELOVL1 (Isokawa et al., 2019; Sassa et al., 2013).

Those lipids (based on the experiments in previous chapters) are expected to be in the high mass and high retention time range. This may point to some form of alteration. However, apart from the down-regulated cluster, also other signals in a similar retention time range can be found. These may not represent VLCFA-lipids, but rather triacylglycerides, which due to their less polar head group and 3 acyl-groups (instead of 2 in phospholipids), have an increased retention time and mass (also compared to the characterization of the lipidomics method in chapter 3.4.5.1, p. 54). Another drawback of this analysis is the suitability of the control group, where we were only able to use plasma from 3 family members of different ages and sex. Due to the invasiveness of a blood draw, it can only occur in children if there is a medical necessity. Other groups have shown clear correlation between age (after adulthood), sex and lipidome (Slade et al., 2021). Although the correlation with age (at least after adulthood) is negligible and effects with appropriate size may still be interpretable, the accessibility of a proper control group is a common problem in pediatric research.

Using an untargeted library, we then identified the lipids bases on their mass/charge ratio and their fragments (where possible). In the two-group comparison, we found a variety of different lipids altered. This included sphingomyelins (SM), (Ether)-phosphatidylcholines (PC), Carnitines (Car) and others. Most curiously, we also found a decrease of SM 24:1, most likely nervonyl-sphingomyelin, in ELOVL1-Mt patient plasma. Interestingly FA 24:1 is one of the most abundant VLCFAs in cells, especially in myelin (Bourre et al., 1986). Fatty acid 24:1, nervonic acid, is synthesized by the elongation of fatty acid 22:1 (erucic acid) with Elov1, thus making SM24:1 decreases a likely result upon deactivation of Elov1.

SM 24:1 also has been found by another study to be one of the most severely altered metabolites (Isokawa et al., 2019). Apart from SM 24:1, we did not detect further altered VLCFA-lipids.

6.0 Applications: A human patient with an ELOVL1-mutation

VLFCA-lipids are primarily found in nervous tissue and skin (Sassa and Kihara, 2014), usually leaving only marginal concentrations of VLFCA-lipids in plasma (Moser et al., 1999). Due to their low concentration, it is reasonable to assume that we were not able to detect further alterations.

The type of mass spectrometers used by our untargeted metabolomics platform is optimized for screening, using a quadrupole time of flight mass spectrometer (QToF). In our specific setup, all ions leaving the liquid chromatography column enter the mass spectrometer detector at very similar times. Due to the high abundance of ions, also higher background noise forms, the lower signal/noise ratio then affects detection and linearity in the lower intensity range (Chindarkar et al., 2015). Thus, making lower intensity ions less likely to appear in our untargeted analysis. Targeted metabolomics typically uses a different mass spectrometer, which is able to select an (often predefined) ion and filters out the rest, so that only one ion with a distinct mass/charge-ratio enters the detector. As described before, for this, a triple-quadrupole mass spectrometer is typically used (see in chapter 1.3.2.3, p. 15 or applied in the targeted metabolomics kit method e.g., chapter 3.3, p. 47). Filtering out single ions, the signal/noise ratio increases and thus better linearity even for lower intensity compounds is reached. While modern mass spectrometers are able to scan hundreds of mass/charge ratios every second, however they still are biased and never truly untargeted.

SM 24:1 is a curious biomarker that we also found to be altered in the previous section in Zellweger syndrome patients upon dysfunction of the peroxisome using the AbsoluteIDQ p180-kit (chapter 4.2.3, p. 92). The origin of SM24:1 decreases in both disorders has different reasons. However, it demonstrates the possibility of also analyzing Elov11 patient plasma with this targeted and quantitative metabolomics method. A parallel analysis of the same samples with the targeted metabolomics kit also may show us parallels in findings of the methods and might also validate our findings of SM 24:1. Using a two-group comparison of Elov11 patient and controls, we found a variety of different lipids changed. Among those most prominently were SM but also ether-phosphatidylcholines (PC ae) and our main compound of interest, SM 24:1. Unfortunately, the control group in the first targeted metabolomics assay was measured in serum, prohibiting a direct comparison with the Elov11-Mt patient plasma.

PC ae and SM were also found to be altered in the untargeted lipidomics method, but the downregulation in these classes appeared to be milder. In the targeted screen, we could also identify most SM to decrease, especially one other most prominent lipid, SM 16:0, was dropped compared to controls. The same SM was not changed in the untargeted method, possibly due to a slightly lower linearity of the method. Overall, the results of both methods were not opposing, but rather adding to each other, proving the suitability of the untargeted method, but also showing additional explanatory power when using the targeted method. Unfortunately, in the previously published patient studies, the plasma lipidome was not studied (Kutkowska-Kaźmierczak et al., 2018; Mueller et al., 2019). The classical GC-MS fatty acid profile did not show any absolute alteration in these patients. As seen in chapter 4.0 (p. 88), metabolomics can provide deeper insights into the pathophysiology and able to find more relevant biomarkers.

6.0 Applications: A human patient with an ELOVL1-mutation

Analysis of patient plasma is a way to observe a patient's direct health status; however, patient plasma also gets influenced by a variety of other factors than just pure genotype. Among those are nutrition, general physical status and even the circadian rhythm, where most factors are hard to control and may even strongly vary in the control group, impairing the ability to find alterations (Johnson and Gonzalez, 2012). Also, plasma may be a pool of the most common lipid classes of the human body, but some lipids are strongly restricted to certain tissues (Harayama and Riezman, 2018; Raghu, 2020).

In this project, we additionally decided to generate a patient-derived fibroblast cell culture model, where we were able to control more parameters in detail.

First, we used the less biased retention time/mass overview graph for untargeted analyses, which we also used in plasma analysis in chapter 3.4.5.1 (p. 54). Under standard cell culture conditions, we were already able to find significant changes in the lipidome of the ELOVL1 patient. More strongly than in the plasma, we found decreased lipidomic features in the higher mass/higher retention time range (upper right corner) compared to controls. This effect even became more apparent when removing fetal calf serum (FCS), the principal source of (esterified) fatty acids, including VLCFA (Bailey et al., 1972). With the decrease of signals in the upper right corner, we can now detect increases in the lower mass/lower retention time region (lower left corner). In the basal state other metabolites such as the previously mentioned triglycerides may overlay decreases in the upper left corner, while also VLCFA fatty acids from the cell culture medium may compensate for lower endogenous production. Thus, upon withdrawing exogenous fatty acid sources, firstly, endogenously stored TG are depleted, but also the cells have to be more reliant on the synthesis of VLCFA. This also becomes very evident when looking at SM 24:1 in fibroblasts. In the basal state, the Elov11-Mt is barely different to the controls, while when removing FCS, the abundance of SM 24:1 strongly decreases while in the controls stays the same. In other cell types, the basal conditions may be more affected. In the publication of Sassa et al. different differentiation stages in keratinocytes showed altered fatty acid accumulation patterns. Fibroblasts are a very common cell culture model of pediatric research, as they can be patient derived using a simple skin biopsy, are low maintenance and long-lived cells (Fernandes et al., 2016). Being cells of the connective tissue, they only replicate the basic metabolism of cells, but often do not show many similarities to very specialized cells such as oligodendrocytes, neurons or keratinocytes. It is also known that, the metabolism of VLFCFA-lipids is different between fibroblast and these cells, as myelin and epithelial skin barrier synthesis require higher amounts of VLFCFA. Thus, additionally to fibroblasts, other groups in Elov11 research have also used keratinocytes (Isokawa et al., 2019; Sassa et al., 2013). In these cells, they could also find more altered lipids such as hydroxy-ceramides and others decreased, even under basal conditions. Given these cell types higher VLFCFA use, a missing production of those also affects the prevalence of this lipids stronger.

While the fibroblast cell model is not an ideal model, a cell culture model would allow for the test of potential treatment options. Here a treatment of the cells using VLFCFA fatty acids might be a possibility to study basic biochemical changes of cells. However when using a neuronal/glia cell culture model, even the study of in vitro myelination would be possible (Kerman et al., 2015). This would require generation of iPSC neurons and oligodendrocytes.

6.0 Applications: A human patient with an ELOVL1-mutation

Possibly a murine model would be more appropriate. Interestingly ELOVL1 also has been in the interest of researchers in X-Adrenoleukodystrophy, where a reduction of its activity was proposed to ameliorate the effects of VLFCA accumulation (Ofman et al., 2010), showing a potential connection between this type of disorders.

In summary, we successfully found lipidomic alterations in an Elov11-Mt patient using our untargeted metabolomics platform. This changes chiefly included the decrease of various sphingolipids, were we most strongly detected a lowering of SM 24:1. We were also able to validate those results using the targeted AbsoluteIDQ p180-kit.

7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1

*This work originates from a collaboration with the Institute of clinical pharmacology, UMG, with Mr. Ole Jensen. This work is part of larger publication dealing with the role of OCT1 and has been published as: Jensen O, Matthaei J, Klemp HG, Meyer MJ, Brockmüller J, Tzvetkov MV. Isobutyrylcarnitine as a Biomarker of OCT1 Activity and Interspecies Differences in its Membrane Transport. *Front Pharmacol.* 2021 May 10;12:674559. doi: 10.3389/fphar.2021.674559.*

7.1 Introduction to the analysis of species differences in the organic cation transporter 1

Small organic molecule transporters are an important cause of neuropediatric disorders (Hediger et al., 2013). One of those disorders is X-Adrenoleukodystrophy (X-ALD), a dysfunction of an ABC-transporter probably transporting very long chain fatty acids into peroxisomes (ABCD1; Mosser et al., 1994), which we also analyzed in a previous section (chapter 5.0, p. 99). Another significant rare disorder is primary systemic carnitine deficiency, a disorder caused by mutations in a Na⁺-dependent organic cation transporter (OCTN2, SLC22A5, Longo et al., 2006). Symptoms include myopathy, arrhythmias, hypoglycemia, hepatic encephalopathies, and sudden death (especially in infants). Compensating the low blood carnitine by replacing it with external carnitine generally improves symptoms (Magoulas and El-Hattab, 2012). Therefore, it seems evident that, transporter deficiencies play a substantial role in our neuropediatric research.

Organic cation transporter 1 (OCT1) is a membrane transporter (Koepsell, 2013) that is known to be chiefly responsible for the transport of the xenobiotics Fenoterol (Tzvetkov et al., 2018), Metformin and Sumatriptan (Matthaei et al., 2016). The gene encoding OCT1, *SLC22A1*, is very diverse in humans, and loss of function mutations are common in the European population (Kerb et al., 2002; Seitz et al., 2015; Shu et al., 2003). While these mutations lead to severe pharmacokinetic consequences for the described xenobiotics, affected humans do not develop any other symptoms. An exclusive endogenous substrate of human OCT1 was not proven. Apart from its function in the physiological state, finding endogenous substrates may also be used as biomarkers for OCT1 activity. For instance tetradecanedioate was identified as a biomarker for OAT1, a transporter for many nonsteroidal anti-inflammatory drugs (Yee et al., 2016). Genome-wide association studies (GWAS) have found OCT1 variant activity to be associated with plasma levels of isobutyrylcarnitine (IBC) a metabolite of valine (Luo et al., 2020; Suhre et al., 2011). A transporter study showed the transporter activity of the murine isoform (m)OCT1 for acyl-carnitines, including IBC (Kim et al., 2017). However, more recent analyses have already shown that human (h)OCT1 and mOCT1 have different specificities for multiple compounds (Meyer et al., 2020). Thus, it stands to question if the association of IBC blood concentrations and OCT1 transport activity is caused by direct transporter effects or by other mechanisms.

To elucidate the mechanism further and study whether other biomarkers specific for hOCT1 can be found, the institute of clinical pharmacology contacted us to elaborate the mechanism using metabolomics. We applied our untargeted metabolomics method for hydrophilic metabolites to study differential efflux of the murine and human OCT1 transporter in a HEK293 cell culture assay in a small cooperation study.

7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1

7.2 Specialized method for analyzing the interspecies difference in OCT1

Cell culture experiments were carried out by Ole Jensen (Institute for Clinical Pharmacology, UMG), and details can be found in the corresponding publication (Jensen et al., 2021). Briefly, murine OCT1 (mOCT1) isoform, human OCT1 (hOCT1) isoform, and an empty vector (pcDNA5) were used for overexpression of OCT1 in HEK293 cells. Cells were plated in 12-well plates, the cell culture medium was removed, and cells were incubated with fresh-frozen pooled human plasma for 5 minutes. Then the supernatant was aliquoted, and cells were washed and mechanically detached into cell lysis buffer. After cell lysis, the lysate was normalized to protein concentration (accord. to chapter 2.2.3 and 2.2.4, p. 23). Hydrophilic metabolites of supernatant and cell lysate were then extracted using the ZIC-SPE method. Metabolites were then analyzed using the ZIC-HILIC-MS method in positive ionization mode (chapter 2.2.9, p. 26). Analysis of this data set was done using Progenesis Q1 software, using automatic processing under the settings recommended by the manufacturer. Certain metabolites were identified using our in-house metabolite identification library (according to chapter 2.2.11, p. 28).

7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1

7.3 Results: mOCT1 but not hOCT1 shows increased accumulation of acyl-carnitines in cell supernatant

The organic cation transporter 1 (OCT1) is a known transporter of many xenobiotic compounds such as different drugs (Koepsell, 2013). Albeit not such much is known of the endogenous target of OCT1, it is even less clear whether this transporter transports the same compounds in different species (Meyer et al., 2020). Based on experiments with the murine OCT1 isoform (mOCT1) and GWAS studies in humans, isobutyryl-carnitine has been implied as a biomarker of OCT1 function in humans (Kim et al., 2017; Luo et al., 2020; Suhre et al., 2011).

This project aimed to compare the murine and human isoforms of OCT1 (mOCT1, hOCT1) cloned into HEK cells using untargeted metabolomics. In order to assess the transporter activity, we first treated the cells with human plasma to load cells with compounds typically found in plasma in a more unbiased manner. Then plasma, as well as a cell lysate, was used for metabolomics measurements. To primarily focus on metabolites deregulated between the plasma and cell culture lysate, we decided to use the intensity ratio of plasma to the cell lysate. As OCT1 is mainly expected to transport small positively charged compounds, we used the untargeted metabolomics ZIC-HILIC method in positive ionization mode.

7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1

Using the 65 metabolites with the highest ANOVA p-value, a stringent clustering of the OCT1 respective isoforms was found (**Figure 64**). We found alterations of multiple metabolic features between the human (hOCT1) and mouse (mOCT1) isoform. Especially, among the identified metabolites, it seemed evident that the hOCT1-overexpressing cells did not show a different profile compared to empty-vector controls.

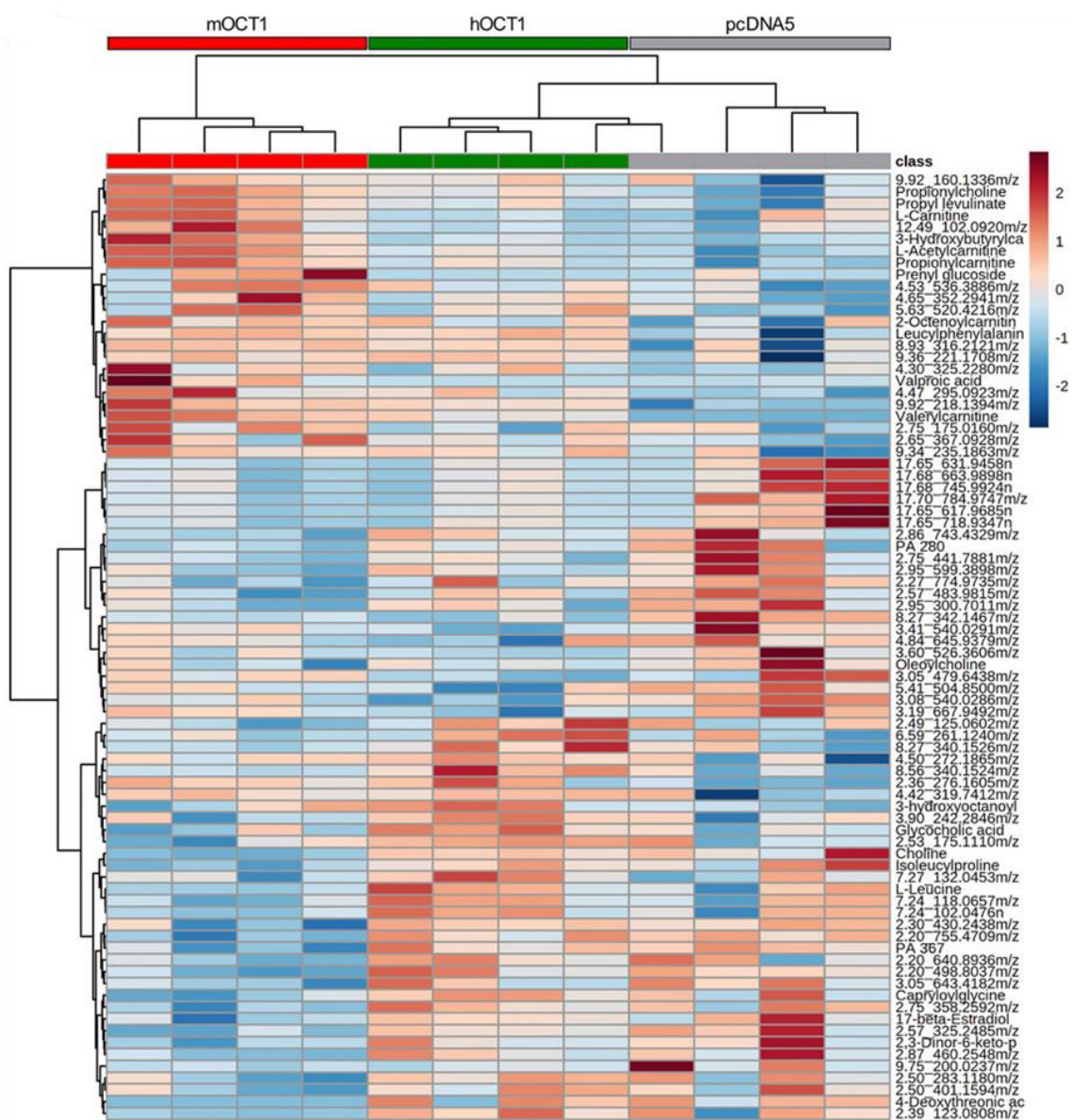


Figure 64 Several metabolites are deregulated between plasma and cell lysate in OCT1 isoforms. Murine and human isoform OCT1 were overexpressed in Hek293 cells and compared to empty vector (pcDNA5) control, n=4. Cell lysates were standardized using protein concentration, lysate and cell supernatant were used to extract hydrophilic metabolites using ZIC-HILIC-SPE. Metabolites were then analyzed using ZIC-HILIC-MS in positive mode. The intensity ratio between plasma and lysate was calculated and features partly identified using the in-house library and progensis software. Unidentified metabolites are displayed with their retention time and mass to charge ratio as the identifiers.

Interestingly among the altered metabolites were also different acyl-carnitine species. Acyl-carnitine transport in the mOCT1 isoform was mentioned by others (Kim et al., 2017).

7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1

Especially acetyl-carnitine, but to a milder extent also free carnitine and hydroxybutyryl-carnitine were increased in mOCT1 compared to the empty vector and hOCT1 cells (**Figure 65**). In the works by Kim et al., free carnitine was also altered, acetyl-carnitine very mildly modified, and hydroxybutyryl-carnitine not reported.

However, regarding the carnitines and a wealth of other molecules, overexpression of the human isoform did not increase the efflux of any detected metabolites in an exclusive way. Other metabolites increased in hOCT1 compared to mOCT1, were also increased in the empty-vector control pcDNA5.

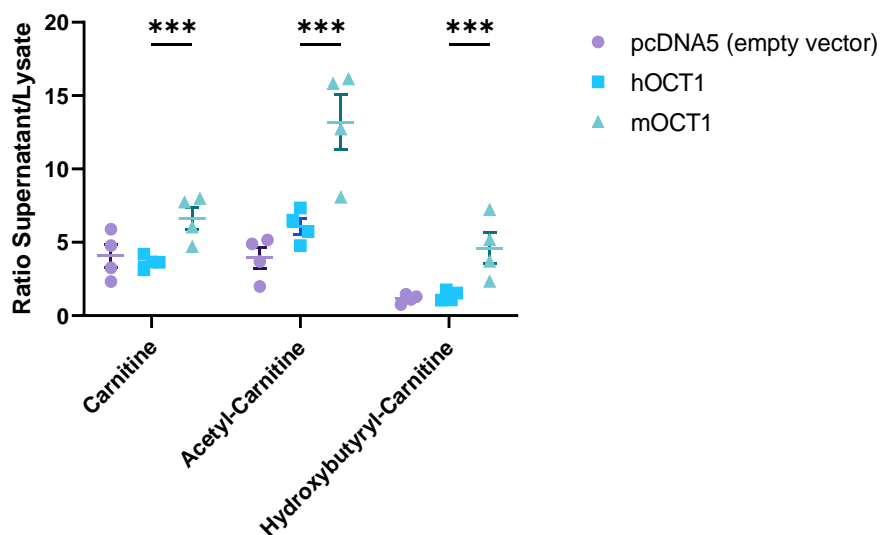


Figure 65 Deregulated (Acyl-)Carnitine in different OCT1 isoforms. Murine and human isoform OCT1 were overexpressed in Hek293 cells and compared to empty vector (pcDNA5) control, n=4. Cell lysates were standardized using protein concentration, lysate, and cell supernatant were used to extract hydrophilic metabolites using ZIC-HILIC-SPE. Metabolites were then analyzed using ZIC-HILIC-MS in positive mode. The intensity ratio between plasma and lysate was calculated and features identified using the in-house library. ANOVA p-value: ***: p<0.001

To summarize, even though a specific metabolite transported by hOCT1 was not found, we were able to show that murine OCT1 expression leads to a strongly different metabolite pattern compared to cellular expression of the human isoform. Additionally, we showed that our metabolomics pipeline was able to reproduce the acyl-carnitine findings of mOCT1 by Kim et al. (Kim et al., 2017).

7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1

7.4 Discussion: Untargeted metabolomics pointed to the transport acyl-carnitines by murine but not the human OCT1 isoform

The organic cation transporter 1 (OCT1) is a membrane transporter that so far is not well understood. While specific xenobiotic targets of OCT1 are known (Matthaei et al., 2016; Tzvetkov et al., 2018), exclusive endogenous metabolites were not found so far. In one study isobutyryl-carnitine (IBC) efflux activity of OCT1 was shown in the murine isoform mOCT1 (Kim et al., 2017). However, more recent results have demonstrated that mOCT activity deviates from the human (hOCT1) isoform (Meyer et al., 2020). Even though GWAS studies showed associations of hOCT1 with the IBC, a direct transporter activity with hOCT1 remains to be proven (Luo et al., 2020; Suhre et al., 2011).

To compare the potential efflux activity of both transporters, Ole Jensen first overexpressed the transporters hOCT1 and mOCT1 in a HEK293 cell culture model. Then he incubated the cells with pooled human patient plasma. Subsequently, we used our untargeted metabolomics method to determine differentially transported metabolites by analyzing the differential metabolome of cell supernatant and cell lysate.

Using heatmap-clustering, we identified a striking difference in the differential metabolome of mOCT1 and hOCT1 cells. The mOCT1 cells showed a particular clustering; the hOCT1 samples clustered similarly to the empty-vector control. Among the most changed metabolites were especially (acyl-)carnitines, as small cationic organic molecules. The highest supernatant/lysate-ratio was found for acetyl-carnitine, followed by hydroxybutyryl-carnitine and carnitine. In our analysis, a high supernatant/lysate ratio may indicate a higher efflux in the mOCT1 overexpressing cells. Efflux activity for acyl-carnitines in mice was shown previously (Kim et al., 2017).

Interestingly, in our analysis we did not find any metabolite that was uniquely increased in hOCT1 cells but not in the empty vector cells. To address this further, Ole Jensen then added deuterium-labeled acyl-carnitines and also observed transport only by mOCT1 (Jensen et al., 2021).

Here it is important to note that the overexpression of mOCT1 and hOCT1 compared to the pcDNA5 was validated previously by RT-PCR, as well as using a functional transport assay of 4-(4-(dimethylamino)styryl)-N-methylpyridinium (Saadatmand et al., 2012). Therefore, it is unlikely that the missing difference between the pcDNA5 and hOCT1 is caused by background activity of endogenous OCT1 in pcDNA5 HEK293 cells.

It is possible that the origin of the differential concentrations in human blood lies not in the cellular influx or efflux but is possibly caused by general alterations in cellular metabolism or even altered intracellular transport between organelles. hOCT1 could be a regulated transporter as well, requiring activation by another factor that might not be expressed in HEK239 cells. A regulation might depend on the metabolic state, leading to a difference in transport to facilitate a switch between the usage of the upstream metabolite valine from protein biosynthesis to energy generation or refueling succinyl-CoA in the citrate cycle. While OCT1 shows the highest expression in the liver, this theory might suggest some muscle participation as an important site for valine catabolism (Holeček, 2018).

7.0 Applications: Use of the ZIC-HILIC method to study species differences in the organic cation transporter 1

Further analyses are needed to delineate the proper function of this enigmatic transporter.

In summary, in cooperation with the Institute of Clinical Pharmacology, we demonstrated the use of our untargeted ZIC-HILIC-MS pipeline for hydrophilic metabolites to study differential transport in a cell culture model. We then used the identification possibilities from our in-house retention time library and were able to identify 3 (acyl-)carnitines that were differentially transported by murine OCT1. These were then confirmed using further studies as well as by researchers from other groups. Thus, we showed the ability of our ZIC-HILIC-MS method to produce relevant results in the study of membrane transporters.

Many rare neuropediatric disorders are caused by transporter deficiencies. Consequently, showing the ability to study cellular transport deficiencies with our metabolomics platform is highly beneficial for our research.

8.0 General discussion, conclusion, and perspectives

Rare disorders are a group of highly diverse disorders that are typically hard to diagnose (Ferreira, 2019). However, the diagnosis is of key importance for the development of potential therapies. A large proportion of rare disorders are inherited metabolism disorders (IMD), especially neurometabolic disorders (Filiano, 2006; Rath et al., 2010; Willemsen et al., 2016). While the introduction of newborn screening has increased the speed of diagnosis for the most common neurometabolic disorders, a large variety is still not diagnosed. Consequently, many patients still spend a long time reaching an accurate diagnosis, hindering proper treatment and decreasing the chance to stop symptom progression.

8.1 Metabolomics can help characterize rare disorders together with genomics methods

Next-generation sequencing has led to a great leap forward for the detection of rare diseases (Turro et al., 2020; Vinkšnel et al., 2021; Wright et al., 2018). Here, techniques such as whole-genome sequencing or whole exome sequencing have led to better sensitivity for rare disorder diagnosis and screen for thousands of different disorders using comprehensive databases. These genomics techniques are successful, albeit they still leave some white spots, offering a typical diagnostic rate of 42 % (Wright et al., 2018). Genomic variants that are not clearly pathogen but also not documented (variants of unknown significance, VUS) cannot be judged for their relevance in the disease origin. Based on the natural variability of the human genome, the occurrence of VUS is common (~10-20, typically up to 4 if filtering by inherited mutations), and genomics techniques will yield multiple different possible causes of this disorder. Even when the variant is known as pathogenic, the disease outbreak or disease strength often does not depend on the accurate genotype. We saw an example of this phenomenon in chapter 5.0 (p. 99) about X-ALD, a peroxisomal neurometabolic disorder, where there is no known correlation between the genotype and the disease outbreak or severity.

One way to investigate the relevance of VUS would be the use of comprehensive functional biomarkers directly relating to the phenotype. This may be fulfilled by the holistic study of metabolites, metabolomics.

The metabolome is linked to both the genotype as well as the phenotype and can dynamically react to changes in the environment (Fiehn, 2002). While there are multiple techniques for approaching metabolomics, liquid chromatography mass spectrometry (LC-MS) typically has a supreme sensitivity, linear range, and a high range of metabolites that it can analyze without the need for derivatization (Dettmer et al., 2007). Consequently, we aimed to create an LC-MS-based metabolomics platform to characterize rare neurometabolic disorders in this thesis. Secondly, we sought to apply this platform to a selected set of neurometabolic diseases to show the platform's utility.

8.2 We successfully created a metabolomics platform based on untargeted and targeted metabolomics

We decided to create a methodic platform based on two pillars: Untargeted metabolomics with an in-house method and secondly targeted metabolomics starting with the commercial AbsoluteIDQ p180 kit. While we could generate valuable data with the targeted metabolomics kit, the untargeted metabolomics methods had to be developed first.

A wide array of untargeted metabolomics methods exist and most methods are very diverse, using different devices and protocols as well as only a few deeply analyze the methodic capacities of their assays (Almontashiri et al., 2020; Breier et al., 2014; Naz et al., 2017). This underlines the necessity to develop in-house methods tailored to the specifications of the institute's research. Unfortunately, untargeted metabolomics methods intensely suffer from misidentifications of metabolites, reducing the scientific output (Xiao et al., 2012). More information on the metabolites, such as retention time and experimental fragmentation data, are needed to reduce misidentifications. Our approach was to directly create a library containing the necessary information of up to 408 metabolites. Using this identification information, we were also able to study essential parameters for method validation on a broader scale than typically seen in metabolomics pipelines. The acquisition of untargeted metabolomics data allows us to identify metabolites of interest even retrospectively. While the compound identification library grows with every new project, every project can also improve the identification of the projects before.

Our untargeted pipeline consisted of a method for hydrophilic compounds and a method for lipophilic compounds, a separation that is often seen in metabolomics (Contrepolis et al., 2015; Naz et al., 2017; Wernisch and Pennathur, 2016). In the field of rare disorder detection with LC-MS, the groups of Coene et al. and Bonte et al. have immensely contributed by creating their method pipelines (Bonte et al., 2019; Coene et al., 2018). Curiously, both platforms use only reversed phase chromatography (RPLC) as their primary analysis method. Albeit RPLC is a very robust technique, these techniques cannot analyze a large variety of metabolites potentially crucial in diseases. Thus, we expect to cover a wider range of different disorders using our combined methods for hydrophilic metabolites and lipids.

Subsequently, using the targeted metabolomics kit and our untargeted method pipeline, we successfully characterized multiple different diseases and for numerous purposes; this included:

- Finding new biomarkers for different peroxisomal disorders (chapter 4.0, p. 88)
- Finding biomarkers for disease progression of the same disease (chapter 5.0, p. 99)
- Proving the cause of a neurometabolic disorder in a single patient (chapter 6.0, p. 112)
- Finding markers for a membrane transporter (chapter 7.0, p.126)

8.3 The metabolomics platform still faces challenges ahead

At the example of these varied types of studies, we were able to show the utility of our metabolomics platform. To further prove the utility of the platform, further research is needed. Extended research and optimization should include four critical areas:

- the generation of better control cohorts
- the selection of a suitable study matrix
- the collection of more rare disorders
- optimization of statistical data analysis.

8.3.1 The access to appropriate control groups is challenging

As mentioned in the introduction, inadequate control groups are among the main problems in pediatric research (chapter 1.1.2, p. 5). Due to ethics regulations, medical procedures (such as blood draws) can only be done if there is a medical indication. Subsequently, most blood or tissue samples in pediatric care can only be acquired from unhealthy patients. This severely limits the sensitivity of diagnostics. Currently, in our institute, most control blood originates from patients of phenylketonuria, where a balanced diet resulted in a normalization of phenylalanine levels. While these may be used (with caution) for the analysis of peroxisomal disorders, it is not recommended for many others. Possible ways to acquire the blood of healthy patients include removing blood from routine blood screens either before planned operations (no extensive inflammation or crushing syndrome should be present) or during the regular checkups of young children of different ages (e.g., the German pediatric examinations U1-J2). However, at neither of those, it is absolutely necessary to draw blood. Thus, the number of samples generated is expected to be fewer as well. Additionally, the use of inconspicuous samples from blood draws of children with psychosomatic symptoms (e.g., diffuse abdominal pain) would be a possibility.

Similarly, for experiments in cell culture, control cohorts are missing. Possible ways to alleviate this would be the collection during routine operations, as well. Collection of material for primary fibroblast cultures would be imaginable during procedures such as circumcision or tooth extractions. For example, the ATCC fibroblast cell line BJ is generated from foreskin fibroblasts (<https://www.atcc.org/products/crl-2522>, accessed 09.06.2021). However, their applicability should be further studied, as fibroblasts from different tissues may have different growth and metabolome profiles (Lynch and Watt, 2018).

Therefore, the generation of a sufficiently sized control cohort requires the collection over more extended periods and possibly from different study centers. Unfortunately, this exceeds the scope of this doctoral thesis, but the generation of a valid control remains a further goal in this institute. Albeit, even using a suboptimal control cohort, we were still able to characterize various disorders based on their strong effect.

8.3.2 The selection of a proper study matrix is challenging

The selection of an appropriate study matrix (e.g., plasma, fibroblasts, CSF, or urine) can significantly influence the detection of different disorders. Based on their function, certain cells (fibroblasts) will not show a metabolome as impacted as other cell types (neurons) might. Thus, specific metabolic aberrations may not be detected. Similarly, blood-derived metabolic profiles present a pool of effects from an extensive range of different cell types. That can offer the advantage that strong aberrations of all cell types can be found, but smaller abnormalities from single tissues may not be found. Also, nutrition can interfere.

Our example of the single patient study of an ELOVL1 defect (chapter 6.0, p. 112) observed similar effects. While we could detect slight aberrations in the plasma and the fibroblast metabolome, they were very mild. Only upon withdrawal of the fatty acid-containing FCS could we see more substantial effects in the fibroblasts (compare to nutrition effects). Other groups have shown that the most significant change in the fatty acid profile was found in myeline and the epidermis, fitting to the phenotype of the patients (Isokawa et al., 2019; Kutkowska-Kaźmierczak et al., 2018; Sassa et al., 2013). While the culture of epidermal keratinocytes would be possible, generating primary patient-derived oligodendrocytes or neuron cell culture is typically not possible. Generating these cell types from patient-derived induced pluripotent stem cells (iPSC) would be a possibility to study these disorders (Dolmetsch and Geschwind, 2011). This would allow access to a significant number of study materials (tissues and cell types) and offer the opportunity to research therapy trials in vitro. New protocols of iPSC generation have enabled researchers to produce cells faster and in larger numbers than ever before, proposing the generation of neuronal stem cells even directly from blood (Chen et al., 2020; Gunhanlar et al., 2018; Sheng et al., 2018). Regrettably, iPSC generation remains expensive, very labor-intensive, and the stability of differentiations may be questionable, making more extensive studies with patients complicated (Doss and Sachinidis, 2019).

Another strategy that could enable better disease detection would be the use of liquid biopsy approaches using extracellular vesicles (EVs). EVs originate from specific cell types that often also carry cell-specific surface proteins and encapsulate cytoplasm from the original cell (Kalra et al., 2016). Using antibody-affinity purification, microvesicles from specific tissue types may be yielded in insufficient purity. This has been shown feasible for genomics and transcriptomics already, and biomarker potential in neurological disorders such as Alzheimer's disease has been shown as well (Badhwar and Haqqani, 2020). As only a low percentage (typically below 1% for brain tissue) of the vesicles have the targeted tissue type (Li et al., 2020), a higher sensitivity for metabolomics studies would be necessary. Especially with targeted metabolomics, this is within the realm of possibility. Curiously, the lipidomic profiling of cell culture-derived EV already resulted in compelling results (Wang et al., 2012), albeit blood-derived tissue-specific EVs were not analyzed using metabolomics methods. Yet, this would indeed be a fascinating application.

8.3.3 The diversity and scarcity of rare disorders is challenging

As rare disorders are rare by definition, acquiring patient samples of a higher number of disorders is a long-term goal and will increase over a long time frame. This also is important to compare the approaches of our pipeline to the other platforms in the field (Bonte et al., 2019; Coene et al., 2018).

However, as the pipeline showed alterations in diverse phenotypes, even including transporter studies (OCT1) and effects by ferroptosis (APP- project), we are confident that this platform may detect a large variety of disorders.

8.3.4 The conception of bioinformatics pipelines is challenging

The fourth complex is the necessity to improve bioinformatics pipelines. Most clinical applications of metabolomics use a minimal set of statistical evaluations, mostly using multiple Student's T-tests, ANOVA, Heatmaps, and principal component analysis (Lamichhane et al., 2018). While we successfully applied these methods also to our pipeline, more advanced techniques may give superior results.

As established in the earlier sections, the metabolism is a complex and emergent system, where the relationships between metabolites are not linear. However, the linearity of relationships between metabolites is exactly one assumption of most of these techniques (Gromski et al., 2015; Hotelling, 1933; Mosconi et al., 2008). Additionally, methods to combine metabolomics with other -omics data should be implemented to further understand disorders more holistically, as proposed by Graham et al. (Graham et al., 2018).

New advancements from the application of other omics techniques such as artificial neural networks may be able to analyze metabolomics data even in a more sensitive manner (Brougham et al., 2010; Mendez et al., 2019; Pomyen et al., 2020). Artificial neuronal networks can classify between multiple groups, even if the data behind the classification is nonlinear. However, this comes with the drawback that many patients are needed in the most straightforward setups, as stated by other authors in metabolomics (Alakwaa et al., 2018). Unfortunately, this is not achievable for many rare disorders due to their low prevalence.

Fortunately, other approaches exist as well. From bank fraud detection and noise removal, a technique called autoencoder has been developed that was even applied to medical image analysis (Chen et al., 2017; Hosny et al., 2018; Jo et al., 2019). An autoencoder is a typically unsupervised neuronal network that has a range of input variables. Over intermediated layers, it compresses the number of variables to a lower number of features before it then uses the same amount of layers (symmetrically) to (re-)predict the input (Bank et al., 2021). These neuronal networks may be trained with data from a normal state. Then a test state (such as a bank fraud case or a single patient) may be predicted. For each of the variables (here: metabolites), the error between the input and predicted value may be calculated. An actual interference may then cause strong prediction errors due to disease. Thus, using a sufficiently sized control group, even single patients may be identified. The application of autoencoders for disease detection with metabolomics data in single patients has not been shown yet but may present a promising target for future research.

8.4 Summary of discussion

The detection of rare disorders in a fast manner is key to enable proper treatment. To enable better and faster diagnosis in cooperation with genomics, we successfully developed a LC-MS based platform for metabolomics analysis of rare disorders. Additionally, we were able to study different research question in the neuropediatric field using our pipeline and thus fulfilled the objective of this thesis project. Further challenges such as optimizing control groups and improving bioinformatic analyses still lie ahead. Steadily including more metabolites into our retention database, increasing our quantification abilities by acquiring other mass spectrometers and optimizations of analytical pipelines will lead to a consistent long-term improvement of our platform.

Especially in combination with other omics techniques, metabolomics will certainly provide deeper insights into rare neuropediatric disorders than ever before.

9.0 References

- Aicheler, F., Li, J., Hoene, M., Lehmann, R., Xu, G., and Kohlbacher, O. (2015). Retention Time Prediction Improves Identification in Nontargeted Lipidomics Approaches. *Anal. Chem.* *87*, 7698–7704.
- Alakwaa, F.M., Chaudhary, K., and Garmire, L.X. (2018). Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *J. Proteome Res.* *17*, 337–347.
- Aldana, J., Romero-Otero, A., and Cala, M.P. (2020). Exploring the Lipidome: Current Lipid Extraction Techniques for Mass Spectrometry Analysis. *Metabolites* *10*.
- Allen, F., Greiner, R., and Wishart, D. (2015). Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* *11*, 98–110.
- Almontashiri, N.A.M., Zha, L., Young, K., Law, T., Kellogg, M.D., Bodamer, O.A., and Peake, R.W.A. (2020). Clinical Validation of Targeted and Untargeted Metabolomics Testing for Genetic Disorders: A 3 Year Comparative Study. *Sci Rep* *10*, 9382.
- Alpert, A.J. (1990). Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr* *499*, 177–196.
- Annesley, T.M. (2003). Ion suppression in mass spectrometry. *Clin Chem* *49*, 1041–1044.
- Aretz, I., and Meierhofer, D. (2016). Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology. *Int J Mol Sci* *17*.
- Åsberg, D., Langborg Weinmann, A., Leek, T., Lewis, R.J., Klarqvist, M., Leško, M., Kaczmarski, K., Samuelsson, J., and Fornstedt, T. (2017). The importance of ion-pairing in peptide purification by reversed-phase liquid chromatography. *J Chromatogr A* *1496*, 80–91.
- Astles, R., Williams, C.P., and Sedor, F. (1994). Stability of plasma lactate in vitro in the presence of antiglycolytic agents. *Clin Chem* *40*, 1327–1330.
- Austin, C.P., Cutillo, C.M., Lau, L.P.L., Jonker, A.H., Rath, A., Julkowska, D., Thomson, D., Terry, S.F., Montleau, B. de, Ardigò, D., et al. (2018). Future of Rare Diseases Research 2017–2027: An IRDiRC Perspective. *Clinical and Translational Science* *11*, 21–27.
- Bacino, C.A., Chao, Y.-H., Seto, E., Lotze, T., Xia, F., Jones, R.O., Moser, A., and Wangler, M.F. (2015). A homozygous mutation in PEX16 identified by whole-exome sequencing ending a diagnostic odyssey. *Molecular Genetics and Metabolism Reports* *5*, 15–18.
- Badhwar, A., and Haqqani, A.S. (2020). Biomarker potential of brain-secreted extracellular vesicles in blood in Alzheimer's disease. *Alzheimers Dement (Amst)* *12*.
- Bailey, J.M., Howard, B.V., Dunbar, L.M., and Tillman, S.F. (1972). Control of lipid metabolism in cultured cells. *Lipids* *7*, 125–134.
- Bank, D., Koenigstein, N., and Giryas, R. (2021). Autoencoders. *ArXiv:2003.05991 [Cs, Stat]*.
- Bansal, S., and DeStefano, A. (2007). Key elements of bioanalytical method validation for small molecules. *AAPS J* *9*, E109–E114.
- Barnes, S., Benton, H.P., Casazza, K., Cooper, S.J., Cui, X., Du, X., Engler, J., Kabarowski, J.H., Li, S., Pathmasiri, W., et al. (2016). Training in metabolomics research. I. Designing the experiment, collecting and extracting samples and generating metabolomics data. *J Mass Spectrom* *51*, 461–475.
- Bartel, J., Krumsiek, J., and Theis, F.J. (2013). Statistical methods for the analysis of high-throughput metabolomics data. *Comput Struct Biotechnol J* *4*.
- van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., and van der Werf, M.J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* *7*, 142.
- Berger, J., and Gärtner, J. (2006). X-linked adrenoleukodystrophy: clinical, biochemical and pathogenetic aspects. *Biochim Biophys Acta* *1763*, 1721–1732.
- Berger, J., Forss-Petter, S., and Eichler, F.S. (2014). Pathophysiology of X-linked adrenoleukodystrophy. *Biochimie* *98*, 135–142.
- Bergeron, A., Furtado, M., and Garofolo, F. (2009). Importance of using highly pure internal standards for successful liquid chromatography/tandem mass spectrometric bioanalytical assays. *Rapid Commun Mass Spectrom* *23*, 1287–1297.
- Beutler, E., and Baluda, M.C. (1966). A simple spot screening test for galactosemia. *J Lab Clin Med* *68*, 137–141.
- Beznan, L., and Moser, H.W. (1998). Incidence of X-linked adrenoleukodystrophy and the relative frequency of its phenotypes. *American Journal of Medical Genetics* *76*, 415–419.
- Bi, H., Krausz, K.W., Manna, S.K., Li, F., Johnson, C.H., and Gonzalez, F.J. (2013). Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI-MS-based metabolomic analysis of adherent mammalian cancer cells. *Anal Bioanal Chem* *405*, 5279–5289.
- Bickel, H., Gerrard, J., and Hickmans, E.M. (1954). The Influence of Phenylalanine Intake on the Chemistry and Behaviour of a Phenylketonuria Child. *Acta Paediatrica* *43*, 64–77.
- Birjandi, A.P., Bojko, B., Ning, Z., Figeys, D., and Pawliszyn, J. (2017). High throughput solid phase microextraction: A new alternative for analysis of cellular lipidome? *Journal of Chromatography B* *1043*, 12–19.

9.0 References

- Björkhem, I., Sisfontes, L., Boström, B., Kase, B.F., and Blomstrand, R. (1986). Simple diagnosis of the Zellweger syndrome by gas-liquid chromatography of dimethylacetals. *J Lipid Res* 27, 786–791.
- Blaise, B.J., Correia, G., Tin, A., Young, J.H., Vergnaud, A.-C., Lewis, M., Pearce, J.T.M., Elliott, P., Nicholson, J.K., Holmes, E., et al. (2016). Power Analysis and Sample Size Determination in Metabolic Phenotyping. *Anal Chem* 88, 5179–5188.
- Blaw, ME (1970). Melanodermic type leukodystrophy (adreno-leukodystrophy). *Handbook of Clinical Neurology*. New York: American Elsevier 128–133.
- Bligh, E.G., and Dyer, W.J. (1959). A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37, 911–917.
- Boelhouwer, C., Knegtel, J.T., and Tels, M. (1967). On the Mechanism of the Thermal Polymerization of Linseed Oil. *Fette, Seifen, Anstrichmittel* 69, 432–436.
- Bonte, R., Bongaerts, M., Demirdas, S., Langendonk, J.G., Huidekoper, H.H., Williams, M., Onkenhout, W., Jacobs, E.H., Blom, H.J., and Ruijter, G.J.G. (2019). Untargeted Metabolomics-Based Screening Method for Inborn Errors of Metabolism using Semi-Automatic Sample Preparation with an UHPLC- Orbitrap-MS Platform. *Metabolites* 9.
- Boulieu, R., Bory, C., Baltassat, P., and Gonnet, C. (1983). Hypoxanthine and xanthine levels determined by high-performance liquid chromatography in plasma, erythrocyte, and urine samples from healthy subjects: the problem of hypoxanthine level evolution as a function of time. *Anal Biochem* 129, 398–404.
- Bourre, J.M., Boiron, F., Cassagne, C., Dumont, O., Leterrier, F., Metzger, H., and Viret, J. (1986). Biochemical and physicochemical determinations in a premyelin fraction obtained by zonal centrifugation in normal mouse and in dysmyelinating mutants (quaking, shiverer, and myelin-deficient). *Neurochem Pathol* 4, 29–42.
- Boysen, A.K., Heal, K.R., Carlson, L.T., and Ingalls, A.E. (2018). Best-Matched Internal Standard Normalization in Liquid Chromatography–Mass Spectrometry Metabolomics Applied to Environmental Samples. *Anal. Chem.* 90, 1363–1369.
- Bradbury, J., Genta-Jouve, G., Allwood, J.W., Dunn, W.B., Goodacre, R., Knowles, J.D., He, S., and Viant, M.R. (2015). MUSCLE: automated multi-objective evolutionary optimization of targeted LC-MS/MS analysis. *Bioinformatics* 31, 975–977.
- Brady, J., and O’Leary, N. (1998). Interference Due to Haemolysis in Routine Photometric Analysis—A Survey. *Ann Clin Biochem* 35, 128–134.
- Braverman, N.E., Raymond, G.V., Rizzo, W.B., Moser, A.B., Wilkinson, M.E., Stone, E.M., Steinberg, S.J., Wangler, M.F., Rush, E.T., Hacia, J.G., et al. (2016). Peroxisome biogenesis disorders in the Zellweger spectrum: An overview of current diagnosis, clinical manifestations, and treatment guidelines. *Mol Genet Metab* 117, 313–321.
- Breier, M., Wahl, S., Prehn, C., Fugmann, M., Ferrari, U., Weise, M., Banning, F., Seissler, J., Grallert, H., Adamski, J., et al. (2014). Targeted Metabolomics Identifies Reliable and Stable Metabolites in Human Serum and Plasma Samples. *PLOS ONE* 9, e89728.
- Bristow, A.W.T., and Webb, K.S. (2003). Intercomparison study on accurate mass measurement of small molecules in mass spectrometry. *J Am Soc Mass Spectrom* 14, 1086–1098.
- Broadhurst, D., Goodacre, R., Reinke, S.N., Kuligowski, J., Wilson, I.D., Lewis, M.R., and Dunn, W.B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* 14, 72.
- Brougham, D.F., Ivanova, G., Gottschalk, M., Collins, D.M., Eustace, A.J., O’Connor, R., and Havel, J. (2010). Artificial Neural Networks for Classification in Metabolomic Studies of Whole Cells Using 1H Nuclear Magnetic Resonance. *Journal of Biomedicine and Biotechnology* 2011, e158094.
- Brunori, P., Masi, P., Faggiani, L., Villani, L., Tronchin, M., Galli, C., Laube, C., Leoni, A., Demi, M., and La Gioia, A. (2011). Evaluation of bilirubin concentration in hemolysed samples, is it really impossible? The altitude-curve cartography approach to interfered assays. *Clinica Chimica Acta* 412, 774–777.
- Burla, B., Arita, M., Arita, M., Bendt, A.K., Cazenave-Gassiot, A., Dennis, E.A., Ekroos, K., Han, X., Ikeda, K., Liebisch, G., et al. (2018). MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines. *J Lipid Res* 59, 2001–2017.
- Buszewski, B., and Noga, S. (2012). Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Anal Bioanal Chem* 402, 231–247.
- Bylda, C., Thiele, R., Kobold, U., and A. Volmer, D. (2014). Recent advances in sample preparation techniques to overcome difficulties encountered during quantitative analysis of small molecules from biofluids using LC-MS/MS. *Analyst* 139, 2265–2276.
- Calderón-Santiago, M., Priego-Capote, F., and Castro, M.D.L. de (2015). Enhancing detection coverage in untargeted metabolomics analysis by solid-phase extraction on-line coupled to LC–MS/MS. *ELECTROPHORESIS* 36, 2179–2187.
- Califf, R.M. (2018). Biomarker definitions and their applications. *Exp Biol Med (Maywood)* 243, 213–221.
- Cambiaghi, A., Ferrario, M., and Masseroli, M. (2017). Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration. *Briefings in Bioinformatics* 18, 498–510.
- Cavalli, G., and Heard, E. (2019). Advances in epigenetics link genetics to the environment and disease. *Nature* 571, 489–499.
- Centerwall, S.A., and Centerwall, W.R. (2000). The Discovery of Phenylketonuria: The Story of a Young Couple, Two Retarded Children, and a Scientist. *Pediatrics* 105, 89–103.

9.0 References

- Chamberlain, C.A., Rubio, V.Y., and Garrett, T.J. (2019). Impact of matrix effects and ionization efficiency in non-quantitative untargeted metabolomics. *Metabolomics* 15, 135.
- Chan, J.P., Wong, B.H., Chin, C.F., Galam, D.L.A., Foo, J.C., Wong, L.C., Ghosh, S., Wenk, M.R., Cazenave-Gassiot, A., and Silver, D.L. (2018). The lysolipid transporter Mfsd2a regulates lipogenesis in the developing brain. *PLoS Biol* 16.
- Chen, M., Shi, X., Zhang, Y., Wu, D., and Guizani, M. (2017). Deep Features Learning for Medical Image Analysis with Convolutional Autoencoder Neural Network. *IEEE Transactions on Big Data* 1–1.
- Chen, M., Maimaitili, M., Habekost, M., Gill, K.P., Mermet-Joret, N., Nabavi, S., Febbraro, F., and Denham, M. (2020). Rapid generation of regionally specified CNS neurons by sequential patterning and conversion of human induced pluripotent stem cells. *Stem Cell Research* 48, 101945.
- Chen, W., Walker, S., and John C, B. (1992). The mechanism of floc formation in protein precipitation by polyelectrolytes. *Chemical Engineering Science* 47, 1039–1045.
- Chindarkar, N.S., Park, H.-D., Stone, J.A., and Fitzgerald, R.L. (2015). Comparison of Different Time of Flight-Mass Spectrometry Modes for Small Molecule Quantitative Analysis. *J Anal Toxicol* 39, 675–685.
- Chirita, R.-I., West, C., Zubrzycki, S., Finaru, A.-L., and Elfakir, C. (2011). Investigations on the chromatographic behaviour of zwitterionic stationary phases used in hydrophilic interaction chromatography. *J Chromatogr A* 1218, 5939–5963.
- Christians, U., Klawitter, J., Hornberger, A., and Klawitter, J. (2011). How unbiased is non-targeted metabolomics and is targeted pathway screening the solution? *Curr Pharm Biotechnol* 12, 1053–1066.
- Ciccimaro, E., and Blair, I.A. (2010). Stable-isotope dilution LC–MS for quantitative biomarker analysis. *Bioanalysis* 2, 311–341.
- Clark, J.D., Lin, L.L., Kriz, R.W., Ramesha, C.S., Sultzman, L.A., Lin, A.Y., Milona, N., and Knopf, J.L. (1991). A novel arachidonic acid-selective cytosolic PLA2 contains a Ca(2+)-dependent translocation domain with homology to PKC and GAP. *Cell* 65, 1043–1051.
- Clements, F.E. (1917). The Ecological Society of America. *Journal of Ecology* 5, 119–128.
- Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 74, 829–836.
- Coene, K.L.M., Kluijtmans, L.A.J., van der Heeft, E., Engelke, U.F.H., de Boer, S., Hoegen, B., Kwast, H.J.T., van de Vorst, M., Huigen, M.C.D.G., Keularts, I.M.L.W., et al. (2018). Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients. *J Inherit Metab Dis* 41, 337–353.
- Constantinescu, C.S., Farooqi, N., O'Brien, K., and Gran, B. (2011). Experimental autoimmune encephalomyelitis (EAE) as a model for multiple sclerosis (MS). *Br J Pharmacol* 164, 1079–1106.
- Contrepois, K., Jiang, L., and Snyder, M. (2015). Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-Mass Spectrometry. *Mol Cell Proteomics* 14, 1684–1695.
- Criscuolo, A., Zeller, M., Cook, K., Angelidou, G., and Fedorova, M. (2019). Rational selection of reverse phase columns for high throughput LC-MS lipidomics. *Chem Phys Lipids* 221, 120–127.
- Croley, T.R., White, K.D., Callahan, J.H., and Musser, S.M. (2012). The Chromatographic Role in High Resolution Mass Spectrometry for Non-Targeted Analysis. *J. Am. Soc. Mass Spectrom.* 23, 1569–1578.
- Cuperlović-Culf, M., Barnett, D.A., Culf, A.S., and Chute, I. (2010). Cell culture metabolomics: applications and future directions. *Drug Discov Today* 15, 610–621.
- D'Alessandro, A., Giardina, B., Gevi, F., Timperio, A.M., and Zolla, L. (2012). Clinical Metabolomics: the next stage of clinical biochemistry. *Blood Transfus* 10, s19–s24.
- De Livera, A.M., Olshansky, M., and Speed, T.P. (2013). Statistical analysis of metabolomics data. *Methods Mol Biol* 1055, 291–307.
- Denihan, N.M., Walsh, B.H., Reinke, S.N., Sykes, B.D., Mandal, R., Wishart, D.S., Broadhurst, D.I., Boylan, G.B., and Murray, D.M. (2015). The effect of haemolysis on the metabolomic profile of umbilical cord blood. *Clin Biochem* 48, 534–537.
- Dettmer, K., Aronov, P.A., and Hammock, B.D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews* 26, 51–78.
- Dettmer, K., Nürnberger, N., Kaspar, H., Gruber, M.A., Almstetter, M.F., and Oefner, P.J. (2011). Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. *Anal Bioanal Chem* 399, 1127–1139.
- Dhopeswarkar, G.A., and Mead, J.F. (1973). Uptake and transport of fatty acids into the brain and the role of the blood-brain barrier system. *Adv Lipid Res* 11, 109–142.
- Djombou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* 8, 61.
- Doff, J., Douce, D., Jones, G., Koroleva, E.V., Skeldon, P., and Thompson, G.E. (2012). Contamination of stainless steel in an electrospray ionization source. *Applied Surface Science* 258, 5270–5282.
- Dolan, J.W., and Snyder, L.R. (2017). *Liquid Chromatography (Second Edition)*, S. Fanali, P.R. Haddad, C.F. Poole, and M.-L. Riekkola, eds. (Elsevier).

9.0 References

- Dolmetsch, R., and Geschwind, D.H. (2011). The human brain in a dish: The promise of iPSC-derived neurons. *Cell* 145, 831–834.
- Dong, M.W. (2006). *Modern HPLC for Practicing Scientists* (John Wiley & Sons).
- Doss, M.X., and Sachinidis, A. (2019). Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications. *Cells* 8.
- Dunn, W.B., and Winder, C.L. (2011). Sample preparation related to the intracellular metabolome of yeast methods for quenching, extraction, and metabolite quantitation. *Methods Enzymol* 500, 277–297.
- Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J.D., Halsall, A., Haselden, J.N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* 6, 1060–1083.
- Ejigu, B.A., Valkenburg, D., Baggerman, G., Vanaerschot, M., Witters, E., Dujardin, J.-C., Burzykowski, T., and Berg, M. (2013). Evaluation of Normalization Methods to Pave the Way Towards Large-Scale LC-MS-Based Metabolomics Profiling Experiments. *OMICS* 17, 473–485.
- Ellis, D.M., Draper, N.P., and Smith, H. (1968). *Applied Regression Analysis*.
- EMA (2012). Bioanalytical method validation. EMEA/CHMP/EWP/192217/2009 Rev. 1 Corr.
- Emwas, A.-H.M. (2015). The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol Biol* 1277, 161–193.
- Epstein, A.A., Narayanasamy, P., Dash, P.K., High, R., Bathena, S.P.R., Gorantla, S., Poluektova, L.Y., Alnouti, Y., Gendelman, H.E., and Boska, M.D. (2013). Combinatorial assessments of brain tissue metabolomics and histopathology in rodent models of human immunodeficiency virus infection. *J Neuroimmune Pharmacol* 8, 1224–1238.
- Ermgren, I., Nestor, M., Pettersson, C., and Hedeland, M. (2021). Improved Sensitivity in Hydrophilic Interaction Liquid Chromatography-Electrospray-Mass Spectrometry after Removal of Sodium and Potassium Ions from Biological Samples. *Metabolites* 11.
- Fan, T.W.-M. (2012). Considerations of Sample Preparation for Metabolomics Investigation. In *The Handbook of Metabolomics*, T.W.-M. Fan, A.N. Lane, and R.M. Higashi, eds. (Totowa, NJ: Humana Press), pp. 7–27.
- Federici, G., and Soddu, S. (2020). Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *Journal of Experimental & Clinical Cancer Research* 39, 46.
- Ferdinandusse, S., Denis, S., Mooyer, P.A.W., Dekker, C., Duran, M., Soorani-Lunsing, R.J., Boltshauser, E., Macaya, A., Gärtner, J., Majoie, C.B.L.M., et al. (2006). Clinical and biochemical spectrum of D-bifunctional protein deficiency. *Annals of Neurology* 59, 92–104.
- Fernandes, I.R., Russo, F.B., Pignatari, G.C., Evangelinellis, M.M., Tavolari, S., Muotri, A.R., and Beltrão-Braga, P.C.B. (2016). Fibroblast sources: Where can we get them? *Cytotechnology* 68, 223–228.
- Ferreira, C.R. (2019). The burden of rare diseases. *American Journal of Medical Genetics Part A* 179, 885–892.
- Ferreira, C.R., Rahman, S., Keller, M., Zschocke, J., and ICIMD Advisory Group (2021). An international classification of inherited metabolic disorders (ICIMD). *J Inher Metab Dis* 44, 164–177.
- Fiehn, O. (2002). Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* 48, 155–171.
- Filiano, J.J. (2006). Neurometabolic Diseases in the Newborn. *Clinics in Perinatology* 33, 411–479.
- Folch, J., Ascoli, I., Lees, M., Meath, J.A., and LeBARON, N. (1951). Preparation of lipide extracts from brain tissue. *J Biol Chem* 191, 833–841.
- Fölling, A. (1934). Über Ausscheidung von Phenylbrenztraubensäure in den Harn als Stoffwechselanomalie in Verbindung mit Imbezillität. 227, 169–181.
- Fonville, J.M., Richards, S.E., Barton, R.H., Boulange, C.L., Ebbels, T.M.D., Nicholson, J.K., Holmes, E., and Dumas, M.-E. (2010). The evolution of partial least squares models and related chemometric approaches in metabolomics and metabolic phenotyping. *Journal of Chemometrics* 24, 636–649.
- Frankel, E.N., Hu, M.-L., and Tappel, A.L. (1989). Rapid headspace gas chromatography of hexanal as a measure of lipid peroxidation in biological samples. *Lipids* 24, 976.
- Fraser, G.W. (2002). Ion detection efficiency of microchannel plates (MCPs). *International Journal of Mass Spectrometry* 215, 13–30.
- Freiburghaus, K., Largiadèr, C.R., Stettler, C., Fiedler, G.M., Bally, L., and Bovet, C. (2019). Metabolomics by UHPLC–MS: benefits provided by complementary use of Q-TOF and QQQ for pathway profiling. *Metabolomics* 15, 120.
- Fu, X., Deja, S., Kucejova, B., Duarte, J.A.G., McDonald, J.G., and Burgess, S.C. (2019). Targeted Determination of Tissue Energy Status by LC-MS/MS. *Anal Chem* 91, 5881–5887.
- Fuhrer, T., Heer, D., Begemann, B., and Zamboni, N. (2011). High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Anal. Chem.* 83, 7074–7080.
- Furey, A., Moriarty, M., Bane, V., Kinsella, B., and Lehane, M. (2013). Ion suppression; a critical review on causes, evaluation, prevention and applications. *Talanta* 115, 104–122.
- Gama, M.R., Collins, C.H., and Bottoli, C.B.G. (2013). Nano-liquid chromatography in pharmaceutical and biomedical research. *J Chromatogr Sci* 51, 694–703.

9.0 References

- Geno, P.W. (1992). Ion Detection in Mass Spectrometry. In *Mass Spectrometry in the Biological Sciences: A Tutorial*, M.L. Gross, ed. (Dordrecht: Springer Netherlands), pp. 133–142.
- German, J.B., Hammock, B.D., and Watkins, S.M. (2005). Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* 1, 3–9.
- Giegel, J.L. (1974). Stability of Triglycerides in Serum. *Clinical Chemistry* 20, 920–920.
- Gika, H.G., Theodoridis, G.A., Plumb, R.S., and Wilson, I.D. (2014). Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis* 87, 12–25.
- Giustarini, D., Dalle-Donne, I., Colombo, R., Milzani, A., and Rossi, R. (2004). Interference of Plasmatic Reduced Glutathione and Hemolysis on Glutathione Disulfide Levels in Human Blood. *Free Radical Research* 38, 1101–1106.
- Godfrey, A.R., Jones, L., Davies, M., and Townsend, R. (2017). Miltefosine: a novel internal standard approach to lysophospholipid quantitation using LC-MS/MS. *Anal Bioanal Chem* 409, 2791–2800.
- González-Domínguez, R., González-Domínguez, Á., Sayago, A., and Fernández-Recamales, Á. (2020). Recommendations and Best Practices for Standardizing the Pre-Analytical Processing of Blood and Urine Samples in Metabolomics. *Metabolites* 10.
- Gower, J.C. (1982). Euclidean Distance Geometry. *Mathematical Scientist* 7, 1–14.
- Graham, E., Lee, J., Price, M., Tarailo-Graovac, M., Matthews, A., Engelke, U., Tang, J., Kluijtmans, L.A.J., Wevers, R.A., Wasserman, W.W., et al. (2018). Integration of genomics and metabolomics for prioritization of rare disease variants: a 2018 literature review. *J Inher Metab Dis* 41, 435–445.
- Griggs, R.C., Batshaw, M., Dunkle, M., Gopal-Srivastava, R., Kaye, E., Krischer, J., Nguyen, T., Paulus, K., Merkel, P.A., and Rare Diseases Clinical Research Network (2009). Clinical research for rare disease: opportunities, challenges, and solutions. *Mol Genet Metab* 96, 20–26.
- Gromski, P.S., Muhamadali, H., Ellis, D.I., Xu, Y., Correa, E., Turner, M.L., and Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Anal Chim Acta* 879, 10–23.
- Gross, J.H. (2013). Einleitung. In *Massenspektrometrie: Ein Lehrbuch*, J.H. Gross, ed. (Berlin, Heidelberg: Springer).
- Gu, C. (2019). LC/MS: Fundamentals, Perspectives, and Applications. In *HPLC and UHPLC for Practicing Scientists*, (John Wiley & Sons, Ltd), pp. 147–175.
- Gunhanlar, N., Shpak, G., van der Kroeg, M., Gouty-Colomer, L.A., Munshi, S.T., Lendemeijer, B., Ghazvini, M., Dupont, C., Hoogendijk, W.J.G., Gribnau, J., et al. (2018). A simplified protocol for differentiation of electrophysiologically mature neuronal networks from human induced pluripotent stem cells. *Mol Psychiatry* 23, 1336–1344.
- Guo, X., and Lankmayr, E. (2011). Phospholipid-based matrix effects in LC-MS bioanalysis. *Bioanalysis* 3, 349–352.
- Guo, B.J., Yang, Z.L., and Zhang, L.J. (2018). Gadolinium Deposition in Brain: Current Scientific Evidence and Future Perspectives. *Front Mol Neurosci* 11, 335.
- Guo, D., Mant, C.T., and Hodges, R.S. (1987). Effects of ion-pairing reagents on the prediction of peptide retention in reversed-phase high-resolution liquid chromatography. *Journal of Chromatography A* 386, 205–222.
- Guthrie, R. (1992). The origin of newborn screening. *Screening* 1, 5–15.
- Gyamfi, D., Ofori Awuah, E., and Owusu, S. (2019). Chapter 1 - Classes, Nomenclature, and Functions of Lipids and Lipid-Related Molecules and the Dietary Lipids. In *The Molecular Nutrition of Fats*, V.B. Patel, ed. (Academic Press), pp. 3–16.
- Haid, M., Muschet, C., Wahl, S., Römisch-Margl, W., Prehn, C., Möller, G., and Adamski, J. (2018). Long-Term Stability of Human Plasma Metabolites during Storage at –80 °C. *J. Proteome Res.* 17, 203–211.
- Hajra, A.K., and Das, A.K. (1996). Lipid Biosynthesis in Peroxisomes. *Annals of the New York Academy of Sciences* 804, 129–141.
- Häkkinen, M.R. (2011). Polyamine analysis by LC-MS. *Methods Mol Biol* 720, 505–518.
- Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.K.P., Fraser, P.D., and Bramley, P.M. (2005). Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot* 56, 219–243.
- Hama, K., Fujiwara, Y., Morita, M., Yamazaki, F., Nakashima, Y., Takei, S., Takashima, S., Setou, M., Shimozawa, N., Imanaka, T., et al. (2018). Profiling and Imaging of Phospholipids in Brains of Abcd1-Deficient Mice. *Lipids* 53, 85–102.
- Hamilton, J.A., and Brunaldi, K. (2007). A Model for Fatty Acid Transport into the Brain. *J Mol Neurosci* 33, 12–17.
- Hanna, V.S., and Hafez, E.A.A. (2018). Synopsis of arachidonic acid metabolism: A review. *J Adv Res* 11, 23–32.
- Harayama, T., and Riezman, H. (2018). Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology* 19, 281–296.
- Hardeman, D., and van den Bosch, H. (1989). Topography of ether phospholipid biosynthesis. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 1006, 1–8.
- Hediger, M.A., Cléménçon, B., Burrier, R.E., and Bruford, E.A. (2013). The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Mol Aspects Med* 34, 95–107.
- Heil, C.S., Wehrheim, S.S., Paithankar, K.S., and Grininger, M. (2019). Fatty Acid Biosynthesis: Chain-Length Regulation and Control. *ChemBioChem* 20, 2298–2321.

9.0 References

- Herzog, K., Pras-Raves, M.L., Vervaart, M.A.T., Luyf, A.C.M., Kampen, A.H.C. van, Wanders, R.J.A., Waterham, H.R., and Vaz, F.M. (2016). Lipidomic analysis of fibroblasts from Zellweger spectrum disorder patients identifies disease-specific phospholipid ratios [S]. *Journal of Lipid Research* 57, 1447–1454.
- Herzog, K., Pras-Raves, M.L., Ferdinandusse, S., Vervaart, M.A.T., Luyf, A.C.M., van Kampen, A.H.C., Wanders, R.J.A., Waterham, H.R., and Vaz, F.M. (2018a). Functional characterisation of peroxisomal β -oxidation disorders in fibroblasts using lipidomics. *J Inherit Metab Dis* 41, 479–487.
- Herzog, K., Pras-Raves, M.L., Ferdinandusse, S., Vervaart, M.A.T., Luyf, A.C.M., van Kampen, A.H.C., Wanders, R.J.A., Waterham, H.R., and Vaz, F.M. (2018b). Plasma lipidomics as a diagnostic tool for peroxisomal disorders. *J Inherit Metab Dis* 41, 489–498.
- Holeček, M. (2018). Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. *Nutrition & Metabolism* 15, 33.
- Honour, J.W. (2011). Development and validation of a quantitative assay based on tandem mass spectrometry. *Ann Clin Biochem* 48, 97–111.
- Hopiavuori, B.R., Anderson, R.E., and Agbaga, M.-P. (2019). ELOVL4: Very long-chain fatty acids serve an eclectic role in mammalian health and function. *Prog Retin Eye Res* 69, 137–158.
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., and Aerts, H.J.W.L. (2018). Artificial intelligence in radiology. *Nat Rev Cancer* 18, 500–510.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24, 417–441.
- Hsu, L.-C., Wen, Z.-H., Chen, H.-M., Lin, H.-T., Chiu, C.-M., and Wu, H.-C. (2013). Evaluation of the Anti-Inflammatory Activities of 5,8,11-cis-Eicosatrienoic Acid. *Food and Nutrition Sciences* 4, 113–119.
- Hu, Q., Tang, H., and Wang, Y. (2020). Challenges in Analysis of Hydrophilic Metabolites Using Chromatography Coupled with Mass Spectrometry. *J. Anal. Test.* 4, 140–162.
- Hubbard, W.C., Moser, A.B., Tortorelli, S., Liu, A., Jones, D., and Moser, H. (2006). Combined liquid chromatography–Tandem mass spectrometry as an analytical method for high throughput screening for X-linked adrenoleukodystrophy and other peroxisomal disorders: Preliminary findings. *Molecular Genetics and Metabolism* 89, 185–187.
- Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* 9, 90–95.
- Ichi, I., Kono, N., Arita, Y., Haga, S., Arisawa, K., Yamano, M., Nagase, M., Fujiwara, Y., and Arai, H. (2014). Identification of genes and pathways involved in the synthesis of Mead acid (20:3n-9), an indicator of essential fatty acid deficiency. *Biochim Biophys Acta* 1841, 204–213.
- Ikeda, M., Kanao, Y., Yamanaka, M., Sakuraba, H., Mizutani, Y., Igarashi, Y., and Kihara, A. (2008). Characterization of four mammalian 3-hydroxyacyl-CoA dehydratases involved in very long-chain fatty acid synthesis. *FEBS Letters* 582, 2435–2440.
- Isokawa, M., Sassa, T., Hattori, S., Miyakawa, T., and Kihara, A. (2019). Reduced chain length in myelin sphingolipids and poorer motor coordination in mice deficient in the fatty acid elongase Elovl1. *FASEB BioAdvances* 1, 747–759.
- Jacyna, J., Kordalewska, M., and Markuszewski, M.J. (2019). Design of Experiments in metabolomics-related studies: An overview. *Journal of Pharmaceutical and Biomedical Analysis* 164, 598–606.
- Jakobsson, A., Westerberg, R., and Jakobsson, A. (2006). Fatty acid elongases in mammals: Their regulation and roles in metabolism. *Progress in Lipid Research* 45, 237–249.
- Jandera, P. (2008). Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts. *J Sep Sci* 31, 1421–1437.
- Jensen, O., Matthaehi, J., Klemp, H.G., Meyer, M.J., Brockmöller, J., and Tzvetkov, M.V. (2021). Isobutyrylcarnitine as a Biomarker of OCT1 Activity and Interspecies Differences in its Membrane Transport. *Front Pharmacol* 12.
- Jervis, G.A. (1947). Studies on phenylpyruvic oligophrenia; the position of the metabolic error. *J Biol Chem* 169, 651–656.
- Jo, T., Nho, K., and Saykin, A.J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Front Aging Neurosci* 11, 220.
- Jobard, E., Trédan, O., Postoly, D., André, F., Martin, A.-L., Elena-Herrmann, B., and Boyault, S. (2016). A Systematic Evaluation of Blood Serum and Plasma Pre-Analytcs for Metabolomics Cohort Studies. *Int J Mol Sci* 17.
- Johnson, C.H., and Gonzalez, F.J. (2012). Challenges and Opportunities of Metabolomics. *J Cell Physiol* 227, 2975–2981.
- Johnson, M.M., Swan, D.D., Surette, M.E., Stegner, J., Chilton, T., Fonteh, A.N., and Chilton, F.H. (1997). Dietary supplementation with gamma-linolenic acid alters fatty acid content and eicosanoid production in healthy humans. *J Nutr* 127, 1435–1444.
- Jones, J.W., Thompson, C.J., Carter, C.L., and Kane, M.A. (2015). Electron-Induced Dissociation (EID) for Structure Characterization of Glycerophosphatidylcholine: Determination of Double Bond Positions and Localization of Acyl Chains. *J Mass Spectrom* 50, 1327–1339.
- Kalra, H., Drummen, G.P.C., and Mathivanan, S. (2016). Focus on Extracellular Vesicles: Introducing the Next Small Big Thing. *Int J Mol Sci* 17.
- Kamlage, B., Maldonado, S.G., Bethan, B., Peter, E., Schmitz, O., Liebenberg, V., and Schatz, P. (2014). Quality Markers Addressing Preanalytical Variations of Blood and Plasma Processing Identified by Broad and Targeted Metabolite Profiling. *Clinical Chemistry* 60, 399–412.

9.0 References

- Kang, Y., Schneider, B.B., and Covey, T.R. (2017). On the Nature of Mass Spectrometer Analyzer Contamination. *J. Am. Soc. Mass Spectrom.* *28*, 2384–2392.
- Karnes, H.T., and March, C. (1991). Calibration and validation of linearity in chromatographic biopharmaceutical analysis. *Journal of Pharmaceutical and Biomedical Analysis* *9*, 911–918.
- Karnovsky, A., and Li, S. (2020). Pathway Analysis for Targeted and Untargeted Metabolomics. In *Computational Methods and Data Analysis for Metabolomics*, S. Li, ed. (New York, NY: Springer US), pp. 387–400.
- Kassmann, C.M., and Nave, K.-A. (2008). Oligodendroglial impact on axonal function and survival - a hypothesis. *Curr Opin Neurol* *21*, 235–241.
- Kassmann, C.M., Lappe-Siefke, C., Baes, M., Brügger, B., Mildner, A., Werner, H.B., Natt, O., Michaelis, T., Prinz, M., Frahm, J., et al. (2007). Axonal loss and neuroinflammation caused by peroxisome-deficient oligodendrocytes. *Nat Genet* *39*, 969–976.
- Katsanis, S.H., and Katsanis, N. (2013). Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet* *14*, 415–426.
- Keller, B.O., Sui, J., Young, A.B., and Whittall, R.M. (2008). Interferences and contaminants encountered in modern mass spectrometry. *Anal Chim Acta* *627*, 71–81.
- Kelley, R.I. (1983). The cerebrohepato-renal syndrome of Zellweger, morphologic and metabolic aspects. *American Journal of Medical Genetics* *16*, 503–517.
- Kerb, R., Brinkmann, U., Chatskaia, N., Gorbunov, D., Gorboulev, V., Mornhinweg, E., Keil, A., Eichelbaum, M., and Koepsell, H. (2002). Identification of genetic variations of the human organic cation transporter hOCT1 and their functional consequences. *Pharmacogenetics* *12*, 591–595.
- Kerman, B.E., Kim, H.J., Padmanabhan, K., Mei, A., Georges, S., Joens, M.S., Fitzpatrick, J.A.J., Jappelli, R., Chandross, K.J., August, P., et al. (2015). In vitro myelin formation using embryonic stem cells. *Development* *142*, 2213–2225.
- Kettwig, M., Klemp, H., Nessler, S., Streit, F., Krätzner, R., Rosewich, H., and Gärtner, J. (2021). Targeted metabolomics revealed changes in phospholipids during the development of neuroinflammation in Abcd1tm1Kds mice and X-linked adrenoleukodystrophy patients. *J Inherit Metab Dis*.
- Khadka, M., Todor, A., Maner-Smith, K.M., Colucci, J.K., Tran, V., Gaul, D.A., Anderson, E.J., Natrajan, M.S., Roupael, N., Mulligan, M.J., et al. (2019). The Effect of Anticoagulants, Temperature, and Time on the Human Plasma Metabolome and Lipidome from Healthy Donors as Determined by Liquid Chromatography-Mass Spectrometry. *Biomolecules* *9*.
- Khamis, M.M., Holt, T., Awad, H., El-Aneed, A., and Adamko, D.J. (2018). Comparative analysis of creatinine and osmolality as urine normalization strategies in targeted metabolomics for the differential diagnosis of asthma and COPD. *Metabolomics* *14*, 115.
- Kim, H.I., Raffler, J., Lu, W., Lee, J.-J., Abbey, D., Saleheen, D., Rabinowitz, J.D., Bennett, M.J., Hand, N.J., Brown, C., et al. (2017). Fine Mapping and Functional Analysis Reveal a Role of SLC22A1 in Acylcarnitine Transport. *Am J Hum Genet* *101*, 489–502.
- Kind, T., and Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* *8*, 105.
- King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., and Olah, T. (2000). Mechanistic investigation of ionization suppression in electrospray ionization. *J Am Soc Mass Spectrom* *11*, 942–950.
- Kirkland, J.J., Schuster, S.A., Johnson, W.L., and Boyes, B.E. (2013). Fused-core particle technology in high-performance liquid chromatography: An overview. *Journal of Pharmaceutical Analysis* *3*, 303–312.
- Klemp, H.G., Kettwig, M., Streit, F., Gärtner, J., Rosewich, H., and Krätzner, R. (2021). LC-MS Based Platform Simplifies Access to Metabolomics for Peroxisomal Disorders. *Metabolites* *11*, 347.
- van der Kloet, F.M., Bobeldijk, I., Verheij, E.R., and Jellema, R.H. (2009). Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. *J. Proteome Res.* *8*, 5132–5141.
- Klouwer, F.C.C., Berendse, K., Ferdinandusse, S., Wanders, R.J.A., Engelen, M., and Poll-The, B.T. (2015). Zellweger spectrum disorders: clinical overview and management approach. *Orphanet Journal of Rare Diseases* *10*, 151.
- Klupczynska, A., Misiura, M., Milyk, W., Oscilowska, I., Palka, J., Kokot, Z.J., and Matysiak, J. (2020). Development of an LC-MS Targeted Metabolomics Methodology to Study Proline Metabolism in Mammalian Cell Cultures. *Molecules* *25*.
- Knittelfelder, O.L., Weberhofer, B.P., Eichmann, T.O., Kohlwein, S.D., and Rechberger, G.N. (2014). A versatile ultra-high performance LC-MS method for lipid profiling. *J Chromatogr B Analyt Technol Biomed Life Sci* *951–952*, 119–128.
- Kochhar, S., Jacobs, D.M., Ramadan, Z., Berruex, F., Fuerholz, A., and Fay, L.B. (2006). Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabolomics. *Anal Biochem* *352*, 274–281.
- Koepsell, H. (2013). The SLC22 family with transporters of organic cations, anions and zwitterions. *Mol Aspects Med* *34*, 413–435.
- Köfeler, H.C., and Gross, M.L. (2005). Correction of accurate mass measurement for target compound verification by quadrupole time-of-flight mass spectrometry. *Journal of the American Society for Mass Spectrometry* *16*, 406–408.
- Kornhuber, M.E., Balabanova, S., Heiligensetzer, G.V., Kornhuber, C., Zettlmeissl, H., and Kornhuber, A.W. (1991). Stability of human blood serum aminoacids after storage at different pH and temperature conditions. *Clinica Chimica Acta* *197*, 189–200.

9.0 References

- Koseoglu, M., Hur, A., Atay, A., and Cuhadar, S. (2011). Effects of hemolysis interference on routine biochemistry parameters. *Biochem Med* 21, 79–85.
- Kostiainen, R., Kotiaho, T., Kuuranne, T., and Auriola, S. (2003). Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. *J Mass Spectrom* 38, 357–372.
- Kushnir, M.M., Rockwood, A.L., Nelson, G.J., Yue, B., and Urry, F.M. (2005). Assessing analytical specificity in quantitative analysis using tandem mass spectrometry. *Clinical Biochemistry* 38, 319–327.
- Kutkowska-Każmierczak, A., Rydzanicz, M., Chlebowski, A., Kłosowska-Kosicka, K., Mika, A., Gruchota, J., Jurkiewicz, E., Kowalewski, C., Pollak, A., Stradomska, T.J., et al. (2018). Dominant ELOVL1 mutation causes neurological disorder with ichthyotic keratoderma, spasticity, hypomyelination and dysmorphic features. *J Med Genet* 55, 408–414.
- Lakka, N.S., and Kuppan, C. (2019). *Principles of Chromatography Method Development* (IntechOpen).
- Lamichhane, S., Sen, P., Dickens, A.M., Hyötyläinen, T., and Orešič, M. (2018). Chapter Fourteen - An Overview of Metabolomics Data Analysis: Current Tools and Future Perspectives. In *Comprehensive Analytical Chemistry*, J. Jaumot, C. Bedia, and R. Tauler, eds. (Elsevier), pp. 387–413.
- Lands, W.E.M. (1958). METABOLISM OF GLYCEROLIPIDES: A COMPARISON OF LECITHIN AND TRIGLYCERIDE SYNTHESIS. *Journal of Biological Chemistry* 231, 883–888.
- Le, A., Ng, A., Kwan, T., Cusmano-Ozog, K., and Cowan, T.M. (2014). A rapid, sensitive method for quantitative analysis of underivatized amino acids by liquid chromatography–tandem mass spectrometry (LC–MS/MS). *Journal of Chromatography B* 944, 166–174.
- Lei, Z., Huhman, D.V., and Sumner, L.W. (2011). Mass Spectrometry Strategies in Metabolomics *. *Journal of Biological Chemistry* 286, 25435–25442.
- Lenz, E.M., Bright, J., Wilson, I.D., Hughes, A., Morrisson, J., Lindberg, H., and Lockton, A. (2004). Metabonomics, dietary influences and cultural differences: a ¹H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J Pharm Biomed Anal* 36, 841–849.
- Leonard, A.E., Pereira, S.L., Sprecher, H., and Huang, Y.-S. (2004). Elongation of long-chain fatty acids. *Progress in Lipid Research* 43, 36–54.
- Li, Y., He, X., Li, Q., Lai, H., Zhang, H., Hu, Z., Li, Y., and Huang, S. (2020). EV-origin: Enumerating the tissue-cellular origin of circulating extracellular vesicles using exLR profile. *Computational and Structural Biotechnology Journal* 18, 2851–2859.
- Licea-Perez, H., Junnotula, V., Zohrabian, S., and Karlinsey, M. (2016). Development of a multi-sugar LC-MS/MS assay using simple chemical derivatization with acetic anhydride. *Anal. Methods* 8, 3023–3033.
- Liebisch, G., Ekroos, K., Hermansson, M., and Ejsing, C.S. (2017). Reporting of lipidomics data should be standardized. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1862, 747–751.
- Liebisch, G., Ahrends, R., Arita, M., Arita, M., Bowden, J.A., Ejsing, C.S., Griffiths, W.J., Holčápek, M., Köfeler, H., Mitchell, T.W., et al. (2019). Lipidomics needs more standardization. *Nature Metabolism* 1, 745–747.
- Lin, Y., Caldwell, G.W., Li, Y., Lang, W., and Masucci, J. (2020). Inter-laboratory reproducibility of an untargeted metabolomics GC–MS assay for analysis of human plasma. *Scientific Reports* 10, 10918.
- Lindon, J.C., Nicholson, J.K., Holmes, E., and Everett, J.R. (2000). Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in Magnetic Resonance* 12, 289–320.
- Linertová, R., García-Pérez, L., and Gorostiza, I. (2017). Cost-of-Illness in Rare Diseases. In *Rare Diseases Epidemiology: Update and Overview*, M. Posada de la Paz, D. Taruscio, and S.C. Groft, eds. (Cham: Springer International Publishing), pp. 283–297.
- Link, A.J., and LaBaer, J. (2011). Trichloroacetic Acid (TCA) Precipitation of Proteins. *Cold Spring Harb Protoc* 2011, pdb.prot5651.
- Loes, D.J., Hite, S., Moser, H., Stillman, A.E., Shapiro, E., Lockman, L., Latchaw, R.E., and Krivit, W. (1994). Adrenoleukodystrophy: a scoring method for brain MR observations. *AJNR Am J Neuroradiol* 15, 1761–1766.
- Löfgren, L., Forsberg, G.-B., and Ståhlman, M. (2016). The BUMÉ method: a new rapid and simple chloroform-free method for total lipid extraction of animal tissue. *Sci Rep* 6, 27688.
- Longo, N., di San Filippo, C.A., and Pasquali, M. (2006). DISORDERS OF CARNITINE TRANSPORT AND THE CARNITINE CYCLE. *Am J Med Genet C Semin Med Genet* 142C, 77–85.
- Lu, J.F., Lawler, A.M., Watkins, P.A., Powers, J.M., Moser, A.B., Moser, H.W., and Smith, K.D. (1997). A mouse model for X-linked adrenoleukodystrophy. *Proc Natl Acad Sci U S A* 94, 9366–9371.
- Lu, W., Su, X., Klein, M.S., Lewis, I.A., Fiehn, O., and Rabinowitz, J.D. (2017). Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. *Annu Rev Biochem* 86, 277–304.
- Luo, L., Ramanathan, R., Horlbogen, L., Mathialagan, S., Costales, C., Vourvahis, M., Holliman, C.L., and Rodrigues, A.D. (2020). A Multiplexed HILIC-MS/HRMS Assay for the Assessment of Transporter Inhibition Biomarkers in Phase I Clinical Trials: Isobutryryl-Carnitine as an Organic Cation Transporter (OCT1) Biomarker. *Anal Chem* 92, 9745–9754.
- Lynch, M.D., and Watt, F.M. (2018). Fibroblast heterogeneity: implications for human disease. *J Clin Invest* 128, 26–35.
- Magnes, C., Fauland, A., Gander, E., Narath, S., Ratzer, M., Eisenberg, T., Madeo, F., Pieber, T., and Sinner, F. (2014). Polyamines in biological samples: Rapid and robust quantification by solid-phase extraction online-coupled to liquid chromatography–tandem mass spectrometry. *J Chromatogr A* 1331, 44–51.

9.0 References

- Magoulas, P.L., and El-Hattab, A.W. (2012). Systemic primary carnitine deficiency: an overview of clinical manifestations, diagnosis, and management. *Orphanet Journal of Rare Diseases* 7, 68.
- Malm, L., Tybring, G., Moritz, T., Landin, B., and Galli, J. (2016). Metabolomic Quality Assessment of EDTA Plasma and Serum Samples. *Biopreserv Biobank* 14, 416–423.
- la Marca, G. (2014). Mass spectrometry in clinical chemistry: the case of newborn screening. *Journal of Pharmaceutical and Biomedical Analysis* 101, 174–182.
- Martin, A.C., Pawlus, A.D., Jewett, E.M., Wyse, D.L., Angerhofer, C.K., and Hegeman, A.D. (2014). Evaluating solvent extraction systems using metabolomics approaches. *RSC Adv.* 4, 26325–26334.
- Masson, P., Alves, A.C., Ebbels, T.M.D., Nicholson, J.K., and Want, E.J. (2010). Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS. *Anal Chem* 82, 7779–7786.
- Matthaei, J., Kuron, D., Faltraco, F., Knoch, T., Dos Santos Pereira, J.N., Abu Abed, M., Prukop, T., Brockmöller, J., and Tzvetkov, M.V. (2016). OCT1 mediates hepatic uptake of sumatriptan and loss-of-function OCT1 polymorphisms affect sumatriptan pharmacokinetics. *Clin Pharmacol Ther* 99, 633–641.
- Matthan, N.R., Ip, B., Resteghini, N., Ausman, L.M., and Lichtenstein, A.H. (2010). Long-term fatty acid stability in human serum cholesteryl ester, triglyceride, and phospholipid fractions. *J Lipid Res* 51, 2826–2832.
- Matyash, V., Liebisch, G., Kurzchalia, T.V., Shevchenko, A., and Schwudke, D. (2008). Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res* 49, 1137–1146.
- Mayatepek, E., Baumann, M., Meissner, T., Hanefeld, F., and Korenke, G.C. (2003). Role of leukotrienes as indicators of the inflammatory demyelinating reaction in x-linked cerebral adrenoleukodystrophy. *J Neurol* 250, 1259–1260.
- Mazzocchi, F. (2012). Complexity and the reductionism–holism debate in systems biology. *WIREs Systems Biology and Medicine* 4, 413–427.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology* 17, 122.
- McMaster (2007). Column Aging, Diagnosis, and Healing. In *HPLC*, (John Wiley & Sons, Ltd), pp. 73–87.
- Melhem, E.R., Loes, D.J., Georgiades, C.S., Raymond, G.V., and Moser, H.W. (2000). X-linked Adrenoleukodystrophy: The Role of Contrast-enhanced MR Imaging in Predicting Disease Progression. *American Journal of Neuroradiology* 21, 839–844.
- Mendez, K.M., Broadhurst, D.I., and Reinke, S.N. (2019). The application of artificial neural networks in metabolomics: a historical perspective. *Metabolomics* 15, 142.
- Merrill, M.H., and Fleisher, M.S. (1932). Factors involved in the use of organic solvents as precipitating and drying agents of immune sera. *J Gen Physiol* 16, 243–256.
- Metherel, A.H., and Stark, K.D. (2016). The stability of blood fatty acids during storage and potential mechanisms of degradation: A review. *Prostaglandins Leukot Essent Fatty Acids* 104, 33–43.
- Metherel, A.H., Hogg, R.C., Buzikievich, L.M., and Stark, K.D. (2013). Butylated hydroxytoluene can protect polyunsaturated fatty acids in dried blood spots from degradation for up to 8 weeks at room temperature. *Lipids in Health and Disease* 12, 22.
- Meyer, M.J., Tuerkova, A., Römer, S., Wenzel, C., Seitz, T., Gaedcke, J., Oswald, S., Brockmöller, J., Zdražil, B., and Tzvetkov, M.V. (2020). Differences in Metformin and Thiamine Uptake between Human and Mouse Organic Cation Transporter 1: Structural Determinants and Potential Consequences for Intrahepatic Concentrations. *Drug Metab Dispos* 48, 1380–1392.
- Micheel, C.M., Nass, S.J., Omenn, G.S., Trials, C. on the R. of O.-B.T. for P.P.O. in C., Services, B. on H.C., Policy, B. on H.S., and Medicine, I. of (2012). *Omics-Based Clinical Discovery: Science, Technology, and Applications* (National Academies Press (US)).
- Mičová, K., Friedecký, D., and Adam, T. (2017). *Mass Spectrometry for the Sensitive Analysis of Intracellular Nucleotides and Analogues* (IntechOpen).
- Mignarri, A., Vinciguerra, C., Giorgio, A., Ferdinandusse, S., Waterham, H., Wanders, R., Bertini, E., Dotti, M.T., and Federico, A. (2012). Zellweger Spectrum Disorder with Mild Phenotype Caused by PEX2 Gene Mutations. *JIMD Rep* 6, 43–46.
- Miller, M.J., Kennedy, A.D., Eckhart, A.D., Burrage, L.C., Wulff, J.E., Miller, L.A.D., Milburn, M.V., Ryals, J.A., Beaudet, A.L., Sun, Q., et al. (2015). Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. *J Inher Metab Dis* 38, 1029–1039.
- Millington, D.S., Kodo, N., Norwood, D.L., and Roe, C.R. (1990). Tandem mass spectrometry: A new method for acylcarnitine profiling with potential for neonatal screening for inborn errors of metabolism. *Journal of Inherited Metabolic Disease* 13, 321–324.
- Misra, B.B. (2020). Data normalization strategies in metabolomics: Current challenges, approaches, and tools. *Eur J Mass Spectrom (Chichester)* 26, 165–174.
- Monnin, C., Ramrup, P., Daigle-Young, C., and Vuckovic, D. (2018). Improving negative liquid chromatography/electrospray ionization mass spectrometry lipidomic analysis of human plasma using acetic acid as a mobile-phase additive. *Rapid Commun Mass Spectrom* 32, 201–211.
- Moon, Y.-A., and Horton, J.D. (2003). Identification of Two Mammalian Reductases Involved in the Two-carbon Fatty Acyl Elongation Cascade *. *Journal of Biological Chemistry* 278, 7335–7343.
- Morris, M., Pierre, T., and Robert K., B. (1994). Characterization of a high-pressure quadrupole collision cell for low-energy collision-induced dissociation. *Journal of the American Society for Mass Spectrometry* 5, 1042–1063.

9.0 References

- Mosconi, F., Julou, T., Desprat, N., Sinha, D.K., Allemand, J.-F., Croquette, V., and Bensimon, D. (2008). Some nonlinear challenges in biology. *Nonlinearity* 21, T131–T147.
- Moser, H.W., and Moser, A.B. (1996). Very long-chain fatty acids in diagnosis, pathogenesis, and therapy of peroxisomal disorders. *Lipids* 31, S141–S144.
- Moser, A.B., Kreiter, N., Bezman, L., Lu, S.-E., Raymond, G.V., Naidu, S., and Moser, H.W. (1999). Plasma very long chain fatty acids in 3,000 peroxisome disease patients and 29,000 controls. *Annals of Neurology* 45, 100–110.
- Mosser, J., Lutz, Y., Stoeckel, M.E., Sarde, C.O., Kretz, C., Douar, A.M., Lopez, J., Aubourg, P., and Mandel, J.L. (1994). The gene responsible for adrenoleukodystrophy encodes a peroxisomal membrane protein. *Hum Mol Genet* 3, 265–271.
- Mueller, N., Sassa, T., Morales-Gonzalez, S., Schneider, J., Salchow, D.J., Seelow, D., Knierim, E., Stenzel, W., Kihara, A., and Schuelke, M. (2019). De novo mutation in ELOVL1 causes ichthyosis, acanthosis nigricans, hypomyelination, spastic paraplegia, high frequency deafness and optic atrophy. *J Med Genet* 56, 164–175.
- Murphy, R.C., and Folco, G. (2019). Lysophospholipid acyltransferases and leukotriene biosynthesis: intersection of the Lands cycle and the arachidonate PI cycle [S]. *Journal of Lipid Research* 60, 219–226.
- Muschet, C., Möller, G., Prehn, C., de Angelis, M.H., Adamski, J., and Tokarz, J. (2016). Removing the bottlenecks of cell culture metabolomics: fast normalization procedure, correlation of metabolites to cell number, and impact of the cell harvesting method. *Metabolomics* 12, 151.
- Naidong, W. (2003). Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. *J Chromatogr B Analyt Technol Biomed Life Sci* 796, 209–224.
- Narváez-Rivas, M., and Zhang, Q. (2016). Comprehensive untargeted lipidomic analysis using core–shell C30 particle column and high field orbitrap mass spectrometer. *J Chromatogr A* 1440, 123–134.
- Natarajan, A., Christopher, R., Netravathi, M., Bhat, M.D., and Chandra, S.R. (2019). Flow injection ionization-tandem mass spectrometry-based estimation of a panel of lysophosphatidylcholines in dried blood spots for screening of X-linked adrenoleukodystrophy. *Clin Chim Acta* 495, 167–173.
- Naz, S., Gallart-Ayala, H., Reinke, S.N., Mathon, C., Blankley, R., Chaleckis, R., and Wheelock, C.E. (2017). Development of a Liquid Chromatography–High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition. *Anal. Chem.* 89, 7933–7942.
- Nessler, S., Boretius, S., Stadelmann, C., Bittner, A., Merkler, D., Hartung, H.-P., Michaelis, T., Brück, W., Frahm, J., Sommer, N., et al. (2007). Early MRI changes in a mouse model of multiple sclerosis are predictive of severe inflammatory tissue damage. *Brain* 130, 2186–2198.
- Nguengang Wakap, S., Lambert, D.M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., and Rath, A. (2020). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet* 28, 165–173.
- Ofman, R., Dijkstra, I.M.E., van Roermund, C.W.T., Burger, N., Turkenburg, M., van Cruchten, A., van Engen, C.E., Wanders, R.J.A., and Kemp, S. (2010). The role of ELOVL1 in very long-chain fatty acid homeostasis and X-linked adrenoleukodystrophy. *EMBO Mol Med* 2, 90–97.
- Ohno, Y., Suto, S., Yamanaka, M., Mizutani, Y., Mitsutake, S., Igarashi, Y., Sassa, T., and Kihara, A. (2010). ELOVL1 production of C24 acyl-CoAs is linked to C24 sphingolipid synthesis. *Proc Natl Acad Sci U S A* 107, 18439–18444.
- Okuno, T., Gijón, M.A., Zarini, S., Martin, S.A., Barkley, R.M., Johnson, C.A., Ohba, M., Yokomizo, T., and Murphy, R.C. (2018). Altered eicosanoid production and phospholipid remodeling during cell culture. *J Lipid Res* 59, 542–549.
- Oliver, S.G., Winson, M.K., Kell, D.B., and Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology* 16, 373–378.
- Opialla, T., Kempa, S., and Pietzke, M. (2020). Towards a More Reliable Identification of Isomeric Metabolites Using Pattern Guided Retention Validation. *Metabolites* 10.
- Ovčáčíková, M., Lisa, M., Cífková, E., and Holčápek, M. (2016). Retention behavior of lipids in reversed-phase ultrahigh-performance liquid chromatography–electrospray ionization mass spectrometry. *Journal of Chromatography A* 1450, 76–85.
- Pabst, M., Grass, J., Fischl, R., Léonard, R., Jin, C., Hinterkörner, G., Borth, N., and Altmann, F. (2010). Nucleotide and Nucleotide Sugar Analysis by Liquid Chromatography–Electrospray Ionization–Mass Spectrometry on Surface-Conditioned Porous Graphitic Carbon. *Anal Chem* 82, 9782–9788.
- Pang, Z., Chong, J., Zhou, G., de Lima Morais, D.A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.-É., Li, S., and Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research*.
- Pearson, K. (1895). Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London* 58, 240–242.
- Peraman, R., Bhadrara, K., and Padmanabha Reddy, Y. (2015). Analytical Quality by Design: A Tool for Regulatory Flexibility and Robust Analytics. *Int J Anal Chem* 2015.
- Pereira, L. (2008). Porous Graphitic Carbon as a Stationary Phase in HPLC: Theory and Applications. *Journal of Liquid Chromatography & Related Technologies* 31, 1687–1731.
- Peris-Vicente, J., Esteve-Romero, J., and Carda-Broch, S. (2015). Validation of Analytical Methods Based on Chromatographic Techniques: An Overview. In *Analytical Separation Science*, (American Cancer Society), pp. 1757–1808.
- Pifferi, F., Laurent, B., and Plourde, M. (2021). Lipid Transport and Metabolism at the Blood-Brain Interface: Implications in Health and Disease. *Front. Physiol.* 12.

9.0 References

- Piras, D., Locci, E., Palmas, F., Ferino, G., Fanos, V., Noto, A., D'aloja, E., and Finco, G. (2016). Rare disease: a focus on metabolomics. *Expert Opinion on Orphan Drugs* 4, 1229–1237.
- Polson, C., Sarkar, P., Incledon, B., Raguvaran, V., and Grant, R. (2003). Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 785, 263–275.
- Pomyen, Y., Wanichthanarak, K., Pounsombat, P., Fahrman, J., Grapov, D., and Khoomrung, S. (2020). Deep metabolome: Applications of deep learning in metabolomics. *Computational and Structural Biotechnology Journal* 18, 2818–2825.
- Poole, C.F. (2003). New trends in solid-phase extraction. *TrAC Trends in Analytical Chemistry* 22, 362–373.
- Rácz, O., Lepej, J., Fodor, B., Lepejová, K., Jarčuška, P., and Kováčová, A. (2012). Pitfalls in the Measurements and Assessment of Glomerular Filtration Rate and How to Escape them. *EJIFCC* 23, 33–40.
- Raghu, P. (2020). Functional diversity in a lipidome. *PNAS* 117, 11191–11193.
- Rampler, E., Schoeny, H., M. Mitic, B., Abiead, Y.E., Schwaiger, M., and Koellensperger, G. (2018). Simultaneous non-polar and polar lipid analysis by on-line combination of HILIC, RP and high resolution MS. *Analyst* 143, 1250–1258.
- Rath, A., Aymé, S., and Bellet, B. (2010). Classification of rare diseases: a worldwide effort to contribute to the International Classification of Diseases. *Orphanet Journal of Rare Diseases* 5, O21.
- Raymond, G.V., Aubourg, P., Paker, A., Escolar, M., Fischer, A., Blanche, S., Baruchel, A., Dalle, J.-H., Michel, G., Prasad, V., et al. (2019). Survival and Functional Outcomes in Boys with Cerebral Adrenoleukodystrophy with and without Hematopoietic Stem Cell Transplantation. *Biol Blood Marrow Transplant* 25, 538–548.
- Reis, A., Rudnitskaya, A., Blackburn, G.J., Fauzi, N.M., Pitt, A.R., and Spickett, C.M. (2013). A comparison of five lipid extraction solvent systems for lipidomic studies of human LDL[S]. *Journal of Lipid Research* 54, 1812–1824.
- Ribbenstedt, A., Ziarrusta, H., and Benskin, J.P. (2018). Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PLOS ONE* 13, e0207082.
- Richmond, P.A., van der Kloet, F., Vaz, F.M., Lin, D., Uzozie, A., Graham, E., Kobor, M., Mostafavi, S., Moerland, P.D., Lange, P.F., et al. (2020). Multi-Omic Approach to Identify Phenotypic Modifiers Underlying Cerebral Demyelination in X-Linked Adrenoleukodystrophy. *Front Cell Dev Biol* 8.
- Richter, T., Nestler-Parr, S., Babela, R., Khan, Z.M., Tesoro, T., Molsen, E., and Hughes, D.A. (2015). Rare Disease Terminology and Definitions—A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. *Value in Health* 18, 906–914.
- Roberts, L.D., Souza, A.L., Gerszten, R.E., and Clish, C.B. (2012). Targeted Metabolomics. *Curr Protoc Mol Biol*, Unit 30.2.
- Rogatsky, E., Jayatilake, H., Goswami, G., Tomuta, V., and Stein, D. (2005). Sensitive LC MS Quantitative Analysis of Carbohydrates by Cs⁺ Attachment. *Journal of the American Society for Mass Spectrometry* 16, 1805–1811.
- Römisch-Margl, W., Prehn, C., Bogumil, R., Röhring, C., Suhre, K., and Adamski, J. (2012). Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 8, 133–142.
- Rosenling, T., Stoop, M.P., Smolinska, A., Muilwijk, B., Coulier, L., Shi, S., Dane, A., Christin, C., Suits, F., Horvatovich, P.L., et al. (2011). The impact of delayed storage on the measured proteome and metabolome of human cerebrospinal fluid. *Clin Chem* 57, 1703–1711.
- Rousseau, F., Giguère, Y., Berthier, M.-T., Guérette, D., Girard, J.-G., and Déry, M. (2012). Newborn Screening By Tandem Mass Spectrometry: Impacts, Implications and Perspectives (IntechOpen).
- Rozet, E., Marini, R.D., Ziemons, E., Boulanger, B., and Hubert, P. (2011). Advances in validation, risk and uncertainty assessment of bioanalytical methods. *J Pharm Biomed Anal* 55, 848–858.
- Ruiz, M., Jové, M., Schlüter, A., Casasnovas, C., Villarroya, F., Guilera, C., Ortega, F.J., Naudí, A., Pamplona, R., Gimeno, R., et al. (2015). Altered glycolipid and glycerophospholipid signaling drive inflammatory cascades in adrenomyeloneuropathy. *Human Molecular Genetics* 24, 6861–6876.
- Saadatmand, A.R., Tadjepisheh, S., Brockmöller, J., and Tzvetkov, M.V. (2012). The prototypic pharmacogenetic drug debrisoquine is a substrate of the genetically polymorphic organic cation transporter OCT1. *Biochemical Pharmacology* 83, 1427–1434.
- Saah, A.J., and Hoover, D.R. (1997). “Sensitivity” and “Specificity” Reconsidered: The Meaning of These Terms in Analytical and Diagnostic Settings. *Ann Intern Med* 126, 91–94.
- Sangster, J. (1997). *Octanol-water partition coefficients: fundamentals and physical chemistry* (Chichester; New York: Wiley).
- Sargsyan, Y., and Thoms, S. (2020). Staying in Healthy Contact: How Peroxisomes Interact with Other Cell Organelles. *Trends in Molecular Medicine* 26, 201–214.
- Sassa, T., and Kihara, A. (2014). Metabolism of Very Long-Chain Fatty Acids: Genes and Pathophysiology. *Biomol Ther (Seoul)* 22, 83–92.
- Sassa, T., Ohno, Y., Suzuki, S., Nomura, T., Nishioka, C., Kashiwagi, T., Hirayama, T., Akiyama, M., Taguchi, R., Shimizu, H., et al. (2013). Impaired Epidermal Permeability Barrier in Mice Lacking Elov11, the Gene Responsible for Very-Long-Chain Fatty Acid Production. *Mol Cell Biol* 33, 2787–2796.
- Saugstad, O.D. (1975). Hypoxanthine as a measurement of hypoxia. *Pediatr Res* 9, 158–161.
- Sawant, P.D., Kumar, S.A., Wankhede, S., and Rao, D.D. (2018). Creatinine as a normalization factor to estimate the representativeness of urine sample - Intra-subject and inter-subject variability studies. *Appl Radiat Isot* 136, 121–126.

9.0 References

- Schmidt, A., Karas, M., and Dülcks, T. (2003). Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *Journal of the American Society for Mass Spectrometry* *14*, 492–500.
- Schrimpe-Rutledge, A.C., Codreanu, S.G., Sherrod, S.D., and McLean, J.A. (2016). Untargeted metabolomics strategies – Challenges and Emerging Directions. *J Am Soc Mass Spectrom* *27*, 1897–1905.
- Schwaiger, M., Schoeny, H., Abiead, Y.E., Hermann, G., Rampler, E., and Koellensperger, G. (2019). Merging metabolomics and lipidomics into one analytical run. *Analyst* *144*, 220–229.
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., and Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol* *48*, 2097–2098.
- Seitz, T., Stalman, R., Dalila, N., Chen, J., Pojar, S., Dos Santos Pereira, J.N., Krätzner, R., Brockmöller, J., and Tzvetkov, M.V. (2015). Global genetic analyses reveal strong inter-ethnic variability in the loss of activity of the organic cation transporter OCT1. *Genome Med* *7*, 56.
- Sergeant, S., Rahbar, E., and Chilton, F.H. (2016). Gamma-linolenic acid, Dihommo-gamma linolenic, Eicosanoids and Inflammatory Processes. *Eur J Pharmacol* *785*, 77–86.
- Sevim Bayrak, C., and Itan, Y. (2020). Identifying disease-causing mutations in genomes of single patients by computational approaches. *Hum Genet* *139*, 769–776.
- Seymour, C.W., Carlbom, D., Cooke, C.R., Watkins, T.R., Bulger, E.M., Rea, T.D., and Baird, G.S. (2011). Temperature and time stability of whole blood lactate: implications for feasibility of pre-hospital measurement. *BMC Research Notes* *4*, 169.
- Shabir, G.A. (2003). Validation of high-performance liquid chromatography methods for pharmaceutical analysis. Understanding the differences and similarities between validation requirements of the US Food and Drug Administration, the US Pharmacopeia and the International Conference on Harmonization. *J Chromatogr A* *987*, 57–66.
- Sheng, C., Jungverdorben, J., Wiethoff, H., Lin, Q., Flitsch, L.J., Eckert, D., Hebisch, M., Fischer, J., Kesavan, J., Weykopf, B., et al. (2018). A stably self-renewing adult blood-derived induced neural stem cell exhibiting patternability and epigenetic rejuvenation. *Nat Commun* *9*.
- Shih, Y.C., Prausnitz, J.M., and Blanch, H.W. (1992). Some characteristics of protein precipitation by salts. *Biotechnol Bioeng* *40*, 1155–1164.
- Shriver, L.P. (2016). Mass Spectrometry-Based Metabolomics: A Practical Guide. *J. Am. Soc. Mass Spectrom.* *27*, 1–2.
- Shu, Y., Leabman, M.K., Feng, B., Mangravite, L.M., Huang, C.C., Stryke, D., Kawamoto, M., Johns, S.J., DeYoung, J., Carlson, E., et al. (2003). Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. *Proc Natl Acad Sci U S A* *100*, 5902–5907.
- Si-Hung, L., Causon, T.J., and Hann, S. (2017). Comparison of fully wettable RPLC stationary phases for LC-MS-based cellular metabolomics. *Electrophoresis* *38*, 2287–2295.
- Silva, L.P., Lorenzi, P.L., Purwaha, P., Yong, V., Hawke, D.H., and Weinstein, J.N. (2013). Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines. *Anal Chem* *85*, 9536–9542.
- Siskos, A.P., Jain, P., Römisch-Margl, W., Bennett, M., Achaintre, D., Asad, Y., Marney, L., Richardson, L., Koulman, A., Griffin, J.L., et al. (2017). Inter-laboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma. *Anal Chem* *89*, 656–665.
- Slade, E., Irvin, M.R., Xie, K., Arnett, D.K., Claas, S.A., Kind, T., Fardo, D.W., and Graf, G.A. (2021). Age and sex are associated with the plasma lipidome: findings from the GOLDN study. *Lipids in Health and Disease* *20*, 30.
- Sleno, L., and Volmer, D.A. (2004). Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry* *39*, 1091–1112.
- Smith, S. (1994). The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J* *8*, 1248–1259.
- Sonnenberg, R.A., Naz, S., Cougnaud, L., and Vuckovic, D. (2019). Comparison of underivatized silica and zwitterionic sulfobetaine hydrophilic interaction liquid chromatography stationary phases for global metabolomics of human plasma. *J Chromatogr A* *1608*, 460419.
- Spagou, K., Tsoukali, H., Raikos, N., Gika, H., Wilson, I.D., and Theodoridis, G. (2010). Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. *Journal of Separation Science* *33*, 716–727.
- Steinman, L., and Zamvil, S.S. (2005). Virtues and pitfalls of EAE for the development of therapies for multiple sclerosis. *Trends in Immunology* *26*, 565–571.
- Stevens, V.L., Hoover, E., Wang, Y., and Zanetti, K.A. (2019). Pre-Analytical Factors that Affect Metabolite Stability in Human Urine, Plasma, and Serum: A Review. *Metabolites* *9*.
- Stevenson, D.A., and Carey, J.C. (2004). Contribution of malformations and genetic disorders to mortality in a children's hospital. *Am J Med Genet A* *126A*, 393–397.
- Stokvis, E., Rosing, H., and Beijnen, J.H. (2005). Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not? *Rapid Commun Mass Spectrom* *19*, 401–407.
- Suhre, K., Shin, S.-Y., Petersen, A.-K., Mohny, R.P., Meredith, D., Wägele, B., Altmair, E., Cardiogram, Deloukas, P., Erdmann, J., et al. (2011). Human metabolic individuality in biomedical and pharmaceutical research. *Nature* *477*, 54–60.
- Sullivan, K.E. (2021). The scary world of variants of uncertain significance (VUS): A hitchhiker's guide to interpretation. *Journal of Allergy and Clinical Immunology* *147*, 492–494.

9.0 References

- Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., et al. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 3, 211–221.
- Suzuki, Y., Shimozawa, N., Yajima, S., Inoue, K., Orii, T., and Kondo, N. (1996). Incidence of peroxisomal disorders in Japan. *Jpn J Hum Genet* 41, 167–175.
- Sysi-Aho, M., Katajamaa, M., Yetukuri, L., and Orešič, M. (2007). Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics* 8, 93.
- Tabor, H. (1954). Metabolic Studies on Histidine, Histamine, and Related Imidazoles. *Pharmacol Rev* 6, 299–343.
- Takeuchi, K., and Reue, K. (2009). Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. *American Journal of Physiology-Endocrinology and Metabolism* 296, E1195–E1209.
- Tang, D.-Q., Zou, L., Yin, X.-X., and Ong, C.N. (2016). HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS. *Mass Spectrom Rev* 35, 574–600.
- Tansey, W.P. (2006). Freeze-thaw lysis for extraction of proteins from Mammalian cells. *CSH Protoc* 2006.
- Therrell, B.L., Padilla, C.D., Loeber, J.G., Kneisser, I., Saadallah, A., Borrajo, G.J.C., and Adams, J. (2015). Current status of newborn screening worldwide: 2015. *Seminars in Perinatology* 39, 171–187.
- Tome, T., Žigart, N., Časar, Z., and Obreza, A. (2019). Development and Optimization of Liquid Chromatography Analytical Methods by Using AQbD Principles: Overview and Recent Advances. *Org. Process Res. Dev.* 23, 1784–1802.
- Trammell, S.A., and Brenner, C. (2013). Targeted, LCMS-based Metabolomics for Quantitative Measurement of NAD⁺ Metabolites. *Comput Struct Biotechnol J* 4.
- Tsakelidou, E., Virgiliou, C., Valianou, L., Gika, H.G., Raikos, N., and Theodoridis, G. (2017). Sample Preparation Strategies for the Effective Quantitation of Hydrophilic Metabolites in Serum by Multi-Targeted HILIC-MS/MS. *Metabolites* 7.
- Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., and Arita, M. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* 12, 523–526.
- Tsugawa, H., Ikeda, K., Tanaka, W., Senoo, Y., Arita, M., and Arita, M. (2017). Comprehensive identification of sphingolipid species by in silico retention time and tandem mass spectral library. *Journal of Cheminformatics* 9, 19.
- Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., Okahashi, N., Yamada, Y., Tada, I., Bonini, P., et al. (2020). A lipidome atlas in MS-DIAL 4. *Nat Biotechnol* 38, 1159–1163.
- Turro, E., Astle, W.J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H.L., Sanchis-Juan, A., Frontini, M., Thys, C., et al. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* 583, 96–102.
- Tzvetkov, M.V., Matthaei, J., Pojar, S., Faltraco, F., Vogler, S., Prukop, T., Seitz, T., and Brockmüller, J. (2018). Increased Systemic Exposure and Stronger Cardiovascular and Metabolic Adverse Reactions to Fenoterol in Individuals with Heritable OCT1 Deficiency. *Clin Pharmacol Ther* 103, 868–878.
- Ulmer, C.Z., Jones, C.M., Yost, R.A., Garrett, T.J., and Bowden, J.A. (2018). Optimization of Folch, Bligh-Dyer, and Matyash Sample-to-Extraction Solvent Ratios for Human Plasma-Based Lipidomics Studies. *Anal Chim Acta* 1037, 351–357.
- Van Damme, T., Lachová, M., Lynen, F., Szucs, R., and Sandra, P. (2014). Solid-phase extraction based on hydrophilic interaction liquid chromatography with acetone as eluent for eliminating matrix effects in the analysis of biological fluids by LC-MS. *Anal Bioanal Chem* 406, 401–407.
- Vasireddy, V., Uchida, Y., Salem, N., Jr, Kim, S.Y., Mandal, M.N.A., Reddy, G.B., Bodepudi, R., Alderson, N.L., Brown, J.C., Hama, H., et al. (2007). Loss of functional ELOVL4 depletes very long-chain fatty acids (\geq C28) and the unique ω -O-acylceramides in skin leading to neonatal death. *Human Molecular Genetics* 16, 471–482.
- Vass, A., Robles-Molina, J., Pérez-Ortega, P., Gilbert-López, B., Dernovics, M., Molina-Díaz, A., and García-Reyes, J.F. (2016). Study of different HILIC, mixed-mode, and other aqueous normal-phase approaches for the liquid chromatography/mass spectrometry-based determination of challenging polar pesticides. *Anal Bioanal Chem* 408, 4857–4869.
- Verhoeven, N.M., Kulik, W., van den Heuvel, C.M.M., and Jakobs, C. (1995). Pre- and postnatal diagnosis of peroxisomal disorders using stable-isotope dilution gas chromatography - mass spectrometry. *J Inherit Metab Dis* 18, 45–60.
- Verhoeven, N.M., Wanders, R.J.A., Poll-The, B.T., Saudubray, J.-M., and Jakobs, C. (1998). The metabolism of phytanic acid and pristanic acid in man: A review. *Journal of Inherited Metabolic Disease* 21, 697–728.
- Vijay Kapoore, R., Coyle, R., A. Staton, C., J. Brown, N., and Vaidyanathan, S. (2017). Influence of washing and quenching in profiling the metabolome of adherent mammalian cells: a case study with the metastatic breast cancer cell line MDA-MB-231. *Analyst* 142, 2038–2049.
- Vijlder, T.D., Valkenburg, D., Lemièrre, F., Romijn, E.P., Laukens, K., and Cuyckens, F. (2018). A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. *Mass Spectrometry Reviews* 37, 607–629.
- Vinayavekhin, N., and Saghatelian, A. (2010). Untargeted Metabolomics. *Current Protocols in Molecular Biology* 90, 30.1.1-30.1.24.
- Vingara, L.K., Yu, H.J., Wagshul, M.E., Serafin, D., Christodoulou, C., Pelczer, I., Krupp, L.B., and Maletić-Savatić, M. (2013). Metabolomic approach to human brain spectroscopy identifies associations between clinical features and the frontal lobe metabolome in multiple sclerosis. *Neuroimage* 82, 586–594.

9.0 References

- Vinkškel, M., Writzl, K., Maver, A., and Peterlin, B. (2021). Improving diagnostics of rare genetic diseases with NGS approaches. *J Community Genet* 12, 247–256.
- Vogel, B.H., Bradley, S.E., Adams, D.J., D'Aco, K., Erbe, R.W., Fong, C., Iglesias, A., Kronn, D., Levy, P., Morrissey, M., et al. (2015). Newborn screening for X-linked adrenoleukodystrophy in New York State: diagnostic protocol, surveillance protocol and treatment guidelines. *Mol Genet Metab* 114, 599–603.
- Vuckovic, D. (2012). Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. *Anal Bioanal Chem* 403, 1523–1548.
- Vuckovic, D. (2013). Chapter 4 - Sample Preparation in Global Metabolomics of Biological Fluids and Tissues. In *Proteomic and Metabolomic Approaches to Biomarker Discovery*, H.J. Issaq, and T.D. Veenstra, eds. (Boston: Academic Press), pp. 51–75.
- Wahrheit, J., and Heinzle, E. (2014). Quenching methods for the analysis of intracellular metabolites. *Methods Mol Biol* 1104, 211–221.
- Walter, N.G., and Engelke, D.R. (2002). Ribozymes: Catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London)* 49, 199–203.
- Wanders, R.J.A. (2004). Metabolic and molecular basis of peroxisomal disorders: A review. *American Journal of Medical Genetics Part A* 126A, 355–375.
- Wanders, R.J.A., and Waterham, H.R. (2006). Peroxisomal disorders: The single peroxisomal enzyme deficiencies. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1763, 1707–1720.
- Wanders, R.J.A., Komen, J., and Ferdinandusse, S. (2011). Phytanic acid metabolism in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1811, 498–507.
- Wang, C., Wang, M., and Han, X. (2015). Applications of Mass Spectrometry for Cellular Lipid Analysis. *Mol Biosyst* 11, 698–713.
- Wang, G., Dinkins, M., He, Q., Zhu, G., Poirier, C., Campbell, A., Mayer-Proschel, M., and Bieberich, E. (2012). Astrocytes Secrete Exosomes Enriched with Proapoptotic Ceramide and Prostate Apoptosis Response 4 (PAR-4). *J Biol Chem* 287, 21384–21395.
- Wang, L., Naser, F.J., Spalding, J.L., and Patti, G.J. (2019). A Protocol to Compare Methods for Untargeted Metabolomics. In *Metabolic Signaling: Methods and Protocols*, S.-M. Fendt, and S.Y. Lunt, eds. (New York, NY: Springer), pp. 1–15.
- Wang, M., Wang, C., and Han, X. (2017). Selection of internal standards for accurate quantification of complex lipid species in biological extracts by electrospray ionization mass spectrometry—What, how and why? *Mass Spectrom Rev* 36, 693–714.
- Wangler, M.F., Hubert, L., Donti, T.R., Ventura, M.J., Miller, M.J., Braverman, N., Gawron, K., Bose, M., Moser, A.B., Jones, R.O., et al. (2018). A metabolomic map of Zellweger spectrum disorders reveals novel disease biomarkers. *Genet Med* 20, 1274–1283.
- Want, E.J., O'Maille, G., Smith, C.A., Brandon, T.R., Uritboonthai, W., Qin, C., Trauger, S.A., and Siuzdak, G. (2006). Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal Chem* 78, 743–752.
- Ward, J.H.Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, 236–244.
- Waterham, H.R., Ferdinandusse, S., and Wanders, R.J.A. (2016). Human disorders of peroxisome metabolism and biogenesis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1863, 922–933.
- Weinhofer, I., Rommer, P., Zierfuss, B., Altmann, P., Foiani, M., Heslegrave, A., Zetterberg, H., Gleiss, A., Musolino, P.L., Gong, Y., et al. (2021). Neurofilament light chain as a potential biomarker for monitoring neurodegeneration in X-linked adrenoleukodystrophy. *Nat Commun* 12, 1816.
- Weller, P.F. (2016). leukocyte lipid bodies — structure and function as “eicososomes.” *Trans Am Clin Climatol Assoc* 127, 328–340.
- Wernisch, S., and Pennathur, S. (2016). Evaluation of coverage, retention patterns, and selectivity of seven liquid chromatographic methods for metabolomics. *Anal Bioanal Chem* 408, 6079–6091.
- Westerberg, R., Tvrdik, P., Undén, A.-B., Månsson, J.-E., Norlén, L., Jakobsson, A., Holleran, W.H., Elias, P.M., Asadi, A., Flodby, P., et al. (2004). Role for ELOVL3 and fatty acid chain length in development of hair and skin function. *J Biol Chem* 279, 5621–5629.
- Willemsen, M.A., Harting, I., and Wevers, R.A. (2016). Neurometabolic disorders. *Neurol Clin Pract* 6, 348–357.
- Wilson, J.M.G and Jungner, G (1968). Principles and practice of screening for disease. *J R Coll Gen Pract* 16, 318.
- Winkler, H. (1920). *Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche* (Jena : G. Fischer,).
- Winterbourn, C.C. (1995). Toxicity of iron and hydrogen peroxide: the Fenton reaction. *Toxicology Letters* 82–83, 969–974.
- Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djombou, Y., Mandal, R., Aziat, F., Dong, E., et al. (2013). HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* 41, D801-807.
- Wong, B.H., and Silver, D.L. (2020). Mfsd2a: A Physiologically Important Lysolipid Transporter in the Brain and Eye. In *Lipid Transfer in Lipoprotein Metabolism and Cardiovascular Disease*, X.-C. Jiang, ed. (Singapore: Springer), pp. 223–234.
- Wright, C.F., FitzPatrick, D.R., and Firth, H.V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet* 19, 253–268.

9.0 References

- Wyss, M., and Kaddurah-Daouk, R. (2000). Creatine and Creatinine Metabolism. *Physiological Reviews* 80, 1107–1213.
- Xia, J., Psychogios, N., Young, N., and Wishart, D.S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* 37, W652–W660.
- Xiao, J.F., Zhou, B., and Ressom, H.W. (2012). Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *Trends Analyt Chem* 32, 1–14.
- Yang, W., Chen, Y., Xi, C., Zhang, R., Song, Y., Zhan, Q., Bi, X., and Abliz, Z. (2013a). Liquid Chromatography–Tandem Mass Spectrometry-Based Plasma Metabonomics Delineate the Effect of Metabolites' Stability on Reliability of Potential Biomarkers. *Anal. Chem.* 85, 2606–2610.
- Yang, Y., Cruickshank, C., Armstrong, M., Mahaffey, S., Reisdorph, R., and Reisdorph, N. (2013b). New sample preparation approach for mass spectrometry-based profiling of plasma results in improved coverage of metabolome. *J Chromatogr A* 1300, 217–226.
- Yao, C.-H., Liu, G.-Y., Yang, K., Gross, R.W., and Patti, G.J. (2016a). Inaccurate quantitation of palmitate in metabolomics and isotope tracer studies due to plastics. *Metabolomics* 12.
- Yao, C.-H., Grider, R.F., Mahieu, N.G., Liu, G.-Y., Chen, Y.-J., Wang, R., Singh, M., Potter, G.S., Gross, R.W., Schaefer, J., et al. (2016b). Exogenous Fatty Acids are the Preferred Source of Membrane Lipids in Proliferating Fibroblasts. *Cell Chem Biol* 23, 483–493.
- Yee, S.W., Giacomini, M.M., Hsueh, C.-H., Weitz, D., Liang, X., Goswami, S., Kinchen, J.M., Coelho, A., Zur, A.A., Mertsch, K., et al. (2016). Metabolomic and Genome-wide Association Studies Reveal Potential Endogenous Biomarkers for OATP1B1. *Clin Pharmacol Ther* 100, 524–536.
- Zeharia, A., Ebberink, M.S., Wanders, R.J.A., Waterham, H.R., Gutman, A., Nissenkorn, A., and Korman, S.H. (2007). A novel PEX12 mutation identified as the cause of a peroxisomal biogenesis disorder with mild clinical phenotype, mild biochemical abnormalities in fibroblasts and a mosaic catalase immunofluorescence pattern, even at 40 degrees C. *J Hum Genet* 52, 599–606.
- Zhang, B., Xie, M., Bruschiweiler-Li, L., and Brüschiweiler, R. (2016). Nanoparticle-Assisted Removal of Protein in Human Serum for Metabolomics Studies. *Anal Chem* 88, 1003–1007.
- Zhang, X., Dong, J., and Raftery, D. (2020). Five Easy Metrics of Data Quality for LC–MS-Based Global Metabolomics. *Anal. Chem.* 92, 12925–12933.
- Zhou, B., Xiao, J.F., Tuli, L., and Ressom, H.W. (2012). LC-MS-based metabolomics. *Mol Biosyst* 8, 470–481.
- Zhu, X., Petrovski, S., Xie, P., Ruzzo, E.K., Lu, Y.-F., McSweeney, K.M., Ben-Zeev, B., Nissenkorn, A., Anikster, Y., Oz-Levi, D., et al. (2015). Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. *Genet Med* 17, 774–781.
- Zöllner, N., and Kirsch, K. (1962). Über die quantitative Bestimmung von Lipoiden (Mikromethode) mittels der vielen natürlichen Lipoiden (allen bekannten Plasmalipoiden) gemeinsamen Sulfo-phosphovanillin-Reaktion. *Z. Gesamte Exp. Med.* 135, 545–561.
- Züllig, T., Trötzmüller, M., and Köfeler, H.C. (2020). Lipidomics from sample preparation to data analysis: a primer. *Anal Bioanal Chem* 412, 2191–2209.

10.0 Supplemental material

10.1 Glossary of mass spectrometry terms

Abundance/signal abundance/signal intensity: Semiquantitative, arbitrary unit, measuring the number of ions detected by the MS detector. The unit can broadly relate to the concentration of analytes in a sample but is dependent on the mass spectrometer and ionization efficiency of the analyte in the particular sample

Chromatogram: A graph of mass spectrometry signal intensity, plotted versus time

Collision cell: A compartment in a mass spectrometer that is used to purposefully fragment molecular ions

Continuous acquisition mode: Data acquisition mode in mass spectrometry, where mass/charge ratios are recorded continuously.

Electrospray ionization/ESI: ESI is a soft ionization method that uses a high-voltage charged metal capillary that can produce a spray of charged ions from analytes.

Extracted mass chromatogram/XIC: A chromatogram of mass spectrometry intensity signals summed over a certain mass range (mass tolerance)

InChIKey: International chemical identifier key; an up to 27 ASCII-character long unique structural chemical identifier developed by NIST

Internal standard/ISTD: a substance added in the same amount into the samples, commonly used to control the quality of measurement or to normalize interferences in the analysis. In mass spectrometry, internal standards are mostly using stable isotopes such as deuterium or Carbon-13.

Ionization efficiency: Degree of how well analytes molecules form molecular ions in an ionization source. May be estimated from the matrix factor.

Ionization source: Compartment of the mass spectrometer, which is responsible for the formation of ions from analytes. In LC-MS, mostly ESI is used.

Ion suppression: Reduction of signal intensity in mass spectrometers, based on the reduction of ionization efficiency, especially caused by compounds such as phospholipids in ionization sources

Isobaric compound: A compound with its molecular ion that has the same mass/charge ratio as another compound within a certain mass tolerance.

Liquid chromatography: Analysis method that uses a hollow column most often filled with a chemically modified silica gel for the separation of analytes dissolved in a liquid mobile phase.

Lockmass/Lockspray: A method for the intra-run mass correction using an additional capillary and infusing a stable compound of a known mass (such as leucine-enkephalin)

LOESS algorithm: Locally weighted scatter-plot smoother; A method for the intensity normalization of metabolomics data, based on quality control samples.

10.0 Supplemental material

logP: decadic logarithm of water octanol partition coefficient, a measure for lipophilicity (positive values) or hydrophilicity (negative value).

Mass spectrometer: An analytic device able to determine the mass/charge ratio composition of an analytical mixture. A mass spectrometer typically consists at least of an ionization source, a mass analyzer, and a detector; some possess a collision cell or an additional mass analyzer.

Mass tolerance: range of mass/charge ratio, within mass spectrometry signals, are extracted to generate an extracted mass chromatogram (XIC)

(Sample) Matrix, matrices: Solution, mixture or solid that contains an analyte (e.g., blood, tissue, cells, methanol). Confounds matrix effects, by contributing to ion suppression.

MSe/All ion fragmentation: Acquisition of data from fragmented analyte ions by unselectively fragmenting all incoming molecular ions at the same time

Quadrupole time of flight mass spectrometer/ QToF: a mass spectrometer type with a quadrupole mass analyzer, followed typically by a collision cell and completed with a time of flight mass analyzer.

(Triple) Quadrupole mass analyzer: A mass analyzer, based on four metal rods applying alternate electromagnetic fields that can either let an ion of a certain mass range pass or not pass. Mostly used as arrangement of a first quadrupole, a sequential collision cell (which is a quadrupole) and a third quadrupole. Commonly used for quantification including targeted metabolomics. This mass analyzer typically has lower mass resolution.

SMILES: Simplified molecular input line entry specification; an ASCII-character based unique molecular identifier predating the InChIKey.

Time of flight mass analyzer/Tof: A mass analyzer that measures mass/charge ratios of ions by measuring their flight time in an evacuated tube. Commonly used for fast high-resolution measurements with higher resolution than quadrupole analyzers, but lower signal/noise ratio, mostly used for untargeted metabolomics.

Total ion chromatogram/TIC: A chromatogram of mass spectrometry intensity signals summed over all masses.

10.2 Lipid nomenclature

Lipids are structured biomolecules that commonly consist of a linker molecule (mostly glycerol), a lipid class defining headgroup, and varying acyl-chains (fatty acids) that are either esterified to the linker (acyl-bond, a), bound by an ether-bond (ethyl-bond, e) or bound by an amide-bond (e.g., in sphingolipids). In some cases, ether-bonds are indicated by “-O-”.

An overview of different lipid structures can be found in **Figure 66**. Lipids that have less acyl groups than typical for their lipid class carry the prefix “lyso”.

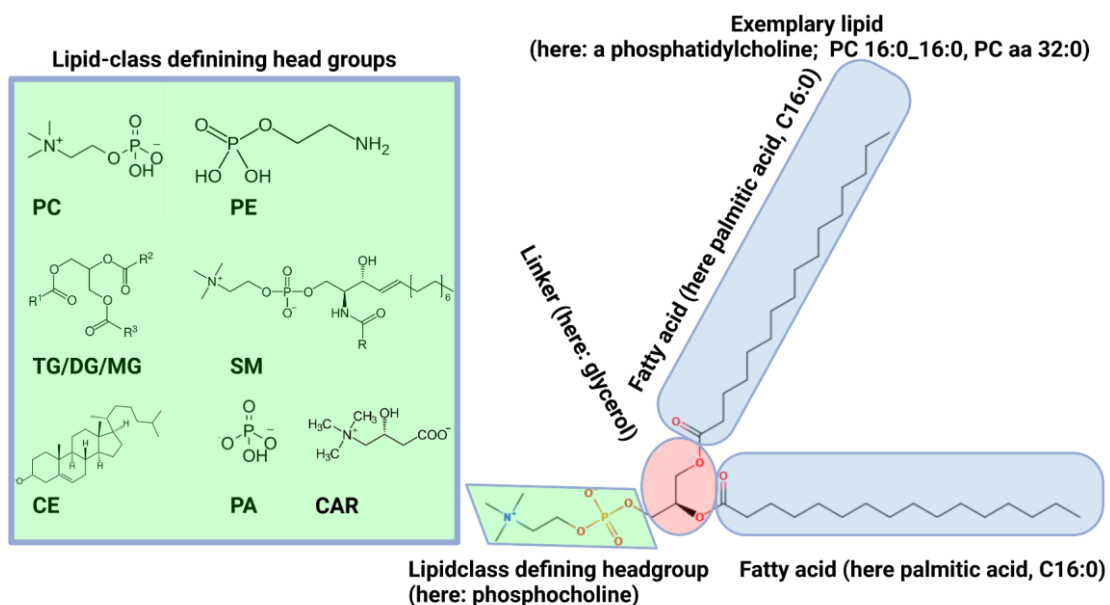


Figure 66 Scheme of typical lipid structures. Created with BioRender.com.

Shorthand lipid nomenclature starts with the class abbreviation (e.g., PC), then follows by the number of carbon atoms in the fatty acid separated by “:” from the number of double bonds in the respective chain (e.g., Dipalmitoyl-phosphatidylcholine=PC 16:0_16:0). In some cases (e.g., in the Biocrates AbsoluteIDQ p180-kit), the sum over all carbon atoms and double bonds is used (where a lipid e.g., PC 16:0_16:0 equals to a shorthand notation of PC aa 32:0).

Modifications by hydroxy groups can be indicated by attaching a “-OH”.

10.0 Supplemental material

Common abbreviation for lipids can be found in **Table 18**.

Table 18 Common abbreviations of lipids

Abbreviation	Name of lipid class
CAR/C	Acyl-carnitines
CE	Cholesteryl-ester
DG/DAG	Diacylglyceride
LPC	Lysophosphatidycholine
LPE	Lysophosphatidylethanolamine
MG/MAG	Monoacylglyceride
PA	Phosphatidylglyceride
PC	Phosphatidycholine
PE	Phosphatidylethanolamine
SM	Sphingomyeline
TG/TAG	Triacylglyceride

10.3 Supplementary Data

10.3.1 Human metabolite mass and retention time library

As described in chapter 2.2.9 (p. 26) this human metabolite library was acquired using human metabolite reference standards from MetaSci. The method used a Waters I-Class UPLC with an G2-S QToF mass spectrometer with an ESI ion source. As mobile phases were used: A: 90% 30mM ammonium acetate pH 7.0 in water, 10% acetonitrile; B: 10% 30mM ammonium acetate pH 7.0 in water, 90% acetonitrile. Mass spectrometry data was acquired using positive and negative ionization sensitivity mode in separate runs from 100 Da to 800 Da in continuous MSe acquisition mode. Thus, fragments were generated by alternating acquisitions of unfragmented ions to fragmenting all incoming ions by ramping fragmentation energy between 10-40 V (MSe mode). Chromatograms, as well as fragmentation spectra, are available inside a Waters UNIFI database, but important identifiers for further use (e.g., in other, open source-based programs) are displayed here. Only metabolites that are detectable in at least the reference standards using our method are displayed here. Metabolites that cannot be distinguished by the method, are displayed as one position in the database. Other data formats such as MSP are available upon request.

10.0 Supplemental material

Table 19 Human metabolite compound library in ZIC-HILIC positive ionization mode. Human metabolite reference standards from MetaSci in a concentration of 100 μ M were acquired using the ZIC-HILIC-MS method in positive ionization mode. Only $[M+H]^+$ molecular ions were used for identification. The table shows an exempt from the UNIFI-Database and only metabolites detected in reference standards are shown. Other dataformats (such as MSP) are also available.

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
1-Methyl Adenosine	Purine nucleosides	282.1202	C11H15N5O4	GFYLSDSUCHVORB-JOLDIKRXSA-N	11.22	130.05029,150.077
2-Amino-4-methylsulfonimidoyl butanoic acid	Amino acids and derivatives	181.0647	C5H12N2O3S	SXTAYKAGBXMACB-UHFFFAOYSA-N	11.26	
2-Hydroxycaproic acid/2-Hydroxyisocaproic acid/2-Ethyl-2-Hydroxybutyric acid	Medium-chain fatty acids	133.0865	C6H12O3	NYHNVHGFPZAZGA-UHFFFAOYSA-N	4.35	
2-Methylbutyrylglycine	Amino acids and derivatives	160.0974	C7H13NO3	HOACIBQKYRHBOW-UHFFFAOYSA-N	4.21	
2-Methylsuccinic acid	Branched fatty acids	133.0501	C5H8O4	WXUQAQHNMJWLGTG-UHFFFAOYSA-N	10.50	
3-Aminoisobutanoic acid	Amino acids and derivatives	104.0712	C4H9NO2	QCHPKSFMHDHPSNR-UHFFFAOYSA-N	9.99	
3-Chlorotyrosine	Amino acids and derivatives	216.0428	C9H10ClNO3	ACWBBAGYTKWBCD-UHFFFAOYSA-N	7.94	170.03628,199.01518
3-Hydroxy-3-methylglutaric acid	Hydroxy fatty acids	163.0607	C6H10O5	NPOAOTPXWNWTSU-UHFFFAOYSA-N	2.39	
3-Indolebutyric acid	3-alkylindoles	204.1025	C12H13NO2	JTEDVYBZBROSJT-UHFFFAOYSA-N	2.78	130.06513,186.09134,117.0573,89.05971,144.08078
3-Methoxytyramine	Methoxyphenols	168.1025	C9H13NO2	DIVQKHQLANKJQO-UHFFFAOYSA-N	9.26	151.07536,123.08037
3-Methyl-L-histidine	Amino acids and derivatives	170.0930	C7H11N3O2	JDHILDINMRGULE-LURJTMIESA-N	13.21	95.06037,81.04472
3-Methylxanthine	Xanthines	167.0569	C6H6N4O2	GMSNIKWQOZHGF-UHFFFAOYSA-N	4.18	124.05054,96.05562
3-Nitro-L- tyrosine	Amino acids and derivatives	227.0668	C9H10N2O5	FBTSQILOGYXGMD-LURJTMIESA-N	8.03	181.06118
3-Phenylpropionylglycine	Ketones	208.0974	C11H13NO3	XHSURMJJKAFELI-UHFFFAOYSA-N	4.14	91.05366,105.06938
4-Aminobutylacetamide	Carboximide acids	131.1185	C6H14N2O	KLZGKIDSEJWEDW-UHFFFAOYSA-N	15.32	114.09075,72.08044
4-Guanidinobutyric acid	Amino acids and derivatives	146.0924	C5H11N3O2	TUHVEAJXIMEOSA-UHFFFAOYSA-N	11.40	86.06004,69.03355,128.08166
4-Pyridoxic acid	Pyridinecarboxylic acids	184.0604	C8H9NO4	HXACOUQIXZGNBF-UHFFFAOYSA-N	2.68	148.03902
5-Aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside	Imidazole ribonucleosides and ribonucleotides	259.1043	C9H14N4O5	RTRQQBHATOEIAF-UUOKFMHZA-N	6.81	110.03414,127.06144,152.04545,82.04132
5-Aminolevulinic acid	Amino acids and derivatives	132.0661	C5H9NO3	ZGXJTSGNOSYLO-UHFFFAOYSA-N	12.51	114.05542,86.0606
5-Aminopentylacetamide	Carboxylic acid amides	145.1341	C7H16N2O	RMOIHAKNOFHOE-UHFFFAOYSA-N	11.85	128.1077,69.07066,86.09713
5-Hydroxy-3-indoleacetic acid	Indole-3-acetic acid derivatives	192.0655	C10H9NO3	DUUGKQCEGZLNZNO-UHFFFAOYSA-N	8.47	146.06026,174.05529
5-Methoxytryptamine	Tryptamines and derivatives	191.1179	C11H14N2O	JTEJPPKMYBDEMY-UHFFFAOYSA-N	7.34	174.09135,143.07272
5-Methoxy-tryptophan	Serotonins	235.1083	C12H14N2O3	KVNPSKDDJARYKK-JTQLQIEISA-N	7.16	218.08085,133.05176
5-Methylcytosine	Hydroxypyrimidines	126.0668	C5H7N3O	LRSASMSXMSNRBT-UHFFFAOYSA-N	6.19	83.05958
5-Methyluridine	Pyrimidine nucleosides	259.0930	C10H14N2O6	DWRXFEITVBNRMK-JXOAFINSAN-N	4.7	134.06094

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
5-Pregnenolone	Glucocorticoids, progestogens and derivatives	317.2481	C21H32O2	JERGUCJJOXJXHF-TVVWXWENSA-N	2.25	139.11174,109.10118,125.09609,151.11174,147.11683
Acetyl-5-hydroxytryptamine	Hydroxyindoles	219.1128	C12H14N2O2	MVAWJSDNICKHF-UHFFFAOYSA-N	2.9	160.07472,142.0642
Acetylcholine	Cholines	146.1181	C7H15NO2	OIPILFWXSMYKGL-UHFFFAOYSA-N	10.64	
Acetyl-D-mannosamine	Aminosaccharides	222.0978	C8H15NO6	OVRNDRQMDRJTHS-ZTVVOAFPASA-N	8.44	135.03148,133.04954
Acetylglucosamine	Aminosaccharides	222.0978	C8H15NO6	OVRNDRQMDRJTHS-DMHSOCPYSA-N	8.06	89.05912,133.08527,138.05495
Acetylglutamic acid	Amino acids and derivatives	190.0716	C7H11NO5	RFMMMVDNIPUKGG-UHFFFAOYSA-N	9.99	84.04439
Acetyl-L-arginine	Amino acids and derivatives	217.1301	C8H16N4O3	SNEIUMQYRCDYCH-LURJTMIESA-N	11.04	70.06513,158.08117,116.07061,71.0478
Acetyl-L-carnitine	Acyl carnitines	204.1236	C9H17NO4	RDHQFKQIGNGIED-UHFFFAOYSA-N	10.64	145.04915,85.0278
Acetyl-L-methionine	Amino acids and derivatives	192.0695	C7H13NO3S	XUYPLNMDZIRQH-UHFFFAOYSA-N	4.62	
Acetyl-L-ornithine	Amino acids and derivatives	175.1083	C7H14N2O3	JRLGPAXAGHMNOL-LURJTMIESA-N	11.39	70.06513
Acetyl-L-Proline	Amino acids and derivatives	158.0817	C7H11NO3	GNMSLDIYJOSUSW-LURJTMIESA-N	7.45	70.06513,85.07252
Acetyl-L-tyrosine	Amino acids and derivatives	224.0923	C11H13NO4	CAHKINHBCWCFCF-JTQLQIEISA-N	6.05	95.04914,147.04406
Acetylneuraminic acid	Sugar acids and derivatives	310.1138	C11H19NO9	SQVRNKJHWKZAKO-PFQGNLYSA-N	9.61	208.08156
Adenine	6-aminopurines	136.0618	C5H5N5	GFFGJBXGBJISGV-UHFFFAOYSA-N	4.93	119.03522
Adenosine	Purine nucleosides	268.1046	C10H13N5O4	QIRDTQYFTABQOQ-KQYNXXCUSA-N	4.93	136.06227
Adenosine diphosphate	Purine ribonucleoside diphosphates	428.0373	C10H15N5O10P2	XTWYTFMLZFPYCI-KQYNXXCUSA-N	10.23	136.06141,410.02613
Adenosine monophosphate	Purine ribonucleoside monophosphates	348.0709	C10H14N5O7P	UDMBCSSLTHHNCDC-KQYNXXCUSA-N	10.07	97.02774,136.06124,232.08216
Adenosine triphosphate	Purine ribonucleoside triphosphates	508.0036	C10H16N5O13P3	ZKHQWZAMRWXGA-KQYNXXCUSA-N	10.42	97.02274,136.06124
All trans-Retinal	Retinoids	285.2219	C20H28O	NCYCYXZNIZJOKI-OVSJKPMPASA-N	2.18	123.11683,135.08044,175.14813,135.11683,161.09609
Allantoic acid	Amino acids and derivatives	177.0618	C4H8N4O4	NUCLJNSWZCHRKL-UHFFFAOYSA-N	10.18	
alpha-Acetyl-L-glutamine	Amino acids and derivatives	189.0875	C7H12N2O4	KSMRODHGGIIXDV-UHFFFAOYSA-N	8.25	
alpha-Acetyl-L-lysine	Amino acids and derivatives	189.1239	C8H16N2O3	VEYYWZRYIYDQJM-UHFFFAOYSA-N	9.70	84.08001,175.10772,126.09134
Androsterone	Androgens and derivatives	291.2324	C19H30O2	QGXBDMJGAMFCBF-XYQQMQERSA-N	2.27	109.10118,123.11683,107.08553,121.10118,95.08553
Arginine	Amino acids and derivatives	175.1195	C6H14N4O2	ODKSFYDXXFIFQN-BYPYZUCNSA-N	19.22	70.0642,116.07061,133.09715
Asparagine	Amino acids and derivatives	133.0613	C4H8N2O3	DCXYFEDJOCDFNAF-REOHCLBNSA-N	11.22	85.02841,74.02365
Bilirubin	Bilirubins	585.2713	C33H36N4O6	BPYKTIZUTYGOLE-IFADSCNNSA-N	3.99	299.13902
Biotin	Biotin and derivatives	245.0960	C10H16N2O3S	YBJHBAHKTGYVGT-ZKWXMUHAASA-N	6.44	97.04014,227.08527

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Caffeine	Xanthines	195.0877	C8H10N4O2	RYYVLZVUVIJVGH-UHFFFAOYSA-N	2.75	123.04271
Calciferol	Vitamin D and derivatives	397.3471	C28H44O	MECHNRXZTMCUDQ-RKHKHRCZSA-N	2.23	119.08553,91.05243,105.06988,109.10118,145.10118
Capryloyl glycine	Amino acids and derivatives	202.1443	C10H19NO3	SAVLIGUQOQSOEP-UHFFFAOYSA-N	4.16	76.0393,158.08117,112.07569,114.05495,98.06004
Carbamazepine	Dibenzazepines	237.1022	C15H12N2O	FFGPTBGBLSHEPO-UHFFFAOYSA-N	2.57	195.06787,197.08352
Carnitine	Carnitines	162.1130	C7H15NO3	PHIQXFUZVPYII-LURJTMIESA-N	12.50	
Carnosine	Hybrid peptides	227.1144	C9H14N4O3	CQOVNPNJLQNMDC-ZETCQYMNSA-N	14.42	83.05965
Cholesterol 3-Sulfate	Cholesterols and derivatives	467.3195	C27H46O4S	BHYOQNUJELFYRT-DPAQBDFSA-N	2.33	109.10118,145.10118
Choline	Cholines	104.1064	C5H14NO	OEYIOHPDSNJKLS-UHFFFAOYSA-N	15.48	
cis-Aconitic acid	Tricarboxylic acids and derivatives	175.0243	C6H6O6	GTZCVFVGUGFEME-IWQZZHSRSA-N	11.13	
Citicoline	Pyrimidine ribonucleoside diphosphates	489.1152	C14H26N4O11P2	RZZPDXZPRHQCG-QJAKKHQSA-N	11.35	184.07296,264.03915
Citrulline	Amino acids and derivatives	176.1035	C6H13N3O3	RHGKLRLOHDJJDR-BYPYZUCNSA-N	11.22	82.03915
Corticosterone	21-hydroxysteroids	347.2222	C21H30O4	OMFXVFTZEKFBZ-HJTSIMOOSA-N	2.60	207.13796,163.11174
Cortisone	21-hydroxysteroids	361.2015	C21H28O5	MFYSYFVPBJMHGN-ZPOLXVWWSA-N	2.52	179.07027,197.08084,165.12739,207.13796
Creatine	Amino acids and derivatives	132.0768	C4H9N3O2	CVSVTCORWBXHQV-UHFFFAOYSA-N	10.89	90.05495
Creatine phosphate	Amino acids and derivatives	212.0436	C4H10N3O5P	DRBBFCLWYRJSJZ-UHFFFAOYSA-N	11.10	78.95939,110.03579,134.04687,177.0978,135.03179
Creatinine	Amino acids and derivatives	114.0662	C4H7N3O	DDRJAANPRJIHGJ-UHFFFAOYSA-N	6.64	69.04472
Cyclic Adenosine monophosphate	3',5'-cyclic purine nucleotides	330.0604	C10H12N5O6P	IVOMOUWHPKRLKQYNXXCUSA-N	6.93	136.0613,89.05906
Cystathionine	Amino acids and derivatives	223.0753	C7H14N2O4S	ILRYLPWNYFXEMH-WHFBIAKZSA-N	12.83	
Cysteic acid	Amino acids and derivatives	170.0123	C3H7NO5S	XVOYSCVBGLVSOL-UHFFFAOYSA-N	10.10	124.00562
Cysteine	Amino acids and derivatives	122.0276	C3H7NO2S	XUJNEKJLAYXESH-REOHCLBNSA-N	10.50	74.02384,88.093,105.00048
Cystine	Amino acids and derivatives	241.0317	C6H12N2O4S2	LEVWYRKDKASIDU-IMJSIDKUSA-N	13.45	206.02085,224.03127,222.01383,204.00327,132.94628
Cytidine	Pyrimidine nucleosides	244.0934	C9H13N3O5	UHDGCWIWMRCDJ-XVFCMESISA-N	7.74	112.05008
Cytidine triphosphate	Monosaccharides	483.9924	C9H16N3O14P3	PCDQPRRSZKQHHS-XVFCMESISA-N	11.00	97.02787
Cytosine	Pyrimidones	112.0505	C4H5N3O	OPTASPLRGRNAP-UHFFFAOYSA-N	7.71	95.02399,69.04386
Dehydroprogesterone	20-oxosteroids	313.2168	C21H28O2	VRRHHTISESGZFN-RKFFNLMNSA-N	2.32	163.11174,191.14304,269.18999
Deoxyadenosine	Purine 2'-deoxyribonucleosides	252.1097	C10H13N5O3	OLXZPDWKRNYJZ-RRKRQDMNSA-N	4.38	136.06177,119.03592,73.02841

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Deoxycytidine	Pyrimidine 2'-deoxyribonucleosides	228.0984	C9H13N3O4	CKTSBUTUHBMZGZ-SHYZEUFSA-N	6.70	112.05008,117.05462
Deoxyguanosine	Purine 2'-deoxyribonucleosides	268.1046	C10H13N5O4	YKBGVTZYEHREMT-KVQBGUIXSA-N	6.97	152.05573,135.02923
Deoxyguanosine monophosphate	Purine deoxyribonucleoside monophosphates	348.0709	C10H14N5O7P	LTFMZDNNPPEQNG-KVQBGUIXSA-N	10.76	135.02932,152.05591
Deoxyinosine	Purine 2'-deoxyribonucleosides	253.0937	C10H12N4O4	VGONTNSXDCQUGY-RRKCRQDMSA-N	5.61	137.04579,117.05462,225.07441
Deoxyuridine	Pyrimidine 2'-deoxyribonucleosides	229.0825	C9H12N2O5	MXHRCPNRJAMMIM-SHYZEUFSA-N	4.42	73.02841,87.04406
Dimethylbenzimidazole	Benzimidazoles	147.0917	C9H10N2	LJUQGASMPRMWW-UHFFFAOYSA-N	2.55	131.06037,132.0682
Dimyristoyl-glycerol-3-phosphoethanolamine	Phosphatidylethanolamines	636.4604	C33H66NO8P	NEZDNQCXEZDCBI-WJOKGBTCSA-N	4.43	184.07332,447.25062,495.44079,488.27717
Diosmin	Flavonoid O-glycosides	609.1820	C28H32O15	GZSOSUNBTXUMFQ-UHFFFAOYSA-N	4.11	301.06982,119.0346,94.03921
Estradiol	Estrogens and derivatives	273.1855	C18H24O2	VOXZDWNPVJITMN-ZBRFXRBCSA-N	2.34	121.06479,107.04914,147.08044,151.11174,93.06988
Estriol	Estrogens and derivatives	289.1804	C18H24O3	PROQIPRRNZUXQM-ZXXIGWHRSA-N	2.44	
Flavin Adenine Dinucleotide	Flavin nucleotides	786.1650	C27H33N9O15P2	VWWQXMAJTJZDQX-UYBVJOGSSA-N	8.40	439.10089,136.06139,348.06967
Flavone	Flavones	223.0759	C15H10O2	VHBFQKBNRFLZ-UHFFFAOYSA-N	2.33	129.03349,77.03858,147.04406,93.03349,79.05423
Folic acid	Amino acids and derivatives	442.1475	C19H19N7O6	OVBPILPVIDEAO-LBPRGKRZSA-N	10.34	295.0938
Folinic acid	Tetrahydroptericoic acids and derivatives	474.1737	C20H23N7O7	VVIAGPKUTFNRDU-UHFFFAOYSA-N	9.99	327.11935,299.12448
Fructosebiphosphate	Monosaccharides	341.0039	C6H14O12P2	WSMBXSQDFPTODV-JGWLITMVSA-N	11.50	97.02841,151.0601
Glucosamine	Monosaccharides	180.0872	C6H13NO5	MSWZFWKMSRAUBD-HOWGCPQDSA-N	13.94	99.044006
Glucosamine 6-phosphate	Monosaccharides	260.0535	C6H14NO8P	XHMJOUIAFHJHBW-UKFBFLRUSA-N	13.05	126.05449,89.05971,206.02129
Glucose 1-phosphate/Fructose-6-phosphate/Glucose 6-phosphate/Galactose-1-phosphate	Monosaccharides	261.0376	C6H13O9P	HXXFSFRBOHSIMQ-GASJEMHNSA-N	10.95	85.02793,145.04893,97.02841,109.02841
Glutamic acid	Amino acids and derivatives	148.0610	C5H9NO4	WHUUTDBJXRKMK-VKHMYEASNA-N	10.56	97.02841,130.04987,132.04171
Glutamine	Amino acids and derivatives	147.0770	C5H10N2O3	ZDXPYRJPNDTMRX-VKHMYEASNA-N	11.26	
Glutaryl-Carnitine	Acyl carnitines	277.1526	C12H22NO6	NXJAXUYOQLTISD-VIFPVBQESA-N	11.47	85.02841,87.04406,187.09649,218.13868,144.10191
Glutathione	Peptides	308.0916	C10H17N3O6S	RWSXRVCMGQZWBV-WDSKDSINSA-N	10.19	97.02841,232.08761
Glutathione oxidized	Peptides	613.1598	C20H32N6O12S2	YPZRWBKMTBYPTK-UHFFFAOYSA-N	11.07	152.05803,231.0434
Glycero-3-Phosphocholine	Glycerophosphocholines	258.1107	C8H20NO6P	SUHOQUVVVLNYQR-QMMMGPBSA-N	11.91	104.10699,124.99982,184.07332
Glycerol Tridecanoate	Triacylglycerols	555.4625	C33H62O6	LADGBHLMCUIINGV-UHFFFAOYSA-N	2.2	383.31559,155.14304
Glycerol Trihexanoate	Triacylglycerols	387.2747	C21H38O6	MAYGICSNZYXLHB-UHFFFAOYSA-N	2.14	

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Glycerol Tri-octanoate	Triacylglycerols	471.3686	C27H50O6	VLPFTAMPNXLGLX-UHFFFAOYSA-N	2.08	127.11174,327.25299
Glycerophosphoric acid/Glycerol 3-Phosphate	Glycerophosphates	173.0215	C3H9O6P	AWUCVROLDVIAJX-GSVOUGTGSA-N	10.53	98.98417
Glycocholic acid	Glycinated bile acids and derivatives	466.3169	C26H43NO6	RFDAIACWWDREDC-FRVQLJFSFA-N	4.98	414.30027,412.28462,430.29519,337.25259,448.30575
Glycoursodeoxycholic acid	Glycinated bile acids and derivatives	450.3220	C26H43NO5	GHCZAJUBVMUEKPP-XROMFQGDASA-N	4.47	424.30027,357.27881,339.26824,107.08553
Guanine	Purines and purine derivatives	152.0567	C5H5N5O	UYTPUPDQBNUYGX-UHFFFAOYSA-N	7.39	
Guanosine	Purine nucleosides	284.0995	C10H13N5O5	NYHBQMYGNKIUIF-UUOKFMHZSA-N	7.97	243.08497,152.05623,250.09347
Guanosine 5-monophosphate	Purine ribonucleoside monophosphates	364.0658	C10H14N5O8P	RQFCJASXJCIDSX-UUOKFMHZSA-N	11.06	97.02841,208.0829,177.03247,152.05669
Guanosine triphosphate	Purine ribonucleoside triphosphates	523.9985	C10H16N5O14P3	XKMLYUALXHKNFU-UUOKFMHZSA-N	11.06	85.02841,152.05669
Histamine	Primary amines	112.0869	C5H9N3	NTYJJOPFIAHURM-UHFFFAOYSA-N	22.00	95.05971,83.06037
Histidine	Amino acids and derivatives	156.0773	C6H9N3O2	HNDVDQJGIGZPNO-YFKPBYRVSA-N	11.85	83.05962
Histidinol	Aralkylamines	142.0981	C6H11N3O	ZQISRDCJNBUVMM-YFKPBYRVSA-N	17.85	82.05286,124.0862,107.02071
Homoarginine	Amino acids and derivatives	189.1352	C7H16N4O2	QUOGESRFPZDMMT-YFKPBYRVSA-N	18.44	84.08078
Homocitrulline	Amino acids and derivatives	190.1192	C7H15N3O3	XIGSAGMEBXLVJJ-YFKPBYRVSA-N	10.88	173.09207
Homocystine	Amino acids and derivatives	269.0630	C8H16N2O4S2	ZTVZLYBCZNMWCF-UHFFFAOYSA-N	12.03	88.02155
Homoserine	Amino acids and derivatives	120.0661	C4H9NO3	UKAUYYFTDYCKQA-VKHYHEASA-N	10.80	74.05959
Hydrocortisone	21-hydroxysteroids	363.2172	C21H30O5	JYGXADMDTFJGBT-VWUMJDOOSA-N	2.51	121.06479,135.08044,123.08044,107.04914
Hydroxylysine	Amino acids and derivatives	163.1083	C6H14N2O3	YSMODUONRAFBET-UHNVWZDZSA-N	18.88	89.05866,128.06977,145.09668
Hydroxyprogesterone	Glucocorticoids, progestogens and derivatives	331.2273	C21H30O3	DBPWSSGDRRHUNT-CEGNMAFCSA-N	2.33	163.11174,121.06479
Hydroxyproline/cis-4-Hydroxy-proline	Amino acids and derivatives	132.0661	C5H9NO3	PMMYEEVYMWASQNDMTCNVIQSA-N	11.00	
Hydroxytryptophan	Serotonins	221.0926	C11H12N2O3	LDCYZAJDBXYCGN-UHFFFAOYSA-N	9.44	204.0648,134.05925
Hypotaurine	Sulfinic acids	110.0276	C2H7NO2S	VVIUBCNACGLLV-UHFFFAOYSA-N	10.68	
Hypoxanthine	Purinones	137.0458	C5H4N4O	FDGQSTZJBJUBT-UHFFFAOYSA-N	5.72	119.03557
Imidazole acetic acid	Substituted imidazoles	127.0508	C5H6N2O2	PRJKNHOMHKJCEJ-UHFFFAOYSA-N	10.65	61.02841,81.04472
Indole	Indoles	118.0657	C8H7N	SIKJAJRHWYJAI-UHFFFAOYSA-N	2.42	
Indolelactic acid	Indolyl carboxylic acids and derivatives	206.0817	C11H11NO3	XGILAAAMKEQXLS-UHFFFAOYSA-N	4.69	130.06536
Inosine	Purine nucleosides	269.0886	C10H12N4O5	UGQMRVRMYASKQ-KQYNXXCUSA-N	6.85	137.04579
Inosine triphosphate	Purine ribonucleoside triphosphates	508.9876	C10H15N4O14P3	HAEJPOIATWHALX-KQYNXXCUSA-N	10.65	97.02848,137.04583

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Isoleucine/Leucine/allo-Isoleucine	Amino acids and derivatives	132.1025	C6H13NO2	AGPKZVBTJJNPAG-WHFBIAKZSA-N	8.40	89.05912,56.09643,86.09643,102.09134,81.03349
Isovaleryl-Carnitine	Acyl carnitines	247.1784	C12H24NO4	IGQBPDJNUXPENT-UHFFFAOYSA-N	8.21	85.02841,144.10191,85.05222,87.04406,84.08078
Isoxanthopterin	Pterins and derivatives	180.0516	C6H5N5O2	GLKCOBIIKZYKFN-UHFFFAOYSA-N	7.5	110.03489
Kynurenic acid	Quinoline carboxylic acids	190.0499	C10H7NO3	HCZHHEIFKROPDY-UHFFFAOYSA-N	4.72	144.04369
Kynurenine	Ketones	209.0926	C10H12N2O3	YGPSJZOEDVAXAB-QMMMGPBSA-N	7.22	192.06545,94.0651
Lactose/Maltose	Fatty acyl glycosides of mono- and disaccharides	343.1241	C12H22O11	DKXBNKWCZMJT-JVCRWLNRSA-N	10.80	163.0601,85.02841,145.04951
L-Dopa	Amino acids and derivatives	198.0766	C9H11NO4	WTDQRDBEARUVNC-LURJTMIESA-N	10.30	
Levulinic acid	Gamma-keto acids and derivatives	117.0552	C5H8O3	JOXCMJARBKPKM-UHFFFAOYSA-N	6.52	
Lithocholic acid	Hydroxy bile acids, alcohols and derivatives	377.3056	C24H40O3	SMEROWZSTRWXGI-WFVDQZAMSA-N	2.27	165.12739,109.10118,123.11683
Luteolin	Flavones	287.0556	C15H10O6	IQNAANSBPBGFQ-UHFFFAOYSA-N	2.51	161.02332,269.04445,179.03389,109.02841,125.02332
Lysine	Amino acids and derivatives	147.1134	C6H14N2O2	KDXKERNBIXSRK-YFKPBYRVSA-N	19.71	84.08078,130.08626
Maltotriose	Oligosaccharides	505.1769	C18H32O16	FYGDTMLNYKFZSV-DZOUCCMSA-N	5.62	253.09325,89.05971
Melibiose	Glycosyl compounds	343.1241	C12H22O11	DLRVLDZNNYCBX-ABXHMFFYSA-N	11.09	85.02841,99.04406
Menadione	Naphthoquinones	173.0603	C11H8O2	MJVAVZPDRWSRRC-UHFFFAOYSA-N	2.24	79.05423,77.03858,159.04406,99.04406,97.02841
Methionine	Amino acids and derivatives	150.0589	C5H11NO2S	FFEARJCKVFRZRR-BYPYZUCNSA-N	8.58	86.06001,104.05285
Methyladenine	6-aminopurines	150.0780	C6H7N5	FSASIHFSFGAIJM-UHFFFAOYSA-N	5.77	108.04305
Methyl-D-aspartic acid	Amino acids and derivatives	148.0610	C5H9NO4	HOKKHZGPKSLGJE-GSVOUGTGSAN	10.28	
Methyl-L-glutamic acid/Aminoadipic acid	Amino acids and derivatives	162.0766	C6H11NO4	OYIFNHCXNCRBQI-BYPYZUCNSA-N	10.30	97.02841,99.04406
Methyltyramine	Amphetamines and derivatives	152.1076	C9H13NO	GIKNHHRFLCDOEU-UHFFFAOYSA-N	7.44	121.06479,136.07569
m-Hydroxyhippuric Acid	Benzamides	196.0610	C9H9NO4	XDOFWFNMYJRHEW-UHFFFAOYSA-N	5.98	
NAD	(5'->5')-dinucleotides	664.1170	C21H27N7O14P2	BAWFJGJZGIEFAR-NNYOXOHSSA-O	10.42	
NADH	(5'->5')-dinucleotides	666.1326	C21H29N7O14P2	BAWFJGJZGIEFAR-NNYOXOHSSA-O	9.61	302.0421,514.051,400.01931,320.05298,649.1055
NADP	(5'->5')-dinucleotides	744.0833	C21H28N7O17P3	XJLXINKUBYWONI-NNYOXOHSSA-O	10.88	136.06141,428.03618,508.0031
Nicotine	Pyrolidinylopyridines	163.1235	C10H14N2	SNICXGAKADSCV-JTQLQIEISA-N	9.71	117.0573,106.06513
Nicotinic acid	Pyridinecarboxylic acids	124.0393	C6H5NO2	PVNIIMVLHYAWGP-UHFFFAOYSA-N	6.30	80.04975,108.04439,77.026
Nicotinoylglycine	Amino acids and derivatives	181.0608	C8H8N2O3	ZBSGKPYXQINNGF-UHFFFAOYSA-N	7.52	135.05443,79.04122

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Norleucine	Amino acids and derivatives	132.1025	C6H13NO2	LRQKBLKVPFOOQJ-YFKPBYRVSA-N	7.75	86.09732
Octanoyl-Carnitine	Acyl carnitines	289.2253	C15H30NO4	CXTATJFJDMJMIY-CYBMUJFWSA-N	6.00	85.02841,144.10191,246.16998,229.14344,85.05222
Oleic acid	Long-chain fatty acids	283.2637	C18H34O2	ZQPPMHVWEC SIRJ-KTKRTIGZSA-N	2.19	95.08553,109.10118,81.06988,123.11683,125.09608
Ornithine	Amino acids and derivatives	133.0977	C5H12N2O2	AHLPHDHMMVZTML-BYPYZUCNSA-N	19.30	70.06513,89.05912,116.07061
Palmitoyl-Carnitine	Acyl carnitines	401.3505	C23H46NO4	XOMRRQXKHYMOC-OAQYLSRUSA-N	4.00	85.02841,288.21693,144.10191,246.16998,229.14344
Pantothenic acid	Amino acids and derivatives	220.1185	C9H17NO5	GHOKWGTUJEAQD-UHFFFAOYSA-N	6.84	
Phenylacetylglucine	Amino acids and derivatives	194.0812	C10H11NO3	UTYVDVLMYQPLQB-UHFFFAOYSA-N	4.50	91.05355,84.04439
Phenylalanine	Amino acids and derivatives	166.0868	C9H11NO2	COLNVLHDHVKWLRT-QMMMGPBSA-N	7.27	120.08078,77.03858,93.06988,79.05423
Phosphocholine	Cholines	184.0728	C5H15NO4P	YHHSOZFOIEMCP-UHFFFAOYSA-O	15.32	
Phosphoenolpyruvic acid	Phosphate esters	168.9897	C3H5O6P	DTBNBXWJWCWCIK-UHFFFAOYSA-N	10.98	122.9843,150.97946
Phosphoethanolamine	Phosphoethanolamines	142.0264	C2H8NO4P	SUHOOTKUPISOBE-UHFFFAOYSA-N	12.96	98.98417,109.02768,106.98926
Pipecolic acid	Amino acids and derivatives	130.0868	C6H11NO2	HXEACLLIILLPRG-UHFFFAOYSA-N	9.73	84.08124
Progesterone	Glucocorticoids, progestins and derivatives	315.2324	C21H30O2	RJKFOVLPORLFTN-LEKSSAKUSA-N	2.28	109.06479,135.08044
Proline	Amino acids and derivatives	116.0712	C5H9NO2	ONIBWKKTOPOVIA-BYPYZUCNSA-N	10.11	70.065
Propionyl-Carnitine	Acyl carnitines	219.1471	C10H20NO4	UFAHZIUFPNHSL-UHFFFAOYSA-N	9.90	85.02841,85.05222,87.04406,84.08078,129.07843
Pyridoxal	Pyridoxals and derivatives	168.0661	C8H9NO3	RADKZDMFGJYCB-UHFFFAOYSA-N	4.89	150.05434,122.06004
Pyridoxal-5-phosphate	Pyridoxals and derivatives	248.0319	C8H10NO6P	NGVDGCNFWLIFO-UHFFFAOYSA-N	9.34	120.04393,137.04713,153.04204
Pyridoxamine	Pyridoxamine 5'-phosphates	169.0972	C8H12N2O2	NHZMQXZHNVTQA-UHFFFAOYSA-N	17.25	134.06056,152.07061,124.07569
Pyridoxine	Pyridoxines	170.0812	C8H11NO3	LXNHXLLTXMVWPM-UHFFFAOYSA-N	4.8	134.06094
Quinaldic acid	Quinoline carboxylic acids	174.0555	C10H7NO2	LOAUVZALPPNFOQ-UHFFFAOYSA-N	4.16	156.04439,128.04948
Reserpine	Yohimbine alkaloids	609.2812	C33H40N2O9	QEVHRUUCFGRFIF-MDEJGZGSSA-N	2.59	195.06519,448.19659,174.09134,397.21218,365.18597
Retinoic Acid	Retinoids	301.2168	C20H28O2	SHGAZHPJJPHSC-YCNIQYBTA-N	2.23	123.11683,135.08044,175.14813,135.11683,161.09609
Riboflavin	Flavins	377.1461	C17H20N4O6	AUNGANRZJHBGPY-SCRDCRAPSAN	4.50	198.0658,243.08715
Riboflavin monophosphate	Flavin nucleotides	457.1125	C17H21N4O9P	FVTCRASFAADXXNN-SCRDCRAPSAN	9.00	359.13432,439.10089,243.08722
Ribose 5-phosphate	Monosaccharides	231.0270	C5H11O8P	KTVPXOYAKDPRHY-SOOFDHNKSA-N	10.75	

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Sarcosine	Amino acids and derivatives	90.0550	C3H7NO2	FSYKKLYZXJSNPZ-UHFFFAOYSA-N	11.00	57.03349
Serotonin	Serotonins	177.1022	C10H12N2O	QZAYGJVTTNCVMB-UHFFFAOYSA-N	10.28	160.07508,146.06004,143.07295,133.05222
Sorbitol/Galactitol	Sugar alcohols	183.0869	C6H14O6	FBPFZTCFMRRESA-JGWLITMVSA-N	9.40	110.03623,85.02841
Tamoxifen	Stilbenes	372.2328	C26H29NO	NKANXQFJJCIGDU-QPLCGJKRSA-N	2.71	72.08078,129.06988,178.0777,128.06205,107.04914
Taurine	Organosulfonic acids	126.0219	C2H7NO3S	XOAAWQZATWQOTB-UHFFFAOYSA-N	10.60	111.01104,96.99539
Taurochenodeoxycholic acid	Taurinated bile acids and derivatives	500.3046	C26H45NO6S	BHTRKEVKTKCXOH-BJLOMENOSA-N	2.42	126.02113,464.28287
Taurodeoxycholic acid	Taurinated bile acids and derivatives	500.3046	C26H45NO6S	AWDRATDZQPNJFN-VAYUFCLWSA-N	2.4	464.28256,126.02143,339.26724
Tauroursodeoxycholic acid	Taurinated bile acids and derivatives	500.3046	C26H45NO6S	BHTRKEVKTKCXOH-VSHSPWMTSA-N	2.4	126.02199,464.28244
Theophylline	Xanthines	181.0720	C7H8N4O2	ZFXFYBGIUFBOJW-UHFFFAOYSA-N	2.8	
Thiamine	Thiamines	265.1123	C12H16N4OS	JZRWCZRTZMZEH-UHFFFAOYSA-N	16.24	122.07068,144.04711
Thiamine monophosphate	Thiamines	345.0787	C12H17N4O4PS	HZSAJDVWZRBGIF-UHFFFAOYSA-N	14.44	
Thiamine pyrophosphate	Thiamines	425.0450	C12H18N4O7P2S	AYEKOFBPNLCAJY-UHFFFAOYSA-O	13.50	
Thioctamide	Lipoamides	206.0673	C8H15NOS2	FCCDDURTIUXBY-UHFFFAOYSA-N	2.49	119.0889,91.0576,105.07325,95.08553,81.06988
Threonine	Amino acids and derivatives	120.0661	C4H9NO3	AYFVYJQAPQTCGC-GBXIJSLDSA-N	10.60	102.05495
Thymidine	Pyrimidine 2'-deoxyribonucleosides	243.0981	C10H14N2O5	IQFYKVMVGFJEH-XLPZGREQSA-N	4.16	89.05917,127.0502
Thymine	Hydroxypyrimidines	127.0508	C5H6N2O2	RWQNBDRDOKXIBV-UHFFFAOYSA-N	4.12	69.03253,113.043,71.04914
Tiglylglycine	Amino acids and derivatives	158.0817	C7H11NO3	WRUSVQOKJIDBLP-HWKANZROSA-N	4.76	
Tocopherol	Vitamin E compounds	431.3889	C29H50O2	QUEDXNHFTDJVIY-UHFFFAOYSA-N	2.15	165.09101,81.06988,95.08553
trans-2-Methyl-2-pentenoic acid	Branched fatty acids	115.0759	C6H10O2	JJYWRQLLQAKNAD-SNAWJCMRSA-N	9.00	
Trihydroxyisoflavone	Isoflavones	271.0601	C15H10O5	TZBJGXHYKVUXJN-UHFFFAOYSA-N	2.39	253.04954
Triiodo-L-thyronine	Amino acids and derivatives	651.7979	C15H12I3NO4	AUYJCJSJGJYCDS-LBPRGKRZSA-N	4.10	605.79232,507.86625
Tryptophan	Indolyl carboxylic acids and derivatives	205.0977	C11H12N2O2	QIVBCDIJAJPQS-VIFPVBQESA-N	7.55	188.07061,118.06513,159.09167
Tyramine	Phenethylamines	138.0913	C8H11NO	DZGWFCGJZKJUFU-UHFFFAOYSA-N	9.05	121.06534,95.04962
Tyrosine	Amino acids and derivatives	182.0817	C9H11NO3	OUYCCASQSFEME-QMMMGPBSA-N	8.91	107.04914
UDP Glucose	Pyrimidine nucleotide sugars	567.0629	C15H24N2O17P2	HSCJRCZDFQWRP-LPTOLDLSA-N	10.07	97.02787,177.2947
UPD-alpha-D-Galactose	Pyrimidine nucleotide sugars	567.0629	C15H24N2O17P2	HSCJRCZDFQWRP-ABVWGUQPSA-N	10.30	97.02274
Uracil	Pyrimidones	113.0346	C4H4N2O2	ISAKRJGGNUQOIC-UHFFFAOYSA-N	4.33	57.03368,72.04439
Ureidosuccinic acid	Amino acids and derivatives	177.0512	C5H8N2O5	HLKXYZVTANABHZ-REOHLBBSA-N	10.61	88.0395

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Uric acid	Xanthines	169.0362	C5H4N4O3	LEHOTFFKMJEONL-UHFFFAOYSA-N	8.09	127.04904,156.04037
Uridine	Pyrimidine nucleosides	245.0774	C9H12N2O6	DRTQHJVPVMBUCF-XVFCMESISA-N	5.7	
Uridine diphosphate	Pyrimidine ribonucleoside diphosphates	405.0100	C9H14N2O12P2	XCCTYIAWTASOJW-XVFCMESISA-N	10.49	97.02841
Uridine monophosphate	Pyrimidine ribonucleoside monophosphates	325.0437	C9H13N2O9P	DJJCXFVJDGTHFX-XVFCMESISA-N	10.52	97.02841
Urocanic acid	Substituted imidazoles	139.0508	C6H6N2O2	LOIYMARKYCTBW-OWOJBTEDSA-N	8.13	121.03935,133.04954,124.99982,259.04783
Vitamin D3	Vitamin D and derivatives	385.3471	C27H44O	QYSXJUFSXHHAJI-QWSSABAFA-SA-N	2.24	91.05423,195.40118,107.08553,105.06988,95.08553
Vitamin K1	Vitamin K compounds	451.3576	C31H46O2	MBWXNTAXLNYFJB-LKUDQCMESA-N	2.56	109.10118,187.07536,123.11683,225.09101,95.08553
Xanthine	Xanthines	153.0413	C5H4N4O2	LRFTYWOQMYALW-UHFFFAOYSA-N	5.9	110.03489
Xanthosine	Purine nucleosides	285.0835	C10H12N4O6	UBORTCNDUKBEOP-UUOKFMHZA-N	7.95	89.0598,153.0407
Xanthurenic acid	Quinoline carboxylic acids	206.0453	C10H7NO4	FBZONXHGPPHHIY-UHFFFAOYSA-N	6.76	160.0393,188.03422

Table 20 Human metabolite compound library in ZIC-HILIC negative ionization mode. Human metabolite reference standards from MetaSci in a concentration of 100 µM were acquired using the ZIC-HILIC-MS method in negative ionization mode. Only [M-H]⁻ molecular ions were used for identification. The table shows an excerpt from the UNIFI-Database and only metabolites detected in reference standards are shown. Other dataformats (such as MSP) are also available.

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
10-Hydroxydecanoic acid	Medium-chain hydroxy acids and derivatives	187.1334	C10H20O3	FYSSBMZUBSBFJL-SECINFHSA-N	2.75	
1-Methyl Adenosine	Purine nucleosides	280.1046	C11H15N5O4	GFYLSDSUCHVORB-JOLDIKRXSA-N	11.22	148.06287
2-Amino-4-methylsulfonimidoyl-butanoic acid	Amino acids and derivatives	179.0490	C5H12N2O3S	SXTAYKAGBXMxCB-UHFFFAOYSA-N	11.26	78.00191
2-Aminoisobutyric acid	Amino acids and derivatives	102.0555	C4H9NO2	FUOOLUPWFVMBKG-UHFFFAOYSA-N	10.03	85.02863,67.01894
2-Deoxyadenosine	Purine 2'-deoxyribonucleosides	250.0940	C10H13N5O3	OLXZPDWKRNYJJZ-RRKCRQDMSA-N	4.36	160.06287
2-Deoxycytidine	Pyrimidine 2'-deoxyribonucleosides	226.0828	C9H13N3O4	CKTSBUTUHBMZGZ-SHYZEUOFA-SA-N	6.70	134.04948
2-Deoxyuridine	Pyrimidine 2'-deoxyribonucleosides	227.0668	C9H12N2O5	MXHRCPNRJAMMIM-SHYZEUOFA-SA-N	4.42	211.04862
2-Hydroxy-3-methylbutyric acid	Hydroxy fatty acids	117.0552	C5H10O3	NGEWQZIDQIYUNV-UHFFFAOYSA-N	4.72	99.04515,83.05024,71.05024,85.0295,101.02442
2-Hydroxybutyric acid	Alpha hydroxy acids and derivatives	103.0395	C4H8O3	AFENDNXGAFYKQO-UHFFFAOYSA-N	6.17	85.0295,71.01385

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
2-Hydroxycaproic acid/2-Ethyl-2-Hydroxybutyric acid/2-Hydroxyisocaproic acid	Medium-chain fatty acids	131.0708	C6H12O3	NYHNVHGFPAZGA-UHFFFAOYSA-N	4.35	85.06589,113.0608,59.01385
2-Hydroxyoctanoic acid	Medium-chain fatty acids	159.1021	C8H16O3	JKRDADVRIYVCCY-UHFFFAOYSA-N	2.58	
2-Isopropylmalic acid	Hydroxy fatty acids	175.0606	C7H12O5	BITYXLXUCSKTJS-ZETCQYMHSA-N	7.05	115.04007,85.06589
2-Methylbutyrylglycine	Amino acids and derivatives	158.0817	C7H13NO3	HOACIBQKYRHBOW-UHFFFAOYSA-N	4.25	74.02475,128.03532
2-Methylglutaric acid	Branched fatty acids	145.0501	C6H10O4	AQYCMVICBNBXNA-UHFFFAOYSA-N	10.11	101.0608,83.05024
2-Methylhippuric acid	Benzamides	192.0661	C10H11NO3	YOEBAVRJHRCKRE-UHFFFAOYSA-N	3.66	148.07679,174.05605
2-Methylsuccinic acid	Branched fatty acids	131.0344	C5H8O4	WXUAQHNMJWJLTG-UHFFFAOYSA-N	10.50	87.04515
2-Octenoic acid	Medium-chain fatty acids	141.0915	C8H14O2	CWMPVPFLSZGCV-SREVYHEPSA-N	3.14	
2-Oxoadipic acid	Medium-chain keto acids and derivatives	159.0293	C6H8O5	FGSBNBBHOZHUBO-UHFFFAOYSA-N	9.92	115.04007
2-Picolinic acid	Pyridinecarboxylic acids	122.0242	C6H5NO2	SIOXPEMLGUPBBT-UHFFFAOYSA-N	4.80	107.03766
2-Propanamidoacetic acid	Amino acids and derivatives	130.0510	C5H9NO3	WOMAZEJKVZLFE-UHFFFAOYSA-N	7.20	74.02475,86.06114
3-Chloro-Tyrosine	Amino acids and derivatives	214.0271	C9H10ClNO3	ACWBBAGYTKWBCD-UHFFFAOYSA-N	7.94	72.0091
3-Hydroxy-3-methylglutaric acid	Hydroxy fatty acids	161.0455	C6H10O5	NPOAOTPXWNWTS-UHFFFAOYSA-N	2.39	103.03967,146.05846,116.04789,118.04789
3-Hydroxybutanoic acid/3-hydroxybutyric acid	Beta hydroxy acids and derivatives	103.0395	C4H8O3	WHBMMWSBFZVSSR-GSVOUGTGSA-N	7.60	
3-Hydroxymandelic acid	1-hydroxy-4-unsubstituted benzenoids	167.0344	C8H8O4	OLSDAJRAVOVKLG-UHFFFAOYSA-N	6.84	121.0295,123.04515,93.03459,92.02676,149.02442
3-Hydroxyphenylacetic acid	1-hydroxy-4-unsubstituted benzenoids	151.0401	C8H8O3	FVMDYYGIDFPZAX-UHFFFAOYSA-N	5.97	119.04989,59.01472,107.05024,92.02676,106.04241
3-Indolebutyric acid	3-alkylindoles	202.0868	C12H13NO2	JTEDVYBZBROSJT-UHFFFAOYSA-N	2.79	116.05057,158.09752,156.08187,184.07679
3-Indolepropionic acid	Indolyl carboxylic acids and derivatives	188.0717	C11H11NO2	GOLXRNDWAUTYKT-UHFFFAOYSA-N	4.13	59.01429,116.05043,129.0571,142.06622,144.08187
3-Methyl-L-histidine	Amino acids and derivatives	168.0773	C7H11N3O2	JDHILDINMRGULE-LURJTMIESA-N	13.21	101.06115,72.0091,151.0513
3-Methylthio propionic acid	Straight chain fatty acids	119.0167	C4H8O2S	CAOMCZAIALVUPA-UHFFFAOYSA-N	5.03	
3-Methylxanthine	Xanthines	165.0412	C6H6N4O2	GMSNIKWWOQHZGF-UHFFFAOYSA-N	4.18	122.0346
3-Nitro-tyrosine	Amino acids and derivatives	225.0511	C9H10N2O5	FBTSQILOGYXGMD-LURJTMIESA-N	8.03	163.05004,136.03923
3-Phenyllactic acid	Phenylpropanoic acids	165.0552	C9H10O3	VOXXWSYKYCBWHO-UHFFFAOYSA-N	3.54	147.04515,119.05024

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
4-Acetamidobutyric acid	Amino acids and derivatives	144.0666	C6H11NO3	UZTFMUBKZQVKLK-UHFFFAOYSA-N	8.42	102.05656,126.05673
4-Guanidinobutyric acid	Amino acids and derivatives	144.0779	C5H11N3O2	TUHVEAJXIMEOSA-UHFFFAOYSA-N	11.40	102.05573,127.05089
4-Hydroxyphenylpyruvic acid	Phenylpyruvic acid derivatives	179.0350	C9H8O4	KKADPXVIOXHVKN-UHFFFAOYSA-N	5.52	107.04956,106.04241
4-Methyl-2-oxovaleric Acid	Short-chain keto acids and derivatives	129.0557	C6H10O3	BKAJNAXTPSGJCU-UHFFFAOYSA-N	3.50	72.99312,85.06589
4-Methylvaleric acid	Branched fatty acids	115.0759	C6H12O2	FGKJLKRYENPLQH-UHFFFAOYSA-N	3.21	97.06589,71.01385
4-Pyridoxic acid	Pyridinecarboxylic acids	182.0459	C8H9NO4	HXACOUQIXZGNBF-UHFFFAOYSA-N	2.68	108.04785,138.05863
5-Aminolevulinic acid	Amino acids and derivatives	130.0504	C5H9NO3	ZGXJTSGNIOSYLO-UHFFFAOYSA-N	12.51	112.0404,99.04515
5-Hydroxy-3-indoleacetic acid	Indole-3-acetic acid derivatives	190.0510	C10H9NO3	DUUGKQCEGLZNO-UHFFFAOYSA-N	8.47	144.04789,131.03974
5-Hydroxylysine	Amino acids and derivatives	161.0926	C6H14N2O3	YSMODUONRAFBE-UHNNVZDZSA-N	18.88	115.08769,125.07204
5-Hydroxymethyl uracil	Pyrimidones	141.0300	C5H6N2O3	JDBGXEHEIRGOBU-UHFFFAOYSA-N	5.80	123.02
5-Hydroxymethyl-2-furancarboxylic acid	Furoic acids	141.0188	C6H6O4	PCSKKIURRTAEM-UHFFFAOYSA-N	7.09	97.0295
5-Hydroxytryptophan	Serotonins	219.0770	C11H12N2O3	LDCYZAJDBXYCGN-UHFFFAOYSA-N	9.44	132.04549,72.0091,175.08769
5-Methoxy-tryptophan	Serotonins	233.0926	C12H14N2O3	KVNPSKDDJARYKK-JTQLQIEISA-N	7.16	158.06114,174.07944
5-Methylcytosine	Hydroxypyrimidines	124.0511	C5H7N3O	LRASMSXMSNRBT-UHFFFAOYSA-N	6.19	
5-Methyluridine	Pyrimidine nucleosides	257.0773	C10H14N2O6	DWRXFEITVBNRMK-JXOAFINSAN-N	4.70	190.05693,142.05287
Aconitic acid	Tricarboxylic acids and derivatives	173.0086	C6H6O6	GTZCVFVGUGFEME-IWQZZHSRAN-N	11.13	129.01933,141.0158,154.9986
Adenosine 5-diphosphate	Purine ribonucleoside diphosphates	426.0216	C10H15N5O10P2	XTWYTFMLZFPYCI-KQYNXXCUSA-N	10.23	158.92771,96.97414
Adenosine 5-triphosphate	Purine ribonucleoside triphosphates	505.9879	C10H16N5O13P3	ZKHQWZAMYRWXGA-KQYNXXCUSA-N	10.42	158.9277
Adenosine-5-monophosphate	Purine ribonucleoside monophosphates	346.0552	C10H14N5O7P	UDMBCSSLTHHNC-DKQYNXXCUSA-N	10.07	96.97152,78.96091
Alanine/Sarcosine	Amino acids and derivatives	88.0398	C3H7NO2	QNAAYBMKLOCPYJ-REOHCLBHSAN-N	10.90	57.03533,58.02984
Allantoic acid	Amino acids and derivatives	175.0473	C4H8N4O4	NUCLJNSWZCHRKL-UHFFFAOYSA-N	10.18	89.03565,132.04146
Allantoin	Imidazoles	157.0362	C4H6N4O3	POJWUDADGALRAB-UHFFFAOYSA-N	8.49	97.00435
alpha-Hydroxyhippuric acid	Benzamides	194.0453	C9H9NO4	XDOFWFNMYJRHEW-UHFFFAOYSA-N	4.81	72.99312,120.04549
alpha-Lipoic acid	Lipoic acids and derivatives	205.0357	C8H14O2S2	AGBQKNBQESQJND-SSDOTTSWSAN-N	3.83	

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Aminoadipic acid	Amino acids and derivatives	160.0610	C6H11NO4	OYIFNHCXNCRBQI-UHFFFAOYSA-N	10.65	67.01932,85.02931,85.0295,71.01385,142.05097
Androsterone	Androgens and derivatives	289.2167	C19H30O2	QGXBDMJGAMFCBF-HLUDHZFRSA-N	2.53	
Anhydro-d-Glucitol	Monosaccharides	163.0606	C6H12O5	MPCAJMNYNOGXPB-SLPGGIOYSA-N	7.79	87.00899,71.05024,99.04515,133.05063,134.05907
Arabinose	Monosaccharides	149.0450	C5H10O5	SRBFZHDQGSBBOR-HWQSCIPKSA-N	16.82	121.05063
Arabitol	Amino acids and derivatives	151.0606	C5H12O5	CKLJMWTZIZZHCS-REOHLBHSAN	9.01	71.01425
Arachidic acid	Long-chain fatty acids	311.2956	C20H40O2	VKOBVWXKNCXXDE-UHFFFAOYSA-N	2.18	269.2486,255.23295
Arachidonic acid	Long-chain fatty acids	303.2324	C20H32O2	YZXBAPSDXZZRGB-DOFZRALJSA-N	2.18	
Arginine	Amino acids and derivatives	173.1038	C6H14N4O2	ODKSFYDXXFIFQN-BYPYZUCNSA-N	19.22	131.08455
Ascorbic acid	Furanones	175.0242	C6H8O6	TYQCGQRIZGCHNB-MVHIGOERSA-N	9.48	87.00763,115.00229,76.96005
Asparagine	Amino acids and derivatives	131.0457	C4H8N2O3	DCXYFEDJOCNFAF-REOHLBHSAN	11.22	72.0091
Aspartic acid	Amino acids and derivatives	132.0297	C4H7NO4	CKLJMWTZIZZHCS-REOHLBHSAN	10.60	72.0091,86.02475
Azelaic acid	Medium-chain fatty acids	187.0970	C9H16O4	BDJRBEYXGGNYIS-UHFFFAOYSA-N	9.58	125.09719,141.0921,97.06589,123.08154
beta-Alanine/Sarcosine	Amino acids and derivatives	88.0398	C3H7NO2	UCMIRNVEIXFBKS-UHFFFAOYSA-N	11.10	57.0353,71.0385,59.01477
beta-Glycerophosphoric acid/n-Glycerol 3-Phosphate	Glycerophosphates	171.0058	C3H9O6P	DHCLVCXQIBBOPH-UHFFFAOYSA-N	10.53	78.95847,152.99436,96.97039
beta-hydroxyisobutyrate	Beta hydroxy acids and derivatives	103.0395	C4H8O3	DBXBTMSZEOQQDU-VKHMYHEASA-N	8.33	73.0295
beta-Hydroxyisovaleric acid	Hydroxy fatty acids	117.0557	C5H10O3	AXFYFNCPONWUHW-UHFFFAOYSA-N	5.55	59.01496
Bilirubin	Bilirubins	583.2556	C33H36N4O6	BPYKTIZUTYGOLE-IFADSCNNSA-N	3.95	285.1247,391.2263,253.13464,213.10334
Biotin	Biotin and derivatives	243.0803	C10H16N2O3S	YBJHBAHKTGYVGT-ZKWXMUHSA-N	6.44	199.09389
Capryloyl glycine	Amino acids and derivatives	200.1287	C10H19NO3	SAVLIIGUQOSOEP-UHFFFAOYSA-N	4.17	74.02475
Carnosine	Hybrid peptides	225.0988	C9H14N4O3	CQOVNPNJLQNMDC-ZETCQYMHSA-N	14.42	110.07363,154.06413
Chenodeoxycholic acid	Hydroxy bile acids, alcohols and derivatives	391.2848	C24H40O4	RUDATBOHQWQJDD-BSWAIDMHSA-N	2.58	373.27482
Cholesterol 3-Sulfate	Cholesterols and derivatives	465.3038	C27H46O4S	BHYOQNUELFTYRT-DPAQBDFSA-N	2.23	96.9601
Cholic acid	Hydroxy bile acids, alcohols and derivatives	407.2797	C24H40O5	BHQCQFFYZLCCQ-OELDTZBJSAN	2.66	343.26425,289.2173

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
cis-4-Hydroxy-D-proline/L-Hydroxyproline	Amino acids and derivatives	130.0504	C5H9NO3	PMMYEEVYMWASQN-QWVZWVQMSA-N	11.00	112.0404,85.0295,71.01385
Citicoline	Pyrimidine ribonucleoside diphosphates	487.0995	C14H26N4O11P2	RZPDXZPRHQCG-OJAKKHQRSA-N	11.35	78.96081
Citrulline	Amino acids and derivatives	174.0879	C6H13N3O3	RHGKLRLOHDJJDR-BYPYZUCNSA-N	11.22	131.0826
Corticosterone	21-hydroxysteroids	345.2066	C21H30O4	OMFXVFTZEKFJBZ-HJTSIMOOSA-N	2.49	329.17589,301.18092,137.0608,311.16527
Cortisone	21-hydroxysteroids	359.1858	C21H28O5	MFYSYFVPBJMHGN-ZPOLXVRWSA-N	2.51	329.17583,301.18092,137.0608,311.16527
Creatine	Amino acids and derivatives	130.0616	C4H9N3O2	CVSVTCORWBXHQV-UHFFFAOYSA-N	10.89	88.0404
Cystathionine	Amino acids and derivatives	221.0596	C7H14N2O4S	ILRYLPWNYFXEMH-WHFBIKZSA-N	12.83	134.02812,120.01247
Cysteic acid	Amino acids and derivatives	167.9967	C3H7NO5S	XVOYSCVBGLVSOL-UHFFFAOYSA-N	10.10	80.96519,94.98084,79.95736,71.01385
Cysteine	Amino acids and derivatives	120.0119	C3H7NO2S	XUJNEKJLAYXESH-REOHCLBNSA-N	10.50	71.01385,86.02475
Cystine	Amino acids and derivatives	239.0160	C6H12N2O4S2	LEVWYRKDKASIDU-IMJSDKUSA-N	13.45	222.01383,204.00327,132.94628
Cytidine	Pyrimidine nucleosides	242.0777	C9H13N3O5	UHDGCVIWMRVCDJ-XVFCMESISA-N	7.74	
Cytidine 5-monophosphate	Pyrimidine ribonucleoside monophosphates	322.0440	C9H14N3O8P	IERHLVCPSMICTF-XVFCMESISA-N	11.04	78.9594,96.96969
Cytidine 5-triphosphate	Monosaccharides	481.9767	C9H16N3O14P3	PCDQPRRSZKQHHS-XVFCMESISA-N	11.21	78.9594,96.96969,158.9248
Deoxyguanosine-5-monophosphate	Purine deoxyribonucleoside monophosphates	346.0552	C10H14N5O7P	LTFMZDNNPPEQNG-KVQBGUIXSA-N	10.76	78.96001
Deoxyinosine	Purine 2'-deoxyribonucleosides	251.0780	C10H12N4O4	VGONTNSXDCQUGY-RRKCRQMSA-N	5.61	108.02217,135.03333
Dihydroxyacetone phosphate	Monosaccharides	168.9907	C3H7O6P	GNGACRATGGDKBX-UHFFFAOYSA-N	11.06	78.95905,96.96962,150.98018,138.98018,73.0295
Dihydroxybenzoic acid	Hydroxybenzoic acid derivatives	153.0188	C7H6O4	UYEMGAFJOZZIFP-UHFFFAOYSA-N	8.40	109.0295,121.0295
Dimethylglutaric acid	Branched fatty acids	159.0657	C7H12O4	DUHQIGLHYXLKAE-UHFFFAOYSA-N	3.57	115.07645
Diosmin	Flavonoid O-glycosides	607.1663	C28H32O15	GZSOSUNBTXMUFQ-YFAPSIMESA-N	4.11	
Dioxoheptanoic acid	Medium-chain keto acids and derivatives	157.0501	C7H10O4	WYEPBHZLDUPOD-UHFFFAOYSA-N	4.47	116.04949,142.06493
Dodecanedioic acid	Medium-chain fatty acids	229.1440	C12H22O4	TVIDDXQYHWJXFK-UHFFFAOYSA-N	3.98	143.10775,123.08154,129.0921,181.15979,109.06589
Dopa	Amino acids and derivatives	196.0610	C9H11NO4	WTRDRQBPEARUVNC-LURJTMIESA-N	10.3	72.0091,122.03733,109.0295

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Eicosatrienoic acid	Long-chain fatty acids	305.2480	C20H34O2	HOBAELRKJCKHQD-QNEBEIHSSA-N	2.18	135.08154,123.08154,109.06589,95.05024
Estradiol	Estrogens and derivatives	271.1698	C18H24O2	VOXZDWNPVJITMN-ZBRFXRBCSA-N	2.35	95.05024,109.06589,107.05024,113.09719,121.06589
Estriol	Estrogens and derivatives	287.1647	C18H24O3	PROQIPRRNZUXQM-ZXXIGWHRSA-N	2.47	
Estrone	Estrogens and derivatives	269.1541	C18H22O2	DNXHEGUUPJUMQT-CBZIJGRNSA-N	2.40	125.09719,163.11284,97.06589,123.08154,109.06589
Ethosuximide	Pyrrolidine-2-ones	140.0711	C7H11NO2	HAPOVYFOVWVWLR-UHFFFAOYSA-N	2.52	
Flavin Adenine Dinucleotide	Flavin nucleotides	784.1493	C27H33N9O15P2	VWWQXMAJ TJZDQX-UYBVJOGSSA-N	8.40	439.10089,366.139,346.05581,437.08677
Flavone	Flavones	221.0602	C15H10O2	VHBFQKKBGNRLFZ-UHFFFAOYSA-N	4.30	
Folic acid	Amino acids and derivatives	440.1318	C19H19N7O6	OVBPIULPVIDEAO-LBPRGKRZSA-N	10.34	175.04958
Folinic acid	Tetrahydroptericoic acids and derivatives	472.1581	C20H23N7O7	VVIAGPKUTFNRDU-UHFFFAOYSA-N	9.99	308.1072,128.0351,272.1153
Formononetine	4'-O-methylated isoflavonoids	267.0663	C16H12O4	HKQYGTCTOTHOMP-UHFFFAOYSA-N	2.31	251.03498,91.01894,107.05024,159.04515
Fructose/Mannose	Glycosyl compounds	179.0555	C6H12O6	RFSUNEUAIZKAJO-ARQDHWQXSA-N	9.22	59.01385,71.01385,99.04515
Fructose-biphosphate	Monosaccharides	338.9882	C6H14O12P2	WSMBXSQDFPTODV-JGWLITMUSA-N	11.50	72.95905,96.96962,211.00131,241.01146,150.97976
Fucose	Monosaccharides	163.0606	C6H12O5	SHZGCJCMOBCMKK-DHVFOXMCSA-N	3.24	118.06354
Fumaric acid	Dicarboxylic acids and derivatives	115.0031	C4H4O4	VZCYOQTPPOCHFL-OWOJBTEDSA-N	10.72	71.01385
Galactonic acid	Medium-chain hydroxy acids and derivatives	195.0505	C6H12O7	RGHNJXZEOKUKBD-MGCNEYSASA-N	10.04	75.0089
Galactose	Monosaccharides	179.0555	C6H12O6	WQZGKKKJJFFOK-PHYPRBDBSA-N	6.04	
Glucosamine 6-phosphate	Monosaccharides	258.0379	C6H14NO8P	XHMJOUIAFHJHBW-UKFBFLRUSA-N	13.05	78.96089,96.96962,240.02749
Glucose 1-phosphate/Galactose-1-phosphate/D-Fructose-6-phosphate/D-Glucose 6-phosphate	Monosaccharides	259.0219	C6H13O9P	HXXFSFRBOHSIMQ-GASJEMHNSA-N	11.00	78.95809,96.9693,183.01041,241.0148
Glucuro-3-6-lactone	Isosorbides	175.0242	C6H8O6	OGLCQHRZUSEXNB-WHDMSYDLSA-N	6.44	113.02442,85.0295
Glucuronic acid	Sugar acids and derivatives	193.0348	C6H10O7	AEMOLEFTQBMLNQ-AQKNRBDQSA-N	10.34	132.04281
Glutamic acid	Amino acids and derivatives	146.0453	C5H9NO4	WHUUTDBJXJRKMK-VKHYHEASA-N	10.56	
Glutamine	Amino acids and derivatives	145.0613	C5H10N2O3	ZDXPYRJPNDTMRX-VKHYHEASA-N	11.26	111.00877,67.01968,85.03.97.02976

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Glyceric acid	Sugar acids and derivatives	105.0188	C3H6O4	RBNPOMFGQQGHHO-UWTATZPHSA-N	8.49	75.00877,72.99312
Glycerol Tributryrate	Triacylglycerols	301.1651	C15H26O6	UYXTWWCETRIEDR-UHFFFAOYSA-N	2.52	
Glycerol Tridecanoate	Triacylglycerols	553.4468	C33H62O6	LADGBHLMCUINGV-UHFFFAOYSA-N	2.25	243.16018,211.13397
Glycocholic acid	Glycinated bile acids and derivatives	464.3012	C26H43NO6	RFDAIACWWDREDC-MZMBZMQMSA-N	4.96	74.02475,402.30137,400.28572,382.27515,446.2912
Glycodeoxycholic acid/Glycochenodeoxycholic acid	Glycinated bile acids and derivatives	448.3063	C26H43NO5	WVULKSPCQVQLCU-ZMBDPXIHSA-N	2.74	125.09719,74.02475,329.23603,402.30137,137.09691
Glycoursodeoxycholic acid	Glycinated bile acids and derivatives	448.3063	C26H43NO5	GHCZAJUBVMUEKKP-XROMFQGDSA-N	4.36	74.02475,386.30645
Guanidinoacetic acid	Amino acids and derivatives	116.0466	C3H7N3O2	BPMFZUMJYQTVII-UHFFFAOYSA-N	11.06	74.02475
Guanine	Purines and purine derivatives	150.0421	C5H5N5O	UYTPUPDQBNUYGX-UHFFFAOYSA-N	7.39	133.01558
Guanosine	Purine nucleosides	282.0838	C10H13N5O5	NYHBQMYGNKIUIF-UUOKFMHZA-N	7.97	241.07258,150.04213,133.01523
Guanosine 5-monophosphate	Purine ribonucleoside monophosphates	362.0502	C10H14N5O8P	RQFCJASXJCIDSX-UUOKFMHZA-N	11.06	96.97141,78.96081,306.07347,179.05745
Guanosine 5-triphosphat	Purine ribonucleoside triphosphates	521.9828	C10H16N5O14P3	XKMLYUALXHKNFU-UUOKFMHZA-N	11.06	158.92771,272.96138,96.9714
Heneicosanoic acid	Long-chain fatty acids	325.3112	C21H42O2	CKDDRHZIAZRDBW-UHFFFAOYSA-N	2.18	253.21678
Heptadecanoic acid	Long-chain fatty acids	269.2480	C17H34O2	KEMQGGTRYUADPNZ-UHFFFAOYSA-N	2.23	255.23295,85.0295,151.11284
Heptanoic acid	Medium-chain fatty acids	129.0915	C7H14O2	MNWFJYAOYHMED-UHFFFAOYSA-N	3.05	
Hippuric acid	Benzamides	178.0504	C9H9NO3	QIAFMBKCNZACKA-UHFFFAOYSA-N	4.18	134.06114
Histidine	Amino acids and derivatives	154.0616	C6H9N3O2	HNDVDQJGZPNO-YFKPBYRVSA-N	11.85	136.05164
Histidinol	Aralkylamines	140.0824	C6H11N3O	ZQISRDCJNBVMM-YFKPBYRVSA-N	17.85	
Homoarginine	Amino acids and derivatives	187.1195	C7H16N4O2	QUOGESRFPZDMMT-YFKPBYRVSA-N	18.44	145.09825,170.0935,128.0717
Homocitrulline	Amino acids and derivatives	188.1035	C7H15N3O3	XIGSAGMEBXLVJJ-YFKPBYRVSA-N	10.88	145.09825
Homocystine	Amino acids and derivatives	267.0473	C8H16N2O4S2	ZTVZLYBCZNMWCF-UHFFFAOYSA-N	11.92	72.00985,132.0123,179.04559,114.9867
Homogentisic acid	2(hydroxyphenyl)acetic acids	167.0350	C8H8O4	IGMNYECMUMZDDF-UHFFFAOYSA-N	5.31	123.04482,121.0295,108.02166
Homoserine	Amino acids and derivatives	118.0504	C4H9NO3	UKAUYYFTDYCKQA-VKHMYHEASA-N	10.80	100.04066
Homovanillic acid	Methoxyphenols	181.0506	C9H10O4	QRMZSPFSDQBLIX-UHFFFAOYSA-N	6.50	137.0608,122.03733,151.04007,163.04007,148.01487

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Hydrocortisone	21-hydroxysteroids	361.2015	C21H30O5	JYGXADMDTFJGBT-VVUMJDOOSA-N	2.83	331.19148,297.14962,315.16018,189.091,135.08154
Hydroxy-3-methoxymandelic acid	Methoxyphenols	197.0450	C9H10O5	CGQCWMI AEPEHNQ-QMMMGPBSA-N	6.44	179.0376,137.02603
Hydroxyglutaric acid	Short-chain hydroxy acids and derivatives	147.0293	C5H8O5	HWXBTVNRSUOJR-GSVOUGTGSA-N	10.22	87.01507,129.01933,85.0295,103.04007
Hydroxyphenyllactic acid	Phenylpropanoic acids	181.0501	C9H10O4	JVGVDSSUAVXRDY-UHFFFAOYSA-N	6.10	87.00926,163.04017,135.04525
Hypotaurine	Sulfinic acids	108.0119	C2H7NO2S	VVIUBCNYACGLLV-UHFFFAOYSA-N	10.68	
Imidazole-acetic Acid	Substituted imidazoles	125.0351	C5H6N2O2	PRJKNHOMHKJCEJ-UHFFFAOYSA-N	10.65	81.04582
Indole-2-carboxylic acid/Indole-3-carboxylic acid	Indolecarboxylic acids and derivatives	160.0398	C9H7NO2	HCUARIEZVDMPT-UHFFFAOYSA-N	3.50	116.05057
Indole-3-lactic acid	Indolyl carboxylic acids and derivatives	204.0661	C11H11NO3	XGILAAAMKEQUXLS-UHFFFAOYSA-N	4.69	158.0624,116.05154
Indole-3-propionic acid	Indolyl carboxylic acids and derivatives	188.0717	C11H11NO2	GOLXRNDWAUTYKT-UHFFFAOYSA-N	4.13	59.01429,116.05043,129.0571,142.06622,144.08187
Indoxyl sulfate	Arylsulfates	212.0023	C8H7NO4S	BXFFHSIDQOFMLE-UHFFFAOYSA-N	3.80	79.95659,132.04433,80.96519
Inosine	Purine nucleosides	267.0729	C10H12N4O5	UGQMRVRMYASKQ-KQYNXCUSA-N	6.85	135.03375,108.02034
Inosine 5-triphosphate	Purine ribonucleoside triphosphates	506.9719	C10H15N4O14P3	HAEJPQIATWHALX-KQYNXCUSA-N	10.65	158.92629
Isocitric acid	Tricarboxylic acids and derivatives	191.0192	C6H8O7	ODBLHEXUDAPZAU-VVJHMBFSA-N	11.16	111.00877,173.00916,73.0295,85.0295
Isoxanthopterin	Pterins and derivatives	178.0370	C6H5N5O2	GLKCOBIIKZYKFN-UHFFFAOYSA-N	7.50	136.01525,161.0105
Ketoleucine	Short-chain keto acids and derivatives	129.0557	C6H10O3	BKAJNAXTPSGJCU-UHFFFAOYSA-N	3.50	72.99312,85.06589
Kynurenic acid	Quinoline carboxylic acids	188.0353	C10H7NO3	HCZHHEIFKROPDY-UHFFFAOYSA-N	4.72	144.04549,142.02984
Kynurenine	Ketones	207.0770	C10H12N2O3	YGFSJZOEDVAXAB-QMMMGPBSA-N	7.22	190.0536,92.0513
Lactate	Alpha hydroxy acids and derivatives	89.0239	C3H6O3	JVTAAEKCFNVCJ-REOHLBHSAN	7.69	71.01385
Lauric acid	Medium-chain fatty acids	199.1704	C12H24O2	POULHZVOKOAJMA-UHFFFAOYSA-N	2.28	95.05024,109.06589,99.04515,81.03459,85.0295
Levulinic acid	Gamma-keto acids and derivatives	115.0395	C5H8O3	JOXCMJARBKPKM-UHFFFAOYSA-N	6.52	85.0295,99.04647,73.0295,97.0295
Linoleic acid	Lineolic acids and derivatives	279.2324	C18H32O2	OYHQOLUKZRVURQ-HZJYTRNSA-N	2.54	125.09719,137.09719,113.09719,155.10775,163.11284
Linolenic Acid	Lineolic acids and derivatives	277.2167	C18H30O2	DTOSIQBPPRVQHS-PDBXOOCHSA-N	2.33	163.11184,181.15979
L-Isoleucine/L-Leucine/L-allo-Isoleucine	Amino acids and derivatives	130.0868	C6H13NO2	AGPKZVBTJJNPAG-WHFBIAKZSA-N	8.5	74.02475,83.05024,113.0608,100.0404

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Lithocholic acid	Hydroxy bile acids, alcohols and derivatives	375.2899	C24H40O3	SMEROWZSTRWXGI-YPLGJCPNSA-N	2.37	143.10775
Luteolin	Flavones	285.0399	C15H10O6	IQPNAANSBPBGFQ-UHFFFAOYSA-N	2.56	107.01385,163.03975,267.02945,109.02942,177.01908
Malic acid	Beta hydroxy acids and derivatives	133.0137	C4H6O5	BJEPYKJPYRNKOW-UHFFFAOYSA-N	11.75	115.00368,71.01385
Maltotriose	Oligosaccharides	503.1612	C18H32O16	FYGDTMLNYKFZSV-DZOUCCMHSA-N	5.62	108.02024,161.04643,251.07807
Melatonin	3-alkylindoles	231.1139	C13H16N2O2	DRLFMBDRBRZALE-UHFFFAOYSA-N	2.50	144.0469,157.05475
Melibiose	Glycosyl compounds	341.1084	C12H22O11	DLRVVLDZNNYCBX-ABXHMFFYSA-N	11.09	179.05611,221.0675
Methionine	Amino acids and derivatives	148.0432	C5H11NO2S	FFEARJCKVFRZRR-BYPYZUCNSA-N	8.58	
m-Hydroxyhippuric Acid	Benzamides	194.0453	C9H9NO4	XDOFWFNMYJRHEW-UHFFFAOYSA-N	5.98	93.03541,150.0575
Myoinositol	Secondary alcohols	179.0555	C6H12O6	CDAIMWEQUEBRE-GPIVLXJGSA-N	11.16	71.01484,87.00877,99.00877
Myristic acid	Long-chain fatty acids	227.2011	C14H28O2	TUNFSRHWOTWDNC-UHFFFAOYSA-N	2.25	157.1234,113.0608,155.10775
N-Acetyl-5-hydroxytryptamine	Hydroxyindoles	217.0983	C12H14N2O2	MVAWJSIDNICKHF-UHFFFAOYSA-N	2.90	58.02967,131.03647,144.04438
N-Acetyl-D-mannosamine	Aminosaccharides	220.0821	C8H15NO6	OVRNDRQMDRJTHS-ZTVVOAFPSA-N	8.44	87.00877,170.04588
N-Acetylglucosamine	Aminosaccharides	220.0821	C8H15NO6	OVRNDRQMDRJTHS-FMDGEEDCSA-N	8.06	
N-Acetylglutamic acid	Amino acids and derivatives	188.0559	C7H11NO5	RFMMMVDNIPUKGG-UHFFFAOYSA-N	9.99	128.03595
N-Acetylglycine	Amino acids and derivatives	116.0353	C4H7NO3	OKJIRPAQVSHGFK-UHFFFAOYSA-N	9.00	74.02475,59.01385
N-Acetyl-L-alanine	Amino acids and derivatives	130.0504	C5H9NO3	KTHDTJVBEPMMGL-VKHYHEASA-N	7.67	88.04092
N-Acetyl-L-arginine	Amino acids and derivatives	215.1144	C8H16N4O3	SNEIUMQYRCDYCH-LURJTMIESA-N	10.99	173.09317,131.0826,129.10334
N-Acetyl-L-aspartic acid	Amino acids and derivatives	174.0402	C6H9NO5	OTCCIMWXFLJLIA-BYPYZUCNSA-N	10.11	156.03256,88.0404,130.05097
N-Acetyl-Leucine	Amino acids and derivatives	172.0974	C8H15NO3	WXNXCEHYYPACJF-ZETCQYMHSA-N	4.21	130.08735
N-Acetyl-L-methionine	Amino acids and derivatives	190.0538	C7H13NO3S	XUYPLNMDZIRQH-UHFFFAOYSA-N	4.62	148.04575,142.05097,98.06114,84.04549
N-Acetyl-L-ornithine	Amino acids and derivatives	173.0926	C7H14N2O3	JRLGPAXAGHMNOL-LURJTMIESA-N	11.39	131.0826
N-Acetyl-L-tyrosine	Amino acids and derivatives	222.0766	C11H13NO4	CAHKINHBCWCHCF-JTQLQIEISA-N	6.05	180.0693,107.052
N-Acetylneuraminic acid	Sugar acids and derivatives	308.0981	C11H19NO9	SQVRNKHJWKZAKO-PFQGNLYSA-N	9.61	170.04588,87.00877
N-Acetylserotonin	Hydroxyindoles	217.0983	C12H14N2O2	MVAWJSIDNICKHF-UHFFFAOYSA-N	2.90	58.02967,131.03647,144.04438
NAD	(5'->5')-dinucleotides	662.1013	C21H27N7O14P2	BAWFJGZGIEFAR-NNYOXOHSSA-O	10.42	

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
NADH	(5'->5')-dinucleotides	664.1169	C21H29N7O14P2	BOPGDPNILDQYTO-UHFFFAOYSA-N	9.87	346.0558,408.0116,158.9253,397.0207,78.9591
NADP	(5'->5')-dinucleotides	742.0676	C21H28N7O17P3	XJLXINKUBYWONI-NNYOXOHOSSA-O	10.88	158.92771,408.01782,620.02016,158.92538
Nicotinic acid	Pyridinecarboxylic acids	122.0248	C6H5NO2	PVNIIMVLHYAWGP-UHFFFAOYSA-N	6.30	78.03942
Nicotinoylglycine	Amino acids and derivatives	179.0462	C8H8N2O3	ZBSGKPYXQINNGF-UHFFFAOYSA-N	7.52	78.03492,135.05552,133.04074
N-Isovalerylglycine	Amino acids and derivatives	158.0823	C7H13NO3	ZRQXMKMBBMNNQC-UHFFFAOYSA-N	5.43	74.02475,114.09244,98.06114
N-Methyl-D-aspartic acid	Amino acids and derivatives	146.0453	C5H9NO4	HOKKHZGPKSLGJE-GSVOUGTGSA-N	10.28	115.00368
N-Methyl-L-glutamic acid	Amino acids and derivatives	160.0610	C6H11NO4	XLBVNMSMFQMKEY-BYPYZUCNSA-N	10.15	128.03532,71.01385,146.04588
Nonadecanoic acid	Long-chain fatty acids	297.2793	C19H38O2	ISYWECDZWTKFF-UHFFFAOYSA-N	2.50	155.10775,139.11284
Norleucine	Amino acids and derivatives	130.0868	C6H13NO2	LRQKBLKVPFOOQJ-YFKPBYRVSAN	7.75	
N-Tiglylglycine	Amino acids and derivatives	156.0661	C7H11NO3	WRUSVQOKJIDBLP-HWKANZROSA-N	4.76	112.07876,74.0264
N-α-Acetyl-L-glutamine	Amino acids and derivatives	187.0719	C7H12N2O4	KSMRODHGGIIXDV-UHFFFAOYSA-N	8.25	127.05201,145.06277,169.06325
Oleic acid	Long-chain fatty acids	281.2480	C18H34O2	ZQPPMHVWEC SIRJ-KTKRTIGZSA-N	2.45	129.0921,141.0921,195.17544,111.08154,99.08154
O-Phosphoethanolamine	Phosphoethanolamines	140.0118	C2H8NO4P	SUHOOTKUPISOBE-UHFFFAOYSA-N	12.96	78.95943,96.96962
Ornithine	Amino acids and derivatives	131.0820	C5H12N2O2	AHLPHDHMMVZTML-BYPYZUCNSA-N	19.30	114.05605
Orotic acid	Pyrimidinecarboxylic acids and derivatives	155.0098	C5H4N2O4	PXQPEWDEAKTCGB-UHFFFAOYSA-N	6.80	111.01883
Oxoglutaric acid/3-Methylglutaric acid/Adipic acid/Acetonedicarboxylic acid	Gamma-keto acids and derivatives	145.0137	C5H6O5	KPGXRSRHYNQIFN-UHFFFAOYSA-N	10.60	83.01385,101.02419
Palmitoleic Acid	Long-chain fatty acids	253.2167	C16H30O2	SECPZKHBENQXJG-FPLPWBNSAN	2.26	81.03476,125.09698,113.09711,155.10751
Pantothenic acid	Amino acids and derivatives	218.1028	C9H17NO5	GHOKWGTUZJEAQD-UHFFFAOYSA-N	6.83	88.04074,71.051
Pentadecanoic acid	Long-chain fatty acids	241.2167	C15H30O2	WQEPLUUGTLDZJY-UHFFFAOYSA-N	2.32	141.0921,195.17544,114.08154
Phenylalanine	Amino acids and derivatives	164.0711	C9H11NO2	COLNVLDHVKWLRT-QMMMGPBSAN	7.27	74.02475,72.0091,91.05532,147.04515
Phosphoenolpyruvic acid	Phosphate esters	166.9751	C3H5O6P	DTBNBXWJWCWCIK-UHFFFAOYSA-N	10.98	78.96077,148.96671
Pimelic acid	Medium-chain fatty acids	159.0663	C7H12O4	WLJVNTCWHIRURA-UHFFFAOYSA-N	10.52	97.06589,85.02981
Pipecolinic acid	Amino acids and derivatives	128.0711	C6H11NO2	HXEACLLIILLPRG-UHFFFAOYSA-N	9.73	
Progesterone	Gluco/mineralocorticoids, progestogens and derivatives	313.2167	C21H30O2	RJKFOVLPORLFTN-LEKSSAKUSA-N	2.35	135.08154,123.08154,109.6589,107.05024,121.06589

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Proline	Amino acids and derivatives	114.0555	C5H9NO2	ONIBWKKTOPOVIA-BYPYZUCNSA-N	10.11	
Pyridoxal	Pyridoxals and derivatives	166.0504	C8H9NO3	RADKZDMFGJYCBB-UHFFFAOYSA-N	4.89	108.04746,138.05841
Pyridoxamine	Pyridoxamine 5'-phosphates	167.0826	C8H12N2O2	NHZMQXZHNVTQA-UHFFFAOYSA-N	16.82	108.04549,121.05531,150.05605
Pyridoxine	Pyridoxines	168.0666	C8H11NO3	LXNHXLLTXMVWPM-UHFFFAOYSA-N	4.80	122.06114
Pyroglutamic acid	Amino acids and derivatives	128.0348	C5H7NO3	ODHCTXKNWHHXJC-VKHYHEASA-N	9.20	82.02984
Quinaldic acid	Quinoline carboxylic acids	172.0398	C10H7NO2	LOAUVZALPPNFOQ-UHFFFAOYSA-N	4.20	
Quinic acid	Cyclic alcohols and derivatives	191.0555	C7H12O6	AAWZDTNXLGCEK-RKGSJAZSA-N	10.54	85.0295,93.03459,127.04007
Quinolinic acid	Pyridinecarboxylic acids	166.0146	C7H5NO4	GJAWHXHKYXBSV-UHFFFAOYSA-N	12.47	122.02475,78.01927
Reserpine	Yohimbine alkaloids	607.2655	C33H40N2O9	QEVHRUUCFGRFIF-MDEJGZGSSA-N	2.75	592.24263,577.21915
Retinoic Acid	Retinoids	299.2011	C20H28O2	SHGAZHPCJPHSC-YCNIQYBTA-N	2.27	255.21182,111.04515,149.0608,85.0295,123.04515
Riboflavin	Flavins	375.1304	C17H20N4O6	AUNGANRZJHBGPY-SCRDCRAPSA-N	4.50	255.0933,255.08875,212.08294
Riboflavin 5-Monophosphate	Flavin nucleotides	455.0968	C17H21N4O9P	FVTCRASFADXXNN-SCRDCRAPSA-N	9.00	96.97139,183.0064,315.12245
Ribose 5-phosphate	Monosaccharides	229.0113	C5H11O8P	KTVPXOYAKDPRHY-SOOFDHNKSA-N	10.75	96.96822,78.9578,183.01094
Sebacic acid	Medium-chain fatty acids	201.1127	C10H18O4	CXMXRPHNRROMY-UHFFFAOYSA-N	8.29	95.05024
sn-glycero-3-Phosphocholine	Glycerophosphocholines	256.0950	C8H20NO6P	SUHOQUVVVLYNQR-QMMMGOBSA-N	11.91	152.9984,197.02205
Sorbitol/Galactitol	Sugar alcohols	181.0712	C6H14O6	FBPFZTCFMRRESA-JGWLITMVA-N	9.32	71.01385,101.02442,85.0295,119.03498,89.02442
Stearic acid	Long-chain fatty acids	283.2643	C18H36O2	QIQXTHQIDYTFRH-UHFFFAOYSA-N	2.25	157.1234,227.20165,155.10775
Suberic acid	Medium-chain fatty acids	173.0814	C8H14O4	TYFQFVWCELRYAO-UHFFFAOYSA-N	8.91	
Succinic acid	Dicarboxylic acids and derivatives	117.0193	C4H6O4	KDYFGRWQOYBRFD-UHFFFAOYSA-N	10.73	73.0295
Sucrose	Glycosyl compounds	341.1084	C12H22O11	CZMRCDWAGMREC-UGDNZRGBSA-N	10.07	179.05611,161.04555
Taurine	Organosulfonic acids	124.0074	C2H7NO3S	XOAAWQZATWQOTB-UHFFFAOYSA-N	10.60	79.95781
Taurochenodeoxycholate/Taurodeoxycholic acid/Tauroursodeoxycholic acid	Taurinated bile acids and derivatives	498.2889	C26H45NO6S	BHTRKEVTKCXOH-BJLOMENOSA-N	2.42	124.00739,106.98084
Tetradecanedioic acid	Long-chain fatty acids	257.1753	C14H26O4	HQHCYKULIHKCEB-UHFFFAOYSA-N	3.81	195.17544,239.16527
Theophylline	Xanthines	179.0574	C7H8N4O2	ZFXYFBGIUFBOJW-UHFFFAOYSA-N	2.80	122.03569,94.04127,164.03324
Thiamine	Thiamines	263.0966	C12H16N4OS	JZRWCGRZTMZEH-UHFFFAOYSA-N	16.24	

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Thiamine pyrophosphate	Thiamines	423.0293	C12H18N4O7P2S	AYEKOFBPNLCAJY-UHFFFAOYSA-O	10.04	176.93595,158.92538
Threonine	Amino acids and derivatives	118.0504	C4H9NO3	AYFVYJQAPQTCCC-GBXIJSLDSA-N	10.60	
Thymidine	Pyrimidine 2'-deoxyribonucleosides	241.0824	C10H14N2O5	IQFYKKMVGJFEH-XLPZGREQSA-N	4.16	225.06156,125.03565,83.05024,162.05605
Thymine	Hydroxypyrimidines	125.0351	C5H6N2O2	RWQNBROKXIBIV-UHFFFAOYSA-N	4.14	
Tocopherol	Vitamin E compounds	429.3732	C29H50O2	GVJHHUAWPYXKBD-IEOSBIPEA-N	2.11	189.6921,327.23295,163.07645,149.0608,151.07645
trans-2-Butene-dicarboxylic acid	Medium-chain fatty acids	143.0344	C6H8O4	YHGNXQAFNHCBTK-OWOJBTEDSA-N	10.16	
Traumatic acid	Medium-chain fatty acids	227.1283	C12H20O4	MAZWDMBBCPDUDJ-VQHVLOKHSAN-N	6.19	183.13905,165.12849
Tricarballic acid	Tricarboxylic acids and derivatives	175.0242	C6H8O6	KQTIICEAUMSDG-UHFFFAOYSA-N	11.18	157.01231,69.03397
Tridecanoic acid	Long-chain fatty acids	213.1854	C13H26O2	SZHOJFHSIKHZHA-UHFFFAOYSA-N	2.28	255.23295,337.3112,143.10775,123.08154
Trihydroxyisoflavone	Isoflavones	269.0455	C15H10O5	TZBJGXHYKVUXJN-UHFFFAOYSA-N	2.43	91.01894,93.03459
Triiodo-thyronine	Amino acids and derivatives	649.7822	C15H12I3NO4	AUYCYJSGJYCDS-LBPRGKRZSA-N	4.10	507.86625,632.75621,576.76638,522.87829
Tryptophan	Indolyl carboxylic acids and derivatives	203.0820	C11H12N2O2	QIVBCDIJIAJPQS-VIFPVBQESA-N	7.55	159.09255,130.06607,74.02535
UDP Glucose/UPDGalactose	Pyrimidine nucleotide sugars	565.0472	C15H24N2O17P2	HSCJRCZFDQWRP-LPTOLDLDSA-N	10.20	408.01782,158.92771,323.02859,272.95708,158.92538,96.96962
Undecanedicarboxylic acid	Long-chain fatty acids	243.1596	C13H24O4	DXNCZXXFRKPEPY-UHFFFAOYSA-N	3.98	225.14962
Undecanoic acid	Medium-chain fatty acids	185.1541	C11H22O2	ZDPHROOEEOARMN-UHFFFAOYSA-N	2.90	
Uracil	Pyrimidones	111.0200	C4H4N2O2	ISAKRJDGNUQOIC-UHFFFAOYSA-N	4.33	98.02383,70.02902
Ureidosuccinic acid	Amino acids and derivatives	175.0355	C5H8N2O5	HLKXYZVTANABHZ-REOHLBHSAN-N	10.61	132.03004,73.03026,115.00368
Uric acid	Xanthines	167.0205	C5H4N4O3	LEHOTFFKMJEONL-UHFFFAOYSA-N	8.09	124.01525,151.00234
Uridine	Pyrimidine nucleosides	243.0617	C9H12N2O6	DRTQHJPVMGBUCF-XVFCMESISA-N	5.70	140.0359,152.0361,200.0582
Uridine 5-diphosphate	Pyrimidine ribonucleoside diphosphates	402.9944	C9H14N2O12P2	XCCTYIAWTASOJW-XVFCMESISA-N	10.49	
Urocanic acid	Substituted imidazoles	137.0351	C6H6N2O2	LOIYMIARKYCTBW-OWOJBTEDSA-N	8.13	93.04622,71.01489,81.04582
Valeric acid	Straight chain fatty acids	101.0602	C5H10O2	NQPDZGIKBAWPEJ-UHFFFAOYSA-N	3.93	
Valproic acid	Branched fatty acids	143.1072	C8H16O2	NIJYAXOARWZEE-UHFFFAOYSA-N	2.57	

10.0 Supplemental material

NAME	Ontology	PRECURSORMZ	FORMULA	INCHIKEY	RETENTIONTIME	Fragmentlist
Vanillic acid	Methoxybenzoic acids and derivatives	167.0350	C8H8O4	WKOLLVMJNQIZCI-UHFFFAOYSA-N	5.94	108.02168,123.04515,152.01151,107.01385
Vitamin K1	Vitamin K compounds	449.3419	C31H46O2	MBWXNTAXLNYFJB-LKUDQCMESA-N	2.24	185.0608,434.31903,171.04515,79.05532,209.0608
Vitamin K2	Vitamin K compounds	443.2950	C31H40O2	DKHGEMRMDICWDU-GHDNBGIDSA-N	2.34	
Xanthine	Xanthines	151.0256	C5H4N4O2	LRFVTYWOQMYALW-UHFFFAOYSA-N	5.90	108.01925
Xanthosine	Purine nucleosides	283.0678	C10H12N4O6	UBORTCNDUKBEOP-UUOKFMHZA-N	7.95	151.02615
Xanthurenic acid	Quinoline carboxylic acids	204.0297	C10H7NO4	FBZONXHGGPHHIY-UHFFFAOYSA-N	6.76	158.02475,159.03258,169.0404

10.0 Supplemental material

10.3.2 CortecsT3 fatty acid identification data

Table 21 Identification information for fatty acid RPLC-MS method. Authentic fatty acid reference standards in a concentration of 100 μ M were measured as $[M-H]^-$ precursor ion in negative ionization mode using the fatty acid RPLC fatty acid method.

Compound name	Abbreviation	Formula	Precursor MZ	Retention time
5-Oxo-eicosatetraenoic acid	5-oxo-EETE	C20H30O3	317.21166	4.38
Arachidic acid	FA C20:0	C20H40O2	311.29496	8.99
Arachidonic acid	FA C20:4	C20H32O2	303.23236	6.31
cis-Eicosenoic acid	FA C20:1	C20H38O2	309.27936	8.33
Docosatetraenoic acid	FA 22:4	C22H36O2	331.26366	7.20
Dodecanoic acid	FA C12:0	C12H24O2	199.16976	4.63
Eicosatrienoic acid	FA C20:3	C20H34O2	305.24806	6.79
Erucic acid	FA C22:1	C22H42O2	337.31066	8.99
Hexadecanoic acid	FA C16:0	C16H32O2	255.23236	7.14
Linoleic acid	FA C18:2	C18H32O2	279.23236	6.47
Myristic acid	FA C14:0	C14H28O2	227.20106	5.94
Nervonic acid	FA C24:1	C24H46O2	365.34196	9.67
Octadecanoic acid	FA C18:0	C18H36O2	283.26366	8.16
Oleic acid	FA C18:1	C18H34O2	281.24806	7.27
Palmitic acid-d31	FA 16:0-d31	C16H1D31O2	286.42696	7.14
Palmitoleic acid	FA C16:1	C16H30O2	253.21676	6.13
Prostaglandin E1	PGE1	C20H34O5	353.23276	1.86
Tetracosanoic acid-d47	FA C24:0-d7	C24HD47O2	414.65256	10.17
Tetracosapentaenoic acid	FA 24:5	C24H38O2	357.27936	7.43
Tetracosatetraenoic acid	FA 24:4	C24H40O2	359.29496	8.08
Tetrahexanoic acid	FA 24:6	C24H36O2	355.26366	6.87

10.0 Supplemental material

10.4 List of MetaSci metabolites for metabolite library

Table 22 List of human metabolite reference standards acquired from MetaSci. Authentic metabolite reference standards acquired from MetaSci in singular glass vials. Metabolites names are oriented on the nomenclature of the supplier.

(+)-Delta-Tocopherol	2-Methyl-4-pentenoic Acid	5,6-Dihydro Thymine
(+)-Pantothenic acid, sodium salt	2-Methylbutyrylglycine	5,6-Dimethylbenzimidazole
(±)-3-Methyl-2-oxovaleric acid sodium salt	2-Methylglutaric acid	5-Aminoimidazole-4-carboxamide-1-β-D-ribofuranoside
(±)-Sodium β-hydroxyisobutyrate	2-Methylheptanoic Acid	5-Aminolevulinic acid hydrochloride
(±)-α-Lipoic acid	2-Methylhexanoic acid	5-Aminovaleric acid
(±)-α-Tocopherol	2-Methylhippuric acid	5-Hydroxyindole-3-acetic acid
(2S)-2-Amino-4-(S-methylsulfonimidoyl)butanoic acid	2-Methylmalonic acid	5-Hydroxymethyl-2-furancarboxylic acid
(R)-3-Hydroxybutanoic acid	2-Methylsuccinic acid	5-Methoxy-DL-tryptophan
(R)-Mevalonolactone	2-Octenoic acid	5-Methoxytryptamine
(S)-(-)-2-Hydroxyisocaproic acid	2-Oxo-3-phenylpropanoic acid	5-Methylcytosine
1,11-Undecanedicarboxylic acid	2-Oxadipic acid	5-Methyluridine
1,2-Dimristoyl-sn-glycero-3-phosphoethanolamine	2-Phenylethanol	5-Pregnen-3β-ol-20-one
1,2-Dipalmitoyl-sn-glycero-3-phosphate monosodium salt	2-Picolinic acid	5β-Cholestan-3α-ol
1,2-Dipalmitoyl-sn-glycero-3-phosphoethanolamine	3-(Methylthio)propionic acid	7-Dehydrocholesterol
1,2-Distearoyl-sn-glycero-3-phosphoethanolamine	3,3',5-Triiodo-L-thyronine sodium salt hydrate	Acetoacetic acid Lithium Salt
1,3-Acetonedicarboxylic acid	3,3-Dimethylglutaric acid	Acetylcholine chloride
1,3-Diaminopropane	3,4-Dihydroxybenzeneacetic acid	Adenine hydrochloride
1,4-Diaminobutane dihydrochloride	3,5-Dihydroxybenzoic acid	Adenosine
1,5-Anhydro-d-Glucitol	3-Aminoisobutanoic acid hydrate	Adenosine 5'-diphosphate sodium salt
1,5-Diaminopentane dihydrochloride	3-Chloro-L-Tyrosine	Adenosine 5'-triphosphate disodium salt
1,6-anhydro-b-D-Glucose	3-Hydroxy-3-methylglutaric acid	Adenosine cyclophosphate
10-Hydroxydecanoic acid	3-hydroxybutyric acid	Adenosine-5'-monophosphate monohydrat
10-Undecen-1-ol	3-Hydroxyisovaleric acid	Adipic acid
16-Dehydroprogesterone	3-Hydroxymandelic acid	Agmatine sulfate
17α-Ethynylestradiol	3-Hydroxyphenylacetic acid	All trans-Retinal
17-α-Hydroxyprogesterone	3-Hydroxypropionic Acid Sodium Salt	Allantoic acid
1-Hexadecanol	3-Indolebutyric acid	Allantoin
1-Methyl Adenosine	3-Indolepropionic acid	Aminoadipic acid
1-Methylnicotinamide chloride	3-Methoxytyramine HCl	Androsterone
1-Octadecanol	3-Methyl-2-oxobutanoic acid	Arachidic acid
2,3-Pyridinedicarboxylic acid	3-Methyladenine	Arachidonic Acid 250 mg in 1mL Ethanol
2,6-Dihydroxypyridine HCl	3-Methyladipic acid	Ascorbic acid
2'-Deoxyadenosine monohydrate	3-Methylglutaric acid	Azelaic acid
2'-Deoxyguanosine	3-Methyl-L-histidine	Betaine
2-Amino-1-phenylethanol	3-Methylvaleric acid	Bilirubin
2-Aminoisobutyric acid	3-Methylxanthine	Biotin
2'-Deoxycytidine	3-Nitro-L- tyrosine	Caffeine
2-Deoxy-D-ribose	3-Ureidopropionic acid	Calciferol
2'-Deoxyguanosine 5'-monophosphate (sodium salt hydrate)	4-(Acetylamino)butanoic acid	Caprylic acid
2'-Deoxyuridine	4,6-Dioxoheptanoic acid	Capryloyl glycine
2-Ethyl-2-Hydroxybutyric acid	4-Ethylactanoic acid	Carbamazepine
2-Ethylbutyric Acid	4-Guanidinobutyric acid	Chenodeoxycholic acid sodium salt

10.0 Supplemental material

2-Hexanone	4-Hydroxy-3-methoxyphenylacetic Acid/ Homovanillic acid	Chitin
2-Hydroxy-2-methylbutyric acid	4-Hydroxyphenylacetic acid	Cholestanol
2-Hydroxy-3-methylbutyric acid	4-Hydroxyphenylpyruvic acid	Cholesterol
2-Hydroxybutyric acid	4-Methyl-n-octanoic Acid	Cholesterol 3-Sulfate Sodium Salt
2-Hydroxycaproic acid	4-Methylvaleric acid	Cholesteryl linoleate
2-Hydroxyoctanoic acid	4-Pentenoic acid	Cholesteryl oleate
2-Isopropylmalic acid	4-Pyridoxic acid	Cholesteryl palmitate
Cholesteryl stearate	DL-4-Hydroxy-3-methoxymandelic acid	Glycochenodeoxycholic acid sodium salt
Cholic acid	DL-5-Hydroxylysine hydrochloride	Glycocholic acid hydrate
Choline chloride	DL-6,8-Thioctamide	Glycodeoxycholic acid sodium salt
cis-10-Nonadecenoic acid	DL-Homocystine	Glycolaldehyde dimer
cis-11-Eicosenoic acid	DL-Indole-3-lactic acid	Glycoursodeoxycholic acid
cis-3-Hexen-1-ol	DL-Isocitric acid trisodium salt hydrate	Guanidine hydrochloride
cis-4,7,10,13,16,19-Docosahexaenoic acid	DL-p-Hydroxyphenyllactic acid	Guanidoacetic acid
cis-4-Hydroxy-D-proline	DL-Threitol	Guanine
cis-8,11,14-Eicosatrienoic Acid 1.5 mg EtOH sln	DL- α -Hydroxyglutaric acid disodium salt	Guanosine
cis-Aconitic acid	DL- β -Leucine hydrochloride	Guanosine 5'-triphosphate sodium salt hydrate
Citicoline sodium salt	D-Mannose	Guanosine 5'-monophosphate (disodium salt)
Citric acid	Docosanoic acid	Heneicosanoic acid
Coenzyme Q10	Dodecanedioic acid diammonium salt	Heptadecanoic acid
Corticosterone	Dodecanoic acid	Heptanoic acid
Cortisone	Dopamine hydrochloride	Hexadecanedioic acid
Creatine monohydrate	D-Ribose 5-phosphate disodium salt hydrate	Hexadecanoic acid
Creatinine	Elaidic Acid	Hexanoic acid
Crotonic acid	Epinephrine	Hexanoyl Glycine
Cyanocobalamin	Ergosterol	Hippuric acid
Cystathionine	Erucic acid	Histamine dihydrochloride
Cysteamine hydrochloride	Estradiol	Homogentisic acid
Cytidine	Estriol	H-Trp-NH ₂ .HCl
Cytidine 5'-triphosphate (disodium salt)	Estrone	Hydrocortisone
Cytidine 5'-monophosphate	Ethanolamine hydrochloride	Hydoxycholic acid
Cytosine	Ethosuximide	Hypotaurine
D-(-)-Arabinose	Ethylmalonic acid	Hypoxanthine
D-(-)-Fructose	Farnesyl Acetate (mixture of isomers)	Imidazole-4(5)-acetic Acid Hydrochloride
D-(-)-Quinic acid	Flavin Adenine Dinucleotide Disodium Salt Hydrate	Indole
D-(+)-Galactose	Flavone	Indole-2-carboxylic acid
D-(+)-Glyceric acid Hemicalcium salt	Folic acid	Indole-3-carboxylic acid
D-(+)-Xylose	Folinic acid calcium salt hydrate	Indoxyl sulfate potassium salt
D-3-Phenyllactic acid	Formononetin	Inosine
D-Arabitol	Fumaric acid	Inosine 5'-triphosphate trisodium salt
Decanoic acid	Galactitol	Isoxanthopterin
Dehydroascorbic acid	Genistein	Kynurenic acid
Deoxyadenosine monophosphate	Glutaconic acid	L-(-)-3-Phenyllactic acid
Deoxycholic acid	Glutaric acid	L-(-)-Malic acid
Deoxyinosine	Glyceraldehyde	L-(-)-Norepinephrine (+)-bitartrate salt monohydrate
D-Fructose-1,6-diphosphate trisodium salt octahydrate	Glyceraldehyde	L-(-)-Pipicolinic acid
D-Fructose-6-phosphate disodium	Glycerol	L-5-Hydroxytryptophan

10.0 Supplemental material

D-Galactonic acid hemicalcium salt	Glycerol Tri-(cis-13)-docosenoate	L-Alanine
D-Glucosamine 6-phosphate	Glycerol Tributryrate	L-allo-Isoleucine
D-Glucosamine hydrochloride	Glycerol Tridecanoate	Lanosterol
D-Glucose 6-phosphate sodium salt	Glycerol Tridocosanoate	L-Anserine nitrate salt
D-Glucuro-3,6-lactone	Glycerol Trieicosanoate	L-Arabinose
D-Glucuronic acid	Glycerol Trihexanoate	L-Arabitol
Dihydrouracil	Glycerol trilinoleate	L-Arginine
Dihydroxyacetone phosphate	Glycerol Tri-n-octanoate	L-Asparagine
Diosmin	Glycerol Tripalmitate	L-Aspartic acid
DL -3,4-Dihydroxymandelic acid	Glycine	L-Carnitine inner salt
L-Carnosine	N-(5-Aminopentyl)acetamide	Palmitoleic Acid
L-Citrulline	N,N-Dimethylglycine hydrochloride	Pentadecanoic acid
L-Cysteic acid monohydrate	N-Acetyl-5-hydroxytryptamine	Petroselinic Acid
L-Cysteine	N-Acetyl-D-galactosamine	Phenyl-Ac-Gln-OH
L-Cystine	N-Acetyl-DL-serine	Phenyl-ac-Gly-OH
L-Dopa	N-Acetyl-DL-Tryptophan	Phospho(enol)pyruvic Acid Monopotassium Salt
Levulinic acid	N-Acetyl-D-mannosamine	Phosphorylcholine chloride calcium salt tetrahydrate
L-Fucose	N-Acetylglucosamine	Phosphoserine
L-Glutamic acid	N-Acetylglutamic acid	Pimelic acid
L-Glutamine	N-Acetyl glycine	Progesterone
L-Glutathione oxidized	N-Acetyl-L-alanine	Propionic acid
L-Glutathione reduced	N-Acetyl-L-arginine dihydrate	Purine
L-Histidine monohydrochloride monohydrate	N-Acetyl-L-aspartic acid	Pyridoxal 5'-phosphate hydrate
L-Histidinol Dihydrochloride	N-Acetyl-L-cysteine	Pyridoxal hydrochloride
L-Homoarginine Hydrochloride	N-Acetyl-L-leucine	Pyridoxamine dihydrochloride
L-Homocitrulline	N-Acetyl-L-methionine	Pyridoxine
L-homocystine	N-Acetyl-L-ornithine	Pyrimidine
L-Homoserine	N-Acetyl-L-phenylalanine	Pyruvic acid
L-Hydroxyproline	N-Acetyl-L-Proline	Pyruvic Aldehyde sln aq 35-45%
Linoleic acid	N-Acetyl-L-tyrosine	Quinaldic acid
L-Isoleucine	N-Acetylneuraminic acid	Raloxifene HCl
Lithocholic acid	NAD	Reserpine
L-Kynurenine	NADH, disodium salt hydrate	Retinoic Acid
L-Leucine	NADP hydrate	Retinol
L-Lysine	n-Butyric acid, Na salt	Riboflavin
L-Methionine	Nervonic acid	Riboflavin 5'-Monophosphate Sodium Salt
L-Nicotine	n-Glycerol 3-Phosphate Bis(cyclohexylammonium) Salt	Ribose
L-Norleucine	Niacinamide	Sarcosine
L-Norvaline	Nicotinic acid	Sebacic acid
L-Phenylalanine	Nicotinic acid	Senecioic acid
L-Proline	N-Iso valerylglycine	Serotonin
L-Pyroglutamic acid	N-Methyl-D-aspartic acid	sn-glycero-3-Phosphocholine
L-Serine	N-Methyl-L-glutamic acid	Sodium alpha-ketoisocaproate
L-Threonine	N'-Methylnicotinamide	Sodium creatine phosphate dibasic tetrahydrate
L-Tryptophan	N-Methyltyramine HCl	Sodium hyaluronate
L-Tyrosine	Nonadecanoic acid	Sodium L-lactate
Luteolin	Nonanoic acid	Sodium taurochenodeoxycholate

10.0 Supplemental material

L-Valine	N-Propionylglycine	Sorbitol
Malonic acid	N-Tiglylglycine	Spermidine
Maltose Monohydrate	N- α -Acetyl-L-glutamine	Spermine tetrahydrochloride
Maltotriose	N- α -Acetyl-L-lysine	Suberic acid
Melatonin	O-Acetyl-L-carnitine HCl	Suberyl Glycine
Melibiose	Octadecanoic acid	Succinic acid
Menadione	Oleic acid	Sucrose
Methylguanidine HCl	O-Phosphorylethanolamine	Tamoxifen citrate
m-Hydroxyhippuric Acid	Ornithine HCl	Taurine
Myoinositol	Orotic acid	Taurodeoxycholic acid sodium salt hydrate
Myristic acid	Ortho-Hydroxyphenylacetic acid	Tauroursodeoxycholic acid, Na salt
N-(3-Phenylpropionyl)glycine	Oxalacetic acid	Tetracosanoic acid
N-(4-Aminobutyl)acetamide hydrochloride	Oxoglutaric acid	Tetradecanedioic acid
Theobromine	Tyramine	Vitamin K1
Theophylline	UDP Glucose (disodium salt)	Vitamin K2
Thiamine hydrochloride	Undecanoic acid	Vitamin K4
Thiamine monophosphate chloride dihydrate	UPD- α -D-Galactose	Xanthine
Thiamine pyrophosphate hydrochloride	Uracil	Xanthosine
Thymidine	Urea	Xanthurenic acid
Thymine	Ureidosuccinic acid	α -D-Galactose-1-phosphate dipotassium salt hydrate
Tiglic acid	Uric acid	α -D-Glucose 1-phosphate disodium salt hydrate
trans-2-Butene-1,4-dicarboxylic Acid	Uridine	α -D-Glucose anhydrous
trans-2-Methyl-2-pentenoic Acid	Uridine 5'-diphosphate disodium salt hydrate	α -Hydroxyhippuric acid
trans-Aconitic acid	Uridine-5'-phosphoric acid disodium salt	α -Lactose
Traumatic acid	Urocanic acid	α -Linolenic Acid
Tricarballic acid	Valeric acid	β -Alanine
Tricosanoic acid	Valproic acid sodium salt	β -Carotene
Tridecanoic acid	Vanillic acid	β -Glycerophosphoric acid, disodium salt pentahydrate
Tryptamine	Vitamin D3	β -Lactose

10.0 Supplemental material

10.5 List of Figures

Figure 1 The metabolome is the interface between genotype and phenotype	4
Figure 2 Planning a metabolomics study requires the definition of multiple steps	5
Figure 3 Liquid chromatography is a separation technique.....	13
Figure 4 Mass spectrometers separate analyte ions based on their mass to charge ratio	14
Figure 5 Metabolomics platforms need to include multiple steps in their pipelines.	32
Figure 6 The AbsoluteIDQ p180-kit from Biocrates is a set of metabolomics method for the analysis of lipids and biogenic amines.....	33
Figure 7 The metabolite identification library is generated from the measurements of metabolite reference standards.....	43
Figure 8 Exemplary chromatograms of acceptable and problematic FIA-MS measurements.....	48
Figure 9 Exemplary picture of the LC-MS method selection process.....	51
Figure 10 Coverage of the metabolite reference library: ZIC-HILIC-MS.....	52
Figure 11 Scheme of a typical, structured lipid.	54
Figure 12 A typical total ion chromatogram of EDTA plasma measured with the lipidomics method shows different lipid signals.	55
Figure 13 A plot of identified lipids and lipid classes from EDTA-plasma over different retention and mass ranges further shows a structure in lipid retention times.....	55
Figure 14 Reference lipids of different lipid classes similarly show patterning.....	56
Figure 15 The mean intensity of different lipid classes is greater in the RPLC-SPE.....	57
Figure 16 The mean group variance of different lipid classes is lower in the RPLC-SPE.	58
Figure 17 The RPLC-SPE has superior intensities and smaller deviations than the Bligh and Dyer method: Exemplary lipids.	58
Figure 18 Signal-to-noise ratios of two SPE methods: ZIC-HILIC MS in positive ionization mode.....	60
Figure 19 Signal-to-noise ratios of two SPE methods: ZIC-HILIC MS in negative ionization mode	60
Figure 20 The RPLC-SPE method has lower deviations, but not higher mean intensities than the ZIC-SPE method in selected hydrophilic metabolites.....	61
Figure 21 Stability of mass deviation to reference standards: ZIC-HILIC-MS.....	62
Figure 22 Stability of intrarun mass deviation: ZIC-HILIC-MS.....	63
Figure 23 Stability of retention error to reference standards: ZIC-HILIC-MS	63
Figure 24 Stability of intra-run retention time error: ZIC-HILIC-MS	64
Figure 25 Retention time distribution of metabolites in ZIC-HILIC MS	65
Figure 26 The ZIC-HILIC-MS method shows a high number of detectable metabolites	65
Figure 27 Different metabolites are detectable in different matrices	66
Figure 28 Using the ZIC-HILIC metabolomics pipeline, the metabolome can be covered in important checkpoints	66
Figure 29 Detectability of metabolites in different plasma concentrations in ZIC-HILIC-MS	67
Figure 30 Limit of detection of 50 metabolites that were not detectable in healthy specimens	69
Figure 31 4 exemplary metabolites that were previously not detected in biospecimen showed high linearity in spiked samples: ZIC-HILIC-MS	70
Figure 32 Distribution of Pearson correlation coefficient of identified metabolites from ZIC-HILIC-MS.....	71
Figure 33 Distribution of Pearson correlation coefficient of identified metabolites from CortecsT3 lipidomics method.....	71
Figure 34 Graph of Pearson correlation coefficient in dependence on the maximum plasma/blank-ratio: An alternative signal to noise ratio.	72
Figure 35 Exemplary metabolites that changed with the concentration of lysed human whole blood: ZIC-HILIC negative mode	74
Figure 36 Exemplary metabolites that changed with the concentration of lysed human whole blood: ZIC-HILIC positive mode	75
Figure 37 Exemplary metabolites that changed with the concentration of lysed human whole blood: CortecsT3 lipidomics method.....	76
Figure 38 Significant alterations in metabolome in different extraction states under various external conditions: ZIC-HILIC-MS	78
Figure 39 Representative metabolites changed by different extraction states under various external conditions: ZIC-HILIC-MS negative mode	78
Figure 40 Representative metabolites changed by different extraction states under various external conditions: ZIC-HILIC-MS positive mode.....	79
Figure 41 Significant changes in metabolome in different extraction states under various external conditions: CortecsT3 lipidomics method	79
Figure 42 Representative metabolites changed by different extraction states under various external conditions: CortecsT3 lipidomics method	80
Figure 43 Our untargeted metabolomics pipeline uses RPLC-SPE and two LC-MS-methods for the analysis of lipids and hydrophilic metabolites.....	85
Figure 44 Volcano-plot showing metabolic changes of Zellweger syndrome spectrum patients (ZSS/PEX1) in serum compared to the control group	91
Figure 45 Membrane lipids indicate differential phenotype between Zellweger syndrome spectrum patients (ZSS) and X-linked adrenoleukodystrophy (X-ALD) patients, as well as controls	94

10.0 Supplemental material

Figure 46 Volcano plot reveals accumulations of VLCFA lipids in <i>Abcd1^{tm1Kds}</i> mice compared to controls	102
Figure 47 LysoPC 26:0 is increased in serum and brain tissue of <i>Abcd1^{tm1Kds}</i> mice. Analysis of serum or brain tissue using the commercial metabolomics kit.	104
Figure 48 <i>Abcd1^{tm1Kds}</i> mice can be clustered based on different reactions to EAE	105
Figure 49 Volcano plot reveals metabolome differences in serum of symptomatic compared to asymptomatic <i>Abcd1^{tm1Kds}</i> EAE animals	106
Figure 50 Two polyunsaturated fatty acid-containing lysophosphatidylcholines are reduced in symptomatic EAE <i>Abcd1^{tm1Kds}</i> mice	106
Figure 51 Two polyunsaturated fatty acid-containing lysophosphatidylcholines are reduced in symptomatic wild-type mice.	107
Figure 52 Linear regression of lysoPC 20:4 serum concentration to phenotypical EAE scores in EAE <i>Abcd1^{tm1Kds}</i> mice hints at a potential connection between lysoPC concentration and disease strenght	107
Figure 53 LysoPC 20:3 and 20:4 are decreased in human patients before and after cALD beginning.	108
Figure 54 Very long chain fatty acid elongation cycle and ELOVL specificity	112
Figure 55 An overview of the plasma lipidome of a patient with an ELOVL1-mutation shows decreased signals in the high retention time and mass region.	115
Figure 56 Volcano-Plot comparing ELOVL1-Mt patient to controls in plasma shows decreased membrane lipids.	116
Figure 57 SM C24:1 is decreased in ELOVL1-Mt patient plasma	117
Figure 58 Volcano-Plot comparing ELOVL1-Mt patient plasma to controls using the targeted metabolomics kit reveals mostly reduction in tested membrane lipids	117
Figure 59 SM C24:1 is altered in patient ELOVL1-deficient patient plasma	118
Figure 60 Sphingomyelin SM C16:0 is found decreased in ELOVL1-Mt plasma using the p180-kit but not using the untargeted lipidomics method	118
Figure 61 Overview of ELOVL1-Mt patient fibroblast lipidome compared to controls under standard culture conditions shows mildly decreased signals in the high retention time and high mass region	119
Figure 62 Overview of ELOVL1-Mt patient fibroblast lipidome compared to controls under FCS withdrawal reveals strongly decreased signals in the high retention time and high mass region.	120
Figure 63 SM C24:1 deficient phenotype in ELOVL1-deficient cells can be exacerbated by FCS-challenge.....	121
Figure 64 Several metabolites are deregulated between plasma and cell lysate in OCT1 isoforms	129
Figure 65 Deregulated (Acyl-)Carnitine in different OCT1 isoforms	130
Figure 69 Scheme of typical lipid structures	156

10.0 Supplemental material

10.6 List of tables

Table 1 Mass spectrometer conditions for ZIC-HILIC-MS method.....	27
Table 2 Mass spectrometer conditions for the RPLC-MS CortecsT3 lipidomics method.....	27
Table 3 MS-Dial processing parameters for ZIC-HILIC-MS method in positive ionization mode.....	29
Table 4 MS-Dial processing parameters for ZIC-HILIC-MS method in negative ionization mode.....	29
Table 5 MS-Dial processing parameter for the CortecsT3 method.....	30
Table 6 UNIFI processing parameters for ZIC-HILIC positive mode.....	45
Table 7 UNIFI processing parameters for ZIC-HILIC negative mode.....	46
Table 8 Exemplary data set from tryptophan in the metabolite identification library.....	53
Table 9 TOP 5 metabolites altered by hemolyzed blood: ZIC-HILIC-MS in negative mode.....	73
Table 10 TOP 5 metabolites altered by hemolyzed blood: ZIC-HILIC-MS in positive mode.....	75
Table 11 TOP 5 metabolites altered by hemolyzed blood: CortecsT3 Lipidomics method.....	76
Table 12 ZSS patients with irregularly high GC-MS plasmalogen markers can still be identified as plasmalogen deficient by the FIA-MS marker PC ae 36:4.....	92
Table 13 VLCFA sphingomyelin species are elevated in the serum of X-ALD and ZWS patients, non-VLCFA sphingomyelin species are reduced in ZSS patients compared to controls.....	93
Table 14 D-BPIII patient with irregularly low GC-MS plasmalogen markers can still be identified as different to ZSS patients, but also controls, by the FIA-MS marker PC ae 36:4.....	95
Table 15 Clinical characteristics of X-ALD patients with confirmed ABCD1 mutation.....	101
Table 16 Significant changes of serum metabolites in Abcd1 ^{tm1Kds} mice compared to controls.....	103
Table 17 Significant changes of metabolites in brain tissue of Abcd1 ^{tm1Kds} mice compared to controls.....	103
Table 19 Common abbreviations of lipids.....	157
Table 20 Human metabolite compound library in ZIC-HILIC positive ionization mode. Human metabolite reference standards from MetaSci in a concentration of 100 µM were acquired using the ZIC-HILIC-MS method in positive ionization mode...	159
Table 21 Human metabolite compound library in ZIC-HILIC negative ionization mode.....	167
Table 22 Identification information for fatty acid RPLC-MS method.....	180
Table 23 List of human metabolite reference standards acquired from MetaSci.....	181