Nutritional metabolomics:

The search for dietary exposure variables

Millie Rådjursöga

Department of Internal Medicine
and Clinical Nutrition,
Institute of Medicine,
Sahlgrenska Academy
University of Gothenburg

Gothenburg, Sweden, 2018

UNIVERSITY OF
GOTHENBURG

" To conquer fear is the beginning of wisdom, in the pursuit of truth as in the endeavor after a worthy manner of life."

Bertrand Russell

Abstract

To establish associations and causation between diet and health, objective and reliable methods are needed to measure dietary exposure. Metabolomics provide an unbiased tool for exploring the modulation of the human metabolome in response to food intake.

The aim of this doctoral thesis was to investigate the postprandial metabolic response in two cross-over meal studies using nuclear magnetic resonance (NMR) metabolomics. In addition, a methodological study with the aim to compare three pre-processing protocols for high-throughput NMR serum metabolomics for large samples series was included in the present work.

The meal studies aimed to investigate:
(1) the postprandial metabolic response to two equicaloric breakfast meals, cereal breakfast (CB) and egg and ham breakfast (EHB) in serum and urine. (2) the postprandial metabolic response to breakfast meals corresponding to vegan (VE), lacto-ovo vegetarian (LOV) and omnivore (OM) diets in serum. Metabolic profiles along with key discriminatory metabolites of biological relevance that largely reflected dietary composition were identified in both meal studies. Tyrosine and proline were found to discriminate for both the CB and LOV. Also valine was higher after the LOV compared to the VE and higher in the CB breakfast that had high dairy content. In turn, creatine, isoleucine, choline and lysine were discriminating for both the EHB and OM breakfasts in serum, that both contained comparably high content of animal protein. This implies that the metabolic response to meals high in dairy and meat can be reflected in metabolite concentrations irrespective of the total food matrix in a meal. In addition, coffee and tea consumption could be identified in urine. Comparing dilution, precipitation (methanol) and ultrafiltration as pre-processing deproteinization methods for serum, the precipitation protocol was found to be the method of choice for high-throughput NMR metabolomics for large sample series. Overall, our results demonstrate NMR metabolomics as an applicable method in the search of dietary exposure variables.

Keywords

Sammanfattning på svenska

Avhandlingen innefattar fyra delarbeten inom området nutritionell metabolomik. Inom ramen för avhandlingen har vi studerat metabolt svar till olika typer av måltider i två kliniska måltidsstudier. Avhandlingen innefattar även ett arbete av metodologisk karaktär med fokus på förbearbetning av serumprover för nukleär magnet resonans (NMR) metabolomik.

Det övergripande syftet med avhandlingsarbetet var att undersöka metabolomik som objektivt verktyg för att utvärdera kostintag. Det specifika syftet för den första måltidsstudien var att studera metabolt svar och reproducerbarhet hos postprandiella metabola profiler samt identifiera metaboliter som gör det möjligt att särskilja dessa måltider i urin och serum. Den andra måltidsstudien syftade likaså till att undersöka skillnaden i postprandiellt metabolt svar men här mellan tre måltider som speglar tre olika typer av kosthållning; vegan-, lakto ovo vegetarisk-, och blandkost. Det metodologiska delarbetet syftade till att undersöka metoder för förbearbetning av serum för NMR metabolomik vid analys av stora provserier.

I den första måltidsstudien (delarbeten I och II) studerade vi den postprandiella metabola responsen i urin och serum till två olika typer av frukostar med samma innehåll av energigivande näringsämnen (makronutrienter) men med olika livsmedelsinnehåll. Här studerade vi även reproducerbarheten hos den metabola responsen vid upprepade tillfällen. Studien inkluderade 24 friska volontärer (12 män, 12 kvinnor) som åt varje frukost vid fyra tillfällen, totalt åtta tillfällen under två veckor. Vi analyserade sedan den metabola responsen i urin och serum med ^{1}H NMR och uni- och multivariat statistik och identifierade diskriminerande metaboliter.

I den andra måltidsstudien (delarbete III) studerade vi den postprandiella metabola responsen i serum till tre olika typer av frukostar baserade på vegan-, lakto ovo vegetarisk-, och blandkost. Här inkluderade studien 32 friska volontärer, fördelat jämnt mellan antal män och kvinnor. Serumprover analyserades med ^{1}H NMR och uni- och multivariat statistik. Variation i koncentration hos identifierade metaboliter mellan prover i fasta och efter måltid samt mellan frukostmåltiderna identifierades.

I den metodologiska studien (delarbete IV) undersöktes tre olika sätt att förbearbeta serumprover för ^{1}H NMR metabolomik, dvs. ut-spädning, ultrafiltrering och metanolfällning, och skillnaden dem emellan med avse-

ende på reproducerbarhet, automatisering, användbarhet vid stora provserier samt identifiering och kvantifiering av metaboliter.

Resultaten från de båda måltidsstudierna visar att reproducerbara metabola profiler från måltider innehållande olika livsmedel men väsentligt samma innehåll av makronutrienter kan identifieras i både serum och urin med hjälp av multivariata modeller. Vidare visar resultaten att skillnaden i metabol respons kan karaktäriseras genom identifiering av diskriminerande metaboliter vilka i stor utsträckning skiljer mellan serum och urin. En svårighet är valet av tidpunkt för provtagning i förhållande till måltiden, då hastigheten för omsättning av olika livsmedel och ämnen skiljer sig åt. De metabola profilerna kännetecknas av både endogena metaboliter och av metaboliter som direkt kan relateras till mängden i de i måltiden ingående livsmedlen.

Metanolfällning av serum, i jämförelse med ultrafiltrering och utspädning, visade sig vara den metod för förbearbetning av serumprover som är att föredra vid NMR-baserad serum-metabolomik vid automatiserat arbetsflöde för stora provserier.

Resultaten från ingående delarbeten tyder på att NMR metabolik är ett fungerande verktyg för att skapa reproducerbara metabola profiler och att identifiera skillnader i metabol respons från måltidsinterventioner. Kontrollerade måltidsstudier är användbara för att bidra med ökad kunskap om metabol respons till olika koster och livsmedel. Detta i sin tur kan leda till att vi med en objektiv metod såsom NMR metabolomik kan identifiera koster/kostvanor for komplementär användning till de traditionellt använda kostundersökningsmetoderna.

List of papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Rådjursöga, M., Karlsson, B. G., Lindqvist, H. M., Pedersen, A., Persson, C., Pinto, R. C., Ellegård, L., & Winkvist, A.

    Metabolic profiles from two different breakfast meals characterized by ^{1}H NMR-based metabolomics.

    Food Chem 2017; 231: 267-274.

II. Rådjursöga, M., Lindqvist, H. M., Pedersen, A., Karlsson, B. G., Malmodin, D., Brunius, C., Ellegård, L., & Winkvist, A.

    The ^{1}H NMR Serum Metabolomics Response of a Two Meal Challenge, a Cross-Over Dietary Intervention Study in Healthy Human Volunteers

    Submitted

III. Rådjursöga, M., Lindqvist, H. M., Pedersen, A., Karlsson, B. G., Malmodin, D., Ellegård, L., & Winkvist, A.

    Nutritional metabolomics: Postprandial Response of Meals Relating to Vegan, Lacto-Ovo Vegetarian, and Omnivore Diets.

    Nutrients 2018, 10(8), E1063; doi: 10.3390/nu10081063

IV. Pedersen, A., Rådjursöga, M., Malmodin, D. & Karlsson, B. G.

    Improving deproteinization pre-processing throughput of NMR-based serum metabolomics.

    Manuscript

All papers were reprinted with the permission of the publishers: Elsevier B.V. (Paper I) and MDPI (Paper III).

Content

Abbreviations

ANOVA	ANalysis Of VAriance
BMI	Body Mass Index
CB	Cereal Breakfast
CPMG	Carr-Purcell-Meiboom-Gill
CV	Cross Validation
CV-ANOVA	Coefficient of Variation- ANalysis of VAriance
DIL	Dilution protocol
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
DWP	Deep Well Plate
EHB	Egg & Ham Breakfast
FFQ	Food Frequency Questionnaire
FoodBAll	Food Biomarkers Alliance
HMDB	Human Metabolome DataBase
HSQC	Heteronuclear multiple quantum correlation
Jres	J-resolved
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid Chromatography - Mass Spectrometry
LOV	Lacto-Ovo Vegetarian
LV	Latent Variable
MET	Metabolic equivalents
MS	Mass Spectrometry
MVA	MultiVariate data Analysis
NAA	N-acetylated-amino acid
NMR	Nuclear Magnetic Resonance
OM	Omnivore
OPLS-DA	Orthogonal Partial Least Squares with Discriminant Analysis
OPLS-EP	Orthogonal Partial Least Squares with Effect Projections
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
ppm	parts per million
PQN	Probabilistic Quotient Normalization
PREC	Methanol precipitation protocol
Q^2	Predictive ability
QC	Quality Control
R^2	Explained variation
S/N	Signal to Noise
SMPD	Small Molecule Pathway Database

SOP	Standard Operation Procedure
TMAO	Trimethylamine-N-oxide
TOCSY	^{1}H-^{1}H total correlation spectroscopy
TSP	3-trimethylsilylpropionic acid
UF	Ultrafiltration protocol
UV	Unit Variance
VE	Vegan
VIP	Variable Importance Plot

1. Background

1.1 The metabolome

The complete set of metabolites in an organism is called the metabolome[1]. The metabolome includes metabolites such as carbohydrates, peptides, amino acids, nucleic acids, minerals, organic acids, vitamins, alkaloids and polyphenols[2]. The metabolome together with the genome (DNA), transcriptome (RNA) and proteome (proteins/enzymes) constitute the main building blocks of a biological system[3] (Figure 1). Simplified, as stated by Dettmer *et al*. (2007) the genome describes what can happen, the transcriptome what appears to be happening, the proteome what makes it happen and the metabolome describes what has happened and is currently happening[1]. A biological system is dynamic and influenced by a variety of endogenous processes and exogenous factors[4] generating a biological phenotype. The metabolome constitutes a metabolic phenotype influenced by a number of factors, diet included[5,6] (Figure 2). Thus, the metabolome is the combination of endogenous and exogenous metabolites as well as products from the microbiota, and as a result the metabolome is of great importance to understanding what effect a certain external factor such as physical activity or diet will have on the human body.
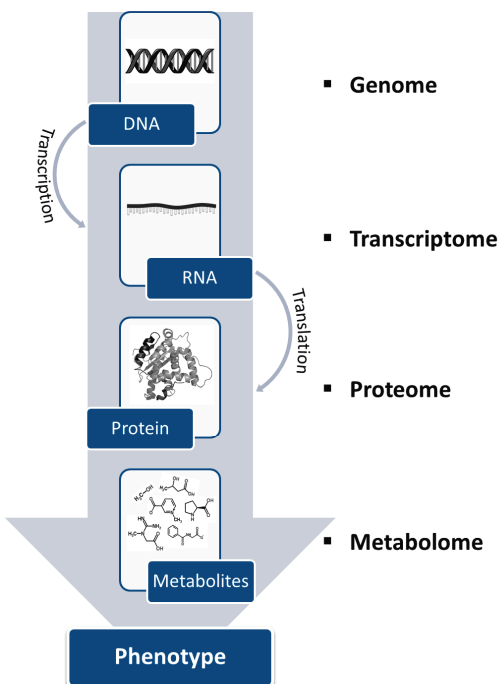
Figure 1 A schematic figure outlining the "omics cascade", the building blocks of a biological system influencing the phenotype. The genome and transcriptome carries the coding information for protein synthesis in DNA and mRNA, respectively, and is effectuated through the processes transcription and translation. The metabolome, being at the endpoint and thus closest to the phenotype, includes metabolites generated from endogenous processes and exogenous factors.
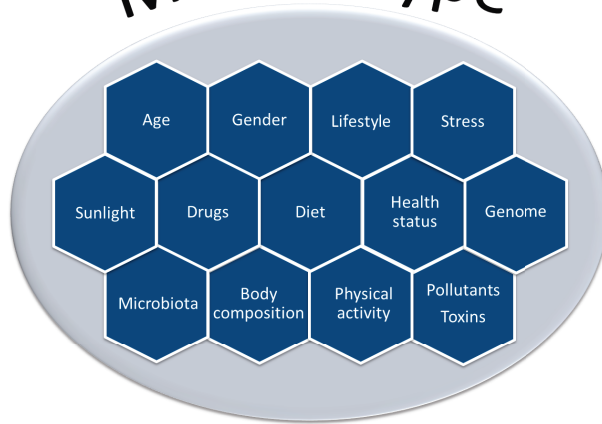
Figure 2 Metabotype. Intrinsic and extrinsic factors influencing the metabolome generating a metabotype (metabolic phenotype).

1.2 Metabolomics in nutrition

To prove causation between diet and health, objective and reliable methods are needed to measure dietary exposure[7]. However, it is a challenge to measure dietary intake with techniques that are both accurate and applicable to free-living individuals. Current dietary assessment methods include dietary records, food diaries, 24-hour dietary recalls, food frequency questionnaires and diet history records. These methods rely on subjects' own reports of their intake and are therefore prone to over- and underreporting of foods[8]. These methods are associated with difficulties, in measuring true dietary intake due to factors such as capturing variation in dietary intake over time, estimation of portion size and difficulties to remember and report all foods consumed[8]. Despite validation efforts random and systematic errors in dietary assessment methods makes it difficult to measure true dietary intake[9,10].

A few validated biomarkers are used alone, or in combination, to validate or to substitute dietary assessment methods[7,10]. Doubly labeled water ($^2H^{17}O$) that measure energy expenditure, urinary nitrogen and potassium to measure protein consumption and potassium intake are widely used biomarkers in nutritional studies. These are examples of recovery biomarkers designed to correlate intake and excretion levels[11]. Other types of dietary biomarkers are predictive biomarkers (sucrose and fructose)[12] and concentration biomarkers (serum vitamins, urinary electrolytes, blood lipids)[13]. Unfortunately, current dietary biomarkers fail to reflect the complex matrix of an overall diet[9]. Providing accu-

rate and reliable measurements of dietary exposure constitutes one of the most challenging problems in nutrition research today[14].

Metabolomics has been defined as the identification and quantification of small (< 1.5kDa) molecules (metabolites) in a biological sample and their systematic and temporal alterations caused by intrinsic and extrinsic factors[2,6,15,16]. With a focus on small molecules and the interaction between them, metabolomics is applied in a number of clinical areas[17] including clinical nutrition[14]. Metabolomics in human nutrition has been described as: "The study of endogenous and gut microbiota metabolic response to food (general diet or intervention) and the identification of metabolites that originate from food and could be used as biomarkers of exposure of these foods"[18].

Using metabolomics, allowing the simultaneous characterization of a large number of metabolites in biological samples, provides the possibility of mapping the complex metabolism of food consumption and the biological consequences of different diets[2]. Applying chemometric tools, i.e. multivariate data analysis, global metabolic profiles/fingerprints can be characterized together with the identification of candidate food biomarkers.

The modern history of metabolomics goes back to the beginning and middle of the 20th century when the development of Mass Spectrometry (MS) and later on Nuclear Magnetic Resonance (NMR) spectroscopy instrumentation began[19]. Already in 1948, MS was used for profiling of biofluids in human subjects and differences between healthy individuals and patients with schizophrenia and alcoholism were identified[19]. During the 1960s and 1970s, technological advances in chromatographic methods enabled researchers to study metabolic profiles and patterns[20-22]. In 1971 Horning and Horning, pioneers in MS metabolomics, coined the concept "metabolic profiles"[21]. The term "metabolome" however, was first stated in 1998 by Oliver *et al*[23]. Since then, a steady progression in resolution and sensitivity in MS and NMR technologies together with the evolution of chromatographic methods have advanced the field of metabolomics. As a result, the human serum[24], plasma[25], urine[26], cerebrospinal fluid[27], saliva[28], and circadian[29] metabolome have been extensively investigated. Metabolomics applied in nutrition followed in the wake of the progression in medical and pharmaceutical sciences[2,30] and articles and reviews in the area started to appear in the early 2000s[31]. However, still only parts of the metabolome related to food consumption have been characterized[32].

Metabolomics analyses generally comprise large volumes of data with high complexity, often noisy and collinear by nature[33]. To be able to extract relevant information and generate robust models from this data, applicable data analytical techniques must be employed[34]. Chemometric approaches i.e. statistics combined with pattern recognition methods, in contrast to traditional univariate statistical tools, are better equipped to tackle the metabolomics data structure (more variables than observations) along with the covariation, interdependence and missing data observed in metabolomics data[33,34]. Multivariate methods such as principal component analysis (PCA), non-linear mapping and factor analysis were first applied for profiling and pattern recognition aspects of MS derived data during the second half of the 1970s[35]. The combinatory use of NMR profiling and pattern recognition software was introduced a decade later, in 1989[36]. Since then the field has progressed and today a number of both commercial and freely available software's for analysis of metabolomics data are accessible[37].

There are two main approaches to metabolomics analysis, non-targeted and targeted[38]. The former is an unbiased hypothesis-generating approach aiming to reproducibly detect as many metabolites as possible in a biological sample with the applied analytical platform[38,39]. Currently, no analytical technology alone can identify all metabolites in a biological sample. Methods used in metabolomics i.e. NMR and MS-based platforms have different characteristics regarding e.g. sample preparation, time of analysis, metabolite range and number of detected metabolites, quantification, resolution, sensitivity and reproducibility and hence can be utilized in a complementary fashion[2,40,41]. When using combined MS techniques, this approach can detect the chromatographic peak area (not concentrations) of up to 3000 metabolites including previously identified and unidentified metabolites[42]. For NMR, generally around 100 metabolites can be identified in a given biofluid. Typically multivariate statistics is applied and metabolic profiles can be used to display the variability between samples and sample groups[43]. The advantage of the non-targeted approach is that it makes no assumptions about candidate molecules[14].

Targeted metabolic profiling, on the other hand, is a hypothesis-testing approach. The approach focuses on quantification and identification of a predefined set of metabolites believed to relate to the studied phenomena[44]. Aims include the identification of differences in concentration between sample groups, turnover, bioavailability and effects of metabolism in these metabolites[38,44]. Typically, internal standards for each targeted metabolite are used and absolute concentrations measured[42]. For NMR, in the case where the resonance frequencies of individual metabolites are known, targeted metabolomics can involve a single 1D sub spectrum per metabolite[45]. However, for NMR, the difference in the ac-

quired data and the number of detected metabolites between targeted and untargeted approaches is, in the normal case, negligible.

1.2.1 The food metabolome

The food metabolome has been defined as "the part of the human metabolome directly derived from the digestion and biotransformation of foods and their constituents"[32]. To investigate the food metabolome, and to detect potential food intake biomarkers, intervention studies are conducted of short- and long-term exposure of different foods and diets. These are combined with observational studies where reported food intake is associated with chemical profiles and identified metabolites.

Proline betaine is an example of a well-investigated food intake biomarker[46,47]. Proline betaine has been identified in both intervention and cross sectional studies as a short-term dietary biomarker for citrus intake[46,48]. Characteristic metabolites (potential food biomarkers) of individual foods has been identified for coffee[49-51], black tea[52,53], chocolate[54], banana[55], red wine[56,57], rye[58-60], whey protein[61], nuts[62], raspberries[63], bluberries[64], cheese[65], milk products[66,67], buckthorn[68], broccoli[63], legumes[69], strawberry[54,68], beetroot[54], lingonberries[70], herring[71], salmon[63], chicken[72], red meat[72-75], seafood[76] and fish[77]. However, it should be mentioned that the level of confidence[78] of the identified characteristic metabolites in these studies varies[32].

In addition, groups of metabolites have been identified to increase or decrease in relation to different diets or food groups. For example, Stella *et al.* (2006) studied the urine metabolic phenotypes of three different diets: vegetarian, low meat and high meat in a cross-over intervention study[79]. Elevated levels of creatine, carnitine, acetyl carnitine and trimethylamine-N-oxide (TMAO) were associated with the high meat diet. In addition, TMAO has also been reported as a potential dietary biomarker for fish consumption[63,72,80]. In a 48h intervention study, Draper *et al.* (2018) investigated the metabolic difference between a vegan and an omnivore diet in plasma[81]. The vegan diet was associated with decreased levels of branched chain amino acids, triglycerides and cholesterol and increased levels of arginine, glycine, two monosaturated acids (C12:1 and C14:1) and three saturated fatty acids (C12:0, C12:0 and C12:0). In 2014 Vázquez-Fresno *et al.* presented, a 3-year intervention study where urine metabolic phenotypes generated from the Mediterranean diet and a low fat diet (control) were studied[82]. Metabolites associating with the Mediterranean diet were related to the metabolism

of carbohydrates, lipids, creatine, creatinine, amino acids and metabolites of microbial origin[82]. Other examples of diets studied in the area of nutritional metabolomics are the new Nordic diet[83,84], western diet[85], vegan, vegetarian and omnivore diets[81,86], diets of varying macronutrient[87], fiber[88], dairy[89] and phytochemical composition[90].

In addition to metabolites derived from food consumption alone, metabolites found in bio fluids can be derived in interplay between food consumption and the gut microbiota[91,92].Gut microbes have been shown to transform food derived nutrients such as polyphenols and fibre[92,93]. Hippuric acid (hippurate) is an example of a potential food biomarker derived from gut microbiota metabolism[94]. The polyphenol content of foods are metabolized by the gut microflora producing phenylpropionic acids that is further transformed to benzoic acids that in turn are metabolized to hippurate in the liver and excreted in the urine[95]. The consumption of polyphenol rich foods and drinks suck as coffee, tea, whole grains, fruit and vegetables are foods high in polyphenols and have been associated with increased levels of hippurate[95,96 97,98].

Metabolites, derived from food consumption, up and down regulate cellular processes and affect health status. As an example, using metabolomics, Vincent *et al*. (2016) showed that the consumption of herring affects amino acid and energy metabolism. In turn, associations between metabolites related to different foods and the development of type two diabetes and glucose tolerant status have been found using metabolomics[74,99].

Although a number of metabolites have been identified as potential food intake biomarkers, validation and identification strategies of these biomarkers constitute a challenge in nutritional metabolomics[32,100]. For this reason, the Food Biomarkers Alliance (FoodBAll), an European project including twenty-four research partners from thirteen countries, was founded[100]. FoodBAll aims to develop a system and clear strategies for discovery, validation, and classification of food intake biomarkers and identify novel biomarkers of foods consumed across Europe[100].

1.3 The NMR metabolomics workflow
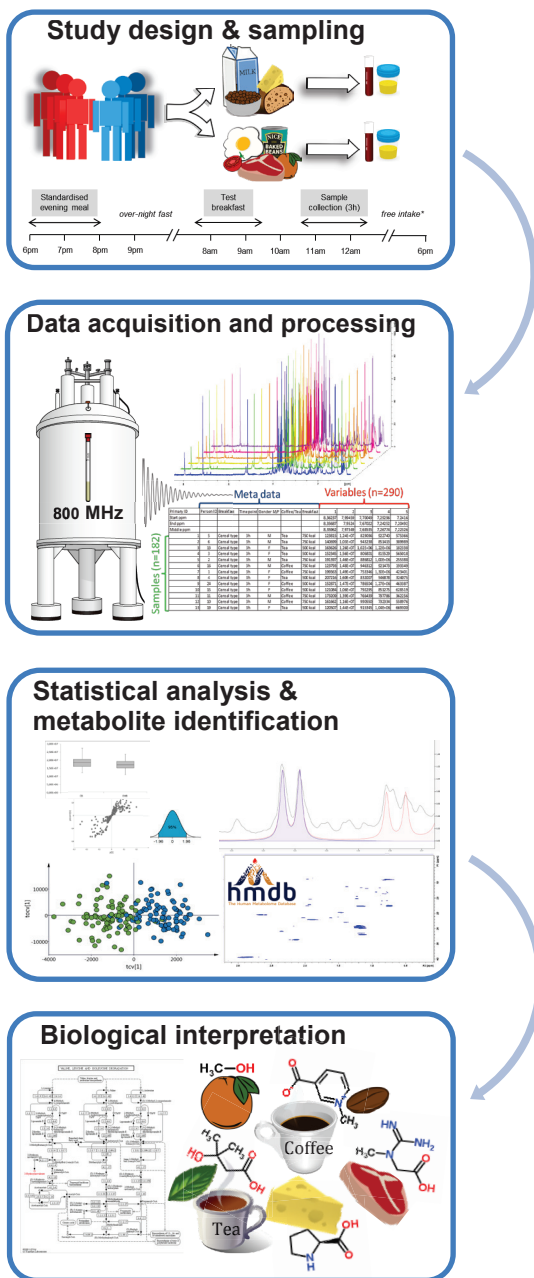
1.3.1 Study design

The aim of the study design is to set up the study in a way that makes it possible to answer the objectives of the study with robust models and to reduce/minimize unwanted variation. The biological variation of interest in metabolomics data is often confounded with unwanted variation[101]. Unwanted variation can occur from biological and experimental variability[102]. Unwanted biological variability includes inter-individual differences depending on intrinsic properties and/or external factors not controlled by the study design[103,104]. Inter-individual variation commonly constitutes a large part of the variation in metabolomics data[105,106]. Unwanted experimental variability can arise from the sample matrix, fluctuations related to the study design, sample collection, pre-analytical handling and pre-processing of samples and instrumental drift. As such, factors of unwanted variation alongside decisions regarding sample size, experiments to perform and methods used are considered in the design of the study[107,108].

In addition, the study design often includes formulation of the aim of the study, which is of utmost importance as considerations regarding which experiments to conduct and how are based on the objectives of the study[34]. Characterization of exclusion and inclusion criteria and information collected regarding study subjects are also included in the study design[34]. The information (meta data) collected about study subjects can be used to identify confounding factors and variation in the data not relating to the aim of the study[107]. Gender, age, medicinal chemistry profile, body mass index (BMI), and lifestyle factors such as level of physical activity, food habits, consumption of alcohol and nicotine and use of supplements and medications are examples of retrieved meta data.

1.3.2 Sampling & pre-analytical handling

The most commonly used biofluids in clinical studies in nutritional metabolomics are blood (serum/plasma) and urine. Sampling of urine has the advantages of large sample volumes, noninvasive and easy collection which can be performed either at the study objects home or at a clinic[109,110]. However, for NMR analyses, in contrast to blood, urine exhibits larger pH fluctuations and variable salt concentrations, which might impair the analysis and obstruct metabolite identification. Blood sampling, in turn, is commonly collected by venipuncture, which is an invasive (and potentially painful) procedure that requires trained staff[109].

Serum is obtained by letting blood samples clot for at least 20-30 minutes before centrifugation where the clot (red and white blood cells along with clotting pro-

teins including fibrinogens) is separated from the serum[111]. Plasma, on the other hand, is collected in tubes containing anticoagulants (heparin, EDTA, sodium citrate etc.)[112]. As no clotting is required, plasma can be processed directly after collection. During centrifugation, red and white blood cells are removed from the plasma that compared to serum contains clotting proteins and relating factors. Plasma benefits from a quicker processing and less hemolysis than serum that can influence metabolite concentrations[111]. In turn, both anticoagulants and polymeric gels might have an effect on metabolomics analysis[112-114]. The reliability of measured metabolites, using liquid chromatography mass spectrometry (LC-MS), has been shown to be higher in serum compared to EDTA plasma[115]. In combination, 101 out of 159 metabolite concentrations displayed significantly higher levels in serum compared to plasma in the same study. In comparison, using NMR, Kaluarachchi *et al*. (2018) found lactate, glutamine, lipids and 37 lipoprotein subclasses higher in serum and glucose higher in plasma[116]. However, when using Carr-Purcell-Meiboom-Gill (CPMG) NMR pulse sequence as T2 filter, Kaluarachchi *et al*. (2018) in conformity with Teahan *et al*. (2006) identified minimal difference between metabolite profiles of serum and heparin plasma[106,116].

Pre-analytical handling has been shown to influence metabolite concentrations and introduce unwanted variation in both blood and urine[109]. Pre-analytical factors that can influence sample quality of serum and plasma samples include time point of sampling, choice of collection tube[112,117,118], centrifugation[119], clotting conditions (time, temperature)[106,111,115,120,121], delayed sample processing and storage[105,111,115,121-123], storage temperature[122], freeze thaw cycles[106,115,122] and shipment[115]. Some metabolites have shown a higher sensitivity to pre-analytical conditions. Among others, lactate, pyruvate, glucose, 5-oxoproline, lysophosphatidylcholines, phosphatidylcholines, lipids, glutamate, cysteine and ornithine/arginine ratios have shown perturbations related to time and temperature[106,122-125].

Urine is less sensitive than blood regarding variation introduced during the pre-analytical phase. Still, factors such as additives, centrifugation, temperature and storage have been shown to affect metabolite concentrations in urine[125-127].

To accomplish uniform pre-analytical handling standard operation procedures (SOPs) for different sample matrices should be developed to minimize variation in between-sample handling[125,128]. In the clinical setting, applicable blood sample collection tubes should be used and centrifugation of samples performed 30 min after sampling[105,118]. In addition, the time between sampling, sample han-

dling and storage should be kept to a minimum and not delayed for more than 2h and thereafter samples should be kept at 4°C to minimize metabolism and transport between intra- and extracellular compartments[106,124,125]. Samples should be stored at minimum -80°C[122]. Preferably, time points of each step in the SOP are recorded for all samples.

1.3.3 NMR-analysis

Sample pre-processing

When analyzing samples with NMR, variation in pH and ionic strength (i.e. salt concentration) between samples affect the signals detected from different protons. In addition, samples of high viscosity, i.e. serum or plasma, hamper the analysis due to slower tumbling rates in more viscous solutions with consequent broadened signal linewidth. Changes in pH results in drift of chemical shifts that may complicate the identification process and give rise to overlapping signals[129]. Furthermore, variation in ionic strength of the samples can influence the performance of the probe as well as induce drift in chemical shifts particularly of metabolites known to chelate metal ions, e.g. citrate[129]. To reduce the influence of variation in pH and ionic strength between samples and to reduce viscosity, addition of buffer and dilution of samples are recommended[128,130]. The most commonly used protocol for serum/plasma and urine pre-processing for NMR metabolomics constitutes dilution with phosphate buffer (pH 7.4) before analysis[128,130-132]. The recommended addition of buffer to serum/plasma samples is 50/50 % v/v while for urine samples it is 90/10 % v/v[125].

A chemical shift referencing standard such as 3-trimethylsilylpropionic acid (TSP) or 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), with a single peak defined as being at a chemical shift of 0.0 ppm, is typically included in the buffer solution. In addition to chemical shift referencing a global internal standard like TSP and DSS can also be used for metabolite quantification[132]. However, depending on the sample matrix, global internal standards like DSS and TSP can influence calculated metabolite concentrations by binding to proteins leading to inaccurate determination of metabolite concentrations[130,133]. If the use of internal standards is not applicable, the anomeric proton glucose signal of α-glucose at 5.20 ppm can be used for chemical shift referencing and alignment as glucose is minimally affected by pH fluctuations.

Serum and plasma samples contain proteins. In addition to the pre-processing procedure, where phosphate buffer is used for dilution and pH correction, SOPs involving ultrafiltration or precipitation with different solvents are applied in the metabolomics field[133]. The two latter pre-processing procedures aim to remove proteins from the samples. The removal of proteins facilitates metabolite identification and quantification and reduces viscosity. However, the additional sample handling may introduce unwanted variation[134].

NMR spectroscopy

NMR spectroscopy is the golden standard of small molecule structure elucidation and is thus very suitable to identify and/or investigate the structure of unknown molecules. The technique utilizes the inherent magnetic properties of nuclei of certain isotopes e.g. ^{1}H and ^{13}C. When placed in a strong magnetic field the nuclei align with or against the field according to the Boltzmann distribution. A subsequent irradiation with electromagnetic energy of appropriate frequency can be absorbed by such 'NMR active' nuclei of a metabolite through a process called magnetic resonance. When the nuclei relax back to equilibrium, an oscillating signal (free induction decay) can be detected in coils situated around the sample. The consequent NMR spectrum is attained by Fourier transformation of the detected time-domain signal to give a frequency spectrum.

The frequency of a given signal is generally transformed to a magnetic field strength-independent entity which is defined as the chemical shift, and this is expressed in parts per million (ppm). The chemical shift of a signal depends on the local environment of the nucleus generating the signal. The local environment is dependent on the density of electrons around the nucleus. The methyl signal from molecules like TSP and DSS is by definition set to 0 ppm and is used as internal standard to calibrate the chemical shift scale. A signal from the same metabolite can be split into several peaks in the spectrum. This phenomenon is called scalar coupling and is caused by the magnetic spin-spin effect from non-equivalent nuclei found two or more bonds from the nucleus/ei producing the signal. The area under the curve (peak) is directly proportional to the concentration of the metabolite in a sample. However, the absolute quantitation is only exactly true if the data acquisition is performed in such a way as to allow for complete magnetic relaxation between individual scans. In practice, for metabolomics analyses such long inter-scan delays (>10 s) are rarely used as the overall sample throughput would be severely hampered. For a comprehensive text on the NMR methodology, see Levitt (2001)[135].

When analyzing serum or plasma samples without prior removal of high molecular weight molecules like proteins and lipo proteins during the pre-processing of samples, an experimental relaxation filter needs to be employed to counter the influence of the components on the NMR spectrum. The signals from proteins and lipoproteins generate broad signals of high intensity in a spectrum and without an experimental filter for suppression, e.g. a T2 filter such as Carr-Purcell-Meiboom-Gill (CPMG) pulse train, these signals influence the detection and resolution of signals the low molecular weight metabolites[130,136]. However, any T2 filter such as CPMG also attenuates, the overall signal to noise (S/N) ratio of the resulting spectrum[130].

1D & 2D NMR

1D ^{1}H NMR experiments generate spectra with protons signals from all metabolites in a sample, including spin-spin dipolar couplings. The chemical shift of signals and couplings from individual protons are distributed along a frequency axis and the area under the curve (integral) is directly proportional to the concentration of a given signal (Figure 3). These spectra are used for deconvolution and quantification of individual metabolites.

The J-resolved (Jres) experiment is used to facilitate metabolite identification by separating the dipolar coupling constant and ^{1}H chemical shift information into two separate dimensions of a 2D NMR spectrum. The projection of the 2D Jres spectrum is essentially a decoupled 1D ^{1}H spectrum[137]. In addition, Jres spectra benefits from attenuated signals from macromolecules.

^{1}H-^{1}H total correlation spectroscopy (TOCSY) is a 2D experiment that generates correlations between all protons in a given spin-system. This experiment is useful for identifying protons relating to the same metabolite and can also be used to verify annotations in 1D ^{1}H spectra.

Heteronuclear multiple quantum correlation (HSQC) is a 2D experiment that correlates chemical shifts of a proton (^{1}H isotope) and the carbon (^{13}C isotope) it is directly bound to. Visualized, the proton spectrum is found on one axis and the carbon spectrum on the other. HSQC experiments uses one bound coupling between ^{1}H and ^{13}C and provide cross-peaks between the corresponding proton and carbon.
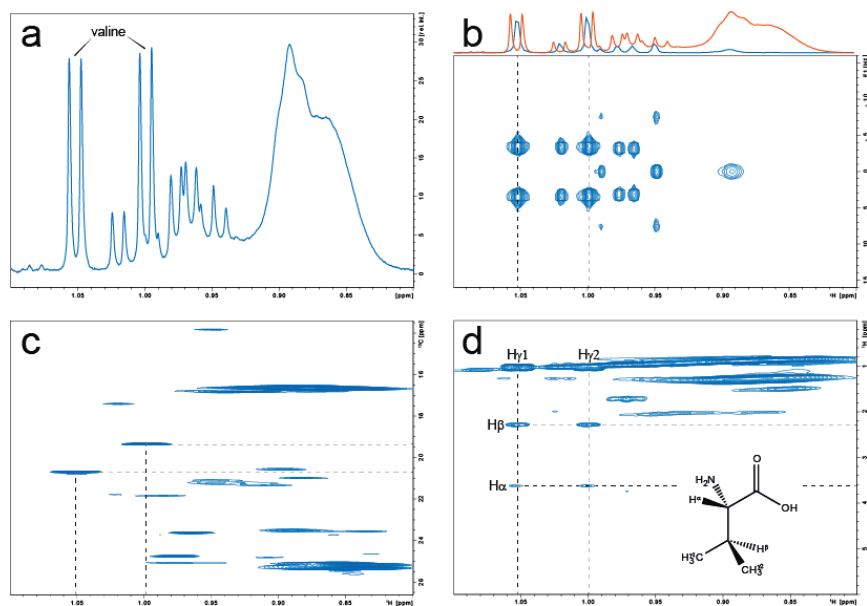
Figure 3. *1D and 2D NMR spectra of valine in a serum sample. a)* ^1H *NMR spectrum. b) 2D J-resolved spectrum with the 1D J-resolved (blue) and 1D* ^1H *NMR (red) projections. The J-resolved projection show single peaks for each of the coupled multiplets while the* ^1H *NMR projection displays the original splitting pattern. The dotted lines denote the splitting pattern of the two valine signals of the* H^γ*. c) HSQC spectrum with* ^{13}C *chemical shift on the y-axis and* ^1H *chemical shift on the x-axis. The* ^{13}C *and* ^1H *correlated chemical shifts can be used to validate annotations of the metabolite using reference spectral databases such as HMDB. The dotted lines indicate the* ^1H *and* ^{13}C *chemical shifts for the valine methyl groups. d) TOCSY spectrum displaying the interrelationship between different protons of the same molecule in the sample. The dotted lines indicate the correlation between the* α, β *and* γ *protons of valine and their chemical shifts.*

1.3.4 Data pre-processing

Alignment and binning

NMR spectra are composed of signals (peaks) related to different metabolites. These peaks differ in intensity, shape and frequency (chemical shift). When using multivariate methods for analysis of NMR data, differences in chemical shift between spectra can impair the analysis. The reason for this is that the data are divided into rows (samples) and columns (variables) where each variable is thought to correspond to the intensity originating from the same peak across all samples[138]. Despite the use of buffer to stabilize pH, variation in chemical shifts of peaks from metabolites affected by pH or various ion concentrations is often seen. To facilitate integration of signal intensities to generate a sensible data table for multivariate analysis, alignment of all spectra in parallel is applied[138]. In combination, binning or bucketing is commonly used to reduce the number of data points (variables) and to correct for misalignments, i.e. placing a given signal originating from a metabolite in the same bin. The reduction of data points is done by grouping spectral features in so-called bins or buckets[139]. The spectral width of each bin normally ranges between 0.01 and 0.05 ppm[140]. In each bin, the area under the spectral curve (intensity) is summed, generating an intensity for the bin instead of the individual data points. Binning can be performed using automated or manual methods[140].

To visualize binning, three different binning methods, manual, conventional and optimized bucketing are shown for a ^{1}H NMR serum spectrum (Figure 4) . Using a manual approach the spectra were aligned for approximately every 3rd peak using icoshift[138] and binning of peaks was performed to a linear baseline on all spectra in parallel using an in-house MatLab (MathWorks, Natick, USA) routine. Peaks were integrated/binned within the chemical shift range of 0.72 – 8.4 ppm. Conventional and optimized binning were performed using the matlab function REF. Here, spectra were aligned to one peak, the glucose signal at 5.20ppm, and with a bucket size of 0.02 ppm. For the optimized binning, a slackness of 0.5 was utilized which corresponds to 0.01 ppm i.e. 50% of the bucket size. Binning was performed in the chemical shift range 14.72 to 5.32 ppm. Buckets in the following ppm-ranges; >8.4, 5.01-4.57, and <-0.65 were discarded for conventional and optimized binning methods. This was based on visual inspection of spectrum overlays of all spectra in Matlab and looking at peak shapes around the water peak. The number of variables/binned regions (data points) for each method is presented in Table 1.

Table 1. Number of data points for different binning methods in 1D ^{1}H NMR serum spectra.

Method	Data points (n)
Raw spectra	65536
Conventional binning	370 (1008 before reduction)[*]
Optimized binning	362 (993 before reduction)[*]
Manual binning	296

[*]Bins in ppm-ranges; >8.4, 5.01-4.57, and <-0.65 were discarded in relation to peak shapes and relevance.

The results of conventional, optimized, and manual binning in the spectral region 3.47 ppm – 3.15 ppm in a ^{1}H NMR serum spectrum are visualized in Figure 4. Manual binning displayed less overlap in single bins, the least number of data points, better alignment, and the possibility to exclude areas only including noise. However, it should be noted that the reproducibility using manual binning is lower in relation to the automated methods and bias concerning small intensity peaks considered as noise or included as a variables is a risk that has to be considered.
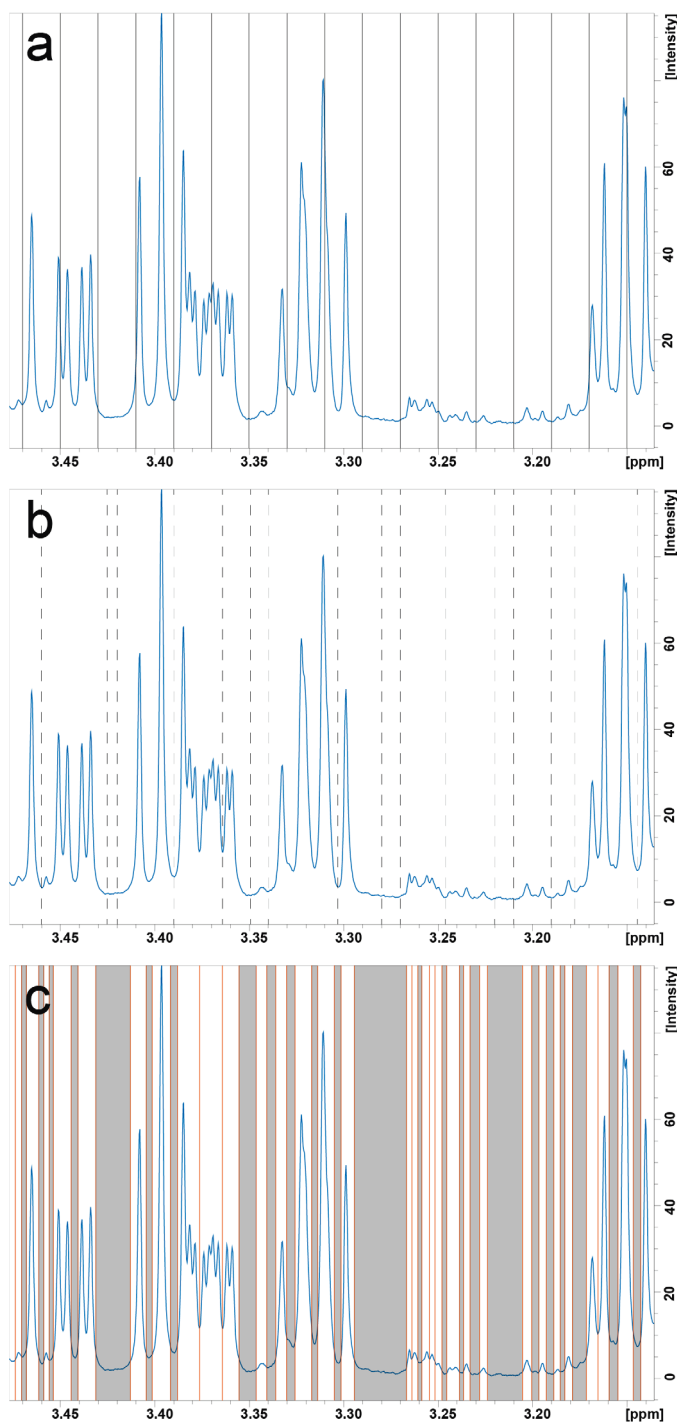
Figure 4. Result of different binning/bucketing methods in a 1D ^1H NMR serum spectrum in the spectral region 3.47 ppm – 3.15 ppm.
a) Conventional binning with bin size 0.02 ppm. Solid black lines denote division of bins. b) Optimized binning with bin size 0.02 ppm and slackness 0.5. Dotted black lines indicate division of bins. c) Manual binning. Red lines indicate division between bins. Grey areas denote areas not included in the analysis.

Normalization and scaling

Normalization is applied to adjust for sample variation ("rows" in the data set) while scaling is applied to battle variation between variables (columns) across samples[141]. The objective of normalization is to reduce differences in overall concentrations between samples arising from experimental variation and variable dilution that makes samples comparable to facilitate data analysis[141-143]. In turn, scaling aims to minimize the influence of variable intensities on multivariate models[144].

Normalization is of special importance in urine samples as the dilution effect is more prevalent than in blood samples where metabolite concentrations are highly regulated[141]. Several normalization algorithms applicable to metabolomics data have been developed[142,143]. As an example, probabilistic quotient normalization (PQN) was introduced in 2006 by Dieterie *et al.*[143]. The method is based on the assumption that changes in the overall concentration of a sample (dilution effects) influence the complete spectrum while changes in concentration of single metabolites that are connected to the biological phenomena studied are assumed to affect only parts of the spectrum[143]. Additional normalization procedures include, among others, the use of quality control (QC) samples, internal or external standards, "non-changing" metabolites, and scaling[102].

Metabolites present in low concentrations are not by necessity of less biological relevance than those found in high concentrations. However, the difference in metabolite concentration across samples in metabolomics data might influence the multivariate analysis in that high concentration metabolites are more likely to be expressed in the model[145]. The reason for this is that projection methods like PCA are based on the maximum variance in the data[145]. To reduce the influence depending on metabolite concentrations (variable intensity) scaling can be applied. Scaling converts the data into differences in concentration relative to a scaling factor[144]. In combination with scaling, mean-centering is often applied where variables are shifted to vary around zero rather than their mean intensities[144]. Centering reduces variation not relating to in-between sample variation[144]. The two most prevalent scaling methods in multivariate data analysis are unit variance (UV) scaling and Pareto scaling. UV scaling divides each variable by the standard deviation of the column and gives all variables equal importance in the model[144,145]. The effects of UV scaling and centering are visualized in figure 5. In turn, Pareto scaling is similar to UV scaling, but each variable is here divided by the square root of the standard deviation that leads to de-emphasis of large size variables, emphasis of medium size variables while small size variables and baseline noise are not amplified[144,145].
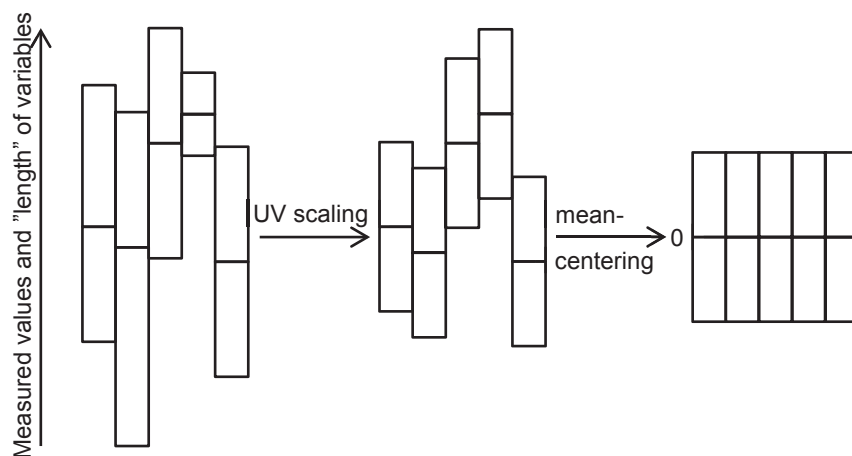
Figure 5 *. Illustration of unit variance (UV) scaling and mean centering. After UV scaling all variables will have the same "length" of intensities and when applying mean centering a mean value of zero. Figure design was inspired by a corresponding figure created by Eriksson et al. (2006)[145].*

1.3.5 Multivariate data analysis

For NMR metabolomics data multivariate statistical methods are often applied for sample comparison analysis[17]. Predominantly principal component analysis (PCA)[146] and algorithms based on partial least squares (PLS) regression and discriminant analysis[147-151] are used for identification of clusters and classification, respectively.

Metabolomics data often comprise of highly correlated variables. Traditional statistical methods such as Students t-test, Analysis Of Variance (ANOVA) and multiple linear regression analyses assume variable independence and are therefore often less suited. In contrast, multivariate projection-based methods benefit from the possibility to compare all variables between samples simultaneously. In addition, the number of variables in a metabolomics data set commonly by far exceeds the number of observations. Multivariate methods use so called dimension reduction where key components containing the maximum amount of information/variation in the data are generated[17]. It is still advisable, however, to use traditional statistics, complementary for validation of single identified metabolites and the robustness of multivariate models.

Multivariate data analysis is often performed using an unsupervised analysis followed by a supervised analysis. PCA is an unsupervised projection-based method that is applied to extract and visualize systematic variation in the data by summarizing the data into underlying trends[34]. Projection based methods like PCA converts multidimensional data into lower dimensional planes called latent variables (LVs) or principal components (PCs). The first PC describes the largest variation in the data, the second PC the second largest variation and so forth. The data matrix is divided into scores and loadings. The scores are shown as "swarms" of data points when PCs are plotted against each other and each data points corresponds to an observation. In turn, loadings refer to the variables (metabolite signals) responsible for the distribution of the scores in the different PCs.

Following PCA, multivariate regression methods are often applied to find variation in the data relating to previously known information i.e. the study question. To identify the variation in the data related to the study question at hand, which usually is only expressed in a small portion of the data, a more focused analysis than PCA is often needed enhancing the separation between groups of observation using a multivariate regression method. Examples of such methods are PLS and orthogonal partial least squares (OPLS), which both can be combined with discriminant analysis (DA). These methods are used for assessing the relationship between the data matrix and one or more response vectors, which normally are based on known meta data for different groups of observations, for example treated individuals and controls[148]. This is accomplished by the rotation of PCs in such a way that maximum separation among classes is attained[17]. In turn, the enhanced separation between groups of classified observations enable improved understanding of which variables are responsible for the separation. The difference between PLS and OPLS lies in that OPLS identifies and filters the orthogonal systematic variation (not related to the response vector) in the data[148]. This property facilitates the interpretation of the data.

Paired samples

Paired data comprise dependent samples that are connected to each other in some way. Examples of dependent samples are pre and post intervention samples or a series of samples from the same individual and matched samples (age, gender etc.). In clinical intervention studies and prospective studies, dependent samples are a common part of the study design. Individuals are acting as their own control or are compared to a matched individual with the aim to measure

the outcome of for example a treatment or a diet. When using paired samples in the study design for metabolomics studies, the statistical tools employed should take into account the dependence of the samples in the analysis. For univariate analysis this problem has been solved using different t-tests appropriate for dependent samples, for example the Wilcoxon signed rank test that also is suitable for skewed data[107,152]. For multivariate statistics common tools like PLS and OPLS-DA are used for modelling and interpretation of metabolomics data. However, PLS and OPLS-DA do not take into account the paired or matched sample information and are therefore similar to independent samples statistical tests. Using independent samples statistical analysis on dependent data will affect the outcome and potentially not find the actual variation, as the individual variation between samples is not considered. Multivariate statistical tools accounting for paired data are however scarce. Multilevel PLSDA[153,154], OPLS-Effect Projections (EP)[151] and the online suite Metaboanalyst[155] are multivariate statistical tools that consider paired data.

1.3.6 Metabolite identification

Identification and quantification of metabolites in NMR spectra can be performed in a manual and/or automated fashion.

When manually identifying metabolites, deconvolution (peak fitting) of peaks in 1D ^{1}H NMR spectra, is generally applied initially. Using the Chenomx NMR suite for deconvolution (Chenomx Inc., Edmonton, Canada) on suitable experimental data, it is also possible to quantify metabolites in parallel if the experiment has been performed according to the parameters recommended by Chenomx. Confirmation of annotations can be done by employing 2D experimental approaches. ^{1}H chemical shifts of 1D NMR spectra can be correlated to ^{13}C chemical shifts in 2D HSQC spectra and the metabolite spin system/-s investigated by ^{1}H-^{1}H correlations in 2D TOCSY. STOCSY is a means of correlating a given ^{1}H peak in a spectrum with other peaks of the dataset, thus allowing the potential of identifying which peaks belong to the same metabolite but can also be a consequence of two or more metabolites that are connected in e.g. a metabolic pathway and thus vary accordingly. 2D J-resolved spectra can be used to elucidate the particular coupling pattern by processing the data so as to put the coupling pattern into a pseudo-second dimension, allowing also the identification of weak peaks otherwise hidden beneath heavy overlap in the regular ^{1}H spectrum[156]. Experimental data like this in combination with spectral information in databases such as the Human Metabolome DataBase (HMDB)[157], the

BioMagResBank [158], and Birmingham Metabolite Library[159] are used to assign and confirm metabolite identities. In addition, spiking with authentic standards can be applied for identification when the methods mentioned above are not sufficient e.g. in the identification of low concentration metabolites[160].

In addition to manual identification of metabolites, there are several automated or semi-automated methods available for NMR data. These include, among others; complex mixture analysis by NMR[161], ccpnmr analysis program produced by collaborative computing project for NMR[162], BAYESIL (identification and quantification)[163], pattern recognition-based assignment for metabolomics[164], Bayesian automated metabolite analyzer for NMR (deconvolution and quantification)[165], the urine shift predictor (identification and quantification)[166], automatic statistical identification in complex spectra (identification and quantification)[167], and automated quantification algorithm quantification of targeted metabolites in human plasma[168].

There are several aspects to consider in the identification and quantification of metabolites in NMR spectra from biofluid samples.

1: The identification of low concentration metabolites. ^{1}H NMR has a lower limit of detection of metabolites in the 1-5 μM range[139]. However, the annotation of low concentration metabolites cannot always be confirmed with ^{13}C chemical shifts in 2D HSQC spectra. This is a consequence of the natural abundance of the NMR-active ^{13}C isotope, which is 1.1 %, essentially meaning that only 1/100th of the potential carbon signal in a given sample generates data. To increase the S/N ratio of peaks in experiments using ^{13}C correlation (e.g. ^{1}H,^{13}C-HSQC), acquisition is typically run with more increments in the indirect dimension (^{13}C) and with an increased number of collected scans. Even so, annotations of low concentration metabolites might be difficult to confirm without authentic standard spiking experiments.

2: The variation in chemical shifts in ^{1}H NMR spectra of biofluid samples caused by differences in pH and salt concentrations that is especially apparent in urine has to be tackled[166,169,170]. This variation complicates not only the identification of peaks in 1D NMR spectra using peak fitting (deconvolution) in individual samples, but also the identification of the same peak over multiple samples[171]. The two doublet peaks of citrate constitute one such an example where its signals at 2.70 ppm and 2.56 ppm changes between samples in relation to pH and interactions with ions in the sample matrix[156]. The use of buffer is applied for both serum/plasma and urine samples to minimize changes in relation

to pH and ionic strength[125,170,172]. However, the use of buffer does not fully resolve the variations in peak positions and therefore complementary methods have been proposed[166,169]. Imidazole can be used as an internal pH indicator and thus allow the identification of shifting peaks if their spectra at different pH values are known[173,174]. However, imidazole is foremost applicable as a pH indicator in the chemical shift range between 8.6 – 5.8 ppm[173]. Takis *et al.* (2017) propose a method where the concentration of metabolites might be used to predict chemical shifts and chemical shifts to predict pH and ionic concentrations. The resulting algorithm uses five navigator signals to predict chemical shifts of 63 metabolites. In combination, a strong relationship between the imidazole signals of L-histidine and pH was shown in the same study[166].
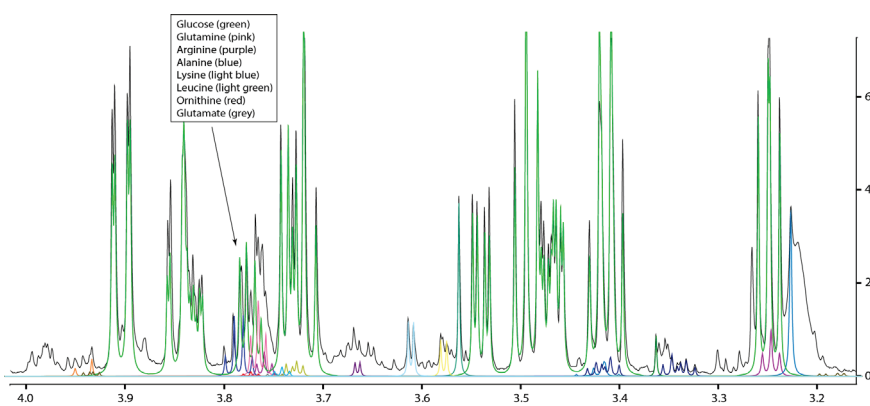

Figure 6. Deconvoluted peaks in the spectral region 4.0 – 3.2 ppm in a ¹H NMR serum spectrum. In this region, overlapping peaks are especially apparent why the quantification and identification of biological relevant metabolites is hampered. The black spectral line indicates the original spectra and differently coloured lines denote fitted peaks from different metabolites.

3: Overlapping peaks in NMR spectra[165]. When peaks of different metabolites are overlapping, although there can be good resolution in 2D spectra, it is difficult to identify metabolites of interest from a biological perspective. The chemical shift ranges between 4.1 and 3.2 ppm in spectra of serum/plasma constitute a significant challenge in regards to peak overlap. Here, peaks of glucose with relative high intensity overlap with a number of other metabolites (Figure 6). The crowding of peaks and the relatively large area under the curve of glucose peaks aggravate identification of other biologically relevant metabolites in this area[163]. If a metabolite displays several peaks in the spectra, the statistical significance of a distinctive peak in relatively uncrowded spectral region might be a

hint of relevance for the biological interpretation. Furthermore, besides generating difficulties in metabolite identification, overlapping peaks is an obstacle for quantification[168]. Owing to this, improvements in accuracy of manual spectral fitting software, relying on experimentally observed signals, have been one approach to tackle quantification of overlapping metabolites[163]. In addition, an automated high throughput method, including quantification of 67 plasma metabolites, using one peak per metabolite for quantification was recently developed by Röhnisch *et al*. (2017), but that method requires sample ultrafiltration and is thus not very suitable for large sample sets[168].

1.3.7 Biological interpretation

Clusters of observations in statistical models (unsupervised and/or supervised) related to previously known information (meta data) of biological relevance for the sample set or in relation to the study question at hand together with the identification of key metabolites changing in concentration between groups are connected to information of metabolite interactions in different pathways. Depending on the metabolite, the interactions could be related to either change in endogenous metabolism in relation to exogenous perturbations or as metabolites of interest as biomarkers of a certain factor. An example is outlined in Figure 7. The connection between key metabolites and their interaction in different pathways can be established by the use of databases like, the Kyoto encyclopedia of genes and genomes (KEGG)[175] and small molecule pathway database (SMPD)[176]. Metabolic pathway information is of particular interest when studying changes in the endogenous metabolism in relation to a certain perturbation connected to potential health outcomes.
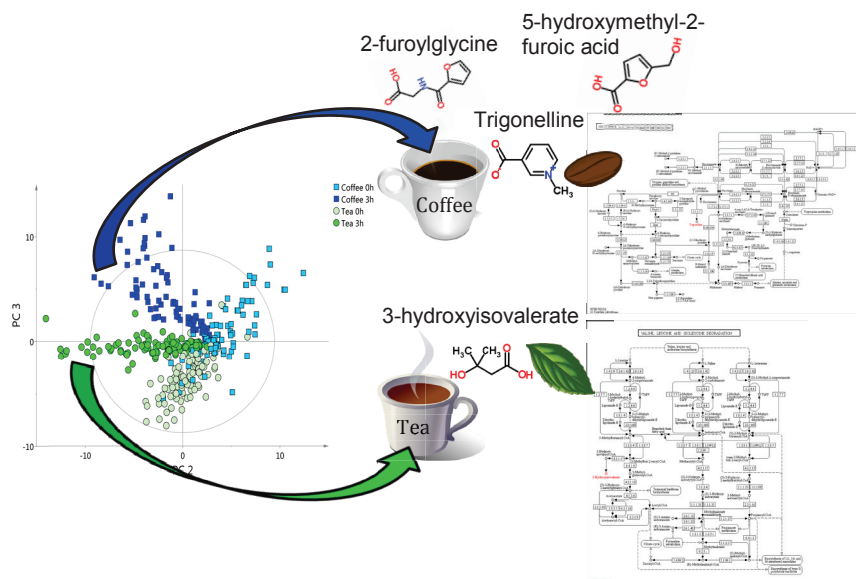
Figure 7. Principal component (PCA) model of urine samples from coffee and tea consumers. In paper I a difference was identified between coffee- and tea drinkers. This was shown in both fasting and postprandial samples. However more so in postprandial samples where 8.0% of the explained variation in the PCA model was related to beverage consumption. Several metabolites increasing in relation to coffee consumption were identified; trigonelline, 2-furoylglyione and 5-hydroxymethyl-2-furoic acid (sumiki's acid). In contrast, 3-hydroxyisovalerate was identified in increased levels in tea drinkers. Investigating 3-hydroxyisovalerate, a component found in tea, further it was found to be a metabolite of leucine degradation generated during fermentation of tealeaves. In addition, trigonelline is generated from nicotinate metabolism.

2. Aims

The long-term goal of this research program is to characterize individual food intake and dietary intake patterns, reflecting acute as well as habitual dietary intake using metabolomics. My PhD project aimed to contribute to this larger picture by investigating how the metabolic phenotype is modulated by different meals.

The overall aim of this thesis was to investigate the acute metabolic responses to food intake in two controlled cross-over dietary intervention studies using ^{1}H NMR metabolomics. In addition, a methodological study investigating pre-analytical deproteinization methods for high-throughput NMR-based serum metabolomics for large sample series was performed.

Aims of the first meal study (paper I & II):

The first meal study was mainly focused on the methodological aspects of the NMR metabolomics workflow for nutritional intervention studies. The overall aim was to assess the discriminative potential of postprandial metabolic profiles between two equicaloric breakfast meals (cereal vs. egg and ham breakfasts) with the same macronutrient distribution, in urine and serum, on several occasions.

Specific aims were:

To assess reproducibility of metabolic profiles between meal occasions.

To identify metabolites responsible for discrimination between meals.

To assess if the metabolic effects of a breakfast meal could be traced one day after ingestion.

To investigate different multivariate analysis tools in relation to the cross-over design.

Specific aims of the second meal study (paper III):

To evaluate if postprandial metabolic profiles could be distinguished between breakfast meals composed to correspond to vegan, lacto-ovo-vegetarian and omnivore diets using ^{1}H NMR serum metabolomics.

To identify metabolites increasing or decreasing in concentration in response to the three meals.

Specific aims methodological study (paper IV)

To compare standard serum dilution protocol for NMR based metabolomics with precipitation and ultrafiltration techniques.

To enable high-throughput processing of serum/plasma samples while keeping metabolite information content and preserving reproducibility.

3. Materials & Methods

3.1 Paper I & II

Ethical approval, recruitment and subject screening
The project was approved by the Regional Ethical Review Board in Gothenburg, Sweden (reference number 561-12), adhered to the Helsinki Declaration, and registered with ClinicalTrials.gov (identifier: NCT02039596).
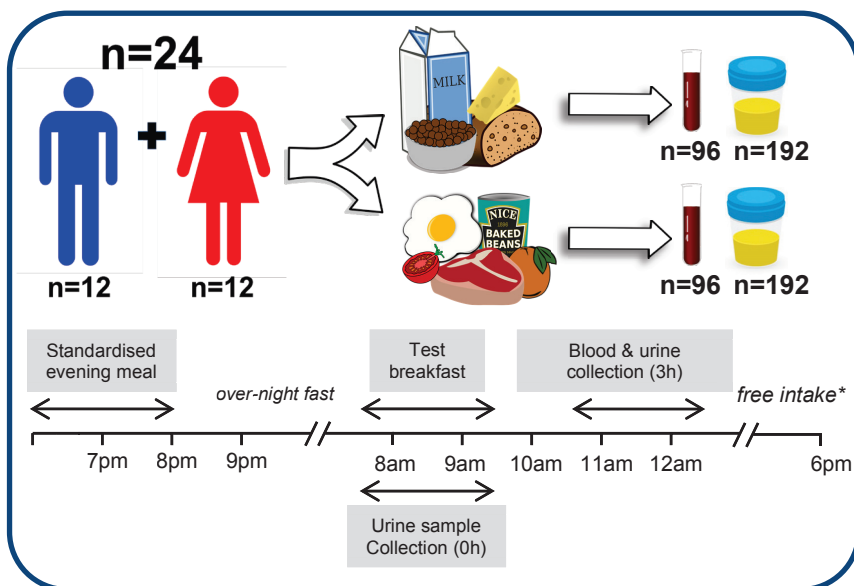

Figure 8. Overview of the study design used in paper I and II. All included volunteers consumed each breakfast meal four times in a stratified order, Tuesday to Friday during two consecutive weeks. Fasting (0h) urine and postprandial (3h) urine and blood samples were collected.

Volunteers were recruited by advertisement at the University of Gothenburg, Sweden, and Chalmers University of Technology in Gothenburg, Sweden. Before entering the study, participants provided written informed consent.
In total, 24 healthy volunteers, 12 males and 12 females, were enrolled in the study (Table 2). Volunteers were considered suitable if apparently healthy (normal serum electrolytes, iron status, creatinine, liver transaminases, bilirubin and alkaline phosphatase, C-reactive protein, plasma glucose, and thyroid status), with no regular use of medications (contraceptives were permitted), and BMI

>18.5 and <30 kg/m^2. Screening included a short lifestyle questionnaire and a three-day weighed-food diary. Body composition was measured with bioimpedance (ImpediMed Bioimp Version 5.3.1.1).

Exclusion criteria included: aged <18 or >65 years, pregnancy or lactation, use of nicotine, natural remedies and/or herbal tea, alcohol consumption higher than 5 units per week (1 unit = 12 g alcohol), allergies to food items included in the study, and the practice of an extreme diet or intent to change physical activity and/or dietary habits before or during the intervention.

Table 2. Anthropometric characteristics of included volunteers (n=24)

Characteristics	Males (n=12)		Females (n=12)	
	mean±SD	min/max	mean±SD	min/max
Age (year)	27.3±11.2	19.0/54.0	24.4±8.2	18.0/46.0
Height (cm)	184.3±6.0	172.0/192.0	169.2±6.1	159.0/177.0
Body weight (kg)	77.5±7.8	66.8/91.3	66.2±7.1	56.6/77.4
BMI (kg/m2)	22.8±2.1	20.6/26.9	23.1±2.3	19.5/26.7
Fat mass (%)	13.9±5.9	6.7/24.7	26.9±5.2	18.4/34.0

Study design

Study participants consumed the two different breakfasts, cereal breakfast (CB) and egg and ham breakfast (EHB), four times each, in the following order; abba/baab, where a = CB and b = EHB, Tuesday to Friday during two consecutive weeks (eight occasions in total). Breakfasts were consumed at the Department of Internal Medicine and Clinical Nutrition, University of Gothenburg, Sweden.

The CB consisted of orange juice, oat puffs with milk, and a rye bread sandwich with hard cheese and fresh tomato. The EHB consisted of orange juice, scrambled eggs, white beans in tomato sauce, fried pork loin, tomato and toasted white bread with orange marmalade. Study participants choose either coffee (male n=6, female n=4) or tea (male n=6, female n=8), both with 20 ml of milk. Participants were also given a choice between a large (750 kcal) (male n=12, female n=1) or a small (500 kcal) (female n=11) breakfast size.

The two breakfast meals had similar composition of protein, fat and carbohydrates (see supplementary information in *paper I*). Two weeks before and during the intervention, study participants were asked to refrain from using dietary supplements and occasional medications. The day before and during the intervention, volunteers were asked to abstain from drinking alcohol, engaging in

strenuous exercise (>2 h moderate intense physical activity, defined as 3–6 MET:s (metabolic equivalents))[177], and eating fish.

To help stabilize background metabolic profiles further, a standardized evening meal of quenelles with tagliatelle in tomato sauce (488 kcal) was provided to be consumed between 18:00 and 20:00 h (Figure 8). Volunteers were instructed to drink water for the evening meal, not eat anything further and only drink water before arriving to the test kitchen between 07.30 and 09.30 h. During the intervention, volunteers noted health status, occasional medications, and exact time of evening meal together with water intake during the overnight fast.

Sample collection and preprocessing
Urine
Fasting (0h) and postprandial (3 h ± 22 min) urine was collected in 500 mL clean plastic cups. In total, 384 samples were collected, n=192 fasting and n=192 postprandial. 10 mL, of the total collected volume of urine, was transferred to a plastic tube and centrifuged at 4°C at 3800 rpm for 10 min to spin down potential cells in the sample. Urine supernatant was aliquoted in 1.8 mL cryo vials and placed at -20 °C within 2 h, and at -80 °C within 4 h. Samples were stored at -80 °C until analysis. ^{1}H nuclear magnetic resonance (NMR) spectroscopy analysis was performed on all urine samples. Prior to ^{1}H NMR analysis, urine samples were thawed at 4°C overnight and centrifuged at 13400 rpm for 60 s. Thereafter, 540 µL of urine supernatant was mixed with 60 µL of buffer solution (1.5 M KH_2PO_4 in D_2O, 0.1% TSP-d4 and 0.02% NaN_3, pD 7.0) in 5.0 mm NMR tubes (Bruker BioSpin, 96 sample racks for SampleJet) using a Gilson 215 liquid handler. Three samples, representative for the dataset were used for two-dimensional (2D) ^{1}H- ^{13}C HSQC and ^{1}H- ^{1}H TOCSY experiments, were manually prepared in the same manner. The buffer solution contained potassium dihydrogen phosphate (KH_2PO_4) and sodium azide (NaN_3) from SigmaAldrich, deuterium oxide (D_2O) from Cambridge Isotopes and 3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt (TSP-d4) from MerckMillipore.

Serum
Postprandial serum samples were collected after breakfast meals (3h 15 min ± 22 min). In total, 192 samples were collected. Venous blood was drawn into 4 mL Z serum Separator Clot activator tubes (VACUETTE® TUBE Greiner Bio-One), allowed to clot at 4°C for 30 min and centrifuged at 4°C at 2600 x g for 10 min. 400 µL serum was aliquoted in 500 µL cryo vials and placed at -20 °C within 1 h (57 min ± 11 min) and at -80 °C within 2 h. Samples were stored at -80 °C until analysis. ^{1}H-nuclear magnetic resonance (NMR) spectroscopy analysis was per-

formed on all serum samples. Prior to ^{1}H-NMR analysis, serum samples were thawed for 60 min at 4°C, 100 µL serum was mixed with 100 µL phosphate buffer (75 mM Na_2HPO_4, 20% D_2O, 0.2 mM imidazole, 4% NaN_3, 0.08% TSP-d_4, pH 7.4) in a deep well plate. 180µL sample mix was transferred to 3.0 mm NMR tubes (Bruker BioSpin, 96 sample racks for SampleJet) using a SamplePro liquid handling robot (Bruker BioSpin, Rheinstetten, Germany). Samples were kept at 6°C until analysis. For quality control three samples with pooled serum from four individuals in the dataset and three buffer samples were used on each 96 sample rack.

NMR spectroscopy analysis

Urine

All ^{1}H NMR spectra were acquired on an Oxford 800 MHz magnet equipped with a Bruker Avance III HD console with a 5mm TCI probe and a cooled (6°C) SampleJet automatic sample changer for sample handling. All ^{1}H NMR experiments were performed at 310 K. NMR data (1D NOESY) were recorded using the standard Bruker pulse sequence 'noesygppr1d'. The spectral width was 30 ppm, the relaxation delay was 4 sec, including presaturation of the residual water resonance. The acquisition time was 1.36 sec. With a total of 32 scans collected into 65k data points, the measuring time for each sample was 3 min 15 sec. All data sets were zero filled to 128 k and an exponential line-broadening of 0.3 Hz was applied before Fourier transformation. All data processing was performed with TopSpin 3.2pl6 (Bruker BioSpin, Rheinstetten, Germany). TSP-d4 was used for referencing. ^{1}H NMR data were acquired for a total of 383 urine samples.

For annotation, three samples representative for the data set were utilized for natural abundance ^{1}H,^{13}C-HSQC ('hsqcedetgpsisp2.4') and ^{1}H,^{1}H-TOCSY ('mlevgpph19') experiments performed on Bruker Avance III HD spectrometers, operating at either 900 or 800 MHz. Spectrometers were equipped with cryoprobes, the 800 MHz with a 3 mm TCI probe and the 900 MHz with a 5 mm TCI probe. The HSQC spectra were measured with 80 ms acquisition time, a 1.2 s pulse delay, 32 scans and acquisition of 2048 data points (for ^{1}H) and 768 increments (for ^{13}C). The ^{1}H and ^{13}C pulse widths were p1 = 8.5 µs and p3 = 9.3 µs, respectively. The ^{1}H and ^{13}C spectral widths were 20 ppm and 180.00 ppm, respectively. ^{1}H,^{1}H-TOCSY spectra were acquired with the same proton pulse width as for the HSQC, spectral widths of 10 ppm in both dimensions. The acquisition time was 115 ms and the pulse delay 2 s. 32 scans, 2048 points and 384 increments were acquired in the direct and indirect dimensions, respectively. All data were acquired and processed with TopSpin 3.2pl6.

Serum

All ^{1}H NMR spectra were measured on an Oxford 800 MHz magnet equipped with a Bruker Avance III HD console and with a 3mm TCI cryoprobe and a cooled (6°C) SampleJet automatic sample changer for sample handling. All ^{1}H NMR experiments were performed at 298 K. NMR data (1D perfect echo with excitation sculpting for water suppression) was recorded using the Bruker pulse sequence 'zgespe'. The spectral width was 20 ppm, the relaxation delay was 1.34 s. The acquisition time was 2.04 s. With a total of 64 scans collected into 64k data points, the measurement time for each sample was 4 min 19 s. All data sets were zero filled to 128 k and an exponential line-broadening of 0.3 Hz was applied before Fourier transformation. All data processing was performed with TopSpin 3.2pl6 (Bruker BioSpin, Rheinstetten, Germany). TSP-d_4 was used for referencing. ^{1}H NMR data were acquired for a total of 192 serum samples. For annotation, pooled serum from all individuals in the dataset was utilized for natural abundance ^{1}H-^{13}C HSQC ('hsqcetgpsisp2.24') and [1]H-^{1}H TOCSY (' mlevgpphw5') experiments. The ^{1}H-^{13}C HSQC spectra were measured with acquisition times of 63.9 ms (^{1}H) and 50.9 ms (^{13}C), a 3 s pulse delay, 8 scans and acquisition of 2048 data points (^{1}H) in 768 increments (^{13}C). The ^{1}H and ^{13}C pulse widths were p1 = 7.44 μs and p3 = 9.3 μs, respectively. The ^{1}H and ^{13}C spectral widths were 20 ppm and 100.00 ppm, respectively. [1]H-^{1}H TOCSY spectra were acquired with the same proton pulse width as for the ^{1}H-^{13}C HSQC. The spectral widths were13.95 ppm in both dimensions, the acquisition times were 183.5 ms (F2) and 229 ms (F1), the [1]H-^{1}H TOCSY mixing times 80 ms and the pulse delay 2 s. 8 scans were used, 2048 points and 512 increments were acquired in the direct and indirect dimensions, respectively.

Sodium phosphate (Na_2HPO_4), imidazole, and sodium azide (NaN_3) were bought from SigmaAldrich, deuterium oxide (D_2O) from Cambridge Isotopes, and 3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt (TSP-d_4) from MerckMillipore.

Metabolite identification

Urine

Metabolite identification was accomplished by using the 1D NOESY, 2D HSQC, and TOCSY spectra of selected samples representative for the data set. Chenomx NMR suite 8.1 (Chenomx Inc., Edmonton, Canada) was used for spectral line fitting of 1D NOESY spectra. Chemical shifts in 1D NOESY, 2D HSQC, and TOCSY spectra were compared reference spectra in HMDB[178].

Serum

1D proton, 2D ^{1}H-^{13}C HSQC, and 2D ^{1}H-^{1}H TOCSY spectra of pooled serum from all individuals in the dataset were used for metabolite identification. Chenomx NMR suite 8.1 (Chenomx Inc., Edmonton, Canada) was used for spectral line fitting of 1D proton spectra. Chemical shifts in 1D proton, 2D ^{1}H-^{13}C HSQC, and 2D ^{1}H-^{1}H TOCSY spectra were compared with reference spectra in HMDB[179].

Pre-processing and statistical analyses

^{1}H NMR spectra were aligned using icoshift and manual integration of peaks was performed to a linear baseline on all spectra in parallel using an in-house developed Matlab routine. For urine, 185 signals were integrated within the chemical shift range of 0.737 - 9.286 ppm and prior to integration, the water region was removed. For serum, 296 peaks were integrated within chemical shift range of 0.721 – 8.362 ppm. Data were normalized using Probabilistic Quotient Normalization (PQN).

Multivariate data analysis (MVA)[34], including principal component analysis (PCA), projection to latent structures (PLS) and orthogonal projections to latent structures with discriminant analysis (OPLS-DA), was performed using SIMCA software v.13 and 14.1 (Umetrics AB, Umeå, Sweden).

Prior to modeling, all data were centered and normalized to unit variance for urnine and the serum datat was pareto scaled. Cross-validation groups were set to 24 (equal to the number of study participants) and assigned observations based on observation ID for each individual, so that all samples from one individual were left out in one cross-validation round. PCA models were used to explore clustering patterns of observations, trends in the data and outliers. PCA models including both postprandial and fasting samples or postprandial samples only, were analyzed separately. Discrimination of classes according to meals CB/EHB and to coffee/tea was achieved using OPLS-DA. To evaluate the metabolic response to breakfast meals and coffee or tea (urine only), postprandial samples (urine, n=192 and serum n=182) were used in OPLS-DA models. For serum, two variables identified as imidazole and four additional, highly abundant signals from unidentified lipids were removed since they, as a consequence of using Pareto scaling, influenced the model merely on account of the magnitude of the peaks. However, these lipids did not differ significantly between breakfast treatments (Mann Whitney U-test p>0.05) leaving 290 variables included for further analysis. The number of latent variables (LVs) in OPLS-DA models was set according to cross-validation and drop of predictive estimation (Q^2). The

validity of OPLS-DA models was assessed using permutation tests (n=999), cross-validated predictive residuals (CV-ANOVA), visual comparison between scores and cross-validated scores, the cumulative amount of explained variation in the data summarized by the model ($R^2X[cum]$ and $R^2Y[cum]$), and the predictive ability of the model ($Q^2[cum]$).

Top-ranked variables in variable importance (VIP) scores and loadings in the OPLS-DA models, including postprandial samples from both breakfast meals, were used to select class-discriminating variables of interest for annotation for both serum and urine. In addition, S-plot was used for serum samples. For urine, loadings from OPLS-DA models including both fasting and postprandial samples were assessed for the respective breakfast meals separately. The objective was to cross out variables relating to a general metabolic response after consuming a meal. Variables in OPLS-DA models discriminating for coffee and tea were selected in the same manner and annotated with the aid of 2D NMR data. For urine, univariate statistical analysis was performed using Student's t-test and Mann Whitney U-test with Benjamini-Hochberg[180] correction at the 95% level were applied. A variable was considered significant if its loadings absolute value was larger than the confidence interval in the OPLS-DA model and also with a p<0.05. For serum, Mann Whitney U-test and Wilcoxon signed rank test both with Benjamini Hochberg correction at the 95% level were used. A variable was considered significant if p<0.05. These calculations were performed in MatLabR2015a.

3.2 Paper III

Ethical Approvement

The project was approved by the Regional Ethical Review Board in Gothenburg (reference number 561-12) and registered with ClinicalTrials.gov (identifier: NCT02039596).
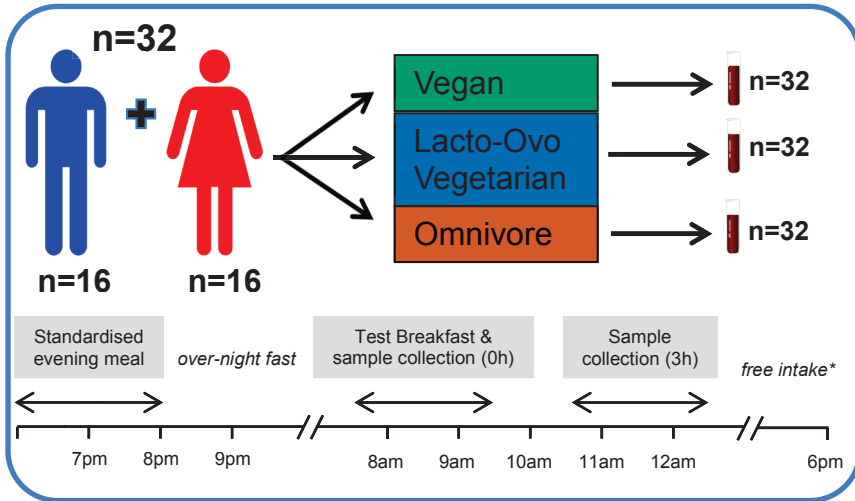

Figure 9. Overview of the study design used in paper III. All included volunteers consumed each breakfast meal in a randomized order during three consecutive days. Fasting (0h) and postprandial (3h) blood samples were collected.

Study Participants

Volunteers were recruited in August to October 2013 by way of advertisement at the University of Gothenburg, Sweden, and Chalmers University of Technology in Gothenburg, Sweden. Before entering the study participants gave written informed consent.

In total, 32 healthy subjects, 16 males and 16 females, were enrolled in the study (Table 3). Volunteers were assessed for suitability by screening, which included a short lifestyle questionnaire, a three-day weighed-food diary, clinical chemistry tests including blood hemoglobin levels, serum electrolytes, iron status, vitamin B12 and folate, calcium, magnesium, creatinine, liver transaminases, bilirubin and alkaline phosphatase, C-reactive protein, plasma glucose and thyroid status. Anthropometrical measurements such as height, weight and body composition were also included in the screening. Body composition was measured with bi-

oimpedance (ImpediMed Bioimp Version 5.3.1.1). Volunteers were considered healthy if they had normal clinical chemistry, were apparently healthy and with no regular use of medications (contraceptives were permitted), and BMI >18.5 and <30 kg/m^2.

Exclusion criteria included: pregnancy or lactation, use of nicotine, dietary supplements, natural remedies and/or herbal tea, alcohol consumption higher than 5 units per week (1 unit = 12 g alcohol), allergies to food items included in the study, unwilling to consume foods included in the meals and the practice of an extreme diet or intent to change physical activity and/or dietary habits before or during the intervention.

Table 3. Demographic characteristics of volunteers (n=32).

	Model set		Prediction set[1]	
	Males (n=11)	Females (n=9)	Males (n=5)	Females (n=7)
Characteristics	mean±SD	mean±SD	mean±SD	mean±SD
Age (year)	27.0±6.6	25.9±10.1	33.2±13.2	31.9±8.2
Height (cm)	184.1±5.5	168.4±4.7	183.0±4.4	170.9±3.8
Body weight (kg)	79.0±11.0	62.0±5.2	74.5±4.4	60.0±5.4
BMI (kg/m2)	23.3±2.5	21.9±1.9	22.2±0.7	20.5±1.6
Fat mass (%)	14.7±5.4	23.5±3.8	15.1±3.0	21.8±5.5

[1]Samples from individuals lacking a full sample set (samples from all breakfast meals) were used for prediction

Study design
Study participants consumed one out of three breakfasts; vegan (VE), lacto-ovo vegetarian (LOV) and omnivore (OM) in a randomized order during three consecutive days (Figure 9). Breakfasts were consumed at the test kitchen of the Department of Internal Medicine and Clinical Nutrition, University of Gothenburg, Sweden, between 07:30 and 09:30 h. VE consisted of tea with oat-milk, soya based yoghurt containing blueberries, two rye bread sandwiches, one with lentil mash and green pepper topping and one with cashew butter and banana. LOV consisted of tea with milk, fruit yoghurt (raspberry/rhubarb), two rye bread sandwiches, one with cottage cheese and apple and the other with hard cheese and tomato. OM consisted of tea with milk, boiled egg and caviar, two rye bread sandwiches, one topped with ham and red pepper and one with liver pâté and cucumber. Study participants could freely choose large (750 kcal) or small (500 kcal) breakfast size; all but two females and one male chose the small breakfast. Composition of both breakfast sizes are found in supplementary material in *Paper III* Tables S1-S3 and macro nutrients content in Table S4. Briefly, amounts

of carbohydrates, fat and proteins for the small breakfast were 74g, 17g, and 19g, respectively for VE, and 60g, 22g and 24g for LOV and 60g, 26g and 26g for OM. In addition the fiber content of the VE, LOV and OM breakfasts were 9.9g, 6.9g and 5.8g respectively.

To help stabilize background metabolic profiles, a standardized evening meal was provided to be consumed between 18:00 – 20:00h on the evening before the breakfast intervention. The standardized evening meal consisted of Pasta al Pomodora (520 kcal). Study participants were instructed to drink water to the evening meal and arrive fasting to the test kitchen on days of the intervention. All days during the intervention, study participants noted health status, occasional medications, time for evening meal and water intake during the overnight fast. The day before and during the intervention, study participants were asked to abstain from eating fish, drinking alcohol, and engaging in strenuous exercise (>2 h moderate intense physical activity, defined as 3–6 MET:s (metabolic equivalents))[177]. Two weeks before and during the intervention, study participants were also asked to refrain from using dietary supplements and occasional medications. These measures were taken in order to reduce inter- and intra-individual variation.

Sample collection
Serum was collected before (0h) and 3h (190 min ± 11 min) postprandial breakfast meals. In total, 188 samples were collected. Venous blood was drawn into 5 mL BD vacutainer glass tube (BD Hemogard[TM] BD Vacutainer®), turned approximately 5 times, allowed to clot at room temperature for 5 minutes and at 4°C for additionally 30 min and centrifuged at 4°C at 3800 rpm for 10 min. 400 µL serum was aliquoted in 500 µL cryo vials and placed at -20 °C within 1 h (1 h ± 8 min) and at -80 °C within 2 h. Samples were stored at -80 °C until analysis.

Sample preparation and NMR spectroscopy analysis
[1]H-nuclear magnetic resonance (NMR) spectroscopy analysis was performed on all serum samples. Prior to ^{1}H-NMR analysis, serum samples were thawed for 60 min at 4°C, 120 µL serum was mixed with 120 µL phosphate buffer (75 mM Na_2HPO_4, 20% D_2O, 0.2 mM imidazole, 4% NaN_3, 0.08% TSP-d$_4$, pH 7.4) in deep well plate and 180 µL transferred to 3.0 mm NMR tubes (Bruker BioSpin, 96 sample racks for SampleJet) using SamplePro (Bruker BioSpin, Rheinstetten, Germany). Samples were kept at 6°C until analysis. Four pooled serum samples, two Sernorm Human and two buffer samples were used on each 96 sample rack for quality control. Pooled serum from each breakfasts were used for two-dimensional (2D) ^{1}H – ^{13}C HSQC and [1]H-^{1}H TOCSY experiments.

All ^{1}H NMR spectra were measured at 800 MHz using BRUKER Avance HDIII spectrometer equipped with a 3mm TCI probe and a cooled (6°C) SampleJet for sample handling. All ^{1}H NMR experiments were performed at 298 K. NMR data were recorded using the pulse sequence 'zgespe', encompassing a perfect echo sequence with excitation sculpting for water suppression. The spectral width was 20 ppm, the relaxation delay was 1.3 sec and the acquisition time was 2.04 sec. With a total of 128 scans collected into 64k data points, the measuring time for each sample was 8 min 9 sec. All data sets were zero filled to 128 k and an exponential line-broadening of 0.3 Hz was applied before Fourier transformation. TSP-d4 was used for referencing. ^{1}H NMR data was acquired for a total of 188 serum samples.

For annotation, pooled serum samples, from all individuals in the dataset, were utilized for natural abundance ^{1}H,^{13}C HSQC ('hsqcedetgpsisp2.2') and ^{1}H-^{1}H TOCSY ('mlevgpphw5') experiments The HSQC spectra were recorded with a 3 s pulse delay, 32 dummy scans, 8 scans and acquisition of 2048 data points (for ^{1}H) and 1024 increments (for ^{13}C). The ^{1}H and ^{13}C pulse widths were p1 = 7.4 μs and p3 = 9.3 μs, respectively. The ^{1}H and ^{13}C spectral widths were 20 ppm and 100.00 ppm, respectively. ^{1}H-^{1}H TOCSY spectra were acquired with the same proton pulse width as for the HSQC, spectral widths of 14 ppm in both dimensions. The pulse delay was 2 s, 32 scans were recorded and 4096 points and 512 increments were acquired in the direct and indirect dimensions, respectively. All data processing was performed with TopSpin 3.2pl6 (Bruker BioSpin, Rheinstetten, Germany). Sodium phosphate (Na_2HPO_4), imidazole and sodium azide (NaN_3) were from SigmaAldrich, deuterium oxide (D_2O) from Cambridge Isotopes and 3-(trimethylsilyl) propionic-2,2,3,3-d$_4$ acid sodium salt (TSP-d$_4$) from MerckMillipore. Seronorm Human was from Sero A/S

Pre-processing and statistical analyses
In total, 14 samples were excluded before data processing and analysis due to poor quality. Hence, 174 samples were further processed. ^{1}H NMR spectra were aligned using icoshift and manual integration of peaks was performed to a defined baseline on all spectra in parallel using an in-house Matlab software application. In total 196 variables were integrated within chemical shift range of 0.860 – 8.447 ppm. Prior to modelling, all data were centered and scaled to unit variance.

Multivariate data analysis (MVA), using principal component analysis (PCA), orthogonal (2) projections to latent structures with discriminant analysis (O2PLS-DA)[150] and orthogonal projections to latent structures with effect pro-

jection (OPLS-EP)[151] were performed using SIMCA software v.15 (Umetrics AB, Umeå, Sweden).

PCA models were used to explore clustering patterns of observations, trends in the data and outliers. Two different models were generated. One including post-prandial (3h) samples and one including samples generated from a calculated effect matrix where the value of each individual and variable at time point 0h were subtracted from time point 3h for each breakfast. These models were also used for O2PLS-DA. No outliers were identified in the model including 3h samples while 3 samples were removed as outliers using Hotellings T2 range (T2Crit 99%) and Distance to Model (DModX) (DCrit 0.05) from the effect matrix. In total 60 and 57 samples were included in the O2PLS-DA models including 3h samples and the effect matrix samples, respectively. Samples from individuals lacking a full sample set (samples from all breakfast meals) were used for prediction (Table 3). Five OPLS-EP models were generated. Using the effect matrix calculated for the O2PLS-DA model, the individual response from each breakfast meal was investigated, each model including 19 individuals. Additionally, effect matrices were calculated from 3h samples between the LOV and the VE breakfast and from the LOV and OM breakfasts. Detailed model statistics are presented in supplementary material in *paper III*, Table S5.

Separation and classification of metabolic profiles between different breakfasts were evaluated using O2PLS-DA. PLS, OPLS and O2PLS divide the variability of the data into a systematic (R^2X) and a residual part. In OPLS and O2PLS the systematic part in turn is divided into a predictive (R^2Xpred) and an uncorrelated orthogonal (R^2Xorth) part. The difference between the two is that OPLS only returns one predictive component and O2PLS returns two. When comparing only two groups OPLS is often used since one component is sufficient to distinguish two groups apart. In the case of three groups it is reasonable to use two predictive components, allowing the samples to distribute over a predictive plane rather than a single line. Therefore O2PLS was used when modelling the VE, LOV and OM groups jointly in the same model.

The validity of the O2PLS-DA models were assessed using default cross validation (every 7th sample), permutation tests (n=999), Coefficient of Variation-ANalysis Of VAriance testing of Cross-Validated predictive residuals (CV-ANOVA), the cumulative amount of explained variation in the data summarized by the model ($R^2X[cum]$ and $R^2Y[cum]$) and the predictive ability of the model ($Q^2[cum]$) visual comparison between scores along the predictive ability of external samples (prediction set).

OPLS-EP models were used to identify discriminating metabolites between fasting and postprandial samples for each breakfast. Additionally, OPLS-EP models were used to identify differences in postprandial (3h) samples between the LOV and VE and between the LOV and OM breakfasts. Prior to modeling, all data were scaled to unit variance. The validity was assessed using default cross validation (every 7th sample), CV-ANOVA, the cumulative amount of explained variation in the data summarized by the model (R^2X[cum] and R^2Y[cum]) and the predictive ability of the model (Q^2[cum]), visual comparison between scores and the response vector (Y). Discriminating variables were selected using loadings (pq >0.1) and top ranked variables in variable importance (VIP) scores in combination.

The average fold change for each variable intensity was defined and calculated. Firstly, a vector I0 having the intensity values for each individual at the first time point, and a corresponding vector I3 having the corresponding intensity values at the second time point, were constructed. Secondly, a complex vector $Z = I0 + i*I3$, where i being the imaginary unit, was calculated. Thirdly, phase angles of each element theta = angle(Z) and the mean of theta, mean(theta) were calculated and mean(theta) defined to be the phase angle of the average fold change giving equal weight for each individuals fold change irrespective of peak intensities, average fold change = tan(mean(theta)). The reason for using the mean of the phase angles rather than for example the mean of log2 values of the fold changes was to avoid numerical problems since for some individuals some measured peak intensities were very small and could not be predicted accurately or had been set to zero. In addition, p-values were calculated using the differences of I3-I0 rather than the angles theta, using Wilcoxon signed rank test with Benjamini Hochberg correction at the 95% level. A variable was considered significant if $p<0.05$. These calculations were performed in MatLabR2015a.

Metabolite identification
1D proton, 2D ^{1}H-^{13}C HSQC, and 2D ^{1}H-^{1}H TOCSY spectra of pooled serum from all individuals in the dataset were used for metabolite identification. Chenomx NMR suite 8.1 (Chenomx Inc., Edmonton, Canada) was used for spectral line fitting of 1D proton spectra. Chemical shifts in 1D proton, 2D ^{1}H-^{13}C HSQC, and 2D ^{1}H-^{1}H TOCSY spectra were compared with reference spectra in HMDB[179].

3.3 Paper IV

Materials and samples

Unless otherwise stated, all chemicals were from Sigma-Aldrich. Ultrafiltration tubes were from GE Healthcare (VivaSpin 500 3k and 10k MWCO) and Millipore (Amicon Ultra 0.5 ml Ultracel 3k and 10k MWCO). Ultrafiltration plates were from Pall LifeSciences (AcroPrep Advance Omega 3k and 10k MWCO). Serum was from a single healthy individual, collected at one occasion, aliquoted, flash-frozen in $N_2(l)$ and stored at -86°C for further use.

Ethical approval

The project was approved by the Regional Ethical Review Board in Gothenburg (reference number 976-16). One female volunteer was recruited and provided informed written consent.

Sampling

Venous blood was collected in silica coated vacuum tubes with gel, allowed to clot at 4°C for 30 min and thereafter centrifuged at 4°C at 3800 rpm for 10 min in a Jouan CR3i centrifuge with a T-40 swing-out rotor. The resulting serum was pooled and 300 µL aliquoted in 500 µL cryo vials and placed at -20 °C within 1 h and at -80 °C within 1.5 h. Samples were stored at -80 °C until analysis.

Sample preprocessing

Serum samples in 500 µL cryo vials were thawed at room temperature for 10 min and were subject to either dilution (DIL), methanol-precipitation (PREC) or ultrafiltration (UF) before transfer to SampleJet NMR 3 mm tube racks using a SamplePro Tube L liquid handler (Bruker BioSpin).

For the DIL protocol, a 50/50% v/v dilution with buffer A (75 mM sodium phosphate, pH 7.4, 20% v/v D_2O, 0.02% NaN_3, 500 µM DSS-d_6, 250 µM imidazole) was used. Using a 96-channel Bravo liquid handler (Agilent) 100 µl thawed serum was mixed with 100 µl buffer A in a 96-well deepwell plate (DWP; Porvair cat. no. 219009, 2 ml polypropylene, square wells) (Figure 10). 180µL sample mix was then transferred to 3.0 mm NMR tubes as described above. Samples were kept at 4-6°C during preprocessing and until analysis. Methanol precipitation was performed essentially as described in[160]. Briefly, a 96-channel Bravo liquid handler (Agilent) was used to mix 150 µl thawed serum with 750 µl cold (-20°C) methanol in a 96-well deepwell plate (DWP; Porvair cat. no. 219009, 2 ml polypropylene, square wells). The Bravo robot operated with 250 µl filter tips, necessitating several pipetting steps for the methanol addi-

tion. The Bravo software procedures used in this work is available upon request. The plate was sealed with a piercable sealing cap (Porvair cat. no. 219004) and shaken at 12°C for 30 min at 800 rpm in an Eppendorf thermomixer, placed at -20°C for 30 min and thereafter spun at maximum speed in an Eppendorf 5804R centrifuge with an A-2-DWP rotor (2250 x g) for 60 min at 4°C. 600 µl of the supernatant was transferred to a new DWP with the Bravo liquid handler. After transfer was finished, the receiver plate was dried in a Labconco Centrivap (Labconco Corp.) lyophilizer at 20°C over night. The resulting pellets were first washed with 50 µl of deuterated methanol, dried in the lyophilizer again for 1h and then dissolved in 200 µl buffer B (75 mM sodium phosphate, pD 6.95, 100% D_2O, 0.02% NaN_3, 500 µm DSS-d_6, 1 mM imidazole) per well by shaking at 12°C, 45 min, 800 rpm in an Eppendorf thermomixer. Samples were spun down briefly before 180µL sample was transferred to 3.0 mm NMR tubes as described above.
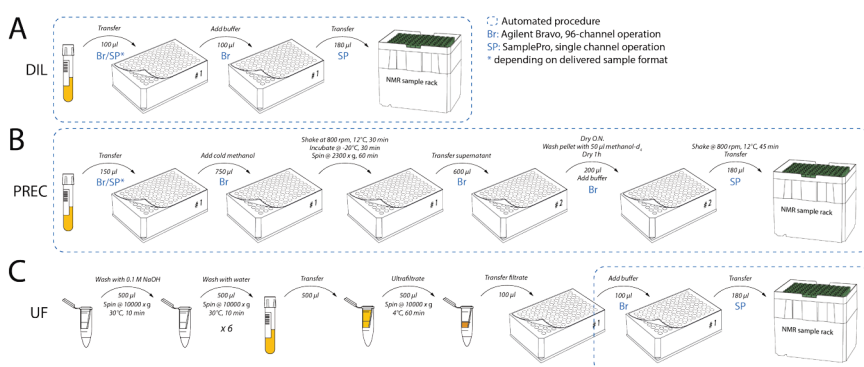

Figure 10. *Workflow for the different methods adapted for automation and large sample series. DIL (dilution), PREC (precipitation) UF (ultra filtration). Whereas all pipetting steps in the DIL and PREC protocols, respectively, are automated, the degree of automation is considerably lower for the UF procedure.*

Ultrafiltration in 0.5 ml tubes, pore filter size 10 k or 3k (Amicon ultra) was performed in an Eppendorf 5804R centrifuge equipped with a fixed-angle FA-45-30-11 rotor. All tubes were washed once with 500 µl 0.1M NaOH by spinning at 14000 x g for 20 min, followed by six times with 300 µl ultra-pure water, all at 30°C. 200 µl thawed serum samples were ultrafiltrated at 14000 x g, 4°C for 1h until 100 µl filtrate could be mixed with 100 µl buffer A and transferred to 3 mm NMR tubes as described above. Ultrafiltration in a 96-well plate format was performed either in a centrifuge, using an Eppendorf 5804R centrifuge equipped

with an A-2-DWP rotor at 2250 x g, 4°C, or with a MultiScreenHTS (Millipore) vacuum manifold operating a 10 inches Hg. Ultrafiltration plates, pore filter size 10 k (Millipore) were washed in the respective format (centrifugation or vacuum). An initial wash using 150 µl 0.1M NaOH was followed by 8 washes with 150 µl ultra-pure water as described above, at 30°C (centrifugation) or at room temperature (vacuum). 275 µl thawed serum and buffer A mix (1:1 ratio) was added to washed wells. Centrifugal ultrafiltration of the plate was run for 2 h. Vacuum was applied and the filtration monitored.

NMR spectroscopy

1D ^{1}H-spectra were acquired on either an Oxford 800 MHz or an Oxford 900 MHz magnet equipped with a Bruker Avance III HD console and a 3mm (800 MHz) TCI cryoprobe, using the pulse sequence 'zgespe', encoding an experiment with perfect-echo J-modulation suppression and excitation sculpting for water suppression including an optional CPMG pulse train of 193 ms as T2 filter. Parameters were 128 scans, a 1.3 s relaxation delay, an acquisition time of 2 s and data was collected in 64k points at 25°C, resulting in a total experimental time of approximately ten minutes. Data processing was performed in TopSpin3.5pl1(Bruker BioSpin), including multiplying the FID with a exponential line broadening function of 0.3 Hz, zero filling, phase correction, baseline correction and referencing to DSS-d$_6$. 2D J-resolved (pulse sequence 'jresgpprqf'), ^{1}H,^{1}H-TOCSY ('mlevgpphw5') and natural abundance ^{13}C-HSQC ('hsqcetgpsisp2.2') were performed for annotation purposes on representative DIL, PREC and UF samples. Parameters for the 2D NMR acquisition are available upon request.

Data analysis

Signal-to-noise measurements were made in TopSpin with the command 'sino', using the region between -1 and -2 ppm as representative of noise. Processed 1D NMR data was imported into MatLab 2016b (MathWorks) and aligned with icoshift[138] and peaks integrated to a linear baseline with an in-house MatLab procedure. Annotation of metabolites was accomplished using ChenomX 8.3 (ChenomX Inc.), 2D NMR data and the HMDB[181]. All multivariate analysis was performed in Simca 14.1 (Sartorius Stedim Umetrics).

4. Results & Discussion

4.1 Methodological aspects

Modeling metabolic profiles
In the intervention studies, metabolic profiles in serum and urine from different breakfast meals were investigated. Using multivariate statistical methods, metabolic profiles were characterized and discriminating metabolites identified. OPLS-DA was used on the urine data and OPLS-DA, OPLS-EP and ANOVA-PLS on the serum data in the first study (*paper I* and *II*), while O2PLS-DA and OPLS-EP were used in the second study (*paper III*).

OPLS-DA does not consider sample dependency, and as such is not adapted for the paired sample structure that is generated by a cross over design. Despite this limitation the results regarding characterization and separation of metabolic profiles between breakfast meals and the selection of discriminating metabolites in serum were fairly comparable to results from the OPLS-EP and ANOVA-PLS models. However, OPLS-DA displayed a less robust model with lower predictive ability (Q^2) and less explained variation in both the X and Y cumulative fractions in relation to the selected latent variables, compared to the OPLS-EP model. In addition, the OPLS-DA model showed lower classification performance compared to the ANOVA-PLS model, which modelled the decomposed variation related to the breakfast meals along with the residual variation. Despite of the above mentioned limitations of OPLS-DA, when modelling both serum and urine, metabolic profiles could be reproducibly characterized and a difference between the EHB and CB with respect to the metabolic profiles could be identified at both the individual and group level. The OPLS-DA model correctly classified 88% and 98% of metabolic profiles from the EHB and CB, respectively, in urine samples while in serum the corresponding numbers were 90% and 92%.

When modelling the VE, LOV and OM groups jointly in the same model, O2PLS-DA was used. This as O2PLS, compared to OPLS, returns two predictive components instead of one and in case of three groups it is reasonable to use two predictive components, allowing the samples to distribute over a predictive plane rather than a single line. In combination, OPLS-EP was used to investigate the individual difference in metabolic response between fasting and postprandial sate. In addition, OPLS-EP was used to capture the difference of the LOV com-

pared to the VE and OM breakfasts. This as OPLS-EP, as shown in *paper II*, effectively manage sample-dependency in cross-over designs when repeated sampling is not applied.

In the PCA model including both fasting and postprandial urine samples the explained variation (42.3%) in PC 1 was related to spectral magnitude (Figure 11). We chose to center and UV scale the data but did not use any computational normalization prior to modelling. Instead we used diet as a normalization factor (standardized evening meal), which has previously been shown to reduce inter-individual variation in urine concentration[182]. Spectral magnitude was to some extent associated with the volume of collected urine. Urine volume has previously been applied as factor for normalization[183]. However, as the urine volume, in our case, did not fully explain the overall concentration it was not used as a factor for normalization. In addition, the total intensity of a spectrum (spectral magnitude) has been used as a normalization factor[141] and this could have been tested in the present case. However, UV scaling was considered sufficient for normalization as this approach generated biologically relevant results. Finally, computational approaches for normalization of urine samples for metabolomics analysis has been developed, and here quantile normalization and cubic spline have been proposed as the preferred method of choice for large (n>50) and small (n<50) datasets respectively[142].
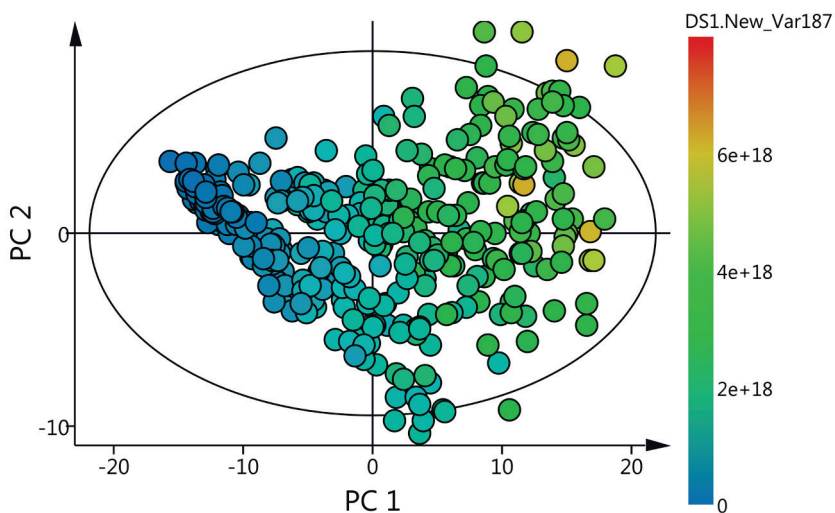

Figure 11. *Principal component analysis model displaying the magnitude of the overall spectra in PC 1. Spectral magnitude (overall concentration) was calculated by the sum of squares of all integrated peaks in the spectra for all samples. The model included 382 urine samples (fasting (n=190) and postprandial (n=192)) from 24 healthy individuals.*

Metabolites in serum and urine

Blood is highly controlled by homeostasis and comprise both endogenous and exogenous metabolites alongside inorganic salts and lipids, lipoproteins and proteins while urine constitute inorganic salts together with water soluble waste products. As such these biofluids may represent different parts of the food metabolome. The advantage of urine collected at one time point can reflect the metabolism over several hours while for blood if not repeated sampling is conducted only a snapshot is captured. Our results imply that the serum and urine metabolomes in response to a meal partially overlap, but are mainly complementary when evaluating the metabolic response to different meals.

Several of the identified metabolites in this study may originate from both endogenous processes and external factors. Creatine is an amino acid synthesized from glycine, arginine and methionine[184]. Creatine is provided equally by diet, where the main dietary sources of creatine are red meat and fish[185], and by endogenous synthesis[186]. Inter-individual variation of creatine levels has been found in relation to difference in muscle mass[187] and gender[188]. Likewise, methanol is considered both an endo- and exogenous metabolite[189]. The origin of endogenous methanol is not fully understood, but has been proposed to originate from gut microbiota metabolism[190] and from a metabolic pathway involving S-adenosylmethionine[191]. Exogenous methanol can originate from metabolic processes depending on the diet[192]. Further, the branched amino acid isoleucine also has both endo- and exogenous origin. It has been associated with inter-individual differences such as age[193] and insulin sensitivity[194]. In addition, differences in both isoleucine and glycine concentrations in blood have been related to inter-individual variation arising from unknown factors[195]. Glycine is a non-chiral amino acid found in high concentrations in legumes, nuts, chicken, and red meat. Glycine has also been identified as an endogenous metabolite in plasma[196]. Notably, metabolites generated from both endogenous and exogenous factors should be used as biomarkers with caution. Levels of inter-individual variation of these metabolites in relation to endogenous processes must be established before changes in concentration of these metabolites may be used as biomarkers of external factors/perturbations.

Here, metabolites are discussed from the perspective of their biological relevance and not as biomarkers of specific foods. This as the study design was intended to identify clusters of metabolites relating to an overall matrix generated from different meals, rather than to identify biomarker of individual foods.

4.2 Paper I & II

In the first intervention study the reproducibility of metabolic profiles in serum and urine from two equicaloric breakfast meals with comparable fat, protein and carbohydrate distribution were investigated. Using multivariate statistical methods, i.e. OPLS-DA, OPLS-EP and ANOVA-PLS, metabolic profiles in both serum and urine were characterized. Here, the presentation is limited to the results from the OPLS-DA model as this is comparable in both the urine and serum sample modelling (Figure 12).
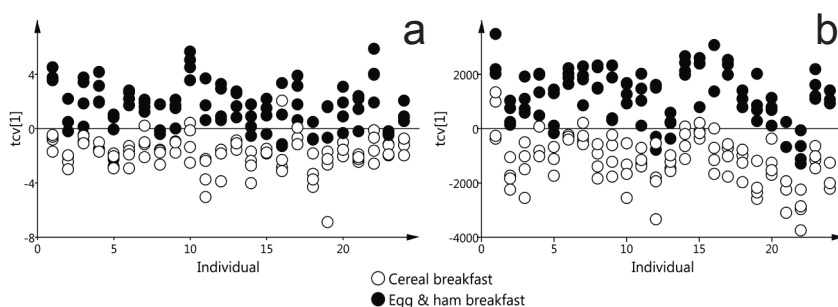

Figure 12. *Breakfast-dependent cross-validated scores (cross-validated x scores (tcv)) in orthogonal projections to latent structures with discriminant analysis (OPLS-DA) models. Models included 24 healthy volunteers who consumed each breakfast at four occasions (eight in total) in a stratified order. a)Model included postprandial (3h) urine samples (n=192) and 185 variables b) Model included postprandial (3 h) serum samples (n=182) and 290 variables.*

Discriminating metabolites
Key discriminatory metabolites were identified in both urine and serum (Table 4). In urine, the EHB displayed overall higher concentration of most variables compared to the CB. Only one identified metabolite, erythrose, displayed higher concentration in urine after consumption of the CB compared to the EHB, while in serum proline, tyrosine, valine, N-acetylated amino acids (NAA) and 3-hydroxybutyrate were identified as discriminating metabolites. For the EHB, creatine, lysine and citrated were higher in urine and creatine, lysine, isoleucine, glycine, choline, glutamine, methanol and 4-aminobutyrate were higher in serum compared to the CB. The relative difference in serum concentrations of a number of selected significantly changed metabolites in postprandial samples from the CB and EHB are presented in Figure 13. Creatine and lysine overlapped between serum and urine as discriminatory metabolites for the EHB.

Table 4. Metabolites positively associated with consumption of breakfast meals and beverages in the multivariate models.

Consumption	Metabolite	Biofluid	Chemical shift (ppm)
CB[1]	Erythrose	Urine	5.29
CB	Tyrosine	Serum	7.19, 6.89
CB	Proline	Serum	4.13-4.15, 3.33-3.35, 2.37-2.33, 2.10, 2.09, 2.07, 2.06, 2.03-1.97
CB	NAA[2]	Serum	2.09, 2.08, 2.06, 2.05, 2.03, 1.99
CB	3-hydroxybutyrate	Serum	1.20
CB	Valine	Serum	1.05, 1.04, 1.00, 0.99
EHB[3]	Creatine	Urine	3.94
EHB	Lysine	Urine	1.68
EHB	Citrate	Urine	2.69, 2.55
EHB	Creatine	Serum	3.94, 3.04
EHB	Lysine	Serum	3.03, 3.02, 1.90-1.92
EHB	Methanol	Serum	3.37
EHB	Isoleucine	Serum	1.97, 1.98, 1.02, 1.01, 0.95-0.93
EHB	4-aminobutyrate	Serum	1.90
EHB	Glycine	Serum	3.57
EHB	Choline	Serum	3.21
EHB	Glutamine	Serum	2.16
Coffee	2-furoylglycine	Urine	6.64
Coffee	Sumiki's acid	Urine	6.48, 4.61
Tea	3 hydroxyisovalerate	Urine	1.28

[1]Cereal Breakfast
[2]N-acetylated-amino acids
[3]Egg & Ham Breakfast

Discriminating metabolites from CB found in urine and serum

For the CB there was no overlap between serum and urine regarding discriminating metabolites, which might have been expected since erythrose and amino acids have previously been detected in both serum and urine. However, high levels of amino acids are not expected in urine, as excess intake of amino acids after deamination are excreted as urea. Erythrose was present at higher concentration in urine samples. The monosaccharide is, in its phosphorylated form (D-

Erythrose 4-phosphate), an intermediate in the pentose phosphate pathway (carbohydrate metabolism). The CB was higher in carbohydrates derived from cereals, i.e. rye bread and oat puffs, compared to the EHB where carbohydrates were mainly derived from white bread and orange marmalade.
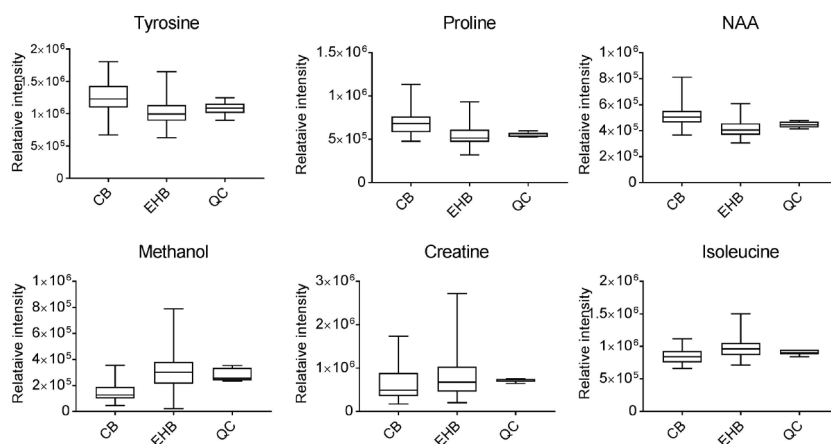

Figure 13. Boxplots of relative concentrations of selected significant (p<0.05) metabolites in paper II. Variation in relative concentrations of tyrosine (signal 7.19 ppm), proline (signal 4.14 ppm), N-acetylated-amino acids (NAA) (signal 2.08 ppm), methanol (signal 3.37 ppm), creatine (signal 3.04 ppm), and isoleucine (signal 0.94 ppm) in postprandial serum samples from healthy volunteers who consumed cereal breakfast (CB) (n=92) and egg & ham breakfast (EHB) (n=90). Variation in quality control (QC) serum samples (n=7) is also shown.

Proline, tyrosine and valine were found at significantly higher concentrations in postprandial serum samples from the CB. Consistently, the CB consisted of a higher content of proline than did the EHB. The hard cheese contributed to the major content of proline in the CB (Supporting information Table S2, *paper II*). 10% of the total amino acid content of cow's milk has previously been shown to be proline [197]. There was little difference in total tyrosine content between the two breakfast meals. Surprisingly, the tyrosine concentration in the postprandial serum samples was significantly higher in the CB. In the CB, hard cheese had the highest content of tyrosine, while in the EHB white beans in tomato sauce and pork loin had the highest content of tyrosine. Further, the EHB had a higher content of valine compared to the CB. Nevertheless, postprandial samples from the CB displayed significantly higher concentrations of valine. The main food that contributed to the valine content in the CB was the hard cheese, whereas in the EHB it was white beans in tomato sauce.

The significantly higher concentration of *N*-acetylated amino acids (NAA) in serum might be related to a higher oxidation rate of fatty acids and/or carbohydrates in the CB. NAA are formed in a reaction with acetyl-CoA where the acetyl group is donated by CoA and attached to the amino acid amide nitrogen. This reaction is catalyzed by the enzyme D-amino-acid N-acetyltransferase (KEGG, reaction R02191). A prerequisite for acetylation of amino acids is the presence of Acetyl-CoA. Acetyl-CoA is mainly produced by the metabolism of carbohydrates and fatty acids (KEGG pathway map01100). The metabolic oxidation rate of fatty acids is dependent on the length of the carbon chain and degree of unsaturation[198]. Short chain fatty acids (12:0) have been shown to be metabolically oxidized more rapidly compared to long chain fatty acids (18:0)[198]. In turn, polyunsaturated fatty acids display an increased oxidation rate in comparison to saturated fatty acids of the same chain length[198].

The CB had comparatively higher content of short chain fatty acids. Consequently, a higher concentration of acetyl-CoA relating to a comparatively increased oxidation rate might be present in serum after consumption of the CB. Another potential explanation for the higher concentration of NAAs after consumption of the CB could be the concentration of free amino acids. In particular glutamate, a major player in amino acid metabolism, might be of interest. *N*-acetylglutamate regulates the urea cycle and the metabolism of amino acids. This is achieved by allosterically activating carbamoyl phosphate synthase, an enzyme that sequesters generated ammonia. An increased synthesis of *N*-acetylglutamate might generate a higher rate of amino acid metabolism and in turn influence nitrogen metabolism. Glutamate was found in two discriminating variables for the CB in serum. However, in our data proline and glutamate are overlapping in the 1D ^{1}H NMR spectra (signals at 2.36 and 2.35 ppm). As variables were integrated (manually binned) in the 1D spectra, it was not possible to separate proline and glutamate. Thus we cannot conclude anything regarding the concentration of glutamate alone. The higher concentration of NAA in postprandial serum samples from the CB, might be explained by the suggestions outlined above. Similar to the glutamate–proline overlap, it was not possible to unequivocally identify which specific N-acetylated amino acids were discriminating and, consequently, the corresponding detailed pathways involved remains to be unraveled.

3-hydroxybutyrate is generated by the conversion of Acetyl-CoA produced by oxidation of fatty acids (KEGG, pathway map00072). This process occurs primarily in the liver mitochondria and is called ketogenesis. Accordingly, the resulting end products (3-hydroxybutyrate and acetoacetate) are referred to as ketones. 3-hydroxybutyrate is transported by the blood and can be used as alter-

native fuels in peripheral tissues where it is converted into Acetyl-CoA. As our breakfast meals were isocaloric and consisted of the same amount of protein, fat, and carbohydrates, the increased concentration of 3-hydroxybutyrate in post-prandial serum samples from the CB is likely linked to the difference in composition of the macronutrients. Saturated medium chain fatty acids such as octanoic acid (8:0) and decanoic acid (10:0) has been shown to produce more ketones per unit of energy than other fatty acids[199]. The CB (500 kcal) included 1.4 g medium chain triglycerides (4:0-10:0) while the corresponding EHB included 0.8 g (Supporting information Table S6, *paper II*). In speculation, the higher content of medium chain fatty acids in the CB might be the factor behind the higher concentration of 3-hydroxybutyrate. 3-hydroxybutyrate has previously been identified in urine[26]. However, here the concentration after three hours did not discriminate between groups which might be related to time of sampling or level of 3-hydroxybutyrate excreted in urine.

Discriminating metabolites from EHB found in urine and serum
Creatine and lysine discriminated for the EHB in both serum and urine. Levels of creatine in plasma and urine have previously been shown to increase after ingestion of red meat[79,200]. Creatine is excreted in urine[201] and elevated levels of creatine in urine have previously been related to dietary consumption of animal proteins[79,202,203]. The increased levels of creatine are believed to originate from the pork loin in that breakfast meal.

The EHB had higher content of all amino acids, except proline, tyrosine, and tryptophan (Supporting information Table S1, *paper II*). White beans alongside the pork loin and scrambled eggs were the main sources of arginine, lysine, alanine, glutamine, methionine, isoleucine and glycine which differed significantly in postprandial serum samples from the EHB compared to the CB while in urine only lysine was found in higher concentration after the EHB. In turn, 4-aminobutyrate is an oxidation product of glutamate that is a hydration product of glutamine and has been identified in both serum and urine.

Methanol was found in increased concentrations in serum samples. Methanol has only been identified in urine after high alcohol consumption. The metabolism of fruit and vegetables, mainly due to the pectin content that is metabolized by the gut microflora, is believed to generate the majority of the serum methanol originating from the diet, other than alcoholic beverages [192,204]. Together with the tomatoes and white beans, the orange marmalade, high in natural as well as added pectin, likely caused the higher methanol concentration after the EHB. In

turn, oranges are high in citrate and this was reflected in the urine. Methanol has only been identified in urine after high alcohol consumption.

Combined, our results suggest that the absorption rate of amino acids differ between individual amino acids and between different foods. This might be related to the ratio between free and bound amino acids as well as or in combination with the metabolic turnover of different foods[71]. In conclusion, proline, isoleucine and glycine were the amino acid with consistent both intake- and serum concentrations. The consumption of orange marmalade was reflected in both urine and serum with increased levels of citrate and methanol respectively; citrate potentially reflecting the citrate content and methanol the pectin content of the marmalade. This implies that a food product can be reflected by different metabolites depending on the biofluid. In contrast, creatine was found in both serum and urine in increased levels after consumption of the EHB, and this might imply a partially similar response to red meat consumption in serum and urine.

Discriminating metabolites from coffee and tea found in urine
The study design of the intervention allowed for study volunteers to freely choose tea or coffee as their hot beverage during the study period. The beverage choice was reflected primarily in the urine metabolome in both fasting and postprandial samples (Figure 7). In the PCA model including only postprandial samples, 8% of the explained variation was found in PC 3 that separated coffee and tea drinkers. 2-furoylglycine and sumiki's acid were found as discriminating metabolites in coffee drinkers while 3-hydroxyisovalerate discriminated for tea consumers (Table 4). The excretion of sumiki's acid (5-hydroxymethyl-2-furoic acid) in urine has previously been associated with consumption of coffee but also with honey, alcohol and dried fruit[205]. 3-hydroxyisovalerate is a degradation product of leucine. Our findings might be explained by the reduction of free leucine in black tea during the manufacturing process (fermentation and drying)[206]. In addition, hippurate and trigonelline were found in both tea and coffee drinkers, however at higher levels in samples from individuals drinking coffee. In support, hippurate has previously been identified in both tea and coffee consumers while trigonelline has been identified in habitual coffee drinkers in both observational and intervention studies[49;51;52]. Likely the turnover rate of these metabolites in blood is high why they did not influence the serum metabolome three hours after breakfast. As all urine was collected the beverage consumption was captured.
These results are of interest as both tea and coffee are commonly consumed beverages. However, when studying the food metabolome in intervention studies,

the study design is advised to reduce the variation introduced by beverage choice.

4.3 Paper III

In the second intervention study we investigated the metabolic response in serum in relation to three different meals corresponding to vegan, lacto-ovo vegetarian and omnivore diets. Metabolic profiles were characterized for individual meals and compared between meals.

Using O2PLS-DA, metabolic profiles between meals were identified (Figure 14). Separation in relation to breakfast meals was seen in both postprandial (3h) samples and samples based on the effect matrix i.e. modeling the change between fasting and postprandial samples.
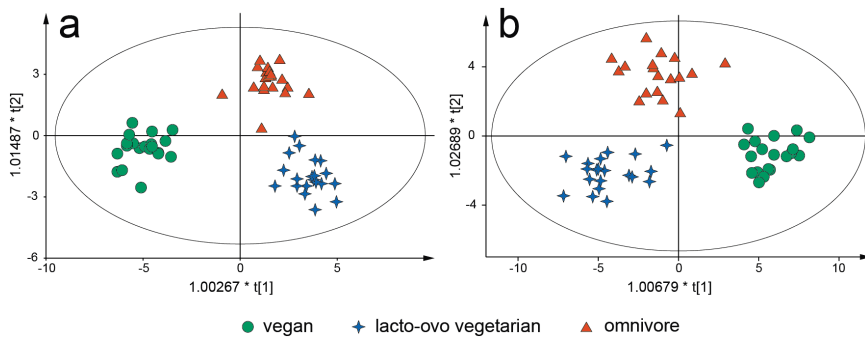

Figure 14. *Orthogonal (2) projections to latent structures with discriminant analysis (O2PLS-DA) models of metabolic profiles from vegan, lacto-ovo vegetarian, and omnivore breakfasts. The distribution of metabolic profiles of a) postprandial (3h) samples (n=60) and b)a calculated effect matrix on the difference between the fasting and the postprandial state (n=57). Both models included 196 variables.*

In addition, OPLS-EP models were used to investigate the individual responses for each breakfast meal (Figure 15). Overall, all individuals showed a positive response in relation to the response vector (predicted effect) (Y>0) between fasting and postprandial state for all breakfast meals. The VE meal showed a fairly homogeneous response in all individuals (Y>0.75) while for the LOV and OM meals the inter-individual response displayed larger fluctuations in individual responses. The predicted effect was <0.6 for three individuals in response to the LOV and <0.6 for five individuals in response to the OM where one individual showed a predicted effect of <0.05. In addition nine and ten individuals showed predicted values of >1 for the OM and LOV breakfast, respectively. In specula-

tion, the difference in the predicted effect between meals might relate to a corresponding inter-individual difference in metabolism of included foods.
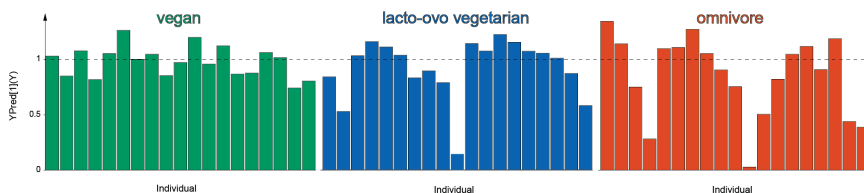

Figure 15. *Predicted values in relation to the response vector (Y) for healthy volunteers consuming vegan, lacto-ovo vegetarian and omnivore breakfasts in orthogonal projections to latent structures with effect matrix (OPLS-EP) models. The dotted line (Y = 1) indicates the response vector value for the models. The magnitude of the predicted effect for each volunteer is given by the height of the corresponding bar. Deviations from the value 1 for a specific volunteer indicate a larger (>1) or smaller (<1) metabolic effect (difference between fasting and postprandial state) in the model direction (metabolic profile) associated with the metabolism of foods included in the different breakfast meals. Each model included 19 observations (equal to number of individuals in the models) and 196 variables. The individual orders of predicted values are not comparable between models.*

In *paper III* the emphasis was concentrated to discriminating metabolites between the different breakfast meals. Here, the metabolic response to individual meals between fasting and postprandial state was also assessed. Identified metabolites found in increased concentration between fasting and postprandial samples in response to the different breakfast meals are listed in Table 5. In short, the VE showed the smallest number of metabolites increasing in concentration while OM showed the largest. This might be connected to the metabolic overturn of foods included in the different meals. However, as we only sampled 3h postprandial we lack the data to asses this further.

Alanine, glutamate, myo-inositol, ornithine and proline were found in increased levels after all breakfast meals. In combination, 3-hydroxyisobutyrate, isoleucine, leucine, lysine, methionine, N-acetylcysteine, threonine, tyrosine and valine were found in increased levels after LOV and OM meals, while asparagine, betaine and carnitine in a variable overlapping with cholines were found in increased levels after the OM and VE breakfasts. Arginine and methanol were only found in increased levels after the VE and glutamine only after the LOV. In turn, ascorbate, choline, creatine and creatinine in an overlapping variable and serine distinguished the OM.

In a cohort study, investigating the relationship between intake and plasma concentration levels of amino acids between different diets, alanine was found in

higher concentrations in vegans compared to meat eaters[207]. In addition, in the same study lysine, methionine, isoleucine, leucine, tyrosine and valine were higher in meat eaters and vegetarians compared to vegans. In a cross-over inter-vention study investigating the metabolic impact in plasma of a 48h vegan vs animal diet, branched chain amino acids (leucine, isoleucine and valine), proline,

Table 5. Metabolites found in increased levels between fasting and 3h postprandial state for vegan, lacto-ovo vegetarian and omnivore breakfasts.

Meal	Metabolite	Chemical shift (ppm)	p-value
VE[1]	Alanine	1.47	0.0004
VE	Arginine	1.64	0.0009
VE	Asparagine	2.95	0.02
VE	Betaine	3.26	0.0003
VE	Carnitine & Cholines	3.22	0.01
VE	Glutamate	2.34	0.02
VE	Methanol	3.35	0.04
VE	Myo-Inositol	3.62	0.0005
VE	Ornithine	3.05	<0.0001
VE	Proline	4.13	<0.0001
LOV[2]	3-Hydroxyisobutyrate	1.06	0.0001
LOV	Alanine	1.47	0.0009
LOV	Glutamate	2.34	<0.0001
LOV	Glutamine	2.44	0.03
LOV	Isoleucine	1.00	0.0006
LOV	Leucine	0.95	0.003
LOV	Lysine	3.02	0.0009
LOV	Methionine	2.63	0.0002
LOV	Myo-Inositol	3.62	0.05
LOV	N-Acetylcysteine	2.93	<0.0001
LOV	Ornithine	3.05	<0.0001
LOV	Proline	4.13	<0.0001
LOV	Threonine	4.24	0.004
LOV	Tyrosine	6.88	<0.0001
LOV	Valine	0.98	<0.0001
OM[3]	3-Hydroxyisobutyrate	1.06	0.003
OM	Alanine	1.47	0.03
OM	Ascorbate	4.51	0.003

OM	Asparagine	2.95	0.04
OM	Betaine	3.26	<0.0001
OM	Carnitine & Cholines	3.22	0.0004
OM	Choline	3.19	<0.0001
OM	Creatine & Creatinine	3.03	0.03
OM	Glutamate	2.34	0.002
OM	Isoleucine	1.00	<0.0001
OM	Leucine	0.95	0.003
OM	Lysine	3.02	0.0002
OM	Methionine	2.63	0.0002
OM	Myo-Inositol	3.62	0.0006
OM	N-Acetylcysteine	2.93	0.001
OM	Ornithine	3.05	<0.00001
OM	Proline	4.13	0.0004
OM	Serine	3.95	0.03
OM	Threonine	4.24	0.04
OM	Tyrosine	6.88	0.0003
OM	Valine	0.98	<0.0001

[1]Vegan
[2]Lacto-ovo vegetarian
[3]Omnivore

tyrosine and valine were higher in concentration after the animal diet which corresponded to the dietary intake of these amino acids[81]. In turn, the same study showed an increase in arginine and glycine after the vegan diet. These findings largely correspond to our results. However, here glycine levels were not increased after the VE and alanine increased for all breakfast meals although comparably more so for the LOV.

Glutamate and ornithine are both metabolites involved in the urea cycle metabolism [208]. Both metabolites increased after consumption of all breakfast meals indicating that this is a reflection of urea cycle metabolism activity independent of food content. This is also supported by Reeds *et al*. (1966), who showed that dietary glutamate is metabolized in the gastrointestinal tract and does not enter the blood stream[209].

Myo-inositol is the most abundant of nine existing isomers of the polyol inositol. Myo-inositol is found free, as phytic acid or bound to phospholipids in both animal and plant cells [210]. As such it is found in a number of foods and especially

high concentrations are present in fresh fruits and vegetables and different seeds (nuts, grains, beans) [211]. In the present study levels of myo-inositol were increased after consumption of all breakfast, however only discriminating for the VE and OM breakfasts in the OPLS-EP models. The increased level of myo-inositol in the VE and OM breakfasts are believed to originate from the soy yoghurt, red and green bell pepper, cashew butter and red bell pepper, respectively. Dairy products are naturally low in myo-inositol [211] which might explain why the metabolite did not discriminate for the LOV that largely consisted of dairy products.

3-hydroxyisobutyrate, a ketone body and by-product of valine catabolism, increased in concentration after both the LOV and OM although comparably more so for the LOV. This implies a higher increase in valine degradation for the LOV 3h postprandial.

Choline, betaine and methionine are involved in betaine metabolism. Betaine is obtained from the diet either as betaine or choline, which is oxidized to betaine in the human body. Red meat, egg, liver, and soy beans are foods rich in choline while high concentrations of betaine are found in wheat products, spinach, beats and shrimps [212]. Here we found increased concentrations of both choline and betaine after consumption of the OM breakfast and increased concentration of betaine after the VE breakfast. As the breakfasts contain food products high in choline, i.e. liver, egg, ham and soy, but no food items high in betaine, potentially the increased concentrations of betaine is a result of choline oxidation. Methionine is generated from homocysteine via a methyl donation from betaine but can also be obtained from the diet. Almonds, sesame seeds, hard cheese, egg, poultry, red meat, sea food and oat are foods high in methionine. Methionine showed an increase in concentration after the OM but a decrease after the VE why the methionine concentration in the OM is believed to be derived from the metabolism of the included foods and not as a result of betaine metabolism.

Limitations of the meal intervention included, difference in macronutrient composition between meals and its potential influence on the results, time point of sample collection, and the absence of continuous postprandial sampling. Also, findings here were identified in a fairly homogenous group of healthy individuals. And thus, present results need to be verified in a larger sample set of a more various group reflecting an overall population. In addition, to these limitations multivariate models from the first meal study were more robust with comparably lower p-values according to CV-ANOVA when modelling 3h samples. This implies that a larger study population and or repeated sampling of the same meal

would have improved the multivariate models. However, determining the size of the study population in metabolomics studies is not a trivial task and with limited resources a pilot study to determine sample size was not an option.

Overlapping metabolites in serum between meal studies
Tyrosine and proline were found discriminating for both the CB and LOV. Also valine was higher after the LOV compared to the VE and higher in the CB breakfast. Both these meals had high dairy content. In turn, creatine, isoleucine, choline and lysine were discriminating for both the EHB and OM breakfast in serum and both breakfasts contained comparably high content of animal protein. This implies that the metabolic response to high dairy meals can be reflected in concentrations of tyrosine, proline and valine, while creatine, isoleucine, choline and lysine reflects high intake of meat irrespective of the total food matrix in a meal.

4.4 Paper IV

In this work, the aim was to evaluate the performance of three different pre-processing protocols: dilution, precipitation and ultrafiltration for high-throughput NMR serum metabolomics. Reproducibility within sample groups, spectral information content, internal standard compatibility, sample stability and high-throughput potential were assessed for each sample group (Table 6).

Table 6. Overall assessment of pre-analytical methods

Preprocessing	Inform-ation	Reprod-ucibility	High-throughput	Internal standard compatibility	Sample stability
Dilution	+	++	++	-	-
Percipitation	++	+	++	++	++
Ultrafiltration	++	+	-	++	++

Assessing reproducibility, clustering of samples in PCA model along with visual inspection of overlaid spectra were used. To further evaluate reproducibility and differences between sample groups regarding metabolite concentrations, additional experiments are performed with spike in standards of different compounds and comparison to a reference sample. The dilution protocol, with the simplest workflow displayed least variation between samples. However, the reproducibility of the precipitated and ultrafiltrated samples was within an acceptable range. The presence of proteins and lipoproteins in serum influence signal-intensities of metabolites interacting with these macromolecules to a various extent. ^{1}H NMR analysis of both ultrafiltrated and precipitated serum samples had a similar and comparably higher information content regarding identified metabolites than did the corresponding analysis of diluted serum samples. As an example, alanine, lactate and 3-hydroxybutyrate was comparably attenuated to various extent in the precipitated sample (Figure 16). In addition, several peaks, including peaks belonging to protons of tryptophan and phenylalanine, in the ppm range 7.9 to 7.4 were present in the precipitated sample but not in the diluted sample. Volatile compounds including ethanol, methanol, trimethylamine, formic acid, acetic acid and acetone are attenuated ore even lost completely during precipitation with methanol and lyophilization. However, still the information content exceeds that of the dilution protocol.

In addition, internal standards such as 3-trimethylsilylpropionic acid (TSP) or 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), used for quantification, exhib-

it strong interactions with proteins, broadening their signals and negating their use as standards in the dilution protocol. The presence of proteins in serum decreases the stability of the sample due to possible enzymatic activity and interaction events between proteins and metabolites, effectively introducing unwanted between-sample variation. When working with large sample series where the time between pre-processing of samples and the subsequent analysis may vary within a timeframe of 24h, the stability of the samples are of particular importance. In conclusion from the issues raised above, for a comprehensive and quantitative analysis of serum metabolites, the protein fraction of the sample needs to be removed.
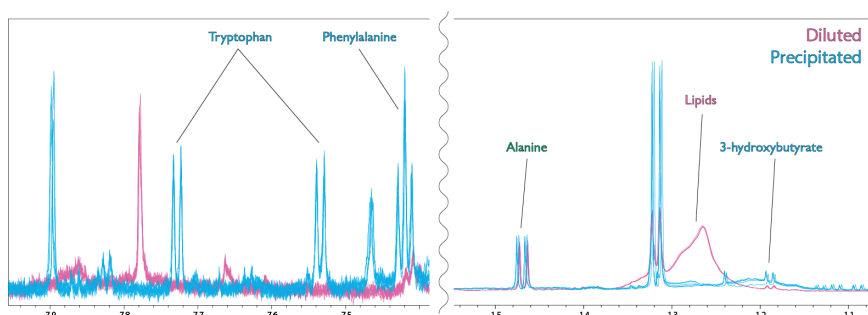

Figure 16. Excerpt from NMR spectra from serum samples that were subjected to the dilution (DIL) protocol and the precipitation (PREC) protocol. A T2-relaxation filter was applied during the NMR data acquisition. Following each of the two protocols, six samples were analyzed and six spectra superimposed. Both protocols give excellent reproducibility. The difference in observable concentration for select metabolites are indicated. Both tryptophan and phenyl alanine are completely suppressed after the DIL protocol, whereas both alanine and lactate are suppressed to various degrees. Despite the T2 relaxation filter, the signal from free and aggregated lipids is clearly seen in the spectra after DIL.

A high degree of automatization implies minimal manual work input. For preparation of large number of samples the use of 96 well plates are preferable for simultaneous processing. For ultrafiltration, the use of microtiter plates was not feasible due to uneven distribution of filtrate when centrifuging the plate driven by the geometric constraints of plate and rotor design. Similarly, vacuum filtration was not practically feasible as the filtrate buildup even at room temperature was exceedingly slow. This left the choice of single tubes. For large sample series, the washing of the filters in these tubes, in addition to the actual serum filtration, is an immense task since each filter had to be washed seven times to remove glycerol. In contrast, the use of deep well plates could be applied for the precipitation protocol enabling automated deproteinization.

So far, according to the factors evaluated here, the precipitation protocol is the preferred choice over dilution and ultrafiltration for high-throughput NMR serum metabolomics of large sample series.

5. Conclusions

In two cross-over intervention studies in healthy human volunteers ^{1}H NMR metabolomics could differentiate acute metabolic profiles between different meals in serum and urine and identify metabolites positively associated with the different meals.

Acute metabolic profiles were reproducibly characterized in serum and urine after the consumption of two equicaloric breakfast meals. NMR metabolomics identified metabolites discriminating between the two breakfast meals. Coffee or tea consumption was clearly distinguished in urine, while the difference in serum metabolites largely mirrored differences in dietary composition or kinetic differences in e.g. absorption rate between foods. Thus, serum and urine can be used complementarily to evaluate food and beverage consumption.

NMR metabolomics could distinguish meals corresponding to vegan, lacto-ovo-vegetarian and omnivore diets, by metabolic profiles along with food derived metabolites.

Tyrosine and proline were found discriminating for both the CB and LOV. Also valine was higher after the LOV compared to the VE and higher in the CB breakfast. Both these meals had high dairy content. In turn, creatine, isoleucine, choline and lysine was discriminating for both the EHB and OM breakfast in serum and containing comparable high content of animal protein. This implies that the metabolic response to high dairy meals can be reflected in concentrations of tyrosine, proline and valine while creatine, isoleucine, choline and lysine reflects high intake of meat irrespectable of the total food matrix in a meal.

Urine collected at one time point can reflect the metabolism over several hours while for blood, if not repeated sampling is conducted, only a snapshot will be captured. This gives urine an advantage if continuous blood sampling cannot be conducted in postprandial studies.

Assessing pre-analytical deproteinization methods for high-throughput NMR-based serum metabolomics for large sample series ultrafiltration in tubes and methanol precipitation were found to fulfill the criteria on information content and reproducibility. In addition, methanol precipitation was compatible in with an atomized workflow. However, it is highly questionable whether ultrafiltration in single tubes is compatible with large sample series and automation.

6. Future perspectives

Globally, diet is the second largest risk factor attributed to the global burden of disease and in middle-income and high-income countries it is the largest risk factor[213]. In light of this, applicable and reliable methods to evaluate dietary intake in connection to health outcomes are of outmost importance. Metabolomics offers the opportunity to identify metabolic phenotypes and dietary intake biomarkers. Metabolomics in nutrition is a relatively new field but holds great promise as a complement to traditional dietary assessment methods like 24 h recalls, food diaries and FFQs. However, to increase the usability of metabolomics in nutrition the field needs to address several challenges such as biomarker validation, standardisation of study designs and the overall workflow.

The main focus in nutritional metabolomics, so far, has been to identify biomarkers of individual foods and metabolites that change in concentration in relation to different diets. As this work progresses, an overall picture of the food metabolome and the metabolic phenotype modulated by food intake is generated. To further advance the field and strengthen associations between diet and health outcomes, clusters of metabolites connected to metabolic phenotypes of different diets should be associated with metabotypes of dietary risk factors. This could be partly accomplished by investigating the long term modulation of the metabolome related to different diets in controlled intervention studies and connecting findings to the metabolic phenotype of various health outcomes. As a complement, population based observational studies could be used to link diet-related disease risk to metabolic phenotypes connected to habitual dietary intake. As a prevention measure, metabolomics in combination with other omics technologies, along with pathway and network analysis, could be useful to identify individuals at risk of developing diet related diseases.

Acknowledgement

Det finns många personer som jag vill tacka och som varit en viktig del av min doktorandtid och bidragit till den här avhandlingen. Speciellt tack vill jag rikta till:

Alla studiedeltagare, STORT tack för ert engagemang, utan er hade denna avhandling inte varit möjlig.

Anna Winkvist, min huvudhandledare, tack för stöd, feedback och engagemang under upp och nergångar. Du har varit en fantastisk mentor! Har varit härligt att jag kunnat vara mig själv med dig och utöver ditt stöd i forskningen har jag verkligen uppskattat våra filosofiska samtal.

Helen Lindqvist, min bihandledare, tack för kreativa och roliga möten där vi skrivit manus och analyserat data. Du har verkligen varit en stor hjälp i att föra arbetet framåt och hitta nya lösningar!

Lars Ellegård, min bihandledare, tack för att du delat med dig av din kunskap och erfarenhet. Ett särskilt tack för presentkortet till Saluhallen, det blev många goda chokladpraliner!

Göran Karlsson, min bihandledare, tack för att du väglett mig och delat med dig av din kunskap inom många områden. I tillägg, så vill jag tacka för att du tog emot mig på SNC och alltid fått mig att känna mig välkommen.

Anders Pedersen, min klippa, utan din hjälp, ditt tålamod och genorositet med din tid och kunskap hade inte detta projekt varit möjligt. Orden räcker inte till för att tacka dig.

Vibeke Malmros, du är en av de mest generösa personer som jag träffat. Du fick mig att känna mig välkommen från dag ett och har varit en klippa under arbetet med interventionerna. Jag har alltid kunna komma till dig och vädra stort som smått. Jag kommer aldrig att glömma chokladen på sängen första gången jag sov hos dig, det fick mig att känna mig väldigt välkommen.

Daniel Malmodin och **Rui Pinto**, tack för värdefull statistiskt hjälp och för att ni haft tålamod med alla mina dumma frågor.

References

1.   Dettmer, K., Aronov, P.A. & Hammock, B.D. Mass spectrometry-based metabolomics. *Mass spectrometry reviews* **26**, 51-78 (2007).
2.   Wishart, D.S. Metabolomics: applications to food science and nutrition research. *Trends in Food Science & Technology* **19**, 482-493 (2008).
3.   Trifonova, O., Lokhov, P. & Archakov, A. Postgenomics diagnostics: metabolomics approaches to human blood profiling. *Omics : a journal of integrative biology* **17**, 550-559 (2013).
4.   Astarita, G. & Langridge, J. An emerging role for metabolomics in nutrition science. *Journal of nutrigenetics and nutrigenomics* **6**, 181-200 (2013).
5.   Zivkovic, A.M. & German, J.B. Metabolomics for assessment of nutritional status. *Current opinion in clinical nutrition and metabolic care* **12**, 501-507 (2009).
6.   Holmes, E., Wilson, I.D. & Nicholson, J.K. Metabolic phenotyping in health and disease. *Cell* **134**, 714-717 (2008).
7.   Bingham, S.A. Biomarkers in nutritional epidemiology. *Public health nutrition* **5**, 821-827 (2002).
8.   Coulston, A.M., Boushey, C.J. & Ferruzzi, G. *Nutrition in the Prevention and Treatment of Disease*, (Academic Press, London, 2013).
9.   Jenab, M., Slimani, N., Bictash, M., Ferrari, P. & Bingham, S.A. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Human genetics* **125**, 507-525 (2009).
10.   Kaaks, R., Ferrari, P., Ciampi, A., Plummer, M. & Riboli, E. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public health nutrition* **5**, 969-976 (2002).
11.   Kaaks, R., Riboli, E. & Sinha, R. Biochemical markers of dietary intake. *IARC scientific publications*, 103-126 (1997).
12.   Tasevska, N., Runswick, S.A., McTaggart, A. & Bingham, S.A. Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **14**, 1287-1294 (2005).
13.   Bingham, S.*, et al.* Associations between dietary methods and biomarkers, and between fruits and vegetables and risk of ischaemic heart disease, in the EPIC Norfolk Cohort Study. *International journal of epidemiology* **37**, 978-987 (2008).
14.   Primrose, S.*, et al.* Metabolomics and human nutrition. *The British journal of nutrition* **105**, 1277-1283 (2011).
15.   Lindon, J.C., Nicholson, J.K. & Holmes, E. *The Handbook of Metabonomics and Metabolomics*, (2007).
16.   Gibbons, H., O'Gorman, A. & Brennan, L. Metabolomics as a tool in nutritional research. *Current opinion in lipidology* **26**, 30-34 (2015).

17.   Wishart, D.S. Computational approaches to metabolomics. *Methods in molecular biology* **593**, 283-313 (2010).

18.   Fave, G., Beckmann, M.E., Draper, J.H. & Mathers, J.C. Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes & nutrition* **4**, 135-141 (2009).

19.   van der Greef, J., van Wietmarschen, H., van Ommen, B. & Verheij, E. Looking back into the future: 30 years of metabolomics at TNO. *Mass spectrometry reviews* **32**, 399-415 (2013).

20.   Dalgliesh, C.E., Horning, E.C., Horning, M.G., Knox, K.L. & Yarger, K. A gas-liquid-chromatographic procedure for separating a wide range of metabolites occuring in urine or tissue extracts. *The Biochemical journal* **101**, 792-810 (1966).

21.   Horning, E.C. & Horning, M.G. Metabolic profiles: gas-phase methods for analysis of metabolites. *Clinical chemistry* **17**, 802-809 (1971).

22.   Pauling, L., Robinson, A.B., Teranishi, R. & Cary, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 2374-2376 (1971).

23.   Oliver, S.G., Winson, M.K., Kell, D.B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends in biotechnology* **16**, 373-378 (1998).

24.   Psychogios, N.*, et al.* The human serum metabolome. *PloS one* **6**, e16957 (2011).

25.   Trabado, S.*, et al.* The human plasma-metabolome: Reference values in 800 French healthy volunteers; impact of cholesterol, gender and age. *PloS one* **12**, e0173615 (2017).

26.   Bouatra, S.*, et al.* The Human Urine Metabolome. *PloS one* **8**, e73076 (2013).

27.   Wishart, D.S.*, et al.* The human cerebrospinal fluid metabolome. *Journal of Chromatography B* **871**, 164-173 (2008).

28.   Dame, Z.*, et al.* The human saliva metabolome. *Metabolomics : Official journal of the Metabolomic Society*, 1-20 (2015).

29.   Dallmann, R., Viola, A.U., Tarokh, L., Cajochen, C. & Brown, S.A. The human circadian metabolome. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2625-2629 (2012).

30.   Gibney, M.J.*, et al.* Metabolomics in human nutrition: opportunities and challenges. *The American journal of clinical nutrition* **82**, 497-503 (2005).

31.   German, J.B., Roberts, M.A., Fay, L. & Watkins, S.M. Metabolomics and individual metabolic assessment: the next great challenge for nutrition. *The Journal of nutrition* **132**, 2486-2487 (2002).

32.   Scalbert, A.*, et al.* The food metabolome: a window over dietary exposure. *The American journal of clinical nutrition* **99**, 1286-1308 (2014).

33.   Eriksson, L*., et al.* Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Analytical and bioanalytical chemistry* **380**, 419-429 (2004).

34.   Trygg, J., Holmes, E. & Lundstedt, T. Chemometrics in metabonomics. *Journal of proteome research* **6**, 469-479 (2007).

35.   Van Der Greef, J. & Smilde, A.K. Symbiosis of chemometrics and metabolomics: Past, present, and future. *Journal of Chemometrics* **19**, 376-386 (2005).

36.   Nicholson, J.K. & Wilson, I.D. High resolution proton magnetic resonance spectroscopy of biological fluids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **21**, 449-501 (1989).

37.   Gardinassi, L.G., Xia, J., Safo, S.E. & Li, S. Bioinformatics Tools for the Interpretation of Metabolomics Data. *Current Pharmacology Reports*, 1-10 (2017).

38.   Lodge, J.K. Symposium 2: Modern approaches to nutritional research challenges: Targeted and non-targeted approaches for metabolite profiling in nutritional research. *The Proceedings of the Nutrition Society* **69**, 95-102 (2010).

39.   Oresic, M. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutrition, metabolism, and cardiovascular diseases : NMCD* **19**, 816-824 (2009).

40.   Courant, F., Antignac, J.P., Dervilly-Pinel, G. & Bizec, B.L. Basics of Mass Spectrometry-Based Metabolomics. *Proteomics* (2014).

41.   Vernocchi, P*., et al.* Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome. *Frontiers in cellular and infection microbiology* **2**, 156 (2012).

42.   Broadhurst, D*., et al.* Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics : Official journal of the Metabolomic Society* **14**, 72 (2018).

43.   Tautenhahn, R*., et al.* An accelerated workflow for untargeted metabolomics using the METLIN database. *Nature biotechnology* **30**, 826-828 (2012).

44.   Bingol, K. Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods. *High-throughput* **7**(2018).

45.   Keun, H.C. *NMR-based Metabolomics*, (Royal Society of Chemistry, 2018).

46.   Heinzmann, S.S*., et al.* Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *The American journal of clinical nutrition* **92**, 436-443 (2010).

47.   Lloyd, A.J., Beckmann, M., Fave, G., Mathers, J.C. & Draper, J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *The British journal of nutrition* **106**, 812-824 (2011).

48.   Pujos-Guillot, E., *et al.* Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. *Journal of proteome research* **12**, 1645-1659 (2013).

49.   Lang, R., Wahl, A., Stark, T. & Hofmann, T. Urinary N-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. *Molecular nutrition & food research* **55**, 1613-1623 (2011).

50.   Lloyd, A.J., *et al.* Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *The American journal of clinical nutrition* **97**, 377-389 (2013).

51.   Rothwell, J.A., *et al.* New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. *PloS one* **9**, e93474 (2014).

52.   Daykin, C.A., *et al.* Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. *Journal of agricultural and food chemistry* **53**, 1428-1434 (2005).

53.   van Velzen, E.J., *et al.* Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. *Journal of proteome research* **8**, 3317-3330 (2009).

54.   Andersen, M.B., *et al.* Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. *Analytical and bioanalytical chemistry* **406**, 1829-1844 (2014).

55.   Wang, Y., *et al.* Untargeted Metabolomics Identifies Novel Potential Biomarkers of Habitual Food Intake in a Cross-Sectional Study of Postmenopausal Women. *The Journal of nutrition* (2018).

56.   van Dorsten, F.A., *et al.* The metabolic fate of red wine and grape juice polyphenols in humans assessed by metabolomics. *Molecular nutrition & food research* **54**, 897-908 (2010).

57.   Jacobs, D.M., *et al.* Impact of short-term intake of red wine and grape polyphenol extract on the human metabolome. *Journal of agricultural and food chemistry* **60**, 3078-3085 (2012).

58.   Bondia-Pons, I., *et al.* UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. *Molecular nutrition & food research* **57**, 412-422 (2013).

59.   Bondia-Pons, I., *et al.* Postprandial differences in the plasma metabolome of healthy Finnish subjects after intake of a sourdough fermented endosperm rye bread versus white wheat bread. *Nutrition journal* **10**, 116 (2011).

60.   Shi, L., *et al.* Targeted metabolomics reveals differences in the extended postprandial plasma metabolome of healthy subjects after intake of whole-grain rye porridges versus refined wheat bread. *Molecular nutrition & food research* (2016).

61.   Stanstrup, J., *et al.* Intakes of whey protein hydrolysate and whole whey proteins are discriminated by LC–MS metabolomics. *Metabolomics : Official journal of the Metabolomic Society* **10**, 719-736 (2013).

62.   Mora-Cubillos, X.*, et al.* Plasma metabolomic biomarkers of mixed nuts exposure inversely correlate with severity of metabolic syndrome. *Molecular nutrition & food research* **59**, 2480-2490 (2015).

63.   Lloyd, A.J.*, et al.* Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *The American journal of clinical nutrition* **94**, 981-991 (2011).

64.   Langer, S., Kennel, A. & Lodge, J.K. The influence of juicing on the appearance of blueberry metabolites 2 h after consumption: a metabolite profiling approach. *The British journal of nutrition* **119**, 1233-1244 (2018).

65.   Hjerpsted, J., Ritz, C., Schou, S., Tholstrup, T. & Dragsted, L. Effect of cheese and butter intake on metabolites in urine using an untargeted metabolomics approach. *Metabolomics : Official journal of the Metabolomic Society* **10**, 1176-1185 (2014).

66.   Pallister, T.*, et al.* Metabolites of milk intake: a metabolomic approach in UK twins with findings replicated in two European cohorts. *European journal of nutrition* (2016).

67.   Pimentel, G.*, et al.* Metabolic footprinting of fermented milk consumption in serum of healthy men. *Journal of Nutrition* **148**, 851-860 (2018).

68.   Cuparencu, C.S.*, et al.* Identification of urinary biomarkers after consumption of sea buckthorn and strawberry, by untargeted LC–MS metabolomics: a meal study in adult men. *Metabolomics : Official journal of the Metabolomic Society* **12**, 1-20 (2016).

69.   Madrid-Gambin, F.*, et al.* Untargeted (1)H-NMR based metabolomics analysis of urine and serum profiles after consumption of lentils, chickpeas and beans: an extended meal study to discover dietary biomarkers of pulses. *Journal of agricultural and food chemistry* (2018).

70.   Lehtonen, H.M.*, et al.* 1H NMR-based metabolic fingerprinting of urine metabolites after consumption of lingonberries (Vaccinium vitis-idaea) with a high-fat meal. *Food Chem* **138**, 982-990 (2013).

71.   Ross, A.B., Svelander, C., Undeland, I., Pinto, R. & Sandberg, A.S. Herring and Beef Meals Lead to Differences in Plasma 2-Aminoadipic Acid, beta-Alanine, 4-Hydroxyproline, Cetoleic Acid, and Docosahexaenoic Acid Concentrations in Overweight Men. *The Journal of nutrition* **145**, 2456-2463 (2015).

72.   Cheung, W.*, et al.* A metabolomic study of biomarkers of meat and fish intake. *The American journal of clinical nutrition* (2017).

73.   O'Sullivan, A., Gibney, M.J. & Brennan, L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *The American journal of clinical nutrition* **93**, 314-321 (2011).

74.   Carrizo, D.*, et al.* Untargeted metabolomic analysis of human serum samples associated with different levels of red meat consumption: A possible indicator of type 2 diabetes? *Food Chem* **221**, 214-221 (2017).

75.   Khodorova, N.V.*, et al.* Urinary Metabolomics Profiles Associated to Bovine Meat Ingestion in Humans. *Molecular nutrition & food research* (2018).

76.   Schmedes, M.*, et al.* Lean-seafood intake decreases urinary markers of mitochondrial lipid and energy metabolism in healthy subjects: metabolomics results from a randomized crossover intervention study. *Molecular nutrition & food research* (2016).

77.   Hanhineva, K.*, et al.* Nontargeted metabolite profiling discriminates diet-specific biomarkers for consumption of whole grains, Fatty fish, and bilberries in a randomized controlled trial. *The Journal of nutrition* **145**, 7-17 (2015).

78.   Sumner, L.W.*, et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics : Official journal of the Metabolomic Society* **3**, 211-221 (2007).

79.   Stella, C.*, et al.* Susceptibility of human metabolic phenotypes to dietary modulation. *Journal of proteome research* **5**, 2780-2788 (2006).

80.   Svensson, B.G., Akesson, B., Nilsson, A. & Paulsson, K. Urinary excretion of methylamines in men with varying intake of fish from the Baltic Sea. *Journal of toxicology and environmental health* **41**, 411-420 (1994).

81.   Draper, C.F.*, et al.* A 48-Hour Vegan Diet Challenge in Healthy Women and Men Induces a BRANCH-Chain Amino Acid Related, Health Associated, Metabolic Signature. *Molecular nutrition & food research* **62**(2018).

82.   Vazquez-Fresno, R.*, et al.* Metabolomic Pattern Analysis after Mediterranean Diet Intervention in a Nondiabetic Population: A 1- and 3-Year Follow-up in the PREDIMED Study. *Journal of proteome research* (2014).

83.   Andersen, M.B.*, et al.* Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *Journal of proteome research* **13**, 1405-1418 (2014).

84.   Khakimov, B.*, et al.* New Nordic diet versus average Danish diet: a randomized controlled trial revealed healthy long-term effects of the new Nordic diet by GC-MS blood plasma metabolomics. *Journal of proteome research* (2016).

85.   Bouchard-Mercier, A., Rudkowska, I., Lemieux, S., Couture, P. & Vohl, M.C. The metabolic signature associated with the Western dietary pattern: a cross-sectional study. *Nutrition journal* **12**, 158 (2013).

86.   Schmidt, J.A.*, et al.* Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. *The American journal of clinical nutrition* **102**, 1518-1526 (2015).

87.   Esko, T.*, et al.* Metabolomic profiles as reliable biomarkers of dietary composition. *The American journal of clinical nutrition* (2017).

88.   Johansson-Persson, A., Barri, T., Ulmius, M., Onning, G. & Dragsted, L.O. LC-QTOF/MS metabolomic profiles in human plasma after a 5-week high dietary fiber intake. *Analytical and bioanalytical chemistry* **405**, 4799-4809 (2013).

89.   Zheng, H., *et al.* Metabolic Effects of a 24-Week Energy-Restricted Intervention Combined with Low or High Dairy Intake in Overweight Women: An NMR-Based Metabolomics Investigation. *Nutrients* **8**(2016).

90.   May, D.H., *et al.* Metabolomic profiling of urine: response to a randomised, controlled feeding study of select fruits and vegetables, and application to an observational study. *The British journal of nutrition* **110**, 1760-1770 (2013).

91.   Martin, F.P., *et al.* Metabolic effects of dark chocolate consumption on energy, gut microbiota, and stress-related metabolism in free-living subjects. *Journal of proteome research* **8**, 5568-5579 (2009).

92.   Sonnenburg, J.L. & Backhed, F. Diet-microbiota interactions as moderators of human metabolism. *Nature* **535**, 56-64 (2016).

93.   Cassidy, A. & Minihane, A.M. The role of metabolism (and the microbiome) in defining the clinical efficacy of dietary flavonoids. *The American journal of clinical nutrition* **105**, 10-22 (2017).

94.   Rechner, A.R., *et al.* The metabolic fate of dietary polyphenols in humans. *Free radical biology & medicine* **33**, 220-235 (2002).

95.   Walsh, M.C., *et al.* Influence of acute phytochemical intake on human urinary metabolomic profiles. *The American journal of clinical nutrition* **86**, 1687-1693 (2007).

96.   Penczynski, K.J., *et al.* Relative validation of 24-h urinary hippuric acid excretion as a biomarker for dietary flavonoid intake from fruit and vegetables in healthy adolescents. *European journal of nutrition* **56**, 757-766 (2017).

97.   Pallister, T., *et al.* Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. *Scientific reports* **7**, 13670 (2017).

98.   Radjursoga, M., *et al.* Metabolic profiles from two different breakfast meals characterized by 1H NMR-based metabolomics. *Food Chem* **231**, 267-274 (2017).

99.   Savolainen, O., *et al.* Biomarkers of food intake and nutrient status are associated with glucose tolerance status and development of type 2 diabetes in older Swedish women. *The American journal of clinical nutrition* (2017).

100.   Brouwer-Brolsma, E.M., *et al.* Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance. *The Proceedings of the Nutrition Society* **76**, 619-627 (2017).

101.   Livera, A.M.D., *et al.* Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Analytical chemistry* **87**, 3606-3615 (2015).

102. De Livera, A.M., Olshansky, M. & Speed, T.P. Statistical analysis of metabolomics data. *Methods in molecular biology* **1055**, 291-307 (2013).

103. Jansen, J.J., Szymanska, E., Hoefsloot, H.C. & Smilde, A.K. Individual differences in metabolomics: individualised responses and between-metabolite relationships. *Metabolomics : Official journal of the Metabolomic Society* **8**, 94-104 (2012).

104. de Roos, B. & Brennan, L. Personalised interventions—A precision approach for the next generation of dietary intervention studies. *Nutrients* **9**(2017).

105. Bervoets, L.*, et al.* Influence of preanalytical sampling conditions on the 1H NMR metabolic profile of human blood plasma and introduction of the Standard PREanalytical Code used in biobanking. *Metabolomics : Official journal of the Metabolomic Society* **11**, 1197-1207 (2015).

106. Teahan, O.*, et al.* Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Analytical chemistry* **78**, 4307-4318 (2006).

107. Broadhurst, D.I. & Kell, D.B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics : Official journal of the Metabolomic Society* **2**, 171-196 (2006).

108. Surowiec, I.*, et al.* Multivariate strategy for the sample selection and integration of multi-batch data in metabolomics. *Metabolomics : Official journal of the Metabolomic Society* **13**, 114 (2017).

109. Ellervik, C. & Vaught, J. Preanalytical variables affecting the integrity of human biospecimens in biobanking. *Clinical chemistry* **61**, 914-934 (2015).

110. Wilson, T.*, et al.* Developing community-based urine sampling methods to facilitate dietary exposure biomarker technology for population assessment. *Proceedings of the Nutrition Society* **75**(2016).

111. Jorgenrud, B.*, et al.* The influence of sample collection methodology and sample preprocessing on the blood metabolic profile. *Bioanalysis* **7**, 991-1006 (2015).

112. Lopez-Bascon, M.A., Priego-Capote, F., Peralbo-Molina, A., Calderon-Santiago, M. & Luque de Castro, M.D. Influence of the collection tube on metabolomic changes in serum and plasma. *Talanta* **150**, 681-689 (2016).

113. Gonzalez-Covarrubias, V., Dane, A., Hankemeier, T. & Vreeken, R.J. The influence of citrate, EDTA, and heparin anticoagulants to human plasma LC-MS lipidomic profiling. *Metabolomics : Official journal of the Metabolomic Society* **9**, 337-348 (2013).

114. Barri, T. & Dragsted, L.O. UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics: effect of experimental artefacts and anticoagulant. *Analytica chimica acta* **768**, 118-128 (2013).

115.   Breier, M*., et al.* Targeted Metabolomics Identifies Reliable and Stable Metabolites in Human Serum and Plasma Samples. *PloS one* **9**, e89728 (2014).

116.   Kaluarachchi, M*., et al.* A comparison of human serum and plasma metabolites using untargeted 1H NMR spectroscopy and UPLC-MS. *Metabolomics : Official journal of the Metabolomic Society* **14**, 32 (2018).

117.   Bowen, R.A.R. & Remaley, A.T. Interferences from blood collection tube components on clinical chemistry assays. *Biochemia Medica* **24**, 31-44 (2014).

118.   Drake, S.K., Bowen, R.A., Remaley, A.T. & Hortin, G.L. Potential interferences from blood collection tubes in mass spectrometric analyses of serum polypeptides. *Clinical chemistry* **50**, 2398-2401 (2004).

119.   Lesche, D*., et al.* Does centrifugation matter? Centrifugal force and spinning time alter the plasma metabolome. *Metabolomics : Official journal of the Metabolomic Society* **12**, 1-6 (2016).

120.   Barton, R.H., Nicholson, J.K., Elliott, P. & Holmes, E. High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *International journal of epidemiology* **37 Suppl 1**, i31-40 (2008).

121.   Brunius, C*., et al.* Prediction and modeling of pre-analytical sampling errors as a strategy to improve plasma NMR metabolomics data. *Bioinformatics* (2017).

122.   Anton, G*., et al.* Pre-analytical sample quality: metabolite ratios as an intrinsic marker for prolonged room temperature exposure of serum samples. *PloS one* **10**, e0121495 (2015).

123.   Kamlage, B*., et al.* Quality markers addressing preanalytical variations of blood and plasma processing identified by broad and targeted metabolite profiling. *Clinical chemistry* **60**, 399-412 (2014).

124.   Jain, M., Kennedy, A.D., Elsea, S.H. & Miller, M.J. Analytes related to erythrocyte metabolism are reliable biomarkers for preanalytical error due to delayed plasma processing in metabolomics studies. *Clinica chimica acta; international journal of clinical chemistry* (2017).

125.   Bernini, P*., et al.* Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *Journal of biomolecular NMR* **49**, 231-243 (2011).

126.   Rasmussen, L.G*., et al.* Standardization of factors that influence human urine metabolomics. *Metabolomics : Official journal of the Metabolomic Society* **7**, 71-83 (2010).

127.   Saude, E. & Sykes, B. Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics : Official journal of the Metabolomic Society* **3**, 19-27 (2007).

128.   Brennan, L. NMR-based metabolomics: From sample preparation to applications in nutrition research. *Progress in Nuclear Magnetic Resonance Spectroscopy* **83**, 42-49 (2014).

129.   Bharti, S.K. & Roy, R. Quantitative 1H NMR spectroscopy. *TrAC Trends in Analytical Chemistry* **35**, 5-26 (2012).

130.   Stolzenburg, S., Lauridsen, M., Toft, H., Zalloua, P. & Baunsgaard, D. Improved quality of 1H NMR spectroscopic data for enhanced metabolic profiling of low molecular weight metabolites in human serum. *Metabolomics : Official journal of the Metabolomic Society* **7**, 270-277 (2011).

131.   Beckonert, O.*, et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2**, 2692-2703 (2007).

132.   Emwas, A.-H.*, et al.* Recommendations and Standardization of Biomarker Quantification Using NMR-Based Metabolomics with Particular Focus on Urinary Analysis. *Journal of proteome research* (2016).

133.   Sheedy, J.R., Ebeling, P.R., Gooley, P.R. & McConville, M.J. A sample preparation protocol for 1H nuclear magnetic resonance studies of water-soluble metabolites in blood and urine. *Analytical biochemistry* **398**, 263-265 (2010).

134.   Nagana Gowda, G.A. & Raftery, D. Quantitating metabolites in protein precipitated serum using NMR spectroscopy. *Analytical chemistry* **86**, 5433-5440 (2014).

135.   Levitt, M.H. *Spin dynamics: basics of nuclear magnetic resonance*, (John Wiley & Sons, 2001).

136.   Tiziani, S.*, et al.* Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. *Analytical biochemistry* **377**, 16-23 (2008).

137.   Rodriguez-Martinez, A.*, et al.* J-Resolved 1H NMR 1D-Projections for Large-Scale Metabolic Phenotyping Studies: Application to Blood Plasma Analysis. *Analytical chemistry* (2017).

138.   Savorani, F., Tomasi, G. & Engelsen, S.B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of magnetic resonance* **202**, 190-202 (2010).

139.   Wishart, D.S. Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry* **27**, 228-237 (2008).

140.   Sousa, S.A.A., Magalhães, A. & Ferreira, M.M.C. Optimized bucketing for NMR spectra: Three case studies. *Chemometrics and Intelligent Laboratory Systems* **122**, 93-102 (2013).

141.   Craig, A., Cloarec, O., Holmes, E., Nicholson, J.K. & Lindon, J.C. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical chemistry* **78**, 2262-2267 (2006).

142.   Kohl, S.M.*, et al.* State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics : Official journal of the Metabolomic Society* **8**, 146-160 (2012).

143.   Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of

complex biological mixtures. Application in 1H NMR metabonomics. *Analytical chemistry* **78**, 4281-4290 (2006).

144.   van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K. & van der Werf, M.J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics* **7**, 142 (2006).

145.   Eriksson, L., Kettaneh-Wold, N., Trygg, J., Wikström, C. & Wold, S. *Multi- and Megavariate Data Analysis : Part I: Basic Principles and Applications*, (Umetrics Inc, 2006).

146.   Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **2**, 37-52 (1987).

147.   Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**, 109-130 (2001).

148.   Bylesjö, M.*, et al.* OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics* **20**, 341-351 (2006).

149.   Trygg, J. & Wold, S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics* **16**, 119-128 (2002).

150.   Trygg, J. O2-PLS for qualitative and quantitative analysis in multivariate calibration. *Journal of Chemometrics* **16**, 283-293 (2002).

151.   Jonsson, P.*, et al.* Constrained randomization and multivariate effect projections improve information extraction and biomarker pattern discovery in metabolomics studies involving dependent samples. *Metabolomics : Official journal of the Metabolomic Society* **11**, 1667-1678 (2015).

152.   Whitley, E. & Ball, J. Statistics review 6: Nonparametric methods. *Critical Care* **6**, 509-513 (2002).

153.   Westerhuis, J.A., van Velzen, E.J., Hoefsloot, H.C. & Smilde, A.K. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics : Official journal of the Metabolomic Society* **6**, 119-128 (2010).

154.   van Velzen, E.J.*, et al.* Multilevel data analysis of a crossover designed human nutritional intervention study. *Journal of proteome research* **7**, 4483-4491 (2008).

155.   Xia, J., Psychogios, N., Young, N. & Wishart, D.S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic acids research* **37**, W652-660 (2009).

156.   Dona, A.C.*, et al.* A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Computational and structural biotechnology journal* **14**, 135-153 (2016).

157.   Wishart, D.S.*, et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research* (2017).

158.   Ulrich, E.L.*, et al.* BioMagResBank. *Nucleic acids research* **36**, D402-D408 (2008).

159.   Ludwig, C.*, et al.* Birmingham Metabolite Library: a publicly accessible database of 1-D 1H and 2-D 1H J-resolved NMR spectra of authentic metabolite standards (BML-NMR). *Metabolomics : Official journal of the Metabolomic Society* **8**, 8-18 (2012).

160.   Nagana Gowda, G.A., Gowda, Y.N. & Raftery, D. Expanding the limits of human blood metabolite quantitation using NMR spectroscopy. *Analytical chemistry* **87**, 706-715 (2015).

161.   Bingol, K., Li, D.-W., Zhang, B. & Brüschweiler, R. Comprehensive Metabolite Identification Strategy Using Multiple Two-Dimensional NMR Spectra of a Complex Mixture Implemented in the COLMARm Web Server. *Analytical chemistry* **88**, 12411-12418 (2016).

162.   Chignola, F.*, et al.* The CCPN Metabolomics Project: a fast protocol for metabolite identification by 2D-NMR. *Bioinformatics* **27**, 885-886 (2011).

163.   Ravanbakhsh, S.*, et al.* Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PloS one* **10**, e0124219 (2015).

164.   Dubey, A., Rangarajan, A., Pal, D. & Atreya, H.S. Pattern Recognition-Based Approach for Identifying Metabolites in Nuclear Magnetic Resonance-Based Metabolomics. *Analytical chemistry* **87**, 7148-7155 (2015).

165.   Hao, J.*, et al.* Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nature protocols* **9**, 1416-1427 (2014).

166.   Takis, P.G., Schafer, H., Spraul, M. & Luchinat, C. Deconvoluting interrelationships between concentrations and chemical shifts in urine provides a powerful analysis tool. *Nature communications* **8**, 1662 (2017).

167.   Tardivel, P.J.C.*, et al.* ASICS: an automatic method for identification and quantification of metabolites in complex 1D 1H NMR spectra. *Metabolomics : Official journal of the Metabolomic Society* **13**, 109 (2017).

168.   Rohnisch, H.E.*, et al.* AQuA: An Automated Quantification Algorithm for High-Throughput NMR-Based Metabolomics and Its Application in Human Plasma. *Analytical chemistry* (2018).

169.   Tredwell, G.D., Bundy, J.G., De Iorio, M. & Ebbels, T.M.D. Modelling the acid/base 1H NMR chemical shift limits of metabolites in human urine. *Metabolomics : Official journal of the Metabolomic Society* **12**, 1-10 (2016).

170.   Xiao, C., Hao, F., Qin, X., Wang, Y. & Tang, H. An optimized buffer system for NMR-based urinary metabonomics with effective pH control, chemical shift consistency and dilution minimization. *Analyst* **134**, 916-925 (2009).

171.   Weljie, A.M., Newton, J., Mercier, P., Carlson, E. & Slupsky, C.M. Targeted Profiling: Quantitative Analysis of 1H NMR Metabolomics Data. *Analytical chemistry* **78**, 4430-4442 (2006).

172. Lauridsen, M., Hansen, S.H., Jaroszewski, J.W. & Cornett, C. Human urine as test material in 1H NMR-based metabonomics: recommendations for sample preparation and storage. *Analytical chemistry* **79**, 1181-1186 (2007).

173. Baryshnikova, O.K., Williams, T.C. & Sykes, B.D. Internal pH indicators for biomolecular NMR. *Journal of biomolecular NMR* **41**, 5-7 (2008).

174. Barnes, S.*, et al.* Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. *Journal of mass spectrometry : JMS* **51**, 535-548 (2016).

175. Kanehisa, M. KEGG Bioinformatics Resource for Plant Genomics and Metabolomics. *Methods in molecular biology* **1374**, 55-70 (2016).

176. Jewison, T.*, et al.* SMPDB 2.0: big improvements to the Small Molecule Pathway Database. *Nucleic acids research* **42**, D478-484 (2014).

177. Ministers, N.C.o. *Nordic Nutrition Recommendations 2012*, (Nordic Council of Ministers).

178. Wishart, D.S.*, et al.* HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic acids research* **41**, D801-807 (2013).

179. Wishart, D.S.*, et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic acids research* **46**, D608-D617 (2018).

180. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

181. Wishart, D.S.*, et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic acids research* **37**, D603-610 (2009).

182. Walsh, M.C., Brennan, L., Malthouse, J.P., Roche, H.M. & Gibney, M.J. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *The American journal of clinical nutrition* **84**, 531-539 (2006).

183. Fave, G.*, et al.* Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples. *Metabolomics : Official journal of the Metabolomic Society* **7**, 469-484 (2011).

184. Wyss, M. & Kaddurah-Daouk, R. Creatine and creatinine metabolism. *Physiological reviews* **80**, 1107-1213 (2000).

185. Brosnan, J.T., da Silva, R.P. & Brosnan, M.E. The metabolic burden of creatine synthesis. *Amino Acids* **40**, 1325-1331 (2011).

186. Joncquel-Chevalier Curt, M.*, et al.* Creatine biosynthesis and transport in health and disease. *Biochimie* **119**, 146-165 (2015).

187. Davies, K.M., Heaney, R.P. & Rafferty, K. Decline in muscle mass with age in women: a longitudinal study using an indirect measure. *Metabolism: clinical and experimental* **51**, 935-939 (2002).

188.   Kochhar, S., *et al.* Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. *Analytical biochemistry* **352**, 274-281 (2006).

189.   Dorokhov, Y.L., Shindyapina, A.V., Sheshukova, E.V. & Komarova, T.V. Metabolic methanol: molecular pathways and physiological roles. *Physiological reviews* **95**, 603-644 (2015).

190.   Komarova, T.V., *et al.* Endogenous methanol regulates mammalian gene activity. *PloS one* **9**, e90239 (2014).

191.   Axelrod, J. & Daly, J. Pituitary gland: enzymic formation of methanol from S-adenosylmethionine. *Science (New York, N.Y.)* **150**, 892-893 (1965).

192.   Lindinger, W., Taucher, J., Jordan, A., Hansel, A. & Vogel, W. Endogenous production of methanol after the consumption of fruit. *Alcoholism, clinical and experimental research* **21**, 939-943 (1997).

193.   Chaleckis, R., Murakami, I., Takada, J., Kondoh, H. & Yanagida, M. Individual variability in human blood Metabolites identifies age-related differences. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 4252-4259 (2016).

194.   Chorell, E., *et al.* Plasma metabolomic response to postmenopausal weight loss induced by different diets. *Metabolomics : Official journal of the Metabolomic Society* **12**(2016).

195.   Assfalg, M., *et al.* Evidence of different metabolic phenotypes in humans. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1420-1424 (2008).

196.   A, J., *et al.* Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome. *Analytical chemistry* **77**, 8086-8094 (2005).

197.   Davis, T.A., *et al.* Amino acid composition of human milk is not unique. *The Journal of nutrition* **124**, 1126-1132 (1994).

198.   DeLany, J.P., Windhauser, M.M., Champagne, C.M. & Bray, G.A. Differential oxidation of individual dietary fatty acids in humans. *The American journal of clinical nutrition* **72**, 905-911 (2000).

199.   Dhamija, R., Eckert, S. & Wirrell, E. Ketogenic diet. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques* **40**, 158-167 (2013).

200.   Harris, R.C., *et al.* Absorption of creatine supplied as a drink, in meat or in solid form. *Journal of sports sciences* **20**, 147-151 (2002).

201.   MacNeil, L., *et al.* Analysis of creatine, creatinine, creatine-d3 and creatinine-d3 in urine, plasma, and red blood cells by HPLC and GC-MS to follow the fate of ingested creatine-d3. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **827**, 210-215 (2005).

202.   Heinzmann, S.S., *et al.* Stability and robustness of human metabolic phenotypes in response to sequential food challenges. *Journal of proteome research* **11**, 643-655 (2012).

203.   Rasmussen, L.G*., et al.* Assessment of the effect of high or low protein diet on the human urine metabolome as measured by NMR. *Nutrients* **4**, 112-131 (2012).

204.   Shindyapina, A.V*., et al.* Dietary methanol regulates human gene activity. *PloS one* **9**, e102837 (2014).

205.   Husøy, T*., et al.* Dietary exposure to 5-hydroxymethylfurfural from Norwegian food and correlations with urine metabolites of short-term exposure. *Food and Chemical Toxicology* **46**, 3697-3702 (2008).

206.   Roberts, G.R. & Sanderson, G.W. Changes undergone by free amino-acids during the manufacture of black tea. *Journal of the science of food and agriculture* **17**, 182-188 (1966).

207.   Schmidt, J.A*., et al.* Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: A cross-sectional analysis in the EPIC-Oxford cohort. *European journal of clinical nutrition* **70**, 306-312 (2016).

208.   Brosnan, J.T. Glutamate, at the interface between amino acid and carbohydrate metabolism. *The Journal of nutrition* **130**, 988s-990s (2000).

209.   Reeds, P.J*., et al.* Enteral glutamate is almost completely metabolized in first pass by the gastrointestinal tract of infant pigs. *The American journal of physiology* **270**, E413-418 (1996).

210.   Croze, M.L. & Soulage, C.O. Potential role and therapeutic interests of myo-inositol in metabolic diseases. *Biochimie* **95**, 1811-1827 (2013).

211.   Clements, R.S., Jr. & Darnell, B. Myo-inositol content of common foods: development of a high-myo-inositol diet. *The American journal of clinical nutrition* **33**, 1954-1967 (1980).

212.   Zeisel, S.H., Mar, M.H., Howe, J.C. & Holden, J.M. Concentrations of choline-containing compounds and betaine in common foods. *The Journal of nutrition* **133**, 1302-1307 (2003).

213.   Gakidou, E. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1345-1422 (2017).