THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Mathematical modeling in precision nutrition

Identification of metabotypes and prediction of metabolic response

VIKTOR SKANTZE

Food and Nutrition Science
Department of Life Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2023

Mathematical modeling in precision nutrition

VIKTOR SKANTZE
ISBN 978-91-7905-980-4

Acknowledgements, dedications, and similar personal statements in this thesis, reflect the author's own views.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5446
ISSN 0346-718X

Department of Life Sciences, Food and Nutrition Science
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Department of Systems and Data Analysis
Fraunhofer-Chalmers Research Centre for Industrial Mathematics
Chalmers Science Park
SE-412 88 Gothenburg
Sweden
Telephone +46 (0)31 309 94 00

Cover: Artwork by Johan Broberg

Printed by Chalmers digitaltryck
Gothenburg, Sweden 2023

Dedication

To my family and friends

Mathematical modeling in precision nutrition

Identification of metabotypes and prediction of metabolic response

VIKTOR SKANTZE

Department of Life Sciences

Chalmers University of Technology

ABSTRACT

Precision nutrition (PN) aims to tailor diets for individuals or groups based on comprehensive data to improve prevention of diseases, such as cardiometabolic diseases. Predicting individual postprandial metabolic responses and identifying individuals with similar metabolic phenotypes (metabotypes) could guide tailored diet strategies. While many metabolic markers are associated with health outcomes, predictive methods for high-dimensional postprandial responses are lacking. Furthermore, metabotyping has mainly been performed using cluster analysis on data from static blood markers or from responses to single dietary challenges. However, methods to incorporate time-resolved data from several dietary challenges or multi-omics (*e.g.*, metabolomics and microbiomics) have not been explored properly. This thesis breaks ground by addressing these challenges using time-resolved and static metabolomics, gut microbiota, dietary, and health status data.

The research presented in this thesis showed successful identification of metabotypes related to different cardiometabolic risks in a free-living population using multiple factor analysis of static microbiota and metabolomics. This led to deeper metabolic characterization compared to using single omics. Furthermore, dynamic mode decomposition was used to investigate the predictability of postprandial metabolic responses using the baseline metabolome and nutritional information of meals. The method was shown to be predictive in both measured (R^2=0.4) and simulated (R^2=0.65) data. It was also used along with the tensor decomposition CANDECOMP/PARAFAC to identify metabotypes relating to amino acid absorption in data from a crossover intervention study using repeated measurements from multiple dietary challenges, showing the utility of performing the two important PN tasks in one method. Finally, kinetic model parameters derived from postprandial plasma glucose dynamics were investigated to identify differential responders to meal challenges. Identified clusters were differently associated with type-2 diabetes risk markers and gut microbiota, which showed that differences in postprandial dynamics relate to type 2 diabetes risk markers and can be used to identify individuals at risk.

In conclusion, the analytical methods developed in this thesis present a versatile toolbox that may be used to improve metabotyping in complex study designs, enable dynamic predictions of postprandial responses, and demonstrate the utility of postprandial dynamics in detecting individuals at risk of disease.

Keywords: Personalized nutrition; precision nutrition; metabotypes; differential responders.

LIST OF PUBLICATIONS

I. **Viktor Skantze**, Carl Brunius, Mikael Wallman, Agnetha Rostgaard-Hansen, Ana Guadall Roldán, Ferran Reverter Comes, Maria Cristina Andrés Lacueva, Marie Palmnäs, Anne Tjønneland, Mats Jirstrand, Rikard Landberg, Jytte Halkjaer, Alejandro Sanchez Plá. Identification of metabotypes related to cardiometabolic risk factors and their association with diet in the Diet, Cancer, and Health-Next Generations MAX cohort. 2023
Manuscript.

II. **Viktor Skantze**, Mikael Wallman, Ann-Sofie Sandberg, Rikard Landberg, Mats Jirstrand, & Carl Brunius. Identification of metabotypes in complex biological data using tensor decomposition. *Chemometrics and Intelligent Laboratory Systems, 2023*
DOI: https://doi.org/10.1016/j.chemolab.2022.104733

III. **Viktor Skantze**, Mats Jirstrand, Carl Brunius, Ann-Sofie Sandberg, Rikard Landberg, & Mikael Wallman. Data-driven analysis and prediction of dynamic postprandial metabolic response to multiple dietary challenges using dynamic mode decomposition. *Frontiers in Nutrition, 2023*
DOI: https://doi.org/10.3389/fnut.2023.1304540
Accepted.

IV. **Viktor Skantze***, Therese Hjorth*, Mikael Wallman, Carl Brunius, Johan Dicksved, Erik A Pelve, Anders Esberg, Marilena Vitale, Rosalba Giacco, Giuseppina Costabile, Robert E Bergia, Mats Jirstrand, Wayne W Campbell, Gabriele Riccardi, Rikard Landberg. Differential Responders to a Mixed Meal Tolerance Test Associated with Type 2 Diabetes Risk Factors and Gut Microbiota—Data from the MEDGI-Carb Randomized Controlled Trial. *Nutrients 2023*.
DOI: https://doi.org/10.3390/nu15204369

*Joint first authorship.

CONTRIBUTION REPORT

Paper I: I was the first author and, performed exploratory analysis, interpreted the data and results, and wrote the manuscript. The co-authors contributed to the design of the study, writing parts of the manuscript, commenting on the writing, and interpreting of results.

Paper II: I was the first author and performed exploratory analysis, interpretation of data and results, and wrote the manuscript. Co-authors contributed to the design of the study, comments on writing, and interpretation of results.

Paper III: I was the first author and co-conceived the design of the study, performed the mathematical analysis and derivations, exploratory analysis, interpreted the data and results, and wrote the manuscript. The coo-authors contributed with analysis, comments on writing, and interpretation of results.

Paper IV: I was the shared first author and performed exploratory analysis, interpretation of data and results, wrote the parts of the manuscript relating to method, results, discussion, and conclusion, and co-wrote the other parts with the other shared first author. The co-authors contributed to the design of the study, commented on the writing, and interpretation of results.

ABBREVIATIONS

BMI	Body mass index
CANDECOMP	Canonical polyadic decomposition
CP	CANDECOMP/PARAFAC
DCH-NG	Diet, Cancer, and Health – Next Generations
DMD	Dynamic mode decomposition
DMDc	Dynamic mode decomposition with control
HbA1c	Glycated hemoglobin
HDL	High-density lipoprotein
HNFI	Healthy Nordic food index
hPDI	Healthy plant-based diet index
hs-CRP	High-sensitivity C-reactive protein
MFA	Multiple factor analysis
MUVR	Multivariate modeling with minimally biased variable selection in R
LDS	Linear dynamical system
PARAFAC	Parallel factor analysis
PCA	Principal component analysis
PDI	Plant-based diet index
pDMD	Parametric dynamic mode decomposition
pDMDc	Parametric dynamic mode decomposition with control
SVD	Singular value decomposition

TABLE OF CONTENTS

1 INTRODUCTION

Cardiometabolic diseases have increased dramatically during the last decades years (1). A recent definition of cardiometabolic diseases describes a collection of conditions and diseases including obesity, dyslipidemia, hypertension, prediabetes, type 2 diabetes, chronic kidney disease, nonalcoholic fatty liver disease, and metabolic syndrome (2). Importantly, there is a crosstalk between metabolic disorders and cardiovascular disease as they share similar biological pathways and intermediate risk factors (3). Cardiovascular diseases, *i.e.*, disorders of the heart and blood vessels, were until recently considered belonging to the cardiometabolic diseases but are now considered as steppingstones towards their development (2). Shared risk factors include central obesity, insulin resistance, reduced high-density lipoprotein (HDL), elevated triglycerides, low-density lipoprotein (LDL), total cholesterol, blood pressure, and fasting blood glucose (4,5). However, cardiometabolic diseases are to some extent preventable with a healthy lifestyle including healthy dietary patterns (6–9) and the effect of diet can be estimated for more efficient preventative advice.

The metabolic state of an organism is affected by exposure to diet and lifestyle and can be characterized using measurements of thousands of variables that collectively describe the molecular basis of human physiology related to health. This is possible due to the recent advances in the comprehensive measurement of omics data (*e.g.*, genomics, proteomics, and metabolomics), representing molecular information from genes to metabolites (10). Omics can provide comprehensive information on disease phenotypes that may be utilized to tailor the prevention of disease (11). Among the different omics data, metabolomics comprises the comprehensive assessment of small molecules, *i.e.*, metabolites that represent a final read-out of endogenous metabolism, the exposome (which includes gut microbiota, diet, drugs, *etc.*), and the interactions between the two (12).

Diet is one of the main modifiable lifestyle factors that is related to health and it is therefore an important target for prevention of cardiometabolic diseases (13). However, dietary recommendations are typically given on a population level and a large portion of individuals do not adhere to them (14–17). In addition, it is well known that individuals exhibit large differences in their responses to diet and in their nutritional requirements, which motivates tailoring of diet to meet the needs of individuals or groups of individuals with similar metabolic features for improved population health, *i.e.*, precision nutrition (18–21). Precision nutrition can be defined as providing the right diet to the right person at the right time *i.e.*, optimizing the dietary intake to the needs of an individual (22). However, precision nutrition could also be tailored to groups of individuals that have similarities in their gut microbiota, metabolism, or health traits (23). Individuals sharing similarities in metabolic traits can be referred to as sharing metabolic phenotypes or metabotypes. To pave the way for group-based precision nutrition based on metabotypes, there is a great need to identify and accurately define metabotypes and to tailor diets accordingly.

Identification of metabotypes has primarily been conducted by clustering static clinical markers measured at one point in time and investigating their links to disease outcomes (13). Metabotypes can also be identified from the metabolome and other omics, which may provide a more detailed and earlier view of the metabolic state than measurements of established

clinical biomarkers. This concept is, however, still underexplored (13,24–26). Furthermore, metabolism is a collection of continuous processes whereas a static measurement only provides a snapshot of the metabolic status. To capture more information about metabolism, time-resolved data (*i.e.*, sampled at several points in time) can be used. Furthermore, since metabotypes can be seen as groups of individuals with different metabolic regulations they may also respond differently to the same dietary exposures (23). This concept has not been explored thoroughly as very few studies have been conducted to relate identified metabotypes with time-resolved differential responses to diet (27). The data from such studies vary in time, metabolites, and subjects and require complex analytical tools that can handle responses to multiple meals. Such dynamic analyses become even more challenging when the molecular responses are measured in large numbers, frequently in thousands, as in omics data.

Even if metabotypes could characterize groups that may respond differently to food intake, a more direct way to identify responders/non-responders is to use predictive models based on recorded postprandial time-resolved responses to different foods and then estimate the response to new foods. This approach has previously been applied to single or few metabolic markers such as glucose levels (28–30), insulin, and triglycerides (31,32). However, the human metabolism is estimated to produce over 200,000 metabolites that reflect biological processes and it is estimated that changes in more than 22,000 of these are related to health and disease (33). Therefore, it is of great importance for the advancement of precision nutrition to predict not only the response of well-established metabolic markers but also to explore the postprandial response in as many metabolites as possible. Moreover, to my knowledge, there have been no reported methods to combine the metabotyping concept with the prediction of responders/non-responders in dynamic data, *i.e*. finding different metabotypes in dynamic metabolite data.

To achieve this, methods that can identify differential responders and metabotypes in complex study designs and that predict high dimensional postprandial response are thus highly warranted. Such designs involve the collection of several omics sources from the same individual (multi-omics) and postprandial time-resolved biological data. Moreover, deeper exploration of using static multi-omics to identify metabotypes is needed as this can facilitate the characterization of the molecular phenotype of individuals and be used for early identification of groups at differential disease risk that may benefit from differential dietary recommendations.

2 OBJECTIVES

This thesis aimed to develop novel algorithmic methods for the identification of metabotypes and prediction of postprandial responses to food, to advance the field of precision nutrition and thereby aid in the prevention of cardiometabolic diseases. Metabolomics and microbiota data as well as clinical measures stemming from intervention trials (crossover and parallel designs) and cohorts of free-living individuals were to be used to unravel interindividual metabolic differences (metabotypes) and to predict metabolic responses to food.

Specific objectives:

A1) To identify metabotypes related to cardiometabolic risk factors in static multi-omics data and investigate their association with diet (paper I)

A2) To develop descriptive and predictive methods to identify metaboptypes in high-dimensional time-resolved postprandial data from dietary interventions (papers II, III)

A3) To develop a method to predict high-dimensional postprandial response to food over time (paper III)

A4) To identify differential responders in blood glucose response to standardized meal tests using dynamical modeling (paper IV)

3 BACKGROUND

3.1 Cardiometabolic diseases

Cardiometabolic diseases (CMD) comprise a group of diseases and conditions including obesity, dyslipidemia, hypertension, prediabetes, type 2 diabetes, chronic kidney disease, nonalcoholic fatty liver disease, and metabolic syndrome, that together constitute the most common cause of death globally (34). Until recent decades, cardiometabolic diseases were mostly increasing in the industrialized world, but the rates are now among the highest in the world in developing countries (35,36).

Many of the cardiometabolic diseases share etiology and determinants including genetics, smoking, excessive drinking, and environmental exposures such as microbiota, lifestyle, and diet (2). They also share intermediate risk factors, such as central obesity, insulin resistance, reduced high-density lipoprotein (HDL), elevated triglycerides, low-density lipoprotein (LDL), total cholesterol, blood pressure, and fasting blood glucose (37). Blood lipid dysregulation, in particular of cholesterol, has also been identified as a major risk factor for cardiometabolic diseases (38). Specifically, low-density lipoprotein (LDL) is involved in the development of coronary heart disease and is an established risk factor for cardiometabolic diseases at large. Hypertension is another important risk factor for stroke, heart failure, and renal failure (38). Elevated fasting blood glucose can be a consequence of the development of insulin resistance and type 2 diabetes (T2D) (39), which together with hypertension are the main causes of chronic kidney disease and nonalcoholic fatty diver disease (40,41).

3.1.1 Prevention of cardiometabolic diseases

Promoting healthy lifestyle habits, including exercise, sleep, and management of stress and diet, are the main preventive actions to combat cardiometabolic diseases (2) since many of the cardiometabolic diseases and their pre-conditions are reversible with a healthy diet and lifestyle (42). Both primary and secondary prevention (*i.e.*, avoiding the disease altogether and preventing relapse, respectively) are of importance in reducing morbidity and mortality inherent to cardiometabolic diseases (43).

Several studies have demonstrated how a healthy diet can be used to prevent cardiometabolic diseases. For example, in the large randomized control trial PREDIMED, the Mediterranean diet supplemented with extra virgin oil or nuts was shown to reduce the incidence of cardiovascular events compared to a reduced fat diet (43). Moreover, several dietary intervention studies with Nordic diet have shown beneficial effects on cardiometabolic disease risk factors (44,45). Also, observational studies have shown that the healthy Nordic diet was associated with lower mortality risk and low-grade inflammation (46,47). Furthermore, the quality of different food items can have a large impact on the risk of developing cardiometabolic diseases. For instance, in an observational study, it was shown that high vs low intake of refined carbohydrates was associated with an increased risk of developing type 2 diabetes while the consumption of dietary fiber showed an inverse association (48). Additionally, saturated and trans fats have been associated with an increased risk of developing coronary heart disease (49). Further preventative advice can possibly be provided when the effect of diet on the cardiometabolic diseases is accurately measured.

3.1.2 Measurements of the metabolism

The metabolism is the collection of chemical processes in the body necessary to sustain life. The exposome, *i.e.*, the collection of exposures to the body, including diet, is mirrored in the metabolome, which consists of thousands of metabolites, *i.e.*, small molecules derived from the interaction of the human or microbial metabolism with the exposome (50). The metabolites serve as intermediate or final products of metabolic reactions that take place by the host or its microbiota. Examples of metabolites include amino acids, sugars, lipids, and vitamins (51). Thus, the metabolome can reflect what is absorbed from the diet, but also the state of the human metabolic system which is linked to the state of health.

In metabolomics, different analytical techniques are applied to measure thousands of metabolites from a single biological sample (51). Two main approaches exist in metabolomics; targeted and untargeted metabolomics. In targeted metabolomics, a set of *a priori* selected metabolites are considered in targeted analysis while when using the untargeted approach, the aim is to measure 'all' the metabolites available in the sample although often not with the same precision as in the targeted case. A limitation of the latter approach is the difficulty in the identification of the large number of unknown features detected. Among these techniques, liquid chromatography-mass spectrometry has become an important tool to obtain a broad coverage of the measurable metabolome (52).

While metabolomics covers a wide range of the metabolome, these techniques are also time-consuming and typically require sample management, advanced equipment, and a high degree of technical skill. Specific metabolites and metabolic markers can be identified with other, faster, and more robust methods. For example, clinical blood markers such as blood glucose, lipids, cholesterol, and specific hormones are easily measured and provide a rough image of the metabolic state. Furthermore, a recent technology called continuous glucose monitoring (CGM) has emerged as a tool for diabetic patients, enabling them to monitor their blood glucose levels themselves (53). The CGM device can be installed in the arm and measure glucose levels every 5 minutes, providing high-resolution data on glucose dynamics throughout the day. This can be particularly useful to monitor postprandial glucose dynamics, which relates to diabetes development (54).

3.1.3 Gut microbiota – a link between diet and cardiometabolic diseases

Gut microbiota has also emerged as an important risk factor for cardiometabolic diseases and appears to play a role in metabolic diseases, immunity, and cardiometabolic health (55–57). Specifically, the effects of the microbiota can change the metabolism and increase the risk of developing cardiometabolic diseases. The gut microbiota includes bacteria, archaea, fungi, and viruses, that reside in the digestive tract. Importantly, food components that are not absorbed in the small intestine, such as dietary fiber, can be fermented by bacteria in the large intestine to produce compounds with health-promoting local and systemic effects, such as short-chain fatty acids and indole propionate (58). However, other nutrients like protein can also be fermented in the lack of carbohydrates, creating potentially toxic metabolites such as sulfides and ammonia (59). The prevalence of these bacteria is often measured in feces as a proxy for the gut, using gene sequencing. Here, DNA reads of bacteria are matched with a reference sequence database to determine the presence of bacteria at different taxonomic levels such as species, genus, family, *etc*. (60).

The interaction between the gut microbiota and human metabolism is still not completely mapped but important information capturing parts of these processes can now be measured using omics. This collection of measurements of the metabolome, microbiome, and clinical markers make up an important image of the metabolic state of the individual that could be used to infer metabolic function, disease risk, and also insights into how to prevent disease.

3.2 Precision nutrition

Current dietary recommendations are adapted to promote health and prevent disease at a population level but they are poorly adhered to (15,17,61). Policies and advice are directed to the general public and to specific targeted groups, such as pregnant women, children, and the elderly, to ensure their nutritional requirements are met (62). Studies that underpin existing recommendations are often large-scale epidemiological studies where the role of specific nutrients, foods, or dietary patterns has been investigated in general populations (63). However, these recommendations fail to accommodate the inter-individual variation in the effects of the diet on metabolism and health (64).

Recent research has shown that different individuals can respond metabolically differently to the same food. Large inter-individual variation have been found for plasma glucose, triglycerides, and insulin (28,32) after the intake of standardized meals, but also in response to salt (65) and caffeine (66). Determinants of the variance in response to food can be attributed to sex, genetics, metabolic profile, exposome (including diet and gut microbiota), and yet unknown factors (64). This implies that dietary recommendations should ideally be given on an individual level to maximize beneficial effects and prevent disease outcomes in *e.g.*, cardiometabolic diseases (13). Further potential benefits of personalized dietary advice include reducing health care costs and improving individual's motivation to change their dietary habits (67,68).

Precision nutrition deals with dietary recommendations tailored to the individual's specific needs based on information on the metabolic state using *e.g.* clinical measures, metabolome, and microbiome (63). Advice can also be tailored to groups of metabolically similar individuals who would benefit from the same dietary pattern. The concepts of precision nutrition and 'personalized nutrition' have been used interchangeably and there is no consensus on terminologies although there have been attempts to provide a distinction between the two concepts (69). In this thesis, precision nutrition is the preferred term because it covers broader aspects of tailoring diet on individual and group levels (63,69).

Precision nutrition has not yet widely been implemented in a clinical setting for the prevention of disease (in the general population) or official dietary guidelines since the field is yet in its infancy with more research needed to establish efficient strategies as well as evidence of their efficacy and effectiveness (13,23,63).

Initial research in precision nutrition was mainly focused on investigating the links between nutrition and genetics (70). Later, the field of nutrigenomics emerged, investigating the interactions between nutrition and the genome using genomic tools (71). It was shown that genetics affect the metabolism and how dietary components act on the signaling and expression of genes (65). Genetics plays a role in lactose tolerance (73), and metabolic syndromes (74) and have been investigated in sports contexts (75). Furthermore, specific genetic predispositions play a role in several other diseases and conditions, such as in weight management (76). This has been further confirmed in a randomized control trial on weight loss

called the Now-trial, where individuals were divided into two groups and given weight management advice based on standard or standard with added tailored nutrigenomic information (77). The group given nutrigenomic advice had a significantly greater reduction in weight compared to the control group after the trial (78). The personalized nutrition study (POINTS) also investigated weight loss in relation to personalized nutrition based on responders to fat and carbohydrates identified using genotypic information. The results from this trial showed no improved weight loss when consuming the genotype concordant diet (79). Furthermore, the ongoing MyGeneMyDiet study aims to investigate the effect of weight loss using genotype information to provide nutrition and lifestyle recommendations (80). Nutrigenomics has also been investigated in relation to other factors in cardiometabolic health (81). In the large Food4Me study, dietary advice based on genotypes and differential phenotypes derived from anthropometry and blood biomarkers was used to infer dietary advice in a randomized controlled study (82). It was concluded that the inclusion of genetic information did not improve clinical measures such as total cholesterol, carotenoids, fatty acids, and vitamin D, compared to getting personalized advice based on the baseline diet. However, it has been shown that adherence to diet has been improved when the diet is tailored, providing promise to the concept of personalization (82,83). An example of this is that individuals carrying the obesity-associated FTO-gene had an increased reduction in body weight when they were informed about their genetic predisposal compared to the non-carriers of the gene (82).

More recent studies aiming to provide precision nutrition on individual and group levels have been conducted focusing on measuring the metabolism in a fasted state, but also in response to meals. The postprandial response has been shown to be important to map the metabolic system since static measurements do not account for dynamic events (84). Zeevi et al. conducted the first large-scale landmark-study in the field of precision nutrition, where postprandial glucose levels were predicted in response to different foods (28). In their algorithm, they included blood parameters, dietary habits, anthropometrics, physical activity, and gut microbiota to predict personalized responses to different meals. It was concluded that there was high interpersonal variability in glucose responses to the same food and that this variability was associated with clinical and microbiome profiles (28). Furthermore, an algorithm was successfully developed to tailor personalized diets to optimize postprandial glucose response, resulting in “good” or “bad” foods identified for the participants. Interestingly, foods that were predicted as “good” in some individuals were predicted as “bad” in other individuals, demonstrating the interindividual variation. Similar studies have subsequently been performed showing that a machine learning model provides better glycemic control than standard-of-care methods (30,85).

Many studies, including the one from Zeevi et al., have used CGM to measure glucose variability. In a recent study, CGM measurements of free-living individuals were used to estimate their glucose variability and to compare its efficacy for the assessment of glucose control with the use of standard clinical measures. Individuals who were classified as having efficient glucose regulation according to standard static metrics actually reached pre-diabetic glucose ranges. These results show the importance of time-resolved data and the potential of wearable devices in precision nutrition. Most PN-studies have focused on glycemic responses, but recently other metabolic factors have also been included. In the PREDICT trials, metabolic responses to food have been more widely assessed. In the PREDICT 1 trial, 1002 individuals

(where 600 were twins) were given test meals, and their glucose, insulin, and triglyceride responses to the food were recorded. These responses were used to train a machine learning model to predict responses to new meals and successfully predict glucose and triglycerides in a separate validation cohort (32). It was shown that genetics did not contribute largely to postprandial response, in concordance with the Food4Me trial and other nutrigenomics studies. Instead, other drivers of the interindividual differences were meal composition, habitual diet, meal context, anthropometry, microbiome, clinical and biochemical parameters. Yet another determinant is the timing of meals which also affects the response to food and fasting for longer periods of time has been associated with beneficial health outcomes (86). Hence, the determinants influencing the metabolic state are many, comprising a complex interaction of factors to be used in precision nutrition (Fig. 1).

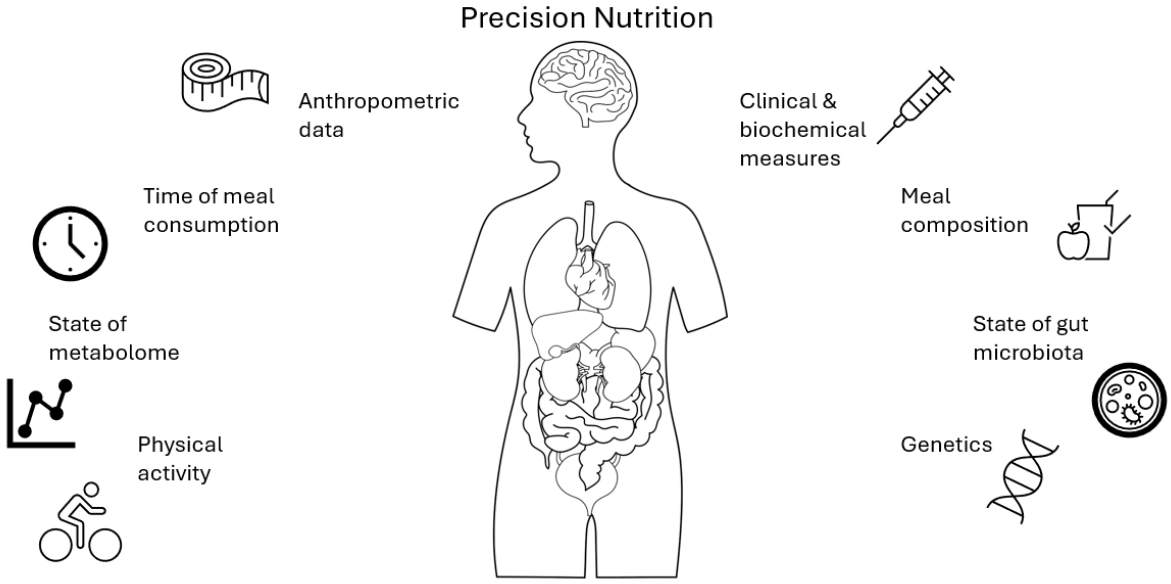

Figure 1: A conceptual visualization of factors that may contribute to inter-individual variance in diet-related health outcomes in precision nutrition.

3.2.1 Metabotypes - metabolic phenotypes

Precision nutrition strategies described above are directed towards individuals and such approaches require comprehensive assessment and measurements of individual traits which are expensive. An alternative approach could be to provide tailored dietary advice on a group level, where individuals with similarities in their metabolic regulation, i.e.metabolic phenotypes or *metabotypes*, can benefit from similar dietary recommendations (23). The existence of metabotypes that are differently associated with disease risk, health conditions, and/or response to diet have been demonstrated in several studies where clustering of static metabolic markers was investigated in relation to disease outcomes. For instance, five metabotypes were identified in Swedish individuals with type 2 diabetics using six variables (glutamate decarboxylase antibodies, age at diagnosis, BMI, HbA1c, and homoeostatic model assessment 2 estimates of β-cell function and insulin resistance) (87). The findings have been replicated in other populations and show that type 2 diabetes is a heterogeneous disease, depending on metabolic factors which were also associated with different complications, suggesting the possibility of tailoring treatment for each group. Furthermore, identifying patients with different metabotypes based on insulin resistance (precondition to type 2 diabetes) and the differential effect of diet tailored to these groups were tested in the PERSON-study where tissue-specific

insulin resistance phenotypes were given hypothesized optimal and suboptimal diets (88). After the 12-week intervention, the group receiving the optimal diet improved in their cardiometabolic health markers in terms of glucose regulation, suggesting the importance of tailored treatment based on specific disease phenotypes (89).

In another study, three metabotypes were identified from biochemical and anthropometrical measures which were associated with differential risk of developing cardiometabolic diseases and after a seven year follow-up, disease incidence was greater in the predetermined unhealthy metabotype (90). From a dietary metabotype perspective, it is hypothesized that individuals belonging to the same metabotype also have similar metabolic regulation and that similar dietary recommendations would be beneficial for them. However, the response to food has typically not been recorded in these studies (since static markers were used for clustering), which may vary within the supposed metabotype. Another way of identifying metabotypes is to expose individuals to dietary interventions and record their metabolic responses over time. Individuals who respond similarly are likely to have similar metabolic regulations. If groups of differential dynamic responders are associated with differential risks of an outcome, it is more likely that these individuals would benefit from similar dietary recommendations. This concept was tested in women (n = 24) exposed to high and low glycemic index meal tests where their metabolic response in terms of blood glucose, insulin, leptin, and non-esterified fatty acids was recorded (91). Three groups of differential responders were found, of which two were associated with subclinical metabolic dysfunctions.

3.2.2 Data analysis tools in precision nutrition

Generally, individuals who have large differences in measured levels of metabolic markers of interest could be considered to have different metabolic profiles. However, challenges to identify such differences arise when the number of measured metabolic markers increases. Thus, it is useful to represent the data in a reduced form where the most distinct differences between individuals are highlighted. A common way to perform data reduction is to arrange the data in a matrix and perform matrix decomposition, *e.g.*, principal component analysis (PCA) (92), which will be discussed in greater detail in Section 3.3. Essentially, the matrix is represented by several pairs of vectors, *i.e.*, scores describing the observations and loadings describing the contribution of each original variable to the component. Scores and loadings thus facilitate interpretation of the major patterns in the data in a low number of components, *i.e.*, a latent space, described by the scores when using PCA. The scores can be used *e.g.* for clustering, where some measure of distance (*e.g.*, Euclidean or Manhattan distance) in latent space is used to assess similarity between individuals and subsequent grouping into metabotypes.

While many clustering methods that can be used for metabotyping from matrix decomposed data exist, they all share a common trait of being unsupervised. This means that ground truth clusters are unknown. The method uses the available data together with the intrinsic heuristics to estimate the best candidate clusters according to a given criteria. This trait also implies that there is no guarantee that the clusters found by the algorithm are clinically meaningful. Therefore, clusters must be investigated in terms of associations with other clinically relevant measures. This has been the procedure to identify metabotypes using static biochemical measures (82,90,93–95).

Further opportunities to identify clusters relating to biologically meaningful phenomena can be attained using combined modalities of data (*e.g.*, several omics data). For example, clusters that are reflected in metabolome, microbiome, and clinical parameters can be expected to provide more information about the metabolic mechanisms behind the clusters than using solely clinical parameters, which has been the main approach (13). However, there is still no standard data reduction method for the combination of such datasets, but this is a highly active research field (96). The different omics datasets are often subject to different inherent data distributions (*e.g.*, metabolomics data is measured on a continuous scale, and microbiota is measured in counts of detections in the sample) depending on their source, leading to several choices for preprocessing and reduction method of choice. Ideally, all omics measurements should be performed on each subject to avoid unmatched data, requiring imputation or filtering of subjects with missing information.

No standard method for identification of differential responders to food exists yet, since the concept still is underexplored (13,63). To assess differential responders, the physiological response of the food must be measured repeatedly in some biologically relevant marker. In static measurements, the response of the markers would accommodate the columns and the individual samples in the rows of a matrix, respectively. However, in time-resolved responses, the time-dependent measurement of the same variable as in adds a third dimension resulting in an array with three dimensions, *i.e.*, a third-order tensor with indices: individuals, metabolites, and time (97). Potential methods to reduce such tensors include CANDECOMP/PARAFAC (CP) which is interpreted similarly to PCA, albeit with loadings for both metabolites and time points (97). Furthermore, in studies investigating differential response, it is relevant to expose the subjects to several dietary challenges to explore more of their metabolic system, in *e.g.* crossover designs. This induces a fourth dimension to the tensor as the same metabolic markers are measured at the same time points in response to different foods. Methods reducing tensors to more interpretable components have so far gained very little attention in precision nutrition, but hold great potential in describing relations between dynamic response and metabolic mechanisms (13,63).

Another way to reduce dynamic metabolic response to food is to identify a predictive model representing a relationship between dietary intake and the metabolic response instead of a purely descriptive one. This relationship can be described as a mathematical function of time and dietary input, which represents the metabolic processes that produce the postprandial response output over time. In pharmacology, the relationship between an administered drug and the physiological response is routinely subjected to mathematical modeling (98). The response to the drug is often measured over time and modeled using differential equations, which describe the rate of change in drug concentration over time (99). The same principles can be applied in response to food. The model is typically built using mathematical functions together with parameters describing different parts of the metabolic processes (Fig. 2). These parameters are estimated when fitting the model to the response data, which yields a reduced representation of the response if the model contains fewer parameters than observations in time, dietary exposures, *etc*. Such a model can be used to predict individual responses to, for instance, new dietary interventions. Previously, dynamic prediction models have been built using machine learning algorithms like random forest and resulted in successful predictions of a few metabolic markers (32). However, random forests are general prediction models and do not provide interpretable parameters in terms of metabolic processes. Additionally, they do not

inherently provide multivariate predictions over time. High-dimensional multivariate predictions have previously been performed using deep learning methods, but not using time-resolved data (100). Deep learning methods are however known to require a lot of training data and generate models that are large and often challenging to interpret (101).

In summary, sound methodological developments regarding the reduction of mixed data sources and dynamic postprandial response to identifying metabotypes, and prediction of high-dimensional metabolic response to food are warranted.

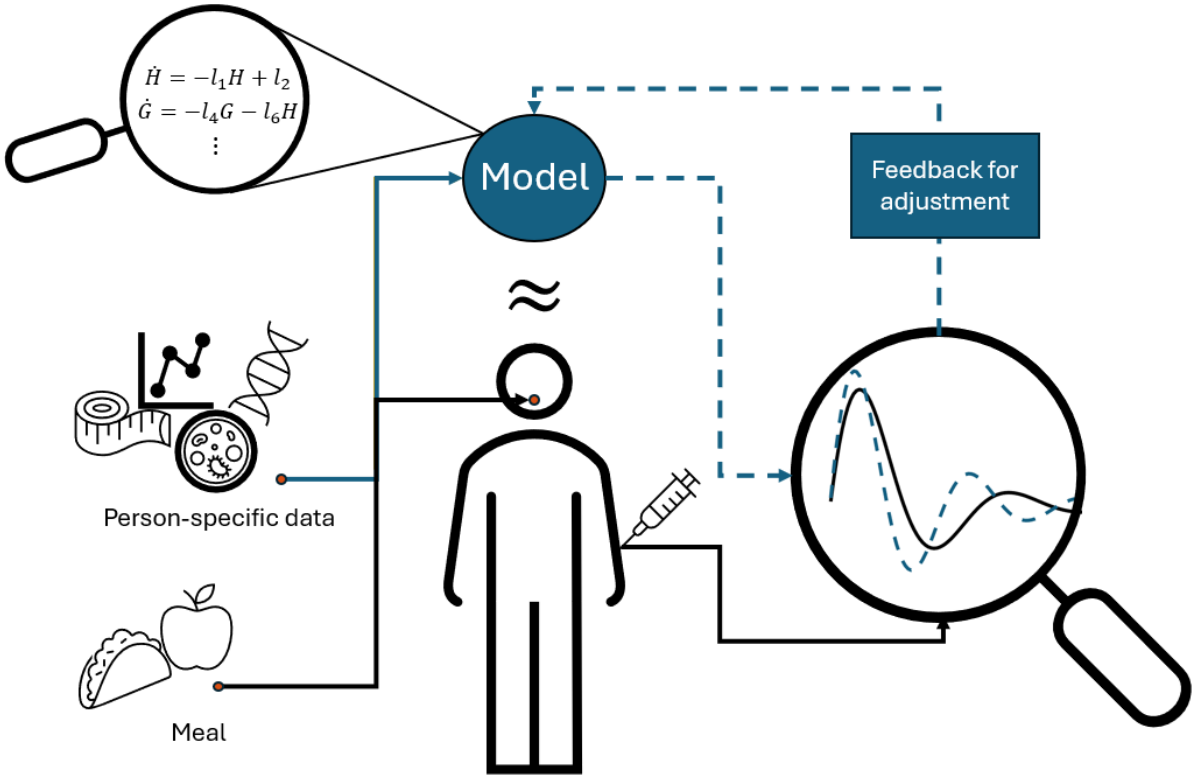

Figure 2: Conceptual visualization of modeling of the metabolic system

4 METHODS AND DATA

4.1 Data

In this thesis, data from three studies have been used for the development of methods and to address the research questions.

4.1.1 Diet, Cancer and Health- Next Generation MAX- a study in free-living men and women

Multi-omics data from the Danish Diet, Cancer and Health- Next Generations MAX study were used to identify metabotypes in a population of free-living men and women (paper I). The Danish free-living cohort study Diet, Cancer and Health-Next Generations (DCH-NG) was conducted between 2015 and 2019, to assess the association between heredity, genetics, diet, and multiple forms of cancer and other non-communicable diseases (102). The DCH-NG cohort included 39 554 individuals and a subset of the population (n = 720) was included in a new study named DCH-NG MAX which was set up to validate and assess the reproducibility of fecal microbiota, diet, saliva and metabolomics in stool, urine, and plasma. The study participants made three visits to the clinic (baseline, 6 months, and 12 months) where measurements were taken. Completion of a lifestyle questionnaire, two 24-hour dietary recall, and a food frequency questionnaire were also done at the given time points. To chart dietary intake, the amount of consumed food subgroup (self-estimated portion sizes in weight or volume) was summarized over two consecutive days. Furthermore, dietary indices were calculated from the 24-hour dietary recalls and included; healthy plant-based diet index, healthy Nordic diet score and unhealthy plant-based diet index, as described in previous studies (47,104,105,106). During the visits the following samples and measurements were obtained: urine, saliva, blood, stool, blood pressure, weight, waist and hip circumference, muscle mass, total fat mass, fat-free mass, visceral fat, and height.

4.1.2 A dietary crossover intervention study to compare metabolic response to three different meals

Data from a crossover intervention trial (106,107) were used to identify metabotypes in postprandial dynamic metabolomics data (papers II and III). The trial comprised of middle-aged overweight men (n =17, BMI 25–30 kg/m^2, 41–67 years of age). All consumed three different diets (pickled herring, baked beef, and baked herring) on separate test occasions with one washout week in between test meals (Fig. 3). Baseline clinical measures along with anthropometric measures were recorded, including alanine aminotransferase (ALAT), aspartate transaminase (AST), gamma-glutamyl transferase (GGT), cholesterol, low-density lipoprotein (LDL), creatinine, thyroid stimulating hormone (TSH), and body mass index (BMI). On each test occasion, blood samples for metabolomics analysis were taken 8 times at one-hour intervals, including a baseline sample just before the meal was consumed. Metabolomics analysis was performed using GC-MS and 79 targeted metabolites were measured, resulting in 79 metabolite trajectories based on 8 time points for three different diets for all individuals. The main metabolites belonged to the following compound classes: amino acids (n=35), carboxyl acids (n=6), lipids (n=8), and carbohydrates (n=16).

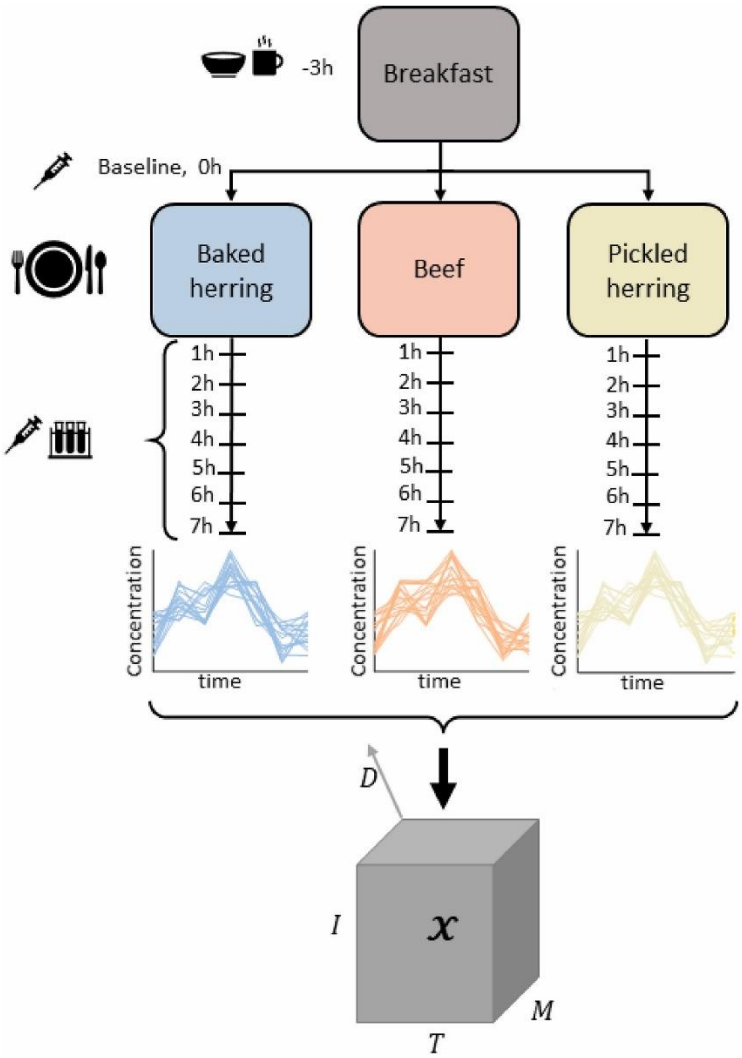

Figure 3: Conceptual overview of the dietary crossover intervention with three test meals

4.1.3 Dietary intervention trial to evaluate the effects of a high vs. low glycemic index diet on cardiometabolic risk factors

Data from the MEDGI-Carb trial (108,109) were used to identify differential responders of postprandial glucose in response to standardized meals served during test days before and after 12-week intervention with either based on high or low GI diets in a context of a healthy Mediterranean diet pattern (paper IV) in participants at risk of developing type 2 diabetes.

The MEDGI-Carb trial is an international multi-center randomized, controlled, parallel-group, 15-week dietary trial, consisting of a 3-week baseline period followed by a 12-week controlled dietary intervention in adults at elevated risk of developing type 2 diabetes. During the 12-week intervention period, participants consumed a Mediterranean-style, controlled, isocaloric, weight-maintenance diet. Furthermore, the participants were instructed to consume either a low-GI or high-GI diet with intervention-specific foods. Half of the daily carbohydrate intake was identical in the two groups, including vegetables and fruit. The other half consisted of carbohydrates with GI < 55 and > 70 in the low and high GI groups, respectively. Markers of glucose metabolism were obtained during standardized testing days by completion of an eight-hour mixed meal tolerance test, an oral glucose tolerance test, and 6 days of 24-hour CGM at baseline and post-testing. Here, anthropometric traits were measured and blood samples were

drawn to estimate HbA1c, insulin, glucose, HDL, triglycerides, and blood pressure. Insulin sensitivity indices such as the quantitative insulin sensitivity check index (QUICKI), Stumvoll, and Matsuda were calculated using data from the OGTT (110).

4.2 Methods used to generate data

The data modalities used in the research presented in this thesis are listed in Table (1).

Table 1: Data sources and types of data used in the research papers presented in this thesis.

Data sources & types	**Paper I**	**Paper II**	**Paper III**	**Paper IV**
Anthropometric traits	✓	✓	✓	✓
Blood glucose				✓
HbA1c	✓			✓
Insulin				✓
Triglycerides				✓
Cholesterol	✓			
hs-CRP	✓			
Blood pressure	✓			
Thyroid-stimulating hormone		✓	✓	
Creatinine		✓	✓	
Gamma-glutamyl transferase		✓	✓	
Alanine aminotransferase		✓	✓	
Aspartate transaminase		✓	✓	
Targeted metabolomics	✓	✓	✓	
Untargeted metabolomics	✓			
Microbiota	✓			✓
Time-resolved data		✓	✓	✓

4.3 Data reduction of high-dimensional biochemical data and multi-omics

4.3.1 Matrix decompositions

To identify metabotypes using multi-omics data, multi-factor analysis was used to reduce the data and subsequently to cluster the reduced representation of individuals (paper I). In this section, the theory of these methodologies is explained.

Static metabolic markers allow for the reduction of dimensionality using matrix decompositions. Several decompositions exist but the most widely used of these techniques is PCA. In PCA, a low-dimensional representation of the data is identified in which global patterns in the data are captured. Essentially, vector pairs (scores representing samples and loadings representing variables) called components denote a new coordinate system that facilitates the identification of the largest variation in the data. This method is popular for data reduction since there always exists a solution to PCA for a real matrix and since calculating the solution is very efficient (111). A very close connection exists between PCA and the matrix decomposition singular value decomposition (SVD) which can be used to calculate PCA. In SVD, data $\boldsymbol{X}$ is decomposed into three matrices according to Eq. 1a, where $\boldsymbol{U}$ denotes the left singular vectors, $\boldsymbol{V}$ the right singular vectors, and $\boldsymbol{\Sigma}$ the singular values in placed in a diagonal matrix. These are used to obtain the standard PCA solution (Eq. 1b), given that the columns of $\boldsymbol{X}$ are mean subtracted, the PCA scores ($\boldsymbol{T} = \boldsymbol{U\Sigma}$) are the left singular vectors $\boldsymbol{U}$ scaled by the singular values in the diagonal matrix $\boldsymbol{\Sigma}$ of the SVD, and the scaled right singular vectors of the SVD $\frac{\boldsymbol{V\Sigma}}{\sqrt{\boldsymbol{n}}}$ (where *n* is the number of rows) are called loadings (111).

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\intercal} \tag{1a}$$
$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{V}^{\intercal} \tag{1b}$$

Additionally, the SVD is useful since computed parameters of the matrix decomposition are the global solution to the optimization problem of finding the components explaining the most variance. Moreover, the algorithms for computing the SVD and PCA are deterministic which facilitates interpretation of the parameters. In contrast, it is more challenging to interpret results if they vary at every run of the algorithm, which is the case for other machine learning methods.

A component-wise description of PCA is shown in Equation 2 where factors (a general term for scaled loadings and scores used in Sec. 3.5.1) t_i and v_i (multiplied using the outer product denoted as $\circ$) together denote F components that are added to equal the data $\boldsymbol{X}$.

$$\boldsymbol{X} = \sum_{i=1}^{F} \boldsymbol{t}_i \boldsymbol{v}_i^{\intercal} = \sum_{i=1}^{F} \boldsymbol{t}_i \circ \boldsymbol{v}_i \tag{2}$$

where all $\boldsymbol{t}_i^{\intercal}\boldsymbol{t}_j = 0$ and $\boldsymbol{v}_i^{\intercal}\boldsymbol{v}_j = 0$ when $j \neq i$, for all 1 to F components, *i.e.*, the components are orthogonal to each other. Here, F denotes the number of components to completely reconstruct the data. Since the components of PCA range in explained variation, we can truncate the decomposition, effectively removing parts of the data that are of little interest for analysis, *e.g.*, estimated noise or non-systematic variance. However, choosing the number of components of systematic interest can be a challenging task, but heuristic methods like the Scree plot can be used (92). Furthermore, a proxy for the measure of noise in the data can be attained using the rank of the matrix which is an analytical property in matrix theory denoting the number of linearly independent vectors the matrix contains (112).

When using matrix decompositions to identify metabotypes, the vectors representing individual samples (scores in PCA) are clustered. However, when clusters are sought in more than one omics source, joint reduction of datasets (matrices) is useful. There are many options to reduce multi-omics jointly. Arguably the most intuitive method is called multiblock analysis, which essentially amounts to concatenating the different datasets into one large matrix and performing a single matrix decomposition using *e.g.*, PCA (113). A common method that partly uses methods from multi-block analysis is called multiple factor analysis (MFA) and is made to include discrete as well as continuous data. Other methods include joint matrix decompositions that essentially decompose each dataset differently by sharing scores between decompositions and allowing for freedom in loading vectors for each individual dataset. However, these methods are computationally costly and no global solution is guaranteed to be found in general (114). A common limitation of the method includes that different distributions of the datasets might not allow for interpretable results as artifacts can arise. In joint matrix decompositions, advances have been made to allow for different distributions per dataset (115).

4.3.2 Preprocessing prior to data reduction

Another way to reduce the discrepancy between the inherent assumptions of the matrix decomposition and the applied data is to apply preprocessing to the data for improved data reduction performance. For instance, in order for PCA components to be independent, it is required that the variables in $\boldsymbol{X}$ follow a multivariate normal distribution, which is not always the case in biologically sampled data (116). Specifically, metabolomics can generate thousands of features that are log-normally distributed (117) where most of the measurements lie in the

lower range of detection and a few measurements lie in the end higher range of detection. Hence, it is often desirable to log-transform metabolomics data prior to applying PCA for data reduction. Similarly, PCA can be used to reduce other biologically sampled data *e.g.*, microbiota where the presence of bacteria is detected in the sample. Each detection of a strain of bacteria via DNA sequencing yields a count on presence, meaning that microbiota data is discrete and not normally distributed. Various ways to process the data prior to reduction have been investigated but a common procedure is to scale the counts for each bacteria strain by the total number of detected counts in the individual sample. Using this method, a ratio of the detected bacteria in each sample is produced to make individuals comparable. However, typically this data row-wise bounded by unity, which precludes normality, and it is common that some strains of bacteria present in the study population are totally absent from a subset of the samples, leading to counts of zero. To counteract these problems, the log-transformation tends to make the distribution deviate less from normality, and bacteria that are present in too few samples can be omitted from the analysis.

Preprocessing techniques can also be used to weigh the importance of the variables used in the data reduction. For instance, it is important to scale the measurements for each omics feature if these are of equal importance to identify metabotypes. For example, when using metabolomics, metabolites will range in their detected scale, but for the identification of metabotypes, each metabolite carries equal importance if no *a priori* information can serve to rank their relative contribution. A standard method to assure that each feature is of equal importance is to subtract the average of all individual measurements of the feature (centering) and to scale them by their standard deviation into so-called z-scores (standardization or auto-scaling). Furthermore, centering can reduce the number of components used to model the data using a matrix decomposition. However, several other options for scaling are available (118).

4.4 Mechanistic model-based data reduction of dynamic data

To reduce postprandial glucose data and identify differential responders, kinetic modeling of the glucose regulatory system was used (paper IV). In this section, the theory and background regarding mathematical modeling and parameter estimation are described.

4.4.1 Dynamical modeling

When both time-resolved data and *a priori* knowledge of the mechanisms of the metabolic system are available, these two components can be integrated into a mechanistic model, allowing for more biologically interpretable results, compared to purely data-driven approaches. This setting typically occurs when a smaller subset of the metabolic system is modeled and only a few metabolic markers are involved. As touched upon in the introduction, a common framework to model time-dependent processes is dynamical systems which have been used extensively in physics, astronomy, and mechanics (119). Models of dynamical systems can be described in continuous or discrete time and common frameworks to use are differential equations and difference equations, respectively. In this context, differential equations typically describe time derivatives of the model state variables using functions of the state variables themselves, often together with external input signals (99). They can be used to model how biological processes occur in the body or how compounds are absorbed from medications or food. In the context of glucose metabolism, several formulations in terms of systems of differential equations have been used, ranging from 4 parameters in the early

minimal glucose model to 10 parameters, modeling oscillations of glycemia due to the euglycemic hyperinsulinemic clamp, used in insulin sensitivity determination (120–124). One of the simplest models representing blood glucose regulation after consumption of glucose is presented in Equation 3.

$$\dot{H}_i = -l_{1,i}H_i + l_{2,i} + l_{3,i}G_i \tag{3a}$$

$$\dot{G}_i = -l_{4,i}G_i + l_{5,i} - l_{6,i}H_i - I_i \tag{3b}$$

Here, G and H represent glucose and insulin concentrations in the blood respectively, while l_1H denotes the average rate of insulin removal independent of glucose, l_2 the average rate of release of insulin by the pancreas, independent of glucose, l_3G the net increase in the rate of release of insulin due to glucose. Additionally, l_4G represents the average rate of glucose removal independent of insulin, l_5 the average rate of release of glucose into the blood, l_6H the average rate of glucose removal dependent on insulin, and I the rate of decrease of blood glucose due to absorption. Finally, i denotes the individual for which the parameters are specific. Differential equations are in general difficult to solve analytically, making it necessary to use numerical methods to obtain solutions. However, the model in Equation 2 has an analytical solution (Eq. 4) under assumptions and lumping of parameters as described in detail in (120).

$$\hat{G}_i(t) = G_{b,i} + A_i \, sin(\omega_i t) \, e^{-\alpha_i t} \tag{4}$$

Here, insulin has been expressed in terms of other mechanisms and a formulation using only glucose ($\hat{G}$) is obtained without dependence on derivatives. Instead, other interpretable parameters have been included where G_b denotes the glucose baseline level, A the sinusoidal amplitude involved in the resulting amplitude of the glucose concentration, ω the sinusoidal frequency relating to the velocity of glucose oscillations, and α damping coefficient determining the rate of glucose decay. When parametrizing the data, the model represents a reduced form of the measured data given that the number of measured samples is larger than the number of estimated parameters. Thus, the model parameters can be used for clustering to identify differentially responding individuals with the benefit that the parameters hopefully provide a meaningful description of physiological processes and a better basis for clustering compared to raw glucose time trajectories.

4.4.2 Parameter estimation

Estimation of the parameters given glucose data can be done using several techniques. In optimization theory, the loss function is a construct describing how similar the model output is to the data. A commonly used function is the mean squared error (MSE) described in Eq. 5.

$$\mathrm{MSE}_i = \frac{1}{T}\sum_{t=1}^{T} \left(\boldsymbol{G}_{i,t} - \hat{\boldsymbol{G}}_{i,t}(\boldsymbol{\varphi}_i)\right)^2 \tag{5}$$

Here, $\boldsymbol{G}_{i,t}$ denotes the collection of T measured glucose samples in time, $\hat{\boldsymbol{G}}_{i,t}$ the model-predicted glucose, i the individual sample at time point t. Since $\hat{\boldsymbol{G}}_{i,t}$ depends on the parameters $\boldsymbol{\varphi}_i = (G_{b,i}, A_i, \alpha_i, \omega_i)$, changing them will change the shape of the model prediction and thus its fit to the data. The objective of optimization algorithms is to find the set of parameters that achieves the best possible fit, the optimal parameter set. There are different methods to

approach optimization in continuous space including using the derivative of the function to find a minimum and using intelligent searches of the fitness landscape without using derivatives. However, the loss function, which in this case is a multidimensional surface, may have many optima, and the algorithm is not guaranteed to find the global optimum. Therefore, it is recommended to use as few parameters as possible to reduce the search space for the optimization algorithm. The reduction of the parameter space can be achieved when data on several individuals of a population is available. It is then reasonable to assume that some systematic variance in the data is shared among the population. The mixed effects modeling framework assumes that some features of the estimated parameters are shared and that some are individual, hence allowing a parameter estimation algorithm to make more efficient use of the data by explicitly acknowledging the similarities and differences between individuals within the study population (125). The framework has been used extensively in pharmacology to estimate differences in drug response between groups of individuals but is equally well suited to model other metabolic processes relating to food intake (126). Here, the parameters are described probabilistically according to Equation 6 where the individual parameters $\boldsymbol{\varphi}_i$ are described by the fixed effects (shared among all individuals) $\boldsymbol{\beta}$, and the random effects $\boldsymbol{\eta}_i \sim \mathcal{N}(0, \boldsymbol{\Psi})$.

$$\boldsymbol{\varphi}_i = \boldsymbol{\beta}\, e^{\boldsymbol{\eta}_i} \tag{6}$$

For the purposes of the study in paper IV, the random effects were modeled using a Gaussian mixture, which allowed simultaneous clustering and parameter estimation, removing the need for post hoc clustering methods. However, this more detailed description of the study population provided by the mixed effects framework comes at a cost, requiring the use of more sophisticated parameter estimation algorithms, able to deal with the hierarchical structure of the parameters and the probabilistic description of the random effects. A number of methods have been developed specifically for this task, including FOCE, SAEM, and Markov Chain Monte Carlo (127–129). Crucially, these tools commonly provide diagnostics of model fit and parameter uncertainty, allowing adaptation of the complexity of the mixed effects formulation to the estimation problem and data at hand (125).

4.5 Data reduction of high-dimensional time-resolved omics data

To identify metabotypes in dynamic postprandial data stemming from dietary crossover intervention trials, tensor decomposition, and dynamical modeling was used (papers II and III). In this section, the theory and background of these methods are explained.

4.5.1 Descriptive methods

Dynamic omics data refers to repeated measurements of the same omics features at sequential points in time. When attempting to reduce this data to identify metabotypes, at least three dimensions are available (individuals, omics features, and time points). As discussed in Section 4.3.1, dynamic omics data could be reduced using matrix decompositions such as PCA by concatenating the time points per omics feature and thus mixing time and omics features in the columns of the matrix (97). The problem with this approach is that each new time point of the same omics feature is interpreted as a new variable in the matrix decomposition, thus breaking the causal link between time and omics features, which makes the interpretation of dynamics

more challenging. To more directly address this issue, tensor decompositions can be used. These methods can be viewed as higher-dimensional extensions to matrix decomposition methods, and are applicable directly to data stored in tensor form (*e.g.*, individuals, omics features, and time points) (130). This approach has the advantage of keeping the relation between time and omics features intact. However, tensor decompositions are still not as commonly used as matrix decompositions and some of the mathematical properties that are given for matrices do not extend to tensors. For example, tensor rank can so far not be determined efficiently by any algorithm and instead, a series of heuristic tools are available depending on the tensor decomposition that is used.

Arguably the most similar tensor decomposition compared to PCA and the most interpretable is CANDECOMP/PARAFAC (CP), which is described in Equation 7 and note the similarity to Equation 2 (97).

$$\boldsymbol{\mathcal{X}} = \sum_{f=1}^{F} \boldsymbol{a}_f \circ \boldsymbol{b}_f \circ \boldsymbol{c}_f \tag{7}$$

Here, the third-order tensor $\boldsymbol{\mathcal{X}}$ **(**with dimensions representing individuals, omics features, and time points) is decomposed using a sum of F components, consisting of factors ($\boldsymbol{a}_f$, $\boldsymbol{b}_f$, and $\boldsymbol{c}_f$) that are multiplied by the outer product. Often, the aim is for F to be equal to the tensor rank or lower in order to reduce dimensionality. In CP, the components are not restricted to be orthogonal (131), but constraints like orthogonality, non-negativity, and regularization can be imposed (132). However, unlike PCA, CP is not guaranteed to fit perfectly to a tensor, which can complicate data reduction. In addition, solutions to CP frequently degenerate, meaning that factors between components correlate strongly, and thus do not independently explain variation of the data. Other tensor decomposition methods can also be used for data reduction without the disadvantage of degeneracy, *e.g.*, the Tucker decomposition. This decomposition fits a larger range of data but is less interpretable since factors can be modeled using different numbers of components (133).

Similar to matrix decomposition, preprocessing of tensors prior to decomposition will affect the results. For tensors, the preprocessing step is not as straightforward as in the case of matrices and their decompositions (134). In the matrix case, centering can remove the need for more components to model the data and this is the case for tensor as well. However, when centering across the individual dimension (analogous to removing the average of features in the matrix case) in a third-order tensor consisting of individuals, time, and features, an average of all the individual dynamics for each feature is removed. Hence, centering across one dimension disturbs the scaling within all other dimensions but most importantly, the dynamics of the resulting data changes when centering. When centering across the individual dimension, the dynamics of the average individual are removed, and the resulting tensor describes the individual deviation from the average dynamics. Theoretically, this reduces the number of needed components when using CP to decompose the tensor while no artifacts are induced to the CP factors. Thus, analysis can be focused on the difference between individuals (since the variance representing the average individual is removed from the data), instead of the general dynamics of the data. However, the data used in practical cases does not always follow the mathematical structure in Eq.7, which leads to the estimation of CP factors not representing the original dynamics. Instead, artifacts of dynamical patterns are found in the CP factors

hampering the analysis of the factors representing the time dimension. Therefore, it is not yet clear how to preprocess dynamical tensor data when attempting to identify differential responders.

4.5.2 Predictive methods

Although tensor decomposition methods like CP preserve the relation between time points and omics features, they do not take the causal link between different time points into account, making them better suited for description than prediction. If time dependency is instead inherently built into the model, then predictive capabilities can be obtained. A common way to model these data is to use differential equations as described in Section 4.4.

Systems of differential equations are state space models, where terms of interest, *e.g.*, rate of change of a metabolic marker, can be denoted as a state. In the discrete case and where all state space equations are linear, one can write the system in terms of matrices and use the properties of linear algebra for analysis (135). These so-called linear dynamical systems (LDS) have been used extensively in control theory because of their mathematical properties that allow for interpretable analysis (135,136). Here we focus on the discrete setting since it underlies the development of dynamic mode decomposition (described in detail further on). We start by describing an LDS for the modeling of dynamic metabolomics data, using only a system matrix $\boldsymbol{A}$ (representing the dynamics of the system) (Eq. 8).

$$\boldsymbol{x_{t+1}} = \boldsymbol{A}\,\boldsymbol{x_t} \tag{8}$$

Here, $\boldsymbol{x_t}$ represents a vector of M metabolite measurements at time t and $\boldsymbol{A} \in \mathbb{R}^{M\times M}$ is the linear operator that evolves the measurements $\boldsymbol{x_t}$ one time step forward in time to $\boldsymbol{x_{t+1}}$. Analysis of the system is facilitated by the unique properties of LDSs. For instance, the dynamics of the modeled dynamical system can be determined by the eigenvalues $\boldsymbol{\lambda}_i \in \mathbb{C}$ of $\boldsymbol{A}$ (112), where eigenvalues are defined as the scaling of a corresponding eigenvector $\boldsymbol{v} \in \mathbb{C}^M$ according to Eq. 9.

$$\boldsymbol{Av} = \boldsymbol{\lambda v} \tag{9}$$

The collection of eigenvalues describes the dynamic properties of $\boldsymbol{A}$ via their placement in the complex plane. In the discrete case $\boldsymbol{A} \in \mathbb{R}^{M\times M}$, the trajectory $\boldsymbol{x_t}$ can have four distinct dynamic behaviours; constant ($|\lambda| = 1$), exponentially increasing ($|\lambda| > 1$), exponentially decreasing or oscillating while decreasing ($|\lambda| < 1$). While these systems are preferably modeled based on a mechanistic understanding, this is often not possible for metabolic systems since they typically consist of a large number of states (metabolites) and comprise interactions and mechanisms that are not fully understood. An alternative approach is to model the metabolic system in a data-driven manner. Dynamic mode decomposition (DMD) has recently emerged as a promising tool for this task(137). In the context of metabolic modeling, the objective of DMD is to estimate the parameters of $\boldsymbol{A}$ using all measured metabolite data (Eq. 10).

$$\boldsymbol{X} := \begin{bmatrix} | & | & & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_{T-1} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{M\times(T-1)} \tag{10a}$$

$$\boldsymbol{X}' := \begin{bmatrix} | & | & & | \\ \boldsymbol{x}_2 & \boldsymbol{x}_3 & \dots & \boldsymbol{x}_{T} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{M\times(T-1)} \tag{10b}$$

$$\boldsymbol{X}' = \boldsymbol{A}\,\boldsymbol{X} \tag{10c}$$

Here $\boldsymbol{x}_t$ represents a vector with *M* metabolite measurements at time *t* and $\boldsymbol{A}$ the linear operator evolving the measurements $\boldsymbol{X}$ one time step forward to $\boldsymbol{X}'$ with T equidistant samples in time. The linear operator is estimated using the least squares solution in Eq. 11.

$$\boldsymbol{X}'\boldsymbol{X}^{\dagger} = \boldsymbol{A} \tag{11}$$

Here, † represents the Moore-Penrose inverse which can be calculated using the SVD of $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\intercal}$ as

$$\boldsymbol{X}'\boldsymbol{V}\boldsymbol{\Sigma}^{\dagger}\boldsymbol{U}^{\intercal} = \boldsymbol{A}. \tag{12}$$

As described in Section 4.3.1, SVD decomposes the data into components of decreasing amounts of variance-accounted-for. This allows it to be used to remove components that likely describe noise and keep only the most informative systematic information in the data, *i.e.*, to truncate the SVD of $\boldsymbol{X} \approx \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}}^{\intercal}$, where $\widetilde{\boldsymbol{U}} \in \mathbb{R}^{M\times S}$, $\widetilde{\boldsymbol{V}} \in \mathbb{R}^{(T-1)\times S}$, $\widetilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{S\times S}$ and S is the number of components of the truncated decomposition (Eq. 13).

$$\boldsymbol{X}'\widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{\Sigma}}^{\dagger}\widetilde{\boldsymbol{U}}^{\intercal} \approx \boldsymbol{A} \tag{13}$$

Additionally, we can project the system matrix onto the subspace of $\widetilde{\boldsymbol{U}}$, to effectively shrink the matrix $\boldsymbol{A}$ to a latent space, describing the essential dynamics of the LDS, to simplify the analysis of the system using *e.g.*, eigenvalue analysis (Eq. 14).

$$\widetilde{\boldsymbol{U}}^{\intercal}\boldsymbol{X}'\widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{\Sigma}}^{\dagger}\widetilde{\boldsymbol{U}}^{\intercal}\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{U}}^{\intercal}\boldsymbol{X}'\widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{\Sigma}}^{\dagger} = \widetilde{\boldsymbol{A}} \tag{14}$$

We then have a reduced LDS describing latent dynamics $\widetilde{\boldsymbol{x}}_{\boldsymbol{t}} = \widetilde{\boldsymbol{U}}^{\intercal}\boldsymbol{x}_t$ (Eq.15).

$$\widetilde{\boldsymbol{x}}_{\boldsymbol{t+1}} = \widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{x}}_{\boldsymbol{t}} \tag{15a}$$
$$\boldsymbol{x}_t = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{x}}_{\boldsymbol{t}} \tag{15b}$$

Standard DMD learns the dynamics of a particular system after one excitation (here corresponding to dietary intake) and does not inherently consider other excitations of the same system. However, an extension of DMD called parametric DMD (pDMD) (138) utilizes a different parameterization of the system to learn the dynamics that govern the different responses, *e.g.*, responses to different dietary intakes. However, DMD and pDMD will not distinguish between the dietary provocation (impulse) and the metabolic regulation underpinning the response to the provocation. The recent development of DMD with control (DMDc) (139) elegantly addresses this by estimating an LDS with input taken into account (Eq.16).

$$\widetilde{\boldsymbol{x}}_{\boldsymbol{t+1}} = \widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{x}}_{\boldsymbol{t}} + \widetilde{\boldsymbol{B}}\boldsymbol{z}_{\boldsymbol{t}} \tag{16}$$

In Eq 11, the input vector is denoted $\boldsymbol{z}$ and DMDc is used to estimate $\widetilde{\boldsymbol{A}}$ and an additional matrix $\widetilde{\boldsymbol{B}}$ mapping from the input space to the metabolite space that $\widetilde{\boldsymbol{x}}_{\boldsymbol{t}}$ exists in. The combination of pDMD and DMDc allows for learning of responses from multiple diets and thus also prediction of response to new diets (paper III). Additionally, the estimated latent states $\widetilde{\boldsymbol{x}}_{\boldsymbol{t}}$ give summary information of individual responses to food in a large number of metabolites, allowing the reduced LDS to be used to identify differential responders.

5 RESULTS AND DISCUSSION

5.1 Identification of metabotypes in multi-omics data

A workflow for the identification of metabotypes related to cardiometabolic health (targeting obesity, dyslipidemia, hyperglycemia, insulin resistance, hypertension, and low-grade inflammation) from static biological markers in a free-living Danish population was developed in paper I.

The anthropometric-, biochemical-, and multi-omics data (microbiota and metabolites from plasma and urine) were collected and processed from the DCH-NG MAX population (described in detail in Section 3.1.3). Data filtering was performed with the aim of selecting features that were related to cardiometabolic clinical markers. To achieve this, the double cross-validation framework MUVR (140) was used for regression of clinical markers from omics datasets. MUVR provides a recursive variable selection when training the model, optimizing the trade-off between parsimony and prediction performance. In this way, MUVR suggests models using varying numbers of features depending on whether parsimony or performance is valued. The most parsimonious models were chosen (*i.e.*, the ones utilizing the least number of features while maintaining performance) and the features selected by the model were used in the following clustering to identify metabotypes. Post filtering, 56, 100, 86, and 174 selected variables from microbiota, targeted plasma metabolomics, targeted urine metabolomics, and untargeted plasma metabolomics, were selected, respectively.

The selected omics features were analyzed using multiple factor analysis (MFA) which involved normalizing each dataset and concatenating them to a large matrix prior to performing a PCA. The scores of the MFA were clustered using Gaussian mixture models. Three clusters were chosen since they could be externally validated using the available data and since loading directions determined by the MFA aligned well with the proposed clusters (Fig. 4). The clusters (Fig. 4A) were predominantly separated by the microbiota, targeted plasma metabolome, and the untargeted plasma metabolome (Fig. 4B).

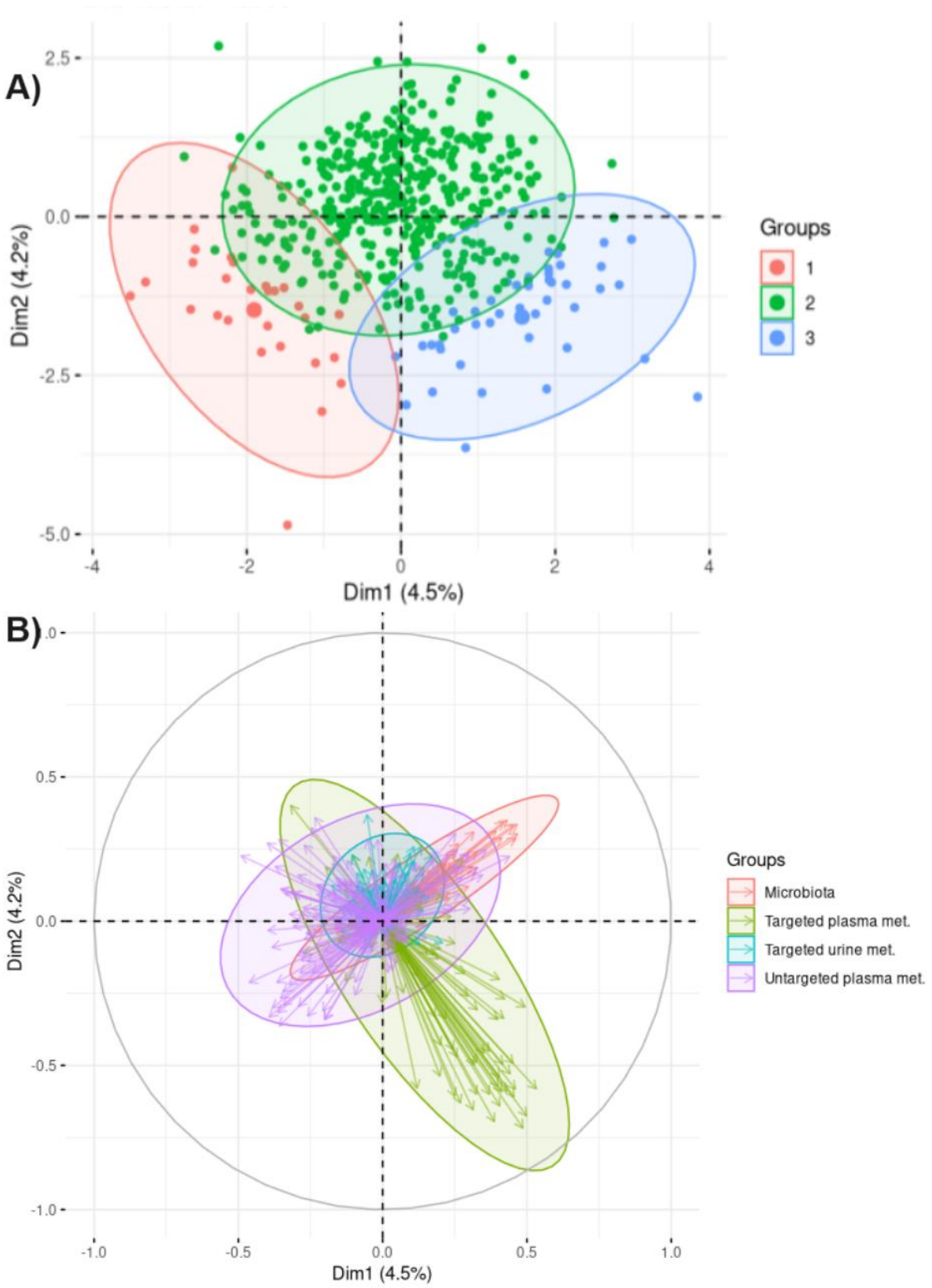

Figure 4: Visualization of MFA using multi-omics. A) MFA scores color-coded by clusters found using Gaussian mixture models. B) MFA loading plot color-coded by omics categories.

When visually inspecting the cardiometabolic profile of the clusters, cluster 1 represents an unhealthier profile, having higher levels of risk factors like BMI, waist circumference,

triglycerides, cholesterol, *etc*. (Fig. 5). This is also reflected in that the same cluster had the highest prevalence of metabolic syndrome (82%), while the other cluster had the least prevalence (16% and 9%, respectively). Here we used the definition and cut-off values described by Alberti et al. to classify metabolic syndrome (141). Moreover, cluster 2 had a poorer cardiometabolic profile compared to cluster 3 which represented the healthiest individuals in the population.

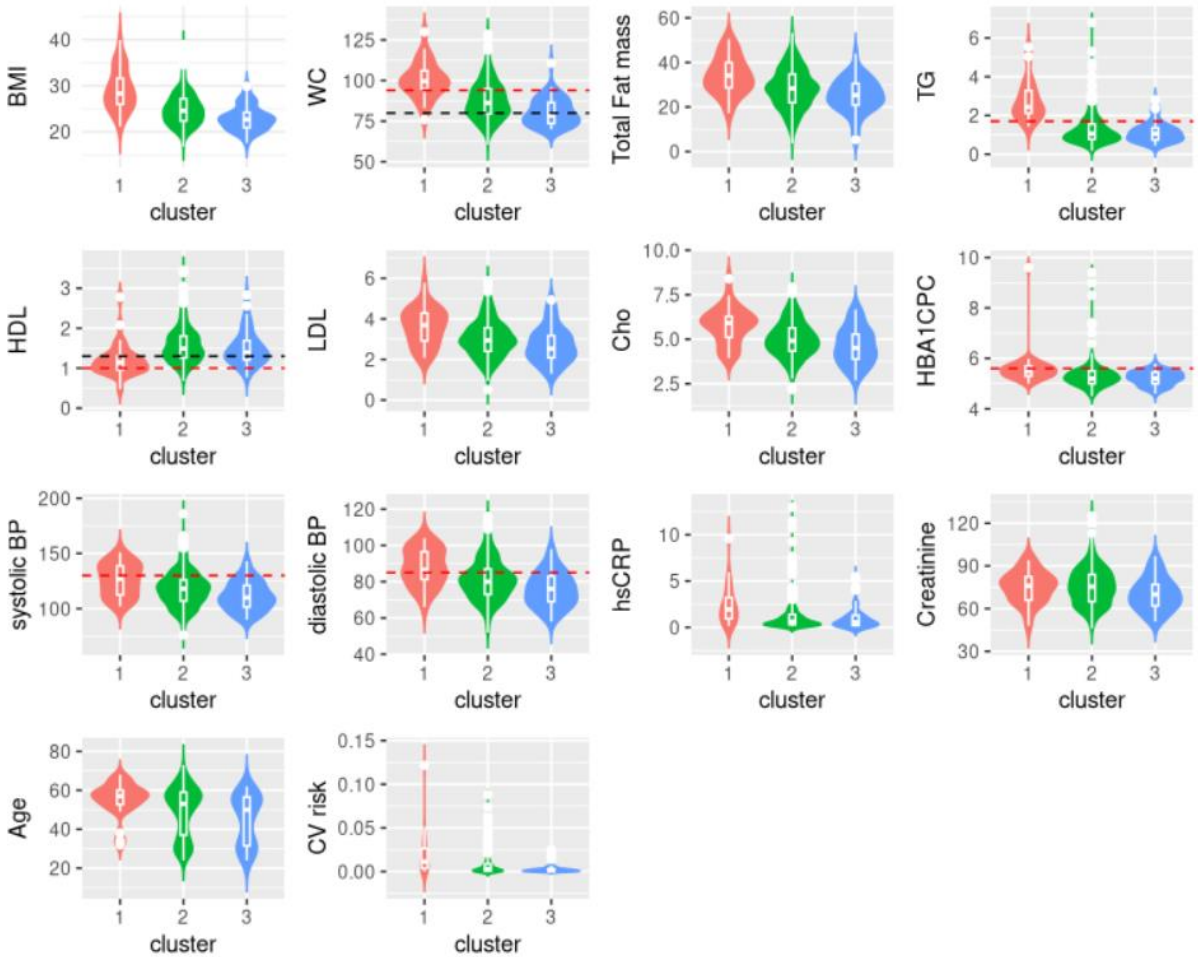

Figure 5: Cluster distribution in clinical cardiometabolic risk factors, age, creatinine, and cardiovascular risk. Dashed lines represent cut-offs for metabolic syndrome (141) in the measures that are considered in the definition (red: cut-offs for men when the definition is gender-specific and for both sexes when nongender-specific, black: cut-offs for women when the definition is gender-specific).

Analysis of the most contributing loadings (Fig. 6) showed that cluster 1 had the lowest abundance of several strains of the *Ruminococcaceae* family which are generally found to be plant polysaccharide degrading bacteria that have been positively associated with gut health and reduced risk of cardiovascular disease (142,143). Furthermore, highly contributing loadings also showed lowered levels of some unidentified metabolites derived from untargeted plasma metabolomics and a lowered abundance of *Christensenellaceae* which has been inversely related to BMI (144). Loadings also showed that cluster 2 had higher levels of hexadecanoyl-L-carnitine (palmitoylcarnitine) together with higher levels of 3 unidentified metabolites derived from untargeted metabolomics in plasma. Lastly, cluster 3 had lower levels of palmitoylcarnitine, but higher levels of some amino acids such as serine, proline, and of tryptophan derived quinolinic acid, indolepropionic acid, and indoleacetic acid (Fig. 6).

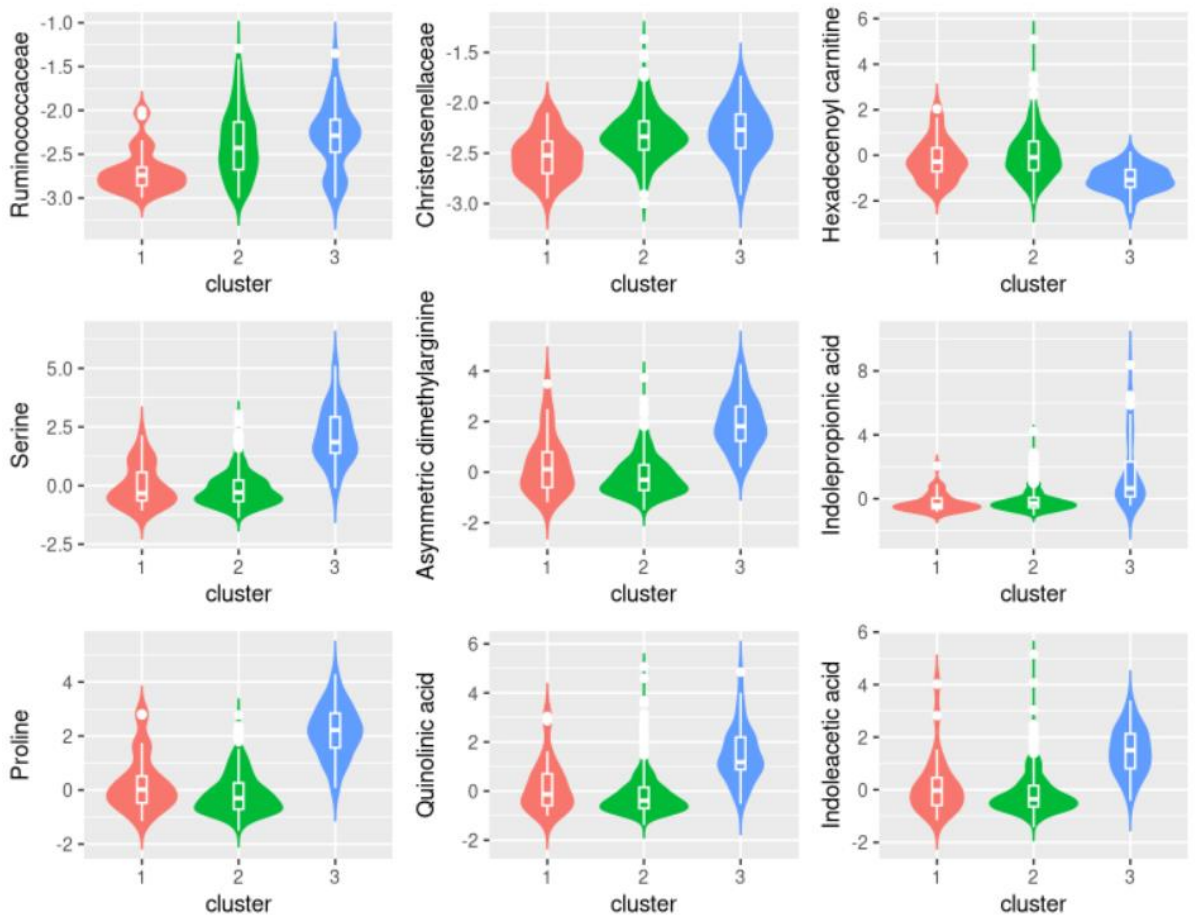

Figure 6: Distribution of clusters in strongly contributing MFA loadings i.e., omics features.

To assess how dietary intake was associated with the different clusters, 65 food subgroups (coffee, milk substitutes, cabbages, *etc*.) consumed over two days were investigated in relation to the clusters. When analyzing the correlation between MFA components and self-reported food intakes, it was clear that cluster 1 was associated with consumption of unhealthy food items such as processed meat, spirits and brandy, soft drinks, beer, and butter (third quadrant in Fig. 7). It is notable that cluster 1 which was low in plant-based foods was also low in the abundance of the family *Ruminococcaceae* (142). In contrast, cluster 2 showed the highest abundance of this family of gut microbiota and also had the highest intake of leafy vegetables, nuts, seeds, and fruits among the three clusters (Fig. 7).

Cluster 2 was characterized by two groups of food items; one that was high in fish (first MFA quadrant in Fig. 7) and one that was high in fruits and nuts (second quadrant in Fig. 7). This may indicate that cluster 2 could be further subdivided into two clusters for better tailoring of food, which is further supported by the large number of outliers in cardiometabolic markers and loadings (Figs. 5&6). However, HDL levels were the highest among all clusters while many individuals also had high levels of the inflammation marker hsCRP, which is contradictory due to the anti-association between inflammation and HDL often observed (145). Noticeably, the second cluster included individuals with the highest levels of HbA1c, which is a risk factor for type 2 diabetes (146) and thus, this group may benefit from lower intakes of high glycemic foods and adhere to a healthy diet plan such as the healthy Mediterranean diet which has been shown to lower HbA1c (147).

The third cluster was undoubtedly the healthiest with the best cardiometabolic risk profile. This cluster also had the highest levels of indole propionic acid which is an anti-inflammatory produced by the gut microbiota (148). Seemingly, this cluster also had a healthy dietary pattern with correlations to food items such as whole grain cereals, lean dairy products, eggs, and coffee.

Diet was further investigated in relation to the clusters using *a priori* defined dietary indices which were calculated from repeated 24h recalls. Cluster 1 adhered the least to the healthy plant-based diet Index, healthy Nordic diet score, and the most to the unhealthy plant-based diet Index.

In summary, the results suggest that three clusters could be derived from metabolomics and gut microbiome data that represented different cardiometabolic risk profiles and that such clusters were plausibly associated with different dietary intakes at the food level as well as to adherence to *a priori* defined food indices.

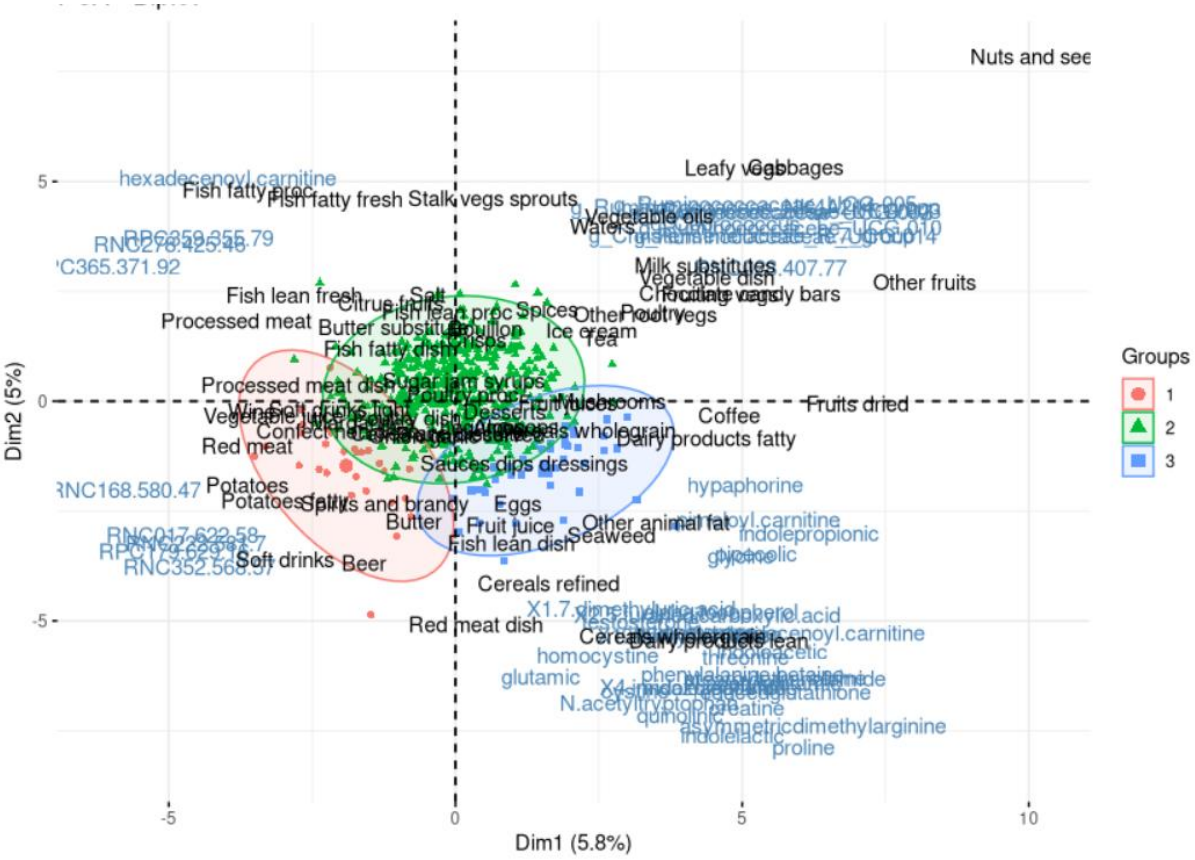

Figure 7: MFA biplot with scores color-coded by clusters and loadings denoted by omics features. Consumption of food items are added as correlations to MFA components to visualize dietary patterns in relation to clusters.

5.2 Identification of metabotypes in high-dimensional dynamic postprandial metabolomics data

Two distinct methods for identifying metabotypes in high-dimensional omics data stemming from crossover interventions with repeated measures were developed (papers II and III).

5.2.1 Using tensor decomposition – a descriptive method

Firstly, when the same metabolites have been measured repeatedly at different time points and for different intervention diets, the data can be viewed as a tensor (paper II). The dataset used in the studies (papers II & III) was obtained from the dietary crossover intervention described in Section 3.1.1. Additionally, a synthetic dataset was generated with two metabolic dynamic patterns and two *a priori*-defined induced metabotypes. The data was generated from factors using the mathematical description of CP (Eq. 7) to evaluate the recovering factors using CP on different preprocessing methods, thus inspecting the performance of each method. Three preprocessing methods were applied to measured and synthetic data to investigate the effect on the interpretation of dynamics and identification of metabotypes. The preprocessing methods denoted P1, P2, and P3 were described as scaling all metabolite data by its standard deviation (P1), centering around the average dynamics per metabolite prior to scaling as in P1 (P2), and centering around global average value per diet prior to scaling as in P1 (P3).

Results on simulated data showed that P2 and P3 recovered the induced metabotypes and dynamics properly while P1 was less successful. Figure 8 shows the recovery of factors that determined the synthetic data using different preprocessing methods.

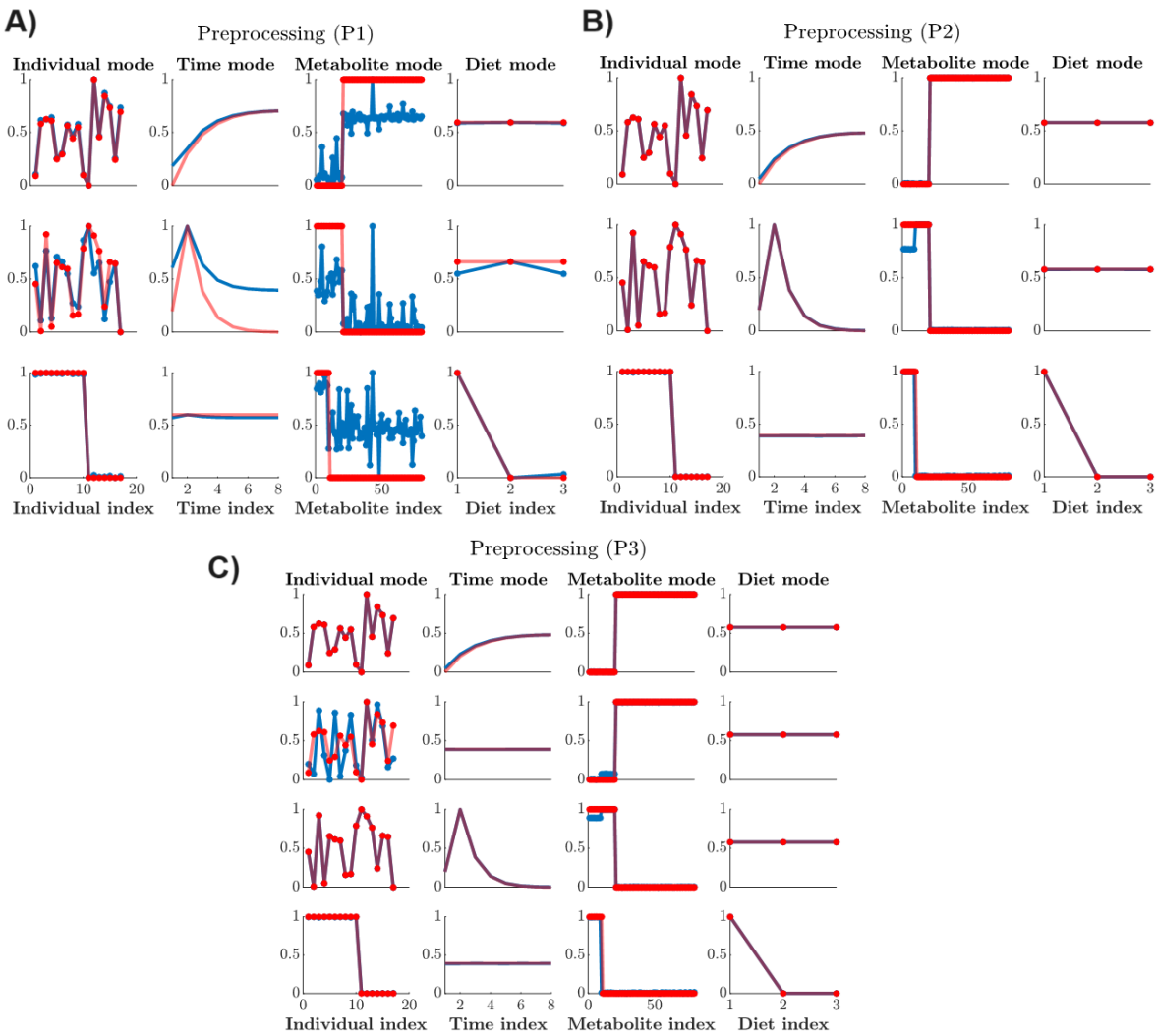

Figure 8: CP applied to synthetic data with three different preprocessing methods P1, P2, and P3 in subfigures A, B, and C, respectively. The blue lines represent the ground truth factors that make up the synthetic data and the red ones represent the estimated factors using CP. The columns represent the modes (individuals, time, metabolites, and diet) and the rows the CP components

A fourth component was necessary to capture all the dynamics when using P3 since that preprocessing did not reduce the effective rank needed to model the data (discussed in Sec. 3.5.1). When applying these preprocessing methods to measured data, it was observed that when using P1 and P3, metabolite dynamics were reflected in the factors but to a lesser degree in P2 (Fig. 9). This indicated that the data cannot be described fully using a multi-linear structure as in Eq. 7.

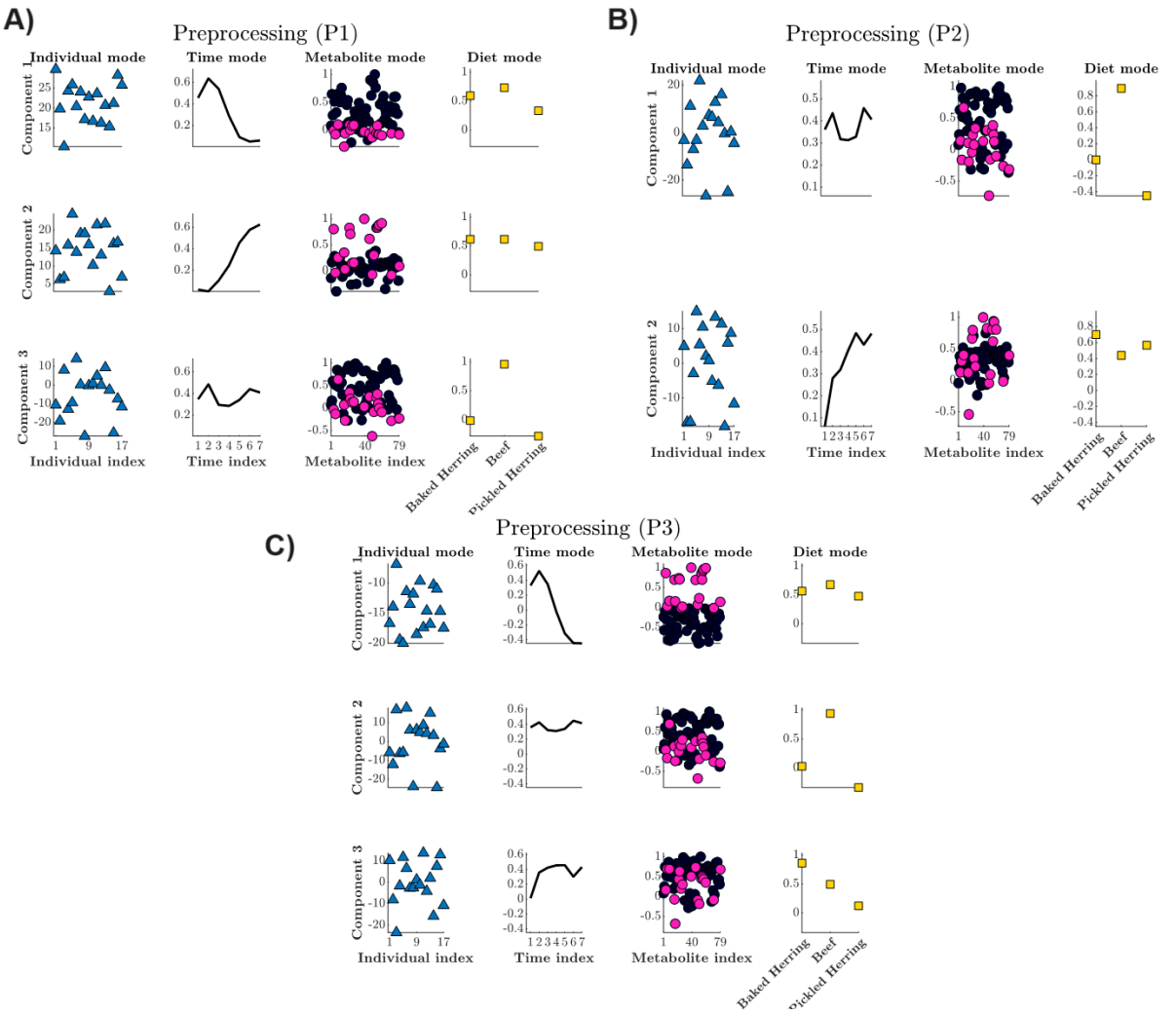

Figure 9: CP applied to the data from the dietary intervention trial described in three Section 3.1.1, using different preprocessing methods P1, P2, and P3 in subfigures A, B, and C, respectively. The columns represent the modes (individuals, time, metabolites, and diet) and the rows of the CP components. The black dots represent the fast dynamic metabolites such as sugars and amino acids, while the pink dots represent the slower dynamic metabolites such as lipids.

Clusters of scores (identified using k-means) representing metabotypes identified using P2 were associated with baseline levels of creatinine and the metabolites contributing the most to the clusters (metabotypes) were amino acids, which have creatinine as a precursor in their production (Fig. 10).

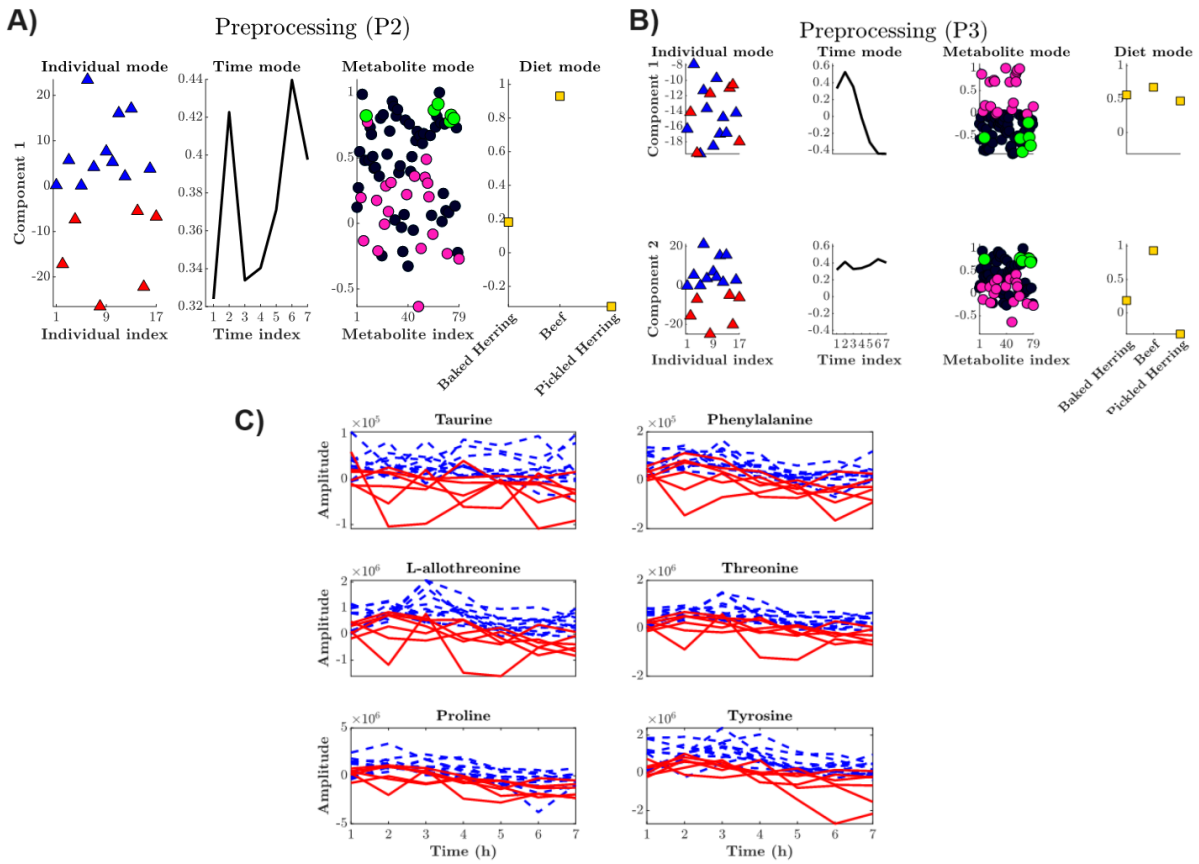

Figure 10: A) 1-component model using preprocessing method (P2) on measured data. B) 2-component model using preprocessing method (P3) on measured data. The pink dots represent the "slow" dynamic metabolites, the black and green dots represent "fast" dynamic metabolites and the green dots represent amino acids. Clusters of scores marked as red and blue triangles were associated with baseline creatinine. C) Amino acids time series color-coded by clustering indices from clustering of scores (red and blue) (no preprocessing method applied).

These results show that the CP could be used for reducing dynamic omics tensor data stemming from crossover or similar studies to identify metabotypes, and that the preprocessing methods could all be applied to the data for different purposes (paper II). P1 extracted more interpretable dynamic patterns from the data while P2 extracted differences between individuals but with less interpretable dynamics. P3 acted as a compromise between P1 and P2, showing individual differences, but also clearer dynamics than P2 but not theoretically reducing the number of CP components for modeling. The results highlighted that CP is a valid tool for the identification of metabotypes in time-resolved omics data and that several choices of preprocessing methods are possible, each with its advantages and disadvantages. Ideally, the choice should ultimately be dictated by the aim of the analysis. Finally, CP gives a comprehensive overview of the data from complex study designs in terms of the estimated factors that can be used for clustering and interpretation of dynamical patterns.

5.2.2 Using pDMDc, a predictive method

The same measured dataset from the crossover intervention trial was also used as a testing ground for the development of a method for metabotype identification using DMD and to predict individual postprandial metabolite responses (paper III). As described in Sec 3.5.2, two separate developments of DMD were combined to accommodate multiple dietary challenges (pDMD) and to separate dietary intake from metabolic response (DMDc), resulting in a new formulation denoted pDMDc. In order to facilitate the evaluation of the new method against a

known ground truth, a synthetic dataset from a virtual metabolic human dynamic model (149) was utilized. The dataset consisted of 50 healthy and 50 diabetic individuals who were generated as postprandial responses to 3 meal interventions.

Results on measured data showed that the same metabotypes as previously identified (Fig. 10) could be identified in the measured data in the first latent state induced by the meat diet (Fig. 11A). The metabotypes were also compared to the clustering using CP (Fig. 11B) and color-coded in raw data of the most contributing metabolites (Fig. 11C).

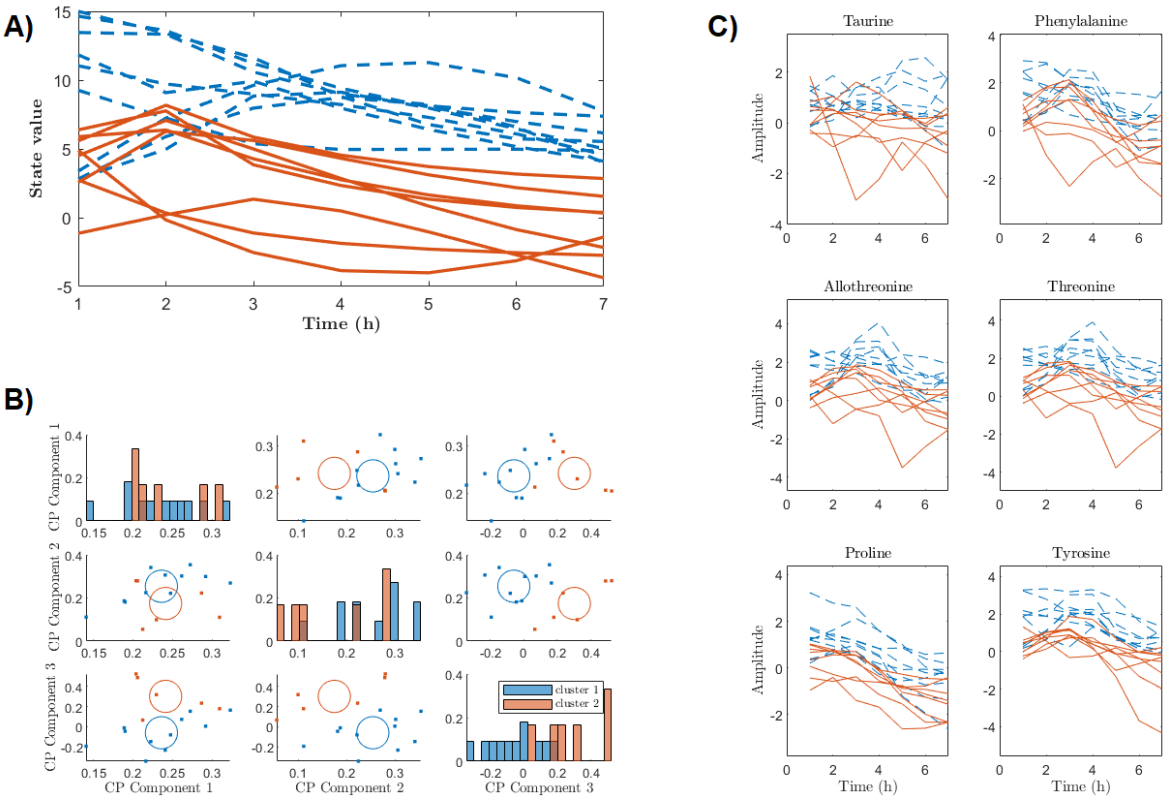

Figure 11: Identification of metabotypes in measured data. A) Clustering of individuals via the first pDMDc latent dynamic states. B) Clustering of CP scores. C) Clusters (blue and red) color-coded in contributing raw data metabolites.

Healthy individuals or individuals with type 2 diabetes could also be distinguished using the same method (Fig. 12). However, although the ground truth clusters were identified to a large extent in the second, third, and fourth latent states, the separation was clearest in the fourth state, indicating that pDMDc might not be as effective for finding clusters as CP.

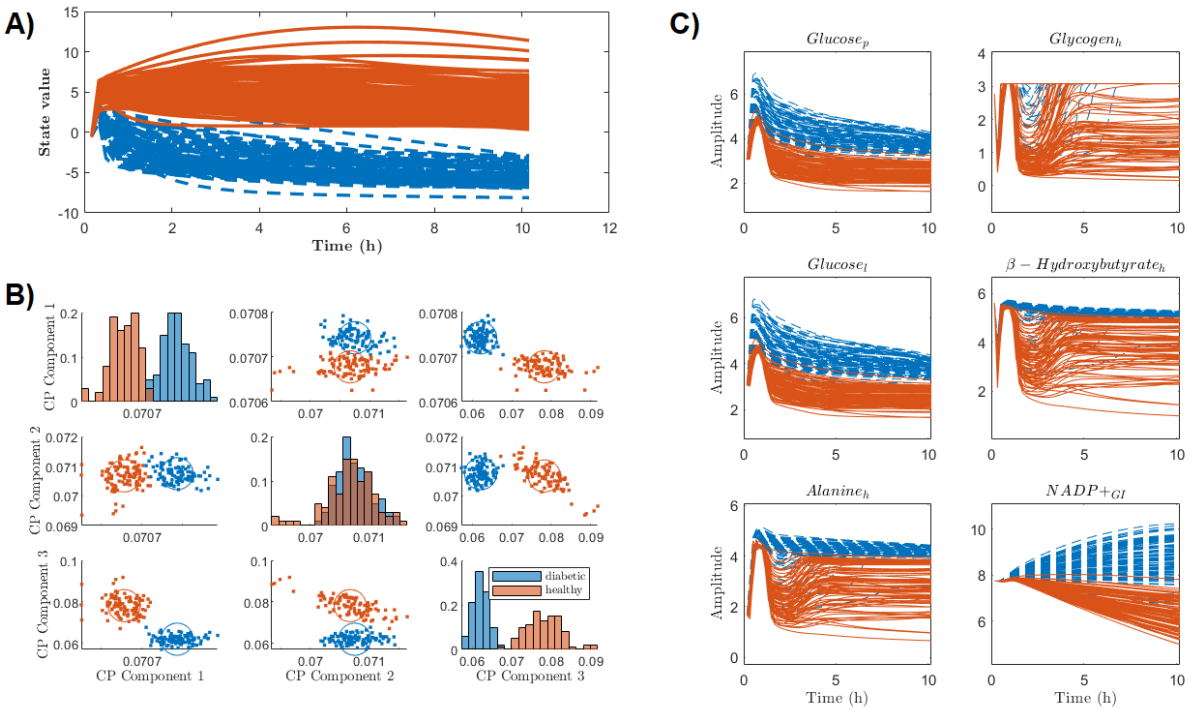

Figure 12: Identification of metabotypes in simulated data. A) Clustering of individuals via the fourth pDMDc latent dynamic states. B) Clustering of CP scores. C) Clusters (blue and red) color-coded in contributing raw data metabolites.

When comparing CP to pDMDc for identification of metabotypes, CP captured a summary of the data more clearly as seen in Fig. 10. However, for pDMDc and LDSs no intuitive way of demonstrating such a summary exists to the best of our knowledge. Secondly, for CP the clusters were found in the first and third components when applied to the simulated data while in pDMDc the clusters were distinctly found in the fourth latent state. On the other hand, pDMDc provided a better fit to the data, since it does not suffer from the problem with degenerating components that can affect CP performance.

Clustering was performed on state and diet level, meaning clustering of single states and diets simultaneously, contrary to CP where multiple components (analogous to states in pDMDc) were clustered. Attempts to try to cluster individual state matrices $\widetilde{\boldsymbol{A}}$ were conducted but the *a priori* metabotypes could not be found using this strategy. This is an interesting approach since the $\widetilde{\boldsymbol{A}}$ matrix for each individual gives a general description of its metabolic dynamic behavior. This suggests that using $\widetilde{\boldsymbol{A}}$ matrices for clustering could be used to discern groups with different metabolic regulations. We further attempted to restrict the dynamics by imposing the same eigenvector basis on all individuals hence making only the eigenvalues inter-individually different. This attempt also failed when trying to identify the given clusters. Hence, so far CP shows a better overview of the data and can be used to cluster general differences among individuals, but may have problems with fitting to the data. pDMDc on the other hand did not give the same type of descriptive overview of the data, but clustering could be more targeted on states and dietary exposures and had fewer problems fitting the data.

5.3 Prediction of high-dimensional postprandial response to food over time

Besides the identification of metabotypes, represented as differential plasma metabolite responders to food, perhaps the main advantage of pDMDc lies in being able to predict postprandial metabolic response to diets (paper III). The model could learn the relationship between dietary information like intake of specific macronutrients, baseline metabolite profile, and postprandial metabolite response over time. Since the measured data (described in Section 4.1.2) only consisted of three dietary challenges, a pooling of the individuals was made to let the model learn from as many postprandial dynamic responses as possible. Ideally, each individual would have been challenged with a large number of diets to measure their postprandial response. By pooling the data, the model learned postprandial responses from all individuals to predict responses to meals in random individuals due to the lack of several dietary challenges per person. The model learned from 60% (training set) of the responses and used 20% (validation set) to determine an appropriate number of model states while avoiding overfitting to the data, and 20% (test set) to test the final predictions. The model was evaluated using a cross-validation procedure where data for each of the training, validation, and test sets were resampled without replacement 100 times, and an average prediction measure (R^2) was used to indicate the true predictive power. Results showed that response to new diets could be predicted with $R^2 = 0.4$ which means that 40% of the variance of the response could be accurately predicted. An example of the predictions of six selected metabolites can be seen in Fig. 13a and the relation between all predictions and data in Fig. 13b.

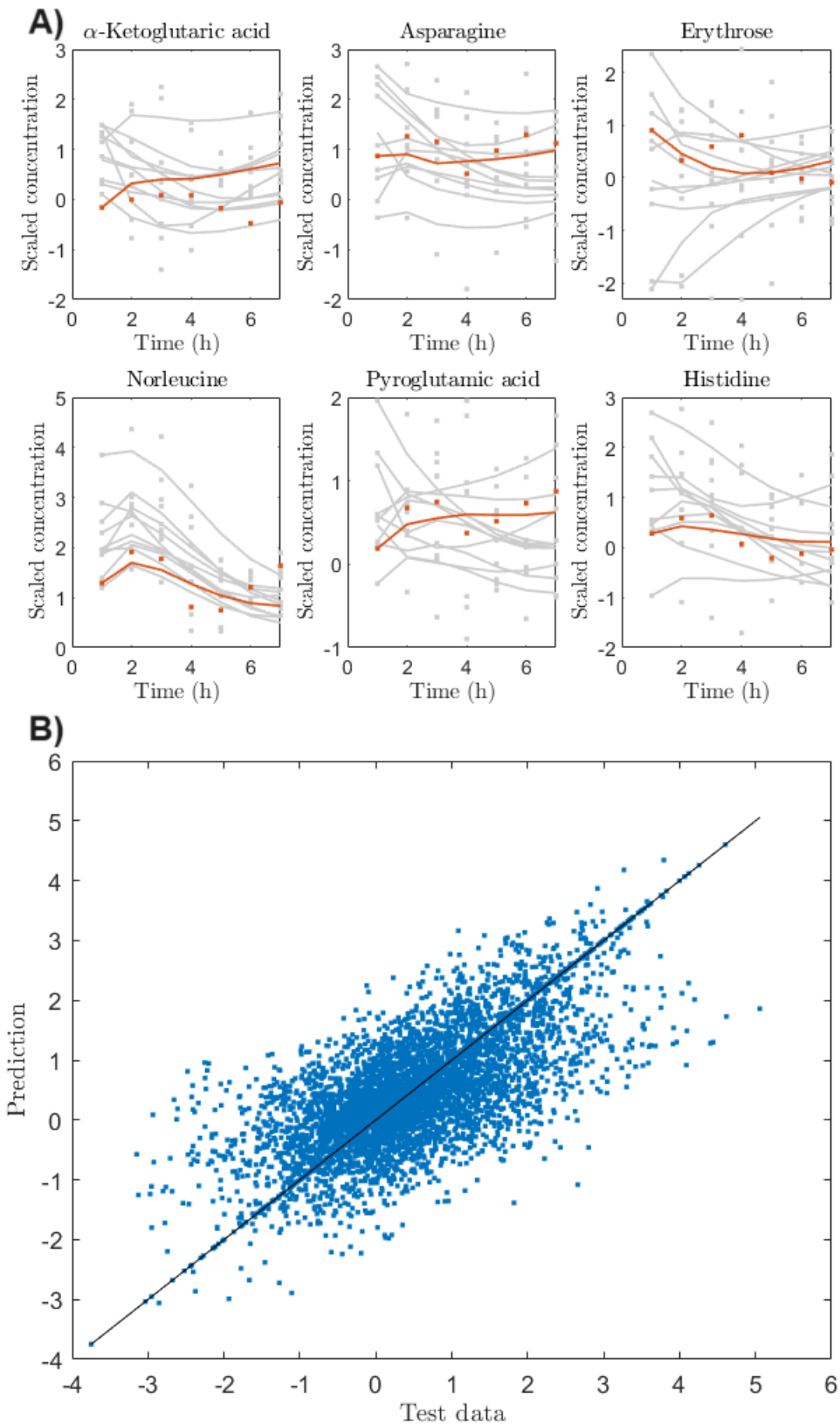

Figure 13: A) Normalized dynamic metabolite trajectories for training (gray) and holdout test (red) observations, exemplified for 6 out of 79 metabolites. Here dots are data and lines are model prediction (red) or reconstruction (grey). B) Entire normalized test data and prediction of test data as a scatter plot for one of the cross-validation iterations. The line represents the perfect match between data and predictions.

A simulation study was conducted to extrapolate the prediction results from the crossover intervention data to situations where measurements from a larger number of diets were available. The simulated data was a subset to match the dimensions of the measured data but with more dietary exposures. The number of dietary exposures included in the training set varied from 3 to 40 to evaluate the effect on prediction accuracy from the inclusion of additional dietary responses. The results showed that using the measured data (three diets) yielded R^2 = 0.4 which was close to what was achieved on average in the simulated data using three diets in the training set. Further, more diets in the simulated training set increased the prediction performance up to R^2 = 0.65 (Fig. 14). The increase in prediction performance in simulated data when increasing diets in the training set indicates that the inclusion of further diets has the potential to greatly improve the performance of the model. However, the performance seems to saturate after 30 diets in the training data.

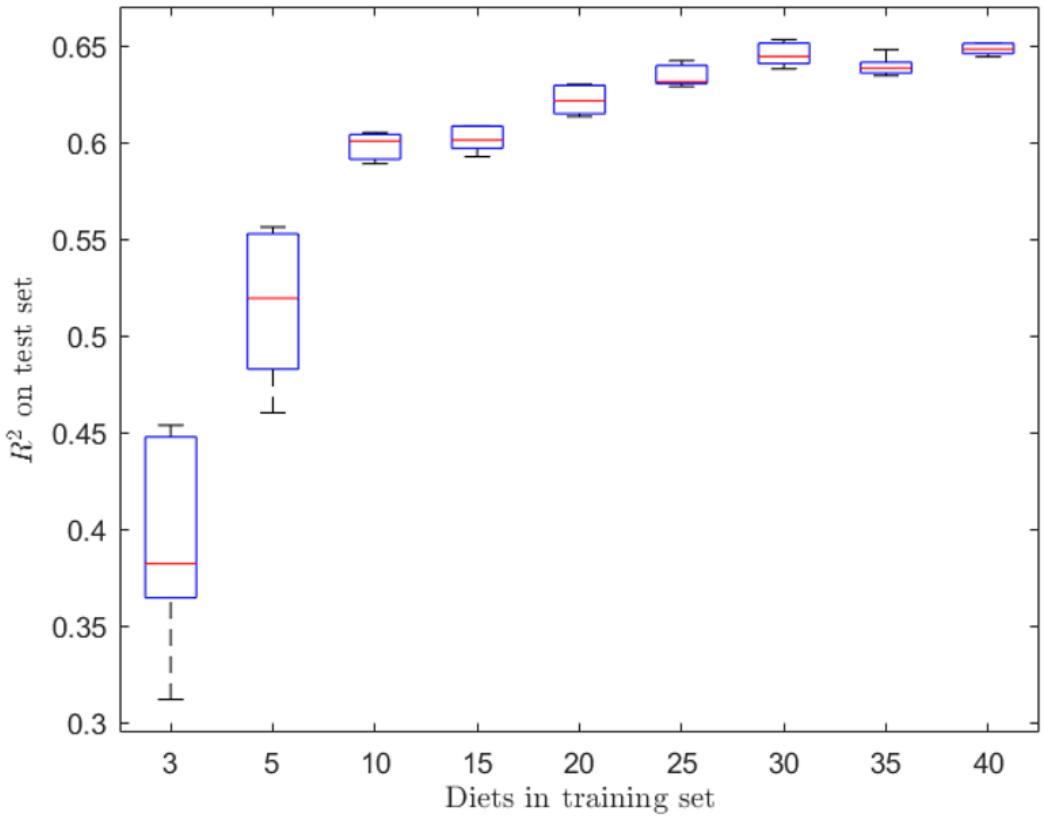

Figure 14: Prediction metric R^2 of a large test set with increasing number of diets in the training set where 5 iterations of scrambling of the examples prior to splitting training and validation were performed.

The results indicate that pDMDc can be a useful tool in precision nutrition as it can predict metabolite response to meals and identify differential responders relating to metabotypes. Moreover, given the potential of pDMDc to predict responses to new diets, it holds great promise as a tool for optimizing dietary input for desired postprandial metabolite trajectories. However, although metabolomics data was used in this study, pDMDc has the potential to be used in other omics data as well, extending to other research fields, such as pharmacology. Furthermore, pDMDc learns LDSs which come with a well-developed theory for automatic control, which is used in technical regulatory systems such as in the heating and cooling of houses. Hence, if the LDS approximates the metabolic system well enough, it might be possible to use the model as a tool for optimizing metabolic health via the control of diet and

retrieve information on what type of dietary input an individual would need to optimize health parameters such as specific risk factors.

5.4 Identification of differential responders in blood glucose in response to standardized meal tests

A simple mechanistic glucose model was applied to glucose responses from standardized meal tests to identify differential responders to food by clustering of estimated parameters (paper IV). Here, clustering was done in unison with parameter estimation of a kinetic model fitting the response data. The data stemmed from a dietary intervention trial testing high versus low glycemic index (GI) foods in participants at risk of developing type 2 diabetes from Sweden, the USA, and Italy described in Section 4.1.3.

Results showed that two clusters (A and B) could be estimated and that they were well separated in parameters governing amplitude and frequency in the model (Fig. 15).

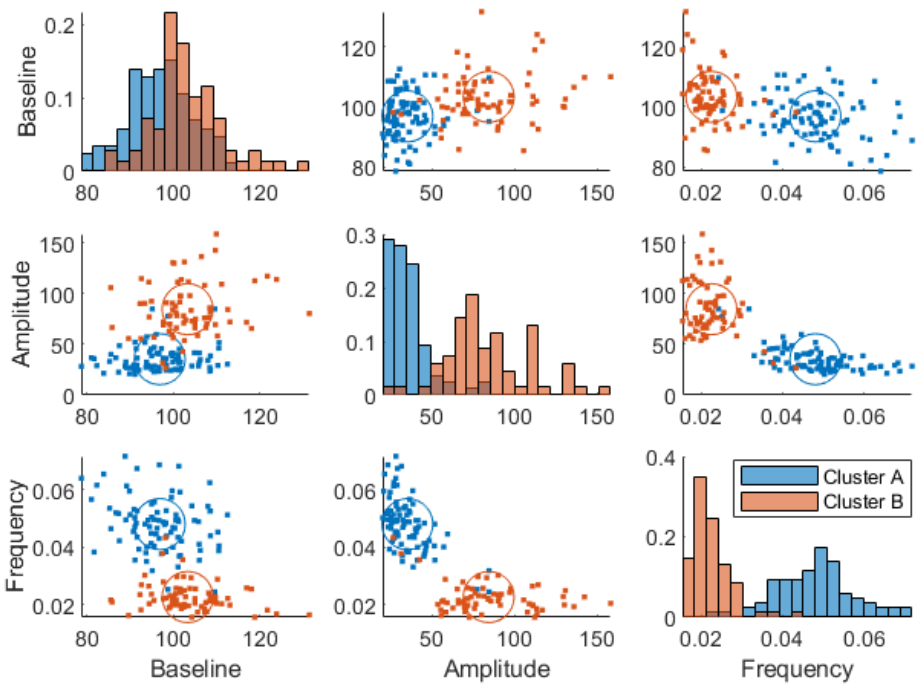

Figure 15: Parameter distribution obtained from fitting the model in Equation (4) to the postprandial breakfast MMTT data. The blue and red colors represent clusters A and B respectively. The diagonal represents histograms of the parameter distribution and the off-diagonal represents pairwise joint distributions.

The clusters were shown to be differentially associated with type 2 diabetic risk markers HbA1c ($p=2.8\cdot10^{-5}$), insulin sensitivity indices (QUICKI ($p=1.4\cdot10^{-6}$), Stumvoll ($p=1.7\cdot10^{-3}$),

Matsuda ($p=1.8\cdot10^{-8}$) (110))and waist circumference ($p=1.1\cdot10^{-6}$) using one-way ANOVA (Fig. 16).

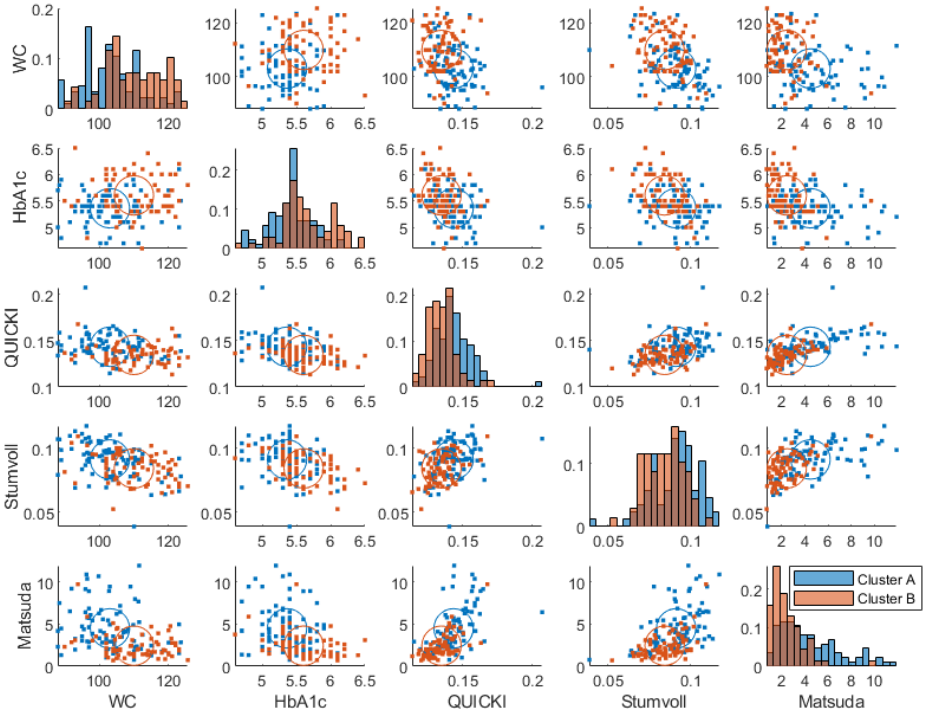

Figure 16: Baseline joint distribution of diabetes risk markers which had significant associations with clusters. The diagonal represents histograms of the parameter distribution and the off-diagonal represents pairwise joint distributions.

The clusters were also shown to be associated differentially with the gut microbiota genera *Clostridium sensu stricto 1* (ANOVA p = 0.007) and *Blautia* (ANOVA p = 0.024), which have been reported to be associated with glucose metabolism. Cluster A had a higher proportion of *Clostridium sensu stricto 1* than cluster B and vice versa for *Blautia*, which was consistent with previously reported associations of these genera with glucose control (150–152). However, it was not obvious what mechanistic relationship these genera had to the postprandial glucose

response. Further, the individuals in cluster A were observed to have a low and early glucose peak in general, indicating that they have a more efficient glucose regulation (Fig. 17).

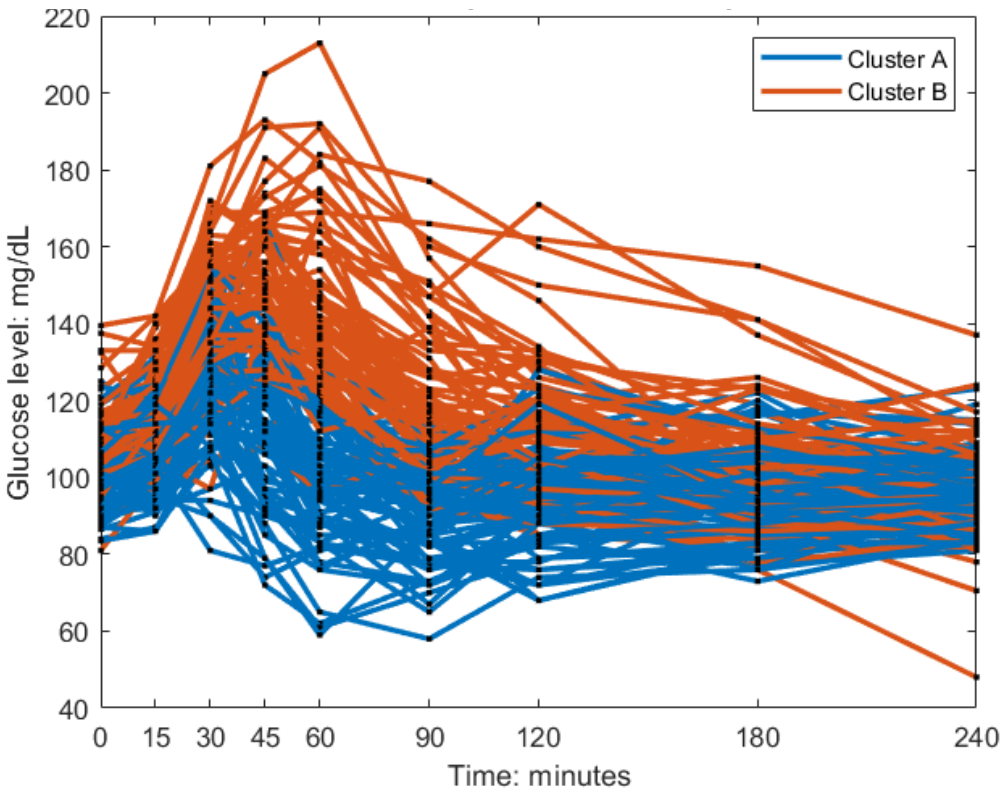

Figure 17: Baseline postprandial breakfast MMTT response color-coded by the clusters.

The clusters were also associated differently with clinical cut-offs for differential glucose control, *i.e.*, prediabetes (fasting HbA1c ≥5.7% and fasting blood glucose>100mg/dl, p=0.01, insulin resistance ($p=6.5\cdot10^{-7}$), (Matsuda index ≤2.5), glucose control ($p=6.6\cdot10^{-5}$) (normal, impaired, or diabetic (153,154) using a Chi-squared test. The same cluster analysis and model fitting were conducted using the data post-intervention after 12 weeks and the clusters were found to be stable (Cohen's kappa = 0.42, moderate stability between clusters (95%CI 0.27-0.56)).

Importantly, the clusters were not associated with diet, meaning that individuals in the unhealthy cluster B showed poor glucose control even when consuming the low GI diet, suggesting that alternative dietary compositions should be investigated to improve the metabolic health of this group. This highlights the importance of precision nutrition for individuals with poor glucose control. With the development of CGM technology, continuous glucose measurements are now easily obtained, but a corresponding method for insulin measurements is lacking. The potential of making clinical use of a model that only requires glucose measurements is therefore greater than one relying on both metabolic markers. Hence, the use of this model, together with the mixed effects framework to jointly estimate the parameters of clusters and glucose dynamics may be used in a home sampling environment with CGM measurements. Furthermore, the advantage of using mixed meal tolerance tests instead of oral glucose tolerance is that the postprandial response will provide more information about glucose regulation as it represents a more normal eating condition.

6 LIMITATIONS

The studies presented in this thesis have several limitations that are discussed below.

When using multi-omics to identify metabotypes relating to cardiometabolic diseases (paper I), filtering was performed to remove features that were not associated with cardiometabolic risk factors. Omics features associated with individual cardiometabolic risk factors were selected by the MUVR algorithm. The selected features were chosen to utilize the potential mechanistic information from all omics sources despite prediction performance being lacking for some sources selected, *i.e.*, $Q^2<0.2$. The included features were used in MFA which identified components of individuals differing in cardiometabolic health, showing that the filtering step was successful despite the criteria of low predictive capability. Further limitations include the choice of the number of clusters as internal clustering metrics (investigating optimal number of clusters k, based on the separation between them) such as Silhouette values and Bayesian information criteria proposed solutions with $k>5$. However, the available data could not differentially explain what the $k>5$ clusters were representing, indicating a low signal-to-noise ratio where the external validation was useful in identifying cardiometabolically meaningful clusters.

Identifying metabotypes as differential responders to food was done using CP and pDMDc (papers II & III), where CP was successful in identifying metabotypes despite being limited in fitting the data (27% explained variance). Using pDMDc, the estimated LDSs were less useful in providing an overview of the data than CP but were observed to have fewer limitations in fitting the data. The LDSs could also be used to predict postprandial metabolomics responses to new food, where prediction performance on simulated data increased when allowing more diets in the training set. The performance saturated at $R^2 = 0.6$ after 10 diets, indicating that the model could not fit the nonlinear simulated dynamics to the full extent. However, pDMDc performs relatively well on nonlinear dynamics despite being a linear method, which comes with benefits such as eigenvalue analysis and the possibility of inferring a data-driven optimal diet to lower cardiometabolic risk.

Differential responders were also identified in time-resolved postprandial glucose measurements using a kinetic mechanistic model (paper IV). However, the current model formulation is simplified and is likely only valid given a fasted state before the meal, since multiple sequential meals can create more complex glucose dynamics. Furthermore, a model assumption was that the glucose levels must return to the estimated baseline after the regulation of glucose is completed and the model did not capture some systematic phenomena like slow undershoots. Despite these limitations, the model described most of the physiological response and clustering of parameters represented comprehensive differences in glucose regulation.

7 CONCLUSIONS

The research presented in this doctoral thesis contributes to the advancement of mathematical modeling in precision nutrition. Specifically, the developed methods allow for the prediction of high-dimensional postprandial dynamical responses and metabotyping in multi-omics and time-resolved metabolomics.

To identify metabotypes related to cardiometabolic risk in multi-omics data from a free-living population, MFA scores were clustered and investigated for association with known risk markers (A1, paper I). The integration of multi-omics was proven useful in identifying metabotypes in static measurements and explaining their metabolic traits, leading to a better understanding of phenotypic characteristics. However, full characterization of the metabotypes was not achieved and further studies should aim to measure markers of deeper metabolic mechanics such as glucose regulation, hormone production, *etc*.

Two methods to identify groups of differential responders (metabotypes) in high-dimensional time-resolved data (A2) were developed for this purpose using CP, (paper I) and DMD (paper II), where CP provided a more interpretable summary of the data like using PCA on matrix data, while DMD could explain a higher proportion of the variation in the data and cluster more detailed dynamical trends. Both these methods can be used to infer metabotypes in data stemming from complex study designs like crossover design with repeated measurements.

Another advantage of using DMD (paper III) was that it could be used to derive predictive models of individual metabolic regulation, allowing for the prediction of dynamic postprandial responses to novel dietary interventions in high-dimensional sets of measurable metabolites (A3). To the best of my knowledge, no other methods have so far accomplished this task. Furthermore, the method is linear, has a unique solution, and is arguably easier to interpret than machine learning approaches. Finally, the method could potentially be used to identify the optimal diet for target levels of metabolites to reduce cardiometabolic risk factors.

To identify differential responders in blood glucose response to standardized meal tests, a simple mechanistic kinetic model was used. Two response clusters were identified in model parameters (A4) and were differently associated with diabetes risk markers (paper IV). The high-risk cluster showed poor glucose control compared to the low-risk cluster even after consuming a low GI meal, highlighting the need for a tailored diet for this group, the interindividual variability in response to food, and the importance of the development of precision nutrition. The mechanistic model used requires only glucose measurements, making it ideal to test in a home setting using CGM.

To conclude, the research presented in this doctoral thesis has contributed to advance the research on metabotyping in static and time-resolved data, identification of differential responders, and prediction of high-dimensional postprandial response. Using these methods, metabotypes can be identified in a free-living and clinical setting using static measurements and time-resolved responses to food, respectively. Tailored diets based on the metabotypes and on predictions of postprandial responses can later be provided with the aim of lowering cardiometabolic risk. However, although the methods primarily have been applied to problems within this field, they are most likely applicable in many other scientific fields, such as precision medicine and pharmacology.

8 FUTURE PERSPECTIVES

The presented research in this thesis provides a stepping stone in the quest to provide precision nutrition. However, more work is still needed, and some important potential developments are listed below.

pDMDc was developed to identify differential responders to food and to predict postprandial response. The method could be improved by finding a more elegant way to estimate $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{B}}$ when responses from more diets are included in the data. The current method essentially concatenates the data from all exposures and finds the least squares solution that fits the data from all diets placed in sequence. Perhaps different ways to estimate the parameters would be relevant for the prediction and description of the metabolic system. Current research is conducted to further improve the estimation of the system when using data from different excitations (155,156).

Further extensions using pDMDc include that dietary intake can be included in the model with more *a priori* knowledge, as it is not realistic that the food is received as an impulse in the body, which is how the current approach suggests. The meal intake could be modeled like a dispersion to accommodate the different properties of the macronutrients. This could then result in a better estimation of the underlying metabolic dynamics. Moreover, the inclusion of micronutrients and food items should be investigated.

The model presented in paper IV is an analytical solution to a system of differential equations describing glucose dynamics, the derivation of which assumes a single impulse of glucose to the metabolic system. However, this model could be extended to include parameters representing the meal composition to accommodate for the prediction of postprandial response to varying meals. By doing so, the method could be used for the identification of differential responders, but also for prediction of response for the responder groups, like was done in paper II for high-dimensional omics data. This could allow for response prediction to various meals using a simple model with interpretable parameters that would also be used to identify differential responders.

To fully achieve precision nutrition, large dietary interventions are needed to measure the interindividual variability in response to food. Studies like this have been conducted, but only measuring a few metabolic markers over time. Here, one needs to cover as many measurable endpoints of the metabolism as possible, including time-resolved multi-omics and clinical measures. Using pDMDc one could chart the individual differences and similarities in response to food, identify metabotypes, and predict individual responses to novel dietary challenges. Having a trained model of the individual metabolisms, one could attempt to identify which metabotypes have an elevated cardiometabolic risk and estimate the optimal diet for them using pDMDc as a control tool, thus providing true precision nutrition.

9 ACKNOWLEDGEMENTS

I would like to thank the funders supporting this PhD thesis; The Swedish Foundation for Strategic Research and Formas, A Research Council for Sustainable Development.

Mats Jirstrand (main supervisor at Fraunhofer-Chalmers Centre), thank you for allowing me to work at Fraunhofer-Chalmers Centre and to learn from all the great minds employed there. Also, thanks for the great technical chats and teachings, and for always being positive and helping me in whatever regard.

Rikard Landberg (main supervisor at Chalmers), many thanks for all the support during late nights of hard work. You are always so positive and very interested in science - a true inspiration!

Mikael Wallman (co-supervisor at Fraunhofer-Chalmers Centre), thank you for all the great support in difficult times and chats about technical details but also the great in-depth discussions about music or other random subjects. I was very lucky to have had you as a supervisor.

Carl Brunius (co-supervisor at Chalmers), thank you so very much for all the help regarding technical details and for your never-ending enthusiasm about music and whatever random stuff we end up talking about! I really appreciated having a beer after work, and talking about Anarcho-syndicalism.

Fredrik Westerlund (examiner), thank you for the help and quick feedback on my work.

Marcus, we started the doctoral studies together at FCC and I'm really lucky to have had you as a colleague and friend. You were always up for a chat, whether about some technical detail or what to do during the weekend ;).

Mattias, a brother from another island! I really enjoyed having you as an office mate and sometimes spending too much time listening to Sörns than we should have. You really made my days better when sending AI-generated images of me writing my thesis!

Ludvig, thanks for all the great laughs during lunch!

Magnus, thanks for the discussions and feedback regarding DMD and other techniques!

FCC colleagues, thanks for always giving great feedback on ideas and for the fun chats on the sofas.

João, muito obrigado por ser o meu amigo mais proximo no Chalmers! Foi ótimo partilhar o escritório com você, cara!

Paula, muito obrigado pelo apoio, ânimo e por todos os risos durante esse período!

Therese, thanks for the nice work on our joint project and for the fun chats in the office.

Anna, thank you for the interesting talks on glucose metabolism and for the help making my own CGM analyses.

Marie, thank you for your support in the MAX project and for smart comments to improve my work.

Agnetha, thank you for the fast support regarding the MAX data.

Ann-Sofie, thanks for always being supportive and interested in my work on your data.

Clemens and Nathalie, thank you for the thorough revision of my thesis!

Elahe, thank you for always being so kind and helpful. I really enjoyed the nice chats over some tea and dried Persian fruit.

Elise, the northerner who does everything! Thank you for your help in work and in calming me down, haha!

Núria, you became a true friend very fast! Thank you for all the great chats and fun moments together. I'm really happy you quit your first PhD, hehe!

Annalisa, a true inspiration in terms of work ethic and memorization skills! Thank you for being a great collaborator in our joint project and for the discussions on cardiometabolic health ;).

Agatha, thanks for the great and fun chats on life and the future!

Sebastian, the beast of FNS! Thanks for the interesting talks on diverse ways of biohacking and consumption of chicken hearts.

Marina, thanks for the fun talks on Greek celebrities!

Thank you to all the coworkers at FNS who helped maintain a fun and exciting working environment!

Thanks to Johan, for crating the wonderful art piece on the cover of this thesis!

Till sist, ett stort tack till min familj som alltid har stöttat mig genom åren!

10 REFERENCES

1. Fact sheets [Internet]. [cited 2023 Nov 9]. Available from: https://www.who.int/news-room/fact-sheets

2. Shi S, Huang H, Huang Y, Zhong VW, Feng N. Lifestyle Behaviors and Cardiometabolic Diseases by Race and Ethnicity and Social Risk Factors Among US Young Adults, 2011 to 2018. Journal of the American Heart Association. 2023 Sep 5;12(17):e028926.

3. Sattar N, Gill JMR, Alazawi W. Improving prevention strategies for cardiometabolic disease. Nat Med. 2020 Mar;26(3):320–5.

4. Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. American Heart Journal. 1991 Jan 1;121(1, Part 2):293–8.

5. Cannon CP. Cardiovascular disease and modifiable cardiometabolic risk factors. Clinical Cornerstone. 2007 Jan 1;8(3):11–28.

6. Drozdz D, Alvarez-Pitti J, Wójcik M, Borghi C, Gabbianelli R, Mazur A, et al. Obesity and Cardiometabolic Risk Factors: From Childhood to Adulthood. Nutrients. 2021 Nov;13(11):4176.

7. Camhi SM, Katzmarzyk PT. Tracking of cardiometabolic risk factor clustering from childhood to adulthood. International Journal of Pediatric Obesity. 2010 Jan 1;5(2):122–9.

8. Assmann G, Buono P, Daniele A, Della Valle E, Farinaro E, Ferns G, et al. Functional foods and cardiometabolic diseases: International Task Force for Prevention of Cardiometabolic Diseases. Nutrition, Metabolism and Cardiovascular Diseases. 2014 Dec 1;24(12):1272–300.

9. Van Camp G. Cardiovascular disease prevention. Acta Clinica Belgica. 2014 Dec 1;69(6):407–11.

10. Vailati-Riboni M, Palombo V, Loor JJ. What Are Omics Sciences? In: Ametaj BN, editor. Periparturient Diseases of Dairy Cows: A Systems Biology Approach [Internet]. Cham: Springer International Publishing; 2017 [cited 2023 Dec 1]. p. 1–7. Available from: https://doi.org/10.1007/978-3-319-43033-1_1

11. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biology. 2017 May 5;18(1):83.

12. Worley B, Powers R. Multivariate Analysis in Metabolomics. Current Metabolomics. 2013 Jan 1;1(1):92–107.

13. Palmnäs M, Brunius C, Shi L, Rostgaard-Hansen A, Torres NE, González-Domínguez R, et al. Perspective: Metabotyping—A Potential Personalized Nutrition Strategy for Precision Prevention of Cardiometabolic Disease. Advances in Nutrition [Internet]. 2019; Available from: https://academic.oup.com/advances/advance-article/doi/10.1093/advances/nmz121/5645624

14. Maiani G, Periago Castón MJ, Catasta G, Toti E, Cambrodón IG, Bysted A, et al. Carotenoids: Actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. Molecular Nutrition & Food Research. 2009;53(S2):S194–218.

15. Adherence to dietary recommendations for Swedish adults across categories of greenhouse gas emissions from food | Public Health Nutrition | Cambridge Core [Internet]. [cited 2023 Dec 2]. Available from: https://www.cambridge.org/core/journals/public-health-nutrition/article/adherence-to-dietary-recommendations-for-swedish-adults-across-categories-of-greenhouse-gas-emissions-from-food/D7FDD847F74D5CF386C69F4BAD6A2B05

16. Nutrient intakes and compliance with nutrient recommendations in children aged 1–4 years in Ireland - Walton - 2017 - Journal of Human Nutrition and Dietetics - Wiley Online Library [Internet]. [cited 2023 Dec 2]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/jhn.12452

17. Krebs-Smith SM, Guenther PM, Subar AF, Kirkpatrick SI, Dodd KW. Americans do not meet federal dietary recommendations. J Nutr. 2010 Oct;140(10):1832–8.

18. de Roos B. Personalised nutrition: ready for practice? Proc Nutr Soc. 2013 Feb;72(1):48–52.

19. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008 Sep 5;134(5):714–7.

20. Nicholson JK, Holmes E. Global systems biology and personalized healthcare solutions. Discov Med. 2006 Apr;6(32):63–70.

21. Beger RD, Dunn W, Schmidt MA, Gross SS, Kirwan JA, Cascante M, et al. Metabolomics enables precision medicine: ‘A White Paper, Community Perspective’. Metabolomics. 2016;12(10):149.

22. Livingstone KM, Ramos-Lopez O, Pérusse L, Kato H, Ordovas JM, Martínez JA. Precision nutrition: A review of current approaches and future endeavors. Trends in Food Science & Technology. 2022 Oct 1;128:253–64.

23. Riedl A, Gieger C, Hauner H, Daniel H, Linseisen J. Metabotyping and its application in targeted nutrition: an overview. British Journal of Nutrition. 2017 Jun;117(12):1631–44.

24. Lappa D, Meijnikman AS, Krautkramer KA, Olsson LM, Aydin Ö, Rijswijk ASV, et al. Self-organized metabotyping of obese individuals identifies clusters responding differently to bariatric surgery. PLOS ONE. 2023 Mar 2;18(3):e0279335.

25. Muniandy M, Velagapudi V, Hakkarainen A, Lundbom J, Lundbom N, Rissanen A, et al. Plasma metabolites reveal distinct profiles associating with different metabolic risk factors in monozygotic twin pairs. Int J Obes. 2019 Mar;43(3):487–502.

26. Li K, Brennan L, McNulty BA, Bloomfield JF, Duff DanJ, Devlin NFC, et al. Plasma fatty acid patterns reflect dietary habits and metabolic health: A cross-sectional study. Molecular Nutrition & Food Research. 2016;60(9):2043–52.

27. Fiamoncini J, Rundle M, Gibbons H, Thomas EL, Geillinger-Kästle K, Bunzel D, et al. Plasma metabolome analysis identifies distinct human metabotypes in the postprandial state with different susceptibility to weight loss-mediated metabolic improvements. The FASEB Journal. 2018;32(10):5447–58.

28. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell. 2015 Nov 19;163(5):1079–94.

29. Seo W, Lee YB, Lee S, Jin SM, Park SM. A machine-learning approach to predict postprandial hypoglycemia. BMC Medical Informatics and Decision Making. 2019 Nov 6;19(1):210.

30. Mendes-Soares H, Raveh-Sadka T, Azulay S, Edens K, Ben-Shlomo Y, Cohen Y, et al. Assessment of a Personalized Approach to Predicting Postprandial Glycemic Responses to Food Among Individuals Without Diabetes. JAMA Network Open. 2019 Feb 8;2(2):e188102.

31. Bell KJ, Petocz P, Colagiuri S, Brand-Miller JC. Algorithms to Improve the Prediction of Postprandial Insulinaemia in Response to Common Foods. Nutrients. 2016 Apr;8(4):210.

32. Berry SE, Valdes AM, Drew DA, Asnicar F, Mazidi M, Wolf J, et al. Human postprandial responses to food and potential for precision nutrition. Nat Med. 2020 Jun;26(6):964–73.

33. Human Metabolome Database: Database Statistics [Internet]. [cited 2023 Dec 2]. Available from: https://hmdb.ca/statistics

34. de Waard AKM, Hollander M, Korevaar JC, Nielen MMJ, Carlsson AC, Lionis C, et al. Selective prevention of cardiometabolic diseases: activities and attitudes of general practitioners across Europe. European Journal of Public Health. 2019 Feb 1;29(1):88–93.

35. Ndisang JF, Rastogi S. Cardiometabolic Diseases and Related Complications: Current Status and Future Perspective. Biomed Res Int. 2013;2013:467682.

36. Miranda JJ, Barrientos-Gutiérrez T, Corvalan C, Hyder AA, Lazo-Porras M, Oni T, et al. Understanding the rise of cardiometabolic diseases in low- and middle-income countries. Nat Med. 2019 Nov;25(11):1667–79.

37. Mezhal F, Oulhaj A, Abdulle A, AlJunaibi A, Alnaeemi A, Ahmad A, et al. High prevalence of cardiometabolic risk factors amongst young adults in the United Arab Emirates: the UAE Healthy Future Study. BMC Cardiovascular Disorders. 2023 Mar 15;23(1):137.

38. Nabel EG. Cardiovascular Disease. New England Journal of Medicine. 2003 Jul 3;349(1):60–72.

39. Tao Z, Shi A, Zhao J. Epidemiological Perspectives of Diabetes. Cell Biochem Biophys. 2015 Sep 1;73(1):181–5.

40. Chronic Kidney Disease in the United States, 2023 [Internet]. 2023 [cited 2023 Oct 10]. Available from: https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html

41. Angulo P. Nonalcoholic Fatty Liver Disease. New England Journal of Medicine. 2002 Apr 18;346(16):1221–31.

42. Rao GH. Prevention or Reversal of Cardiometabolic Diseases. Journal of Clinical and Preventive Cardiology. 2018 Mar;7(1):22.

43. Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Arós F, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. N Engl J Med. 2018 Jun 21;378(25):e34.

44. Adamsson V, Reumark A, Fredriksson IB, Hammarström E, Vessby B, Johansson G, et al. Effects of a healthy Nordic diet on cardiovascular risk factors in hypercholesterolaemic subjects: a randomized controlled trial (NORDIET). J Intern Med. 2011 Feb;269(2):150–9.

45. de Mello VDF, Schwab U, Kolehmainen M, Koenig W, Siloaho M, Poutanen K, et al. A diet high in fatty fish, bilberries and wholegrain products improves markers of endothelial function and inflammation in individuals with impaired glucose metabolism in a randomised controlled trial: the Sysdimet study. Diabetologia. 2011 Nov;54(11):2755–67.

46. Olsen A, Egeberg R, Halkjær J, Christensen J, Overvad K, Tjønneland A. Healthy aspects of the Nordic diet are related to lower total mortality. J Nutr. 2011 Apr 1;141(4):639–44.

47. Lankinen M, Uusitupa M, Schwab U. Nordic Diet and Inflammation—A Review of Observational and Intervention Studies. Nutrients. 2019 Jun;11(6):1369.

48. Meyer KA, Kushi LH, Jacobs DR, Slavin J, Sellers TA, Folsom AR. Carbohydrates, dietary fiber, and incident type 2 diabetes in older women. Am J Clin Nutr. 2000 Apr;71(4):921–30.

49. Willett WC. Dietary fats and coronary heart disease. Journal of Internal Medicine. 2012;272(1):13–24.

50. Beecher CWW. The Human Metabolome. In: Harrigan GG, Goodacre R, editors. Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis [Internet]. Boston, MA: Springer US; 2003 [cited 2023 Dec 11]. p. 311–9. Available from: https://doi.org/10.1007/978-1-4615-0333-0_17

51. Idle JR, Gonzalez FJ. Metabolomics. Cell Metabolism. 2007 Nov;6(5):348–51.

52. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrometry Reviews. 2007;26(1):51–78.

53. Systematic review and meta-analysis of the effectiveness of continuous glucose monitoring (CGM) on glucose control in diabetes | Diabetology & Metabolic Syndrome [Internet]. [cited 2023 Dec 3]. Available from: https://link.springer.com/article/10.1186/1758-5996-5-39

54. Nutrients | Free Full-Text | Determination of Postprandial Glycemic Responses by Continuous Glucose Monitoring in a Real-World Setting [Internet]. [cited 2023 Dec 3]. Available from: https://www.mdpi.com/2072-6643/11/10/2305

55. Hur KY, Lee MS. Gut Microbiota and Metabolic Disorders. Diabetes Metab J. 2015 Jun;39(3):198–203.

56. Hou K, Wu ZX, Chen XY, Wang JQ, Zhang D, Xiao C, et al. Microbiota in health and diseases. Sig Transduct Target Ther. 2022 Apr 23;7(1):1–28.

57. Gabriel CL, Ferguson JF. Gut Microbiota and Microbial Metabolism in Early Risk of Cardiometabolic Disease. Circulation Research. 2023 Jun 9;132(12):1674–91.

58. Tan J, McKenzie C, Potamitis M, Thorburn AN, Mackay CR, Macia L. Chapter Three - The Role of Short-Chain Fatty Acids in Health and Disease. In: Alt FW, editor. Advances in Immunology [Internet]. Academic Press; 2014 [cited 2023 Dec 3]. p. 91–119. Available from: https://www.sciencedirect.com/science/article/pii/B9780128001004000039

59. Relevance of protein fermentation to gut health - PubMed [Internet]. [cited 2023 Dec 11]. Available from: https://pubmed.ncbi.nlm.nih.gov/22121108/

60. Bokulich NA, Ziemski M, Robeson MS, Kaehler BD. Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods. Computational and Structural Biotechnology Journal. 2020 Jan 1;18:4048–62.

61. Ewers B, Trolle E, Jacobsen SS, Vististen D, Almdal TP, Vilsbøll T, et al. Dietary habits and adherence to dietary recommendations in patients with type 1 and type 2 diabetes compared with the general population in Denmark. Nutrition. 2019 May 1;61:49–55.

62. Nordic Nutrition Recommendations 2023 [Internet]. 2023 [cited 2023 Dec 3]. Available from: https://www.norden.org/en/publication/nordic-nutrition-recommendations-2023

63. Kirk D, Catal C, Tekinerdogan B. Precision nutrition: A systematic literature review. Computers in Biology and Medicine. 2021 Jun 1;133:104365.

64. Cecil JE, Barton KL. Inter-individual differences in the nutrition response: from research to recommendations. Proceedings of the Nutrition Society. 2020 May;79(2):171–3.

65. Grillo A, Salvi L, Coruzzi P, Salvi P, Parati G. Sodium Intake and Hypertension. Nutrients. 2019 Sep;11(9):1970.

66. Cornelis MC, El-Sohemy A, Campos H. Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption2. The American Journal of Clinical Nutrition. 2007 Jul 1;86(1):240–4.

67. Horne J, Madill J, O'Connor C, Shelley J, Gilliland J. A Systematic Review of Genetic Testing and Lifestyle Behaviour Change: Are We Using High-Quality Genetic Interventions and Considering Behaviour Change Theory? Lifestyle Genomics. 2018 Apr 10;11(1):49–63.

68. Fallaize R, Macready AL, Butler LT, Ellis JA, Lovegrove JA. An insight into the public acceptance of nutrigenomic-based personalised nutrition. Nutrition Research Reviews. 2013 Jun;26(1):39–48.

69. Ordovas JM, Ferguson LR, Tai ES, Mathers JC. Personalised nutrition and health. BMJ. 2018 Jun 13;361:bmj.k2173.

70. Roper JA. Genetic determination of nutritional requirements. Proceedings of the Nutrition Society. 1960 Mar;19(1):39–45.

71. Mathers JC. Nutrigenomics in the modern era. Proceedings of the Nutrition Society. 2017 Aug;76(3):265–75.

72. Bordoni L, Gabbianelli R. Primers on nutrigenetics and nutri(epi)genomics: Origins and development of precision nutrition. Biochimie. 2019 May 1;160:156–71.

73. Mattar R, Mazo DF de C, Carrilho FJ. Lactose intolerance: diagnosis, genetic, and clinical factors. CEG. 2012 Jul 5;5:113–21.

74. Perez-Martinez P, Garcia-Rios A, Delgado-Lista J, Perez-Jimenez F, Lopez-Miranda J. Metabolic syndrome: Evidences for a personalized nutrition. Molecular Nutrition & Food Research. 2012;56(1):67–76.

75. Guest NS, Horne J, Vanderhout SM, El-Sohemy A. Sport Nutrigenomics: Personalized Nutrition for Athletic Performance. Front Nutr. 2019 Feb 19;6:8.

76. Goni L, Cuervo M, Milagro FI, Martínez JA. Future Perspectives of Personalized Weight Loss Interventions Based on Nutrigenetic, Epigenetic, and Metagenomic Data123. The Journal of Nutrition. 2016 Apr 1;146(4):905S-912S.

77. Horne J, Gilliland J, O'Connor C, Seabrook J, Hannaberg P, Madill J. Study protocol of a pragmatic randomized controlled trial incorporated into the Group Lifestyle Balance™ program: the nutrigenomics, overweight/obesity and weight management trial (the NOW trial). BMC Public Health. 2019 Mar 15;19(1):310.

78. Horne JR, Gilliland JA, O'Connor CP, Seabrook JA, Madill J. Change in Weight, BMI, and Body Composition in a Population-Based Intervention Versus Genetic-Based Intervention: The NOW Trial. Obesity (Silver Spring). 2020 Aug;28(8):1419–27.

79. Höchsmann C, Yang S, Ordovás JM, Dorling JL, Champagne CM, Apolzan JW, et al. The Personalized Nutrition Study (POINTS): evaluation of a genetically informed weight loss approach, a Randomized Clinical Trial. Nat Commun. 2023 Oct 9;14(1):6321.

80. Nacis JS, Labrador JPH, Ronquillo DGD, Rodriguez MP, Dablo AMFD, Frane RD, et al. A study protocol for a pilot randomized controlled trial to evaluate the effectiveness of a gene-based nutrition and lifestyle recommendation for weight management among adults: the MyGeneMyDiet® study. Frontiers in Nutrition [Internet]. 2023 [cited 2023 Dec 8];10. Available from: https://www.frontiersin.org/articles/10.3389/fnut.2023.1238234

81. Celis-Morales C, Livingstone KM, Marsaux CFM, Forster H, O'Donovan CB, Woolhead C, et al. Design and baseline characteristics of the Food4Me study: a web-based

randomised controlled trial of personalised nutrition in seven European countries. Genes Nutr. 2014 Dec 10;10(1):450.

82. Celis-Morales C, Livingstone KM, Marsaux CF, Macready AL, Fallaize R, O'Donovan CB, et al. Effect of personalized nutrition on health-related behaviour change: evidence from the Food4Me European randomized controlled trial. Int J Epidemiol. 2017 Apr 1;46(2):578–88.

83. Celis-Morales C, Lara J, Mathers JC. Personalising nutritional guidance for more effective behaviour change. Proceedings of the Nutrition Society. 2015 May;74(2):130–8.

84. Hall H, Perelman D, Breschi A, Limcaoco P, Kellogg R, McLaughlin T, et al. Glucotypes reveal new patterns of glucose dysregulation. PLoS Biol. 2018 Jul;16(7):e2005143.

85. Mendes-Soares H, Raveh-Sadka T, Azulay S, Ben-Shlomo Y, Cohen Y, Ofek T, et al. Model of personalized postprandial glycemic response to food developed for an Israeli cohort predicts responses in Midwestern American individuals. Am J Clin Nutr. 2019 Jul 1;110(1):63–75.

86. Paoli A, Tinsley G, Bianco A, Moro T. The Influence of Meal Frequency and Timing on Health in Humans: The Role of Fasting. Nutrients. 2019 Apr;11(4):719.

87. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. 2018 May;6(5):361–9.

88. Gijbels A, Trouwborst I, Jardon KM, Hul GB, Siebelink E, Bowser SM, et al. The PERSonalized Glucose Optimization Through Nutritional Intervention (PERSON) Study: Rationale, Design and Preliminary Screening Results. Front Nutr. 2021;8:694568.

89. Trouwborst I, Gijbels A, Jardon KM, Siebelink E, Hul GB, Wanders L, et al. Cardiometabolic health improvements upon dietary intervention are driven by tissue-specific insulin resistance phenotype: A precision nutrition trial. Cell Metab. 2023 Jan 3;35(1):71-83.e5.

90. Riedl A, Wawro N, Gieger C, Meisinger C, Peters A, Roden M, et al. Identification of Comprehensive Metabotypes Associated with Cardiometabolic Diseases in the Population-Based KORA Study. Mol Nutr Food Res. 2018 Aug;62(16):e1800117.

91. Krishnan S, Newman JW, Hembrooke TA, Keim NL. Variation in metabolic responses to meal challenges differing in glycemic index in healthy women: Is it meaningful? Nutr Metab (Lond). 2012 Mar 29;9:26.

92. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning [Internet]. New York, NY: Springer; 2009 [cited 2023 Sep 26]. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-0-387-84858-7

93. Botelho PB, Fioratti CO, Abdalla DSP, Bertolami MC, Castro IA. Classification of individuals with dyslipidaemia controlled by statins according to plasma biomarkers of oxidative stress using cluster analysis. British Journal of Nutrition. 2010 Jan;103(2):256–65.

94. Wilcox M, Li Q, Sun Y, Stang P, Berlin J, Wang D. Genome-wide association study for empirically derived metabolic phenotypes in the Framingham Heart Study offspring cohort. BMC Proc. 2009 Dec 15;3 Suppl 7(Suppl 7):S53.

95. Zubair N, Kuzawa CW, McDade TW, Adair LS. Cluster Analysis Reveals Important Determinants of Cardiometabolic Risk Patterns in Filipino Women. Asia Pac J Clin Nutr. 2012;21(2):271–81.

96. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinform Biol Insights. 2020 Jan 1;14:1177932219899051.

97. Bro R. PARAFAC. Tutorial and applications. Chemometrics and Intelligent Laboratory Systems. 1997;

98. Hedaya MA. Basic Pharmacokinetics, Second Edition. CRC Press; 2012. 599 p.

99. Tricomi FG, McHarg EA. Differential Equations. Courier Corporation; 2012. 290 p.

100. Wang T, Holscher HD, Maslov S, Hu FB, Weiss ST, Liu YY. Predicting metabolic response to dietary intervention using deep learning. bioRxiv. 2023 Mar 15;2023.03.14.532589.

101. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. Journal of Big Data. 2015 Feb 24;2(1):1.

102. Petersen KEN, Halkjær J, Loft S, Tjønneland A, Olsen A. Cohort profile and representativeness of participants in the Diet, Cancer and Health-Next Generations cohort study. Eur J Epidemiol. 2022 Jan;37(1):117–27.

103. Martínez-González MA, Sánchez-Tainta A, Corella D, Salas-Salvadó J, Ros E, Arós F, et al. A provegetarian food pattern and reduction in total mortality in the Prevención con Dieta Mediterránea (PREDIMED) study. Am J Clin Nutr. 2014 Jul;100 Suppl 1:320S-8S.

104. Hansen CP, Overvad K, Kyrø C, Olsen A, Tjønneland A, Johnsen SP, et al. Adherence to a Healthy Nordic Diet and Risk of Stroke: A Danish Cohort Study. Stroke. 2017 Feb;48(2):259–64.

105. Satija A, Bhupathiraju SN, Rimm EB, Spiegelman D, Chiuve SE, Borgi L, et al. Plant-Based Dietary Patterns and Incidence of Type 2 Diabetes in US Men and Women: Results from Three Prospective Cohort Studies. PLoS Med. 2016 Jun;13(6):e1002039.

106. Ross AB, Svelander C, Undeland I, Pinto R, Sandberg AS. Herring and Beef Meals Lead to Differences in Plasma 2-Aminoadipic Acid, β-Alanine, 4-Hydroxyproline, Cetoleic Acid, and Docosahexaenoic Acid Concentrations in Overweight Men. The Journal of Nutrition. 2015;145(11):2456–63.

107. Svelander C, Gabrielsson BG, Almgren A, Gottfries J, Olsson J, Undeland I, et al. Postprandial lipid and insulin responses among healthy, overweight men to mixed

meals served with baked herring, pickled herring or baked, minced beef. Eur J Nutr. 2015 Sep;54(6):945–58.

108. Bergia RE, Giacco R, Hjorth T, Biskup I, Zhu W, Costabile G, et al. Differential Glycemic Effects of Low- versus High-Glycemic Index Mediterranean-Style Eating Patterns in Adults at Risk for Type 2 Diabetes: The MEDGI-Carb Randomized Controlled Trial. Nutrients. 2022 Jan;14(3):706.

109. Bergia RE, Biskup I, Giacco R, Costabile G, Gray S, Wright A, et al. The MEDGICarb-Study: Design of a multi-center randomized controlled trial to determine the differential health-promoting effects of low- and high-glycemic index Mediterranean-style eating patterns. Contemporary Clinical Trials Communications. 2020 Sep 1;19:100640.

110. Gutch M, Kumar S, Razi SM, Gupta KK, Gupta A. Assessment of insulin sensitivity/resistance. Indian J Endocrinol Metab. 2015;19(1):160–4.

111. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems. 1987 Aug 1;2(1):37–52.

112. Horn RA, Johnson CR. Matrix Analysis. Cambridge University Press; 2012. 663 p.

113. Mishra P, Roger JM, Jouan-Rimbaud-Bouveresse D, Biancolillo A, Marini F, Nordon A, et al. Recent trends in multi-block data analysis in chemometrics for multi-source data integration. TrAC Trends in Analytical Chemistry. 2021 Apr 1;137:116206.

114. Liu J, Wang C, Gao J, Han J. Multi-View Clustering via Joint Nonnegative Matrix Factorization. In: Proceedings of the 2013 SIAM International Conference on Data Mining (SDM) [Internet]. Society for Industrial and Applied Mathematics; 2013 [cited 2023 Dec 12]. p. 252–60. (Proceedings). Available from: https://epubs.siam.org/doi/abs/10.1137/1.9781611972832.28

115. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Molecular Systems Biology. 2018 Jun;14(6):e8124.

116. Comon P. Independent component analysis, A new concept? Signal Processing. 1994 Apr 1;36(3):287–314.

117. Abram KJ, McCloskey D. A Comprehensive Evaluation of Metabolomics Data Preprocessing Methods for Deep Learning. Metabolites. 2022 Feb 24;12(3):202.

118. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006 Jun 8;7:142.

119. Moser J. Dynamical systems, theory and applications. Dynamical Systems, Theory and Applications [Internet]. 1975 Jan 1 [cited 2022 Jun 27];38. Available from: https://ui.adsabs.harvard.edu/abs/1975LNP....38.....M

120. Ackerman E, Rosevear JW, McGuckin WF. A Mathematical Model of the Glucose-tolerance test. Phys Med Biol. 1964 Apr;9(2):203.

121. Bolie VW. Coefficients of normal blood glucose regulation. Journal of Applied Physiology. 1961 Sep;16(5):783–8.

122. Bergman RN, Ider YZ, Bowden CR, Cobelli C. Quantitative estimation of insulin sensitivity. American Journal of Physiology-Endocrinology and Metabolism. 1979 Jun 1;236(6):E667.

123. Palumbo P, Ditlevsen S, Bertuzzi A, De Gaetano A. Mathematical modeling of the glucose–insulin system: A review. Mathematical Biosciences. 2013 Aug 1;244(2):69–81.

124. Picchini U, De Gaetano A, Panunzi S, Ditlevsen S, Mingrone G. A mathematical model of the euglycemic hyperinsulinemic clamp. Theoretical Biology and Medical Modelling. 2005 Nov 3;2(1):44.

125. Lindstrom ML, Bates DM. Nonlinear mixed effects models for repeated measures data. Biometrics. 1990 Sep;46(3):673–87.

126. Davidian M. Nonlinear Models for Repeated Measurement Data. New York: Routledge; 2017. 360 p.

127. Wang Y. Derivation of various NONMEM estimation methods. J Pharmacokinet Pharmacodyn. 2007 Oct 1;34(5):575–93.

128. The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model - Samson - 2007 - Statistics in Medicine - Wiley Online Library [Internet]. [cited 2023 Dec 3]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2950?casa_token=Pa8Khz1iAl0AAAAA:4irkBAQ0HCiCuEQ0hYvq0mHDxQCrExuvyV-SVZSI7uZqf5KSsRR5BPsv3HBhuWj6SdRQORjoTcfJNOna

129. Markov chain Monte Carlo method and its application - Brooks - 1998 - Journal of the Royal Statistical Society: Series D (The Statistician) - Wiley Online Library [Internet]. [cited 2023 Dec 3]. Available from: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9884.00117?casa_token=ZA6ZSAZmfEkAAAAA:VG__0jwt-ZRFUi0MCHUO5gyjl3Qq4-4_XZG3WIbJQ2Yqv_WfMbYt4ZAPnTZ6q8RFxMqmBXMCzN4ANF7x

130. Foundations of the PARAFAC procedure: Models and conditions for an ‘explanatory’ multi-model factor analysis. 1970 Jan 1;16:1–84.

131. Kolda TG, Bader BW. Tensor Decompositions and Applications. SIAM Rev. 2009 Aug 6;51(3):455–500.

132. Li L, Hoefsloot H, de Graaf AA, Acar E, Smilde AK. Exploring dynamic metabolomics data with multiway data analysis: a simulation study. BMC Bioinformatics. 2022 Jan 10;23:31.

133. Bhatt V, Kumar S, Saini S. Tucker decomposition and applications. Materials Today: Proceedings. 2021 Jan 1;46:10787–92.

134. Centering and scaling in component analysis - Bro - 2003 - Journal of Chemometrics - Wiley Online Library [Internet]. [cited 2023 Dec 12]. Available from: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/cem.773

135. Glad T, Ljung L. Control Theory. CRC Press; 2000. 500 p.

136. Ljung L. Perspectives on system identification. Annual Reviews in Control. 2010 Apr 1;34(1):1–12.

137. Schmid PJ. Dynamic mode decomposition of numerical and experimental data. Journal of Fluid Mechanics. 2010 Aug;656:5–28.

138. Andreuzzi F, Demo N, Rozza G. A dynamic mode decomposition extension for the forecasting of parametric dynamical systems. arXiv:211009155 [cs, math] [Internet]. 2021 Oct 18 [cited 2022 Feb 23]; Available from: http://arxiv.org/abs/2110.09155

139. Proctor JL, Brunton SL, Kutz JN. Dynamic mode decomposition with control [Internet]. arXiv; 2014 [cited 2023 Jan 12]. Available from: http://arxiv.org/abs/1409.6358

140. Shi L, Westerhuis JA, Rosén J, Landberg R, Brunius C. Variable selection and validation in multivariate modelling. Bioinformatics. 2019 Mar 15;35(6):972–80.

141. Alberti K g. m. m., Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the Metabolic Syndrome. Circulation. 2009 Oct 20;120(16):1640–5.

142. Biddle A, Stewart L, Blanchard J, Leschine S. Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. Diversity. 2013 Sep;5(3):627–40.

143. Lakshmanan AP, Al Zaidan S, Bangarusamy DK, Al-Shamari S, Elhag W, Terranegra A. Increased Relative Abundance of Ruminoccocus Is Associated With Reduced Cardiovascular Risk in an Obese Population. Frontiers in Nutrition [Internet]. 2022 [cited 2023 Dec 4];9. Available from: https://www.frontiersin.org/articles/10.3389/fnut.2022.849005

144. Waters JL, Ley RE. The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. BMC Biology. 2019 Oct 28;17(1):83.

145. Feingold KR, Grunfeld C. Effect of inflammation on HDL structure and function. Curr Opin Lipidol. 2016 Oct;27(5):521–30.

146. Pieske B, Wachter R. Impact of diabetes and hypertension on the heart. Curr Opin Cardiol. 2008 Jul;23(4):340–9.

147. Itsiopoulos C, Brazionis L, Kaimakamis M, Cameron M, Best JD, O'Dea K, et al. Can the Mediterranean diet lower HbA1c in type 2 diabetes? Results from a randomized cross-over study. Nutr Metab Cardiovasc Dis. 2011 Sep;21(9):740–7.

148. Negatu DA, Gengenbacher M, Dartois V, Dick T. Indole Propionic Acid, an Unusual Antibiotic Produced by the Gut Microbiota, With Anti-inflammatory and Antioxidant Properties. Frontiers in Microbiology [Internet]. 2020 [cited 2023 Nov 30];11. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2020.575586

149. Kurata H. Virtual metabolic human dynamic model for pathological analysis and therapy design for diabetes. iScience. 2021 Feb 19;24(2):102101.

150. The human gut microbiota and glucose metabolism: a scoping review of key bacteria and the potential role of SCFAs - PubMed [Internet]. [cited 2023 Dec 6]. Available from: https://pubmed.ncbi.nlm.nih.gov/36026526/

151. Maskarinec G, Raquinio P, Kristal BS, Setiawan VW, Wilkens LR, Franke AA, et al. The gut microbiome and type 2 diabetes status in the Multiethnic Cohort. PLoS One. 2021;16(6):e0250855.

152. Shen N, Dimitrova N, Ho CH, Torres PJ, Camacho FR, Cai Y, et al. Gut microbiome activity predicts risk of type 2 diabetes and metformin control in a large human cohort [Internet]. medRxiv; 2021 [cited 2023 Jun 15]. p. 2021.08.13.21262051. Available from: https://www.medrxiv.org/content/10.1101/2021.08.13.21262051v1

153. Petersmann A, Müller-Wieland D, Müller UA, Landgraf R, Nauck M, Freckmann G, et al. Definition, Classification and Diagnosis of Diabetes Mellitus. Exp Clin Endocrinol Diabetes. 2019 Dec;127(S 01):S1–7.

154. Placzkowska S, Pawlik-Sobecka L, Kokot I, Piwowar A. Indirect insulin resistance detection: Current clinical trends and laboratory limitations. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2019 Sep;163(3):187–99.

155. Li CY, Chen Z, Tse TKT, Weerasuriya AU, Zhang X, Fu Y, et al. A parametric and feasibility study for data sampling of the dynamic mode decomposition: range, resolution, and universal convergence states. Nonlinear Dyn. 2022 Mar 1;107(4):3683–707.

156. Huhn QA, Tano ME, Ragusa JC, Choi Y. Parametric dynamic mode decomposition for reduced order modeling. Journal of Computational Physics. 2023 Feb 15;475:111852.

PAPERS I-IV