Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies

Marynka M. Ulaszewska[1,†], Christoph H. Weinert[2,†], Alessia Trimigno[3,†], Reto Portmann[4,†], Cristina Andres Lacueva[5], René Badertscher[4], Lorraine Brennan[6], Carl Brunius[7], Achim Bub[8], Francesco Capozzi[3], Marta Cialiè Rosso[9], Chiara E. Cordero[9], Hannelore Daniel[10], Stephanie Durand[11], Bjoern Egert[2], Paola G. Ferrario[8], Edith J.M. Feskens[12], Pietro Franceschi[13], Mar Garcia-Aloy[5], Franck Giacomoni[11], Pieter Giesbertz[14], Raúl González-Domínguez[5], Kati, Hanhineva[15], Lieselot Y. Hemeryck[16], Joachim Kopka[17], Sabine Kulling[2], Rafael Llorach[5], Claudine Manach[18], Fulvio Mattivi[1,19] Carole Migné[11], Linda H. Münger[20], Beate Ott[23], Gianfranco Picone[3], Grégory Pimentel[20], Estelle Pujos-Guillot[11], Samantha Riccadonna[13], Manuela Rist[8], Caroline Rombouts[16], Josep Rubert[1], Thomas Skurk[23], Pedapati S. C. Sri Harsha[6], Lieven Van Meulebroek[16], Lynn Vanhaecke[16], Rosa Vázquez-Fresno[22], David Wishard[22], and Guy Vergères[20,]*

[1]Department of Food Quality and Nutrition, Fondazione Edmund Mach, Research and Innovation Centre, San Michele all'Adige, Italy

[2]Department of Safety and Quality of Fruit and Vegetables, Max Rubner-Institut, Karlsruhe, Germany

[3]Department of Agricultural and Food Science, University of Bologna, Italy

[4]Method Development and Analytics Research Division, Agroscope, Federal Office for Agriculture, Berne, Switzerland

[5]Biomarkers & Nutrimetabolomics Laboratory, Department of Nutrition, Food Sciences and Gastronomy, XaRTA, INSA, Faculty of Pharmacy and Food Sciences, Campus Torribera, University of Barcelona, Barcelona, Spain. CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona, Spain

[6]School of Agriculture and Food Science, Institute of Food and Health, University College Dublin, Dublin, Ireland

[7]Department of Biology and Biological Engineering, Food and Nutrition Science, Chalmers University of Technology, Gothenburg, Sweden

[8]Department of Physiology and Biochemistry of Nutrition, Max Rubner-Institut, Karlsruhe, Germany

[9]Dipartimento di Scienza e Tecnologia del Farmaco Università degli Studi di Torino, Turin, Italy

[10]Nutritional Physiology, Technische Universität München, Freising, Germany

[11]Plateforme d'Exploration du Métabolisme, MetaboHUB-Clermont, INRA, UNH, Université Clermont Auvergne, Clermont-Ferrand, France

[12]Division of Human Nutrition, Wageningen University, Wageningen, The Netherlands

[13]Computational Biology Unit, Fondazione Edmund Mach, Research and Innovation Centre, San Michele all'Adige, Italy

[14]Molecular Nutrition Unit, Technische Universität München, Freising, Germany

[15]Institute of Public Health and Clinical Nutrition, Department of Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

[16]Laboratory of Chemical Analysis, Department of Veterinary Public Health and Food Safety, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium

[17]Department of Molecular Physiology, Applied Metabolome Analysis, Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

[18]INRA, UMR 1019, Human Nutrition Unit, Université Clermont Auvergne, Clermont-Ferrand, France

[19]Center Agriculture Food Environment, University of Trento, San Michele all'Adige, Italy

[20]Food Microbial Systems Research Division, Agroscope, Federal Office for Agriculture, Berne, Switzerland

[21]Else Kröner Fresenius Center for Nutritional Medicine, Technisal University of Munich, Munich, Germany

[22]Departments of Biological Sciences and Computing Science, University of Alberta, Edmonton, Canada

[23]ZIEL Institute for Food and Health, Core Facility Human Studies, Technical University of Munich, Freising, Germany

[†]First authors

*Correspondence: Dr. Guy Vergères, Schwarzenburgstrasse 161, 3003 Bern, Switzerland, guy.vergeres@agroscope.admin.ch

Keywords: GC-MS, LC-MS, metabolomics, NMR, nutrition

The life sciences are currently being transformed by an unprecedented wave of developments in molecular analysis, which include important advances in instrumental analysis as well as biocomputing. In light of the central role played by metabolism in nutrition, metabolomics is rapidly being established as a key analytical tool in human nutritional studies. Consequently, an increasing number of nutritionists integrate metabolomics into their study designs. Within this dynamic landscape, the potential of nutritional metabolomics (nutrimetabolomics) to be translated into a science, which can impact on health policies, still needs to be realized.

A key element to reach this goal is the ability of the research community to join, to collectively make the best use of the potential offered by nutritional metabolomics. This article, therefore, provides a methodological description of nutritional metabolomics that reflects on the state-of-the-art techniques used in the laboratories of the Food Biomarker Alliance (FoodBAll) (funded by the European Joint Programming Initiative a Healthy Diet for a Healthy Life (JPI HDHL)) as well as points of reflections to harmonize this field. It is not intended to be exhaustive but, rather, to present a pragmatic guidance on metabolomic methodologies, providing readers with useful "tips and tricks" along the analytical workflow.

Accepted Article

Graphical abstract. Nutritional metabolomics (nutrimetabolomics) requests a transdisciplinary effort spanning the fields of food science, human nutrition, analytical chemistry, and bioinformatics.
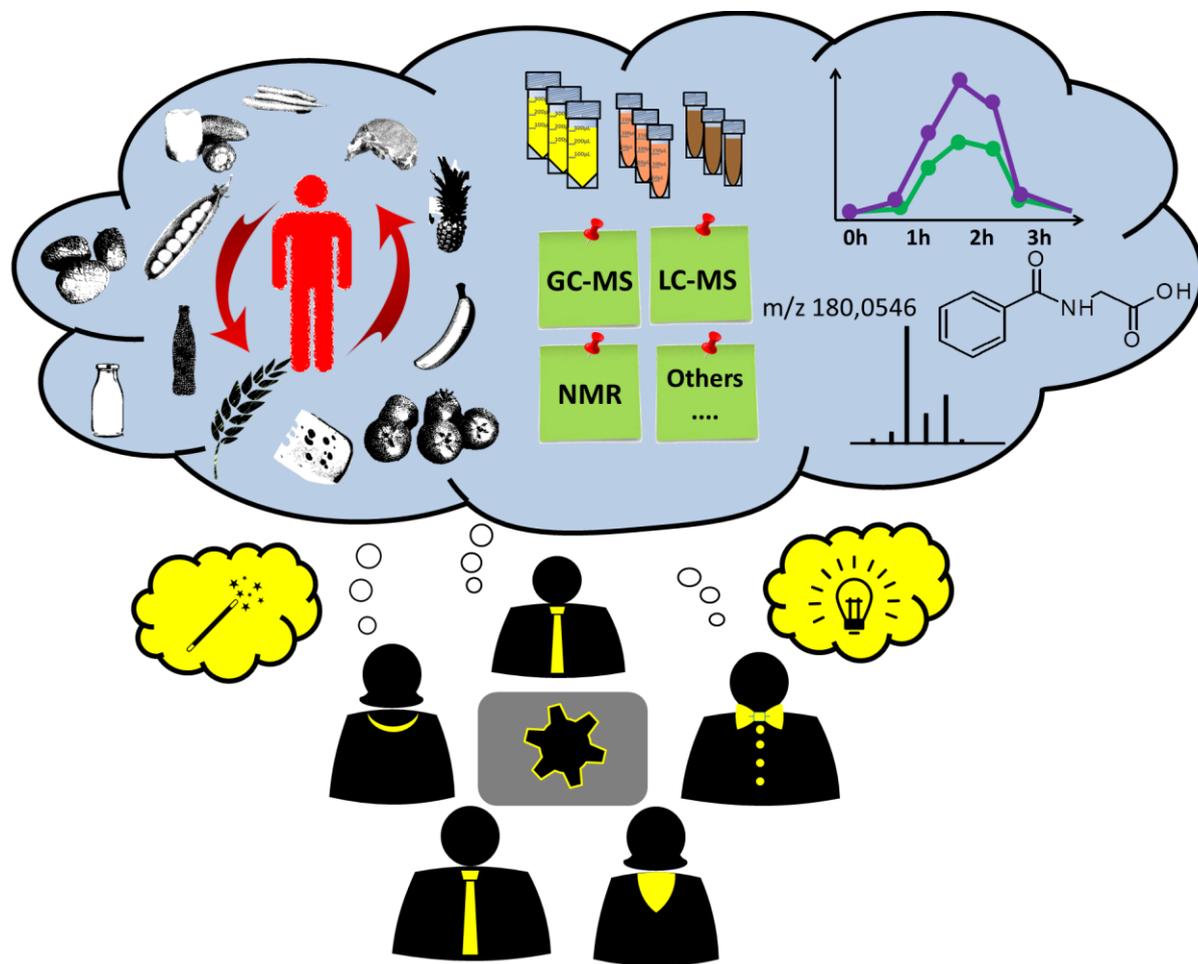
1. Introduction

The rapidly growing field of metabolomics focuses on the analysis of many hundreds of metabolites in complex specimens that include biofluids, tissues or cells. Recent advances in high-throughput metabolomic approaches have provided an improved understanding of altered metabolic pathways, new gene functions or the regulation of important enzymes. At the same time, the integration of metabolomics with nutritional science enhances current clinical and research practices by providing a deeper insight into the relationships between various metabolites and health status. Both society and industry insist and push for a mechanistic understanding of the impact of diet on human health. In nutritional metabolomics, the aim is usually to investigate perturbations of the human metabolome by specific diets, foods, nutrients, micro-organisms or bioactive compounds. Although nutrition research encompasses studies on the interaction of foods and nutrients with a range of biological model systems (human, animal, cellular, enzymatic) the concept of nutritional metabolomics employed in this article specifically refers to the application of metabolomics to the analysis of samples derived from human nutritional studies. As metabolomics is fundamentally phenotype driven, nutritional metabolomics provides better and more individualized biomarkers than most other techniques and is expected to furnish better indicators of dietary effects on a target population or patients. Ultimately, the intertwining of nutrition and metabolomics in nutritional metabolomics aims to achieve personalized prognostic and diagnostic nutrition – making nutrimetabolomics one of the most promising avenues for improving the nutritional care and dietary treatment of patients in the future.

There is a widely acknowledged need for standardization in both nutrition and metabolomics; however, the agreement and implementation of adequate standards is often difficult. These needs become even more evident when considering the overall goals of nutritional metabolomics: health improvement, disease prevention, better clinical practice, formulation of healthy foods, and substantiation of health claims. Fortunately, recent efforts by the international community have allowed agreements on data sharing and a consensus around the proposed standards to be reached [1, 2]. Advances in the "omics" fields have already led to the publication of validated, highly standardized datasets suitable for data mining/sharing and for inclusion in informative systems complying with the FAIR (Findable, Accessible, Interoperable, and Re-usable) protocol [3]. In that context, the international metabolomics community has made efforts to determine standard procedures [4], to evaluate the analytical performance of untargeted metabolomics [5], and to identify the needs for infrastructure [6], including data sharing [7]. The Food Biomarker Alliance (FoodBAll), which includes the authors of this article,

participates in this effort by fostering the application of novel metabolomic techniques for human nutrition studies as complement to traditional dietary assessment methods such as food frequency questionnaire and 24 hour recall [8].

There are several challenges in standardizing nutritional metabolomics. Among the most important issues, which greatly influence the validity and usefulness of results, are experimental design and sample definitions (e.g. type and number). Proper design of the experiment is a pre-requisite for obtaining a clean dataset, in which systematic errors, random errors, and artefacts are minimized. This supports the production of reliable and reproducible results. Secondly, sample preparation prior to analysis also represents an important step that should be executed with caution and under standardized and validated conditions, to avoid and/or reduce systematic errors. This is an essential requirement, because many biological samples change over time and are thus unstable during storage. Consequently, their chemical composition can vary greatly. The lack of standardization in the protocols used for sample handling and for acquiring mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectral data (e.g. controlling the pH of a biological sample in aqueous solution before recording the NMR spectrum) could lead to poor reproducibility and/or incorrect interpretation of the recorded data. A third challenge is the diversity of procedures involved in metabolomic data analysis. In untargeted metabolomic experiments, the choice of the algorithms (and their parameters) for metabolic feature extraction (hereafter referred as to 'feature') and data processing have a fundamental role in determining the outcomes of the analysis. An incorrect choice at this level will result in an erroneous representation of the chemical information encoded in the raw data (spectra). Problems may also arise in later stages of the analysis if advanced data analysis tools are inappropriately applied or proper validation strategies are not implemented. Indeed, whereas advances in analytical technologies have made MS and NMR particularly attractive and useful to the nutritional metabolomics community, the data analysis protocols used in many metabolomic approaches are still unsatisfying. Today's ultrasensitive instruments provide analytical scientists an unprecedented opportunity to detect more compounds at lower concentrations, resulting in a larger and larger proportion of unknown compounds to be identified, most of which are not found in available databases. This is particularly true in the field of nutritional metabolomics, where the focus is not only on human endogenous metabolites, but also on food-derived compounds introduced with the diet. Finally, the future evolution of nutritional metabolomics lies in integrating with other evolving "omics" techniques and with other well-defined tools used in disease diagnostics.

Nutritional metabolomics experiments should be thoroughly discussed by all participating stakeholders, such as nutritionists, analytical chemists, statisticians, medical doctors, data analysts, ethical experts and legislators. These discussions should be undertaken to address or resolve common issues regarding, for example, the expected level of intra- and inter-individual variation (metabotype), the uncertainty of measurements, the power of the investigation, the requirements for issuing health claims, etc. There are many opinions on these matters and widely differing approaches – but relatively little guidance or consensus. The aim of this article is therefore to conduct a critical review of this field that combines information available in the scientific literature and the opinions of experts from different scientific areas to provide some guidance regarding nutritional metabolomics methods – from beginning to end.

The structure of this article follows the workflow of an untargeted metabolomic experiment (**Table 1**). Chapters 2 and 3 describe the study design as well as provide considerations on sampling issues, two generic tasks that are essential to the success of nutritional studies. Chapters 4 to 7 and 9 describe experimental methodologies that require specific considerations depending on the selected (untargeted) technology (GC-MS, LC-MS, NMR). These chapters are summarized in the main text and covered in detail in the Appendices. In particular, they cover the pre-analytical component (sample preparation) as well as four sections on instrumental processing (chromatography, data acquisition, raw data file processing, and compound identification). Finally, Chapters 8 and 10 are devoted to data analysis and biological interpretations, since these topics are common to all three instrumental techniques. Both data analysis and biological interpretation are essential to derive information and to complete a full picture of metabolic changes in the body after a nutritional intervention. As reflected in the structure of this review, the metabolomic workflow in nutritional studies is essentially a linear process spanning from the study design to the biological interpretation of the data. At the same time, this workflow is also a complex process integrating branching into its structure in order to allow for the necessary flexibility associated with the complex analysis of human biology as well as with the wealth of technological solutions made available to researchers. This review, therefore, does not provide a single standardized metabolomic workflow but, rather, presents guidelines and expert opinions for each of the modules building this workflow. Of note, metabolomic technologies are currently also transforming food sciences by allowing an in depth characterization of food composition. Combing the food metabolome with the metabolome of biospecimens from human studies will further foster nutritional sciences. Food metabolomics is, however, beyond the scope of this article and the reader is referred to reviews on this topic [9-12].

2.   Study design

The fundamental prerequisite to ensure robust and trustworthy conclusions from study data is a suitable study outline. By definition, experimental design is the process of planning data acquisition in order to meet predefined objectives and providing the power for a statistically significant answer to the research question. Nutritional metabolomic investigations test a hypothesis dependent on a nutritional pattern or a defined food intake. These investigations range in type from double blind, randomized, placebo-controlled, nutrient supplementation studies to community based lifestyle intervention or population-based fortification projects. The study design process should therefore include careful consideration of the hypothesis, study duration, intervention, control and blinding, primary and secondary outcome measures (including assessment of background diet), acceptance criteria, statistical power, eligibility criteria, data collection methodology, and ways of measuring and encouraging compliance [13].

2.1.   Research hypothesis and experimental planning

Experimental design is about maximizing the possibilities of obtaining significant results in relation to the hypothesis being investigated. There are three initial steps in deciding how to proceed with a research idea. The first is to state a precise research question, including the concepts of interest. The second involves brainstorming about the primary and confounding variables and appropriate study subjects. Finally, it is necessary to state a specific, measurable hypothesis from which research designs and methods can be defined [14-16]. The research question and hypothesis to be tested will directly influence all other aspects of the study, including the study design and duration, the treatment/exposure and control concept, the identity of the subjects, and the eligibility/exclusion criteria.

The main difference between nutritional metabolomics and more traditional nutritional studies is the multivariate data output. This is considered to be an important advantage, because multivariate responses allow much more rigorous result validation and give a broader coverage of what is happening during a nutritional intervention. On the other hand, the fact that large metabolomic datasets, in general, consist of much more variables than observational trials makes proper statistical analyses challenging. For this reason, strong input from data analytics at the beginning of study planning is essential.

As with all highly sensitive methodologies, minimization of any type of unwanted variation is important. Unwanted variation can be broadly summarized as biological and technical (pre-analytical and analytical). Variation can arise from differences in sample collection, storage conditions, metabolite extraction, and instrument/stability. Given that biological variance is part of the outcome, failure to minimize unwanted technical variation can have a negative impact on the outcome of any study, thus resulting in the identification of fewer biomarkers or of false (positive or negative) biomarkers. This is especially important for biomarker candidates of low abundance [17, 18].

2.2. Study design and duration

Studies need to be sufficiently powered to produce meaningful measurement of specificity and sensitivity, and the use of cross-validation procedures assumes that test subjects need to be carefully selected. In this way, a properly planned study with collection of sufficient metadata (dietary, anthropometric, medication, etc.) should make it possible to avoid the need for a pilot study and to use metabolomics to its full potential.

The selection of the study design depends on the central scientific hypothesis, as well as on the availability of time, ethical issues, economical resources, subject compliance, and the possible role of confounding factors, such as seasonal variations. The study duration should be sufficiently long to allow the tracking of changes in the primary outcome measurements, supported by data from exploratory studies or pilot experiments. At the same time, it is also important to limit the study duration to prevent participant fatigue leading to non-compliance or withdrawal [17].

In principle, there are two main categories of study designs to assess research outcomes: randomized control intervention (clinical trials) and observation (e.g. cohort – either longitudinal or cross-sectional, case–control) studies. These two groups are distinguished by the intervention (providing of a particular nutrient, whole food, food group, whole diet, bioactive compound, and/or placebo). Observational studies help generating hypotheses and exploring associations between diet and health outcomes. These studies, however, cannot categorically prove cause-and-effect relationships [19-21]. Interventional study designs can vary from short-term studies, when the immediate effect of the intervention is investigated within few hours to as long as 24 or 48, hours postprandial (i.e. meal/food item consumed once), to long-term studies that evaluate the effects of the intervention over a period of weeks, months, or years [22-24]. The study setting can also vary, from those in free-

living populations to studies conducted in controlled facilities (i.e. hospitals, nursing homes). Nutritional interventions encompass the consumption of a wide range of specific products, including nutrients, food extracts, foods, dietary supplements, etc, which are given to participants at single occasions (e.g. at the beginning of acute postprandial studies) or delivered at home at regular intervals (e.g. every week in case of long term dietary interventions) with instructions on how to prepare and preserve them. The handling of these products (storage, extraction, analysis…) is, however, beyond the scope of this article and the reader is referred to reviews and books on this topic [25-29].

If no information on a defined output variable is available from literature, pilot studies (early exploratory studies) are sometimes necessary to generate essential data for a power calculation. These studies tend to be single-arm, with or without a control group. Pilot studies are a cost- and time-effective way of assessing the potential effects of a food/biomolecules and are considered as a forerunner to controlled studies. Moreover, they help identify challenges in implementing a full-scale trial to evaluate food matrix issues or to ascertain the dose or quantity of food to be consumed. Further, pilot studies form the basis of power calculations for subsequent controlled studies [30, 31].

Randomized controlled trials (RCTs) are considered the "gold standard" in nutritional studies. In the case of RCTs, the subjects are randomly assigned to treatment groups (allocation concealment). Additional quality steps in RCTs include the use of blinding, where the subject does not know whether he or she is receiving treatment, and double-blinding, where neither the subject nor the investigators know which group is receiving a respective treatment [9]. Blinding is often not possible in nutritional studies, in particular, and evidently, when foods or meals are tested. Two basic RCT study designs are commonly used: parallel group studies and crossover studies. In parallel group designs, each participant receives only one of the nutritional interventions (product A or B, active intervention vs control) studied. Comparisons between groups are done on a between-participant basis. Parallel studies are usually preferred for experiments that require longer-term intervention, because of their shorter overall time frame. Furthermore, parallel studies are essential when a washout period may be ineffective at returning outcome measures to the baseline or when returning to the baseline may be unethical (for example if body weight may be affected). In studies using cross-over designs, participants receive all the interventions (one or more) to be compared and the randomization specifies the order of testing. Between the different experimental phases a wash-out period is necessary. This has the advantage that comparisons between interventions can be made on a within-participant basis, with a consequent improvement in the precision of the

comparison and therefore the effectiveness of the study. With this design, confounding factors are kept to a minimum, because each subject acts as his or her own control (**Box 1**).

A crucial aspect for the experimentation is the metabolic status of the participants. Fasting level is defined to be 12 hours without any caloric intake, which affects primarily hepatic metabolism. When studying non-caloric food components after an overnight fast, one has to be aware that the interventions takes place in a catabolic situation. This is not advisable for different reasons (e.g. rare inborn errors of metabolism) as long as fasting is not the focus of the research question. Therefore, a defined background diet might be administered providing a sufficient number of calories to achieve a balanced metabolism. However, this background should not interfere with proposed food components.

The habitual diet of participants might interfere with the intervention trial. Therefore, several days of a run-in period are recommended. During this period, participants should be instructed to avoid certain foods relevant for biomarker identification. Providing a detailed "allowed" and "not allowed" food list is helpful to insure compliance of the study subjects with the study protocol. Frequently, in order to reduce intra- and inter-subject variability, standardized meals are provided to participants 24 h or 12 h before the intervention. Standardized meals provide a sufficient and similar number of calories to achieve a balanced metabolism among all participants. During the wash-out period in a crossover design, participants usually return to common dietary habits during a specific period of time awaiting for the second/third etc. part of the trial. This period of time, in turn, might again interfere with the incoming treatment and for this reason, several days of a run-in period are required [32-36].

2.2.1.   *Inter- and intra-individual variability*

Humans are extremely diverse and a number of human studies have found that metabolomic results are strongly influenced by inter- and intra-individual variation [36-38]. Biological variability arises from many factors, for example sex, age, genetic background, circadian rhythm, seasonal differences, menstrual cycle, stress, and gut microbiota. These biological factors may introduce systematic bias [32, 33, 36]. Inter- and intra-individual variations are considerable for each biofluid. Experiments from Rasmussen et al. [34, 38] with urine samples have shown that diet standardization during a three-day period reduced inter- and intra-individual variations and that effects of diet culture and cohabitation were significantly attenuated following diet standardization. On the other hand, Welch et al. [35] showed that a one-day standard diet followed for 24 h before biofluid collection appeared to reduce variation in first void urine but not in fasting plasma samples. Therefore, in the study of human nutrition,

it is important to control and understand the factors that contribute to normal physiological variation. This is done so that normal metabolic fluctuations are not confused with biomarkers representing a metabolic change due to nutritional intervention. In particular, the experimental design should consider the possible presence of subgroups of individuals responding differently to the same nutritional intervention, i.e. having a different metabotype [39]. This consideration directly impacts on the number of participants enrolled in nutritional studies as a higher number of biological replicates is then required. Therefore it is important to understand the factors underlying this variability and adapt the study design accordingly.

2.3. Sample size

As with any good study design, an estimation of the number of participants (sample size, biological replicates) is essential. The usual methods for sample size estimation require specification of the magnitude of the smallest meaningful difference in the outcome variable. The study must be sufficiently large to have acceptable power to detect this difference as statistically significant and must take into account possible non-compliance and the anticipated drop-out rate. Except of that, the number of participants must also consider intra- and inter-individual variations. Higher numbers of biological replicates thus allow for a better characterization of the studied population. However, in metabolomic studies, it is often not known which variables will change and the classical approaches to sample size cannot normally be applied [40-43].

Nyamundanda et al. [40] proposed a method known as MetSizeR for sample size estimation in metabolomic experiments, considering factors such as FDR (False Discovery Rate) and the availability of pilot studies. The main advantages of this approach include the ability to determine sample size, even when experimental pilot data are not available, and technique specificity (NMR and MS), which can improve the power of the study. More recently, Blaise and colleagues [42] developed a method for performing power calculations for metabolic phenotyping. The method is composed of 3 steps: i) modelling of the distribution of pilot data; ii) introducing an artificial effect; and iii) estimating confidence intervals for performance metrics. Furthermore, a recent update for MetaboAnalyst, a comprehensive web-based tool designed to perform metabolomic data analysis, includes a power analysis module based on an approach originally used for gene expression and requires pilot data [43]. Using either of these approaches can help researchers estimate sample size for metabolomics data.

2.4.   Data collection and sampling methodology

As one would expect from international best practices, studies should follow the CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting and should accurately define recruitment, inclusion and exclusion criteria [44, 45]. Clear guidelines on how the samples are labelled should be established from the outset [46]. From a nutritional metabolomic viewpoint, it is advisable to collect background dietary information on participants in order to characterize their habitual diet in terms of foods and diet composition. Furthermore, detailed characteristics of the population are important to collect but should not be limited to age, sex, BMI, anthropometric data, level of medication use, physical activity, and smoking habits (**Table 2**). Sample management is a part of process control, and one of the essentials of a quality management system. The quality of the work an analytical laboratory produces is only as good as the quality of the samples it uses for testing. Also, particular care should be given to the labelling of samples. The laboratory must be proactive in ensuring that the samples received meet all of the requirements needed to produce accurate test results. In fact, sample handling, such as collection, labelling, processing, aliquoting, storage, and transportation, may significantly affect the results of the study. For example, if test samples are handled differently from control samples, biased misclassification may occur.

The main information that has to be associated with each sample can be divided into two categories: the information regarding the participant and the specific sample information. This information should be reported in a database that each lab should keep. Each sample, especially in intervention studies, should also report a computer-printed label on the container with the relevant information. Although this seems obvious, there is a large percentage of laboratories employing handwriting. Handwritten labels are difficult to read and are unreliable. The ink can smear and the information can be easily misread, leading to repetitive errors. Some laboratories are now employing barcodes. This labelling method certainly avoids any possible smearing or misreading problem, though it requires special equipment and software.

Blood samples are generally collected after overnight fasting (10-12 hours) for biochemical and metabolome analysis, while in the case of postprandial studies, blood samples are taken according to a specific design (for example, at 0 min, 30 min, 60 min, 120 min, 180 min, 240 min, and 300 min). It is imperative that all blood samples are collected in a similar fashion and according to the correct procedures (**Boxes 2-4**). Urine samples are usually collected before blood donation. Depending on the study design and the research hypothesis first void, spot, or 24 h urines can be collected. Moreover, pooled urine samples can also be collected parallel to

postprandial blood samples, either by spot sampling at the time of the blood sampling or by pooling the urine samples between consecutive time points. For urine collection it is essential to pay attention to the chilling procedure and strict protocols should be adhered to. Fecal samples, as a result of human physiology, can be collected less often than urine or blood. If fecal samples are included for analysis by the study design, their collection may occur once per day, usually at the beginning of the study, to reflect the microbiome and metabolome status before intervention, and at the end of study, to observe eventual changes after intervention [47]. The protocols and recommendations for samples collections and preparation are discussed in section 3.2 as well as in **Appendix 4**. The hypothesis and objectives may differ substantially from one research work to another, and for this reason the sampling time is set according to the aim of the research. Samples can be collected as: i) random samples collected at any time; ii) timed samples used to study time-related trends in nutritional studies, e.g. fasting samples and consecutive time intervals (every half hour or every hour); iii) samples collected over 24 h to understand the overall status of the individual. At the sample collection stage due attention to the number and size of aliquots is necessary in order to avoid unwanted freeze-thaw cycles (see section 3.5 for more details).

3.   Sample characteristics and pre-analytical processing

3.1.   Type of biological samples

Blood plasma, serum, urine, feces, saliva, muscle, liver, sweat, exhaled breath, or gastrointestinal fluid all reflect different responses to dietary intakes and, thus, can potentially be used for metabolome exploration. The rationale for selecting a particular biofluid for further analysis will be essential for appropriate physiological and biological interpretation of the observed metabolome. Whereas the selected fluids and tissues report up to a certain point on overlapping physiological phenomena, each of them provide some specific insights: plasma and serum on the bioavailability of nutrients, saliva as well as gastrointestinal fluids on food digestion, feces on the interaction of the gut microbiome with food, and urine on clearance effects. Plasma, serum, urine, and feces are frequently used for metabolomic profiling in nutrition, while other fluids and tissues are not yet well explored. The salivary metabolome has been measured to investigate the effect of standardizing diet on reducing intra- and interindividual variability in nutritional studies [48] as well as to identify biomarkers of compliance to the Mediterranean diet [49], whereas breath was used to investigate biomarkers of intake of garlic [50, 51].

There is clear evidence that a single biological fluid is not sufficient to decipher the impact of foods or diets on human physiology and health. Consequently, an increasing number of investigations use multi-compartment metabolomic analyses. Chapter 3 will focus on the most popular biological fluids, namely plasma, serum, urine, and feces.

3.1.1. Serum/ Plasma

Blood-derived serum and plasma are the most common matrices used in human metabolomic studies, both because they are relatively easy to collect and because the blood metabolome reflects individual changes in metabolism. A decision on the choice of the blood sample (plasma or serum) to be collected during the entire duration of the study should be taken at an early stage of the study design. The essential difference between serum and plasma is that serum is collected after a process of clotting, while plasma is collected without clotting. Serum and plasma are separated from macroparticles in whole blood, including coagulated material and cells, by centrifugation. Serum is less viscous than plasma due to the lack of fibrinogen, prothrombin, and other clotting proteins. However, serum is more prone to ex-vivo protein degradation, which generates large numbers of small peptides that can significantly complicate the task of extracting meaningful information [52]. General comparisons of serum vs. plasma metabolomes, in terms of reproducibility, discriminative ability, and coverage indicate that they offer similar analytical opportunities [52-54] (**Box 5**). However, metabolite concentrations are generally higher in serum, yet still highly correlated between the two matrices. Furthermore, serum reveals more potential biomarkers than plasma when comparing different phenotypes [55-57]. Therefore, a pragmatic solution must be adopted according to practical considerations.

3.1.2. Urine

Urine is an easy-to-access biological fluid composed of endogenous and exogenous metabolites. Urine contains over 95% water. Major urinary solutes are sodium, ammonia, phosphate, sulphate, urea, and creatinine. Furthermore, urine contains a large palette of compounds derived from metabolism in tissues and by microbiota as well as some proteins (globulins). The number of conjugated metabolites in urine is higher than in human blood or serum [58-60], because urine reflects the human metabolism at the end of the ADME (absorption, distribution, metabolism, excretion) nutrikinetic process. The main advantages of urine compared to other biological fluids are that large quantities can be obtained using non-invasive procedures under normal circumstances (self-collected) and sample preparation is easier due to lower protein content. A disadvantage of urine, compared to blood, is that urine volume and thus the overall concentration of urine metabolites may vary

by up to a factor of 20 [58-60]. This necessitates sample-specific normalization, either to urine volume or osmolarity. Furthermore, urine samples can contain cell debris, crystals, and other particles, which may form sediments [61, 62]. Fresh urine samples should, therefore, be gently centrifuged to remove solid particles (see **Boxes 2-3**). Nevertheless, despite this pre-treatment, thawed urine samples often still show substantial sediments, which only dissolve if the samples are gently heated or diluted. Sediment particles may impair instrumental analysis. On the other hand, urine sediments may contain a number of relevant metabolites, such as uric acid, cysteine, oxalate, lipid species, and probably many others. Depending on the focus of the study and the intended analysis, it must be decided whether the sediment of thawed urine samples should be removed by strong centrifugation or whether measures should be taken to release the metabolites contained in or adsorbed to the sediment particles.

3.1.3.   Feces

Fecal material is a complex biological matrix with a diverse metabolic composition, containing both endogenous and exogenous metabolites with variant polarity. These metabolites include both nonpolar lipids (e.g. fatty acids, triglycerides, and phosphoglycerolipids) as well as polar compounds (e.g. short chain fatty acids (SCFAs), amino acids, bile acids, and carbohydrates). In addition, feces contain both microbial and mammalian cells with numerous enzymes and biological processes continuing during sample collection, storage, and transport. Fecal samples provide direct information about interactions between host and gut microbiota since they carry numerous biochemical compounds derived from the host, the host's microbiota, and food residuals. With straight links to the dietary pattern and state of health (i.e. gastrointestinal pathologies), feces are an interesting biological matrix for studying metabolic alterations and gaining nutritional-clinical insight [63]. Hence, fecal metabolomics offers significant potential for investigating the impact of the environment (exposome) on the regulation of host metabolism and for understanding the underlying mechanisms controlling the colonic-systemic axis in the metabolism of various food components. An advantage of the fecal sample is the fact that it can be obtained non-invasively in relative large quantities (see section 3.2.3 and **Boxes 2 and 4**).

3.2.   Sample collection and pre-storage sample preparation

The pre-analytical phase, comprising sample collection, post-collection sample processing, and storage, is a major source of variability in metabolomics [64-67]. This is particularly relevant in large-scale multicenter studies or when using samples coming from biobanks. Therefore, the implementation of standard operating procedures (SOPs) throughout the entire metabolomic pipeline in order to ensure good reproducibility and repeatability is

mandatory. In this sense, some general considerations, regardless of the type of sample investigated, should include the choice of the correct sample-collection container and the metabolic quenching protocol. Sampling devices usually contain additives that may generate confounding peaks in metabolomic profiles and cause ion suppression when MS is used as detector. Thus, it is recommended that sample tubes/containers from a single manufacturing batch be used for a complete study, whenever possible. Moreover, sample tubes/containers should be checked before initiating the study to ensure that no chemicals are present or leaking from the tubes into the samples, as this may affect the subsequent analysis. Blood collection tubes can introduce numerous exogenous interfering compounds into blood samples, mainly when anticoagulants are employed for obtaining plasma samples. Therefore collection kits must be carefully selected in order to minimize matrix effects. Urine samples are usually collected using bare polypropylene containers of the required volume.

The application of a metabolic quenching protocol after sample collection is crucial to inactivate enzymatic reactions which is necessary to obtain an accurate picture of the metabolome at the time of sampling. It has been demonstrated that residual enzymatic activities and chemical reactions induced by air oxidation and light can induce significant metabolic alterations during sample processing, including depletion of easily oxidizable species [68], conversion of labile metabolites (e.g. energy-related metabolites, nucleotides) [69, 70], and increased lipid hydrolysis [71], among others. The metabolic quenching of biofluids is usually achieved by snap-freezing in liquid nitrogen, since many metabolites have a very short half-life. The use of dry ice for this purpose is not recommended, since the solubilization of carbon dioxide may cause non-reproducible changes in the sample pH, which has important consequences on subsequent chemical analysis [72].

3.2.1.   Serum/ plasma

Plasma is obtained from non-clotted whole blood by separating all blood cells from the liquid part by centrifugation, for which several anticoagulants can be used as described below. On the contrary, serum is obtained from clotted whole blood by sedimenting the clot containing blood cells and clotting proteins. Independently of the sample type, it is noteworthy that a consistent blood sampling site must be used throughout the whole study, since venous and arterial blood have shown different metabolic compositions [73] (see **Box 5**).

There are two systems for blood collection: vacuum (e.g. Vacutainer® tubes) and aspiration (e.g. Monovettes®). In many metabolomic studies, Vacutainer® tubes are used due to the safe and standardized way of blood collection. The vacuum in the tubes ensures a constant volume of blood collected that cannot be influenced. Other researchers prefer the use of Monovettes®, with which an experienced phlebotomist can

control the vacuum to adapt to the veins of the volunteer and reduce the risk of hemolysis (i.e. lysis of red blood cells). In any case, the blood collection tubes should not be shaken whilst handling and the time of collection should be recorded. Each tube should be filled up as much as possible (to ensure a reproducible concentration of anticoagulant in the samples) and inverted gently at least 8 times after collection. Plasma tubes should be placed horizontally *on* ice (not *in* ice) to prevent red cell lysis and to reduce protease activity, whereas serum tubes should be incubated for 30 minutes at room temperature (RT) to allow for clotting [72]. If other tubes are collected, for example to characterize the subject's nutritional status (measurement of serum glucose, lipids, etc.), the tubes intended for metabolomic measurement should be collected first, to prevent potentially lysed cells in the needle from contaminating the sample. Plasma samples can immediately be processed after blood collection, but serum requires a previous clotting step at RT that can introduce great metabolic variability. The time needed for clot formation (30 min) facilitates the enzymatic conversion, degradation, and loss of several metabolites, as well as the release of numerous compounds into serum from activated platelets, as previously reported [55-57]. Therefore, the implementation of SOPs is particularly important when serum samples are collected, paying particular attention to the use of the same clotting time for all samples along the study. The delay time and storage temperature between blood collection and centrifugation also have a large impact on metabolomic profiles. Boyanton et al. [74] analyzed the stability of 24 compounds in plasma and serum stored at RT in contact with cells and after centrifugation. The authors found significant changes in the metabolome when cells were not removed. Similarly, Kamlage et al. [75] demonstrated that a high number of metabolites are significantly altered as a consequence of prolonged processing times at different temperatures, including signaling metabolites, lipids, carbohydrates, and amino acids. However, the most important changes can be associated with energy-related metabolites (i.e. increased lactate, decreased glucose, and pyruvate), which could be attributable to the anaerobic metabolism of erythrocytes [76-78]. Thus, as a general recommendation, blood processing times should be reduced to the minimum (preferably no longer than 2 h) and samples should be kept cool in the meantime.

Although manufacturers recommend centrifugation time and speed, still different centrifugation protocols alter the biochemistry of the samples through different amount of platelets remaining in the plasma. Indeed, significant differences in platelet count, together with alterations in NMR and HRMS metabolomes, related to the different centrifugation protocols were recently found [79-81]. Yazigi Junior et al. [79] reported that the greatest platelet enrichment was obtained after two rounds of centrifugation at $400 \times g$ and 10 min. In addition to the

number of centrifugation cycles, the centrifugal force significantly influenced platelet enrichment, with a centrifugation protocol at 1500 ×g for 10 min producing a higher platelet enrichment than at 3000 × g for 5 min. As a protocol at 3000 × g for 5 minutes resulted in lower platelet counts, the authors assumed that the centrifugal force recompenses the effect of the centrifugation time. Moreover, in the NMR analyses, the plasma content of free glutamine, as well as of different lipid classes (i.e. glycerophosphocholines and sphingomyelins) in untargeted UPLC-QTOF MS analyses, was associated with the centrifugation conditions [81].

The above findings highlight the importance of well controlled and constant conditions for the preparation of plasma. Since there are no harmonized recommendations regarding centrifugation conditions for untargeted metabolomics, the choice of spinning conditions should balance between gentle cell handling and shorter turnaround time. Centrifugation conditions vary between laboratories, and this factor becomes critical in multicenter studies or long-term sample storage in a biobanks. Concluding, for plasma immediately after collection, for serum after clotting, blood collection tubes should be centrifuged according to the manufacturer's instructions (e.g. for Vacutainer® tubes at 1,200 x g for 10 minutes at 4°C) and consistent for entire study. Moreover the time between whole blood collection and centrifugation, combined with the temperature and sunlight exposure, should be carefully considered as these are also important pre-analytical factors. Prolonged exposure to room temperature and sunlight should be avoided as whole blood in such conditions is metabolically active, what may affect several metabolites including lactate and arginine/ornithine or ascorbic acid [82-85]. In this context, a new factor to evaluate plasma quality was established, the LacaScore, which is based on the ascorbic acid/lactate ratio in plasma [86].

The anticoagulant used for blood plasma preparation may also affect the resulting metabolic profile, due to the different mechanisms involved in reducing coagulation by various agents (e.g. heparin, EDTA, citrate). For instance, the presence of cations can cause problems in metabolomic and lipidomic analysis by binding to negatively charged phospholipids thereby causing ion enhancement [87]. In this line, it has been observed that lithium ions from heparin may increase the ionization efficiency of many metabolites including phospholipids and triglycerides, but lithium also exacerbates matrix effects by increasing the signals of plastic polymers [66, 88]. Different anticoagulants can also have an effect on various aspects of sample preparation, such as the extraction procedure or the derivatization process (for GC-MS metabolomics). Moreover, if the anticoagulant agent is present in the final extract, it can also affect the final MS analysis by forming sodium and potassium formate clusters, leading to ion suppression or enhancement. As shown in the work of Jørgenrud et al. [89], who analyzed

plasma and serum using an untargeted approach, different types of anticoagulants exerted an effect on levels of amino acids, carboxylic acids, and sugar alcohols. EDTA was poorly suited for the analysis of polar metabolites and sodium citrate caused problems in determining citric acid and its derivatives, while serum proved to be the best option for the metabolic profiling of polar compounds. A couple of years earlier, Barri and Dragsted [56] found that differences between plasma preparation types were primarily related to ion suppression or enhancement caused by citrate and EDTA anticoagulants. Mass spectral characteristics of sodium formate and potassium formate ion clusters were detected in citrate and EDTA plasma samples respectively. These originate from formate in the mobile phase and Na^+ (in Na-citrate tubes) and K^+ (in K-EDTA tubes). It was also reported that the anticoagulant counter cation (Na^+ or K^+) can make some metabolites more dominant in electrospray ionization (ESI)-MS. Moreover, polymeric material residues originating from blood collection tubes for serum preparation were only observed in serum samples [56]. Hence, the choice of the best anticoagulant in metabolomic research is still a subject of debate.

Finally important pre-analytical issue to consider during blood collection and processing is the hemolysis of blood samples. This results in the contamination of plasma and serum samples through the release of hemoglobin and other intracellular components from erythrocytes. Common causes of hemolysis include long application of the tourniquet, strong aspiration, vigorous shaking of the blood collection tube, excessive centrifugation speed, and exposure to inappropriate temperature. Thus, special care must be taken during drawing and handling of blood samples. The breakdown of blood cells strongly alters metabolic profiles of blood-derived samples by increasing the concentrations of numerous metabolites coming from the intracellular space as well as by inducing the degradation of some compounds by the action of released enzymes [66, 75, 90]. For this reason, the use of hemolyzed samples should be avoided in metabolomic studies.

3.2.2. *Urine*

Urine samples should be collected in containers that do not to release plasticizers or other compounds into the sample (see **Box 2**). Generally, for metabolomic analyses no preservative should be added to the container, but urines should be kept cool (<10°C) until processing.

Different sampling modes can be used to collect urine samples depending on the purpose of the experiment, including spot sampling, timed sampling at pre-set intervals, or collection of 24-h urine [73]. The sampling time, however, may affect sample collection. Spot urine samples can be directly collected in an appropriate container at any time of day and put on ice or in the fridge until processing. For this purpose, the collection of mid-stream

urine is normally preferred in order to minimize the presence of contaminants (e.g. bacteria, particles). On the other hand, timed sampling can be used to investigate the excretion pattern of metabolites during intervention studies or to monitor metabolic oscillations associated to the circadian rhythm. In that case, if subjects need to urinate before the time point for the next collection, they should collect this urine in a jug and transfer it to a 2.5 L urine container that must be kept on ice or refrigerated until urine collection at the subsequent collection time point. These urine samples should then be pooled as one sample before processing. Finally, the use of 24-hour urine allows one to obtain an overall picture of metabolic excretion, by eliminating the large variability observed over shorter sampling times. However, the collection of 24-hour samples is cumbersome, since all urines from the day must be collected in jugs and transferred to 2.5 L urine containers, which must be kept cool by placing them on ice, in a refrigerator or in a cool bag. After 24-hour collection is completed, the volume of the whole urine samples should be measured and recorded.

The addition of preservatives offers another route to enhance the stability of samples such as urine, which are prone to bacterial contamination during collection and storage. In practice, the most common preservatives are boric acid and sodium azide, but their use in metabolomics research is under debate. In this context, it has been demonstrated that borate causes slight alterations in NMR urinary metabolomics profiles as a result of the formation of complexes with hydroxyl and carboxylate groups from some metabolites (e.g. mannitol, citrate, and hydroxyisobutyrate) [91]. However, these changes were negligible in comparison with inter-individual variations and the authors of this study proposed that this method of preservation is fit for purpose in metabolomic studies. Saude and Sykes [92] proposed that sodium azide should be added to all urine samples in order to stabilize metabolomic profiles acquired by NMR. However, other authors note that the use of this preservative is not mandatory if urine samples are correctly stored [93, 94]. In this line, Bernini et al. [76] observed that the combined application of a mild pre-centrifugation step and filtration shows higher performance for bacterial removal than the use of azide. Accordingly, the general recommendation of the European Consensus Expert Group is to avoid the use of additives to urine [95].

For processing of all types of urine samples, a centrifugation or filtration step is recommended in order to remove sediments: primary (cell debris, bacteria, small molecules) and secondary (artefacts created due to freezing and thawing), as well as to eliminate other non-cellular components and materials in suspension [62]. In this regard, Saude and Sykes [92] found that filtration is effective in conserving the urinary metabolome, although caution must be taken due to possible losses of larger metabolites. More recently, it has been demonstrated that

the application of a mild pre-centrifugation step combined with filtration is the best protocol to avoid urine contamination with compounds derived from cellular components [76]. However, filtration can lead to adsorptive losses of some metabolites; so the most common procedure for urine processing is simple centrifugation. For this, the required volume of urine should be transferred into appropriate centrifuge tubes and centrifuged at 1,800 x g for 10 min at 4°C. The supernatant can then be decanted into the required number of pre-labelled Micro-plastic tubes or Falcon® tubes and the rest of the sample discarded, storing all aliquots immediately at -80°C.

3.2.3. Feces

Fecal material a relatively new matrix for metabolomic profiling [96, 97]. As a non-invasive matrix, feces can be self-collected by volunteers using collection kits that consist of a sterilized wide-mouth plastic bag and a plastic container (i.e. Exakt Pak ® canister), comfortable special stool specimen collection units (i.e. Fecotainer®, Suesse, Labortechnik, Gudensberg, Germany, Fisherbrand™ Commode Specimen Collection System, Diagnolab) or even a freezing toilet operating at -30 °C (see **Boxes 2 and 4**). Participants must be carefully instructed to avoid stool contamination with water, urine, or other materials (e.g. toilet paper). The samples are collected directly inside sterilized plastic bags or a Fecotainer®, sealed with a clip or cover, and placed immediately in sealed insulated containers. Frequently stool samples are produced in participants homes, thus intermediate storage in a domestic freezer is a practical solution prior to pick-up from participants. Even in cases where a common household freezer is available, in studies where human participants are asked to self-sample and ship to a laboratory, samples may experience temperature fluctuations during storage (as home freezers undergo automatic defrost cycles; typical temperature ranges vary from -20°C to -2°C) or during shipping, allowing samples to thaw. In regions with extreme temperatures or in cases of freezer failure, samples may additionally be exposed to extreme heat. However, it is unknown how a domestic freezer and temperature fluctuations affect the accuracy of further metabolomic analysis. As a general rule, containers should be delivered to the laboratory as soon as possible for further storage. The elimination of microbial activity in fecal samples prior to metabolomic analysis is crucial for achieving good metabolomic coverage and warranted by the storage of fecal samples at -80°C and lyophilization prior to extraction. For example, SCFAs may undergo bacterial fermentation and the results may be drastically affected by time and temperature fluctuations. For this reason, the adherence to well defined pre-analytical protocols is definitely required for this particular matrix.

Fecal samples reflect diet-gut microbiota-host interactions covering a time span from several hours to few weeks. Sometimes, it can be difficult to collect fecal samples together with urine samples (i.e. 24-h) or plasma samples (i.e. fasting blood) due to the physiological characteristics of participants (sever/mild constipation, fecal impaction, mild diarrhea). The volume of fecal samples may vary from few grams up to several hundreds grams. Likewise, the state and water content can show large variation, in particular with transit time (seven stool types according to the Bristol Stool Chart). Moreover, Gratton et al. [98] and Santiago et al. [99] found significant differences in metabolite profiles and microbial composition between the top, middle, bottom, and edge positions of stool samples. For this reason the homogenization of the sample and selection of a representative part of the stool sample are important factors prior to aliquoting and freezing [98, 100, 101].

Several methods for fecal pre-preparation exist for storage, including: i) simple freezing; ii) direct centrifugation supernatant collection–fecal water storage; (iii) extraction and centrifugation and subsequent supernatant collection; or (iv) freeze-drying [96, 97, 102, 103].

The easiest method for fecal storage is homogenization of the fresh stool sample by stirring with a sterile spatula directly in the delivery bag [104-106] and then aliquoting a few milligrams to a few grams into feces tubes (i.e. Sartstedt) and storing at -80°C [97, 105, 107, 108]. Le Galle et al. [108] described a method that involved collecting multiple fecal aliquots of 20 mg from the same area below the surface of the stool, thereby improving representativeness of the sample. For high volumes, where homogenizing manually by spatula is difficult, a stomacher blender can be used that automatically homogenizes samples in the delivery bag for a certain amount of time (i.e. 2 min).

For the extraction of polar metabolites, feces can be transformed into fecal water and then stored at -80°C. Fecal samples should be homogenized with a buffer for few minutes and then centrifuged for 2 h at 35,000 x g at 4°C. The supernatant is usually decanted, sterile filtered (Millipore, 0.8/0.2 mm), aliquoted in 1-2 mL portions, and stored at -80°C until analysis. A very important aspect in the preparation of fecal water is the selection of an appropriate feces volume over weight ratio. It is strongly recommended to select relatively large quantities of stool (e.g. 15 g) for fecal water preparation, in order to compensate for the eventual inefficiency of fecal homogenization due to the significant variation of the types of human fecal matter.

Pre-storage treatment and homogenization of fecal material is frequently done directly in ice-cold phosphate-buffered saline (PBS) by mixing representative volumes of feces with 2 volumes of PBS buffer [96, 97, 102]. It has been shown that the addition of 95% ethanol, water/methanol, or twice distilled water tends to improve the

overall recovery of fecal metabolites [96, 97, 105, 107, 109, 110]. Such prepared fecal slurry (fecal water) should be vortexed and ultracentrifuged at 35,000 x g for 2 h at 4°C and the supernatant filtered, aliquoted, and stored at -80°C. Ultracentrifugation of feces and fecal slurries is crucial for sedimentating particles.

As an alternative, fresh feces can be transformed in fecal powder. Gorzelak et al. [100] successfully applied this technique starting from ground freezing (-80°C) in liquid nitrogen of entire stool samples using a mortar and pestle until the sample was a fine powder. From the powder a collection of subsamples were prepared and stored at -80°C [100]. Similarly, Vanden Bussche et al. [63] and Van Meulebroek et al. [111] proposed protocols for untargeted polar metabolomic and lipidomic analysis of feces, where fecal samples were frozen at -80°C (for one or two days) and lyophilized, prior to aliquoting and final storage at -80°C. Freeze drying in sample preparation may remove the water from the sample, thus minimizing the effect of confounding factors for NMR spectroscopy [112]. However, freeze drying may also potentially decrease the content of volatile compounds such as SCFAs [102], which may be of interest as potential biomarkers [109]. In this context, proper execution of the freeze drying process and the absence of any defrosting is crucial as freeze drying was shown to deliver the richest fecal metabolomes (over 9,000 compounds) in recent untargeted studies [63, 111].

3.3.   Aliquoting

Aliquots should be prepared at the moment of sampling, when serum/plasma and urine (see **Box 3**) or feces (see **Box 4**) are obtained from volunteers, in order to minimize subsequent freeze-thaw cycles, which affect sample quality. As previously mentioned, all samples collected throughout a metabolomic study must be aliquoted in tubes from the same manufacturer to avoid any analytical bias. Aliquots should be prepared according to the study design and number and types of tests to be performed. It is strongly recommended to have at least three aliquots dedicated to each test (i.e. three aliquots for an LC-MS assay, three aliquots for a GC-MS assay, etc.) (**Box 6**). The generation of a number of extra aliquots is highly recommended in order to allow for eventual additional analyses (e.g. identification of newly discover biomarkers) that may turn out to be necessary or interesting at later stages. The volume of one aliquot should fit the volume of biological fluid used for extraction or analysis. In particular, for LC-MS assays involving between 50 and 100 μL of biofluid for extraction, the aliquot should be 150-300 μl. This allows for precise pipetting and QC preparation. Storage containers should be chosen according to the chosen storage temperature, volume of aliquots, and sample processing procedures. For example, not all sample tubes are suited for storage in liquid nitrogen. Cryovials, however, have different

sizes than standard microvials and cannot be centrifuged in the same rotors. In any case, it is important to ensure that the same tubes are used throughout the study or project and that the tubes do not leak substances that can interfere with metabolomic analyses.

In preparing aliquots we suggest filling storage vials at ca 80-90 % of their volume, in order to avoid air storage and the prompt conservation of the aliquots at the lowest possible temperature (e.g. -80 ºC) (**Boxes 7 and 8**). Alternatively, prompt freezing at -20°C and transferring samples to lower temperatures for long-term storage within one week does not negatively affect sample integrity. Of note, the CO_2 sublimating from the dry ice can acidify samples, which is highly undesirable in NMR-based metabolomics as this technology is sensitive to shifts in pH [72]. In situations where samples are not aliquoted immediately after collection or in situations where samples need to be thawed in order to divide already frozen samples into smaller aliquots for several different analyses/procedures, or for pooled QC sample preparation, thawing samples on ice, followed by prompt redistribution of the larger sample into several smaller aliquots and prompt refreezing until analysis, should minimize any variability that could be introduced by additional handling. Recording this deviation from the SOP is required as such information may later allow one to account for unexpected variance in the dataset.

3.4.   Storage

In general, urine, plasma/serum and fecal samples should be transferred into cryovials immediately after collection and kept at -80ºC in an ultra-cold freezer or in liquid nitrogen tanks until analysis with continuous monitoring of temperature and storage data. In case of 24-hour urine or fecal samples collected at the volunteer's home, when the sample collection is not immediate as in case of plasma/serum, samples in container should be stored in volunteer's refrigerator at 4°C in order to lower enzymatic and bacterial activities. Once the collection of 24-hour or fecal samples is accomplished, the samples should be transported to the laboratory, preferably on ice, or in a cooling bag [98]. In the laboratory, samples should be prepared according to their destination, and stored at -80ºC (see **Boxes 7 and 8**).

For long-term stability of biological fluids for untargeted metabolomics, freeze-thaw cycles need to be avoided. However, in many cases the availability of -80ºC freezers is limited in clinical settings, so temporary freezing at -20ºC is often necessary before the shipment of samples to biobanks. For this reason, numerous studies have been conducted over the past few years with the aim of evaluating the stability of the entire

metabolome under different storage conditions. Some authors have demonstrated that the storage of urine samples in the fridge (4ºC) or by using cool packs (10ºC) is possible for 24-72 hours without significant degradation of urinary components [94, 113-115]. It has also been shown that long-term storage of urine at either -20°C or -80°C, for up to 6 months, usually yields comparable metabolomic profiles [93, 113, 115]. On the other hand, Živković et al. [116] found in a recent study that levels of volatile metabolites show a time dependent decrease over time if urine samples are not deep-frozen, which is of great importance in GC-MS-based metabolomics. In any case, it should be noted that even urine stored at -80 ºC can suffer slight metabolic alterations, especially during the first days of storage. As a result, some authors recommend a 1-week minimum period of storage to ensure a consistent degradation of urinary components in all samples across the study cohort in order to facilitate their comparison [92]. Regarding plasma and serum, there is a clear consensus about the metabolome's instability in non-frozen samples because of the high concentration of enzymes. Clark et al. [117, 118] reported changes in fat-soluble vitamins and various plasma analytes (albumin, apolipoproteins A1 and B, cholesterol, HDL, total protein, triglycerides, alanine transaminase, creatine kinase, creatinine and γ-glutamyl transferase) when whole blood was stored over several days at room temperature (RT). Similarly, a recent NMR-based metabolomics study revealed a large impact on lipoproteins and choline compounds in plasma samples stored at RT for more than 2.5 hours [119], while sample refrigeration at 4ºC only preserves metabolite profiles for up to 24 hours [75, 120, 121]. Anton et al. [121] also examined the effects of different storage conditions (RT, dry ice, wet ice, for 12 h, 24 h and 36 h) on the concentrations of 127 metabolites in serum. These authors found a clear signature of degradation in samples kept at RT (lysophosphatidylcholines, phosphatidylcholines, decadienylcarnitine, arginine, glycine, ornithine, phenylalanine, serine, leucine, isoleucine, ornithine, phenylalanine, serine), while storage on wet ice led to less pronounced concentration changes. Therefore, the most common procedure for long-term storage of blood-derived samples and urine is deep-freezing at -80ºC [119, 120, 122] although some authors have described that samples stored at -20 ºC can also be stable for short periods (i.e. a few weeks or months) [119, 123, 124]. Finally, some studies investigated the stability of plasma samples at -80ºC over multiple years (5 to 17 years), concluding that some metabolites, especially amino acids, acylcarnitines, glycerophospholipids, sphingomyelins, vitamins, and hexoses suffer from long term storage, with changes in concentration varying between +13.7% and -14.5% [125, 126], while analysis of plasma samples after 14 to 17 years showed no significant differences using an untargeted Ultra Performance Liquid Chromatography (UPLC)-TOF approach [127].

3.5. Thawing

Before analysis, samples need to be thawed. Plasma or serum samples are thawed by many researchers at 4°C overnight (see **Box 8**). This often yields lumps of heat shock proteins. Thawing blood samples at RT reduces the formation of lumps/clots but increases the risk of enzymatic reactions or other changes to the samples. Feces and fecal water should be thawed in controlled and gradual manner to avoid extensive exposure to RT. Gradual thawing in a fridge overnight or by placing samples on ice reduces fermentative processes and enzymatic activities. The choice of the thawing method in this case depends on size and quantity of aliquots. A few grams of feces requires longer thawing time than, for instance, 500 µl of fecal water, which can be defrosted within 30 minutes. Additionally, it is important to note that the metabolic composition of feces suffers more from changes in temperature than fecal water [98].

Not all metabolites are equally sensitive to pre-analytical handling conditions. Therefore sample stability during thawing must be ensured. Several authors have assessed the impact of freeze-thaw cycles on the quantity of different metabolites in plasma and serum samples, including lipids and some amino acids. For example, Anton and co-authors [121] performed a series of experiments in which several freeze-thaw cycles were followed by a targeted metabolomic approach (180 metabolites) to assess its impact on the serum metabolome. Except for minor increases in glycine, methionine, tryptophan, phenylalanine, and tyrosine, no major changes were found in concentrations of the 127 metabolites in serum after four freeze-thaw cycles. The slight increase in phenylalanine and other amino acids may indicate some protein degradation, which occurs during thawing and refreezing and reflects some ongoing metabolism. In this context, Wood et al. [128] found two fatty acids, namely docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA), are stable for up to three freeze-thaw cycles, while arachidonic acid increased significantly at the third freeze-thaw cycle. Zivkovic and co-authors [116] also studied changes in serum lipid composition, with up to three freeze-thaw cycles. Minor alterations were observed following three cycles (<1% of monitored metabolites), which could be caused by enzymatic hydrolysis and synthesis, enzymatic transfer of lipids between lipoproteins, and non-enzymatic oxidation. Similarly, other authors also found that no more than 3-4 freeze-thaw cycles are tolerated, as beyond this number significant alterations can be introduced in metabolic profiles [119, 129]. Interestingly, the effect of thawing is strongly sample-dependent, with a larger impact on lipid-rich samples [119]. In particular, Fliniaux et al. [18] observed a visible impact of 5 or 10 freeze-thaw cycles on the metabolic profiles of serum samples,

measured using an untargeted proton NMR spectroscopy approach. By contrast, Cuhadar et al. [123] reported good stability of 17 routine chemistry analytes in serum for up to 10 freeze-thaw cycles (aspartate aminotransferase, alanine aminotransferase, creatine kinase, gamma-glutamyl transferase, direct bilirubin, glucose, creatinine, cholesterol, triglycerides, HDL) while others changed significantly (blood urea nitrogen, uric acid, total protein, albumin, total bilirubin, calcium, lactate dehydrogenase).

As a general rule, urine stability upon repetitive freeze-thaw cycles seems to be higher compared to serum and plasma. LC or UPLC-MS metabolomic profiles of urine are not affected by the thawing of samples for up to nine freeze-thaw cycle [113]. However, Saude and Sykes [92] previously found that urines thawed twice a week during 4 weeks show reduced metabolic stability. More recently, the stability of urinary volatile metabolites was investigated with regard to different storage conditions and freeze-thaw processes. The authors of these studies found that more than two freeze-thaw cycles may impact GC-MS metabolite profiles [116, 130]. To ensure good reproducibility it is advisable to thaw urine on ice or overnight in the fridge and to keep samples on ice during extraction. The latter process should be initiated as soon as the samples melt in order to avoid microbial growth.

Neither fresh feces nor fecal water stored for further metabolomic analysis should be subjected to unnecessary freeze−thaw cycles. Fresh fecal samples thawed to RT risk enzymatic activity that can affect the metabolic profile of the sample. Gratton et al. [98] found that alanine, lysine, leucine, isoleucine, and uracil were elevated after the second freeze−thaw cycle of fecal water and these changes persisted after the third freeze−thaw cycle. They also noticed increased levels of phenylalanine and decreased levels of N6-acetyllysine.

It is generally recommended to keep the number of freeze-thaw cycles of all biological fluids to a minimum, for which proper aliquoting of samples after their collection is crucial. However, if re-aliquoting of frozen samples is unavoidable, stepwise thawing at 4°C and immediate refreezing is advisable. Furthermore, a correct study design should consider the use of samples with the same number of freeze-thaw cycles for comparison purposes.

3.6.   Quality control

3.6.1.   Quality control samples for GC-MS and LC-MS analyses

A major challenge in GC-MS and LC-MS metabolomic analyses is the handling of analytical variability. To date, test mixtures, internal standards, and pooled quality control (QC) samples have commonly been used in order to monitor metabolomic workflows [58] (**Boxes 9 and 10**). QC samples in the context of metabolomics

have been described in earlier studies [131-133]. The repeated analysis of QC samples serves several purposes: i) the general monitoring of the performance of the analytical system, for example concerning retention time (t_R) and signal intensity stability, mass calibration etc.; ii) the determination of a method's overall precision (including sample preparation, intra-day and inter-day) if several QC sample aliquots are prepared independently per batch; iii) the calculation of QC sample-based drift correction functions aiming to remove systematic trends and batch effects (see more information regarding batch effect and signal drifts in Chapter 8).

The QC samples should be representative of the qualitative and quantitative composition of the samples being analyzed in the study. A pooled QC sample can be prepared at the time of sampling, directly from the vacutainer/container while other aliquots are created. Alternately the sample can be prepared at the time of sample extraction, by taking a part of the biological fluid or extract from each sample. The pooled QC sample can be preserved in a single vial/container. However, it is strongly advisable to split the pooled QC sample into several aliquots of sufficient volume (i.e. 5 mL aliquots) [131-133] to avoid too many freeze-thaw cycles (see **Box 8**) and to allow for multiple extractions at later stages (e.g. future analyses). The pooled QC sample should be prepared and extracted in the same way as the rest of samples from the study according to the SOP. For small to medium studies (up to 500 samples) aliquoting of 50 or 100 μl from each sample guarantees a sufficient volume of pooled QC sample. Extraction of study samples, QCs, and blanks must be organized carefully, and depends on complexity and duration of extraction method. In LC-MS assays, where 96 well plates can be used, a high number of samples can be prepared at once (i.e. 200 per day). It is thus possible to extract the pooled QC sample i.e. 20-30 times at the beginning of the first well plate and combine extracts into a single vial. Such a strategy allows one to obtain several mL's of pooled QC extracts. This QC extract is then injected multiple times at regular intervals during instrumental analysis run (e.g. two QC injections every 5/8 samples) and algorithms aimed at discarding noisy features or reducing sample-to-sample or batch-to-batch variations in signal intensity from these injections are then applied during data processing [17, 131-141] (see **Box 9**). High volumes of the extracted pooled QC sample are particularly important when the full sample set is being injected more than once, for example separately in the positive and negative ionization mode. Furthermore, if the QCs are prepared independently several times per day or batch, intra-day/intra-batch and inter-day/inter-batch method precision can be calculated for all the analytes detectable in the QCs. In case of long extraction methods (i.e. liquid-liquid extraction or derivatization procedures), where a limited number of samples can be extracted per day (i.e. 10-15/day) it is impossible to prepare a high volume QC extract at the beginning of whole extraction process. In

these cases the pooled QC sample must be extracted every day, together with blanks, and consecutively injected on a daily basis. This is particularly true for GC-MS extraction after derivatization, where repeated injection of derivatized samples is not recommended, as such extracts have limited stability (see **Box 10)**. In addition, derivatized extracts can be affected by the increased formation of artefacts such as siloxanes after penetration of the GC vial septa by the injection needle. For these reasons, an appropriate organization of QC sample preparation and injection is of key importance to reduce noise intensity, to reduce differences between samples or batches, and for a correct determination of variations in all processes involved in data acquisition (e.g. t_R and abundance) and data pre-processing (e.g. feature extraction) [17, 131-139, 142] (see **Boxes 9 and 10**). We recommend a double QC sample injection along the injection queue, as the trend estimation based on a mean of two injections is more robust.

If it is not possible to create a pooled QC sample due to limited sample volume or if the study involves thousands of samples to be collected over several months or years, then an alternative QC sample should be used. In the case of a large study set (i.e. greater than 500 samples) the QC may be prepared from the first batch of samples collected. However, the recruitment of subjects should be randomized and the samples should be representative of the entire study group. Alternatively, a commercially available QC sample may be used, for example pooled human serum and urine can be purchased from commercial suppliers. Preparation of the QCs should follow the same sample procedure performed in the preparation of the biological study samples, and the number of freeze-thaw cycles should be standardized for the QC and biological study samples [131-133]. If neither a pooled QC nor a commercial alternative is available, (i.e. for samples with low volumes such as tears or bile) then a synthetic substitute may be used. The synthetic substitute should comprise a metabolite mix that includes multiple representatives from each class of metabolites expected in the study samples. The synthetic QC should be prepared under identical conditions as the study samples.

The frequent injection of QC samples has proven to be quite efficient for correcting small variations like batch effects. But this only solves part of the problem since some types of variation, caused by reductions in sensitivity, or the fact that different compounds or compound classes exhibit different drift patterns, are almost impossible to correct afterwards [17, 131-139, 142]. To account for or to minimize this variation, action must be taken prior to the analytical run. For example, a systematic scheme for the sample run order should be created by combining the experimental design with randomization.

As mentioned earlier, QC samples should be analyzed intermittently throughout the analytical experiment. As these correspond to identical samples being analyzed multiple times, the data acquired from the QC samples can be used to determine the within-experiment precision. Following signal correction, calculation of the RSD (relative standard deviation) for each feature and the percentage of QC samples in which the feature was detected is applied for quality assessment. The percentage detection rate defines whether the feature is consistently detected or not, with an acceptance criterion of 50% being applied. If the feature is detected in fewer than five out of ten QC samples, then the data for that feature is removed [123]. In terms of RSD limits, the cut-off level depends on the analytical platform, i.e. an RSD < 20-30% for UPLC-MS and an RSD < 30% for GC-MS are generally accepted. The higher acceptance limit for GC-MS reflects the greater number of processing steps for GC-MS studies (e.g. chemical derivatization), lower injection volumes, and the reproducibility of GC injectors that creates a greater between-sample variation [131, 132]. This feature filtering step removes non-reproducible features before data analysis.

3.6.2. Randomization

One challenge in metabolomics is the analysis of large sample cohorts or samples measured at different points in time. This is because analytical drift often introduces a bias that obscures the data analysis. This includes common situations such as comparing matched sample pairs (e.g. control versus case or pre- versus post-intervention) and matched sample series (e.g. individual subjects over time or the duration of a process). This analytical drift can confound interpretation and biomarker detection [131-134]. Thus, an additional necessary step in the experimental design is randomization of the sample extraction and sample analysis sequence (see **Boxes 9 and 10**). This procedure minimizes the bias introduced when preparing and analyzing replicate samples jointly. An example of an already existing and accepted approach is to use the experimental design to create sample run order schemes, e.g. to obtain a balanced distribution between cases and controls in all separate batches or well plates [17, 131-139, 142]. In principle, two randomization schemes are possible: complete randomization and partial or groupwise randomization. Complete randomization means that all samples are randomized over all batches or the entire measurement series. This means that the samples from all groups are exposed to all sources of analytical errors (random or systematic) to approximately the same extent. This randomization scheme is appropriate if – in principle – a comparison between all samples or all groups is the aim. Sometimes full randomization is not possible due to many reasons. The most frequent reasons are when high numbers of samples need to be analyzed (i.e. several thousand), when a technical intervention needs to be conducted during

the instrumental analysis (e.g. cleaning of mass spectrometer), as well as when the duration of the experimental design is very long (population screening of several years or follow up studies). Partial randomization means that samples belonging to particular subgroups are analyzed in a batchwise fashion (equal amount of samples in each batch). For example, in a kinetic intervention study, one batch could comprise all timed samples from one volunteer. This approach minimizes analytical errors within the group but accepts potentially higher systematic offsets between the groups [139]. Other factors that should be balanced when dividing samples between several batches are: diet (equal number of control vs intervention), sex (equal number of males and females), age, BMI, etc. The choice of the randomization scheme may thus depend on the research question being asked.

3.6.3   Internal standards and retention index markers

The variability of a metabolomic measurement can be determined with internal standards. Internal standards serve to control entire measurement series, to correct for t_R shifts during chromatography, and to help correct drift or batch effects happening during measurement. How many and what kind of internal standards should be used depends on the analytical question. Internal standards can either be added to the sample or they can be taken from the analysis itself (endogenous compounds). In the first case, chemicals, which normally are not present in the biofluid or an isotopically labelled compound, can be added. Both approaches are used for t_R (GC-MS, LC-MS) or reference frequency (NMR) alignment between analyses and batches. Isotopically labelled metabolites can also be used for absolute quantification of their corresponding endogenous metabolites. It is advisable to choose a labelling level with care, especially in case of deuterated compounds. Compounds with one to three deuterium atoms can overlap partially with the isotopes of native compounds, thereby creating identification and quantification problems. It is often difficult to decide what isotopically labelled metabolites should be added for non-targeted metabolomic analysis. As a general rule, it is advisable to add more than one chemical for t_R verification. The more standards that are added, the less dependent the correction is for unexplained effects of single chemicals. Internal standards should be introduced into the analysis process as early as possible.

In case of ^{1}H-NMR spectroscopy internal reference compounds such as a 3-(trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt (TMSP or TSP) and 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) are used in most studies. DSS is similar to TMSP, traditionally employed in NMR analyses, but the former is characterized by a much higher water solubility [143]. In GC analysis some compounds are used to determine the retention indices such as n-alkanes or saturated fatty acid methyl esters (FAMEs) [144]. Retention index markers can be added to

all samples or only to the blanks (depending on their use in data processing). For LC-MS assays, a variety of deuterated compounds such as tryptophan-d5, hippuric acid-d5, phenylalanine-d5, bile acids (any kind)-d5, or creatinine-d5 are frequently used, as they are characterized by a good ionization response in the ion source for both ionization modes. In Chapter 8 more information is given regarding the use of internal standards in explorative analysis and quality assessment.

3.7.   Normalization: pre- or post-analytical processing

Normalization within the metabolomic workflows can occur during the early stages of the sample analysis phase (pre-analytical normalization) or, later, during the data analysis or data processing phase (post-analytical normalization) (**Table 3**) (see Chapter 8 for post-acquisition data normalization). Normalization of samples themselves is key as, depending on the biofluid analyzed, their composition can greatly vary depending on day time, health status, food and water intake, gender, and age (see section 2.2.1 on inter and intra-individual variability).

The human body tightly controls blood volume and composition [145]. Thus, blood samples – whether serum or plasma – often do not have to be normalized in human studies. However in the case of urine, overall concentration of metabolites may differ by a factor of more than 20 as a result of the physiological control mechanisms responsible for the maintenance of water homeostasis [146]. In order to enable a proper comparison of metabolite levels in different urine samples, a normalization step is mandatory [147-149].

Pre-analytical normalization means one adjusts the overall sample concentration by differential dilution during sample preparation. In principle, this can be done according to, e.g. creatinine content, osmolality, specific gravity, or urine volume. As demonstrated by Edmands et al. [150] the discovery of discriminating features of dietary intake could be improved by pre-analytical normalization based on urine specific gravity. In accordance to these outcomes, the ability to identify discriminating markers based on multivariate data analysis was improved when pre-analytical normalization to osmolality was compared with non-normalized datasets [151]. Pre-analytical normalization has both advantages and disadvantages. The advantages are that i) more analytes can be consistently detected within the linear dynamic range (which is especially relevant in the case of MS-based metabolomics); ii) inter-sample matrix effects caused by largely different concentrations of matrix components, for example during derivatization or evaporation in the GC sample inlet, are reduced. The

disadvantages are that: i) it is time-consuming; ii) it can be an additional source of error during sample preparation; iii) it excludes the possibility to change the normalization later on; iv) it might cause low abundance metabolites to fall below their detection limit. Normalization on creatine content should only be considered if the study population consists of only men or only women, but not both (**Box 11**) [47, 132, 152, 153]. If reference data for pre-analytical normalization are missing, post-analytical normalization is required.

Post-analytical normalization is done by calculation after the instrumental data has been collected and enables a comparison of the effects of different normalization references (more details are given in Chapter 8). However, post-analytical normalization means that one has to scale the intensities of the analytes using a global sample-specific factor, which ignores the existence of analyte-specific response factors. [47, 132, 152-154]Thus, as equal solute concentrations among urine samples cannot be taken for granted, we suggest pre-analytical normalization be used if information regarding osmolality or specific gravity are available, otherwise post-analytical normalization is required. Details regarding these issues are available in Chapter 8.

4.   Sample preparation

Beyond considerations base on the characteristics of the samples to be analyzed, a proper nutritional metabolomic experiment requests a careful preparation of the samples, which is specific to the instrument being used for the analysis. Detailed information on these aspects can be found in Chapter 4 (Appendix 1) devoted to the sample preparation as detailed for GC-MS (section 4.1 – Sample preparation for GC-MS), LC-MS (section 4.2 – Sample preparation for LC-MS), and NMR (section 4.3 - Sample preparation for NMR).

5.   Chromatography

While the use of NMR for metabolomics usually does not require chromatographic fractionation of the samples, this step is crucial to all MS-based metabolomics [155, 156]. Chapter 5 (Appendix 2) provides detailed information on the organization of the samples for chromatography, which is common to both GC-MS and LC-MS (section 5.1). The reader interested in GC-MS may then proceed to a special section devoted to the injection of GC samples (section 5.2). The two major chromatographic techniques in GC-MS are then presented in section 5.3 (One-dimensional gas chromatography) and section 5.4 (Two-dimensional gas chromatography). For readers interested in LC-MS they may continue reading section 5.1 with section 5.5, devoted to the injection of LC samples, and section 5.6, describing LC.

6.   Data acquisition

Data acquisition, presented in Chapter 6 (Appendix 3) is the core process in metabolomics allowing the detection and quantification of relevant features. Two key spectrometric technologies, MS and NMR are used for data acquisition. Section 6.1 provides details on data acquisition in MS whereas Section 6.2 describes data acquisition in NMR spectroscopy.

7.   Raw data files conversion and feature extraction

Before a proper statistical analysis of the data can be performed, data acquired by MS and NMR spectroscopy first needs to be processed via a technology-dependent manner. Chapter 7 (Appendix 4) provides a review of data conversion and feature extraction methods in MS-based metabolomics (section 7.1) whereas section 7.2 provides the corresponding information for NMR-based metabolomics.

8.   Data visualization, pre-processing, and analysis

Once the features have been extracted, the data is typically presented in a "raw" data matrix, which normally needs to be processed before data analysis. We refer to this phase as pre-processing. This step encompasses several operations, such as imputation of missing values, sample normalization, and transformation/scaling of variables. The subsequent data analysis includes a broad range of techniques used to extract relevant information from the data. Nutritional studies often rely on repeated measure designs, which can be either parallel or crossover according to the aim of the particular experiment [157] (see Chapter 2). Classical statistical modeling offers established approaches to effectively manage these designs. However, explicit inclusion of such a design or factor information in multivariate analysis and machine learning is less established and is still an active area of research. Regardless of whether one is using classical or multivariate data analytical approaches, modelling, validation, and testing are crucial aspects that have to be carefully considered. Above all, we want to stress that the selection of the most appropriate approaches for data analysis and validation require an early and continuous interaction between the data analysts and the researchers when defining the study design and its goals. **Box 12** shows Tips & Tricks on a typical workflow for data analysis and statistical modelling.

8.1.   Data visualization

The efficient use of visualization tools to capture the complexity of a nutritional metabolomic dataset does often not receive sufficient attention in the design of the data analysis protocol. However, results should be both

visualized and visible at the level of raw data, regardless the complexity of the data and the analysis strategy. In particular we recommend that: i) biomarkers should e.g. show different concentrations in strip-plots; ii) correlated variables should show non-random patterns in xy-plots; iii) outliers should be clearly distinguishable from the bulk of the data [158, 159]. If these criteria are not met, an error has most likely occurred in the data analysis pipeline. Unfortunately, it is not always easy to visualize metabolomic datasets, especially in complex experimental designs. First, the sheer number of variables does not allow for an efficient visual inspection. Second, a multivariate visualization is generally more informative in the presence of correlated variables. Appropriate statistical or other data analytical models, suitable for the study design and for addressing the aims of the investigation, should only be applied after careful data inspection and data quality assessment.

8.1.1.   *Multivariate data visualization*

Principal Component Analysis (PCA) is a multivariate statistical approach that is particularly suited for data exploration [160, 161]. It projects a multidimensional dataset onto a reduced number of (usually few) latent variables that correspond to the main variance components in the data. In PCA, variance is a proxy for the information content and this tool is thus useful to identify the dominant sources of variability in the dataset (outliers, analytical trends, batch effects, etc.). In the PCA approach, the data are approximated with a linear combination of observation scores and variable loadings. The scores are the coordinates of the observations (i.e. samples) in the lower-dimensional PCA space, while the loadings quantify the contribution of the measured variables to each component. The PCA model is then represented graphically with score plots, loading plots, or biplots. The score plot (usually) gives a 2D representation of the sample distributions, according to the score values of the principal components. Since PCA is built without using information from the study design or other underlying factors (such as age, gender, BMI, etc.), any pattern observable among the samples is due only to the trends, similarities and dissimilarities of the inherent metabolomic fingerprints [162, 163]. Thus, PCA can be used to inspect relationships among the samples (**Box 13**), such as the presence of obvious clusters in the data or outliers, and to identify possible confounding factors [160]. However, since the inter-individual variability often exceeds the systematic treatment variability, the first principal component(s) may not reveal the class separation of relevance to the research question. This, however, does not *per se* imply the absence of intervention effects [162]. Other tools, such as probabilistic principal component analysis (PPCA), may be used to visualize metabolomic data conditional on covariates such as age, gender, and body mass index [163, 164].

PCA loading plots can be used for exploring relations among variables. However, interpretation of loading plots can be complex, especially in the presence of many variables. Finally, biplots are a superimposition of the previous two representations (score plots and loading plots), which can be used to give a general view of the relations between the variables and the samples. The biplot can be useful to identify variables "driving" the separation between sample classes or variables relating to gradients of outcome or confounding variables. PCA is normally performed on mean centered data [165, 166] and is sensitive to the variable scaling, which affects the relative potential contribution of variables to the overall variability and thus to the PCA [167]. Different approaches for variable scaling can be adopted [168] with "unit variance" scaling (i.e. scaling by standard deviation) [166] and Pareto scaling (i.e. scaling by the square root of the standard deviation, which reduces the impact of low intensity signals) [165, 169-172], being the most frequent scaling techniques used in nutritional metabolomics. For a detailed review on centering and scaling see van den Berg et al. [168].

For complex experimental designs, ANOVA-Simultaneous Component Analysis (ASCA) and ANOVA-PCA can be used to model the structure of the study, including multiple factors (both study and confounding factors) and time series designs, into modelling and visualization [173, 174]. These methods decompose the original metabolomics data matrix by factor levels into several submatrices. These "effect matrices" can then be analyzed by PCA to examine their overall variability [174].

8.1.2. Outliers

Outliers can be seen as the (few) observations that clearly stand out from the bulk of data [175, 176] and they can be the result of biological or analytical variability [159, 176]. Removing these observations can often remove focus in the data analysis from differences between bulk and outlier observations to systematic differences between treatment effects. However, outlier removal should not be performed without careful consideration. Outliers due to biological variation can, indeed, highlight relevant phenomena that have been under-considered in the experimental design and therefore potentially highly informative. On the contrary, "odd" samples, which can be clearly attributed to analytical issues, should be removed in order to obtain unbiased results [176].

In untargeted metabolomics, outliers are normally detected using PCA, where outlying samples indeed will contribute strongly to systematic variability and, thus, be clearly distinguished in the first principal components [159] (see **Box 13**). Outliers can also be identified in a hierarchical clustering [175]. At the univariate level, outliers are often identified using box plots, where they are located far away from the interquartile range (IQR).

However, it is important to note that univariate approaches can be misleading in untargeted experiments since they do not necessarily reflect the complexity of the dataset.

8.2.   Data matrix pre-processing

The data matrix may require pre-processing to account for undesired sources of variability, e.g. pre-analytical or analytical variability. Pre-analytical variability needs to be minimized during sampling and sample management, but it may also be reduced post-analysis using various computational approaches [177]. The impact of undesired analytical variability is especially pronounced in MS-based data and can be measured and reduced by proper organization of the analytical sequence and by including suitable QC samples for modelling and correcting instrumental drift (see section 8.2.2). Care should be taken to keep analytical factors (batches and sequences) orthogonal to the study factors to reduce systematic instrumental analysis bias to a minimum.

Pre-processing also includes the removal of insignificant features, as well as sample normalization and variable transformation. Careful curation of data is essential to ensure high quality data analysis and to effectively focus on the relevant study questions.

8.2.1.   *Removal of insignificant features: blanks, noise*

The current practice for data cleaning is to use blank samples to subtract chemical noise from derivatization reagents, solvents, contaminants, column bleeding, or carry-over. The identification and elimination of noisy features, commonly present in MS data and NMR spectroscopy, allows one to focus on relevant features present in the data. For instance, it is likely that variables showing small variance (across all the samples) are not affected by the dietary treatment under study and thus can be considered just noise and filtered out [178].

Filtering can also be performed on the basis of analytical criteria, by considering the feature variance across repeated injection of QC samples [179, 180]. In a similar way, it is also possible to set up a dilution series of QC samples, thereby excluding features not exhibiting linear trends. However, as discussed by Naz et al. [171], this approach may lead to the loss of relevant features. Blank samples (run at the beginning of each analytical batch) may also be used to identify spurious signals, which can be excluded from the subsequent data analysis [165, 179].

8.2.2.   *Batch effect inspection and removal*

In addition to the biological variability brought on by the study design, variability also arises from undesired factors, such as pre-analytical issues in sample management [177] or instrumental variability [134] within and

between batches. These batch effects have to be carefully monitored and managed as they may otherwise significantly determine the outcome of the study.

In many cases, the importance of the batch effect can be inspected using PCA (see **Box 13**). Ideally, biological variability should dominate the score plot, indicating that the analytical pipeline is sufficiently stable to address the biological problem under investigation. If this is not the case, systematic variability can be modelled and therefore be used for signal improvement.

Instrument stability is widely considered to be higher for NMR than for MS. Therefore batch correction in NMR-based metabolomics may not be necessary. On the contrary, for LC-MS metabolomics, systematic and random variability in signal sensitivity, mass accuracy (m/z), and t_R between samples, both within and between batches [132], contribute to noise and increased risk for misalignment. This can have a negative impact on statistical analysis [181, 182]. Misalignment is, in general, larger between batches than within and systematic misalignment between batches is especially problematic. However, efforts to address this issue are still extremely limited [134].

Due to the lower analytical stability of MS compared to NMR, awareness of the need to correct for signal drift is much higher and several options are available. For targeted analysis, the use of internal standards is recommended [183, 184]. For untargeted metabolomics, instead, internal standards may not cover all signal fluctuations adequately, since the large number of detectable analytes in general cannot be adequately covered by a limited number of internal standards [64]. Following a seminal paper published in 2011 [132], QC sample strategies have become commonly applied in signal drift management. These strategies use QC feature intensities to model and normalize sample-to-sample variation in signal intensity. The current literature is rife with QC-correction algorithms (e.g. [65]). For instance, it is possible to correct the signal drift by local regression (LOESS) of the intensity trend in pooled QC samples [132]. Several regression models have been proposed with different susceptibility/tolerance to outliers. In particular, van der Kloet et al. [17] proposed an adjustment of offset differences between the analytical batches by using the average of the pooled intensities in each batch. Population-based, instead of QC-based, correction procedures are also an interesting option, especially for such features that are not well captured by QC samples [139]. The efficacy of the batch correction algorithm should be clearly visible by examining the trajectory of QC injections in a PCA plot. In fact, the trajectories should be substantially less apparent or even removed after drift correction. There is, thus far, no consensus in the metabolomic field on which algorithms produce the most robust correction. While awaiting proper benchmark

testing, the choice of which algorithm to apply is thus second to the choice of incorporating drift correction in the analytical pipeline [185].

8.2.3.   *Missing value imputation*

With the term "missing value" we refer to a hole in the data matrix, which indicates the absence of a specific feature inside a sample. A typical metabolomic dataset may contain 20% of missing data in up to 80% of all variables [186]. Missingness can occur from analytical, computational, or biological causes [186-188], but, regardless of origin, a principled strategy to deal with them is required. Features showing a high degree of missingness can, for example, be removed from the data matrix whereas remaining missing values can be imputed, i.e. filled in with reasonable numbers. Unfortunately, a consensus definition of the high degree of missingness-threshold does not exist. A feature can be missing because its associated compound is absent in the biological sample or because its concentration is below the detection limit of the equipment [186]. Missing values of this type are frequently concentrated in one of the study groups, reflecting biological variability induced by the study question [186]. This can be particularly frequent in nutritional studies, where some metabolites appear in the biological samples only after eating specific foods (i.e. intervention groups), while they are absent if these foods are not consumed (i.e. control groups). When this happens, the threshold for variable removal should be evaluated taking into account the design of the study to avoid substantial biases in the analysis [186] e.g. by evaluating missingness on a per-class-basis. On the other hand, when missing values are seemingly randomly distributed among study groups, the situation is less critical and the analyst has more freedom to remove highly-missing variables, while imputing the rest. Some of the most popular strategies for handling missing data in metabolomics are zero imputation, k-nearest neighbors (kNN) and random forest (RF) imputation [189]. Several other strategies have been proposed and implemented [175, 187, 189, 190], although none of them is considered universally optimal [190]. In MS it has been shown that single-value imputation methods (such as zero, half-minimum, mean, or median imputation) risk to artificially reduce and skew variable distributions and therefore should not be the first choice. Moreover, several imputation techniques are sensitive to outliers [72], wherefore outlier detection is required prior to the application of imputation. It is important to remark that imputation strategies should not be guided by known factors or covariates to avoid overfitting.

8.2.4.   *Data transformation: normalization and scaling*

In addition to analytical variability from instrumentation, biological samples may also represent different dilution conditions. Metabolite concentrations in urine, for example, can vary greatly between samples. In such

cases, a straight comparison of metabolic profiles without correction could result in biased or unclear results [191]. Therefore it is often necessary to normalize between urine samples to improve comparability. Sample normalization methods do indeed have an impact on the outcome of metabolomic experiments and this topic has received much attention in the last decades [168, 192, 193]. Normalization is often performed by using sample specific scaling factors like total ion current (TIC), mean or median intensity, thus assuming that all signals within the samples should scale similarly. Another aspect of this assumption, however, is an implied self-averaging. To maintain a constant overall signal, a reduction of some metabolites should be compensated by an equal increase of other metabolites, which is not always the case. To address this limitation, several alternatives have been developed, such as Probabilistic Quotient Normalization (PQN) [194]. This technique was originally developed for NMR and represents an attractive alternative for LC-MS [195]. All the above-mentioned methods are "data driven" in that normalization is performed based only on signals arising from the actual sample without using external information, such as intrinsic matrix properties. For example, creatinine or osmolality can be used for normalization of urine both pre- and post-analysis [151, 191, 196]. It should, however, be noted that normalizing by creatinine or osmolarity has been criticized since these levels are not necessarily constant for various diseases or other physiological conditions [197]. As a general recommendation, several normalization methods should be tried and critically assessed (see **Boxes 11 and 13**). Heuristics for the performance comparison of normalization strategies were recently suggested, based on statistical criteria such as the maximization of biological inter-group effects, the reduction of intra-group effects, and p-value distributions [198].

Between-sample normalization thus aims to make comparisons between samples less sensitive to pre-analytical or analytical variability. On the other hand, operations such as transformation, centering and scaling can instead be performed on the variables to adjust how they contribute to data modelling and data analytical results [168]. For models with underlying assumptions of homoscedasticity (i.e. where the variance is independent on the signal intensity) or where extreme differences in intensity between samples may occur, logarithmic (or square root) transformations can be employed to achieve more statistically valid and reliable results [70]. Multivariate, component-based methods, such as PCA and PLS (partial least squares) are particularly sensitive to the scale of the variables. Therefore mean centering and scaling are frequently performed prior to applying these methods. The two most frequently applied methods for scaling include scaling to unit variance (achieved by dividing intensities by the standard deviation) and Pareto scaling (achieved by dividing intensities by the

square root of the standard deviation). A frequent intuitive interpretation is that unit variance scaling allows the variables equal opportunity to contribute to the multivariate model, whereas Pareto scaling down-weighs the contributions from signals with lower intensity. It is important to note that the increase in importance of low intensity variables makes the overall results more sensitive to analytical issues because low intensity signals are more difficult to measure reliably and often show a higher degree of missingness. On the other hand, down-weighing of low intensity signals may result in loss of biologically relevant information. Traditionally, unit variance scaling has been the standard choice for MS-based approaches, whereas Pareto scaling is more frequent in NMR metabolomics, although the edges are blurring. It should furthermore be noted that several other options for transformation, centering and scaling are available and the default choices described above are not always the best choice [168]. In addition, there are also multivariate methods that are scale invariant and therefore insensitive to variable transformation, centering and scaling, such as random forest [199].

8.3.   Data analysis

The last part of the data analysis protocol comprises the extraction of useful information from the data. At this point the data analyst should be quite familiar with not only with the scientific question and the experimental design, but also with the peculiar characteristics of the dataset (outliers, trends, batches, possible confounding factors). All this information will be critical to performing a fruitful data analysis and avoiding mistakes or biased conclusions.

In principle, the experimental design already defines the most appropriate statistical approach [157], but also the application of the most suitable statistical model will not be straightforward because many models rely on specific assumptions on data distribution that could often not be met in practice. In addition, the experimental variables often largely outnumber the samples and this poses further challenges to many well-established statistical tools [200]. Data mining or machine learning tools can help in such scenarios because these algorithms usually have limited requirements about the data distribution. Unfortunately, the introduction of external knowledge (such as information about the experimental design) in these tools can be difficult and this can lead to inconclusive results. Moreover, the interpretation of the outcomes of the analysis in terms of the initial variables is often not straightforward. In other words, machine learning algorithms are often very efficient for prediction, but can be unsatisfactory for visualization, interpretation, or search for a mechanistic understanding of the process under investigation. However, machine learning methods can be hyphenated with pathway or network visualization tools to improve interpretability.

8.3.1. *Statistical modelling*

Classical statistics provide researchers with a well-established set of tools for addressing several research questions, such as comparing the response to different treatments or modelling the effects of a set of variables on the concentration of a metabolite. Statistical models can be constructed with high flexibility and in the process of model building and checking it is possible to include external knowledge from the experimental design or knowledge of confounding variables. Building and fitting a model is normally an iterative procedure where the data analyst will have to decide which variables should be included. This usually leads to well-fitting, interpretable models. Due to multiple collinearities, poor signal-to-noise, or lack of relevant information in individual variables, it will rarely happen that more than a minority of all measured features are included in the model. This may facilitate (or even over-facilitate) the interpretation. Linear models (LM) are the basis for comparing outcome averages (both for numerical and categorical variables) across subpopulations [201] and they can be extended to cope with a variety of more complex study designs. When the outcome variable is restricted (e.g. count or binary) or when variable variances depend on their signal intensities, such as for most chemical analyses, the analytical framework should be extended to generalized linear models (GLM) [157]. In the presence of grouped data, LMs and GLMs should be made hierarchical (mixed models) to fully take advantage of the groups defined in the study design. Such grouping includes repeated measures (both parallel and crossover), blocked designs, and multilevel data.

In the specific area of nutritional studies, statistical models can also be fruitfully used to analyze time course data [201]. When the interest is to model trends in time, generalized additive models (GAM) [202] are a possible solution with the caveat that, due to their data driven nature, they require a sufficient amount of time points to give reliable results. In presence of a less frequent sampling, kinetic models can be an attractive alternative, because less data are required to reliably fit a model of well-defined mathematical form [203]. Unfortunately, the application of statistical models in multivariable omics is not always easy, with the major issue being unstable modelling arising from multiple variable collinearities, which are always present in untargeted metabolomic experiments (e.g. ions resulting from the ionization of the same metabolite or NMR signals arising from the same molecule). Consequently, a preliminary step of variable selection is always necessary, which may result in loss of information. See **Box 12** for an overview of different statistical methods.

8.3.2. *Data mining methods*

In contrast to linear models that strain under the pressure of collinear variables, multivariate methods will typically thrive in the presence of correlated variables. These approaches have indeed been developed with the explicit objective to exploit variable correlation and identify hidden structures in the data. The hypothesis is that -omic experiments are not actually reflecting thousands of measured variables but, rather, that these measured entities are in fact the proxies of much fewer, relevant latent variables.

Two extremely popular multivariate approaches in metabolomics are Partial Least Squares (or Projection to Latent Structures (PLS) and its classification version PLS Discriminant Analysis (PLSDA). The interested reader can find a recent review tutorial on PLSDA in the work of Triba et al. [204]. A variant of PLS, Orthogonal Projection to Latent Structures (OPLS) has also been developed to facilitate model interpretation in terms of the original experimental design variables [205, 206]. Being supervised techniques, PLS-based approaches identify spectral signals that co-vary with the modelled variable, e.g. class membership (PLSDA) or actual dietary exposure (PLS regression). In other words, taking *a priori* knowledge of samples into account allows one to filter out metabolic information that is not correlated to the nutritional treatment or dietary pattern under consideration [169, 207-210]. Recently, PLS-based methods have also been proposed in data fusion applications. Despite their popularity, these tools are not the be-all and end-all of supervised, multivariate data analysis. One fundamental problem is that correlation among variables arises by chance as the sample-to-variable ratio decreases and the multivariate tool can be fooled by random correlation resulting in overfitting and false positive findings. In this condition, multivariate models can return attractive results, which unfortunately cannot be generalized. In order to avoid this situation, a thorough validation of multivariate models is absolutely mandatory (see section 8.3.3). This fundamental aspect is often not given enough attention, which has resulted in a plethora of overoptimistic results (i.e. false positive results) in the metabolomic literature.

As already mentioned, it is not easy to include the experimental design (dependent samples, cross-over designs, etc.) in multivariate algorithms. However, specific solutions for PLS-based methods have recently been proposed. These multilevel, multivariate tools have been first suggested for unsupervised PCA [211] and later also for supervised PLS [205, 212]. The general idea behind these techniques is to apply multivariate modeling not on the original data, but rather on the effect (or difference) matrix between dependent samples, thus effectively managing the dependency between samples. A version of the multilevel technique, OPLS-Effect Projection, has been recently suggested, which provides analytically identical results as standard multilevel PLS [213]. OPLS-

Effect Projection is more easily implemented in commercial software for PLS analysis, since most commercial software does not allow cross-validation conditional on dependent samples.

In broad terms, a standard PLSDA can be seen as a multivariate equivalent of an unpaired t-test between two groups and multilevel analysis, primarily developed to manage the dependency between two samples per repeating unit, corresponds to a multivariate paired t-test. Following this line of thought, one could envisage the development of multivariate tools able to incorporate multiple (> 1) factors and/or multiple (> 2) levels per factor like in the abovementioned extensions of the linear model family. These types of tools would be extremely useful for the analysis of time series data. Time series data is frequently collected in studies on postprandial metabolic regulation or biomarker discovery, especially when the most opportune time to look for relevant biomarkers of dietary (or other) exposure is *a priori* unknown. Along this line of development, it has already been suggested to analyze ANOVA-decomposed matrices in a supervised fashion by PLS [214].

Going beyond PLS, other data mining tools, like Support Vector Machines (SVM), Random Forest (RF), Genetic algorithms and kNN (k-nearest neighbor), to name a few, are routinely applied in other domains. These are receiving increasing attention for the analysis of untargeted metabolomic data [215]. Unfortunately, even the most powerful approach will be confronted with the need to develop a strategy for including knowledge about the experimental design in the analysis, which is not obvious for many of the above-mentioned algorithms. To obtain an optimal performance machine learning tools usually require the tuning of (sometimes many) algorithm parameters; this requires access to a sufficient number of samples. This is always a problem in untargeted metabolomic investigations. Moreover, the interpretation of machine learning outcomes is generally less easy than PLS family of algorithms that can make the quest for a meaningful biological interpretation a daunting task.

8.3.3. *Validation and statistical testing*

Untargeted metabolomics will, in general, produce "wide" data, i.e. dataset in which the number of variables (or features) outnumbers the number of observations. This results in mathematically underdetermined systems and produces the connected "curse of dimensionality", also known as the "large p, small n" problem [200]. This huge number of partially collinear variables poses challenges not only in statistical modelling, but also for the validation of models and the generalization of scientific findings. Any findings from a statistical analysis should ideally be validated using independent samples. The experimental design should then allow the (iterative) splitting of the samples in training and test sets: the former to be used for model tuning, the latter for honest

estimation of its model performance. Unfortunately, researchers in nutritional metabolomics can seldom afford to keep samples out of model construction due to the loss in power combined to the high resource cost to generate data. Several procedures to address the issue of validation have been proposed: they aim to reduce overfitting and false positive findings using only available data [200].

In the case of univariate approaches, the most common strategy is to assess the general validity of the results using significance (p-value) as a proxy. Variables are analyzed one by one using classical statistical tools (e.g. Student's t-test, ANOVA, or mixed models) and only the variables showing a sufficiently small p-value are considered relevant (for historical reasons, 0.05 and 0.01 are commonly considered as acceptable significance thresholds, even though they should not be taken as sharp boundaries). The multitude of variables in -omic sciences result, however, in the well-known multiplicity issue [216]. Thus, to assess statistical significance it is necessary to adopt strategies to reduce false positive discoveries using p-value correction approaches. These include, for instance, the most conservative Bonferroni correction [164, 217], the Benjamini-Hochberg ("false discovery rate" or FDR) correction [170, 180, 210, 217], or the Holm step-down procedure [217].

Multivariate models, on the other hand, will in general not generate p-values for modelling outcomes, so validation is the necessary step to check if the outcome of a specific experiment can be generalized. Three different levels of validation can be used: (internal) cross-validation to decrease the level of overfitting [205, 218], permutation analysis to assess overfitting during model construction and to formally test the actual model performance [205, 219], and external validation to examine the generalizability of observed findings.

In (internal) cross-validation, the data are divided into segments and submodels are created by holding out test segments and training on the remaining data. Examples of strategies for splitting the data include K-fold, leave one out, Monte-Carlo, and repeated double cross-validation [220-222]. Optimal parameter values can then be obtained from predictions for several parameter settings, by comparing fitness metrics. Among them, Q^2 (predictive ability of the model) is particularly popular, although several other fitness metrics are available [223]. Despite its usefulness, cross-validation does not safeguard against overfitting so more complex validation schemes have been developed to attain more unbiased prediction estimates [205, 218]. To gauge model performance further, permutation analysis is used to compare the fitness metric from the actual model against a null distribution of fitness metrics from models constructed using randomly permuted data. For instance, the Q^2 from the original dataset will be compared with the distribution of Q^2 values when the original response values are randomly assigned to the observations [163, 171, 219]. This type of validation will tell the analyst if the

information content associated with the "right" labels is higher than the one obtained only by chance. However, it will not guarantee the general validity of observed findings. Independent validation remains the only honest way to evaluate general validity of observed findings. This should be increasingly taken into consideration in the study design phase and networks for nutritional metabolomics, such as the FoodBAll consortium [8], could help overcome present shortcomings in the validation procedure by providing access to study material for external validation.

8.3.4.   *Selection of discriminative features*

To facilitate the interpretation of the experimental results, it is often necessary to identify the features, which contribute the most to the performance of a multivariate model. In general, the identification of the subset of the most important features (potential biomarker candidates) depends heavily on the analysis pipeline. In the case of statistical modelling, the selection of the most interesting variables is included into, and indistinguishable from, the complex process of building statistical models [83]. On the contrary, if a multivariate approach has been chosen, inner characteristics of the model (like weights, loadings, or regression coefficients) are usually inspected for identifying the most important features. In the case of projection-based methods, the Variable Importance in Projection (VIP) score represents the relative importance of a variable within the PLS model. For OPLS, S-plots can also be used for variable importance visualization and variable selection [171, 224, 225]. This representation combines the contribution of a feature to the variance of the observations with the reliability of its predictive contribution. In case of two models, the Shared and Unique Structure (SUS)-plot, an extension of the S-plot, can be used to identify the features that are important for both models or unique to one of the two [225]. The SUS-plot can be useful to analyze a design including a control, placebo, and intervention group [164]. Variable importance can also be generally assessed by Jack-knifed approaches [226] although this requires increased biostatistical and computational expertise from the data analyst. Other examples of feature selection approaches rely on the receiver operating characteristic (ROC) curve (which combines the sensitivity and specificity of the individual biomarkers or of the biomarker panel) [169, 227], K-means clustering with Pearson or Spearman rank correlation [180], and hierarchical clustering [170]. Examples of tool implementing these approaches are MetaboAnalyst, GENE-E, and the ROCCET web-based tool [208].

8.3.5.   *Data analysis and outcome reproducibility*

Making data and data processing workflows available for the community is essential to guarantee computational and scientific reproducibility according to the FAIR protocol [3]. Metabolomic data can be documented with the

ISAtab format [228] and then stored into Metabolights [7] or in the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI), which delivers the Data Sharing In Nutrition (DASH-IN) infrastructure (http://www.enpadasi.eu/), or other similar dedicated cross-technique repositories. The data analyst can ensure the reproducibility of the data analysis workflow by relying on scripting languages such as R (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/) or Matlab (MATLAB and Statistics Toolbox, The MathWorks, Inc., Natick, Massachusetts, United States) and sharing the code needed for repeating the analysis (for instance through Github [229] or similar platforms). An interesting alternative is represented by web-based workflow management solutions. In the field of metabolomics, Workflow4Metabolomics (W4M; http://workflow4metabolomics.org) permits the building and running of complete data analysis pipelines, which can be stored and documented together with the associated data, published on W4M (i.e. shared with the community) and referenced with a digital object identifier (DOI), which can be cited in publications [230]. Another example is PhenoMeNal (http://phenomenal-h2020.eu/home/), a comprehensive and standardized e-infrastructure that supports the data processing and analysis pipelines for molecular phenotype data generated by metabolomics applications. Also, XCMS Online, a web-based version of the widely used XCMS software, allows users to easily upload and process LC-MS data, providing a solution for the complete untargeted metabolomic workflow including feature detection, t_R correction, alignment, annotation, statistical analysis, and data visualization (https://xcmsonline.scripps.edu) [231]. Finally the processing of NMR data can be done with NMRProcFlow an open source software that greatly helps spectra processing (https://nmrprocflow.org/). NMRProcFlow has been developed as two separate applications (NMRspec and NMRviewer), each of them embedded in Docker images, sharing a data volume and communicating together through AJAX web services by making intensively use of JavaScript functionalities [232].

9.   Compound identification

Compound identification, presented in Chapter 9 (Appendix 5), is clearly the major bottleneck in metabolomics. This issue is even more relevant in nutritional studies, which deal with complex interactions between the human organism and foods or diets that are characterized by a large compositional diversity. In this context, particular care needs to be given to clearly identifying the level of identification of the metabolites (section 9.1). Section

9.2 provides information on the databases available to researchers conducting metabolomic experiments, including a database focusing on the food metabolome. Finally, specific aspects of compound identification are detailed for each of the three metabolomic technologies, namely GC-MS (section 9.3), LC-MS (section 9.4), and NMR (section 9.5).

10. Biological interpretations

10.1. Defining biomarkers

Once the compounds of interest in a nutritional intervention study have been identified by a particular analytical platform, the next step is the biological interpretation of the findings. However, the research community is often confronted with a semantic issue in as much as biological effects are tightly associated with biomarkers and the mere definition of biomarkers is poorly harmonized, even understood. Therefore the term 'biomarker' needs to be defined because it is used with different meanings in different scientific communities. The FoodBAll consortium has recently defined biomarkers as "objective measure used to characterize the current condition of a biological system" [233]. In particular, three major categories of markers are presented in this position paper: i) exposure and intake biomarkers, reflecting the level of extrinsic variables that humans are exposed to, such as diets and food compounds, including nutrients and non-nutrients; ii) effect biomarkers, referring to the functional response of the human body to an exposure; and, finally (iii) susceptibility or host factor biomarkers, representing the individual susceptibility or resilience to an exposure predicting the intensity of its effect on the individual. Gao et al. [233] propose a scheme with six subclasses of biomarkers based on their intended use, rather than the technology or outcomes. These include: i) food compound intake biomarkers (FCIBs); ii) food or food component intake biomarkers (FIBs); iii) dietary pattern biomarkers (DPBs); iv) food compound status biomarkers (FCSBs); v) effect biomarkers; and vi) physiological or health state biomarkers. Further initiatives to define biomarkers have been conducted elsewhere. In the context of nutritional metabolomic applications, the use of 'monitoring biomarker' was suggested by a FDA-NIH working group [234]. A monitoring biomarker was defined as "a biomarker measured serially for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) a medical product or an environmental agent". The FDA-NIH working group proposed and defined "diagnostic biomarkers, pharmacodynamic/response biomarkers, predictive and prognostic biomarkers, and safety and risk biomarkers". Taken together, the definition of the various terms

enveloping the concept of biomarkers is still dynamic and awaiting international harmonization. Confusion in terminology should nonetheless be avoided and an appropriate use of the terms of the 'biomarker' family across the life sciences requires from experts in food, diet, and nutrition to be precise in defining what is meant when using this generic term. In line with the recommendation of the FoodBAll consortium, the remaining of this section will use the term food intake biomarker (FIB) and effect biomarker [233].

FIBs are interesting from a nutritional point of view as they provide information on metabolites, or metabolite ratio, identified in human samples (eventually in animal samples) in response to acute or previous ingestion of distinct food constituents or certain food items. FIBs, however, do not undergo major biological modification, if any, in the mammalian system but clearly change in the biosamples in concentration over time in response to the ingestion of test foods. Such metabolites could also be called "flow-through entities" and their dynamic nature make them suboptimal as FIBs. FIBs may respond to dietary intake within hours (acute biomarkers) or time periods extending from days to years, a clear terminology defining medium-term from long-term biomarkers being not harmonized among nutritionists. A well-known example for a FIB is proline-betaine found in citrus fruits and products [235]. Like other betaine-derivatives, proline-betaine is an important osmolyte in the plant, is absorbed in the human intestine, and is excreted without any metabolism via the kidneys. An effect biomarker, in contrast to a pure FIB, would be a metabolite that is a product of a major biological process in the mammalian system associated with enzymatic/chemical modification. This will, in most cases and for almost all plant foods, be the transformation of a secondary plant metabolite by phase I and phase II processes, which result in a variety of conjugates, for example with sulfate or glucuronic acid (or both) for better excretion from the mammalian system.

10.2.   Markers of intake

Food metabolomics has a number of application areas that can be seen as game-changers in food and nutrition science. This applies, for example, to using metabolomics as a tool for food authenticity [236] or for following the transformation of a food as it moves through the food chain [10, 237, 238]. When it comes to the food-diet-health relationship in humans, science needs to assess food consumption by calculating nutrient or calorie intake; that is done mainly via dietary surveys such as food frequency questionnaires or 24-h recalls. Here, food metabolomics comes into play by providing qualitative insights on what happens to food constituents as they move through the human body. Of course, it would be fantastic if food metabolomics could provide quantitative

measures of consumption; which for selected food items seems feasible. What also appears to be reasonable as another achievement is to classify food intake patterns (diet styles) based on corresponding metabolite signatures. With a rapidly growing number of excellent studies that identify food-specific metabolite markers, there is the perspective that food metabolomics will transform the sciences relying on food consumption data. The FoodBall consortium will certainly have its share here [8, 233].

There are of course some intrinsic problems when using metabolite analysis to assess the response to intake of distinct food items. Most foods contain only tiny amounts of ingredients that qualify as markers since the bulk ingredients are water, carbohydrates, proteins, and fats. These macronutrients are either stored in glycogen or lipids in liver and fat tissue or are transformed into CO_2, H_2O, and some nitrogen and sulphur appearing in breath, urine, and feces. That leaves mainly minerals and trace elements, vitamins, and compounds from secondary metabolism that mammary species excrete after ingestion of food as being potential markers. Actually, minerals and trace elements can now easily be measured by means of inductively coupled plasma mass spectrometry (ICP-MS) and urine is the prime site of excretion. Unforunately only a few studies have used urinary mineral profiling for food intake marker discovery. When we talk about foods of animal origin, we have currently a limited number of markers of intake. This is not surprising since all mammals, including humans, produce and contain essentially the same nutrients and metabolites. Therefore it is a question of whether the consumption of a given animal product brings plasma levels and, in turn, levels in urine above the levels contained in blood and urine as a result of endogenous production in human metabolism [239]. In this respect it is important to look at the ratio of individual metabolites before and after food intake. However, at the same time, we need to know better under which conditions certain metabolites change in biofluids, independently of food intake. Some animal products have unique constituents such as anserine (ß-alanine-1-methylhistidine) that is found in very high levels in poultry and is not synthetized in human metabolism [240]. The related peptide in the human system and in pork and beef is carnosine (ß-alanine-histidine). Both dipeptides are quickly hydrolyzed in circulating blood and the constituent amino acids are found in blood and urine. To use these dipeptides and/or their degradation products for assessing intake the separate analysis of pi-methylhistidine (for anserine) and tau-methylhistidine (for carnosine) next to anserine and carnosine is required. Much simpler appears the use of lactose and the constituent monosaccharides galactose and fructose as markers of intake of dairy products. Although lactose can be synthetized in humans (at least in the lactating mammary gland for breast-feeding), the intake of lactose is usually so high that this disaccharide and its constituent monosaccharides increase

substantially in plasma and urine and can thus be classified as good markers [241]. What remains to be seen is why disaccharides (or even larger oligosaccharides) contained in food can cross into blood for excretion into urine. There is no known transport system that allows uptake of intact disaccharides in the intestine and more surprising is that those sugars are not completely hydrolyzed by the enzymes maltase and glucoamylase, sucrase, or even lactase. The latter has of course a huge difference in activity at the intestinal mucosa based on whether humans carry the mutations in the minichromosome-maintenance gene that determines whether an individual is lactose-tolerant or not and that depends on whether the mutations are found on just one or on both alleles [242]. The best explanation currently is that the disaccharides pass passively through the intestinal epithelium (in magnitude depending on the oral load of the sugar). This passive transport could be higher in some physiological conditions (i.e. stress) or states of impaired intestinal functions such as intestinal bowel syndromes, inflammatory bowel diseases, celiac disease, and others [242]. Once in blood, there are no enzymes in the extracellular space that can hydrolyze the disaccharides and they are thus filtered in kidney and appear in urine since they are also not reabsorbed in the renal system. Although not systematically assessed, human urine can contain unusual disaccharides and larger oligosaccharides since thousands of different hetero-oligosaccharides are synthesized in the human system and those are mainly bound to proteins passing through the secretory pathway. Although these glycosylated proteins are mainly degraded in intracellular compartments, some obviously escape complete hydrolysis, appearing in plasma and urine.

The plant kingdom is also represented by thousands of different organic acids, polyphenolic compounds, and terpenoids. These compounds have the advantage over all other metabolites with regard to their uniqueness either by individual compounds or their patterns of metabolites for a given plant or plant family (here called a lead compound). These entities are not synthesized in mammals and are in most cases are not degraded by any enzymes in the human system – except when they reach the large intestine and are subjected to bacterial fermentation with degradation products found in blood and urine. When looking at polyphenols as a subfamily of secondary plant metabolites, it is still not known by which mechanism they are absorbed in the gut, although there are marked differences in the kinetics of appearance in systemic circulation. Since numerous compounds are found as glycosides in the plant material, deglycosylation seems to be required for efficient uptake of the aglycone into intestinal cells. This may also require enzymatic processes provided by gut microbiota, which in turn leads to absorption from large intestine. This is in contrast to uptake in the small intestine - that usually brings compounds into plasma with peaks occurring at one to two hours after intake – which may need six to

eight hours. That means that the time of collection of samples after ingestion of a food needs to be long enough to cover full absorption of the test dose provided. For most of the polyphenols plasma levels rarely exceed 1 µmol/L at normal amounts of a test food [243]. Already in intestinal epithelial cells, but mostly in liver, lead compounds undergo – depending on their chemical nature –xenobiotic phase I and/or phase II transformation with conjugation to sulfate and/or glucuronic acid to increase water solubility for efficient renal excretion. For most polyphenols ingested by humans only metabolites are found in plasma with the lead compound not modified rarely exceeding 1 to 5% of the sum of all metabolites in plasma. In most cases clearance from the plasma also occurs quickly with negligible retention in kidney. There are, however, exceptions since some compounds show efficient binding to plasma proteins that increases their retention in the body drastically. This means a sufficient time frame for urine collection must be built into the study design to recover as much of the dose in urine as possible. Genetic heterogeneity in enzymes that mediate phase I and phase II processes, as well as differences in conjugation capacity, add another layer of complexity to urinary metabolite profiling of plant material in the mammalian system. Heterogeneity in biotransformation becomes even more of an issue as previous intake of plant material with "bioactives" can have pronounced effects on enzymes of xenobiotic metabolism. This means that the extent by which an administered food with given constituents is converted into metabolites and conjugates will vary depending on previous exposure to the same or different plant food. The most impressive examples for this phenomenon is the induction but also inhibition of cytochrome P450 oxidases such as CYP3A4 in the intestine and liver generated by a single dose (glass) of grapefruit juice and more so by repeated doses [244]. This issue has become a highly relevant research area since this interaction of plant food constituents with the human system have major effects of the metabolism of drugs that utilize the same pathways. A large number of drugs from almost all important medical indication areas are metabolized by CYP3A4 with the production of metabolites that usually have no or much lower activity than the parent compound. Changes in CYP3A4 activity therefore alter the plasma levels of the active compound drastically. This can lead to unwanted side effects or even increased toxicity [245]. Based on hundreds of human studies, measures were taken to inform users of the drugs aware of these grapefruit-drug interactions by advising them not to consume grapefruit juice when taking the drugs. Although this may only be the best indicator reaction, it can be anticipated that many more plant compounds can alter enzymes and transporters in xenobiotic metabolism; developed by nature to excrete these natural xenobiotics [246]. Such effects may need to be taken into account when marker metabolites are defined since the spectrum and quantity of the metabolite may be different, depending on previous exposure to plant material.

The gut microbiota with its huge diversity in species and in its metabolic capacity adds an extra layer of complexity to metabolite analysis. The latter is an emerging field and for only few compounds/compound classes has the metabolic activity of the gut microbiota been assessed as an important contributor to metabolite transformation and urinary metabolite profiles. The production of equol from daizein as a lead in soy and soy products is the paradigm [247], with only around 30-40% of the European population able to produce equol (that carries likely the relevant bioactivity) in their guts.

Next to the question of sensitivity and selectivity of a marker metabolite, a key question for food intake assessment is whether food metabolomic applications will be able to provide quantitative information on the amount of the food consumed. Studies published recently have provided the first evidence that quantification appears feasible with proline-betaine as a key compound for citrus intake [248] or anserine and methyl-histidines for estimation of chicken meat consumption [249]. Another important question is whether metabolite signatures can serve as a surrogate of a diet style; first studies suggest that this may be possible [172]. Of course, each food will have a different quantity of constituents depending on where and how it was produced, transported, processed, and consumed. Portion size and inter-individual or even intra-individual differences in metabolism brings in additional variability beyond other physiological parameters such as gastrointestinal processing and differences in absorption and renal clearance. That leaves researchers with a huge challenge on how quantitative information from food metabolomics can be expected for the majority of the ordinary foods consumed in free living individuals. Despite this limitation, patterns or signatures of metabolites that come with different types of diets comprising variable quantities of food of plant or animal origin will help epidemiologists to improve data quality when assessing the diet-health relationship.

10.3. Markers reflecting the health status

Although this research strategy focusing metabolomics on health status sounds appealing, it is almost impossible to define an individual's health status based on metabolite profiles of human biofluids. The best example for the difficulties associated with this is the reference value for a "healthy" plasma level of cholesterol (LDL-cholesterol) or a critical level for medical intervention. It is still controversial despite decades of discussions in different science communities. Although the World Health Organization (WHO) defined health in 1948, this definition is not applicable to any issue when metabolomic tools are being used to assess the health to disease transition. The current state-of-the-art is the use of huge cohorts of seemingly healthy humans to obtain

metabolite profiles in blood (plasma or serum) or urine or any other accessible biosample. The next step is to analyze in retrograde perspective whether any disease occurred in the population under study for the time of observation associated with alterations of individual or multiple metabolites observed in the biosample. It is interesting to note that, in hundreds of such studies, metabolite concentrations are almost exclusively provided as "relative levels" or fold-changes, despite the fact that metabolomics could measure real concentrations. Remarkably, any statistical analysis applied to cohorts comprising hundreds or thousands, now even tens of thousands, of people brings back differences for subgroups with disease or at risk of disease with impressive p-values despite little differences between the measured mean values. That challenges the proposal that metabolomics could be an important tool in personalizing medical treatment or for individualizing nutrition. If projected from a mega-cohort back onto the individual, the cohort-derived data are questionable in value. Yet, metabolomics as a new screening tool in human cohorts for association studies on disease markers and employing genetic, epigenetic, transcriptomic (usually by RNAseq), and proteomic applications has become a gold standard. Giving meaning to the data and providing an answer of what may have caused the changes in metabolite levels or patterns is still the biggest challenge. Unfortunately, it appears that more data and more data points will not necessarily increase our understanding of the underlying biology. The same holds true for the question of whether the marker metabolites are just reporters of a health to disease trajectory or also causative for the disease. There are hundreds of papers published with claims that disease-specific biomarkers have been identified in blood or urine profiles by non-targeted or targeted approaches. However, they all have in common that they measure in compartments that in most cases are not the origin of the metabolic perturbation. These are, of course, specialized cells in specialized organs and based on organ mass, adipose tissue, muscle, and liver may dominate the metabolite levels in plasma. But plasma is extracellular space and extra- and intracellular compartments are separated by cell membranes with low intrinsic permeability. However, a huge diversity of proteins transport in unidirectional or bidirectional manner literally every metabolite into or out of the cell. Based on these ion gradients cells maintain distinct metabolite levels and patterns and those can be different by one to two orders of magnitude in concentration across a plasma cell membrane. The same holds true for the renal epithelium. Although all low molecular compounds are filtered by the glomerulum in the kidney and reach the tubular system, hundreds of transport proteins in renal tubular cells are responsible for reabsorption of nutrients/metabolites or are involved in secreting compounds across the renal epithelium into urine. This also creates huge concentration differences and different metabolite patterns when comparing plasma and urinary samples in the same volunteer. Although a bit speculative, the (genetic) make-up of renal transporters and their

expression levels seem to provide or at least contribute to an individual fingerprint of metabolite patterns as shown for urine samples in which individuals are easily identified based on few discriminating amino acids [250]. Most interestingly, there is no correlation between these amino acids measured in plasma and urine in the same humans.

Metabolomic approaches to identify metabolites that associate with the development of type 2 diabetes in humans, used here as an example, have identified around 50 to 80 metabolites that characterize an insulin-resistant state or fully developed diabetes. Surprisingly, most of the metabolites are amino acids and intermediates from their degradation pathways followed by lipids, mainly from the phospholipid/lysophosphatide subclasses [251]. Although diabetes is considered to be a metabolic disease with relevance to carbohydrate metabolism, so far only few sugars and metabolites derived from these substrates have been reported as markers. This also sheds some light on the bias that the methods bring into these discovery approaches. Analytes that can easily be measured are easily measured on whatever platform and thus are frequently overrepresented in the ranking of disease-specific markers that may finally turn out not to be so specific. It is also not too surprising that a number of these marker-metabolites have been known for decades as changed in states of obesity, insulin resistance, or the metabolic syndrome. What is disappointing is the fact that the diabetes-specific marker-metabolites identified so far do not provide a significant improvement in the diagnosis of type 2 diabetes when compared to the classical diagnostic measures such as fasting glucose level, a family history, or a calculated diabetes risk score [252, 253].

In summary, metabolic profiling approaches used in food research for identifying markers of dietary exposure or used as discovery tools to identify disease-specific markers are still in the early developmental phase and that leaves hope that they could really turn out to be game-changers. What we need is to better harmonize the use of the term "biomarker" [233]. We also need more experts in food, diet, and health to give meaning to the wealth of data generated by high throughput and high density technologies. While the science of classical physiology and biochemistry of intermediate metabolism was largely left behind when genetics and the genomics revolution arrived, it is the hope of the authors of this document that metabolomics may be able to bring back the interest in and attraction to study metabolism. This would certainly give greater meaning and provide more sound biological interpretation of the analytical data that can now be measured.

References

[1]   S. A. Sansone, P. Rocca-Serra, D. Field, E. Maguire, C. Taylor, O. Hofmann, H. Fang, S. Neumann, W. Tong, L. Amaral-Zettler, K. Begley, T. Booth, L. Bougueleret, G. Burns, B. Chapman, T. Clark, L. A. Coleman, J. Copeland, S. Das, A. De Daruvar, P. De Matos, I. Dix, S. Edmunds, C. T. Evelo, M. J. Forster, P. Gaudet, J. Gilbert, C. Goble, J. L. Griffin, D. Jacob, J. Kleinjans, L. Harland, K. Haug, H. Hermjakob, S. J. Ho Sui, A. Laederach, S. Liang, S. Marshall, A. Mcgrath, E. Merrill, D. Reilly, M. Roux, C. E. Shamu, C. A. Shang, C. Steinbeck, A. Trefethen, B. Williams-Jones, K. Wolstencroft, I. Xenarios, W. Hide, *Nat Genet* **2012**, 44, 121.

[2]   D. R. Fiehn Oliver , Jules Griffin, Mariet Van Der Werf, Basil Nikolau, Norman Morrison, Lloyd W. Sumner, Roy Goodacre, Nigel W. Hardy, Chris Taylor, Jennifer Fostel, Bruce Kristal, Rima Kaddurah-Daouk, Pedro Mendes, Ben Van Ommen, John C. Lindon & Susanna-Assunta Sansone, *Metabolomics* **2007**, 3.

[3]   M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. T Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Van Schaik, S. A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, *Sci Data* **2016**, 3, 160018.

[4]   S. A. Sansone, T. Fan, R. Goodacre, J. L. Griffin, N. W. Hardy, R. Kaddurah-Daouk, B. S. Kristal, J. Lindon, P. Mendes, N. Morrison, B. Nikolau, D. Robertson, L. W. Sumner, C. Taylor, M. Van Der Werf, B. Van Ommen, O. Fiehn, *Nat Biotechnol* **2007**, 25, 846.

[5]   J. C. Martin, M. Maillot, G. Mazerolles, A. Verdu, B. Lyan, C. Migne, C. Defoort, C. Canlet, C. Junot, C. Guillou, C. Manach, D. Jabob, D. J. Bouveresse, E. Paris, E. Pujos-Guillot, F. Jourdan, F. Giacomoni, F. Courant, G. Fave, G. Le Gall, H. Chassaigne, J. C. Tabet, J. F. Martin, J. P. Antignac, L. Shintu, M. Defernez, M. Philo, M. C. Alexandre-Gouaubau, M. J. Amiot-Carlin, M. Bossis, M. N. Triba, N. Stojilkovic, N. Banzet, R. Molinie, R. Bott, S. Goulitquer, S. Caldarelli, D. N. Rutledge, *Metabolomics* **2015**, 11, 807.

[6]   M. Van Rijswijk, C. Beirnaert, C. Caron, M. Cascante, V. Dominguez, W. B. Dunn, T. M. D. Ebbels, F. Giacomoni, A. Gonzalez-Beltran, T. Hankemeier, K. Haug, J. L. Izquierdo-Garcia, R. C. Jimenez, F. Jourdan, N. Kale, M. I. Klapa, O. Kohlbacher, K. Koort, K. Kultima, G. Le Corguille, P. Moreno, N. K. Moschonas, S. Neumann, C. O'donovan, M. Reczko, P. Rocca-Serra, A. Rosato, R. M. Salek, S. A. Sansone, V. Satagopam, D. Schober, R.

Shimmo, R. A. Spicer, O. Spjuth, E. A. Thevenot, M. R. Viant, R. J. M. Weber, E. L. Willighagen, G. Zanetti, C. Steinbeck, *F1000Res* **2017**, 6.

[7]   K. Haug, R. M. Salek, P. Conesa, J. Hastings, P. De Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. Gonzalez-Beltran, S. A. Sansone, J. L. Griffin, C. Steinbeck, *Nucleic Acids Res* **2013**, 41, D781.

[8]   E. M. Brouwer-Brolsma, L. Brennan, C. A. Drevon, H. Van Kranen, C. Manach, L. O. Dragsted, H. M. Roche, C. Andres-Lacueva, S. J. L. Bakker, J. Bouwman, F. Capozzi, S. De Saeger, T. E. Gundersen, M. Kolehmainen, S. E. Kulling, R. Landberg, J. Linseisen, F. Mattivi, R. P. Mensink, C. Scaccini, T. Skurk, I. Tetens, G. Vergeres, D. S. Wishart, A. Scalbert, E. J. M. Feskens, *Proc Nutr Soc* **2017**, 76, 619.

[9]   S. D. Johanningsmeier, G. K. Harris, C. M. Klevorn, *Annu Rev Food Sci Technol* **2016**, 7, 413.

[10]   S. Kim, J. Kim, E. J. Yun, K. H. Kim, *Curr Opin Biotechnol* **2016**, 37, 16.

[11]   J. Rubert, M. Zachariasova, J. Hajslova, *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **2015**, 32, 1685.

[12]   J. M. Cevallos-Cevallos, J. I. Reyes-De-Corcuera, *Adv Food Nutr Res* **2012**, 67, 1.

[13]   M. Bouhifd, R. Beger, T. Flynn, L. Guo, G. Harris, H. Hogberg, R. Kaddurah-Daouk, H. Kamp, A. Kleensang, A. Maertens, S. Odwin-Dacosta, D. Pamies, D. Robertson, L. Smirnova, J. Sun, L. Zhao, T. Hartung, *ALTEX* **2015**, 32, 319.

[14]   C. Boushey, J. Harris, B. Bruemmer, S. L. Archer, L. Van Horn, *J Am Diet Assoc* **2006**, 106, 89.

[15]   J. V. Woodside, B. V. Koletzko, C. C. Patterson, R. W. Welch, *World Rev Nutr Diet* **2013**, 108, 18.

[16]   C. J. Boushey, J. Harris, B. Bruemmer, S. L. Archer, *J Am Diet Assoc* **2008**, 108, 679.

[17]   F. M. Van Der Kloet, Ivana Bobeldijk, Elwin R. Verheij, R. H. Jellema, *Journal of proteome research* **2009**, 8.

[18]   O. Fliniaux, G. Gaillard, A. Lion, D. Cailleu, F. Mesnard, F. Betsou, *J Biomol NMR* **2011**, 51, 457.

[19]   N. Keum, D. H. Lee, N. Marchand, H. Oh, H. Liu, D. Aune, D. C. Greenwood, E. L. Giovannucci, *Br J Nutr* **2015**, 114, 1099.

[20]   L. Schwingshackl, B. Missbach, J. Konig, G. Hoffmann, *Public Health Nutr* **2015**, 18, 1292.

[21]   E. Ros, M. A. Martinez-Gonzalez, R. Estruch, J. Salas-Salvado, M. Fito, J. A. Martinez, D. Corella, *Adv Nutr* **2014**, 5, 330s.

[22]   M. B. Andersen, A. Rinnan, C. Manach, S. K. Poulsen, E. Pujos-Guillot, T. M. Larsen, A. Astrup, L. O. Dragsted, *J Proteome Res* **2014**, 13, 1405.

[23]   J. Wittwer, I. Rubio-Aliaga, B. Hoeft, I. Bendik, P. Weber, H. Daniel, *Mol Nutr Food Res* **2011**, 55, 341.

[24]   D. H. May, S. L. Navarro, I. Ruczinski, J. Hogan, Y. Ogata, Y. Schwarz, L. Levy, T. Holzman, M. W. Mcintosh, J. W. Lampe, *Br J Nutr* **2013**, 110, 1760.

[25]   F. Devlieghere, L. Vermeiren, J. Debevere, *International Dairy Journal* **2004**, 14, 273.

[26]   M. S. Rahman, *Handbook of food preservation*, CRC Press 2008.

[27]   S. Osborn, W. Morley, *Developing food products for consumers with specific dietary needs* Elsevier 2016.

[28]   N. Garti, *Delivery and controlled release of bioactives in foods and nutraceuticals*, Elsevier 2008.

[29]   P. B. Ottaway, *Food fortification and supplementation*, 2008.

[30]   K. N. Avery, P. R. Williamson, C. Gamble, E. O'connell Francischetto, C. Metcalfe, P. Davidson, H. Williams, J. M. Blazeby, R. Members of the Internal Pilot Trials Workshop Supported by the Hubs for Trials Methodology, *BMJ Open* **2017**, 7, e013537.

[31]   J. Lovegrove, *Nutrition Research Methodologies*, Wiley-Blackwell 2015.

[32]   A. H. Emwas, C. Luchinat, P. Turano, L. Tenori, R. Roy, R. M. Salek, D. Ryan, J. S. Merzaban, R. Kaddurah-Daouk, A. C. Zeri, G. A. Nagana Gowda, D. Raftery, Y. Wang, L. Brennan, D. S. Wishart, *Metabolomics* **2015**, 11, 872.

[33]   G. Fave, M. Beckmann, A. J. Lloyd, S. Zhou, G. Harold, W. Lin, K. Tailliart, L. Xie, J. Draper, J. C. Mathers, *Metabolomics* **2011**, 7, 469.

[34]   L. G. Rasmussen, F. Savorani, T. M. Larsen, L. O. Dragsted, A. Astrup, S. B. Engelsen, *Metabolomics* **2010**, 7, 71.

[35]   R. W. Welch, J. M. Antoine, J. L. Berta, A. Bub, J. De Vries, F. Guarner, O. Hasselwander, H. Hendriks, M. Jakel, B. V. Koletzko, C. C. Patterson, M. Richelle, M. Skarp, S. Theis, S. Vidry, J. V. Woodside, F. International Life Sciences Institute Europe Functional Foods Task, *Br J Nutr* **2011**, 106 Suppl 2, S3.

[36]   J. H. Winnike, M. G. Busby, P. B. Watkins, T. M. O'connell, *Am J Clin Nutr* **2009**, 90, 1496.

[37]   S. Krug, G. Kastenmuller, F. Stuckler, M. J. Rist, T. Skurk, M. Sailer, J. Raffler, W. Romisch-Margl, J. Adamski, C. Prehn, T. Frank, K. H. Engel, T. Hofmann, B. Luy, R. Zimmermann, F. Moritz, P. Schmitt-Kopplin, J. Krumsiek, W. Kremer, F. Huber, U. Oeh, F. J. Theis, W. Szymczak, H. Hauner, K. Suhre, H. Daniel, *FASEB J* **2012**, 26, 2607.

[38] L. G. Rasmussen, H. Winning, F. Savorani, C. Ritz, S. B. Engelsen, A. Astrup, T. M. Larsen, L. O. Dragsted, *Genes Nutr* **2012**, 7, 281.

[39] A. Riedl, C. Gieger, H. Hauner, H. Daniel, J. Linseisen, *Br J Nutr* **2017**, 117, 1631.

[40] G. Nyamundanda, I. C. Gormley, Y. Fan, W. M. Gallagher, L. Brennan, *BMC Bioinformatics* **2013**, 14, 338.

[41] D. Trutschel, S. Schmidt, I. Grosse, S. Neumann, *Front Bioeng Biotechnol* **2015**, 3, 129.

[42] B. J. Blaise, G. Correia, A. Tin, J. H. Young, A. C. Vergnaud, M. Lewis, J. T. Pearce, P. Elliott, J. K. Nicholson, E. Holmes, T. M. Ebbels, *Anal Chem* **2016**, 88, 5179.

[43] J. Xia, I. V. Sinelnikov, B. Han, D. S. Wishart, *Nucleic Acids Res* **2015**, 43, W251.

[44] K. F. Schulz, D. G. Altman, D. Moher, D. Fergusson, *Lancet* **2010**, 375, 1144.

[45] D. Moher, S. Hopewell, K. F. Schulz, V. Montori, P. C. Gotzsche, P. J. Devereaux, D. Elbourne, M. Egger, D. G. Altman, *Bmj* **2010**, 340, c869.

[46] V. W. Madurasinghe, M. R. C. S. G. Sandra Eldridge on Behalf Of, S. E. C. G. Gordon Forbes on Behalf of The, *Trials* **2016**, 17, 27.

[47] A. Bub, A. Kriebel, C. Dorr, S. Bandt, M. Rist, A. Roth, E. Hummel, S. Kulling, I. Hoffmann, B. Watzl, *JMIR Res Protoc* **2016**, 5, e146.

[48] M. C. Walsh, L. Brennan, J. P. Malthouse, H. M. Roche, M. J. Gibney, *Am J Clin Nutr* **2006**, 84, 531.

[49] G. B. Proctor, P. André, E. Lopez-Garcia, D. Gomez Cabrero Lopez, E. Neyraud, C. Feart, F. Rodriguez Artalejo, E. Garcia-Esquinas, M. Morzel, *Nutrition Bulletin* **2017**, 42, 369.

[50] J. Taucher, A. Hansel, A. Jordan, W. Lindinger, *J Agric Food Chem* **1996**, 44, 3778.

[51] L. D. Lawson, Z. J. Wang, *J Agric Food Chem* **2005**, 53, 1974.

[52] D. C. Wedge, J. W. Allwood, W. Dunn, A. A. Vaughan, K. Simpson, M. Brown, L. Priest, F. H. Blackhall, A. D. Whetton, C. Dive, R. Goodacre, *Anal Chem* **2011**, 83, 6689.

[53] O. Teahan, S. Gamble, E. Holmes, Jonathan Waxman, J. K. Nicholson, C. Bevan, H. C. Keun, *Analytical Chemistry* **2006**, 78.

[54] L. Liu, J. Aa, G. Wang, B. Yan, Y. Zhang, X. Wang, C. Zhao, B. Cao, J. Shi, M. Li, T. Zheng, Y. Zheng, G. Hao, F. Zhou, J. Sun, Z. Wu, *Anal Biochem* **2010**, 406, 105.

[55]   Z. Yu, G. Kastenmuller, Y. He, P. Belcredi, G. Moller, C. Prehn, J. Mendes, S. Wahl, W. Roemisch-Margl, U.
        Ceglarek, A. Polonikov, N. Dahmen, H. Prokisch, L. Xie, Y. Li, H. E. Wichmann, A. Peters, F. Kronenberg, K.
        Suhre, J. Adamski, T. Illig, R. Wang-Sattler, *PLoS One* **2011**, 6, e21230.

[56]   T. Barri, L. O. Dragsted, *Anal Chim Acta* **2013**, 768, 118.

[57]   J. R. Denery, A. A. Nunes, T. J. Dickerson, *Anal Chem* **2011**, 83, 1040.

[58]   E. J. Want, I. D. Wilson, H. Gika, G. Theodoridis, R. S. Plumb, J. Shockcor, E. Holmes, J. K. Nicholson, *Nat Protoc*
        2010, 5, 1005.

[59]   B. Álvarez-Sánchez, F. Priego-Capote, M. D. Luque De Castro, *TrAC Trends in Analytical Chemistry* **2010**, 29, 111.

[60]   B. Álvarez-Sánchez, F. Priego-Capote, M. D. L. D. Castro, *TrAC Trends in Analytical Chemistry* **2010**, 29, 120.

[61]   S. R. Khan, Glenton, P. A., Backov, R., & Talham, D. R, *Kidney International* **2002**, 62.

[62]   A. Jiye, Q. Huang, G. Wang, W. Zha, B. Yan, H. Ren, S. Gu, Y. Zhang, Q. Zhang, F. Shao, L. Sheng, J. Sun,
        Analytical Biochemistry **2008**, 379, 20.

[63]   J. Vanden Bussche, M. Marzorati, D. Laukens, L. Vanhaecke, *Anal Chem* **2015**, 87, 10927.

[64]   A. Coppens, M. Speeckaert, J. Delanghe, *Acta Clinica Belgica* **2010**, 65.

[65]   J. Godzien, V. Alonso-Herranz, C. Barbas, A. E. Grace, *Metabolomics* **2015**, 11, 518.

[66]   P. Yin, R. Lehmann, G. Xu, *Anal Bioanal Chem* **2015**, 407, 4879.

[67]   J. Godzien, V. Alonso-Herranz, C. Barbas, E. G. Armitage, *Metabolomics* **2014**, 11, 518.

[68]   N. C. Van De Merbel, *TrAC Trends in Analytical Chemistry* **2008**, 27, 924.

[69]   A. Gil, D. Siegel, H. Permentier, D. J. Reijngoud, F. Dekker, R. Bischoff, *Electrophoresis* **2015**.

[70]   D. Siegel, H. Permentier, D. J. Reijngoud, R. Bischoff, *J Chromatogr B Analyt Technol Biomed Life Sci* **2014**, 966,
        21.

[71]   Deprez Stephanie, B. C. Sweatman, S. C. Connor, J. N. Haselden, C. J. Waterfield, *Journal of Pharmaceutical and
        Biomedical Analysis* **2002**, 30.

[72]   M. J. Rist, C. Muhle-Goll, B. Gorling, A. Bub, S. Heissler, B. Watzl, B. Luy, *Metabolites* **2013**, 3, 243.

[73]   M. A. Fernández-Peralbo, M. D. Luque De Castro, *TrAC Trends in Analytical Chemistry* **2012**, 41, 75.

[74]   B. Boyanton, K. Blick, *Clinical Chemistry* **2002**, 48, 5.

[75]   B. Kamlage, S. G. Maldonado, B. Bethan, E. Peter, O. Schmitz, V. Liebenberg, P. Schatz, *Clin Chem* **2014**, 60, 399.

Accepted Article

[76] P. Bernini, I. Bertini, C. Luchinat, P. Nincheri, S. Staderini, P. Turano, *J Biomol NMR* **2011**, 49, 231.

[77] L. Bervoets, E. Louis, G. Reekmans, L. Mesotten, M. Thomeer, P. Adriaensens, L. Linsen, *Metabolomics* **2015**, 11, 1197.

[78] E. Jobard, O. Tredan, D. Postoly, F. Andre, A. L. Martin, B. Elena-Herrmann, S. Boyault, *Int J Mol Sci* **2016**, 17.

[79] J. A. Yazigi Junior, J. B. Dos Santos, B. R. Xavier, M. Fernandes, S. G. Valente, V. M. Leite, *Rev Bras Ortop* **2015**, 50, 729.

[80] A. G. Perez, J. F. Lana, A. A. Rodrigues, A. C. Luzo, W. D. Belangero, M. H. Santana, *ISRN Hematol* **2014**, 2014, 176060.

[81] D. Lesche, R. Geyer, D. Lienhard, C. T. Nakas, G. Diserens, P. Vermathen, A. B. Leichtle, *Metabolomics* **2016**, 12, 159.

[82] J. A. Kirwan, L. Brennan, D. Broadhurst, O. Fiehn, M. Cascante, W. B. Dunn, M. A. Schmidt, V. Velagapudi, *Clin Chem* **2018**.

[83] M. K. Townsend, Y. Bao, E. M. Poole, K. A. Bertrand, P. Kraft, B. M. Wolpin, C. B. Clish, S. S. Tworoger, *Cancer Epidemiol Biomarkers Prev* **2016**, 25, 823.

[84] L. Malm, G. Tybring, T. Moritz, B. Landin, J. Galli, *Biopreserv Biobank* **2016**, 14, 416.

[85] M. Jain, A. D. Kennedy, S. H. Elsea, M. J. Miller, *Clin Chim Acta* **2017**, 466, 105.

[86] J. P. Trezzi, A. Bulla, C. Bellora, M. Rose, P. Lescuyer, M. Kiehntopf, K. Hiller, F. Betsou, *Metabolomics* **2016**, 12, 96.

[87] V. Gonzalez-Covarrubias, A. Dane, T. Hankemeier, R. J. Vreeken, *Metabolomics* **2012**, 9, 337.

[88] H. Mei, Y. Hsieh, C. Nardo, X. Xu, S. Wang, K. Ng, W. A. Korfmacher, *Rapid Commun Mass Spectrom* **2003**, 17, 97.

[89] B. Jørgenrud, S. Jäntti, I. Mattila, P. Paivi, K. S. Rønningen, Y.-J. Hannele, O. Matej, T. Hyötyläinen, *Bioanalysis* **2015**, 7.

[90] N. M. Denihan, B. H. Walsh, S. N. Reinke, B. D. Sykes, R. Mandal, D. S. Wishart, D. I. Broadhurst, G. B. Boylan, D. M. Murray, *Clin Biochem* **2015**, 48, 534.

[91] Leon M. Smith, Anthony D. Maher, Elizabeth J. Want, Paul Elliott, Jeremiah Stamler, G. E. Hawkes, Elaine Holmes, J. C. Lindon, J. K. Nicholson, *Analytical Chemistry* **2009**, 81.

[92] E. J. Saude, B. D. Sykes, *Metabolomics* **2007**, 3, 19.

[93]	M. Lauridsen, S. H. Hansen, J. W. Jaroszewski, C. Cornett, *Analytical Chemistry* **2007**, 79.

[94]	A. Roux, E. A. Thevenot, F. Seguin, M. F. Olivier, C. Junot, *Metabolomics* **2015**, 11, 1095.

[95]	M. Yuille, T. Illig, K. Hveem, G. Schmitz, J. Hansen, M. Neumaier, G. Tybring, E. Wichmann, B. Ollier, *Biopreserv Biobank* **2010**, 8, 65.

[96]	O. Deda, H. G. Gika, I. D. Wilson, G. A. Theodoridis, *J Pharm Biomed Anal* **2015**, 113, 137.

[97]	S. Matysik, C. I. Le Roy, G. Liebisch, S. P. Claus, *Trends in Food Science & Technology* **2016**, 57, 244.

[98]	J. Gratton, J. Phetcharaburanin, B. H. Mullish, H. R. Williams, M. Thursz, J. K. Nicholson, E. Holmes, J. R. Marchesi, J. V. Li, *Anal Chem* **2016**, 88, 4661.

[99]	A. Santiago, S. Panda, G. Mengels, X. Martinez, F. Azpiroz, J. Dore, F. Guarner, C. Manichanh, *BMC Microbiology* **2014**, 14.

[100]	M. A. Gorzelak, S. K. Gill, N. Tasnim, Z. Ahmadi-Vand, M. Jay, D. L. Gibson, *PLoS One* **2015**, 10, e0134802.

[101]	J. Kumar, M. Kumar, S. Gupta, V. Ahmed, M. Bhambi, R. Pandey, N. S. Chauhan, *Genomics Proteomics Bioinformatics* **2016**, 14, 371.

[102]	J. Saric, Y. Wang, J. Li, M. Coen, J. Utzinger, J. R. Marchesi, J. Keiser, K. Veselkov, J. C. Lindon, J. K. Nicholson, E. Holmes, *Journal of proteome research* **2008**.

[103]	A. Klinder, P. C. Karlsson, Y. Clune, R. Hughes, M. Glei, J. J. Rafter, I. Rowland, J. K. Collins, B. L. Pool-Zobel, *Nutr Cancer* **2007**, 57, 158.

[104]	C. L. Lauber, N. Zhou, J. I. Gordon, R. Knight, N. Fierer, *FEMS Microbiol Lett* **2010**, 307, 80.

[105]	E. Loftfield, E. Vogtmann, J. N. Sampson, S. C. Moore, H. Nelson, R. Knight, N. Chia, R. Sinha, *Cancer Epidemiol Biomarkers Prev* **2016**, 25, 1483.

[106]	R. Sinha, J. Chen, A. Amir, E. Vogtmann, J. Shi, K. S. Inman, R. Flores, J. Sampson, R. Knight, N. Chia, *Cancer Epidemiol Biomarkers Prev* **2016**, 25, 407.

[107]	S. J. Song, A. Amir, J. L. Metcalf, K. R. Amato, Z. Z. Xu, G. Humphrey, R. Knight, *mSystems* **2016**, 1.

[108]	G. Le Gall, S. O. Noor, K. Ridgway, L. Scovell, C. Jamieson, I. T. Johnson, I. J. Colquhoun, E. K. Kemsley, A. Narbad, *J Proteome Res* **2011**, 10, 4208.

[109]	D. Monleon, J. M. Morales, A. Barrasa, J. A. Lopez, C. Vazquez, B. Celda, *NMR Biomed* **2009**, 22, 342.

[110]	K. E. Gregory, S. S. Bird, V. S. Gross, V. R. Marur, A. V. Lazarev, W. A. Walker, B. S. Kristal, *Anal Chem* **2013**, 85, 1114.

[111] L. Van Meulebroek, E. De Paepe, V. Vercruysse, B. Pomian, S. Bos, B. Lapauw, L. Vanhaecke, *Anal Chem* **2017**, 89, 12502.

[112] T. Bezabeh, R. Somorjai, B. Dolenko, N. Bryskina, B. Levin, C. N. Bernstein, E. Jeyarajah, A. H. Steinhart, D. T. Rubin, I. C. Smith, *NMR Biomed* **2009**, 22, 593.

[113] H. G. Gika, G. A. Theodoridis, I. D. Wilson, *J Sep Sci* **2008**, 31, 1598.

[114] K. Budde, O. N. Gok, M. Pietzner, C. Meisinger, M. Leitzmann, M. Nauck, A. Kottgen, N. Friedrich, *Arch Biochem Biophys* **2016**, 589, 10.

[115] J. Laparre, Z. Kaabia, M. Mooney, T. Buckley, M. Sherry, B. Le Bizec, G. Dervilly-Pinel, *Anal Chim Acta* **2017**, 951, 99.

[116] T. Zivkovic Semren, I. Brcic Karaconji, T. Safner, N. Brajenovic, B. Tariba Lovakovic, A. Pizent, *Talanta* **2018**, 176, 537.

[117] V. A. Clarke, H. Lovegrove, A. Williams, M. Machperson, *J Behav Med* **2000**, 23, 367.

[118] S. D. Clarke, S. K. Kim, *J Nutr* **1998**, 128, 2036.

[119] J. Pinto, M. R. Domingues, E. Galhano, C. Pita, C. Almeida Mdo, I. M. Carreira, A. M. Gil, *Analyst* **2014**, 139, 1168.

[120] W. Yang, Y. Chen, C. Xi, R. Zhang, Y. Song, Q. Zhan, X. Bi, Z. Abliz, *Anal Chem* **2013**, 85, 2606.

[121] G. Anton, R. Wilson, Z. H. Yu, C. Prehn, S. Zukunft, J. Adamski, M. Heier, C. Meisinger, W. Romisch-Margl, R. Wang-Sattler, K. Hveem, B. Wolfenbuttel, A. Peters, G. Kastenmuller, M. Waldenberger, *PLoS One* **2015**, 10, e0121495.

[122] T. Moriya, Y. Satomi, H. Kobayashi, *Metabolomics* **2016**, 12.

[123] S. Cuhadar, M. Koseoglu, A. Atay, A. Dirican, *Biochemia Medica* **2013**, 70.

[124] A. M. Zivkovic, M. M. Wiest, U. T. Nguyen, R. Davis, S. M. Watkins, J. B. German, *Metabolomics* **2009**, 5, 507.

[125] M. Haid, C. Muschet, S. Wahl, W. Romisch-Margl, C. Prehn, G. Moller, J. Adamski, *J Proteome Res* **2017**.

[126] J. E. Lee, Y. Y. Kim, *OMICS* **2017**, 21, 499.

[127] D. G. Hebels, P. Georgiadis, H. C. Keun, T. J. Athersuch, P. Vineis, R. Vermeulen, L. Portengen, I. A. Bergdahl, G. Hallmans, D. Palli, B. Bendinelli, V. Krogh, R. Tumino, C. Sacerdote, S. Panico, J. C. Kleinjans, T. M. De Kok, M. T. Smith, S. A. Kyrtopoulos, C. Envirogenomarkers Project, *Environ Health Perspect* **2013**, 121, 480.

[128] J. T. Wood, J. S. Williams, L. Pandarinathan, A. Courville, M. R. Keplinger, D. R. Janero, P. Vouros, A. Makriyannis, C. J. Lammi-Keefe, *Clin Chem Lab Med* **2008**, 46, 1289.

[129] P. Yin, A. Peter, H. Franken, X. Zhao, S. S. Neukamm, L. Rosenbaum, M. Lucio, A. Zell, H. U. Haring, G. Xu, R. Lehmann, *Clin Chem* **2013**, 59, 833.

[130] W. B. Dunn, D. Broadhurst, D. I. Ellis, M. Brown, A. Halsall, S. O'hagan, I. Spasic, A. Tseng, D. B. Kell, *International Journal of Epidemiology* **2008**, 37, i23.

[131] W. Dunn, I. Wilson, A. Nicholls, D. Broadhurst, *Bioanalysis* **2012**, 4.

[132] W. B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-Mcintyre, N. Anderson, M. Brown, J. D. Knowles, A. Halsall, J. N. Haselden, A. W. Nicholls, I. D. Wilson, D. B. Kell, R. Goodacre, *Nat Protoc* **2011**, 6, 1060.

[133] T. Sangster, H. Major, R. Plumb, A. J. Wilson, I. D. Wilson, *Analyst* **2006**, 131, 1075.

[134] C. Brunius, L. Shi, R. Landberg, *Metabolomics* **2016**, 12, 173.

[135] F. Fernandez-Albert, Rafael Llorach, Mar Garcia-Aloy, Andrey Ziyatdinov, Cristina Andres-Lacueva, A. Perera, *Bioinformatics* **2014**, 30.

[136] M. A. Kamleh, T. M. Ebbels, K. Spagou, P. Masson, E. J. Want, *Anal Chem* **2012**, 84, 2670.

[137] J. A. Kirwan, D. I. Broadhurst, R. L. Davidson, M. R. Viant, *Anal Bioanal Chem* **2013**, 405, 5147.

[138] M. R. Nezami Ranjbar, Y. Zhao, M. G. Tadesse, Y. Wang, H. W. Ressom, *Proteome Sci* **2013**, 11, S13.

[139] R. Wehrens, J. A. Hageman, F. Van Eeuwijk, R. Kooke, P. J. Flood, E. Wijnker, J. J. Keurentjes, A. Lommen, H. D. Van Eekelen, R. D. Hall, R. Mumm, R. C. De Vos, *Metabolomics* **2016**, 12, 88.

[140] P. Yin, G. Xu, *J Chromatogr A* **2014**, 1374, 1.

[141] J. Zhou, Y. Yin, *Analyst* **2016**, 141, 6362.

[142] P. Yin, H. Chen, X. Liu, Q. Wang, Y. Jiang, R. Pan, *Analytical Letters* **2014**, 47, 1579.

[143] J. R. Sheedy, P. R. Ebeling, P. R. Gooley, M. J. Mcconville, *Analytical biochemistry* **2010**, 398, 263.

[144] M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, O. Yanes, *TrAC Trends in Analytical Chemistry* **2016**, 78, 23.

[145] B. M. Brenner, F. C. Rector, *Brenner & Rector's the Kidney*, Elsevier/Saunders 2012.

[146] Y. Tsuchiya, Y. Takahashi, T. Jindo, K. Furuhama, K. T. Suzuki, *European Journal of Pharmacology* **2003**, 475, 119.

[147] M. F. Boeniger, L. K. Lowry, J. Rosenberg, *Am Ind Hyg Assoc J* **1993**, 54, 615.

[148] P. Jatlow, S. Mckee, S. S. O'malley, *Clin Chem* **2003**, 49, 1932.

[149] A. H. Garde, A. M. Hansen, J. Kristiansen, L. E. Knudsen, *Ann Occup Hyg* **2004**, 48, 171.

[150] W. M. B. Edmands, P. Ferrari, A. Scalbert, *Analytical Chemistry* **2014**, 86, 10925.

[151] A. J. Chetwynd, A. Abdul-Sada, S. G. Holt, E. M. Hill, *Journal of Chromatography A* **2016**, 1431, 103.

[152] M. J. Rist, A. Roth, L. Frommherz, C. H. Weinert, R. Kruger, B. Merz, D. Bunzel, C. Mack, B. Egert, A. Bub, B. Gorling, P. Tzvetkova, B. Luy, I. Hoffmann, S. E. Kulling, B. Watzl, *PLoS One* **2017**, 12, e0183228.

[153] C. H. Weinert, M. T. Empl, R. Kruger, L. Frommherz, B. Egert, P. Steinberg, S. E. Kulling, *Mol Nutr Food Res* **2017**, 61.

[154] C. H. Weinert, B. Egert, S. E. Kulling, *J Chromatogr A* **2015**, 1405, 156.

[155] I. D. Wilson, R. Plumb, J. Granger, H. Major, R. Williams, E. M. Lenz, *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, 817, 67.

[156] D. L. Callahan, D. De Souza, A. Bacic, U. Roessner, *J Sep Sci* **2009**, 32, 2273.

[157] J. Lawson, *Design and analysis of experiments with R*, Chapman and Hall/CRC 2014.

[158] E. Saccenti, H. C. J. Hoefsloot, A. K. Smilde, J. A. Westerhuis, M. M. W. B. Hendriks, *Metabolomics* **2014**, 10, 361.

[159] P. Filzmoser, R. Maronna, M. Werner, *Computational Statistics & Data Analysis* **2008**, 52, 1694.

[160] S. Riccadonna, P. Franceschi, in: Komsta, Ł., Vander Heyden, Y., Sherma, J. (Eds.), *Chemometrics in chromatography*, CRC Press 2018.

[161] R. G. Brereton, *Chemometrics for pattern recognition*, John Wiley & Sons, Ltd 2009.

[162] M. B. Andersen, M. Kristensen, C. Manach, E. Pujos-Guillot, S. K. Poulsen, T. M. Larsen, A. Astrup, L. Dragsted, *Anal Bioanal Chem* **2014**, 406, 1829.

[163] L. Brennan, *Biochem Soc Trans* **2013**, 41, 670.

[164] M.-B. S. Andersen, H. C. Reinbach, A. Rinnan, T. Barri, C. Mithril, L. O. Dragsted, *Metabolomics* **2013**, 9, 984.

[165] J. L. Preau, Jr., L. Y. Wong, M. J. Silva, L. L. Needham, A. M. Calafat, *Environ Health Perspect* **2010**, 118, 1748.

[166] C. S. Cuparencu, M.-B. S. Andersen, G. Gurdeniz, S. S. Schou, M. W. Mortensen, A. Raben, A. Astrup, L. O. Dragsted, *Metabolomics* **2016**, 12, 31.

[167] J. Pekkinen, N. Rosa-Sibakov, V. Micard, P. Keski-Rahkonen, M. Lehtonen, K. Poutanen, H. Mykkanen, K. Hanhineva, *Mol Nutr Food Res* **2015**, 59, 1550.

[168] R. A. Van Den Berg, H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde, M. J. Van Der Werf, *BMC Genomics* **2006**, 7, 142.

[169] X. Mora-Cubillos, S. Tulipani, M. Garcia-Aloy, M. Bullo, F. J. Tinahones, C. Andres-Lacueva, *Mol Nutr Food Res* **2015**, 59, 2480.

[170] K. Hanhineva, M. A. Lankinen, A. Pedret, U. Schwab, M. Kolehmainen, J. Paananen, V. De Mello, R. Sola, M. Lehtonen, K. Poutanen, M. Uusitupa, H. Mykkanen, *J Nutr* **2015**, 145, 7.

[171] S. Naz, M. Vallejo, A. Garcia, C. Barbas, *J Chromatogr A* **2014**, 1353, 99.

[172] H. Gibbons, E. Carr, B. A. Mcnulty, A. P. Nugent, J. Walton, A. Flynn, M. J. Gibney, L. Brennan, *Mol Nutr Food Res* **2017**, 61.

[173] A. K. Smilde, J. J. Jansen, H. C. Hoefsloot, R. J. Lamers, J. Van Der Greef, M. E. Timmerman, *Bioinformatics* **2005**, 21, 3043.

[174] G. Zwanenburg, H. C. J. Hoefsloot, J. A. Westerhuis, J. J. Jansen, A. K. Smilde, *Journal of Chemometrics* **2011**, 25, 561.

[175] J. Xia, D. S. Wishart, *Nat Protoc* **2011**, 6, 743.

[176] J. Godzien, M. Ciborowski, S. Angulo, C. Barbas, *Electrophoresis* **2013**, 34, 2812.

[177] C. Brunius, A. Pedersen, D. Malmodin, B. G. Karlsson, L. I. Andersson, G. Tybring, R. Landberg, *Bioinformatics* **2017**, 33, 3567.

[178] K. Hanhineva, C. Brunius, A. Andersson, M. Marklund, R. Juvonen, P. Keski-Rahkonen, S. Auriola, R. Landberg, *Mol Nutr Food Res* **2015**, 59, 2315.

[179] M. J. Cichon, K. M. Riedl, L. Wan, J. M. Thomas-Ahner, D. M. Francis, S. K. Clinton, S. J. Schwartz, *Mol Nutr Food Res* **2017**, 61.

[180] M. Katajamaa, M. Oresic, *J Chromatogr A* **2007**, 1158, 318.

[181] K. A. Veselkov, L. K. Vingara, P. Masson, S. L. Robinette, E. Want, J. V. Li, R. H. Barton, C. Boursier-Neyret, B. Walther, T. M. Ebbels, I. Pelczer, E. Holmes, J. C. Lindon, J. K. Nicholson, *Anal Chem* **2011**, 83, 5864.

[182] R. Smith, D. Ventura, J. T. Prince, *Brief Bioinform* **2015**, 16, 104.

[183] S. Bijlsma, I. Bobeldijk, E. R. Verheij, R. Ramaker, S. Kochhar, I. A. Macdonald, B. Van Ommen, A. K. Smilde, *Anal Chem* **2006**, 78, 567.

[184] M. Sysi-Aho, M. Katajamaa, L. Yetukuri, M. Oresic, *BMC Bioinformatics* **2007**, 8, 93.

[185] W. Dunn, D. Broadhurst, A. Edison, M. Viant, D. W. Bearden, R. Beger, C. Guillou, *Metabolomics* **2017**, 13.

[186] O. Hrydziuszko, M. R. Viant, *Metabolomics* **2012**, 8.

[187] R. Wei, J. Wang, M. Su, E. Jia, S. Chen, T. Chen, Y. Ni, *Sci Rep* **2018**, 8, 663.

[188] P. Gromski, Y. Xu, H. Kotze, E. Correa, D. Ellis, E. Armitage, M. Turner, R. Goodacre, *Metabolites* **2014**, 4, 433.

[189] N. Kumar, M. A. Hoque, M. Shahjaman, S. M. Islam, M. N. Mollah, *Biomed Res Int* **2017**, 2017, 2437608.

[190] S. L. Taylor, L. R. Ruhaak, K. Kelly, R. H. Weiss, K. Kim, *Brief Bioinform* **2017**, 18, 312.

[191] Y. Gagnebin, D. Tonoli, P. Lescuyer, B. Ponte, S. De Seigneux, P. Y. Martin, J. Schappler, J. Boccard, S. Rudaz, *Anal Chim Acta* **2017**, 955, 27.

[192] S. M. Kohl, M. S. Klein, J. Hochrein, P. J. Oefner, R. Spang, W. Gronwald, *Metabolomics* **2012**, 8, 146.

[193] A. M. De Livera, M. Sysi-Aho, L. Jacob, J. A. Gagnon-Bartsch, S. Castillo, J. A. Simpson, T. P. Speed, *Anal Chem* **2015**, 87, 3606.

[194] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, *Analytical Chemistry* **2006**, 78, 4281.

[195] R. Di Guida, J. Engel, J. W. Allwood, R. J. M. Weber, M. R. Jones, U. Sommer, M. R. Viant, W. B. Dunn, *Metabolomics* **2016**, 12, 93.

[196] A.-H. Emwas, R. Roy, R. T. Mckay, D. Ryan, L. Brennan, L. Tenori, C. Luchinat, X. Gao, A. C. Zeri, G. N. Gowda, *Journal of proteome research* **2016**, 15, 360.

[197] F. C. Vogl, S. Mehrl, L. Heizinger, I. Schlecht, H. U. Zacharias, L. Ellmann, N. Nurnberger, W. Gronwald, M. F. Leitzmann, J. Rossert, K. U. Eckardt, K. Dettmer, P. J. Oefner, *Anal Bioanal Chem* **2016**, 408, 8483.

[198] B. Li, J. Tang, Q. Yang, S. Li, X. Cui, Y. Li, Y. Chen, W. Xue, X. Li, F. Zhu, *Nucleic Acids Research* **2017**.

[199] R. Wehrens, *Chemometrics with R - Multivariate data analysis in the natural sciences and life sciences*, Springer, Heidelberg 2011.

[200] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2009.

[201] A. Gelman, J. Hill, *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press 2012.

[202]   N. Saw, C. Moser, S. Martens, P. Franceschi, *Hortic Res* **2017**, 4, 17038.

[203]   K. Sriyudthsak, F. Shiraishi, M. Y. Hirai, *Front Mol Biosci* **2016**, 3, 15.

[204]   M. N. Triba, L. Le Moyec, R. Amathieu, C. Goossens, N. Bouchemal, P. Nahon, D. N. Rutledge, P. Savarin, *Mol Biosyst* **2015**, 11, 13.

[205]   J. A. Westerhuis, H. C. J. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. J. Van Velzen, J. P. M. Van Duijnhoven, F. A. Van Dorsten, *Metabolomics* **2008**, 4, 81.

[206]   M. Stocchero, S. Riccadonna, P. Franceschi, *Journal of chemometrics* **2018**, e2987.

[207]   J. Trygg, E. Holmes, T. Lundstedt, *Journal of Proteome Research* **2007**, 6, 469.

[208]   E. Pujos-Guillot, J. Hubert, J. F. Martin, B. Lyan, M. Quintana, S. Claude, B. Chabanas, J. A. Rothwell, C. Bennetau-Pelissero, A. Scalbert, B. Comte, S. Hercberg, C. Morand, P. Galan, C. Manach, *J Proteome Res* **2013**, 12, 1645.

[209]   R. Llorach, S. Medina, C. Garcia-Viguera, P. Zafrilla, J. Abellan, O. Jauregui, F. A. Tomas-Barberan, A. Gil-Izquierdo, C. Andres-Lacueva, *Electrophoresis* **2014**, 35, 1599.

[210]   J. A. Rothwell, Y. Fillatre, J. F. Martin, B. Lyan, E. Pujos-Guillot, L. Fezeu, S. Hercberg, B. Comte, P. Galan, M. Touvier, C. Manach, *PLoS One* **2014**, 9, e93474.

[211]   J. J. Jansen, H. C. J. Hoefsloot, J. Van Der Greef, M. E. Timmerman, A. K. Smilde, *Analytica Chimica Acta* **2005**, 530, 173.

[212]   E. J. Van Velzen, J. A. Westerhuis, J. P. Van Duynhoven, F. A. Van Dorsten, H. C. Hoefsloot, D. M. Jacobs, S. Smit, R. Draijer, C. I. Kroner, A. K. Smilde, *J Proteome Res* **2008**, 7, 4483.

[213]   P. W. Jonsson Par, Anna Wuolikainen, Henrik Antti, *Metabolomics* **2015**, 11.

[214]   U. Thissen, S. Wopereis, S. A. Van Den Berg, I. Bobeldijk, R. Kleemann, T. Kooistra, K. W. Van Dijk, B. Van Ommen, A. K. Smilde, *BMC Bioinformatics* **2009**, 10, 52.

[215]   A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B. S. Kristal, B. Van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart, S. Wopereis, *Metabolomics* **2009**, 5, 435.

[216]   P. Franceschi, M. Giordan, R. Wehrens, *TrAC Trends in Analytical Chemistry* **2013**, 50, 11.

[217]   T. Esko, J. N. Hirschhorn, H. A. Feldman, Y. H. Hsu, A. A. Deik, C. B. Clish, C. B. Ebbeling, D. S. Ludwig, *Am J Clin Nutr* **2017**, 105, 547.

[218]   P. Filzmoser, K. Hron, C. Reimann, *Environmetrics* **2009**, 20, 621.

[219] L. Fredrik, B. Hansen, W. Karcher, *Journal of chemometrics* **1996**, 10, 521.

[220] D.-S. Cao, Bing Wang, Mao-Mao Zeng, Yi-Zeng Liang, Qing-Song Xu, Liang-Xiao Zhang, Hong-Dong Li; Qian-Nan Hu, *Analyst* **2011**, 136.

[221] I. Garcia-Perez , Joram M Posma, Rachel Gibson, Edward S Chambers, Tue H Hansen, Henrik Vestergaard, Torben Hansen, Manfred Beckmann, Oluf Pedersen, Paul Elliott, Jeremiah Stamler, Jeremy K Nicholson, John Draper, John C Mathers, Elaine Holmes, Gary Frost, *Lancet Diabetes Endocrinol* **2017**, 5.

[222] L. Brennan, *Essays Biochem* **2016**, 60, 451.

[223] A. K. S. Szymanska Ewa , Edoardo Saccenti, Johan A. Westerhuis, *Metabolomics* **2012**.

[224] F. Tugizimana, P. A. Steenkamp, L. A. Piater, I. A. Dubery, *Metabolites* **2016**, 6.

[225] S. Wiklund, Erik Johansson, Lina Sjostrom, Ewa J. Mellerowicz, Ulf Edlund, John P. Shockcor, Johan Gottfries,Thomas Moritz, Johan Trygg, *Analytical Chemistry* **2008**, 80, 7.

[226] G. Nyamundanda, L. Brennan, I. C. Gormley, *BMC Bioinformatics* **2010**, 11, 571.

[227] E. Chambers, D. M. Wagrowski-Diehl, Z. Lu, J. R. Mazzeo, *J Chromatogr B Analyt Technol Biomed Life Sci* **2007**, 852, 22.

[228] R.-S. P. Sansone Sa, Field D, Maguire E, Taylor C, Hofmann O, Fang H, Neumann S, Tong W, Amaral-Zettler L, Begley K, Booth T, Bougueleret L, Burns G, Chapman B, Clark T, Coleman La, Copeland J, Das S, De Daruvar a, De Matos P, Dix I, Edmunds S, Evelo Ct, Forster Mj, Gaudet P, Gilbert J, Goble C, Griffin Jl, Jacob D, Kleinjans J, Harland L, Haug K, Hermjakob H, Ho Sui Sj, Laederach a, Liang S, Marshall S, Mcgrath a, Merrill E, Reilly D, Roux M, Shamu Ce, Shang Ca, Steinbeck C, Trefethen a, Williams-Jones B, Wolstencroft K, Xenarios I, Hide W., *Nat Genet* **2012**, 44.

[229] D. Blankenberg, A. Gordon, G. Von Kuster, N. Coraor, J. Taylor, A. Nekrutenko, *Bioinformatics* **2010**, 26, 1783.

[230] Y. Guitton, M. Tremblay-Franco, G. Le Corguillé, J.-F. Martin, M. Pétéra, P. Roger-Mele, A. Delabrière, S. Goulitquer, M. Monsoor, C. Duperier, C. Canlet, R. Servien, P. Tardivel, C. Caron, F. Giacomoni, E. A. Thévenot, *The International Journal of Biochemistry & Cell Biology* **2017**.

[231] R. Tautenhahn, G. J. Patti, D. Rinehart, G. Siuzdak, *Anal Chem* **2012**, 84, 5035.

[232] D. Jacob, C. Deborde, M. Lefebvre, M. Maucourt, A. Moing, *Metabolomics* **2017**, 13, 36.

[233] Q. Gao, G. Pratico, A. Scalbert, G. Vergeres, M. Kolehmainen, C. Manach, L. Brennan, L. A. Afman, D. S. Wishart, C. Andres-Lacueva, M. Garcia-Aloy, H. Verhagen, E. J. M. Feskens, L. O. Dragsted, *Genes Nutr* **2017**, 12, 34.

[234] F.-N. B. W. Group, *BEST (Biomarkers, EndpointS, and other Tools) Resource*, Food and Drug Administration (US), Silver Spring (MD) 2016.

[235] S. S. Heinzmann, I. J. Brown, Q. Chan, M. Bictash, M. E. Dumas, S. Kochhar, J. Stamler, E. Holmes, P. Elliott, J. K. Nicholson, *Am J Clin Nutr* **2010**, 92, 436.

[236] J. Riedl, S. Esslinger, C. Fauhl-Hassek, *Anal Chim Acta* **2015**, 885, 17.

[237] L. Laghi, G. Picone, F. Capozzi, *Trends in Analytical Chemistry* **2014**, 59, 93.

[238] F. Capozzi, A. Trimigno, in: Sebedio, J. L., Brennan, L. (Eds.), *Metabolomics as a Tool in Nutrition Research*, Woodhead Publishing 2015, pp. 203-228.

[239] G. Pimentel, K. J. Burton, F. P. Pralong, N. Vionnet, R. Portmann, G. Vergères, *Current Opinion in Food Science* **2017**, 16, 67.

[240] T. M. Bayless, E. Brown, D. M. Paige, *Curr Gastroenterol Rep* **2017**, 19, 23.

[241] G. Pimentel, K. J. Burton, M. Rosikiewicz, C. Freiburghaus, U. Von Ah, L. H. Münger, F. P. Pralong, N. Vionnet, G. Greub, R. Badertscher, G. Vergeres, *Br J Nutr* **2017**, 118, 1070.

[242] A. Y. Del Valle-Pinero, H. E. Van Deventer, N. H. Fourie, A. C. Martino, N. S. Patel, A. T. Remaley, W. A. Henderson, *Clin Chim Acta* **2013**, 418, 97.

[243] Manach C, Williamson G, Morand C, Scalbert A, R. C, *Am J Clin Nutr* **2005**, 81.

[244] A. Dahan, H. Altman, *Eur J Clin Nutr* **2004**, 58, 1.

[245] G. An, J. K. Mukker, H. Derendorf, R. F. Frye, *J Clin Pharmacol* **2015**, 55, 1313.

[246] S. Mouly, C. Lloret-Linares, P. O. Sellier, D. Sene, J. F. Bergmann, *Pharmacol Res* **2017**, 118, 82.

[247] A. A. Franke, J. F. Lai, B. M. Halm, *Arch Biochem Biophys* **2014**, 559, 24.

[248] H. Gibbons, C. J. R. Michielsen, M. Rundle, G. Frost, B. A. Mcnulty, A. P. Nugent, J. Walton, A. Flynn, M. J. Gibney, L. Brennan, *Mol Nutr Food Res* **2017**, 61.

[249] X. Yin, H. Gibbons, M. Rundle, G. Frost, B. A. Mcnulty, A. P. Nugent, J. Walton, A. Flynn, M. J. Gibney, L. Brennan, *J Nutr* **2017**.

[250] Assfalg M, Bertini I, Colangiuli D, Luchinat C, Schäfer H, Schütz B, S. M., *Proceedings of the National Academy of Sciences* **2008**, 105, 4.

[251] Guasch-Ferré M, Hruby A, Toledo E, Clish Cb, Martínez-González Ma, Salas-Salvadó J, H. Fb, *Diabetes Care* **2016**, 39.

[252]   Floegel A, Y. Z. Stefan N, Mühlenbruch K, Drogan D, Joost Hg, Fritsche A, Häring Hu, Hrabě De Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze Mb, Adamski J, Boeing H, P. T, *Diabetes* **2013**, 62.

[253]   T. C. Carter, D. Rein, I. Padberg, E. Peter, U. Rennefahrt, D. E. David, V. Mcmanus, E. Stefanski, S. Martin, P. Schatz, S. J. Schrodi, *Metabolism* **2016**, 65, 1399.

[254]   R. C. Miller, E. Brindle, D. J. Holman, J. Shofer, N. A. Klein, M. R. Soules, K. A. O'connor, *Clin Chem* **2004**, 50, 924.

[255]   N. Di Ferrante, H. S. Lipscomb, *Clin Chim Acta* **1970**, 30, 69.

[256]   D. L. Heavner, W. T. Morgan, S. B. Sears, J. D. Richardson, G. D. Byrd, M. W. Ogden, *J Pharm Biomed Anal* **2006**, 40, 928.

[257]   K. J. Boudonck, M. W. Mitchell, L. Nemet, L. Keresztes, A. Nyska, D. Shinar, M. Rosenstock, *Toxicol Pathol* **2009**, 37, 280.

[258]   V. Chadha, U. Garg, U. S. Alon, *Pediatr Nephrol* **2001**, 16, 374.

[259]   G. G. Gyamlani, E. J. Bergstralh, J. M. Slezak, T. S. Larson, *Am J Kidney Dis* **2003**, 42, 685.

[260]   W. Richmond, G. Colgan, S. Simon, M. Stuart-Hilgenfeld, N. Wilson, U. S. Alon, *Clin Nephrol* **2005**, 64, 264.

[261]   B. M. Warrack, S. Hnatyshyn, K. H. Ott, M. D. Reily, M. Sanders, H. Zhang, D. M. Drexler, *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, 877, 547.

Author contributions

All authors contributed significantly to the writing and revising of the manuscript. Apart from the four first authors and the last, corresponding author, all other authors appear in alphabetical order.

Acknowledgements

Conflict of interest

All authors declare no conflict of interest.

Table 1. Table of content. The colored dots indicate the metabolomic technology discussed in each section (blue: generic tasks, red: GC-MS, green: LC-MS, orange; NMR)

Table 2. Collection and labeling of biological samples.

Information on the participant
Gender
Age
Recent diet and supplement use
Physiological conditions (e.g. menstrual cycle)
Recent smoking
Current medication use
Recent medical illness
Specific information on the sample
Laboratory identification number (anonymized)
Type of sample (blood/urine/feces)
Study name
Time and date of collection
Volume of aliquot
Storage conditions
Tests to be performed
Precise storage location
Labeling of samples
Study name
Participant ID
Visit number
Sample Type
Time-point

Table 3. Normalization methods.

Normalization method	Pros	Cons
Urine volume	Easy to perform	Not very accurate
Creatinine [254, 255]	Standard clinical laboratory technique	Can vary up to 5 fold
Specific gravity [256, 257]	Easy standard laboratory method, correlates well with osmolality	Sensitive to protein or glucose containing samples
Osmolality [258-261]	Standard clinical laboratory technique, very reliable absolute quantification of urine concentration	-

Supporting information

Appendix 1. Chapter 4 - Sample preparation

Appendix 2. Chapter 5 - Chromatography

Appendix 3. Chapter 6 - Data acquisition

Appendix 4. Chapter 7 - Raw data files conversion and feature extraction

Appendix 5. Chapter 9 - Compound identification

Box titles and captions

Box 1: Nutritional study designs.

Box 2: Urine and feces collection. Different collection devices and containers are shown for sampling of spot urine and 24h urine, as well as faces.

Box 3: Pre-storage and aliquoting of plasma/serum and urine.

This figure shows volcano plots **(panels A+B)** based on urinary metabolite profiles of >300 healthy volunteers (men and women) of the KarMeN study [30] as determined by untargeted GC×GC-MS [123, 125]. Urine samples were normalized according to urinary creatinine or osmolality by differential dilution before the analysis. Each dot in the volcano plots stands for one analyte. The log2-transformed mean ratio (x-axis), i.e. the fold change between men and women, is plotted against significance (y-axis), here expressed as the −log10-transformed p-values of a Welch t-test (higher numbers correspond to small p-values, with the green dashed line indicating the p=0.05 threshold).

Both plots reveal a significant difference between men and women for many metabolites. However, a systematic bias was obviously introduced by the creatinine normalization: The entire plot is left-skewed (significant and non-significant parts) towards women, leading to the (erroneous) conclusion that urine samples of women are on average more concentrated. In comparison, the volcano plot after normalization on osmolality is much more symmetrical. Thus, although creatinine content and osmolality of urine samples are highly correlated (**see panel C**), the application of creatinine normalization on a study population comprising men and women may introduce a strong bias. See also [103].

Box 4: Pre-storage and aliquoting of feces. Three different ways of feces preparation are presented: fresh

feces freezing, fecal water preparation, and fecal powder preparation.

DATA ANALYSIS AND STATISTICAL MODELLING FLOW

OVERVIEW OF STATISTICAL METHODS

	Visualisation	Regression	Independent 2-group test	Dependent 2-group test	Multiple factors	Time series
Parametric	Plot	Linear models	t-test	Paired t-test	ANOVA Mixed models	Repeated measures
Non-parametric	Rank plot	Rank regr. Kernel LOESS	MW U-test	Wilcoxon signed-rank	Kruskal-Wallis Friedman rank ANOVA	Friedman rank-repeated
Multivariate	PCA	PLS; RF; SVM	PLS-DA; RF; SVM	ML-PLS; ML-RF; ML-SVM	ANOVA decomposition	ANOVA decomposition

Box 5: Differences between plasma and serum.

VERIFICATION OF QUALITY CONTROL SAMPLES AND OUTLIERS WITH PCA

Example of Batch Effect – samples divided into two clouds due to a prompt problem during injections (RAW, not normalized data)

The same sample set normalized by:
1) median of each feature/plate = 1
2) Intensity of creatinine *m/z* feature

Example of Drift due to loss of signal intensity– Samples and QCs suffered from drop in the signal intensity during injections (RAW, not normalized data)

The same sample set normalized by:
1) median of each feature/plate = 1
2) Intensity of creatinine *m/z* feature

Example of QC prepared from study samples – a tight QCs cloud is located nearby a wide cloud of real samples (RAW, not normalized data)

Example of QC prepared from commercial biological fluid (i.e. plasma)– a tight QCs cloud is located far from a wider cloud of real samples (RAW, not normalized data). Three samples (outliers) are separated from the rest of samples (blue circle)

● QC samples   ● Study samples   ○ Outliers samples

Box 6: Different types of possible analysis for biospecimens.

UNTARGETED ANALYSIS

➤ LC-MS: polar and medium-polar metabolites (i.e. polyphenols and its metabolites, aminoacids)
➤ LC-MS: non-polar metabolites (i.e lipidomics)
➤ GC-MS: volatile metabolites
➤ GC-MS: metabolites after derivatization
➤ NMR

for analysis
to be stored

MINIMUM REQUIREMENT

3 ALIQUOTS FOR ONE TYPE OF ANALYSIS (i.e. GC-MS)

TARGETED ANALYSIS

➤ LC-MS (kits, etc)
➤ GC-MS quantitative analysis with/without derivatization (i.e. SCFA, sugars etc)
➤ Clinical analysis (multiple choices)

In general multiple aliquots of smaller volume (i.e. 200 µL) are recommended, rather than few aliquots of high volume.

Higher number of aliquots is recommended, which allows for additional analysis (i.e. additional analytical technique that arises in the future)

RECOMMENDED

Tips & Tricks

Box 7: Preparation and sustainable storage of vials and containers.

Use containers adapted for the volume of matrix that will be stored. Matrix should fill possibly 80%-90% of container with minimum amount of air above its surface. Use containers with low adherence in the interior of the tube, and twisted taps that avoid splashing while opening

Tips & Tricks

Conical/skirted vials for better recovery of matrix, twisted taps.
Recommended cryovial

-80 °C and more

RECOMMENDED

matrix residues on the cap

AIR

DON'T STORE AIR

11 cm

7 cm

PROJECT NAME....., DATE...., MATRIX......., ALIQUOT NR., AFFILIATION...
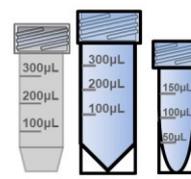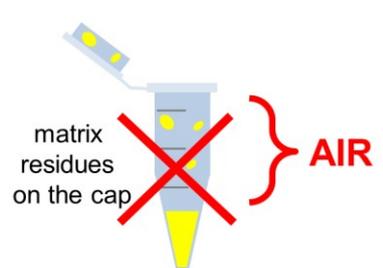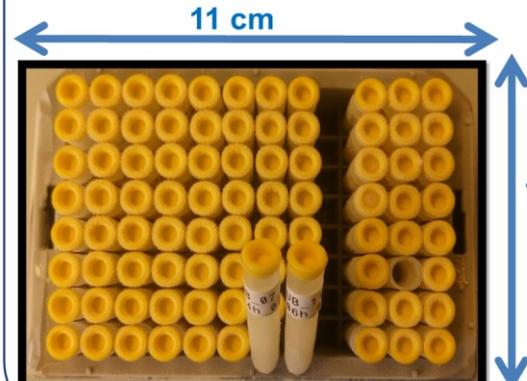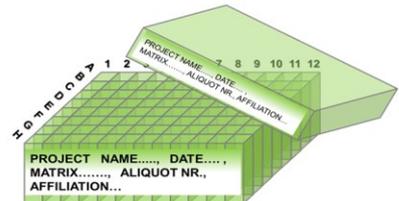
Optimization of storage capacity in freezers of small aliquotes in tight boxes. Avoid long term storage in paper boxes.

Box 8: Different storage and thawing conditions for biospecimens.

Box 9: LC-MS analysis: preparation of QC samples and injection queue.

Box 10: GC-MS analysis: preparation of QC samples and injection queue.

Box 11: Normalization of nutrimetabolomic data: possible gender bias.

NORMALIZATION TO CREATININE AND OSMOLALITY: POSSIBLE GENDER BIAS

This figure shows volcano plots (**panels A+B**) based on urinary metabolite profiles of >300 healthy volunteers (men and women) of the KarMeN study [30] as determined by untargeted GC×GC-MS [123, 125]. Urine samples were normalized according to urinary creatinine or osmolality by differential dilution before the analysis. Each dot in the volcano plots stands for one analyte. The log2-transformed mean ratio (x-axis), i.e. the fold change between men and women, is plotted against significance (y-axis), here expressed as the –log10-transformed p-values of a Welch t-test (higher numbers correspond to small p-values, with the green dashed line indicating the p=0.05 threshold).

Both plots reveal a significant difference between men and women for many metabolites. However, a systematic bias was obviously introduced by the creatinine normalization: The entire plot is left-skewed (significant and non-significant parts) towards women, leading to the (erroneous) conclusion that urine samples of women are on average more concentrated. In comparison, the volcano plot after normalization on osmolality is much more symmetrical. Thus, although creatinine content and osmolality of urine samples are highly correlated (**see panel C**), the application of creatinine normalization on a study population comprising men and women may introduce a strong bias. See also [103].

Box 12: General modelling flow for data analysis and selection of statistical methods.

DATA ANALYSIS AND STATISTICAL MODELLING FLOW

OVERVIEW OF STATISTICAL METHODS

	Visualisation	Regression	Independent 2-group test	Dependent 2-group test	Multiple factors	Time series
Parametric	Plot	Linear models	t-test	Paired t-test	ANOVA Mixed models	Repeated measures
Non-parametric	Rank plot	Rank regr. Kernel LOESS	MW U-test	Wilcoxon signed-rank	Kruskal-Wallis Friedman rank ANOVA	Friedman rank-repeated
Multivariate	PCA	PLS; RF; SVM	PLS-DA; RF; SVM	ML-PLS; ML-RF; ML-SVM	ANOVA decomposition	ANOVA decomposition

Box 13: Use of PCA for verification of data quality. PCA visualization of study samples and QC samples affected by: batch effect (left site), drift due to loss of intensity (middle), presence of outliers (right site).