TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt
Lehrstuhl für Ernährungsmedizin

Identification of comprehensive metabotypes and their application in the exploration of diet-diabetes associations in the KORA study

Anna Katharina Riedl

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Martin Klingenspor
Prüfende der Dissertation:	1. Prof. Dr. Johann J. Hauner
	2. Prof. Dr. Jakob Linseisen

Die Dissertation wurde am 12.12.2018 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 12.03.2019 angenommen.

Table of content

Summary

The concept of ‘metabotyping’ denotes the formation of metabolically or phenotypically homogeneous subgroups of individuals, so-called ‘metabotypes’. In nutritional research, this concept appears frequently in conjunction with ‘targeted’ nutrition, which describes the delivery of dietary advice tailored to groups of similar individuals such as metabotype subgroups. This approach seems to be more effective than generalized advice regarding individual dietary behavior change. It is probably easier to implement in large populations than personalized advice for every single individual. Therefore, it is essential to identify distinct metabotypes that show differences in diet-disease associations and/or in their response to dietary interventions.

This dissertation comprises four manuscripts. As previous studies showed heterogeneous applications of the metabotyping concept, the first manuscript gives a broad literature overview and suggestions for its future application. Based on these suggestions, the second manuscript describes the identification of comprehensive metabotypes by an extensive parameter set available in the large population-based KORA (Cooperative Health Research in the Region of Augsburg) study. These metabotypes were significantly linked to cardiometabolic disease occurrence such as diabetes. In the third manuscript, differences between the metabotype subgroups were identified in cross-sectional associations of dietary factors with diabetes. This may be relevant for targeted dietary advice in diabetes prevention. Further, these results propose that previous inconsistent findings on diet-diabetes associations may be partially due to inter-individual metabolic diversity. The inconsistencies in diet-diabetes associations were further investigated in the fourth manuscript using a detailed breakdown of glucose tolerance status, which resulted in the identification of new associations of dietary factors with prediabetes.

In conclusion, comprehensive metabotypes were identified in the KORA study, the validity of which needs to be confirmed in other populations. The results of this dissertation further indicate the presence of differences between metabotype subgroups in diet-diabetes associations. However, their causality needs to be verified in prospective and interventional studies. In addition, both other diet-related cardiometabolic diseases, such as gout, and early disease stages, such as prediabetes, may be worth testing for distinct metabotype subgroups in their associations with diet. Finally, by considering these results, effective targeted dietary recommendations may be developed at the metabotype subgroup level for successful prevention and management of diet-related cardiometabolic diseases.

Zusammenfassung

Das ‚Metabotyping'-Konzept bezeichnet die Identifizierung metabolisch oder phänotypisch homogener Subgruppen von Individuen, sogenannter ‚Metabotypen'. Dieses Konzept taucht in der Ernährungsforschung häufig im Zusammenhang mit dem Begriff ‚Targeted Nutrition' auf, der die Bereitstellung von Ernährungsempfehlungen zugeschnitten auf Gruppen ähnlicher Individuen, wie Metabotyp-Subgruppen, beschreibt. Dieser Ansatz erscheint effektiver als allgemeine Empfehlungen zur Veränderung des individuellen Ernährungsverhaltens und in der Bevölkerung einfacher umsetzbar als personalisierte Empfehlungen für jedes einzelne Individuum. Dafür ist es erforderlich, eindeutige Metabotypen zu identifizieren, die Unterschiede in Ernährungs-Krankheits-Assoziationen und/oder in ihrer Reaktion auf Ernährungsinterventionen zeigen.

Diese Dissertation umfasst vier Manuskripte. Da das Metabotyping-Konzept in bisherigen Studien uneinheitlich angewandt wurde, gibt das erste Manuskript eine umfassende Literaturübersicht und macht Empfehlungen für die zukünftige Anwendung dieses Konzepts. Basierend auf diesen Empfehlungen beschreibt das zweite Manuskript die Identifizierung von umfassenden Metabotypen anhand zahlreicher Parameter, die in der großen bevölkerungsbasierten KORA-Studie (Kooperative Gesundheitsforschung in der Region Augsburg) zur Verfügung stehen. Diese Metabotypen waren signifikant mit dem (Neu-)Auftreten kardiometabolischer Krankheiten, wie beispielsweise Diabetes, verknüpft. Im dritten Manuskript wurden in einer Querschnittsanalyse Unterschiede in den Assoziationen von Ernährungsfaktoren und Diabetes zwischen Metabotyp-Subgruppen identifiziert, die relevant für zielgerichtete Ernährungsempfehlungen zur Prävention von Diabetes sein könnten. Außerdem deuten diese Ergebnisse darauf hin, dass bisher inkonsistente Ergebnisse in Ernährungs-Diabetes-Assoziationen teilweise durch interindividuelle metabolische Unterschiede hervorgerufen wurden. Die Inkonsistenzen in Ernährungs-Diabetes-Assoziationen wurden im vierten Manuskript mithilfe einer genaueren Aufgliederung des Glukosetoleranzstatus weiter untersucht, sodass neue Assoziationen zwischen Ernährungsfaktoren und Prädiabetes identifiziert werden konnten.

Zusammenfassend ist festzuhalten, dass umfassende Metabotypen in der KORA-Studie identifiziert wurden, deren Validität nun in anderen Populationen bestätigt werden sollte. Die Ergebnisse dieser Dissertation weisen zudem auf Unterschiede in Ernährungs-Diabetes-Assoziationen zwischen Metabotyp-Subgruppen hin. Deren Kausalität muss noch in prospektiven Beobachtungsstudien und Interventionsstudien überprüft werden.

Zusätzlich sollten sowohl andere ernährungsbezogene kardiometabolische Krankheiten, wie zum Beispiel Gicht, als auch Krankheitsvorstufen, wie Prädiabetes, in ihrer Assoziation mit Ernährung für unterschiedliche Metabotyp-Subgruppen untersucht werden. Auf Basis dieser Ergebnisse können effektive, zielgerichtete Ernährungsempfehlungen auf der Metabotyp-Subgruppenebene für die erfolgreiche Prävention und Behandlung von ernährungsbezogenen kardiometabolischen Krankheiten entwickelt werden.

Abbreviations

24HFL	24-hour food list
ADA	American Diabetes Association
BMI	Body mass index
BVSII	Bavarian Food Consumption Survey II
CI	Confidence interval
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DIfE	German Institute of Human Nutrition Potsdam-Rehbruecke (Deutsches Institut für Ernährungsforschung Potsdam-Rehbrücke)
EDTA	Ethylenediaminetetraacetic acid
EFPQ	European Food Propensity Questionnaire
EPIC	European Prospective Investigation into Cancer and Nutrition
FFQ	Food frequency questionnaire
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
KORA	Cooperative Health Research in the Region of Augsburg
MONICA	Multinational Monitoring of Trends and Determinants in Cardiovascular disease
MSM	Multiple Source Method
NCI	National Cancer Institute
NGT	Normal glucose tolerance
OGTT	Oral glucose tolerance test
OPTICS	Ordering Points To Identify the Clustering Structure
OR	Odds ratio
SD	Standard deviation
SNP	Single nucleotide polymorphism
SSB	Sugar sweetened beverages
T2DM	Type 2 diabetes mellitus
UDM	Undetected diabetes mellitus
WHO	World Health Organization

Tables and Figures

Tables

Figures

1. Introduction

The term ‘metabotype’ was first defined in 2000 by Gavaghan et al. [1] as “a probabilistic multiparametric description of an organism in a given physiological state based on analysis of its cell types, biofluids or tissues” in a rodent study. Afterwards, this concept was frequently used in further rodent studies based on biochemical parameters of the bio-specimens plasma, urine or feces to identify strain-specific metabotypes or metabolically homogeneous strain subgroups [2-13]. The latter were mainly found for age [9], sex [6,10-12], diet [7,13] and time of sample collection [5,8,12]. This concept allowed in rodent models, for example, the identification of specific biomarkers for diseases and the investigation of drug efficacy [1,3,4,6,13].

Later, ‘metabotyping’ or ‘metabolic phenotyping’ was also described in human studies as the process of identifying metabotypes or metabolic phenotypes, i.e. metabolically or phenotypically homogeneous subgroups of individuals [14-19]. It also arose especially in medical literature with the overall objective to better characterize individuals and, thus, to improve the prevention, diagnosis and management of diseases using optimized approaches [14,20]. Meanwhile, it was also applied in numerous nutritional studies, frequently in connection with ‘targeted’ nutrition. Targeted nutrition describes the delivery of tailored dietary recommendations at a group level of similar individuals, such as the metabotype subgroup level [14]. It expands the idea of personalized nutrition, which was first introduced in the 1970s and means tailoring dietary advice to an individual based on the individual’s characteristics in contrast to general advice for whole populations (**Figure 1**) [14,21,22].

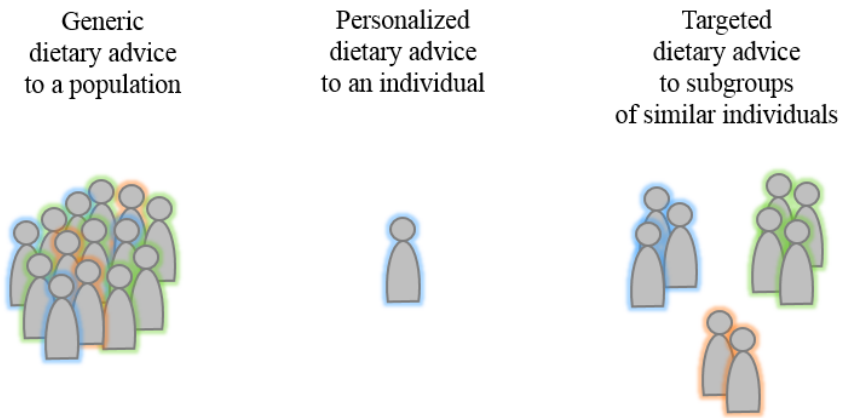

Figure 1: Schematic description of generic, personalized and targeted dietary advice. Generic advice is provided at the population level, personalized advice at the individual’s level and targeted advice at a subgroup level e.g. to metabolically similar individuals.

Personalized nutrition was originally designed based on genetic information [23], for example for the treatment of hypolactasia [24], celiac disease [25] or phenylketonuria [26]. By decoding the complete human genome sequence in 2001, great hopes have been placed on preventing and treating chronic diseases using personalized healthcare approaches such as personalized nutrition based on genetic information at the population level [27,28]. However, dietary intervention by genotype subgroups had only little success in most studies [29-31]. Thus, the identification of risk groups for diseases solely based on genetic data seems to be an incomplete approach for personalized nutrition recommendations, as genes explain only a small part of the risk for most chronic diseases [32,33]. In contrast, the dietary reference values of the German Nutrition Society (Deutsche Gesellschaft für Ernährung) are mainly tailored based on age, sex and particular circumstances like pregnancy and lactation [34]. Likewise, risk scores for diseases are often based on standard clinical markers, sex, age, anthropometric measures and lifestyle (e.g. smoking, alcohol consumption, diet, physical activity). An example is the German Diabetes Risk Score, which includes sex, age, height, weight, waist circumference, family history of diabetes mellitus, hypertension, smoking, physical activity and diet in the risk prediction of developing type 2 diabetes mellitus (T2DM) within the next 5 years [35]. As a consequence, a broader approach was established in the last few years for the delivery of personalized nutrition on three different levels [36]: level 1 describes personalized advice based on an individual's diet; level 2 personalized advice is additionally based on an individual's metabolic or phenotypic characteristics like anthropometry and biochemical parameters; and level 3 additionally includes genetic information in the development of personalized dietary recommendations [37].

In general, it has been shown that personalized nutrition is more effective than generic dietary advice regarding motivation for dietary behavior change and subsequent changes in phenotypic characteristics such as body weight, and may thus have distinct benefits for the individual's health and the health care system (**Figure 2**) [14,20,38-41].

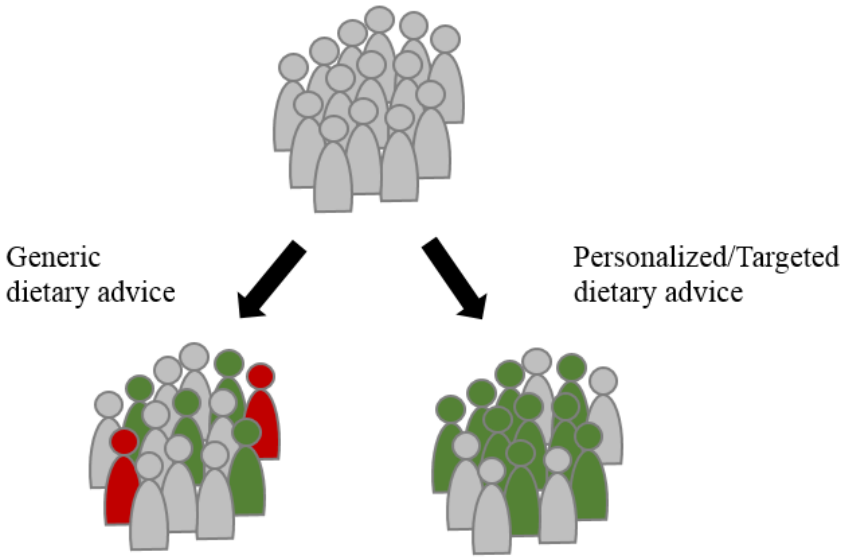

Figure 2: Schematic depiction of the possible effectiveness of generic dietary advice vs. personalized/targeted dietary advice regarding motivation for dietary behavior change and subsequent changes in phenotypic characteristics.
Grey: no effect; red: adverse effect; green: positive effect.

However, it is still not fully clear which level of personalization yields the greatest success [40,41]. It is well known that metabolic diversity originates from inter-individual differences in genetic, epigenetic, transcriptional and post-transcriptional factors, the gut microbiome and environmental factors like diet and further lifestyle factors [42,43]. This inter-individual metabolic variation leads to differences in nutrient requirements, responses to dietary and medical interventions as well as to differences in disease development between individuals [20,38,39,43]. Thus, including metabolic characteristics, as described in level 2 and level 3, in personalized nutritional advice seems to be a promising approach. Grouping metabolically similar individuals in the metabotyping approach may facilitate the consideration of metabolic characteristics in personalized nutrition and may enable the implementation at a population level, e.g. in the general practice setting [14].

Unhealthy dietary behavior and physical inactivity are strong risk factors for many chronic diseases such as T2DM, cardiovascular diseases and cancer [44,45]. Targeted dietary advice (possibly combined with targeted strategies for the increase of physical activity) at the metabotype subgroup level may help to prevent or delay the progression in healthy individuals as well as to manage chronic diet-related cardiometabolic diseases in patients [46]. So far, there were some studies defining patient subgroups of metabolic diseases by

metabotyping [47-55] or investigating associations between metabotypes and selected diseases in population-based samples [56-65]. However, only little is known about differential responsiveness of metabotypes to dietary interventions regarding disease-specific outcomes, which provide the basis for the development of targeted dietary disease prevention strategies [66-68].

To select one cardiometabolic disease, diabetes mellitus is an increasing global public health concern with an estimated prevalence of 8.8% in 2015 and a projected prevalence of 10.4% in 2040 in the adult population aged 20-79 years worldwide [69]. With 1.6 million deaths directly attributable to diabetes, this disease was ranked on position 7 of all causes of death worldwide in 2016 [70]. This implies a major challenge for healthcare systems all over the world and a large burden for the individual's health due to diabetes itself but also due to its frequent accompanied adverse health consequences [71]. These include macro- and microvascular diseases like myocardial infarction, stroke, kidney failure or blindness [71-73].

In Germany, there is also an increasing trend in the prevalence of known T2DM, the most common type of diabetes, from 8.5% in 2009 to 9.5% in 2015 [74]. Additionally, there is a large number of individuals estimated to have blood glucose levels above the threshold for T2DM but who have not been diagnosed [75,76]. Furthermore, there is an estimated 18 to 24% of the population in Germany with prediabetes, i.e. elevated blood glucose levels but still under the threshold for T2DM, who will very likely progress to T2DM [77-79]. Prediabetes was also shown to be associated with vascular complications [78,80].

The main reason for the rising T2DM and prediabetes prevalence is the growing proportion of older adults in the population following an unhealthy lifestyle [71]. Thus, targeted lifestyle modifications e.g. in physical activity or dietary behavior, are starting points to prevent a further increase in the prevalence of prediabetes and T2DM [78,81-83]. So far, little research has been conducted on associations between diet and prediabetes [84-88] and the relationship between diet and T2DM is not yet fully clear due to inconsistent findings [89,90]. One reason for these contradicting findings might be inter-individual metabolic differences that may cause differences in the responsiveness to diet between individuals regarding diabetes. The consideration of metabolic differences between individuals by investigating metabotypes in diet-diabetes associations may, as a result, help to understand these inconsistencies across studies. Finally, differences between metabotypes may be exploited to prevent diabetes by developing and establishing targeted dietary recommendations at the metabotype subgroup level.

Aim of the dissertation

Metabotyping is a relatively new concept, but already applied in various human studies [14]. This concept may be useful for the development of targeted dietary recommendations in disease prevention. The global aim of this dissertation is to give an overview of the previous literature on this topic and to investigate the identification of comprehensive metabotypes and their applicability in epidemiological and nutritional research in a large population-based study.

Therefore, the specific objectives of this dissertation are:

1. To summarize the existing literature on metabotyping in human studies in a literature review and to discuss the potential of metabotypes in the development of targeted and precise strategies in disease prevention and management.

2. To identify comprehensive and valid metabotypes in a large population-based cohort study using statistical clustering methods based on an extensive set of biochemical and anthropometric parameters. Subsequently, the aim is a detailed characterization of metabotypes by sociodemographic and lifestyle variables, metabolite profiles as well as cardiometabolic disease occurrence.

3. To analyze diet-disease associations exemplified by diabetes in a large cross-sectional population-based study in total and stratified by metabotype. This is to investigate whether inconsistent results across previous studies occurred partially due to metabolic differences between individuals.

Figure 3 shows an overview and the chronological sequence of these objectives.

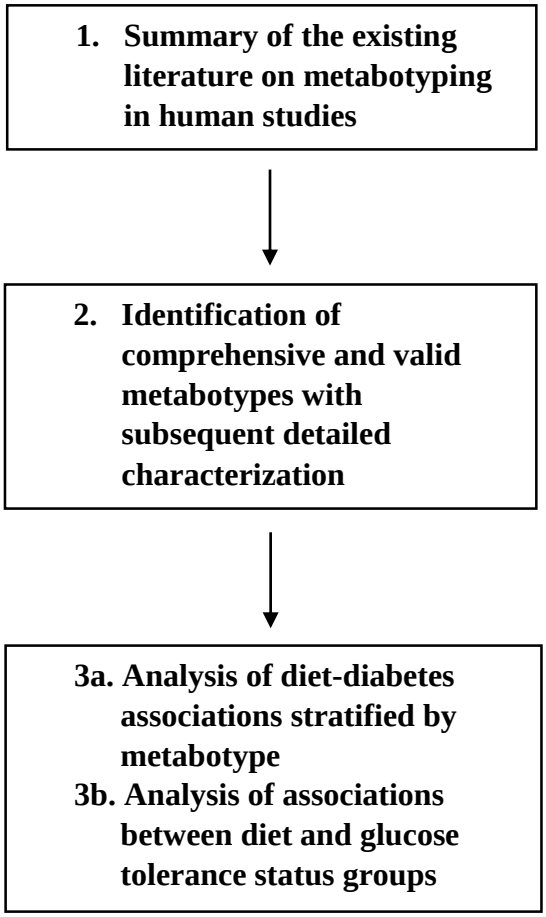

Figure 3: Overview and chronological sequence of the dissertation objectives

The objectives 1 and 2 each resulted in a manuscript with the doctoral candidate as the first author. Both manuscripts are published in international, peer-reviewed journals (see [91] and [92]).

The objective 3a resulted in a manuscript with the doctoral candidate as the first author and this manuscript is submitted to, but not yet accepted in an international, peer-reviewed journal. The objective 3b resulted in a manuscript with the doctoral candidate as the second author and this manuscript is accepted in an international, peer-reviewed journal.

The dissertation is based on these four manuscripts and the term ‘manuscript’ is used in the following for published, accepted and submitted manuscripts for reasons of simplification.

2. Methods

2.1 Literature search

The first manuscript gives a broad overview of the previous literature on the identification of metabotypes in humans until May 2016. Not all criteria have been met for a systematic literature review outlined, e.g., by the Cochrane Collaboration [93]. A comprehensive literature search was conducted using PubMed, Google and Google Scholar to find all articles on metabotyping in population-based samples, samples restricted to healthy individuals and samples restricted to patients with frequent, chronic diet-related metabolic diseases. Therefore, search terms from four different categories, namely 'Metabotyping keywords', 'Grouping keywords', 'Extended keywords' and 'Disease keywords' listed in **Table 1** were combined as follows:

- To find all articles on metabotyping in population-based samples and samples restricted to healthy individuals:
  "Metabotyping keyword" AND "Grouping keyword" (AND optional "Extended keyword");

- To find all articles on metabotyping in samples restricted to patients with metabolic diseases:
  "Metabotyping keyword" AND "Grouping keyword" AND "Disease keyword" (AND optional "Extended keyword").

Table 1: Search terms for the literature review

Metabotyping keywords	Grouping keywords	Extended keywords to consider underlying factors for metabolic variability [20]	Disease keywords
"metabotype"	"cluster"	"genetics"	"obesity"
"metabolic phenotype"	"pattern"	"genotype"	"adiposity"
"metabolomic phenotype"	"subgroup"	"single nucleotide	"metabolic syndrome"
"molecular phenotype"	"subtype"	polymorphism (SNP)"	"diabetes"
"clinical phenotype"	"cluster analysis"	"epigenetics"	"dyslipidemia"
"biochemical phenotype"	"principal component	"transcriptomics"	"hyperlipidemia"
"metabolic profile"	analysis"	"gut microbiota"	"hyperuricemia"
"metabolomic profile"		"enterotype"	"gout"
"metabolic pattern"			"hypertension"
"nutritional phenotype"			
"nutritype"			
"metabolome"			
"metabolomics"			
"metabolism"			
"metabolic response"			

Relevant articles were selected based on the following inclusion criteria:

- English original research articles on human studies not restricted to any study design or sample size;
- Metabotyping by means of statistical methods (and not only based on the combination of specific cut-off points of metabolites as e.g. in the definition of the metabolic syndrome);
- Metabotyping based at least on metabolites measured in blood or urine (but also in combination with genetic, epigenetic, transcriptomic, microbiome, anthropometric or lifestyle data);
- Metabotyping in samples of patients restricted to common chronic diet-related metabolic diseases. Reasons for the selection of these diseases were their worldwide growing prevalence and the possibility to prevent them by targeted dietary intervention [94].

Finally, 34 research articles were included in the literature review.

2.2 Study population

The analyses in the second, third and fourth manuscript are all based on data from the population-based epidemiological Cooperative Health Research in the Region of Augsburg (KORA) health research platform, which was established in the city of Augsburg and the two adjacent counties Augsburg and Aichach-Friedberg in Southern Germany. Originally started as the international World Health Organization (WHO) Multinational Monitoring of Trends and Determinants in Cardiovascular Disease (MONICA) Augsburg project and continued as KORA since 1996, four independent cross-sectional health surveys (S1-S4) were carried out between 1984 and 2001 in intervals of 5 years, including about 18000 participants aged 25-64 years in S1 and 25-74 years in S2-S4. The participation rate varied between 79% in the S1 survey and 67% in the S4 survey. [95]

In detail, the analyses of the three manuscripts were conducted using data of the KORA F4 (2006-2008) and FF4 (2013/2014) studies, which are both follow-ups of the KORA S4 health survey (1999-2001). In the baseline KORA S4 survey, a total of 6640 community-dwelling adults aged 25-74 years living in the study region were randomly selected from the local population registries using the same sampling method of the MONICA project [96]. Due to deaths before the KORA S4 examination (n = 51), moving to an unknown location (n = 40) or moving outside the study region (n = 132), or insufficient German language skills (n = 37), only 6380 individuals were eligible to participate in KORA S4. As 536 individuals were too ill or too busy, 1409 refused to participate, 172 could not be contacted and two had too many variables with missing values, the final KORA S4 sample size comprised 4261 individuals (67% of all eligible individuals).

Of these 4261 individuals, 3080 individuals and 2279 individuals, respectively, participated again in the subsequent first follow-up KORA F4 study after 7 years and in the second follow-up KORA FF4 study after further 7 years. A flow diagram of the study populations used in the second (KORA F4 and FF4), third (KORA FF4) and fourth (KORA FF4) manuscript is shown in **Figure 4**.

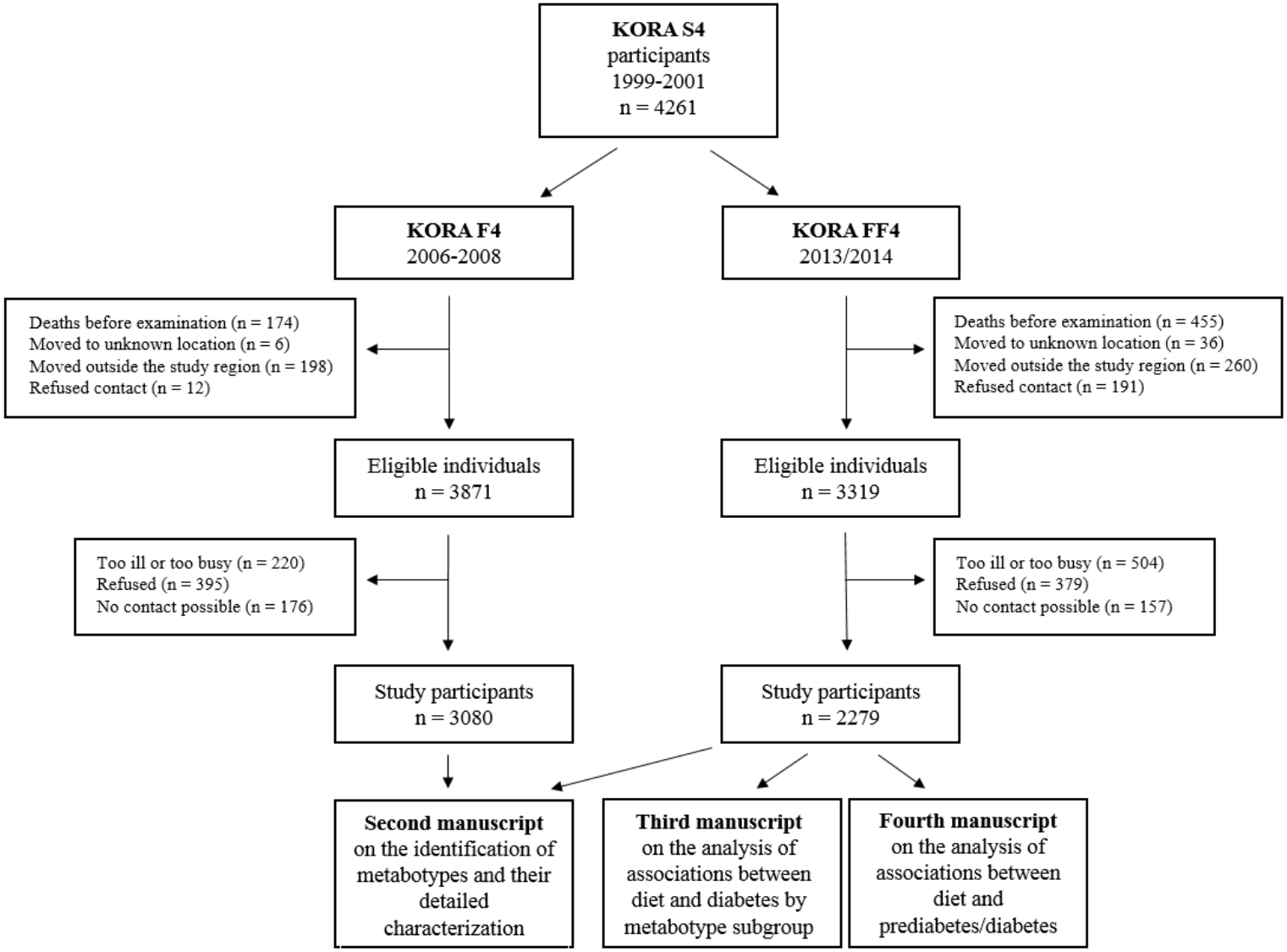

Figure 4: Flow diagram of the KORA F4 and FF4 study populations used in the second, third and fourth manuscript

Of the 3080 participants in KORA F4, 2161 participated again in KORA FF4, and consequently in both follow-ups. Reduction of the study population between both studies occurred due to 168 deaths before the KORA FF4 examination, 11 participants moving to an unknown location, 86 moving outside the study region and 67 refusing to be contacted. Of the remaining 2748 eligible participants for KORA FF4, 332 were too ill or too busy, 207 refused to participate and 48 could not be contacted. Thus, 118 individuals participated in KORA FF4 but not in KORA F4. Detailed information on the participation response has been provided elsewhere [97].

In both studies, all participants completed self-administered questionnaires and were invited to the study center for a standardized investigation program. This included a computer-assisted personal interview and physical examinations like blood examination, blood pressure and anthropometric measurements. All investigations were conducted by trained medical staff. A detailed description has been given previously [76]. Prior to the examinations, all participants gave their written informed consent. Both studies conform to the criteria laid down in the Declaration of Helsinki and were approved by the Ethics Committees of the Bavarian Medical Association and the Bavarian Chamber of Physicians in Munich.

2.3 Bio-specimens collection and measurement of biochemical parameters

In the second and third manuscript, metabotypes were identified, inter alia, based on biochemical parameters. Fasting (≥ 8 hours) venous blood samples were collected in both studies, KORA F4 and FF4, in the morning of study center visits generally from the left arm of the sitting participant by trained medical staff according to a standardized procedure. Participants were asked to avoid physical exercise and smoking before the examination. Whole blood was gathered in S-Monovette ethylenediaminetetraacetic acid (EDTA) tubes [Sarstedt, Nümbrecht, Germany], which were inverted two times, mixed by a universal shaker [Sarstedt] and cooled at 4°C. For serum collection, blood was drawn into S-Monovette serum gel tubes [Sarstedt] which were subsequently inverted two times and incubated upright for 30 minutes at room temperature (18-25°C) to attain complete coagulation. For plasma collection, blood was drawn into S-Monovette EDTA tubes [Sarstedt], which were inverted two times and mixed by a universal shaker [Sarstedt]. After centrifugation for 10 minutes (4000 rpm at 15°C), serum, plasma and urine were filled into Nunc cryotubes [Thermo Fisher Scientific, Waltham, MA, USA] and kept for a

maximum of 6 hours at 4°C. Then, all samples were stored frozen at −80°C until the standardized analysis of biochemical parameters.

All the biochemical parameters listed in **Table 2** were determined in KORA F4 participants. The measurement was described in more detail in the second manuscript, where all these 33 available fasting biochemical parameters were used in addition to body mass index (BMI) in the identification of comprehensive metabotypes in the KORA F4 study population. The new identification of metabotypes in KORA FF4 in the third manuscript was performed in addition to BMI based on a restricted set of 15 out of the originally 33 biochemical parameters in KORA F4, as only these were measured in this study. These parameters are highlighted in bold in **Table 2**. Details on biochemical parameter measurements in KORA FF4 participants are provided elsewhere [98].

Table 2: Biochemical parameters measured in KORA and applied in metabotyping

Whole EDTA blood	Serum	Plasma	Urine
- Glycated hemoglobin	**- Glucose**	- Lipoprotein(a)	- Creatinine
- Leukocytes	**- Total cholesterol**	- Apolipoprotein A-IV	- Albumin
- Average telomere length in leukocytes	**- High density lipoprotein**	- Afamin	
	- Total cholesterol/high density lipoprotein ratio	- Leptin	
	- Low density lipoprotein	- Sex-hormone-binding globulin	
	- Triglycerides	- Renin	
	- Uric acid	- Aldosterone	
	- Insulin	- Non-esterified fatty acids	
	- High-sensitive C-reactive protein	- Insulin-like growth factor-I	
	- Gamma-glutamyltransferase	- Insulin-like-growth-factor-binding-protein-3	
	- Glutamate-pyruvate transaminase		
	- Glutamate-oxaloacetate transaminase		
	- Alkaline phosphatase		
	- Cystatin C		
	- Interleukin-18		
	- Thyroid peroxidase antibodies		
	- Thyroid-stimulating hormone		
	- Free thyroxine		

All parameters were measured in KORA F4 participants; only the **bold parameters** were measured in KORA FF4 participants.

2.4 Assessment of anthropometric measures

The anthropometric measures BMI and waist circumference were used either as a clustering variable in the identification of metabotypes (second and third manuscript) and/or as a covariable in the analysis of diet-diabetes associations (third and fourth manuscript). In both studies, KORA F4 and FF4, weight (kg) was measured with an electronic scale, standing height (cm) with a stadiometer and waist circumference (cm) with an inelastic measuring tape in the middle between the lowest rib and the upper margin of the iliac crest. All three were determined in light clothing without shoes by trained investigators in a physical examination at the study center. The BMI was calculated as the ratio of weight in kilograms and height in meters squared. It was used both continuously and categorized according to the WHO thresholds into underweight (BMI < 18.5 kg/m^2), normal weight (BMI 18.5 kg/m^2 to < 25 kg/m^2), overweight (BMI 25 kg/m^2 to < 30 kg/m^2) and obesity (BMI ≥ 30 kg/m^2) [45].

2.5 Assessment of sociodemographic and lifestyle variables

Sociodemographic and lifestyle information of participants was used for the characterization of participants (second, third and fourth manuscript) and as covariables in the analysis of diet-diabetes associations (third and fourth manuscript). This information was assessed in the KORA F4 and FF4 studies through questionnaires and the computer-assisted personal interview. This included age (years), sex (male, female), education (years), physical activity (active, inactive) and smoking status (never, former, current).
Participants were classified as physically active if they reported leisure time physical activity during summer and winter and if they were active ≥ 1 hour per week in either of these seasons. Years of education included both the highest school degree and vocational training. They were classified into three groups either into < 10 years, 10 - < 12 years and ≥ 12 years for the detailed characterization of metabotypes in KORA F4 in the second manuscript or into < 10 years, 10 - < 13 years, ≥ 13 years for the characterization of participants or the adjustment of diet-diabetes associations in KORA FF4 in the third and fourth manuscript. For the sex-specific analyses in the fourth manuscript, education years were condensed to < 13 years and ≥ 13 years due to low frequencies in the < 10 years group within each sex group.

2.6 Assessment of chronic cardiometabolic diseases

In the second manuscript, the prevalence (in KORA F4) and incidence (in KORA FF4, during the 7-year follow-up from KORA F4 to FF4) of cardiometabolic diseases was compared between the metabotype subgroups identified in the KORA F4 study population. In the third and fourth manuscript, diabetes was used as the outcome variable and hypertension and family history of diabetes (father/mother; yes, no, don't know) were used as covariables in the analysis of diet-diabetes associations in KORA FF4 participants. The presence (yes/no) of diseases was assessed in KORA F4 and FF4 participants via questionnaire, the personal interview or in the physical examination at the study center. In detail, the systolic and diastolic blood pressure were measured automatically with an oscillometric digital blood pressure monitor three times at intervals of at least 3 minutes at the right arm of the sitting participant under standardized conditions by trained staff. Hypertension was specified as mean blood pressure ≥ 140/90 mmHg of the second and third measurement, or as the existence of a drug controlled, known hypertension. Dyslipidemia and hyperuricemia/gout were determined by the assessment of current specific medication intake. Lipid lowering medication in dyslipidemia treatment included statins, fibrates and other lipid modifying agents such as plant-based and alternative agents. Therefore, participants were asked to bring the product packaging of their medications used during the last 7 days before the study examination to the study center. The intake of medications was recorded through a database-supported computer software [99]. Previous cancer history and inpatient care of myocardial infarction and stroke were based on self-reports of participants. The self-reported diabetes diagnosis was supplemented by asking the participants for current antidiabetic medication intake, and both were verified by consulting the respective treating physician.

In the third and fourth manuscript, where diabetes was used as the outcome variable, glucose tolerance status was determined by a standard 75 g oral glucose tolerance test (OGTT; Dextro OGT, Boehringer Mannheim, Germany). It was performed in the morning between 7:00 and 11:00 a.m. in all participants, who were not previously diagnosed with diabetes and who were fasting overnight for at least 8 hours. By taking blood samples as described under 2.3, the glucose levels were measured in the fasting state and 2 hours after the OGTT. The glucose tolerance status was classified according to the 2003 American Diabetes Association (ADA) criteria [100] as described in **Table 3**. Prediabetes was defined as having isolated impaired fasting glucose (IFG), isolated impaired glucose tolerance (IGT) or combined IFG/IGT.

Table 3: Classification of the glucose tolerance status according to the ADA criteria

Glucose tolerance status	**ADA criteria** [100]
Normal glucose tolerance (NGT)	< 5.6 mmol/l fasting and < 7.8 mmol/l 2-h glucose
Impaired fasting glucose (IFG)	5.6–6.9 mmol/l fasting glucose
Impaired glucose tolerance (IGT)	7.8–11.0 mmol/l 2-h glucose
Undetected diabetes mellitus (UDM)	≥ 7.0 mmol/l fasting or ≥ 11.1 mmol/l 2-h glucose

2.7 Assessment of dietary intake

In the third and fourth manuscript, dietary intake data, which were extensively assessed in KORA FF4 participants (and not in KORA F4 participants), were investigated in their association with diabetes in the total study population, stratified by sex and stratified by metabotype subgroup.

Usual dietary intake data were collected via a food frequency questionnaire (FFQ) and up to three 24-hour food lists (24HFL) and were finally available from 1602 KORA FF4 participants (652 (40.7%) participants completed two 24HFLs; 826 (51.6%) completed three 24HFLs). The FFQ, which is based on the German version of the multilingual European Food Propensity Questionnaire (EFPQ) [101], was used to assess the type, usual consumption frequency and usual consumption amount of 148 food items over the last 12 months. Pictures were used to help the participants to estimate portion sizes. The 24HFLs are structured questionnaires restricted to > 300 food items and were used to determine the type of food consumed within the last 24 hours. Neither meals nor portion sizes were assessed. A detailed description of the 24HFLs is given elsewhere [102]. The participants were asked to complete the first 24HFL during their study center visit, and the FFQ as well as the other two 24HFLs at home, all preferably web-based to minimize the submission of incomplete data. Otherwise, paper versions of the questionnaires were offered. Concerning the 24HFLs, the study participants were contacted by the study center on 2 randomly selected days within the following 3 months after their study center visit to record their diet of the previous day. Two 24HFLs should be completed on working days and the other one on a weekend day.

The usual dietary intake of all food items was estimated using an advanced blended procedure, which is based on the National Cancer Institute (NCI) method [103,104] and the Multiple Source Method (MSM) [105] to minimize the error of traditional dietary assessment tools [106,107]. Therefore, information of both dietary assessment tools used

in KORA FF4, 24HFL and FFQ, was combined using a two-step approach. The consumption probability and the usual consumption amount on consumption days were estimated separately. In the first step, the consumption probability was calculated for each food item for each individual based on 24HFLs with logistic mixed models including the covariables age, sex, BMI, physical activity, smoking, education and food consumption frequency estimated by the FFQ. In the second step, the consumption amount on consumption days was calculated with models containing the same covariables to link both parts. As the 24HFLs do not assess portion sizes, the consumption amounts were predicted using data from 24-hour dietary recalls completed from participants of an external study, the Bavarian Food Consumption Survey II (BVSII). This is a population-based, cross-sectional study on dietary and lifestyle habits of the Bavarian population conducted in 2002/2003 [108]. The consumption amount for each food item on a consumption day was computed with the BVSII data using linear mixed models. These were adjusted for age, sex, BMI, physical activity, education and smoking. The usual consumption amount of each food item on a consumption day was then predicted for each individual in KORA FF4 using the β-estimates of the model parameters. Finally, the usual dietary intake of all food items was calculated for each participant by the multiplication of the estimated consumption probability of a certain food item by the estimated consumption amount on a consumption day. Thereof, 16 food groups and 21 food subgroups were built in accordance with the European Prospective Investigation into Cancer and Nutrition (EPIC)-Soft classification system [109]. Nutrients were estimated from the usual dietary intake data using the German Food Composition Database (Bundeslebensmittelschlüssel BLS 3.02). The 15 food groups and subgroups listed in **Table 4** as well as the nutrients alcohol (g/d) and fiber (g/d) were selected to be analyzed in the third and fourth manuscript in their association with diabetes, as they were previously shown to be associated either positively or negatively with T2DM [89,90,110,111]. Insoluble fiber (g/d) was additionally considered in the fourth manuscript as a substitute for whole grain intake, which is not contained in the EPIC-Soft classification system.

Table 4: Food groups and subgroups selected for the analysis of diet-diabetes associations

- Fruits (g/d)	- Poultry (g/d)	- Yogurt (g/d)
- Vegetables (g/d)	- Processed meat (g/d)	- Cheese (g/d)
- Potatoes (g/d)	- Eggs (g/d)	- Coffee (g/d)
- Total meat (g/d)	- Total dairy (g/d)	- Fruit and vegetable juice (g/d)
- Red meat (beef and pork, g/d)	- Milk (g/d)	- Sugar sweetened beverages (SSB, g/d)

Total energy intake (kcal/d) was used in both manuscripts as a covariable and was obtained from the usual dietary intake data considering all available food groups and subgroups.

In KORA F4, where extensive dietary intake data were not available, alcohol consumption was assessed differently for the description of the total KORA F4 study population and the metabotype subgroups in the second manuscript. It was based on the average daily consumption amount of alcoholic beverages (beer, wine and spirits) estimated using an effective recall-method concerning last week (last working day and last weekend) [112]. Alcohol consumption was grouped by sex into no (0 g/d for men and women), moderate ($> 0 - < 40$ g/d for men and $> 0 - < 20$ g/d for women) and high (≥ 40 g/d for men and ≥ 20 g/d for women).

2.8 Statistical analyses

All statistical analyses were carried out using the statistical software package RStudio versions 1.0.136 and 1.0.143 that use R versions 3.2.2 and 3.4.0 [R Development Core Team, 2010, http://www.r-project.org]. P-values < 0.05 were considered statistically significant.

Metabotyping

The concept of metabotyping was applied in the KORA F4 study population in the second manuscript and in the KORA FF4 study population in the third manuscript.
The aim was to identify comprehensive metabotypes. This is assumed to be achieved by a detailed metabolic characterization of individuals and, thus, by the inclusion of a broad range of biochemical parameters of different metabolic pathways and anthropometric measures in the grouping process [91,113]. Metabotyping was initially performed in the KORA F4 study population, where an extensive set of 34 biochemical and anthropometric

parameters was measured as described under section 2.3. A variety of statistical methods exists to group individuals and, thus, to identify metabotypes [113-118]. Thereof, the following unsupervised learning methods were considered [114]:

- k-means cluster analysis as a partitioning clustering algorithm and
- hierarchical cluster analysis, which are the both most commonly used methods;
- self-organizing maps and
- Gaussian mixture models as model-based clustering algorithms; but also
- 'Density-Based Spatial Clustering of Applications with Noise (DBSCAN)' and
- 'Ordering Points To Identify the Clustering Structure (OPTICS)' as density-based clustering algorithms, which are innovative machine learning methods preferably used in bioinformatics [119].

Each of these methods has individual assumptions and requirements regarding, for example, sample size and scale level of variables, and has consequently advantages and disadvantages. In addition, depending on the applied method, different data pre-processing steps like outlier exclusion or data transformations such as standardization are required. For example, k-means cluster analysis is assumed to be superior in large datasets compared to hierarchical cluster analysis, but the number of clusters needs to be determined in advance. [118]

K-means cluster analysis [120] was selected as the method of choice, as it resulted in significantly distinct biochemical parameter concentrations and anthropometric measures across the metabotypes identified and, thus, in metabolically homogeneous metabotypes. As the k-means cluster algorithm requires the prior determination of the cluster number [118], a series of cluster analyses with different cluster numbers ranging between two and eight clusters was performed and evaluated by 26 cluster validity criteria of the R package 'NbClust' version 3.0 [121]. Two and three clusters were identified as the most appropriate cluster numbers and three clusters were selected as the final solution due to a more precise metabolic classification of participants.

The analysis of diet-diabetes associations described in the third and fourth manuscript had to be performed in KORA FF4 participants because data on usual dietary intake were not available from KORA F4 participants. Despite some studies showing that plasma metabolite profiles of individuals remain quite stable over a few years [122,123], it could not be assumed that the KORA F4 participants would be clustered to the same metabotypes after 7-years of follow-up in KORA FF4 due to diseases or the decrease of metabolic flexibility during the aging process [124]. Thus, the metabotype assignment of KORA F4

participants was not simply transferred to KORA FF4 for the analysis of diet-diabetes associations stratified by metabotype in the third manuscript. In addition, the sample size would have become smaller as well, since not all KORA F4 participants took again part in KORA FF4 as described previously (section 2.2). Therefore, metabotypes were newly identified in KORA FF4 participants using the same methods as applied in KORA F4, however, based on a subset of 16 out of 34 originally used biochemical and anthropometric parameters in KORA F4 for reasons of availability (see 2.3). Repeating the identification of metabotypes in KORA F4 participants using the same 16 parameters as available in KORA FF4 resulted in a similar cluster allocation of individuals (1513 of 1729 or 87.5% of participants) and consequently in a similar separation of parameter values across clusters compared to the use of all 34 parameters. Therefore, it is assumed that the 16 parameters available in KORA FF4 participants were appropriate to identify comprehensive metabotypes again in KORA FF4 for which diet-diabetes analyses were stratified in the third manuscript.

Pre-processing of data was required in both studies prior to clustering of individuals with information on biochemical and anthropometric parameters available (1768 of 3080 KORA F4 participants; 2279 KORA FF4 participants) into metabotypes. This included the exclusion of participants not fasting at least 8 hours before blood sampling (KORA F4: n = 23; KORA FF4: n = 54) and participants with more than 10% missing values of the respective biochemical and anthropometric parameters used for clustering (KORA F4: n = 16; KORA FF4: n = 7). Multivariate imputation by chained equations from the R package ‘mice’ version 2.25 [125] was used to impute the remaining missing values and generated respectively five complete datasets with ten iterations each (KORA F4: n = 1729; KORA FF4: n = 2218), as complete data are required for clustering. Further, all biochemical and anthropometric parameters used for clustering were previously z-transformed to standardize different scales and units and, thus, to avoid bias [115,116]. As the number of clustering variables did not exceed the proposed threshold of one per ten individuals due to the large study populations, there was no problem of over-adjustment so that data reduction methods such as principal component analysis, multiple correspondence analysis or factor analysis were not necessary [126]. Likewise, as the aim was to identify comprehensive metabotypes not optimized for the prediction of a specific disease or outcome, the clustering variables were not restricted to disease-specific parameters using methods such as the random forest algorithm [127].

Finally, k-means cluster algorithm of the R package ‘miclust’ (multiple imputation in cluster analysis) version 1.2.5 [128] was performed simultaneously on the respective five imputed datasets based on 34 (in KORA F4) or 16 biochemical and anthropometric parameters available (in KORA FF4) to establish three clusters respectively in KORA F4 and FF4. This is illustrated in **Figure 5** for the identification of metabotypes in KORA F4 participants in the second manuscript. These analyses were repeated in KORA F4 limited to a subsample of older participants aged ≥ 60 years to examine the effect of age on the clustering results.

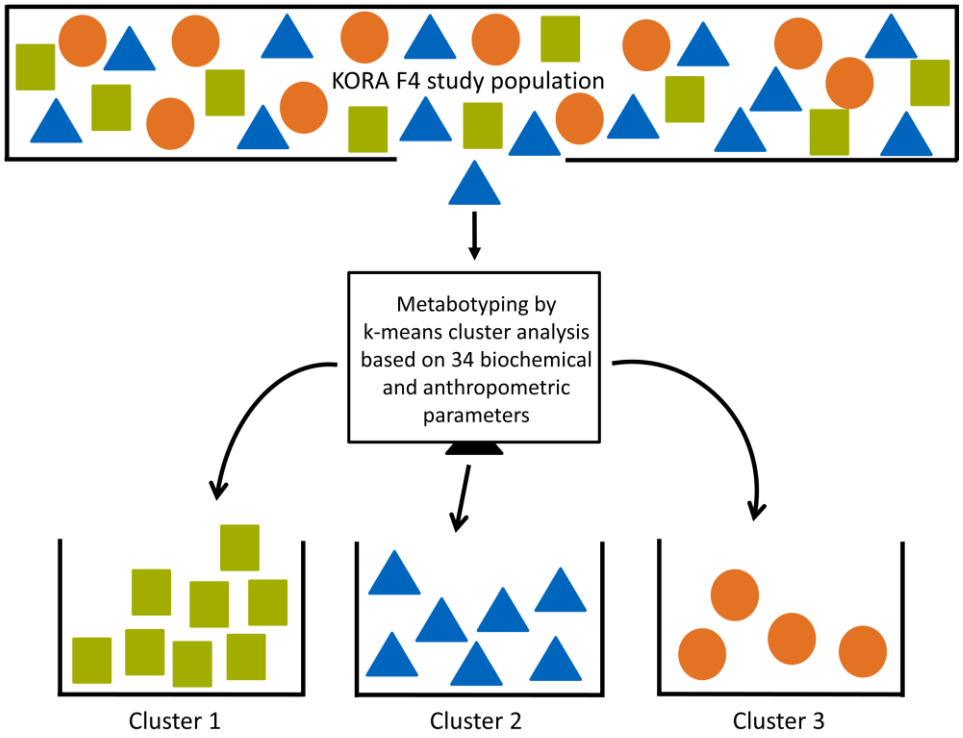

Figure 5: Identification of three clusters representing metabotypes in the KORA F4 study population (described in detail in the second manuscript).
[Riedl A, Wawro N, Gieger C et al. (2018) Identification of comprehensive metabotypes associated with cardiometabolic diseases in the population-based KORA study. Mol Nutr Food Res 62(16):e1800117. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.] [92]

Descriptive statistics

In the second, third and fourth manuscript, the total KORA F4 and FF4 study populations and subgroups by metabotype, sex and glucose tolerance status were characterized depending on the respective study aim in detail by socio-demographic and lifestyle variables, anthropometric and metabolic characteristics, cardiometabolic disease prevalence and incidence or usual dietary intake.

Due to deviations from normal distribution, medians, 25^{th} and 75^{th} percentiles were calculated for continuous data. Total (n) and relative frequencies (%) were computed for categorical variables. Differences in characteristics between subgroups of individuals were detected by Kruskal-Wallis test (followed by Kruskal-Wallis post-hoc test with Bonferroni correction) for the non-normally distributed continuous data. Pearson's chi-squared test or Fisher's exact test (followed by chi-squared post-hoc test with Bonferroni correction) were used for the categorical variables.

Analysis of diet-diabetes associations

Various usual dietary intake variables (listed under section 2.7) were analyzed with respect to their cross-sectional association with glucose tolerance status in KORA FF4: in the total KORA FF4 study population in the third and fourth manuscript, stratified by sex in the fourth manuscript (n = 2279), as well as stratified by metabotype subgroup in the third manuscript (using the newly identified metabotypes based on 16 biochemical and anthropometric parameters in 2218 KORA FF4 participants). Prior to the analyses, participants with type 1 diabetes (third manuscript: n = 4; fourth manuscript: n = 6), unclear glucose tolerance status (third manuscript: n = 67; fourth manuscript: n = 93) or missing dietary intake data (third manuscript: n = 628; fourth manuscript: n = 638) or covariable data (third manuscript: n = 2; fourth manuscript: n = 0) were excluded. Thus, the final samples for the analyses of diet-diabetes associations comprised n = 1542 individuals in the fourth and n = 1517 individuals in the third manuscript.

The glucose tolerance status was classified in four categories (normal glucose tolerance (NGT, = reference), prediabetes, undetected diabetes mellitus (UDM), prevalent T2DM) in the fourth manuscript. These were condensed to two categories (NGT/prediabetes (= reference), UDM/prevalent T2DM) in the third manuscript to increase the sample sizes for the analyses stratified by metabotype subgroup. In addition, the metabotype clusters 1 and 2 had to be combined and analyzed together compared to cluster 3 in these analyses

due to a small number of diabetes cases in both clusters. All food groups and food subgroups were rescaled by 50 g/d and fiber by 10 g/d in the third manuscript. They were divided by their respective standard deviations (SD) in the fourth manuscript to facilitate the comparability of risk estimates between different dietary intake variables. In both manuscripts, alcohol was divided by sex into low (< 5 g/d for men and < 2 g/d for women), moderate (≥ 5 - < 20 g/d for men and ≥ 2 - < 10 g/d for women) and high intake (≥ 20 g/d for men and ≥ 10 g/d for women) according to the nutrient reference values of the German Nutrition Society (Deutsche Gesellschaft für Ernährung) to distinguish between moderate and heavy alcohol intake [34]. Low alcohol consumption defined as < 2 g/d or < 5 g/d was used instead of no consumption (0 g/d), as the calculation of usual dietary intake does not allow for an intake of 0 g/d.

Depending on the number of categories of the outcome variable 'glucose tolerance status', multinomial logistic regression (R function 'mlogit') was performed in the fourth manuscript and binary logistic regression (R function 'glm') was performed in the third manuscript [129]. In both manuscripts, for each dietary intake variable, two regression models with different sets of covariables were fitted. These were selected based on theoretical considerations and previous literature on diet-diabetes associations [89,90]. The basic model was adjusted for age, sex and energy intake. The fully adjusted model was additionally adjusted for waist circumference, BMI (continuously, only in the fourth manuscript), family history of diabetes, physical activity, smoking, education, hypertension and metabotype (dichotomized into combined clusters 1 and 2, and cluster 3; only in the third manuscript). The regression analyses were performed and odds ratio (OR) estimates with 95% confidence intervals (CI) calculated for the respective dietary intake variables in the total study population (third and fourth manuscript), stratified by sex (fourth manuscript) and for two distinct metabotype subgroups (cluster 1 and 2 compared to cluster 3; third manuscript). To investigate metabolic influences on diet-diabetes associations, likelihood ratio tests were applied to identify possible interaction effects (by p-values for interaction < 0.05) between metabotypes and the respective dietary intake variables in the fully adjusted model for the total study population in the third manuscript. Additionally, to consider age-specific metabolic differences, the analyses stratified by metabotype were performed in a subpopulation of older adults aged ≥ 60 years.

3. Short description of publications and publication manuscripts

3.1 Publication 1: Metabotyping and its application in targeted nutrition: an overview

Inter-individual metabolic diversity originates from intrinsic factors such as genetics and from environmental factors such as lifestyle and diet. Due to these metabolic differences, individuals show different nutrient requirements and responses to dietary interventions or medical treatments. To better understand the role of metabolic characteristics in nutritional and medical research, the concept of metabotyping emerged. It describes the process of clustering metabolically similar individuals into so-called metabotypes. Such metabolically homogeneous subgroups are proposed to enable targeted dietary and medical approaches in disease prevention at a group level, which are more feasible and potentially similarly effective as personalized strategies at the individual's level in large populations. So far, there are various human studies on the identification of metabotypes in nutritional and medical research. A total of 34 research articles on metabotyping, which was implemented with statistical methods in population-based samples and samples restricted to healthy individuals or patients with chronic metabolic diseases, were identified in a comprehensive literature search and summarized in this manuscript. Large heterogeneities were discovered in the metabotyping process between these studies. This manuscript gives a broad overview on the use of metabotyping in the previous literature and discusses the potential of metabotypes in the development of targeted and precise strategies in disease prevention and management.

The doctoral candidate conceived the literature overview, conducted the literature search and drafted the manuscript.

[Riedl A, Gieger C, Hauner H, Daniel H, Linseisen J (2017) Metabotyping and its application in targeted nutrition: an overview. *Br J Nutr* 117(12):1631-1644.] [91]

3.2 Publication 2: Identification of comprehensive metabotypes associated with cardiometabolic diseases in the population-based KORA study

Due to the broad definition of the term ‘metabotype’ as metabolically or phenotypically homogeneous subgroup, previous studies on metabotyping showed a large heterogeneity. This included, among other things, variations in the applied statistical methods, the number of identified metabotypes, and particularly in the amount and type of metabolic and/or phenotypic variables included in the metabotyping process. To obtain comprehensive metabotypes by the inclusion of numerous parameters of diverse metabolic pathways and anthropometry, the large German population-based KORA F4 study (2006-2008) was used. It has a large set of 34 biochemical and anthropometric parameters available. By the application of k-means cluster analysis based on these parameters, 1729 participants aged 32-77 years were divided into three significantly metabolically different clusters. These metabotype subgroups showed clear differences in the prevalence (in KORA F4) and incidence (within the 7-year follow-up between KORA F4 and FF4, determined in KORA FF4 in 2013/2014) of cardiometabolic diseases, namely hypertension, T2DM, hyperuricemia/gout, dyslipidemia, myocardial infarction, stroke and cancer. In brief, cluster 3 showed the most unfavorable metabolic characteristics and the highest cardiometabolic disease occurrence at KORA F4 and FF4, followed by cluster 2, and finally by cluster 1 with the most favorable metabolic characteristics and the lowest cardiometabolic disease prevalence and incidence. By identifying differences between these metabotypes in diet-disease associations in further observational studies or dietary intervention studies, these comprehensive metabotypes may be used for the development of targeted and precise dietary disease prevention and management strategies at this metabotype group level. Especially cluster 3 may benefit from it due to the highest prevalence and incidence of cardiometabolic diseases in this cluster.

The doctoral candidate conceived and conducted the data analyses, interpreted the data, and drafted the manuscript.

[Riedl A, Wawro N, Gieger C, Meisinger C, Peters A, Roden M, Kronenberg F, Herder C, Rathmann W, Völzke H, Reincke M, Koenig W, Wallaschofski H, Hauner H, Daniel H, Linseisen J (2018) Identification of comprehensive metabotypes associated with cardiometabolic diseases in the population-based KORA study. *Mol Nutr Food Res* 62(16):e1800117.] [92]

3.3 Submitted manuscript: Diet-diabetes associations stratified by metabotype

The relationship between diet and diabetes is still not fully clear, as shown by inconsistent results in the previous literature. Inter-individual metabolic differences may be a major reason for different responses to diet between individuals. To examine this hypothesis, various food intake variables were investigated in their cross-sectional association with UDM/prevalent T2DM for two metabolically homogeneous subgroups of the large German population-based KORA FF4 study population (2013/2014). These were identified using k-means cluster analysis based on 16 available biochemical and anthropometric parameters. The negative association of fruit intake and the positive associations of consumption of total meat, processed meat and sugar sweetened beverages (SSB) with UDM/prevalent T2DM found in the total KORA FF4 study sample of 1517 individuals aged 38-87 years, remained significant only in one of two metabotypes after stratification. Total meat (OR: 1.67, 95% CI: 1.04 – 2.67) and processed meat (OR: 2.23, 95% CI: 1.24 – 4.04) were positively associated with UDM/prevalent T2DM in the rather favorable metabotype; fruits was negatively (OR: 0.83, 95% CI: 0.68 – 0.99) and SSB was positively (OR: 1.21, 95% CI: 1.09 – 1.35) associated with UDM/prevalent T2DM in the rather unfavorable metabotype (respectively per increase in the intake amount by 50 g/d). In addition, the association between SSB and UDM/prevalent T2DM was significantly different between the metabotype subgroups indicated by a p-value for the interaction of 0.01. These results suggest a metabolic influence on diet-diabetes associations, but prospective and interventional studies are needed to investigate causal relationships. This may be relevant for the development of targeted dietary strategies at the metabotype level for diabetes prevention.

The doctoral candidate conceived and conducted the data analyses, interpreted the data, and drafted the manuscript.

Riedl A, Wawro N, Gieger C, Meisinger C, Peters A, Rathmann W, Koenig W, Strauch K, Quante A, Thorand B, Huth C, Daniel H, Hauner H, Linseisen J. Diet-diabetes associations stratified by metabotype. Submitted to: *Eur J Nutr* (2018 Sep 03).

3.4 Accepted manuscript: Differential associations between diet and prediabetes or diabetes in the KORA FF4 study

Diabetes represents a major public health burden, which may be reduced by lifestyle changes such as in dietary behavior. However, besides inconsistent results in diet-diabetes associations in the previous literature, even less is known about the association between diet and prediabetes. To further clarify these relationships, numerous dietary factors including food groups, food subgroups and nutrients were investigated in their association with prevalent T2DM, previously UDM and especially prediabetes in a large population-based sample of 1542 participants of the German cross-sectional KORA FF4 study (2013/2014). Low fruit intake, high intake of total meat, processed meat and SSB, and moderate alcohol intake showed significant associations with prevalent T2DM and/or UDM either in the total study population or in at least one of both sex-stratified analyses. Coffee was negatively (OR: 0.86, 95% CI: 0.76 – 0.98) and heavy alcohol positively (OR: 1.78, 95% CI: 1.23 – 2.59) associated with prediabetes (per increase in the intake amount by the respective SD) but remained only significant for men after stratification by sex. The causality of the newly discovered associations between diet and prediabetes needs to be confirmed in prospective studies. These findings may be useful for the early prevention of diabetes, as prediabetes was shown to be a strong risk factor for diabetes and is already associated with adverse health effects.

The doctoral candidate was strongly involved in the formulation of the research question and the design and interpretation of the data analyses. Finally, the candidate revised and approved the final manuscript.

Breuninger TA, Riedl A, Wawro N, Rathmann W, Strauch K, Quante A, Peters A, Thorand B, Meisinger C, Linseisen J. Differential associations between diet and prediabetes or diabetes in the KORA FF4 study. Accepted in: *J Nutr Sci* (2018 Nov 09).

4. Discussion

Main findings

The first manuscript gives a unique overview of available human studies using the concept of metabotyping in the field of nutrition and in medical research on frequent chronic metabolic diseases with relation to diet. Due to the wide definition of the metabotyping concept, i.e. the identification of metabolically or phenotypically homogeneous subgroups in the previous literature [14-19], this review included a total of 34 rather heterogeneous studies. These were conducted in different populations worldwide with great variations in sample size, age range and sex. Even more importantly, studies showed large differences in the metabotyping process, i.e. in the statistical methods and the parameters used in the identification of metabotypes. Consequently, different kinds of metabotypes were established. These are difficult to compare and, thus, it is difficult to postulate coherent and meaningful conclusions for the identification of metabotypes.

Over the time, the assumption arose that the inclusion of various parameters from different metabolic pathways and additional phenotypic data in the metabotyping process may enable a more detailed characterization of individuals and, thus, the identification of more comprehensive metabotypes [91,113]. Consequently, the aim of the second manuscript was to identify comprehensive and valid metabotypes initially not related to any disease in the large German population-based KORA F4 study, wherein an extensive set of biochemical and anthropometric data was available. These metabotypes were characterized in detail and associations with cardiometabolic diseases were tested. Applying k-means cluster analysis, one of the most common grouping methods used for metabotyping, based on BMI and all 33 biochemical parameters available in KORA F4 participants, three metabolically distinctly different clusters were identified. Variation of metabolites within each cluster was lower than between clusters, supporting the aim of generation of metabolically rather homogeneous clusters representing valid metabotypes. In brief, cluster 3 represented a metabotype with rather unfavorable metabolic characteristics, followed by cluster 2 describing an intermediate metabotype and cluster 1 with the most favorable metabolite profile. These were significantly associated with the prevalence (in KORA F4) and incidence (in the 7-year follow-up between KORA F4 and FF4, determined in KORA FF4) of cardiometabolic diseases decreasing clearly from cluster 3 to cluster 2 and finally to cluster 1. Thus, a ‘high-risk’ cluster 3, an ‘intermediate-risk’ cluster 2 and a ‘low-risk’ cluster 1 for cardiometabolic disease occurrence were identified. By this metabotype-based risk classification, additionally physically active

individuals, younger individuals or individuals with normal BMI and, thus, apparently healthy individuals could be identified as 'at risk' and may, thus, benefit from targeted disease prevention strategies. This is the first study that examined the cross-sectional and longitudinal associations of metabotypes with such a large number of diseases. Thereby, it expanded the existing studies investigating only single or few diseases mostly such as diabetes, hypertension or the metabolic syndrome [56-65].

To investigate if these comprehensive metabotypes may be useful for the development of targeted dietary strategies in disease prevention, diet-disease associations were analyzed stratified by metabotype subgroup by the example of diabetes in the third manuscript. Differences across metabotypes may also contribute to the understanding of previously not fully conclusive results in diet-diabetes associations in the literature [89,90]. Such differences in associations with UDM/prevalent T2DM between metabotypes were found in the KORA FF4 study for the intake of fruits, total meat, processed meat and SSB, which were all found to be significantly associated with UDM/prevalent T2DM in the total study population. After stratification by metabotype, higher intake of total meat and processed meat was significantly associated with UDM/prevalent T2DM only in the more favorable metabotype. This metabotype combined the 'low-risk' cluster 1 and the 'intermediate-risk' cluster 2 due to the low diabetes prevalence in these groups. On the contrary, for lower intake of fruits and higher intake of SSB significant associations with UDM/prevalent T2DM were only found in the 'high-risk' cluster 3, representing a more unfavorable metabotype. The association between SSB and UDM/prevalent T2DM was further significantly different between both metabotypes. These results show that inconsistent findings of previous studies, mainly for SSB, may partially be due to inter-individual metabolic differences. These results may thus be relevant for the development of targeted dietary advice in diabetes prevention, which needs to be confirmed in interventional and prospective studies. This holds especially for the 'high-risk' cluster 3, which represents an unfavorable metabotype concerning its metabolic and anthropometric characteristics, the high prevalence of prediabetes, UDM and T2DM, and an unhealthier dietary pattern with higher intake of energy, total meat, red meat, processed meat and SSB and lower intake of vegetables, total dairy, milk, yogurt and fruit and vegetable juice. However, the more favorable metabotype may also benefit from targeted strategies due to its high number of prediabetic individuals, which are known to progress very likely to diabetes [77-79] and may already be affected by vascular complications [78,80].

The previously inconsistent findings in diet-diabetes associations were further clarified in the fourth manuscript by the analysis of associations between diet and prediabetes, for which little is known so far [84-88]. Thus, it was the first time to our knowledge that prediabetes was investigated in regards to its association with such a large number of food items. After further subdividing the glucose tolerance status (into NGT, prediabetes, UDM and prevalent T2DM), the associations of fruits, total meat, processed meat and SSB with UDM/prevalent diabetes found in the third manuscript were also seen for the subgroups UDM and/or prevalent T2DM in the total KORA FF4 study population. In addition, sex-specific differences were found, e.g. an additional positive association between moderate alcohol and UDM only in men. New insights were gained concerning coffee consumption being inversely and high alcohol consumption being positively associated with prediabetes. To examine also the role of metabolic characteristics in diet-prediabetes associations, larger studies are needed to enable stratification by metabotype subgroup with sufficient sample size.

As metabotypes were identified heterogeneously in the previous literature, as shown in the first manuscript, the most important aspects in metabotyping are discussed in the next sections to be considered in future studies.

Role of population characteristics in the identification of metabotypes

Studies on metabotyping were conducted in countries all over the world. Due to metabolic differences between populations by intrinsic characteristics such as genetics and environmental factors such as lifestyle [20,130-132] it is not surprising that different metabotypes were identified in particular between Western countries and East Asian countries. Whether metabotypes may be transferable between different ethnicities remains to be investigated. However, metabotypes identified in different Western countries were more comparable [130] and this may enable the identification and subsequent reproduction and validation of metabotypes in different Western populations.

To achieve sufficient sample sizes for subsequent analyses, both men and women were included simultaneously in the identification of metabotypes in KORA F4 and FF4, as done in most previous studies on metabotyping. Despite metabotyping without metabolites showing substantial sex-specific differences, such as steroid hormones or branched chain amino acids [130,133,134], the metabotypes identified in the second and third manuscript differed by sex with higher proportions of men in the more unfavorable metabotypes. The subsequent analyses on diet-diabetes associations stratified by metabotype were adjusted

for sex, but not further stratified by sex due to resulting insufficient sample sizes and due to missing significant interaction effects between sex and dietary factors.

Likewise, participants with all adult age ranges were taken together for the identification of metabotypes in KORA F4 (32-77 years) and FF4 (38-87 years) participants to ensure sufficient sample sizes. However, it is well known that aging is associated with large changes in metabolic characteristics and metabolic plasticity, which describes the flexibility of the metabolism to react to certain influences [124]. Comparable to sex (described above), age was not used as a clustering variable in the metabotyping process in the second and third manuscript, i.e. clusters were not built by direct influence of age but indirectly based on anthropometric and biochemical parameters associated with age. Nevertheless, the metabotypes identified were significantly different in their median age with the highest in the most unfavorable metabotype. Similar clustering results were obtained by limiting the sample to participants aged ≥ 60 years. Subsequent analyses on diet-diabetes associations stratified by metabotype were adjusted for age. Further, a sensitivity analysis restricted to older adults aged ≥ 60 years in the third manuscript revealed similar results in diet-diabetes associations compared to the whole study population aged 38-87 years, a finding which attenuates the importance of age for metabotyping.

Parameter selection for the identification of metabotypes

Previous studies showed large differences in the type and number of clustering variables used in the metabotyping process due to the broad definition of the term ‘metabotype’. However, the selection of clustering variables is the most important part of metabotyping, as these are responsible for the grouping of individuals and finally for the established metabotypes. So far, there is disagreement on the need of a stricter metabotype definition regarding the clustering variables. On the one hand, a uniform definition may facilitate the comparability of metabotypes identified in different studies. On the other hand, depending on the respective study aim, different metabolites are important. For example, specific lipids seem to be sufficient for the identification of plasma lipoprotein or plasma fatty acid clusters [56,58,135], specific response variables seem to be relevant in the metabotype identification by means of dietary interventions [15,136,137] or disease-specific variables seem to be adequate in diagnosing or subgrouping of patients [52]. However, it could be suggested to develop more specific sub-definitions of metabotypes such as for the

metabolism of lipids or carbohydrates. A stricter definition may also be interesting for the identification of comprehensive metabotypes. It is presumed that a more precise characterization of individuals by means of extensive metabolic parameters from different metabolic pathways and additional phenotypic data such as anthropometry may enable the identification of more comprehensive metabotypes [91,113]. Thus, all available 33 (KORA F4) or 15 (KORA FF4) biochemical fasting parameters were included in addition to BMI in the metabotyping process in the second and third manuscripts, which were among the first to use such a large set of parameters in the identification of metabotypes. However, this could be further expanded in future investigations using full ^{1}H NMR spectra or '-omics' data, which are becoming more often available [43,138-140]. These include primarily metabolomics data but also genomics, epigenomics, transcriptomics, proteomics or gut microbiome data, which are known to influence metabolism. Such extensive data were already included in few studies [54,65,67,68,141,142]. Further, besides fasting metabolites, metabolic response data to interventions might also be included, as some metabolic differences between individuals could only be identified by challenging the metabolism [143]. An example of this is the 2-h glucose value after an OGTT. There are some studies identifying metabotypes based on response values of parameters to dietary interventions [15,91,136,137,141,144].

Otherwise, comparing the classification of KORA F4 participants, by using on the one hand all 34 parameters and on the other hand only the subset of 16 parameters available in KORA FF4 participants in metabotyping, revealed relatively similar results. This indicates that a smaller subset of potentially routinely measured parameters may be already sufficient for the identification of comprehensive metabotypes. It remains to be investigated, which parameters are most important for the identification of comprehensive and distinct metabotypes. Based on these, a risk score might be developed to facilitate the implementation of targeted dietary recommendations in large populations. Therefore, the measurement of fasting metabolite data is easier than metabolic response data to interventions in the general population. Further, individual metabotypes seem to be relatively robust towards metabolite variations due to diurnal time, stress, latent illness and sample handling [20,132,145-147] and were shown to be quite stable over at least a few years [122,123]. This may enable the development of targeted disease prevention strategies on a reliable basis.

Choice of a statistical strategy for the identification of metabotypes

As described in chapter 2.1, the literature review in the first manuscript included many studies, which used different statistical methods in the identification of metabotypes. All studies identifying metabotypes based on the combination of cut-off values of anthropometric and biochemical parameters were excluded, as statistical methods provide more precise cluster solutions. For the identification of comprehensive metabotypes initially independent of an outcome or disease, as it was the aim in the second manuscript, unsupervised learning algorithms are preferred to supervised learning algorithms like regression analysis [113-119]. There are numerous unsupervised learning algorithms available, of which several were tested (see 2.8), including k-means and hierarchical cluster analyses as the most prominent ones [14,91,114]. With all of them having their individual strengths and weaknesses, there are different views on the most suitable method to get the best cluster separation [114,118]. Therefore, the most appropriate method and its specifically required pre-processing steps have to be chosen individually depending on the structure and properties of the data. For the large data sets available in KORA F4 and FF4, k-means cluster analysis [120] was used and resulted in the identification of metabolically distinctly different clusters.

Application of metabotypes in nutritional research

Despite the novelty of the metabotyping concept, numerous human studies were conducted on the identification and detailed characterization (by sociodemographic and lifestyle factors, metabolic profiles and disease occurrence) of metabotypes in healthy individuals or patients of various diseases [14,91]. Individuals clustered into metabotypes may benefit from targeted and precise disease prevention or management strategies, which are supposed to be, like personalized healthcare strategies, higher in their effectiveness than generalized approaches [14,20,38-41]. As an unhealthy diet is a strong risk factor, especially for cardiometabolic diseases [44,45], dietary strategies at the metabotype subgroup level may be relevant for disease prevention and management [46]. However, only little is known so far about differential responsiveness of metabotypes to dietary interventions regarding disease-specific outcomes, which provide the basis for the development of targeted dietary strategies [66-68]. In brief, O'Sullivan et al. [66] detected a metabotype responsive to vitamin D supplementation regarding metabolic syndrome markers, Vázquez-Fresno et al. [67] found metabotypes responsive to red wine

polyphenols in cardiovascular risk patients, and Moazzami et al. [68] could show different postprandial metabolic responses to interventions with different types of bread between metabotype subgroups. Besides these few intervention studies investigating for a differential responsiveness of metabotypes to single dietary factors, the third manuscript explored for the first time the possible influence of metabolic characteristics by means of metabotypes on the associations of various food groups, food subgroups and nutrients with diabetes in the large population-based cross-sectional KORA FF4 study. Diabetes was chosen as the outcome due to its rising prevalence worldwide [69], its preventability or onset delay by dietary behavior change [78,81-83], and as inconsistent results in previous studies on diet-diabetes associations [89,90] might be partially triggered by metabolic differences between individuals. Further, this research might be deepened by the investigation of diet-prediabetes associations stratified by metabotype, which may be feasible in larger studies with sufficient sample sizes. In addition, as the comprehensive metabotypes of the second and third manuscripts were identified initially independent of diseases, it may also be interesting to test associations of dietary factors with other diet-related diseases such as hypertension, hyperuricemia/gout or dyslipidemia for differences between metabotypes. Such differences may be used for the development of targeted dietary recommendations in disease prevention and management, which was carried out so far only by O'Donovan et al. [57,148]. They report a high agreement between targeted dietary recommendations, given at a metabotype subgroup level according to decision trees, and personalized recommendations at the individual's level. This indicated a high effectiveness of targeted dietary strategies by a simultaneous simplified implementation in large populations, whereby the individual's health status may be improved and enormous costs for the health care systems may be avoided. However, more studies are needed to further strengthen the development of targeted dietary recommendations for disease prevention and management at the metabotype subgroup level and to confirm their effectiveness. This could be further increased by other targeted strategies concerning e.g. physical activity, which is also a well-known risk factor for cardiometabolic diseases [44,45].

Associations of diet with prediabetes and diabetes

The results on the relationship between diet and diabetes in this work were generally consistent with findings reported in meta-analyses and review articles [149-159], despite some previous studies identifying no or only marginal relations between these food intake

variables and diabetes [88,160-168]. However, all other dietary factors, namely vegetables, potatoes, red meat, poultry, eggs, total dairy, milk, yogurt, cheese, fruit and vegetable juice, total fiber and insoluble fiber, analyzed here due to previously shown associations with diabetes in the literature [89,90,110,111], were not associated with diabetes or prediabetes in the fully adjusted models in the KORA FF4 study population.

In comparison, the German Diabetes Risk Score, which was conceived by the German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE) included the dietary intake variables red meat, whole grain and coffee in the risk prediction of developing diabetes within the next 5 years [35]. Despite the associations of total meat and processed meat with diabetes in the KORA FF4 study population, there was no association found between red meat and diabetes. This may be explained by the relatively small consumption amount in this study population [169,170]. Insoluble fiber was investigated in the fourth manuscript as a substitute for whole grain intake, which is not included in the EPIC-Soft classification system [109], but showed no significant association with diabetes. This is consistent with results of a meta-analysis on total, cereal, vegetable and fruit fiber after the adjustment for BMI [111], but stands in contrast to other meta-analyses showing a negative association of dietary fiber and/or whole grain intake with diabetes [171-173]. Concerning the intake of coffee, no significant association was found with diabetes in contrast to the literature [174-176]. However, a significant inverse association was identified with prediabetes, which is in line with a study investigating the combined risk for prediabetes and T2DM in individuals with known T2DM-related single nucleotide polymorphisms (SNPs) [86]. Heavy alcohol intake was also associated with prediabetes replicating previously findings in the literature [88]. Since little is known so far on associations between food items and prediabetes [84-88], the results on coffee and alcohol consumption, especially seen in men, need to be confirmed as potential new prevention targets for prediabetes. With prediabetes as a strong risk factor for the development of diabetes [77,78], these dietary factors may be also relevant in diabetes risk prediction.

It was the first time to our knowledge that metabolic characteristics by means of metabotypes were considered in associations of various food groups, food subgroups and nutrients with diabetes. This required the combination of the metabotype clusters 1 and 2 and their joint analysis in comparison to metabotype cluster 3 because of the low diabetes prevalence in both groups. Despite these interesting results, analyses by more detailed glucose tolerance status could not be performed stratified by metabotype due to insufficient sample sizes in the groups.

Strengths and limitations of the manuscripts

One strength of these manuscripts is the use of large population-based samples for the identification of comprehensive metabotypes and the analysis of diet-diabetes associations. This enabled the identification of meaningful metabolically homogeneous subgroups of the KORA F4 and FF4 study populations and allowed to consider these metabolic differences by stratifying diet-diabetes associations by metabotype subgroup. Nevertheless, the sample size was too small to additionally analyze associations of diet with prediabetes and UDM (as described in the fourth manuscript) for metabolically homogeneous subgroups. Further, the results may be affected by considerably different sample sizes between the metabotype subgroups as well as by selection bias, as not all participants in KORA S4, which is a representative sample of the population living in the study region, took part in the follow-up studies KORA F4 and FF4 [97]. Another limitation is that the causality of diet-diabetes associations could not be established due to the cross-sectional study design.

A further strength is the availability of extensive data in the KORA F4 and FF4 studies. Many biochemical and anthropometric parameters assessed by standardized methods once in the fasting state were available for the identification of comprehensive metabotypes. Repeated measurements over time were not assumed to change the results substantially, as intra-individual parameter variations [20,145,146] were demonstrated to be smaller than variations between individuals [147]. However, the inclusion of response data to interventions or '-omics' data may improve the identification of comprehensive metabotypes. Dietary intake was extensively assessed in KORA FF4 participants using an advanced method combining information of a FFQ and repeated 24HFLs so that a full range of food items could be investigated in their associations with the glucose tolerance status. Further, additional information was assessed in both studies to enable the adjustment of diet-diabetes associations by important confounders and to enable a detailed characterization of metabotypes by socio-demographic and lifestyle factors in KORA F4 as well as by cardiometabolic disease status in KORA F4 and FF4 with a long follow-up period of 7 years. However, dietary intake as well as information on socio-demographics, lifestyle and disease status were assessed mainly based on self-reports, thus misreporting cannot be excluded. Nevertheless, the glucose tolerance status was accurately determined for the analysis of diet-diabetes associations either by physician-based validated self-reported diabetes diagnosis or antidiabetic medication intake, or by an OGTT classifying individuals according to the ADA criteria [100].

Conclusions and future perspectives

The literature review on metabotyping in the first manuscript shows that the identification of metabotypes was performed heterogeneously. Thus, it is proposed to consider stricter metabotype definitions for specific metabolic pathways in future research to improve the comparability of studies. For the identification of comprehensive metabotypes, it is suggested to use an as detailed as possible characterization of individuals. This has been carried out in the second and third manuscripts by using extensive sets of biochemical and anthropometric parameters. These results need to be replicated and validated in other cohorts to prove generalizability. Future research should investigate on the one hand whether the metabotype definition could be further refined using an even more precise characterization of individuals by means of the additional consideration of metabolic response data to interventions or '-omics' data. On the other hand, a risk score may be developed based on a reduced set of the most important parameters, potentially routinely measured in clinical practice. This may facilitate the transfer of metabotypes to larger populations for targeted interventions at the metabotype subgroup level to improve disease prevention and management.

The fundament for the development of such targeted strategies is the identification of differences between metabotypes. Concerning diet as a modifiable risk factor for many cardiometabolic diseases, little is known so far on differential responsiveness of metabotypes to dietary factors [66-68]. Therefore, the different cross-sectional associations of various dietary factors with diabetes between metabotypes identified in the analyses of the third manuscript indicate a potential influence of metabolic characteristics on diet-diabetes associations. These may partially explain inconsistent results in the previous literature. Future studies may further investigate analyses stratified by metabotype for other diet-related cardiometabolic diseases. This would also be interesting for early disease stages, such as prediabetes, for which relatively new associations with diet were found in the fourth manuscript in the total KORA FF4 study population; due to sample size limitations, stratification by metabotype was not possible. It may be also interesting to examine dietary patterns in addition to single dietary intake variables [177] and other risk factors for cardiometabolic diseases such as physical inactivity [44,45]. These research activities may be highly relevant to strengthen and deepen the understanding of the role of inter-individual metabolic characteristics and for the development of comprehensive and targeted recommendations for better disease prevention and treatment at the metabotype subgroup level. Therefore, especially

interventional and large prospective cohort studies should be conducted to identify causal relationships.

Finally, the results of this dissertation in combination with findings from the previous literature and future studies on differences between metabotype subgroups should be used for the development of targeted and, thus, precise dietary strategies at the metabotype subgroup level. This was rarely performed previously [57,148] but may be highly relevant to improve success of disease prevention and management. Targeted strategies may facilitate the implementation of tailored dietary strategies in large populations, but its effectiveness compared to personalized and generalized recommendations regarding individual health benefits and health care costs needs to be investigated.

5. References

1. Gavaghan CL, Holmes E, Lenz E *et al.* (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett* 484(3):169-174.

2. Pontoizeau C, Fearnside JF, Navratil V *et al.* (2011) Broad-ranging natural metabotype variation drives physiological plasticity in healthy control inbred rat strains. *J Proteome Res* 10(4):1675-1689.

3. Clayton TA, Lindon JC, Cloarec O *et al.* (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature* 440(7087):1073-1077.

4. Holmes E, Nicholls AW, Lindon JC *et al.* (2000) Chemometric models for toxicity classification based on NMR spectra of biofluids. *Chem Res Toxicol* 13(6):471-478.

5. Gavaghan CL, Wilson ID, Nicholson JK (2002) Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. *FEBS Lett* 530(1-3):191-196.

6. Gavaghan McKee CL, Wilson ID, Nicholson JK (2006) Metabolic phenotyping of nude and normal (Alpk:ApfCD, C57BL10J) mice. *J Proteome Res* 5(2):378-384.

7. Fearnside JF, Dumas ME, Rothwell AR *et al.* (2008) Phylometabonomic patterns of adaptation to high fat diet feeding in inbred mice. *PLoS One* 3(2):e1668.

8. Bollard ME, Holmes E, Lindon JC *et al.* (2001) Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution (1)H NMR spectroscopy of urine and pattern recognition. *Anal Biochem* 295(2):194-202.

9. Calvani R, Brasili E, Praticò G *et al.* (2014) Fecal and urinary NMR-based metabolomics unveil an aging signature in mice. *Exp Gerontol* 49:5-11.

10. Garcia-Perez I, Villasenor A, Wijeyesekera A *et al.* (2012) Urinary metabolic phenotyping the slc26a6 (chloride-oxalate exchanger) null mouse model. *J Proteome Res* 11(9):4425-4435.

11. Qiao Q, Li T, Sun J *et al.* (2011) Metabolomic analysis of normal (C57BL/6J, 129S1/SvImJ) mice by gas chromatography-mass spectrometry: detection of strain and gender differences. *Talanta* 85(1):718-724.

12. Plumb R, Granger J, Stumpf C *et al.* (2003) Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst* 128(7):819-823.

13. Yang Y, Liu Y, Zheng L *et al.* (2014) Serum metabonomic analysis of apoE(-/-) mice reveals progression axes for atherosclerosis based on NMR spectroscopy. *Mol Biosyst* 10(12):3170-3178.

14. O'Donovan CB, Walsh MC, Gibney MJ *et al.* (2016) Can metabotyping help deliver the promise of personalised nutrition? *Proc Nutr Soc* 75(1):106-114.

15. Morris C, O'Grada C, Ryan M *et al.* (2013) Identification of differential responses to an oral glucose tolerance test in healthy adults. *PLoS One* 8(8):e72890.

16. Nicholson JK, Holmes E, Kinross JM *et al.* (2012) Metabolic phenotyping in clinical and surgical environments. *Nature* 491(7424):384-392.

17. Kaput J (2008) Nutrigenomics research for personalized nutrition and medicine. *Curr Opin Biotechnol* 19(2):110-120.

18. Brennan L (2008) Session 2: personalised nutrition. Metabolomic applications in nutritional research. *Proc Nutr Soc* 67(4):404-408.

19. Brennan L (2017) Use of metabotyping for optimal nutrition. *Curr Opin Biotechnol* 44:35-38.

20. Holmes E, Wilson ID, Nicholson JK (2008) Metabolic phenotyping in health and disease. *Cell* 134(5):714-717.

21. de Toro-Martín J, Arsenault BJ, Després JP *et al.* (2017) Precision nutrition: a review of personalized nutritional approaches for the prevention and management of metabolic syndrome. *Nutrients* 9(8).

22. Nizel AE (1972) Personalized nutrition counseling. *ASDC J Dent Child* 39(5):353-360.

23. Williams RJ, Beerstecher E, Jr., Berry LJ (1950) The concept of genetotrophic disease. *Lancet* 1(6599):287-289.

24. Rasinperä H, Savilahti E, Enattah NS *et al.* (2004) A genetic test which can be used to diagnose adult-type hypolactasia in children. *Gut* 53(11):1571-1576.

25. Ludvigsson JF, Bai JC, Biagi F *et al.* (2014) Diagnosis and management of adult coeliac disease: guidelines from the British Society of Gastroenterology. *Gut* 63(8):1210-1228.

26. DiLella AG, Huang WM, Woo SL (1988) Screening for phenylketonuria mutations by DNA amplification with the polymerase chain reaction. *Lancet* 1(8584):497-499.

27. Venter JC, Adams MD, Myers EW *et al.* (2001) The sequence of the human genome. *Science* 291(5507):1304-1351.

28. Lander ES, Linton LM, Birren B *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921.

29. Heni M, Herzberg-Schäfer S, Machicao F *et al.* (2012) Dietary fiber intake modulates the association between variants in TCF7L2 and weight loss during a lifestyle intervention. *Diabetes Care* 35(3):e24.

30. Rauhio A, Uusi-Rasi K, Nikkari ST *et al.* (2013) Association of the FTO and ADRB2 genes with body composition and fat distribution in obese women. *Maturitas* 76(2):165-171.

31. Fallaize R, Celis-Morales C, Macready AL *et al.* (2016) The effect of the apolipoprotein E genotype on response to personalized dietary advice intervention: findings from the Food4Me randomized controlled trial. *Am J Clin Nutr* 104(3):827-836.

32. Rappaport SM (2016) Genetic factors are not the major causes of chronic diseases. *PLoS One* 11(4):e0154387.

33. Carlsten C, Brauer M, Brinkman F *et al.* (2014) Genes, the environment and personalized medicine: We need to harness both environmental and genetic data to maximize personal and population health. *EMBO Rep* 15(7):736-739.

34. Deutsche Gesellschaft für Ernährung (DGE), Österreichische Gesellschaft für Ernährung (ÖGE), Schweizerische Gesellschaft für Ernährungsforschung (SGE) *et al.* (2008) *Referenzwerte für die Nährstoffzufuhr 1. Auflage*. Neuer Umschau Buchverlag, Neustadt an der Weinstraße.

35. Mühlenbruch K, Ludwig T, Jeppesen C *et al.* (2014) Update of the German Diabetes Risk Score and external validation in the German MONICA/KORA study. *Diabetes Res Clin Pract* 104(3):459-466.

36. Ronteltap A, van Trijp H, Berezowska A *et al.* (2013) Nutrigenomics-based personalised nutritional advice: in search of a business model? *Genes Nutr* 8(2):153-163.

37. Gibney MJ, Walsh MC (2013) The future direction of personalised nutrition: my diet, my phenotype, my genes. *Proc Nutr Soc* 72(2):219-225.

38. de Roos B (2013) Personalised nutrition: ready for practice? *Proc Nutr Soc* 72(1):48-52.

39. Nicholson JK, Holmes E (2006) Global systems biology and personalized healthcare solutions. *Discov Med* 6(32):63-70.

40. Celis-Morales C, Livingstone KM, Marsaux CF *et al.* (2017) Effect of personalized nutrition on health-related behaviour change: evidence from the Food4me European randomized controlled trial. *Int J Epidemiol* 46(2):578-588.

41. Livingstone KM, Celis-Morales C, Navas-Carretero S *et al.* (2016) Effect of an Internet-based, personalized nutrition randomized trial on dietary changes associated with the Mediterranean diet: the Food4Me Study. *Am J Clin Nutr* 104(2):288-297.

42. Nicholson JK (2006) Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* 2:52.

43. Beger RD, Dunn W, Schmidt MA *et al.* (2016) Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics* 12(10):149.

44. Ezzati M, Riboli E (2013) Behavioral and dietary risk factors for noncommunicable diseases. *N Engl J Med* 369(10):954-964.

45. World Health Organization (2009) Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization, Geneva.

46. Ordovas JM, Ferguson LR, Tai ES *et al.* (2018) Personalised nutrition and health. *BMJ* 361:bmj k2173.

47. Zák A, Burda M, Vecka M *et al.* (2014) Fatty acid composition indicates two types of metabolic syndrome independent of clinical and laboratory parameters. *Physiol Res* 63(Suppl 3):S375-385.

48. Schader LM (2015) Comparison of genetic risk factors between two type II diabetes subtypes. Bachelor Thesis, University of Arizona.

49. Li L, Cheng WY, Glicksberg BS *et al.* (2015) Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med* 7(311):311ra174.

50. Amato MC, Pizzolanti G, Torregrossa V *et al.* (2016) Phenotyping of type 2 diabetes mellitus at onset on the basis of fasting incretin tone: Results of a two-step cluster analysis. *J Diabetes Investig* 7(2):219-225.

51. Frei F, Lessa Bde S, Nogueira JC *et al.* (2013) Cluster analysis for classification of patients submitted to Fobi-Capella bariatric surgery. *Arq Bras Cir Dig* 26(Suppl 1):33-38.

52. Arguelles W, Llabre MM, Sacco RL *et al.* (2015) Characterization of metabolic syndrome among diverse Hispanics/Latinos living in the United States: Latent class analysis from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Int J Cardiol* 184:373-379.

53. Kim E, Oh W, Pieczkiewicz DS *et al.* (2014) Divisive hierarchical clustering towards identifying clinically significant pre-diabetes subpopulations. *AMIA Annu Symp Proc* 2014:1815-1824.

54. Mäkinen VP, Soininen P, Forsblom C *et al.* (2008) 1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. *Mol Syst Biol* 4:167.

55. Botelho PB, Fioratti CO, Abdalla DS *et al.* (2010) Classification of individuals with dyslipidaemia controlled by statins according to plasma biomarkers of oxidative stress using cluster analysis. *Br J Nutr* 103(2):256-265.

56. van Bochove K, van Schalkwijk DB, Parnell LD *et al.* (2012) Clustering by plasma lipoprotein profile reveals two distinct subgroups with positive lipid response to fenofibrate therapy. *PLoS One* 7(6):e38072.

57. O'Donovan CB, Walsh MC, Nugent AP *et al.* (2015) Use of metabotyping for the delivery of personalised nutrition. *Mol Nutr Food Res* 59(3):377-385.

58. Frazier-Wood AC, Glasser S, Garvey WT *et al.* (2011) A clustering analysis of lipoprotein diameters in the metabolic syndrome. *Lipids Health Dis* 10:237.

59. Zubair N, Kuzawa CW, McDade TW *et al.* (2012) Cluster analysis reveals important determinants of cardiometabolic risk patterns in Filipino women. *Asia Pac J Clin Nutr* 21(2):271-281.

60. Zubair N, Kuzawa CW, Lee NR *et al.* (2014) Clustering and determinants of cardiometabolic risk factors among Filipino young adults. *Asia Pac J Clin Nutr* 23(1):148-158.

61. Wilcox MA, Wyszynski DF, Panhuysen CI *et al.* (2003) Empirically derived phenotypic subgroups - qualitative and quantitative trait analyses. *BMC Genet* 4(Suppl 1):S15.

62. Wilcox M, Li Q, Sun Y *et al.* (2009) Genome-wide association study for empirically derived metabolic phenotypes in the Framingham Heart Study offspring cohort. *BMC Proc* 3(Suppl 7):S53.

63. Tzeng CR, Chang YC, Chang YC *et al.* (2014) Cluster analysis of cardiovascular and metabolic risk factors in women of reproductive age. *Fertil Steril* 101(5):1404-1410.

64. Andreeva-Gateva PA, Simeonov VD, Georrgieva-Nikolova RT *et al.* (2014) Cluster analysis of the components of the metabolic syndrome in subjects with an increased risk for diabetes mellitus type 2. *Endocrinologya* 19(3):168-185.

65. Qureshi W, Wagenknecht L, Watkins S *et al.* (2014) Abstract 23: metabolomic features associated with incident hypertension - results from Insulin Resistance Atherosclerosis Study. *Circulation* 129(Suppl 1):A23.

66. O'Sullivan A, Gibney MJ, Connor AO *et al.* (2011) Biochemical and metabolomic phenotyping in the identification of a vitamin D responsive metabotype for markers of the metabolic syndrome. *Mol Nutr Food Res* 55(5):679-690.

67. Vázquez-Fresno R, Llorach R, Perera A *et al.* (2016) Clinical phenotype clustering in cardiovascular risk patients for the identification of responsive metabotypes after red wine polyphenol intake. *J Nutr Biochem* 28:114-120.

68. Moazzami AA, Shrestha A, Morrison DA *et al.* (2014) Metabolomics reveals differences in postprandial responses to breads and fasting metabolic characteristics associated with postprandial insulin demand in postmenopausal women. *J Nutr* 144(6):807-814.

69. Ogurtsova K, da Rocha Fernandes JD, Huang Y *et al.* (2017) IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract* 128:40-50.

70. World Health Organization (2018) The top 10 causes of death. World Health Organization, Geneva. http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 28 September 2018.

71. World Health Organization (2016) Global report on diabetes. World Health Organization, Geneva.

72. Wu Y, Ding Y, Tanaka Y *et al.* (2014) Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. *Int J Med Sci* 11(11):1185-1200.

73. Fowler MJ (2008) Microvascular and macrovascular complications of diabetes. *Clinical Diabetes* 26(2):77-82.

74. Goffrier B, Schulz M, Bätzing-Feigenbaum J (2017) Administrative Prävalenzen und Inzidenzen des Diabetes mellitus von 2009 bis 2015. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi). Versorgungsatlas-Bericht Nr. 17/03. Berlin. doi: 10.20364/VA-17.03. http://www.versorgungsatlas.de/themen/alle-analysen-nach-datum-sortiert/?tab=6&uid=79. Accessed 28 September 2018.

75. Meisinger C, Strassburger K, Heier M *et al.* (2010) Prevalence of undiagnosed diabetes and impaired glucose regulation in 35-59-year-old individuals in Southern Germany: the KORA F4 Study. *Diabet Med* 27(3):360-362.

76. Rathmann W, Haastert B, Icks A *et al.* (2003) High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. *Diabetologia* 46(2):182-189.

77. Buysschaert M, Bergman M (2011) Definition of prediabetes. *Med Clin North Am* 95(2):289-297, vii.

78. Tabák AG, Herder C, Rathmann W *et al.* (2012) Prediabetes: a high-risk state for diabetes development. *Lancet* 379(9833):2279-2290.

79. Heidemann C, Du Y, Paprott R *et al.* (2016) Temporal changes in the prevalence of diagnosed diabetes, undiagnosed diabetes and prediabetes: findings from the German Health Interview and Examination Surveys in 1997-1999 and 2008-2011. *Diabet Med* 33(10):1406-1414.

80. Brannick B, Wynn A, Dagogo-Jack S (2016) Prediabetes as a toxic environment for the initiation of microvascular and macrovascular complications. *Exp Biol Med (Maywood)* 241(12):1323–1331.

81. Walker KZ, O'Dea K, Gomez M *et al.* (2010) Diet and exercise in the prevention of diabetes. *J Hum Nutr Diet* 23(4):344-352.

82. Bazzano LA, Serdula M, Liu S (2005) Prevention of type 2 diabetes by diet and lifestyle modification. *J Am Coll Nutr* 24(5):310-319.

83. Thomas GN, Jiang CQ, Taheri S *et al.* (2010) A systematic review of lifestyle modification and glucose intolerance in the prevention of type 2 diabetes. *Curr Diabetes Rev* 6(6):378-387.

84. Hruby A, Ma J, Rogers G *et al.* (2017) Associations of dairy intake with incident prediabetes or diabetes in middle-aged adults vary by both dairy type and glycemic status. *J Nutr* 147(9):1764-1775.

85. Wirström T, Hilding A, Gu HF *et al.* (2013) Consumption of whole grain reduces risk of deteriorating glucose tolerance, including progression to prediabetes. *Am J Clin Nutr* 97(1):179-187.

86. Lee JK, Kim K, Ahn Y *et al.* (2015) Habitual coffee intake, genetic polymorphisms, and type 2 diabetes. *Eur J Endocrinol* 172(5):595-601.

87. Ma J, Jacques PF, Meigs JB *et al.* (2016) Sugar-sweetened beverage but not diet soda consumption is positively associated with progression of insulin resistance and prediabetes. *J Nutr* 146(12):2544-2550.

88. Cullmann M, Hilding A, Östenson CG (2012) Alcohol consumption and risk of pre-diabetes and type 2 diabetes development in a Swedish population. *Diabet Med* 29(4):441-452.

89. Schwingshackl L, Hoffmann G, Lampousi AM *et al.* (2017) Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. *Eur J Epidemiol* 32(5):363-375.

90. Xi P, Liu RH (2016) Whole food approach for type 2 diabetes prevention. *Mol Nutr Food Res* 60(8):1819-1836.

91. Riedl A, Gieger C, Hauner H *et al.* (2017) Metabotyping and its application in targeted nutrition: an overview. *Br J Nutr* 117(12):1631-1644.

92. Riedl A, Wawro N, Gieger C *et al.* (2018) Identification of comprehensive metabotypes associated with cardiometabolic diseases in the population-based KORA study. *Mol Nutr Food Res* 62(16):e1800117.

93. Higgins JPT, Green S (editors) (2011) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration. http://handbook.cochrane.org/. Accessed 25 August 2018.

94. World Health Organization (2003) Diet, nutrition and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation. *WHO Technical Report Series* no. 916. World Health Organization, Geneva.

95. Holle R, Happich M, Löwel H *et al.* (2005) KORA--a research platform for population based health research. *Gesundheitswesen* 67(Suppl 1):S19-S25.

96. Chambless L, Cairns V, Herbold M *et al.* (1986) MONICA Project Augsburg, Survey Sampling. GSF-Bericht 31/86. Gesellschaft für Strahlen- und Umweltforschung, Munich.

97. Holle R, Hochadel M, Reitmeir P *et al.* (2006) Prolonged recruitment efforts in health surveys: effects on response, costs, and potential bias. *Epidemiology* 17(6):639-643.

98. Rospleszcz S, Schafnitzel A, Koenig W *et al.* (2018) Association of glycemic status and segmental left ventricular wall thickness in subjects without prior cardiovascular disease: a cross-sectional study. *BMC Cardiovasc Disord* 18(1):162.

99. Mühlberger N, Behrend C, Stark RG *et al.* (2003) Datenbankgestützte Online-Erfassung von Arzneimitteln im Rahmen gesundheitswissenschaftlicher Studien - Erfahrungen mit der IDOM-Software. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 34:601-611. Urban & Fischer, Munich.

100. American Diabetes Association (2011) Diagnosis and classification of diabetes mellitus. *Diabetes Care* 34(Suppl 1):S62–S69.

101. Illner AK, Harttig U, Tognon G *et al.* (2011) Feasibility of innovative dietary assessment in epidemiological studies using the approach of combining different assessment instruments. *Public Health Nutr* 14(6):1055-1063.

102. Freese J, Feller S, Harttig U *et al.* (2014) Development and evaluation of a short 24-h food list as part of a blended dietary assessment strategy in large-scale cohort studies. *Eur J Clin Nutr* 68(3):324-329.

103. Tooze JA, Midthune D, Dodd KW *et al.* (2006) A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc* 106(10):1575-1587.

104. Tooze JA, Kipnis V, Buckman DW *et al.* (2010) A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. *Stat Med* 29(27):2857-2868.

105. Haubrock J, Nöthlings U, Volatier JL *et al.* (2011) Estimating usual food intake distributions by using the multiple source method in the EPIC-Potsdam Calibration Study. *J Nutr* 141(5):914-920.

106. Michels KB (2001) A renaissance for measurement error. *Int J Epidemiol* 30(3):421-422.

107. Rosner B, Willett WC, Spiegelman D (1989) Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med* 8(9):1051-1069; discussion 1071-1073.

108. Himmerich S, Gedrich K, Karg G (2004) Bayerische Verzehrsstudie (BVS) II: Abschlussbericht. Forschungsbericht im Auftrag des Bayerischen Staatsministeriums für Gesundheit, Ernährung und Verbraucherschutz, Freising.

109. Slimani N, Deharveng G, Charrondière RU *et al.* (1999) Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. European Prospective Investigation into Cancer and Nutrition. *Comput Methods Programs Biomed* 58(3):251-266.

110. Mozaffarian D (2016) Dietary and policy priorities for cardiovascular disease, diabetes, and obesity: a comprehensive review. *Circulation* 133(2):187-225.

111. InterAct Consortium (2015) Dietary fibre and incidence of type 2 diabetes in eight European countries: the EPIC-InterAct Study and a meta-analysis of prospective studies. *Diabetologia* 58(7):1394-1408.

112. Keil U, Chambless LE, Döring A *et al.* (1997) The relation of alcohol intake to coronary heart disease and all-cause mortality in a beer-drinking population. *Epidemiology* 8(2):150-156.

113. Bartel J, Krumsiek J, Theis FJ (2013) Statistical methods for the analysis of high-throughput metabolomics data. *Comput Struct Biotechnol J* 4:e201301009.

114. Andreopoulos B, An A, Wang X *et al.* (2009) A roadmap of clustering algorithms: finding a match for a biomedical application. *Brief Bioinform* 10(3):297-314.

115. Everitt BS, Landau S, Leese M *et al.* (2011) *Cluster analysis*. 5th edition [Balding DJ, Cressie NAC, Fitzmaurice GM et al., editors]. John Wiley & Sons Ltd, London.

116. Kaufman L, Rousseeuw PJ (2005) *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons Inc, Hoboken.

117. Jain AK, Topchy A, Law MHC *et al.* (2004) Landscape of clustering algorithms. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)* 1:260-263.

118. Bacher J, Pöge A, Wenzig K (2010) *Clusteranalyse – Anwendungsorientierte Einführung in Klassifikationsverfahren (Cluster analysis – application-oriented introduction to classification methods).* 3rd edition. Oldenbourg Wissenschaftsverlag GmbH, Munich.

119. Dua S, Chowriappa P (2013) *Data Mining for Bioinformatics*. CRC Press Taylor & Francis Group, Boca Raton.

120. Steinley D (2006) K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* 59(Pt 1):1-34.

121. Charrad M, Ghazzali N, Boiteau V *et al.* (2014) NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw* 61(6):1-36.

122. Bernini P, Bertini I, Luchinat C *et al.* (2009) Individual human phenotypes in metabolic space and time. *J Proteome Res* 8(9):4264-4271.

123. Yousri NA, Kastenmüller G, Gieger C *et al.* (2014) Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics* 10(5):1005-1017.

124. Chaleckis R, Murakami I, Takada J *et al.* (2016) Individual variability in human blood metabolites identifies age-related differences. *Proc Natl Acad Sci U S A* 113(16):4252-4259.

125. van Buuren S, Groothuis-Oudshoorn K (2011) mice: multivariate imputation by chained equations in R. *J Stat Softw* 45(3):1-67.

126. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 22(1):4-37.

127. Genuer R, Poggi JM, Tuleau-Malot C (2010) Variable selection using random forests. *Pattern Recognition Letters* 31(14):2225-2236.

128. Basagana X, Barrera-Gómez J, Benet M *et al.* (2013) A framework for multiple imputation in cluster analysis. *Am J Epidemiol* 177(7):718-725.

129. Stoltzfus JC (2011) Logistic regression: a brief primer. *Acad Emerg Med* 18(10):1099-1104.

130. Holmes E, Loo RL, Stamler J *et al.* (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453(7193):396-400.

131. Lenz EM, Bright J, Wilson ID *et al.* (2004) Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J Pharm Biomed Anal* 36(4):841-849.

132. Dumas ME, Maibaum EC, Teague C *et al.* (2006) Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem* 78(7):2199-2208.

133. Krumsiek J, Mittelstrass K, Do KT *et al.* (2015) Gender-specific pathway differences in the human serum metabolome. *Metabolomics* 11(6):1815-1833.

134. Kochhar S, Jacobs DM, Ramadan Z *et al.* (2006) Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. *Anal Biochem* 352(2):274-281.

135. Li K, Brennan L, McNulty BA *et al.* (2016) Plasma fatty acid patterns reflect dietary habits and metabolic health: a cross-sectional study. *Mol Nutr Food Res* 60(9):2043-2052.

136. Wang TT, Edwards AJ, Clevidence BA (2013) Strong and weak plasma response to dietary carotenoids identified by cluster analysis and linked to beta-carotene 15,15'-monooxygenase 1 single nucleotide polymorphisms. *J Nutr Biochem* 24(8):1538-1546.

137. Krishnan S, Newman JW, Hembrooke TA *et al.* (2012) Variation in metabolic responses to meal challenges differing in glycemic index in healthy women: Is it meaningful? *Nutr Metab (Lond)* 9:26.

138. Valdes AM, Glass D, Spector TD (2013) Omics technologies and the study of human ageing. *Nat Rev Genet* 14(9):601-607.

139. Kurland IJ, Accili D, Burant C *et al.* (2013) Application of combined omics platforms to accelerate biomedical discovery in diabesity. *Ann N Y Acad Sci* 1287:1-16.

140. Corella D, Ordovás JM (2018) [The role of omics in precision nutrition: strengths and weaknesses]. *Nutr Hosp* 35(Spec No4):10-18.

141. Bouwman J, Vogels JT, Wopereis S *et al.* (2012) Visualization and identification of health space, based on personalized molecular phenotype and treatment response to relevant underlying biological processes. *BMC Med Genomics* 5:1.

142. Chua EC, Shui G, Lee IT *et al.* (2013) Extensive diversity in circadian regulation of plasma lipids and evidence for different circadian metabolic phenotypes in humans. *Proc Natl Acad Sci U S A* 110(35):14468-14473.

143. Krug S, Kastenmüller G, Stückler F *et al.* (2012) The dynamic range of the human metabolome revealed by challenges. *FASEB J* 26(6):2607-2619.

144. Fiamoncini J, Rundle M, Gibbons H *et al.* (2018) Plasma metabolome analysis identifies distinct human metabotypes in the postprandial state with different susceptibility to weight loss-mediated metabolic improvements. *FASEB J* 32(10):5447-5458.

145. Rezzi S, Martin FP, Alonso C *et al.* (2009) Metabotyping of biofluids reveals stress-based differences in gut permeability in healthy individuals. *J Proteome Res* 8(10):4799-4809.

146. Ghini V, Saccenti E, Tenori L *et al.* (2015) Allostasis and resilience of the human individual metabolic phenotype. *J Proteome Res* 14(7):2951-2962.

147. Assfalg M, Bertini I, Colangiuli D *et al.* (2008) Evidence of different metabolic phenotypes in humans. *Proc Natl Acad Sci U S A* 105(5):1420-1424.

148. O'Donovan CB, Walsh MC, Woolhead C *et al.* (2017) Metabotyping for the development of tailored dietary advice solutions in a European population: the Food4Me study. *Br J Nutr* 118(8):561-569.

149. Li M, Fan Y, Zhang X *et al.* (2014) Fruit and vegetable intake and risk of type 2 diabetes mellitus: meta-analysis of prospective cohort studies. *BMJ Open* 4(11):e005497.

150. Wang PY, Fang JC, Gao ZH *et al.* (2016) Higher intake of fruits, vegetables or their fiber reduces the risk of type 2 diabetes: A meta-analysis. *J Diabetes Investig* 7(1):56-69.

151. Li S, Miao S, Huang Y *et al.* (2015) Fruit intake decreases risk of incident type 2 diabetes: an updated meta-analysis. *Endocrine* 48(2):454-460.

152. Aune D, Ursin G, Veierod MB (2009) Meat consumption and the risk of type 2 diabetes: a systematic review and meta-analysis of cohort studies. *Diabetologia* 52(11):2277-2287.

153. Micha R, Wallace SK, Mozaffarian D (2010) Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: a systematic review and meta-analysis. *Circulation* 121(21):2271-2283.

154. Feskens EJ, Sluik D, van Woudenbergh GJ (2013) Meat consumption, diabetes, and its complications. *Curr Diab Rep* 13(2):298-306.

155. Ley SH, Hamdy O, Mohan V *et al.* (2014) Prevention and management of type 2 diabetes: dietary components and nutritional strategies. *Lancet* 383(9933):1999-2007.

156. Imamura F, O'Connor L, Ye Z *et al.* (2015) Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2

diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. *BMJ* 351:h3576.

157. InterAct Consortium, Romaguera D, Norat T *et al.* (2013) Consumption of sweet beverages and type 2 diabetes incidence in European adults: results from EPIC-InterAct. *Diabetologia* 56(7):1520-1530.

158. Malik VS, Popkin BM, Bray GA *et al.* (2010) Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. *Diabetes Care* 33(11):2477-2483.

159. Knott C, Bell S, Britton A (2015) Alcohol consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of more than 1.9 million individuals from 38 observational studies. *Diabetes Care* 38(9):1804-1812.

160. Cooper AJ, Forouhi NG, Ye Z *et al.* (2012) Fruit and vegetable intake and type 2 diabetes: EPIC-InterAct prospective study and meta-analysis. *Eur J Clin Nutr* 66(10):1082-1092.

161. Hamer M, Chida Y (2007) Intake of fruit, vegetables, and antioxidants and risk of type 2 diabetes: systematic review and meta-analysis. *J Hypertens* 25(12):2361-2369.

162. Carter P, Gray LJ, Troughton J *et al.* (2010) Fruit and vegetable intake and incidence of type 2 diabetes mellitus: systematic review and meta-analysis. *BMJ* 341:c4229.

163. Villegas R, Shu XO, Gao YT *et al.* (2006) The association of meat intake and the risk of type 2 diabetes may be modified by body weight. *Int J Med Sci* 3(4):152-159.

164. Montonen J, Järvinen R, Heliövaara M *et al.* (2005) Food consumption and the incidence of type II diabetes mellitus. *Eur J Clin Nutr* 59(3):441-448.

165. Hodge AM, English DR, O'Dea K *et al.* (2007) Dietary patterns and diabetes incidence in the Melbourne Collaborative Cohort Study. *Am J Epidemiol* 165(6):603-610.

166. Montonen J, Järvinen R, Knekt P *et al.* (2007) Consumption of sweetened beverages and intakes of fructose and glucose predict type 2 diabetes occurrence. *J Nutr* 137(6):1447-1454.

167. Paynter NP, Yeh HC, Voutilainen S *et al.* (2006) Coffee and sweetened beverage consumption and the risk of type 2 diabetes mellitus: the atherosclerosis risk in communities study. *Am J Epidemiol* 164(11):1075-1084.

168. Nettleton JA, Lutsey PL, Wang Y *et al.* (2009) Diet soda intake and risk of incident metabolic syndrome and type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care* 32(4):688-694.

169. Micha R, Michas G, Mozaffarian D (2012) Unprocessed red and processed meats and risk of coronary artery disease and type 2 diabetes--an updated review of the evidence. *Curr Atheroscler Rep* 14(6):515-524.

170. Pan A, Sun Q, Bernstein AM *et al.* (2013) Changes in red meat consumption and subsequent risk of type 2 diabetes mellitus: three cohorts of US men and women. *JAMA Intern Med* 173(14):1328-1335.

171. Yao B, Fang H, Xu W *et al.* (2014) Dietary fiber intake and risk of type 2 diabetes : a dose-response analysis of prospective studies. *Eur J Epidemiol* 29(2):79-88.

172. Aune D, Norat T, Romundstad P *et al.* (2013) Whole grain and refined grain consumption and the risk of type 2 diabetes : a systematic review and dose-response meta-analysis of cohort studies. *Eur J Epidemiol* 28(11):845-858.

173. Ye E, Chacko S, Chou E *et al.* (2012) Greater whole-grain intake is associated with lower risk of type 2 diabetes , cardiovascular disease, and weight gain. *J Nutr* 142(7):1304-1313.

174. Lee JH, Oh MK, Lim JT *et al.* (2016) Effect of coffee consumption on the progression of type 2 diabetes mellitus among prediabetic individuals. *Korean J Fam Med* 37(1):7-13.

175. Yarmolinsky J, Mueller NT, Duncan BB *et al.* (2015) Coffee consumption, newly diagnosed diabetes, and other alterations in glucose homeostasis: a cross-sectional analysis of the Longitudinal Study of Adult Health (ELSA-Brasil). *PLoS One* 10(5):e0126469.

176. Bhupathiraju SN, Pan A, Manson JE *et al.* (2014) Changes in coffee intake and subsequent risk of type 2 diabetes: three large cohorts of US men and women. *Diabetologia* 57(7):1346-1354.

177. Schulze MB, Martínez-González MA, Fung TT *et al.* (2018) Food based dietary patterns and chronic disease prevention. *BMJ* 361:k2396.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen bedanken, die mich während der Promotion unterstützt und an mich geglaubt haben.

An erster Stelle danke ich Herrn Prof. Dr. Jakob Linseisen sowohl für die Möglichkeit dieses interessante Thema im Rahmen des enable-Kompetenzclusters der Ernährungsforschung zu bearbeiten als auch für die hervorragende Betreuung am Helmholtz Zentrum München. Ich danke ihm auch für die Möglichkeit an vielen Veranstaltungen und Konferenzen teilzunehmen, um mit anderen Wissenschaftlern und Doktoranden in Kontakt zu treten. Ganz besonders danke ich ihm für die Ermöglichung eines Forschungsaufenthalts am University College Dublin bei Prof. Lorraine Brennan zum Aufbau einer internationalen Kooperation in diesem Forschungsbereich.

Herrn Prof. Dr. med. Hans Hauner danke ich für die Übernahme der Betreuung am Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der TU München, sowie als Mitglied meines HELENA Thesis Committees und als Kooperationspartner im enable-Teilprojekt für die Unterstützung und für die gute Zusammenarbeit.

Herrn Prof. Dr. Martin Klingenspor danke ich, dass er sich bereit erklärt hat den Prüfungsvorsitz zu übernehmen.

Weiterer Dank gilt dem enable-Kompetenzcluster der Ernährungsforschung mit allen enable-Partnern und enable-Doktoranden für die erfolgreiche Zusammenarbeit. Auch hier möchte ich mich für die Möglichkeit der Teilnahme an zahlreichen Veranstaltungen bedanken, besonders für die Reise zur Konferenz Nutrition 2018 in Boston.

Ich möchte mich auch bei allen Koautoren für die hilfreichen Kommentare und die gute Zusammenarbeit bei den Publikationen bedanken.

Außerdem möchte ich mich bei meinen Kolleginnen und Kollegen am Helmholtz Zentrum München und am Lehrstuhl für Epidemiologie der LMU München am UNIKA-T Augsburg bedanken für die Hilfsbereitschaft bei fachlichen und administrativen Angelegenheiten, die Motivation und die angenehme Arbeitsatmosphäre im ‚Container'.

Ganz herzlich danke ich auch meiner Familie und meinen Freunden für die Unterstützung bei allen Hochs und Tiefs der Doktorandenzeit.